Lectures on Stochastic Programming: Modeling and Theory

Alexander Shapiro Darinka Dentcheva Andrzej Ruszczyński

Second edition SIAM, Philadelphia

Darinka Dentcheva Department of Mathematical Sciences Stevens Institute of Technology Hoboken, NJ 07030, USA

Andrzej Ruszczyński Department of Management Science and Information Systems Rutgers University Piscataway, NJ 08854, USA

Alexander Shapiro School of Industrial and Systems Engineering Georgia Institute of Technology Atlanta, GA 30332, USA

The authors dedicate this book:

to Julia, Benjamin, Daniel, Natan and Yael; to Tsonka, Konstantin and Marek; and to the Memory of Feliks, Maria, and Dentcho.

Preface

Preface to the second edition

In the second edition, we introduced new material to reflect recent developments in the area of stochastic programming. Chapter 6 underwent substantial revision. In sections 6.3.4–6.3.6, we extended the discussion of law invariant coherent risk measures and their Kusuoka representations. In sections 6.8.2–6.8.6, we provided in-depth analysis of dynamic risk measures and concepts of time consistency, including several new results.

In Chapter 4, we provided analytical description of the tangent and normal cones of chance constrained sets in section 4.3.3. We extended the analysis of optimality conditions in section 4.3.4 to non-convex problems.

In Chapter 5, we added section 5.10 with a discussion of the Stochastic Dual Dynamic Programming method, which became popular in power generation planning. We also made corrections and small additions in Chapters 3 and 7, and we updated the bibliography.

Preface to the first edition

The main topic of this book are optimization problems involving uncertain parameters, for which stochastic models are available. Although many ways have been proposed to model uncertain quantities, stochastic models have proved their flexibility and usefulness in diverse areas of science. This is mainly due to solid mathematical foundations and theoretical richness of the theory of probability and stochastic processes, and to sound statistical techniques of using real data.

Optimization problems involving stochastic models occur in almost all areas of science and engineering, so diverse as telecommunication, medicine, or finance, to name just a few. This stimulates interest in rigorous ways of formulating, analyzing, and solving such problems. Due to the presence of random parameters in the model, the theory combines concepts of the optimization theory, the theory of probability and statistics, and functional analysis. Moreover, in recent years the theory and methods of stochastic programming have undergone major advances. All these factors motivated us to present in an accessible and rigorous form contemporary models and ideas of stochastic programming. We hope that the book will encourage other researchers to apply stochastic programming models and to undertake further studies of this fascinating and rapidly developing area.

We do not try to provide a comprehensive presentation of all aspects of stochastic

programming, but we rather concentrate on theoretical foundations and recent advances in selected areas. The book is organized in 7 chapters. The first chapter addresses modeling issues. The basic concepts, such as recourse actions, chance (probabilistic) constraints and the nonanticipativity principle, are introduced in the context of specific models. The discussion is aimed at providing motivation for the theoretical developments in the book, rather than practical recommendations.

Chapters 2 and 3 present detailed development of the theory of two- and multistage stochastic programming problems. We analyze properties of the models, develop optimality conditions and duality theory in a rather general setting. Our analysis covers general distributions of uncertain parameters, and also provides special results for discrete distributions, which are relevant for numerical methods. Due to specific properties of two- and multi-stage stochastic programming problems, we were able to derive many of these results without resorting to methods of functional analysis.

The basic assumption in the modeling and technical developments is that the probability distribution of the random data is not influenced by our actions (decisions). In some applications this assumption could be unjustified. However, dependence of probability distribution on decisions typically destroys the convex structure of the optimization problems considered, and our analysis exploits convexity in a significant way.

Chapter 4 deals with chance (probabilistic) constraints, which appear naturally in many applications. The chapter presents the current state of the theory, focusing on the structure of the problems, optimality theory, and duality. We present generalized convexity of functions and measures, differentiability, and approximations of probability functions. Much attention is devoted to problems with separable chance constraints and problems with discrete distributions. We also analyze problems with first order stochastic dominance constraints, which can be viewed as problems with continuum of probabilistic constraints. Many of the presented results are relatively new and were not previously available in standard textbooks.

Chapter 5 is devoted to statistical inference in stochastic programming. The starting point of the analysis is that the probability distribution of the random data vector is approximated by an empirical probability measure. Consequently the "true" (expected value) optimization problem is replaced by its sample average approximation (SAA). Origins of this statistical inference are going back to the classical theory of the maximum likelihood method routinely used in statistics. Our motivation and applications are somewhat different, because we aim at solving stochastic programming problems by Monte Carlo sampling techniques. That is, the sample is generated in the computer and its size is only constrained by the computational resources needed to solve the constructed SAA problem. One of the byproducts of this theory is the complexity analysis of two and multistage stochastic programming. Already in the case of two stage stochastic programming the number of scenarios (discretization points) grows exponentially with the increase of the number of stages.

In Chapter 6 we outline the modern theory of risk averse approaches to stochastic programming. We focus on the analysis of the models, optimality theory, and duality. Static and two-stage risk-averse models are analyzed in much detail. We also outline a risk-averse approach to multistage problems, using conditional risk mappings and the principle of "time consistency."

Chapter 7 contains formulations of technical results used in the other parts of the book. For some of these, less known, results we give proofs, while others are referred to the literature. The subject index can help the reader to find quickly a required definition or formulation of a needed technical result.

Several important aspects of stochastic programming have been left out. We do not discuss numerical methods for solving stochastic programming problems, with exception of section 5.9 where the Stochastic Approximation method, and its relation to complexity estimates, is considered. Of course, numerical methods is an important topic which deserves careful analysis. This, however, is a vast and separate area which should be considered in a more general framework of modern optimization methods and to large extent would lead outside the scope of this book.

We also decided not to include a thorough discussion of stochastic integer programming. The theory and methods of solving stochastic integer programming problems draw heavily from the theory of general integer programming. Their comprehensive presentation would entail discussion of many concepts and methods of this vast field, which would have little connection with the rest of the book.

At the beginning of each chapter we indicate the authors who were primarily responsible for writing the material, but the book is the creation of all three of us, and we share equal responsibility for errors and inaccuracies that escaped our attention.

We thank Stevens Institute of Technology and Rutgers University for granting sabbatical leaves to Darinka Dentcheva and Andrzej Ruszczyński, during which a large portion of this work was written. Andrzej Ruszczyński is also thankful to the Department of Operations Research and Financial Engineering of Princeton University for providing him with excellent conditions for his stay during the sabbatical leave.

Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński

Contents

iii

Preface

1	Stochastic Programming Models			1
	1.1	Introducti	on	. 1
	1.2	Inventory		. 1
		1.2.1	The Newsvendor Problem	. 1
		1.2.2	Chance Constraints	. 5
		1.2.3	Multistage Models	. 6
	1.3	Multi-Pro	duct Assembly	. 9
		1.3.1	Two Stage Model	. 9
		1.3.2	Chance Constrained Model	. 11
		1.3.3	Multistage Model	. 12
	1.4	Portfolio S	Selection	. 13
		1.4.1	Static Model	. 13
		1.4.2	Multistage Portfolio Selection	. 17
		1.4.3	Decision Rules	. 21
	1.5	Supply Cl	nain Network Design	. 22
	Exerc	sises		. 25
2	Two	Two Stage Problems		
-	2.1	Linear Tw	o Stage Problems	27
	2.1	2.1.1	Basic Properties	27
		2.1.2	The Expected Recourse Cost for Discrete Distributions	30
		2.1.3	The Expected Recourse Cost for General Distributions	. 33
		2.1.4	Optimality Conditions	. 38
	2.2	Polyhedra	1 Two-Stage Problems	. 42
		2.2.1	General Properties	. 42
		2.2.2	Expected Recourse Cost	. 44
		2.2.3	Optimality Conditions	. 47
	2.3	General Two-Stage Problems		. 48
		2.3.1	Problem Formulation. Interchangeability	. 48
		2.3.2	Convex Two-Stage Problems	. 50
	2.4 Nonanticipativity		pativity	. 53
		2.4.1	Scenario Formulation	. 53

Contents

		2.4.2	Dualization of Nonanticipativity Constraints	. 55	
		2.4.3	Nonanticipativity Duality for General Distributions	. 56	
		2.4.4	Value of Perfect Information	. 60	
	Exerc	ises		. 61	
3	Multistage Problems				
	3.1	Problem 1	Formulation	. 63	
		3.1.1	The General Setting	. 63	
		3.1.2	The Linear Case	. 68	
		3.1.3	Scenario Trees	. 71	
		3.1.4	Filtration Interpretation	. 74	
		3.1.5	Algebraic Formulation of Nonanticipativity Constraints	. 75	
		3.1.6	Piecewise Affine Policies	. 80	
	3.2	Duality .		. 83	
		3.2.1	Convex Multistage Problems	. 83	
		3.2.2	Optimality Conditions	. 84	
		3.2.3	Dualization of Feasibility Constraints	. 87	
		3.2.4	Dualization of Nonanticipativity Constraints	. 89	
	Exerc	ises		. 91	
4	Ontimization Models with Probabilistic Constraints				
-	4.1	Introducti	ion	. 95	
	4.2	Convexity	v in probabilistic optimization	102	
		421	Generalized concavity of functions and measures	102	
		422	Convexity of probabilistically constrained sets	115	
		423	Connectedness of probabilistically constrained sets	122	
	43	Senarable	probabilistic constraints	122	
	1.5	4 3 1	Continuity and differentiability properties of distribution	. 125	
		1.5.1	functions	123	
		432	n-Efficient points	125	
		433	The tangent and normal cones of $conv \mathcal{T}$	133	
		4.3.5	Ontimality conditions and duality theory	136	
	4.0.4 Optimization problems with non-senerable probabilistic const		tion problems with non-separable probabilistic constraints	152	
	4.4 Optim		Differentiability of probability functions and optimality	. 132	
		4.4.1	conditions	. 153	
		4.4.2	Approximations of non-separable probabilistic constraint	s 156	
	4.5	Semi-infi	nite probabilistic problems	. 164	
	Exerc	ises	· · · · · · · · · · · · · · · · · · ·	. 171	
5	Statistical Inference 17				
	5.1	Statistical	Properties of SAA Estimators	. 175	
		5.1.1	Consistency of SAA Estimators	. 177	
		5.1.2	Asymptotics of the SAA Optimal Value	. 183	
		513	Second Order Asymptotics	186	
		514	Minimax Stochastic Programs	190	
	52	Stochasti	c Generalized Fountions	10/	
	5.4	Stochasti		. 174	

6

	5.2.1	Consistency of Solutions of the SAA Generalized Equa-	106
		tions	. 196
5.0	5.2.2	Asymptotics of SAA Generalized Equations Estimators	. 198
5.3	Monte Ca	arlo Sampling Methods	. 200
	5.3.1	Exponential Rates of Convergence and Sample Size Es-	202
	5 2 2	timates in Case of Finite Feasible Set	. 202
	5.3.2	Sample Size Estimates in General Case	. 206
5 1	5.5.5 Oueri M	Finite Exponential Convergence	. 212
5.4	Quasi-Mo	Daduation Techniques	. 213
5.5		Letin Hypersuba Sampling	. 219
	5.5.1	Laun Hypercube Sampling	. 219
	5.5.2	Importance Sempling and Likelihood Patic Methods	. 221
56	J.J.J Validation	n A nelvoie	. 221
5.0	5 6 1	Estimation of the Ontimelity Con	. 225
	5.6.2	Statistical Testing of Optimality Conditions	. 224
57	Chance (Statistical Testing of Optimality Conditions	. 220
5.7	5 7 1	Monte Carlo Sampling Approach	. 231
	572	Validation of an Optimal Solution	. 232
5 8	SAA Met	thed Applied to Multistage Stochastic Programming	. 230
5.8	5 Q 1	Statistical Properties of Multistage SAA Estimators	. 242
	582	Complexity Estimates of Multistage Programs	. 243
5.0	J.0.2 Stochasti	c Approximation Method	. 240
5.9	5 0 1	Classical Approach	. 252
	502	Robust SA Approach	. 252
	5.9.2	Mirror Descent SA Method	. 255
	5.9.5	Accuracy Certificates for Mirror Descent SA Solutions	. 250
5 10	Stochasti	c Dual Dynamic Programming Method	. 200
5.10	5 10 1	Approximate Dynamic Programming Approach	271
	5 10 2	The SDDP algorithm	274
	5.10.2	Convergence Properties of the SDDP Algorithm	274
	5 10 4	Risk Averse SDDP Method	279
Exerci	5.10.4		285
Exerci			. 205
Risk A	Averse Opt	imization	289
6.1	Introduct	ion	. 289
6.2	Mean-ris	k models	. 290
	6.2.1	Main ideas of mean–risk analysis	. 290
	6.2.2	Semideviations	. 291
	6.2.3	Weighted mean deviations from quantiles	. 292
	6.2.4	Average Value-at-Risk	. 293
6.3	Coherent	Risk Measures	. 297
	6.3.1	Differentiability Properties of Risk Measures	. 303
	6.3.2	Examples of Risk Measures	. 308
	6.3.3	Law Invariant Risk Measures	. 318
	6.3.4	Spectral Risk Measures	. 323

Contents

	6.3.5	Kusuoka Representations	. 327
	6.3.6	Probability Spaces with Atoms	. 334
	6.3.7	Stochastic Orders	. 337
6.4	Ambiguo	us Chance Constraints	. 339
6.5	Optimiza	tion of Risk Measures	. 347
	6.5.1	Dualization of Nonanticipativity Constraints	. 350
	6.5.2	Interchangeability Principle for Risk Measures	. 351
	6.5.3	Examples	. 354
6.6	Statistica	Properties of Risk Measures	. 359
	6.6.1	Average Value-at-Risk	. 359
	6.6.2	Absolute Semideviation Risk Measure	. 364
	6.6.3	Von Mises Statistical Functionals	. 366
6.7	The Prob	lem of Moments	. 369
6.8	Multistag	e Risk Averse Optimization	. 374
	6.8.1	Scenario Tree Formulation	. 374
	6.8.2	Conditional Risk Mappings	. 381
	6.8.3	Dynamic Risk Measures	. 386
	6.8.4	Risk Averse Multistage Stochastic Programming	. 390
	6.8.5	Time Consistency of Multiperiod Problems	. 393
	6.8.6	Minimax Approach to Risk-Averse Multistage Program-	
		ming	. 400
	6.8.7	Portfolio Selection and Inventory Model Examples	. 403
Exerc	ises		. 407
Back	oround Ma	terial	413
7 1	Optimiza	tion and Convex Analysis	414
/11	7.1.1	Directional Differentiability	. 414
	7.1.2	Elements of Convex Analysis	. 417
	7.1.3	Optimization and Duality	420
	714	Optimality Conditions	426
	715	Perturbation Analysis	432
	716	Eniconvergence	440
72	Probabili	tv	441
/.2	7 2 1	Prohability Spaces and Random Variables	441
	7.2.1	Conditional Probability and Conditional Expectation	446
	7.2.2	Measurable Multifunctions and Random Functions	. 440
	7.2.3	Expectation Functions	. 451
	7.2.4	Uniform Laws of Large Numbers	. 451
	7.2.5	Laws of Large Numbers for Risk Measures	. 457
	7.2.0	Law of Large Numbers for Random Sets and Subdiffer-	. +02
	1.2.1	entials	467
	728	Delta Method	. 407
	7.2.8	Exponential Bounds of the Large Deviations Theory	. 470
	7.2.9 7.2.10	Uniform Exponential Bounds	. +/0 /Q1
73	7.2.10 Flements	of Functional Analysis	. 401 107
1.5		Conjugate Duality and Differentiability	. 40/
	1.3.1	Conjugate Duanty and Differentiability	. 491

7

	7.3.2 7.3.3 7.3.4 Exercises	Paired Locally Convex Topological Vector Spaces Lattice Structure	494 495 496 497	
8 Bibliographical Remarks		emarks	499	
Bibliography				
Index				

Chapter 1 Stochastic Programming Models

Andrzej Ruszczyński and Alexander Shapiro

1.1 Introduction

The readers familiar with the area of optimization can easily name several classes of optimization problems, for which advanced theoretical results exist and efficient numerical methods have been found. In that respect we can mention linear programming, quadratic programming, convex optimization, nonlinear optimization, etc. *Stochastic programming* sounds similar, but no specific formulation plays the role of the generic stochastic programming problem. The presence of random quantities in the model under consideration opens the door to wealth of different problem settings, reflecting different aspects of the applied problem at hand. The main purpose of this chapter is to illustrate the main approaches that can be followed when developing a suitable stochastic optimization model. For the purpose of presentation, these are very simplified versions of problems encountered in practice, but we hope that they will still help us to convey our main message.

1.2 Inventory

1.2.1 The Newsvendor Problem

Suppose that a company has to decide about order quantity x of a certain product to satisfy demand d. The cost of ordering is c > 0 per unit. If the demand d is larger than x, then the company makes an additional order for the unit price $b \ge 0$. The cost of this is equal to b(d-x) if d > x, and is zero otherwise. On the other hand, if d < x, then holding cost of

 $h(x-d) \ge 0$ is incurred. The total cost is then equal to¹

$$F(x,d) = cx + b[d-x]_{+} + h[x-d]_{+}.$$
(1.1)

We assume that b > c, i.e., the back order penalty cost is *larger* than the ordering cost.

The objective is to minimize the total cost F(x, d). Here x is the decision variable and the demand d is a parameter. Therefore, if the demand is known, the corresponding optimization problem can be formulated in the form

$$\min_{x \ge 0} F(x, d). \tag{1.2}$$

The objective function F(x, d) can be rewritten as

$$F(x,d) = \max\{(c-b)x + bd, (c+h)x - hd\},$$
(1.3)

which is a piecewise linear function with a minimum attained at $\bar{x} = d$. That is, if the demand d is known, then (as expected) the best decision is to order exactly the demand quantity d.

Consider now the case when the ordering decision should be made *before* a realization of the demand becomes known. One possible way to proceed in such situation is to view the demand D as a *random variable*. By capital D we denote the demand when viewed as a random variable in order to distinguish it from its particular realization d. We assume, further, that the probability distribution of D is *known*. This makes sense in situations where the ordering procedure repeats itself and the distribution of D can be estimated from historical data. Then it makes sense to talk about the expected value, denoted $\mathbb{E}[F(x, D)]$, of the total cost viewed as a function of the order quantity x. Consequently, we can write the corresponding optimization problem

$$\min_{x \ge 0} \{ f(x) := \mathbb{E}[F(x, D)] \}.$$
(1.4)

The above formulation approaches the problem by optimizing (minimizing) the total cost *on average*. What would be a possible justification of such approach? If the process repeats itself, then by the Law of Large Numbers, for a given (fixed) x, the average of the total cost, over many repetitions, will converge (with probability one) to the expectation $\mathbb{E}[F(x, D)]$, and, indeed, in that case the solution of problem (1.4) will be optimal on average.

The above problem gives a very simple example of a *two-stage problem* or a problem with a *recourse action*. At the first stage, before a realization of the demand D is known, one has to make a decision about the ordering quantity x. At the second stage, after a realization d of demand D becomes known, it may happen that d > x. In that case the company takes the recourse action of ordering the required quantity d-x at the higher cost of b > c.

The next question is how to solve the expected value problem (1.4). In the present case it can be solved in a closed form. Consider the cumulative distribution function (cdf) $H(x) := \Pr(D \le x)$ of the random variable D. Note that H(x) = 0 for all x < 0,

¹For a number $a \in \mathbb{R}$, $[a]_+$ denotes the maximum $\max\{a, 0\}$.

because the demand cannot be negative. The expectation $\mathbb{E}[F(x, D)]$ can be written in the following form

$$\mathbb{E}[F(x,D)] = b \,\mathbb{E}[D] + (c-b)x + (b+h) \int_0^x H(z)dz.$$
(1.5)

Indeed, the expectation function $f(x) = \mathbb{E}[F(x, D)]$ is a *convex* function. Moreover, since it is assumed that f(x) is well defined and finite values, it is continuous. Consequently for $x \ge 0$ we have

$$f(x) = f(0) + \int_0^x f'(z)dz,$$

where at nondifferentiable points the derivative f'(z) is understood as the right side derivative. Since $D \ge 0$ we have that $f(0) = b\mathbb{E}[D]$. Also we have that

$$f'(z) = c + \mathbb{E}\left[\frac{\partial}{\partial z} \left(b[D-z]_{+} + h[z-D]_{+}\right)\right]$$
$$= c - b\Pr(D \ge z) + h\Pr(D \le z)$$
$$= c - b(1 - H(z)) + hH(z)$$
$$= c - b + (b + h)H(z).$$

Formula (1.5) then follows.

We have that $\frac{d}{dx} \int_0^x H(z)dz = H(x)$, provided that $H(\cdot)$ is continuous at x. In this case, we can take the derivative of the right hand side of (1.5) with respect to x and equate it to zero. We conclude that the optimal solutions of problem (1.4) are defined by the equation (b+h)H(x) + c - b = 0, and hence an optimal solution of problem (1.4) is equal to the quantile

$$\bar{x} = H^{-1}(\kappa), \quad \text{with} \quad \kappa = \frac{b-c}{b+h}.$$
 (1.6)

Remark 1. Recall that for $\kappa \in (0, 1)$ the left side κ -quantile of the cdf $H(\cdot)$ is defined as $H^{-1}(\kappa) := \inf\{t : H(t) \ge \kappa\}$. In a similar way, the right side κ -quantile is defined as $\sup\{t : H(t) \le \kappa\}$. If the left and right κ -quantiles are the same, then problem (1.4) has unique optimal solution $\bar{x} = H^{-1}(\kappa)$. Otherwise the set of optimal solutions of problem (1.4) is given by the whole interval of κ -quantiles.

Suppose for the moment that the random variable D has a finitely supported distribution, i.e., it takes values $d_1, ..., d_K$ (called *scenarios*) with respective probabilities $p_1, ..., p_K$. In that case its cdf $H(\cdot)$ is a step function with jumps of size p_k at each d_k , k = 1, ..., K. Formula (1.6) for an optimal solution still holds with the corresponding left side (right side) κ -quantile, coinciding with one of the points d_k , k = 1, ..., K. For example, the scenarios may represent historical data collected over a period of time. In such case the corresponding cdf is viewed as the *empirical* cdf giving an approximation (estimation) of the true cdf, and the associated κ -quantile is viewed as the sample estimate of the κ -quantile associated with the true distribution. It is instructive to compare the quantile solution \bar{x} with a solution corresponding to one specific demand value $d := \bar{d}$, where \bar{d} is, say, the mean (expected value) of D. As it was mentioned earlier, the optimal solution of such (deterministic) problem is \bar{d} . The mean \bar{d} can be very different from the κ -quantile $\bar{x} = H^{-1}(\kappa)$. It is also worthwhile to mention that sample quantiles are typically much less sensitive than sample mean to random perturbations of the empirical data.

In applications, closed form solutions for stochastic programming problems such as (1.4) are rarely available. In the case of finitely many scenarios it is possible to model the stochastic program as a deterministic optimization problem, by writing the expected value $\mathbb{E}[F(x, D)]$ as the weighted sum:

$$\mathbb{E}[F(x,D)] = \sum_{k=1}^{K} p_k F(x,d_k).$$

The deterministic formulation (1.2) corresponds to *one* scenario *d* taken with probability one. By using the representation (1.3), we can write problem (1.2) as the linear programming problem

$$\begin{array}{l} \min_{x \ge 0, v} v \\ \text{s.t. } v \ge (c-b)x + bd, \\ v \ge (c+h)x - hd. \end{array}$$
(1.7)

Indeed, for fixed x, the optimal value of (1.7) is equal to $\max\{(c-b)x+bd, (c+h)x-hd\}$, which is equal to F(x, d). Similarly, the expected value problem (1.4), with scenarios d_1, \dots, d_K , can be written as the linear programming problem:

$$\underset{\substack{x \ge 0, v_1, \dots, v_K}{\text{subject to } v_k \ge (c-b)x + bd_k, \ k = 1, \dots, K, \\ v_k \ge (c+h)x - hd_k, \ k = 1, \dots, K.}{}$$
(1.8)

It is worth noting here the almost separable structure of problem (1.8). For a fixed x, problem (1.8) separates into the sum of optimal values of problems of the form (1.7) with $d = d_k$. As we shall see later such decomposable structure is typical for two-stage stochastic programming problems.

Worst Case Approach

One can also consider the worst case approach. That is, suppose that there are known lower and upper bounds for the demand, i.e., it is unknown that $d \in [l, u]$, where $l \leq u$ are given (nonnegative) numbers. Then the worst case formulation is

$$\underset{x \ge 0}{\operatorname{Min}} \max_{d \in [l,u]} F(x,d).$$
(1.9)

That is, while making decision x one is prepared for the worst possible outcome of the maximal cost. By (1.3) we have that

$$\max_{d\in [l,u]} F(x,d) = \max\{F(x,l),F(x,u)\}.$$

Clearly we should look at the optimal solution in the interval [l, u], and hence problem (1.9) can be written as

$$\min_{x \in [l,u]} \left\{ \psi(x) := \max \left\{ cx + h[x-l]_+, cx + b[u-x]_+ \right\} \right\}.$$

The function $\psi(x)$ is a piecewise linear convex function. Assuming that b > c we have that the optimal solution of problem (1.9) is attained at the point where h(x - l) = b(u - x). That is, the optimal solution of problem (1.9) is

$$x^* = \frac{hl + bu}{h + b}.$$

The "worst case" solution x^* can be quite different from the solution \bar{x} which is optimal on average (given in (1.6)), and could be overall conservative. For instance, if h = 0, i.e., the holding cost is zero, then $x^* = u$. On the other hand, the optimal on average solution \bar{x} depends on the distribution of the demand D which could be unavailable.

Suppose now that in addition to the lower and upper bounds of the demand we know its mean (expected value) $\bar{d} = \mathbb{E}[D]$. Of course, we have that $\bar{d} \in [l, u]$. Then we can consider the following worst case formulation

$$\underset{x \ge 0}{\operatorname{Min}} \sup_{H \in \mathfrak{M}} \mathbb{E}_{H}[F(x, D)], \qquad (1.10)$$

where \mathfrak{M} denotes the set of probability measures supported on the interval [l, u] and having mean \overline{d} , and the notation $\mathbb{E}_H[F(x, D)]$ emphasizes that the expectation is taken with respect to the cumulative distribution function (probability measure) $H(\cdot)$ of D. We study minimax problems of the form (1.10) in section 6.7 (see also problem 6.8 on page 409).

1.2.2 Chance Constraints

We have already observed that for a particular realization of the demand D, the cost $F(\bar{x}, D)$ can be quite different from the optimal-on-average cost $\mathbb{E}[F(\bar{x}, D)]$. Therefore a natural question is whether we can control the risk of the cost F(x, D) to be not "too high." For example, for a chosen value (threshold) $\tau > 0$, we may add to problem (1.4) the constraint $F(x, D) \leq \tau$ to be satisfied for *all* possible realizations of the demand D. That is, we want to make sure that the total cost will not be larger than τ in all possible circumstances. Assuming that the demand can vary in a specified uncertainty set $\mathfrak{D} \subset \mathbb{R}$, this means that the inequalities $(c - b)x + bd \leq \tau$ and $(c + h)x - hd \leq \tau$ should hold for all possible realizations $d \in \mathfrak{D}$ of the demand. That is, the ordering quantity x should satisfy the following inequalities

$$\frac{bd-\tau}{b-c} \le x \le \frac{hd+\tau}{c+h}, \quad \forall d \in \mathfrak{D}.$$
(1.11)

This could be quite restrictive if the uncertainty set \mathfrak{D} is large. In particular, if there is at least one realization $d \in \mathfrak{D}$ greater than τ/c , then the system (1.11) is inconsistent, i.e., the corresponding problem has no feasible solution.

In such situations it makes sense to introduce the constraint that the probability of F(x, D) being larger than τ is less than a specified value (significance level) $\alpha \in (0, 1)$.

This leads to a *chance* (also called *probabilistic*) constraint which can be written in the form

$$\Pr\{F(x,D) > \tau\} < \alpha, \tag{1.12}$$

or equivalently

$$\Pr\{F(x,D) \le \tau\} \ge 1 - \alpha. \tag{1.13}$$

By adding the chance constraint (1.13) to the optimization problem (1.4) we want to minimize the total cost on average, while making sure that the risk of the cost to be excessive (i.e., the probability that the cost is larger than τ) is small (i.e., less than α). We have that

$$\Pr\{F(x,D) \le \tau\} = \Pr\left\{\frac{(c+h)x-\tau}{h} \le D \le \frac{(b-c)x+\tau}{b}\right\}.$$
(1.14)

For $x \le \tau/c$ the inequalities on the right hand side of (1.14) are consistent and hence for such x (assuming that $H(\cdot)$ is continuous),

$$\Pr\{F(x,D) \le \tau\} = H\left(\frac{(b-c)x+\tau}{b}\right) - H\left(\frac{(c+h)x-\tau}{h}\right).$$
(1.15)

The chance constraint (1.13) becomes:

$$H\left(\frac{(b-c)x+\tau}{b}\right) - H\left(\frac{(c+h)x-\tau}{h}\right) \ge 1 - \alpha.$$
(1.16)

Even for small (but positive) values of α , it can be a significant relaxation of the corresponding worst case constraints (1.11).

1.2.3 Multistage Models

Suppose now that the company has a planning horizon of T periods. We model the demand as a random process D_t indexed by the time t = 1, ..., T. At the beginning, at t = 1, there is (known) inventory level y_1 . At each period t = 1, ..., T the company first observes the current inventory level y_t and then places an order to replenish the inventory level to x_t . This results in order quantity $x_t - y_t$ which clearly should be nonnegative, i.e., $x_t \ge y_t$. After the inventory is replenished, demand d_t is realized² and hence the next inventory level, at the beginning of period t + 1, becomes $y_{t+1} = x_t - d_t$. We allow backlogging and the inventory level y_t may become negative. The total cost incurred in period t is

$$c_t(x_t - y_t) + b_t[d_t - x_t]_+ + h_t[x_t - d_t]_+$$

where c_t, b_t, h_t are the ordering, backorder penalty, and holding costs per unit, respectively, at time t. We assume that $b_t > c_t > 0$ and $h_t \ge 0, t = 1, ..., T$. The objective is to minimize the expected value of the total cost over the planning horizon. This can be written as the following optimization problem

$$\underset{x_t \ge y_t}{\min} \sum_{t=1}^{T} \mathbb{E} \left\{ c_t (x_t - y_t) + b_t [D_t - x_t]_+ + h_t [x_t - D_t]_+ \right\}$$
s.t. $y_{t+1} = x_t - D_t, \ t = 1, ..., T - 1.$
(1.17)

²As before, we denote by d_t a particular realization of the random variable D_t .

For T = 1 problem (1.17) is essentially the same as the (static) problem (1.4) (the only difference is the assumption here of the initial inventory level y_1). However, for T > 1 the situation is more subtle. It is not even clear what is the exact meaning of the formulation (1.17). There are several equivalent ways to give precise meaning to the above problem. One possible way is to write equations describing the dynamics of the corresponding optimization process. That is what we discuss next.

Consider the demand process D_t , t = 1, ..., T. We denote by $D_{[t]} := (D_1, ..., D_t)$ the history of the demand process up to time t, and by $d_{[t]} := (d_1, ..., d_t)$ its particular realization. At each period (stage) t, our decision about the inventory level x_t should depend only on information available at the time of the decision, i.e., on an observed realization $d_{[t-1]}$ of the demand process, and not on future observations. This principle is called the *nonanticipativity* constraint. We assume, however, that the probability distribution of the demand process is known. That is, the conditional probability distribution of D_t , given $D_{[t-1]} = d_{[t-1]}$, is assumed to be known.

At the last stage t = T, for observed inventory level y_T , we need to solve the problem:

$$\underset{x_T \ge y_T}{\operatorname{Min}} c_T(x_T - y_T) + \mathbb{E} \left\{ b_T [D_T - x_T]_+ + h_T [x_T - D_T]_+ \left| D_{[T-1]} = d_{[T-1]} \right\}. (1.18)$$

The expectation in (1.18) is conditional on the realization $d_{[T-1]}$ of the demand process prior to the considered time T. The optimal value (and the set of optimal solutions) of problem (1.18) depends on y_T and $d_{[T-1]}$, and is denoted $Q_T(y_T, d_{[T-1]})$. At stage t = T - 1 we solve the problem

$$\underset{x_{T-1} \ge y_{T-1}}{\min} c_{T-1}(x_{T-1} - y_{T-1}) \\
+ \mathbb{E} \Big\{ b_{T-1}[D_{T-1} - x_{T-1}]_{+} + h_{T-1}[x_{T-1} - D_{T-1}]_{+} \\
+ Q_{T} \left(x_{T-1} - D_{T-1}, D_{[T-1]} \right) \Big| D_{[T-2]} = d_{[T-2]} \Big\}.$$
(1.19)

Its optimal value is denoted $Q_{T-1}(y_{T-1}, d_{[T-2]})$. Proceeding in this way backwards in time we write the following *dynamic programming* equations

$$Q_t(y_t, d_{[t-1]}) = \min_{x_t \ge y_t} c_t(x_t - y_t) + \mathbb{E} \Big\{ b_t [D_t - x_t]_+ \\ + h_t [x_t - D_t]_+ + Q_{t+1} \left(x_t - D_t, D_{[t]} \right) \big| D_{[t-1]} = d_{[t-1]} \Big\},$$
(1.20)

t = T - 1, ..., 2. Finally, at the first stage we need to solve problem

$$\min_{x_1 \ge y_1} c_1(x_1 - y_1) + \mathbb{E} \Big\{ b_1[D_1 - x_1]_+ + h_1[x_1 - D_1]_+ + Q_2 (x_1 - D_1, D_1) \Big\}.$$
(1.21)

Let us take a closer look at the above decision process. We need to understand how the dynamic programming equations (1.19)–(1.21) could be solved and what is the meaning of the solutions. Starting with the last stage t = T, we need to calculate the value functions $Q_t(y_t, d_{[t-1]})$ going backwards in time. In the present case the value functions cannot be calculated in a closed form and should be approximated numerically. For a generally distributed demand process this could be very difficult or even impossible. The situation simplifies dramatically if we assume that the random process D_t is *stagewise independent*, that is, if D_t is independent of $D_{[t-1]}$, t = 2, ..., T. Then the conditional expectations in equations (1.18)–(1.19) become the corresponding unconditional expectations. Consequently, the value functions $Q_t(y_t)$ do not depend on demand realizations and become functions of the respective univariate variables y_t only. In that case by discretization of y_t and the (one-dimensional) distribution of D_t , these value functions can be calculated in a recursive way.

Suppose now that somehow we can solve the dynamic programming equations (1.19)-(1.21). Let \bar{x}_t be a corresponding optimal solution, i.e., \bar{x}_T is an optimal solution of (1.18), \bar{x}_t is an optimal solution of the right hand side of (1.20) for t = T - 1, ..., 2, and \bar{x}_1 is an optimal solution of (1.21). We see that \bar{x}_t is a function of y_t and $d_{[t-1]}$, for t = 2, ..., T, while the first stage (optimal) decision \bar{x}_1 is independent of the data. Under the assumption of stagewise independence, $\bar{x}_t = \bar{x}_t(y_t)$ becomes a function of y_t alone. Note that y_t , in itself, is a function of $d_{t-1} = (d_1, ..., d_{t-1})$ and decisions $(x_1, ..., x_{t-1})$. Therefore we may think about a sequence of possible decisions $x_t = x_t(d_{t-1}), t = 1, ..., T$, as functions of realizations of the demand process available at the time of the decision (with the convention that x_1 is independent of the data). Such a sequence of decisions $x_t(d_{t-1})$ is called an *implementable policy*, or simply a *policy*. That is, an implementable policy is a rule which specifies our decisions, based on information available at the current stage, for any possible realization of the demand process. By definition, an implementable policy $x_t = x_t(d_{[t-1]})$ satisfies the nonanticipativity constraint. A policy is said to be *feasible* if it satisfies other constraints with probability one (w.p.1). In the present case a policy is feasible if $x_t \ge y_t$, t = 1, ..., T, for almost every realization of the demand process.

We can now formulate the optimization problem (1.17) as the problem of minimization of the expectation in (1.17) with respect to all implementable feasible policies. An optimal solution of such problem will give us an optimal policy. We have that a policy \bar{x}_t is optimal if it is given by optimal solutions of the respective dynamic programming equations. Note again that under the assumption of stagewise independence, an optimal policy $\bar{x}_t = \bar{x}_t(y_t)$ is a function of y_t alone. Moreover, in that case it is possible to give the following characterization of the optimal policy. Let x_t^* be an (unconstrained) minimizer of

$$c_t x_t + \mathbb{E}\left\{b_t [D_t - x_t]_+ + h_t [x_t - D_t]_+ + Q_{t+1} (x_t - D_t)\right\}, \ t = T, \dots, 1,$$
(1.22)

with the convention that $Q_{T+1}(\cdot) = 0$. Since $Q_{t+1}(\cdot)$ is nonnegative valued and $c_t + h_t > 0$, we have that the function in (1.22) tends to $+\infty$ if $x_t \to +\infty$. Similarly, as $b_t > c_t$, it also tends to $+\infty$ if $x_t \to -\infty$. Moreover, this function is convex and continuous (as long as it is real valued) and hence attains its minimal value. Then by using convexity of the value functions it is not difficult to show that $\bar{x}_t = \max\{y_t, x_t^*\}$ is an optimal policy. Such policy is called the *basestock* policy. A similar result holds without the assumption of stagewise independence, but then the critical values x_t^* depend on realizations of the demand process up to time t - 1.

As it was mentioned above, if the stagewise independence condition is satisfied, then each value function $Q_t(y_t)$ is a function of the variable y_t . In that case we can accurately represent $Q_t(\cdot)$ by discretization, i.e., by specifying its values at a finite number of points on the real line. Consequently, the corresponding dynamic programming equations can be accurately solved recursively going backwards in time. The situation starts to change dramatically with increase of the number of variables on which the value functions depend, like in the example discussed in the next section. The discretization approach may still work with several state variables, but it quickly becomes impractical when the dimension of the state vector increases. This is called the "curse of dimensionality." As we shall see it later, stochastic programming approaches the problem in a different way, by exploring convexity of the underlying problem, and thus attempting to solve problems with a state vector of high dimension. This is achieved by means of discretization of the random process D_t in a form of a scenario tree, which may also become prohibitively large.

1.3 Multi-Product Assembly

1.3.1 Two Stage Model

Consider a situation where a manufacturer produces n products. There are in total m different parts (or subassemblies) which have to be ordered from third party suppliers. A unit of product i requires a_{ij} units of part j, where i = 1, ..., n and j = 1, ..., m. Of course, a_{ij} may be zero for some combinations of i and j. The demand for the products is modeled as a random vector $D = (D_1, ..., D_n)$. Before the demand is known, the manufacturer may pre-order the parts from outside suppliers, at the cost of c_j per unit of part j. After the demand D is observed, the manufacturer may decide which portion of the demand is to be satisfied, so that the available numbers of parts are not exceeded. It costs additionally l_i to satisfy a unit of demand for product i, and the unit selling price of this product is q_i . The parts which are not used are assessed salvage values $s_j < c_j$. The unsatisfied demand is lost.

Suppose the numbers of parts ordered are equal to x_j , j = 1, ..., m. After the demand D becomes known, we need to determine how much of each product to make. Let us denote the numbers of units produced by z_i , i = 1, ..., n, and the numbers of parts left in inventory by y_j , j = 1, ..., m. For an observed value (a realization) $d = (d_1, ..., d_n)$ of the random demand vector D, we can find the best production plan by solving the following linear programming problem

$$\underset{z,y}{\min} \sum_{i=1}^{n} (l_i - q_i) z_i - \sum_{j=1}^{n} s_j y_j$$
s.t. $y_j = x_j - \sum_{i=1}^{n} a_{ij} z_i, \quad j = 1, \dots, m,$
 $0 \le z_i \le d_i, \quad i = 1, \dots, n, \quad y_j \ge 0, \quad j = 1, \dots, m.$

Introducing the matrix A with entries a_{ij} , where i = 1, ..., n and j = 1, ..., m, we can write this problem compactly as follows:

$$\begin{aligned}
& \underset{z,y}{\min} (l-q)^{\mathsf{T}} z - s^{\mathsf{T}} y \\
& \text{s.t. } y = x - A^{\mathsf{T}} z, \\
& 0 \le z \le d, \quad y \ge 0.
\end{aligned}$$
(1.23)

Observe that the solution of this problem, that is, the vectors z and y, depend on realization d of the demand vector D as well as on x. Let Q(x, d) denote the optimal value of problem (1.23). The quantities x_j of parts to be ordered can be determined from the following optimization problem

$$\underset{x>0}{\operatorname{Min}} \ c^{\mathsf{T}}x + \mathbb{E}[Q(x,D)],$$
(1.24)

where the expectation is taken with respect to the probability distribution of the random demand vector D. The first part of the objective function represents the ordering cost, while the second part represents the expected cost of the optimal production plan, given ordered quantities x. Clearly, for realistic data with $q_i > l_i$, the second part will be negative, so that some profit will be expected.

Problem (1.23)–(1.24) is an example of a *two stage stochastic programming problem*, with (1.23) called the *second stage problem*, and (1.24) called the *first stage problem*. As the second stage problem contains random data (random demand D), its optimal value Q(x, D) is a random variable. The distribution of this random variable depends on the first stage decisions x, and therefore the first stage problem cannot be solved without understanding of the properties of the second stage problem.

In the special case of finitely many demand scenarios d^1, \ldots, d^K occurring with positive probabilities p_1, \ldots, p_K , with $\sum_{k=1}^{K} p_k = 1$, the two stage problem (1.23)–(1.24) can be written as one large scale linear programming problem:

$$\operatorname{Min} c^{\mathsf{T}} x + \sum_{k=1}^{K} p_{k} \left[(l-q)^{\mathsf{T}} z^{k} - s^{\mathsf{T}} y^{k} \right]$$

s.t. $y^{k} = x - A^{\mathsf{T}} z^{k}, \quad k = 1, \dots, K,$
 $0 \le z^{k} \le d^{k}, \quad y^{k} \ge 0, \quad k = 1, \dots, K,$
 $x \ge 0,$
(1.25)

where the minimization is performed over vector variables x and z^k , y^k , k = 1, ..., K. We have integrated the second stage problem (1.23) into this formulation, but we had to allow for its solution (z^k, y^k) to depend on the scenario k, because the demand realization d^k is different in each scenario. Because of that, problem (1.25) has the numbers of variables and constraints roughly proportional to the number of scenarios K.

It is worthwhile to notice the following. There are three types of decision variables here. Namely, the numbers of ordered parts (vector x), the numbers of produced units (vector z) and the numbers of parts left in the inventory (vector y). These decision variables are naturally classified as the *first* and the *second* stage decision variables. That is, the first stage decisions x should be made *before* a realization of the random data becomes available and hence should be independent of the random data, while the second stage decision variables z and y are made *after* observing the random data and are functions of the data. The first stage decision variables are often referred to as "*here-and-now*" decisions (solution), and second stage decisions are referred to as "*wait-and-see*" decisions (solution). It can also be noticed that the second stage problem (1.23) is feasible for every possible realization of the random data; for example, take z = 0 and y = x. In such a situation we say that the problem has *relatively complete recourse*.

1.3.2 Chance Constrained Model

Suppose now that the manufacturer is concerned with the possibility of losing demand. The manufacturer would like the probability that all demand be satisfied to be larger than some fixed service level $1 - \alpha$, where $\alpha \in (0, 1)$ is small. In this case the problem changes in a significant way.

Observe that if we want to satisfy demand $D = (D_1, \ldots, D_n)$, we need to have $x \ge A^T D$. If we have the parts needed, there is no need for the production planning stage, as in problem (1.23). We simply produce $z_i = D_i$, $i = 1, \ldots, n$, whenever it is feasible. Also, the production costs and salvage values do not affect our problem. Consequently, the requirement of satisfying the demand with probability at least $1 - \alpha$ leads to the following formulation of the corresponding problem

$$\begin{aligned}
& \underset{x \ge 0}{\operatorname{Min}} c^{\mathsf{T}} x \\
& \text{ s.t. } \operatorname{Pr} \left\{ A^{\mathsf{T}} D < x \right\} > 1 - \alpha.
\end{aligned}$$
(1.26)

The chance (also called probabilistic) constraint in the above model is more difficult than in the case of the newsvendor model considered in section 1.2.2, because it involves a random vector $W = A^{\mathsf{T}}D$ rather than a univariate random variable.

Owing to the separable nature of the chance constraint in (1.26), we can rewrite this constraint in the following way:

$$H_W(x) \ge 1 - \alpha, \tag{1.27}$$

where $H_W(x) := \Pr(W \le x)$ is the cumulative distribution function of the *n*-dimensional random vector $W = A^T D$. Observe that if n = 1 and c > 0, then an optimal solution \bar{x} of (1.27) is given by the left side $(1 - \alpha)$ -quantile of W, that is, $\bar{x} = H_W^{-1}(1 - \alpha)$. On the other hand, in the case of multidimensional vector W its distribution has many "smallest (left side) $(1 - \alpha)$ -quantiles," and the choice of \bar{x} will depend on the relative proportions of the cost coefficients c_j . It is also worth mentioning that even when the coordinates of the demand vector D are independent, the coordinates of the vector W can be dependent, and thus the chance constraint of (1.27) cannot be replaced by a simpler expression featuring one dimensional marginal distributions.

The feasible set

$$\left\{ x \in \mathbb{R}^m_+ : \Pr\left(A^\mathsf{T} D \le x\right) \ge 1 - \alpha \right\}$$

of problem (1.26) can be written in the following equivalent form

$$\left\{x \in \mathbb{R}^m_+ : A^\mathsf{T} d \le x, \ d \in \mathfrak{D}, \ \mathsf{Pr}(\mathfrak{D}) \ge 1 - \alpha\right\}.$$
(1.28)

In the above formulation (1.28) the set \mathfrak{D} can be *any* measurable subset of \mathbb{R}^n such that probability of $D \in \mathfrak{D}$ is at least $1 - \alpha$. A considerable simplification can be achieved by choosing a *fixed* set \mathfrak{D}_{α} in such a way that $\Pr(\mathfrak{D}_{\alpha}) \geq 1 - \alpha$. In that way we obtain a simplified version of problem (1.26):

$$\begin{aligned} & \underset{x \ge 0}{\operatorname{Min}} c^{\mathsf{T}} x \\ & \text{s.t. } A^{\mathsf{T}} d \le x, \quad \forall \ d \in \mathfrak{D}_{\alpha}. \end{aligned} \tag{1.29}$$

The set \mathfrak{D}_{α} in this formulation is sometimes referred to as the *uncertainty set*, and the whole formulation as the *robust optimization problem*. Observe that in our case we can solve this problem in the following way. For each part type j we determine x_j to be the minimum number of units necessary to satisfy *every* demand $d \in \mathfrak{D}_{\alpha}$, that is

$$x_j = \max_{d \in \mathfrak{D}_{\alpha}} \sum_{i=1}^n a_{ij} d_i, \quad j = 1, \dots, n.$$

In this case the solution is completely determined by the uncertainty set \mathfrak{D}_{α} and it does not depend on the cost coefficients c_j .

The choice of the uncertainty set, satisfying the corresponding chance constraint, is not unique and often is governed by computational convenience. In this book we shall be mainly concerned with stochastic models, and we shall not discuss models and methods of robust optimization.

1.3.3 Multistage Model

Consider now the situation when the manufacturer has a planning horizon of T periods. The demand is modeled as a stochastic process D_t , t = 1, ..., T, where each $D_t = (D_{t1}, ..., D_{tn})$ is a random vector of demands for the products. The unused parts can be stored from one period to the next, and holding one unit of part j in inventory costs h_j . For simplicity, we assume that all costs and prices are the same in all periods.

It would not be reasonable to plan specific order quantities for the entire planning horizon T. Instead, one has to make orders and production decisions at successive stages, depending on the information available at the current stage. We use symbol $D_{[t]} := (D_1, \ldots, D_t)$ to denote the history of the demand process in periods $1, \ldots, t$. In every multistage decision problem it is very important to specify which of the decision variables may depend on which part of the past information.

Let us denote by $x_{t-1} = (x_{t-1,1}, \ldots, x_{t-1,n})$ the vector of quantities ordered at the beginning of stage t, before the demand vector D_t becomes known. The numbers of unites produced in stage t, will be denoted by z_t and the inventory level of parts at the end of stage t by y_t , for $t = 1, \ldots, T$. We use the subscript t - 1 for the order quantity to stress that it may depend on the past demand realizations $D_{[t-1]}$, but not on D_t , while the production and storage variables at stage t may depend on $D_{[t]}$, which includes D_t . In the special case of T = 1, we have the two-stage problem discussed in section 1.3.1; the variable x_0 corresponds to the first stage decision vector x, while z_1 and y_1 correspond to the second stage decision vectors z and y, respectively.

Suppose T > 1 and consider the last stage t = T, after the demand D_T has been observed. At this time all inventory levels y_{T-1} of the parts, as well as the last order quantities x_{T-1} are known. The problem at stage T is therefore identical to the second stage problem (1.23) of the two stage formulation:

$$\begin{aligned}
& \underset{z_{T}, y_{T}}{\text{Min}} (l-q)^{\mathsf{T}} z_{T} - s^{\mathsf{T}} y_{T} \\
& \text{s.t. } y_{T} = y_{T-1} + x_{T-1} - A^{\mathsf{T}} z_{T}, \\
& 0 \le z_{T} \le d_{T}, \quad y_{T} \ge 0,
\end{aligned} \tag{1.30}$$

where d_T is the observed realization of D_T . Denote by $Q_T(x_{T-1}, y_{T-1}, d_T)$ the optimal value of (1.30). This optimal value depends on the latest inventory levels, order quantities, and on the present demand. At stage T-1 we know realization $d_{[T-1]}$ of $D_{[T-1]}$, and thus we are concerned with the conditional expectation of the last stage cost, that is, the function

$$\mathcal{Q}_T(x_{T-1}, y_{T-1}, d_{[T-1]}) := \mathbb{E} \{ Q_T(x_{T-1}, y_{T-1}, D_T) \mid D_{[T-1]} = d_{[T-1]} \}.$$

At stage T - 1 we solve the problem

$$\underset{z_{T-1}, y_{T-1}, x_{T-1}}{\min} (l-q)^{\mathsf{T}} z_{T-1} + h^{\mathsf{T}} y_{T-1} + c^{\mathsf{T}} x_{T-1} + \mathcal{Q}_{T} (x_{T-1}, y_{T-1}, d_{[T-1]})$$
s.t. $y_{T-1} = y_{T-2} + x_{T-2} - A^{\mathsf{T}} z_{T-1},$
 $0 \le z_{T-1} \le d_{T-1}, \quad y_{T-1} \ge 0.$
(1.31)

Its optimal value is denoted by $Q_{T-1}(x_{T-2}, y_{T-2}, d_{[T-1]})$. Generally, the problem at stage $t = T - 1, \ldots, 1$ has the form

$$\begin{aligned}
& \underset{z_{t}, y_{t}, x_{t}}{\text{Min}} \left(l - q \right)^{\mathsf{T}} z_{t} + h^{\mathsf{T}} y_{t} + c^{\mathsf{T}} x_{t} + \mathcal{Q}_{t+1}(x_{t}, y_{t}, d_{[t]}) \\
& \text{s.t. } y_{t} = y_{t-1} + x_{t-1} - A^{\mathsf{T}} z_{t}, \\
& 0 \le z_{t} \le d_{t}, \quad y_{t} \ge 0,
\end{aligned} \tag{1.32}$$

with

$$\mathcal{Q}_{t+1}(x_t, y_t, d_{[t]}) := \mathbb{E} \{ Q_{t+1}(x_t, y_t, D_{[t+1]}) \mid D_{[t]} = d_{[t]} \}$$

The optimal value of problem (1.32) is denoted by $Q_t(x_{t-1}, y_{t-1}, d_{[t]})$, and the backward recursion continues. At stage t = 1, the symbol y_0 represents the initial inventory levels of the parts, and the optimal value function $Q_1(x_0, d_1)$ depends only on the initial order x_0 and realization d_1 of the first demand D_1 .

The initial problem is to determine the first order quantities x_0 . It can be written as

$$\min_{x_0 \ge 0} c^{\mathsf{T}} x_0 + \mathbb{E}[Q_1(x_0, D_1)].$$
(1.33)

Although the first stage problem (1.33) looks similar to the first stage problem (1.24) of the two stage formulation, it is essentially different since the function $Q_1(x_0, d_1)$ is not given in a computationally accessible form, but in itself is a result of recursive optimization.

1.4 Portfolio Selection

1.4.1 Static Model

Suppose that we want to invest capital W_0 in n assets, by investing an amount x_i in asset i, for i = 1, ..., n. Suppose, further, that each asset has a respective return rate R_i (per one period of time), which is unknown (uncertain) at the time we need to make our decision. We address now a question of how to distribute our wealth W_0 in an optimal way. The total wealth resulting from our investment after one period of time equals

$$W_1 = \sum_{i=1}^n \xi_i x_i,$$

where $\xi_i := 1 + R_i$. We have here the balance constraint $\sum_{i=1}^n x_i \le W_0$. Suppose, further, that one of possible investments is cash, so that we can write this balance condition as the equation $\sum_{i=1}^n x_i = W_0$. Viewing returns R_i as random variables, one can try to maximize the expected return on an investment. This leads to the following optimization problem

$$\underset{x \ge 0}{\operatorname{Max}} \mathbb{E}[W_1] \text{ subject to } \sum_{i=1}^n x_i = W_0.$$
(1.34)

We have here that

$$\mathbb{E}[W_1] = \sum_{i=1}^n \mathbb{E}[\xi_i] x_i = \sum_{i=1}^n \mu_i x_i,$$

where $\mu_i := \mathbb{E}[\xi_i] = 1 + \mathbb{E}[R_i]$ and $x = (x_1, ..., x_n) \in \mathbb{R}^n$. Therefore, problem (1.34) has a simple optimal solution of investing everything into an asset with the largest expected return rate and has the optimal value of $\mu^* W_0$, where $\mu^* := \max_{1 \le i \le n} \mu_i$. Of course, from the practical point of view such solution is not very appealing. Putting everything into one asset can be very dangerous, because if its realized return rate will be bad, then one can lose much money.

An alternative approach is to maximize expected utility of the wealth represented by a concave nondecreasing function $U(W_1)$. This leads to the following optimization problem

$$\operatorname{Max}_{x \ge 0} \mathbb{E}[U(W_1)] \text{ subject to } \sum_{i=1}^n x_i = W_0.$$
(1.35)

This approach requires specification of the utility function. For instance, let U(W) be defined as

$$U(W) := \begin{cases} (1+q)(W-a), & \text{if } W \ge a, \\ (1+r)(W-a), & \text{if } W \le a, \end{cases}$$
(1.36)

with r > q > 0 and a > 0. We can view the involved parameters as follows: a is the amount that we have to pay after return on the investment, q is the interest rate at which we can invest the additional wealth W - a, provided that W > a, and r is the interest rate at which we will have to borrow if W is less than a. For the above utility function, problem (1.35) can be formulated as the following two stage stochastic linear program

$$\operatorname{Max}_{x \ge 0} \mathbb{E}[Q(x,\xi)] \text{ subject to } \sum_{i=1}^{n} x_i = W_0,$$
(1.37)

where $Q(x,\xi)$ is the optimal value of the problem

$$\max_{y,z \in \mathbb{R}_+} (1+q)y - (1+r)z \text{ subject to } \sum_{i=1}^n \xi_i x_i = a + y - z.$$
 (1.38)

We can view the above problem (1.38) as the second stage program. Given a realization $\xi = (\xi_1, ..., \xi_n)$ of random data, we make an optimal decision by solving the corresponding optimization problem. Of course, in the present case the optimal value $Q(x, \xi)$ is a function of $W_1 = \sum_{i=1}^n \xi_i x_i$ and can be written explicitly as $U(W_1)$.

Yet another possible approach is to maximize the expected return while controlling the involved risk of the investment. There are several ways how the concept of risk could be formalized. For instance, we can evaluate risk by variability of W measured by its *variance* $\mathbb{V}ar[W] = \mathbb{E}[W^2] - (\mathbb{E}[W])^2$. Since W_1 is a linear function of the random variables ξ_i , we have that

$$\mathbb{V}\mathrm{ar}[W_1] = x^\mathsf{T} \Sigma x = \sum_{i,j=1}^n \sigma_{ij} x_i x_j,$$

where $\Sigma = [\sigma_{ij}]$ is the covariance matrix of the random vector ξ (note that the covariance matrices of the random vectors $\xi = (\xi_1, ..., \xi_n)$ and $R = (R_1, ..., R_n)$ are identical). This leads to the optimization problem of maximizing the expected return subject to the additional constraint $\mathbb{V}ar[W_1] \leq \nu$, where $\nu > 0$ is a specified constant. This problem can be written as

$$\max_{x \ge 0} \sum_{i=1}^{n} \mu_i x_i \text{ subject to } \sum_{i=1}^{n} x_i = W_0, \ x^{\mathsf{T}} \Sigma x \le \nu.$$
(1.39)

Since the covariance matrix Σ is positive semidefinite, the constraint $x^{\mathsf{T}}\Sigma x \leq \nu$ is convex quadratic, and hence (1.39) is a convex problem. Note that problem (1.39) has at least one feasible solution of investing everything in cash, in which case $\mathbb{V}ar[W_1] = 0$, and since its feasible set is compact, the problem has an optimal solution. Moreover, since problem (1.39) is convex and satisfies the Slater condition, there is no duality gap between this problem and its dual:

$$\underset{\lambda \ge 0}{\operatorname{Min}} \quad \underset{\substack{\Sigma_{i=1}^{n} x_i = W_0 \\ x \ge 0}}{\operatorname{Max}} \left\{ \sum_{i=1}^{n} \mu_i x_i - \lambda \left(x^{\mathsf{T}} \Sigma x - \nu \right) \right\}.$$
(1.40)

Consequently, there exists Lagrange multiplier $\overline{\lambda} \ge 0$ such that problem (1.39) is equivalent to the problem

$$\max_{x\geq 0} \sum_{i=1}^{n} \mu_i x_i - \bar{\lambda} x^{\mathsf{T}} \varSigma x \text{ subject to } \sum_{i=1}^{n} x_i = W_0,$$
(1.41)

The equivalence here means that the optimal value of problem (1.39) is equal to the optimal value of problem (1.41) plus the constant $\lambda \nu$, and that any optimal solution of problem (1.39) is also an optimal solution of problem (1.41). In particular, if problem (1.41) has unique optimal solution \bar{x} , then \bar{x} is also the optimal solution of problem (1.39). The corresponding Lagrange multiplier $\bar{\lambda}$ is given by an optimal solution of the dual problem (1.40). We can view the objective function of the above problem as a compromise between the expected return and its variability measured by its variance.

Another possible formulation is to minimize $\mathbb{V}ar[W_1]$ keeping the expected return $\mathbb{E}[W_1]$ above a specified value τ . That is,

$$\underset{x \ge 0}{\operatorname{Min}} x^{\mathsf{T}} \Sigma x \text{ subject to } \sum_{i=1}^{n} x_i = W_0, \quad \sum_{i=1}^{n} \mu_i x_i \ge \tau.$$
(1.42)

For appropriately chosen constants ν , $\overline{\lambda}$ and τ , problems (1.39)–(1.42) are equivalent to each other. Problems (1.41) and (1.42) are quadratic programming problems, while problem (1.39) can be formulated as a conic quadratic problem. These optimization problems

can be efficiently solved. Note finally that these optimization problems are based on the first and second order moments of random data ξ and do not require complete knowledge of the probability distribution of ξ .

We can also approach risk control by imposing chance constraints. Consider the problem

$$\max_{x \ge 0} \sum_{i=1}^{n} \mu_i x_i \text{ subject to } \sum_{i=1}^{n} x_i = W_0, \ \mathsf{Pr}\left\{\sum_{i=1}^{n} \xi_i x_i \ge b\right\} \ge 1 - \alpha.$$
(1.43)

That is, we impose the constraint that with probability at least $1 - \alpha$ our wealth $W_1 = \sum_{i=1}^{n} \xi_i x_i$ should not fall below a chosen amount *b*. Suppose the random vector ξ has a multivariate normal distribution with mean vector μ and covariance matrix Σ , written $\xi \sim \mathcal{N}(\mu, \Sigma)$. Then W_1 has normal distribution with mean $\sum_{i=1}^{n} \mu_i x_i$ and variance $x^T \Sigma x$, and

$$\Pr\{W_1 \ge b\} = \Pr\left\{Z \ge \frac{b - \sum_{i=1}^n \mu_i x_i}{\sqrt{x^{\mathsf{T}} \Sigma x}}\right\} = \Phi\left(\frac{\sum_{i=1}^n \mu_i x_i - b}{\sqrt{x^{\mathsf{T}} \Sigma x}}\right),\tag{1.44}$$

where $Z \sim \mathcal{N}(0, 1)$ has the standard normal distribution and $\Phi(z) = \Pr(Z \leq z)$ is the cdf of Z.

Therefore we can write the chance constraint of problem (1.43) in the form³

$$b - \sum_{i=1}^{n} \mu_i x_i + z_\alpha \sqrt{x^{\mathsf{T}} \Sigma x} \le 0, \qquad (1.45)$$

where $z_{\alpha} := \Phi^{-1}(1-\alpha)$ is the $(1-\alpha)$ -quantile of the standard normal distribution. Note that since matrix Σ is positive semidefinite, $\sqrt{x^{\mathsf{T}}\Sigma x}$ defines a seminorm on \mathbb{R}^n and is a convex function. Consequently, if $0 < \alpha \leq 1/2$, then $z_{\alpha} \geq 0$ and the constraint (1.45) is convex. Therefore, provided that problem (1.43) is feasible, there exists a Lagrange multiplier $\gamma \geq 0$ such that problem (1.43) is equivalent to the problem

$$\underset{x \ge 0}{\operatorname{Max}} \sum_{i=1}^{n} \mu_{i} x_{i} - \eta \sqrt{x^{\mathsf{T}} \Sigma x} \text{ subject to } \sum_{i=1}^{n} x_{i} = W_{0}, \qquad (1.46)$$

where $\eta = \gamma z_{\alpha}/(1+\gamma)$.

In financial engineering the (left side) $(1 - \alpha)$ -quantile of a random variable Y (representing losses) is called *Value-at-Risk*, i.e.,

$$V@\mathsf{R}_{\alpha}(Y) := H^{-1}(1-\alpha), \tag{1.47}$$

where $H(\cdot)$ is the cdf of Y. The chance constraint of problem (1.43) can be written in the form of a Value at Risk constraint

$$\mathsf{V}@\mathsf{R}_{\alpha}\left(b-\sum_{i=1}^{n}\xi_{i}x_{i}\right) \leq 0.$$
(1.48)

³Note that if $x^{\mathsf{T}} \Sigma x = 0$, i.e., $\mathbb{V}ar(W_1) = 0$, then the chance constraint of problem (1.43) holds iff $\sum_{i=1}^{n} \mu_i x_i \ge b$. In that case equivalence to the constraint (1.45) obviously holds.

It was possible to write chance (Value-at-Risk) constraint here in a closed form because of the assumption of joint normal distribution. Note that in the present case the random variables ξ_i cannot be negative, which indicates that the assumption of normal distribution is not very realistic.

1.4.2 Multistage Portfolio Selection

Suppose we are allowed to rebalance our portfolio in time periods t = 1, ..., T - 1, but without injecting additional cash into it. At each period t we need to make a decision about distribution of our current wealth W_t among n assets. Let $x_0 = (x_{10}, ..., x_{n0})$ be initial amounts invested in the assets. Recall that each x_{i0} is nonnegative and that the balance equation $\sum_{i=1}^{n} x_{i0} = W_0$ should hold.

We assume now that respective return rates $R_{1t}, ..., R_{nt}$, at periods t = 1, ..., T, form a random process with a known distribution. Actually we will work with the (vector valued) random process $\xi_1, ..., \xi_T$, where $\xi_t = (\xi_{1t}, ..., \xi_{nt})$ and $\xi_{it} := 1 + R_{it}, i = 1, ..., n$, t = 1, ..., T. At time period t = 1 we can rebalance the portfolio by specifying the amounts $x_1 = (x_{11}, ..., x_{n1})$ invested in the respective assets. At that time we already know the actual returns in the first period, so it is reasonable to use this information in the rebalancing decisions. Thus, our second stage decisions, at time t = 1, are actually functions of realizations of the random data vector ξ_1 , i.e., $x_1 = x_1(\xi_1)$. And so on, at time t our decision $x_t = (x_{1t}, ..., x_{nt})$ is a function $x_t = x_t(\xi_{[t]})$ of the available information given by realization $\xi_{[t]} = (\xi_1, ..., \xi_t)$ of the data process up to time t. A sequence of specific functions $x_t = x_t(\xi_{[t]}), t = 0, 1, ..., T - 1$, with x_0 being constant, defines an *implementable policy* of the decision process. It is said that such policy is *feasible* if it satisfies w.p.1 the model constraints, i.e., the nonnegativity constraints $x_{it}(\xi_{[t]}) \ge 0, i =$ 1, ..., n, t = 0, ..., T - 1, and the balance of wealth constraints

$$\sum_{i=1}^{n} x_{it}(\xi_{[t]}) = W_t$$

At period t = 1, ..., T, our wealth W_t depends on the realization of the random data process and our decisions up to time t, and is equal to

$$W_t = \sum_{i=1}^n \xi_{it} x_{i,t-1}(\xi_{[t-1]}).$$

Suppose our objective is to maximize the expected utility of this wealth at the last period, that is, we consider the problem

$$\operatorname{Max} \mathbb{E}[U(W_T)]. \tag{1.49}$$

It is a multistage stochastic programming problem, where stages are numbered from t = 0 to t = T - 1. Optimization is performed over all implementable and feasible policies.

Of course, in order to complete the description of the problem, we need to define the probability distribution of the random process $R_1, ..., R_T$. This can be done in many different ways. For example, one can construct a particular scenario tree defining time evolution of the process. If at every stage the random return of each asset is allowed to have just two continuations, independently of other assets, then the total number of scenarios is 2^{nT} . It also should be ensured that $1 + R_{it} \ge 0$, i = 1, ..., n, t = 1, ..., T, for all possible realizations of the random data.

In order to write dynamic programming equations let us consider the above multistage problem backwards in time. At the last stage t = T - 1 a realization $\xi_{[T-1]} = (\xi_1, ..., \xi_{T-1})$ of the random process is known and x_{T-2} has been chosen. Therefore, we have to solve the problem

$$\max_{\substack{x_{T-1} \ge 0, W_T}} \mathbb{E} \Big\{ U[W_T] \big| \xi_{[T-1]} \Big\}
s.t. W_T = \sum_{i=1}^n \xi_{iT} x_{i,T-1}, \sum_{i=1}^n x_{i,T-1} = W_{T-1},$$
(1.50)

where $\mathbb{E}\{U[W_T]|\xi_{[T-1]}\}\$ denotes the conditional expectation of $U[W_T]$ given $\xi_{[T-1]}$. The optimal value of the above problem (1.50) depends on W_{T-1} and $\xi_{[T-1]}$, and is denoted $Q_{T-1}(W_{T-1},\xi_{[T-1]})$.

Continuing in this way, at stage t = T - 2, ..., 1, we consider the problem

$$\max_{\substack{x_t \ge 0, W_{t+1} \\ x_t \ge 0, W_{t+1}}} \mathbb{E} \Big\{ Q_{t+1}(W_{t+1}, \xi_{[t+1]}) \big| \xi_{[t]} \Big\}$$

s.t. $W_{t+1} = \sum_{i=1}^n \xi_{i,t+1} x_{i,t}, \ \sum_{i=1}^n x_{i,t} = W_t,$ (1.51)

whose optimal value is denoted $Q_t(W_t, \xi_{[t]})$. Finally, at stage t = 0 we solve the problem

$$\max_{x_0 \ge 0, W_1} \mathbb{E}[Q_1(W_1, \xi_1)]$$

s.t. $W_1 = \sum_{i=1}^n \xi_{i1} x_{i0}, \ \sum_{i=1}^n x_{i0} = W_0.$ (1.52)

For a general distribution of the data process ξ_t it may be hard to solve these dynamic programming equations. The situation simplifies dramatically if the process ξ_t is stagewise independent, i.e., ξ_t is (stochastically) independent of $\xi_1, ..., \xi_{t-1}$, for t = 2, ..., T. Of course, the assumption of stagewise independence is not very realistic in financial models, but it is instructive to see the dramatic simplifications it allows. In that case the corresponding conditional expectations become unconditional expectations and the cost-to-go (value) function $Q_t(W_t)$, t = 1, ..., T - 1, does not depend on $\xi_{[t]}$. That is, $Q_{T-1}(W_{T-1})$ is the optimal value of the problem

$$\max_{x_{T-1} \ge 0, W_T} \mathbb{E} \{ U[W_T] \}$$

s.t. $W_T = \sum_{i=1}^n \xi_{iT} x_{i,T-1}, \ \sum_{i=1}^n x_{i,T-1} = W_{T-1},$

and $Q_t(W_t)$ is the optimal value of

$$\max_{\substack{x_t \ge 0, W_{t+1}}} \mathbb{E}\{Q_{t+1}(W_{t+1})\}$$

s.t. $W_{t+1} = \sum_{i=1}^n \xi_{i,t+1} x_{i,t}, \ \sum_{i=1}^n x_{i,t} = W_t,$

for t = T - 2, ..., 1.

The other relevant question is what utility function to use. Let us consider the *log-arithmic* utility function $U(W) := \ln W$. Note that this utility function is defined for W > 0. For positive numbers a and w and for $W_{T-1} = w$ and $W_{T-1} = aw$, there is a one-to-one correspondence $x_{T-1} \leftrightarrow ax_{T-1}$ between the feasible sets of the corresponding problem (1.50). For the logarithmic utility function this implies the following relation between the optimal values of these problems

$$Q_{T-1}(aw,\xi_{[T-1]}) = Q_{T-1}(w,\xi_{[T-1]}) + \ln a.$$
(1.53)

That is, at stage t = T - 1 we solve the problem

$$\max_{x_{T-1} \ge 0} \mathbb{E} \left\{ \ln \left(\sum_{i=1}^{n} \xi_{i,T} x_{i,T-1} \right) \left| \xi_{[T-1]} \right\} \text{ s.t. } \sum_{i=1}^{n} x_{i,T-1} = W_{T-1}.$$
 (1.54)

By (1.53) its optimal value is

$$Q_{T-1}\left(W_{T-1},\xi_{[T-1]}\right) = \nu_{T-1}\left(\xi_{[T-1]}\right) + \ln W_{T-1},$$

where $\nu_{T-1}(\xi_{[T-1]})$ denotes the optimal value of (1.54) for $W_{T-1} = 1$. At stage t = T-2 we solve problem

$$\max_{x_{T-2} \ge 0} \mathbb{E} \left\{ \nu_{T-1} \left(\xi_{[T-1]} \right) + \ln \left(\sum_{i=1}^{n} \xi_{i,T-1} x_{i,T-2} \right) \left| \xi_{[T-2]} \right. \right\}$$
s.t.
$$\sum_{i=1}^{n} x_{i,T-2} = W_{T-2}.$$
(1.55)

Of course, we have that

$$\mathbb{E}\left\{\nu_{T-1}\left(\xi_{[T-1]}\right) + \ln\left(\sum_{i=1}^{n}\xi_{i,T-1}x_{i,T-2}\right)\left|\xi_{[T-2]}\right\}\right\}$$
$$= \mathbb{E}\left\{\nu_{T-1}\left(\xi_{[T-1]}\right)\left|\xi_{[T-2]}\right\} + \mathbb{E}\left\{\ln\left(\sum_{i=1}^{n}\xi_{i,T-1}x_{i,T-2}\right)\left|\xi_{[T-2]}\right.\right\},\$$

and hence by arguments similar to (1.53), the optimal value of (1.55) can be written as

$$Q_{T-2}\left(W_{T-2},\xi_{[T-2]}\right) = \mathbb{E}\left\{\nu_{T-1}\left(\xi_{[T-1]}\right) \left|\xi_{[T-2]}\right\} + \nu_{T-2}\left(\xi_{[T-2]}\right) + \ln W_{T-2},\right\}$$

where $\nu_{T-2}(\xi_{[T-2]})$ is the optimal value of the problem

$$\max_{x_{T-2} \ge 0} \mathbb{E} \left\{ \ln \left(\sum_{i=1}^{n} \xi_{i,T-1} x_{i,T-2} \right) \left| \xi_{[T-2]} \right\} \text{ s.t. } \sum_{i=1}^{n} x_{i,T-2} = 1.$$

An identical argument applies at earlier stages. Therefore, it suffices to solve at each stage t = T - 1, ..., 1, 0, the corresponding optimization problem

$$\max_{x_t \ge 0} \mathbb{E} \left\{ \ln \left(\sum_{i=1}^n \xi_{i,t+1} x_{i,t} \right) \Big| \xi_{[t]} \right\} \text{ s.t. } \sum_{i=1}^n x_{i,t} = W_t,$$
(1.56)

in a completely myopic fashion.

By definition, we set ξ_0 to be constant, so that for the first stage problem, at t = 0, the corresponding expectation is unconditional. An optimal solution $\bar{x}_t = \bar{x}_t(W_t, \xi_{[t]})$ of problem (1.56) gives an optimal policy. In particular, first stage optimal solution \bar{x}_0 is given by an optimal solution of the problem:

$$\max_{x_0 \ge 0} \mathbb{E}\left\{ \ln\left(\sum_{i=1}^n \xi_{i1} x_{i0}\right) \right\} \text{ s.t. } \sum_{i=1}^n x_{i0} = W_0.$$
(1.57)

We also have here that the optimal value, denoted ϑ^* , of the optimization problem (1.49) can be written as follows

$$\vartheta^* = \ln W_0 + \nu_0 + \sum_{t=1}^{T-1} \mathbb{E} \left[\nu_t(\xi_{[t]}) \right], \qquad (1.58)$$

where $\nu_t(\xi_{[t]})$ is the optimal value of problem (1.56) for $W_t = 1$. Note that $\nu_0 + \ln W_0$ is the optimal value of problem (1.57) with ν_0 being the (deterministic) optimal value of (1.57) for $W_0 = 1$.

If the random process ξ_t is stagewise independent, then conditional expectations in (1.56) are the same as the corresponding unconditional expectations, and hence optimal values $\nu_t(\xi_{[t]}) = \nu_t$ do not depend on $\xi_{[t]}$ and are given by the optimal value of the problem

$$\max_{x_t \ge 0} \mathbb{E} \left\{ \ln \left(\sum_{i=1}^n \xi_{i,t+1} x_{i,t} \right) \right\} \quad \text{s.t.} \quad \sum_{i=1}^n x_{i,t} = 1.$$
(1.59)

Also in the stagewise independent case the optimal policy can be described as follows. Let $x_t^* = (x_{1t}^*, ..., x_{nt}^*)$ be the optimal solution of (1.59), t = 0, ..., T - 1. Such optimal solution is unique by strict concavity of the logarithm function. Then

$$\bar{x}_t(W_t) := W_t x_t^*, \ t = 0, ..., T - 1,$$

defines the optimal policy.

Consider now the *power* utility function $U(W) := W^{\gamma}$, with $1 \ge \gamma > 0$, defined for $W \ge 0$. Suppose again that the random process ξ_t is *stagewise independent*. Recall that this condition implies that the cost-to-go function $Q_t(W_t)$, t = 1, ..., T - 1, depends only

on W_t . By using arguments similar to the analysis for the logarithmic utility function it is not difficult to show that $Q_{T-1}(W_{T-1}) = W_{T-1}^{\gamma}Q_{T-1}(1)$, and so on. The optimal policy $\bar{x}_t = \bar{x}_t(W_t)$ is obtained in a myopic way as an optimal solution of the problem

$$\max_{x_t \ge 0} \mathbb{E}\left\{ \left(\sum_{i=1}^n \xi_{i,t+1} x_{it} \right)^{\gamma} \right\} \quad \text{s.t.} \quad \sum_{i=1}^n x_{it} = W_t.$$
(1.60)

That is, $\bar{x}_t(W_t) = W_t x_t^*$, where x_t^* is an optimal solution of problem (1.60) for $W_t = 1$, t = 0, ..., T - 1. In particular, the first stage optimal solution \bar{x}_0 is obtained in a myopic way by solving the problem

$$\max_{x_0 \ge 0} \mathbb{E} \left\{ \left(\sum_{i=1}^n \xi_{i1} x_{i0} \right)^{\gamma} \right\} \quad \text{s.t. } \sum_{i=1}^n x_{i0} = W_0$$

The optimal value ϑ^* of the corresponding multistage problem (1.49) is

$$\vartheta^* = W_0^\gamma \prod_{t=0}^{T-1} \eta_t, \tag{1.61}$$

where η_t is the optimal value of problem (1.60) for $W_t = 1$.

The above myopic behavior of multistage stochastic programs is rather exceptional. A more realistic situation occurs in the presence of transaction costs. These are costs associated with the changes in the numbers of units (stocks, bonds) held. Introduction of transaction costs will destroy such myopic behavior of optimal policies.

1.4.3 Decision Rules

Consider the following policy. Let $x_t^* = (x_{1t}^*, ..., x_{nt}^*)$, t = 0, ..., T - 1, be vectors such that $x_t^* \ge 0$ and $\sum_{i=1}^n x_{it}^* = 1$. Define the *fixed mix* policy

$$x_t(W_t) := W_t x_t^*, \quad t = 0, ..., T - 1.$$
 (1.62)

As it was discussed above, under the assumption of stagewise independence, such policies are optimal for the logarithmic and power utility functions provided that x_t^* are optimal solutions of the respective problems (problem (1.59) for the logarithmic utility function and problem (1.60) with $W_t = 1$ for the power utility function). In other problems, a policy of form (1.62) may be non-optimal. However, it is readily implementable, once the current wealth W_t is observed. As it was mentioned before, rules for calculating decisions as functions of the observations gathered up to time t, similarly to (1.62), are called policies, or alternatively *decision rules*.

We analyze now properties of the decision rule (1.62) under the simplifying assumption of stagewise independence. We have

$$W_{t+1} = \sum_{i=1}^{n} \xi_{i,t+1} x_{it}(W_t) = W_t \sum_{i=1}^{n} \xi_{i,t+1} x_{it}^*.$$
 (1.63)

Since the random process $\xi_1, ..., \xi_T$ is stagewise independent, by independence of ξ_{t+1} and W_t we have

$$\mathbb{E}[W_{t+1}] = \mathbb{E}[W_t] \mathbb{E}\left(\sum_{i=1}^n \xi_{i,t+1} x_{it}^*\right) = \mathbb{E}[W_t] \underbrace{\sum_{i=1}^n \mu_{i,t+1} x_{it}^*}_{x_t^{*^\top} \mu_{t+1}}, \quad (1.64)$$

where $\mu_t := \mathbb{E}[\xi_t]$. Consequently, by induction,

$$\mathbb{E}[W_t] = \prod_{\tau=1}^t \left(\sum_{i=1}^n \mu_{i\tau} x_{i,\tau-1}^* \right) = \prod_{\tau=1}^t \left(x_{\tau-1}^{*\mathsf{T}} \mu_{\tau} \right).$$

In order to calculate the variance of W_t we use the formula

$$\mathbb{V}\mathrm{ar}(Y) = \mathbb{E}(\underbrace{\mathbb{E}[(Y - \mathbb{E}(Y|X))^2|X]}_{\mathbb{V}\mathrm{ar}(Y|X)}) + \underbrace{\mathbb{E}([\mathbb{E}(Y|X) - \mathbb{E}Y]^2)}_{\mathbb{V}\mathrm{ar}[\mathbb{E}(Y|X)]},$$
(1.65)

where X and Y are random variables. Applying (1.65) to (1.63) with $Y := W_{t+1}$ and $X := W_t$ we obtain

$$\mathbb{V}\mathrm{ar}[W_{t+1}] = \mathbb{E}[W_t^2] \mathbb{V}\mathrm{ar}\left(\sum_{i=1}^n \xi_{i,t+1} x_{it}^*\right) + \mathbb{V}\mathrm{ar}[W_t] \left(\sum_{i=1}^n \mu_{i,t+1} x_{it}^*\right)^2.$$
(1.66)

Recall that $\mathbb{E}[W_t^2] = \mathbb{V}ar[W_t] + (\mathbb{E}[W_t])^2$ and $\mathbb{V}ar(\sum_{i=1}^n \xi_{i,t+1} x_{it}^*) = x_t^* \mathcal{I} \Sigma_{t+1} x_t^*$, where Σ_{t+1} is the covariance matrix of ξ_{t+1} .

It follows from (1.64) and (1.66) that

$$\frac{\mathbb{V}\mathrm{ar}[W_{t+1}]}{(\mathbb{E}[W_{t+1}])^2} = \frac{x_t^{*\mathsf{T}} \Sigma_{t+1} x_t^{*}}{(x_t^{*\mathsf{T}} \mu_{t+1})^2} + \frac{\mathbb{V}\mathrm{ar}[W_t]}{(\mathbb{E}[W_t])^2},\tag{1.67}$$

and hence

$$\frac{\operatorname{Var}[W_t]}{(\mathbb{E}[W_t])^2} = \sum_{\tau=1}^t \frac{\operatorname{Var}\left(\sum_{i=1}^n \xi_{i,\tau} x_{i,\tau-1}^*\right)}{\left(\sum_{i=1}^n \mu_{i\tau} x_{i,\tau-1}^*\right)^2} = \sum_{\tau=1}^t \frac{x_{\tau-1}^{*\mathsf{T}} \Sigma_\tau x_{\tau-1}^*}{(x_{\tau-1}^{*\mathsf{T}} \mu_{\tau})^2}, \quad t = 1, ..., T.$$
(1.68)

This shows that if the terms $\frac{x_{\tau-1}^{*T}\Sigma_{\tau}x_{\tau-1}^{*}}{(x_{\tau-1}^{*T}\mu_{\tau})^2}$ are of the same order for $\tau = 1, ..., T$, then the ratio of the standard deviation $\sqrt{\mathbb{Var}[W_T]}$ to the expected wealth $\mathbb{E}[W_T]$ is of order $O(\sqrt{T})$ with increase of the number of stages T.

1.5 Supply Chain Network Design

In this section we discuss a stochastic programming approach to modeling a supply chain network design. A supply chain is a network of suppliers, manufacturing plants, warehouses, and distribution channels organized to acquire raw materials, convert these raw materials to finished products, and distribute these products to customers. Let us first describe a deterministic mathematical formulation for the supply chain design problem.

Denote by S, \mathcal{P} and C the respective (finite) sets of suppliers, processing facilities and customers. The union $\mathcal{N} := S \cup \mathcal{P} \cup C$ of these sets is viewed as the set of nodes of a directed graph $(\mathcal{N}, \mathcal{A})$, where \mathcal{A} is a set of arcs (directed links) connecting these nodes in a way representing flow of the products. The processing facilities include manufacturing centers \mathcal{M} , finishing facilities \mathcal{F} and warehouses \mathcal{W} , i.e., $\mathcal{P} = \mathcal{M} \cup \mathcal{F} \cup \mathcal{W}$. Further, a manufacturing center $i \in \mathcal{M}$ or a finishing facility $i \in \mathcal{F}$ consists of a set of manufacturing or finishing machines \mathcal{H}_i . Thus the set \mathcal{P} includes the processing centers as well as the machines in these centers. Let \mathcal{K} be the set of products flowing through the supply chain.

The supply chain configuration decisions consist of deciding which of the processing centers to build (major configuration decisions), and which processing and finishing machines to procure (minor configuration decisions). We assign a binary variable $x_i = 1$, if a processing facility *i* is built or machine *i* is procured, and $x_i = 0$ otherwise. The operational decisions consist of routing the flow of product $k \in \mathcal{K}$ from the supplier to the customers. By y_{ij}^k we denote the flow of product *k* from a node *i* to a node *j* of the network where $(i, j) \in \mathcal{A}$. A deterministic mathematical model for the supply chain design problem can be written as follows

$$\underset{x,y}{\operatorname{Min}} \sum_{i \in \mathcal{P}} c_i x_i + \sum_{k \in \mathcal{K}} \sum_{(i,j) \in \mathcal{A}} q_{ij}^k y_{ij}^k$$
(1.69)

s.t.
$$\sum_{i\in\mathcal{N}} y_{ij}^k - \sum_{\ell\in\mathcal{N}} y_{j\ell}^k = 0, \ j\in\mathcal{P}, \ k\in\mathcal{K},$$
(1.70)

$$\sum_{i \in \mathcal{N}} y_{ij}^k \ge d_j^k, \ j \in \mathcal{C}, \ k \in \mathcal{K},$$
(1.71)

$$\sum_{i \in \mathcal{N}} y_{ij}^k \le s_j^k, \ j \in \mathcal{S}, \ k \in \mathcal{K},$$
(1.72)

$$\sum_{k \in \mathcal{K}} r_j^k \left(\sum_{i \in \mathcal{N}} y_{ij}^k \right) \le m_j x_j, \ j \in \mathcal{P},$$
(1.73)

$$x \in \mathcal{X}, \ y \ge 0. \tag{1.74}$$

Here c_i denotes the investment cost for building facility i or procuring machine i, q_{ij}^k denotes the per-unit cost of processing product k at facility i and/or transporting product k on arc $(i, j) \in \mathcal{A}$, d_j^k denotes the demand of product k at node j, s_j^k denotes the supply of product k at node j, r_j^k denotes per-unit processing requirement for product k at node j, m_j denotes capacity of facility j, $\mathcal{X} \subset \{0, 1\}^{|\mathcal{P}|}$ is a set of binary variables and $y \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{K}|}$ is a vector with components y_{ij}^k . All cost components are annualized.

The objective function (1.69) is aimed at minimizing total investment and operational costs. Of course, a similar model can be constructed for maximizing the profits. The set \mathcal{X} represents logical dependencies and restrictions, such as $x_i \leq x_j$ for all $i \in \mathcal{H}_j$ and $j \in \mathcal{P}$ or $j \in \mathcal{F}$, i.e., machine $i \in \mathcal{H}_j$ should only be procured if facility j is built (since x_i are binary, the constraint $x_i \leq x_j$ means that $x_i = 0$ if $x_j = 0$). Constraints (1.70) enforce the flow conservation of product k across each processing node j. Constraints (1.71) require that the total flow of product k to a customer node j, should exceed the

demand d_j^k at that node. Constraints (1.72) require that the total flow of product k from a supplier node j, should be less than the supply s_j^k at that node. Constraints (1.73) enforce capacity constraints of the processing nodes. The capacity constraints then require that the total processing requirement of all products flowing into a processing node j should be smaller than the capacity m_j of facility j if it is built $(x_j = 1)$. If facility j is not built $(x_j = 0)$ the constraint will force all flow variables $y_{ij}^k = 0$ for all $i \in \mathcal{N}$. Finally, constraint (1.74) enforces feasibility constraint $x \in \mathcal{X}$ and the nonnegativity of the flow variables corresponding to an arc $(ij) \in \mathcal{A}$ and product $k \in \mathcal{K}$.

It will be convenient to write problem (1.69)–(1.74) in the following compact form

$$\min_{\mathbf{x}\in\mathcal{X},\ y\geq 0} c^{\mathsf{T}}x + q^{\mathsf{T}}y \tag{1.75}$$

s.t.
$$Ny = 0,$$
 (1.76)

$$Cy \ge d, \tag{1.77}$$

$$Sy \le s, \tag{1.78}$$

$$Ry \le Mx, \tag{1.79}$$

where vectors c, q, d and s correspond to investment costs, processing/transportation costs, demands, and supplies, respectively, matrices N, C and S are appropriate matrices corresponding to the summations on the left-hand-side of the respective expressions. The notation R corresponds to a matrix of r_j^k , and the notation M corresponds to a matrix with m_j along the diagonal.

It is realistic to assume that at time a decision about vector $x \in \mathcal{X}$ should be made, i.e., which facilities to built and machines to procure, there is an uncertainty about parameters involved in operational decisions represented by vector $y \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{K}|}$. This naturally classifies decision variables x as the first stage decision variables and y as the second stage decision variables. Note that problem (1.75)–(1.79) can be written in the following equivalent form as a two stage program:

$$\operatorname{Min}_{x \in \mathcal{X}} c^{\mathsf{T}} x + Q(x,\xi), \tag{1.80}$$

where $Q(x,\xi)$ is the optimal value of the second stage problem

$$\underset{y \ge 0}{\operatorname{Min}} q^{\mathsf{T}} y \tag{1.81}$$

s.t.
$$Ny = 0,$$
 (1.82)

$$Cy \ge d,\tag{1.83}$$

$$Sy \le s,$$
 (1.84)

$$Ry \le Mx,\tag{1.85}$$

with $\xi = (q, d, s, R, M)$ being vector of the involved parameters. Of course, the above optimization problem depends on the data vector ξ . If some of the data parameters are uncertain, then the deterministic problem (1.80) does not make much sense since it depends on unknown parameters.

Suppose now that we can model uncertain components of the data vector ξ as random variables with a specified joint probability distribution. Then we can formulate the
following stochastic programming problem

$$\underset{x \in \mathcal{X}}{\operatorname{Min}} c^{\mathsf{T}} x + \mathbb{E}[Q(x,\xi)],$$
(1.86)

where the expectation is taken with respect to the probability distribution of the random vector ξ . That is, the cost of the second stage problem enters the objective of the first stage problem *on average*. A distinctive feature of the stochastic programming problem (1.86) is that the first stage problem here is a combinatorial problem with binary decision variables and finite feasible set \mathcal{X} . On the other hand, the second stage problem (1.81)–(1.85) is a linear programming problem and its optimal value $Q(x, \xi)$ is convex in x (if x is viewed as a vector in $\mathbb{R}^{|\mathcal{P}|}$).

It could happen that for some $x \in \mathcal{X}$ and some realizations of the data ξ the corresponding second stage problem (1.81)–(1.85) is infeasible, i.e., the constraints (1.82)–(1.85) define an empty set. In that case, by definition, $Q(x,\xi) = +\infty$, i.e., we apply an infinite penalization for infeasibility of the second stage problem. For example, it could happen that demand d is not satisfied, i.e., $Cy \leq d$ with some inequalities strict, for any $y \geq 0$ satisfying constraints (1.82), (1.84) and (1.85). Sometimes this can be resolved by a recourse action. That is, if demand is not satisfied, then there is a possibility of supplying the deficit d - Cy at a penalty cost. This can be modelled by writing the second stage problem in the form

$$\underset{y \ge 0, z \ge 0}{\operatorname{Min}} q^{\mathsf{T}} y + h^{\mathsf{T}} z$$
(1.87)

s.t.
$$Ny = 0,$$
 (1.88)

$$Cy + z \ge d,\tag{1.89}$$

$$Sy \le s, \tag{1.90}$$

$$Ry \le Mx,\tag{1.91}$$

where h represents the vector of (positive) recourse costs. Note that the above problem (1.87)–(1.91) is always feasible, for example y = 0 and $z \ge d$ clearly satisfy the constraints of this problem.

Exercises

- 1.1. Consider the expected value function f(x) := E[F(x, D)], where function F(x, d) is defined in (1.1). (i) Show that function F(x, d) is convex in x, and hence f(x) is also convex. (ii) Show that f(·) is differentiable at a point x > 0 iff the cdf H(·) of D is continuous at x.
- 1.2. Let H(z) be the cdf of a random variable Z and $\kappa \in (0, 1)$. Show that the minimum in the definition $H^{-1}(\kappa) = \inf\{t : H(t) \ge \kappa\}$ of the left side quantile is always attained.
- 1.3. Consider the chance constrained problem discussed in section 1.2.2. (i) Show that system (1.11) has no feasible solution if there is a realization of d greater than τ/c. (ii) Verify equation (1.15). (iii) Assume that the probability distribution of the demand D is supported on an interval [l, u] with 0 ≤ l ≤ u < +∞. Show that if the

significance level $\alpha = 0$, then the constraint (1.16) becomes

$$\frac{bu-\tau}{b-c} \le x \le \frac{hl+\tau}{c+h},$$

and hence is equivalent to (1.11) for $\mathfrak{D} = [l, u]$.

- 1.4. Show that the optimal value functions $Q_t(y_t, d_{[t-1]})$, defined in (1.20), are convex in y_t .
- 1.5. Assuming the stagewise independence condition, show that the basestock policy $\bar{x}_t = \max\{y_t, x_t^*\}$, for the inventory model, is optimal (recall that x_t^* denotes a minimizer of (1.22)).
- 1.6. Consider the assembly problem discussed in section 1.3.1 in the case when all demand has to be satisfied, by making additional orders of the missing parts. In this case, the cost of each additionally ordered part j is $r_j > c_j$. Formulate the problem as a linear two stage stochastic programming problem.
- 1.7. Consider the assembly problem discussed in section 1.3.3 in the case when all demand has to be satisfied, by backlogging the excessive demand, if necessary. In this case, it costs b_i to delay delivery of a unit of product *i* by one period. Additional orders of the missing parts can also be made after the last demand D_T becomes known. Formulate the problem as a linear multistage stochastic programming problem.
- 1.8. Show that for utility function U(W), of the form (1.36), problems (1.35) and (1.37)–(1.38) are equivalent.
- 1.9. Show that variance of the random return $W_1 = \xi^T x$ is given by formula $\mathbb{V}ar[W_1] = x^T \Sigma x$, where $\Sigma = \mathbb{E}\left[(\xi \mu)(\xi \mu)^T\right]$ is the covariance matrix of the random vector ξ and $\mu = \mathbb{E}[\xi]$.
- 1.10. Show that the optimal value function $Q_t(W_t, \xi_{[t]})$, defined in (1.51), is concave in W_t .
- 1.11. Let D be a random variable with cdf $H(t) = \Pr(D \le t)$ and $D^1, ..., D^N$ be an iid random sample of D with the corresponding empirical cdf $\hat{H}_N(\cdot)$. Let $a = H^{-1}(\kappa)$ and $b = \sup\{t : H(t) \le \kappa\}$ be respective left and right-side κ -quantiles of $H(\cdot)$. Show that min $\left\{ \left| \hat{H}_N^{-1}(\kappa) a \right|, \left| \hat{H}_N^{-1}(\kappa) b \right| \right\}$ tends w.p.1 to 0 as $N \to \infty$.

Chapter 2 Two Stage Problems

Andrzej Ruszczyński and Alexander Shapiro

2.1 Linear Two Stage Problems

2.1.1 Basic Properties

In this section we discuss two-stage stochastic linear programming problems of the form

$$\begin{aligned}
& \underset{x \in \mathbb{R}^n}{\min} c^{-1} x + \mathbb{E}[Q(x,\xi)] \\
& \text{s.t. } Ax = b, \ x \ge 0,
\end{aligned}$$
(2.1)

where $Q(x,\xi)$ is the optimal value of the second stage problem

$$\begin{aligned}
&\underset{y \in \mathbb{R}^m}{\min} \ q^\mathsf{T}y \\
&\text{s.t. } Tx + Wy = h, \ y \ge 0.
\end{aligned}$$
(2.2)

Here $\xi := (q, h, T, W)$ are the data of the second stage problem. We view some or all elements of vector ξ as random and the expectation operator at the first stage problem (2.1) is taken with respect to the probability distribution of ξ . Often, we use the same notation ξ to denote a random vector and its particular realization. Which one of these two meanings will be used in a particular situation will usually be clear from the context. If in doubt, then we will write $\xi = \xi(\omega)$ to emphasize that ξ is a random vector defined on a corresponding probability space. We denote by $\Xi \subset \mathbb{R}^d$ the support of the probability distribution of ξ .

If for some x and $\xi \in \Xi$ the second stage problem (2.2) is infeasible, then by definition $Q(x,\xi) = +\infty$. It could also happen that the second stage problem is unbounded from below and hence $Q(x,\xi) = -\infty$. This is somewhat pathological situation, meaning that

for some value of the first stage decision vector and a realization of the random data, the value of the second stage problem can be improved indefinitely. Models exhibiting such properties should be avoided (we discuss this later).

The second stage problem (2.2) is a linear programming problem. Its dual problem can be written in the form

$$\begin{aligned}
& \underset{\pi}{\operatorname{Max}} \pi^{\mathsf{T}}(h - Tx) \\
& \text{s.t. } W^{\mathsf{T}}\pi \leq q.
\end{aligned}$$
(2.3)

By the theory of linear programming, the optimal values of problems (2.2) and (2.3) are equal to each other, unless both problems are infeasible. Moreover, if their common optimal value is finite, then each problem has a nonempty set of optimal solutions.

Consider the function

$$s_q(\chi) := \inf \{ q^{\mathsf{T}} y : Wy = \chi, \ y \ge 0 \}.$$
 (2.4)

Clearly, $Q(x,\xi) = s_q(h - Tx)$. By the duality theory of linear programming, if the set

$$\Pi(q) := \left\{ \pi : W^{\mathsf{T}} \pi \le q \right\}$$
(2.5)

is nonempty, then

$$s_q(\chi) = \sup_{\pi \in \Pi(q)} \pi^\mathsf{T} \chi, \tag{2.6}$$

i.e., $s_q(\cdot)$ is the support function of the set $\Pi(q)$. The set $\Pi(q)$ is convex, closed, and polyhedral. Hence, it has a finite number of extreme points. (If, moreover, $\Pi(q)$ is bounded, then it coincides with the convex hull of its extreme points.) It follows that if $\Pi(q)$ is nonempty, then $s_q(\cdot)$ is a positively homogeneous *polyhedral* function. If the set $\Pi(q)$ is empty, then the infimum on the right hand side of (2.4) may take only two values: $+\infty$ or $-\infty$. In any case it is not difficult to verify directly that the function $s_q(\cdot)$ is convex.

Proposition 2.1. For any given ξ , the function $Q(\cdot,\xi)$ is convex. Moreover, if the set $\{\pi : W^{\mathsf{T}}\pi \leq q\}$ is nonempty and problem (2.2) is feasible for at least one x, then the function $Q(\cdot,\xi)$ is polyhedral.

Proof. Since $Q(x,\xi) = s_q(h - Tx)$, the above properties of $Q(\cdot,\xi)$ follow from the corresponding properties of the function $s_q(\cdot)$. \Box

Differentiability properties of the function $Q(\cdot,\xi)$ can be described as follows.

Proposition 2.2. Suppose that for given $x = x_0$ and $\xi \in \Xi$, the value $Q(x_0, \xi)$ is finite. Then $Q(\cdot, \xi)$ is subdifferentiable at x_0 and

$$\partial Q(x_0,\xi) = -T^{\mathsf{T}}\mathfrak{D}(x_0,\xi), \qquad (2.7)$$

where

$$\mathfrak{D}(x,\xi) := \arg \max_{\pi \in \Pi(q)} \pi^{\mathsf{T}}(h - Tx)$$

is the set of optimal solutions of the dual problem (2.3).

Proof. Since $Q(x_0, \xi)$ is finite, the set $\Pi(q)$ defined in (2.5) is nonempty, and hence $s_q(\chi)$ is its support function. It is straightforward to see from the definitions that the support function $s_q(\cdot)$ is the conjugate function of the indicator function

$$\mathbb{I}_q(\pi) := \begin{cases} 0, & \text{if } \pi \in \Pi(q), \\ +\infty, & \text{otherwise.} \end{cases}$$

Since the set $\Pi(q)$ is convex and closed, the function $\mathbb{I}_q(\cdot)$ is convex and lower semicontinuous. It follows then by the Fenchel–Moreau Theorem (Theorem 7.5) that the conjugate of $s_q(\cdot)$ is $\mathbb{I}_q(\cdot)$. Therefore, for $\chi_0 := h - Tx_0$, we have (see (7.24))

$$\partial s_q(\chi_0) = \arg \max_{\pi} \left\{ \pi^\mathsf{T} \chi_0 - \mathbb{I}_q(\pi) \right\} = \arg \max_{\pi \in \Pi(q)} \pi^\mathsf{T} \chi_0.$$
(2.8)

Since the set $\Pi(q)$ is polyhedral and $s_q(\chi_0)$ is finite, it follows that $\partial s_q(\chi_0)$ is nonempty. Moreover, the function $s_0(\cdot)$ is piecewise linear, and hence formula (2.7) follows from (2.8) by the chain rule of subdifferentiation. \Box

It follows that if the function $Q(\cdot, \xi)$ has a finite value in at least one point, then it is subdifferentiable at that point, and hence is proper. Its domain can be described in a more explicit way.

The positive hull of a matrix W is defined as

$$pos W := \{ \chi : \chi = Wy, \ y \ge 0 \}.$$
(2.9)

It is a convex polyhedral cone generated by the columns of W. Directly from the definition (2.4) we see that dom $s_q = \text{pos } W$. Therefore,

$$\operatorname{dom} Q(\cdot,\xi) = \{x : h - Tx \in \operatorname{pos} W\}.$$

Suppose that x is such that $\chi = h - Tx \in \text{pos } W$ and let us analyze formula (2.7). The recession cone of $\Pi(q)$ is equal to

$$\Pi_0 := \Pi(0) = \left\{ \pi : W^{\mathsf{T}} \pi \le 0 \right\}.$$
(2.10)

Then it follows from (2.6) that $s_q(\chi)$ is finite iff $\pi^T \chi \leq 0$ for every $\pi \in \Pi_0$, that is, iff χ is an element of the polar cone to Π_0 . This polar cone is nothing else but pos W, i.e.,

$$\Pi_0^* = \operatorname{pos} W. \tag{2.11}$$

If $\chi_0 \in \operatorname{int}(\operatorname{pos} W)$, then the set of maximizers in (2.6) must be bounded. Indeed, if it was unbounded, there would exist an element $\pi_0 \in \Pi_0$ such that $\pi_0^T \chi_0 = 0$. By perturbing χ_0 a little to some χ , we would be able to keep χ within pos W and get $\pi_0^T \chi > 0$, which is a contradiction, because pos W is the polar of Π_0 . Therefore the set of maximizers in (2.6) is the convex hull of the vertices v of $\Pi(q)$ for which $v^T \chi = s_q(\chi)$. Note that $\Pi(q)$ must have vertices in this case, because otherwise the polar to Π_0 would have no interior.

If χ_0 is a boundary point of pos W, then the set of maximizers in (2.6) is unbounded. Its recession cone is the intersection of the recession cone Π_0 of $\Pi(q)$ and of the subspace $\{\pi : \pi^T \chi_0 = 0\}$. This intersection is nonempty for boundary points χ_0 and is equal to the normal cone to pos W at χ_0 . Indeed, let π_0 be normal to pos W at χ_0 . Since both χ_0 and $-\chi_0$ are feasible directions at χ_0 , we must have $\pi_0^T \chi_0 = 0$. Next, for every $\chi \in \text{pos } W$ we have $\pi_0^T \chi = \pi_0^T (\chi - \chi_0) \leq 0$, so $\pi_0 \in \Pi_0$. The converse argument is similar.

2.1.2 The Expected Recourse Cost for Discrete Distributions

Let us consider now the expected value function

$$\phi(x) := \mathbb{E}[Q(x,\xi)]. \tag{2.12}$$

As before the expectation here is taken with respect to the probability distribution of the random vector ξ . Suppose that the distribution of ξ has finite support. That is, ξ has a finite number of realizations (called *scenarios*) $\xi_k = (q_k, h_k, T_k, W_k)$ with respective (positive) probabilities p_k , k = 1, ..., K, i.e., $\Xi = \{\xi_1, ..., \xi_K\}$. Then

$$\mathbb{E}[Q(x,\xi)] = \sum_{k=1}^{K} p_k Q(x,\xi_k).$$
(2.13)

For a given x, the expectation $\mathbb{E}[Q(x,\xi)]$ is equal to the optimal value of the linear programming problem

$$\operatorname{Min}_{y_1,\ldots,y_K} \sum_{k=1}^{K} p_k q_k^{\mathsf{T}} y_k$$
s.t. $T_k x + W_k y_k = h_k,$
 $y_k \ge 0, \quad k = 1, \ldots, K.$

$$(2.14)$$

If for at least one $k \in \{1, ..., K\}$ the system $T_k x + W_k y_k = h_k, y_k \ge 0$, has no solution, i.e., the corresponding second stage problem is infeasible, then problem (2.14) is infeasible, and hence its optimal value is $+\infty$. From that point of view the sum in the right hand side of (2.13) equals $+\infty$ if at least one of $Q(x, \xi_k) = +\infty$. That is, we assume here that $+\infty + (-\infty) = +\infty$.

The whole two stage problem is equivalent to the following large-scale linear programming problem:

$$\underset{x,y_{1},...,y_{K}}{\operatorname{Min}} c^{\mathsf{T}} x + \sum_{k=1}^{K} p_{k} q_{k}^{\mathsf{T}} y_{k}$$
s.t. $T_{k} x + W_{k} y_{k} = h_{k}, \quad k = 1, ..., K,$

$$Ax = b,$$

$$x \ge 0, \quad y_{k} \ge 0, \quad k = 1, ..., K.$$
(2.15)

Properties of the expected recourse cost follow directly from properties of parametric linear programming problems.

Proposition 2.3. Suppose that the probability distribution of ξ has finite support $\Xi = \{\xi_1, \ldots, \xi_K\}$ and that the expected recourse cost $\phi(\cdot)$ has a finite value in at least one point $\bar{x} \in \mathbb{R}^n$. Then the function $\phi(\cdot)$ is polyhedral, and for any $x_0 \in \text{dom } \phi$,

$$\partial \phi(x_0) = \sum_{k=1}^{K} p_k \partial Q(x_0, \xi_k).$$
(2.16)

Proof. Since $\phi(\bar{x})$ is finite, all values $Q(\bar{x}, \xi_k)$, k = 1, ..., K, are finite. Consequently, by Proposition 2.2, every function $Q(\cdot, \xi_k)$ is polyhedral. It is not difficult to see that a linear combination of polyhedral functions with positive weights is also polyhedral. Therefore, it follows that $\phi(\cdot)$ is polyhedral. We also have that dom $\phi = \bigcap_{k=1}^{K} \text{dom } Q_k$, where $Q_k(\cdot) :=$ $Q(\cdot, \xi_k)$, and for any $h \in \mathbb{R}^n$, the directional derivatives $Q'_k(x_0, h) > -\infty$ and

$$\phi'(x_0,h) = \sum_{k=1}^{K} p_k Q'_k(x_0,h).$$
(2.17)

Formula (2.16) then follows from (2.17) by duality arguments. Note that equation (2.16) is a particular case of the Moreau–Rockafellar Theorem (Theorem 7.4). Since the functions Q_k are polyhedral, there is no need here for an additional regularity condition for (2.16) to hold true.

The subdifferential $\partial Q(x_0, \xi_k)$ of the second stage optimal value function is described in Proposition 2.2. That is, if $Q(x_0, \xi_k)$ is finite, then

$$\partial Q(x_0,\xi_k) = -T_k^{\mathsf{T}} \arg \max \left\{ \pi^{\mathsf{T}}(h_k - T_k x_0) : W_k^{\mathsf{T}} \pi \le q_k \right\}.$$
 (2.18)

It follows that the expectation function ϕ is differentiable at x_0 iff for every $\xi = \xi_k$, $k = 1, \ldots, K$, the maximum in the right hand side of (2.18) is attained at a unique point, i.e., the corresponding second stage dual problem has a unique optimal solution.

Example 2.4 (Capacity Expansion) We have a directed graph with node set \mathcal{N} and arc set \mathcal{A} . With each arc $a \in \mathcal{A}$ we associate a decision variable x_a and call it the *capacity* of a. There is a cost c_a for each unit of capacity of arc a. The vector x constitutes the vector of first stage variables. They are restricted to satisfy the inequalities $x \ge x^{\min}$, where x^{\min} are the existing capacities.

At each node n of the graph we have a random demand ξ_n for shipments to n (if ξ_n is negative, its absolute value represents shipments from n and we have $\sum_{n \in \mathcal{N}} \xi_n = 0$). These shipments have to be sent through the network and they can be arbitrarily split into pieces taking different paths. We denote by y_a the amount of the shipment sent through arc a. There is a unit cost q_a for shipments on each arc a.

Our objective is to assign the arc capacities and to organize the shipments in such a way that the expected total cost, comprising the capacity cost and the shipping cost, is minimized. The condition is that the capacities have to be assigned *before* the actual demands ξ_n become known, while the shipments can be arranged *after* that.

Let us define the second stage problem. For each node n denote by $\mathcal{A}_+(n)$ and $\mathcal{A}_-(n)$ the sets of arcs entering and leaving node i. The second stage problem is the network flow problem

$$\operatorname{Min} \sum_{a \in \mathcal{A}} q_a y_a \tag{2.19}$$

s.t.
$$\sum_{a \in \mathcal{A}_{+}(n)} y_a - \sum_{a \in \mathcal{A}_{-}(n)} y_a = \xi_n, \quad n \in \mathcal{N},$$
 (2.20)

$$0 \le y_a \le x_a, \quad a \in \mathcal{A}. \tag{2.21}$$

This problem depends on the random demand vector ξ and on the arc capacities, x. Its optimal value is denoted by $Q(x,\xi)$.

Suppose that for a given $x = x_0$ the second stage problem (2.19)–(2.21) is feasible. Denote by μ_n , $n \in \mathcal{N}$, the optimal Lagrange multipliers (node potentials) associated with the node balance equations (2.20), and by π_a , $a \in \mathcal{A}$ the (nonnegative) Lagrange multipliers associated with the constraints (2.21). The dual problem has the form

$$\operatorname{Max} - \sum_{n \in \mathcal{N}} \xi_n \mu_n - \sum_{(i,j) \in \mathcal{A}} x_{ij} \pi_{ij}$$

s.t. $-\pi_{ij} + \mu_i - \mu_j \leq q_{ij}, \quad (i,j) \in \mathcal{A}$
 $\pi > 0.$

As $\sum_{n \in \mathcal{N}} \xi_n = 0$, the values of μ_n can be translated by a constant without any change in the objective function, and thus without any loss of generality we can assume that $\mu_{n_0} = 0$ for some fixed node n_0 . For each arc a = (i, j) the multiplier π_{ij} associated with the constraint (2.21) has the form

$$\pi_{ij} = \max\{0, \mu_i - \mu_j - q_{ij}\}.$$

Roughly, if the difference of node potentials $\mu_i - \mu_j$ is greater than q_{ij} , the arc is saturated and the capacity constraint $y_{ij} \le x_{ij}$ becomes relevant. The dual problem becomes equivalent to:

$$\operatorname{Max} - \sum_{n \in \mathcal{N}} \xi_n \mu_n - \sum_{(i,j) \in \mathcal{A}} x_{ij} \max\{0, \mu_i - \mu_j - q_{ij}\}.$$
 (2.22)

Let us denote by $\mathcal{M}(x_0,\xi)$ the set of optimal solutions of this problem satisfying the condition $\mu_{n_0} = 0$. Since $T^{\mathsf{T}} = [0 - I]$ in this case, formula (2.18) provides the description of the subdifferential of $Q(\cdot,\xi)$ at x_0 :

$$\partial Q(x_0,\xi) = -\left\{ \left(\max\{0,\mu_i - \mu_j - q_{ij}\} \right)_{(i,j)\in\mathcal{A}} : \mu \in \mathcal{M}(x_0,\xi) \right\}.$$

The first stage problem has the form

$$\underset{x \ge x^{\min}}{\min} \sum_{(i,j) \in \mathcal{A}} c_{ij} x_{ij} + \mathbb{E}[Q(x,\xi)].$$
(2.23)

If ξ has finitely many realizations ξ^k attained with probabilities p_k , k = 1, ..., K, the subdifferential of the overall objective can be calculated by (2.16):

$$\partial f(x_0) = c + \sum_{k=1}^{K} p_k \partial Q(x_0, \xi^k).$$

2.1.3 The Expected Recourse Cost for General Distributions

Let us discuss now the case of a general distribution of the random vector $\xi \in \mathbb{R}^d$. The recourse cost $Q(\cdot, \cdot)$ is the minimum value of the integrand which is a random lower semicontinuous function (see section 7.2.3). Therefore, it follows by Theorem 7.42 that $Q(\cdot, \cdot)$ is measurable with respect to the Borel sigma algebra of $\mathbb{R}^n \times \mathbb{R}^d$. Also for every ξ the function $Q(\cdot, \xi)$ is lower semicontinuous. It follows that $Q(x, \xi)$ is a random lower semicontinuous function. Recall that in order to ensure that the expectation $\phi(x)$ is well defined we have to verify two conditions:

- (i) $Q(x, \cdot)$ is measurable (with respect to the Borel sigma algebra of \mathbb{R}^d);
- (ii) either $\mathbb{E}[Q(x,\xi)_+]$ or $\mathbb{E}[(-Q(x,\xi))_+]$ are finite.

The function $Q(x, \cdot)$ is measurable, as the optimal value of a linear programming problem. We only need to verify condition (ii). We describe below some important particular situations where this condition is satisfied.

The two-stage problem (2.1)–(2.2) is said to have *fixed* recourse if the matrix W is fixed (not random). Moreover, we say that the recourse is *complete* if the system $Wy = \chi$ and $y \ge 0$ has a solution for every χ . In other words, the positive hull of W is equal to the corresponding vector space. By duality arguments, the fixed recourse is complete iff the feasible set $\Pi(q)$ of the dual problem (2.3) is bounded (in particular, it may be empty) for every q. Then its recession cone, $\Pi_0 = \Pi(0)$, must contain only the point 0, provided that $\Pi(q)$ is nonempty. Therefore, another equivalent condition for complete recourse is that $\pi = 0$ is the only solution of the system $W^T \pi \le 0$.

A particular class of problems with fixed and complete recourse are *simple recourse* problems, in which W = [I; -I], the matrix T and the vector q are deterministic, and the components of q are positive.

It is said that the recourse is *relatively complete* if for every x in the set

$$\mathcal{X} := \{ x : Ax = b, \ x \ge 0 \}$$

the feasible set of the second stage problem (2.2) is nonempty for almost every (a.e.) $\omega \in \Omega$. That is, the recourse is relatively complete if for every feasible first stage point x the inequality $Q(x,\xi) < +\infty$ holds true for a.e. $\xi \in \Xi$, or in other words $Q(x,\xi(\omega)) < +\infty$ with probability one (w.p.1). This definition is in accordance with the general principle that an event which happens with zero probability is irrelevant for the calculation of the corresponding expected value. For example, the capacity expansion problem of Example 2.4 is *not* a problem with relatively complete recourse, unless x^{\min} is so large that every demand $\xi \in \Xi$ can be shipped over the network with capacities x^{\min} .

The following condition is sufficient for relatively complete recourse.

For every
$$x \in \mathcal{X}$$
 the inequality $Q(x,\xi) < +\infty$ holds true for all $\xi \in \Xi$. (2.24)

In general, condition (2.24) is not necessary for relatively complete recourse. It becomes necessary and sufficient in the following two cases:

(i) the random vector ξ has a finite support; or

(ii) the recourse is fixed.

Indeed, sufficiency is clear. If ξ has a finite support, i.e., the set Ξ is finite, then the necessity is also clear. To show the necessity in the case of fixed recourse, suppose the recourse is relatively complete. This means that if $x \in \mathcal{X}$ then $Q(x,\xi) < +\infty$ for all ξ in Ξ , except possibly for a subset of Ξ of probability zero. We have that $Q(x,\xi) < +\infty$ iff $h - Tx \in \text{pos } W$. Let $\Xi_0(x) = \{(h,T,q) : h - Tx \in \text{pos } W\}$. The set pos W is convex and closed and thus $\Xi_0(x)$ is convex and closed as well. By assumption, $P[\Xi_0(x)] = 1$ for every $x \in \mathcal{X}$. Thus $\bigcap_{x \in \mathcal{X}} \Xi_0(x)$ is convex, closed, and has probability 1. The support of ξ must be its subset.

Example 2.5 Consider

$$Q(x,\xi) := \inf\{y : \xi y = x, \ y \ge 0\},\$$

with $x \in [0, 1]$ and ξ being a random variable whose probability density function is $p(z) := 2z, 0 \le z \le 1$. For all $\xi > 0$ and $x \in [0, 1], Q(x, \xi) = x/\xi$, and hence

$$\mathbb{E}[Q(x,\xi)] = \int_0^1 \left(\frac{x}{z}\right) 2z dz = 2x.$$

That is, the recourse here is relatively complete and the expectation of $Q(x,\xi)$ is finite. On the other hand, the support of $\xi(\omega)$ is the interval [0, 1], and for $\xi = 0$ and x > 0 the value of $Q(x,\xi)$ is $+\infty$, because the corresponding problem is infeasible. Of course, probability of the event " $\xi = 0$ " is zero, and from the mathematical point of view the expected value function $\mathbb{E}[Q(x,\xi)]$ is well defined and finite for all $x \in [0, 1]$. Note, however, that arbitrary small perturbation of the probability distribution of ξ may change that. Take, for example, some discretization of the distribution of ξ with the first discretization point t = 0. Then, no matter how small the assigned (positive) probability at t = 0 is, $Q(x,\xi) = +\infty$ with positive probability. Therefore, $\mathbb{E}[Q(x,\xi)] = +\infty$, for all x > 0. That is, the above problem is extremely unstable and is not well posed. As it was discussed above, such behavior cannot occur if the recourse is fixed.

Let us consider the support function $s_q(\cdot)$ of the set $\Pi(q)$. We want to find sufficient conditions for the existence of the expectation $\mathbb{E}[s_q(h-Tx)]$. By Hoffman's lemma (Theorem 7.12) there exists a constant κ , depending on W, such that if for some q_0 the set $\Pi(q_0)$ is nonempty, then for every q the following inclusion is satisfied

$$\Pi(q) \subset \Pi(q_0) + \kappa ||q - q_0||B,$$
(2.25)

where $B := \{\pi : \|\pi\| \le 1\}$ and $\|\cdot\|$ denotes the Euclidean norm. This inclusion allows to derive an upper bound for the support function $s_q(\cdot)$. Since the support function of the unit ball B is the norm $\|\cdot\|$, it follows from (2.25) that if the set $\Pi(q_0)$ is nonempty, then

$$s_q(\cdot) \le s_{q_0}(\cdot) + \kappa \|q - q_0\| \|\cdot\|.$$
(2.26)

Consider $q_0 = 0$. The support function $s_0(\cdot)$ of the cone Π_0 has the form

$$s_0(\chi) = \begin{cases} 0, & \text{if } \chi \in \text{pos } W, \\ +\infty, & \text{otherwise.} \end{cases}$$

Therefore, (2.26) with $q_0 = 0$ implies that if $\Pi(q)$ is nonempty, then $s_q(\chi) \le \kappa ||q|| ||\chi||$ for all $\chi \in \text{pos } W$, and $s_q(\chi) = +\infty$ for all $\chi \notin \text{pos } W$. Since $\Pi(q)$ is polyhedral, if it is nonempty then $s_q(\cdot)$ is piecewise linear on its domain, which coincides with pos W, and

$$|s_q(\chi_1) - s_q(\chi_2)| \le \kappa ||q|| \, ||\chi_1 - \chi_2||, \quad \forall \chi_1, \chi_2 \in \text{pos} \, W.$$
(2.27)

Proposition 2.6. Suppose that the recourse is fixed and

$$\mathbb{E}[\|q\| \|h\|] < +\infty \ and \ \mathbb{E}[\|q\| \|T\|] < +\infty.$$
(2.28)

Consider a point $x \in \mathbb{R}^n$. Then $\mathbb{E}[Q(x,\xi)_+]$ is finite if and only if the following condition holds w.p.1:

$$h - Tx \in \text{pos}\,W.\tag{2.29}$$

Proof. We have that $Q(x,\xi) < +\infty$ iff condition (2.29) holds. Therefore, if condition (2.29) does not hold w.p.1, then $Q(x,\xi) = +\infty$ with positive probability, and hence $\mathbb{E}[Q(x,\xi)_+] = +\infty$.

Conversely, suppose that condition (2.29) holds w.p.1. Then $Q(x,\xi) = s_q(h - Tx)$ with $s_q(\cdot)$ being the support function of the set $\Pi(q)$. By (2.26) there exists a constant κ such that for any χ ,

$$s_q(\chi) \le s_0(\chi) + \kappa \|q\| \, \|\chi\|.$$

Also for any $\chi \in \text{pos } W$ we have that $s_0(\chi) = 0$, and hence w.p.1,

$$s_q(h - Tx) \le \kappa \|q\| \|h - Tx\| \le \kappa \|q\| (\|h\| + \|T\| \|x\|).$$

It follows then by (2.28) that $\mathbb{E}[s_q(h-Tx)_+] < +\infty$. \Box

Remark 2. If q and (h, T) are independent and have finite first moments¹, then

$$\mathbb{E}\big[\|q\| \, \|h\|\big] = \mathbb{E}\big[\|q\|\big] \, \mathbb{E}\big[\|h\|\big] \text{ and } \mathbb{E}\big[\|q\| \, \|T\|\big] = \mathbb{E}\big[\|q\|\big] \, \mathbb{E}\big[\|T\|\big],$$

and hence condition (2.28) follows. Also condition (2.28) holds if (h, T, q) has finite second moments.

We obtain that, under the assumptions of Proposition 2.6, the expectation $\phi(x)$ is well defined and $\phi(x) < +\infty$ iff condition (2.29) holds w.p.1. If, moreover, the recourse is complete, then (2.29) holds for any x and ξ , and hence $\phi(\cdot)$ is well defined and is less than $+\infty$. Since the function $\phi(\cdot)$ is convex, we have that if $\phi(\cdot)$ is less than $+\infty$ on \mathbb{R}^n and is finite valued in at least one point, then $\phi(\cdot)$ is finite valued on the entire space \mathbb{R}^n .

Proposition 2.7. Suppose that: (i) the recourse is fixed, (ii) for a.e. q the set $\Pi(q)$ is nonempty, (iii) condition (2.28) holds.

¹We say that a random variable $Z = Z(\omega)$ has a finite *r*-th moment if $\mathbb{E}[|Z|^r] < +\infty$. It is said that $\xi(\omega)$ has finite *r*-th moments if each component of $\xi(\omega)$ has a finite *r*-th moment.

Then the expectation function $\phi(x)$ is well defined and $\phi(x) > -\infty$ for all $x \in \mathbb{R}^n$. Moreover, ϕ is convex, lower semicontinuous and Lipschitz continuous on dom ϕ , and its domain is a convex closed subset of \mathbb{R}^n given by

$$\operatorname{dom} \phi = \left\{ x \in \mathbb{R}^n : h - Tx \in \operatorname{pos} W \text{ w.p.1} \right\}.$$
(2.30)

Proof. By assumption (ii) the feasible set $\Pi(q)$ of the dual problem is nonempty w.p.1. Thus $Q(x,\xi)$ is equal to $s_q(h-Tx)$ w.p.1 for every x, where $s_q(\cdot)$ is the support function of the set $\Pi(q)$. Let $\pi(q)$ be the element of the set $\Pi(q)$ that is closest to 0. It exists, because $\Pi(q)$ is closed. By Hoffman's lemma (see (2.25)) there is a constant κ such that $\|\pi(q)\| \leq \kappa \|q\|$. Then for every x the following holds w.p.1:

$$s_q(h - Tx) \ge \pi(q)^{\mathsf{T}}(h - Tx) \ge -\kappa \|q\| (\|h\| + \|T\| \|x\|).$$
(2.31)

Owing to condition (2.28), it follows from (2.31) that $\phi(\cdot)$ is well defined and $\phi(x) > -\infty$ for all $x \in \mathbb{R}^n$. Moreover, since $s_q(\cdot)$ is lower semicontinuous, the lower semicontinuity of $\phi(\cdot)$ follows by Fatou's lemma. Convexity and closedness of dom ϕ follow from the convexity and lower semicontinuity of ϕ . We have by Proposition 2.6 that $\phi(x) < +\infty$ iff condition (2.29) holds w.p.1. This implies (2.30).

Consider two points $x, x' \in \text{dom } \phi$. Then by (2.30) the following holds true w.p.1:

$$h - Tx \in \text{pos } W$$
 and $h - Tx' \in \text{pos } W$. (2.32)

By (2.27), if the set $\Pi(q)$ is nonempty and (2.32) holds, then

$$|s_q(h - Tx) - s_q(h - Tx')| \le \kappa ||q|| \, ||T|| \, ||x - x'||.$$

It follows that

$$|\phi(x) - \phi(x')| \le \kappa \mathbb{E}[||q|| ||T||] ||x - x'||.$$

Together with condition (2.28) this implies the Lipschitz continuity of ϕ on its domain.

Denote by Σ the support² of the probability distribution (measure) of (h, T). Formula (2.30) means that a point x belongs to dom ϕ iff the probability of the event $\{h-Tx \in \text{pos } W\}$ is one. Note that the set $\{(h, T) : h - Tx \in \text{pos } W\}$ is convex and polyhedral, and hence is closed. Consequently x belongs to dom ϕ iff for every $(h, T) \in \Sigma$ it follows that $h - Tx \in \text{pos } W$. Therefore, we can write formula (2.30) in the form

$$\operatorname{dom} \phi = \bigcap_{(h,T)\in\Sigma} \left\{ x : h - Tx \in \operatorname{pos} W \right\}.$$
(2.33)

It should be noted that we assume that the recourse is fixed.

Let us observe that for any set \mathcal{H} of vectors h, the set $\bigcap_{h \in \mathcal{H}} (-h + \text{pos } W)$ is convex and polyhedral. Indeed, we have that pos W is a convex polyhedral cone, and hence can

²Recall that the support of the probability measure is the smallest closed set such that the probability (measure) of its complement is zero.

be represented as the intersection of a finite number of half spaces $A_i = \{\chi : a_i^\mathsf{T} \chi \leq 0\}$, $i = 1, \ldots, \ell$. Since the intersection of any number of half spaces of the form $b + A_i$, with $b \in B$, is still a half space of the same form (provided that this intersection is nonempty), we have the set $\bigcap_{h \in \mathcal{H}} (-h + \text{pos } W)$ can be represented as the intersection of half spaces of the form $b_i + A_i$, $i = 1, \ldots, \ell$, and hence is polyhedral. It follows that if T and W are fixed, then the set at the right hand side of (2.33) is convex and polyhedral.

Let us discuss now the differentiability properties of the expectation function $\phi(x)$. By Theorem 7.52 and formula (2.7) of Proposition 2.2 we have the following result.

Proposition 2.8. Suppose that the expectation function $\phi(\cdot)$ is proper and its domain has a nonempty interior. Then for any $x_0 \in \text{dom } \phi$,

$$\partial \phi(x_0) = -\mathbb{E}\left[T^{\mathsf{T}}\mathfrak{D}(x_0,\xi)\right] + \mathcal{N}_{\operatorname{dom}\phi}(x_0), \qquad (2.34)$$

where

$$\mathfrak{D}(x,\xi) := \arg \max_{\pi \in \Pi(q)} \pi^{\mathsf{T}}(h - Tx).$$

Moreover, ϕ is differentiable at x_0 if and only if x_0 belongs to the interior of dom ϕ and the set $\mathfrak{D}(x_0, \xi)$ is a singleton w.p.1.

As we discussed earlier, when the distribution of ξ has a finite support (i.e., there is a finite number of scenarios) the expectation function ϕ is piecewise linear on its domain and is differentiable everywhere only in the trivial case if it is linear.³ In the case of a continuous distribution of ξ the expectation operator smoothes the piecewise linear function $Q(\cdot, \xi)$ out.

Proposition 2.9. Suppose the assumptions of Proposition 2.7 are satisfied, and the conditional distribution of h, given (T, q), is absolutely continuous for almost all (T, q). Then ϕ is continuously differentiable on the interior of its domain.

Proof. By Proposition 2.7, the expectation function $\phi(\cdot)$ is well defined and greater than $-\infty$. Let x be a point in the interior of dom ϕ . For fixed T and q, consider the multifunction

$$\mathfrak{Z}(h) := \arg \max_{\pi \in \Pi(q)} \, \pi^{\mathsf{T}}(h - Tx).$$

Conditional on (T, q), the set $\mathfrak{D}(x, \xi)$ coincides with $\mathfrak{Z}(h)$. Since $x \in \operatorname{dom} \phi$, relation (2.30) implies that $h - Tx \in \operatorname{pos} W$ w.p.1. For every $h - Tx \in \operatorname{pos} W$, the set $\mathfrak{Z}(h)$ is nonempty and forms a face of the polyhedral set $\Pi(q)$. Moreover, there exists a set A given by the union of a finite number of linear subspaces of \mathbb{R}^m (where m is the dimension of h), which are perpendicular to the faces of sets $\Pi(q)$, such that if $h - Tx \in (\operatorname{pos} W) \setminus A$ then $\mathfrak{Z}(h)$ is a singleton. Since an affine subspace of \mathbb{R}^m has Lebesgue measure zero, it follows that the Lebesgue measure of A is zero. As the conditional distribution of h, given (T, q), is absolutely continuous, the probability that $\mathfrak{Z}(h)$ is not a singleton is zero. By integrating this probability over the marginal distribution of (T, q), we obtain that probability of the

³By linear we mean here that it is of the form $a^{T}x + b$. It is more accurate to call such a function affine.

event " $\mathfrak{D}(x,\xi)$ is not a singleton" is zero. By Proposition 2.8, this implies the differentiability of $\phi(\cdot)$. Since $\phi(\cdot)$ is convex, it follows that for every $x \in \operatorname{int}(\operatorname{dom} \phi)$ the gradient $\nabla \phi(x)$ coincides with the (unique) subgradient of ϕ at x, and that $\nabla \phi(\cdot)$ is continuous at x. \Box

Of course, if h and (T,q) are independent, then the conditional distribution of h given (T,q) is the same as the unconditional (marginal) distribution of h. Therefore, if h and (T,q) are independent, then it suffices to assume in the above proposition that the (marginal) distribution of h is absolutely continuous.

2.1.4 Optimality Conditions

We can now formulate optimality conditions and duality relations for linear two stage problems. Let us start from the problem with discrete distributions of the random data in (2.1)–(2.2). The problem takes on the form:

$$\operatorname{Min}_{x} c^{\mathsf{T}} x + \sum_{k=1}^{K} p_{k} Q(x, \xi_{k})$$
s.t. $Ax = b, \ x > 0,$

$$(2.35)$$

where $Q(x,\xi)$ is the optimal value of the second stage problem, given by (2.2).

Suppose the expectation function $\phi(\cdot) := \mathbb{E}[Q(\cdot,\xi)]$ has a finite value in at least one point $\bar{x} \in \mathbb{R}^n$. It follows from Propositions 2.2 and 2.3 that for every $x_0 \in \operatorname{dom} \phi$,

$$\partial \phi(x_0) = -\sum_{k=1}^{K} p_k T_k^{\mathsf{T}} \mathfrak{D}(x_0, \xi_k), \qquad (2.36)$$

where

$$\mathfrak{D}(x_0,\xi_k) := \arg\max\left\{\pi^{\mathsf{T}}(h_k - T_k x_0) : W_k^{\mathsf{T}} \pi \le q_k\right\}.$$

As before, we denote $\mathcal{X} := \{x : Ax = b, x \ge 0\}.$

Theorem 2.10. Let \bar{x} be a feasible solution of problem (2.1)–(2.2), i.e., $\bar{x} \in \mathcal{X}$ and $\phi(\bar{x})$ is finite. Then \bar{x} is an optimal solution of problem (2.1)–(2.2) iff there exist $\pi_k \in \mathfrak{D}(\bar{x}, \xi_k)$, $k = 1, \ldots, K$, and $\mu \in \mathbb{R}^m$ such that

$$\sum_{k=1}^{K} p_k T_k^{\mathsf{T}} \pi_k + A^{\mathsf{T}} \mu \le c,$$

$$\bar{x}^{\mathsf{T}} \left(c - \sum_{k=1}^{K} p_k T_k^{\mathsf{T}} \pi_k - A^{\mathsf{T}} \mu \right) = 0.$$
(2.37)

Proof. Necessary and sufficient optimality conditions for minimizing $c^{\mathsf{T}}x + \phi(x)$ over $x \in \mathcal{X}$ can be written as

$$0 \in c + \partial \phi(\bar{x}) + \mathcal{N}_{\mathcal{X}}(\bar{x}), \tag{2.38}$$

where $\mathcal{N}_{\mathcal{X}}(\bar{x})$ is the normal cone to the feasible set \mathcal{X} . Note that condition (2.38) implies that the sets $\mathcal{N}_{\mathcal{X}}(\bar{x})$ and $\partial \phi(\bar{x})$ are nonempty and hence $\bar{x} \in \mathcal{X}$ and $\phi(\bar{x})$ is finite. Note also that there is no need here for additional regularity conditions since $\phi(\cdot)$ and \mathcal{X} are convex and polyhedral. Using the characterization of the subgradients of $\phi(\cdot)$, given in (2.36), we conclude that (2.38) is equivalent to existence of $\pi_k \in \mathfrak{D}(\bar{x}, \xi_k)$ such that

$$0 \in c - \sum_{k=1}^{K} p_k T_k^{\mathsf{T}} \pi_k + \mathcal{N}_{\mathcal{X}}(\bar{x}).$$

Observe that

$$\mathcal{N}_{\mathcal{X}}(\bar{x}) = \{ A^{\mathsf{T}} \mu - h : h \ge 0, \ h^{\mathsf{T}} \bar{x} = 0 \}.$$
(2.39)

The last two relations are equivalent to conditions (2.37).

Conditions (2.37) can also be obtained directly from the optimality conditions for the large scale linear programming formulation

$$\underset{x,y_1,\ldots,y_K}{\operatorname{Min}} c^{\mathsf{T}} x + \sum_{k=1}^{K} p_k q_k^{\mathsf{T}} y_k$$
s.t. $T_k x + W_k y_k = h_k, \quad k = 1, \ldots, K,$

$$Ax = b,$$

$$x \ge 0,$$

$$y_k \ge 0, \quad k = 1, \ldots, K.$$

$$(2.40)$$

By minimizing, with respect to $x \ge 0$ and $y_k \ge 0, k = 1, ..., K$, the Lagrangian

$$c^{\mathsf{T}}x + \sum_{k=1}^{K} p_k q_k^{\mathsf{T}} y_k - \mu^{\mathsf{T}} (Ax - b) - \sum_{k=1}^{K} p_k \pi_k^{\mathsf{T}} (T_k x + W_k y_k - h_k) = \left(c - A^{\mathsf{T}} \mu - \sum_{k=1}^{K} p_k T_k^{\mathsf{T}} \pi_k\right)^{\mathsf{T}} x + \sum_{k=1}^{K} p_k \left(q_k - W_k^{\mathsf{T}} \pi_k\right)^{\mathsf{T}} y_k + b^{\mathsf{T}} \mu + \sum_{k=1}^{K} p_k h_k^{\mathsf{T}} \pi_k,$$

we obtain the following dual of the linear programming problem (2.40):

$$\max_{\mu,\pi_{1},...,\pi_{K}} b^{\mathsf{T}}\mu + \sum_{k=1}^{K} p_{k}h_{k}^{\mathsf{T}}\pi_{k}$$

s.t. $c - A^{\mathsf{T}}\mu - \sum_{k=1}^{K} p_{k}T_{k}^{\mathsf{T}}\pi_{k} \ge 0,$
 $q_{k} - W_{k}^{\mathsf{T}}\pi_{k} \ge 0, \ k = 1,...,K.$

Therefore optimality conditions of Theorem 2.10 can be written in the following equivalent

form

$$\sum_{k=1}^{K} p_k T_k^{\mathsf{T}} \pi_k + A^{\mathsf{T}} \mu \le c,$$

$$\bar{x}^{\mathsf{T}} \left(c - \sum_{k=1}^{K} p_k T_k^{\mathsf{T}} \pi_k - A^{\mathsf{T}} \mu \right) = 0,$$

$$q_k - W_k^{\mathsf{T}} \pi_k \ge 0, \ k = 1, \dots, K,$$

$$\bar{y}_k^{\mathsf{T}} \left(q_k - W_k^{\mathsf{T}} \pi_k \right) = 0, \ k = 1, \dots, K.$$

× 1

The last two of the above conditions correspond to feasibility and optimality of multipliers π_k as solutions of the dual problems.

If we deal with general distributions of problem's data, additional conditions are needed, to ensure the subdifferentiability of the expected recourse cost, and the existence of Lagrange multipliers.

Theorem 2.11. Let \bar{x} be a feasible solution of problem (2.1)–(2.2). Suppose that the expected recourse cost function $\phi(\cdot)$ is proper, $\operatorname{int}(\operatorname{dom} \phi) \cap \mathcal{X}$ is nonempty and $\mathcal{N}_{\operatorname{dom} \phi}(\bar{x}) \subset \mathcal{N}_{\mathcal{X}}(\bar{x})$. Then \bar{x} is an optimal solution of problem (2.1)–(2.2) iff there exist a measurable function $\pi(\omega) \in \mathfrak{D}(\bar{x}, \xi(\omega)), \omega \in \Omega$, and a vector $\mu \in \mathbb{R}^m$ such that

$$\mathbb{E}[T^{\mathsf{T}}\pi] + A^{\mathsf{T}}\mu \le c,$$

$$\bar{x}^{\mathsf{T}}(c - \mathbb{E}[T^{\mathsf{T}}\pi] - A^{\mathsf{T}}\mu) = 0.$$

Proof. Since $int(dom \phi) \cap \mathcal{X}$ is nonempty, we have by the Moreau-Rockafellar Theorem that

$$\partial \left(c^{\mathsf{T}} \bar{x} + \phi(\bar{x}) + \mathbb{I}_{\mathcal{X}}(\bar{x}) \right) = c + \partial \phi(\bar{x}) + \partial \mathbb{I}_{\mathcal{X}}(\bar{x}).$$

Also $\partial \mathbb{I}_{\mathcal{X}}(\bar{x}) = \mathcal{N}_{\mathcal{X}}(\bar{x})$. Therefore we have here that (2.38) are necessary and sufficient optimality conditions for minimizing $c^{\mathsf{T}}x + \phi(x)$ over $x \in \mathcal{X}$. Using the characterization of the subdifferential of $\phi(\cdot)$ given in (2.34), we conclude that (2.38) is equivalent to existence of a measurable function $\pi(\omega) \in \mathfrak{D}(\bar{x}, \xi(\omega))$, such that

$$0 \in c - \mathbb{E}[T^{\mathsf{T}}\pi] + \mathcal{N}_{\operatorname{dom}\phi}(\bar{x}) + \mathcal{N}_{\mathcal{X}}(\bar{x}).$$
(2.41)

Moreover, because of the condition $\mathcal{N}_{\operatorname{dom}\phi}(\bar{x}) \subset \mathcal{N}_{\mathcal{X}}(\bar{x})$, the term $\mathcal{N}_{\operatorname{dom}\phi}(\bar{x})$ can be omitted. The proof can be completed now by using (2.41) together with formula (2.39) for the normal cone $\mathcal{N}_{\mathcal{X}}(\bar{x})$. \Box

The additional technical condition $\mathcal{N}_{\operatorname{dom}\phi}(\bar{x}) \subset \mathcal{N}_{\mathcal{X}}(\bar{x})$ was needed in the above derivations in order to eliminate the term $\mathcal{N}_{\operatorname{dom}\phi}(\bar{x})$ in (2.41). In particular, this condition holds if $\bar{x} \in \operatorname{int}(\operatorname{dom}\phi)$, in which case $\mathcal{N}_{\operatorname{dom}\phi}(\bar{x}) = \{0\}$, or in case of relatively complete recourse, i.e., when $\mathcal{X} \subset \operatorname{dom}\phi$. If the condition of relatively complete recourse is not satisfied, we may need to take into account the normal cone to the domain of $\phi(\cdot)$. In general, this requires application of techniques of functional analysis, which are beyond

the scope of this book. However, in the special case of a deterministic matrix T we can carry out the analysis directly.

Theorem 2.12. Let \bar{x} be a feasible solution of problem (2.1)–(2.2). Suppose that the assumptions of Proposition 2.7 are satisfied, $\operatorname{int}(\operatorname{dom} \phi) \cap \mathcal{X}$ is nonempty and the matrix T is deterministic. Then \bar{x} is an optimal solution of problem (2.1)–(2.2) iff there exist a measurable function $\pi(\omega) \in \mathfrak{D}(\bar{x}, \xi(\omega)), \omega \in \Omega$, and a vector $\mu \in \mathbb{R}^m$ such that

$$T^{\mathsf{T}}\mathbb{E}[\pi] + A^{\mathsf{T}}\mu \le c,$$

$$\bar{x}^{\mathsf{T}}(c - T^{\mathsf{T}}\mathbb{E}[\pi] - A^{\mathsf{T}}\mu) = 0.$$

Proof. Since T is deterministic, we have that $\mathbb{E}[T^{\mathsf{T}}\pi] = T^{\mathsf{T}}\mathbb{E}[\pi]$, and hence the optimality conditions (2.41) can be written as

$$0 \in c - T^{\mathsf{T}} \mathbb{E}[\pi] + \mathcal{N}_{\operatorname{dom} \phi}(\bar{x}) + \mathcal{N}_{\mathcal{X}}(\bar{x}).$$

Now we need to calculate the cone $\mathcal{N}_{\operatorname{dom}\phi}(\bar{x})$. Recall that under the assumptions of Proposition 2.7 (in particular, that the recourse is fixed and $\Pi(q)$ is nonempty w.p.1), we have that $\phi(\cdot) > -\infty$ and formula (2.30) holds true. We have here that only q and h are random while both matrices W and T are deterministic, and (2.30) simplifies to

dom
$$\phi = \left\{ x : -Tx \in \bigcap_{h \in \Sigma} \left(-h + \operatorname{pos} W \right) \right\},\$$

where Σ is the support of the distribution of the random vector h. The tangent cone to dom ϕ at \bar{x} has the form

$$\mathcal{T}_{\operatorname{dom}\phi}(\bar{x}) = \left\{ d: -Td \in \bigcap_{h \in \Sigma} \left(\operatorname{pos} W + \operatorname{lin}(-h + T\bar{x}) \right) \right\}$$
$$= \left\{ d: -Td \in \operatorname{pos} W + \bigcap_{h \in \Sigma} \operatorname{lin}(-h + T\bar{x}) \right\}.$$

Defining the linear subspace

$$L := \bigcap_{h \in \varSigma} \ln(-h + T\bar{x}),$$

we can write the tangent cone as

$$\mathcal{T}_{\operatorname{dom}\phi}(\bar{x}) = \{d : -Td \in \operatorname{pos} W + L\}.$$

Therefore the normal cone equals

$$\mathcal{N}_{\operatorname{dom}\phi}(\bar{x}) = \left\{ -T^{\mathsf{T}}v : v \in (\operatorname{pos} W + L)^* \right\} = -T^{\mathsf{T}} \left[(\operatorname{pos} W)^* \cap L^{\perp} \right].$$

Here we used the fact that pos W is polyhedral and no interior condition is needed for calculating $(pos W + L)^*$. Recalling equation (2.11) we conclude that

$$\mathcal{N}_{\operatorname{dom}\phi}(\bar{x}) = -T^{\mathsf{T}}(\Pi_0 \cap L^{\perp}).$$

Observe that if $\nu \in \Pi_0 \cap L^{\perp}$ then ν is an element of the recession cone of the set $\mathfrak{D}(\bar{x}, \xi)$, for all $\xi \in \Xi$. Thus $\pi(\omega) + \nu$ is also an element of the set $\mathfrak{D}(x, \xi(\omega))$, for almost all $\omega \in \Omega$. Consequently

$$-T^{\mathsf{T}}\mathbb{E}\big[\mathfrak{D}(\bar{x},\xi)\big] + \mathcal{N}_{\operatorname{dom}\phi}(\bar{x}) = -T^{\mathsf{T}}\mathbb{E}\big[\mathfrak{D}(\bar{x},\xi)\big] - T^{\mathsf{T}}\big(\Pi_0 \cap L^{\perp}\big)$$
$$= -T^{\mathsf{T}}\mathbb{E}\big[\mathfrak{D}(\bar{x},\xi)\big],$$

and the result follows. \Box

Example 2.13 (Capacity Expansion (continued)) Let us return to Example 2.13 and suppose the support Ξ of the random demand vector ξ is compact. Only the right hand side ξ in the second stage problem (2.19)–(2.21) is random and for a sufficiently large x the second stage problem is feasible for all $\xi \in \Xi$. Thus conditions of Theorem 2.11 are satisfied. It follows from Theorem 2.11 that \bar{x} is an optimal solution of problem (2.23) iff there exist measurable functions $\mu_n(\xi)$, $n \in \mathcal{N}$, such that for all $\xi \in \Xi$ we have $\mu(\xi) \in \mathcal{M}(\bar{x}, \xi)$, and for all $(i, j) \in \mathcal{A}$ the following conditions are satisfied:

$$c_{ij} \ge \int_{\Xi} \max\{0, \mu_i(\xi) - \mu_j(\xi) - q_{ij}\} P(d\xi),$$
(2.42)

$$\left(\bar{x}_{ij} - x_{ij}^{\min}\right) \left(c_{ij} - \int_{\Xi} \max\{0, \mu_i(\xi) - \mu_j(\xi) - q_{ij}\} P(d\xi)\right) = 0.$$
(2.43)

In particular, for every $(i, j) \in \mathcal{A}$ such that $\bar{x}_{ij} > x_{ij}^{\min}$ we have equality in equation (2.42). Each function $\mu_n(\xi)$ can be interpreted as a random potential of node $n \in \mathcal{N}$.

2.2 Polyhedral Two-Stage Problems

2.2.1 General Properties

Let us consider a slightly more general formulation of a two-stage stochastic programming problem:

$$\operatorname{Min} f_1(x) + \mathbb{E}[Q(x,\omega)], \qquad (2.44)$$

where $Q(x, \omega)$ is the optimal value of the second stage problem

$$\begin{aligned} &\underset{y}{\min} \ f_2(y,\omega) \\ &\text{s.t.} \ T(\omega)x + W(\omega)y = h(\omega). \end{aligned}$$
(2.45)

We assume in this section that the above two-stage problem is *polyhedral*. That is, the following holds.

• The function $f_1(\cdot)$ is *polyhedral* (compare with Definition 7.1). This means that there exist vectors c_j and scalars α_j , $j = 1, ..., J_1$, vectors a_k and scalars b_k , k =

 $1, \ldots, K_1$, such that $f_1(x)$ can be represented as follows:

$$f_1(x) = \begin{cases} \max_{1 \le j \le J_1} \alpha_j + c_j^\mathsf{T} x, & \text{if } a_k^\mathsf{T} x \le b_k, \quad k = 1, \dots, K_1, \\ +\infty, & \text{otherwise,} \end{cases}$$

and its domain dom $f_1 = \{x : a_k^T x \leq b_k, k = 1, ..., K_1\}$ is nonempty. (Note that any polyhedral function is convex and lower semicontinuous.)

• The function f_2 is *random polyhedral*. That is, there exist random vectors $q_j = q_j(\omega)$ and random scalars $\gamma_j = \gamma_j(\omega)$, $j = 1, \ldots, J_2$, random vectors $d_k = d_k(\omega)$ and random scalars $r_k = r_k(\omega)$, $k = 1, \ldots, K_2$, such that $f_2(y, \omega)$ can be represented as follows:

$$f_2(y,\omega) = \begin{cases} \max_{1 \le j \le J_2} \gamma_j(\omega) + q_j(\omega)^\mathsf{T} y, & \text{if } d_k(\omega)^\mathsf{T} y \le r_k(\omega), \quad k = 1, \dots, K_2, \\ +\infty, & \text{otherwise,} \end{cases}$$

and for a.e. ω the domain of $f_2(\cdot, \omega)$ is nonempty.

Note that (linear) constraints of the second stage problem which are independent of x, as for example $y \ge 0$, can be absorbed into the objective function $f_2(y, \omega)$. Clearly, the linear two-stage model (2.1)–(2.2) is a special case of a polyhedral two-stage problem. The converse is also true, that is, every polyhedral two-stage model can be re-formulated as a linear two-stage model. For example, the second stage problem (2.45) can be written as follows:

$$\begin{array}{ll}
\underset{y,v}{\operatorname{Min}} & v \\
\text{s.t.} & T(\omega)x + W(\omega)y = h(\omega), \\
& \gamma_j(\omega) + q_j(\omega)^{\mathsf{T}}y \leq v, \quad j = 1, \dots, J_2, \\
& d_k(\omega)^{\mathsf{T}}y \leq r_k(\omega), \quad k = 1, \dots, K_2.
\end{array}$$

Here both v and y play the role of the second stage variables, and the data (q, T, W, h) in (2.2) have to be re-defined in an appropriate way. In order to avoid all these manipulations and unnecessary notational complications that come together with such a conversion, we shall address polyhedral problems in a more abstract way. This will also help us to deal with multistage problems and general convex problems.

Consider the Lagrangian of the second stage problem (2.45):

$$L(y,\pi;x,\omega) := f_2(y,\omega) + \pi^{\mathsf{T}} (h(\omega) - T(\omega)x - W(\omega)y)$$

We have

$$\begin{split} \inf_{y} L(y,\pi;x,\omega) &= \pi^{\mathsf{T}} \big(h(\omega) - T(\omega)x \big) + \inf_{y} \big[f_2(y,\omega) - \pi^{\mathsf{T}} W(\omega)y \big] \\ &= \pi^{\mathsf{T}} \big(h(\omega) - T(\omega)x \big) - f_2^*(W(\omega)^{\mathsf{T}} \pi, \omega), \end{split}$$

where $f_2^*(\cdot, \omega)$ is the conjugate⁴ of $f_2(\cdot, \omega)$. We obtain that the dual of problem (2.45) can be written as follows

$$\operatorname{Max}\left[\pi^{\mathsf{T}}(h(\omega) - T(\omega)x) - f_{2}^{*}(W(\omega)^{\mathsf{T}}\pi, \omega)\right].$$
(2.46)

By the duality theory of linear programming, if, for some (x, ω) , the optimal value $Q(x, \omega)$ of problem (2.45) is less than $+\infty$ (i.e., problem (2.45) is feasible), then it is equal to the optimal value of the dual problem (2.46).

Let us denote, as before, by $\mathfrak{D}(x, \omega)$ the set of optimal solutions of the dual problem (2.46). We then have an analogue of Proposition 2.2.

Proposition 2.14. Let $\omega \in \Omega$ be given and suppose that $Q(\cdot, \omega)$ is finite in at least one point \bar{x} . Then the function $Q(\cdot, \omega)$ is polyhedral (and hence convex). Moreover, $Q(\cdot, \omega)$ is subdifferentiable at every x at which the value $Q(x, \omega)$ is finite, and

$$\partial Q(x,\omega) = -T(\omega)^{\mathsf{T}}\mathfrak{D}(x,\omega).$$
(2.47)

Proof. Let us define the function $\psi(\pi) := f_2^*(W^T\pi)$ (for simplicity we suppress the argument ω). We have that if $Q(x,\omega)$ is finite, then it is equal to the optimal value of problem (2.46), and hence $Q(x,\omega) = \psi^*(h - Tx)$. Therefore $Q(\cdot,\omega)$ is a polyhedral function. Moreover, it follows by the Fenchel-Moreau Theorem that

$$\partial \psi^*(h - Tx) = \mathfrak{D}(x, \omega),$$

and the chain rule for subdifferentiation yields formula (2.47). Note that we do not need here additional regularity conditions because of the polyhedricity of the considered case. \Box

If $Q(x,\omega)$ is finite, then the set $\mathfrak{D}(x,\omega)$ of optimal solutions of problem (2.46) is a nonempty convex closed polyhedron. If, moreover, $\mathfrak{D}(x,\omega)$ is bounded, then it is the convex hull of its finitely many vertices (extreme points), and $Q(\cdot,\omega)$ is finite in a neighborhood of x. If $\mathfrak{D}(x,\omega)$ is unbounded, then its recession cone (which is polyhedral) is the normal cone to the domain of $Q(\cdot,\omega)$ at the point x.

2.2.2 Expected Recourse Cost

Let us consider the expected value function $\phi(x) := \mathbb{E}[Q(x, \omega)]$. Suppose that the probability measure P has a finite support, i.e., there exists a finite number of scenarios ω_k with respective (positive) probabilities $p_k, k = 1, ..., K$. Then

$$\mathbb{E}[Q(x,\omega)] = \sum_{k=1}^{K} p_k Q(x,\omega_k).$$

⁴Note that since $f_2(\cdot, \omega)$ is polyhedral, so is $f_2^*(\cdot, \omega)$.

For a given x, the expectation $\mathbb{E}[Q(x,\omega)]$ is equal to the optimal value of the problem

$$\underset{y_{1},...,y_{K}}{\min} \sum_{k=1}^{K} p_{k} f_{2}(y_{k},\omega_{k})
s.t. T_{k}x + W_{k}y_{k} = h_{k}, \ k = 1,...,K,$$
(2.48)

where $(h_k, T_k, W_k) := (h(\omega_k), T(\omega_k), W(\omega_k))$. Similarly to the linear case, if for at least one $k \in \{1, ..., K\}$ the set

$$\operatorname{dom} f_2(\cdot, \omega_k) \cap \{y : T_k x + W_k y = h_k\}$$

is empty, i.e., the corresponding second stage problem is infeasible, then problem (2.48) is infeasible, and hence its optimal value is $+\infty$.

Proposition 2.15. Suppose that the probability measure P has a finite support and that the expectation function $\phi(\cdot) := \mathbb{E}[Q(\cdot, \omega)]$ has a finite value in at least one point $x \in \mathbb{R}^n$. Then the function $\phi(\cdot)$ is polyhedral, and for any $x_0 \in \text{dom } \phi$,

$$\partial \phi(x_0) = \sum_{k=1}^{K} p_k \partial Q(x_0, \omega_k).$$
(2.49)

The proof is identical to the proof of Proposition 2.3. Since the functions $Q(\cdot, \omega_k)$ are polyhedral, formula (2.49) follows by the Moreau–Rockafellar Theorem.

The subdifferential $\partial Q(x_0, \omega_k)$ of the second stage optimal value function is described in Proposition 2.14. That is, if $Q(x_0, \omega_k)$ is finite, then

$$\partial Q(x_0,\omega_k) = -T_k^{\mathsf{T}} \arg \max \left\{ \pi^{\mathsf{T}} \left(h_k - T_k x_0 \right) - f_2^* (W_k^{\mathsf{T}} \pi, \omega_k) \right\}.$$
(2.50)

It follows that the expectation function ϕ is differentiable at x_0 iff for every ω_k , $k = 1, \ldots, K$, the maximum at the right hand side of (2.50) is attained at a unique point, i.e., the corresponding second stage dual problem has a unique optimal solution.

Let us now consider the case of a general probability distribution P. We need to ensure that the expectation function $\phi(x) := \mathbb{E}[Q(x, \omega)]$ is well defined. General conditions are complicated, so we resort again to the case of fixed recourse.

We say that the two-stage polyhedral problem has *fixed recourse* if the matrix W and the set⁵ $\mathcal{Y} := \text{dom } f_2(\cdot, \omega)$ are fixed, i.e., do not depend on ω . In that case,

$$f_2(y,\omega) = \begin{cases} \max_{1 \le j \le J_2} \gamma_j(\omega) + q_j(\omega)^{\mathsf{T}} y, & \text{if } y \in \mathcal{Y}, \\ +\infty, & \text{otherwise.} \end{cases}$$

Denote $W(\mathcal{Y}) := \{Wy : y \in \mathcal{Y}\}$. Let x be such that

$$h(\omega) - T(\omega)x \in W(\mathcal{Y})$$
 w.p.1. (2.51)

⁵Note that since it is assumed that $f_2(\cdot, \omega)$ is polyhedral, it follows that the set \mathcal{Y} is nonempty and polyhedral.

This means that for a.e. ω the system

$$y \in \mathcal{Y}, \quad Wy = h(\omega) - T(\omega)x$$
 (2.52)

has a solution. Let for some $\omega_0 \in \Omega$, y_0 be a solution of the above system, i.e., $y_0 \in \mathcal{Y}$ and $h(\omega_0) - T(\omega_0)x = Wy_0$. Since system (2.52) is defined by linear constraints, we have by Hoffman's lemma that there exists a constant κ such that for almost all ω we can find a solution $\bar{y}(\omega)$ of the system (2.52) with

$$\|\bar{y}(\omega) - y_0\| \le \kappa \|(h(\omega) - T(\omega)x) - (h(\omega_0) - T(\omega_0)x)\|.$$

Therefore the optimal value of the second stage problem can be bounded from above as follows:

$$Q(x,\omega) \leq \max_{1 \leq j \leq J_2} \left\{ \gamma_j(\omega) + q_j(\omega)^{\mathsf{T}} \bar{y}(\omega) \right\}$$

$$\leq Q(x,\omega_0) + \sum_{j=1}^{J_2} |\gamma_j(\omega) - \gamma_j(\omega_0)|$$

$$+ \kappa \sum_{j=1}^{J_2} ||q_j(\omega)|| (||h(\omega) - h(\omega_0)|| + ||x|| ||T(\omega) - T(\omega_0)||).$$
(2.53)

Proposition 2.16. Suppose that the recourse is fixed and

$$\mathbb{E}|\gamma_j| < +\infty, \ \mathbb{E}[\|q_j\| \|h\|] < +\infty \ and \ \mathbb{E}[\|q_j\| \|T\|] < +\infty, \ j = 1, \dots, J_2.$$
(2.54)

Consider a point $x \in \mathbb{R}^n$. Then $\mathbb{E}[Q(x, \omega)_+]$ is finite if and only if condition (2.51) holds.

Proof. The proof uses (2.53), similarly to the proof of Proposition 2.6.

Let us now formulate conditions under which the expected recourse cost is bounded from below. Let C be the recession cone of \mathcal{Y} , and C^* be its polar. Consider the conjugate function $f_2^*(\cdot, \omega)$. It can be verified that

$$\operatorname{dom} f_2^*(\cdot, \omega) = \operatorname{conv} \{ q_j(\omega), \ j = 1, \dots, J_2 \} + \mathcal{C}^*.$$
(2.55)

Indeed, by the definition of the function $f_2(\cdot, \omega)$ and its conjugate, we have that $f_2^*(z, \omega)$ is equal to the optimal value of the

$$\underset{y,v}{\operatorname{Max}} v$$

s.t. $z^{\mathsf{T}}y - \gamma_j(\omega) - q_j(\omega)^{\mathsf{T}}y \ge v, \ j = 1, \dots, J_2, \ y \in \mathcal{Y}.$

Since it is assumed that the set \mathcal{Y} is nonempty, the above problem is feasible, and since \mathcal{Y} is polyhedral, it is linear. Therefore its optimal value is equal to the optimal value of its dual. In particular, its optimal value is less than $+\infty$ iff the dual problem is feasible. Now the dual problem is feasible iff there exist $\pi_j \ge 0$, $j = 1, \ldots, J_2$, such that $\sum_{j=1}^{J_2} \pi_j = 1$ and

$$\sup_{y\in\mathcal{Y}}y^{\mathsf{T}}\left(z-\sum_{j=1}^{J_2}\pi_jq_j(\omega)\right)<+\infty.$$

The last condition holds iff $z - \sum_{j=1}^{J_2} \pi_j q_j(\omega) \in \mathcal{C}^*$, which completes the argument. Let us define the set

$$\Pi(\omega) := \left\{ \pi : W^{\mathsf{T}} \pi \in \operatorname{conv} \left\{ q_j(\omega), \ j = 1, \dots, J_2 \right\} + \mathcal{C}^* \right\}.$$

We may remark that in the case of a linear two stage problem the above set coincides with the one defined in (2.5).

Proposition 2.17. Suppose that: (i) the recourse is fixed, (ii) the set $\Pi(\omega)$ is nonempty w.p.1, (iii) condition (2.54) holds.

Then the expectation function $\phi(x)$ is well defined and $\phi(x) > -\infty$ for all $x \in \mathbb{R}^n$. Moreover, ϕ is convex, lower semicontinuous and Lipschitz continuous on dom ϕ , its domain dom ϕ is a convex closed subset of \mathbb{R}^n and

dom
$$\phi = \{x \in \mathbb{R}^n : h - Tx \in W(\mathcal{Y}) \text{ w.p.1}\}.$$
 (2.56)

Furthermore, for any $x_0 \in \operatorname{dom} \phi$ *,*

$$\partial \phi(x_0) = -\mathbb{E}\left[T^{\mathsf{T}}\mathfrak{D}(x_0,\omega)\right] + \mathcal{N}_{\mathrm{dom}\,\phi}(x_0),\tag{2.57}$$

Proof. Note that the dual problem (2.46) is feasible iff $W^{\mathsf{T}}\pi \in \text{dom} f_2^*(\cdot, \omega)$. By formula (2.55) assumption (ii) means that problem (2.46) is feasible, and hence $Q(x, \omega)$ is equal to the optimal value of (2.46), for a.e. ω . The remainder of the proof is similar to the linear case (Proposition 2.7 and Proposition 2.8). \Box

2.2.3 Optimality Conditions

The optimality conditions for polyhedral two-stage problems are similar to those for linear problems. For completeness we provide the appropriate formulations. Let us start from the problem with finitely many elementary events ω_k occurring with probabilities p_k , $k = 1, \ldots, K$.

Theorem 2.18. Suppose that the probability measure P has a finite support. Then a point \bar{x} is an optimal solution of the first stage problem (2.44) iff there exist $\pi_k \in \mathfrak{D}(\bar{x}, \omega_k)$, $k = 1, \ldots, K$, such that

$$0 \in \partial f_1(\bar{x}) - \sum_{k=1}^K p_k T_k^\mathsf{T} \pi_k.$$
(2.58)

Proof. Since $f_1(x)$ and $\phi(x) = \mathbb{E}[Q(x, \omega)]$ are convex functions, a necessary and sufficient condition for a point \bar{x} to be a minimizer of $f_1(x) + \phi(x)$ reads

$$0 \in \partial |f_1(\bar{x}) + \phi(\bar{x})|. \tag{2.59}$$

In particular, the above condition requires for $f_1(\bar{x})$ and $\phi(\bar{x})$ to be finite valued. By the Moreau–Rockafellar Theorem we have that $\partial [f_1(\bar{x}) + \phi(\bar{x})] = \partial f_1(\bar{x}) + \partial \phi(\bar{x})$. Note

that there is no need here for additional regularity conditions because of the polyhedricity of functions f_1 and ϕ . The proof can be completed now by using formula for $\partial \phi(\bar{x})$ given in Proposition 2.15. \Box

In the case of general distributions, the derivation of optimality conditions requires additional assumptions.

Theorem 2.19. Suppose that: (i) the recourse is fixed and relatively complete, (ii) the set $\Pi(\omega)$ is nonempty w.p.1, (iii) condition (2.54) holds.

Then a point \bar{x} is an optimal solution of problem (2.44)–(2.45) iff there exists a measurable function $\pi(\omega) \in \mathfrak{D}(\bar{x}, \omega), \omega \in \Omega$, such that

$$0 \in \partial f_1(\bar{x}) - \mathbb{E}[T^{\mathsf{T}}\pi].$$
(2.60)

Proof. The result follows immediately from the optimality condition (2.59) and formula (2.57). Since the recourse is relatively complete, we can omit the normal cone to the domain of $\phi(\cdot)$.

If the recourse is not relatively complete, the analysis becomes complicated. The normal cone to the domain of $\phi(\cdot)$ enters the optimality conditions. For the domain described in (2.56) this cone is rather difficult to describe in a closed form. Some simplification can be achieved when T is deterministic. The analysis then mirrors the linear case, as in Theorem 2.12.

2.3 General Two-Stage Problems

2.3.1 Problem Formulation, Interchangeability

In a general way two-stage stochastic programming problems can be written in the following form

$$\operatorname{Min}_{x \in \mathcal{X}} \left\{ f(x) := \mathbb{E}[F(x, \omega)] \right\},$$
(2.61)

where $F(x, \omega)$ is the optimal value of the second stage problem

$$\underset{y \in \mathcal{G}(x,\omega)}{\operatorname{Min}} g(x, y, \omega).$$
(2.62)

Here $\mathcal{X} \subset \mathbb{R}^n$, $g : \mathbb{R}^n \times \mathbb{R}^m \times \Omega \to \mathbb{R}$ and $\mathcal{G} : \mathbb{R}^n \times \Omega \rightrightarrows \mathbb{R}^m$ is a multifunction. In particular, the linear two-stage problem (2.1)–(2.2) can be formulated in the above form with $g(x, y, \omega) := c^{\mathsf{T}}x + q(\omega)^{\mathsf{T}}y$ and

$$\mathcal{G}(x,\omega) := \{ y : T(\omega)x + W(\omega)y = h(\omega), \ y \ge 0 \}.$$

We also use notation $g_{\omega}(x, y) = g(x, y, \omega)$ and $\mathcal{G}_{\omega}(x) = \mathcal{G}(x, \omega)$.

Of course, the second stage problem (2.62) can be also written in the following equivalent form

$$\underset{y \in \mathbb{R}^m}{\operatorname{Min}} \bar{g}(x, y, \omega),$$
(2.63)

where

$$\bar{g}(x, y, \omega) := \begin{cases} g(x, y, \omega), & \text{if } y \in \mathcal{G}(x, \omega) \\ +\infty, & \text{otherwise.} \end{cases}$$
(2.64)

We assume that the function $\overline{g}(x, y, \omega)$ is random lower semicontinuous. Recall that if $g(x, y, \cdot)$ is measurable for every $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ and $g(\cdot, \cdot, \omega)$ is continuous for a.e. $\omega \in \Omega$, i.e., $g(x, y, \omega)$ is a Carathéodory function, then $g(x, y, \omega)$ is random lsc. Random lower semicontinuity of $\overline{g}(x, y, \omega)$ implies that the optimal value function $F(x, \cdot)$ is measurable (see Theorem 7.42). Moreover, if for a.e. $\omega \in \Omega$ function $F(\cdot, \omega)$ is continuous, then $F(x, \omega)$ is a Carathéodory function, and hence is random lsc. The indicator function $\mathbb{I}_{\mathcal{G}_{\omega}(x)}(y)$ is random lsc if for every $\omega \in \Omega$ the multifunction $\mathcal{G}_{\omega}(\cdot)$ is closed and $\mathcal{G}(x, \omega)$ is measurable with respect to the sigma algebra of $\mathbb{R}^n \times \Omega$ (see Theorem 7.41). Of course, if $g(x, y, \omega)$ and $\mathbb{I}_{\mathcal{G}_{\omega}(x)}(y)$ are random lsc, then their sum $\overline{g}(x, y, \omega)$ is also random lsc.

Now let \mathfrak{Y} be a linear *decomposable* space of measurable mappings from Ω to \mathbb{R}^m . For example, we can take $\mathfrak{Y} := \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^m)$ with $p \in [1, +\infty]$. Then by the interchangeability principle we have

$$\mathbb{E}\underbrace{\left[\inf_{y\in\mathbb{R}^m}\bar{g}(x,y,\omega)\right]}_{F(x,\omega)} = \inf_{\boldsymbol{y}\in\mathfrak{Y}}\mathbb{E}\big[\bar{g}(x,\boldsymbol{y}(\omega),\omega)\big],\tag{2.65}$$

provided that the right hand side of (2.65) is less than $+\infty$ (see Theorem 7.92). This implies the following interchangeability principle for two-stage programming.

Theorem 2.20. *The two-stage problem* (2.61)–(2.62) *is equivalent to the following problem:*

$$\underset{x \in \mathbb{R}^{n}, \, \boldsymbol{y} \in \mathfrak{Y}}{\min} \mathbb{E}\left[g(x, \, \boldsymbol{y}(\omega), \omega)\right]$$
s.t. $x \in \mathcal{X}, \, \boldsymbol{y}(\omega) \in \mathcal{G}(x, \omega)$ a.e. $\omega \in \Omega$.
$$(2.66)$$

The equivalence is understood in the sense that optimal values of problems (2.61) and (2.66) are equal to each other, provided that the optimal value of problem (2.66) is less than $+\infty$. Moreover, assuming that the common optimal value of problems (2.61) and (2.66) is finite, we have at if (\bar{x}, \bar{y}) is an optimal solution of problem (2.66), then \bar{x} is an optimal solution of the first stage problem (2.61) and $\bar{y} = \bar{y}(\omega)$ is an optimal solution of the second stage problem (2.62) for $x = \bar{x}$ and a.e. $\omega \in \Omega$; conversely, if \bar{x} is an optimal solution of the first stage problem (2.61) and for $x = \bar{x}$ and a.e. $\omega \in \Omega$ the second stage problem (2.62) has an optimal solution $\bar{y} = \bar{y}(\omega)$ such that $\bar{y} \in \mathfrak{Y}$, then (\bar{x}, \bar{y}) is an optimal solution of problem (2.62).

Note that optimization in the right hand side of (2.65) and in (2.66) is performed over mappings $\boldsymbol{y} : \Omega \to \mathbb{R}^m$ belonging to the space \mathfrak{Y} . In particular, if $\Omega = \{\omega_1, \ldots, \omega_K\}$ is finite, then by setting $y_k := \boldsymbol{y}(\omega_k), k = 1, \ldots, K$, every such mapping can be identified with a vector (y_1, \ldots, y_K) and the space \mathfrak{Y} with the finite dimensional space \mathbb{R}^{mK} . In that case problem (2.66) takes the form (compare with (2.15)):

$$\underset{x,y_1,\ldots,y_K}{\operatorname{Min}} \sum_{k=1}^{K} p_k g(x, y_k, \omega_k) \\
\text{s.t. } x \in \mathcal{X}, \ y_k \in \mathcal{G}(x, \omega_k), \ k = 1, \ldots, K.$$
(2.67)

2.3.2 Convex Two-Stage Problems

We say that the two-stage problem (2.61)–(2.62) is *convex* if the set \mathcal{X} is convex (and closed) and for every $\omega \in \Omega$ the function $\bar{g}(x, y, \omega)$, defined in (2.64), is convex in $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$. We leave this as an exercise to show that in such case the optimal value function $F(\cdot, \omega)$ is convex, and hence (2.61) is a convex problem. It could be useful to understand what conditions will guarantee convexity of the function $\bar{g}_{\omega}(x, y) = \bar{g}(x, y, \omega)$. We have that $\bar{g}_{\omega}(x, y) = g_{\omega}(x, y) + \mathbb{I}_{\mathcal{G}_{\omega}(x)}(y)$. Therefore $\bar{g}_{\omega}(x, y)$ is convex if $g_{\omega}(x, y)$ is convex and the indicator function $\mathbb{I}_{\mathcal{G}_{\omega}(x)}(y)$ is convex in (x, y). It is not difficult to see that the indicator function $\mathbb{I}_{\mathcal{G}_{\omega}(x)}(y)$ is convex iff the following condition holds for any $t \in [0, 1]$:

$$y \in \mathcal{G}_{\omega}(x), \ y' \in \mathcal{G}_{\omega}(x') \ \Rightarrow ty + (1-t)y' \in \mathcal{G}_{\omega}(tx + (1-t)x').$$
(2.68)

Equivalently this condition can be written as

$$t\mathcal{G}_{\omega}(x) + (1-t)\mathcal{G}_{\omega}(x') \subset \mathcal{G}_{\omega}(tx + (1-t)x'), \quad \forall x, x' \in \mathbb{R}^n, \ \forall t \in [0,1].$$
(2.69)

Multifunction \mathcal{G}_{ω} satisfying the above condition (2.69) is called *convex*. By taking x = x' we obtain that if the multifunction \mathcal{G}_{ω} is convex, then it is convex valued, i.e., the set $\mathcal{G}_{\omega}(x)$ is convex for every $x \in \mathbb{R}^n$.

In the remainder of this section we assume that the multifunction $\mathcal{G}(x,\omega)$ is defined in the following form

$$\mathcal{G}(x,\omega) := \{ y \in Y : T(x,\omega) + W(y,\omega) \in -C \},$$
(2.70)

where Y is a nonempty convex closed subset of \mathbb{R}^m and $T = (t_1, \ldots, t_\ell) : \mathbb{R}^n \times \Omega \to \mathbb{R}^\ell$, $W = (w_1, \ldots, w_\ell) : \mathbb{R}^m \times \Omega \to \mathbb{R}^\ell$ and $C \subset \mathbb{R}^\ell$ is a closed convex cone. Cone C defines a partial order, denoted " \preceq_C ", on the space \mathbb{R}^ℓ . That is, $a \preceq_C b$ iff $b - a \in C$. In that notation the constraint $T(x, \omega) + W(y, \omega) \in -C$ can be written as $T(x, \omega) + W(y, \omega) \preceq_C$ 0. For example, if $C := \mathbb{R}^\ell_+$, then the constraint $T(x, \omega) + W(y, \omega) \preceq_C 0$ means that $t_i(x, \omega) + w_i(y, \omega) \leq 0, i = 1, \ldots, \ell$. We assume that $t_i(x, \omega)$ and $w_i(y, \omega), i = 1, \ldots, \ell$, are Carathéodory functions and that for every $\omega \in \Omega$, mappings $T_\omega(\cdot) = T(\cdot, \omega)$ and $W_\omega(\cdot) = W(\cdot, \omega)$ are convex with respect to the cone C. A mapping $G : \mathbb{R}^n \to \mathbb{R}^\ell$ is said to be *convex with respect* to C if the multifunction $\mathcal{M}(x) := G(x) + C$ is convex. Equivalently, mapping G is convex with respect to C if

$$G(tx + (1-t)x') \preceq_C tG(x) + (1-t)G(x') \ \forall x, x' \in \mathbb{R}^n, \ \forall t \in [0,1].$$

For example, mapping $G(\cdot) = (g_1(\cdot), \ldots, g_\ell(\cdot))$ is convex with respect to $C := \mathbb{R}^{\ell}_+$ iff all its components $g_i(\cdot), i = 1, \ldots, \ell$, are convex functions. Convexity of T_{ω} and W_{ω} implies convexity of the corresponding multifunction \mathcal{G}_{ω} .

We assume, further, that $g(x, y, \omega) := c(x) + q(y, \omega)$, where $c(\cdot)$ and $q(\cdot, \omega)$ are real valued convex functions. For $\mathcal{G}(x, \omega)$ of the form (2.70), and given x, we can write the second stage problem, up to the constant c(x), in the form

$$\begin{aligned} &\underset{y \in Y}{\min} \ q_{\omega}(y) \\ &\text{s.t.} \ W_{\omega}(y) + \chi_{\omega} \preceq_{C} 0 \end{aligned}$$
(2.71)

with $\chi_{\omega} := T(x, \omega)$. Let us denote by $\vartheta(\chi, \omega)$ the optimal value of problems (2.71). Note that $F(x, \omega) = c(x) + \vartheta(T(x, \omega), \omega)$. The (Lagrangian) dual of problem (2.71) can be written in the form

$$\max_{\pi \succeq_C 0} \left\{ \pi^{\mathsf{T}} \chi_{\omega} + \inf_{y \in Y} L_{\omega}(y, \pi) \right\},$$
(2.72)

where

$$L_{\omega}(y,\pi) := q_{\omega}(y) + \pi^{\mathsf{T}} W_{\omega}(y)$$

is the Lagrangian of problem (2.71). We have the following results (see Theorems 7.8 and 7.9).

Proposition 2.21. Let $\omega \in \Omega$ and χ_{ω} be given and suppose that the specified above convexity assumptions are satisfied. Then the following statements hold true:

- (i) The functions $\vartheta(\cdot, \omega)$ and $F(\cdot, \omega)$ are convex.
- (ii) Suppose that problem (2.71) is subconsistent. Then there is no duality gap between problem (2.71) and its dual (2.72) if and only if the optimal value function $\vartheta(\cdot, \omega)$ is lower semicontinuous at χ_{ω} .
- (iii) There is no duality gap between problems (2.71) and (2.72) and the dual problem (2.72) has a nonempty set of optimal solutions if and only if the optimal value function θ(·, ω) is subdifferentiable at χ_ω.
- (iv) Suppose that the optimal value of (2.71) is finite. Then there is no duality gap between problems (2.71) and (2.72) and the dual problem (2.72) has a nonempty and bounded set of optimal solutions if and only if $\chi_{\omega} \in int(\operatorname{dom} \vartheta(\cdot, \omega))$.

The regularity condition $\chi_{\omega} \in \operatorname{int}(\operatorname{dom} \vartheta(\cdot, \omega))$ means that for all small perturbations of χ_{ω} the corresponding problem (2.71) remains feasible.

We can also characterize the differentiability properties of the optimal value functions in terms of the dual problem (2.72). Let us denote by $\mathfrak{D}(\chi, \omega)$ the set of optimal solutions of the dual problem (2.72). This set may be empty, of course.

Proposition 2.22. Let $\omega \in \Omega$, $x \in \mathbb{R}^n$ and $\chi = T(x, \omega)$ be given. Suppose that the specified convexity assumptions are satisfied and that problems (2.71) and (2.72) have finite and equal optimal values. Then

$$\partial \vartheta(\chi, \omega) = \mathfrak{D}(\chi, \omega).$$
 (2.73)

Suppose, further, that functions $c(\cdot)$ and $T_{\omega}(\cdot)$ are differentiable, and

$$0 \in \inf\{T_{\omega}(x) + \nabla T_{\omega}(x)\mathbb{R}^{\ell} - \operatorname{dom}\vartheta(\cdot,\omega)\}.$$
(2.74)

Then

$$\partial F(x,\omega) = \nabla c(x) + \nabla T_{\omega}(x)^{\mathsf{T}} \mathfrak{D}(\chi,\omega).$$
(2.75)

Corollary 2.23. Let $\omega \in \Omega$, $x \in \mathbb{R}^n$ and $\chi = T(x, \omega)$ and suppose that the specified convexity assumptions are satisfied. Then $\vartheta(\cdot, \omega)$ is differentiable at χ if and only if $\mathfrak{D}(\chi, \omega)$ is a singleton. Suppose, further, that the functions $c(\cdot)$ and $T_{\omega}(\cdot)$ are differentiable. Then the function $F(\cdot, \omega)$ is differentiable at every x at which $\mathfrak{D}(\chi, \omega)$ is a singleton.

Proof. If $\mathfrak{D}(\chi, \omega)$ is a singleton, then the set of optimal solutions of the dual problem (2.72) is nonempty and bounded, and hence there is no duality gap between problems (2.71) and (2.72). Thus formula (2.73) holds. Conversely, if $\partial \vartheta(\chi, \omega)$ is a singleton and hence is nonempty, then again there is no duality gap between problems (2.71) and (2.72), and hence formula (2.73) holds.

Now if $\mathfrak{D}(\chi, \omega)$ is a singleton, then $\vartheta(\cdot, \omega)$ is continuous at χ and hence the regularity condition (2.74) holds. It follows then by formula (2.75) that $F(\cdot, \omega)$ is differentiable at x and formula

$$\nabla F(x,\omega) = \nabla c(x) + \nabla T_{\omega}(x)^{\mathsf{T}} \mathfrak{D}(\chi,\omega)$$
(2.76)

holds true.

Let us focus on the expectation function $f(x) := \mathbb{E}[F(x,\omega)]$. If the set Ω is finite, say $\Omega = \{\omega_1, \ldots, \omega_K\}$ with corresponding probabilities $p_k, k = 1, \ldots, K$, then $f(x) = \sum_{k=1}^{K} p_k F(x, \omega_k)$ and subdifferentiability of f(x) is described by the Moreau-Rockafellar Theorem (Theorem 7.4) together with formula (2.75). In particular, $f(\cdot)$ is differentiable at a point x if the functions $c(\cdot)$ and $T_{\omega}(\cdot)$ are differentiable at x and for every $\omega \in \Omega$ the corresponding dual problem (2.72) has a unique optimal solution.

Let us consider the general case, when Ω is not assumed to be finite. By combining Proposition 2.22 and Theorem 7.52 we obtain that, under appropriate regularity conditions ensuring for a.e. $\omega \in \Omega$ formula (2.75) and interchangeability of the subdifferential and expectation operators, it follows that $f(\cdot)$ is subdifferentiable at a point $\bar{x} \in \text{dom } f$ and

$$\partial f(\bar{x}) = \nabla c(\bar{x}) + \int_{\Omega} \nabla T_{\omega}(\bar{x})^{\mathsf{T}} \mathfrak{D}(T_{\omega}(\bar{x}), \omega) \, dP(\omega) + \mathcal{N}_{\mathrm{dom}\,f}(\bar{x}).$$
(2.77)

In particular, it follows from the above formula (2.77) that $f(\cdot)$ is differentiable at \bar{x} iff $\bar{x} \in int(\operatorname{dom} f)$ and $\mathfrak{D}(T_{\omega}(\bar{x}), \omega) = \{\pi(\omega)\}$ is a singleton w.p.1, in which case

$$\nabla f(\bar{x}) = \nabla c(\bar{x}) + \mathbb{E}\left[\nabla T_{\omega}(\bar{x})^{\mathsf{T}} \pi(\omega)\right].$$
(2.78)

We obtain the following conditions for optimality.

Proposition 2.24. Let $\bar{x} \in \mathcal{X} \cap \operatorname{int}(\operatorname{dom} f)$ and assume that formula (2.77) holds. Then \bar{x} is an optimal solution of the first stage problem (2.61) iff there exists a measurable selection

 $\pi(\omega) \in \mathfrak{D}(T(\bar{x},\omega),\omega)$ such that

$$-c(\bar{x}) - \mathbb{E}\left[\nabla T_{\omega}(\bar{x})^{\mathsf{T}}\pi(\omega)\right] \in \mathcal{N}_{\mathcal{X}}(\bar{x}).$$
(2.79)

Proof. Since $\bar{x} \in \mathcal{X} \cap \operatorname{int}(\operatorname{dom} f)$, we have that $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$ and \bar{x} is an optimal solution iff $0 \in \partial f(\bar{x}) + \mathcal{N}_{\mathcal{X}}(\bar{x})$. By formula (2.77) and since $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$, this is equivalent to condition (2.79). \Box

2.4 Nonanticipativity

2.4.1 Scenario Formulation

An additional insight into the structure and properties of two-stage problems can be gained by introducing the concept of *nonanticipativity*. Consider the first stage problem (2.61). Assume that the number of scenarios is finite, i.e., $\Omega = \{\omega_1, \ldots, \omega_K\}$ with respective (positive) probabilities p_1, \ldots, p_K . Let us relax the first stage problem by replacing vector x with K vectors x_1, x_2, \ldots, x_K , one for each scenario. We obtain the following relaxation of problem (2.61):

$$\operatorname{Min}_{x_1,\ldots,x_K} \sum_{k=1}^K p_k F(x_k,\omega_k) \text{ subject to } x_k \in \mathcal{X}, \ k = 1,\ldots,K.$$
(2.80)

We observe that problem (2.80) is separable in the sense that it can be split into K smaller problems, one for each scenario:

$$\underset{x_k \in X}{\operatorname{Min}} F(x_k, \omega_k), \quad k = 1, \dots, K,$$
(2.81)

and that the optimal value of problem (2.80) is equal to the weighted sum, with weights p_k , of the optimal values of problems (2.81), k = 1, ..., K. For example, in the case of two-stage linear program (2.15), relaxation of the form (2.80) leads to solving K smaller problems

$$\begin{aligned} \min_{\substack{x_k \ge 0, y_k \ge 0}} c^{\mathsf{T}} x_k + q_k^{\mathsf{T}} y_k \\ \text{s.t. } A x_k = b, \ T_k x_k + W_k y_k = h_k \end{aligned}$$

Problem (2.80), however, is not suitable for modeling a two stage decision process. This is because the first stage decision variables x_k in (2.80) are now allowed to depend on a realization of the random data at the second stage. This can be fixed by introducing the additional constraint

$$(x_1, \dots, x_K) \in \mathfrak{L}, \tag{2.82}$$

where $\mathfrak{L} := \{ \boldsymbol{x} = (x_1, \dots, x_K) : x_1 = \dots = x_K \}$ is a linear subspace of the *nK*dimensional vector space $\mathfrak{X} := \mathbb{R}^n \times \dots \times \mathbb{R}^n$. Due to the constraint (2.82), all realizations $x_k, k = 1, \dots, K$, of the first stage decision vector are equal to each other, that is, they do not depend on the realization of the random data. The constraint (2.82) can be written in different forms, which can be convenient in various situations, and will be referred to as the *nonanticipativity constraint*. Together with the nonanticipativity constraint (2.82), problem (2.80) becomes

$$\underset{x_1,\ldots,x_K}{\operatorname{Min}} \sum_{k=1}^{K} p_k F(x_k,\omega_k)$$
s.t. $x_1 = \cdots = x_K, \ x_k \in \mathcal{X}, \ k = 1,\ldots,K.$

$$(2.83)$$

Clearly, the above problem (2.83) is equivalent to problem (2.61). Such nonanticipativity constraints are especially important in multistage modeling which we discuss later.

A way to write the nonanticipativity constraint is to require that

$$x_k = \sum_{i=1}^{K} p_i x_i, \quad k = 1, \dots, K,$$
 (2.84)

which is convenient for extensions to the case of a continuous distribution of problem data. Equations (2.84) can be interpreted in the following way. Consider the space \mathfrak{X} equipped with the scalar product

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{i=1}^{K} p_i x_i^{\mathsf{T}} y_i.$$
 (2.85)

Define linear operator $P:\mathfrak{X} \to \mathfrak{X}$ as

$$\boldsymbol{P}\boldsymbol{x} := \left(\sum_{i=1}^{K} p_i x_i, \dots, \sum_{i=1}^{K} p_i x_i\right).$$

Constraint (2.84) can be compactly written as

$$x = Px$$

It can be verified that P is the orthogonal projection operator of \mathfrak{X} , equipped with the scalar product (2.85), onto its subspace \mathfrak{L} . Indeed, P(Px) = Px, and

$$\langle \boldsymbol{P}\boldsymbol{x}, \boldsymbol{y} \rangle = \left(\sum_{i=1}^{K} p_i x_i\right)^{\mathsf{T}} \left(\sum_{k=1}^{K} p_k y_k\right) = \langle \boldsymbol{x}, \boldsymbol{P}\boldsymbol{y} \rangle.$$
 (2.86)

The range space of P, which is the linear space \mathfrak{L} , is called the nonanticipativity subspace of \mathfrak{X} .

Another way to algebraically express nonanticipativity, which is convenient for numerical methods, is to write the system of equations

$$x_1 = x_2,$$

 $x_2 = x_3,$
 \vdots
 $x_{K-1} = x_K.$
(2.87)

This system is very sparse: each equation involves only two variables, and each variable appears in at most two equations, which is convenient for many numerical solution methods.

2.4.2 Dualization of Nonanticipativity Constraints

We discuss now a dualization of problem (2.80) with respect to the nonanticipativity constraints (2.84). Assigning to these nonanticipativity constraints Lagrange multipliers $\lambda_k \in \mathbb{R}^n$, $k = 1, \ldots, K$, we can write the Lagrangian

$$L(\boldsymbol{x},\boldsymbol{\lambda}) := \sum_{k=1}^{K} p_k F(x_k,\omega_k) + \sum_{k=1}^{K} p_k \lambda_k^{\mathsf{T}} \left(x_k - \sum_{i=1}^{K} p_i x_i \right).$$

Note that since P is an orthogonal projection, I - P is also an orthogonal projection (onto the space orthogonal to \mathfrak{L}), and hence

$$\sum_{k=1}^{K} p_k \lambda_k^{\mathsf{T}} \left(x_k - \sum_{i=1}^{K} p_i x_i \right) = \langle \boldsymbol{\lambda}, (\boldsymbol{I} - \boldsymbol{P}) \boldsymbol{x} \rangle = \langle (\boldsymbol{I} - \boldsymbol{P}) \boldsymbol{\lambda}, \boldsymbol{x} \rangle.$$

Therefore the above Lagrangian can be written in the following equivalent form

$$L(\boldsymbol{x},\boldsymbol{\lambda}) = \sum_{k=1}^{K} p_k F(x_k,\omega_k) + \sum_{k=1}^{K} p_k \left(\lambda_k - \sum_{j=1}^{K} p_j \lambda_j\right)^{\mathsf{T}} x_k.$$

Let us observe that shifting the multipliers λ_k , k = 1, ..., K, by a constant vector does not change the value of the Lagrangian, because the expression $\lambda_k - \sum_{j=1}^{K} p_j \lambda_j$ is invariant to such shifts. Therefore, with no loss of generality we can assume that

$$\sum_{j=1}^{K} p_j \lambda_j = 0.$$

or, equivalently, that $P\lambda = 0$. Dualization of problem (2.80) with respect to the nonanticipativity constraints takes the form of the following problem

$$\operatorname{Max}_{\boldsymbol{\lambda}} \left\{ D(\boldsymbol{\lambda}) := \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \right\} \text{ subject to } \boldsymbol{P}\boldsymbol{\lambda} = 0.$$
(2.88)

By general duality theory we have that the optimal value of problem (2.61) is greater than or equal to the optimal value of problem (2.88). These optimal values are equal to each other under some regularity conditions, we will discuss a general case in the next section. In particular, if the two stage problem is linear and since the nonanticipativity constraints are linear, we have in that case that there is no duality gap between problem (2.61) and its dual problem (2.88) unless both problems are infeasible.

Let us take a closer look at the dual problem (2.88). Under the condition $P\lambda = 0$, the Lagrangian can be written simply as

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = \sum_{k=1}^{K} p_k \big(F(x_k, \omega_k) + \lambda_k^{\mathsf{T}} x_k \big).$$

We see that the Lagrangian can be split into K components:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = \sum_{k=1}^{K} p_k L_k(x_k, \lambda_k),$$

where $L_k(x_k, \lambda_k) := F(x_k, \omega_k) + \lambda_k^{\mathsf{T}} x_k$. It follows that

$$D(\boldsymbol{\lambda}) = \sum_{j=1}^{K} p_k \mathcal{D}_k(\lambda_k),$$

where

$$\mathcal{D}_k(\lambda_k) := \inf_{x_k \in X} L_k(x_k, \lambda_k).$$

For example, in the case of the two-stage linear program (2.15), $\mathcal{D}_k(\lambda_k)$ is the optimal value of the problem

$$\begin{split} \min_{x_k, y_k} (c + \lambda_k)^{\mathsf{T}} x_k + q_k^{\mathsf{T}} y_k \\ \text{s.t.} \ Ax_k &= b, \\ T_k x_k + W_k y_k &= h_k, \\ x_k &\geq 0, \quad y_k \geq 0. \end{split}$$

We see that value of the dual function $D(\lambda)$ can be calculated by solving K independent scenario subproblems.

Suppose that there is no duality gap between problem (2.61) and its dual (2.88) and their common optimal value is finite. This certainly holds true if the problem is linear, and both problems, primal and dual, are feasible. Let $\bar{\lambda} = (\bar{\lambda}_1, \dots, \bar{\lambda}_K)$ be an optimal solution of the dual problem (2.88). Then the set of optimal solutions of problem (2.61) is contained in the set of optimal solutions of the problem

$$\min_{x_k \in \mathcal{X}} \sum_{k=1}^{K} p_k L_k(x_k, \bar{\lambda}_k)$$
(2.89)

This inclusion can be strict, i.e., the set of optimal solutions of (2.89) can be larger than the set of optimal solutions of problem (2.61) (see an example of linear program defined in (7.32)). Of course, if problem (2.89) has *unique* optimal solution $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_K)$, then $\bar{x} \in \mathfrak{L}$, i.e., $\bar{x}_1 = \ldots = \bar{x}_K$, and this is also the optimal solution of problem (2.61) with \bar{x} being equal to the common value of $\bar{x}_1, \ldots, \bar{x}_K$. Note also that the above problem (2.89) is separable, i.e., \bar{x} is an optimal solution of (2.89) iff for every $k = 1, \ldots, K$, \bar{x}_k is an optimal solution of the problem

$$\operatorname{Min}_{x_k \in \mathcal{X}} L_k(x_k, \bar{\lambda}_k)$$

2.4.3 Nonanticipativity Duality for General Distributions

In this section we discuss dualization of the first stage problem (2.61) with respect to nonanticipativity constraints in general (not necessarily finite scenarios) case. For the sake of convenience we write problem (2.61) in the form

$$\min_{x \in \mathbb{R}^n} \left\{ \bar{f}(x) := \mathbb{E}[\bar{F}(x,\omega)] \right\},$$
(2.90)

where $\bar{F}(x,\omega) := F(x,\omega) + \mathbb{I}_{\mathcal{X}}(x)$, i.e., $\bar{F}(x,\omega) = F(x,\omega)$ if $x \in \mathcal{X}$ and $\bar{F}(x,\omega) = +\infty$ otherwise. Let \mathfrak{X} be a linear *decomposable* space of measurable mappings from Ω to \mathbb{R}^n .

Unless stated otherwise we use $\mathfrak{X} := \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ for some $p \in [1, +\infty]$ such that for every $\boldsymbol{x} \in \mathfrak{X}$ the expectation $\mathbb{E}[\bar{F}(\boldsymbol{x}(\omega), \omega)]$ is well defined. Then we can write problem (2.90) in the following equivalent form

$$\min_{\boldsymbol{x}\in\mathcal{S}} \mathbb{E}[\bar{F}(\boldsymbol{x}(\omega),\omega)],$$
(2.91)

where \mathfrak{L} is a linear subspace of \mathfrak{X} formed by mappings $\boldsymbol{x} : \Omega \to \mathbb{R}^n$ which are constant almost everywhere, i.e.,

$$\mathfrak{L} := \left\{ \boldsymbol{x} \in \mathfrak{X} : \boldsymbol{x}(\omega) \equiv x \text{ for some } x \in \mathbb{R}^n \right\},\$$

where $\boldsymbol{x}(\omega) \equiv x$ means that $\boldsymbol{x}(\omega) = x$ for a.e. $\omega \in \Omega$.

Consider the dual⁶ $\mathfrak{X}^* := \mathcal{L}_q(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ of the space \mathfrak{X} and define the scalar product (bilinear form)

$$\langle \boldsymbol{\lambda}, \boldsymbol{x} \rangle := \mathbb{E} \left[\boldsymbol{\lambda}^{\mathsf{T}} \boldsymbol{x} \right] = \int_{\Omega} \boldsymbol{\lambda}(\omega)^{\mathsf{T}} \boldsymbol{x}(\omega) dP(\omega), \ \boldsymbol{\lambda} \in \mathfrak{X}^{*}, \ \boldsymbol{x} \in \mathfrak{X}.$$

Also, consider the projection operator $P : \mathfrak{X} \to \mathfrak{L}$ defined as $[Px](\omega) \equiv \mathbb{E}[x]$. Clearly the space \mathfrak{L} is formed by such $x \in \mathfrak{X}$ that Px = x. Note that

$$\langle oldsymbol{\lambda}, oldsymbol{P} oldsymbol{x}
angle = \mathbb{E}\left[oldsymbol{\lambda}
ight]^{\mathsf{I}} \mathbb{E}\left[oldsymbol{x}
ight] = \langle oldsymbol{P}^*oldsymbol{\lambda}, oldsymbol{x}
angle,$$

where P^* is a projection operator $[P^*\lambda](\omega) \equiv \mathbb{E}[\lambda]$ from \mathfrak{X}^* onto its subspace formed by constant almost everywhere mappings. In particular, if p = 2, then $\mathfrak{X}^* = \mathfrak{X}$ and $P^* = P$.

With problem (2.91) is associated the following Lagrangian

$$L(\boldsymbol{x},\boldsymbol{\lambda}) := \mathbb{E}[\bar{F}(\boldsymbol{x}(\omega),\omega)] + \mathbb{E}\left[\boldsymbol{\lambda}^{\mathsf{T}}(\boldsymbol{x}-\mathbb{E}[\boldsymbol{x}])\right].$$

Note that

$$\mathbb{E}\left[oldsymbol{\lambda}^{\mathsf{T}}(oldsymbol{x}-\mathbb{E}[oldsymbol{x}])
ight]=\langleoldsymbol{\lambda},oldsymbol{x}-oldsymbol{P}oldsymbol{x}
angle=\langleoldsymbol{\lambda}-oldsymbol{P}^*oldsymbol{\lambda},oldsymbol{x}
angle,$$

and $\lambda - P^* \lambda$ does not change by adding a constant to $\lambda(\cdot)$. Therefore we can set $P^* \lambda = 0$, in which case

$$L(\boldsymbol{x},\boldsymbol{\lambda}) = \mathbb{E}\left[\bar{F}(\boldsymbol{x}(\omega),\omega) + \boldsymbol{\lambda}(\omega)^{\mathsf{T}}\boldsymbol{x}(\omega)\right] \text{ for } \mathbb{E}[\boldsymbol{\lambda}] = 0.$$
(2.92)

This leads to the following dual of problem (2.90):

$$\max_{\boldsymbol{\lambda} \in \mathfrak{X}^*} \left\{ D(\boldsymbol{\lambda}) := \inf_{\boldsymbol{x} \in \mathfrak{X}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \right\} \text{ subject to } \mathbb{E}[\boldsymbol{\lambda}] = 0.$$
(2.93)

In case of finitely many scenarios the above dual is the same as the dual problem (2.88).

⁶Recall that 1/p + 1/q = 1 for $p, q \in (1, +\infty)$. If p = 1, then $q = +\infty$. Also for $p = +\infty$ we use q = 1. This results in a certain abuse of notation since the space $\mathfrak{X} = \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ is not reflexive and $\mathfrak{X}^* = \mathcal{L}_1(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ is smaller than its dual. Note also that if $\boldsymbol{x} \in \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^n)$, then its expectation $\mathbb{E}[\boldsymbol{x}] = \int_{\Omega} \boldsymbol{x}(\omega) dP(\omega)$ is well defined and is an element of vector space \mathbb{R}^n .

By the interchangeability principle (Theorem 7.92) we have

$$\inf_{\boldsymbol{x}\in\mathfrak{X}} \mathbb{E}\left[\bar{F}(\boldsymbol{x}(\omega),\omega) + \boldsymbol{\lambda}(\omega)^{\mathsf{T}}\boldsymbol{x}(\omega)\right] = \mathbb{E}\left[\inf_{\boldsymbol{x}\in\mathbb{R}^{n}} \left(\bar{F}(\boldsymbol{x},\omega) + \boldsymbol{\lambda}(\omega)^{\mathsf{T}}\boldsymbol{x}\right)\right].$$

Consequently

$$D(\boldsymbol{\lambda}) = \mathbb{E}[\mathcal{D}_{\omega}(\boldsymbol{\lambda}(\omega))],$$

where $\mathcal{D}_{\omega}: \mathbb{R}^n \to \overline{\mathbb{R}}$ is defined as

$$\mathcal{D}_{\omega}(\lambda) := \inf_{x \in \mathbb{R}^n} \left(\lambda^{\mathsf{T}} x + \bar{F}_{\omega}(x) \right) = -\sup_{x \in \mathbb{R}^n} \left(-\lambda^{\mathsf{T}} x - \bar{F}_{\omega}(x) \right) = -\bar{F}_{\omega}^*(-\lambda).$$
(2.94)

That is, in order to calculate the dual function $D(\lambda)$ one needs to solve for every $\omega \in \Omega$ the finite dimensional optimization problem (2.94), and then to integrate the optimal values obtained.

By the general theory we have that the optimal value of problem (2.91), which is the same as the optimal value of problem (2.90), is greater than or equal to the optimal value of its dual (2.93). We also have that there is no duality gap between problem (2.91) and its dual (2.93) and both problems have optimal solutions \bar{x} and $\bar{\lambda}$, respectively, iff $(\bar{x}, \bar{\lambda})$ is a saddle point of the Lagrangian defined in (2.92). By definition a point $(\bar{x}, \bar{\lambda}) \in \mathfrak{X} \times \mathfrak{X}^*$ is a saddle point of the Lagrangian iff

$$\bar{\boldsymbol{x}} \in \arg\min_{\boldsymbol{x}\in\mathfrak{L}} L(\boldsymbol{x},\bar{\boldsymbol{\lambda}}) \text{ and } \bar{\boldsymbol{\lambda}} \in \operatorname*{arg\,max}_{\boldsymbol{\lambda}:\mathbb{E}[\boldsymbol{\lambda}]=0} L(\bar{\boldsymbol{x}},\boldsymbol{\lambda}).$$
 (2.95)

By the interchangeability principle (see equation (7.277) of Theorem 7.92) we have that first condition in (2.95) can be written in the following equivalent form

$$\bar{\boldsymbol{x}}(\omega) \equiv \bar{x} \text{ and } \bar{x} \in \arg\min_{x \in \mathbb{R}^n} \left\{ \bar{F}(x,\omega) + \bar{\boldsymbol{\lambda}}(\omega)^{\mathsf{T}} x \right\} \text{ a.e. } \omega \in \Omega.$$
 (2.96)

Since $\bar{x}(\omega) \equiv \bar{x}$, the second condition in (2.95) means that $\mathbb{E}[\bar{\lambda}] = 0$.

Let us assume now that the considered problem is *convex*, i.e., the set \mathcal{X} is convex (and closed) and $F_{\omega}(\cdot)$ is a convex function for a.e. $\omega \in \Omega$. It follows that $\bar{F}_{\omega}(\cdot)$ is a convex function for a.e. $\omega \in \Omega$. Then the second condition in (2.96) holds iff $\bar{\lambda}(\omega) \in -\partial \bar{F}_{\omega}(\bar{x})$ for a.e. $\omega \in \Omega$. Together with condition $\mathbb{E}[\bar{\lambda}] = 0$ this means that

$$0 \in \mathbb{E}\left[\partial \bar{F}_{\omega}(\bar{x})\right]. \tag{2.97}$$

It follows that the Lagrangian has a saddle point iff there exists $\bar{x} \in \mathbb{R}^n$ satisfying condition (2.97). We obtain the following result.

Theorem 2.25. Suppose that the function $F(x, \omega)$ is random lower semicontinuous, the set X is convex and closed and for a.e. $\omega \in \Omega$ the function $F(\cdot, \omega)$ is convex. Then there is no duality gap between problems (2.90) and (2.93) and both problems have optimal solutions iff there exists $\bar{x} \in \mathbb{R}^n$ satisfying condition (2.97). In that case \bar{x} is an optimal solution of (2.90) and a measurable selection $\bar{\lambda}(\omega) \in -\partial \bar{F}_{\omega}(\bar{x})$ such that $\mathbb{E}[\bar{\lambda}] = 0$ is an optimal solution solution of (2.93).

Recall that the inclusion $\mathbb{E}\left[\partial \bar{F}_{\omega}(\bar{x})\right] \subset \partial \bar{f}(\bar{x})$ always holds (see equation (7.133) in the proof of Theorem 7.52). Therefore, condition (2.97) implies that $0 \in \partial \bar{f}(\bar{x})$, which in turn implies that \bar{x} is an optimal solution of (2.90). Conversely, if \bar{x} is an optimal solution of (2.90), then $0 \in \partial \bar{f}(\bar{x})$, and if in addition $\mathbb{E}\left[\partial \bar{F}_{\omega}(\bar{x})\right] = \partial \bar{f}(\bar{x})$, then (2.97) follows. Therefore, Theorems 2.25 and 7.52 imply the following result.

Theorem 2.26. Suppose that: (i) the function $F(x, \omega)$ is random lower semicontinuous, (ii) the set \mathcal{X} is convex and closed, (iii) for a.e. $\omega \in \Omega$ the function $F(\cdot, \omega)$ is convex, (iv) problem (2.90) possesses an optimal solution \bar{x} such that $\bar{x} \in int(dom f)$. Then there is no duality gap between problems (2.90) and (2.93) and the dual problem (2.93) has an optimal solution $\bar{\lambda}$, and the constant mapping $\bar{x}(\omega) \equiv \bar{x}$ is an optimal solution of the problem

$$\operatorname{Min}_{\boldsymbol{x}\in\mathfrak{X}} \mathbb{E}\left[\bar{F}(\boldsymbol{x}(\omega),\omega) + \bar{\boldsymbol{\lambda}}(\omega)^{\mathsf{T}}\boldsymbol{x}(\omega)\right].$$

Proof. Since \bar{x} is an optimal solution of problem (2.90) we have that $\bar{x} \in \mathcal{X}$ and $f(\bar{x})$ is finite. Moreover, since $\bar{x} \in \text{int}(\text{dom} f)$ and f is convex, it follows that f is proper and $\mathcal{N}_{\text{dom} f}(\bar{x}) = \{0\}$. Therefore, it follows by Theorem 7.52 that $\mathbb{E}\left[\partial F_{\omega}(\bar{x})\right] = \partial f(\bar{x})$. Furthermore, since $\bar{x} \in \text{int}(\text{dom} f)$ we have that $\partial \bar{f}(\bar{x}) = \partial f(\bar{x}) + \mathcal{N}_{\mathcal{X}}(\bar{x})$, and hence $\mathbb{E}\left[\partial \bar{F}_{\omega}(\bar{x})\right] = \partial \bar{f}(\bar{x})$. By optimality of \bar{x} , we also have that $0 \in \partial \bar{f}(\bar{x})$. Consequently, $0 \in \mathbb{E}\left[\partial \bar{F}_{\omega}(\bar{x})\right]$, and hence the proof can be completed by applying Theorem 2.25. \Box

If \mathcal{X} is a subset of $\operatorname{int}(\operatorname{dom} f)$, then any point $x \in \mathcal{X}$ is an interior point of $\operatorname{dom} f$. In that case condition (iv) of the above theorem is reduced to existence of an optimal solution. The condition $\mathcal{X} \subset \operatorname{int}(\operatorname{dom} f)$ means that $f(x) < +\infty$ for every x in a neighborhood of the set \mathcal{X} . This requirement is slightly stronger than the condition of relatively complete recourse.

Example 2.27 (Capacity Expansion (continued)) Let us consider the capacity expansion problem of Examples 2.4 and 2.13. Suppose that \bar{x} is the optimal first stage decision and let $\bar{y}_{ij}(\xi)$ be the corresponding optimal second stage decisions. The scenario problem has the form

$$\begin{split} &\operatorname{Min} \ \sum_{\substack{(i,j)\in\mathcal{A}\\(i,j)\in\mathcal{A}_+(n)}} \left[(c_{ij} + \lambda_{ij}(\xi)) x_{ij} + q_{ij} y_{ij} \right] \\ & \text{s.t.} \ \sum_{\substack{(i,j)\in\mathcal{A}_+(n)\\0\leq y_{ij}\leq x_{ij}, \quad (i,j)\in\mathcal{A}_-(n)}} y_{ij} = \xi_n, \quad n\in\mathcal{N}, \end{split}$$

From Example 2.13 we know that there exist random node potentials $\mu_n(\xi)$, $n \in \mathcal{N}$, such that for all $\xi \in \Xi$ we have $\mu(\xi) \in \mathcal{M}(\bar{x}, \xi)$, and conditions (2.42)–(2.43) are satisfied. Also, the random variables $g_{ij}(\xi) = -\max\{0, \mu_i(\xi) - \mu_j(\xi) - q_{ij}\}$ are the corresponding subgradients of the second stage cost. Define

$$\lambda_{ij}(\xi) = \max\{0, \mu_i(\xi) - \mu_j(\xi) - q_{ij}\} - \int_{\Xi} \max\{0, \mu_i(\xi) - \mu_j(\xi) - q_{ij}\} P(d\xi), \quad (i, j) \in \mathcal{A}.$$

We can easily verify that $x_{ij}(\xi) = \bar{x}_{ij}$ and $\bar{y}_{ij}(\xi)$, $(i, j) \in \mathcal{A}$, are optimal solution of the scenario problem, because the first term of λ_{ij} cancels with the subgradient $g_{ij}(\xi)$, while the second term satisfies the optimality conditions (2.42)–(2.43). Moreover, $\mathbb{E}[\lambda] = 0$ by construction.

2.4.4 Value of Perfect Information

Consider the following relaxation of the two-stage problem (2.61)–(2.62):

$$\min_{\boldsymbol{x}\in\mathfrak{X}} \mathbb{E}[\bar{F}(\boldsymbol{x}(\omega),\omega)].$$
(2.98)

This relaxation is obtained by removing the nonanticipativity constraint from the formulation (2.91) of the first stage problem. By the interchangeability principle (Theorem 7.92) we have that the optimal value of the above problem (2.98) is equal to $\mathbb{E}\left[\inf_{x \in \mathbb{R}^n} \bar{F}(x,\omega)\right]$. The value $\inf_{x \in \mathbb{R}^n} \bar{F}(x,\omega)$ is equal to the optimal value of problem

$$\underset{x \in \mathcal{X}, y \in \mathcal{G}(x,\omega)}{\operatorname{Min}} g(x, y, \omega).$$
(2.99)

That is, the optimal value of problem (2.98) is obtained by solving problems of the form (2.99), one for each $\omega \in \Omega$, and then taking the expectation of the calculated optimal values.

Solving problems of the form (2.99) makes sense if we have a perfect information about the data, i.e., the scenario $\omega \in \Omega$ is known at the time when the first stage decision should be made. The problem (2.99) is deterministic, e.g., in the case of two-stage linear program (2.1)–(2.2) it takes the form

$$\min_{x \ge 0, y \ge 0} c^{\mathsf{T}} x + q^{\mathsf{T}} y \text{ subject to } Ax = b, \ Tx + Wy = h.$$

An optimal solution of the second stage problem (2.99) depends on $\omega \in \Omega$ and is called the *wait-and-see* solution.

We have that for any $x \in \mathcal{X}$ and $\omega \in \Omega$, the inequality $F(x, \omega) \ge \inf_{x \in \mathcal{X}} F(x, \omega)$ clearly holds, and hence $\mathbb{E}[F(x, \omega)] \ge \mathbb{E}[\inf_{x \in \mathcal{X}} F(x, \omega)]$. It follows that

$$\inf_{x \in \mathcal{X}} \mathbb{E}[F(x,\omega)] \ge \mathbb{E}\left[\inf_{x \in \mathcal{X}} F(x,\omega)\right].$$
(2.100)

Another way to view the above inequality is to observe that problem (2.98) is a relaxation of the corresponding two-stage stochastic problem, which of course implies (2.100).

Suppose that the two-stage problem has an optimal solution $\bar{x} \in \arg \min_{x \in \mathcal{X}} \mathbb{E}[F(x, \omega)]$. As $F(\bar{x}, \omega) - \inf_{x \in \mathcal{X}} F(x, \omega) \ge 0$ for all $\omega \in \Omega$, we conclude that

$$\mathbb{E}[F(\bar{x},\omega)] = \mathbb{E}\left[\inf_{x\in\mathcal{X}}F(x,\omega)\right]$$
(2.101)

iff $F(\bar{x}, \omega) = \inf_{x \in \mathcal{X}} F(x, \omega)$ w.p.1. That is, equality in (2.101) holds iff

$$F(\bar{x},\omega) = \inf_{x \in \mathcal{X}} F(x,\omega) \text{ for a.e. } \omega \in \Omega.$$
(2.102)
In particular, this happens if $\overline{F}_{\omega}(x)$ has a minimizer independent of $\omega \in \Omega$. This, of course, may happen only in rather specific situations.

The difference $F(\bar{x}, \omega) - \inf_{x \in \mathcal{X}} F(x, \omega)$ represents the value of perfect information of knowing ω . Consequently

$$\mathrm{EVPI} := \inf_{x \in \mathcal{X}} \mathbb{E}[F(x, \omega)] - \mathbb{E}\left[\inf_{x \in \mathcal{X}} F(x, \omega)\right]$$

is called the *expected value of perfect information*. It follows from (2.100) that EVPI is always nonnegative and EVPI=0 iff condition (2.102) holds.

Exercises

- 2.1. Consider the assembly problem discussed in section 1.3.1 in two cases:
 - (i) The demand which is not satisfied from the pre-ordered quantities of parts is lost.
 - (ii) All demand has to be satisfied, by making additional orders of the missing parts. In this case, the cost of each additionally ordered part j is $r_j > c_j$.

For each of these cases describe the subdifferential of the recourse cost and of the expected recourse cost.

- 2.2. A transportation company has n depots among which they send cargo. The demand for transportation between depot i and depot $j \neq i$ is modeled as a random variable D_{ij} . The total capacity of vehicles currently available at depot i is denoted s_i , i = 1, ..., n. The company considers repositioning their fleet to better prepare to the uncertain demand. It costs c_{ij} to move a unit of capacity from location i to location j. After repositioning, the realization of the random vector D is observed, and the demand is served, up to the limit determined by the transportation capacity available at each location. The profit from transporting a unit of cargo from location i to location j is equal q_{ij} . If the total demand at location i exceeds the capacity available at location i, the excessive demand is lost. It is up to the company to decide how much of each demand D_{ij} be served, and which part will remain unsatisfied. For simplicity, we consider all capacity and transportation quantities as continuous variables.
 - (a) Formulate the problem of maximizing the expected profit as a two stage stochastic programming problem.
 - (b) Describe the subdifferential of the recourse cost and the expected recourse cost.
- 2.3. Show that the function $s_q(\cdot)$, defined in (2.4), is convex.
- 2.4. Consider the optimal value $Q(x,\xi)$ of the second stage problem (2.2). Show that $Q(\cdot,\xi)$ is differentiable at a point x iff the dual problem (2.3) has unique optimal solution $\bar{\pi}$, in which case $\nabla_x Q(x,\xi) = -T^{\mathsf{T}}\bar{\pi}$.

- 2.5. Consider the two-stage problem (2.1)–(2.2) with fixed recourse. Show that the following conditions are equivalent: (i) problem (2.1)–(2.2) has complete recourse, (ii) the feasible set $\Pi(q)$ of the dual problem is bounded for every q, (iii) the system $W^{\mathsf{T}}\pi \leq 0$ has only one solution $\pi = 0$.
- 2.6. Show that if random vector ξ has a finite support, then condition (2.24) is necessary and sufficient for relatively complete recourse.
- 2.7. Show that the conjugate function of a polyhedral function is also polyhedral.
- 2.8. Show that if $Q(x, \omega)$ is finite, then the set $\mathfrak{D}(x, \omega)$ of optimal solutions of problem (2.46) is a nonempty convex closed polyhedron.
- 2.9. Consider problem (2.63) and its optimal value $F(x, \omega)$. Show that $F(x, \omega)$ is convex in x if $\overline{g}(x, y, \omega)$ is convex in (x, y). Show that the indicator function $\mathbb{I}_{\mathcal{G}_{\omega}(x)}(y)$ is convex in (x, y) iff condition (2.68) holds for any $t \in [0, 1]$.
- 2.10. Show that equation (2.86) implies that $\langle \boldsymbol{x} \boldsymbol{P} \boldsymbol{x}, \boldsymbol{y} \rangle = 0$ for any $\boldsymbol{x} \in \mathfrak{X}$ and $\boldsymbol{y} \in \mathfrak{L}$, i.e., that \boldsymbol{P} is the orthogonal projection of \mathfrak{X} onto \mathfrak{L} .
- 2.11. Derive the form of the dual problem for the linear two stage stochastic programming problem in form (2.80) with nonanticipativity constraints (2.87).

Chapter 3 Multistage Problems

Andrzej Ruszczyński and Alexander Shapiro

3.1 **Problem Formulation**

3.1.1 The General Setting

The two-stage stochastic programming models can be naturally extended to a multistage setting. We already discussed examples of such decision processes in sections 1.2.3 and 1.4.2 for a multistage inventory model and a multistage portfolio selection problem, respectively. In the multistage setting the uncertain data ξ_1, \ldots, ξ_T is revealed gradually over time, in T periods, and our decisions should be adapted to this process. The decision process has the form:

decision $(x_1) \rightsquigarrow$ observation $(\xi_2) \rightsquigarrow$ decision $(x_2) \rightsquigarrow$ \rightsquigarrow observation $(\xi_T) \rightsquigarrow$ decision (x_T) .

We view the sequence $\xi_t \in \mathbb{R}^{d_t}$, t = 1, ..., T, of data vectors as a *stochastic process*, i.e., as a sequence of random variables with a specified probability distribution. We use notation $\xi_{[t]} := (\xi_1, ..., \xi_t)$ to denote the history of the process up to time t.

The values of the decision vector x_t , chosen at stage t, may depend on the information (data) $\xi_{[t]}$ available up to time t, but not on the results of future observations. This is the basic requirement of *nonanticipativity*. As x_t may depend on $\xi_{[t]}$, the sequence of decisions is a stochastic process as well.

We say that the process $\{\xi_t\}$ is *stagewise independent* if ξ_t is stochastically independent of $\xi_{[t-1]}, t = 2, ..., T$. It is said that the process is *Markovian* if for every t = 2, ..., T, the conditional distribution of ξ_t given $\xi_{[t-1]}$ is the same as the conditional distribution of ξ_t given ξ_{t-1} . It is evident that if a process is stagewise independent, then it is Markovian.

As before, we often use the same notation ξ_t to denote a random vector and its particular realization. Which of these two meanings will be used in a particular situation will be clear from the context.

A T-stage stochastic programming problem can be written in the following general nested formulation

$$\underset{x_1 \in \mathcal{X}_1}{\operatorname{Min}} f_1(x_1) + \mathbb{E} \left[\inf_{x_2 \in \mathcal{X}_2(x_1, \xi_2)} f_2(x_2, \xi_2) + \mathbb{E} \left[\cdots + \mathbb{E} \left[\inf_{x_T \in \mathcal{X}_T(x_{T-1}, \xi_T)} f_T(x_T, \xi_T) \right] \right] \right],$$
(3.1)

driven by the random data process $\xi_1, \xi_2, \ldots, \xi_T$. Here $x_t \in \mathbb{R}^{n_t}, t = 1, \ldots, T$, are decision variables, $f_t : \mathbb{R}^{n_t} \times \mathbb{R}^{d_t} \to \mathbb{R}$ are continuous functions and $\mathcal{X}_t : \mathbb{R}^{n_{t-1}} \times \mathbb{R}^{d_t} \rightrightarrows \mathbb{R}^{n_t}$, $t = 2, \ldots, T$, are measurable closed valued multifunctions. The first stage data, i.e., the vector ξ_1 , the function $f_1 : \mathbb{R}^{n_1} \to \mathbb{R}$, and the set $\mathcal{X}_1 \subset \mathbb{R}^{n_1}$ are deterministic. It is said that the multistage problem is *linear* if the objective functions and the constraint functions are linear (of affine). In a typical formulation,

$$f_t(x_t, \xi_t) := c_t^{\mathsf{I}} x_t, \quad \mathcal{X}_1 := \{x_1 : A_1 x_1 = b_1, \ x_1 \ge 0\}, \\ \mathcal{X}_t(x_{t-1}, \xi_t) := \{x_t : B_t x_{t-1} + A_t x_t = b_t, \ x_t \ge 0\}, \ t = 2, \dots, T,$$
(3.2)

Here, $\xi_1 := (c_1, A_1, b_1)$ is known at the first stage (and hence is not random), and $\xi_t := (c_t, B_t, A_t, b_t) \in \mathbb{R}^{d_t}$, t = 2, ..., T, are data vectors¹ some (or all) elements of which may be random.

There are basically two equivalent ways to make this formulation precise. One approach is to consider decision variables $x_t = x_t(\xi_{[t]}), t = 1, ..., T$, as functions² of the data process $\xi_{[t]}$ up to time t. Such a sequence of (measurable) mappings $x_t :$ $\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_t} \to \mathbb{R}^{n_t}, t = 1, ..., T$, is called an *implementable policy* (or simply a *policy*, or a *decision rule*) (recall that ξ_1 is deterministic). An implementable policy is nonanticipative in the sense that decisions $x_t = x_t(\xi_{[t]})$ are functions of a realization of the data process available at time of decision and do not depend on future observations. An implementable policy is said to be *feasible* if it satisfies the feasibility constraints for almost every realization of the random data, i.e.,

$$\boldsymbol{x}_t(\xi_{[t]}) \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\xi_{[t-1]}), \xi_t), \ t = 2, \dots, T, \ \text{w.p.1.}$$
(3.3)

The multistage problem (3.1) can be formulated in the following equivalent form

$$\begin{array}{ll} \underset{x_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T}{\text{Min}} & \mathbb{E} \left[f_1(x_1) + f_2(\boldsymbol{x}_2(\xi_{[2]}), \xi_2) + \dots + f_T\left(\boldsymbol{x}_T(\xi_{[T]}), \xi_T \right) \right] \\ \text{s.t.} & x_1 \in \mathcal{X}_1, \ \boldsymbol{x}_t(\xi_{[t]}) \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\xi_{[t-1]}), \xi_t), \ t = 2, \dots, T. \end{array}$$
(3.4)

Note that optimization in (3.4) is performed over *implementable* and *feasible* policies $(x_1, x_2, ..., x_T)$. That is, $x_2, ..., x_T$ are *functions* of the data process, and hence are elements of appropriate functional spaces of measurable mappings (compare with discussion of section 2.3.1), while $x_1 \in \mathbb{R}^{n_1}$ is a deterministic vector. Therefore, unless the data process $\xi_1, ..., \xi_T$ has a finite number of realizations, formulation (3.4) leads to an infinite

¹If data involves matrices, then their elements can be stacked columnwise to make it a vector.

²As before we use bold notation $x_t(\cdot)$ to emphasize that this is a function rather than a deterministic vector.

dimensional optimization problem. This is a natural extension of the formulation (2.66) of the two-stage problem.

Nested formulation (3.1) allows to write the corresponding *dynamic programming* equations. That is, consider the last stage problem

$$\underset{x_T \in \mathcal{X}_T(x_{T-1},\xi_T)}{\operatorname{Min}} f_T(x_T,\xi_T).$$
(3.5)

The optimal value of this problem, denoted $Q_T(x_{T-1}, \xi_T)$, depends on the decision vector x_{T-1} and data ξ_T . At stage t = 2, ..., T - 1, we formulate³ the problem:

$$\underset{x_t \in \mathbb{R}^{n_t}}{\min} f_t(x_t, \xi_t) + \mathbb{E} \left\{ Q_{t+1}\left(x_t, \xi_{[t+1]}\right) |\xi_{[t]} \right\}$$
s.t. $x_t \in \mathcal{X}_t(x_{t-1}, \xi_t).$
(3.6)

Its optimal value depends on the decision $x_{t-1} \in \mathbb{R}^{n_{t-1}}$ at the previous stage and realization of the data process $\xi_{[t]}$, and is denoted $Q_t(x_{t-1}, \xi_{[t]})$. The idea is to calculate the *cost-to-go* functions⁴ $Q_t(x_{t-1}, \xi_{[t]})$, recursively, going backward in time. At the first stage we finally need to solve the problem:

$$\min_{x_1 \in \mathcal{X}_1} f_1(x_1) + \mathbb{E}\left[Q_2\left(x_1, \xi_2\right)\right].$$
(3.7)

The optimal value of (3.7) gives the optimal value of the corresponding multistage problem (3.1).

For t = 2, ..., T the corresponding dynamic programming equations are

$$Q_t\left(x_{t-1},\xi_{[t]}\right) = \inf_{x_t \in \mathcal{X}_t(x_{t-1},\xi_t)} \left\{ f_t(x_t,\xi_t) + \mathcal{Q}_{t+1}\left(x_t,\xi_{[t]}\right) \right\},$$
(3.8)

where

$$\mathcal{Q}_{t+1}\left(x_t,\xi_{[t]}\right) := \mathbb{E}\left\{Q_{t+1}\left(x_t,\xi_{[t+1]}\right) \left|\xi_{[t]}\right\}\right\}.$$
(3.9)

For t = T the term Q_{T+1} is omitted, and for t = 1 the set \mathcal{X}_1 depends only on ξ_1 and hence is deterministic. An implementable policy $\bar{x}_t(\xi_{[t]}), t = 1, ..., T$, is *optimal* iff \bar{x}_1 is an optimal solution of the first stage problem (3.7) and for t = 2, ..., T:

$$\bar{\boldsymbol{x}}_{t}(\xi_{[t]}) \in \operatorname*{argmin}_{x_{t} \in \mathcal{X}_{t}\left(\bar{\boldsymbol{x}}_{t-1}(\xi_{[t-1]}), \xi_{t}\right)} \left\{ f_{t}(x_{t}, \xi_{t}) + \mathcal{Q}_{t+1}\left(x_{t}, \xi_{[t]}\right) \right\}, \text{ w.p.1.}$$
(3.10)

It could happen that the optimization problem in (3.10) has more than one optimal solution. In that case $\bar{x}_t(\cdot)$ should be a *measurable* selection of the right hand side of (3.10) (see section 7.2.3 for a discussion of measurability concepts).

In the dynamic programming formulation the problem is reduced to solving a family of finite dimensional problems, indexed by t and by $\xi_{[t]}$. It can be viewed as an extension of the formulation (2.61)–(2.62) of the two-stage problem.

Remark 3. As it was pointed above, the formulations (3.1) and (3.4) are equivalent. The nested formulation (3.1) can be written in the form of dynamic programming equations

³For random variables X and Y we denote by $\mathbb{E}[X|Y]$ or $\mathbb{E}_{|Y}[X]$ the conditional expectation of X given Y. ⁴The cost-to-go functions are also called *value* functions.

(3.8)–(3.9), while formulation (3.4) leads to optimization over implementable policies. The equivalence of these two formulations is based on the tower property $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$ of the expectation operator (here *X* and *Y* are two random variables). It is more accurate to write the nested formulation (3.1) as

$$\underset{x_{1}\in\mathcal{X}_{1}}{\operatorname{Min}} f_{1}(x_{1}) + \mathbb{E}_{|\xi_{1}|} \left[\inf_{x_{2}\in\mathcal{X}_{2}(x_{1},\xi_{2})} f_{2}(x_{2},\xi_{2}) + \dots + \mathbb{E}_{|\xi_{[T-1]}|} \left[\inf_{x_{T}\in\mathcal{X}_{T}(x_{T-1},\xi_{T})} f_{T}(x_{T},\xi_{T}) \right] \right],$$
(3.11)

by emphasizing the conditional structure of this formulation. Recall that ξ_1 is deterministic, therefore the conditional expectation $\mathbb{E}_{|\xi_1}$ is the same as the corresponding unconditional expectation; we write $\mathbb{E}_{|\xi_1}$ for uniformity of the notation.

Consider now problem (3.4). We can write its objective function as

$$\mathbb{E}\left[f_{1}(x_{1}) + f_{2}(\boldsymbol{x}_{2}(\xi_{[2]}), \xi_{2}) + \ldots + f_{T}\left(\boldsymbol{x}_{T}(\xi_{[T]}), \xi_{T}\right)\right] = \\
\mathbb{E}\left[\mathbb{E}_{|\xi_{[T-1]}}\left[f_{1}(x_{1}) + f_{2}(\boldsymbol{x}_{2}(\xi_{[2]}), \xi_{2}) + \ldots + f_{T}\left(\boldsymbol{x}_{T}(\xi_{[T]}), \xi_{T}\right)\right]\right] = \\
\mathbb{E}\left[f_{1}(x_{1}) + f_{2}(\boldsymbol{x}_{2}(\xi_{[2]}), \xi_{2}) + \ldots + \mathbb{E}_{|\xi_{[T-1]}}\left[f_{T}(\boldsymbol{x}_{T}(\xi_{[T]}), \xi_{T})\right]\right].$$
(3.12)

First we perform minimization in (3.4) with respect to x_T , which in view of (3.12) can be written as

$$\begin{array}{ll}
\operatorname{Min}_{\boldsymbol{x}_{T}} & \mathbb{E}\left[f_{1}(x_{1}) + f_{2}(\boldsymbol{x}_{2}(\xi_{[2]}), \xi_{2}) + \ldots + \mathbb{E}_{|\xi_{[T-1]}}[f_{T}(\boldsymbol{x}_{T}(\xi_{[T]}), \xi_{T})]\right] \\
\operatorname{s.t.} & \boldsymbol{x}_{T}(\xi_{[T]}) \in \mathcal{X}_{T}(\boldsymbol{x}_{T-1}(\xi_{[T-1]}), \xi_{T}).
\end{array}$$
(3.13)

Consequently problem (3.4) is reduced to minimization with respect to the remaining variables $x_1, ..., x_{T-1}$ of the optimal value of the above problem (3.13); that is to the problem

$$\underset{x_{1},\boldsymbol{x}_{2},...,\boldsymbol{x}_{T-1}}{\underset{\text{s.t.}}{\text{Min}}} \quad \mathbb{E}\Big[f_{1}(x_{1}) + \dots + f_{T-1}\left(\boldsymbol{x}_{T}(\xi_{[T-1]}),\xi_{T-1}\right) + V_{T}(\boldsymbol{x}_{T-1}(\xi_{[T-1]}),\xi_{[T-1]})\Big] \\
\underset{x_{1} \in \mathcal{X}_{1}, \ \boldsymbol{x}_{t}(\xi_{[t]}) \in \mathcal{X}_{t}(\boldsymbol{x}_{t-1}(\xi_{[t-1]}),\xi_{t}), \ t = 2, \dots, T-1, \tag{3.14}$$

where $V_T(\boldsymbol{x}_{T-1}(\xi_{[T-1]}),\xi_{[T-1]})$ is the optimal value of the problem

$$\min_{\boldsymbol{x}_T} \mathbb{E}_{|\xi_{[T-1]}}[f_T(\boldsymbol{x}_T(\xi_{[T]}), \xi_T)] \text{ s.t. } \boldsymbol{x}_T(\xi_{[T]}) \in \mathcal{X}_T(\boldsymbol{x}_{T-1}(\xi_{[T-1]}), \xi_T).$$
(3.15)

By interchanging⁵ the expectation and minimization operators in (3.15) (compare with the discussion in section 2.3.1 of two stage stochastic programming) we can write

$$V_T(\boldsymbol{x}_{T-1}(\xi_{[T-1]}),\xi_{[T-1]}) = \mathcal{Q}_T(\boldsymbol{x}_{T-1}(\xi_{[T-1]}),\xi_{[T-1]}),$$

where

$$Q_T(x_{T-1},\xi_{[T-1]}) = \mathbb{E}_{|\xi_{[T-1]}}[Q_T(x_{T-1},\xi_T)]$$

with $Q_T(x_{T-1},\xi_T)$ being the cost-to-go function at stage T, i.e., $Q_T(x_{T-1},\xi_T)$ is given by the optimal value of the problem

$$\underset{x_T \in \mathbb{R}^{n_T}}{\operatorname{Min}} f_T(x_T, \xi_T) \quad \text{s.t.} \ x_T \in \mathcal{X}_T(x_{T-1}, \xi_T).$$
(3.16)

⁵In order to justify such interchangeability we need to assume that functions $x_t(\cdot)$ belong to appropriate decomposable spaces of measurable mappings, see Theorem 7.92.

Also for a given $\bar{x}_{T-1}(\xi_{[T-1]})$, the corresponding optimal solutions $x_T(\xi_{[T]})$ are given by measurable selections (compare with (3.10))

$$\bar{\boldsymbol{x}}_T(\boldsymbol{\xi}_{[T]}) \in \operatorname*{argmin}_{\boldsymbol{x}_T \in \mathcal{X}_T(\bar{\boldsymbol{x}}_{T-1}(\boldsymbol{\xi}_{[T-1]}), \boldsymbol{\xi}_T)} f_T(\boldsymbol{x}_T, \boldsymbol{\xi}_T), \text{ w.p.1.}$$
(3.17)

Next we can perform minimization in (3.14) with respect to x_{T-1} , and so on by continuing this process backward in time we can derive the dynamic programming equations (3.8)–(3.10), and can write the corresponding nested formulation (3.11).

Remark 4. Suppose that the process ξ_1, \ldots, ξ_T is *stagewise independent*. Then ξ_T is independent of $\xi_{[T-1]}$ and hence the conditional expectation in the definition (3.9) of the cost-to-go function Q_T becomes the unconditional expectation. That is, $Q_T(x_{T-1})$ does not depend on $\xi_{[T-1]}$. By induction, going backward in time, it can be shown that if the stagewise independence condition holds, then each expected value cost-to-go function $Q_t(x_{t-1})$ does not depend on realizations of the random process and the dynamic programming equations (3.8)–(3.9) can be written as

$$Q_t(x_{t-1},\xi_t) = \inf_{x_t \in \mathcal{X}_t(x_{t-1},\xi_t)} \left\{ f_t(x_t,\xi_t) + \mathcal{Q}_{t+1}(x_t) \right\},$$
(3.18)

with $Q_{t+1}(x_t) = \mathbb{E} \{Q_{t+1}(x_t, \xi_{t+1})\}$. Also the optimal policy can be defined in the recursive way

$$\bar{\boldsymbol{x}}_{t} \in \operatorname*{argmin}_{\boldsymbol{x}_{t} \in \mathcal{X}_{t}(\bar{\boldsymbol{x}}_{t-1}, \xi_{t})} \left\{ f_{t}(\boldsymbol{x}_{t}, \xi_{t}) + \mathcal{Q}_{t+1}\left(\boldsymbol{x}_{t}\right) \right\}, \text{ w.p.1},$$
(3.19)

with \bar{x}_t being a function of \bar{x}_{t-1} and ξ_t , t = 2, ..., T, and x_1 being an optimal solution of the first stage problem (3.7).

If the process ξ_1, \ldots, ξ_T is Markovian, then conditional distributions in the above equations, given $\xi_{[t]}$, are the same as the respective conditional distributions given ξ_t . In that case each cost-to-go function Q_t depends on ξ_t rather than the whole $\xi_{[t]}$ and we can write it as $Q_t (x_{t-1}, \xi_t)$. Also the expected value cost-to-go function

$$\mathcal{Q}_{t+1}\left(x_{t},\xi_{t}\right) = \mathbb{E}\left\{Q_{t+1}\left(x_{t},\xi_{t+1}\right)\left|\xi_{t}\right.\right\}$$

is a function of ξ_t rather than the whole history $\xi_{[t]}$ of the data process.

Remark 5. In some cases with stagewise dependent data it is possible to reformulate the problem to make it stagewise independent by increasing the number of decision variables. As an example, consider the linear multistage problem (3.2). Suppose only the right hand side vectors b_t are random and they can be modeled as a (first order) autoregressive process

$$b_t = \mu + \Phi b_{t-1} + \varepsilon_t. \tag{3.20}$$

Here μ and Φ are (deterministic) vector and regression matrix, respectively, and the error process ε_t , t = 1, ..., T, is stagewise independent. The corresponding feasibility constraints can be written in terms of x_t and b_t as

$$B_t x_{t-1} + A_t x_t - b_t = 0, \ x_t \ge 0, \Phi b_{t-1} - b_t + \mu + \varepsilon_t = 0.$$
(3.21)

That is, in terms of decision variables (x_t, b_t) this becomes a linear multistage stochastic programming problem governed by the stagewise independent random process $\varepsilon_1, ..., \varepsilon_T$.

3.1.2 The Linear Case

We discuss linear multistage problems in more detail. Let x_1, \ldots, x_T be decision vectors corresponding to time periods (stages) $1, \ldots, T$. Consider the following linear programming problem

We can view this problem as a multiperiod stochastic programming problem where c_1 , A_1 and b_1 are known, but some (or all) the entries of the cost vectors c_t , matrices B_t and A_t , and right hand side vectors b_t , t = 2, ..., T, are random. In the multistage setting the values (realizations) of the random data become known in the respective time periods (stages), and we have the following sequence of actions:

decision
$$(x_1)$$

observation $\xi_2 := (c_2, B_2, A_2, b_2)$
decision (x_2)
 \vdots
observation $\xi_T := (c_T, B_T, A_T, b_T)$
decision (x_T) .

Our objective is to design the decision process in such a way that the expected value of the total cost is minimized while optimal decisions are allowed to be made at *every* time period t = 1, ..., T.

Let us denote by ξ_t the data vector, realization of which becomes known at time period t. In the setting of the multiperiod problem (3.22), ξ_t is assembled from the components of c_t , B_t , A_t , b_t , some (or all) of which can be random, while the data $\xi_1 = (c_1, A_1, b_1)$ at the first stage of problem (3.22) is assumed to be known. The important condition in the above *multistage decision process* is that every decision vector x_t may depend on the information available at time t (that is, $\xi_{[t]}$), but *not* on the results of observations to be made at later stages. This differs multistage stochastic programming problems from deterministic multiperiod problems, in which all the information is assumed to be available at the beginning of the decision process.

As it was outlined in section 3.1.1, there are two equivalent ways to formulate multistage stochastic programs in a precise mathematical form. In one such formulation $x_t = x_t(\xi_{[t]}), t = 2, ..., T$, is viewed as a function of $\xi_{[t]}$, and the minimization in (3.22) is performed over appropriate functional spaces of such functions. If the number of scenarios is finite, this leads to a formulation of the linear multistage stochastic program as one large (deterministic) linear programming problem. We discuss that further in section 3.1.5. Another possible approach is to write dynamic programming equations, which in a general form we already derived in section 3.1.1. We are going to discuss this next.

Let us look at our problem from the perspective of the last stage T. At that time the values of all problem data, $\xi_{[T]}$, are already known, and the values of the earlier decision vectors, x_1, \ldots, x_{T-1} , have been chosen. Our problem is, therefore, a simple linear programming problem

$$\begin{aligned}
& \underset{x_T}{\operatorname{Min}} c_T^{\mathsf{I}} x_T \\
& \text{s.t. } B_T x_{T-1} + A_T x_T = b_T, \\
& x_T \ge 0.
\end{aligned}$$
(3.23)

The optimal value of this problem depends on the earlier decision vector $x_{T-1} \in \mathbb{R}^{n_{T-1}}$ and data $\xi_T = (c_T, B_T, A_T, b_T)$, and is denoted by $Q_T(x_{T-1}, \xi_T)$.

At stage T-1 we know x_{T-2} and $\xi_{[T-1]}$. We face, therefore, the following stochastic programming problem

$$\begin{aligned}
& \underset{x_{T-1}}{\min} c_{T-1}^{\mathsf{T}} x_{T-1} + \mathbb{E} \left[Q_T(x_{T-1}, \xi_T) \, \big| \, \xi_{[T-1]} \right] \\
& \text{s.t. } B_{T-1} x_{T-2} + A_{T-1} x_{T-1} = b_{T-1}, \\
& x_{T-1} \ge 0.
\end{aligned}$$
(3.24)

The optimal value of the above problem depends on $x_{T-2} \in \mathbb{R}^{n_{T-2}}$ and data $\xi_{[T-1]}$, and is denoted $Q_{T-1}(x_{T-2}, \xi_{[T-1]})$.

Generally, at stage t = 2, ..., T - 1, we have the problem

$$\begin{aligned}
& \min_{x_t} c_t^{\mathsf{T}} x_t + \mathbb{E} \left[Q_{t+1}(x_t, \xi_{[t+1]}) \, \big| \, \xi_{[t]} \right] \\
& \text{s.t. } B_t x_{t-1} + A_t x_t = b_t, \\
& x_t \ge 0.
\end{aligned} \tag{3.25}$$

Its optimal value, called *cost-to-go* function, is denoted $Q_t(x_{t-1}, \xi_{[t]})$.

On top of all these problems is the problem to find the first decisions, $x_1 \in \mathbb{R}^{n_1}$,

$$\begin{aligned}
& \min_{x_1} c_1^{\mathsf{I}} x_1 + \mathbb{E} \left[Q_2(x_1, \xi_2) \right] \\
& \text{s.t. } A_1 x_1 = b_1, \\
& x_1 \ge 0.
\end{aligned}$$
(3.26)

The optimal value of problem (3.26) gives the optimal value of the corresponding multistage program. Note that all subsequent stages t = 2, ..., T are absorbed in the above multistage problem into the function $Q_2(x_1, \xi_2)$ through the corresponding expected values. Note also that since ξ_1 is not random, the optimal value $Q_2(x_1, \xi_2)$ does not depend on ξ_1 . In particular, if T = 2, then (3.26) coincides with the formulation (2.1) of a two-stage linear problem. The dynamic programming equations here take the form (compare with (3.8)):

$$Q_t\left(x_{t-1},\xi_{[t]}\right) = \inf_{x_t} \left\{ c_t^\mathsf{T} x_t + \mathcal{Q}_{t+1}\left(x_t,\xi_{[t]}\right) : B_t x_{t-1} + A_t x_t = b_t, \ x_t \ge 0 \right\}, \quad (3.27)$$

where

$$\mathcal{Q}_{t+1}\left(x_t,\xi_{[t]}\right) := \mathbb{E}\left\{Q_{t+1}\left(x_t,\xi_{[t+1]}\right)\left|\xi_{[t]}\right\}\right\}.$$
(3.28)

Also an implementable policy $\bar{x}_t(\xi_{[t]}), t = 1, ..., T$, is *optimal* iff for t = 1, ..., T the condition

$$\bar{\boldsymbol{x}}_{t}(\xi_{[t]}) \in \arg\min_{x_{t}} \left\{ c_{t}^{\mathsf{T}} x_{t} + \mathcal{Q}_{t+1} \left(x_{t}, \xi_{[t]} \right) : A_{t} x_{t} = b_{t} - B_{t} \bar{\boldsymbol{x}}_{t-1}(\xi_{[t-1]}), \ x_{t} \ge 0 \right\},$$
(3.29)

holds for almost every (a.e.) realization of the random process (for t = T the term Q_{T+1} is omitted and for t = 1 the term $B_t \bar{x}_{t-1}$ is omitted).

If the process ξ_t is Markovian, then each cost-to-go function depends on ξ_t rather than $\xi_{[t]}$, and we can simply write $Q_t(x_{t-1}, \xi_t)$, t = 2, ..., T. If, moreover, the stagewise independence condition holds, then each expectation function Q_t does not depend on realizations of the random process, and we can write it as $Q_t(x_{t-1})$, t = 2, ..., T (see Remark 4 on page 67).

The *nested formulation* of the linear multistage problem can be written as follows (compare with (3.1) and (3.11)):

$$\underset{\substack{A_1x_1=b_1\\x_1\geq 0}}{\operatorname{Min}} c_1^{\mathsf{T}}x_1 + \mathbb{E} \left[\min_{\substack{B_2x_1+A_2x_2=b_2\\x_2\geq 0}} c_2^{\mathsf{T}}x_2 + \dots + \mathbb{E}_{|\xi_{[T-1]}} \left[\min_{\substack{B_Tx_{T-1}+A_Tx_T=b_T\\x_T\geq 0}} c_T^{\mathsf{T}}x_T \right] \right].$$
(3.30)

Suppose now that we deal with an underlying model with a full lower block triangular constraint matrix:

Min	$c_1^T x_1$	+	$c_2^T x_2$	+	$c_3^{T} x_3$	+		+	$c_T^T x_T$		
s.t.	$A_{11}x_1$									=	$b_1,$
	$A_{21}x_1$	+	$A_{22}x_2$							=	$b_2,$
	$A_{31}x_1$	+	$A_{32}x_2$	+	$A_{33}x_3$					=	b_3 ,
	$A_{T1}x_1$	+	$A_{T2}x_2$	+		+	$A_{T,T-1}x_{T-1}$	+	$A_{TT}x_T$	=	b_T ,
	$x_1 \ge 0,$		$x_2 \ge 0,$		$x_3 \ge 0,$				$x_T \ge 0.$		
											(3.31)

In the constraint matrix of (3.22) the respective blocks $A_{t1}, \ldots, A_{t,t-2}$ were assumed to be zeros. This allowed us to express there the optimal value Q_t of (3.25) as a function of the immediately preceding decision, x_{t-1} , rather than all earlier decisions x_1, \ldots, x_{t-1} . In the case of problem (3.31), each respective subproblem of the form (3.25) depends on the entire history of our decisions, $x_{[t-1]} := (x_1, \ldots, x_{t-1})$. It takes on the form

$$\begin{aligned}
& \underset{x_{t}}{\min} c_{t}^{\mathsf{T}} x_{t} + \mathbb{E} \left[Q_{t+1}(x_{[t]}, \xi_{[t+1]}) \, \big| \, \xi_{[t]} \right] \\
& \text{s.t. } A_{t1} x_{1} + \dots + A_{t,t-1} x_{t-1} + A_{t,t} x_{t} = b_{t}, \\
& x_{t} \ge 0.
\end{aligned}$$
(3.32)

Its optimal value (i.e., the cost-to-go function) $Q_t(x_{[t-1]}, \xi_{[t]})$ is now a function of the whole history $x_{[t-1]}$ of the decision process rather than its last decision vector x_{t-1} .

Sometimes it is convenient to convert such a lower triangular formulation into the staircase formulation from which we started our presentation. This can be accomplished by introducing additional variables r_t which summarize the relevant history of our decisions. We shall call these variables the *model state* variables (to distinguish from information states discussed before). The relations that describe the next values of the state variables as a function of the current values of these variables, current decisions and current random parameters are called *model state equations*.

For the general problem (3.31) the vectors $x_{[t]} = (x_1, \ldots, x_t)$ are sufficient model state variables. They are updated at each stage according to the state equation $x_{[t]} = (x_{[t-1]}, x_t)$ (which is linear), and the constraint in (3.32) can be formally written as

$$[A_{t1} A_{t2} \dots A_{t,t-1}] x_{[t-1]} + A_{t,t} x_t = b_t.$$

Although it looks a little awkward in this general case, in many problems it is possible to define model state variables of reasonable size. As an example let us consider the structure

Min	$c_1^T x_1$	+	$c_2^T x_2$	+	$c_3^{T} x_3$	+		+	$c_T^T x_T$		
s.t.	$A_{11}x_1$									=	$b_1,$
	$B_1 x_1$	+	$A_{22}x_2$							=	b_2 ,
	$B_1 x_1$	+	$B_2 x_2$	+	$A_{33}x_{3}$					=	b_3 ,
	$B_1 x_1$	+	$B_2 x_2$	+		+	$B_{T-1}x_{T-1}$	+	$A_{TT}x_T$	=	b_T ,
	$x_1 \ge 0,$		$x_2 \ge 0,$		$x_3 \ge 0,$				$x_T \ge 0,$		

in which all blocks A_{it} , i = t + 1, ..., T are identical and observed at time t. Then we can define the state variables r_t , t = 1, ..., T recursively by the state equation $r_t = r_{t-1} + B_t x_t$, t = 1, ..., T - 1, where $r_0 = 0$. Subproblem (3.32) simplifies substantially:

$$\begin{split} \min_{x_{t}, r_{t}} c_{t}^{\mathsf{T}} x_{t} + \mathbb{E} \left[Q_{t+1}(r_{t}, \xi_{[t+1]}) \, \big| \, \xi_{[t]} \right] \\ \text{s.t.} \, r_{t-1} + A_{t,t} x_{t} = b_{t}, \\ r_{t} = r_{t-1} + B_{t} x_{t}, \\ x_{t} > 0. \end{split}$$

Its optimal value depends on r_{t-1} and is denoted $Q_t(r_{t-1}, \xi_{[t]})$.

Let us finally remark that the simple nonnegativity constraints $x_t \ge 0$ can be replaced in our model by a general constraint $x_t \in X_t$, where X_t is a convex polyhedron defined by some linear equations and inequalities (local for stage t). The set X_t may be random, too, but has to become known at stage t.

3.1.3 Scenario Trees

In order to proceed with numerical calculations one needs to make a discretization of the underlying random process. It is useful and instructive to discuss this in detail. That is, we consider in this section the case where the random process ξ_1, \ldots, ξ_T has a finite number of

realizations. It is useful to depict the possible sequences of data in a form of *scenario tree*. It has nodes organized in levels which correspond to stages $1, \ldots, T$. At level t = 1 we have only one *root node*, and we associate with it the value of ξ_1 (which is known at stage t = 1). At level t = 2 we have as many nodes as many different realizations of ξ_2 may occur. Each of them is connected with the root node by an arc. For each node ι of level t = 2 (which corresponds to a particular realization ξ_2^{ι} of ξ_2) we create at least as many nodes at level 3 as different values of ξ_3 may follow ξ_2^{ι} , and we connect them with the node ι , etc.

Generally, nodes at level t correspond to possible values of $\xi_{[t]}$ that may occur. Each of them is connected to a unique node at level t - 1, called the *ancestor node*, which corresponds to the identical first t - 1 parts of the process $\xi_{[t]}$, and is also connected to nodes at level t + 1, called *children nodes*, which correspond to possible continuations of history $\xi_{[t]}$. Note that, in general, realizations ξ_t^{ι} are vectors and it may happen that some of the values ξ_t^{ι} , associated with nodes at a given level t, are equal to each other. Nevertheless, such equal values may be represented by different nodes, because they may correspond to different histories of the process (see Figure 3.1 in Example 3.1 of the next section).

We denote by Ω_t the set of all nodes at stage t = 1, ..., T. In particular, Ω_1 consists of the unique root node, Ω_2 has as many elements as many different realizations of $\xi_{[2]}$ may occur, etc. For a node $\iota \in \Omega_t$ we denote by $C_{\iota} \subset \Omega_{t+1}, t = 1, ..., T - 1$, the set of all children nodes of ι , and by $a(\iota) \in \Omega_{t-1}, t = 2, ..., T$, the ancestor node of ι . We have that $\Omega_{t+1} = \bigcup_{\iota \in \Omega_t} C_{\iota}$ and the sets C_{ι} are disjoint, i.e., $C_{\iota} \cap C_{\iota'} = \emptyset$ if $\iota \neq \iota'$. Note again that with different nodes at stage $t \ge 3$ may be associated the same numerical values (realizations) of the corresponding data process ξ_t . Scenario is a path from the root node at stage t = 1 to a node at the last stage T. Each scenario represents a history of the process ξ_1, \ldots, ξ_T . By construction, there is one-to-one correspondence between scenarios and the set Ω_T , i.e., $K = |\Omega_T|$.

Next, we define a probability distribution on a scenario tree. In order to deal with the nested structure of the decision process we need to specify the conditional distribution of ξ_{t+1} given $\xi_{[t]}, t = 1, ..., T - 1$. That is, if we are currently at a node $\iota \in \Omega_t$, we need to specify probability of moving from ι to a node $\eta \in C_{\iota}$. Let us denote this probability by $\rho_{\iota\eta}$. Note that $\rho_{\iota\eta} \ge 0$ and $\sum_{\eta \in C_{\iota}} \rho_{\iota\eta} = 1$. Probabilities $\rho_{\iota\eta}, \eta \in C_{\iota}$, represent conditional distribution of ξ_{t+1} given that the path of the process ξ_1, \ldots, ξ_t ended at the node ι .

Every scenario can be defined by its nodes ι_1, \ldots, ι_T , arranged in the chronological order, i.e., node ι_2 (at level t = 2) is connected to the root node, ι_3 is connected to the node ι_2 , etc. The probability of that scenario is then given by the product $\rho_{\iota_1\iota_2}\rho_{\iota_2\iota_3}\cdots\rho_{\iota_T-1\iota_T}$. That is, the conditional probabilities define a probability distribution on the set of scenarios. Conversely, it is possible to derive these conditional probabilities from scenario probabilities $p_k, k = 1, \ldots, K$, as follows. Let us denote by $S^{(\iota)}$ the set of scenarios passing through node ι (at level t) of the scenario tree, and let $p^{(\iota)} := \Pr[S^{(\iota)}]$, i.e., $p^{(\iota)}$ is the sum of probabilities of all scenarios passing through node ι . If $\iota_1, \iota_2, \ldots, \iota_t$, with ι_1 being the root node and $\iota_t = \iota$, is the history of the process up to node ι , then the probability $p^{(\iota)}$

⁶We denote by $|\Omega|$ the number of elements in a (finite) set Ω .

the product

$$p^{(\iota)} = \rho_{\iota_1 \iota_2} \rho_{\iota_2 \iota_3} \cdots \rho_{\iota_{t-1} \iota_t}$$

of the corresponding conditional probabilities. In another way we can write this in the recursive form $p^{(\iota)} = \rho_{a\iota} p^{(a)}$, where $a = a(\iota)$ is the ancestor of the node ι . This equation defines the conditional probability $\rho_{a\iota}$ from the probabilities $p^{(\iota)}$ and $p^{(a)}$. Note that if $a = a(\iota)$ is the ancestor of the node ι , then $S^{(\iota)} \subset S^{(a)}$ and hence $p^{(\iota)} \leq p^{(a)}$. Consequently if $p^{(a)} > 0$, then $\rho_{a\iota} = p^{(\iota)}/p^{(a)}$. Otherwise $S^{(a)}$ is empty, i.e., no scenario is passing through the node ι .

If the process ξ_1, \ldots, ξ_T is *stagewise independent*, then the conditional distribution of ξ_{t+1} given $\xi_{[t]}$ is the same as the unconditional distribution of $\xi_{t+1}, t = 1, \ldots, T-1$. In that case at every stage $t = 1, \ldots, T-1$, with every node $\iota \in \Omega_t$ is associated an identical set of children, with the same set of respective conditional probabilities and with the same respective numerical values.

Recall that a stochastic process Z_t , t = 1, 2, ..., that can take a finite number $\{z_1, ..., z_m\}$ of different values, is a *Markov chain* if

$$\Pr\left\{Z_{t+1} = z_j \mid Z_t = z_i, Z_{t-1} = z_{i_{t-1}}, \dots, Z_1 = z_{i_1}\right\} = \Pr\left\{Z_{t+1} = z_j \mid Z_t = z_i\right\},\$$

for all states $z_{i_{t-1}}, \ldots, z_{i_1}, z_i, z_j$ and all $t = 1, 2, \ldots$. Denote

$$p_{ij} := \Pr \left\{ Z_{t+1} = z_j \mid Z_t = z_i \right\}, \ i, j = 1, \dots, m.$$

In some situations it is natural to model the data process as a Markov chain with the corresponding state space⁷ { ζ^1, \ldots, ζ^m } and probabilities p_{ij} of moving from state ζ^i to state $\zeta^j, i, j = 1, \ldots, m$. We can model such a process by a scenario tree. At stage t = 1 there is one root node to which is assigned one of the values from the state space, say ζ^i . At stage t = 2 there are m nodes to which are assigned values ζ^1, \ldots, ζ^m with the corresponding probabilities p_{i1}, \ldots, p_{im} . At stage t = 3 there are m^2 nodes, such that each node at stage t = 2, associated with a state $\zeta^a, a = 1, \ldots, m$, is the ancestor of m nodes at stage t = 3 to which are assigned values ζ^1, \ldots, ζ^m with the corresponding conditional probabilities p_{a1}, \ldots, p_{am} . At stage t = 4 there are m^3 nodes, etc. At each stage t of such T-stage Markov chain process there are m^{t-1} nodes, the corresponding random vector (variable) ξ_t can take values ζ^1, \ldots, ζ^m with respective probabilities which can be calculated from the history of the process up to time t, and the total number of scenarios is m^{T-1} . We have here that the random vectors (variables) ξ_1, \ldots, ξ_T are independently distributed iff $p_{ij} = p_{i'j}$ for any $i, i', j = 1, \ldots, m$, i.e., the conditional probability p_{ij} of moving from state ζ^i to state ζ^j does not depend on i.

In the above formulation of the Markov chain the corresponding scenario tree represents the total history of the process with the number of scenarios growing exponentially with the number of stages. Now if we approach the problem by writing the cost-to-go functions $Q_t(x_{t-1}, \xi_t)$ of the dynamic programming equations, going backwards, then we do not need to keep track of the history of the process. That is, at every stage t the cost-to-go function $Q_t(\cdot, \xi_t)$ depends only on the current state (realization) $\xi_t = \zeta^i$, i = 1, ..., m, of the process (compare with Remark 4 on page 63). On the other hand, if we want to write the corresponding optimization problem (in the case of a finite number of scenarios)

⁷In our model, the values ζ^1, \ldots, ζ^m may be numbers or vectors.

as one large linear programming problem, we still need the scenario tree formulation. This is the basic difference between the stochastic and dynamic programming approaches to the problem. That is, the stochastic programming approach does not necessarily rely on the Markovian structure of the process considered. This makes it more general at the price of considering a possibly very large number of scenarios.

3.1.4 Filtration Interpretation

We can interpret the scenario tree approach from a somewhat different and more abstract point of view. Let us denote by Ω the set of all scenarios. As Ω is finite, each its subset can be viewed as an event. The set of all subsets of Ω forms a sigma algebra, we denote it by the symbol \mathcal{F} . Particularly important are the events that can be observed up to time t, that is, the events about which we know at time t whether they happened or not. The set of such events is denoted by \mathcal{F}_t , where $t = 1, \ldots, T$. Each \mathcal{F}_t forms a sigma algebra, and $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_T = \mathcal{F}$. The sequence of nested sigma algebras $\mathcal{F}_t, t = 1, \ldots, T$, is called a *filtration*. The corresponding scenario tree is a representation of this filtration.

We can view a node of the tree at level t as an elementary event in the subalgebra \mathcal{F}_t . At time t = 1 we have one root node corresponding to the sure event Ω . This is the only nonempty event in \mathcal{F}_1 ; the other event is the empty set \emptyset . At time t = 2 we have K_2 nodes $\iota_1, \ldots, \iota_{K_2}$; each node ι can be identified with the set of scenarios Ω_2^t passing through this node in the tree. Observe that $\Omega_2^i \cap \Omega_2^j = \emptyset$ for $i \neq j$, and that $\Omega = \Omega_2^{\iota_1} \cup \cdots \cup \Omega_2^{\iota_{K_2}}$. The sub-algebra \mathcal{F}_2 contains all possible unions of these sets and the empty set. Generally, at time $t = 1, \ldots, T - 1$ we have the set Ω_t of possible nodes of cardinality $K_t = |\Omega_t|$. Each node $\iota \in \Omega_t$ is identified with a set of scenarios $\Omega_t^\iota \subset \Omega$, which pass through node ι . The tree structure specifies the set $C_\iota \subset \Omega_{t+1}$ of children nodes of node ι . This is equivalent to saying that the set of scenarios Ω_t^ι is the union of the sets of scenarios Ω_{t+1}^η for $\eta \in C_\iota$. On the other hand, if we have a filtration \mathcal{F}_t , $t = 1, \ldots, T$, then every elementary event $\Omega_t^i \in \mathcal{F}_t$ is also a random event in \mathcal{F}_{t+1} . This means that it is a union of some elementary events $\Omega_{t+1}^j \in \mathcal{F}_{t+1}$, where j is such that $\Omega_{t+1}^j \subset \Omega_t^i$. We can thus define C_i as the set of all these j. Consequently, a finite filtration implies a tree structure.

So far no probabilistic or numerical structure was assumed for the constructed scenario tree, and it simply represents possible states of the "nature" affecting our uncertain process. In order to assign a probabilistic structure to the tree, we have to specify probabilities of the scenarios $\omega \in \Omega = \Omega_T$. This means that we specify the corresponding probability measure (distribution) P on the set Ω_T , and hence $(\Omega_T, \mathcal{F}_T, P)$ becomes a probability space. Every such probability measure implies conditional probabilities of moving from a node $\iota \in \Omega_t$ to a node $\eta \in \Omega_{t+1}$ as follows: $p_{\iota\eta} = P[\Omega_{t+1}^{\eta}]/P[\Omega_t^{\iota}]$. Conversely, conditional probabilities imply a probability measure on Ω .

A random variable $X : \Omega_T \to \mathbb{R}$ is \mathcal{F}_t -measurable iff $X(\cdot)$ is constant on every elementary event of the sub-algebra \mathcal{F}_t . Therefore there is one-to-one correspondence between the spaces of \mathcal{F}_t -measurable variables $X : \Omega_T \to \mathbb{R}$ and functions $Y : \Omega_t \to \mathbb{R}$. In particular, recall that $\mathcal{F}_1 = \{\emptyset, \Omega_T\}$, and hence $X : \Omega_T \to \mathbb{R}$ is \mathcal{F}_1 -measurable iff $X(\cdot)$ is constant on Ω_T . We can view decisions of the corresponding multistage problem as a sequence $\boldsymbol{x}_t : \Omega_T \to \mathbb{R}^{n_t}, t = 1, ..., T$, of random vectors defined on the probability space $(\Omega_T, \mathcal{F}_T, P)$. The nonanticipativity condition means here that \boldsymbol{x}_t is \mathcal{F}_t -measurable, t = 1, ..., T. That is, a sequence of mappings $\boldsymbol{x}_t : \Omega_T \to \mathbb{R}^{n_t}, t = 1, ..., T$, is an implementable policy if it is adapted to the filtration $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T$.

The corresponding multistage stochastic program (compare with formulation (3.4)) can be written here as

$$\underset{x_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T}{\underset{\text{s.t.}}{\text{Min}}} \quad \mathbb{E} \Big[f_1(x_1) + f_2(\boldsymbol{x}_2(\omega), \omega) + \dots + f_T(\boldsymbol{x}_T(\omega), \omega) \Big] \\
\underset{x_1 \in \mathcal{X}_1, \ \boldsymbol{x}_t(\omega) \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\omega), \omega), \ t = 2, \dots, T,$$
(3.33)

with the optimization performed over \mathcal{F}_t -measurable mappings $\boldsymbol{x}_t : \Omega_T \to \mathbb{R}^{n_t}, t = 1, ..., T$. Since the set Ω_T is finite, the feasibility constraints in (3.33) should be satisfied for every $\omega \in \Omega_T$. Formulation (3.33) can also be considered for a general probability space (Ω, \mathcal{F}, P) and a specified filtration $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T$ of sigma algebras, with $\mathcal{F}_1 = \{\emptyset, \Omega\}$ and $\mathcal{F}_T = \mathcal{F}$. For a general probability space the feasibility constraints should be satisfied with probability one.

Consider, for example, the linear case discussed in section 3.1.2. Suppose that the data process $\xi_t(\omega) = (c_t(\omega), B_t(\omega), A_t(\omega), b_t(\omega))$ is defined on a probability space (Ω, \mathcal{F}, P) and is adapted to a considered filtration $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T$, i.e., $\xi_t(\omega)$ is \mathcal{F}_t measurable, t = 1, ..., T. Since $\mathcal{F}_1 = \{\emptyset, \Omega\}$, this implies that ξ_1 is deterministic. Then the cost-to-go function $Q_t(x_{t-1}, \omega), t = 2, ..., T$, of the corresponding dynamic programming equations (compare with (3.25)) is given by the optimal value of the problem⁸

$$\begin{aligned}
& \underset{x_t}{\operatorname{Min}} c_t^{\mathsf{T}}(\omega) x_t + \mathbb{E} \left[Q_{t+1}(x_t, \omega) \, \middle| \, \mathcal{F}_t \right] \\
& \text{s.t. } B_t(\omega) x_{t-1} + A_t(\omega) x_t = b_t(\omega), \\
& x_t > 0.
\end{aligned}$$
(3.34)

where as before, $Q_{T+1}(\cdot, \cdot) \equiv 0$. It follows that $Q_t(x_{t-1}, \cdot)$ is a function of \mathcal{F}_t -measurable data and hence is \mathcal{F}_t -measurable.

At the first stage the following problem should be solved.

$$\begin{aligned}
& \min_{x_1} c_1^{\mathsf{T}} x_1 + \mathbb{E} \left[Q_2(x_1, \omega) \right] \\
& \text{s.t. } A_1 x_1 = b_1, \\
& x_1 \ge 0.
\end{aligned}$$
(3.35)

3.1.5 Algebraic Formulation of Nonanticipativity Constraints

Suppose that in our basic problem (3.22) there are only finitely many, say K, different scenarios the problem data can take. Recall that each scenario can be considered as a path of the respective scenario tree. With each scenario, numbered k, is associated probability p_k and the corresponding sequence of decisions⁹ $\mathbf{x}^k = (x_1^k, x_2^k, \ldots, x_T^k)$. That is, with each possible scenario $k = 1, \ldots, K$ (i.e., a realization of the data process) we associate a sequence of decisions \mathbf{x}^k . Of course, it would not be appropriate to try to find the optimal

⁸See section 7.2.2 for a definition of conditional expectation with respect to a sigma subalgebra.

⁹To avoid ugly collisions of subscripts we change our notation a little and we put the index of the scenario, k, as a superscript.

values of these decisions by solving the relaxed version of (3.22):

The reason is the same as in the two-stage case. That is, in problem (3.36) all parts of the decision vector are allowed to depend on *all* parts of the random data, while each part x_t should be allowed to depend only on the data known up to stage t. In particular, problem (3.36) may suggest different values of x_1 , one for each scenario k, while our first stage decision should be independent of possible realizations of the data process.

In order to correct this problem we enforce the constraints

$$x_1^k = x_1^\ell \text{ for all } k, \ell \in \{1, \dots, K\},$$
(3.37)

similarly to the two stage case (2.83). But this is not sufficient, in general. Consider the second part of the decision vector, x_2 . It should be allowed to depend only on $\xi_{[2]} = (\xi_1, \xi_2)$, so it has to have the same value for all scenarios k for which $\xi_{[2]}^k$ are identical. We must, therefore, enforce the constraints

$$x_2^k = x_2^\ell$$
 for all k,ℓ for which $\xi_{[2]}^k = \xi_{[2]}^\ell$

Generally, at stage t = 1, ..., T, the scenarios that have the same history $\xi_{[t]}$ cannot be distinguished, so we need to enforce the *nonanticipativity constraints*:

$$x_t^k = x_t^\ell \text{ for all } k, \ell \text{ for which } \xi_{[t]}^k = \xi_{[t]}^\ell, \quad t = 1, \dots, T.$$
 (3.38)

Problem (3.36) together with the nonanticipativity constraints (3.38) becomes equivalent to our original formulation (3.22).

Remark 6. Let us observe that if in problem (3.36) only the constraints (3.37) are enforced, then from the mathematical point of view the problem obtained becomes a two-stage stochastic linear program with K scenarios. In this two-stage program the first stage decision vector is x_1 , the second stage decision vector is (x_2, \ldots, x_K) , the technology matrix is B_2 and the recourse matrix is the block matrix

$$\begin{bmatrix} A_2 & 0 & \dots & 0 & 0 \\ B_3 & A_3 & \dots & 0 & 0 \\ & & & & \\ 0 & 0 & \dots & B_T & A_T \end{bmatrix}$$

v

Since the two-stage problem obtained is a relaxation of the multistage problem (3.22), its optimal value gives a lower bound for the optimal value of problem (3.22) and in that sense it may be useful. Note, however, that this model does not make much sense in any application, because it assumes that at the end of the process, when all realizations of the random data become known, one can go back in time and make all decisions x_2, \ldots, x_{K-1} .

Example 3.1 (Scenario Tree) As it was discussed in section 3.1.3 it can be useful to depict the possible sequences of data $\xi_{[t]}$ in a form of a *scenario tree*. An example of such scenario tree is given in Figure 3.1. Numbers along the arcs represent conditional probabilities of moving from one node to the next. The associated process $\xi_t = (c_t, B_t, A_t, b_t), t =$ $1, \ldots, T$, with T = 4, is defined as follows. All involved variables are assumed to be one dimensional, with $c_t, B_t, A_t, t = 2, 3, 4$, being fixed and only right hand side variables b_t being random. The values (realizations) of the random process b_1, \ldots, b_T are indicated by the bold numbers at the nodes of the tree. (The numerical values of c_t, B_t, A_t are not written explicitly, although, of course, they also should be specified.) That is, at level $t = 1, b_1$ has the value 36. At level $t = 2, b_2$ has two values 15 and 50 with respective probabilities 0.4 and 0.6. At level t = 3 we have 5 nodes with which are associated the following numerical values (from left to right) 10, 20, 12, 20, 70. That is, b_3 can take 4 different values with respective probabilities $Pr\{b_3 = 10\} = 0.4 \cdot 0.1$, $Pr\{b_3 = 20\} =$ $0.4 \cdot 0.4 + 0.6 \cdot 0.4$, $\Pr\{b_3 = 12\} = 0.4 \cdot 0.5$ and $\Pr\{b_3 = 70\} = 0.6 \cdot 0.6$. At level t = 4, the numerical values associated with 8 nodes are defined, from left to right, as 10, 10, 30, 12, 10, 20, 40, 70. The respective probabilities can be calculated by using the corresponding conditional probabilities. For example,

$$\Pr\{b_4 = 10\} = 0.4 \cdot 0.1 \cdot 1.0 + 0.4 \cdot 0.4 \cdot 0.5 + 0.6 \cdot 0.4 \cdot 0.4$$

Note that although some of the realizations of b_3 , and hence of ξ_3 , are equal to each other, they are represented by different nodes. This is necessary in order to identify different histories of the process corresponding to different scenarios. The same remark applies to b_4 and ξ_4 . Altogether, there are eight scenarios in this tree. Figure 3.2 illustrates the way in which sequences of decisions are associated with scenarios from Figure 3.1.

The process b_t (and hence the process ξ_t) in this example is not Markovian. For instance,

$$\Pr \{b_4 = 10 \mid b_3 = 20, b_2 = 15, b_1 = 36\} = 0.5,$$

while

$$\Pr\{b_4 = 10 \mid b_3 = 20\} = \frac{\Pr\{b_4 = 10, b_3 = 20\}}{\Pr\{b_3 = 20\}}$$
$$= \frac{0.5 \cdot 0.4 \cdot 0.4 + 0.4 \cdot 0.4 \cdot 0.6}{0.4 \cdot 0.4 + 0.4 \cdot 0.6} = 0.44.$$
$$b_4 = 10 \mid b_3 = 20\} \neq \Pr\{b_4 = 10 \mid b_3 = 20, b_2 = 15, b_1 = 36\}.$$

That is, $\Pr\{b_4 = 10 \mid b_3 = 20\} \neq \Pr\{b_4 = 10 \mid b_3 = 20, b_2 = 15, b_1 = 36\}.$

Relaxing the nonanticipativity constraints means that decisions $x_t = x_t(\omega)$ are viewed as functions of all possible realizations (scenarios) of the data process. This was the case in formulation (3.36) where the problem was separated into K different problems, one for each scenario $\omega_k = (\xi_1^k, \dots, \xi_T^k), k = 1, \dots, K$. There are several ways how the



Figure 3.1. Scenario tree. Nodes represent information states. Paths from the root to leaves represent scenarios. Numbers along the arcs represent conditional probabilities of moving to the next node. Bold numbers represent numerical values of the process.



Figure 3.2. Sequences of decisions for scenarios from Figure 3.1. Horizontal dotted lines represent the equations of nonanticipativity.

corresponding nonanticipativity constraints can be written. One possible way is to write them, similarly to (2.84) for two stage models, as

$$x_t = \mathbb{E}\left[x_t | \xi_{[t]}\right], \quad t = 1, \dots, T.$$
(3.39)

Another way is to use *filtration* associated with the data process. Let \mathcal{F}_t be the sigma algebra generated by $\xi_{[t]}, t = 1, ..., T$. That is, \mathcal{F}_t is the minimal subalgebra of the sigma algebra \mathcal{F} such that $\xi_1(\omega), ..., \xi_t(\omega)$ are \mathcal{F}_t -measurable, denoted $\mathcal{F}_t := \sigma(\xi_1, ..., \xi_t)$. Since ξ_1 is not random, \mathcal{F}_1 contains only two sets: \emptyset and Ω . We have that $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_T \subset \mathcal{F}$. We discussed construction of such a filtration in section 3.1.4. We can write

(3.39) in the following equivalent form¹⁰

$$x_t = \mathbb{E}\left[x_t \middle| \mathcal{F}_t\right], \quad t = 1, \dots, T.$$
(3.40)

Condition (3.40) holds iff $x_t(\omega)$ is measurable with respect \mathcal{F}_t , t = 1, ..., T. As it was discussed in section 3.1.4, one can use this measurability requirement as a definition of the nonanticipativity constraints.

Suppose, for the sake of simplicity, that there is a finite number K of scenarios. To each scenario corresponds a sequence (x_1^k, \ldots, x_T^k) of decision vectors which can be considered as an element of a vector space of dimension $n_1 + \cdots + n_T$. The space of all such sequences (x_1^k, \ldots, x_T^k) , $k = 1, \ldots, K$, is a vector space, denoted \mathfrak{X} , of dimension $(n_1 + \cdots + n_T)K$. The nonanticipativity constraints (3.38) define a linear subspace of \mathfrak{X} , denoted \mathfrak{L} . Define the following scalar product on the space \mathfrak{X} ,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{k=1}^{K} \sum_{t=1}^{T} p_k (x_t^k)^{\mathsf{T}} y_t^k, \qquad (3.41)$$

and let P be the orthogonal projection of \mathfrak{X} onto \mathfrak{L} with respect to this scalar product. Then

$$x = Px$$

is yet another way to write the nonanticipativity constraints.

A computationally convenient way of writing the nonanticipativity constraints (3.38) can be derived by using the following construction, which extends to the multistage case the system (2.87).

Let Ω_t be the set of nodes at level t. For a node $\iota \in \Omega_t$ we denote by $S^{(\iota)}$ the set of scenarios that pass through node ι and are, therefore, indistinguishable on the basis of the information available up to time t. As explained before, the sets $S^{(\iota)}$ for all $\iota \in \Omega_t$ are the atoms of the sigma-subalgebra \mathcal{F}_t associated with the time stage t. We order them and denote them by $S_t^1, \ldots, S_t^{\gamma_t}$.

Let us assume that all scenarios $1, \ldots, K$ are ordered in such a way that each set S_t^{ν} is a set of consecutive numbers $l_t^{\nu}, l_t^{\nu} + 1, \ldots, r_t^{\nu}$. Then nonanticipativity can be expressed by the system of equations

$$x_t^s - x_t^{s+1} = 0, \quad s = l_t^{\nu}, \dots, r_t^{\nu} - 1, \quad t = 1, \dots, T - 1, \quad \nu = 1, \dots, \gamma_t.$$
 (3.42)

In other words, each decision is related to its neighbors from the left and from the right, if they correspond to the same node of the scenario tree.

The coefficients of constraints (3.42) define a giant matrix

$$M = [M^1 \dots M^K],$$

whose rows have two nonzeros each: 1 and -1. Thus, we obtain an algebraic description of the nonanticipativity constraints:

$$M^{1}x^{1} + \dots + M^{K}x^{K} = 0. ag{3.43}$$

Owing to the sparsity of the matrix M, this formulation is very convenient for various numerical methods for solving linear multistage stochastic programming problems: the simplex method, interior point methods and decomposition methods.

¹⁰See section 7.2.2 for a definition of conditional expectation with respect to a sigma subalgebra.

Figure 3.3. The nonanticipativity constraint matrix M corresponding to the scenario tree from Figure 3.1. The subdivision corresponds to the scenario submatrices M^1, \ldots, M^8 .

Example 3.2 Consider the scenario tree depicted in Figure 3.1. Let us assume that the scenarios are numbered from the left to the right. Our nonanticipativity constraints take on the form:

$$\begin{aligned} x_1^1 - x_1^2 &= 0, \quad x_1^2 - x_1^3 &= 0, \quad \dots \quad , \quad x_1^7 - x_1^8 &= 0, \\ x_2^1 - x_2^2 &= 0, \quad x_2^2 - x_2^3 &= 0, \quad x_2^3 - x_2^4 &= 0, \\ x_2^5 - x_2^6 &= 0, \quad x_2^6 - x_2^7 &= 0, \quad x_2^7 - x_2^8 &= 0, \\ x_3^2 - x_3^3 &= 0, \quad x_3^3 - x_4^3 &= 0, \quad x_3^6 - x_3^7 &= 0. \end{aligned}$$

Using *I* to denote the identity matrix of an appropriate dimension, we may write the constraint matrix *M* as shown in Figure 3.3. *M* is always a very sparse matrix: each row of it has only two nonzeros, each column at most two nonzeros. Moreover, all nonzeros are either 1 or -1, which is also convenient for numerical methods.

3.1.6 Piecewise Affine Policies

In some cases it is possible to give a qualitative characterization of optimal policies (decision rules). We already discussed such a situation in the inventory model in section 1.2.3, where we showed, under the assumption of stagewise independence, that the basestock policy is optimal. In this section we discuss forms of optimal policies of linear multistage stochastic problems.

Two stage problems. Let us start with the two stage problem, discussed in section 2.1.1:

$$\underset{x \in \mathcal{X}}{\operatorname{Min}} c^{\mathsf{T}} x + \mathbb{E}[Q(x,\xi)], \qquad (3.44)$$

where $\mathcal{X} := \{x : Ax = b, x \ge 0\}$ and $Q(x, \xi)$ is the optimal value of the second stage problem

$$\underset{y \in \mathbb{R}^m}{\operatorname{Min}} q^{\mathsf{T}} y \quad \text{s.t.} \ T x + W y = h, \ y \ge 0.$$
(3.45)

Denoting u := h - Tx, we can write the second stage problem (3.45) as

$$\underset{y \in \mathbb{R}^m}{\operatorname{Min}} q^{\mathsf{T}} y \quad \text{s.t. } Wy = u, \ y \ge 0,$$
(3.46)

and its dual (compare with (2.3))

$$\max_{\pi} u^{\mathsf{T}} \pi \quad \text{s.t. } W^{\mathsf{T}} \pi \le q. \tag{3.47}$$

We make the following assumptions.

- (i) The $m \times 1$ vector q and $\ell \times m$ matrix W are deterministic (not random), i.e., the recourse is *fixed*, and matrix W has rank ℓ , i.e., rows of W are linearly independent.
- (ii) The feasible set $\{\pi : W^{\mathsf{T}}\pi \leq q\}$ of the dual problem is nonempty.
- (iii) The feasible set of the second stage problem (3.45) is nonempty for every $x \in \mathcal{X}$ and $\xi \in \Xi$, where $\Xi \subset \mathbb{R}^d$ is the support of the distribution of the random data vector $\xi = (h, T)$, i.e., the recourse is *relatively complete*.

It follows from condition (ii) that $Q(x,\xi) > -\infty$ for all x and ξ , and from condition (iii) that $Q(x,\xi) < +\infty$ for all $x \in \mathcal{X}$ and $\xi \in \Xi$; the function $Q(x,\xi)$ is finite valued and the linear programming problem (3.45) has an optimal solution for all $x \in \mathcal{X}$ and $\xi \in \Xi$. Moreover, the linear programming problem (3.46) has an optimal solution which is an extreme point of its feasible set $\{y : Wy = u, y \ge 0\}$. It is possible to choose the optimal solutions in such a way that they form a continuous piecewise linear function of u. This is a known result in the theory of parametric linear programming; let us quickly outline the arguments.

For an index set $\mathcal{I} \subset \{1, ..., m\}$ denote by $W_{\mathcal{I}}$ the submatrix of W formed by columns of W indexed by \mathcal{I} , and by $y_{\mathcal{I}}$ the subvector of y formed by components y_i , $i \in \mathcal{I}$. By a standard result of linear programming we have that a feasible point y is an extreme point (basic solution) of the feasible set iff there exists an index set $\mathcal{I} \subset \{1, ..., m\}$ of cardinality ℓ such that the matrix $W_{\mathcal{I}}$ is nonsingular and $y_i = 0$ for $i \in \{1, ..., m\} \setminus \mathcal{I}$. Consider the optimal set of problem (3.46), denoted $\mathfrak{O}(u)$. It follows that if the optimal set $\mathfrak{O}(u)$ is nonempty, then it contains an extreme point $\bar{y} = \bar{y}(u)$ and there is an index set $\mathcal{I} = \mathcal{I}(u)$ such that $W_{\mathcal{I}}\bar{y}_{\mathcal{I}} = u$ and $\bar{y}_i = 0$ for $i \in \{1, ..., m\} \setminus \mathcal{I}$. This can be written as $\bar{y}(u) = R_{\mathcal{I}}u$, where $R_{\mathcal{I}}$ is $m \times \ell$ matrix with the rows $[R_{\mathcal{I}}]_i = [W_{\mathcal{I}}^{-1}]_i$ for $i \in \mathcal{I}$, and $[R_{\mathcal{I}}]_i = 0$ for $i \in \{1, ..., m\} \setminus \mathcal{I}$.

We have that y and π are optimal solutions of problems (3.46) and (3.47), respectively, iff these points are feasible and the complementarity condition holds, i.e.,

$$Wy = u, y \ge 0, W^{\mathsf{T}}\pi - q \le 0, y^{\mathsf{T}}(W^{\mathsf{T}}\pi - q) = 0.$$
 (3.48)

Thus for a given optimal solution $\bar{\pi} = \bar{\pi}(u)$ of the dual problem, the set $\mathcal{D}(u)$ is defined by a finite number of linear constraints. We can choose a finite number of optimal solutions of the dual problem (3.47) corresponding to different values of u. It follows by Hoffman's lemma (Theorem 7.12) that the multifunction $\mathcal{D}(\cdot)$ is Lipschitz continuous in the following sense: there is a constant $\kappa > 0$, depending only on matrix W, such that

$$\operatorname{dist}(y, \mathfrak{O}(u')) \leq \kappa \|u - u'\|, \ \forall u, u' \in \operatorname{dom} \mathfrak{O}, \ \forall y \in \mathfrak{O}(u).$$

In other words $\mathfrak{O}(\cdot)$ is Lipschitz continuous on its domain with respect to the Hausdorff metric. If for some \bar{u} the optimal set $\mathfrak{O}(\bar{u}) = \{\bar{y}\}$ is a singleton, then $\bar{y}(u)$ is continuous at \bar{u} for any choice of $\bar{y}(u) \in \mathfrak{O}(u)$. In any case it follows that it is possible to choose an extreme point $\bar{y}(u) \in \mathfrak{O}(u)$ such that $\bar{y}(u)$ is continuous on dom \mathfrak{O} . Indeed, it is possible to choose a vector $a \in \mathbb{R}^m$ such that $a^T y$ has unique minimizer over $y \in \mathfrak{O}(u)$ for all $u \in \operatorname{dom} \mathfrak{O}$. This minimizer is an extreme point of $\mathfrak{O}(u)$ and is continuous in $u \in \operatorname{dom} \mathfrak{O}$.

It follows that we can choose a continuous policy $\bar{y}(u)$, $u \in \text{dom }\mathfrak{O}$, and a finite collection \mathfrak{I} of index sets $\mathcal{I} \subset \{1, ..., m\}$, of cardinality ℓ , such that $\bar{y}(u) = R_{\mathcal{I}}u$ for some $\mathcal{I} = \mathcal{I}(u) \in \mathfrak{I}$. At an optimal solution \bar{x} of the first stage problem (3.44), setting $u = h - T\bar{x}$, we have a continuous optimal function of $\xi = (h, T)$ consisting of a finite number of linear pieces: $\bar{y}(\xi) = R_{\mathcal{I}}(h - T\bar{x}), \mathcal{I} \in \mathfrak{I}$.

We say that a policy $y(\xi)$, $\xi \in \Xi$, is *piecewise linear (piecewise affine)* if $y(\cdot)$ is continuous on Ξ and there is a finite number of linear (affine¹¹) functions $\psi_{\mathcal{I}} : \mathbb{R}^d \to \mathbb{R}^m$, $\mathcal{I} \in \mathfrak{I}$, such that for every $\xi \in \Xi$, $\bar{y}(\xi) = \psi_{\mathcal{I}}(\xi)$ for some $\mathcal{I} \in \mathfrak{I}$. By the above discussion we have the following result.

Proposition 3.3. Suppose that the assumptions (i)–(iii) hold and the first stage problem (3.44) has an optimal solution \bar{x} . Then the two stage problem (3.44)–(3.45) possesses a piecewise linear optimal policy.

Note that if only the right hand side vector h is random, while q, W and T are deterministic, then the corresponding policy $\bar{y}(h) = R_{\mathcal{I}}(h - T\bar{x}), \mathcal{I} \in \mathfrak{I}$, is a piecewise affine rather than a piecewise linear function of h.

Multistage problems. Let us consider now the multistage linear problem (3.30) and the corresponding dynamic programming equations (3.27)–(3.28). We say that a policy $x_t(\xi_{[t]}), t = 1, ..., T$, of the problem (3.30), is *piecewise affine* if each $x_t(\cdot)$ is a piecewise affine function of the data.

Proposition 3.4. Suppose that: (i) the optimal value of problem (3.30) is finite, (ii) only the right hand sides $b_2, ..., b_T$ are random, (iii) the random process $b_2, ..., b_T$ is stagewise independent, (iv) the number of scenarios is finite. Then the multistage linear problem (3.30) possesses a piecewise affine optimal policy.

Proof. The proof proceeds in a way similar to the arguments for the two stage case. Since the number of scenarios is finite, problem (3.30) can be written as one large linear program-

¹¹Recall that a function (mapping) $\psi : \mathbb{R}^d \to \mathbb{R}^m$ is said to be *affine* if it can be written as a linear function plus a constant vector.

ming problem. Since its optimal value is finite, it has an optimal solution, i.e., it possesses an optimal policy. Because of the stagewise independence condition (iii), the expectation cost-to-go functions $Q_{t+1}(x_t)$ do not depend on the data process (see Remark 4 on page 63), and because of (iv) they are piecewise linear. Consequently, problem (3.30) has an optimal policy $\bar{x}_1, \bar{x}_2, ..., \bar{x}_T$ with \bar{x}_1 being an optimal solution of the first stage problem and $\bar{x}_t, t = 2, ..., T$, being an optimal solution of the problem

$$\min_{x_t \in \mathbb{R}^{n_t}} c^{\mathsf{T}} x_t + \mathcal{Q}_{t+1}(x_t) \quad \text{s.t.} \ A_t x_t = b_t - B_t \bar{x}_{t-1}, \ x_t \ge 0,$$
(3.49)

and can be considered as a function of \bar{x}_{t-1} and b_t . Since the function $\mathcal{Q}_{t+1}(\cdot)$ is piecewise linear and convex, problem (3.49) can be written as a linear programming problem. Hence similar to the arguments for two stage problems, \bar{x}_t can be taken to be a continuous piecewise linear function of $b_t - B_t \bar{x}_{t-1}$, and hence is a piecewise affine function of \bar{x}_{t-1} and b_t . This completes the proof. \Box

Compared to the case of two stage problems, the above analysis requires the additional condition of finite number of scenarios. This is needed in order to ensure that the expectation cost-to-go functions $Q_{t+1}(x_t)$ are piecewise linear.

3.2 Duality

3.2.1 Convex Multistage Problems

In this section we consider multistage problems of the form (3.1) with

$$\mathcal{X}_t(x_{t-1},\xi_t) := \{x_t : B_t x_{t-1} + A_t x_t = b_t\}, \ t = 2, \dots, T,$$
(3.50)

 $\mathcal{X}_1 := \{x_1 : A_1x_1 = b_1\}$ and $f_t(x_t, \xi_t), t = 1, \dots, T$, being random lower semicontinuous functions. We assume that functions $f_t(\cdot, \xi_t)$ are *convex* for a.e. ξ_t . In particular, if

$$f_t(x_t, \xi_t) := \begin{cases} c_t^\mathsf{T} x_t, & \text{if } x_t \ge 0, \\ +\infty, & \text{otherwise,} \end{cases}$$
(3.51)

then the problem becomes the linear multistage problem given in the nested formulation (3.30). All constraints involving only variables and quantities associated with stage t are absorbed in the definition of the functions f_t . It is implicitly assumed that the data $(A_t, B_t, b_t) = (A_t(\xi_t), B_t(\xi_t), b_t(\xi_t)), t = 1, ..., T$, form a random process.

Dynamic programming equations take here the form:

$$Q_t \left(x_{t-1}, \xi_{[t]} \right) = \inf_{x_t} \left\{ f_t(x_t, \xi_t) + \mathcal{Q}_{t+1} \left(x_t, \xi_{[t]} \right) : B_t x_{t-1} + A_t x_t = b_t \right\},$$
(3.52)

where

$$\mathcal{Q}_{t+1}\left(x_t,\xi_{[t]}\right) := \mathbb{E}\left\{Q_{t+1}\left(x_t,\xi_{[t+1]}\right)\left|\xi_{[t]}\right\}\right\}$$

For every t = 1, ..., T, and $\xi_{[t]}$, the function $Q_t(\cdot, \xi_{[t]})$ is convex. Indeed,

$$Q_T(x_{T-1},\xi_T) = \inf_{x_T} \phi(x_T, x_{T-1},\xi_T),$$

where

$$\phi(x_T, x_{T-1}, \xi_T) := \begin{cases} f_T(x_T, \xi_T), & \text{if } B_T x_{T-1} + A_T x_T = b_T, \\ +\infty, & \text{otherwise.} \end{cases}$$

It follows from the convexity of $f_T(\cdot, \xi_T)$, that $\phi(\cdot, \cdot, \xi_T)$ is convex, and hence the optimal value function $Q_T(\cdot, \xi_T)$ is also convex. Convexity of functions $Q_t(\cdot, \xi_{[t]})$ can be shown in the same way by induction in t = T, ..., 1. Moreover, if the number of scenarios is finite and functions $f_t(x_t, \xi_t)$ are random *polyhedral*, then the cost-to-go functions $Q_t(x_{t-1}, \xi_{[t]})$ are also random polyhedral.

3.2.2 Optimality Conditions

Consider the cost-to-go functions $Q_t(x_{t-1}, \xi_{[t]})$ defined by the dynamic programming equations (3.52). With the optimization problem on the right hand side of (3.52) is associated the following Lagrangian,

$$L_t(x_t, \pi_t) := f_t(x_t, \xi_t) + \mathcal{Q}_{t+1}(x_t, \xi_{[t]}) + \pi_t^{\mathsf{T}}(b_t - B_t x_{t-1} - A_t x_t).$$

This Lagrangian also depends on $\xi_{[t]}$ and x_{t-1} which we omit for brevity of the notation. Denote

$$\psi_t(x_t, \xi_{[t]}) := f_t(x_t, \xi_t) + \mathcal{Q}_{t+1}(x_t, \xi_{[t]})$$

The dual functional is

$$D_t(\pi_t) := \inf_{x_t} L_t(x_t, \pi_t)$$

= $-\sup_{x_t} \left\{ \pi_t^{\mathsf{T}} A_t x_t - \psi_t(x_t, \xi_{[t]}) \right\} + \pi_t^{\mathsf{T}} (b_t - B_t x_{t-1})$
= $-\psi_t^* (A_t^{\mathsf{T}} \pi_t, \xi_{[t]}) + \pi_t^{\mathsf{T}} (b_t - B_t x_{t-1}),$

where $\psi_t^*(\cdot, \xi_{[t]})$ is the conjugate function of $\psi_t(\cdot, \xi_{[t]})$. Therefore we can write the Lagrangian dual of the optimization problem on the right hand side of (3.52) as follows

$$\max_{\pi_t} \left\{ -\psi_t^* (A_t^{\mathsf{T}} \pi_t, \xi_{[t]}) + \pi_t^{\mathsf{T}} (b_t - B_t x_{t-1}) \right\}.$$
(3.53)

Both optimization problems, (3.52) and its dual (3.53), are convex. Under various regularity conditions there is no duality gap between problems (3.52) and (3.53). In particular, we can formulate the following two conditions.

- (D1) The functions $f_t(x_t, \xi_t)$, t = 1, ..., T, are random polyhedral and the number of scenarios is finite.
- (D2) For all sufficiently small perturbations of the vector b_t the corresponding optimal value $Q_t(x_{t-1}, \xi_{[t]})$ is finite, i.e., there is a neighborhood of b_t such that for any b'_t in that neighborhood the optimal value of the right hand side of (3.52) with b_t replaced by b'_t is finite.

We denote by $\mathfrak{D}_t(x_{t-1},\xi_{[t]})$ the set of optimal solutions of the dual problem (3.53). All subdifferentials in the subsequent formulas are taken with respect to x_t for an appropriate $t = 1, \ldots, T$.

Proposition 3.5. Suppose that either condition (D1) holds and $Q_t(x_{t-1}, \xi_{[t]})$ is finite, or condition (D2) holds. Then:

(i) there is no duality gap between problems (3.52) and (3.53), i.e.,

$$Q_t\left(x_{t-1},\xi_{[t]}\right) = \sup_{\pi_t} \left\{ -\psi_t^*(A_t^\mathsf{T}\pi_t,\xi_{[t]}) + \pi_t^\mathsf{T}(b_t - B_t x_{t-1}) \right\},\tag{3.54}$$

(ii) \bar{x}_t is an optimal solution of (3.52) iff there exists $\bar{\pi}_t = \bar{\pi}_t(\xi_{[t]})$ such that $\bar{\pi}_t \in \mathfrak{D}(x_{t-1}, \xi_{[t]})$ and

$$0 \in \partial L_t \left(\bar{x}_t, \bar{\pi}_t \right), \tag{3.55}$$

(iii) the function $Q_t(\cdot, \xi_{[t]})$ is subdifferentiable at x_{t-1} and

$$\partial Q_t \left(x_{t-1}, \xi_{[t]} \right) = -B_t^\mathsf{T} \mathfrak{D}_t \left(x_{t-1}, \xi_{[t]} \right).$$
(3.56)

Proof. Consider the optimal value function

$$\vartheta(y) := \inf_{x_t} \big\{ \psi_t(x_t, \xi_{[t]}) : A_t x_t = y \big\}.$$

Since $\psi_t(\cdot, \xi_{[t]})$ is convex, the function $\vartheta(\cdot)$ is also convex. Condition (D2) means that $\vartheta(y)$ is finite valued for all y in a neighborhood of $\bar{y} := b_t - B_t x_{t-1}$. It follows that $\vartheta(\cdot)$ is continuous and subdifferentiable at \bar{y} . By conjugate duality (see Theorem 7.8) this implies assertion (i). Moreover, the set of optimal solutions of the corresponding dual problem coincides with the subdifferential of $\vartheta(\cdot)$ at \bar{y} . Formula (3.56) then follows by the chain rule. Condition (3.55) means that \bar{x}_t is a minimizer of $L(\cdot, \bar{\pi}_t)$, and hence the assertion (ii) follows by (i).

If condition (D1) holds, then the functions $Q_t(\cdot, \xi_{[t]})$ are polyhedral, and hence $\vartheta(\cdot)$ is polyhedral. It follows that $\vartheta(\cdot)$ is lower semicontinuous and subdifferentiable at any point where it is finite valued. Again, the proof can be completed by applying the conjugate duality theory. \Box

Note that condition (D2) actually implies that the set $\mathfrak{D}_t(x_{t-1}, \xi_{[t]})$ of optimal solutions of the dual problem is nonempty and *bounded*, while condition (D1) only implies that $\mathfrak{D}_t(x_{t-1}, \xi_{[t]})$ is nonempty.

Now let us look at the optimality conditions (3.10), which in the present case can be written as follows:

$$\bar{\boldsymbol{x}}_{t}(\xi_{[t]}) \in \arg\min_{\boldsymbol{x}_{t}} \left\{ f_{t}(\boldsymbol{x}_{t},\xi_{t}) + \mathcal{Q}_{t+1}\left(\boldsymbol{x}_{t},\xi_{[t]}\right) : A_{t}\boldsymbol{x}_{t} = b_{t} - B_{t}\bar{\boldsymbol{x}}_{t-1}(\xi_{[t-1]}) \right\}, \quad (3.57)$$

Since the optimization problem on the right hand side of (3.57) is convex, subject to linear constraints, we have that a feasible policy is optimal iff it satisfies the following optimality conditions: for t = 1, ..., T and a.e. $\xi_{[t]}$ there exists $\bar{\pi}_t(\xi_{[t]})$ such that the following condition holds

$$0 \in \partial \left[f_t(\bar{\boldsymbol{x}}_t(\xi_{[t]}), \xi_t) + \mathcal{Q}_{t+1}\left(\bar{\boldsymbol{x}}_t(\xi_{[t]}), \xi_{[t]} \right) \right] - A_t^{\mathsf{T}} \bar{\boldsymbol{\pi}}_t(\xi_{[t]}).$$
(3.58)

Recall that all subdifferentials are taken with respect to x_t , and for t = T the term Q_{T+1} is omitted.

We shall use the following regularity condition.

(D3) For t = 2, ..., T and a.e. $\xi_{[t]}$ the function $Q_t(\cdot, \xi_{[t-1]})$ is finite valued.

The above condition implies, of course, that $Q_t(\cdot, \xi_{[t]})$ is finite valued for a.e. $\xi_{[t]}$ conditional on $\xi_{[t-1]}$, which in turn implies relatively complete recourse. Note also that condition (D3) does not necessarily imply condition (D2), because in the latter the function $Q_t(\cdot, \xi_{[t]})$ is required to be finite for all small perturbations of b_t .

Proposition 3.6. Suppose that either conditions (D2) and (D3) or condition (D1) are satisfied. Then a feasible policy $\bar{x}_t(\xi_{[t]})$ is optimal iff there exist mappings $\bar{\pi}_t(\xi_{[t]})$, t = 1, ..., T, such that the following condition holds true:

$$0 \in \partial f_t(\bar{x}_t(\xi_{[t]}), \xi_t) - A_t^{\mathsf{T}} \bar{\pi}_t(\xi_{[t]}) + \mathbb{E} \left[\partial Q_{t+1} \left(\bar{x}_t(\xi_{[t]}), \xi_{[t+1]} \right) \big| \xi_{[t]} \right]$$
(3.59)

for a.e. $\xi_{[t]}$ and t = 1, ..., T. Moreover, multipliers $\bar{\pi}_t(\xi_{[t]})$ satisfy (3.59) iff for a.e. $\xi_{[t]}$ it holds that

$$\bar{\pi}_t(\xi_{[t]}) \in \mathfrak{D}(\bar{x}_{t-1}(\xi_{[t-1]}), \xi_{[t]}). \tag{3.60}$$

Proof. Suppose that condition (D3) holds. Then by the Moreau–Rockafellar Theorem (Theorem 7.4) we have that at $\bar{x}_t = \bar{x}_t(\xi_{[t]})$,

$$\partial \left[f_t(\bar{x}_t, \xi_t) + \mathcal{Q}_{t+1}\left(\bar{x}_t, \xi_{[t]} \right) \right] = \partial f_t(\bar{x}_t, \xi_t) + \partial \mathcal{Q}_{t+1}\left(\bar{x}_t, \xi_{[t]} \right).$$

Also by Theorem 7.52 the subdifferential of $Q_{t+1}(\bar{x}_t, \xi_{[t]})$ can be taken inside the expectation to obtain the last term in the right hand side of (3.59). Note that conditional on $\xi_{[t]}$ the term $\bar{x}_t = \bar{x}_t(\xi_{[t]})$ is fixed. Optimality conditions (3.59) then follow from (3.58). Suppose, further, that condition (D2) holds. Then there is no duality gap between problems (3.52) and (3.53), and the second assertion follows by (3.57) and Proposition 3.5(ii).

If condition (D1) holds, then functions $f_t(x_t, \xi_t)$ and $Q_{t+1}(x_t, \xi_{t})$ are random polyhedral, and hence the same arguments can be applied without additional regularity conditions. \Box

Formula (3.56) makes it possible to write optimality conditions (3.59) in the following form.

Theorem 3.7. Suppose that either conditions (D2) and (D3) or condition (D1) are satisfied. Then a feasible policy $\bar{x}_t(\xi_{[t]})$ is optimal iff there exist measurable $\bar{\pi}_t(\xi_{[t]})$, t = 1, ..., T, such that

$$0 \in \partial f_t(\bar{x}_t(\xi_{[t]}), \xi_t) - A_t^{\mathsf{T}} \bar{\pi}_t(\xi_{[t]}) - \mathbb{E}\left[B_{t+1}^{\mathsf{T}} \bar{\pi}_{t+1}(\xi_{[t+1]}) \middle| \xi_{[t]}\right]$$
(3.61)

for a.e. $\xi_{[t]}$ and t = 1, ..., T, where for t = T the corresponding term T + 1 is omitted.

Proof. By Proposition 3.6 we have that a feasible policy $\bar{x}_t(\xi_{[t]})$ is optimal iff conditions (3.59) and (3.60) hold true. For t = 1 this means the existence of $\bar{\pi}_1 \in \mathfrak{D}_1$ such that

$$0 \in \partial f_1(\bar{x}_1) - A_1^{\mathsf{T}} \bar{\pi}_1 + \mathbb{E} \left[\partial Q_2(\bar{x}_1, \xi_2) \right].$$
(3.62)

Recall that ξ_1 is known, and hence the set \mathfrak{D}_1 is fixed. By (3.56) we have

$$\partial Q_2(\bar{x}_1,\xi_2) = -B_2^{\mathsf{I}}\mathfrak{D}_2(\bar{x}_1,\xi_2).$$
(3.63)

Formulas (3.62) and (3.63) mean that there exists a measurable selection

$$\bar{\boldsymbol{\pi}}_2(\xi_2) \in \mathfrak{D}_2\left(\bar{x}_1, \xi_2\right)$$

such that (3.61) holds for t = 1. By the second assertion of Proposition 3.6, the same selection $\bar{\pi}_2(\xi_2)$ can be used in (3.59) for t = 2. Proceeding in that way we obtain existence of measurable selections

$$\bar{\boldsymbol{\pi}}_t(\xi_t) \in \mathfrak{D}_t\left(\bar{\boldsymbol{x}}_{t-1}(\xi_{[t-1]}), \xi_{[t]}\right)$$

satisfying (3.61). □

In particular, consider the multistage linear problem given in the nested formulation (3.30). That is, functions $f_t(x_t, \xi_t)$ are defined in the form (3.51) which can be written as

$$f_t(x_t, \xi_t) = c_t^{\mathsf{I}} x_t + \mathbb{I}_{\mathbb{R}^{n_t}}(x_t).$$

Then $\partial f_t(x_t, \xi_t) = \{c_t + \mathcal{N}_{\mathbb{R}^{n_t}_+}(x_t)\}$ at every point $x_t \ge 0$, and hence optimality conditions (3.61) take the form

$$0 \in \mathcal{N}_{\mathbb{R}^{n_t}_+} \left(\bar{x}_t(\xi_{[t]}) \right) + c_t - A_t^\mathsf{T} \bar{\pi}_t(\xi_{[t]}) - \mathbb{E} \left[B_{t+1}^\mathsf{T} \bar{\pi}_{t+1}(\xi_{[t+1]}) \middle| \xi_{[t]} \right].$$

3.2.3 Dualization of Feasibility Constraints

Consider the linear multistage program given in the nested formulation (3.30). In this section we discuss dualization of that problem with respect to the feasibility constraints. As it was discussed before, we can formulate that problem as an optimization problem with respect to decision variables $x_t = x_t(\xi_{[t]})$ viewed as functions of the history of the data process. Recall that the vector ξ_t of the data process of that problem is formed from some (or all) elements of (c_t, B_t, A_t, b_t) , t = 1, ..., T. As before, we use the same symbols c_t, B_t, A_t, b_t to denote random variables and their particular realization. It will be clear from the context which of these meanings is used in a particular situation.

With problem (3.30) we associate the following Lagrangian:

$$L(\boldsymbol{x}, \boldsymbol{\pi}) := \mathbb{E} \left\{ \sum_{t=1}^{T} \left[c_{t}^{\mathsf{T}} x_{t} + \pi_{t}^{\mathsf{T}} \left(b_{t} - B_{t} x_{t-1} - A_{t} x_{t} \right) \right] \right\}$$
$$= \mathbb{E} \left\{ \sum_{t=1}^{T} \left[c_{t}^{\mathsf{T}} x_{t} + \pi_{t}^{\mathsf{T}} b_{t} - \pi_{t}^{\mathsf{T}} A_{t} x_{t} - \pi_{t+1}^{\mathsf{T}} B_{t+1} x_{t} \right] \right\}$$
$$= \mathbb{E} \left\{ \sum_{t=1}^{T} \left[b_{t}^{\mathsf{T}} \pi_{t} + \left(c_{t} - A_{t}^{\mathsf{T}} \pi_{t} - B_{t+1}^{\mathsf{T}} \pi_{t+1} \right)^{\mathsf{T}} x_{t} \right] \right\},$$

with the convention that $x_0 = 0$ and $B_{T+1} = 0$. Here the multipliers $\pi_t = \pi_t(\xi_{[t]})$, as well as decisions $x_t = \mathbf{x}_t(\xi_{[t]})$, are functions of the data process up to time t.

The dual functional is defined as

$$D(\boldsymbol{\pi}) := \inf_{\boldsymbol{x} > 0} L(\boldsymbol{x}, \boldsymbol{\pi}),$$

where the minimization is performed over variables $x_t = x_t(\xi_{[t]}), t = 1, ..., T$, in an appropriate functional space subject to the nonnegativity constraints. The Lagrangian dual of (3.30) is the problem

$$\operatorname{Max} D(\boldsymbol{\pi}), \tag{3.64}$$

where π lives in an appropriate functional space. Since, for a given π , the Lagrangian $L(\cdot, \pi)$ is separable in $x_t = x_t(\cdot)$, by the interchangeability principle (Theorem 7.92) we can move the operation of minimization with respect to x_t inside the conditional expectation $\mathbb{E}\left[\cdot |\xi_{[t]}\right]$. Therefore, we obtain

$$D(\boldsymbol{\pi}) = \mathbb{E}\left\{\sum_{t=1}^{T} \left[b_t^{\mathsf{T}} \pi_t + \inf_{x_t \in \mathbb{R}^{n_t}_+} \left(c_t - A_t^{\mathsf{T}} \pi_t - \mathbb{E} \left[B_{t+1}^{\mathsf{T}} \pi_{t+1} \big| \xi_{[t]} \right] \right)^{\mathsf{T}} x_t \right] \right\}.$$

Clearly we have that $\inf_{x_t \in \mathbb{R}^{n_t}_+} (c_t - A_t^\mathsf{T} \pi_t - \mathbb{E} \left[B_{t+1}^\mathsf{T} \pi_{t+1} | \xi_{[t]} \right] \right)^\mathsf{T} x_t$ is equal to zero if $A_t^\mathsf{T} \pi_t + \mathbb{E} \left[B_{t+1}^\mathsf{T} \pi_{t+1} | \xi_{[t]} \right] \leq c_t$, and to $-\infty$ otherwise. It follows that in the present case the dual problem (3.64) can be written as

$$\max_{\boldsymbol{\pi}} \mathbb{E}\left[\sum_{t=1}^{T} b_t^{\mathsf{T}} \pi_t\right]$$
s.t. $A_t^{\mathsf{T}} \pi_t + \mathbb{E}\left[B_{t+1}^{\mathsf{T}} \pi_{t+1} | \xi_{[t]}\right] \le c_t, \ t = 1, \dots, T,$

$$(3.65)$$

where for the uniformity of notation we set all "T + 1" terms equal zero. Each multiplier vector $\pi_t = \pi_t(\xi_{[t]}), t = 1, ..., T$, of problem (3.65) is a function of $\xi_{[t]}$. In that sense these multipliers form a dual implementable policy. Optimization in (3.65) is performed over all implementable and feasible dual policies.

If the data process has a finite number of scenarios, then implementable policies $x_t(\cdot)$ and $\pi_t(\cdot)$, $t = 1, \ldots, T$, can be identified with finite dimensional vectors. In that case the primal and dual problems form a pair of mutually dual linear programming problems. Therefore, the following duality result is a consequence of the general duality theory of linear programming.

Theorem 3.8. Suppose that the data process has a finite number of possible realizations (scenarios). Then the optimal values of problems (3.30) and (3.65) are equal unless both problems are infeasible. If the (common) optimal value of these problems is finite, then both problems have optimal solutions.

If the data process has a general distribution with an infinite number of possible realizations, then some regularity conditions are needed to ensure zero duality gap between problems (3.30) and (3.65).

3.2.4 Dualization of Nonanticipativity Constraints

In this section we deal with a problem which is slightly more general than linear problem (3.36). Let $f_t(x_t, \xi_t)$, t = 1, ..., T, be *random polyhedral* functions, and consider the problem

Here ξ_1^k, \ldots, ξ_T^k , $k = 1, \ldots, K$, is a particular realization (scenario) of the corresponding data process, $f_t^k(x_t^k) := f_t(x_t^k, \xi_t^k)$ and $(B_t^k, A_t^k, b_t^k) := (B_t(\xi_t^k), A_t(\xi_t^k), b_t(\xi_t^k))$, $t = 2, \ldots, T$. This problem can be formulated as a multistage stochastic programming problem by enforcing the corresponding nonanticipativity constraints.

As it was discussed in section 3.1.5, there are many ways to write nonanticipativity constraints. For example, let \mathfrak{X} be the linear space of all sequences (x_1^k, \ldots, x_T^k) , $k = 1, \ldots, K$, and \mathfrak{L} be the linear subspace of \mathfrak{X} defined by the nonanticipativity constraints (these spaces were defined above equation (3.41)). We can write the corresponding multistage problem in the following lucid form

$$\underset{\boldsymbol{x}\in\mathfrak{X}}{\operatorname{Min}} f(\boldsymbol{x}) := \sum_{k=1}^{K} \sum_{t=1}^{T} p_k f_t^k(x_t^k) \quad \text{subject to} \quad \boldsymbol{x}\in\mathfrak{L}.$$
(3.66)

Clearly, $f(\cdot)$ is a polyhedral function, so if problem (3.66) has a finite optimal value, then it has an optimal solution and the optimality conditions and duality relations hold true. Let us introduce the Lagrangian associated with (3.66):

$$L(\boldsymbol{x},\boldsymbol{\lambda}) := f(\boldsymbol{x}) + \langle \boldsymbol{\lambda}, \boldsymbol{x} \rangle,$$

with the scalar product $\langle \cdot, \cdot \rangle$ defined in (3.41). By the definition of the subspace \mathfrak{L} , every point $x \in \mathfrak{L}$ can be viewed as an implementable policy. By $\mathfrak{L}^{\perp} := \{ y \in \mathfrak{X} : \langle y, x \rangle = 0, \forall x \in \mathfrak{L} \}$ we denote the orthogonal subspace to the subspace \mathfrak{L} .

Theorem 3.9. A policy $\bar{x} \in \mathfrak{L}$ is an optimal solution of (3.66) if and only if there exist a multiplier vector $\bar{\lambda} \in \mathfrak{L}^{\perp}$ such that

$$\bar{\boldsymbol{x}} \in \arg\min_{\boldsymbol{x}\in\mathfrak{X}} L(\boldsymbol{x}, \bar{\boldsymbol{\lambda}}).$$
 (3.67)

Proof. Let $\bar{\lambda} \in \mathfrak{L}^{\perp}$ and $\bar{x} \in \mathfrak{L}$ be a minimizer of $L(\cdot, \bar{\lambda})$ over \mathfrak{X} . Then by the first order optimality conditions we have that $0 \in \partial_x L(\bar{x}, \bar{\lambda})$. Note that there is no need here for a

constraint qualification since the problem is polyhedral. Now $\partial_x L(\bar{x}, \bar{\lambda}) = \partial f(\bar{x}) + \bar{\lambda}$. Since $\mathcal{N}_{\mathfrak{L}}(\bar{x}) = \mathfrak{L}^{\perp}$, it follows that $0 \in \partial f(\bar{x}) + \mathcal{N}_{\mathfrak{L}}(\bar{x})$, which is a sufficient condition for \bar{x} to be an optimal solution of (3.66). Conversely, if \bar{x} is an optimal solution of (3.66), then necessarily $0 \in \partial f(\bar{x}) + \mathcal{N}_{\mathfrak{L}}(\bar{x})$. This implies existence of $\bar{\lambda} \in \mathfrak{L}^{\perp}$ such that $0 \in \partial_x L(\bar{x}, \bar{\lambda})$. This, in turn, implies that $\bar{x} \in \mathfrak{L}$ is a minimizer of $L(\cdot, \bar{\lambda})$ over \mathfrak{X} . \Box

Also, we can define the dual function

$$D(\boldsymbol{\lambda}) := \inf_{\boldsymbol{x} \in \mathfrak{X}} L(\boldsymbol{x}, \boldsymbol{\lambda}),$$

and the dual problem

$$\max_{\boldsymbol{\lambda} \in \mathfrak{L}^{\perp}} D(\boldsymbol{\lambda}). \tag{3.68}$$

Since the problem considered is polyhedral, we have by the standard theory of linear programming the following results.

Theorem 3.10. The optimal values of problems (3.66) and (3.68) are equal unless both problems are infeasible. If their (common) optimal value is finite, then both problems have optimal solutions.

The crucial role in our approach is played by the requirement that $\lambda \in \mathfrak{L}^{\perp}$. Let us decipher this condition. For $\lambda = (\lambda_t^k)_{t=1,\ldots,T, k=1,\ldots,K}$, the condition $\lambda \in \mathfrak{L}^{\perp}$ is equivalent to

$$\sum_{t=1}^{T} \sum_{k=1}^{K} p_k \langle \lambda_t^k, x_t^k \rangle = 0, \quad \text{for all } \boldsymbol{x} \in \mathfrak{L}$$

We can write this in a more abstract form as

$$\mathbb{E}\left[\sum_{t=1}^{T} \langle \lambda_t, x_t \rangle\right] = 0, \quad \text{for all } \boldsymbol{x} \in \mathfrak{L}.$$
(3.69)

Since¹² $\mathbb{E}_{|t}x_t = x_t$ for all $x \in \mathfrak{L}$, and $\langle \lambda_t, \mathbb{E}_{|t}x_t \rangle = \langle \mathbb{E}_{|t}\lambda_t, x_t \rangle$, we obtain from (3.69) that

$$\mathbb{E}\left[\sum_{t=1}^{T} \langle \mathbb{E}_{|t} \lambda_t, x_t \rangle\right] = 0, \quad \text{for all } \boldsymbol{x} \in \mathfrak{L}.$$

which is equivalent to

$$\mathbb{E}_{|t}[\lambda_t] = 0, \quad t = 1, \dots, T.$$
(3.70)

We can now rewrite our necessary conditions of optimality and duality relations in a more explicit form. We can write the dual problem in the form

$$\max_{\boldsymbol{\lambda} \in \mathfrak{X}} D(\boldsymbol{\lambda}) \quad \text{subject to} \quad \mathbb{E}_{|t}[\lambda_t] = 0, \quad t = 1, \dots, T.$$
(3.71)

¹²In order to simplify notation we denote in the remainder of this section by $\mathbb{E}_{|t|}$ the conditional expectation, conditional on $\xi_{|t|}$.

Corollary 3.11. A policy $\bar{x} \in \mathfrak{L}$ is an optimal solution of (3.66) if and only if there exist multipliers vector $\bar{\lambda}$ satisfying (3.70) such that

$$\bar{\boldsymbol{x}} \in \arg\min_{\boldsymbol{x}\in\mathfrak{X}} L(\boldsymbol{x}, \bar{\boldsymbol{\lambda}}).$$

Moreover, problem (3.66) has an optimal solution if and only if problem (3.71) has an optimal solution. The optimal values of these problems are equal unless both are infeasible.

There are many different ways to express the nonanticipativity constraints, and thus there are many equivalent ways to formulate the Lagrangian and the dual problem. In particular, a dual formulation based on (3.42) is quite convenient for dual decomposition methods. We leave it to the reader to develop the particular form of the dual problem in this case.

Exercises

- 3.1. Consider the inventory model of section 1.2.3.
 - (a) Specify for this problem the variables, the data process, the functions, and the sets in the general formulation (3.1). Describe the sets $\mathcal{X}_t(x_{t-1}, \xi_t)$ as in formula (3.50).
 - (b) Transform the problem to an equivalent linear multistage stochastic programming problem.
- 3.2. Consider the cost-to-go function $Q_t(x_{t-1}, \xi_{[t]})$, t = 2, ..., T, of the linear multistage problem defined as the optimal value of problem (3.25). Show that $Q_t(x_{t-1}, \xi_{[t]})$ is convex in x_{t-1} .
- 3.3. Consider the assembly problem discussed in section 1.3.3 in the case when all demand has to be satisfied, by backlogging the orders. It costs b_i to delay delivery of a unit of product *i* by one period. Additional orders of the missing parts can also be made after the last demand D(T) is known. Write the dynamic programming equations of the problem. How they can be simplified, if the demand is stage-wise independent?
- 3.4. A transportation company has n depots among which they move cargo. They are planning their operation in the next T days. The demand for transportation between depot i and depot j ≠ i on day t, where t = 1, 2..., T, is modeled as a random variable D_{ij}(t). The total capacity of vehicles currently available at depot i is denoted s_i, i = 1,...,n. Before each day t, the company considers repositioning their fleet to better prepare to the uncertain demand on the coming day. It costs c_{ij} to move a unit of capacity from location i to location j. After repositioning, the realization of the random variables D_{ij}(t) is observed, and the demand is served, up to the limit determined by the transportation capacity available at each location. The profit from transporting a unit of cargo from location i to location j is equal q_{ij}. If the total demand at location i exceeds the capacity available at this location, the excessive demand is lost. It is up to the company to decide how much of each

demand D_{ij} will be served, and which part will remain unsatisfied. For simplicity, we consider all capacity and transportation quantities as continuous variables. After the demand is served, the transportation capacity of the vehicles at each location changes, as a result of the arrivals of vehicles with cargo from other locations. Before the next day, the company may choose to reposition some of the vehicles to prepare for the next demand. On the last day, the vehicles are repositioned so that initial quantities s_i , i = 1, ..., n, are restored.

- (a) Formulate the problem of maximizing the expected profit as a multistage stochastic programming problem.
- (b) Write the dynamic programming equations for this problem. Assuming that the demand is stage-wise independent, identify the state variables and simplify the dynamic programming equations.
- (c) Develop a scenario tree based formulation of the problem.
- 3.5. Derive the dual problem to the linear multistage stochastic programming problem (3.36) with nonanticipativity constraints in the form (3.42).
- 3.6. You have initial capital C_0 which you may invest in a stock or keep in cash. You plan your investments for the next T periods. The return rate on cash is deterministic and equals r per each period. The price of the stock is random and equals S_t in period $t = 1, \ldots, T$. The current price S_0 is known to you and you have a model of the price process S_t in the form of a scenario tree. At the beginning, several American options on the stock prize are available. There are n put options with strike prices p_1, \ldots, p_n and corresponding costs c_1, \ldots, c_n . For example, if you buy one put option i, at any time $t = 1, \ldots, T$ you have the right to exercise the option and cash $p_i - S_t$ (this makes sense only when $p_i > S_t$). Also, m call options are available, with strike prices π_1, \ldots, π_m and corresponding costs q_1, \ldots, q_m . For example, if you buy one call option j, at any time t = 1, ..., T you may exercise it and cash $S_t - \pi_i$ (this makes sense only when $\pi_i < S_t$). The options are available only at t = 0. At any time period t you may buy or sell the underlying stock. Borrowing cash and short selling, that is, selling shares which are not actually owned (with the hope of repurchasing them later with profit), are not allowed. At the end of period T all options expire. There are no transaction costs, and shares and options can be bought, sold (in the case of shares), or realized (in the case of options) in any quantities (not necessarily whole numbers). The amounts gained by exercising options are immediately available for purchasing shares. Consider two objective functions.
 - (i) The expected value of your holdings at the end of period T;
 - (ii) The expected value of a piecewise linear utility function evaluated at the value of your final holdings. Its form is

$$u(C_T) = \begin{cases} C_T, & \text{if } C_T \ge 0, \\ (1+R)C_T, & \text{if } C_T < 0, \end{cases}$$

where R > 0 is some known constant.

For both objective functions:

- (a) Develop a linear multistage stochastic programming model.
- (b) Derive the dual problem by dualizing with respect to feasibility constraints.

Chapter 4 Optimization Models with Probabilistic Constraints

Darinka Dentcheva

4.1 Introduction

In this chapter, we discuss stochastic optimization problems with *probabilistic* (also called *chance*) constraints of the form

$$\begin{array}{ll}
\text{Min } c(x) \\
\text{s.t. } \Pr\{g_j(x, Z) \le 0, \ j \in \mathcal{J}\} \ge p, \\
& x \in \mathcal{X}.
\end{array}$$
(4.1)

Here $\mathcal{X} \subset \mathbb{R}^n$ is a nonempty set, $c : \mathbb{R}^n \to \mathbb{R}$, $g_j : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}$, $j \in \mathcal{J}$, with \mathcal{J} being an index set, Z is an s-dimensional random vector, and $p \in (0, 1)$ is a modeling parameter. the notation P_Z stands for the probability measure (probability distribution) induced by the random vector Z on \mathbb{R}^s . The event $A(x) = \{g_j(x, Z) \leq 0, j \in \mathcal{J}\}$ in (4.1) depends on the decision vector x, and its probability, $\Pr\{A(x)\}$ is calculated with respect to the probability distribution P_Z .

This model is in harmony with the statistical decisions, i.e., for a given point x, we do not reject the statistical hypothesis of the constraints $g_j(x, Z) \le 0$, $j \in \mathcal{J}$, being satisfied. We discussed this modeling approach in Chapter 1 in the contexts of inventory, multiproduct and portfolio selection models. Furthermore, imposing constraints on probability of events is particularly appropriate whenever high uncertainty is involved and reliability is a central issue. In those cases, constraints on the expected value may not reflect our attitude to undesirable outcomes.

We also note that the objective function c(x) can represent an expected value function or a measure of risk, i.e., $c(x) = \mathbb{E}[f(x, Z)]$ or $c(x) = \varrho(f(x, Z))$. By virtue of Theorem



Figure 4.1. Vehicle routing network

7.48, if the function $f(\cdot, Z)$ is continuous at x_0 and an integrable random variable \hat{Z} exists such that $|f(x, Z(\omega))| \leq \hat{Z}(\omega)$ for *P*-almost every $\omega \in \Omega$ and for all x in a neighborhood of x_0 , then for all x in a neighborhood of x_0 the expected value function c(x) is well defined and continuous at x_0 . Furthermore, convexity of $f(\cdot, z)$ implies convexity of the expectation function c(x). Therefore, we can carry out the analysis of probabilistically constrained problems using a general objective function c(x) with the understanding that in some cases it may be defined as an expectation function. In this chapter, we focus on the analysis of the probabilistic constraints.

Note that the probability $Pr{A(x)}$ ca be represented as the expected value of the indicator function of the event A(x), i.e., $Pr{A(x)} = \mathbb{E} [\mathbf{1}_{A(x)}]$. The discontinuity of the indicator function and the complexity of the event A(x) make the constraints on probability qualitatively different from the constraints on expected values of continuous functions. Let us consider two examples.

Example 4.1 (Vehicle Routing Problem) Consider a network with m arcs on which a random transportation demand arises. A set of n routes in the network is described by the incidence matrix T. More precisely, T is an $m \times n$ dimensional matrix such that

$$t_{ij} = \begin{cases} 1 & \text{if route } j \text{ contains arc } i, \\ 0 & \text{otherwise.} \end{cases}$$

We have to allocate vehicles to the routes so as satisfy transportation demand. Figure 4.1 depicts a small network, and the table provides the incidence information for 19 routes on this network. For example, route 5 consists of the arcs AB, BC, and CA.

Our aim is to satisfy the demand with high prescribed probability $p \in (0, 1)$. Let x_j be the number of vehicles assigned to route j, j = 1, ..., n. The demand for transportation on each arc is given by the random variables $Z_i, i = 1, ..., m$. We set $Z = (Z_1, ..., Z_m)^T$. A cost c_j is associated with operating a vehicle on route j. Setting $c = (c_1, ..., c_n)^T$, the
Arc	Route																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
AB	1				1						1				1		1		
AC		1				1	1						1					1	
AD			1					1	1			1							
AE				1						1				1		1			1
BA	1					1						1				1		1	
BC					1						1				1		1		
CA		1			1			1						1			1		
CB						1						1				1		1	
CD							1				1		1		1		1	1	
DA			1				1			1	1								
DC								1				1		1		1	1	1	
DE									1				1		1				1
EA				1					1				1		1				1
ED										1				1		1			1

Figure 4.2. Vehicle routing incidence matrix

model can be formulated as follows¹:

$$\operatorname{Min} c^{\mathsf{T}} x \tag{4.2}$$

s.t. $\Pr\{Tx \ge Z\} \ge p,$ (4.3)

$$x \in \mathbb{Z}_+^n. \tag{4.4}$$

In practical applications, we may have heterogeneous fleet of vehicles with different capacities; we may consider imposing constraints on transportation time or other requirements.

In the context of portfolio optimization, probabilistic constraints arise in a natural way, as discussed in Chapter 1.

Example 4.2 (Portfolio Optimization with Value-at-Risk Constraint) We consider n investment opportunities, with random return rates R_1, \ldots, R_n in the next time period. Our aim is to invest our capital in such a way that the expected value of our investment is maximized, under the condition that the chance of losing no more than a given fraction of this amount is at least p, where $p \in (0, 1)$. This requirement is called a *Value-at-Risk* (V@R) constraint.

Denoting the fractions of our capital invested in the *n* instruments by x_1, \ldots, x_n , th value change of our investment changes in the next period of time is:

$$g(x,R) = \sum_{i=1}^{n} R_i x_i.$$

¹The notation \mathbb{Z}_+ used to denote the set of nonnegative integer numbers.

The following stochastic optimization problem constructs a portfolio maximizing the expected return rate under a V@R constraint :

$$\operatorname{Max} \sum_{i=1}^{n} \mathbb{E}[R_i] x_i$$

s.t.
$$\operatorname{Pr}\left\{\sum_{i=1}^{n} R_i x_i \ge \eta\right\} \ge p,$$

$$\sum_{i=1}^{n} x_i = 1,$$

$$x \ge 0.$$

(4.5)

For example, $\eta = -0.1$ may be chosen, if we aim at protecting against losses larger than 10%.

The constraint

$$\Pr\{g_j(x,Z) \le 0, j \in \mathcal{J}\} \ge p$$

is called a joint probabilistic constraint, while the constraints

$$\Pr\{g_j(x,Z) \leq 0\} \geq p_j, \ j \in \mathcal{J}, \text{ where } p_j \in [0,1],$$

are called individual probabilistic constraints.

In the vehicle routing example, we have a joint probabilistic constraint. If we were to cover the demand on each arc separately with high probability, then the constraints would be formulated as follows:

$$\Pr\{T^i x \ge Z_i\} \ge p_i, \quad i = 1, \dots, m,$$

where T^i denotes the *i*th row of the matrix T. However, the latter formulation would not ensure reliability of the network as a whole.

Infinitely many individual probabilistic constraints appear naturally in the context of stochastic orders. For an integrable random variable X, we consider its distribution function $F_X(\cdot)$.

Definition 4.3. A random variable X dominates in the first order a random variable Y (denoted $X \succeq_{(1)} Y$) if

$$F_X(\eta) \le F_Y(\eta), \ \forall \eta \in \mathbb{R}.$$

The left-continuous inverse $F_X^{(-1)}$ of the cumulative distribution function of a random variable X is defined as follows:

$$F_X^{(-1)}(p) = \inf \{\eta : F_1(X;\eta) \ge p\}, \ p \in (0,1).$$

Given $p \in (0, 1)$, the number q = q(X; p) is called a *p*-quantile of the random variable X if

$$\Pr\{X < q\} \le p \le \Pr\{X \le q\}.$$

For $p \in (0, 1)$ the set of *p*-quantiles is a closed interval and $F_X^{(-1)}(p)$ represents its left end. Directly from the definition of the first order dominance we see that

$$X \succeq_{(1)} Y \quad \Leftrightarrow \quad F_X^{(-1)}(p) \ge F_Y^{(-1)}(p) \quad \text{for all} \quad 0
$$(4.6)$$$$

The first order dominance constraint can be interpreted as a continuum of probabilistic (chance) constraints.

Denoting $F_X^{(1)}(\eta) = F_X(\eta)$, we define higher order distribution functions of a random variable $X \in \mathcal{L}_{k-1}(\Omega, \mathcal{F}, P)$ as follows:

$$F_X^{(k)}(\eta) = \int_{-\infty}^{\eta} F_X^{(k-1)}(t) dt$$
 for $k = 2, 3, 4, \dots$

We can express the integrated distribution function $F_X^{(2)}$ as the expected shortfall function. Integrating by parts, for each value η , we have the following formula²:

$$F_X^{(2)}(\eta) = \int_{-\infty}^{\eta} F_X(\alpha) \, d\alpha = \mathbb{E}[(\eta - X)_+].$$
(4.7)

The function $F_X^{(2)}(\cdot)$ is well defined and finite for every integrable random variable. It is continuous, nonnegative and nondecreasing. The function $F_X^{(2)}(\cdot)$ is also convex because its derivative is nondecreasing as it is a cumulative distribution function. By the same arguments, the higher order distribution functions are continuous, nonnegative, nondecreasing, and convex as well.

Due to (4.7) second order dominance relation can be expressed in an equivalent way as follows:

$$X \succeq_{(2)} Y \text{ iff } \mathbb{E}[(\eta - X)_+] \le \mathbb{E}[(\eta - Y)_+] \text{ for all } \eta \in \mathbb{R}.$$
 (4.8)

The stochastic dominance relation generalizes to higher orders as follows:

Definition 4.4. Given two random variables X and Y in $\mathcal{L}_{k-1}(\Omega, \mathcal{F}, P)$ we say that X dominates Y in the kth order if

$$F_X^{(k)}(\eta) \le F_Y^{(k)}(\eta), \ \forall \eta \in \mathbb{R}.$$

We denote this relation by $X \succeq_{(k)} Y$.

We call the following semi-infinite (probabilistic) problem a *stochastic optimization* problem with a stochastic-order constraint:

$$\begin{aligned}
& \underset{x}{\operatorname{Min}} \quad \mathbb{E}[f(x,Z)] \\
& \text{s.t.} \quad \Pr\left\{g(x,Z) \leq \eta\right\} \leq F_{Y}(\eta), \quad \eta \in [a,b], \\
& \quad x \in \mathcal{X}.
\end{aligned}$$
(4.9)

²Recall that $[a]_{+} = \max\{a, 0\}.$

Here the dominance relation is restricted to an interval $[a, b] \subset \mathbb{R}$. There are technical reasons for this restriction, which will become apparent later. In the case of discrete distributions with finitely many realizations, we can assume that the interval [a, b] contains the entire support of the probability measures.

In general, we formulate the following semi-infinite probabilistic problem, which we refer to as a *stochastic optimization problem with a stochastic dominance constraint* of order $k \ge 2$:

$$\begin{aligned}
& \underset{x}{\operatorname{Min}} c(x) \\
& \text{s.t. } F_{g(x,Z)}^{(k)}(\eta) \leq F_{Y}^{(k)}(\eta), \quad \eta \in [a,b], \\
& \quad x \in \mathcal{X}.
\end{aligned}$$
(4.10)

Example 4.5 (Portfolio Selection Problem with Stochastic-Order Constraints)

Returning to Example 4.2, we can require that the net profit on our investment dominates certain benchmark outcome Y, which may be the return rate of our current portfolio or the return rate of some index. Then the Value-at-Risk constraint has to be satisfied at a continuum of points $\eta \in \mathbb{R}$. Setting $\Pr{Y \leq \eta} = p_{\eta}$, we formulate the following model:

$$\operatorname{Max} \sum_{i=1}^{n} \mathbb{E}[R_i] x_i$$

s.t.
$$\operatorname{Pr} \left\{ \sum_{i=1}^{n} R_i x_i \leq \eta \right\} \leq p_{\eta}, \quad \forall \eta \in \mathbb{R},$$

$$\sum_{i=1}^{n} x_i = 1,$$

$$x \geq 0.$$

(4.11)

Using higher order stochastic dominance relations, we formulate a portfolio optimization model of form:

$$\operatorname{Max} \sum_{i=1}^{n} \mathbb{E}[R_i] x_i$$

s.t.
$$\sum_{i=1}^{n} R_i x_i \succeq_{(k)} Y,$$
$$\sum_{i=1}^{n} x_i = 1,$$
$$x \ge 0.$$
$$(4.12)$$

A second order dominance constraint on the portfolio return rate represents a constraint on the shortfall function:

$$\sum_{i=1}^{n} R_{i} x_{i} \succeq_{(2)} Y \quad \Longleftrightarrow \quad \mathbb{E}\left[\left(\eta - \sum_{i=1}^{n} R_{i} x_{i}\right)_{+}\right] \leq \mathbb{E}\left[(\eta - Y)_{+}\right], \ \forall \eta \in \mathbb{R}.$$

The second order dominance constraint can also be viewed as a continuum of *Average Value-at-Risk*³ (AV@R) constraints. We refer to Dentcheva and Ruszczyński [66] for more information about this connection.

If a = b, then the semi-infinite model (4.9) reduces to a problem with a single probabilistic constraint, and problem (4.10) for k = 2 becomes a problem with a single shortfall constraint.

We shall pay special attention to problems with *separable* functions g_i , i = 1, ..., m, that is, functions of form $g_i(x, z) = \hat{g}_i(x) + h_i(z)$. The probabilistic constraint becomes

$$\Pr\{\hat{g}_i(x) \ge -h_i(Z), \ i = 1, \dots, m\} \ge p.$$

We can view the inequalities under the probability as a deterministic vector function \hat{g} : $\mathbb{R}^n \to \mathbb{R}^m, \hat{g} = [\hat{g}_1, \dots, \hat{g}_m]^\mathsf{T}$ constrained from below by a random vector Y with $Y_i = -h_i(Z), i = 1, \dots, m$. The problem can be formulated as follows:

$$\begin{array}{l} \underset{x}{\operatorname{Min}} & c(x) \\ \text{s.t.} & \operatorname{Pr}\left\{\hat{g}(x) \geq Y\right\} \geq p, \\ & x \in \mathcal{X}, \end{array}$$
(4.13)

where the inequality $a \leq b$ for two vectors $a, b \in \mathbb{R}^n$ is understood componentwise.

Problems with separable probabilistic constraints arise frequently in the context of serving certain demand, as in the vehicle routing Example 4.1. Another type of examples is an inventory problem, as the following one.

Example 4.6 (Cash Matching with Probabilistic Liquidity Constraint)

We have random liabilities L_t in periods t = 1, ..., T. We consider an investment in a bond portfolio from a basket of n bonds. The payment of bond i in period t is denoted by a_{it} . It is zero for the time periods t before purchasing of the bond is possible, as well as for t greater than the maturity time of the bond. At the time period of purchase, a_{it} is the negative of the price of the bond. At the following periods, a_{it} is equal to the coupon payment, and at the time of maturity it is equal to the face value plus the coupon payment. All prices of bonds and coupon payments are deterministic and no default is assumed. Our initial capital equals c_0 .

The objective is to design a bond portfolio such that the probability of covering the liabilities over the entire period $1, \ldots, T$ is at least p. Subject to this condition, we want to maximize the final cash on hand, guaranteed with probability p.

Let us introduce the cumulative liabilities

$$Z_t = \sum_{\tau=1}^t L_\tau, \quad t = 1, \dots, T.$$

Denoting by x_i the amount invested in bond *i*, we observe that the cumulative cash flows up to time *t*, denoted c_t , can be expressed as follows:

$$c_t = c_{t-1} + \sum_{i=1}^n a_{it} x_i, \quad t = 1, \dots, T.$$

³Average Value-at-Risk is also called Conditional Value-at-Risk.

Using cumulative cash flows and cumulative liabilities permits the carry-over of capital from one stage to the next one, while keeping the random quantities at the right hand side of the constraints. We represent the cumulative cash flow during the entire period by the vector $c = (c_1, \ldots, c_T)^T$. Let us assume that we quantify our preferences by using concave utility function $U : \mathbb{R} \to \mathbb{R}$. We would like to maximize the final capital at hand in a risk-averse manner. The problem takes on the form

$$\begin{aligned} & \underset{x,c}{\operatorname{Max}} \mathbb{E}\left[U(c_T - Z_T)\right] \\ & \text{s.t. } \mathsf{Pr}\left\{c_t \ge Z_t, \ t = 1, \dots, T\right\} \ge p, \\ & c_t = c_{t-1} + \sum_{i=1}^n a_{it} x_i, \quad t = 1, \dots, T, \\ & x \ge 0. \end{aligned}$$

This optimization problem has the structure of model (4.13). The first constraint can be called a *probabilistic liquidity constraint*.

4.2 Convexity in probabilistic optimization

Fundamental questions for every optimization model concern convexity of the feasible set, as well as continuity and differentiability of the constraint functions. The analysis of models with probability functions is based on specific properties of the underlying probability distributions. In particular, the *generalized concavity* theory plays a central role in probabilistic optimization as it facilitates the application of powerful tools of convex analysis.

4.2.1 Generalized concavity of functions and measures

We consider various nonlinear transformations of functions $f(x) : \Omega \to \mathbb{R}_+$ defined on a convex set $\Omega \subset \mathbb{R}^s$.

Definition 4.7. A nonnegative function f(x) defined on a convex set $\Omega \subset \mathbb{R}^s$ is said to be α -concave, where $\alpha \in [-\infty, +\infty]$, if for all $x, y \in \Omega$ and all $\lambda \in [0, 1]$, the following inequality holds

$$f(\lambda x + (1 - \lambda)y) \ge m_{\alpha}(f(x), f(y), \lambda),$$

where $m_{\alpha} : \mathbb{R}_+ \times \mathbb{R}_+ \times [0,1] \to \mathbb{R}$ is defined as follows:

$$m_{\alpha}(a,b,\lambda) = 0, \quad if \quad ab = 0,$$

and if a > 0, b > 0, $0 \le \lambda \le 1$, then

$$m_{\alpha}(a,b,\lambda) = \begin{cases} a^{\lambda}b^{1-\lambda}, & \text{if } \alpha = 0, \\ \max\{a,b\}, & \text{if } \alpha = \infty, \\ \min\{a,b\}, & \text{if } \alpha = -\infty, \\ (\lambda a^{\alpha} + (1-\lambda)b^{\alpha})^{1/\alpha}, & \text{otherwise.} \end{cases}$$

In the case of $\alpha = 0$, the function f is called *logarithmically concave* or *log-concave* because $\ln f(\cdot)$ is a concave function. In the case of $\alpha = 1$, the function f is simply *concave*.

It is important to note that if f and g are two measurable functions, then the function $m_{\alpha}(f(\cdot), g(\cdot), \lambda)$ is a measurable function for all α and for all $\lambda \in (0, 1)$. Furthermore, $m_{\alpha}(a, b, \lambda)$ has the following monotonicity property:

Lemma 4.8. The mapping $\alpha \mapsto m_{\alpha}(a, b, \lambda)$ is nondecreasing and continuous.

Proof. First we show the continuity of the mapping at $\alpha = 0$. We have the following chain of equations:

$$\ln m_{\alpha}(a,b,\lambda) = \ln(\lambda a^{\alpha} + (1-\lambda)b^{\alpha})^{1/\alpha} = \frac{1}{\alpha}\ln\left(\lambda e^{\alpha\ln a} + (1-\lambda)e^{\alpha\ln b}\right)$$
$$= \frac{1}{\alpha}\ln\left(1 + \alpha\left(\lambda\ln a + (1-\lambda)\ln b\right) + o(\alpha^{2})\right).$$

Applying the l'Hospital rule to the right hand side in order to calculate its limit when $\alpha \rightarrow 0$, we obtain

$$\lim_{\alpha \to 0} \ln m_{\alpha}(a, b, \lambda) = \lim_{\alpha \to 0} \frac{\lambda \ln a + (1 - \lambda) \ln b + o(\alpha)}{1 + \alpha \left(\lambda \ln a + (1 - \lambda) \ln b\right) + o(\alpha^2)}$$
$$= \lim_{\alpha \to 0} \frac{\ln(a^{\lambda} b^{(1 - \lambda)}) + o(\alpha)}{1 + \alpha \ln(a^{\lambda} b^{(1 - \lambda)}) + o(\alpha^2)} = \ln(a^{\lambda} b^{(1 - \lambda)}).$$

Now, we turn to the monotonicity of the mapping. First, let us consider the case of $0 < \alpha < \beta$. We set

$$h(\alpha) = m_{\alpha}(a, b, \lambda) = \exp\left(\frac{1}{\alpha}\ln\left[\lambda a^{\alpha} + (1 - \lambda)b^{\alpha}\right]\right).$$

Calculating its derivative, we obtain:

$$h'(\alpha) = h(\alpha) \Big(\frac{1}{\alpha} \cdot \frac{\lambda a^{\alpha} \ln a + (1-\lambda)b^{\alpha} \ln b}{\lambda a^{\alpha} + (1-\lambda)b^{\alpha}} - \frac{1}{\alpha^2} \ln \left[\lambda a^{\alpha} + (1-\lambda)b^{\alpha} \right] \Big).$$

We have to demonstrate that the expression on the right hand side is nonnegative. Substituting $x = a^{\alpha}$ and $y = b^{\alpha}$, we obtain

$$h'(\alpha) = \frac{1}{\alpha^2} h(\alpha) \Big(\frac{\lambda x \ln x + (1 - \lambda)y \ln y}{\lambda x + (1 - \lambda)y} - \ln \left[\lambda x + (1 - \lambda)y \right] \Big).$$

Using the fact that the function $z \mapsto z \ln z$ is convex for z > 0 and that both x, y > 0, we have that

$$\frac{\lambda x \ln x + (1 - \lambda)y \ln y}{\lambda x + (1 - \lambda)y} - \ln \left[\lambda x + (1 - \lambda)y\right] \ge 0.$$

As $h(\alpha) > 0$, we conclude that $h(\cdot)$ is nondecreasing in this case. If $\alpha < \beta < 0$, we have the following chain of relations:

$$m_{\alpha}(a,b,\lambda) = \left[m_{-\alpha}\left(\frac{1}{a},\frac{1}{b},\lambda\right)\right]^{-1} \le \left[m_{-\beta}\left(\frac{1}{a},\frac{1}{b},\lambda\right)\right]^{-1} = m_{\beta}(a,b,\lambda).$$

In the case of $0 = \alpha < \beta$, we can select a sequence $\{\alpha_k\}$, such that $\alpha_k > 0$ and $\lim_{k\to\infty} \alpha_k = 0$. We use the monotonicity of $h(\cdot)$ for positive arguments and the continuity at 0 to obtain the desired assertion. In the case $\alpha < \beta = 0$, we proceed in the same way, choosing appropriate sequence approaching 0.

If $\alpha < 0 < \beta$, then the inequality

$$m_{\alpha}(a, b, \lambda) \leq m_0(a, b, \lambda) \leq m_{\beta}(a, b, \lambda)$$

follows from the previous two cases. It remains to investigate how the mapping behaves when $\alpha \to \infty$ or $\alpha \to -\infty$. We observe that

$$\max\{\lambda^{1/\alpha}a, (1-\lambda)^{1/\alpha}b\} \le m_{\alpha}(a,b,\lambda) \le \max\{a,b\}.$$

Passing to the limit, we obtain that

$$\lim_{\alpha \to \infty} m_{\alpha}(a, b, \lambda) = \max\{a, b\}.$$

We also infer that

$$\lim_{\alpha \to -\infty} m_{\alpha}(a, b, \lambda) = \lim_{\alpha \to -\infty} [m_{-\alpha}(1/a, 1/b, \lambda)]^{-1} = [\max\{1/a, 1/b\}]^{-1} = \min\{a, b\}.$$

This completes the proof. \Box

This statement has the very important implication that α -concavity entails β -concavity for all $\beta \leq \alpha$. Therefore, all α -concave functions are $(-\infty)$ -concave, that is, *quasi-concave*.

Example 4.9 Consider the density of a nondegenerate multivariate normal distribution on \mathbb{R}^{s} :

$$\theta(x) = \frac{1}{\sqrt{(2\pi)^{s} \det(\Sigma)}} \exp\left\{-\frac{1}{2}(x-\mu)^{\mathsf{T}} \Sigma^{-1}(x-\mu)\right\},\,$$

where Σ is a positive definite symmetric matrix of dimension $s \times s$, $det(\Sigma)$ denotes the determinant of the matrix Σ , and $\mu \in \mathbb{R}^s$. We observe that

$$\ln \theta(x) = -\frac{1}{2}(x-\mu)^{\mathsf{T}} \Sigma^{-1}(x-\mu) - \ln \left(\sqrt{(2\pi)^s \det(\Sigma)}\right)$$

is a concave function. Therefore, we conclude that θ is 0-concave, or log-concave.

,

Example 4.10 Consider a convex body (a convex compact set with nonempty interior) $\Omega \subset \mathbb{R}^s$. The uniform distribution on this set has a density defined as follows:

$$\theta(x) = \begin{cases} \frac{1}{V_s(\Omega)}, & x \in \Omega, \\ 0, & x \notin \Omega, \end{cases}$$

where $V_s(\Omega)$ denotes the Lebesgue measure of Ω . The function $\theta(x)$ is quasi-concave on \mathbb{R}^s and $+\infty$ -concave on Ω .

We point out that for two Borel measurable sets A, B in \mathbb{R}^s , the Minkowski sum $A + B = \{x + y : x \in A, y \in B\}$ is Lebesgue measurable in \mathbb{R}^s . Recall that for two sets, A and B and a number $\lambda \in (0, 1)$, the convex combination is define using the Minkowski sum in the following way:

$$\lambda A + (1 - \lambda)B = \{\lambda x + (1 - \lambda)y : x \in A, y \in B\}.$$

Definition 4.11. A probability measure P defined on the Lebesgue subsets of a convex set $\Omega \subset \mathbb{R}^s$ is said to be α -concave if for any Borel measurable sets $A, B \subset \Omega$ and for all $\lambda \in [0, 1]$, the following inequality holds

$$P(\lambda A + (1 - \lambda)B) \ge m_{\alpha}(P(A), P(B), \lambda).$$

We say that a random vector Z with values in \mathbb{R}^s has an α -concave distribution if the probability measure P_Z induced by Z on \mathbb{R}^s is α -concave.

Lemma 4.12. If a random vector Z has an α -concave probability distribution on \mathbb{R}^s , then its cumulative distribution function F_Z is an α -concave function.

Proof. Indeed, for given points $z^1, z^2 \in \mathbb{R}^s$ and $\lambda \in [0, 1]$, we define

 $A = \{ z \in \mathbb{R}^s : z_i \le z_i^1, i = 1, \dots, s \} \text{ and } B = \{ z \in \mathbb{R}^s : z_i \le z_i^2, i = 1, \dots, s \}.$

Then the inequality for F_Z follows from the inequality in Definition 4.11.

Lemma 4.13. If a random vector Z has independent components with log-concave marginal distributions, then Z has a log-concave distribution.

Proof. For two Borel sets $A, B \subset \mathbb{R}^s$ with $P_Z(A) > 0$ and $P_Z(B) > 0$ and for a number $\lambda \in (0, 1)$, we define the set $C = \lambda A + (1 - \lambda)B$. Each Borel measurable set can be approximated by rectangular sets. Let the sets $A_i^k, B_i^k, i = 1, \ldots, s$, be Borel sets on \mathbb{R} such that

$$A^k = \times_{i=1}^s A_i^k, \quad B^k = \times_{i=1}^s B_i^k$$

and $A^k \subset A$, $B^k \subset B$. For any k = 1, 2..., we can choose the sequences $\{A_i^k\}$, $\{B_i^k\}$ to be non-decreasing, i.e., $A_i^k \subseteq A_i^{k+1}$, $B_i^k \subseteq B_i^{k+1}$ so that $\bigcap_{k=1}^{\infty} A^k = A$, $\bigcap_{k=1}^{\infty} B^k = B$. We define the sets

$$C_i^k = \lambda A_i^k + (1 - \lambda) B_i^k, \quad C^k = \times_{i=1}^s C_i^k.$$

Evidently, $C^k \subset C$ for all k = 1, 2... and $\bigcap_{k=1}^{\infty} C^k = C$. We obtain

$$\ln[P_Z(C)] \ge \ln[P_Z(C^k)] = \sum_{i=1}^s \ln[P_{Z_i}(C_i^k)] = \sum_{i=1}^s \ln[P_{Z_i}(\lambda A_i^k + (1-\lambda)B_i^k)]$$
$$\ge \sum_{i=1}^s \left(\lambda \ln[P_{Z_i}(A_i^k)] + (1-\lambda)\ln[P_{Z_i}(B_i^k)]\right)$$
$$= \lambda \ln[P_Z(A^k)] + (1-\lambda)\ln[P_Z(B^k)]$$

We conclude that

$$\ln[P_Z(C)] \ge \limsup_{k \to \infty} \left(\lambda \ln[P_Z(A^k)] + (1-\lambda) \ln[P_Z(B^k)]\right)$$
$$= \lambda \ln[P_Z(A)] + (1-\lambda) \ln[P_Z(B)].$$

As usually, concavity properties of a function imply certain continuity of the function.

We formulate without proof two theorems addressing this issue.

Theorem 4.14 (Borell [28]). If P is a quasi-concave measure on \mathbb{R}^s and the dimension of its support is s, then P has a density with respect to the Lebesgue measure.

The α -concavity property of a measure can be related to a generalized concavity property of its density (see Brascamp and Lieb [30], Prékopa [195], Rinott[204].)

Theorem 4.15. Let Ω be a convex subset of \mathbb{R}^s and let m > 0 be the dimension of the smallest affine subspace L containing Ω . The probability measure P on Ω is γ -concave with $\gamma \in [-\infty, 1/m]$, if and only if its probability density function with respect to the Lebesgue measure on L is α -concave with

$$\alpha = \begin{cases} \gamma/(1 - m\gamma), & \text{if } \gamma \in (-\infty, 1/m), \\ -1/m, & \text{if } \gamma = -\infty, \\ +\infty, & \text{if } \gamma = 1/m. \end{cases}$$

Corollary 4.16. Let an integrable function $\theta(x)$ be defined and positive on a non-degenerated convex set $\Omega \subset \mathbb{R}^s$. Denote $c = \int_{\Omega} \theta(x) dx$. If $\theta(x)$ is α -concave with $\alpha \in [-1/s, \infty]$ and positive on the interior of Ω , then the measure P on Ω defined by setting

$$P(A) = \frac{1}{c} \int_{A} \theta(x) \, dx, \quad A \subset \Omega,$$

is γ -concave with

$$\gamma = \begin{cases} \alpha/(1+s\alpha), & \text{if } \alpha \in (-1/s, \infty), \\ 1/s, & \text{if } \alpha = \infty, \\ -\infty, & \text{if } \alpha = -1/s. \end{cases}$$

In particular, if a measure P on \mathbb{R}^s has a density function $\theta(x)$ such that $\theta^{-1/s}$ is convex, then P is quasi-concave.

Example 4.17 In Example 4.10, we have observed that the density of the unform distribution on a convex body Ω is an ∞ -concave function. Hence, it generates an 1/s-concave measure on Ω . On the other hand, the density of the normal distribution (Example 4.9) is log-concave, and, therefore, it generates a log-concave probability measure.

Example 4.18 Consider positive numbers $\alpha_1, \ldots, \alpha_s$ and the simplex

$$S = \left\{ x \in \mathbb{R}^s : \sum_{i=1}^s x_i \le 1, \ x_i \ge 0, \ i = 1, \dots, s \right\}.$$

The density function of the *Dirichlet distribution* with parameters $\alpha_1, \ldots, \alpha_s$ is defined as follows:

$$\theta(x) = \begin{cases} \frac{\Gamma(\alpha_1 + \dots + \alpha_s)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_s)} x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} \cdots x_s^{\alpha_s - 1}, & \text{if } x \in \text{int } S, \\ 0, & \text{otherwise.} \end{cases}$$

Here $\Gamma(\cdot)$ stands for the Gamma function $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$. Assuming that $x \in \text{int } S$, we consider

$$\ln \theta(x) = \sum_{i=1}^{s} (\alpha_i - 1) \ln x_i + \ln \Gamma(\alpha_1 + \dots + \alpha_s) - \sum_{i=1}^{s} \ln \Gamma(\alpha_i).$$

If $\alpha_i \ge 1$ for all i = 1, ..., s, then $\ln \theta(\cdot)$ is a concave function on the interior of S and, therefore, $\theta(x)$ is log-concave on cl S. If all parameters satisfy $\alpha_i \le 1$, then $\theta(x)$ is log-convex on cl (S). For other sets of parameters, this density function does not have any generalized concavity properties.

The next results provide calculus rules for α -concave functions.

Theorem 4.19. If the function $f : \mathbb{R}^s \to \mathbb{R}_+$ is α -concave and the function $g : \mathbb{R}^s \to \mathbb{R}_+$ is β -concave, where $\alpha, \beta \ge 1$, then the function $h : \mathbb{R}^s \to \mathbb{R}$, defined as h(x) = f(x) + g(x) is γ -concave with $\gamma = \min{\{\alpha, \beta\}}$.

Proof. Given points $x, y \in \mathbb{R}^s$ and a scalar $\lambda \in (0, 1)$, we set $z = \lambda x + (1 - \lambda)y$. Both functions f and g are γ -concave by virtue of Lemma 4.8. Using the γ -concavity together with the Minkowski inequality, which holds for $\gamma \ge 1$, we obtain:

$$f(z) + g(z)$$

$$\geq \left[\lambda (f(x))^{\gamma} + (1 - \lambda) (f(y))^{\gamma}\right]^{\frac{1}{\gamma}} + \left[\lambda (g(x))^{\gamma} + (1 - \lambda) (g(y))^{\gamma}\right]^{\frac{1}{\gamma}}$$

$$\geq \left[\lambda (f(x) + g(x))^{\gamma} + (1 - \lambda) (f(y) + g(y))^{\gamma}\right]^{\frac{1}{\gamma}}.$$

This completes the proof. \Box

Theorem 4.20. Let f be a concave function defined on a convex set $C \subset \mathbb{R}^s$ and $g : \mathbb{R} \to \mathbb{R}$ be a nonnegative nondecreasing α -concave function, $\alpha \in [-\infty, \infty]$. Then the function $g \circ f$ is α -concave.

Proof. Given $x, y \in \mathbb{R}^s$ and a scalar $\lambda \in (0, 1)$, we consider $z = \lambda x + (1 - \lambda)y$. We have $f(z) \ge \lambda f(x) + (1 - \lambda)f(y)$. By monotonicity and α -concavity of g, we obtain the

following chain of inequalities:

$$[g \circ f](z) \ge g(\lambda f(x) + (1 - \lambda)f(y)) \ge m_{\alpha}(g(f(x)), g(f(y)), \lambda).$$

This proves the assertion. \Box

Theorem 4.21. Let the function $f : \mathbb{R}^s \times \mathbb{R}^m \to \mathbb{R}_+$ be such that for all $y \in Y \subset \mathbb{R}^m$ the function $f(\cdot, y)$ is α -concave ($\alpha \in [-\infty, \infty]$) on the convex set $X \subset \mathbb{R}^s$. Then the function $\varphi(x) = \inf_{y \in Y} f(x, y)$ is α -concave on X.

Proof. Let $x_1, x_2 \in X$ and a scalar $\lambda \in (0, 1)$ be given. We set $z = \lambda x_1 + (1 - \lambda)x_2$. A sequence of points $y_k \in Y$ exists such that

$$\varphi(z) = \inf_{y \in Y} f(z, y) = \lim_{k \to \infty} f(z, y_k).$$

Using the α -concavity of the function $f(\cdot, y)$, we infer that

$$f(z, y_k) \ge m_\alpha \big(f(x_1, y_k), f(x_2, y_k), \lambda \big).$$

The mapping $(a, b) \mapsto m_{\alpha}(a, b, \lambda)$ is monotonic for nonnegative a and b and $\lambda \in (0, 1)$. Therefore, the following inequality holds

$$f(z, y_k) \ge m_{\alpha} (\varphi(x_1), \varphi(x_2), \lambda).$$

Passing to the limit, we obtain the desired relation. \Box

Lemma 4.22. If $\alpha_i > 0$, i = 1, ..., m, and $\sum_{i=1}^m \alpha_i = 1$, then the function $f : \mathbb{R}^m_+ \to \mathbb{R}$, defined as $f(x) = \prod_{i=1}^m x_i^{\alpha_i}$ is concave.

Proof. We shall show the statement for the case of m = 2. For points $x, y \in \mathbb{R}^2_+$ and a scalar $\lambda \in (0, 1)$, we consider $\lambda x + (1 - \lambda)y$. Define the quantities:

$$a_1 = (\lambda x_1)^{\alpha_1}, \quad a_2 = ((1 - \lambda)y_1)^{\alpha_1}, \quad b_1 = (\lambda x_2)^{\alpha_2}, \quad b_2 = ((1 - \lambda)y_2)^{\alpha_2},$$

Using Hölder's inequality, we obtain the following:

$$f(\lambda x + (1 - \lambda)y) = \left(a_1^{\frac{1}{\alpha_1}} + a_2^{\frac{1}{\alpha_1}}\right)^{\alpha_1} \left(b_1^{\frac{1}{\alpha_2}} + b_2^{\frac{1}{\alpha_2}}\right)^{\alpha_2}$$

$$\geq a_1b_1 + a_2b_2 = \lambda x_1^{\alpha_1}x_2^{\alpha_2} + (1 - \lambda)y_1^{\alpha_1}y_2^{\alpha_2}.$$

The assertion in the general case follows by induction. \Box

Theorem 4.23. If the functions $f_i : \mathbb{R}^n \to \mathbb{R}_+$, i = 1, ..., m are α_i -concave and α_i are such that $\sum_{i=1}^m \alpha_i^{-1} > 0$, then the function $g : \mathbb{R}^{nm} \to \mathbb{R}_+$, defined as $g(x) = \prod_{i=1}^m f_i(x_i)$

is
$$\gamma$$
-concave with $\gamma = \left(\sum_{i=1}^{m} \alpha_i^{-1}\right)^{-1}$

Proof. Fix points $x_1, x_2 \in \mathbb{R}^n_+$, a scalar $\lambda \in (0, 1)$ and set $x_\lambda = \lambda x_1 + (1 - \lambda)x_2$. By the generalized concavity of the functions $f_i, i = 1, ..., m$, we have the following inequality:

$$\prod_{i=1}^{m} f_i(x_{\lambda}) \ge \prod_{i=1}^{m} \left(\lambda f_i(x_1)^{\alpha_i} + (1-\lambda) f_i(x_2)^{\alpha_i} \right)^{1/\alpha_i}$$

We denote $y_{ij} = f_i(x_j)^{\alpha_i}$, j = 1, 2. Substituting into the last displayed inequality and raising both sides to power γ , we obtain

$$\left(\prod_{i=1}^m f_i(x_\lambda)\right)^{\gamma} \ge \prod_{i=1}^m \left(\lambda y_{i1} + (1-\lambda)y_{i2}\right)^{\gamma/\alpha_i}.$$

We continue the chain of inequalities using Lemma 4.22:

$$\prod_{i=1}^{m} \left(\lambda y_{i1} + (1-\lambda) y_{i2} \right)^{\gamma/\alpha_i} \ge \lambda \prod_{i=1}^{m} \left[y_{i1} \right]^{\gamma/\alpha_i} + (1-\lambda) \prod_{i=1}^{m} \left[y_{i2} \right]^{\gamma/\alpha_i}.$$

Putting the inequalities together and using the substitutions at the right hand side of the last inequality, we conclude that

$$\prod_{i=1}^{m} \left[f_1(x_{\lambda}) \right]^{\gamma} \ge \lambda \prod_{i=1}^{m} \left[f_i(x_1) \right]^{\gamma} + (1-\lambda) \prod_{i=1}^{m} \left[f_i(x_2) \right]^{\gamma},$$

as required.

In the special case, when the functions $f_i : \mathbb{R}^n \to \mathbb{R}$, i = 1, ..., k are concave, we can apply Theorem 4.23 consecutively, to conclude that $f_1 f_2$ is $\frac{1}{2}$ -concave and $f_1 ... f_k$ is $\frac{1}{k}$ -concave.

Lemma 4.24. If A is a symmetric positive definite matrix of size $n \times n$, then the function $A \mapsto \det(A)$ is $\frac{1}{n}$ -concave.

Proof. Consider two $n \times n$ symmetric positive definite matrices A, B and $\gamma \in (0, 1)$. We note that for every eigenvalue λ of A, $\gamma\lambda$ is an eigenvalue of γA , and, hence, $\det(\gamma A) = \gamma^n \det(A)$. We could apply the following Minkowski inequality for matrices:

$$\left[\det\left(A+B\right)\right]^{\frac{1}{n}} \ge \left[\det(A)\right]^{\frac{1}{n}} + \left[\det(B)\right]^{\frac{1}{n}},\tag{4.14}$$

which implies the $\frac{1}{n}$ -concavity of the function. As inequality (4.14) is not well-known, we provide a proof of it. First, we consider the case of diagonal matrices. In this case the determinants of A and B are products of their diagonal elements and inequality (4.14) follows from Lemma 4.22.

In the general case, let $A^{1/2}$ stand for the symmetric positive definite square-root of A and let $A^{-1/2}$ be its inverse. We have

$$\det (A+B) = \det \left(A^{1/2} A^{-1/2} (A+B) A^{-1/2} A^{1/2} \right)$$
$$= \det \left(A^{-1/2} (A+B) A^{-1/2} \right) \det(A)$$
$$= \det \left(I + A^{-1/2} B A^{-1/2} \right) \det(A)$$
(4.15)

Notice that $A^{-1/2}BA^{-1/2}$ is symmetric positive definite and, therefore, we can choose an $n \times n$ orthogonal matrix R, which diagonalizes it. We obtain

$$\det \left(I + A^{-1/2} B A^{-1/2} \right) = \det \left(R^{\mathsf{T}} \left(I + A^{-1/2} B A^{-1/2} \right) R \right)$$
$$= \det \left(I + R^{\mathsf{T}} A^{-1/2} B A^{-1/2} R \right).$$

At the right hand side of the equation, we have a sum of two diagonal matrices and we can apply inequality (4.14) for this case. We conclude that

$$\left[\det\left(I + A^{-1/2}BA^{-1/2}\right)\right]^{\frac{1}{n}} = \left[\det\left(I + R^{\mathsf{T}}A^{-1/2}BA^{-1/2}R\right)\right]^{\frac{1}{n}}$$
$$\geq 1 + \left[\det\left(R^{\mathsf{T}}A^{-1/2}BA^{-1/2}R\right)\right]^{\frac{1}{n}}$$
$$= 1 + \left[\det(B)\right]^{\frac{1}{n}}\left[\det(A)\right]^{-\frac{1}{n}}.$$

Combining this inequality with equation (4.15), we obtain (4.14) in the general case.

Example 4.25 (Dirichlet distribution continued) We return to Example 4.18. We see that the functions $x_i \mapsto x_i^{\beta_i}$ are $1/\beta_i$ -concave, provided that $\beta_i > 0$. Therefore, the density function of the Dirichlet distribution is a product of $\frac{1}{\alpha_i - 1}$ -concave functions, given that all parameters $\alpha_i > 1$. By virtue of Theorem 4.23, we obtain that this density is γ -concave with $\gamma = (\alpha_1 + \cdots + \alpha_m - s)^{-1}$ provided that $\alpha_i > 1$, $i = 1, \ldots, m$. Due to Corollary 4.16, the Dirichlet distribution is a $(\alpha_1 + \cdots + \alpha_m)^{-1}$ -concave probability measure.

Theorem 4.26. If the s-dimensional random vector Z has an α -concave probability distribution, $\alpha \in [-\infty, +\infty]$, and T is a constant $m \times s$ matrix, then the m-dimensional random vector Y = TZ has an α -concave probability distribution.

Proof. Let $A \subset \mathbb{R}^m$ and $B \subset \mathbb{R}^m$ be two Borel sets. We define

 $A_1 = \{ z \in \mathbb{R}^s : Tz \in A \} \text{ and } B_1 = \{ z \in \mathbb{R}^s : Tz \in B \}.$

The sets A_1 and B_1 are Borel measurable as well due to the continuity properties of the linear mapping $z \mapsto Tz$. Furthermore, for $\lambda \in [0, 1]$, the following inclusion holds:

$$\lambda A_1 + (1 - \lambda)B_1 \subset \left\{ z \in \mathbb{R}^s : Tz \in \lambda A + (1 - \lambda)B \right\}.$$

Denoting P_Z and P_Y the probability measure of Z and Y respectively, we obtain,

$$P_Y \{ \lambda A + (1 - \lambda)B \} \ge P_Z \{ \lambda A_1 + (1 - \lambda)B_1 \}$$
$$\ge m_\alpha (P_Z \{A_1\}, P_Z \{B_1\}, \lambda)$$
$$= m_\alpha (P_Y \{A\}, P_Y \{B\}, \lambda).$$

This completes the proof. \Box

Example 4.27 A univariate *Gamma distribution* is given by the following probability density function

$$f(z) = \begin{cases} \frac{\lambda^{\vartheta} z^{\vartheta - 1} e^{-\lambda z}}{\Gamma(\vartheta)}, & \text{for } z > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $\lambda > 0$ and $\vartheta > 0$ are constants. For $\lambda = 1$ the distribution is the standard gamma distribution. If a random variable Y has the gamma distribution, then ϑY has the standard gamma distribution. It is not difficult to check that this density function is log-concave, provided $\vartheta \ge 1$.

A multivariate gamma distribution can be defined by a certain linear transformation of m independent random variables Z_1, \ldots, Z_m $(1 \le m \le 2^s - 1)$, that have the standard gamma distribution. Let an $s \times m$ matrix A with 0–1 elements be given. Setting $Z = (Z_1, \ldots, Z_{2^s-1})$, we define

Y = AZ.

The random vector Y has a multivariate standard gamma distribution.

We observe that the distribution of the vector Z is log-concave by virtue of Lemma 4.13. Hence, the *s*-variate standard gamma distribution is log-concave by virtue of Theorem 4.26.

Example 4.28 The Wishart distribution arises in estimation of covariance matrices and can be considered as a multidimensional version of the χ^2 - distribution. More precisely, let us assume that Z is an s-dimensional random vector having multivariate normal distribution with a nonsingular covariance matrix Σ and expectation μ . Given a sample with observed values z^1, \ldots, z^N from this distribution, we consider the matrix:

$$\sum_{i=1}^{N} (z^i - \bar{z})(z^i - \bar{z})^{\mathsf{T}},$$

where \bar{z} is the sample mean. This matrix has the Wishart distribution with N - 1 degrees of freedom. We denote the trace of a matrix A by tr(A).

If N > s, the Wishart distribution is a continuous distribution on the space of symmetric square matrices with probability density function defined by

$$f(A) = \begin{cases} \frac{\det(A)^{\frac{N-s-2}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}(\Sigma^{-1}A)\right)}{2^{\frac{N-1}{2}s} \pi^{\frac{s(s-1)}{4}} \det(\Sigma)^{\frac{N-1}{2}} \prod_{i=1}^{s} \Gamma\left(\frac{N-i}{2}\right)}, & \text{for } A \text{ positive definite,} \\ 0, & \text{otherwise.} \end{cases}$$

If s = 1 and $\Sigma = 1$ this density becomes the χ^2 - distribution density with N - 1 degrees of freedom.

If A_1 and A_2 are two positive definite matrices and $\lambda \in (0, 1)$, then the matrix $\lambda A_1 + (1 - \lambda)A_2$ is positive definite as well. Using Lemma 4.24 and Lemma 4.8 we conclude that function $A \mapsto \ln \det(A)$, defined on the set of positive definite Hermitian matrices, is concave. This implies that, if $N \ge s + 2$, then f is a log-concave function on the set of symmetric positive definite matrices. If N = s + 1, then f is a log-convex on the convex set of symmetric positive definite matrices.

Recall that a function $f : \mathbb{R}^s \to \mathbb{R}$ is called *regular in the sense of Clarke* or Clarke-regular if

$$f'(x;d) = \lim_{t \downarrow 0} \frac{1}{t} \left[f(x+td) - f(x) \right] = \lim_{y \to x, t \downarrow 0} \frac{1}{t} \left[f(y+td) - f(y) \right].$$

It is known that convex functions are regular in this sense. We call a concave function f regular with the understanding that the regularity requirement applies to -f. In this case, we have $\partial^{\circ}(-f)(x) = -\partial^{\circ}f(x)$, where $\partial^{\circ}f(x)$ refers to the collection of the Clarke generalized gradients of f at the point x. For convex functions $\partial^{\circ}f(x) = \partial f(x)$.

Theorem 4.29. If $f : \mathbb{R}^s \to \mathbb{R}$ is α -concave ($\alpha \in \mathbb{R}$) on some open set $U \subset \mathbb{R}^s$ and f(x) > 0 for all $x \in U$, then f(x) is locally Lipschitz continuous, directionally differentiable, and Clarke-regular. Its Clarke generalized gradients are given by the formula:

$$\partial^{\circ} f(x) = \begin{cases} \frac{1}{\alpha} [f(x)]^{1-\alpha} \partial [(f(x))^{\alpha}], & \text{if } \alpha \neq 0, \\ f(x) \partial (\ln f(x)), & \text{if } \alpha = 0. \end{cases}$$

Proof. If f is an α -concave function, then an appropriate transformation of f, is a concave function on U. We define

$$\bar{f}(x) = \begin{cases} (f(x))^{\alpha}, & \text{ if } \alpha \neq 0\\ \ln f(x), & \text{ if } \alpha = 0. \end{cases}$$

If $\alpha < 0$, then $f^{\alpha}(\cdot)$ is convex. This transformation is well-defined on the open subset U since f(x) > 0 for $x \in U$, and, thus, $\overline{f}(x)$ is subdifferentiable at any $x \in U$. Further, we represent f as follows:

$$f(x) = \begin{cases} \left(\bar{f}(x)\right)^{1/\alpha}, & \text{if } \alpha \neq 0, \\ \exp(\bar{f}(x)), & \text{if } \alpha = 0. \end{cases}$$

In this representation, f is a composition of a continuously differentiable function and a concave function. By virtue of Clarke [45, Theorem 2.3.9(3)], the function f is locally Lipschitz continuous, directionally differentiable, and Clarke-regular. Its Clarke generalized gradient set is given by the formula:

$$\partial^{\circ} f(x) = \begin{cases} \frac{1}{\alpha} \left(\bar{f}(x) \right)^{1/\alpha - 1} \partial \bar{f}(x), & \text{if } \alpha \neq 0, \\ \exp(\bar{f}(x)) \partial \bar{f}(x), & \text{if } \alpha = 0. \end{cases}$$

Substituting the definition of \overline{f} yields the result. \Box

For a function $f : \mathbb{R}^n \to \mathbb{R}$, we consider the set of points at which it takes positive values. It is denoted by dompos f, i.e.,

dompos
$$f := \{x \in \mathbb{R}^n : f(x) > 0\}.$$

Recall that $\mathcal{N}_X(x)$ denotes the normal cone to the set X at $x \in X$.

Definition 4.30. We call a point $\hat{x} \in \mathbb{R}^n$ a stationary point of an α -concave function f, if there is a neighborhood U of \hat{x} such that f is Lipschitz continuous on U, and $0 \in \partial^{\circ} f(\hat{x})$. Furthermore, for a convex set $X \subset$ domposf, we call $\hat{x} \in X$ a stationary point of fon X, if there is a neighborhood U of \hat{x} such that f is Lipschitz continuous on U and $0 \in \partial^{\circ} f_X(\hat{x}) + \mathcal{N}_X(\hat{x})$.

We observe that certain properties of the maxima of concave functions extend to generalized concave functions.

Theorem 4.31. Let f be an α -concave function f and the set $X \subset$ dompos f be convex. Then all the stationary points of f on X are global maxima and the set of global maxima of f on X is convex.

Proof. First, assume that $\alpha = 0$. Let \hat{x} be a stationary point of f on X. This implies that

$$0 \in f(\hat{x})\partial(\ln f(\hat{x})) + \mathcal{N}_X(\hat{x}), \tag{4.16}$$

Using that $f(\hat{x}) > 0$, we obtain

$$0 \in \partial \left(\ln f(\hat{x}) \right) + \mathcal{N}_X(\hat{x}), \tag{4.17}$$

As the function $f(x) = \ln f(x)$ is concave, this inclusion implies that \hat{x} is a global maximal point of \bar{f} on X. By the monotonicity of $\ln(\cdot)$, we conclude that \hat{x} is a global maximal point of f on X. If a point $\tilde{x} \in X$ is a maximal point of \bar{f} on X, then inclusion (4.17) is satisfied. It entails (4.16) as $X \subset \text{dompos} f$, and, therefore, \tilde{x} is a stationary point of f on X. Therefore, the set of maximal points of f on X is convex because this is the set of maximal points of the concave function \bar{f} .

In the case of $\alpha \neq 0$, the statement follows by the same line of argument using the function $\bar{f}(x) = [f(x)]^{\alpha}$. \Box

Another important property of α -concave measures is the existence of so-called floating body for all probability levels $p \in (\frac{1}{2}, 1)$.

Definition 4.32. A measure P on \mathbb{R}^s has a floating body at level p > 0 if there exists a convex body $C_p \subset \mathbb{R}^s$ such for all vectors $z \in \mathbb{R}^s$:

$$P\left\{x \in \mathbb{R}^s : z^\mathsf{T} x \ge \mathsf{S}_{C_n}(z)\right\} = 1 - p,$$

where $S_{C_p}(\cdot)$ is the support function of the set C_p . The set C_p is called the floating body of P at level p.

Symmetric log-concave measures have floating bodies. We formulate this result of Meyer and Reisner [155] without proof.

Theorem 4.33. Any non-degenerate probability measure with symmetric log-concave density function has a floating body C_p at all levels $p \in (\frac{1}{2}, 1)$.

We see that α -concavity as introduced so far implies continuity of the distribution function. As empirical distributions are very important in practical applications, we would

like to find a suitable generalization of this notion applicable to discrete distributions. For this purpose, we introduce the following notion.

Definition 4.34. A distribution function F is called α -concave on the set $\mathcal{A} \subset \mathbb{R}^s$ with $\alpha \in [-\infty, \infty]$, if

 $F(z) \ge m_{\alpha} (F(x), F(y), \lambda)$

for all $z, x, y \in A$ and $\lambda \in (0, 1)$ such that $z \ge \lambda x + (1 - \lambda)y$.

Observe that if $\mathcal{A} = \mathbb{R}^s$, then this definition coincides with the usual definition of α -concavity of a distribution function.

To illustrate the relation between Definition 4.7 and Definition 4.34 let us consider the case of integer random vectors which are roundups of continuously distributed random vectors.

Remark 7. If the distribution function of a random vector Z is α -concave on \mathbb{R}^s then the distribution function of Y = [Z] is α -concave on \mathbb{Z}^s .

This property follows from the observation that at integer points both distribution functions coincide.

Example 4.35 Every distribution function of an *s*-dimensional *binary random vector* is α -concave on \mathbb{Z}^s for all $\alpha \in [-\infty, \infty]$.

Indeed, let x and y be binary vectors, $\lambda \in (0, 1)$, and $z \ge \lambda x + (1 - \lambda)y$. As z is integer and x and y binary, then $z \ge x$ and $z \ge y$. Hence, $F(z) \ge \max\{F(x), F(y)\}$ by the monotonicity of the cumulative distribution function. Consequently, F is ∞ -concave. Using Lemma 4.8 we conclude that F_Z is α -concave for all $\alpha \in [-\infty, \infty]$.

For a random vector with independent components, we can relate concavity of the marginal distribution functions to the concavity of the joint distribution function. Note that the statement applies not only to discrete distributions, as we can always assume that the set A is the whole space or some convex subset of it.

Theorem 4.36. Consider the s-dimensional random vector $Z = (Z^1, \dots, Z^L)$, where the sub-vectors Z^l , $l = l, \dots, L$, are s_l -dimensional and $\sum_{l=1}^L s_l = s$. Assume that Z^l , $l = l, \dots, L$ are independent and that their marginal distribution functions $F_{Z^l} : \mathbb{R}^{s_l} \to [0, 1]$ are α_l -concave on the sets $\mathcal{A}_l \subset \mathbb{Z}^{s_l}$. Then the following statements hold true:

- 1. If $\sum_{l=1}^{L} \alpha_l^{-1} > 0$, $l = 1, \dots, L$, then F_Z is α -concave on $\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_L$ with $\alpha = (\sum_{l=1}^{L} \alpha_l^{-1})^{-1}$;
- 2. If $\alpha_l = 0$, $l = 1, \dots, L$, then F_Z is log-concave on $\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_L$.

Proof. The proof of the first statement follows by virtue of Theorem 4.23 using the monotonicity of the cumulative distribution function.

For the second statement consider $\lambda \in (0,1)$ and points $x = (x^1, \dots, x^L) \in \mathcal{A}$, $y = (y^1, \dots, y^L) \in \mathcal{A}$, and $z = (z^1, \dots, z^L) \in \mathcal{A}$, such that $z \ge \lambda x + (1 - \lambda)y$. Using

the monotonicity of the function $\ln(\cdot)$ and of $F_Z(\cdot)$, along with the log-concavity of the marginal distribution functions, we obtain the following chain of inequalities:

$$\ln[F_{Z}(z)] \geq \ln[F_{Z}(\lambda x + (1 - \lambda)y)] = \sum_{l=1}^{L} \ln\left[F_{Z^{l}}(\lambda x^{l} + (1 - \lambda)y^{l})\right]$$
$$\geq \sum_{l=1}^{L} \left[\lambda \ln[F_{Z^{l}}(x^{l})] + (1 - \lambda)\ln[F_{Z^{l}}(y_{l})]\right]$$
$$= \lambda \sum_{l=1}^{L} \ln[F_{Z^{l}}(x^{l})] + (1 - \lambda)\sum_{l=1}^{L} \ln[F_{Z^{l}}(y^{l})]$$
$$= \lambda[F_{Z}(x)] + (1 - \lambda)[F_{Z}(y)].$$

This concludes the proof. \Box

For integer random variables our definition of α -concavity is related to log-concavity of sequences.

Definition 4.37. A sequence p_k , $k \in \mathbb{Z}$, is called log-concave, if

$$p_k^2 \ge p_{k-1}p_{k+1}$$
 for all $k \in \mathbb{Z}$.

The following property is established in Prékopa [195, Thm 4.7.2].

Theorem 4.38. Suppose that for an integer random variable Y the probabilities $p_k = \Pr\{Y = k\}, k \in \mathbb{Z} \text{ form a log-concave sequence. Then the distribution function of Y is <math>\alpha$ -concave on \mathbb{Z} for every $\alpha \in [-\infty, 0]$.

4.2.2 Convexity of probabilistically constrained sets

One of the most general results in the convexity theory of probabilistic optimization is the following theorem.

Theorem 4.39. Let the functions $g_j : \mathbb{R}^n \times \mathbb{R}^s$, $j \in \mathcal{J}$, be quasi-concave. If $Z \in \mathbb{R}^s$ is a random vector that has an α -concave probability distribution, then the function

$$G(x) = \Pr\{g_j(x, Z) \ge 0, \ j \in \mathcal{J}\}\tag{4.18}$$

is α -concave on the set

$$D = \{ x \in \mathbb{R}^n : \exists z \in \mathbb{R}^s \text{ such that } g_j(x, z) \ge 0, \ j \in \mathcal{J} \}.$$

Proof. Given the points $x_1, x_2 \in D$ and $\lambda \in (0, 1)$, we define the sets

$$A_i = \{ z \in \mathbb{R}^s : g_i(x_i, z) \ge 0, j \in \mathcal{J} \}, \quad i = 1, 2,$$

and $B = \lambda A_1 + (1 - \lambda)A_2$. We consider

$$G(\lambda x_1 + (1-\lambda)x_2) = \mathsf{Pr}\{g_j(\lambda x_1 + (1-\lambda)x_2, Z) \ge 0, \ j \in \mathcal{J}\}.$$

If $z \in B$, then points $z_i \in A_i$, i = 1, 2, exist such that $z = \lambda z_1 + (1 - \lambda)z_2$. Die to the quasi-concavity of g_j , we obtain that

$$g_j(\lambda x_1 + (1 - \lambda)x_2, \lambda z_1 + (1 - \lambda)z_2) \ge \min\{g_j(x_1, z_1), g_j(x_2, z_2)\} \ge 0$$
 for all $j \in \mathcal{J}$.

These inequalities imply that $z \in \{z \in \mathbb{R}^s : g_j(\lambda x_1 + (1 - \lambda)x_2, z) \ge 0, j \in \mathcal{J}\}$, which entails that $\lambda x_1 + (1 - \lambda)x_2 \in D$ and that

$$G(\lambda x_1 + (1 - \lambda)x_2) \ge \mathsf{Pr}\{B\}.$$

Using the α -concavity of the measure, we conclude that

$$G(\lambda x_1 + (1 - \lambda)x_2) \ge \Pr\{B\} \ge m_{\alpha}\{\Pr\{A_1\}, \Pr\{A_2\}, \lambda\} = m_{\alpha}\{G(x_1), G(x_2), \lambda\},$$

as desired.

Example 4.40 (The log-normal distribution) The probability density function of the onedimensional log-normal distribution with parameters μ and σ is given by

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma x}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), & \text{if } x > 0, \\ 0, & \text{otherwise} \end{cases}$$

This density is neither log-concave, nor log-convex. However, we can show that the cumulative distribution function is log-concave. We consider the multidimensional case. The *m*-dimensional random vector *Z* has the log-normal distribution, if the vector $Y = (\ln Z_1, \ldots, \ln Z_m)^T$ has a multivariate normal distribution. Recall that the normal distribution is log-concave. The distribution function of *Z* at a point $z \in \mathbb{R}^m$, z > 0, can be written as

$$F_Z(z) = \Pr\{Z_1 \le z_1, \dots, Z_m \le z_m\} = \Pr\{z_1 - e_1^Y \ge 0, \dots, z_m - e_m^Y \ge 0\}.$$

We observe that the assumptions of Theorem 4.39 are satisfied for the probability function on the right hand side. Thus, F_Z is a log-concave function.

As a consequence, under the assumptions of Theorem 4.39, we obtain convexity statements for sets described by probabilistic constraints.

Corollary 4.41. Assume that the functions $g_j(\cdot, \cdot)$, $j \in \mathcal{J}$, are quasi-concave jointly in both arguments, and that $Z \in \mathbb{R}^s$ is a random variable that has an α -concave probability distribution. Then the following set is convex and closed:

$$X_0 = \{ x \in \mathbb{R}^n : \Pr\{g_i(x, Z) \ge 0, \ i = 1, \dots, m\} \ge p \}.$$
(4.19)

Proof. Let G(x) be defined as in (4.18), and let $x_1, x_2 \in X_0, \lambda \in [0, 1]$. We have

$$G(\lambda x_1 + (1 - \lambda)x_2) \ge m_{\alpha}\{G(x_1), G(x_2), \lambda\} \ge \min\{G(x_1), G(x_2)\} \ge p.$$

The closedness of the set follows from the continuity of α -concave functions.

We consider the case of a separable mapping g, when the random quantities appear only on the right hand side of the inequalities. Observe that the function $(x, z) \mapsto g(x) - z$ is quasi-concave in both arguments whenever the function g is quasi-concave. Directly from Corollary 4.41, we obtain

Corollary 4.42. Let the mapping $g : \mathbb{R}^n \to \mathbb{R}^m$ be such that each component g_i is a quasi-concave function. Assume that the random vector Z has α -concave distribution for some $\alpha \in \mathbb{R}$. Then the set

$$X_0 = \left\{ x \in \mathbb{R}^n : \Pr\{g(x) \ge Z\} \ge p \right\}$$
(4.20)

is convex.

We could adopt a weaker requirement on the probability distribution of the random vector Z at the expense of a concavity assumption for the mapping g.

Theorem 4.43. Assume that each component g_i of the mapping $g : \mathbb{R}^n \to \mathbb{R}^m$ is a concave function and let the random vector Z have α -concave distribution function for some $\alpha \in \overline{\mathbb{R}}$. Then the set X_0 , defined in (4.20) is convex.

Proof. Indeed, the probability function appearing in the definition of the set X_0 can be described as follows:

$$G(x) = \Pr\{g(x) \ge Z\} = F_Z(g(x))$$

Due to Theorem 4.20, the function $F_Z \circ g$ is α -concave. The convexity of X_0 follows the same way as in Corollary 4.41. \Box

For vectors with independent components, we can state the following result.

Corollary 4.44. Assume that each component g_i of the mapping $g : \mathbb{R}^n \to \mathbb{R}^m$ is a concave function. If the random vector Z has independent components and the one-dimensional marginal distribution functions F_{Z_i} , i = 1, ..., m are α_i -concave with $\sum_{i=1}^k \alpha_i^{-1} > 0$, then the set X_0 , defined in (4.20) is convex.

Proof. We can express the distribution function $\mathcal{F}_Z(g(x)) = prod_{i=1}^m F_{Z_i}(g_i(x_i))$. The functions $F_{Z_i} \circ g_i$ are α_i -concave due to Theorem 4.20. Using Theorem 4.23, we conclude that $F_Z \circ g$ is γ -concave with $\gamma = \left(\sum_{i=1}^k \alpha_i^{-1}\right)^{-1}$. Hence, the set X_0 is convex. \Box

Under the same assumptions the set determined by the first order stochastic dominance constraint with respect to any random variable Y is convex and closed. **Theorem 4.45.** Assume that $g(\cdot, \cdot)$ is a quasi-concave function jointly in both arguments, and that Z has an α -concave distribution. Then the following sets are convex and closed:

$$X_d = \left\{ x \in \mathbb{R}^n : g(x, Z) \succeq_{(1)} Y \right\},$$

$$X_c = \left\{ x \in \mathbb{R}^n : \Pr\{g(x, Z) \ge \eta\} \ge \Pr\{Y \ge \eta\}, \ \forall \eta \in [a, b] \right\}$$

Proof. Let us fix $\eta \in \mathbb{R}$ and observe that the relation $g(x, Z) \succeq_{(1)} Y$ can be formulated in the following equivalent way:

$$\Pr\{g(x,Z) \ge \eta\} \ge \Pr\{Y \ge \eta\}$$
 for all $\eta \in \mathbb{R}$.

Therefore, the first set can be defined as follows:

$$X_d = \left\{ x \in \mathbb{R}^n : \Pr\{g(x, Z) - \eta \ge 0\} \ge \Pr\{Y \ge \eta\} \ \forall \eta \in \mathbb{R} \right\}.$$

For any $\eta \in \mathbb{R}$, we define the set

$$X(\eta) = \left\{ x \in \mathbb{R}^n : \Pr\{g(x, Z) - \eta \ge 0\} \ge \Pr\{Y \ge \eta\} \right\}.$$

This set is convex and closed by virtue of Corollary 4.41. The set X_d is the intersection of the sets $X(\eta)$ for all $\eta \in \mathbb{R}$, and, therefore, it is convex and closed as well. Analogously, the set X_c is convex and closed as $X_c = \bigcap_{\eta \in [a,b]} X(\eta)$. \Box

Let us observe that affine in each argument functions $g_i(x, z) = z^T x + b_i$ are not necessarily quasi-concave in both arguments (x, z). We can apply Theorem 4.39 to conclude that the set

$$X_{l} = \left\{ x \in \mathbb{R}^{n} : \Pr\{x^{\mathsf{T}}a_{i} \le b_{i}(Z), \ i = 1, \dots, m\} \ge p \right\}$$
(4.21)

is convex if $a_i, i = 1, ..., m$ are deterministic vectors. We have the following.

Corollary 4.46. The set X_l is convex whenever $b_i(\cdot)$ are quasi-concave functions and Z has a quasi-concave probability distribution function.

Example 4.47 (Vehicle Routing continued) We return to Example 4.1. The probabilistic constraint (4.3) has the form:

$$\Pr\{Tx \ge Z\} \ge p.$$

If the vector Z of a random demand has an α -concave distribution function, then this constraint defines a convex set. For example, this is the case if each component Z_i has a uniform distribution and the components (the demand on each arc) are independent of each other. If the vector Z has a multi-variate log-normal distribution, then this constraint defines a convex set as well.

If the functions g_i are not separable, we can invoke Theorem 4.33.

Theorem 4.48. Let $p_i \in (0.5, 1)$ for all i = 1, ..., n. The following set is convex

$$X_p = \{ x \in \mathbb{R}^n : P_{Z_i} \{ x^{\mathsf{T}} Z_i \le b_i \} \ge p_i, \, i = 1, \dots, m \},$$
(4.22)

whenever the vectors Z_i have non-degenerated log-concave probability distribution, symmetric around some point $\mu_i \in \mathbb{R}^n$, i = 1, ..., m.

Proof. If the random vector Z_i has a non-degenerated log-concave probability distribution, which is symmetric around some point $\mu_i \in \mathbb{R}^n$, then the vector $Y_i = Z_i - \mu_i$ has a symmetric and non degenerate log-concave distribution.

Given points $x_1, x_2 \in X_p$ and a number $\lambda \in [0, 1]$, we define

$$K_i(x) = \{ a \in \mathbb{R}^n : a^\mathsf{T} x \le b_i \}, \quad i = 1, \dots, n.$$

Let us fix an index *i*. The probability distribution of Y_i satisfies the assumptions of Theorem 4.33. Thus, there is a convex set C_{p_i} such that any supporting plane defines a half-plane containing probability p_i :

$$P_{Y_i}\left\{y \in \mathbb{R}^n : y^{\mathsf{T}} x \le \mathsf{s}_{C_{p_i}}(x)\right\} = p_i, \quad \forall x \in \mathbb{R}^n.$$

Thus,

$$P_{Z_i}\left\{z \in \mathbb{R}^n : z^\mathsf{T} x \le \mathbf{s}_{C_{p_i}}(x) + \mu_i^\mathsf{T} x\right\} = p_i, \quad \forall x \in \mathbb{R}^n.$$
(4.23)

Since $P_{Z_i}\{K_i(x_1)\} \ge p_i$ and $P_{Z_i}\{K_i(x_2)\} \ge p_i$ by assumption, then

$$K_i(x_j) \subset \left\{ z \in \mathbb{R}^n : z^\mathsf{T} x \leq \mathsf{s}_{C_{p_i}}(x) + \langle \mu_i, x \rangle \right\}, \quad j = 1, 2,$$
$$b_i \geq \mathsf{s}_{C_{p_i}}(x_1) + \mu_i^\mathsf{T} x_j, \quad j = 1, 2.$$

The properties of the support function entail that

$$\begin{split} b_i &\geq \lambda \big[\mathsf{S}_{_{Cp_i}}(x_1) + \mu_i^\mathsf{T} x_1 \big] + (1-\lambda) \big[\mathsf{S}_{_{Cp_i}}(x_2) + \mu_i^\mathsf{T} x_2 \big] \\ &= \mathsf{S}_{_{Cp_i}}(\lambda x_1) + \mathsf{S}_{_{Cp_i}}((1-\lambda) x_2) + \mu_i^\mathsf{T} \lambda x_1 + (1-\lambda) x_2 \\ &\geq \mathsf{S}_{_{Cp_i}}(\lambda x_1 + (1-\lambda) x_2) + \mu_i^\mathsf{T} \lambda x_1 + (1-\lambda) x_2. \end{split}$$

Consequently, the set $K_i(x_\lambda)$ with $x_\lambda = \lambda x_1 + (1 - \lambda)x_2$ contains the set

$$\left\{z \in \mathbb{R}^n : z^\mathsf{T} x_\lambda \le \mathsf{S}_{C_{p_i}}(x_\lambda) + \mu_i^\mathsf{T} x_\lambda\right\}$$

and, therefore, using equation (4.23) we obtain that

$$P_{Z_i}\left\{K_i(\lambda x_1 + (1-\lambda)x_2)\right\} \ge p_i.$$

Since *i* was arbitrary, we obtain that $\lambda x_1 + (1 - \lambda)x_2 \in X_p$. \Box

Example 4.49 (Portfolio optimization continued) Let us consider the portfolio example 4.2 and assume that the random vector $R = (R_1, \ldots, R_n)^T$ has a multidimensional normal distribution or a uniform distribution. The value-at-risk constraint can be written as

$$\Pr\left\{-R^{\mathsf{T}}x \le \eta\right\} \ge p$$

with $p \in (0.5, 1)$. The feasible set in this example is convex by virtue of Theorem 4.48 because both distributions are symmetric and log-concave.

An important relation is established between the sets constrained by first and second order stochastic dominance relation to a benchmark random variable (see Dentcheva and Ruszczyński [63]). We denote the space of integrable random variables by $\mathcal{L}_1(\Omega, \mathcal{F}, P)$ and set:

$$A_1(Y) = \{ X \in \mathcal{L}_1(\Omega, \mathcal{F}, P) : X \succeq_{(1)} Y \}, A_2(Y) = \{ X \in \mathcal{L}_1(\Omega, \mathcal{F}, P) : X \succeq_{(2)} Y \}.$$

Proposition 4.50. For every $Y \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$ the set $A_2(Y)$ is convex and closed.

Proof. By changing the order of integration in the definition of the second order function $F^{(2)}$, we obtain

$$F_X^{(2)}(\eta) = \mathbb{E}[(\eta - X)_+]. \tag{4.24}$$

Therefore, an equivalent representation of the second order stochastic dominance relation is given by the relation

$$\mathbb{E}[(\eta - X)_{+}] \le \mathbb{E}[(\eta - Y)_{+}] \quad \text{for all } \eta \in \mathbb{R}.$$
(4.25)

For every $\eta \in \mathbb{R}$ the functional $X \to \mathbb{E}[(\eta - X)_+]$ is convex and continuous in $\mathcal{L}_1(\Omega, \mathcal{F}, P)$, as a composition of a linear function, the "max" function and the expectation operator. Consequently, the set $A_2(Y)$ is convex and closed. \Box

The set $A_1(Y)$ is closed, because convergence in \mathcal{L}_1 implies convergence in probability, but it is not convex in general.

Example 4.51 Suppose that $\Omega = \{\omega_1, \omega_2\}$, $P\{\omega_1\} = P\{\omega_2\} = 1/2$ and $Y(\omega_1) = -1$, $Y(\omega_2) = 1$. Then $X_1 = Y$ and $X_2 = -Y$ both dominate Y in the first order. However, $X = (X_1 + X_2)/2 = 0$ is not an element of $A_1(Y)$ and, thus, the set $A_1(Y)$ is not convex. We notice that X dominates Y in the second order.

First order dominance relation implies the second order dominance as seen directly from the definitions of these relations. Hence, $A_1(Y) \subset A_2(Y)$. We have demonstrated that the set $A_2(Y)$ is convex; therefore, we also have

$$\operatorname{conv}(A_1(Y)) \subset A_2(Y). \tag{4.26}$$

We find sufficient conditions for the opposite inclusion.

Theorem 4.52. Assume that $\Omega = \{\omega_1, \ldots, \omega_N\}$, \mathcal{F} contains all subsets of Ω , and $P\{\omega_k\} = 1/N$, $k = 1, \ldots, N$. If $Y : (\Omega, \mathcal{F}, P) \to \mathbb{R}$ is a random variable, then

$$\operatorname{conv}(A_1(Y)) = A_2(Y).$$

Proof. We only need prove the inverse inclusion to (4.26). For this purpose, suppose $X \in A_2(Y)$. Under the assumptions of the theorem, we can identify X and Y with vectors $x = (x_1, \ldots, x_N)$ and $y = (y_1, \ldots, y_N)$ such that $x_i = X(\omega_i)$ and $y_i = Y(\omega_i)$,

i = 1, ..., N. As the probabilities of all elementary events are equal, the second order stochastic dominance relation coincides with the concept of *weak majorization*. Denoting the *k*th smallest component of *x* by $x_{[k]}$, the vector *x* majorizes the vector *y* weakly if the following system of inequalities is satisfied:

$$\sum_{k=1}^{l} x_{[k]} \ge \sum_{k=1}^{l} y_{[k]}, \quad l = 1, \dots, N.$$

As established by Hardy, Littlewood and Polya [96], weak majorization is equivalent to the existence of a doubly stochastic matrix Π such that

$$x \ge \Pi y$$

By virtue of Birkhoff's Theorem [23], permutation matrices Q^1, \ldots, Q^M and nonnegative reals $\alpha_1, \ldots, \alpha_M$ totaling 1 exist, such that

$$\Pi = \sum_{j=1}^{M} \alpha_j Q^j.$$

Setting $z^j = Q^j y$, we conclude that

$$x \ge \sum_{j=1}^{M} \alpha_j z^j$$

Identifying random variables Z^j on (Ω, \mathcal{F}, P) with the vectors z^j , we also see that

$$X(\omega) \ge \sum_{j=1}^{M} \alpha_j Z^j(\omega),$$

for all $\omega \in \Omega$. Since each vector z^j is a permutation of y and the probabilities are equal, the distribution of Z^j is identical to the distribution of Y. Thus

$$Z^j \succeq_{(1)} Y, \quad j = 1, \dots, M$$

Let us define

$$\hat{Z}^{j}(\omega) = Z^{j}(\omega) + \left(X(\omega) - \sum_{k=1}^{M} \alpha_{k} Z^{k}(\omega)\right), \quad \omega \in \Omega, \quad j = 1, \dots, M$$

Then the last two inequalities render $\hat{Z}^j \in A_1(Y)$, j = 1, ..., M, and

$$X(\omega) = \sum_{j=1}^{M} \alpha_j \hat{Z}^j(\omega),$$

as required. \Box

This result does not extend to general probability spaces, as the following example illustrates.

Example 4.53 We consider the probability space $\Omega = \{\omega_1, \omega_2\}, P\{\omega_1\} = 1/3, P\{\omega_2\} = 2/3$. The benchmark variable Y is defined as $Y(\omega_1) = -1, Y(\omega_2) = 1$. It is easy to see that $X \succeq_{(1)} Y$ if and only if $X(\omega_1) \ge -1$ and $X(\omega_2) \ge 1$. Thus, $A_1(Y)$ is a convex set.

Now, consider the random variable $Z = \mathbb{E}[Y] = 1/3$. It dominates Y in the second order, but it does not belong to conv $A_1(Y) = A_1(Y)$.

It follows from this example that the probability space must be sufficiently rich to observe our phenomenon. If we could define a new probability space $\Omega' = \{\omega_1, \omega_{21}, \omega_{22}\}$, in which the event ω_2 is split in two equally likely events ω_{21}, ω_{22} , then we could use Theorem 4.52 to obtain the equality conv $A_1(Y) = A_2(Y)$. In the context of optimization however, the probability space has to be fixed at the outset and we are interested in sets of random variables as elements of $\mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^n)$, rather than in sets of their distributions.

Theorem 4.54. Assume that the probability space (Ω, \mathcal{F}, P) is atomless. Then

$$A_2(Y) = \operatorname{cl}\{\operatorname{conv}(A_1(Y))\}.$$

Proof. If the space (Ω, \mathcal{F}, P) is atomless, we can partition Ω into N disjoint subsets, each of the same P-measure 1/N, and we verify the postulated equation for random variables which are piecewise constant on such partitions. This reduces the problem to the case analyzed in Theorem 4.52. Passing to the limit with $N \to \infty$, we obtain the desired result. We refer the interested reader to Dentcheva and Ruszczyński [65] for technical details of the proof. \Box

4.2.3 Connectedness of probabilistically constrained sets

Let $\mathcal{X} \subset \mathbb{R}^n$ be a closed convex set. In this section we focus on the following set:

$$X = \left\{ x \in \mathcal{X} : \Pr\left[g_j(x, Z) \ge 0, \ j \in \mathcal{J}\right] \ge p \right\},\tag{4.27}$$

where \mathcal{J} is an arbitrary index set. The functions $g_i : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}$ are continuous, Z is an *s*-dimensional random vector, and $p \in (0, 1)$ is a prescribed probability. It will be demonstrated later (Lemma 4.63) that the probabilistically constrained set X with separable functions g_j is a union of cones intersected by \mathcal{X} . Thus, the set X could be disconnected. The following result provides a sufficient condition for X to be topologically connected. A more general version of this result is proved in Henrion [100].

Theorem 4.55. Assume that the functions $g_j(\cdot, Z)$, $j \in \mathcal{J}$, are quasi-concave and that they satisfy the following condition: for all $x^1, x^2 \in \mathbb{R}^n$, a point $x^* \in \mathcal{X}$ exists such that

$$g_j(x^*, z) \ge \min\{g_j(x^1, z), g_j(x^2, z)\} \quad \forall z \in \mathbb{R}^s, \ \forall j \in \mathcal{J}.$$

Then the set X defined in (4.27) is connected.

Proof. Let $x^1, x^2 \in X$ be arbitrary points. We construct a path joining the two points, which is contained entirely in X. Let $x^* \in \mathcal{X}$ be the point that exists according to the

assumption. We set

$$\pi(t) = \begin{cases} (1-2t)x^1 + 2tx^*, & \text{for } 0 \le t \le 1/2, \\ 2(1-t)x^* + (2t-1)x^2, & \text{for } 1/2 < t \le 1. \end{cases}$$

First, we observe that $\pi(t) \in \mathcal{X}$ for every $t \in [0, 1]$ since $x^1, x^2, x^* \in \mathcal{X}$ and the set \mathcal{X} is convex. Furthermore, the quasi-concavity of $g_j, j \in \mathcal{J}$, and the assumptions of the theorem imply for every j and for $0 \le t \le 1/2$ the following inequality:

$$g_j((1-2t)x^1+2tx^*,z) \ge \min\{g_j(x^1,z),g_j(x^*,z)\} = g_j(x^1,z).$$

Therefore,

$$\Pr\{g_j(\pi(t), Z) \ge 0, \ j \in \mathcal{J}\} \ge \Pr\{g(x^1) \ge 0, \ j \in \mathcal{J}\} \ge p \quad \text{for} \quad 0 \le t \le 1/2.$$

Similar argument applies for $1/2 < t \le 1$. Consequently, $\pi(t) \in X$, and this proves the assertion. \Box

4.3 Separable probabilistic constraints

We focus our attention on problems with separable probabilistic constraints. The problem that we analyze in this section, is the following:

$$\begin{aligned}
& \underset{x}{\operatorname{Min}} c(x) \\
& \text{s.t. } \Pr\{g(x) \ge Z\} \ge p, \\
& x \in \mathcal{X}.
\end{aligned}$$
(4.28)

We assume that $c : \mathbb{R}^n \to \mathbb{R}$ is a convex function and $g : \mathbb{R}^n \to \mathbb{R}^m$ is such that each component $g_i : \mathbb{R}^n \to \mathbb{R}$ is a concave function. We assume that the deterministic constraints are expressed by a closed convex set $\mathcal{X} \subset \mathbb{R}^n$. The vector Z is an m-dimensional random vector.

4.3.1 Continuity and differentiability properties of distribution functions

When the probabilistic constraint involves inequalities with random variables on the right hand side only as in problem (4.28), we can express it as a constraint on a distribution function:

$$\Pr\{g(x) \ge Z\} \ge p \iff F_Z(g(x)) \ge p.$$

Therefore, it is important to analyze the continuity and differentiability properties of distribution functions. These properties are relevant to the numerical solution of probabilistic optimization problems.

Suppose that Z has an α -concave distribution function with $\alpha \in \mathbb{R}$ and that the support of it, supp P_Z , has nonempty interior in \mathbb{R}^s . Then $F_Z(\cdot)$ is locally Lipschitz continuous on int supp P_Z by virtue of Theorem 4.29.

Example 4.56 We consider the following density function

$$\theta(z) = \begin{cases} \frac{1}{2\sqrt{z}}, & \text{for } z \in (0,1), \\ 0, & \text{otherwise.} \end{cases}$$

The corresponding cumulative distribution function is

$$F(z) = \begin{cases} 0, & \text{for } z \le 0, \\ \sqrt{z}, & \text{for } z \in (0, 1), \\ 1, & \text{for } z \ge 1. \end{cases}$$

The density θ is unbounded. We observe that F is continuous but it is not Lipschitz continuous at z = 0. The density θ is also not (-1)-concave and that means that the corresponding probability distribution is not quasi-concave.

Theorem 4.57. Suppose that all one-dimensional marginal distribution functions of an *s*-dimensional random vector Z are locally Lipschitz continuous. Then F_Z is locally Lipschitz continuous as well.

Proof. The statement can be proved by straightforward estimation of the distribution function by its marginals for s = 2 and induction on the dimension of the space.

It should be noted that even if the multivariate probability measure P_Z has a continuous and bounded density, then the distribution function F_Z is not necessarily Lipschitz continuous.

Theorem 4.58. Assume that the random vector Z has a continuous density $\theta : \mathbb{R}^s \to \mathbb{R}_+$ and $\forall i = 1, ..., s$ and for all points $z \in \mathbb{R}^s$, $z = (z_1, ..., z_s)$, the functions

$$z_i \mapsto \int_{-\infty}^{z_1} \dots \int_{-\infty}^{z_{i-1}} \int_{-\infty}^{z_{i+1}} \dots \int_{-\infty}^{z_s} \theta(t_1 \dots, z_i, \dots, t_s) dt_1 \dots dt_{i-1} dt_{i+1} \dots dt_s$$

are continuous as well. Then the distribution function F_Z is continuously differentiable.

Proof. In order to simplify notation, we demonstrate the statement for s = 2. It will be clear how to extend the proof for s > 2. The following identities hold:

$$F_Z(z_1, z_2) = \Pr(Z_1 \le z_1, Z_2 \le z_2) = \int_{-\infty}^{z_1} \int_{-\infty}^{z_2} \theta(t_1, t_2) dt_2 dt_1 = \int_{-\infty}^{z_1} \psi(t_1, z_2) dt_1,$$

where $\psi(t_1, z_2) = \int_{-\infty}^{z_2} \theta(t_1, t_2) dt_2$. Since $\psi(\cdot, z_2)$ is continuous by assumption, we invoke the Newton-Leibnitz Theorem to obtain

$$\frac{\partial F_Z}{\partial z_1}(z_1, z_2) = \psi(z_1, z_2) = \int_{-\infty}^{z_2} \theta(z_1, t_2) dt_2.$$

In a similar way

$$\frac{\partial F_Z}{\partial z_2}(z_1, z_2) = \int_{-\infty}^{z_1} \theta(t_1, z_2) dt_1.$$

Let us show continuity of $\frac{\partial F_Z}{\partial z_1}(z_1, z_2)$. Given the points $z \in \mathbb{R}^2$ and $y^k \in \mathbb{R}^2$, such that $\lim_{k\to\infty} y^k = z$, we have:

$$\begin{aligned} \left| \frac{\partial F_Z}{\partial z_1}(z) - \frac{\partial F_Z}{\partial z_1}(y^k) \right| &= \Big| \int_{-\infty}^{z_2} \theta(z_1, t) dt - \int_{-\infty}^{y_2^k} \theta(y_1^k, t) dt \\ &\leq \Big| \int_{z_2}^{y_2^k} \theta(y_1^k, t) dt \Big| + \Big| \int_{-\infty}^{z_2} [\theta(z_1, t) - \theta(y_1^k, t)] dt \Big|. \end{aligned}$$

First, we observe that the mapping $(z_1, z_2) \mapsto \int_a^{z_2} \theta(z_1, t) dt$ is continuous for every $a \in \mathbb{R}$ by the uniform continuity of $\theta(\cdot)$ on compact sets in \mathbb{R}^2 . Therefore, $\left|\int_{z_2}^{y_2^k} \theta(y_1^k, t) dt\right| \to 0$ whenever $k \to \infty$. Furthermore, $\left|\int_{-\infty}^{z_2} [\theta(z_1, t) - \theta(y_1^k, t)] dt\right| \to 0$ as well, due to the continuity assumptions. This proves that $\frac{\partial F_Z}{\partial z_1}(z)$ is continuous.

The continuity of the second partial derivative follows by the same line of argument. As both partial derivatives exist and are continuous, the function F_Z is continuously differentiable. \Box

Note that the continuity assumptions of the theorem are needed for the continuity of the function

$$\psi(\cdot, z_2) = \int_{-\infty}^{z_2} \theta(\cdot, t_2) dt_2.$$

Assume that the density function θ is locally Hölder-continuous of power $\alpha > 0$ with respect to each argument and the corresponding Hölder-multipliers are integrable functions, i.e., for any point $\bar{t} \in \mathbb{R}$, a constant $\varepsilon > 0$ and an integrable function L(z) > 0, $z \in \mathbb{R}$, exist such that

$$|\theta(t,z) - \theta(\tau,z)| \le L(z)|t - \tau|^{\alpha} \quad \forall t,\tau \in [\bar{t} - \varepsilon, \bar{t} + \varepsilon].$$

In that case, whenever $\tau \rightarrow t$, we obtain

$$|\psi(\tau, z) - \psi(t, z)| \le |t - \tau|^{\alpha} \int_{-\infty}^{z} L(r) dr \to 0.$$

The stronger continuity assumption about θ is a sufficient condition for the assumptions of Theorem 4.58. Another sufficient condition, in addition to the continuity of θ , would be its local domination by an integrable function. That is, for any point $\bar{t} \in \mathbb{R}$, a constant $\varepsilon > 0$ and an integrable function L(z) > 0 exist such that

$$|\theta(t,z)| \le L(z) \quad \forall t \in [\bar{t} - \varepsilon, \bar{t} + \varepsilon].$$

Then, the application of the Lebesgue dominated convergence theorem entails

$$\psi(t_k, z) = \int_{-\infty}^{z} \theta(t_k, r) dr \to \int_{-\infty}^{z} \theta(t, r) dr = \psi(t, z) \quad \text{whenever } t_k \to t.$$

4.3.2 p-Efficient points

In this section, we derive an algebraic description of the feasible set of problem (4.28).

The *p*-level set of the distribution function $F_Z(z) = \Pr\{Z \le z\}$ of Z is defined as follows:

$$\mathcal{Z}_p = \left\{ z \in \mathbb{R}^m : F_Z(z) \ge p \right\}.$$
(4.29)

Problem (4.28) can be formulated as

$$\begin{aligned}
& \underset{x}{\operatorname{Min}} c(x) \\
& \text{s.t. } g(x) \in \mathcal{Z}_p, \\
& x \in \mathcal{X}.
\end{aligned}$$
(4.30)

Lemma 4.59. For every $p \in (0, 1)$ the level set \mathcal{Z}_p is nonempty and closed.

Proof. The statement follows from the monotonicity and the right continuity of the distribution function. \Box

We introduce the key concept of a *p*-efficient point.

Definition 4.60. Let $p \in (0,1)$. A point $v \in \mathbb{R}^m$ is called a *p*-efficient point of the probability distribution function F, if $F(v) \ge p$ and there is no $z \le v$, $z \ne v$ such that $F(z) \ge p$.

The *p*-efficient points are minimal points of the level set Z_p with respect to the partial order in \mathbb{R}^m generated by the nonnegative cone \mathbb{R}^m_+ .

Note that for any scalar random variable Z and for all $p \in (0, 1)$, exactly one *p*-efficient point exists, which is the smallest v such that $F_Z(v) \ge p$, i.e., $v = F_Z^{(-1)}(p)$.

Lemma 4.61. Given $p \in (0, 1)$, we define

$$l = \left(F_{Z_1}^{(-1)}(p), \dots, F_{Z_m}^{(-1)}(p)\right).$$
(4.31)

Then every $v \in \mathbb{R}^m$ such that $F_Z(v) \ge p$ satisfies the inequality $v \ge l$.

Proof. Let $v_i = F_{Z_i}^{(-1)}(p)$ be the *p*-efficient point of the *i*th marginal distribution function. We observe that $F_Z(v) \leq F_{Z_i}(v_i)$ for every $v \in \mathbb{R}^m$ and $i = 1, \ldots, m$, and, therefore, we obtain that the set of *p*-efficient points is bounded from below. \Box

Let $p \in (0, 1)$ and let v^j , $j \in \mathcal{E}$, be *all* p-efficient points of Z. Here \mathcal{E} is an arbitrary index set. We define the cones

$$K_j = v^j + \mathbb{R}^m_+, \quad j \in \mathcal{E}.$$

The following result can be derived from Phelps theorem [183, Lemma 3.12] about the existence of conical support points, but we can easily prove it directly.

Theorem 4.62. It holds that $\mathcal{Z}_p = \bigcup_{j \in \mathcal{E}} K_j$.

Proof. If $y \in \mathbb{Z}_p$ then either y is p-efficient or there exists a vector w such that $w \leq y$, $w \neq y, w \in \mathbb{Z}_p$. By Lemma 4.61, one must have $l \leq w \leq y$. The set $Z_1 = \{z \in \mathbb{Z}_p : l \leq z \leq y\}$ is compact because the set \mathbb{Z}_p is closed by virtue of Lemma 4.59. Thus, there exists $w^1 \in \mathbb{Z}_1$ with the minimal first coordinate. If w^1 is a p-efficient point, then $y \in w^1 + \mathbb{R}_+^m$, what had to be shown. Otherwise, we define $\mathbb{Z}_2 = \{z \in \mathbb{Z}_p : l \leq z \leq w^1\}$, and choose a point $w^2 \in \mathbb{Z}_2$ with the minimal second coordinate. Proceeding in the same way, we shall find the minimal element w^m in the set \mathbb{Z}_p with $w^m \leq w^{m-1} \leq \cdots \leq y$. Therefore, $y \in w^m + \mathbb{R}_+^m$, and this completes the proof. \Box

By virtue of Theorem 4.62, assuming 0 , we obtain the following*disjunctive*semi-infinite formulation of problem (4.30) :

$$\begin{aligned}
& \underset{x}{\operatorname{Min}} c(x) \\
& \text{s.t. } g(x) \in \bigcup_{j \in \mathcal{E}} K_j, \\
& x \in \mathcal{X}.
\end{aligned}$$
(4.32)

This formulation provides an insight into the structure of the feasible set and the nature of its non-convexity. The main difficulty here is the implicit character of the disjunctive constraint.

Let S stand for the simplex in \mathbb{R}^{m+1} ,

$$S = \left\{ \alpha \in \mathbb{R}^{m+1} : \sum_{i=1}^{m+1} \alpha_i = 1, \, \alpha_i \ge 0 \right\}.$$

Denote the convex hull of the *p*-efficient points by *E*, i.e., $E = \text{conv}\{v^j, j \in \mathcal{E}\}$. We obtain a semi-infinite disjunctive representation of the convex hull of Z_p .

Lemma 4.63. The following representation holds

$$\operatorname{conv}\mathcal{Z}_p = E + \mathbb{R}^m_+.$$

Proof. By Theorem 4.62 every point $y \in \operatorname{conv} \mathbb{Z}_p$ can be represented as a convex combination of points in the cones K_j . By the theorem of Caratheodory the number of these points is no more than m + 1. Thus, we can write $y = \sum_{i=1}^{m+1} \alpha_i (v^{j_i} + w^i)$, where $w^i \in \mathbb{R}^m_+, \alpha \in S$ and $j_i \in \mathcal{E}$. The vector $w = \sum_{i=1}^{m+1} \alpha_i w^i$ belongs to \mathbb{R}^m_+ . Therefore, $y \in \sum_{i=1}^{m+1} \alpha_i v^{j_i} + \mathbb{R}^m_+$. \Box

We also may set
$$E = \left\{ \sum_{i=1}^{m+1} \alpha_i v^{j_i} : \alpha \in S, \ j_i \in \mathcal{E} \right\}.$$

Theorem 4.64. For every $p \in (0, 1)$ the set $\operatorname{conv} \mathbb{Z}_p$ is closed.

Proof. Consider a sequence $\{z^k\}$ of points of $\operatorname{conv} \mathcal{Z}_p$ which is convergent to a point \overline{z} . Using Caratheodory's theorem again, we have

$$z^k = \sum_{i=1}^{m+1} \alpha_i^k y_i^k,$$

with $y_i^k \in \mathbb{Z}_p$, $\alpha_i^k \ge 0$, and $\sum_{i=1}^{m+1} \alpha_i^k = 1$. By passing to a subsequence, if necessary, we can assume that the limits

$$\bar{\alpha}_i = \lim_{k \to \infty} \alpha_i^k$$

exist for all i = 1, ..., m + 1. By Lemma 4.61 all points y_i^k are bounded below by some vector l. For simplicity of notation we may assume that l = 0.

Let $I = \{i : \bar{\alpha}_i > 0\}$. Clearly, $\sum_{i \in I} \bar{\alpha}_i = 1$. We obtain

$$z^k \ge \sum_{i \in I} \alpha_i^k y_i^k. \tag{4.33}$$

We observe that $0 \le \alpha_i^k y_i^k \le z^k$ for all $i \in I$ and all k. Since $\{z^k\}$ is convergent and $\alpha_i^k \to \bar{\alpha}_i > 0$, each sequence $\{y_i^k\}$, $i \in I$, is bounded. Therefore, we can assume that each of them is convergent to some limit \bar{y}_i , $i \in I$. By virtue of Lemma 4.59 $\bar{y}_i \in \mathcal{Z}_p$. Passing to the limit in inequality (4.33), we obtain

$$\bar{z} \ge \sum_{i \in I} \bar{\alpha}_i \bar{y}_i \in \operatorname{conv} \mathcal{Z}_p.$$

Due to Lemma 4.63, we conclude that $\bar{z} \in \text{conv} \mathcal{Z}_p$.

In general the set of *p*-efficient points might be unbounded and not closed, as illustrated on Figure 4.3.

The next example demonstrates some properties of the set of p-efficient points and its convex hull.

Example 4.65 Let Z be a random vector in \mathbb{R}^2 whose components Z_1 and Z_2 are independent. The random variable Z_1 is uniformly distributed in the interval [0, 1], and the probability distribution function of Z_2 is the following

$$F_2(y) = \begin{cases} 0 & \text{if } z \le 0\\ 4z & \text{if } 0 < z < 0.2\\ 0.8 & \text{if } 0.2 \le z < 0.7\\ \frac{2z+1}{3} & \text{if } 0.7 \le z < 1\\ 1 & \text{if } y \ge 1. \end{cases}$$

The set of 0.8-efficient points is illustrated in Fig 4.65. The point (1, 0.7) is not a *p*-efficient point as it dominates the point (1, 0.2). However, the points of form $v^t = (t, 1.2/t - 0.5)$ with $t \in [0.8, 1)$ are *p*-efficient and $(1, 0.7) = \lim_{t\uparrow 1} v^t$. This shows that the set of *p*-efficient points in the same example is not closed. The convex hull of *p*-efficient points, *E* in this example is not closed either, because, it does not contain the points $w_{\lambda} = \lambda(1, 0.7) + (1 - \lambda)(1, 0.2)$ for any $\lambda \in (0, 1)$, while $w_{\lambda} = \lim_{t\uparrow 1} w_{\lambda}^t$ with $w_{\lambda}^t = \lambda v^t + (1 - \lambda)(1, 0.2)$ belonging to *E*.



Figure 4.3. *Example of a set* Z_p *with p-efficient points v.*

Figure 4.4. Set of 0.8-efficient points of the random variable Z.



We encounter also a relation between the *p*-efficient points and the extreme points of the convex hull of \mathcal{Z}_p .

Theorem 4.66. For every $p \in (0, 1)$ the set of extreme points of $\operatorname{conv} \mathbb{Z}_p$ is nonempty and it is contained in the set of *p*-efficient points.

Proof. Consider the lower bound l defined in (4.31). The set $\operatorname{conv} \mathbb{Z}_p$ is included in $l + \mathbb{R}^m_+$, by virtue of Lemma 4.61 and Lemma 4.63. Therefore, it does not contain any line. Since $\operatorname{conv} \mathbb{Z}_p$ is closed by Theorem 4.64, it has at least one extreme point.

Let w be an extreme point of $\operatorname{conv} \mathbb{Z}_p$. Suppose that w is not a p-efficient point. Then

Theorem 4.62 implies that there exists a *p*-efficient point $v \le w, v \ne w$. Since $w + \mathbb{R}^m_+ \subset \operatorname{conv} \mathbb{Z}_p$, the point *w* is a convex combination of *v* and w + (w - v). Consequently, *w* cannot be extreme. \Box

Example 4.65 continued. We notice that not every *p*-efficient point is an extreme point of $\operatorname{conv} \mathcal{Z}_p$. Indeed, we observe that the points (0.8, 1) and (1, 0.2) are *p*-efficient. We shall show that all other *p*-efficient points dominate some convex combination of these two points. For every point v^t with 0.8 < t < 1, we consider the point $v_{\lambda}^t = \lambda(0.8, 1) + (1 - \lambda)(1, 0.2)$ with $\lambda = \frac{3}{2t} - \frac{7}{8}$. Notice that $\frac{3}{2t} - \frac{7}{8} \in (0, 1)$ whenever $t \in (0.8, 1)$. The second coordinates of v^t and v_{λ}^t are the same. The first coordinates satisfy the inequality $t > 1 - 0.2\lambda$, which means that $v^t \ge v_{\lambda}^t$. This proves that the only extreme points of $\operatorname{conv} \mathcal{Z}_p$ are the points (0.8, 1) and (1, 0.2). On the other hand, $\operatorname{cl} E = \operatorname{conv}\{(0.8, 1), (1, 0.2), (1, 0.7)\}$ in this example.

The representations (4.62) and (4.63) become very helpful when the vector Z has a discrete distribution on \mathbb{Z}^m , in particular, if the problem is of form (4.78). We shall discuss this special case in more detail. Let us emphasize that our investigations extend to the case when the random vector Z has a discrete distribution with values on a grid. Our further study can be adapted to the case of distributions on non-uniform grids for which a uniform lower bound on the distance of grid points in each coordinate exists. In this presentation, we assume that $Z \in \mathbb{Z}^m$. In this case, we can establish that the distribution function F_Z has finitely many p-efficient points.

Theorem 4.67. For any $p \in (0, 1)$, the set of *p*-efficient points of an integer random vector is nonempty and finite.

Proof. First we shall show that at least one *p*-efficient point exists. Since p < 1, there exists a point *y* such that $F_Z(y) \ge p$. By Lemma 4.61, the level set \mathcal{Z}_p is bounded from below by the vector *l* of *p*-efficient points of one-dimensional marginals. Therefore, if *y* is not *p*-efficient, one of finitely many integer points *v* such that $l \le v \le y$ must be *p*-efficient.

Now we prove the finiteness of the set of p-efficient points. Suppose that there exists an infinite sequence of different p-efficient points v^j , j = 1, 2, ... Since they are integer, and the first coordinate v_1^j is bounded from below by l_1 , with no loss of generality we may select a subsequence which is non-decreasing in the first coordinate. By a similar token, we can select further subsequences which are non-decreasing in the first k coordinates (k = 1, ..., m). Since the dimension m is finite, we obtain a subsequence of different p-efficient points which is non-decreasing in all coordinates. This contradicts the definition of a p-efficient point. \Box

Note the crucial role of Lemma 4.61 in this proof. In conclusion, we have obtained that the disjunctive formulation (4.32) of problem (4.30) has a *finite* index set \mathcal{E} .

Figure 4.5 illustrates the structure of the probabilistically constrained set for a discrete random variable.

The concept of α -concavity on a set can be used at this moment to obtain an equivalent representation of the set \mathcal{Z}_p for a random vector with a discrete distribution.



Figure 4.5. *Example of a discrete set* Z_p *with p-efficient points* v_1, \ldots, v_5 *.*

Theorem 4.68. Let \mathcal{Z}_p be the set of all possible values of an integer random vector Z. If the distribution function F_Z of Z is α -concave on $\mathcal{Z}_p + \mathbb{Z}_+^m$, for some $\alpha \in [-\infty, \infty]$, then for every $p \in (0, 1)$ one has

$$\mathcal{Z}_p = \left\{ y \in \mathbb{R}^m : y \ge z \ge \sum_{j \in \mathcal{E}} \lambda_j v^j, \sum_{j \in \mathcal{E}} \lambda_j = 1, \, \lambda_j \ge 0, \, z \in \mathbb{Z}^m \right\},\$$

where v^j , $j \in \mathcal{E}$, are the *p*-efficient points of *F*.

Proof. The representation (4.32) implies that

$$\mathcal{Z}_p \subset \left\{ y \in \mathbb{R}^m : y \ge z \ge \sum_{j \in \mathcal{E}} \lambda_j v^j, \sum_{j \in \mathcal{E}} \lambda_j = 1, \, \lambda_j \ge 0, \, z \in \mathbb{Z}^m \right\},\$$

We have to show that every point y from the set at the right hand side belongs to \mathbb{Z}_p . By the monotonicity of the distribution function F_Z , we have $F_Z(y) \ge F_Z(z)$ whenever $y \ge z$. Therefore, it is sufficient to show that $\Pr\{Z \le z\} \ge p$ for all $z \in \mathbb{Z}^m$ such that $z \ge \sum_{j \in \mathcal{E}} \lambda_j v^j$ with $\lambda_j \ge 0$, $\sum_{j \in \mathcal{E}} \lambda_j = 1$. We consider five cases with respect to α . Case 1: $\alpha = \infty$. It follows from the definition of α -concavity that

$$F_Z(z) \ge \max\{F_Z(v^j), j \in \mathcal{E} : \lambda_j \neq 0\} \ge p_j$$

Case 2: $\alpha = -\infty$. Since $F_Z(v^j) \ge p$ for each index $j \in \mathcal{E}$ such that $\lambda_j \ne 0$, the assertion follows as in Case 1.

Case 3: $\alpha = 0$. By the definition of α -concavity, we have the following inequalities:

$$F_Z(z) \ge \prod_{j \in \mathcal{E}} [F_Z(v^j)]^{\lambda_j} \ge \prod_{j \in \mathcal{E}} p^{\lambda_j} = p.$$

Case 4: $\alpha \in (-\infty, 0)$. By the definition of α -concavity,

$$[F_Z(z)]^{\alpha} \le \sum_{j \in \mathcal{E}} \lambda_j [F_Z(v^j)]^{\alpha} \le \sum_{j \in \mathcal{E}} \lambda_j p^{\alpha} = p^{\alpha}.$$

Since $\alpha < 0$, we obtain $F_Z(z) \ge p$.

Case 5: $\alpha \in (0, \infty)$. By the definition of α -concavity,

$$[F_Z(z)]^{\alpha} \ge \sum_{j \in \mathcal{E}} \lambda_j [F_Z(v^j)]^{\alpha} \ge \sum_{j \in \mathcal{E}} \lambda_j p^{\alpha} = p^{\alpha},$$

concluding that $z \in \mathbb{Z}_p$, as desired. \Box

The consequence of this theorem is that under the α -concavity assumption all integer point contained in $\operatorname{conv} \mathbb{Z}_p = E + \mathbb{R}^m_+$ satisfy the probabilistic constraint. This demonstrates the importance of the notion of α -concavity for discrete distribution functions as introduced in Definition 4.34. For example, the set \mathbb{Z}_p illustrated in Figure 4.5 does not correspond to any α -concave distribution function, because its convex hull contains integer points which do not belong to \mathbb{Z}_p . These are the points (3,6), (4,5) and (6,2).

Using Theorems 4.68 and 4.43, we obtain the following statement.

Corollary 4.69. If the distribution function F_Z of Z is α -concave on the set A for some $\alpha \in [-\infty, \infty]$, then

$$\mathcal{Z}_p \cap \mathcal{A} = \operatorname{conv} \mathcal{Z}_p \cap \mathcal{A}$$

Under the conditions of Theorem 4.68, problem (4.30) can be formulated in the following equivalent way:

$$\underset{x,z,\lambda}{\operatorname{Min}} c(x) \tag{4.34}$$

s.t.
$$g(x) \ge z$$
, (4.35)

$$z \ge \sum_{j \in \mathcal{E}} \lambda_j v^j, \tag{4.36}$$

$$z \in \mathcal{A},\tag{4.37}$$

$$\sum_{j \in \mathcal{E}} \lambda_j = 1, \tag{4.38}$$

$$\lambda_j \ge 0, \, j \in \mathcal{E},\tag{4.39}$$

$$x \in \mathcal{X}.\tag{4.40}$$

In this way, we have replaced the probabilistic constraint by algebraic equations and inequalities, together with the requirement (4.37). This condition cannot be dropped, in
general. However, if other conditions of the problem imply that g(x) has values on the grid A, then we may remove z entirely form the problem formulation. In this case, we replace constraints (4.35)–(4.37) with

$$g(x) \geq \sum_{j \in \mathcal{E}} \lambda_j v^j$$

For example, if $\mathcal{A} = \mathbb{Z}^n$, the definition of \mathcal{X} contains the constraint $x \in \mathbb{Z}^n$, and T is a matrix with integer elements, then we can dispose of the variable z.

4.3.3 The tangent and normal cones of $conv Z_p$

We use the notation $\mathcal{T}_A(z)$, $\mathcal{N}_A(z)$, and $\mathcal{K}_A(z)$ for the tangent cone, the normal cone, and the cone of feasible directions of a set A at a point $z \in A$, respectively. For a set A, we denote the convex conical hull of A by coneA.

The set $\operatorname{conv} \mathbb{Z}_p$ has extreme points due to the existence of a lower bound l for all points in it. We observe that the recession cone of $\operatorname{conv} \mathbb{Z}_p$ is \mathbb{R}^m_+ . Indeed, Lemma 4.63 implies that \mathbb{R}^m_+ is a subset of the recession cone. Assume that there exists a direction $d \in \mathbb{R}_{\operatorname{conv} \mathbb{Z}_p} \setminus \mathbb{R}^m_+$, then d has a negative coordinate d_{i_0} . Fixing any point p-efficient point v, we can find $\tau > 0$ big enough such that $v_{i_0} + \tau d_{i_0} < l_{i_0}$, which is a contradiction.

v, we can find $\tau > 0$ big enough such that $v_{i_0} + \tau d_{i_0} < l_{i_0}$, which is a contradiction. Every $z^0 \in \operatorname{conv} \mathcal{Z}_p$ has a representation $z^0 = w^0 + y^0$ for some $w^0 \in E$ and $y^0 \in \mathbb{R}^m_+$ by virtue of Theorem 4.66. The cone of feasible directions can be described by the following formula:

$$\begin{aligned} \mathcal{K}_{\operatorname{conv}\mathcal{Z}_{p}}(z^{0}) &= \left\{ d \in \mathbb{R}^{m} : d = \beta(z - z^{0}), \ z \in \operatorname{conv}\mathcal{Z}_{p}, \ \beta \geq 0 \right\} \\ &= \left\{ \beta(w + y - w^{0} - y^{0})), \ w \in E, \ y \in \mathbb{R}^{m}_{+}, \ \beta \geq 0 \right\} \\ &= \left\{ \beta(w - w^{0}), \ w \in E, \ \beta \geq 0 \right\} + \left\{ \beta(y - y^{0}), \ y \in \mathbb{R}^{m}_{+}, \ \beta \geq 0 \right\} \\ &= \mathcal{K}_{E}(w^{0}) + \mathcal{K}_{\mathbb{R}^{m}_{+}}(y^{0}). \end{aligned}$$
(4.41)

For a point $z \in E$, we have $z = \sum_{j \in J} \alpha_j v^j$, where v^j are *p*-efficient points, and $\alpha_j \ge 0$, $\sum_{j \in J} \alpha_j = 1$. We denote

$$J_{\alpha}(z) = \{ j \in J : \alpha_j = 0 \}.$$

Proposition 4.70. Let $z^0 \in \operatorname{conv} \mathcal{Z}_p$ be represented as $z^0 = w^0 + y^0 = \sum_{j \in J} \alpha_j v^j + y^0$, where $w^0 \in E$, $y^0 \in \mathbb{R}^m_+$, v^j are *p*-efficient points, $\alpha_j \ge 0$, $\sum_{j \in J} \alpha_j = 1$. The cone of feasible directions of $\operatorname{conv} \mathcal{Z}_p$ at z^0 has the following representation

$$\mathcal{K}_{\operatorname{conv}\mathcal{Z}_p}(z^0) = \operatorname{cone}\{v^j - w^0 : v^j \in J\} + \{\tau y^0, \tau \in \mathbb{R}\} + \mathbb{R}^m_+$$
$$= \left\{ \sum_{j \in J} \gamma_j v^j, \sum_{j \in J} \gamma_j = 0, \gamma_j \ge 0 \ \forall j \in J_\alpha(w^0) \right\}$$
(4.42)

Proof. The cone $\mathcal{K}_E(w^0)$ can be described in the following way:

$$\mathcal{K}_{E}(w^{0}) = \left\{ d \in \mathbb{R}^{m} : d = \beta (\sum_{i=1}^{m+1} \alpha_{i} v^{j_{i}} - w^{0}), \, \alpha \in S_{m+1}, j_{i} \in J, \beta \ge 0 \right\}$$
$$= \left\{ \beta (\sum_{i=1}^{m+1} \alpha_{i} (v^{j_{i}} - w^{0})), \, \alpha \in S_{m+1}, j_{i} \in J, \beta \ge 0 \right\}$$
$$= \left\{ \sum_{i=1}^{m+1} \gamma_{i} (v^{j_{i}} - w^{0}), \gamma_{i} \ge 0, j_{i} \in J \right\}$$
(4.43)

We observe that $\mathcal{K}_{\mathbb{R}^m_+}(y^0) = \mathbb{R}^m_+ + \{\tau y^0, \tau \in \mathbb{R}\}$. Using (4.41), we obtain that

$$\mathcal{K}_{\operatorname{conv}\mathcal{Z}_p}(z^0) = \mathcal{K}_E(w^0) + \mathcal{K}_{\mathbb{R}^m_+}(y^0) = \mathcal{K}_E(w^0) + \mathbb{R}^m_+ + \{\tau y^0, \, \tau \in \mathbb{R}\}.$$

Now, the representation of $\mathcal{K}_E(w^0)$ yields the first equation in (4.42). Substituting the representation of w^0 by *p*-efficient points at the right-hand side of formula (4.43), we obtain that every direction $w \in \mathcal{K}_E(w^0)$ has the following form:

$$w = \sum_{i \in J} \bar{\gamma}_i \left(v^i - \sum_{j \in J} \alpha_j v^j \right) = \sum_{i \in J} \bar{\gamma}_i v^i - \sum_{i \in J} \sum_{j \in J} \bar{\gamma}_i \alpha_j v^j.$$

Denoting $\beta = \sum_{i \in J} \bar{\gamma}_i$, we obtain

$$w = \sum_{i \in J} \bar{\gamma}_i v^i - \sum_{j \in J} \beta \alpha_j v^j = \sum_{j \in J} \left(\bar{\gamma}_j - \beta \alpha_j \right) v^j = \sum_{j \in J} \gamma_j v^j.$$

Notice that $\gamma_j \ge 0$ for all $j \in J_{\alpha}(w^0)$. Furthermore, the coefficients γ_i satisfy the equation

$$\sum_{j \in J} \gamma_j = \sum_{j \in J} \bar{\gamma}_j - \beta \sum_{j \in J} \alpha_j = 0$$

We infer that

$$\mathcal{K}_E(w^0) \subseteq \Big\{ d = \sum_{j \in J} \gamma_j v^j, \sum_{j \in J} \gamma_j = 0, \, \gamma_j \ge 0 \,\forall j \in J_\alpha(w^0) \Big\}.$$
(4.44)

Assume now, that $w = \sum_{j \in J} \gamma_j v^j$ with $\sum_{j=1}^{m+1} \gamma_j = 0$ and $\gamma_j \ge 0$ for all $j \in J_\alpha(w^0)$. We determine

$$\beta = \max\left\{-\frac{\gamma_j}{\alpha_j} : j \in J \text{ and } \alpha_j > 0\right\}$$

As the set $\{j \in J : \gamma_j < 0, \alpha_j > 0\}$ is non-empty, $\beta > 0$. Setting $\bar{\gamma}_j = \gamma_j + \beta \alpha_j$, we have $\bar{\gamma}_j \ge 0$ for all $j \in J$. Summing up, we have

$$\sum_{j\in J}\bar{\gamma}_j=\sum_{j\in J}\gamma_j+\beta=\beta.$$

Using this relation, we can represent w as follows:

$$w = \sum_{j \in J} \gamma_j v^j = \sum_{j \in J} \left(\bar{\gamma}_j - \beta \alpha_j \right) v^j = \sum_{j \in J} \left(\bar{\gamma}_j - \sum_{i \in J} \bar{\gamma}_i \alpha_j \right) v^j = \sum_{j \in J} \bar{\gamma}_j \left(v^j - \sum_{i \in J} \alpha_i v^i \right).$$

We conclude, that $w \in \mathcal{K}_E(w^0)$. Thus, we obtain a second representation of the cone of feasible directions of E:

$$\mathcal{K}_E(w^0) = \left\{ d = \sum_{j \in J} \gamma_j v^j, \ \sum_{j \in J} \gamma_j = 0 \ \gamma_j \ge 0 \ \forall j \in J_\alpha(w^0) \right\}.$$

If z is an extreme point of $conv Z_p$, then z is a p-efficient point as well, and the representation of the cone of feasible directions simplifies considerably:

$$\mathcal{K}_{\operatorname{conv}\mathcal{Z}_p}(z) = \operatorname{cone}\{v^j - z : v^j \in J\} + \mathbb{R}^m_+.$$
(4.45)

We note that for a random vector Y with discrete distribution, the cone $\mathcal{K}_E(w^0)$ is polyhedral. Define the vectors

$$d^i = \sum_{j \in J \setminus i} \alpha_j (v^i - v^j) \quad \text{for all} \quad i \in J$$

The first part of formula (4.42) takes on the form $\mathcal{K}_E(w^0) = \operatorname{cone}\{d^i, i \in J\}$. If Y has a discrete distribution then the cone of feasible directions and the tangent cones coincide:

$$\mathcal{T}_{\operatorname{conv}\mathcal{Z}_p}(z^0) = \operatorname{cone}\{d^i, i \in J\} + \{\tau y^0, \, \tau \in \mathbb{R}\} + \mathbb{R}^m_+.$$
(4.46)

Proposition 4.71. The normal cone of a point $z^0 \in \operatorname{conv} \mathbb{Z}_p$ with representation $z^0 = w^0 + y^0 = \sum_{j \in J} \alpha_j v^j + y^0$, where $w^0 = \sum_{j \in J} \alpha_j v^j \in E$, $y^0 \in \mathbb{R}^m_+$, and $\alpha_j \geq 0$, $\sum_{j \in J} \alpha_j = 1$, is the following set:

$$\mathcal{N}_{\operatorname{conv}\mathcal{Z}_p}(z^0) = \left\{ y \in \mathbb{R}^m_- : \langle v^j, y \rangle = \langle v^i, y \rangle, \forall i, j \notin J_\alpha(z^0), \\ \langle v^j, y \rangle \ge \langle v^i, y \rangle, \forall i \notin J_\alpha(z^0) \text{ and } j \in J_\alpha(z^0) \\ y_i.y_i^0 = 0, i = 1, \dots, m \right\}$$
(4.47)

Proof. The normal cone at some $z^0 \in \operatorname{conv} \mathcal{Z}_p$ is defined as $\mathcal{N}_{\operatorname{conv} \mathcal{Z}_p}(z^0) = [T_{\operatorname{conv} \mathcal{Z}_p}(z^0)]^\circ$. Using relation (4.41) an having in mind that the tangent cone is the closure of the cone of feasible directions, we infer that the cone $\mathcal{N}_{\operatorname{conv} \mathcal{Z}_p}(z^0)$ has the form:

$$\mathcal{N}_{\operatorname{conv}\mathcal{Z}_p}(z^0) = \left[\mathcal{K}_E(w^0) + \mathcal{K}_{\mathbb{R}^m_+}(y^0)\right]^\circ = \left[\mathcal{K}_E(w^0)\right]^\circ \cap \left[\mathcal{K}_{\mathbb{R}^m_+}(y^0)\right]^\circ$$

$$= \mathcal{N}_E(w^0) \cap \mathcal{N}_{\mathbb{R}^m_+}(y^0).$$
(4.48)

Using (4.42), we obtain that $\mathcal{N}_E(w^0)$ is given by:

$$\mathcal{N}_E(w^0) = \left[\mathcal{K}_E(w^0)\right]^\circ = \{y \in \mathbb{R}^m : \langle v^j, y \rangle \le \langle w^0, y \rangle, \forall v^j \in J\}.$$

Setting $\gamma_i = 1$ and $\gamma_j = -1$ for all pairs of indices $i, j \notin J_\alpha(z^0)$ and all other coefficient $\gamma_\ell = 0$, we obtain the condition $\langle v^j, y \rangle = \langle v^i, y \rangle$ for all $i, j \notin J_\alpha(z^0)$. Furthermore, for each pair of indices (i, j) with $i \in J_\alpha(z^0)$ and $j \notin J_\alpha(z^0)$, we set $\gamma_i = 1, \gamma_j = -1$, and $\gamma_\ell = 0$ for all other indices. Using $\langle v^j, y \rangle \leq \langle w^0, y \rangle$, we obtain $\langle v^j, y \rangle \geq \langle v^i, y \rangle, \forall i \notin J_\alpha(z^0)$ and $j \in J_\alpha(z^0)$ as stated. Additionally, $[\mathcal{K}_{\mathbb{R}^m_+}(y^0)]^\circ = \{y \in \mathbb{R}^m_- : \langle y, y^0 \rangle = 0\}$. Putting this observations together completes the proof. \Box

4.3.4 Optimality conditions and duality theory

In this section, we shall develop optimality conditions for problem (4.30). We shall consider two sets of assumptions: the first set of assumptions will allow us to apply convex analysis results and techniques, while the second one will rely on certain differentiability properties and will allows to obtain necessary conditions of local optimality.

Optimality conditions and duality theory under convexity assumptions

We assume that $c : \mathbb{R}^n \to \mathbb{R}$ is a convex function. The mapping $g : \mathbb{R}^n \to \mathbb{R}^m$ has concave components $g_i : \mathbb{R}^n \to \mathbb{R}$. The set $\mathcal{X} \subset \mathbb{R}^n$ is closed and convex; the random vector Ztakes values in \mathbb{R}^m . The set \mathcal{Z}_p is defined as in (4.29). We split variables and consider the following formulation of the problem:

$$\begin{aligned}
& \underset{x,z}{\operatorname{Min}} c(x) \\
& g(x) \ge z, \\
& x \in \mathcal{X}, \\
& z \in \mathcal{Z}_p.
\end{aligned}$$
(4.49)

Associating a Lagrange multiplier $u \in \mathbb{R}^m_+$ with the constraint $g(x) \ge z$, we obtain the Lagrangian function:

$$L(x, z, u) = c(x) + u'(z - g(x)).$$

The dual functional has the form

$$\Psi(u) = \inf_{(x,z)\in\mathcal{X}\times\mathcal{Z}_p} L(x,z,u) = h(u) + d(u),$$

where

$$h(u) = \inf\{c(x) - u^{\mathsf{T}}g(x) : x \in \mathcal{X}\},$$
(4.50)

$$d(u) = \inf\{u^{\mathsf{T}}z \ : \ z \in \mathcal{Z}_p\}. \tag{4.51}$$

For any $u \in \mathbb{R}^m_+$ the value of $\Psi(u)$ is a lower bound on the optimal value c^* of the original problem. The best Lagrangian lower bound will be given by the optimal value Ψ^* of the problem:

$$\sup_{u \ge 0} \Psi(u). \tag{4.52}$$

We call (4.52) the dual problem to problem (4.49). For $u \ge 0$ one has $d(u) = -\infty$, because the set \mathbb{Z}_p contains a translation of \mathbb{R}^m_+ . The function $d(\cdot)$ is concave. Note that

 $d(u) = -S_{Z_p}(-u)$, where $S_{Z_p}(\cdot)$ is the support function of the set Z_p . By virtue of Theorem 4.64 and Hiriart-Urruty and Lemaréchal [107, Chapter V, Prop.2.2.1], we have

$$d(u) = \inf\{u^{\mathsf{T}}z : z \in \operatorname{conv}\mathcal{Z}_p\}.$$
(4.53)

Let us consider the *convex hull problem*:

$$\begin{aligned}
& \underset{x,z}{\operatorname{Min}} c(x) \\
& g(x) \ge z, \\
& x \in \mathcal{X}, \\
& z \in \operatorname{conv} \mathcal{Z}_{p}.
\end{aligned}$$
(4.54)

We impose the following constraint qualification condition:

There exist points
$$x^0 \in \mathcal{X}$$
 and $z^0 \in \operatorname{conv} \mathcal{Z}_p$ *such that* $g(x^0) > z^0$. (4.55)

If this constraint qualification condition is satisfied, then the duality theory in convex programming Rockafellar [210, Corollary 28.2.1] implies that there exists $\hat{u} \ge 0$ at which the minimum in (4.52) is attained, and $\Psi^* = \Psi(\hat{u})$ is the optimal value of the convex hull problem (4.54).

Now, we study in detail the structure of the dual functional Ψ . We shall characterize the solution sets of the two subproblems (4.50) and (4.51), which provide the values of the dual functional. Observe that the normal cone to the positive orthant at a point $u \ge 0$ is the following:

$$\mathcal{N}_{\mathbb{R}^m_+}(u) = \{ d \in \mathbb{R}^m_- : d_i = 0 \text{ if } u_i > 0, \ i = 1, \dots, m \}.$$
(4.56)

We define the set

$$V(u) = \{ v \in \mathbb{R}^m : u^{\mathsf{T}}v = d(u) \text{ and } v \text{ is a } p \text{-efficient point} \}.$$
(4.57)

Lemma 4.72. For every u > 0 the solution set of (4.51) is nonempty. For every $u \ge 0$ it has the following form: $\hat{Z}(u) = V(u) - \mathcal{N}_{\mathbb{R}^m_{\perp}}(u)$.

Proof. First we consider the case u > 0. Then every recession direction q of Z_p satisfies $u^{\mathsf{T}}q > 0$. Since Z_p is closed, a solution to (4.51) must exist. Suppose that a solution z to (4.51) is not a p-efficient point. By virtue of Theorem 4.62, there is a p-efficient $v \in Z_p$ such that $v \leq z$, and $v \neq z$. Thus, $u^{\mathsf{T}}v < u^{\mathsf{T}}z$, which is a contradiction. Therefore, we conclude that there is a p-efficient point v, which solves problem (4.51).

Consider the general case $u \ge 0$ and assume that the solution set of problem (4.51) is nonempty. In this case, the solution set always contains a *p*-efficient point. Indeed, if a solution *z* is not *p*-efficient, we must have a *p*-efficient point *v* dominated by *z*, and $u^{\mathsf{T}}v \le u^{\mathsf{T}}z$ holds by the nonnegativity of *u*. Consequently, $u^{\mathsf{T}}v = u^{\mathsf{T}}z$ for all *p*-efficient $v \le z$, which is equivalent to $z \in \{v\} - \mathcal{N}_{\mathbb{R}_{T}^{m}}(u)$, as required.

If the solution set of (4.51) is empty, then $V(u) = \emptyset$ by definition and the assertion is true as well. \Box

The last result allows us to calculate the subdifferential of the function d in a closed form.

Lemma 4.73. For every $u \ge 0$ one has $\partial d(u) = \operatorname{conv}(V(u)) - \mathcal{N}_{\mathbb{R}^m_+}(u)$. If u > 0 then $\partial d(u)$ is nonempty.

Proof. From (4.51) we obtain $d(u) = -s_{z_p}(-u)$, where $S_{z_p}(\cdot)$ is the support function of \mathcal{Z}_p and, consequently, of conv \mathcal{Z}_p . Consider the indicator function $\mathbb{I}_{\text{conv} \mathcal{Z}_p}(\cdot)$ of the set conv \mathcal{Z}_p . By virtue of Corollary 16.5.1 in Rockafellar [210], we have

$$\mathsf{S}_{\mathcal{Z}_p}(u) = \mathbb{I}^*_{\operatorname{conv}\mathcal{Z}_p}(u),$$

where the latter function is the conjugate of the indicator function $\mathbb{I}_{\text{conv} \mathcal{Z}_p}(\cdot)$. Thus,

$$\partial d(u) = -\partial \mathbb{I}^*_{\operatorname{conv} \mathcal{Z}_n}(-u).$$

Recall that conv Z_p is closed, by Theorem 4.64. Using Rockafellar [210, Thm 23.5], we observe that $y \in \partial \mathbb{I}^*_{\operatorname{conv} Z_p}(-u)$ if and only if $\mathbb{I}^*_{\operatorname{conv} Z_p}(-u) + \mathbb{I}_{\operatorname{conv} Z_p}(y) = -y^{\mathsf{T}}u$. It follows that $y \in \operatorname{conv} Z_p$ and $\mathbb{I}^*_{\operatorname{conv} Z_p}(-u) = -y^{\mathsf{T}}u$. Consequently,

$$y^{\mathsf{T}}u = d(u). \tag{4.58}$$

Since $y \in \operatorname{conv} \mathcal{Z}_p$ we can represent it as follows:

$$y = \sum_{j=1}^{m+1} \alpha_j e^j + w_j$$

where e^j , j = 1, ..., m + 1, are extreme points of conv Z_p and $w \ge 0$. Using Theorem 4.66 we conclude that e^j are *p*-efficient points. Moreover, applying *u*, we obtain

$$y^{\mathsf{T}}u = \sum_{j=1}^{m+1} \alpha_j u^{\mathsf{T}} e^j + u^{\mathsf{T}} w \ge d(u),$$
(4.59)

because $u^{\mathsf{T}}e^j \ge d(u)$ and $u^{\mathsf{T}}w \ge 0$. Combining (4.58) and (4.59) we conclude that $u^{\mathsf{T}}e^j = d(u)$ for all j, and $u^{\mathsf{T}}w = 0$. Thus $y \in \operatorname{conv} V(u) - \mathcal{N}_{\mathbb{R}^m_{\perp}}(u)$.

Conversely, if $y \in \operatorname{conv} V(u) - \mathcal{N}_{\mathbb{R}^m_+}(u)$ then (4.58) holds true by the definitions of the set V(u) and the normal cone. This implies that $y \in \partial d(u)$, as required.

Furthermore, the set $\partial d(u)$ is nonempty for u > 0 due to Lemma 4.72.

Now, we analyze the function $h(\cdot)$. Define the set of minimizers in (4.50),

$$X(u) = \{ x \in \mathcal{X} : c(x) - u^{\mathsf{T}}g(x) = h(u) \}.$$
(4.60)

Since the set \mathcal{X} is convex and the objective function of problem (4.50) is convex for all $u \ge 0$, we conclude that the solution set X(u) is convex for all $u \ge 0$.

Lemma 4.74. Assume that the set \mathcal{X} is compact. For every $u \in \mathbb{R}^m$, the subdifferential of the function h is described as follows:

$$\partial h(u) = \operatorname{conv} \{-g(x) : x \in X(u)\}.$$

Proof. The function h is concave on \mathbb{R}^m . Since the set \mathcal{X} is compact, c is convex, and g_i , $i = 1, \ldots, m$ are concave, the set X(u) is compact. Therefore, the subdifferential of h(u) for every $u \in \mathbb{R}^m$ is the closure of conv $\{-g(x) : x \in X(u)\}$ (see Hiriart-Urruty and C. Lemaréchal [107, Chapter VI, Lemma 4.4.2]). By the compactness of X(u) and concavity of g, the set $\{-g(x) : x \in X(u)\}$ is closed. Therefore, we can omit taking the closure in the description of the subdifferential of h(u). \Box

The last two results along with Moreau-Rockafellar's theorem provide a description of the subdifferential of the dual function.

Corollary 4.75. Assuming that the set \mathcal{X} is compact, the subdifferential of the dual function is nonempty for any $u \ge 0$ and has the form

$$\partial \Psi(u) = \operatorname{conv} \left\{ -g(x) : x \in X(u) \right\} + \operatorname{conv} V(u) - \mathcal{N}_{\mathbb{R}^m_+}(u)$$

Proof. Since int dom $d \neq \emptyset$ and dom $h = \mathbb{R}^m$, Moreau-Rockafellar's theorem yields $\partial \Psi(u) = \partial h(u) + \partial d(u)$. applying Lemma 4.73 and Lemma 4.74 we obtain the stated formula. \Box

The following necessary and sufficient optimality conditions for problem (4.52) are established.

Theorem 4.76. Assume that the constraint qualification condition (4.55) is satisfied and that the set \mathcal{X} is compact. A vector $u \ge 0$ is an optimal solution of (4.52) if and only if there exists a point $x \in X(u)$, points $v^1, \ldots, v^{m+1} \in V(u)$ and scalars $\beta_1, \ldots, \beta_{m+1} \ge 0$ with $\sum_{j=1}^{m+1} \beta_j = 1$, such that

$$\sum_{j=1}^{m+1} \beta_j v^j - g(x) \in \mathcal{N}_{\mathbb{R}^m_+}(u).$$
(4.61)

Proof. Using Rockafellar [210, Thm. 27.4], the necessary and sufficient optimality condition for (4.52) has the form

$$0 \in -\partial \Psi(u) + \mathcal{N}_{\mathbb{R}^m_+}(u). \tag{4.62}$$

Applying Corollary 4.75, we conclude that there exist

$$p\text{-efficient points } v^{j} \in V(u), \quad j = 1, \dots, m+1,$$

$$\beta^{j} \ge 0, \quad j = 1, \dots, m+1, \quad \sum_{j=1}^{m+1} \beta_{j} = 1,$$

$$x^{j} \in X(u), \quad j = 1, \dots, m+1, \quad \sum_{j=1}^{m+1} \alpha_{j} = 1,$$

$$\alpha^{j} \ge 0, \quad j = 1, \dots, m+1, \quad \sum_{j=1}^{m+1} \alpha_{j} = 1,$$

(4.63)

such that

$$\sum_{j=1}^{m+1} \alpha_j g(x^j) - \sum_{j=1}^{m+1} \beta_j v^j \in -\mathcal{N}_{\mathbb{R}^m_+}(u).$$
(4.64)

If the function c was strictly convex, or g was strictly concave, then the set X(u) would be a singleton. In this case, all x^j would be identical and the above relation would immediately imply (4.61). Otherwise, let us define

$$x = \sum_{j=1}^{m+1} \alpha_j x^j$$

By the convexity of X(u) we have $x \in X(u)$. Consequently,

$$c(x) - \sum_{i=1}^{m} u_i g_i(x) = h(u) = c(x^j) - \sum_{i=1}^{m} u_i g_i(x^j), \quad j = 1, \dots, m+1.$$
(4.65)

Multiplying the last equation by α_j and adding, we obtain

$$c(x) - \sum_{i=1}^{m} u_i g_i(x) = \sum_{j=1}^{m+1} \alpha_j \left[c(x^j) - \sum_{i=1}^{m} u_i g_i(x^j) \right] \ge c(x) - \sum_{i=1}^{m} u_i \sum_{j=1}^{m+1} \alpha_j g_i(x^j).$$

The last inequality follows from the convexity of c. We have the following inequality:

$$\sum_{i=1}^{m} u_i \Big[g_i(x) - \sum_{j=1}^{m+1} \alpha_j g_i(x^j) \Big] \le 0.$$

Since the functions g_i are concave, we have $g_i(x) \ge \sum_{j=1}^{m+1} \alpha_j g_i(x^j)$. Therefore, we conclude that $u_i = 0$ whenever $g_i(x) > \sum_{j=1}^{m+1} \alpha_j g_i(x^j)$. This implies that

$$g(x) - \sum_{j=1}^{m+1} \alpha_j g(x^j) \in -\mathcal{N}_{\mathbb{R}^m_+}(u).$$

Since $\mathcal{N}_{\mathbb{R}^m_+}(u)$ is a convex cone, we can combine the last relation with (4.64) and obtain (4.61), as required.

Now, we prove the converse implication. Assume that we have $x \in X(u)$, points $v^1, \ldots, v^{m+1} \in V(u)$ and scalars $\beta_1 \ldots, \beta_{m+1} \ge 0$ with $\sum_{j=1}^{m+1} \beta_j = 1$, such that (4.61) holds true. By Lemma 4.73 and Lemma 4.74 we have

$$-g(x) + \sum_{j=1}^{m+1} \beta_j v^j \in \partial \Psi(u).$$

Thus (4.61) implies (4.62), which is a necessary and sufficient optimality condition for problem (4.52). \Box

Using these optimality conditions we obtain the following duality result.

Theorem 4.77. Assume that the constraint qualification condition (4.55) for problem (4.49) is satisfied, the probability distribution function of the vector Z is α -concave for some $\alpha \in [-\infty, \infty]$, and the set \mathcal{X} is compact. If a point (\hat{x}, \hat{z}) is an optimal solution of (4.49), then there exists a vector $\hat{u} \ge 0$, which is an optimal solution of (4.52) and the optimal values of both problems are equal. If \hat{u} is an optimal solution of problem (4.52), then there exist a point \hat{x} , such that $(\hat{x}, g(\hat{x}))$ is a solution of problem (4.49), and the optimal values of both problems are equal.

Proof. The α -concavity assumption implies that problems (4.49) and (4.54) are the same. If \hat{u} is optimal solution of problem (4.52), we obtain the existence of points $\hat{x} \in X(\hat{u})$, $v^1, \ldots, v^{m+1} \in V(u)$ and scalars $\beta_1 \ldots, \beta_{m+1} \ge 0$ with $\sum_{j=1}^{m+1} \beta_j = 1$, such that the optimality conditions in Theorem 4.76 are satisfied. Setting $\hat{z} = g(\hat{x})$ we have to show that (\hat{x}, \hat{z}) is an optimal solution of problem (4.49) and that the optimal values of both problems are equal. First we observe that this point is feasible. We choose $y \in -\mathcal{N}_{\mathbb{R}^m_+}(\hat{u})$ such that $y = g(\hat{x}) - \sum_{j=1}^{m+1} \beta_j v^j$. From the definitions of $X(\hat{u})$, $V(\hat{u})$, and the normal cone, we obtain

$$h(\hat{u}) = c(\hat{x}) - \hat{u}^{\mathsf{T}} g(\hat{x}) = c(\hat{x}) - \hat{u}^{\mathsf{T}} \Big(\sum_{j=1}^{m+1} \beta_j v^j + y \Big)$$
$$= c(\hat{x}) - \sum_{j=1}^{m+1} \beta_j d(\hat{u}) - \hat{u}^{\mathsf{T}} y = c(\hat{x}) - d(\hat{u}).$$

Thus,

$$c(\hat{x}) = h(\hat{u}) + d(\hat{u}) = \Psi^* \ge c^*,$$

which proves that (\hat{x}, \hat{z}) is an optimal solution of problem (4.49) and $\Psi^* = c^*$.

If (\hat{x}, \hat{z}) is a solution of (4.49), then by Rockafellar [210, Thm. 28.4] there is a vector $\hat{u} \ge 0$ such that $\hat{u}_i(\hat{z}_i - g_i(\hat{x})) = 0$ and

$$0 \in \partial c(\hat{x}) + \partial \hat{u}^{\mathsf{T}} g(\hat{x}) - \hat{z} + \mathcal{N}_{\mathcal{X} \times \mathcal{Z}_p}(\hat{x}, \hat{z}).$$

This means that

$$0 \in \partial c(\hat{x}) - \partial u^{\mathsf{T}} g(\hat{x}) + \mathcal{N}_{\mathcal{X}}(\hat{x})$$
(4.66)

and

$$0 \in \hat{u} + \mathcal{N}_{\mathcal{Z}_n}(\hat{z}). \tag{4.67}$$

The first inclusion (4.66) is optimality condition for problem (4.50), and thus $x \in X(\hat{u})$. By virtue of Rockafellar [210, Thm. 23.5] the inclusion (4.67) is equivalent to $\hat{z} \in \partial \mathbb{I}^*_{\mathcal{Z}_p}(\hat{u})$. Using Lemma 4.73 we obtain that

$$\hat{z} \in \partial d(\hat{u}) = \operatorname{conv} V(\hat{u}) - \mathcal{N}_{\mathbb{R}^m_+}(\hat{u}).$$

Thus, there exists points $v^1, \ldots, v^{m+1} \in V(u)$ and scalars $\beta_1 \ldots, \beta_{m+1} \ge 0$ with $\sum_{j=1}^{m+1} \beta_j = 1$, such that

$$\hat{z} - \sum_{j=1}^{m+1} \beta_j v^j \in -\mathcal{N}_{\mathbb{R}^m_+}(\hat{u}).$$

Using the complementarity condition $\hat{u}_i(\hat{z}_i - g_i(\hat{x})) = 0$ we conclude that the optimality conditions of Theorem 4.76 are satisfied. Thus, \hat{u} is an optimal solution of problem (4.52).

Our duality theory finds interesting interpretation in the context of the cash matching problem in Example 4.6.

Example 4.78 (Cash Matching continued) Recall the problem formulation

$$\begin{aligned} \max_{x,c} \mathbb{E} \big[U(c_T - Z_T) \big] \\ \text{s.t. } \mathsf{Pr} \big\{ c_t \ge Z_t, \ t = 1, \dots, T \big\} \ge p, \\ c_t = c_{t-1} + \sum_{i=1}^n a_{it} x_i, \quad t = 1, \dots, T, \\ x \ge 0. \end{aligned}$$

If the vector Z has a quasi-concave distribution (e.g., joint normal distribution), the resulting problem is convex.

The convex hull problem (4.54) can be written as follows:

$$\max_{x,\lambda,c} \mathbb{E} \left[U(c_T - Z_T) \right] \tag{4.68}$$

s.t.
$$c_t = c_{t-1} + \sum_{i=1}^n a_{it} x_i, \quad t = 1, \dots, T,$$
 (4.69)

$$c_t \ge \sum_{j=1}^{T+1} \lambda_j v_t^j, \quad t = 1, \dots, T,$$
(4.70)

$$\sum_{j=1}^{T+1} \lambda_j = 1, \tag{4.71}$$

$$\lambda \ge 0, \ x \ge 0. \tag{4.72}$$

In constraint (4.70) the vectors $v^j = (v_1^j, \ldots, v_T^j)$, for $j = 1, \ldots, T + 1$, are *p*-efficient trajectories of the cumulative liabilities $Z = (Z_1, \ldots, Z_T)$. Constraints (4.70)–(4.72) require that the cumulative cash flows are greater than or equal to some convex combination of *p*-efficient trajectories. Recall that by Lemma 4.63, no more than T + 1 *p*-efficient trajectories are needed. Unfortunately, we do not know the optimal collection of these trajectories.

Let us assign non-negative Lagrange multipliers $u = (u_1, \ldots, u_T)$ to the constraint (4.70), multipliers $w = (w_1, \ldots, w_T)$ to the constraints (4.69), and a multiplier $\rho \in \mathbb{R}$ to the constraint (4.71). To simplify notation, we define the function $\overline{U} : \mathbb{R} \to \mathbb{R}$ as follows:

$$U(y) = \mathbb{E}[U(y - Z_T)].$$

It is a concave nondecreasing function of y due to the properties of $U(\cdot)$. We make the convention that its conjugate is defined as follows:

$$\bar{U}^*(u) = \inf_y \{ uy - \bar{U}(y) \}.$$

Consider the dual function of the convex hull problem:

$$D(w, u, \rho) = \min_{x \ge 0, \lambda \ge 0, c} \left\{ -\bar{U}(c_T) + \sum_{t=1}^T w_t \left(c_t - c_{t-1} - \sum_{i=1}^n a_{it} x_i \right) \right. \\ \left. + \sum_{t=1}^T u_t \left(\sum_{j=1}^{T+1} \lambda_j v_t^j - c_t \right) + \rho \left(1 - \sum_{j=1}^{T+1} \lambda_j \right) \right\} \\ = -\max_{x \ge 0} \sum_{i=1}^n \sum_{t=1}^T a_{it} w_t x_i + \min_{\lambda \ge 0} \sum_{j=1}^{T+1} \left(\sum_{t=1}^T v_t^j u_t - \rho \right) \lambda_j + \rho \\ \left. + \min_c \left\{ \sum_{t=1}^{T-1} c_t \left(w_t - u_t - w_{t+1} \right) - w_1 c_0 + c_T (w_T - u_T) - \bar{U}(c_T) \right\} \\ = \rho - w_1 c_0 + \bar{U}^* (w_T - u_T). \right\}$$

The dual problem becomes:

$$\min_{u,w,\rho} - \bar{U}^*(w_T - u_T) + w_1 c_0 - \rho \tag{4.73}$$

s.t.
$$w_t = w_{t+1} + u_t, \quad t = T - 1, \dots, 1,$$
 (4.74)

$$\sum_{t=1}^{l} w_t a_{it} \le 0, \quad i = 1, \dots, n,$$
(4.75)

$$\rho \le \sum_{t=1}^{\mathsf{T}} u_t v_t^j, \quad j = 1, \dots, T+1.$$
(4.76)

$$u \ge 0. \tag{4.77}$$

We can observe that each dual variable u_t is the cost of borrowing a unit of cash for one time period, t. The amount u_t is to be paid at the end of the planning horizon. It follows from (4.74) that each multiplier w_t is the amount that has to be returned at the end of the planning horizon if a unit of cash is borrowed at t and held until time T.

The constraints (4.75) represent the *non-arbitrage condition*. For each bond i we can consider the following operation: borrow money to buy the bond and lend away its coupon payments, according to the rates implied by w_t 's. At the end of the planning horizon, we collect all loans and pay off the debt. The profit from this operation should be non-positive for each bond in order to comply with the "no free lunch" condition, which is expressed via (4.75).

Let us observe that each product $u_t v_t^j$ is the amount that has to be paid at the end, for having a debt in the amount v_t^j in period t. Recall that v_t^j is the p-efficient cumulative liability up to time t. Denote the implied one-period liabilities by

$$L_t^j = v_t^j - v_{t-1}^j, \quad t = 2, \dots, T,$$

$$L_1^j = v_1^j.$$

Changing the order of summation, we obtain

$$\sum_{t=1}^{T} u_t v_t^j = \sum_{t=1}^{T} u_t \sum_{\tau=1}^{t} L_{\tau}^j = \sum_{\tau=1}^{T} L_{\tau}^j \sum_{t=\tau}^{T} u_t = \sum_{\tau=1}^{T} L_{\tau}^j (w_{\tau} + u_T - w_T).$$

It follows that the sum appearing on the right hand side of (4.76) can be viewed as the extra cost of covering the *j*th *p*-efficient liability sequence by borrowed money, that is, the difference between the amount that has to be returned at the end of the planning horizon, and the total liability discounted by $w_T - u_T$.

If we consider the special case of a linear expected utility:

$$\hat{U}(c_T) = c_T - \mathbb{E}[Z_T],$$

Then we can skip the constant $\mathbb{E}[Z_T]$ in the formulation of the optimization problem. The dual function of the convexified cash matching problem becomes:

$$D(w, u, \rho) = -\max_{x \ge 0} \sum_{i=1}^{n} \sum_{t=1}^{T} a_{it} w_t x_i + \min_{\lambda \ge 0} \sum_{j=1}^{T+1} \left(\sum_{t=1}^{T} v_t^j u_t - \rho \right) \lambda_j + \rho$$
$$+ \min_c \left\{ \sum_{t=1}^{T-1} c_t \left(w_t - u_t - w_{t+1} \right) - w_1 c_0 + c_T (w_T - u_T - 1) \right\}$$
$$= \rho - w_1 c_0.$$

The objective function of the dual problem takes on the form:

$$\mathop{\rm Min}_{u,w,\rho} w_1 c_0 - \rho$$

and the constraints (4.74) extends to all time periods:

$$w_t = w_{t+1} + u_t, \quad t = T, T - 1, \dots, 1$$

with the convention $w_{T+1} = 1$.

In this case, the sum on the right hand side of (4.76) is the difference between the cost of covering the *j*th *p*-efficient liability sequence by borrowed money and the total liability.

The variable ρ represents the minimal cost of this form, for all *p*-efficient trajectories. This allows us to interpret the dual objective function in this special case as the amount obtained at *T* for lending away our capital c_0 decreased by the extra cost of covering a *p*-efficient liability sequence by borrowed money. By duality this quantity is the same as c_T , which implies that both ways of covering the liabilities are equally profitable. In the case of a general utility function, the dual objective function contains an additional adjustment term.

Optimality conditions for linear chance-constrained problems

For the special case of discrete distribution and linear constraints under the probability sign, we can obtain more specific necessary and sufficient optimality conditions.

In the *linear* probabilistic optimization problem, we have g(x) = Tx, where T is an $m \times n$ matrix, and $c(x) = c^{\mathsf{T}}x$ with $c \in \mathbb{R}^n$. Furthermore, we assume that \mathcal{X} is a closed convex polyhedral set, defined by a system of linear inequalities. The problem reads as follows:

$$\begin{array}{ll}
\operatorname{Min}_{x} & c^{\mathsf{T}}x \\
\text{s.t.} & \operatorname{Pr}\{Tx \geq Z\} \geq p, \\
& Ax \geq b, \\
& x \geq 0.
\end{array}$$
(4.78)

Here A is an $s \times n$ matrix and $b \in \mathbb{R}^s$.

We say that problem (4.78) satisfies the dual feasibility condition if

$$\Lambda = \{(u, w) \in \mathbb{R}^{m+s}_+ : A^T w + T^T u \le c\} \neq \emptyset.$$
(4.79)

Theorem 4.79. Assume that the feasible set of problem (4.78) is nonempty and that Z has a discrete distribution on \mathbb{Z}^m . Then problem (4.78) has an optimal solution if and only if it satisfies condition (4.79).

Proof. If (4.78) has an optimal solution, then for some $j \in \mathcal{E}$ the linear optimization problem

$$\begin{array}{l} \underset{x}{\operatorname{Min}} c^{\mathsf{T}} x \\ \text{s.t.} \ T x \ge v^{j}, \\ A x \ge b, \\ x \ge 0, \end{array} \tag{4.80}$$

has an optimal solution. By duality in linear programming, its dual problem

$$\begin{aligned}
& \underset{u,w}{\operatorname{Max}} u^{\mathsf{T}} v^{j} + b^{\mathsf{T}} w \\
& \text{s.t. } T^{\mathsf{T}} u + A^{\mathsf{T}} w \leq c, \\
& u, w \geq 0,
\end{aligned} \tag{4.81}$$

has an optimal solution and the optimal values of both programs are equal. Thus, the dual feasibility condition (4.79) must be satisfied. On the other hand, if the dual feasibility condition is satisfied, all dual programs (4.81) for $j \in \mathcal{E}$ have nonempty feasible sets, so the objective values of all primal problems (4.80) for $j \in \mathcal{E}$ are bounded from below. At least one of them has a nonempty feasible set by assumption and, hence, an optimal solution of problem (4.78) exist. \Box

Example 4.80 (Vehicle Routing continued) We return to the vehicle routing Example 4.1,

introduced at the beginning of the chapter. The convex hull problem reads:

$$\begin{aligned}
& \underset{x,\lambda}{\min} c^{\mathsf{T}}x \\
& \text{s.t. } \sum_{i=1}^{n} t_{il}x_{i} \geq \sum_{j \in \mathcal{E}} \lambda_{j}v^{j}, \\
\end{aligned}$$
(4.82)

$$\sum_{j \in \mathcal{E}} \lambda_j = 1,$$

$$x \ge 0, \ \lambda \ge 0.$$
(4.83)

We assign a Lagrange multiplier u to constraint (4.82) and a multiplier u to constraint

$$\begin{split} \max_{u,\mu} \mu \\ \text{s.t.} & \sum_{l=1}^{m} t_{il} u_l \leq c_i, \quad i = 1, 2, \dots, n, \\ & \mu \leq u^{\mathsf{T}} v^j, \quad j \in \mathcal{E}, \\ & u \geq 0. \end{split}$$

We see that u_l provides the increase of routing cost if the demand on arc l increases by 1 unit, μ is the minimum cost for covering the demand with probability p, and the p-efficient points v^{j} correspond to critical demand levels, that have to be covered. The auxiliary problem $\min_{z \in T} u^{\mathsf{T}} z$ identifies *p*-efficient points, which represent critical demand levels. The $z \in \mathcal{Z}_p$

optimal value of this problem provides the minimum total cost of a critical demand.

Optimality conditions under differentiability assumptions

Marr

We discuss the first and second order local optimality conditions for the following problem:

$$\min c(x)$$
subject to $g(x) \in \operatorname{conv} \mathcal{Z}_p$, (4.84)
$$x \in \mathcal{X}.$$

We assume that c and g are continuously differentiable and denote the gradient of the function c calculated at x by $\nabla c(x)$ and the Jacobian of q calculated at x by q'(x). We do not assume convexity or concavity for these functions.

First, we formulate a suitable constraint qualification condition at a given feasible point \hat{x} . Let $g(\hat{x}) = \hat{w} + \hat{y}$ with $\hat{w} \in E$ and $\hat{y} \in \mathbb{R}^m_+$. Let $I_0(\hat{x})$ be the set of active constraints such that $I_0(\hat{x}) = \{1 \le i \le m : \hat{y}_i = 0\}.$

A vector
$$\delta \in \mathcal{K}_{\mathcal{X}}(\hat{x})$$
 and a convex combination of *p*-efficient points w^0 exist:

$$\forall i \in I_0(\hat{x}) \quad g_i(\hat{x}) + \langle \nabla g_i(\hat{x}), \delta \rangle > w_i^0.$$
(4.85)

Recall that Robinson's constraint qualification condition at \hat{x} for problem (4.84) is the requirement

$$0 \in \inf\{g'(\hat{x})(x-\hat{x}) - (z-g(\hat{x})) : x \in \mathcal{X}, z \in \operatorname{conv}\mathcal{Z}_p\}.$$
(4.86)

Lemma 4.81. Condition (4.85) is equivalent to Robinson's constraint qualification condition for problem (4.84) at a feasible point \hat{x} .

Proof. Assume that the Robinson's constraint qualification holds at \hat{x} for problem (4.84). Then we can find a vector $\xi \in \mathbb{R}^m$ with $\xi_i > 0$, as well as points $x^0 \in \mathcal{X}$ and $z^0 \in \text{conv}\mathcal{Z}_p$ such that

$$g'(\hat{x})(x^0 - \hat{x}) - (z^0 - g(\hat{x})) = \xi.$$

We represent $z^0 = \sum_{j \in J} \alpha_j v^j + y^0$, where $y^0 \in \mathbb{R}^m_+$, $\alpha_j \ge 0$, $\sum_{j \in J} \alpha_j = 1$, and v^j are *p*-efficient points. Substituting z^0 in the last displayed equation, we obtain

$$g'(\hat{x})(x^0 - \hat{x}) + g(\hat{x}) - \sum_{j \in J} \alpha_j v^j = y^0 + \xi.$$

As all components of $y^0 + \xi$ are positive, we infer that relation (4.85) holds.

Consider the opposite direction and assume that relation (4.85) holds, i.e., for some points $x^0 \in \mathcal{X}$ and $w^0 \in \varepsilon$

$$g'(\hat{x})(x^0 - \hat{x}) + g(\hat{x}) - w^0 = \eta$$
 with $\eta_i > 0, \, \forall i \in I_0(\hat{x})$

Let $g(\hat{x}) = \hat{w} + \hat{y}$ with $\hat{w} \in \varepsilon$ and $\hat{y} \in \mathbb{R}^m_+$. Consider the functions

$$y_i(\alpha) = (1 - \alpha)\hat{y}_i + \alpha\eta_i, \quad i = 1, \dots, m.$$

We have $y_i(\alpha) > 0$ for all $i \in I_0(\hat{x})$ for all $\alpha \in (0, 1]$. Furthermore, for $\alpha = 0$, we have $y_i(\alpha) > 0$ for all $i \notin I_0(\hat{x})$. Thus, a number $\bar{\alpha} \in (0, 1)$ exists such that $y_i(\bar{\alpha}) > 0$ for all $i = 1, \ldots, m$. We set $\varepsilon = \min_{1 \le i \le m} y_i(\bar{\alpha})$. We shall show that

$$B(0,\varepsilon) \subset \{g'(\hat{x})(x-\hat{x}) - (z-g(\hat{x})) : x \in \mathcal{X}, z \in \operatorname{conv} \mathbb{Z}_p\},\$$

where $B(0,\varepsilon)$ denotes the open ball centered at $0 \in \mathbb{R}^m$ with radius ε . Define $\bar{x} = (1 - \bar{\alpha})\hat{x} + \bar{\alpha}x^0$. It is sufficient to show that we can represent every vector $\xi \in B(0,\varepsilon)$ as follows:

$$\xi = g'(\hat{x})(\bar{x} - \hat{x}) - (z_{\xi} - g(\hat{x})) = g'(\hat{x})\bar{\alpha}(x^0 - \hat{x}) - (z_{\xi} - g(\hat{x}))$$

for some $z_{\xi} \in \operatorname{conv} \mathcal{Z}_p$. The equation is satisfied with

$$z_{\xi} = \bar{\alpha}w^0 + (1-\bar{\alpha})\hat{w} + \bar{\alpha}\eta + (1-\bar{\alpha})\hat{y} - \xi.$$

Notice that $\bar{\alpha}\eta + (1 - \bar{\alpha})\hat{y} - \xi \in \mathbb{R}^m_+$ due to the choice of ε . This entails that $z_{\xi} \in \operatorname{conv} \mathcal{Z}_p$ by virtue of Lemma 4.63. Consequently, Robinson's constraint qualification condition is satisfied at \hat{x} . \Box

We denote the feasible set of problem (4.84) by Φ , i.e,

$$\Phi = \{ x \in \mathcal{X} : g(x) \in \operatorname{conv} \mathcal{Z}_p \}.$$

Proposition 4.82. Under constraint qualification (4.85) at a point $\hat{x} \in \Phi$, the cone:

$$\mathcal{T}_{\Phi}(\hat{x}) = \left\{ \delta \in \mathcal{T}_{\mathcal{X}}(\hat{x}) : \exists w \in E, \exists \beta > 0 : g_i(\hat{x}) + \langle \nabla g_i(\hat{x}), \beta \delta \rangle \ge w_i \; \forall i \in I_0(\hat{x}) \right\}$$

$$(4.87)$$

is the tangent cone of the feasible set Φ at \hat{x} .

Proof. Under constraint qualification condition (4.85) at any feasible point \hat{x} , the tangent cone has the form

$$\mathcal{T}_{\Phi}(\hat{x}) = \left\{ \delta \in \mathcal{T}_{\mathcal{X}}(\hat{x}) : g'(\hat{x})\delta \in \mathcal{T}_{\operatorname{conv}\mathcal{Z}_{p}}(g(\hat{x})) \right\}.$$

We use the representation $g(\hat{x}) = \hat{w} + \hat{y}$ with $\hat{w} \in E$ an $\hat{y} \ge 0$.

If δ belongs to the set at the right hand side of equation (4.87), we can verify that $g'(\hat{x})\delta \in \mathcal{T}_{\operatorname{conv}\mathcal{Z}_p}(g(\hat{x}))$ in the following way. For all $i \in I_0(\hat{x})$, we define

$$y_i = \langle \nabla g_i(\hat{x}), \delta \rangle - \frac{1}{\beta} (w_i - g_i(\hat{x})) > 0.$$

For all $i \notin I_0(\hat{x})$, we define y_i such that

$$y_i = \langle \nabla g_i(\hat{x}), \delta \rangle - \frac{1}{\beta} (w_i - \hat{w}) - \tau \hat{y}_i.$$

We choose $\tau \in \mathbb{R}$ such that $y_i > 0$. Consequently,

$$g'(\hat{x})\delta = \frac{1}{\beta}(w - g(\hat{x})) + \tau_1\hat{y} + y,$$

where $\tau_1 = \tau + \frac{1}{\beta}$. Due to Proposition 4.70, $g'(\hat{x})\delta \in \mathcal{K}_{\operatorname{conv}\mathcal{Z}_p}(\hat{x})$.

In order to show the converse, we consider the case of $g'(\hat{x})\delta \in \mathcal{K}_{\operatorname{conv}\mathcal{Z}_p}(g(\hat{x}))$ first. Due to Proposition 4.70, we can write

$$g'(\hat{x})\delta = \sum_{j \in J} \gamma_j (v^j - \hat{w}) + \tau \hat{y} + y.$$

for some $\gamma_j \ge 0, j \in J$ not all equal to zero, a real number τ , and some vector $y \in \mathbb{R}^m_+$. For $i \in I_0(\hat{x})$, it holds that

$$\langle \nabla g_i(\hat{x}), \delta \rangle \ge \sum_{j \in J} \gamma_j(v_i^j - \hat{w}_i) \quad \text{and} \quad \hat{w}_i = g_i(\hat{x}).$$

Setting $\beta = [\sum_{j \in J} \gamma_j]^{-1} > 0$ and $\alpha_j = \gamma_j \beta$, we conclude that

$$g_i(\hat{x}) + \langle \nabla g_i(\hat{x}), \beta \delta \rangle \ge \sum_{j \in J} \alpha_j v_i^j,$$

as required.

Now, assume that $g'(\hat{x})\beta\delta = \lim_{k\to\infty} z^k$, where $z^k \in \mathcal{K}_{\operatorname{conv}\mathcal{Z}_p}(g(\hat{x}))$. For any $\varepsilon > 0$, using the previous considerations, for all $i \in I_0(\hat{x})$ the following inequality holds:

$$g_i(\hat{x}) + \langle \nabla g_i(\hat{x}), \beta \delta \rangle + \varepsilon \ge w_i^{\varepsilon}, \tag{4.88}$$

where $w^{\varepsilon} \in E$. We notice that the sequence $\{w_{\varepsilon}\}$ is bounded from above due to (4.88) as well as from below due to the existence of a lower bound for all *p*-efficient points. Therefore, passing to the limit with $\varepsilon \to 0$, $\{w_{\varepsilon}\}$ has an accumulation point, which we denote by \tilde{w} . The point $\tilde{w} \in \text{conv} \mathcal{Z}_p$ because this set is closed. Thus, a point $\bar{v} \in E$ exists such that $\tilde{w} \geq \bar{v}$. This combined with (4.88) after limit passage yields:

$$g_i(\hat{x}) + \langle \nabla g_i(\hat{x}), \beta \delta \rangle \ge \bar{v}_i \quad \forall i \in I_0(\hat{x}).$$

We define the partial Lagrangian function:

$$L_r(x,u) = c(x) - \langle u, g(x) \rangle.$$
(4.89)

We obtain the following form of first order necessary conditions of optimality for problem (4.84).

Theorem 4.83. Assume that \hat{x} is a local optimal solution of problem (4.84) and that condition (4.85) is satisfied at \hat{x} . Then a vector $\hat{u} \in \mathbb{R}^m_+$, p-efficient points v^j , $j \in \hat{J}$, and positive reals α_j , $j \in \hat{J}$ exists such that $\sum_{j \in \hat{J}} \alpha_j = 1$ such that the following conditions are satisfied:

$$-\nabla_{x}L_{r}(\hat{x},\hat{u}) \in \mathcal{N}_{\mathcal{X}}(\hat{x}),$$

$$\langle g(\hat{x}),\hat{u}\rangle = \langle v^{j},\hat{u}\rangle \leq \langle v^{i},\hat{u}\rangle, \forall j \in \hat{J}, \forall i \in J$$

$$g(\hat{x}) \geq \sum_{j \in \hat{J}} \alpha_{j}v_{i}^{j}.$$
(4.90)

If the function c is convex, g is a concave mapping, and a point $\hat{x} \in \mathcal{X}$ satisfies conditions (4.90) with some vector $\hat{u} \in \mathbb{R}^m_+$, then \hat{x} is global optimal solution of problem (4.84).

Proof. Condition (4.85) is equivalent to Robinson's constraint qualification by virtue of Lemma 4.81. The theory of smooth non-convex optimization problem of form (4.84) implies that $-\nabla_x L_r(\hat{x}, \hat{u}) \in \mathcal{N}_{\mathcal{X}}(\hat{x})$ and $-\hat{u} \in \mathcal{N}_{\text{conv}\mathcal{Z}_p}(g(\hat{x}))$. The form of the normal cone is given in Proposition 4.71. It implies that a point $\hat{w} \in E$ exists, so that $g(\hat{x}) \geq \hat{w}$ and $\hat{u}_i = 0$ for all *i* such that $g(\hat{x}) > \hat{w}$. The remaining conditions follow from the description of the normal cone.

Observe that the condition $-\nabla_x L_r(\hat{x}, \hat{u}) \in \mathcal{N}_{\mathcal{X}}(\hat{x})$ implies that the point $\hat{x} \in X(\hat{u})$, where the set $X(\hat{u})$ is defined in (4.60). The remaining two conditions of (4.90) imply that \hat{x} is feasible and

$$g(\hat{x}) - \sum_{j \in \hat{J}} \alpha_j v_i^j \in -\mathcal{N}_{\mathbb{R}^m_+}(\hat{u}).$$

Under the convexity assumptions of the theorem, the sufficiency of the conditions (4.90) follows by virtue of Theorem 4.76. \Box

The second order tangent set of $\operatorname{conv} \mathcal{Z}_p$ at a point z is

$$\mathcal{T}_{\operatorname{conv}\mathcal{Z}_p}^2(z,\delta) = \limsup_{t\downarrow 0} \frac{\operatorname{conv}\mathcal{Z}_p - z - t\delta}{\frac{1}{2}t^2}.$$

We define also the *critical cone*:

$$\mathfrak{C}(x^0) = \{ \delta \in \mathcal{T}_{\mathcal{X}}(\hat{x}) : g'(\hat{x})\delta \in \mathcal{T}_{\operatorname{conv}\mathcal{Z}_n}(g(\hat{x})), \, \langle \nabla c(\hat{x}), \delta \rangle = 0 \}.$$

For a point \hat{x} that satisfies (4.90), we denote the set of vectors \hat{u} appearing in (4.90) by $\Lambda(\hat{x})$. Local second order necessary conditions of optimality for problem (4.84) are given by the following theorem.

Theorem 4.84. Assume that a point $\hat{x} \in \Phi$ satisfies conditions (4.85) and (4.90). Then for every nonzero $\delta \in \mathfrak{C}(\hat{x})$ and for every convex set $C \subset \mathcal{T}^2_{\mathcal{X} \times \operatorname{conv} \mathcal{Z}_p}((\hat{x}, g(\hat{x})), (\delta, g'(\hat{x})\delta)),$ the following inequality is satisfied:

$$\sup_{u \in \Lambda(\hat{x})} \left\{ \langle \delta, \nabla_x^2 L_r(\hat{x}, u) \delta \rangle + \sup_{z \in C} \langle -(\nabla_x L_r(\hat{x}, u), u), z \rangle \right\} \ge 0.$$
(4.91)

Proof. The statement follows from [26, Theorem 3.49] after an obvious adaptation with the particular form of the constraint qualification condition, the form of cones and using the first order optimality conditions (4.90). \Box

If we assume that the random vector Y has a discrete distribution such that a strictly positive lower bound on the distance between any two realizations of Y exists, then the set $\operatorname{conv} \mathbb{Z}_p$ is polyhedral. The second order necessary conditions simplifies considerably.

Theorem 4.85. Assume that a point \hat{x} satisfies conditions (4.85) and (4.90) with points v^j , $j \in \hat{J}$ and positive numbers α_j . Then for all nonzero $\delta \in \mathcal{T}_{\mathcal{X}}(\hat{x})$ such that $\langle \nabla c(\hat{x}), \delta \rangle = 0$ and $g_i(\hat{x}) + \langle \nabla g_i(\hat{x}), \delta \rangle \geq \sum_{i \in \hat{J}} \alpha_j v_i^j$ for all $i \in I_0(\hat{x})$ the following inequality holds

$$\sup_{u \in \Lambda(\hat{x})} \langle \delta, \nabla_x^2 L_r(\hat{x}, u) \delta \rangle \ge 0.$$
(4.92)

Proof. Under the assumption of discrete distribution, we can apply Proposition 4.82 to obtain the following representation of the set $\mathfrak{C}(\hat{x})$:

$$\mathfrak{C}(\hat{x}) = \{ \delta \in \mathcal{T}_{\mathcal{X}}(\hat{x}) : g_i(\hat{x}) + \langle \nabla g_i(\hat{x}), \beta \delta \rangle \ge w_i \ \forall i \in I_0(\hat{x}), \ \langle \nabla c(\hat{x}), \delta \rangle = 0 \}.$$

Evidently, the requirement (4.92) can be restricted to $\delta \in \mathfrak{C}(\hat{x})$ such that the corresponding coefficient $\beta = 1$. Then the statement follows from [26, Theorem 3.53].

Second order sufficient conditions of optimality can be established for general probability distributions. **Theorem 4.86.** Assume that $\hat{x} \in \Phi$ satisfies conditions (4.85) and (4.90) and for some $u \in \Lambda(\hat{x})$

$$\langle \delta, \nabla_x^2 L_r(\hat{x}, u) \delta \rangle > 0 \tag{4.93}$$

for every nonzero $\delta \in \mathcal{T}_{\mathcal{X}}(\hat{x})$ such that $\langle \nabla c(\hat{x}), \delta \rangle = 0$ and $\langle u, g'(\hat{x})\delta \rangle = 0$. Then \hat{x} is a local minimizer of (4.84).

Proof. Suppose that \hat{x} is not a local minimum of problem (4.84). Then, points $x^t \in \mathcal{X}$ exist such that $\hat{x} = \lim_{t \downarrow 0} x^t$, $g(x^t) \in \operatorname{conv} \mathcal{Z}_p$, and $c(x^t) < c(\hat{x})$. This implies, that $\delta \in \mathcal{T}_{\mathcal{X}}(\hat{x})$ exists such that $x^t = \hat{x} + t\delta + o(t)$ and $g(x^t) - g(\hat{x}) \in \mathcal{K}_{\operatorname{conv} \mathcal{Z}_p}(g(\hat{x}))$. As $c(x^t) < c(\hat{x})$, we obtain that

$$\langle \nabla c(\hat{x}), \delta \rangle \le 0. \tag{4.94}$$

Due to the first relation in (4.90), the following relation holds for all $\delta \in \mathcal{T}_{\mathcal{X}}(\hat{x})$:

$$\langle \nabla c(\hat{x}), \delta \rangle - \langle [g'(\hat{x})]^{\mathsf{T}} u, \delta \rangle \ge 0.$$

Using this inequality together with (4.94), we obtain

$$\langle u, g'(\hat{x})\delta \rangle = \langle [g'(\hat{x})]^{\mathsf{T}}u, \delta \rangle \le \langle \nabla c(\hat{x}), \delta \rangle \le 0.$$

On the other hand, conditions (4.90) imply that $-u \in \mathcal{N}_{\operatorname{conv}\mathcal{Z}_p}(g(\hat{x}))$ due to Proposition 4.71. On the other hand, $g'(\hat{x})\delta \in \mathcal{T}_{\operatorname{conv}\mathcal{Z}_p}(g(\hat{x}))$ due to (4.85), which entails

$$\langle u, g'(\hat{x})\delta \rangle \ge 0.$$

Consequently, $\langle u, g'(\hat{x})\delta \rangle = 0$. Using (4.90), we obtain

$$0 \le \langle \nabla L_r(\hat{x}, u), \delta \rangle = \langle \nabla c(\hat{x}), \delta \rangle$$

This and (4.94) yield $\langle \nabla c(\hat{x}), \delta \rangle = 0$.

We consider the second order Taylor expansion of the function $L_r(\cdot, \hat{u})$ around \hat{x} and use (4.94) again :

$$L_r(x^t, u) = L_r(\hat{x}, u) + t \langle \nabla L_r(\hat{x}, u), \delta \rangle + \frac{t^2}{2} \langle \delta, \nabla_x^2 L_r(\hat{x}, u) \delta \rangle + o(t^2)$$

$$\geq c(\hat{x}) - \langle u, g(\hat{x}) \rangle + \frac{t^2}{2} \langle \delta, \nabla_x^2 L_r(\hat{x}, u) \delta \rangle + o(t^2)$$

Rearranging terms, we obtain

$$c(x^t) \ge c(\hat{x}) + \langle u, g(x^t) - g(\hat{x}) \rangle + \frac{t^2}{2} \langle \delta, \nabla_x^2 L_r(\hat{x}, u) \delta \rangle + o(t^2)$$

Using the second order condition (4.93) and the fact that $-u \in \mathcal{N}_{\operatorname{conv}\mathcal{Z}_p}(g(\hat{x}))$, we obtain the inequality $c(x^t) > c(\hat{x})$ in contradiction to our assumption. \Box

Notice that if c is a convex function and g is concave mapping, then condition (4.93) imply that x is a global solution of (4.84).

4.4 Optimization problems with non-separable probabilistic constraints

In this section, we concentrate on the following problem:

$$\begin{array}{l} \underset{x}{\operatorname{Min}} \quad c(x) \\ \text{s.t.} \quad \Pr\{g(x, Z) \ge 0\} \ge p, \\ \quad x \in \mathcal{X}. \end{array}$$
(4.95)

The parameter $p \in (0,1)$ denotes some probability level. We assume that the functions $c : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}$ and $g : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^m$ are continuous and the set $\mathcal{X} \subset \mathbb{R}^n$ is a closed convex set. We define the constraint function as follows:

$$G(x) = \mathsf{Pr}\{g(x, Z) \ge 0\}.$$

Recall that if $G(\cdot)$ is α -concave function, $\alpha \in \mathbb{R}$, then a transformation of it is a concave function. In this case, we define

$$\bar{G}(x) = \begin{cases} \ln p - \ln[G(x)], & \text{if } \alpha = 0, \\ p^{\alpha} - [G(x)]^{\alpha}, & \text{if } \alpha > 0, \\ [G(x)]^{\alpha} - p^{\alpha}, & \text{if } \alpha < 0. \end{cases}$$
(4.96)

We obtain the following equivalent formulation of problem (4.95):

$$\begin{array}{l} \underset{x}{\operatorname{Min}} \quad c(x) \\ \text{s.t.} \quad \bar{G}(x) \leq 0, \\ \quad x \in \mathcal{X}. \end{array}$$
(4.97)

Assuming that $c(\cdot)$ is convex, we have a convex problem.

Recall that Slater's condition is satisfied for problem (4.95) if there is a point $x^s \in int \mathcal{X}$ such that $\overline{G}(x^s) > 0$. Using optimality conditions for convex optimization problems, we can infer the following conditions for problem (4.95):

Theorem 4.87. Assume that $c(\cdot)$ is a continuous convex function, the functions $g : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^m$ are quasi-concave, Z has an α -concave distribution, and the set $\mathcal{X} \subset \mathbb{R}^n$ is closed and convex. Furthermore, let Slater's condition be satisfied and int dom $G \neq \emptyset$.

A point $\hat{x} \in \mathcal{X}$ is an optimal solution of problem (4.95) if and only if there is a number $\lambda \in \mathbb{R}_+$ such that $\lambda[G(\hat{x}) - p] = 0$ and

$$0 \in \partial c(\hat{x}) + \lambda \frac{1}{\alpha} G(\hat{x})^{1-\alpha} \partial G(\hat{x})^{\alpha} + \mathcal{N}_{\mathcal{X}}(\hat{x}) \qquad \text{if } \alpha \neq 0$$

or

$$0 \in \partial c(\hat{x}) + \lambda G(\hat{x}) \partial \left(\ln G(\hat{x}) \right) + \mathcal{N}_{\mathcal{X}}(\hat{x}) \qquad \text{if } \alpha = 0.$$

Proof. Under the assumptions of the theorem problem (4.95) can be reformulated in form (4.97), which is a convex optimization problem. The optimality conditions follow from the optimality conditions for convex optimization problems using Theorem 4.29. Due to Slater's condition, we have that G(x) > 0 on a set with non-empty interior, and therefore, the assumptions of Theorem 4.29 are satisfied. \Box

4.4.1 Differentiability of probability functions and optimality conditions

We can avoid concavity assumptions and replace them by differentiability requirements. Under certain assumptions, we can differentiate the probability function and obtain optimality conditions in a differential form. For this purpose, we assume that Z has a probability density function $\theta(z)$, and that the support of P_Z is a closed set with a piece-wise smooth boundary such that supp $P_Z = cl\{int(supp P_Z)\}$. For example, it can be the union of several disjoint sets, but cannot contain isolated points, or surfaces of zero Lebesgue measure.

Consider the multifunction $H : \mathbb{R}^n \rightrightarrows \mathbb{R}^s$, defined as follows:

$$H(x) = \{ z \in \mathbb{R}^s : g_i(x, z) \ge 0, i = 1, \dots, m \}.$$

We denote the boundary of a set H(x) by bdH(x). For an open set $U \subset \mathbb{R}^n$ containing the origin, we set:

$$H_U = \operatorname{cl}\left(\bigcup_{x \in U} H(x)\right) \quad \text{and} \quad \triangle H_U = \operatorname{cl}\left(\bigcup_{x \in U} \operatorname{bd} H(x)\right),$$
$$V_U = \operatorname{cl} U \times H_U \quad \text{and} \quad \triangle V_U = \operatorname{cl} U \times \triangle H_U.$$

For any of these sets, we indicate with upper subscript r its restriction to the supp P_Z , e.g., $H_U^r = H_U \cap \text{supp} P_Z$. Let

$$S_i(x) = \{ z \in \text{supp}P_Z : g_i(x, z) = 0, \ g_j(x, z) \ge 0, \ j \ne i \}, \quad i = 1, \dots, m.$$

We use the notation

$$S(x) = \bigcup_{i=1}^{M} S_i(x), \quad \triangle H_i = \operatorname{int} \big(\bigcup_{x \in U} \big(\partial \{ g_i(x, z) \ge 0 \} \cap H^r(x) \big) \big).$$

The (m-1)-dimensional Lebesgue measure is denoted by P_{m-1} . We assume that the functions $g_i(x, z), i = 1, ..., m$, are continuously differentiable, and such that $\operatorname{bd} H(x) = S(x)$ with S(x) being the (s-1)-dimensional surface of the set $H(x) \subset \mathbb{R}^s$. The set H_U is the union of all sets H(x) when $x \in U$, and, correspondingly, ΔH_U contains all surfaces S(x)when $x \in U$.

First we formulate and prove a result about the differentiability of the probability function for a single constraint function g(x, z), that is, m = 1. In this case we omit the index for the function g, as well as for the set S(x).

Theorem 4.88. Assume that:

- (i) the vector functions $\nabla_x g(x,z)$ and $\nabla_z g(x,z)$ are continuous on $\triangle V_U^r$;
- (ii) the vector functions $\nabla_z g(x, z) > 0$ (componentwise) on the set $\triangle V_U^r$;
- (iii) the function $\|\nabla_x g(x, z)\| > 0$ on $\triangle V_U^r$.

Then the probability function $G(x) = \Pr\{g(x, Z) \ge 0\}$ has partial derivatives for almost all $x \in U$ that can be represented as a surface integral

$$\left(\frac{\partial G(x)}{\partial x_i}\right)_{i=1}^n = \int\limits_{\mathrm{bd}H(x)\cap\mathrm{supp}P_Z} \frac{\theta(z)}{\|\nabla_z g(x,z)\|} \nabla_x g(x,z) dS.$$

Proof. Without loss of generality, we shall assume that $x \in U \subset \mathbb{R}$.

For two points $x, y \in U$, we consider the difference:

$$G(x) - G(y) = \int_{H(x)} \theta(z) dz - \int_{H(y)} \theta(z) dz$$

=
$$\int_{H^{r}(x) \setminus H^{r}(y)} \theta(z) dz - \int_{H^{r}(y) \setminus H^{r}(x)} \theta(z) dz.$$
 (4.98)

By the Implicit Function Theorem, the equation g(x, z) = 0 determines a differentiable function x(z) such that

$$g(x(z), z) = 0$$
 and $\nabla_z x(z) = -\frac{\nabla_z g(x, z)}{\nabla_x g(x, z)}\Big|_{x=x(z)}$

Moreover, the constraint $g(x, z) \ge 0$ is equivalent to $x \ge x(z)$ for all $(x, z) \in U \times \triangle H^r_U$, because the function $g(\cdot, z)$ strictly increases on this set due to the assumption (iii). Thus, for all points $x, y \in U$ such that x < y, we can write:

$$\begin{aligned} H^{r}(x) \setminus H^{r}(y) &= \{ z \in \mathbb{R}^{s} : g(x,z) \geq 0, \ g(y,z) < 0 \} = \{ z \in \mathbb{R}^{s} : x \geq x(z) > y \} = \emptyset, \\ H^{r}(y) \setminus H^{r}(x) &= \{ z \in \mathbb{R}^{s} : g(y,z) \geq 0, \ g(x,z) < 0 \} = \{ z \in \mathbb{R}^{s} : y \geq x(z) > x \}. \end{aligned}$$

Hence, we can continue our representation of the difference (4.98) as follows:

$$G(x) - G(y) = -\int_{\{z \in \mathbb{R}^s : y \ge x(z) > x\}} \theta(z) dz.$$

Now, we apply Schwarz [232, Vol. 1, Theorem 108] and obtain

$$G(x) - G(y) = -\int_{x}^{y} \int_{\{z \in \mathbb{R}^{s} : x(z) = t\}} \frac{\theta(z)}{\|\nabla_{z} x(z)\|} dS dt$$
$$= \int_{y}^{x} \int_{\mathrm{bd}H(x)^{r}} \frac{|\nabla_{x} g(t, z)|\theta(z)}{\|\nabla_{z} g(x, z)\|} dS dt.$$

By Fubini's theorem [232, Vol. 1, Theorem 77], the inner integral converges almost everywhere with respect to the Lebesgue measure. Therefore, we can apply Schwarz [232, Vol. 1, Theorem 90] to conclude that the difference G(x) - G(y) is differentiable almost everywhere with respect to $x \in U$ and we have:

155

$$\frac{\partial}{\partial x}G(x) = \int\limits_{\mathrm{bd}H^r(x)} \frac{\nabla_x g(x,z)\theta(z)}{\|\nabla_z g(x,z)\|} dS.$$

We have used assumption (ii) to set $|\nabla_x g(x,z)| = \nabla_x g(x,z)$. \Box

Obviously, the statement remains valid, if the assumption (ii) is replaced by the opposite strict inequality, so that the function g(x, z) would be strictly decreasing on $U \times \triangle H_U^r$.

We note, that this result does not imply the differentiability of the function G at any fixed point $x_0 \in U$. However, this type of differentiability is sufficient for many applications as it is elaborated in Ermoliev [76] and Usyasev [261].

The conditions of this theorem can be slightly modified, so that the result and the formula for the derivative are valid for piece-wise smooth function.

Theorem 4.89 (Raik [202]). Given a bounded open set $U \subset \mathbb{R}^n$, we assume that:

- (i) the density function $\theta(\cdot)$ is continuous and bounded on the set ΔH_i for each $i = 1, \ldots, m$;
- (ii) the vector functions $\nabla_z g_i(x, z)$ and $\nabla_x g_i(x, z)$ are continuous and bounded on the set $U \times \triangle H_i$ for each i = 1, ..., m;
- (iii) the function $\|\nabla_x g_i(x, z)\| \ge \delta > 0$ on the set $U \times \triangle H_i$ for each $i = 1, \dots, m$;
- (iv) the following conditions are satisfied for all i = 1, ..., m and all $x \in U$:

$$P_{m-1}\{S_i(x) \cap S_j(x)\} = 0, \quad i \neq j, \quad P_{m-1}\{\mathrm{bd}(\mathrm{supp}P_Z \cap S_i(x))\} = 0.$$

Then the probability function G(x) is differentiable on U and

$$\nabla G(x) = \sum_{i=1}^{m} \int_{S_i(x)} \frac{\theta(z)}{\|\nabla_z g_i(x,z)\|} \nabla_x g_i(x,z) dS.$$
(4.99)

The precise proof of this theorem is omitted. We refer to Kibzun and Tretyakov [126], and Kibzun and Uryasev [127] for more information on this topic.

For example, if $g(x, Z) = x^{\mathsf{T}}Z$, m = 1, and Z has a nondegenerate multivariate normal distribution $\mathcal{N}(\bar{z}, \Sigma)$, then $g(x, Z) \sim \mathcal{N}(x^{\mathsf{T}}\bar{z}, x^{\mathsf{T}}\Sigma x)$, and hence the probability function $G(x) = \mathsf{Pr}\{g(x, Z) \ge 0\}$ can be written in the form

$$G(x) = \Phi\left(\frac{x^{\mathsf{T}}\bar{z}}{\sqrt{x^{\mathsf{T}}\Sigma x}}\right),$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution. In this case G(x) is continuously differentiable at every $x \neq 0$.

For problem (4.95), we impose the following constraint qualification at a point $\hat{x} \in \mathcal{X}$. There exists a point $x^r \in \mathcal{X}$ such that:

$$\sum_{i=1}^{m} \int_{S_{i}(x)} \frac{\theta(z)}{\|\nabla_{z}g_{i}(\hat{x}, z)\|} (x^{r} - \hat{x})^{\mathsf{T}} \nabla_{x}g_{i}(\hat{x}, z) dS < 0.$$
(4.100)

This condition implies Robinson's condition. We obtain the following necessary optimality conditions.

Theorem 4.90. Under the assumption of Theorem 4.89, let the constraint qualification (4.100) be satisfied, the function $c(\cdot)$ be continuously differentiable, and let $\hat{x} \in \mathcal{X}$ be an optimal solution of problem (4.95). Then there is a multiplier $\lambda \geq 0$ such that:

$$0 \in \nabla c(\hat{x}) - \lambda \sum_{i=1}^{m} \int_{S_i(x)} \frac{\theta(z)}{\|\nabla_z g_i(x, z)\|} \nabla_x g_i(x, z) dS + \mathcal{N}_{\mathcal{X}}(\hat{x}),$$
(4.101)

$$\lambda \big[G(\hat{x}) - p \big] = 0. \tag{4.102}$$

Proof. The statement follows from the necessary optimality conditions for smooth optimization problems and formula (4.99).

4.4.2 Approximations of non-separable probabilistic constraints

Smoothing approximation via Steklov transformation

In order to apply the optimality conditions formulated in Theorem 4.87 we need to calculate the subdifferential of the probability function \overline{G} defined by the formula (4.96). The calculation involves the subdifferential of the probability function and the characteristic function of the event

$$\{g_i(x,z) \ge 0, i = 1,\ldots,m\}.$$

The latter function may be discontinuous. To alleviate this difficulty, we shall approximate the function G(x) by smooth functions.

Let $k : \mathbb{R} \to \mathbb{R}$ be a nonnegative integrable symmetric function such that

$$\int_{-\infty}^{+\infty} k(t)dt = 1.$$

It can be used as a density function of a random variable K, and, thus,

$$F_K(\tau) = \int_{-\infty}^{\tau} k(t) dt.$$

Taking the characteristic function of the interval $[0, \infty)$, we consider the Steklov-Sobolev average functions for $\varepsilon > 0$:

$$F_K^{\varepsilon}(\tau) = \int_{-\infty}^{+\infty} \mathbf{1}_{[0,\infty)}(\tau + \varepsilon t)k(t)dt = \frac{1}{\varepsilon} \int_{-\infty}^{+\infty} \mathbf{1}_{[0,\infty)}(t)k\left(\frac{t-\tau}{\varepsilon}\right)dt.$$
(4.103)

We see that by the definition of F_K^{ε} and $\mathbf{1}_{[0,\infty)}$ and by the symmetry of $k(\cdot)$ we have:

$$F_{K}^{\varepsilon}(\tau) = \int_{-\infty}^{+\infty} \mathbf{1}_{[0,\infty)}(\tau + \varepsilon t)k(t)dt = \int_{-\tau/\varepsilon}^{+\infty} k(t)dt$$
$$= \int_{-\infty}^{\tau/\varepsilon} k(-t)dt = \int_{-\infty}^{\tau/\varepsilon} k(t)dt$$
$$= F_{K}\left(\frac{\tau}{\varepsilon}\right).$$
(4.104)

Setting

$$g_M(x,z) = \min_{1 \le i \le m} g_i(x,z),$$

we note that g_M is quasi-concave, provided that all g_i are quasi-concave functions. If the functions $g_i(\cdot, z)$ are continuous, then $g_M(\cdot, z)$ is continuous as well.

Using (4.104), we can approximate the constraint function $G(\cdot)$ by the function:

$$G_{\varepsilon}(x) = \int_{\mathbb{R}^{s}} F_{K}^{\varepsilon} \left(g_{M}(x,z) - c \right) dP_{z}$$

$$= \int_{\mathbb{R}^{s}} F_{K} \left(\frac{g_{M}(x,z) - c}{\varepsilon} \right) dP_{z}$$

$$= \frac{1}{\varepsilon} \int_{\mathbb{R}^{s}} \int_{-\infty}^{-c} k \left(\frac{t + g_{M}(x,z)}{\varepsilon} \right) dt dP_{z}.$$
 (4.105)

Now, we show that the functions $G_{\varepsilon}(\cdot)$ uniformly converge to $G(\cdot)$ when ε converges to zero.

Theorem 4.91. Assume that Z has a continuous distribution, the functions $g_i(\cdot, z)$ are continuous for almost all $z \in \mathbb{R}^s$ and that, for certain constant $c \in \mathbb{R}$, we have

$$\mathsf{Pr}\{z \in \mathbb{R}^s : g_M(x,z) = c\} = 0.$$

Then for any compact set $\mathbf{C} \subset \mathcal{X}$ the functions G_{ε} uniformly converge on \mathbf{C} to G when $\varepsilon \to 0$, i.e.,

$$\lim_{\varepsilon \downarrow 0} \max_{x \in \mathbf{C}} \left| G_{\varepsilon}(x) - G(x) \right| = 0.$$

Proof. Defining $\delta(\varepsilon) = \varepsilon^{1-\beta}$ with $\beta \in (0,1)$, we have

$$\lim_{\varepsilon \to 0} \delta(\varepsilon) = 0 \quad \text{and} \quad \lim_{\varepsilon \to 0} F_K\left(\frac{\delta(\varepsilon)}{\varepsilon}\right) = 1, \ \lim_{\varepsilon \to 0} F_K\left(\frac{-\delta(\varepsilon)}{\varepsilon}\right) = 0. \tag{4.106}$$

Define for any $\delta > 0$ the sets:

$$A(x,\delta) = \{z \in \mathbb{R}^s : g_M(x,z) - c \le -\delta\},\ B(x,\delta) = \{z \in \mathbb{R}^s : g_M(x,z) - c \ge \delta\},\ C(x,\delta) = \{z \in \mathbb{R}^s : |g_M(x,z) - c| \le \delta\}.$$

On the set $A(x, \delta(\varepsilon))$ we have $\mathbf{1}_{[0,\infty)}(g_M(x,z)-c)=0$ and, using (4.104), we obtain

$$F_K^{\varepsilon}(g_M(x,z)-c) = F_K\left(\frac{g_M(x,z)-c}{\varepsilon}\right) \le F_K\left(\frac{-\delta(\varepsilon)}{\varepsilon}\right).$$

On the set $B(x, \delta(\varepsilon))$ we have $\mathbf{1}_{[0,\infty)}(g_M(x, z) - c) = 1$ and

$$F_K^{\varepsilon}(g_M(x,z)-c) = F_K\left(\frac{g_M(x,z)-c}{\varepsilon}\right) \ge F_K\left(\frac{\delta(\varepsilon)}{\varepsilon}\right).$$

On the set $C(\delta(\varepsilon))$ we use the fact that $0 \leq \mathbf{1}_{[0,\infty)}(t) \leq 1$ and $0 \leq F_K(t) \leq 1$. We obtain the following estimate:

$$\begin{aligned} \left| G(x) - G_{\varepsilon}(x) \right| &\leq \int_{\mathbb{R}^{s}} \left| \mathbf{1}_{[0,\infty)} \left(g_{M}(x,z) - c \right) - F_{K}^{\varepsilon} \left(g_{M}(x,z) - c \right) \right| dP_{z} \\ &\leq F_{K} \left(\frac{-\delta(\varepsilon)}{\varepsilon} \right) \int_{A(x,\delta(\varepsilon))} dP_{Z} + \left(1 - F_{K} \left(\frac{\delta(\varepsilon)}{\varepsilon} \right) \right) \int_{B(x,\delta(\varepsilon))} dP_{Z} + 2 \int_{C(x,\delta(\varepsilon))} dP_{Z} \\ &\leq F_{K} \left(\frac{-\delta(\varepsilon)}{\varepsilon} \right) + \left(1 - F_{K} \left(\frac{\delta(\varepsilon)}{\varepsilon} \right) \right) + 2P_{Z}(C(x,\delta(\varepsilon))). \end{aligned}$$

The first two terms on the right hand side of the inequality converge to zero when $\varepsilon \to 0$ by the virtue of (4.106). It remains to show that $\lim_{\varepsilon \to 0} P_Z\{C(x, \delta(\varepsilon))\} = 0$ uniformly with respect to $x \in \mathbf{C}$. The function $(x, z, \delta) \mapsto |g_M(x, z) - c| - \delta$ is continuous in (x, δ) and measurable in z. Thus, it is uniformly continuous with respect to (x, δ) on any compact set $\mathbf{C} \times [-\delta_0, \delta_0]$ with $\delta_0 > 0$. The probability measure P_Z is continuous, and, therefore, the function

$$\Theta(x,\delta) = P_Z\{|g_M(x,z) - c| - \delta \le 0\} = P_Z\{\cap_{\beta > \delta} C(x,\beta)\}$$

is uniformly continuous with respect to (x, δ) on $\mathbb{C} \times [-\delta_0, \delta_0]$. By the assumptions of the theorem

$$\Theta(x,0) = P_Z\{z \in \mathbb{R}^s : |g_M(x,z) - c| = 0\} = 0,$$

and, thus,

$$\lim_{\varepsilon \to 0} P_Z \{ z \in \mathbb{R}^s : |g_M(x, z) - c| \le \delta(\varepsilon) \} = \lim_{\delta \to 0} \Theta(x, \delta) = 0.$$

As $\Theta(\cdot, \delta)$ is continuous, the convergence is uniform on compact sets with respect to the first argument. \Box

Now, we derive a formula for the Clarke generalized gradients of the approximation G_{ε} . We define the index set:

$$I(x, z) = \{i : g_i(x, z) = g_M(x, z), 1 \le i \le m\}.$$

Theorem 4.92. Assume that the density function $k(\cdot)$ is nonnegative, bounded, and continuous. Furthermore, let the functions $g_i(\cdot, z)$ be concave for every $z \in \mathbb{R}^s$, and their subgradients be uniformly bounded as follows:

$$\sup\{s \in \partial g_i(y, z), \|y - x\| \le \delta\} \le l_{\delta}(x, z), \, \delta > 0, \, \text{for all } i = 1, \dots, m,$$

where $l_{\delta}(x, z)$ is an integrable function of z for all $x \in \mathcal{X}$. Then $G_{\varepsilon}(\cdot)$ is Lipschitz continuous, Clarke regular and its Clarke generalized gradient set is given by

$$\partial^{\circ}G_{\varepsilon}(x) = \frac{1}{\varepsilon} \int_{\mathbb{R}^{s}} k\left(\frac{g_{M}(x,z) - c}{\varepsilon}\right) \operatorname{conv}\left\{\partial g_{i}(x,z) : i \in I(x,z)\right\} \, dP_{Z}.$$

Proof. Under the assumptions of the theorem, the function $F_K(\cdot)$ is monotonic and continuously differentiable. The function $g_M(\cdot, z)$ is concave for every $z \in \mathbb{R}^s$ and its subdifferential are given by the formula:

$$\partial g_M(y,z) = \operatorname{conv}\{s_i \in \partial g_i(y,z) : g_i(y,z) = g_M(y,z)\}.$$

Thus the subgradients of g_M are uniformly bounded:

$$\sup\{s \in \partial g_M(y, z), \|y - x\| \le \delta\} \le l_{\delta}(x, z), \, \delta > 0$$

Therefore, the composite function $F_K(\frac{g_M(x,z)-c}{\varepsilon})$ is subdifferentiable and its subdifferential can be calculated as:

$$\partial^{\circ} F_K\left(\frac{g_M(x,z)-c}{\varepsilon}\right) = \frac{1}{\varepsilon} k\left(\frac{g_M(x,z)-c}{\varepsilon}\right) \cdot \partial g_M(x,z).$$

The mathematical expectation function

$$G_{\varepsilon}(x) = \int_{\mathbb{R}^s} F_K^{\varepsilon}(g_M(x,z) - c) dP_z = \int_{\mathbb{R}^s} F_K\left(\frac{g_M(x,z) - c}{\varepsilon}\right) dP_z$$

is regular by Clarke [45, Theorem 2.7.2] and its Clarke generalized gradient set has the form:

$$\partial^{\circ}G_{\varepsilon}(x) = \int_{\mathbb{R}^{s}} \partial^{\circ}F_{K}\left(\frac{g_{M}(x,z) - c}{\varepsilon}\right) dP_{Z} = \frac{1}{\varepsilon} \int_{\mathbb{R}^{s}} k\left(\frac{g_{M}(x,z) - c}{\varepsilon}\right) \cdot \partial g_{M}(x,z) dP_{Z}$$

Using the formula for the subdifferential of $g_M(x, z)$, we obtain the statement. \Box

Now we show that if we choose K to have an α -concave distribution, and all assumptions of Theorem 4.39 are satisfied, the generalized concavity property of the approximated probability function is preserved.

Theorem 4.93. If the density function k is α -concave ($\alpha \ge 0$), Z has γ -concave distribution ($\gamma \ge 0$), the functions $g_i(\cdot, z)$, i = 1, ..., m are quasi-concave, then the approximate probability function G_{ε} has a β -concave distribution, where

$$\beta = \begin{cases} (\gamma^{-1} + (1 + s\alpha)/\alpha)^{-1}, & \text{if } \alpha + \gamma > 0, \\ 0, & \text{if } \alpha + \gamma = 0. \end{cases}$$

Proof. If the density function k is α -concave ($\alpha \ge 0$), then K has a γ - concave distribution with $\gamma = \alpha/(1 + s\alpha)$. If Z has γ' -concave distribution ($\gamma \ge 0$), then the random vector $(Z, K)^{\mathsf{T}}$ has a β -concave distribution according to Theorem 4.36, where

$$\beta = \begin{cases} (\gamma^{-1} + \gamma'^{-1})^{-1}, & \text{if } \gamma + \gamma' > 0, \\ 0, & \text{if } \gamma + \gamma' = 0. \end{cases}$$

Using the definition $G_{\varepsilon}(x)$ of (4.105), we can write:

$$G_{\varepsilon}(x) = \int_{\mathbb{R}^{s}} F_{K}^{\varepsilon} \left(\frac{g_{M}(x,z) - c}{\varepsilon} \right) dP_{Z} = \int_{\mathbb{R}^{s}} \int_{-\infty}^{(g_{M}(x,z) - c)/\varepsilon} k(t) dt \, dP_{Z}$$
$$= \int_{\mathbb{R}^{s}} \int_{-\infty}^{\infty} \mathbf{1}_{\{(g_{M}(x,z) - c)/\varepsilon > t\}} dP_{K} \, dP_{Z} = \int_{\mathbb{R}^{s}} \int_{H_{\varepsilon}(x)} dP_{K} \, dP_{Z}, \qquad (4.107)$$

where

$$H_{\varepsilon}(x) = \{(z,t) \in \mathbb{R}^{s+1} : g_M(x,z) - \varepsilon t \ge c\}.$$

Since $g_M(\cdot, z)$ is quasi-concave, the set $H_{\varepsilon}(x)$ is convex. Representation (4.107) of G_{ε} and the β -concavity of (Z, K) imply the assumptions of Theorem 4.39, and, thus, the function G_{ε} is β -concave. \Box

This theorem shows that if the random vector Z has a generalized concave distribution, we can choose a suitable generalized concave density function $k(\cdot)$ for smoothing and obtain an approximate convex optimization problem.

Theorem 4.94. In addition to the assumptions of Theorems 4.91, 4.92 and 4.93. Then on the set $\{x \in \mathbb{R}^n : G(x) > 0\}$, the function G_{ε} is Clarke-regular and the set of Clarke generalized gradients $\partial^{\circ}G_{\varepsilon}(x^{\varepsilon})$ converge to the set of Clarke generalized gradients of G, $\partial^{\circ}G(x)$, in the following sense: if for any sequences $\varepsilon \downarrow 0$, $x^{\varepsilon} \to x$ and $s^{\varepsilon} \in \partial^{\circ}G_{\varepsilon}(x^{\varepsilon})$ such that $s^{\varepsilon} \to s$, then $s \in \partial^{\circ}G(x)$.

Proof. Consider a point x such that G(x) > 0 and points $x^{\varepsilon} \to x$ as $\varepsilon \downarrow 0$. All points x^{ε} can be included in some compact set containing x in its interior. The function G is generalized concave by virtue of Theorem 4.39. It is locally Lipschitz continuous, directionally differentiable, and Clarke-regular due to Theorem 4.29. It follows that G(y) > 0 for all point y in some neighborhood of x. By virtue of Theorem 4.91, this neighborhood can be chosen small enough, so that $G_{\varepsilon}(y) > 0$ for all ε small enough, as well. The functions G_{ε} are generalized concave by virtue of Theorem 4.93. It follows that G_{ε} are locally Lipschitz continuous, directionally differentiable, and Clarke-regular due to Theorem 4.93. It follows that G_{ε} are locally Lipschitz continuous, directionally differentiable, and Clarke-regular due to Theorem 4.29. Using the uniform convergence of G_{ε} on compact sets and the definition of Clarke generalized gradient, we can pass to the limit with $\varepsilon \downarrow 0$ in the inequality

$$\lim_{t \downarrow 0, y \to x^{\varepsilon}} \frac{1}{t} \left[G_{\varepsilon}(y + td) - G_{\varepsilon}(y) \right] \ge d^{\mathsf{T}} s^{\varepsilon} \quad \text{for any } d \in \mathbb{R}^{n}.$$

Consequently, $s \in \partial^{\circ} G(x)$. \Box

Using the approximate probability function we can solve the following approximation of problem (4.95):

$$\begin{split} & \underset{x}{\min} \quad c(x) \\ & \text{s.t.} \quad G_{\varepsilon}(x) \geq p, \\ & \quad x \in \mathcal{X}. \end{split} \tag{4.108}$$

Under the conditions of Theorem 4.94 the function G_{ε} is β -concave for some $\beta \ge 0$. We can specify the necessary and sufficient optimality conditions for the approximate problem.

Theorem 4.95. In addition to the assumptions of Theorem 4.94, assume that $c(\cdot)$ is a convex function, the Slater condition for problem (4.108) is satisfied, and $\operatorname{int} G_{\varepsilon} \neq \emptyset$. A point $\hat{x} \in \mathcal{X}$ is an optimal solution of problem (4.108) if and only if a nonpositive number λ exists such that

$$0 \in \partial c(\hat{x}) + s\lambda \int_{\mathbb{R}^s} k\left(\frac{g_M(\hat{x}, z) - c}{\varepsilon}\right) \operatorname{conv}\left\{\partial g_i(\hat{x}, z) : i \in I(\hat{x}, z)\right\} dP_Z + \mathcal{N}_{\mathcal{X}}(\hat{x}),$$
$$\lambda[G_{\varepsilon}(\hat{x}) - p] = 0.$$

Here

$$s = \begin{cases} \alpha \varepsilon^{-1} \left[G_{\varepsilon}(\hat{x}) \right]^{\alpha - 1}, & \text{if } \beta \neq 0, \\ \left[\varepsilon G_{\varepsilon}(\hat{x}) \right]^{-1}. & \text{if } \beta = 0. \end{cases}$$

Proof. We shall show the statement for $\beta = 0$. The proof for the other case is analogous. Setting $\bar{G}_{\varepsilon}(x) = \ln G_{\varepsilon}(x)$, we obtain a concave function \bar{G}_{ε} , and formulate the problem:

$$\begin{aligned}
& \underset{x}{\min} c(x) \\
& \text{s.t. } \ln p - \bar{G}_{\varepsilon}(x) \leq 0, \\
& x \in \mathcal{X}.
\end{aligned}$$
(4.109)

Clearly, \hat{x} is a solution of the problem (4.109) if and only if it is a solution of problem (4.108). Problem (4.109) is a convex problem and Slater's condition is satisfied for it as well. Therefore, we can write the following optimality conditions for it. The point $\hat{x} \in \mathcal{X}$ is a solution if and only if a number $\lambda_0 > 0$ exists such that

$$0 \in \partial c(x) + \lambda_0 \partial \left[-\bar{G}_{\varepsilon}(\hat{x}) \right] + \mathcal{N}_{\mathcal{X}}(\hat{x}), \tag{4.110}$$

$$\lambda_0[G_\varepsilon(\hat{x}) - p] = 0. \tag{4.111}$$

We use the formula for the Clarke generalized gradients of generalized concave functions to obtain

$$\partial^{\circ} \bar{G}_{\varepsilon}(\hat{x}) = \frac{1}{G_{\varepsilon}(\hat{x})} \partial^{\circ} G_{\varepsilon}(\hat{x}).$$

Moreover, we have a representation of the Clarke generalized gradient set of G_{ε} , which yields

$$\partial^{\circ} \bar{G}_{\varepsilon}(\hat{x}) = \frac{1}{\varepsilon G_{\varepsilon}(\hat{x})} \int_{\mathbb{R}^{s}} k\left(\frac{g_{M}(\hat{x}, z) - c}{\varepsilon}\right) \cdot \partial g_{M}(\hat{x}, z) dP_{Z}.$$

Substituting the last expression into (4.110), we obtain the result. \Box

Normal approximation

In this section we analyze approximation for problems with individual probabilistic constraints, defined by linear inequalities. In this setting it is sufficient to consider a problem with a single probabilistic constraint of form:

Max
$$c(x)$$

s.t. $\Pr\{x^{\mathsf{T}}Z \ge \eta\} \ge p$, (4.112)
 $x \in \mathcal{X}$.

Before developing the normal approximation for this problem, let us illustrate its potential on an example. We return to our Example 4.2, in which we have formulated a portfolio optimization problem under a Value-at-Risk constraint.

$$\operatorname{Max} \sum_{i=1}^{n} \mathbb{E}[R_i] x_i$$

s.t. $\operatorname{Pr}\left\{\sum_{i=1}^{n} R_i x_i \ge -\eta\right\} \ge p,$
$$\sum_{i=1}^{n} x_i \le 1,$$

$$x \ge 0.$$

(4.113)

We denote the net increase of the value of our investment after a period of time by

$$G(x,R) = \sum_{i=1}^{n} \mathbb{E}[R_i]x_i.$$

Let us assume that the random return rates R_1, \ldots, R_n have a joint normal probability distribution. Recall that the normal distribution is log-concave and the probabilistic constraint in problem (4.113) determines a convex feasible set, according to Theorem 4.39.

Another direct way to see that the last transformation of the probabilistic constraint results in a convex constraint is the following. We denote $\bar{r}_i = \mathbb{E}[R_i]$, $\bar{r} = (\bar{r}_1, \dots, \bar{r}_n)^{\mathsf{T}}$, and assume that \bar{r} is not the zero-vector. Further, let Σ be the covariance matrix of the joint distribution of the return rates. We observe that the total profit (or loss) G(x, R) is a normally distributed random variable with expected value $\mathbb{E}[G(x, R)] = \bar{r}^{\mathsf{T}}x$ and variance $\mathbb{Var}[G(x, R)] = x^{\mathsf{T}}\Sigma x$. Assuming that Σ is positive definite, the probabilistic constraint

$$\mathsf{Pr}\{G(x,R) \ge -\eta\} \ge p$$

can be written in the form (see the discussion on page 16)

$$z_p \sqrt{x^\mathsf{T} \Sigma x} - \bar{r}^\mathsf{T} x \le \eta.$$

Hence problem (4.113) can be written in the following form:

$$\text{Max } \bar{r}^{\mathsf{T}} x \\ \text{s.t. } z_p \sqrt{x^{\mathsf{T}} \Sigma x} - \bar{r}^{\mathsf{T}} x \leq \eta, \\ \sum_{i=1}^n x_i \leq 1, \\ x \geq 0.$$
 (4.114)

163

Note that $\sqrt{x^{\mathsf{T}}\Sigma x}$ is a convex function, of x, and $z_p = \Phi^{-1}(p)$ is positive for p > 1/2, and hence (4.114) is a convex programming problem.

Now, we consider the general optimization problem (4.112). Assuming that the *n*-dimensional random vector Z has independent components and the dimension n is relatively large, we may invoke the central limit theorem. Under mild additional assumptions, we can conclude that the distribution of $x^T Z$ is approximately normal and convert the probabilistic constraint into an algebraic constraint in a similar manner. Note, that this approach is appropriate if Z has a substantial number of components and the vector x has appropriately large number of non-zero components, so that the central limit theorem would be applicable to $x^T Z$. Furthermore, we assume that the probability parameter p is not too close to one, such as 0.9999.

We recall several versions of the Central Limit Theorem (CLT). Let Z_i , i = 1, 2, ... be a sequence of independent random variables defined on the same probability space. We assume that each Z_i has finite expected value $\mu_i = \mathbb{E}[Z_i]$ and finite variance $\sigma_i^2 = \mathbb{V}ar[Z_i]$. Setting

$$s_n^2 = \sum_{i=1}^n \sigma_i^2$$
, and $r_n^3 = \sum_{i=1}^n \mathbb{E}(|Z_i - \mu_i|^3)$,

we assume that r_n^3 is finite for every n, and that

$$\lim_{n \to \infty} \frac{r_n}{s_n} = 0. \tag{4.115}$$

Then, the distribution of the random variable

$$\frac{\sum_{i=1}^{n} (Z_i - \mu_i)}{s_n}$$
(4.116)

converges towards the standard normal distribution as $n \to \infty$.

The condition (4.115) is called Lyapunov's condition. In the same setting, we can replace the Lyapunov's condition with the following weaker condition, proposed by Lindeberg. For every $\varepsilon > 0$ we define

$$Y_i = \begin{cases} (X_i - \mu_i)^2 / s_n^2, & \text{if } |X_i - \mu_i| > \varepsilon s_n, \\ 0, & \text{otherwise.} \end{cases}$$

The Lindeberg's condition reads:

$$\lim_{n \to \infty} \sum_{i=1}^{n} \mathbb{E}(Y_i) = 0.$$

Let us denote $\bar{z} = (\mu_1, \dots, \mu_n)^T$. Under the conditions of the Central Limit Theorem, the distribution of our random variable $x^T Z$ is close to the normal distribution with expected value $x^T \bar{z}$ and variance $\sum_{i=1}^n \sigma_i^2 x_i^2$ for problems of large dimensions. Our probabilistic constraint takes on the form:

$$\frac{\bar{z}^{\mathsf{T}}x - \eta}{\sqrt{\sum_{i=1}^{n} \sigma_i^2 x_i^2}} \ge z_p.$$

Define $\mathcal{X} = \left\{ x \in \mathbb{R}^n_+ : \sum_{i=1}^n x_i \leq 1 \right\}$. Denoting the matrix with diagonal elements σ_1 , ..., σ_n by D, problem (4.112) can be replaced by the following approximate problem:

$$\begin{split} \min_{x} & c(x) \\ \text{s.t.} & z_p \|Dx\| \leq \bar{z}^\mathsf{T} x - \eta, \\ & x \in \mathcal{X}. \end{split}$$

The probabilistic constraint in this problem is approximated by an algebraic convex constraint. Due to the independence of the components of the random vector Z, the matrix Dhas a simple diagonal form. There are versions of the central limit theorem which treat the case of sums of dependent variables, for instance the *n*-dependent CLT, the martingale CLT and the CLT for mixing processes. These statements will not be presented here. One can follow the same line of argument to formulate a normal approximation of the probabilistic constraint, which is very accurate for problems with large decision space.

4.5 Semi-infinite probabilistic problems

In this section, we concentrate on the semi-infinite probabilistic problem (4.9). We recall its formulation:

$$\begin{split} & \underset{x}{\operatorname{Min}} c(x) \\ & \text{s.t.} \ \mathsf{Pr}\big\{g(x, Z) \geq \eta\big\} \geq \mathsf{Pr}\big\{Y \geq \eta\big\}, \quad \eta \in [a, b], \\ & x \in \mathcal{X}. \end{split}$$

Our goal is to derive necessary and sufficient optimality conditions for this problem. Denote the space of regular countably additive measures on [a, b] having finite variation by $\mathcal{M}([a, b])$, and its subset of nonnegative measures by $\mathcal{M}_+([a, b])$.

We define the constraint function $G(x, \eta) = \Pr\{z : g(x, z) \ge \eta\}$. As we shall develop optimality conditions in differential form, we impose additional assumptions on problem (4.9):

- (i) The function c is continuously differentiable on \mathcal{X} ;
- (ii) The constraint function $G(\cdot, \cdot)$ is continuous with respect to the second argument and continuously differentiable with respect to the first argument;
- (iii) The reference random variable Y has a continuous distribution.

The differentiability assumption on G may be enforced taking into account the results in section 4.4.1. For example, if the vector Z has a probability density $\theta(\cdot)$, the function $g(\cdot, \cdot)$ is continuously differentiable with nonzero gradient $\nabla_z g(x, z)$ and such that the quantity $\frac{\theta(z)}{\|\nabla_z g(x,z)\|} \nabla_x g(x,z)$ is uniformly bounded (in a neighborhood of x) by an integrable function, then the function G is differentiable. Moreover, we can express its gradient with respect to x a follows:

$$\nabla_x G(x,\eta) = \int_{\text{bd } H(z,\eta)} \frac{\theta(z)}{\|\nabla_z g(x,z)\|} \nabla_x g(x,z) \, dP_{m-1},$$

where $\operatorname{bd} H(z,\eta)$ is the surface of the set $H(z,\eta) = \{z : g(x,z) \ge \eta\}$ and P_{m-1} refers to Lebesgue measure on the m-1-dimensional surface.

We define the set $\mathcal{U}([a, b])$ of functions $u(\cdot)$ satisfying the following conditions:

 $u(\cdot)$ is nondecreasing and right continuous; u(t) = 0 for all $t \le a$; u(t) = u(b), for all $t \ge b$.

It is evident that $\mathcal{U}([a, b])$ is a convex cone.

First we derive a useful formula.

Lemma 4.96. For any real random variable Z and any measure $\mu \in \mathcal{M}([a, b])$ we have

$$\int_{a}^{b} \Pr\{Z \ge \eta\} d\mu(\eta) = \mathbb{E}[u(Z)], \qquad (4.117)$$

where $u(z) = \mu([a, z])$.

Proof. We extend the measure μ to the entire real line, by assigning measure 0 to sets not intersecting [a, b]. Using the probability measure P_Z induced by Z on \mathbb{R} and applying Fubini's theorem we obtain

$$\begin{split} \int_{a}^{b} \Pr\{Z \ge \eta\} \, d\mu(\eta) &= \int_{a}^{\infty} \Pr\{Z \ge \eta\} \, d\mu(\eta) = \int_{a}^{\infty} \int_{\eta}^{\infty} dP_{Z}(z) \, d\mu(\eta) \\ &= \int_{a}^{\infty} \int_{a}^{z} \, d\mu(\eta) \, dP_{Z}(z) = \int_{a}^{\infty} \mu([a, z]) \, dP_{Z}(z) = \mathbb{E}\big[\mu([a, Z])\big]. \end{split}$$

We define $u(z) = \mu([a, z])$ and obtain the stated result. \Box

Let us observe that if the measure μ in the above lemma is nonnegative, then $u \in \mathcal{U}([a, b])$. Indeed, $u(\cdot)$ is nondecreasing since for $z_1 > z_2$ we have

$$u(z_1) = \mu([a, z_1]) = \mu([a, z_2]) + \mu((z_1, z_2]) \ge \mu([a, z_2]) = u(z_2).$$

Furthermore, $u(z) = \mu([a, z]) = \mu([a, b]) = u(b)$ for $z \ge b$.

We introduce the functional $L : \mathbb{R}^n \times \mathcal{U} \to \mathbb{R}$ associated with problem (4.9):

$$L(x,u) = c(x) + \mathbb{E}\left[u(g(x,Z)) - u(Y)\right].$$

We shall see that the functional L plays the role of a Lagrangian of the problem.

We also set $v(\eta) = \Pr\{Y \ge \eta\}.$

Definition 4.97. *Problem* (4.9) *satisfies the* differential uniform dominance condition *at the point* $\hat{x} \in \mathcal{X}$ *if there exists* $x^0 \in \mathcal{X}$ *such that*

$$\min_{a \le \eta \le b} \left[G(\hat{x}, \eta) + \nabla_x G(\hat{x}, \eta) (x^0 - \hat{x}) - v(\eta) \right] > 0$$

Theorem 4.98. Assume that \hat{x} is an optimal solution of problem (4.9) and that the differential uniform dominance condition is satisfied at the point \hat{x} . Then there exists a function $\hat{u} \in \mathcal{U}$, such that

$$-\nabla_x L(\hat{x}, \hat{u}) \in \mathcal{N}_{\mathcal{X}}(\hat{x}), \tag{4.118}$$

$$\mathbb{E}[\hat{u}(g(\hat{x}, Z))] = \mathbb{E}[\hat{u}(Y)]. \tag{4.119}$$

Proof. We consider the mapping $\Gamma : \mathcal{X} \to \mathcal{C}([a, b])$ defined as follows:

$$\Gamma(x)(\eta) = \Pr\{g(x, Z) \ge \eta\} - v(\eta), \quad \eta \in [a, b].$$
(4.120)

We define K as the cone of nonnegative functions in C([a, b]). Problem (4.9) can be formulated as follows:

$$\begin{array}{l} \underset{x}{\operatorname{Min}} & c(x) \\ \text{s.t.} & \Gamma(x) \in K, \\ & x \in \mathcal{X}. \end{array}$$
(4.121)

At first we observe that the functions $c(\cdot)$ and $\Gamma(\cdot)$ are continuously differentiable by the assumptions made at the beginning of this section. Secondly, the differential uniform dominance condition is equivalent to Robinson's constraint qualification condition:

$$0 \in \inf \left\{ \Gamma(\hat{x}) + \nabla_x \Gamma(\hat{x}) (\mathcal{X} - \hat{x}) - K \right\}.$$
(4.122)

Indeed, it is easy to see that the uniform dominance condition implies Robinson's condition. On the other hand, if Robinson's condition holds, then there exists $\varepsilon > 0$ such that the function identically equal to ε is an element of the set at the right hand side of (4.122). Then we can find x^0 such that

$$\Gamma(\hat{x})(\eta) + \left[\nabla_x \Gamma(\hat{x})(\eta)\right](x^0 - \hat{x}) \ge \varepsilon \quad \text{for all} \quad \eta \in [a, b].$$

Consequently, the uniform dominance condition is satisfied.

By the Riesz representation theorem, the space dual to C([a, b]) is the space $\mathcal{M}([a, b])$ of regular countably additive measures on [a, b] having finite variation. The Lagrangian $\Lambda : \mathbb{R}^n \times \mathcal{M}([a, b]) \to \mathbb{R}$, for problem (4.121) is defined as follows:

$$\Lambda(x,\mu) = c(x) + \int_a^b \Gamma(x)(\eta) \, d\mu(\eta). \tag{4.123}$$

The necessary optimality conditions for problem (4.121) have the form: *there exists a measure* $\hat{\mu} \in \mathcal{M}_+([a, b])$, such that

$$-\nabla_x \Lambda(\hat{x}, \hat{\mu}) \in \mathcal{N}_{\mathcal{X}}(\hat{x}), \tag{4.124}$$

$$\int_{a}^{b} \Gamma(\hat{x})(\eta) \, d\hat{\mu}(\eta) = 0. \tag{4.125}$$

Using Lemma 4.96, we obtain the equation for all x:

$$\begin{split} \int_{a}^{b} \Gamma(x)(\eta) \, d\hat{\mu}(\eta) &= \int_{a}^{b} \left(\mathsf{Pr}\big\{g(x,z) \geq \eta\big\} - \mathsf{Pr}\big\{Y \geq \eta\big\} \right) d\hat{\mu}(\eta) \\ &= \mathbb{E}\big[\hat{u}(g(x,Z))\big] - \mathbb{E}\big[\hat{u}(Y)\big], \end{split}$$

where $\hat{u}(\eta) = \hat{\mu}([a, \eta])$. Since $\hat{\mu}$ is nonnegative, the corresponding utility function \hat{u} is an element of $\mathcal{U}([a, b])$. The correspondence between nonnegative measures $\mu \in \mathcal{M}([a, b])$ and utility functions $u \in \mathcal{U}$ and the last equation imply that (4.125) is equivalent to (4.119). Moreover,

$$\Lambda(x,\mu) = L(x,u)$$

and, therefore, (4.124) is equivalent to (4.118).

We note that the functions $u \in \mathcal{U}([a, b])$ can be interpreted as von Neumann-Morgenstern utility functions of rational decision makers. The theorem demonstrates that one can view the maximization of expected utility as a dual model to the model with stochastic dominance constraints. Utility functions of decision makers are very difficult to elicit. This task becomes even more complicated when there is a group of decision-makers who have to come to a consensus. The model (4.9) avoids these difficulties, by requiring that a benchmark random outcome, considered reasonable, be specified. Our analysis, departing from the benchmark outcome, generates the utility function of the decision maker. It is implicitly defined by the benchmark used, and by the problem under consideration.

We will demonstrate that it is sufficient to consider only the subset of $\mathcal{U}([a, b]$ containing piecewise constant utility functions.

Theorem 4.99. Under the assumptions of Theorem 4.98 there exist piecewise constant utility function $w(\cdot) \in \mathcal{U}$ satisfying the necessary optimality conditions (4.118)–(4.119). Moreover, the function $w(\cdot)$ has at most n + 2 jump points: there exist numbers $\eta_i \in [a,b], \quad i = 1, \ldots, k$, such that the function $w(\cdot)$ is constant on the intervals $(-\infty, \eta_1], (\eta_1, \eta_2], \ldots, (\eta_k, \infty)$, and $0 \le k \le n + 2$.

Proof. Consider the mapping Γ defined by (4.120). As already noted in the proof of the previous theorem, it is continuously differentiable due to the assumptions about the probability function. Therefore, the derivative of the Lagrangian has the form

$$\nabla_x \Lambda(\hat{x}, \hat{\mu}) = \nabla_x c(\hat{x}) + \int_a^b \nabla_x \Gamma(\hat{x})(\eta) \ d\hat{\mu}(\eta).$$

The necessary condition of optimality (4.124) can be rewritten as follows

$$-\nabla_x c(\hat{x}) - \int_a^b \nabla_x \Gamma(\hat{x})(\eta) \ d\hat{\mu}(\eta) \in \mathcal{N}_{\mathcal{X}}(\hat{x}).$$

Considering the vector

$$g = \nabla_x c(\hat{x}) - \nabla_x \Lambda(\hat{x}, \hat{\mu}),$$

we observe that the optimal values of multipliers $\hat{\mu}$ have to satisfy the equation

$$\int_{a}^{b} \nabla_{x} \Gamma(\hat{x})(\eta) \ d\mu(\eta) = g. \tag{4.126}$$

At the optimal solution \hat{x} we have $\Gamma(\hat{x})(\cdot) \le 0$ and $\hat{\mu} \ge 0$. Therefore, the complementarity condition (4.125) can be equivalently expressed as the equation

$$\int_{a}^{b} \Gamma(\hat{x})(\eta) \ d\mu(\eta) = 0.$$
(4.127)

Every nonnegative solution μ of equations (4.126)–(4.127) can be used as the Lagrange multiplier satisfying conditions (4.124)–(4.125) at \hat{x} . Define

$$a = \int_a^b d\hat{\mu}(\eta).$$

We can add to equations (4.126)–(4.127) the condition

$$\int_{a}^{b} d\mu(\eta) = a. \tag{4.128}$$

The system of three equations (4.126)–(4.128) still has at least one nonnegative solution, namely $\hat{\mu}$. If $\hat{\mu} \equiv 0$, then the dominance constraint is not active. In this case, we can set $w(\eta) \equiv 0$ and the statement of the theorem follows from the fact that conditions (4.126)–(4.127) are equivalent to (4.124)–(4.125).

Now, consider the case of $\hat{\mu} \neq 0$. In this case, we have a > 0. Normalizing by a, we notice that equations (4.126)–(4.128) are equivalent to the following inclusion:

$$\begin{bmatrix} g/a \\ 0 \end{bmatrix} \in \operatorname{conv} \left\{ \begin{bmatrix} \nabla_x \Gamma(\hat{x})(\eta) \\ \Gamma(\hat{x})(\eta) \end{bmatrix} : \eta \in [a, b] \right\} \subset \mathbb{R}^{n+1}.$$

By Carathéodory's Theorem, there exist numbers $\eta_i \in [a, b]$, and $\alpha_i \ge 0, i = 1, ..., k$ such that

$$\begin{bmatrix} g/a \\ 0 \end{bmatrix} = \sum_{i=1}^{k} \alpha_i \begin{bmatrix} \nabla_x \Gamma(\hat{x})(\eta_i) \\ \Gamma(\hat{x})(\eta_i) \end{bmatrix},$$
$$\sum_{i=1}^{k} \alpha_i = 1,$$

and

$$1 \le k \le n+2.$$
We define atomic measure ν having atoms of mass $c\alpha_i$ at points η_i , i = 1, ..., k. It satisfies equations (4.126)–(4.127):

$$\int_{a}^{b} \nabla_{x} \Gamma(\hat{x})(\eta) \, d\nu(\eta) = \sum_{i=1}^{k} \nabla_{x} \Gamma(\hat{x})(\eta_{i}) c\alpha_{i} = g,$$
$$\int_{a}^{b} \Gamma(\hat{x})(\eta) \, d\nu(\eta) = \sum_{i=1}^{k} \Gamma(\hat{x})(\eta_{i}) c\alpha_{i} = 0.$$

Recall that equations (4.126)–(4.127) are equivalent to (4.124)–(4.125). Now, applying Lemma 4.96, we obtain the utility functions

$$w(\eta) = \nu[a,\eta], \quad \eta \in \mathbb{R}.$$

It is straightforward to check that $w \in \mathcal{U}([a, b])$ and the assertion of the theorem holds.

It follows from Theorem 4.99 that if the dominance constraint is active, then there exist at least one and at most n + 2 target values η_i and target probabilities $v_i = \Pr\{Y \ge \eta_i\}, i = 1, ..., k$ which are critical for problem (4.9). They define a relaxation of (4.9) involving finitely many probabilistic constraints:

$$\begin{aligned}
& \underset{x}{\operatorname{Min}} \quad c(x) \\
& \text{s.t. } \operatorname{Pr} \left\{ g(x, Z) \geq \eta_i \right\} \geq v_i, \quad i = 1, \dots, k, \\
& x \in \mathcal{X}.
\end{aligned}$$

The necessary conditions of optimality for this relaxation yield a solution of the optimality conditions of the original problem (4.9). Unfortunately, the target values and the target probabilities are not known in advance.

A particular situation, in which the target values and the target probabilities can be specified in advance, occurs when Y has a discrete distribution with finite support. Denote the realizations of Y by

$$\eta_1 < \eta_2 < \cdots < \eta_k,$$

and the corresponding probabilities by p_i , i = 1, ..., k. Then the dominance constraint is equivalent to

$$\Pr\{g(x,Z) \ge \eta_i\} \ge \sum_{j=i}^k p_j, \quad i = 1, \dots, k.$$

Here, we use the fact that the probability distribution function of g(x, Z) is continuous and nondecreasing.

Now, we shall derive sufficient conditions of optimality for problem (4.9). We assume additionally that the function g is jointly quasi-concave in both arguments, and Z has an α -concave probability distribution.

Theorem 4.100. Assume that a point \hat{x} is feasible for problem (4.9). Suppose that there exists a function $\hat{u} \in \mathcal{U}$, $\hat{u} \neq 0$, such that conditions (4.118)–(4.119) are satisfied. If the functions c is convex, the function g satisfies the concavity assumptions above and the variable Z has an α -concave probability distribution, then \hat{x} is an optimal solution of problem (4.9).

Proof. By virtue of Theorem 4.45, the feasible set of problem (4.121)) is convex and closed.

Let the operator Γ and the cone K be defined as in the proof of Theorem 4.98. Using Lemma 4.96, we observe that optimality conditions (4.124)–(4.125) for problem (4.121) are satisfied. Consider a feasible direction d at the point \hat{x} . As the feasible set is convex, we conclude that

$$\Gamma(\hat{x} + \tau d) \in K,$$

for all sufficiently small $\tau > 0$. Since Γ is differentiable, we have

$$\frac{1}{\tau} \big[\Gamma(\hat{x} + \tau d) - \Gamma(\hat{x}) \big] \to \nabla_x \Gamma(\hat{x})(d) \quad \text{whenever} \quad \tau \downarrow 0.$$

This implies, that

$$\nabla_x \Gamma(\hat{x})(d) \in T_K(\Gamma(\hat{x})),$$

where $\mathcal{T}_K(\gamma)$ denotes the tangent cone to K at γ . Since

$$\mathcal{T}_K(\gamma) = K + \{t\gamma : t \in \mathbb{R}\},\$$

there exists $t \in \mathbb{R}$ such that

$$\nabla_x \Gamma(\hat{x})(d) + t \Gamma(\hat{x}) \in K. \tag{4.129}$$

Condition (4.124) implies that there exists $q \in \mathcal{N}_{\mathcal{X}}(\hat{x})$ such that

$$\nabla_x c(\hat{x}) + \int_a^b \nabla_x \Gamma(\hat{x})(\eta) \, d\mu(\eta) = -q$$

Applying both sides of this equation to the direction d and using the fact that $q \in \mathcal{N}_{\mathcal{X}}(\hat{x})$ and $d \in \mathcal{T}_{\mathcal{X}}(\hat{x})$, we obtain:

$$\nabla_x c(\hat{x})(d) + \int_a^b \left(\nabla_x \Gamma(\hat{x})(\eta) \right)(d) \, d\mu(\eta) \ge 0.$$
(4.130)

Condition (4.125), relation (4.129), and the nonnegativity of μ imply that

$$\int_{a}^{b} \left(\nabla_{x} \Gamma(\hat{x})(\eta) \right)(d) \, d\mu(\eta) = \int_{a}^{b} \left[\left(\nabla_{x} \Gamma(\hat{x})(\eta) \right)(d) + t \left(\Gamma(\hat{x}) \right)(\eta) \right] d\mu(\eta) \le 0.$$

Substituting into (4.130) we conclude that

 $d^{\mathsf{T}} \nabla_x c(\hat{x}) \ge 0,$

for every feasible direction d at \hat{x} . By the convexity of c, for every feasible point x we obtain the inequality

$$c(x) \ge c(\hat{x}) + d^{\mathsf{T}} \nabla_x c(\hat{x}) \ge c(\hat{x}),$$

as stated. \Box

Exercises

- 4.1. Are the following density functions α -concave and do they define a γ -concave probability measure? What are α and γ ?
 - (a) If the *m*-dimensional random vector Z has the normal distribution with expected value $\mu = 0$ and covariance matrix Σ , the random variable Y is independent of Z and has the χ_k^2 distribution, then the distribution of the vector X with components

$$X_i = \frac{Z_i}{\sqrt{Y/k}}, \quad i = 1, \dots, m,$$

is called a *multivariate Student distribution*. Its density function is defined as follows:

$$\theta_m(x) = \frac{\Gamma(\frac{m+k}{2})}{\Gamma(\frac{k}{2})\sqrt{(2\pi)^m \det(\varSigma)}} \left(1 + \frac{1}{k}x^{\mathsf{T}}\varSigma^{\frac{1}{2}}x\right)^{-(m+k)/2},$$

If m = k = 1, then this function reduces to the well-known univariate Cauchy density

$$\theta_1(x) = \frac{1}{\pi} \frac{1}{1+x^2}, -\infty < x < \infty.$$

(b) The density function of the *m*-dimensional *F*-distribution with parameters n_0, \ldots, n_m , and $n = \sum_{i=1}^m n_i$, defined as follows:

$$\theta(x) = c \prod_{i=1}^{m} x_i^{n_i/2-1} \left(n_0 + \sum_{i=1}^{m} n_i x_i \right)^{-n/2}, \quad x_i \ge 0, \ i = 1, \dots, m,$$

where c is an appropriate normalizing constant.

(c) Consider another multivariate generalization of the *beta distribution*, which is obtained in the following way. Let S_1 and S_2 be two independent sampling covariance matrices corresponding to two independent samples of sizes $s_1 + 1$ and $s_2 + 1$, respectively, taken from the same q-variate normal distribution with covariance matrix Σ . The joint distribution of the elements on and above the main diagonal of the random matrix

$$(S_1 + S_2)^{\frac{1}{2}} S_2 (S_1 + S_2)^{-\frac{1}{2}}$$

is continuous if $s_1 \ge q$ and $s_2 \ge q$. The probability density function of this distribution is defined by

$$\theta(X) = \begin{cases} \frac{c(s_1,q)c(s_2,q)}{c(s_1+s_2,q)} \ \det(X)^{\frac{1}{2}(s_2-q-1)} \ \det(I-X)^{\frac{1}{2}(s_1-q-1)}, \\ & \text{for } X, I-X \text{positive definite,} \\ 0, & \text{otherwise.} \end{cases}$$

Here *I* stands for the identity matrix and the function $c(\cdot, \cdot)$ is defined as follows:

$$\frac{1}{c(k,q)} = 2^{qk/2} \pi^{q(q-1)/2} \prod_{i=1}^{q} \Gamma\left(\frac{k-i+1}{2}\right)$$

The number of independent variables in X is $s = \frac{1}{2}q(q+1)$.

(d) The probability density function of the Pareto distribution is

$$\theta(x) = a(a+1)\dots(a+s-1)\left(\prod_{j=1}^{s}\Theta_{j}\right)^{-1}\left(\sum_{j=1}^{s}\Theta_{j}^{-1}x_{j}-s+1\right)^{-(a+s)}$$

for $x_i > \Theta_i$, i = 1, ..., s and $\theta(x) = 0$ otherwise. Here Θ_i , i = 1, ..., s are positive constants.

- 4.2. Assume that P is an α -concave probability distribution and $A \subset \mathbb{R}^n$ is a convex set. Prove that the function f(x) = P(A + x) is α -concave.
- 4.3. Prove that if $\theta : \mathbb{R} \to \mathbb{R}$ is a log-concave probability density function, then the functions

$$F(x) = \int_{t \le x} \theta(t) dt$$
 and $\bar{F}(x) = 1 - F(x)$

are log-concave as well.

- 4.4. Check that the binomial, the Poisson, the geometric, and the hypergeometric onedimensional probability distributions satisfy the conditions of Theorem 4.38 and are, therefore, log-concave.
- 4.5. Let Z_1, Z_2 , and Z_3 be independent exponentially distributed random variables with parameters λ_1, λ_2 , and λ_3 , respectively. We define $Y_1 = \min\{Z_1, Z_3\}$ and $Y_2 = \min\{Z_2, Z_3\}$. Describe $G(\eta_1, \eta_2) = P(Y_1 \ge \eta_1, Y_2 \ge \eta_2)$ for nonnegative scalars η_1 and η_2 and prove that $G(\eta_1, \eta_2)$ is log-concave on \mathbb{R}^2 .
- 4.6. Let Z be a standard normal random variable, W be a χ^2 -random variable with one degree of freedom, and A be an $n \times n$ positive definite matrix. Is the set

$$\left\{x \in \mathbb{R}^n : \Pr\left(Z - \sqrt{(x^{\mathsf{T}}Ax)W} \ge 0\right) \ge 0.9\right\}$$

convex?

4.7. If Y is an m-dimensional random vector with a log-normal distribution, and $g : \mathbb{R}^n \to \mathbb{R}^m$ is such that each component g_i is a concave function, the show that the set

$$C = \left\{ x \in \mathbb{R}^n : \Pr(g(x) \ge Y) \ge 0.9 \right\}$$

is convex.

- (a) Find the set of *p*-efficient points for m = 1, p = 0.9 and write an equivalent algebraic description of *C*.
- (b) Assume that m = 2 and the components of Y are independent. Find a disjunctive algebraic formulation for the set C.

4.8. Consider the following optimization problem:

$$\begin{split} & \underset{x}{\operatorname{Max}} c^{\mathsf{T}} x \\ & \text{s.t. } \mathsf{Pr} \big\{ g_i(x) \geq Y_i, \, i = 1, 2 \big\} \geq 0.9, \\ & x \geq 0. \end{split}$$

Here $c \in \mathbb{R}^n$, $g_i : \mathbb{R}^n \to \mathbb{R}$, i = 1, 2 are concave functions, Y_1 and Y_2 are independent random variables that have the log-normal distribution with parameters $\mu = 0$, $\sigma = 2.$

Formulate necessary and sufficient optimality conditions for this problem.

4.9. Assuming that Y and Z are independent exponentially distributed random variables, show that the following set is convex:

$$\Big\{x \in \mathbb{R}^3 : \Pr\left(x_1^2 + x_2^2 + x_2x_3 + x_3^2 + Y(x_2 + x_3) + Y^2 \le Z\right) \ge 0.9\Big\}.$$

4.10. Assume that the random variable Z is uniformly distributed in the interval [-1, 1]and $u = (1, ..., 1)^{\mathsf{T}}$. Prove that the following set is convex:

$$\Big\{x \in \mathbb{R}^n : \Pr\left(e^{x^{\mathsf{T}}y} \ge (u^{\mathsf{T}}y)Z, \ \forall y \in \mathbb{R}^n : \|y\| \le 1\right) \ge 0.95\Big\}.$$

4.11. Let Z be a two-dimensional random vector with Dirichlet distribution. Show that the following set is convex:

$$\left\{x \in \mathbb{R}^2 : \Pr\left(\min(x_1 + 2x_2 + Z_1, x_1Z_2 - x_1^2 - Z_2^2) \ge y\right) \ge e^{-y}, \ \forall y \in [\frac{1}{4}, 4]\right\}.$$

4.12. Let Z be an n-dimensional random vector uniformly distributed on a set A. Check whether the set

$$\left\{ x \in \mathbb{R}^n : \Pr\left(x^{\mathsf{T}}Z \le 1\right) \ge 0.95 \right\}$$

is convex for the following cases:

- (a) $A = \{ z \in \mathbb{R}^n : ||z|| \le 1 \}.$
- (b) $A = \{z \in \mathbb{R}^n : -i \le z_i \le i, i = 1, ..., m\}.$ (c) $A = \{z \in \mathbb{R}^n : Tz = 0, -1 \le z_i \le 1, i = 1, ..., m\}$, where T is an $(n-1) \times n$ matrix.
- 4.13. Assume that the two-dimensional random vector Z has independent components, which have the Poisson distribution with parameters $\lambda_1 = \lambda_2 = 2$. Find all pefficient points of F_Z for p = 0.8.

Chapter 5 Statistical Inference

Alexander Shapiro

5.1 Statistical Properties of SAA Estimators

Consider the following stochastic programming problem

$$\min_{x \in \mathcal{X}} \left\{ f(x) := \mathbb{E}[F(x,\xi)] \right\}.$$
(5.1)

Here \mathcal{X} is a nonempty closed subset of \mathbb{R}^n , ξ is a random vector whose probability distribution P is supported on a set $\Xi \subset \mathbb{R}^d$, and $F : \mathcal{X} \times \Xi \to \mathbb{R}$. In the framework of two-stage stochastic programming the objective function $F(x,\xi)$ is given by the optimal value of the corresponding second stage problem. Unless stated otherwise we assume in this chapter that the expectation function f(x) is well defined and *finite valued* for all $x \in \mathcal{X}$. This implies, of course, that for every $x \in \mathcal{X}$ the value $F(x,\xi)$ is finite for a.e. $\xi \in \Xi$. In particular, for two-stage programming this implies that the recourse is relatively complete.

Suppose that we have a sample $\xi^1, ..., \xi^N$ of N realizations of the random vector ξ . This random sample can be viewed as historical data of N observations of ξ , or it can be generated in the computer by Monte Carlo sampling techniques. For any $x \in \mathcal{X}$ we can estimate the expected value f(x) by averaging values $F(x, \xi^j), j = 1, ..., N$. This leads to the following, so-called *sample average approximation* (SAA):

$$\operatorname{Min}_{x \in \mathcal{X}} \left\{ \hat{f}_N(x) := \frac{1}{N} \sum_{j=1}^N F(x, \xi^j) \right\}$$
(5.2)

of the "true" problem (5.1). Let us observe that we can write the sample average function as the expectation

$$\hat{f}_N(x) = \mathbb{E}_{P_N}[F(x,\xi)] \tag{5.3}$$

taken with respect to the *empirical distribution*¹ (measure) $P_N := N^{-1} \sum_{j=1}^N \delta(\xi^j)$. Therefore, for a given sample, the SAA problem (5.2) can be considered as a stochastic programming problem with respective scenarios $\xi^1, ..., \xi^N$, each taken with probability 1/N.

As with data vector ξ , the sample $\xi^1, ..., \xi^N$ can be considered from two points of view, as a sequence of random vectors or as a particular realization of that sequence. Which one of these two meanings will be used in a particular situation will be clear from the context. The SAA problem is a function of the considered sample, and in that sense is random. For a particular realization of the random sample the corresponding SAA problem is a stochastic programming problem with respective scenarios $\xi^1, ..., \xi^N$ each taken with probability 1/N. We always assume that each random vector ξ^j , in the sample, has the same (marginal) distribution P as the data vector ξ . If, moreover, each $\xi^j, j = 1, ..., N$, is distributed independently of other sample vectors, we say that the sample is *independently identically distributed* (iid).

By the Law of Large Numbers we have that, under some regularity conditions, $f_N(x)$ converges pointwise w.p.1 to f(x) as $N \to \infty$. In particular, by the classical LLN this holds if the sample is iid. Moreover, under mild additional conditions the convergence is uniform (see section 7.2.5). We also have that $\mathbb{E}[\hat{f}_N(x)] = f(x)$, i.e., $\hat{f}_N(x)$ is an *unbiased* estimator of f(x). Therefore, it is natural to expect that the optimal value and optimal solutions of the SAA problem (5.2) converge to their counterparts of the true problem (5.1) as $N \to \infty$. We denote by ϑ^* and S the optimal value and the set of optimal solutions, respectively, of the true problem (5.1), and by $\hat{\vartheta}_N$ and \hat{S}_N the optimal value and the set of optimal solutions, respectively, of the SAA problem (5.2).

We can view the sample average functions $f_N(x)$ as defined on a common probability space (Ω, \mathcal{F}, P) . For example, in the case of iid sample, a standard construction is to consider the set $\Omega := \Xi^{\infty}$ of sequences $\{(\xi_1, ...)\}_{\xi_i \in \Xi, i \in \mathbb{N}}$, equipped with the product of the corresponding probability measures. Assume that $F(x, \xi)$ is a *Carathéodory function*, i.e., continuous in x and measurable in ξ . Then $\hat{f}_N(x) = \hat{f}_N(x, \omega)$ is also a Carathéodory function and hence is a random lsc function. It follows (see section 7.2.3 and Theorem 7.42 in particular) that $\hat{\vartheta}_N = \hat{\vartheta}_N(\omega)$ and $\hat{S}_N = \hat{S}_N(\omega)$ are measurable. We also consider a particular optimal solution \hat{x}_N of the SAA problem, and view it as a measurable selection $\hat{x}_N(\omega) \in \hat{S}_N(\omega)$. Existence of such measurable selection is ensured by the Measurable Selection Theorem (Theorem 7.39). This takes care of the measurability questions.

We are going to discuss next statistical properties of the SAA estimators ϑ_N and S_N . Let us make the following useful observation.

Proposition 5.1. Let $f : \mathcal{X} \to \mathbb{R}$ and $f_N : \mathcal{X} \to \mathbb{R}$ be a sequence of (deterministic) real valued functions. Then the following two properties are equivalent: (i) for any $\bar{x} \in \mathcal{X}$ and any sequence $\{x_N\} \subset \mathcal{X}$ converging to \bar{x} it follows that $f_N(x_N)$ converges to $f(\bar{x})$, (ii) the function $f(\cdot)$ is continuous on \mathcal{X} and $f_N(\cdot)$ converges to $f(\cdot)$ uniformly on any compact subset of \mathcal{X} .

Proof. Suppose that property (i) holds. Consider a point $\bar{x} \in \mathcal{X}$, a sequence $\{x_N\} \subset \mathcal{X}$ converging to \bar{x} and a number $\varepsilon > 0$. By taking a sequence with each element equal x_1 , we have by (i) that $f_N(x_1) \to f(x_1)$. Therefore there exists N_1 such that $|f_{N_1}(x_1) - f(x_1)| < \varepsilon$

¹Recall that $\delta(\xi)$ denotes measure of mass one at the point ξ .

 $\varepsilon/2$. Similarly there exists $N_2 > N_1$ such that $|f_{N_2}(x_2) - f(x_2)| < \varepsilon/2$, and so on. Consider now a sequence, denoted x'_N , constructed as follows, $x'_i = x_1$, $i = 1, ..., N_1$, $x'_i = x_2$, $i = N_1 + 1, ..., N_2$, and so on. We have that this sequence x'_N converges to \bar{x} and hence $|f_N(x'_N) - f(\bar{x})| < \varepsilon/2$ for all N large enough. We also have that $|f_{N_k}(x'_{N_k}) - f(x_k)| < \varepsilon/2$, and hence $|f(x_k) - f(\bar{x})| < \varepsilon$ for all k large enough. This shows that $f(x_k) \to f(\bar{x})$ and hence $f(\cdot)$ is continuous at \bar{x} .

Now let C be a compact subset of \mathcal{X} . Arguing by a contradiction suppose that $f_N(\cdot)$ does not converge to $f(\cdot)$ uniformly on C. Then there exists a sequence $\{x_N\} \subset C$ and $\varepsilon > 0$ such that $|f_N(x_N) - f(x_N)| \ge \varepsilon$ for all N. Since C is compact, we can assume that $\{x_N\}$ converges to a point $\bar{x} \in C$. We have

$$|f_N(x_N) - f(x_N)| \le |f_N(x_N) - f(\bar{x})| + |f(x_N) - f(\bar{x})|.$$
(5.4)

The first term in the right hand side of (5.4) tends to zero by (i) and the second term tends to zero since $f(\cdot)$ is continuous, and hence these terms are less that $\varepsilon/2$ for N large enough. This gives a designed contradiction.

Conversely, suppose that property (ii) holds. Consider a sequence $\{x_N\} \subset \mathcal{X}$ converging to a point $\bar{x} \in \mathcal{X}$. We can assume that this sequence is contained in a compact subset of \mathcal{X} . By employing the inequality

$$|f_N(x_N) - f(\bar{x})| \le |f_N(x_N) - f(x_N)| + |f(x_N) - f(\bar{x})|$$
(5.5)

and noting that the first term in the right hand side of this inequality tends to zero because of the uniform convergence of f_N to f and the second term tends to zero by continuity of f, we obtain that property (i) holds. \Box

5.1.1 Consistency of SAA Estimators

In this section we discuss convergence properties of the SAA estimators $\hat{\vartheta}_N$ and \hat{S}_N . It is said that an estimator $\hat{\theta}_N$ of a parameter θ is *consistent* if $\hat{\theta}_N$ converges w.p.1 to θ as $N \to \infty$. Let us consider first consistency of the SAA estimator of the optimal value. We have that for any fixed $x \in \mathcal{X}$, $\hat{\vartheta}_N \leq \hat{f}_N(x)$ and hence if the pointwise LLN holds, then

$$\limsup_{N \to \infty} \hat{\vartheta}_N \le \lim_{N \to \infty} \hat{f}_N(x) = f(x) \text{ w.p.1.}$$

It follows that if the pointwise LLN holds, then

$$\limsup_{N \to \infty} \hat{\vartheta}_N \le \vartheta^* \quad \text{w.p.1.}$$
(5.6)

Without some additional conditions, the inequality in (5.6) can be strict.

Proposition 5.2. Suppose that $\hat{f}_N(x)$ converges to f(x) w.p.1, as $N \to \infty$, uniformly on \mathcal{X} . Then $\hat{\vartheta}_N$ converges to ϑ^* w.p.1 as $N \to \infty$.

Proof. The uniform convergence w.p.1 of $\hat{f}_N(x) = \hat{f}_N(x,\omega)$ to f(x) means that for any $\varepsilon > 0$ and a.e. $\omega \in \Omega$ there is $N^* = N^*(\varepsilon, \omega)$ such that the following inequality holds for

all $N \ge N^*$:

$$\sup_{x \in \mathcal{X}} \left| \hat{f}_N(x, \omega) - f(x) \right| \le \varepsilon.$$
(5.7)

It follows then that $|\hat{\vartheta}_N(\omega) - \vartheta^*| \leq \varepsilon$ for all $N \geq N^*$, which completes the proof. \Box

In order to establish consistency of the SAA estimators of optimal solutions we need slightly stronger conditions. Recall that $\mathbb{D}(A, B)$ denotes the deviation of set A from set B (see equation (7.4) for the corresponding definition).

Theorem 5.3. Suppose that there exists a compact set $C \subset \mathbb{R}^n$ such that: (i) the set S of optimal solutions of the true problem is nonempty and is contained in C, (ii) the function f(x) is finite valued and continuous on C, (iii) $\hat{f}_N(x)$ converges to f(x) w.p.1, as $N \to \infty$, uniformly in $x \in C$, (iv) w.p.1 for N large enough the set \hat{S}_N is nonempty and $\hat{S}_N \subset C$. Then $\hat{\vartheta}_N \to \vartheta^*$ and $\mathbb{D}(\hat{S}_N, S) \to 0$ w.p.1 as $N \to \infty$.

Proof. Assumptions (i) and (iv) imply that both the true and SAA problems can be restricted to the set $\mathcal{X} \cap C$. Therefore we can assume without loss of generality that the set \mathcal{X} is compact. The assertion that $\hat{\vartheta}_N \to \vartheta^*$ w.p.1 follows by Proposition 5.2. It suffices show now that $\mathbb{D}(\hat{S}_N(\omega), S) \to 0$ for every $\omega \in \Omega$ such that $\hat{\vartheta}_N(\omega) \to \vartheta^*$ and assumptions (iii) and (iv) hold. This basically a deterministic result, therefore we omit ω for the sake of notational convenience.

We argue now by a contradiction. Suppose that $\mathbb{D}(\hat{S}_N, \mathcal{S}) \not\to 0$. Since \mathcal{X} is compact, by passing to a subsequence if necessary, we can assume that there exists $\hat{x}_N \in \hat{S}_N$ such that $\operatorname{dist}(\hat{x}_N, \mathcal{S}) \ge \varepsilon$ for some $\varepsilon > 0$, and that \hat{x}_N tends to a point $x^* \in \mathcal{X}$. It follows that $x^* \notin \mathcal{S}$ and hence $f(x^*) > \vartheta^*$. Moreover, $\hat{\vartheta}_N = \hat{f}_N(\hat{x}_N)$ and

$$\hat{f}_N(\hat{x}_N) - f(x^*) = [\hat{f}_N(\hat{x}_N) - f(\hat{x}_N)] + [f(\hat{x}_N) - f(x^*)].$$
(5.8)

The first term in the right hand side of (5.8) tends to zero by the assumption (iii) and the second term by continuity of f(x). That is, we obtain that $\hat{\vartheta}_N$ tends to $f(x^*) > \vartheta^*$, a contradiction. \Box

Recall that, by Proposition 5.1, assumptions (ii) and (iii) in the above theorem are equivalent to the condition that for any sequence $\{x_N\} \subset C$ converging to a point \bar{x} it follows that $\hat{f}_N(x_N) \to f(\bar{x})$ w.p.1. The last assumption (iv) in the above theorem holds, in particular, if the feasible set \mathcal{X} is closed, the functions $\hat{f}_N(x)$ are lower semicontinuous and for some $\alpha > \vartheta^*$ the level sets $\{x \in \mathcal{X} : \hat{f}_N(x) \le \alpha\}$ are uniformly bounded w.p.1. This condition is often referred to as the *inf-compactness condition* (compare with Definition 7.22). Conditions ensuring the uniform convergence of $\hat{f}_N(x)$ to f(x) (assumption (iii)) are given in Theorems 7.53 and 7.55, for example.

The assertion that $\mathbb{D}(\hat{S}_N, S) \to 0$ w.p.1 means that for any (measurable) selection $\hat{x}_N \in \hat{S}_N$, of an optimal solution of the SAA problem, it holds that $\operatorname{dist}(\hat{x}_N, S) \to 0$ w.p.1. If, moreover, $S = \{\bar{x}\}$ is a singleton, i.e., the true problem has unique optimal solution \bar{x} , then this means that $\hat{x}_N \to \bar{x}$ w.p.1. The inf-compactness condition ensures that \hat{x}_N cannot escape to infinity as N increases.

If the problem is convex, it is possible to relax the required regularity conditions. In the following theorem we assume that the integrand function $F(x,\xi)$ is an extended real valued function, i.e., can also take values $\pm \infty$. Denote

$$\bar{F}(x,\xi) := F(x,\xi) + \mathbb{I}_{\mathcal{X}}(x), \ \bar{f}(x) := f(x) + \mathbb{I}_{\mathcal{X}}(x), \ \bar{f}_N(x) := \bar{f}_N(x) + \mathbb{I}_{\mathcal{X}}(x), \ (5.9)$$

i.e., $\bar{f}(x) = f(x)$ if $x \in \mathcal{X}$, and $\bar{f}(x) = +\infty$ if $x \notin \mathcal{X}$, and similarly for functions $F(\cdot, \xi)$ and $\hat{f}_N(\cdot)$. Clearly $\bar{f}(x) = \mathbb{E}[\bar{F}(x,\xi)]$ and $\tilde{f}_N(x) = N^{-1} \sum_{j=1}^N \bar{F}(x,\xi^j)$. Note that if the set \mathcal{X} is convex, then the above 'penalization' operation preserves convexity of respective functions.

Theorem 5.4. Suppose that: (i) the integrand function F is random lower semicontinuous, (ii) for almost every $\xi \in \Xi$ the function $F(\cdot, \xi)$ is convex, (iii) the set \mathcal{X} is closed and convex, (iv) the expected value function f is lower semicontinuous and there exists a point $\bar{x} \in \mathcal{X}$ such that $f(x) < +\infty$ for all x in a neighborhood of \bar{x} , (v) the set S of optimal solutions of the true problem is nonempty and bounded, (vi) the LLN holds pointwise. Then $\hat{\vartheta}_N \to \vartheta^*$ and $\mathbb{D}(\hat{S}_N, S) \to 0$ w.p.1 as $N \to \infty$.

Proof. Clearly we can restrict both the true and the SAA problems to the affine space generated by the convex set \mathcal{X} . Relative to that affine space the set \mathcal{X} has a nonempty interior. Therefore, without loss of generality we can assume that the set \mathcal{X} has a nonempty interior. Since it is assumed that f(x) possesses an optimal solution, we have that ϑ^* is finite and hence $f(x) \geq \vartheta^* > -\infty$ for all $x \in \mathcal{X}$. Since f(x) is convex and is greater than $-\infty$ on an open set (e.g., interior of \mathcal{X}), it follows that $f(\cdot)$ is subdifferentiable at any point $x \in int(\mathcal{X})$ such that f(x) is finite. Consequently $f(x) > -\infty$ for all $x \in \mathbb{R}^n$, and hence f is proper.

Observe that the pointwise LLN for $F(x,\xi)$ (assumption (vi)) implies the corresponding pointwise LLN for $\overline{F}(x,\xi)$. Since \mathcal{X} is convex and closed, it follows that \overline{f} is convex and lower semicontinuous. Moreover, because of the assumption (iv) and since the interior of \mathcal{X} is nonempty, we have that $\mathrm{dom}\overline{f}$ has a nonempty interior. By Theorem 7.54 it follows then that $\widetilde{f}_N \stackrel{\mathrm{e}}{\to} \overline{f}$ w.p.1. Consider a compact set K with a nonempty interior and such that it does not contain a boundary point of $\mathrm{dom}\overline{f}$, and $\overline{f}(x)$ is finite valued on K. Since $\mathrm{dom}\overline{f}$ has a nonempty interior such set exists. Then it follows from $\widetilde{f}_N \stackrel{\mathrm{e}}{\to} \overline{f}$, that $\widetilde{f}_N(\cdot)$ converge to $\overline{f}(\cdot)$ uniformly on K, all w.p.1 (see Theorem 7.31). It follows that w.p.1 for N large enough the functions $\widetilde{f}_N(x)$ are finite valued on K, and hence are proper.

Now let C be a compact subset of \mathbb{R}^n such that the set S is contained in the interior of C. Such set exists since it is assumed that the set S is bounded. Consider the set \widetilde{S}_N of minimizers of $\tilde{f}_N(x)$ over C. Since C is nonempty and compact and $\tilde{f}_N(x)$ is lower semicontinuous and proper for N large enough, and because by the pointwise LLN we have that for any $x \in S$, $\tilde{f}_N(x)$ is finite w.p.1 for N large enough, the set \widetilde{S}_N is nonempty w.p.1 for N large enough. Let us show that $\mathbb{D}(\widetilde{S}_N, S) \to 0$ w.p.1. Let $\omega \in \Omega$ be such that $\tilde{f}_N(\cdot, \omega) \stackrel{e}{\to} \bar{f}(\cdot)$. We have that this happens for a.e. $\omega \in \Omega$. We argue now by a contradiction. Suppose that there exists a minimizer $\tilde{x}_N = \tilde{x}_N(\omega)$ of $\tilde{f}_N(x, \omega)$ over C such that $\operatorname{dist}(\tilde{x}_N, S) \geq \varepsilon$ for some $\varepsilon > 0$. Since C is compact, by passing to a subsequence if necessary, we can assume that \tilde{x}_N tends to a point $x^* \in C$. It follows that $x^* \notin S$. On the other hand we have by Proposition 7.30 that $x^* \in \arg\min_{x \in C} \bar{f}(x)$. Since $\arg \min_{x \in C} \overline{f}(x) = S$, we obtain a contradiction.

Now because of the convexity assumptions, any minimizer of $f_N(x)$ over C which lies inside the interior of C, is also an optimal solution of the SAA problem (5.2). Therefore, w.p.1 for N large enough we have that $\tilde{S}_N = \hat{S}_N$. Consequently we can restrict both the true and the SAA optimization problems to the compact set C, and hence the assertions of the above theorem follow. \Box

Let us make the following observations. Lower semicontinuity of $f(\cdot)$ follows from lower semicontinuity $F(\cdot,\xi)$, provided that $F(x,\cdot)$ is bounded from below by an integrable function (see Theorem 7.47 for a precise formulation of this result). It was assumed in the above theorem that the LLN holds pointwise for all $x \in \mathbb{R}^n$. Actually it suffices to assume that this holds for all x in some neighborhood of the set S. Under the assumptions of the above theorem we have that $f(x) > -\infty$ for every $x \in \mathbb{R}^n$. The above assumptions do not prevent, however, for f(x) to take value $+\infty$ at some points $x \in \mathcal{X}$. Nevertheless, it was possible to push the proof through because in the considered convex case local optimality implies global optimality. There are two possible reasons why f(x) can be $+\infty$. Namely, it can be that $F(x, \cdot)$ is finite valued but grows sufficiently fast so that its integral is $+\infty$, or it can be that $F(x, \cdot)$ is equal $+\infty$ on a set of positive measure, and of course it can be both. For example, in the case of two-stage programming it may happen that for some $x \in \mathcal{X}$ the corresponding second stage problem is infeasible with a positive probability p. Then w.p.1 for N large enough, for at least one of the sample points ξ^{j} the corresponding second stage problem will be infeasible, and hence $f_N(x) = +\infty$. Of course, if the probability p is very small, then the required sample size for such event to happen could be very large.

We assumed so far that the feasible set \mathcal{X} of the SAA problem is fixed, i.e., independent of the sample. However, in some situations it also should be estimated. Then the corresponding SAA problem takes the form

$$\underset{\in \mathcal{X}_N}{\min} \hat{f}_N(x),$$
(5.10)

where \mathcal{X}_N is a subset of \mathbb{R}^n depending on the sample, and therefore is random. As before we denote by $\hat{\vartheta}_N$ and $\hat{\mathcal{S}}_N$ the optimal value and the set of optimal solutions, respectively, of the SAA problem (5.10).

Theorem 5.5. Suppose that in addition to the assumptions of Theorem 5.3 the following conditions hold:

(a) If x_N ∈ X_N and x_N converges w.p.1 to a point x, then x ∈ X.
(b) For some point x ∈ S there exists a sequence x_N ∈ X_N such that x_N → x w.p.1. Then ϑ_N → ϑ* and D(Ŝ_N, S) → 0 w.p.1 as N → ∞.

Proof. Consider an $\hat{x}_N \in \hat{S}_N$. By compactness arguments we can assume that \hat{x}_N converges w.p.1 to a point $x^* \in \mathbb{R}^n$. Since $\hat{S}_N \subset \mathcal{X}_N$, we have that $\hat{x}_N \in \mathcal{X}_N$, and hence it follows by condition (a) that $x^* \in \mathcal{X}$. We also have (see Proposition 5.1) that $\hat{\vartheta}_N = \hat{f}_N(\hat{x}_N)$ tends w.p.1 to $f(x^*)$, and hence $\liminf_{N\to\infty} \hat{\vartheta}_N \geq \vartheta^*$ w.p.1. On the other hand, by condition (b), there exists a sequence $x_N \in \mathcal{X}_N$ converging to a point $x \in \mathcal{S}$ w.p.1. Consequently, $\hat{\vartheta}_N \leq \hat{f}_N(\hat{x}_N) \to f(x) = \vartheta^*$ w.p.1, and hence $\limsup_{N\to\infty} \hat{\vartheta}_N \leq \vartheta^*$. It follows

that $\hat{\vartheta}_N \to \vartheta^*$ w.p.1. The remainder of the proof can be completed by the same arguments as in the proof of Theorem 5.3. \Box

The SAA problem (5.10) is convex if the functions $f_N(\cdot)$ and the sets \mathcal{X}_N are convex w.p.1. It is also possible to show consistency of the SAA estimators of problem (5.10) under the assumptions of Theorem 5.4 together with conditions (a) and (b), of the above Theorem 5.5, and convexity of the set \mathcal{X}_N .

Suppose, for example, that the set \mathcal{X} is defined by the constraints

$$\mathcal{X} := \{ x \in \mathcal{X}_0 : g_i(x) \le 0, \ i = 1, ..., p \},$$
(5.11)

where \mathcal{X}_0 is a nonempty closed subset of \mathbb{R}^n and the constraint functions are given as the expected value functions

$$g_i(x) := \mathbb{E}[G_i(x,\xi)], \quad i = 1, ..., p,$$
(5.12)

with $G_i(x,\xi)$, i = 1, ..., p, being random lsc functions. Then the set \mathcal{X} can be estimated by

$$\mathcal{X}_N := \{ x \in \mathcal{X}_0 : \hat{g}_{iN}(x) \le 0, \ i = 1, ..., p \},$$
(5.13)

where

$$\hat{g}_{iN}(x) := \frac{1}{N} \sum_{j=1}^{N} G_i(x, \xi^j).$$

If for a given point $x \in \mathcal{X}_0$, every function \hat{g}_{iN} converges uniformly to g_i w.p.1 on a neighborhood of x and the functions g_i are continuous, then condition (a) of Theorem 5.5 holds.

Remark 8. Let us note that the samples used in construction of the SAA functions f_N and \hat{g}_{iN} , i = 1, ..., p, can be the same or can be different, independent of each other. That is, for random samples $\xi^{i1}, ..., \xi^{iN_i}$, possibly of different sample sizes N_i , i = 1, ..., p, and independent of each other and of the random sample used in \hat{f}_N , the corresponding SAA functions are

$$\hat{g}_{iN_i}(x) := \frac{1}{N_i} \sum_{j=1}^{N_i} G_i(x, \xi^{ij}), \ i = 1, ..., p$$

The question of how to generate the respective random samples is especially relevant for Monte Carlo sampling methods discussed later. For consistency type results we only need to verify convergence w.p.1 of the involved SAA functions to their true (expected value) counterparts, and this holds under appropriate regularity conditions in both cases – of the same and independent samples. However, from a variability point of view it is advantageous to use independent samples (see Remark 12 on page 194).

In order to ensure condition (b) of Theorem 5.5 one needs to impose a constraint qualification (on the true problem). Consider, for example, $\mathcal{X} := \{x \in \mathbb{R} : g(x) \leq 0\}$ with $g(x) := x^2$. Clearly $\mathcal{X} = \{0\}$, while an arbitrary small perturbation of the function $g(\cdot)$ can result in the corresponding set \mathcal{X}_N being empty. It is possible to show that if a

constraint qualification for the true problem is satisfied at x, then condition (b) follows. For instance, if the set \mathcal{X}_0 is convex and for every $\xi \in \Xi$ the functions $G_i(\cdot, \xi)$ are convex, and hence the corresponding expected value functions $g_i(\cdot)$, i = 1, ..., p, are also convex, then such a simple constraint qualification is the Slater condition. Recall that it is said that the *Slater condition* holds if there exists a point $x^* \in \mathcal{X}_0$ such that $g_i(x^*) < 0$, i = 1, ..., p.

As another example suppose that the feasible set is given by probabilistic (chance) constraints in the form

$$\mathcal{X} = \left\{ x \in \mathbb{R}^n : \Pr\left(C_i(x,\xi) \le 0\right) \ge 1 - \alpha_i, \ i = 1, ..., p \right\},\tag{5.14}$$

where $\alpha_i \in (0,1)$ and $C_i : \mathbb{R}^n \times \Xi \to \mathbb{R}$, i = 1, ..., p, are Carathéodory functions. Of course, we have that²

$$\mathsf{Pr}\big(C_i(x,\xi) \le 0\big) = \mathbb{E}\left[\mathbf{1}_{(-\infty,0]}\big(C_i(x,\xi)\big)\right].$$
(5.15)

Consequently, we can write the above set \mathcal{X} in the form (5.11)–(5.12) with $\mathcal{X}_0 := \mathbb{R}^n$ and

$$G_i(x,\xi) := 1 - \alpha_i - \mathbf{1}_{(-\infty,0]} (C_i(x,\xi)).$$
(5.16)

The corresponding set \mathcal{X}_N can be written as

$$\mathcal{X}_{N} = \left\{ x \in \mathbb{R}^{n} : \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{(-\infty,0]} \left(C_{i}(x,\xi^{j}) \right) \ge 1 - \alpha_{i}, \ i = 1, ..., p \right\}.$$
 (5.17)

Note that $\sum_{j=1}^{N} \mathbf{1}_{(-\infty,0]}(C_i(x,\xi^j))$, in the above formula, counts the number of times that the event " $C_i(x,\xi^j) \leq 0$ ", j = 1, ..., N, happens. The additional difficulty here is that the (step) function $\mathbf{1}_{(-\infty,0]}(t)$ is discontinuous at t = 0. Nevertheless, suppose that the sample is iid and for every x in a neighborhood of the set \mathcal{X} and i = 1, ..., p, the event " $C_i(x,\xi) = 0$ " happens with probability zero, and hence $G_i(\cdot,\xi)$ is continuous at x for a.e. ξ . By Theorem 7.53 this implies that the expectation function $g_i(x)$ is continuous and $\hat{g}_{iN}(x)$ converge uniformly w.p.1 on compact neighborhoods to $g_i(x)$, and hence condition (a) of Theorem 5.5 holds. Condition (b) could be verified by ad hoc methods.

Remark 9. As it was pointed out in the above Remark 8, it is possible to use different, independent of each other, random samples $\xi^{i1}, ..., \xi^{iN_i}$, possibly of different sample sizes N_i , i = 1, ..., p, for constructing the corresponding SAA functions. That is, constraints $\Pr(C_i(x,\xi) > 0) \le \alpha_i$ are approximated by

$$\frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{1}_{(0,\infty)} \left(C_i(x,\xi^{ij}) \right) \le \alpha_i, \ i = 1, ..., p.$$
(5.18)

From the point of view of reducing variability of the respective SAA estimators it could be preferable to use this approach of independent, rather than the same, samples.

²Recall that $\mathbf{1}_{(-\infty,0]}(t) = 1$ if $t \le 0$, and $\mathbf{1}_{(-\infty,0]}(t) = 0$ if t > 0.

5.1.2 Asymptotics of the SAA Optimal Value

Consistency of the SAA estimators gives a certain assurance that the error of the estimation approaches zero in the limit as the sample size grows to infinity. Although important conceptually, this does not give any indication of the magnitude of the error for a given sample. Suppose for the moment that the sample is iid and let us fix a point $x \in \mathcal{X}$. Then we have that the sample average estimator $\hat{f}_N(x)$, of f(x), is unbiased and has variance $\sigma^2(x)/N$, where $\sigma^2(x) := \mathbb{V}ar [F(x,\xi)]$ is supposed to be finite. Moreover, by the Central Limit Theorem (CLT) we have that

$$N^{1/2}\left[\hat{f}_N(x) - f(x)\right] \xrightarrow{\mathcal{D}} Y_x,$$
(5.19)

where " $\xrightarrow{\mathcal{D}}$ " denotes convergence in *distribution* and Y_x has a normal distribution with mean 0 and variance $\sigma^2(x)$, written $Y_x \sim \mathcal{N}(0, \sigma^2(x))$. That is, $\hat{f}_N(x)$ has asymptotically normal distribution, i.e., for large N, $\hat{f}_N(x)$ has approximately normal distribution with mean f(x) and variance $\sigma^2(x)/N$.

This leads to the following (approximate) $100(1-\alpha)\%$ confidence interval for f(x):

$$\left[\hat{f}_N(x) - \frac{z_{\alpha/2}\hat{\sigma}(x)}{\sqrt{N}}, \hat{f}_N(x) + \frac{z_{\alpha/2}\hat{\sigma}(x)}{\sqrt{N}}\right],\tag{5.20}$$

where $z_{\alpha/2} := \Phi^{-1}(1 - \alpha/2)$ and³

$$\hat{\sigma}^2(x) := \frac{1}{N-1} \sum_{j=1}^N \left[F(x,\xi^j) - \hat{f}_N(x) \right]^2$$
(5.21)

is the sample variance estimate of $\sigma^2(x)$. That is, the error of estimation of f(x) is (stochastically) of order $O_p(N^{-1/2})$.

Consider now the optimal value $\hat{\vartheta}_N$ of the SAA problem (5.2). Clearly we have that for any $x' \in \mathcal{X}$ the inequality $\hat{f}_N(x') \ge \inf_{x \in \mathcal{X}} \hat{f}_N(x)$ holds. By taking the expected value of both sides of this inequality and minimizing the left hand side over all $x' \in \mathcal{X}$ we obtain

$$\inf_{x \in \mathcal{X}} \mathbb{E}\left[\hat{f}_N(x)\right] \ge \mathbb{E}\left[\inf_{x \in \mathcal{X}} \hat{f}_N(x)\right].$$
(5.22)

Note that the inequality (5.22) holds even if $f(x) = +\infty$ or $f(x) = -\infty$ for some $x \in \mathcal{X}$. Since $\mathbb{E}[\hat{f}_N(x)] = f(x)$, it follows that $\vartheta^* \ge \mathbb{E}[\hat{\vartheta}_N]$. In fact, typically, $\mathbb{E}[\hat{\vartheta}_N]$ is strictly less than ϑ^* , i.e., $\hat{\vartheta}_N$ is a downwards *biased* estimator of ϑ^* . As the following result shows this bias decreases monotonically with increase of the sample size N.

Proposition 5.6. Let $\hat{\vartheta}_N$ be the optimal value of the SAA problem (5.2), and suppose that the sample is iid. Then $\mathbb{E}[\hat{\vartheta}_N] \leq \mathbb{E}[\hat{\vartheta}_{N+1}] \leq \vartheta^*$ for any $N \in \mathbb{N}$.

³Here $\Phi(\cdot)$ denotes the cdf of the standard normal distribution. For example, to 95% confidence intervals corresponds $z_{0.025} = 1.96$.

Proof. It was already shown above that $\mathbb{E}[\hat{\vartheta}_N] \leq \vartheta^*$ for any $N \in \mathbb{N}$. We can write

$$\hat{f}_{N+1}(x) = \frac{1}{N+1} \sum_{i=1}^{N+1} \left[\frac{1}{N} \sum_{j \neq i} F(x, \xi^j) \right].$$

Moreover, since the sample is iid we have

$$\begin{split} \mathbb{E}[\hat{\vartheta}_{N+1}] &= \mathbb{E}\left[\inf_{x \in \mathcal{X}} \hat{f}_{N+1}(x)\right] \\ &= \mathbb{E}\left[\inf_{x \in \mathcal{X}} \frac{1}{N+1} \sum_{i=1}^{N+1} \left(\frac{1}{N} \sum_{j \neq i} F(x, \xi^j)\right)\right] \\ &\geq \mathbb{E}\left[\frac{1}{N+1} \sum_{i=1}^{N+1} \left(\inf_{x \in \mathcal{X}} \frac{1}{N} \sum_{j \neq i} F(x, \xi^j)\right)\right] \\ &= \frac{1}{N+1} \sum_{i=1}^{N+1} \mathbb{E}\left[\inf_{x \in \mathcal{X}} \frac{1}{N} \sum_{j \neq i} F(x, \xi^j)\right] \\ &= \frac{1}{N+1} \sum_{i=1}^{N+1} \mathbb{E}[\hat{\vartheta}_N] = \mathbb{E}[\hat{\vartheta}_N], \end{split}$$

which completes the proof. \Box

First Order Asymptotics of the SAA Optimal Value

We use the following assumptions about the integrand F:

- (A1) For some point $\tilde{x} \in \mathcal{X}$ the expectation $\mathbb{E}[F(\tilde{x},\xi)^2]$ is finite.
- (A2) There exists a measurable function $C : \Xi \to \mathbb{R}_+$ such that $\mathbb{E}[C(\xi)^2]$ is finite and

$$|F(x,\xi) - F(x',\xi)| \le C(\xi) ||x - x'||,$$
(5.23)

for all $x, x' \in \mathcal{X}$ and a.e. $\xi \in \Xi$.

The above assumptions imply that the expected value f(x) and variance $\sigma^2(x)$ are finite valued for all $x \in \mathcal{X}$. Moreover, it follows from (5.23) that

$$|f(x) - f(x')| \le \kappa ||x - x'||, \quad \forall x, x' \in \mathcal{X},$$

where $\kappa := \mathbb{E}[C(\xi)]$, and hence f(x) is Lipschitz continuous on \mathcal{X} . If \mathcal{X} is compact, we have then that the set \mathcal{S} , of minimizers of f(x) over \mathcal{X} , is nonempty.

Let Y_x be random variables defined in (5.19). These variables depend on $x \in \mathcal{X}$ and we also use notation $Y(x) = Y_x$. By the (multivariate) CLT we have that for any finite set $\{x_1, ..., x_m\} \subset \mathcal{X}$, the random vector $(Y(x_1), ..., Y(x_m))$ has a multivariate normal distribution with zero mean and the same covariance matrix as the covariance matrix of $(F(x_1, \xi), ..., F(x_m, \xi))$. Moreover, by assumptions (A1) and (A2), compactness of \mathcal{X} and since the sample is iid, we have that $N^{1/2}(\hat{f}_N - f)$ converges in distribution to Y viewed as a *random element*⁴ of $C(\mathcal{X})$. This is a so-called functional CLT (see, e.g., Araujo and Giné [4, Corollary 7.17]).

⁴Recall that $C(\mathcal{X})$ denotes the space of continuous functions equipped with the sup-norm. A random element of $C(\mathcal{X})$ is a mapping $Y : \Omega \to C(\mathcal{X})$ from a probability space (Ω, \mathcal{F}, P) into $C(\mathcal{X})$ which is measurable with respect to the Borel sigma algebra of $C(\mathcal{X})$, i.e., $Y(x) = Y(x, \omega)$ can be viewed as a random function.

Theorem 5.7. Let $\hat{\vartheta}_N$ be the optimal value of the SAA problem (5.2). Suppose that the sample is iid, the set \mathcal{X} is compact and assumptions (A1) and (A2) are satisfied. Then the following holds:

$$\hat{\vartheta}_N = \inf_{x \in \mathcal{S}} \hat{f}_N(x) + o_p(N^{-1/2}),$$
(5.24)

$$N^{1/2}\left(\hat{\vartheta}_N - \vartheta^*\right) \xrightarrow{\mathcal{D}} \inf_{x \in \mathcal{S}} Y(x).$$
 (5.25)

If, moreover, $S = \{\bar{x}\}$ is a singleton, then

$$N^{1/2}\left(\hat{\vartheta}_N - \vartheta^*\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(\bar{x})).$$
(5.26)

Proof. Proof is based on the functional CLT and Delta Theorem (Theorem 7.67). Consider Banach space $C(\mathcal{X})$ of continuous functions $\psi : \mathcal{X} \to \mathbb{R}$ equipped with the supnorm $\|\psi\| := \sup_{x \in \mathcal{X}} |\psi(x)|$. Define the min-value function $V(\psi) := \inf_{x \in \mathcal{X}} \psi(x)$. Since \mathcal{X} is compact, the function $V : C(\mathcal{X}) \to \mathbb{R}$ is real valued and measurable (with respect to the Borel sigma algebra of $C(\mathcal{X})$). Moreover, it is not difficult to see that $|V(\psi_1) - V(\psi_2)| \le \|\psi_1 - \psi_2\|$ for any $\psi_1, \psi_2 \in C(\mathcal{X})$, i.e., $V(\cdot)$ is Lipschitz continuous with Lipschitz constant one. By Danskin Theorem (Theorem 7.25), $V(\cdot)$ is directionally differentiable at any $\mu \in C(\mathcal{X})$ and

$$V'_{\mu}(\delta) = \inf_{x \in \bar{\mathcal{X}}(\mu)} \delta(x), \quad \forall \delta \in C(\mathcal{X}),$$
(5.27)

where

$$\bar{\mathcal{X}}(\mu) := \operatorname*{argmin}_{x \in \mathcal{X}} \mu(x).$$

Since $V(\cdot)$ is Lipschitz continuous, directional differentiability in the Hadamard sense follows (see Proposition 7.65). As it was discussed above, we also have here under assumptions (A1) and (A2) and since the sample is iid that $N^{1/2}(\hat{f}_N - f)$ converges in distribution to the random element Y of $C(\mathcal{X})$. Noting that $\hat{\vartheta}_N = V(\hat{f}_N)$, $\vartheta^* = V(f)$ and $\bar{X}(f) = S$, and by applying Delta Theorem to the min-function $V(\cdot)$ at $\mu := f$ and using (5.27) we obtain (5.25) and that

$$\hat{\vartheta}_N - \vartheta^* = \inf_{x \in \mathcal{S}} \left[\hat{f}_N(x) - f(x) \right] + o_p(N^{-1/2}).$$
(5.28)

Since $f(x) = \vartheta^*$ for any $x \in S$, we have that assertions (5.24) and (5.28) are equivalent. Finally, (5.26) follows from (5.25). \Box

Under mild additional conditions (see Remark 57 in section 7.2.8 on page 470) it follows from (5.25) that $N^{1/2}\mathbb{E}[\hat{\vartheta}_N - \vartheta^*]$ tends to $\mathbb{E}[\inf_{x \in S} Y(x)]$ as $N \to \infty$, that is

$$\mathbb{E}[\hat{\vartheta}_N] - \vartheta^* = N^{-1/2} \mathbb{E}\left[\inf_{x \in \mathcal{S}} Y(x)\right] + o(N^{-1/2}).$$
(5.29)

In particular, if $S = \{\bar{x}\}$ is a singleton, then by (5.26) the SAA optimal value $\hat{\vartheta}_N$ has asymptotically normal distribution and, since $\mathbb{E}[Y(\bar{x})] = 0$, we obtain that in this case

the bias $\mathbb{E}[\hat{\vartheta}_N] - \vartheta^*$ is of order $o(N^{-1/2})$. On the other hand, if the true problem has more than one optimal solution, then the right hand side of (5.25) is given by the minimum of a number of random variables. Although each Y(x) has mean zero, their minimum ' $\inf_{x \in S} Y(x)$ ' typically has a negative mean if the set S has more than one element. Therefore, if S is not a singleton, then the bias $\mathbb{E}[\hat{\vartheta}_N] - \vartheta^*$ typically is strictly less than zero and is of order $O(N^{-1/2})$. Moreover, the bias tends to be bigger the larger the set S is. For a further discussion of the bias issue see Remark 10 on page 188, of the next section.

5.1.3 Second Order Asymptotics

Formula (5.24) gives a first order expansion of the SAA optimal value $\hat{\vartheta}_N$. In this section we discuss a second order term in an expansion of $\hat{\vartheta}_N$. It turns out that the second order analysis of $\hat{\vartheta}_N$ is closely related to deriving (first order) asymptotics of optimal solutions of the SAA problem. We assume in this section that the true (expected value) problem (5.1) has unique optimal solution \bar{x} and denote by \hat{x}_N an optimal solution of the corresponding SAA problem. In order to proceed with the second order analysis we need to impose considerably stronger assumptions.

Our analysis is based on the second order Delta Theorem 7.70 and second order perturbation analysis of section 7.1.5. As in section 7.1.5 we consider a convex compact set $U \subset \mathbb{R}^n$ such that $\mathcal{X} \subset \operatorname{int}(U)$, and work with the space $W^{1,\infty}(U)$ of Lipschitz continuous functions $\psi : U \to \mathbb{R}$ equipped with the norm

$$\|\psi\|_{1,U} := \sup_{x \in U} |\psi(x)| + \sup_{x \in U'} \|\nabla\psi(x)\|,$$
(5.30)

where $U' \subset int(U)$ is the set of points where $\psi(\cdot)$ is differentiable.

We make the following assumptions about the true problem.

- (S1) The function f(x) is Lipschitz continuous on U, has unique minimizer \bar{x} over $x \in \mathcal{X}$, and is twice continuously differentiable at \bar{x} .
- (S2) The set \mathcal{X} is second order regular at \bar{x} .
- (S3) The quadratic growth condition (7.71) holds at \bar{x} .

Let \mathcal{K} be the subset of $W^{1,\infty}(U)$ formed by differentiable at \bar{x} functions. Note that the set \mathcal{K} forms a closed (in the norm topology) linear subspace of $W^{1,\infty}(U)$. Assumption (S1) ensures that $f \in \mathcal{K}$. In order to ensure that $\hat{f}_N \in \mathcal{K}$ w.p.1, we make the following assumption.

(S4) Function $F(\cdot,\xi)$ is Lipschitz continuous on U and differentiable at \bar{x} for a.e. $\xi \in \Xi$.

We view \hat{f}_N as a random element of $W^{1,\infty}(U)$, and assume, further, that $N^{1/2}(\hat{f}_N - f)$ converges in distribution to a random element Y of $W^{1,\infty}(U)$.

Consider the min-function $V: W^{1,\infty}(U) \to \mathbb{R}$ defined as

$$V(\psi) := \inf_{x \in \mathcal{X}} \psi(x), \ \psi \in W^{1,\infty}(U).$$

By Theorem 7.27, under the above assumptions (S1)–(S3), the min-function $V(\cdot)$ is second order Hadamard directionally differentiable at f tangentially to the set \mathcal{K} and we have the following formula for the second order directional derivative in a direction $\delta \in \mathcal{K}$:

$$V_f''(\delta) = \inf_{h \in C(\bar{x})} \left\{ 2h^{\mathsf{T}} \nabla \delta(\bar{x}) + h^{\mathsf{T}} \nabla^2 f(\bar{x})h - \mathsf{s} \left(-\nabla f(\bar{x}), \mathcal{T}_{\mathcal{X}}^2(\bar{x}, h) \right) \right\}.$$
(5.31)

Here $C(\bar{x})$ is the critical cone of the true problem, $\mathcal{T}^2_{\mathcal{X}}(\bar{x}, h)$ is the second order tangent set to \mathcal{X} at \bar{x} and $S(\cdot, A)$ denotes the support function of set A (see page 475 for the definition of second order directional derivatives).

Moreover, suppose that the set \mathcal{X} is given in the form

$$\mathcal{X} := \{ x \in \mathbb{R}^n : G(x) \in K \},\tag{5.32}$$

where $G : \mathbb{R}^n \to \mathbb{R}^m$ is a twice continuously differentiable mapping and $K \subset \mathbb{R}^m$ is a closed convex cone. Then, under Robinson constraint qualification, the optimal value of the right hand side of (5.31) can be written in a dual form (compare with (7.89)), which results in the following formula for the second order directional derivative in a direction $\delta \in \mathcal{K}$:

$$V_f''(\delta) = \inf_{h \in C(\bar{x})} \sup_{\lambda \in \Lambda(\bar{x})} \left\{ 2h^{\mathsf{T}} \nabla \delta(\bar{x}) + h^{\mathsf{T}} \nabla_{xx}^2 L(\bar{x}, \lambda) h - \mathsf{s}(\lambda, \mathfrak{T}(h)) \right\}.$$
(5.33)

Here

$$\mathfrak{T}(h) := \mathcal{T}_{K}^{2}(G(\bar{x}), [\nabla G(\bar{x})]h), \qquad (5.34)$$

and $L(x, \lambda)$ is the Lagrangian and $\Lambda(\bar{x})$ is the set of Lagrange multipliers of the true problem.

Theorem 5.8. Suppose that the assumptions (S1)–(S4) hold and $N^{1/2}(\hat{f}_N - f)$ converges in distribution to a random element Y of $W^{1,\infty}(U)$. Then

$$\hat{\vartheta}_N = \hat{f}_N(\bar{x}) + \frac{1}{2} V_f''(\hat{f}_N - f) + o_p(N^{-1}),$$
(5.35)

and

$$N\left[\hat{\vartheta}_N - \hat{f}_N(\bar{x})\right] \xrightarrow{\mathcal{D}} \frac{1}{2} V_f''(Y).$$
(5.36)

Moreover, suppose that for every $\delta \in \mathcal{K}$ the problem in the right hand side of (5.31) has unique optimal solution $\bar{h} = \bar{h}(\delta)$. Then

$$N^{1/2}(\hat{x}_N - \bar{x}) \xrightarrow{\mathcal{D}} \bar{h}(Y).$$
(5.37)

Proof. By the second order Delta Theorem 7.70 we have that

$$\hat{\vartheta}_N = \vartheta^* + V'_f(\hat{f}_N - f) + \frac{1}{2}V''_f(\hat{f}_N - f) + o_p(N^{-1}),$$

and

$$N\left[\hat{\vartheta}_N - \vartheta^* - V_f'(\hat{f}_N - f)\right] \xrightarrow{\mathcal{D}} \frac{1}{2}V_f''(Y).$$

We also have (compare with formula (5.27)) that

$$V'_f(\hat{f}_N - f) = \hat{f}_N(\bar{x}) - f(\bar{x}) = \hat{f}_N(\bar{x}) - \vartheta^*,$$

and hence (5.35) and (5.36) follow.

Now consider a (measurable) mapping $\mathfrak{x}: W^{1,\infty}(U) \to \mathbb{R}^n$, such that

$$\mathfrak{x}(\psi) \in \arg\min_{x \in \mathcal{X}} \psi(x), \ \psi \in W^{1,\infty}(U).$$

We have that $\mathfrak{x}(f) = \overline{x}$ and by equation (7.87) of Theorem 7.27, we have that $\mathfrak{x}(\cdot)$ is Hadamard directionally differentiable at f tangentially to \mathcal{K} , and for $\delta \in \mathcal{K}$ the directional derivative $\mathfrak{x}'(f, \delta)$ is equal to the optimal solution in the right hand side of (5.31), provided that it is unique. By applying Delta Theorem 7.69 this completes the proof of (5.37). \Box

One of the difficulties in applying the above theorem is verification of convergence in distribution of $N^{1/2}(\hat{f}_N - f)$ in the space $W^{1,\infty}(\mathcal{X})$. Actually it could be easier to prove asymptotic results (5.35) – (5.37) by direct methods. Note that formulas (5.31) and (5.33), for the second order directional derivatives $V''_f(\hat{f}_N - f)$, involve statistical properties of $\hat{f}_N(x)$ only at the (fixed) point \bar{x} . Note also that by the (finite dimensional) CLT we have that $N^{1/2} [\nabla \hat{f}_N(\bar{x}) - \nabla f(\bar{x})]$ converges in distribution to normal $\mathcal{N}(0, \Sigma)$ with the covariance matrix

$$\Sigma = \mathbb{E}\left[\left(\nabla F(\bar{x},\xi) - \nabla f(\bar{x})\right)\left(\nabla F(\bar{x},\xi) - \nabla f(\bar{x})\right)^{\mathsf{T}}\right],\tag{5.38}$$

provided that this covariance matrix is well defined and $\mathbb{E}[\nabla F(\bar{x},\xi)] = \nabla f(\bar{x})$, i.e., the differentiation and expectation operators can be interchanged (see Theorem 7.49).

Let Z be a random vector having normal distribution, $Z \sim \mathcal{N}(0, \Sigma)$, with covariance matrix Σ defined in (5.38), and let the set \mathcal{X} be given in the form (5.32). Then by the above discussion and formula (5.33), we have that under appropriate regularity conditions,

$$N\left[\hat{\vartheta}_N - \hat{f}_N(\bar{x})\right] \xrightarrow{\mathcal{D}} \frac{1}{2}\mathfrak{v}(Z),$$
(5.39)

where $\mathfrak{v}(Z)$ is the optimal value of the problem:

$$\lim_{h \in C(\bar{x})} \sup_{\lambda \in \Lambda(\bar{x})} \left\{ 2h^{\mathsf{T}} Z + h^{\mathsf{T}} \nabla_{xx}^2 L(\bar{x}, \lambda) h - \mathsf{s}(\lambda, \mathfrak{T}(h)) \right\},$$
(5.40)

with $\mathfrak{T}(h)$ being the second order tangent set defined in (5.34). Moreover, if for all Z problem (5.40) possesses unique optimal solution $\bar{h} = \mathfrak{h}(Z)$, then

$$N^{1/2}(\hat{x}_N - \bar{x}) \xrightarrow{\mathcal{D}} \mathfrak{h}(Z).$$
 (5.41)

Recall also that if the cone K is polyhedral, then the curvature term $s(\lambda, \mathfrak{T}(h))$ vanishes.

Remark 10. Note that $\mathbb{E}[\hat{f}_N(\bar{x})] = f(\bar{x}) = \vartheta^*$. Therefore, under the respective regularity conditions, in particular under the assumption that the true problem has unique optimal solution \bar{x} , we have by (5.39) that the expected value of the term $\frac{1}{2}N^{-1}\mathfrak{v}(Z)$ can be viewed

as the asymptotic bias of $\hat{\vartheta}_N$. This asymptotic bias is of order $O(N^{-1})$. This can be compared with formula (5.29) for the asymptotic bias of order $O(N^{-1/2})$ when the set of optimal solutions of the true problem is not a singleton. Note also that $\mathfrak{v}(\cdot)$ is nonpositive; in order to see that just take h = 0 in (5.40).

As an example, consider the case where the set \mathcal{X} is defined by a finite number of constraints:

$$\mathcal{X} := \{ x \in \mathbb{R}^n : g_i(x) = 0, \ i = 1, ..., q, \ g_i(x) \le 0, \ i = q + 1, ..., p \},$$
(5.42)

with the functions $g_i(x)$, i = 1, ..., p, being twice continuously differentiable. This is a particular form of (5.32) with $G(x) := (g_1(x), ..., g_p(x))$ and $K := \{0_q\} \times \mathbb{R}^{p-q}_-$. Denote

$$\mathcal{I}(\bar{x}) := \{i : g_i(\bar{x}) = 0, \ i = q+1, ..., p\}$$

the index set of active at \bar{x} inequality constraints. Suppose that the linear independence constraint qualification (LICQ) holds at \bar{x} , i.e., the gradient vectors $\nabla g_i(\bar{x}), i \in \{1, ..., q\} \cup \mathcal{I}(\bar{x})$, are linearly independent. Then the corresponding set of Lagrange multipliers is a singleton, $\Lambda(\bar{x}) = \{\bar{\lambda}\}$. In that case

$$C(\bar{x}) = \left\{ h : h^{\mathsf{T}} \nabla g_i(\bar{x}) = 0, \ i \in \{1, ..., q\} \cup \mathcal{I}_+(\bar{\lambda}), \ h^{\mathsf{T}} \nabla g_i(\bar{x}) \le 0, i \in \mathcal{I}_0(\bar{\lambda}) \right\},\$$

where

$$\mathcal{I}_0(\bar{\lambda}) := \left\{ i \in \mathcal{I}(\bar{x}) : \bar{\lambda}_i = 0 \right\} \text{ and } \mathcal{I}_+(\bar{\lambda}) := \left\{ i \in \mathcal{I}(\bar{x}) : \bar{\lambda}_i > 0 \right\}.$$

Consequently problem (5.40) takes the form

$$\begin{array}{ll}
\underset{h \in \mathbb{R}^n}{\text{Min}} & 2h^{\mathsf{T}}Z + h^{\mathsf{T}} \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})h \\
\text{s.t.} & h^{\mathsf{T}} \nabla g_i(\bar{x}) = 0, \ i \in \{1, ..., q\} \cup \mathcal{I}_+(\bar{\lambda}), \ h^{\mathsf{T}} \nabla g_i(\bar{x}) \le 0, i \in \mathcal{I}_0(\bar{\lambda}).
\end{array}$$
(5.43)

This is a quadratic programming problem. The linear independence constraint qualification implies that problem (5.43) has a unique vector $\alpha(Z)$ of Lagrange multipliers, and that it has a unique optimal solution $\mathfrak{h}(Z)$ if the Hessian matrix $H := \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})$ is positive definite over the linear space defined by the first $q + |\mathcal{I}_+(\bar{\lambda})|$ (equality) linear constraints in (5.43).

If, furthermore, the strict complementarity condition holds, i.e., $\bar{\lambda}_i > 0$ for all $i \in \mathcal{I}_+(\bar{\lambda})$, or in other words $\mathcal{I}_0(\bar{\lambda}) = \emptyset$, then $h = \mathfrak{h}(Z)$ and $\alpha = \alpha(Z)$ can be obtained as solutions of the following system of linear equations

$$\begin{bmatrix} H & A \\ A^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} h \\ \alpha \end{bmatrix} = \begin{bmatrix} Z \\ 0 \end{bmatrix}.$$
 (5.44)

Here $H = \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})$ and A is the $n \times (q + |\mathcal{I}(\bar{x})|)$ matrix whose columns are formed by vectors $\nabla g_i(\bar{x}), i \in \{1, ..., q\} \cup \mathcal{I}(\bar{x})$. Then

$$N^{1/2} \begin{bmatrix} \hat{x}_N - \bar{x} \\ \hat{\lambda}_N - \bar{\lambda} \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, J^{-1} \Upsilon J^{-1}),$$
(5.45)

where

$$J := \left[\begin{array}{cc} H & A \\ A^{\mathsf{T}} & 0 \end{array} \right] \text{ and } \Upsilon := \left[\begin{array}{cc} \varSigma & 0 \\ 0 & 0 \end{array} \right].$$

provided that the matrix J is nonsingular.

Under the linear independence constraint qualification and strict complementarity condition, we have by the second order necessary conditions that the Hessian matrix $H = \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})$ is positive semidefinite over the linear space $\{h : A^T h = 0\}$. Note that this linear space coincides here with the critical cone $C(\bar{x})$. It follows that the matrix J is nonsingular iff H is positive definite over this linear space. That is, here the nonsingularity of the matrix J is equivalent to the second order sufficient conditions at \bar{x} .

Remark 11. As it was mentioned earlier, the curvature term $S(\lambda, \mathfrak{T}(h))$ in the auxiliary problem (5.40) vanishes if the cone K is polyhedral. In particular, this happens if $K = \{0_q\} \times \mathbb{R}^{p-q}$, and hence the feasible set \mathcal{X} is given in the form (5.42). This curvature term can be also written in an explicit form for some nonpolyhedral cones; in particular for the cone of positive semidefinite matrices (see [26, section 5.3.6]).

5.1.4 Minimax Stochastic Programs

Sometimes it is worthwhile to consider minimax stochastic programs of the form

$$\underset{x \in \mathcal{X}}{\operatorname{Min}} \sup_{y \in \mathcal{Y}} \left\{ f(x, y) := \mathbb{E}[F(x, y, \xi)] \right\},$$
(5.46)

where $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{Y} \subset \mathbb{R}^m$ are closed sets, $F : \mathcal{X} \times \mathcal{Y} \times \Xi \to \mathbb{R}$ and $\xi = \xi(\omega)$ is a random vector whose probability distribution is supported on set $\Xi \subset \mathbb{R}^d$. The corresponding SAA problem is obtained by using the sample average as an approximation of the expectation f(x, y), that is

$$\operatorname{Min}_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \left\{ \hat{f}_N(x, y) := \frac{1}{N} \sum_{j=1}^N F(x, y, \xi^j) \right\}.$$
(5.47)

As before denote by ϑ^* and $\hat{\vartheta}_N$ the optimal values of (5.46) and (5.47), respectively, and by $S_x \subset \mathcal{X}$ and $\hat{S}_{x,N} \subset \mathcal{X}$ the respective sets of optimal solutions. Recall that $F(x, y, \xi)$ is said to be a *Carathéodory function*, if $F(x, y, \xi(\cdot))$ is measurable for every (x, y) and $F(\cdot, \cdot, \xi)$ is continuous for a.e. $\xi \in \Xi$. We make the following assumptions.

- (A'1) $F(x, y, \xi)$ is a Carathéodory function.
- (A'2) The sets \mathcal{X} and \mathcal{Y} are nonempty and compact.
- (A'3) $F(x, y, \xi)$ is dominated by an integrable function, i.e., there is an open set $N \subset \mathbb{R}^{n+m}$, containing the set $\mathcal{X} \times \mathcal{Y}$, and an integrable, with respect to the probability distribution of the random vector ξ , function $h(\xi)$ such that $|F(x, y, \xi)| \leq h(\xi)$ for all $(x, y) \in N$ and a.e. $\xi \in \Xi$.

By Theorem 7.48 it follows that the expected value function f(x, y) is continuous on $\mathcal{X} \times \mathcal{Y}$. Since Y is compact, this implies that the max-function

$$\phi(x) := \sup_{y \in \mathcal{Y}} f(x, y)$$

is continuous on \mathcal{X} . It also follows that the function $\hat{f}_N(x, y) = \hat{f}_N(x, y, \omega)$ is a Carathéodory function. Consequently, the sample average max-function

$$\hat{\phi}_N(x,\omega) := \sup_{y \in \mathcal{Y}} \hat{f}_N(x,y,\omega)$$

is a Carathéodory function. Since $\hat{\vartheta}_N = \hat{\vartheta}_N(\omega)$ is given by the minimum of the Carathéodory function $\hat{\phi}_N(x,\omega)$, it follows that it is measurable.

Theorem 5.9. Suppose that assumptions (A'1) - (A'3) hold and the sample is iid. Then $\hat{\vartheta}_N \to \vartheta^*$ and $\mathbb{D}(\hat{\mathcal{S}}_{x,N}, \mathcal{S}_x) \to 0$ w.p.1 as $N \to \infty$.

Proof. By Theorem 7.53 we have that, under the specified assumptions, $\hat{f}_N(x, y)$ converges to f(x, y) w.p.1 uniformly on $\mathcal{X} \times \mathcal{Y}$. That is, $\Delta_N \to 0$ w.p.1 as $N \to \infty$, where

$$\Delta_N := \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \left| \hat{f}_N(x,y) - f(x,y) \right|.$$

Consider $\hat{\phi}_N(x) := \sup_{y \in \mathcal{Y}} \hat{f}_N(x, y)$ and $\phi(x) := \sup_{y \in \mathcal{Y}} f(x, y)$. We have that

$$\sup_{x \in \mathcal{X}} \left| \hat{\phi}_N(x) - \phi(x) \right| \le \Delta_N,$$

and hence $|\hat{\vartheta}_N - \vartheta^*| \leq \Delta_N$. It follows that $\hat{\vartheta}_N \to \vartheta^*$ w.p.1.

The function $\phi(x)$ is continuous and $\hat{\phi}_N(x)$ is continuous w.p.1. Consequently, the set S_x is nonempty and $\hat{S}_{x,N}$ is nonempty w.p.1. Now to prove that $\mathbb{D}(\hat{S}_{x,N}, \mathcal{S}_x) \to 0$ w.p.1, one can proceed exactly in the same way as in the proof of Theorem 5.3. \Box

We discuss now asymptotics of $\hat{\vartheta}_N$ in the convex-concave case. We make the following additional assumptions.

(A'4) The sets X and Y are convex, and for a.e. ξ ∈ Ξ the function F(·, ·, ξ) is convexconcave on X × Y, i.e., the function F(·, y, ξ) is convex on X for every y ∈ Y, and the function F(x, ·, ξ) is concave on Y for every x ∈ X.

It follows that the expected value function f(x, y) is convex concave and continuous on $\mathcal{X} \times \mathcal{Y}$. Consequently, problem (5.46) and its dual

$$\underset{y \in \mathcal{Y}}{\operatorname{Max}} \inf_{x \in \mathcal{X}} f(x, y) \tag{5.48}$$

have nonempty and bounded sets of optimal solutions $S_x \subset \mathcal{X}$ and $S_y \subset \mathcal{Y}$, respectively. Moreover, the optimal values of problems (5.46) and (5.48) are equal to each other and $S_x \times S_y$ forms the set of saddle points of these problems.

(A'5) For some point $(x, y) \in \mathcal{X} \times \mathcal{Y}$, the expectation $\mathbb{E}[F(x, y, \xi)^2]$ is finite, and there exists a measurable function $C : \Xi \to \mathbb{R}_+$ such that $\mathbb{E}[C(\xi)^2]$ is finite and the inequality

$$|F(x, y, \xi) - F(x', y', \xi)| \le C(\xi) \big(||x - x'|| + ||y - y'|| \big)$$
(5.49)

holds for all $(x, y), (x', y') \in \mathcal{X} \times \mathcal{Y}$ and a.e. $\xi \in \Xi$.

The above assumption implies that f(x, y) is Lipschitz continuous on $\mathcal{X} \times \mathcal{Y}$ with Lipschitz constant $\kappa = \mathbb{E}[C(\xi)]$.

Theorem 5.10. Consider the minimax stochastic problem (5.46) and the SAA problem (5.47) based on an iid sample. Suppose that assumptions (A'1) - (A'2) and (A'4) - (A'5) hold. Then

$$\hat{\vartheta}_N = \inf_{x \in S_x} \sup_{y \in S_y} \hat{f}_N(x, y) + o_p(N^{-1/2}).$$
(5.50)

Moreover, if the sets $S_x = {\bar{x}}$ and $S_y = {\bar{y}}$ are singletons, then $N^{1/2}(\hat{\vartheta}_N - \vartheta^*)$ converges in distribution to normal with zero mean and variance $\sigma^2 = \operatorname{Var}[F(\bar{x}, \bar{y}, \xi)]$.

Proof. Consider the space $C(\mathcal{X}, \mathcal{Y})$, of continuous functions $\psi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ equipped with the sup-norm $\|\psi\| = \sup_{x \in \mathcal{X}, y \in \mathcal{Y}} |\psi(x, y)|$, and set $\mathcal{K} \subset C(\mathcal{X}, \mathcal{Y})$ formed by convexconcave on $\mathcal{X} \times \mathcal{Y}$ functions. It is not difficult to see that the set \mathcal{K} is a closed (in the norm topology of $C(\mathcal{X}, \mathcal{Y})$) and convex cone. Consider the optimal value function V : $C(\mathcal{X}, \mathcal{Y}) \to \mathbb{R}$ defined as

$$V(\psi) := \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \psi(x, y), \text{ for } \psi \in C(\mathcal{X}, \mathcal{Y}).$$
(5.51)

Recall that it is said that $V(\cdot)$ is Hadamard directionally differentiable at $f \in \mathcal{K}$, tangentially to the set \mathcal{K} , if the following limit exists for any $\gamma \in \mathcal{T}_{\mathcal{K}}(f)$:

$$V'_f(\gamma) := \lim_{\substack{t \downarrow 0, \eta \to \gamma \\ f + \eta \in \mathcal{K}}} \frac{V(f + t\eta) - V(f)}{t}.$$
(5.52)

By Theorem 7.28 we have that the optimal value function $V(\cdot)$ is Hadamard directionally differentiable at f tangentially to the set \mathcal{K} and

$$V'_f(\gamma) = \inf_{x \in \mathcal{S}_x} \sup_{y \in \mathcal{S}_y} \gamma(x, y),$$
(5.53)

for any $\gamma \in \mathcal{T}_{\mathcal{K}}(f)$.

By the assumption (A'5) we have that $N^{1/2}(\hat{f}_N - f)$, considered as a sequence of random elements of C(X, Y), converges in distribution to a random element of $C(\mathcal{X}, \mathcal{Y})$. Then by noting that $\vartheta^* = f(x^*, y^*)$ for any $(x^*, y^*) \in S_x \times S_y$ and using Hadamard directional differentiability of the optimal value function, tangentially to the set \mathcal{K} , together with formula (5.53) and a version of the Delta method given in Theorem 7.69, we can complete the proof. \Box

Suppose now that the feasible set \mathcal{X} is defined by constraints in the form (5.11). The Lagrangian function of the true problem is

$$L(x,\lambda) := f(x) + \sum_{i=1}^{p} \lambda_i g_i(x).$$

Suppose also that the problem is *convex*, that is the set \mathcal{X}_0 is convex and for all $\xi \in \Xi$ the functions $F(\cdot, \xi)$ and $G_i(\cdot, \xi)$, i = 1, ..., p, are convex. Suppose, further, that the functions

f(x) and $g_i(x)$ are finite valued on a neighborhood of the set S (of optimal solutions of the true problem) and the Slater condition holds. Then with every optimal solution $\bar{x} \in S$ is associated a nonempty and bounded set Λ of Lagrange multipliers vectors $\lambda = (\lambda_1, ..., \lambda_p)$ satisfying the optimality conditions:

$$\bar{x} \in \arg\min_{x \in \mathcal{X}_0} L(x, \lambda), \ \lambda_i \ge 0 \ \text{and} \ \lambda_i g_i(\bar{x}) = 0, \ i = 1, ..., p.$$
 (5.54)

The set Λ coincides with the set of optimal solutions of the dual of the true problem, and therefore is the same for any optimal solution $\bar{x} \in S$.

Let $\hat{\vartheta}_N$ be the optimal value of the SAA problem (5.10) with \mathcal{X}_N given in the form (5.13). That is, $\hat{\vartheta}_N$ is the optimal value of the problem

$$\underset{x \in \mathcal{X}_0}{\operatorname{Min}} \hat{f}_N(x) \text{ subject to } \hat{g}_{iN}(x) \le 0, \ i = 1, \dots, p,$$
(5.55)

with $f_N(x)$ and $\hat{g}_{iN}(x)$ being the SAA functions of the respective integrands $F(x,\xi)$ and $G_i(x,\xi)$, i = 1, ..., p. Assume that conditions (A1) and (A2), formulated on page 184, are satisfied for the integrands F and G_i , i = 1, ..., p, i.e., finiteness of the corresponding second order moments and the Lipschitz continuity condition of assumption (A2) hold for each function. It follows that the corresponding expected value functions f(x) and $g_i(x)$ are finite valued and continuous on X. As in Theorem 5.7, we denote by Y(x) random variables which are normally distributed and have the same covariance structure as $F(x,\xi)$. We also denote by $Y_i(x)$ random variables which are normally distributed and have the same covariance structure as $G_i(x,\xi)$, i = 1, ..., p.

Theorem 5.11. Let $\hat{\vartheta}_N$ be the optimal value of the SAA problem (5.55). Suppose that the sample is iid, the problem is convex and the following conditions are satisfied: (i) the set S, of optimal solutions of the true problem, is nonempty and bounded, (ii) the functions f(x) and $g_i(x)$ are finite valued on a neighborhood of S, (iii) the Slater condition for the true problem holds, (iv) the assumptions (A1) and (A2) hold for the integrands F and G_i , i = 1, ..., p. Then

$$N^{1/2}\left(\hat{\vartheta}_N - \vartheta^*\right) \xrightarrow{\mathcal{D}} \inf_{x \in \mathcal{S}} \sup_{\lambda \in \Lambda} \left[Y(x) + \sum_{i=1}^p \lambda_i Y_i(x) \right].$$
(5.56)

If, moreover, $S = \{\bar{x}\}$ and $\Lambda = \{\bar{\lambda}\}$ are singletons, then

$$N^{1/2}\left(\hat{\vartheta}_N - \vartheta^*\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$
 (5.57)

with

$$\sigma^2 := \mathbb{V}\operatorname{ar}\left[F(\bar{x},\xi) + \sum_{i=1}^p \bar{\lambda}_i G_i(\bar{x},\xi)\right].$$
(5.58)

Proof. Since the problem is convex and the Slater condition (for the true problem) holds, we have that ϑ^* is equal to the optimal value of the (Lagrangian) dual

$$\underset{\lambda \ge 0}{\operatorname{Max}} \inf_{x \in \mathcal{X}_0} L(x, \lambda), \tag{5.59}$$

and the set of optimal solutions of (5.59) is nonempty and compact and coincides with the set of Lagrange multipliers Λ . Since the problem is convex and S is nonempty and bounded, the problem can be considered on a bounded neighborhood of S, i.e., without loss of generality it can be assumed that the set X is compact. The proof can be now completed by applying Theorem 5.10. \Box

Remark 12. There are two possible approaches to generating random samples in construction of SAA problems of the form (5.55) by Monte Carlo sampling techniques. One is to use the same sample $\xi^1, ..., \xi^N$ for estimating the functions f(x) and $g_i(x)$, i = 1, ..., p, by their SAA counterparts, or to use independent samples, possibly of different sizes, for each of these functions (see Remark 8 on page 181). The asymptotic results of the above Theorem 5.11 are for the case of the same sample. The (asymptotic) variance σ^2 , given in (5.58), is equal to the sum of the variances of $F(\bar{x}, \xi)$ and $\bar{\lambda}_i G_i(\bar{x}, \xi)$, i = 1, ..., p, and all their covariances. If we use the independent samples construction, then a similar result holds but without the corresponding covariance terms. Since in the case of the same sample these covariance terms could be expected to be positive, it would be advantageous to use the independent, rather than the same, samples approach in order to reduce variability of the SAA estimates.

5.2 Stochastic Generalized Equations

In this section we discuss the following so-called *stochastic generalized equations*. Consider a random vector ξ , whose distribution is supported on a set $\Xi \subset \mathbb{R}^d$, a mapping $\Phi : \mathbb{R}^n \times \Xi \to \mathbb{R}^n$ and a multifunction $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. Suppose that the expectation $\phi(x) := \mathbb{E}[\Phi(x, \xi)]$ is well defined and finite valued. We refer to

$$\phi(x) \in \Gamma(x) \tag{5.60}$$

as true, or expected value, generalized equation and say that a point $\bar{x} \in \mathbb{R}^n$ is a solution of (5.60) if $\phi(\bar{x}) \in \Gamma(\bar{x})$.

The above abstract setting includes the following cases. If $\Gamma(x) = \{0\}$ for every $x \in \mathbb{R}^n$, then (5.60) becomes the ordinary equation $\phi(x) = 0$. As another example, let $\Gamma(\cdot) := \mathcal{N}_{\mathcal{X}}(\cdot)$, where \mathcal{X} is a nonempty closed convex subset of \mathbb{R}^n and $\mathcal{N}_{\mathcal{X}}(x)$ denotes the (outwards) normal cone to \mathcal{X} at x. Recall that, by the definition, $\mathcal{N}_{\mathcal{X}}(x) = \emptyset$ if $x \notin \mathcal{X}$. In that case \bar{x} is a solution of (5.60) iff $\bar{x} \in \mathcal{X}$ and the following, so-called *variational inequality*, holds

$$(x - \bar{x})^{\mathsf{T}} \phi(\bar{x}) \le 0, \quad \forall x \in \mathcal{X}.$$
 (5.61)

Since the mapping $\phi(x)$ is given in the form of the expectation, we refer to such variational inequalities as *stochastic variational inequalities*. Note that if $\mathcal{X} = \mathbb{R}^n$, then $\mathcal{N}_{\mathcal{X}}(x) = \{0\}$ for any $x \in \mathbb{R}^n$, and hence in that case the above variational inequality is reduced to the equation $\phi(x) = 0$. Let us also remark that if $\Phi(x,\xi) := -\nabla_x F(x,\xi)$, for some real valued function $F(x,\xi)$, and the interchangeability formula $\mathbb{E}[\nabla_x F(x,\xi)] = \nabla f(x)$ holds, i.e., $\phi(x) = -\nabla f(x)$ where $f(x) := \mathbb{E}[F(x,\xi)]$, then (5.61) represents first order necessary, and if f(x) is convex, sufficient conditions for \bar{x} to be an optimal solution for the optimization problem (5.1).

If the feasible set \mathcal{X} of the optimization problem (5.1) is defined by constraints in the form

$$\mathcal{X} := \{ x \in \mathbb{R}^n : g_i(x) = 0, \ i = 1, ..., q, \ g_i(x) \le 0, \ i = q + 1, ..., p \},$$
(5.62)

with $g_i(x) := \mathbb{E}[G_i(x,\xi)], i = 1, ..., p$, then the corresponding first-order (KKT) optimality conditions can be written in a form of variational inequality. That is, let $z := (x, \lambda) \in \mathbb{R}^{n+p}$ and

$$\begin{array}{lll} L(z,\xi) & := & F(x,\xi) + \sum_{i=1}^{p} \lambda_i G_i(x,\xi), \\ \ell(z) & := & \mathbb{E}[L(z,\xi)] = f(x) + \sum_{i=1}^{p} \lambda_i g_i(x) \end{array}$$

be the corresponding Lagrangians. Define

$$\Phi(z,\xi) := \begin{bmatrix} \nabla_x L(z,\xi) \\ G_1(x,\xi) \\ \vdots \\ G_p(x,\xi) \end{bmatrix} \text{ and } \Gamma(z) := \mathcal{N}_K(z),$$
(5.63)

where $K := \mathbb{R}^n \times \mathbb{R}^q \times \mathbb{R}^{p-q}_+ \subset \mathbb{R}^{n+p}$. Note that if $z \in K$, then

$$\mathcal{N}_{K}(z) = \left\{ (v,\gamma) \in \mathbb{R}^{n+p} : \begin{array}{l} v = 0 \text{ and } \gamma_{i} = 0, \ i = 1, \dots, q, \\ \gamma_{i} = 0, \ i \in \mathcal{I}_{+}(\lambda), \ \gamma_{i} \le 0, \ i \in \mathcal{I}_{0}(\lambda) \end{array} \right\},$$
(5.64)

where

$$\begin{aligned} \mathcal{I}_0(\lambda) &:= \{i : \lambda_i = 0, \ i = q + 1, ..., p\}, \\ \mathcal{I}_+(\lambda) &:= \{i : \lambda_i > 0, \ i = q + 1, ..., p\}, \end{aligned}$$
 (5.65)

and $\mathcal{N}_K(z) = \emptyset$ if $z \notin K$. Consequently, assuming that the interchangeability formula holds, and hence $\mathbb{E}[\nabla_x L(z,\xi)] = \nabla \ell_x(z) = \nabla f(x) + \sum_{i=1}^p \lambda_i \nabla g_i(x)$, we have that

$$\phi(z) := \mathbb{E}[\Phi(z,\xi)] = \begin{bmatrix} \nabla_x \ell(z) \\ g_1(x) \\ \vdots \\ g_p(x) \end{bmatrix}$$
(5.66)

and variational inequality $\phi(z) \in \mathcal{N}_K(z)$ represents the KKT optimality conditions for the true optimization problem.

We make the following assumption about the multifunction $\Gamma(x)$.

(E1) The multifunction $\Gamma(x)$ is *closed*, that is, the following holds: if $x_k \to x$, $y_k \in \Gamma(x_k)$ and $y_k \to y$, then $y \in \Gamma(x)$.

The above assumption implies that the multifunction $\Gamma(x)$ is closed valued, i.e., for any $x \in \mathbb{R}^n$ the set $\Gamma(x)$ is closed. For variational inequalities assumption (E1) always holds, i.e., the multifunction $x \mapsto \mathcal{N}_{\mathcal{X}}(x)$ is closed.

Now let ξ^1, \dots, ξ^N be a random sample of N realizations of the random vector ξ , and $\hat{\phi}_N(x) := N^{-1} \sum_{j=1}^N \Phi(x, \xi^j)$ be the corresponding sample average estimate of $\phi(x)$. We refer to $\hat{\phi}$

$$b_N(x) \in \Gamma(x) \tag{5.67}$$

as the SAA generalized equation. There are standard numerical algorithms for solving nonlinear equations which can be applied to (5.67) in the case $\Gamma(x) \equiv \{0\}$, i.e., when (5.67) is reduced to the ordinary equation $\hat{\phi}_N(x) = 0$. There are also numerical procedures for solving variational inequalities. We are not going to discuss such numerical algorithms but rather concentrate on statistical properties of solutions of SAA equations. We denote by Sand \hat{S}_N the sets of (all) solutions of the true (5.60) and SAA (5.67) generalized equations, respectively.

5.2.1 Consistency of Solutions of the SAA Generalized Equations

In this section we discuss convergence properties of the SAA solutions.

Theorem 5.12. Let C be a compact subset of \mathbb{R}^n such that $S \subset C$. Suppose that: (i) the multifunction $\Gamma(x)$ is closed (assumption (E1)), (ii) the mapping $\phi(x)$ is continuous on C, (iii) w.p.1 for N large enough the set \hat{S}_N is nonempty and $\hat{S}_N \subset C$, (iv) $\hat{\phi}_N(x)$ converges to $\phi(x)$ w.p.1 uniformly on C as $N \to \infty$. Then $\mathbb{D}(\hat{S}_N, S) \to 0$ w.p.1 as $N \to \infty$.

Proof. The above result basically is deterministic in the sense that if we view $\hat{\phi}_N(x) = \hat{\phi}_N(x,\omega)$ as defined on a common probability space, then it should be verified for a.e. ω . Therefore we omit saying "w.p.1". Consider a sequence $\hat{x}_N \in \hat{S}_N$. Because of the assumption (iii), by passing to a subsequence if necessary, we only need to show that if \hat{x}_N converges to a point x^* , then $x^* \in S$ (compare with the proof of Theorem 5.3). Now since it is assumed that $\phi(\cdot)$ is continuous and $\hat{\phi}_N(x)$ converges to $\phi(x)$ uniformly, it follows that $\hat{\phi}_N(\hat{x}_N) \to \phi(x^*)$ (see Proposition 5.1). Since $\hat{\phi}_N(\hat{x}_N) \in \Gamma(\hat{x}_N)$, it follows by assumption (E1) that $\phi(x^*) \in \Gamma(x^*)$, which completes the proof. \Box

A few remarks about the assumptions involved in the above consistency result are now in order. By Theorem 7.53 we have that, in the case of iid sampling, the assumptions (ii) and (iv) of the above proposition are satisfied for any compact set C if the following assumption holds.

(E2) For every ξ ∈ Ξ the function Φ(·, ξ) is continuous on C and ||Φ(x, ξ)||_{x∈C} is dominated by an integrable function.

There are two parts to the assumption (iii) of Theorem 5.12, namely, that the SAA generalized equations do not have a solution which escapes to infinity, and that they possess at least one solution w.p.1 for N large enough. The first of these assumptions can be often verified by ad hoc methods. The second assumption is more subtle. We are going to discuss it next. The following concept of strong regularity is due to Robinson [206].

Definition 5.13. Suppose that the mapping $\phi(x)$ is continuously differentiable. We say that a solution $\bar{x} \in S$ is strongly regular if there exist neighborhoods \mathcal{N}_1 and \mathcal{N}_2 of $0 \in \mathbb{R}^n$ and \bar{x} , respectively, such that for every $\delta \in \mathcal{N}_1$ the following (linearized) generalized equation

$$\delta + \phi(\bar{x}) + \nabla \phi(\bar{x})(x - \bar{x}) \in \Gamma(x)$$
(5.68)

has a unique solution in \mathcal{N}_2 , denoted $\tilde{x} = \tilde{x}(\delta)$, and $\tilde{x}(\cdot)$ is Lipschitz continuous on \mathcal{N}_1 .

Note that it follows from the above conditions that $\tilde{x}(0) = \bar{x}$. In case $\Gamma(x) \equiv \{0\}$, strong regularity simply means that the $n \times n$ Jacobian matrix $J := \nabla \phi(\bar{x})$ is invertible, or in other words nonsingular. Also in the case of variational inequalities, the strong regularity condition was investigated extensively, we discuss this later.

Let \mathcal{V} be a convex compact neighborhood of \bar{x} , i.e., $\bar{x} \in \text{int}(\mathcal{V})$. Consider the space $C^1(\mathcal{V}, \mathbb{R}^n)$ of continuously differentiable mappings $\psi : \mathcal{V} \to \mathbb{R}^n$ equipped with the norm:

$$\|\psi\|_{1,\mathcal{V}} := \sup_{x\in\mathcal{V}} \|\phi(x)\| + \sup_{x\in\mathcal{V}} \|\nabla\phi(x)\|.$$

The following (deterministic) result is essentially due to Robinson [207].

Suppose that $\phi(x)$ is continuously differentiable on \mathcal{V} , i.e., $\phi \in C^1(\mathcal{V}, \mathbb{R}^n)$. Let \bar{x} be a strongly regular solution of the generalized equation (5.60). Then there exists $\varepsilon > 0$ such that for any $u \in C^1(\mathcal{V}, \mathbb{R}^n)$ satisfying $||u - \phi||_{1,\mathcal{V}} \leq \varepsilon$, the generalized equation $u(x) \in \Gamma(x)$ has a unique solution $\hat{x} = \hat{x}(u)$ in a neighborhood of \bar{x} , such that $\hat{x}(\cdot)$ is Lipschitz continuous (with respect the norm $||\cdot||_{1,\mathcal{V}}$), and

$$\hat{x}(u) = \tilde{x}(u(\bar{x}) - \phi(\bar{x})) + o(||u - \phi||_{1,\mathcal{V}}).$$
(5.69)

Clearly, we have that $\hat{x}(\phi) = \bar{x}$ and $\hat{x}(\hat{\phi}_N)$ is a solution, in a neighborhood of \bar{x} , of the SAA generalized equation provided that $\|\hat{\phi}_N - \phi\|_{1,\mathcal{V}} \leq \varepsilon$. Therefore, by employing the above results for the mapping $u(\cdot) := \hat{\phi}_N(\cdot)$ we immediately obtain the following.

Theorem 5.14. Let \bar{x} be a strongly regular solution of the true generalized equation (5.60), and suppose that $\phi(x)$ and $\hat{\phi}_N(x)$ are continuously differentiable in a neighborhood \mathcal{V} of \bar{x} and $\|\hat{\phi}_N - \phi\|_{1,\mathcal{V}} \to 0$ w.p.1 as $N \to \infty$. Then w.p.1 for N large enough the SAA generalized equation (5.67) possesses a unique solution \hat{x}_N in a neighborhood of \bar{x} , and $\hat{x}_N \to \bar{x}$ w.p.1 as $N \to \infty$.

The assumption that $\|\hat{\phi}_N - \phi\|_{1,\mathcal{V}} \to 0$ w.p.1, in the above theorem, means that $\hat{\phi}_N(x)$ and $\nabla \hat{\phi}_N(x)$ converge w.p.1 to $\phi(x)$ and $\nabla \phi(x)$, respectively, uniformly on \mathcal{V} . By Theorem 7.53, in the case of iid sampling this is ensured by the following assumption.

(E3) For a.e. ξ the mapping Φ(·, ξ) is continuously differentiable on V, and ||Φ(x, ξ)||_{x∈V} and ||∇_xΦ(x, ξ)||_{x∈V} are dominated by an integrable function.

Note that the assumption that $\Phi(\cdot,\xi)$ is continuously differentiable on a neighborhood of \bar{x} is essential in the above analysis. By combining Theorems 5.12 and 5.14 we obtain the following result.

Theorem 5.15. Let C be a compact subset of \mathbb{R}^n and \bar{x} be a unique in C solution of the true generalized equation (5.60). Suppose that: (i) the multifunction $\Gamma(x)$ is closed (assumption (E1)), (ii) for a.e. ξ the mapping $\Phi(\cdot, \xi)$ is continuously differentiable on C, and $\|\Phi(x,\xi)\|_{x\in C}$ and $\|\nabla_x \Phi(x,\xi)\|_{x\in C}$ are dominated by an integrable function, (iii) the solution \bar{x} is strongly regular, (iv) $\hat{\phi}_N(x)$ and $\nabla \hat{\phi}_N(x)$ converge w.p.1 to $\phi(x)$ and $\nabla \phi(x)$, respectively, uniformly on C. Then w.p.1 for N large enough the SAA generalized equation possesses unique in C solution \hat{x}_N converging to \bar{x} w.p.1 as $N \to \infty$.

Note again that if the sample is iid, then the assumption (iv) in the above theorem is implied by the assumption (ii) and hence is redundant.

5.2.2 Asymptotics of SAA Generalized Equations Estimators

By using the first order approximation (5.69) it is also possible to derive asymptotics of \hat{x}_N . Suppose for the moment that $\Gamma(x) \equiv \{0\}$. Then strong regularity means that the Jacobian matrix $J := \nabla \phi(\bar{x})$ is nonsingular, and $\tilde{x}(\delta)$ is the solution of the corresponding linear equations and hence can be written in the form

$$\tilde{x}(\delta) = \bar{x} - J^{-1}\delta. \tag{5.70}$$

By using (5.70) and (5.69) with $u(\cdot) := \hat{\phi}_N(\cdot)$, we obtain under certain regularity conditions which ensure that the remainder in (5.69) is of order $o_p(N^{-1/2})$, that

$$N^{1/2}(\hat{x}_N - \bar{x}) = -J^{-1}Y_N + o_p(1),$$
(5.71)

where $Y_N := N^{1/2} \left[\hat{\phi}_N(\bar{x}) - \phi(\bar{x}) \right]$. Moreover, in the case of iid sample, we have by the CLT that $Y_N \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma)$, where Σ is the covariance matrix of the random vector $\Phi(\bar{x}, \xi)$. Consequently, \hat{x}_N has asymptotically normal distribution with mean vector \bar{x} and the covariance matrix $N^{-1}J^{-1}\Sigma J^{-1}$.

Suppose now that $\Gamma(\cdot) := \mathcal{N}_{\mathcal{X}}(\cdot)$, with the set \mathcal{X} being nonempty closed convex and *polyhedral*, and let \bar{x} be a strongly regular solution of (5.60). Let $\tilde{x}(\delta)$ be the (unique) solution, of the corresponding linearized variational inequality (5.68), in a neighborhood of \bar{x} . Consider the cone

$$\mathcal{C}_{\mathcal{X}}(\bar{x}) := \left\{ y \in \mathcal{T}_{\mathcal{X}}(\bar{x}) : y^{\mathsf{T}} \phi(\bar{x}) = 0 \right\},$$
(5.72)

called the *critical cone*, and the Jacobian matrix $J := \nabla \phi(\bar{x})$. Then for all δ sufficiently close to $0 \in \mathbb{R}^n$, we have that $\tilde{x}(\delta) - \bar{x}$ coincides with the solution $\tilde{d}(\delta)$ of the variational inequality

$$\delta + Jd \in \mathcal{N}_{\mathcal{C}_{\mathcal{X}}(\bar{x})}(d). \tag{5.73}$$

Note that the mapping $\tilde{d}(\cdot)$ is positively homogeneous, i.e., for any $\delta \in \mathbb{R}^n$ and $t \ge 0$, it follows that $\tilde{d}(t\delta) = t\tilde{d}(\delta)$. Consequently, under the assumption that the solution \bar{x} is strongly regular, we obtain by (5.69) that $\tilde{d}(\cdot)$ is the directional derivative of $\hat{x}(u)$, at $u = \phi$, in the Hadamard sense. Therefore, under appropriate regularity conditions ensuring functional CLT for $N^{1/2}(\hat{\phi}_N - \phi)$ in the space $C^1(\mathcal{V}, \mathbb{R}^n)$, it follows by the Delta Theorem that

$$N^{1/2}(\hat{x}_N - \bar{x}) \xrightarrow{\mathcal{D}} \tilde{d}(Y), \tag{5.74}$$

where $Y \sim \mathcal{N}(0, \Sigma)$ and Σ is the covariance matrix of $\Phi(\bar{x}, \xi)$. Consequently, \hat{x}_N is asymptotically normal iff the mapping $\tilde{d}(\cdot)$ is linear. This, in turn, holds if the cone $\mathcal{C}_{\mathcal{X}}(\bar{x})$ is a linear space.

In the case $\Gamma(\cdot) := \mathcal{N}_{\mathcal{X}}(\cdot)$, with the set \mathcal{X} being nonempty closed convex and polyhedral, there is a complete characterization of the strong regularity in terms of the so-called *coherent orientation* associated with the matrix (mapping) $J := \nabla \phi(\bar{x})$ and the critical cone $\mathcal{C}_{\mathcal{X}}(\bar{x})$. The interested reader is referred to [208],[93] for a discussion of this topic. Let us just remark that if $\mathcal{C}_{\mathcal{X}}(\bar{x})$ is a linear subspace of \mathbb{R}^n , then the variational inequality (5.73) can be written in the form

$$P\delta + PJd = 0, (5.75)$$

where P denotes the orthogonal projection matrix onto the linear space $C_{\mathcal{X}}(\bar{x})$. Then \bar{x} is strongly regular iff the matrix (mapping) PJ restricted to the linear space $C_{\mathcal{X}}(\bar{x})$ is invertible, or in other words nonsingular.

Suppose now that $S = \{\bar{x}\}$ is such that $\phi(\bar{x})$ belongs to the interior of the set $\Gamma(\bar{x})$. Then, since $\hat{\phi}_N(\bar{x})$ converges w.p.1 to $\phi(\bar{x})$, it follows that the event " $\hat{\phi}_N(\bar{x}) \in \Gamma(\bar{x})$ " happens w.p.1 for N large enough. Moreover, by the LD principle (see (7.216)) we have that this event happens with probability approaching one exponentially fast. Of course, $\hat{\phi}_N(\bar{x}) \in \Gamma(\bar{x})$ means that $\hat{x}_N = \bar{x}$ is a solution of the SAA generalized equation (5.67). Therefore, in such case one may compute an exact solution of the true problem (5.60), by solving the SAA problem, with probability approaching one exponentially fast with increase of the sample size. Note that if $\Gamma(\cdot) := \mathcal{N}_{\mathcal{X}}(\cdot)$ and $\bar{x} \in S$, then $\phi(\bar{x}) \in \operatorname{int} \Gamma(\bar{x})$ iff the critical cone $\mathcal{C}_{\mathcal{X}}(\bar{x})$ is equal to $\{0\}$. In that case the variational inequality (5.73) has solution $\bar{d} = 0$ for any δ , i.e., $\tilde{d}(\delta) \equiv 0$.

The above asymptotics can be applied, in particular, to the generalized equation (variational inequality) $\phi(z) \in \mathcal{N}_K(z)$, where $K := \mathbb{R}^n \times \mathbb{R}^q \times \mathbb{R}^{p-q}$ and $\mathcal{N}_K(z)$ and $\phi(z)$ are given in (5.64) and (5.66), respectively. Recall that this variational inequality represents the KKT optimality conditions of the expected value optimization problem (5.1) with the feasible set \mathcal{X} given in the form (5.62). (We assume that the expectation functions f(x) and $g_i(x)$, i = 1, ..., p, are continuously differentiable.) Let \bar{x} be an optimal solution of the (expected value) problem (5.1). It is said that the *linear independence constraint qualification* (LICQ) holds at the point \bar{x} if the gradient vectors $\nabla g_i(\bar{x})$, $i \in \{i : g_i(\bar{x}) = 0, i = 1, ..., p\}$, (of active at \bar{x} constraints) are linearly independent. Under the LICQ, to \bar{x} corresponds a unique vector $\bar{\lambda}$ of Lagrange multipliers, satisfying the KKT optimality conditions. Let $\bar{z} = (\bar{x}, \bar{\lambda})$ and $\mathcal{I}_0(\lambda)$ and $\mathcal{I}_+(\lambda)$ be the index sets defined in (5.65). Then

$$\mathcal{T}_{K}(\bar{z}) = \mathbb{R}^{n} \times \mathbb{R}^{q} \times \left\{ \gamma \in \mathbb{R}^{p-q} : \gamma_{i} \ge 0, \ i \in \mathcal{I}_{0}(\bar{\lambda}) \right\}.$$
(5.76)

In order to simplify notation let us assume that *all* constraints are active at \bar{x} , i.e., $g_i(\bar{x}) = 0$, i = 1, ..., p. Since for sufficiently small perturbations of x, inactive constraints remain inactive, we do not lose generality in the asymptotic analysis by considering only active at \bar{x} constraints. Then $\phi(\bar{z}) = 0$, and hence $C_K(\bar{z}) = \mathcal{T}_K(\bar{z})$.

Assuming, further, that f(x) and $g_i(x)$, i = 1, ..., p are twice continuously differentiable, we have that the following second order necessary conditions hold at \bar{x} :

$$h^{\mathsf{T}} \nabla_{xx}^2 \ell(\bar{z}) h \ge 0, \quad \forall h \in C_{\mathcal{X}}(\bar{x}), \tag{5.77}$$

where

$$C_{\mathcal{X}}(\bar{x}) := \left\{ h : h^{\mathsf{T}} \nabla g_i(\bar{x}) = 0, \ i \in \{1, ..., q\} \cup \mathcal{I}_+(\bar{\lambda}), \ h^{\mathsf{T}} \nabla g_i(\bar{x}) \le 0, \ i \in \mathcal{I}_0(\bar{\lambda}) \right\}.$$

The corresponding second order sufficient conditions are

$$h^{\mathsf{T}} \nabla^2_{xx} \ell(\bar{z}) h > 0, \quad \forall h \in C_{\mathcal{X}}(\bar{x}) \setminus \{0\}.$$
(5.78)

Moreover, \bar{z} is a strongly regular solution of the corresponding generalized equation iff the LICQ holds at \bar{x} and the following (strong) form of second order sufficient conditions is satisfied:

$$h^{\mathsf{T}} \nabla^2_{xx} \ell(\bar{z}) h > 0, \quad \forall h \in \lim(C_{\mathcal{X}}(\bar{x})) \setminus \{0\},$$
(5.79)

where

$$\ln(C_{\mathcal{X}}(\bar{x})) := \left\{ h : h^{\mathsf{T}} \nabla g_i(\bar{x}) = 0, \ i \in \{1, ..., q\} \cup \mathcal{I}_+(\bar{\lambda}) \right\}.$$
 (5.80)

Under the LICQ, the set defined in the right hand side of (5.80) is, indeed, the linear space generated by the cone $C_{\mathcal{X}}(\bar{x})$. We also have here

$$J := \nabla \phi(\bar{z}) = \begin{bmatrix} H & A \\ A^{\mathsf{T}} & 0 \end{bmatrix},$$
(5.81)

where $H := \nabla_{xx}^2 \ell(\bar{z})$ and $A := [\nabla g_1(\bar{x}), ..., \nabla g_p(\bar{x})].$

It is said that the *strict complementarity condition* holds at \bar{x} if the index set $\mathcal{I}_0(\bar{\lambda})$ is empty, i.e., all Lagrange multipliers corresponding to active at \bar{x} inequality constraints are strictly positive. We have here that $\mathcal{C}_K(\bar{z})$ is a linear space, and hence the SAA estimator $\hat{z}_N = \begin{bmatrix} \hat{x}_N \\ \hat{\lambda}_N \end{bmatrix}$ is asymptotically normal, iff the strict complementarity condition holds. If the strict complementarity condition holds, then $\mathcal{C}_K(\bar{z}) = \mathbb{R}^{n+p}$ (recall that it is assumed that all constraints are active at \bar{x}), and hence the normal cone to $\mathcal{C}_K(\bar{z})$, at every point, is $\{0\}$. Consequently, the corresponding variational inequality (5.73) takes the form $\delta + Jd = 0$. Under the strict complementarity condition, \bar{z} is strongly regular iff the matrix J is nonsingular. It follows that under the above assumptions together with the strict complementarity condition, the following asymptotics hold (compare with (5.45))

$$N^{1/2}\left(\hat{z}_N - \bar{z}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, J^{-1} \Sigma J^{-1}\right), \qquad (5.82)$$

where Σ is the covariance matrix of the random vector $\Phi(\bar{z}, \xi)$ defined in (5.63).

5.3 Monte Carlo Sampling Methods

In this section we assume that a random sample $\xi^1, ..., \xi^N$, of N realizations of the random vector ξ , can be generated in the computer. In the Monte Carlo sampling method this is accomplished by generating a sequence $U^1, U^2, ...$, of independent random (or rather pseudorandom) numbers uniformly distributed on the interval [0,1], and then constructing the sample by an appropriate transformation. In that way we can consider the sequence $\omega := \{U^1, U^2, ...\}$ as an element of the probability space equipped with the corresponding product probability measure, and the sample $\xi^j = \xi^j(\omega), i = 1, 2, ...$, as a function of ω . Since computer is a finite deterministic machine, sooner or later the generated sample will start to repeat itself. However, modern random numbers generators have a very large cycle period and this method was tested in numerous applications. We view now the corresponding SAA problem (5.2) as a way of *approximating* the true problem (5.1) while drastically reducing the number of generated scenarios. For a statistical analysis of the constructed SAA problems a particular numerical algorithm applied to solve these problems is irrelevant.

Let us also remark that values of the sample average function $f_N(x)$ can be computed in two somewhat different ways. The generated sample $\xi^1, ..., \xi^N$ can be stored in the computer memory, and called every time a new value (at a different point x) of the sample average function should be computed. Alternatively, the same sample can be generated by using a common seed number in an employed pseudorandom numbers generator (this is why this approach is called the *common random number generation* method).

The idea of common random number generation is well known in simulation. That is, suppose that we want to compare values of the objective function at two points $x_1, x_2 \in \mathcal{X}$. In that case we are interested in the difference $f(x_1) - f(x_2)$ rather than in the individual values $f(x_1)$ and $f(x_2)$. If we use sample average estimates $\hat{f}_N(x_1)$ and $\hat{f}_N(x_2)$ based on *independent* samples, both of size N, then $\hat{f}_N(x_1)$ and $\hat{f}_N(x_2)$ are uncorrelated and

$$\mathbb{V}\mathrm{ar}[\hat{f}_{N}(x_{1}) - \hat{f}_{N}(x_{2})] = \mathbb{V}\mathrm{ar}[\hat{f}_{N}(x_{1})] + \mathbb{V}\mathrm{ar}[\hat{f}_{N}(x_{2})].$$
(5.83)

On the other hand, if we use the *same* sample for the estimators $\hat{f}_N(x_1)$ and $\hat{f}_N(x_2)$, then

$$\operatorname{Var}[\hat{f}_{N}(x_{1}) - \hat{f}_{N}(x_{2})] = \operatorname{Var}[\hat{f}_{N}(x_{1})] + \operatorname{Var}[\hat{f}_{N}(x_{2})] - 2\operatorname{Cov}(\hat{f}_{N}(x_{1}), \hat{f}_{N}(x_{2})).$$
(5.84)

In both cases, $\hat{f}_N(x_1) - \hat{f}_N(x_2)$ is an unbiased estimator of $f(x_1) - f(x_2)$. However, in the case of the same sample the estimators $\hat{f}_N(x_1)$ and $\hat{f}_N(x_2)$ tend to be positively correlated with each other, in which case the variance in (5.84) is smaller than the one in (5.83). The difference between the independent and the common random number generated estimators of $f(x_1) - f(x_2)$ can be especially dramatic when the points x_1 and x_2 are close to each other and hence the common random number generated estimators are highly positively correlated.

By the results of section 5.1.1 we have that, under mild regularity conditions, the optimal value and optimal solutions of the SAA problem (5.2) converge w.p.1, as the sample size increases, to their true counterparts. These results, however, do not give any indication of quality of solutions for a given sample of size N. In the next section we discuss *exponential* rates of convergence of optimal and nearly optimal solutions of the SAA problem (5.2). This allows to give an estimate of the sample size which is required to solve the true problem with a given accuracy by solving the SAA problem. Although such estimates of the sample size typically are too conservative for a practical use, they give an insight into the *complexity* of solving the true (expected value) problem.

Unless stated otherwise, we assume in this section that the random sample $\xi^1, ..., \xi^N$ is iid, and make the following assumption:

(M1) The expectation function f(x) is well defined and finite valued for all $x \in \mathcal{X}$.

For $\varepsilon \geq 0$ we denote by

$$\mathcal{S}^{\varepsilon} := \left\{ x \in \mathcal{X} : f(x) \le \vartheta^* + \varepsilon \right\} \text{ and } \hat{\mathcal{S}}_N^{\varepsilon} := \left\{ x \in \mathcal{X} : \hat{f}_N(x) \le \hat{\vartheta}_N + \varepsilon \right\}$$

the sets of ε -optimal solutions of the true and the SAA problems, respectively.

5.3.1 Exponential Rates of Convergence and Sample Size Estimates in Case of Finite Feasible Set

In this section we assume that the feasible set \mathcal{X} is finite, although its cardinality $|\mathcal{X}|$ can be very large. Since \mathcal{X} is finite, the sets $\mathcal{S}^{\varepsilon}$ and $\hat{\mathcal{S}}_{N}^{\varepsilon}$ are nonempty and finite. For parameters $\varepsilon \geq 0$ and $\delta \in [0, \varepsilon]$, consider the event $\{\hat{\mathcal{S}}_{N}^{\delta} \subset \mathcal{S}^{\varepsilon}\}$. This event means that any δ -optimal solution of the SAA problem is an ε -optimal solution of the true problem. We estimate now the probability of that event.

We can write

$$\left\{\hat{\mathcal{S}}_{N}^{\delta} \not\subset \mathcal{S}^{\varepsilon}\right\} = \bigcup_{x \in \mathcal{X} \setminus \mathcal{S}^{\varepsilon}} \bigcap_{y \in \mathcal{X}} \left\{\hat{f}_{N}(x) \leq \hat{f}_{N}(y) + \delta\right\},\tag{5.85}$$

and hence

$$\Pr\left(\hat{\mathcal{S}}_{N}^{\delta} \not\subset \mathcal{S}^{\varepsilon}\right) \leq \sum_{x \in \mathcal{X} \setminus \mathcal{S}^{\varepsilon}} \Pr\left(\bigcap_{y \in \mathcal{X}} \left\{\hat{f}_{N}(x) \leq \hat{f}_{N}(y) + \delta\right\}\right).$$
(5.86)

Consider a mapping $u : \mathcal{X} \setminus S^{\varepsilon} \to \mathcal{X}$. If the set $X \setminus S^{\varepsilon}$ is empty, then any feasible point $x \in \mathcal{X}$ is an ε -optimal solution of the true problem. Therefore we assume that this set is nonempty. It follows from (5.86) that

$$\Pr\left(\hat{\mathcal{S}}_{N}^{\delta} \not\subset \mathcal{S}^{\varepsilon}\right) \leq \sum_{x \in \mathcal{X} \setminus \mathcal{S}^{\varepsilon}} \Pr\left\{\hat{f}_{N}(x) - \hat{f}_{N}(u(x)) \leq \delta\right\}.$$
(5.87)

We assume that the mapping $u(\cdot)$ is chosen in such a way that

$$f(u(x)) \le f(x) - \varepsilon^* \text{ for all } x \in \mathcal{X} \setminus \mathcal{S}^{\varepsilon},$$
(5.88)

and for some $\varepsilon^* \ge \varepsilon$. Note that such a mapping always exists. For example, if we use a mapping $u : \mathcal{X} \setminus S^{\varepsilon} \to S$, then (5.88) holds with

$$\varepsilon^* := \min_{x \in \mathcal{X} \setminus \mathcal{S}^{\varepsilon}} f(x) - \vartheta^*, \tag{5.89}$$

and that $\varepsilon^* > \varepsilon$ since the set \mathcal{X} is finite. Different choices of $u(\cdot)$ give a certain flexibility to the following derivations.

For each $x \in \mathcal{X} \setminus \mathcal{S}^{\varepsilon}$, define

$$Y(x,\xi) := F(u(x),\xi) - F(x,\xi).$$
(5.90)

Note that $\mathbb{E}[Y(x,\xi)] = f(u(x)) - f(x)$, and hence $\mathbb{E}[Y(x,\xi)] \leq -\varepsilon^*$ for all $x \in \mathcal{X} \setminus S^{\varepsilon}$. The corresponding sample average is

$$\hat{Y}_N(x) := \frac{1}{N} \sum_{j=1}^N Y(x,\xi^j) = \hat{f}_N(u(x)) - \hat{f}_N(x).$$

By (5.87) we have

$$\Pr\left(\hat{\mathcal{S}}_{N}^{\delta} \not\subset \mathcal{S}^{\varepsilon}\right) \leq \sum_{x \in \mathcal{X} \setminus \mathcal{S}^{\varepsilon}} \Pr\left\{\hat{Y}_{N}(x) \geq -\delta\right\}.$$
(5.91)

Let $I_x(\cdot)$ denote the (large deviations) rate function of the random variable $Y(x,\xi)$. The inequality (5.91) together with the LD upper bound (7.198) implies

$$1 - \Pr\left(\hat{\mathcal{S}}_{N}^{\delta} \subset \mathcal{S}^{\varepsilon}\right) \leq \sum_{x \in \mathcal{X} \setminus \mathcal{S}^{\varepsilon}} e^{-NI_{x}(-\delta)}.$$
(5.92)

Note that the above inequality (5.92) is valid for any random sample of size N. Let us make the following assumption.

(M2) For every $x \in \mathcal{X} \setminus S^{\varepsilon}$ the moment generating function $\mathbb{E}\left[e^{tY(x,\xi)}\right]$, of the random variable $Y(x,\xi) = F(u(x),\xi) - F(x,\xi)$, is finite valued in a neighborhood of t = 0.

The above assumption (M2) holds, for example, if the support Ξ of ξ is a bounded subset of \mathbb{R}^d , or if $Y(x, \cdot)$ grows at most linearly and ξ has a distribution from an exponential family.

Theorem 5.16. Let ε and δ be nonnegative numbers. Then

$$1 - \Pr(\hat{\mathcal{S}}_N^{\delta} \subset \mathcal{S}^{\varepsilon}) \le |\mathcal{X}| e^{-N\eta(\delta,\varepsilon)}, \tag{5.93}$$

where

$$\eta(\delta,\varepsilon) := \min_{x \in \mathcal{X} \setminus S^{\varepsilon}} I_x(-\delta).$$
(5.94)

Moreover, if $\delta < \varepsilon^*$ *and assumption* (M2) *holds, then* $\eta(\delta, \varepsilon) > 0$ *.*

Proof. The inequality (5.93) is an immediate consequence of the inequality (5.92). If $\delta < \varepsilon^*$, then $-\delta > -\varepsilon^* \ge \mathbb{E}[Y(x,\xi)]$, and hence it follows by assumption (M2) that $I_x(-\delta) > 0$ for every $x \in \mathcal{X} \setminus S^{\varepsilon}$ (see discussion above equation (7.203)). This implies that $\eta(\delta, \varepsilon) > 0$. \Box

The following asymptotic result is an immediate consequence of inequality (5.93),

$$\limsup_{N \to \infty} \frac{1}{N} \ln \left[1 - \Pr(\hat{\mathcal{S}}_N^{\delta} \subset \mathcal{S}^{\varepsilon}) \right] \le -\eta(\delta, \varepsilon).$$
(5.95)

It means that the probability of the event that any δ -optimal solution of the SAA problem provides an ε -optimal solution of the true problem approaches one *exponentially fast* as $N \to \infty$. Note that since it is possible to employ a mapping $u : \mathcal{X} \setminus S^{\varepsilon} \to S$, with $\varepsilon^* > \varepsilon$ (see (5.89)), this exponential rate of convergence holds even if $\delta = \varepsilon$, and in particular if $\delta = \varepsilon = 0$. However, if $\delta = \varepsilon$ and the difference $\varepsilon^* - \varepsilon$ is small, then the constant $\eta(\delta, \varepsilon)$ could be close to zero. Indeed, for δ close to $-\mathbb{E}[Y(x, \xi)]$, we can write by (7.203) that

$$I_x(-\delta) \approx \frac{\left(-\delta - \mathbb{E}[Y(x,\xi)]\right)^2}{2\sigma_x^2} \ge \frac{(\varepsilon^* - \delta)^2}{2\sigma_x^2},$$
(5.96)

where

$$\sigma_x^2 := \mathbb{V}\mathrm{ar}[Y(x,\xi)] = \mathbb{V}\mathrm{ar}[F(u(x),\xi) - F(x,\xi)].$$
(5.97)

Let us make now the following assumption.

(M3) There is a constant $\sigma > 0$ such that for any $x \in \mathcal{X} \setminus \mathcal{S}^{\varepsilon}$ the moment generating function $M_x(t)$ of the random variable $Y(x,\xi) - \mathbb{E}[Y(x,\xi)]$ satisfies

$$M_x(t) \le \exp\left(\sigma^2 t^2/2\right), \quad \forall t \in \mathbb{R}.$$
 (5.98)

It follows from the above assumption (M3) that

$$\ln \mathbb{E}\left[e^{tY(x,\xi)}\right] - t\mathbb{E}[Y(x,\xi)] = \ln M_x(t) \le \sigma^2 t^2/2, \tag{5.99}$$

and hence the rate function $I_x(\cdot)$, of $Y(x,\xi)$, satisfies

$$I_x(z) \ge \sup_{t \in \mathbb{R}} \left\{ t(z - \mathbb{E}[Y(x,\xi)]) - \sigma^2 t^2 / 2 \right\} = \frac{\left(z - \mathbb{E}[Y(x,\xi)]\right)^2}{2\sigma^2}, \quad \forall z \in \mathbb{R}.$$
 (5.100)

In particular, it follows that

$$I_x(-\delta) \ge \frac{\left(-\delta - \mathbb{E}[Y(x,\xi)]\right)^2}{2\sigma^2} \ge \frac{(\varepsilon^* - \delta)^2}{2\sigma^2} \ge \frac{(\varepsilon - \delta)^2}{2\sigma^2}.$$
 (5.101)

Consequently the constant $\eta(\delta, \varepsilon)$ satisfies

$$\eta(\delta,\varepsilon) \ge \frac{(\varepsilon-\delta)^2}{2\sigma^2},\tag{5.102}$$

\$ 2

and hence the bound (5.93) of Theorem 5.16 takes the form

$$1 - \Pr(\hat{\mathcal{S}}_N^{\delta} \subset \mathcal{S}^{\varepsilon}) \le |\mathcal{X}| \, e^{-N(\varepsilon - \delta)^2 / (2\sigma^2)}.$$
(5.103)

This leads to the following result giving an estimate of the sample size which guarantees that any δ -optimal solution of the SAA problem is an ε -optimal solution of the true problem with probability at least $1 - \alpha$.

Theorem 5.17. Suppose that assumptions (M1) and (M3) hold. Then for $\varepsilon > 0$, $0 \le \delta < \varepsilon$ and $\alpha \in (0, 1)$, and for the sample size N satisfying

$$N \ge \frac{2\sigma^2}{(\varepsilon - \delta)^2} \ln\left(\frac{|\mathcal{X}|}{\alpha}\right),\tag{5.104}$$

it follows that

$$\Pr(\hat{\mathcal{S}}_N^{\delta} \subset \mathcal{S}^{\varepsilon}) \ge 1 - \alpha.$$
(5.105)

Proof. By setting the right hand side of the estimate (5.103) to $\leq \alpha$ and solving the obtained inequality, we obtain (5.104).

Remark 13. A key characteristic of the estimate (5.104) is that the required sample size N depends *logarithmically* both on the size (cardinality) of the feasible set \mathcal{X} and on the tolerance probability (significance level) α . The constant σ , postulated in assumption (M3),
measures, in a sense, variability of a considered problem. If, for some $x \in \mathcal{X}$, the random variable $Y(x,\xi)$ has a normal distribution with mean μ_x and variance σ_x^2 , then its moment generating function is equal to $\exp\left(\mu_x t + \sigma_x^2 t^2/2\right)$, and hence the moment generating function $M_x(t)$, specified in assumption (M3), is equal to $\exp\left(\sigma_x^2 t^2/2\right)$. In that case $\sigma^2 := \max_{x \in \mathcal{X} \setminus S^{\varepsilon}} \sigma_x^2$ gives the smallest possible value for the corresponding constant in assumption (M3). If $Y(x,\xi)$ is bounded w.p.1, i.e., there is constant b > 0 such that

 $\left|Y(x,\xi) - \mathbb{E}[Y(x,\xi)]\right| \le b \text{ for all } x \in \mathcal{X} \text{ and a.e. } \xi \in \Xi,$

then by Hoeffding inequality (see Proposition 7.71 and estimate (7.211)) we have that $M_x(t) \leq \exp(b^2 t^2/2)$. In that case we can take $\sigma^2 := b^2$.

In any case for small $\varepsilon > 0$ we have by (5.96) that $I_x(-\delta)$ can be approximated from below by $(\varepsilon - \delta)^2/(2\sigma_x^2)$.

Remark 14. For, say, $\delta := \varepsilon/2$, the right hand side of the estimate (5.104) is proportional to $(\sigma/\varepsilon)^2$. For Monte Carlo sampling based methods such dependence on σ and ε seems to be unavoidable. In order to see that consider a simple case when the feasible set $\mathcal X$ consists of just two elements, i.e., $\mathcal{X} = \{x_1, x_2\}$, with $f(x_2) - f(x_1) > \varepsilon > 0$. By solving the corresponding SAA problem we make the (correct) decision that x_1 is the ε optimal solution if $f_N(x_2) - f_N(x_1) > 0$. If the random variable $F(x_2,\xi) - F(x_1,\xi)$ has a normal distribution with mean $\mu = f(x_2) - f(x_1)$ and variance σ^2 , then $\hat{f}_N(x_2)$ – $\hat{f}_N(x_1) \sim \mathcal{N}(\mu, \sigma^2/N)$ and the probability of the event $\{\hat{f}_N(x_2) - \hat{f}_N(x_1) > 0\}$ (i.e., of the correct decision) is $\Phi(\mu\sqrt{N}/\sigma)$, where $\Phi(z)$ is the cumulative distribution function of $\mathcal{N}(0,1)$. We have that $\Phi(\varepsilon\sqrt{N}/\sigma) < \Phi(\mu\sqrt{N}/\sigma)$, and in order to make the probability of the incorrect decision less than α we have to take the sample size $N > z_{\alpha}^2 \sigma^2 / \varepsilon^2$, where $z_{\alpha} := \Phi^{-1}(1-\alpha)$. Even if $F(x_2,\xi) - F(x_1,\xi)$ is not normally distributed, the sample size of order σ^2/ε^2 could be justified asymptotically, say by applying the CLT. It also could be mentioned that if $F(x_2,\xi) - F(x_1,\xi)$ has a normal distribution (with known variance), then the uniformly most powerful test for testing $H_0: \mu \leq 0$ versus $H_a: \mu > 0$ is of the form: "reject H_0 if $\hat{f}_N(x_2) - \hat{f}_N(x_1)$ is bigger than a specified critical value" (this is a consequence of the Neyman-Pearson Lemma). In other words, in such situations if we only have an access to a random sample, then solving the corresponding SAA problem is in a sense a best way to proceed.

Remark 15. Condition (5.98) of assumption (M3) can be replaced by a more general condition

$$M_x(t) \le \exp(\psi(t)), \quad \forall t \in \mathbb{R},$$
(5.106)

where $\psi(t)$ is a convex even function with $\psi(0) = 0$. Then, similar to (5.100), we have

$$I_x(z) \ge \sup_{t \in \mathbb{R}} \left\{ t(z - \mathbb{E}[Y(x,\xi)]) - \psi(t) \right\} = \psi^* \left(z - \mathbb{E}[Y(x,\xi)] \right), \quad \forall z \in \mathbb{R},$$
(5.107)

where ψ^* is the conjugate of function ψ . Consequently the estimate (5.93) takes the form

$$1 - \Pr(\hat{\mathcal{S}}_{N}^{\delta} \subset \mathcal{S}^{\varepsilon}) \le |\mathcal{X}| e^{-N\psi^{*}(\varepsilon - \delta)}, \qquad (5.108)$$

and hence the estimate (5.104) takes the form

$$N \ge \frac{1}{\psi^*(\varepsilon - \delta)} \ln\left(\frac{|\mathcal{X}|}{\alpha}\right).$$
(5.109)

For example, instead of assuming that condition (5.98) of assumption (M3) holds for all $t \in \mathbb{R}$, we may assume that this holds for all t in a finite interval [-a, a], where a > 0 is a given constant. That is, we can take $\psi(t) := \sigma^2 t^2/2$ if $|t| \le a$, and $\psi(t) := +\infty$ otherwise. In that case $\psi^*(z) = z^2/(2\sigma^2)$ for $|z| \le a\sigma^2$, and $\psi^*(z) = a|z| - a^2\sigma^2/2$ for $|z| > a\sigma^2$. Consequently the estimate (5.104) of Theorem 5.17 still holds provided that $0 < \varepsilon - \delta \le a\sigma^2$.

5.3.2 Sample Size Estimates in General Case

Suppose now that \mathcal{X} is a bounded, not necessarily finite, subset of \mathbb{R}^n . and that f(x) is finite valued for all $x \in \mathcal{X}$. Then we can proceed in a way similar to the derivations of section 7.2.10. Let us make the following assumptions.

(M4) For any $x', x \in \mathcal{X}$ there exists constant $\sigma_{x',x} > 0$ such that the moment generating function $M_{x',x}(t) = \mathbb{E}\left[e^{tY_{x',x}}\right]$, of random variable $Y_{x',x} := [F(x',\xi) - f(x')] - [F(x,\xi) - f(x)]$, satisfies

$$M_{x',x}(t) \le \exp\left(\sigma_{x',x}^2 t^2/2\right), \quad \forall t \in \mathbb{R}.$$
(5.110)

(M5) There exists a (measurable) function $\kappa : \Xi \to \mathbb{R}_+$ such that its moment generating function $M_{\kappa}(t)$ is finite valued for all t in a neighborhood of zero and

$$|F(x',\xi) - F(x,\xi)| \le \kappa(\xi) ||x' - x||$$
(5.111)

for a.e. $\xi \in \Xi$ and all $x', x \in \mathcal{X}$.

Of course, it follows from (5.110) that

$$M_{x',x}(t) \le \exp\left(\sigma^2 t^2/2\right), \quad \forall x', x \in \mathcal{X}, \ \forall t \in \mathbb{R},$$
(5.112)

where

$$\sigma^2 := \sup_{x', x \in \mathcal{X}} \sigma^2_{x', x}. \tag{5.113}$$

The above assumption (M4) is slightly stronger than assumption (M3), i.e., assumption (M3) follows from (M4) by taking x' = u(x). Note that $\mathbb{E}[Y_{x',x}] = 0$ and recall that if $Y_{x',x}$ has a normal distribution, then equality in (5.110) holds with $\sigma_{x',x}^2 := \mathbb{V}ar[Y_{x',x}]$.

The assumption (M5) implies that the expectation $\mathbb{E}[\kappa(\xi)]$ is finite and the function f(x) is Lipschitz continuous on \mathcal{X} with Lipschitz constant $L = \mathbb{E}[\kappa(\xi)]$. It follows that the optimal value ϑ^* of the true problem is finite, provided the set \mathcal{X} is bounded (recall that it was assumed that \mathcal{X} is nonempty and closed). Moreover, by Cramér's Large Deviation Theorem we have that for any $L' > \mathbb{E}[\kappa(\xi)]$ there exists a positive constant $\beta = \beta(L')$ such that

$$\Pr\left(\hat{\kappa}_N > L'\right) \le \exp(-N\beta),\tag{5.114}$$

where $\hat{\kappa}_N := N^{-1} \sum_{j=1}^N \kappa(\xi^j)$. Note that it follows from (5.111) that w.p.1

$$\left|\hat{f}_N(x') - \hat{f}_N(x)\right| \le \hat{\kappa}_N \|x' - x\|, \quad \forall x', x \in \mathcal{X},$$
(5.115)

i.e., $\hat{f}_N(\cdot)$ is Lipschitz continuous on \mathcal{X} with Lipschitz constant $\hat{\kappa}_N$.

By $D := \sup_{x,x' \in \mathcal{X}} ||x' - x||$ we denote the diameter of the set \mathcal{X} . Of course, the set \mathcal{X} is bounded iff its diameter is finite. We also use notation $a \lor b := \max\{a, b\}$ for numbers $a, b \in \mathbb{R}$.

Theorem 5.18. Suppose that assumptions (M1) and M(4)–(M5) hold, with the corresponding constant σ^2 defined in (5.113) being finite, the set X has a finite diameter D, and let $\varepsilon > 0, \delta \in [0, \varepsilon), \alpha \in (0, 1), L' > L := \mathbb{E}[\kappa(\xi)]$ and $\beta = \beta(L')$ be the corresponding constants and $\varrho > 0$ be a constant specified below in (5.118). Then for the sample size N satisfying

$$N \ge \frac{8\sigma^2}{(\varepsilon - \delta)^2} \left[n \ln\left(\frac{8\varrho L'D}{\varepsilon - \delta}\right) + \ln\left(\frac{2}{\alpha}\right) \right] \bigvee \left[\beta^{-1} \ln\left(\frac{2}{\alpha}\right) \right], \tag{5.116}$$

it follows that

$$\Pr(\hat{\mathcal{S}}_N^{\delta} \subset \mathcal{S}^{\varepsilon}) \ge 1 - \alpha. \tag{5.117}$$

Proof. Let us set $\nu := (\varepsilon - \delta)/(4L')$, $\varepsilon' := \varepsilon - L'\nu$ and $\delta' := \delta + L'\nu$. Note that $\nu > 0$, $\varepsilon' = 3\varepsilon/4 + \delta/4 > 0$, $\delta' = \varepsilon/4 + 3\delta/4 > 0$ and $\varepsilon' - \delta' = (\varepsilon - \delta)/2 > 0$. Let $\bar{x}_1, ..., \bar{x}_M \in \mathcal{X}$ be such that for every $x \in \mathcal{X}$ there exists \bar{x}_i , $i \in \{1, ..., M\}$, such that $||x - \bar{x}_i|| \le \nu$, i.e., the set $\mathcal{X}' := \{\bar{x}_1, ..., \bar{x}_M\}$ forms a ν -net in \mathcal{X} . We can choose this net in such a way that

$$M \le \left(\varrho D/\nu\right)^n \tag{5.118}$$

for a constant $\rho > 0$. If the $\mathcal{X}' \setminus S^{\varepsilon'}$ is empty, then any point of \mathcal{X}' is an ε' -optimal solution of the true problem. Otherwise, choose a mapping $u : \mathcal{X}' \setminus S^{\varepsilon'} \to S$ and consider the sets $\tilde{S} := \bigcup_{x \in \mathcal{X}'} \{u(x)\}$ and $\tilde{\mathcal{X}} := \mathcal{X}' \cup \tilde{S}$. Note that $\tilde{\mathcal{X}} \subset \mathcal{X}$ and $|\tilde{\mathcal{X}}| \leq (2\rho D/\nu)^n$. Now let us replace the set \mathcal{X} by its subset $\tilde{\mathcal{X}}$. We refer to the obtained true and SAA problems as respective reduced problems. We have that $\tilde{S} \subset S$, any point of the set \tilde{S} is an optimal solutions of the true reduced problem and the optimal value of the true reduced problem is equal to the optimal value of the true (unreduced) problem. By Theorem 5.17 we have that with probability at least $1 - \alpha/2$ any δ' -optimal solution of the reduced SAA problem is an ε' -optimal solutions of the reduced (and hence unreduced) true problem provided that

$$N \ge \frac{8\sigma^2}{(\varepsilon - \delta)^2} \left[n \ln\left(\frac{8\varrho L'D}{\varepsilon - \delta}\right) + \ln\left(\frac{2}{\alpha}\right) \right]$$
(5.119)

(note that the right hand side of (5.119) is greater than or equal to the estimate

$$\frac{2\sigma^2}{(\varepsilon' - \delta')^2} \ln\left(\frac{2|\tilde{\mathcal{X}}|}{\alpha}\right)$$

required by Theorem 5.17). We also have by (5.114) that for

$$N \ge \beta^{-1} \ln\left(\frac{2}{\alpha}\right),\tag{5.120}$$

the Lipschitz constant $\hat{\kappa}_N$ of the function $\hat{f}_N(x)$ is less than or equal to L' with probability at least $1 - \alpha/2$.

Now let \hat{x} be a δ -optimal solution of the (unreduced) SAA problem. Then there is a point $x' \in \tilde{\mathcal{X}}$ such that $||\hat{x} - x'|| \leq \nu$, and hence $\hat{f}_N(x') \leq \hat{f}_N(\hat{x}) + L'\nu$, provided that $\hat{\kappa}_N \leq L'$. We also have that the optimal value of the (unreduced) SAA problem is smaller than or equal to the optimal value of the reduced SAA problem. It follows that x' is a δ' -optimal solution of the reduced SAA problem, provided that $\hat{\kappa}_N \leq L'$. Consequently, we have that x' is an ε' -optimal solution of the true problem with probability at least $1 - \alpha$ provided that N satisfies both inequalities (5.119) and (5.120). It follows that

$$f(\hat{x}) \le f(x') + L\nu \le f(x') + L'\nu \le \vartheta^* + \varepsilon' + L'\nu = \vartheta^* + \varepsilon.$$

We obtain that if N satisfies both inequalities (5.119) and (5.120), then with probability at least $1 - \alpha$ any δ -optimal solution of the SAA problem is an ε -optimal solution of the true problem. The required estimate (5.116) follows.

It is also possible to derive sample size estimates of the form (5.116) directly from the uniform exponential bounds derived in section 7.2.10, see Theorem 7.75 in particular.

Remark 16. If instead of assuming that condition (5.110), of assumption (M4), holds for all $t \in \mathbb{R}$, we assume that it holds for all $t \in [-a, a]$, where a > 0 is a given constant, then the estimate (5.116) of the above theorem still holds provided that $0 < \varepsilon - \delta \le a\sigma^2$ (see Remark 15 on page 205).

In a sense the above estimate (5.116) of the sample size gives an estimate of *complex*ity of solving the corresponding true problem by the SAA method. Suppose, for instance, that the true problem represents the first stage of a two-stage stochastic programming problem. For decomposition type algorithms the total number of iterations, required to solve the SAA problem, typically is independent of the sample size N (this is an empirical observation) and the computational effort at every iteration is proportional to N. Anyway size of the SAA problem grows linearly with increase of N. For $\delta \in [0, \varepsilon/2]$, say, the right hand side of (5.116) is proportional to σ^2/ε^2 , which suggests complexity of order σ^2/ε^2 with respect to the desirable accuracy. This is in a sharp contrast with deterministic (convex) optimization where complexity usually is bounded in terms of $\ln(\varepsilon^{-1})$. It seems that such dependence on σ and ε is unavoidable for Monte Carlo sampling based methods. On the other hand, the estimate (5.116) is *linear* in the dimension n of the first-stage problem. It also depends linearly on $\ln(\alpha^{-1})$. This means that by increasing confidence, say, from 99% to 99.99% we need to increase the sample size by the factor of $\ln 100 \approx 4.6$ at most. Assumption (M4) requires for the probability distribution of the random variable $F(x,\xi) - F(x',\xi)$ to have sufficiently light tails. In a sense, the constant σ^2 can be viewed as a bound reflecting variability of the random variables $F(x,\xi) - F(x',\xi)$, for $x, x' \in \mathcal{X}$. Naturally, larger variability of the data should result in more difficulty in solving the problem (see Remark 14 on page 205).

This suggests that by using Monte Carlo sampling techniques one can solve twostage stochastic programs with a reasonable accuracy, say with relative accuracy of 1% or 2%, in a reasonable time, provided that: (a) its variability is not too large, (b) it has relatively complete recourse, and (c) the corresponding SAA problem can be solved efficiently. And, indeed, this was verified in numerical experiments with two-stage problems having a linear second stage recourse. Of course, the estimate (5.116) of the sample size is far too conservative for actual calculations. For practical applications there are techniques which allow to estimate (statistically) the error of a considered feasible solution \bar{x} for a chosen sample size N, we will discuss this in section 5.6.

We are going to discuss next some modifications of the sample size estimate. It will be convenient in the following estimates to use notation O(1) for a generic constant independent of the data. In that way we avoid denoting many different constants throughout the derivations.

(M6) There exists constant $\lambda > 0$ such that for any $x', x \in \mathcal{X}$ the moment generating function $M_{x',x}(t)$, of random variable $Y_{x',x} := [F(x',\xi) - f(x')] - [F(x,\xi) - f(x)]$, satisfies

$$M_{x',x}(t) \le \exp\left(\lambda^2 \|x' - x\|^2 t^2/2\right), \ \forall t \in \mathbb{R}.$$
 (5.121)

The above assumption (M6) is a particular case of assumption (M4) with

$$\sigma_{x',x}^2 = \lambda^2 \|x' - x\|^2,$$

and we can set the corresponding constant $\sigma^2 = \lambda^2 D^2$. The following corollary follows from Theorem 5.18.

Corollary 5.19. Suppose that assumptions (M1) and (M5)–(M6) hold, the set \mathcal{X} has a finite diameter D, and let $\varepsilon > 0$, $\delta \in [0, \varepsilon)$, $\alpha \in (0, 1)$ and $L = \mathbb{E}[\kappa(\xi)]$ be the corresponding constants. Then for the sample size N satisfying

$$N \ge \frac{O(1)\lambda^2 D^2}{(\varepsilon - \delta)^2} \left[n \ln \left(\frac{O(1)LD}{\varepsilon - \delta} \right) + \ln \left(\frac{1}{\alpha} \right) \right], \tag{5.122}$$

it follows that

$$\Pr(\hat{\mathcal{S}}_N^\delta \subset \mathcal{S}^\varepsilon) \ge 1 - \alpha.$$
 (5.123)

For example, suppose that the Lipschitz constant $\kappa(\xi)$, in the assumption (M5), can be taken independent of ξ . That is, there exists a constant L > 0 such that

$$|F(x',\xi) - F(x,\xi)| \le L ||x' - x||$$
(5.124)

for a.e. $\xi \in \Xi$ and all $x', x \in \mathcal{X}$. It follows that the expectation function f(x) is also Lipschitz continuous on \mathcal{X} with Lipschitz constant L, and hence the random variable $Y_{x',x}$, of assumption (M6), can be bounded as $|Y_{x',x}| \leq 2L ||x' - x||$ w.p.1. Moreover, we have that $\mathbb{E}[Y_{x',x}] = 0$, and hence it follows by Hoeffding's inequality (see the estimate (7.211)) that

$$M_{x',x}(t) \le \exp\left(2L^2 \|x' - x\|^2 t^2\right), \quad \forall t \in \mathbb{R}.$$
(5.125)

Consequently we can take $\lambda = 2L$ in (5.121) and the estimate (5.122) takes the form

$$N \ge \left(\frac{O(1)LD}{\varepsilon - \delta}\right)^2 \left[n \ln\left(\frac{O(1)LD}{\varepsilon - \delta}\right) + \ln\left(\frac{1}{\alpha}\right)\right].$$
 (5.126)

Remark 17. It was assumed in Theorem 5.18 that the set \mathcal{X} has a finite diameter, i.e., that \mathcal{X} is bounded. For convex problems this assumption can be relaxed. Assume that the

problem is convex, the optimal value ϑ^* of the true problem is finite and for some $a > \varepsilon$ the set \mathcal{S}^a has a finite diameter D_a^* (recall that $\mathcal{S}^a := \{x \in \mathcal{X} : f(x) \leq \vartheta^* + a\}$). We refer here to the respective true and SAA problems, obtained by replacing the feasible set \mathcal{X} by its subset \mathcal{S}^a , as reduced problems. Note that the set $\mathcal{S}^{\varepsilon}$, of ε -optimal solutions, of the reduced and original true problems are the same. Let N^* be an integer satisfying the inequality (5.116) with D replaced by D_a^* . Then, under the assumptions of Theorem 5.18, we have that with probability at least $1 - \alpha$ all δ -optimal solutions of the reduced SAA problem are ε -optimal solutions of the true problem. Let us observe now that in this case the set of δ -optimal solutions of the reduced SAA problem coincides with the set of δ -optimal solutions of the original SAA problem. Indeed, suppose that the original SAA problem has a δ -optimal solution $x^* \in \mathcal{X} \setminus \mathcal{S}^a$. Let $\bar{x} \in \arg \min_{x \in \mathcal{S}^a} f_N(x)$, such a minimizer does exist since S^a is compact and $\hat{f}_N(x)$ is real valued convex and hence continuous. Then $\bar{x} \in S^{\varepsilon}$ and $\hat{f}_N(x^*) \leq \hat{f}_N(\bar{x}) + \delta$. By convexity of $\hat{f}_N(x)$ it follows that $\hat{f}_N(x) \leq \max{\{\hat{f}_N(\bar{x}), \hat{f}_N(x^*)\}}$ for all x on the segment joining \bar{x} and x^* . This segment has a common point \hat{x} with the set $S^a \setminus S^{\varepsilon}$. We obtain that $\hat{x} \in S^a \setminus S^{\varepsilon}$ is a δ -optimal solutions of the reduced SAA problem, a contradiction.

That is, with such sample size N^* we are guaranteed with probability at least $1 - \alpha$ that any δ -optimal solution of the SAA problem is an ε -optimal solution of the true problem. Also assumptions (M4) and (M5) should be verified for x, x' in the set S^a only.

Remark 18. Suppose that the set S, of optimal solutions of the true problem, is nonempty. Then it follows from the proof of Theorem 5.18 that it suffices in the assumption (M4) to verify condition (5.110) only for every $x \in \mathcal{X} \setminus S^{\varepsilon'}$ and x' := u(x), where $u : \mathcal{X} \setminus S^{\varepsilon'} \to S$ and $\varepsilon' := 3/4\varepsilon + \delta/4$. If the set S is closed, we can use, for instance, a mapping u(x)assigning to each $x \in \mathcal{X} \setminus S^{\varepsilon'}$ a point of S closest to x. If, moreover, the set S is convex and the employed norm is strictly convex (e.g., the Euclidean norm), then such mapping (called metric projection onto S) is defined uniquely. If, moreover, the assumption (M6) holds, then for such x and x' we have $\sigma_{x',x}^2 \leq \lambda^2 \bar{D}^2$, where $\bar{D} := \sup_{x \in \mathcal{X} \setminus S^{\varepsilon'}} \operatorname{dist}(x, S)$. Suppose, further, that the problem is convex. Then (see Remark 17 on page 209), for any $a > \varepsilon$, we can use S^a instead of \mathcal{X} . Therefore, if the problem is convex and the assumption (M6) holds, we can write the following estimate of the required sample size:

$$N \ge \frac{O(1)\lambda^2 \bar{D}_{a,\varepsilon}^2}{\varepsilon - \delta} \left[n \ln \left(\frac{O(1)LD_a^*}{\varepsilon - \delta} \right) + \ln \left(\frac{1}{\alpha} \right) \right], \tag{5.127}$$

where D_a^* is the diameter of \mathcal{S}^a and $\overline{D}_{a,\varepsilon} := \sup_{x \in \mathcal{S}^a \setminus \mathcal{S}^{\varepsilon'}} \operatorname{dist}(x, \mathcal{S})$.

Corollary 5.20. Suppose that assumptions (M1) and (M5)–(M6) hold, the problem is convex, the "true" optimal set S is nonempty and for some $\gamma \ge 1$, c > 0 and r > 0, the following growth condition holds

$$f(x) \ge \vartheta^* + c \left[\operatorname{dist}(x, \mathcal{S}) \right]^{\gamma}, \quad \forall x \in \mathcal{S}^r.$$
(5.128)

Let $\alpha \in (0, 1)$, $\varepsilon \in (0, r)$ and $\delta \in [0, \varepsilon/2]$ and suppose, further, that for $a := \min\{2\varepsilon, r\}$ the diameter D_a^* of S^a is finite.

Then for the sample size N satisfying

$$N \ge \frac{O(1)\lambda^2}{c^{2/\gamma}\varepsilon^{2(\gamma-1)/\gamma}} \left[n \ln\left(\frac{O(1)LD_a^*}{\varepsilon}\right) + \ln\left(\frac{1}{\alpha}\right) \right],\tag{5.129}$$

it follows that

$$\Pr(\hat{\mathcal{S}}_N^{\delta} \subset \mathcal{S}^{\varepsilon}) \ge 1 - \alpha.$$
(5.130)

Proof. It follows from (5.128) that for any $a \leq r$ and $x \in S^a$, the inequality dist $(x, S) \leq (a/c)^{1/\gamma}$ holds. Consequently, for any $\varepsilon \in (0, r)$, by taking $a := \min\{2\varepsilon, r\}$ and $\delta \in [0, \varepsilon/2]$ we obtain from (5.127) the required sample size estimate (5.129). \Box

Note that since $a = \min\{2\varepsilon, r\} \le r$, we have that $S^a \subset S^r$, and if $S = \{x^*\}$ is a singleton, then it follows from (5.128) that $D_a^* \le 2(a/c)^{1/\gamma}$. In particular, if $\gamma = 1$ and $S = \{x^*\}$ is a singleton (in that case it is said that the optimal solution x^* is *sharp*), then D_a^* can be bounded by $4c^{-1}\varepsilon$ and hence we obtain the following estimate

$$N \ge O(1)c^{-2}\lambda^2 \left[n \ln \left(O(1)c^{-1}L \right) + \ln \left(\alpha^{-1} \right) \right],$$
(5.131)

which does not depend on ε . For $\gamma = 2$ condition (5.128) is called the "second-order" or "quadratic" growth condition. Under the quadratic growth condition the first term in the right hand side of (5.129) becomes of order $c^{-1}\varepsilon^{-1}\lambda^2$.

The following example shows that the estimate (5.116) of the sample size cannot be significantly improved for the class of convex stochastic programs.

Example 5.21 Consider the true problem with $F(x,\xi) := ||x||^{2m} - 2m \xi^T x$, where *m* is a positive constant, $|| \cdot ||$ is the Euclidean norm and $\mathcal{X} := \{x \in \mathbb{R}^n : ||x|| \le 1\}$. Suppose, further, that random vector ξ has normal distribution $\mathcal{N}(0, \sigma^2 I_n)$, where σ^2 is a positive constant and I_n is the $n \times n$ identity matrix, i.e., components ξ_i of ξ are independent and $\xi_i \sim \mathcal{N}(0, \sigma^2)$, i = 1, ..., n. It follows that $f(x) = ||x||^{2m}$, and hence for $\varepsilon \in [0, 1]$ the set of ε -optimal solutions of the true problem is given by $\{x : ||x||^{2m} \le \varepsilon\}$. Now let $\xi^1, ..., \xi^N$ be an iid random sample of ξ and $\overline{\xi_N} := (\xi^1 + ... + \xi^N)/N$. The corresponding sample average function is

$$\hat{f}_N(x) = \|x\|^{2m} - 2m\,\bar{\xi}_N^{\mathsf{T}}x,\tag{5.132}$$

and the optimal solution \hat{x}_N of the SAA problem is $\hat{x}_N = \|\bar{\xi}_N\|^{-b}\bar{\xi}_N$, where

$$b := \begin{cases} \frac{2m-2}{2m-1}, & \text{if } \|\bar{\xi}_N\| \le 1, \\ 1, & \text{if } \|\bar{\xi}_N\| > 1. \end{cases}$$

It follows that, for $\varepsilon \in (0,1)$, the optimal solution of the corresponding SAA problem is an ε -optimal solution of the true problem iff $\|\bar{\xi}_N\|^{\nu} \leq \varepsilon$, where $\nu := \frac{2m}{2m-1}$. We have that $\bar{\xi}_N \sim \mathcal{N}(0, \sigma^2 N^{-1} I_n)$, and hence $N \|\bar{\xi}_N\|^2 / \sigma^2$ has a chi-square distribution with *n* degrees of freedom. Consequently, the probability that $\|\bar{\xi}_N\|^{\nu} > \varepsilon$ is equal to the probability $\Pr(\chi_n^2 > N \varepsilon^{2/\nu} / \sigma^2)$. Moreover, $\mathbb{E}[\chi_n^2] = n$ and the probability $\Pr(\chi_n^2 > n)$ increases and tends to 1/2 as *n* increases. Consequently, for $\alpha \in (0, 0.3)$ and $\varepsilon \in (0, 1)$, for example, the sample size *N* should satisfy

$$N > \frac{n\sigma^2}{\varepsilon^{2/\nu}} \tag{5.133}$$

in order to have the property: "with probability $1 - \alpha$ an (exact) optimal solution of the SAA problem is an ε -optimal solution of the true problem". Compared with (5.116), the lower bound (5.133) also grows linearly in n and is proportional to $\sigma^2/\varepsilon^{2/\nu}$. It remains to note that the constant ν decreases to one as m increases.

Note that in this example the growth condition (5.128) holds with $\gamma = 2m$, and that the power constant of ε in the estimate (5.133) is in accordance with the estimate (5.129). Note also that here

$$[F(x',\xi) - f(x')] - [F(x,\xi) - f(x)] = 2m\xi^{\mathsf{T}}(x - x')$$

has normal distribution with zero mean and variance $4m^2\sigma^2 ||x' - x||^2$. Consequently assumption (M6) holds with $\lambda^2 = 4m^2\sigma^2$.

Of course, in this example the "true" optimal solution is $\bar{x} = 0$, and one does not need sampling in order to solve this problem. Note, however, that the sample average function $\hat{f}_N(x)$ here depends on the random sample only through the data average vector $\bar{\xi}_N$. Therefore, any numerical procedure based on averaging will need a sample of size Nsatisfying the estimate (5.133) in order to produce an ε -optimal solution.

5.3.3 Finite Exponential Convergence

We assume in this section that the problem is *convex* and the expectation function f(x) is finite valued.

Definition 5.22. It is said that $x^* \in \mathcal{X}$ is a sharp (optimal) solution, of the true problem (5.1), if there exists constant c > 0 such that

$$f(x) \ge f(x^*) + c ||x - x^*||, \quad \forall x \in \mathcal{X}.$$
 (5.134)

The above condition (5.134) corresponds to the growth condition (5.128) with the power constant $\gamma = 1$ and $S = \{x^*\}$. Since $f(\cdot)$ is convex finite valued we have that the directional derivatives $f'(x^*, h)$ exist for all $h \in \mathbb{R}^n$, $f'(x^*, \cdot)$ is (locally Lipschitz) continuous and formula (7.17) holds. Also by convexity of the set \mathcal{X} we have that the tangent cone $\mathcal{T}_{\mathcal{X}}(x^*)$, to \mathcal{X} at x^* , is given by the topological closure of the corresponding radial cone. By using these facts it is not difficult to show that condition (5.134) is equivalent to:

$$f'(x^*, h) \ge c \|h\|, \quad \forall h \in \mathcal{T}_{\mathcal{X}}(x^*).$$
(5.135)

Since the above condition (5.135) is local in nature, we have that it actually suffices to verify (5.134) for all $x \in \mathcal{X}$ in a neighborhood of x^* .

Theorem 5.23. Suppose that the problem is convex and assumption (M1) holds, and let $x^* \in \mathcal{X}$ be a sharp optimal solution of the true problem. Then $\hat{S}_N = \{x^*\}$ w.p.1 for N large enough. Suppose, further, that assumption (M4) holds. Then there exist constants C > 0 and $\beta > 0$ such that

$$1 - \Pr(\hat{S}_N = \{x^*\}) \le Ce^{-N\beta}, \tag{5.136}$$

i.e., probability of the event that " x^* is the unique optimal solution of the SAA problem" converges to one exponentially fast with increase of the sample size N.

Proof. By convexity of $F(\cdot,\xi)$ we have that $\hat{f}'_N(x^*,\cdot)$ converges to $f'(x^*,\cdot)$ w.p.1 uniformly on the unit sphere (see the proof of Theorem 7.62). It follows w.p.1 for N large enough that

$$\hat{f}'_N(x^*,h) \ge (c/2) \|h\|, \ \forall h \in \mathcal{T}_{\mathcal{X}}(x^*),$$
(5.137)

which implies that x^* is the sharp optimal solution of the corresponding SAA problem.

Now under the assumptions of convexity and (M1) and (M4), we have that $f'_N(x^*, \cdot)$ converges to $f'(x^*, \cdot)$ exponentially fast on the unit sphere (see inequality (7.244) of Theorem 7.77). By taking $\varepsilon := c/2$ in (7.244), we can conclude that (5.136) follows. \Box

It is also possible to consider the growth condition (5.128) with $\gamma = 1$ and the set S not necessarily being a singleton. That is, it is said that the set S of optimal solutions of the true problem is *sharp* if for some c > 0 the following condition holds

$$f(x) \ge \vartheta^* + c \left[\operatorname{dist}(x, \mathcal{S}) \right], \quad \forall x \in \mathcal{X}.$$
(5.138)

Of course, if $S = \{x^*\}$ is a singleton, then conditions (5.134) and (5.138) do coincide. The set of optimal solutions of the true problem is always nonempty and sharp if its optimal value is finite and the problem is *piecewise linear* in the sense that the following conditions hold:

(P1) The set \mathcal{X} is a convex closed polyhedron.

- (P2) The support set $\Xi = \{\xi_1, ..., \xi_K\}$ is finite.
- (P3) For every $\xi \in \Xi$ the function $F(\cdot, \xi)$ is polyhedral.

The above conditions (P1)–(P3) hold in case of two-stage linear stochastic programming problems with a finite number of scenarios.

Under conditions (P1)–(P3) the true and SAA problems are polyhedral, and hence their sets of optimal solutions are polyhedral. By using polyhedral structure and finiteness of the set Ξ it is possible to show the following result (cf., [251]).

Theorem 5.24. Suppose that conditions (P1)–(P3) hold and the set S is nonempty and bounded. Then S is polyhedral and there exist constants C > 0 and $\beta > 0$ such that

$$1 - \Pr(\hat{\mathcal{S}}_N \neq \emptyset \text{ and } \hat{\mathcal{S}}_N \text{ is a face of } \mathcal{S}) \le Ce^{-N\beta}, \tag{5.139}$$

i.e., probability of the event that " \hat{S}_N is nonempty and forms a face of the set S" converges to one exponentially fast with increase of the sample size N.

5.4 Quasi-Monte Carlo Methods

In the previous section we discussed an approach to evaluating (approximating) expectations by employing random samples generated by Monte Carlo techniques. It should be understood, however, that when dimension d (of the random data vector ξ) is small, the Monte Carlo approach may be not a best way to proceed. In this section we give a brief discussion of the so-called quasi-Monte Carlo methods. It is beyond the scope of this book to give a detailed discussion of that subject. This section is based on Niederreiter [167], to which the interested reader is referred for a further reading on that topic. Let us start our discussion by considering a one dimensional case (of d = 1).

Let ξ be a real valued random variable having cdf $H(z) = \Pr(\xi \leq z)$. Suppose that we want to evaluate the expectation

$$\mathbb{E}[F(\xi)] = \int_{-\infty}^{+\infty} F(z)dH(z), \qquad (5.140)$$

where $F : \mathbb{R} \to \mathbb{R}$ is a measurable function. Let $U \sim U[0, 1]$, i.e., U is a random variable uniformly distributed on [0, 1]. Then random variable⁵ $H^{-1}(U)$ has cdf $H(\cdot)$. Therefore, by making change of variables we can write the expectation (5.140) as

$$\mathbb{E}[\psi(U)] = \int_0^1 \psi(u) du, \qquad (5.141)$$

where $\psi(u) := F(H^{-1}(u))$.

Evaluation of the above expectation by Monte Carlo method is based on generating an iid sample $U^1, ..., U^N$, of N replications of $U \sim U[0, 1]$, and consequently approximating $\mathbb{E}[\psi(U)]$ by the average $\bar{\psi}_N := N^{-1} \sum_{j=1}^N \psi(U^j)$. Alternatively, one can employ the Riemann sum approximation

$$\int_{0}^{1} \psi(u) du \approx \frac{1}{N} \sum_{j=1}^{N} \psi(u_j)$$
 (5.142)

by using some points $u_j \in [(j-1)/N, j/N], j = 1, ..., N$, e.g., taking midpoints $u_j := (2j-1)/(2N)$ of equally spaced partition intervals [(j-1)/N, j/N], j = 1, ..., N. If the function $\psi(u)$ is Lipschitz continuous on [0,1], then the error of the Riemann sum approximation⁶ is of order $O(N^{-1})$, while the Monte Carlo sample average error is of (stochastic) order $O_p(N^{-1/2})$. An explanation of this phenomenon is rather clear, an iid sample $U^1, ..., U^N$ will tend to cluster in some areas while leaving other areas of the interval [0,1] uncovered.

One can argue that the Monte Carlo sampling approach has an advantage of the possibility of estimating the approximation error by calculating the sample variance

$$s^{2} := (N-1)^{-1} \sum_{j=1}^{N} \left[\psi(U^{j}) - \bar{\psi}_{N} \right]^{2},$$

and consequently constructing a corresponding confidence interval. It is possible, however, to employ a similar procedure for the Riemann sums by making them random. That is,

⁵Recall that $H^{-1}(u) := \inf\{z : H(z) \ge u\}.$

⁶If $\psi(u)$ is continuously differentiable, then, e.g., the trapezoidal rule gives even a slightly better approximation error of order $O(N^{-2})$. Also one should be careful in making the assumption of Lipschitz continuity of $\psi(u)$. If the distribution of ξ is supported on the whole real line, e.g., is normal, then $H^{-1}(u)$ tends to ∞ as utends to 0 or 1. In that case $\psi(u)$ typically will be discontinuous at u = 0 and u = 1.

each point u_j , in the right hand side of (5.142), is generated randomly, say uniformly distributed, on the corresponding interval [(j-1)/N, j/N], independently of other points $u_k, k \neq j$. This will make the right hand side of (5.142) a random variable. Its variance can be estimated by using several independently generated batches of such approximations.

It does not make sense to use Monte Carlo sampling methods in case of one dimensional random data. The situation starts to change quickly with increase of the dimension d. By making an appropriate transformation we may assume that the random data vector is distributed uniformly on the d-dimensional cube $I^d = [0,1]^d$. For d > 1 we denote by (bold-faced) U a random vector uniformly distributed on I^d . Suppose that we want to evaluate the expectation $\mathbb{E}[\psi(U)] = \int_{I^d} \psi(u) du$, where $\psi: I^d \to \mathbb{R}$ is a measurable function. We can partition each coordinate of I^d into M equally spaced intervals, and hence partition I^d into the corresponding $N = M^d$ subintervals⁷ and use a corresponding Riemann sum approximation $N^{-1} \sum_{j=1}^{N} \psi(u_j)$. The resulting error is of order $O(M^{-1})$, provided that the function $\psi(u)$ is Lipschitz continuous. In terms of the total number N of function evaluations, this error is of order $O(N^{-1/d})$. For d = 2 it is still compatible with the Monte Carlo sample average approximation approach. However, for larger values of d the Riemann sums approach quickly becomes unacceptable. On the other hand, the rate of convergence (error bounds) of the Monte Carlo sample average approximation of $\mathbb{E}[\psi(U)]$ does not depend directly on dimensionality d, but only on the corresponding variance $\mathbb{V}ar[\psi(U)]$. Yet the problem of uneven covering of I^d by an iid sample U^j , j = 1, ..., N, remains persistent.

Quasi-Monte Carlo methods employ the approximation

$$\mathbb{E}[\psi(\boldsymbol{U})] \approx \frac{1}{N} \sum_{j=1}^{N} \psi(\boldsymbol{u}_j)$$
(5.143)

for a carefully chosen (deterministic) sequence of points $u_1, ..., u_N \in I^d$. From the numerical point of view it is important to be able to generate such a sequence iteratively as an infinite sequence of points $u_j, j = 1, ..., \text{ in } I^d$. In that way one does not need to recalculate already calculated function values $\psi(u_j)$ with increase of N. A basic requirement for this sequence is that the right hand side of (5.143) converges to $\mathbb{E}[\psi(U)]$ as $N \to \infty$. It is not difficult to show that this holds (for any Riemann-integrable function $\psi(u)$) if

$$\lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_A(\boldsymbol{u}_j) = V_d(A)$$
(5.144)

for any interval $A \subset I^d$. Here $V_d(A)$ denotes the *d*-dimensional Lebesgue measure (volume) of set $A \subset \mathbb{R}^d$.

Definition 5.25. The star discrepancy of a point set $\{u_1, ..., u_N\} \subset I^d$ is defined by

$$\mathcal{D}^*(\boldsymbol{u}_1,...,\boldsymbol{u}_N) := \sup_{A \in \mathcal{I}} \left| \frac{1}{N} \sum_{j=1}^N \mathbf{1}_A(\boldsymbol{u}_j) - V_d(A) \right|,$$
(5.145)

⁷A set $A \subset \mathbb{R}^d$ is said to be a (*d*-dimensional) interval if $A = [a_1, b_1] \times \cdots \times [a_d, b_d]$.

where \mathcal{I} is the family of all subintervals of I^d of the form $\prod_{i=1}^d [0, b_i)$.

It is possible to show that for a sequence $u_j \in I^d$, j = 1, ..., condition (5.144) holds iff $\lim_{N\to\infty} \mathcal{D}^*(u_1, ..., u_N) = 0$. A more important property of the star discrepancy is that it is possible to give error bounds in terms of $\mathcal{D}^*(u_1, ..., u_N)$ for quasi-Monte Carlo approximations. Let us start with one dimensional case. Recall that variation of a function $\psi : [0,1] \to \mathbb{R}$ is the $\sup \sum_{i=1}^m |\psi(t_i) - \psi(t_{i-1})|$, where the supremum is taken over all partitions $0 = t_0 < t_1 < ... < t_m = 1$ of the interval [0,1]. It is said that ψ has bounded variation if its variation is finite.

Theorem 5.26 (Koksma). If $\psi : [0,1] \to \mathbb{R}$ has bounded variation $V(\psi)$, then for any $u_1, ..., u_N \in [0,1]$ we have

$$\left|\frac{1}{N}\sum_{j=1}^{N}\psi(u_j) - \int_0^1\psi(u)du\right| \le V(\psi)\mathcal{D}^*(u_1,...,u_N).$$
(5.146)

Proof. We can assume that the sequence $u_1, ..., u_N$ is arranged in the increasing order, and set $u_0 = 0$ and $u_{N+1} = 1$. That is, $0 = u_0 \le u_1 \le ... \le u_{N+1} = 1$. Using integration by parts we have

$$\int_0^1 \psi(u) du = u\psi(u) \Big|_0^1 - \int_0^1 u d\psi(u) = \psi(1) - \int_0^1 u d\psi(u),$$

and summation by parts

$$\frac{1}{N}\sum_{j=1}^{N}\psi(u_j) = \psi(u_{N+1}) - \sum_{j=0}^{N}\frac{j}{N}[\psi(u_{j+1}) - \psi(u_j)],$$

we can write

$$\begin{array}{rcl} \frac{1}{N} \sum_{j=1}^{N} \psi(u_j) - \int_0^1 \psi(u) du &=& -\sum_{j=0}^{N} \frac{j}{N} [\psi(u_{j+1}) - \psi(u_j)] + \int_0^1 u d\psi(u) \\ &=& \sum_{j=0}^{N} \int_{u_j}^{u_{j+1}} \left(u - \frac{j}{N}\right) d\psi(u). \end{array}$$

Also for any $u \in [u_j, u_{j+1}], j = 0, ..., N$, we have

$$\left|u-\frac{j}{N}\right| \leq \mathcal{D}^*(u_1,...,u_N).$$

It follows that

$$\left| \frac{1}{N} \sum_{j=1}^{N} \psi(u_j) - \int_0^1 \psi(u) du \right| \leq \sum_{j=0}^{N} \int_{u_j}^{u_{j+1}} \left| u - \frac{j}{N} \right| d\psi(u) \\ \leq \mathcal{D}^*(u_1, ..., u_N) \sum_{j=0}^{N} \left| \psi(u_{j+1}) - \psi(u_j) \right|,$$

and, of course, $\sum_{j=0}^{N} |\psi(u_{j+1}) - \psi(u_j)| \le V(\psi)$. This completes the proof. \Box

This can be extended to a multidimensional setting as follows. Consider a function $\psi: I^d \to \mathbb{R}$. The variation of ψ , in the sense of Vitali, is defined as

$$V^{(d)}(\psi) := \sup_{\mathcal{P} \in \mathcal{J}} \sum_{A \in \mathcal{P}} |\Delta_{\psi}(A)|, \qquad (5.147)$$

where \mathcal{J} denotes the family of all partitions \mathcal{P} of I^d into subintervals, and for $A \in \mathcal{P}$ the notation $\Delta_{\psi}(A)$ stands for an alternating sum of the values of ψ at the vertices of A (i.e., function values at adjacent vertices have opposite signs). The variation of ψ , in the sense of Hardy and Krause, is defined as

$$V(\psi) := \sum_{k=1}^{d} \sum_{1 \le i_1 < i_2 < \dots + i_k \le d} V^{(k)}(\psi; i_1, \dots, i_k),$$
(5.148)

where $V^{(k)}(\psi; i_1, ..., i_k)$ is the variation in the sense of Vitali of restriction of ψ to the k-dimensional face of I^d defined by $u_j = 1$ for $j \notin \{i_1, ..., i_k\}$.

Theorem 5.27 (Hlawka). If $\psi : I^d \to \mathbb{R}$ has bounded variation $V(\psi)$ on I^d in the sense of Hardy and Krause, then for any $u_1, ..., u_N \in I^d$ we have

$$\left|\frac{1}{N}\sum_{j=1}^{N}\psi(\boldsymbol{u}_{j})-\int_{I^{d}}\psi(\boldsymbol{u})d\boldsymbol{u}\right|\leq V(\psi)\mathcal{D}^{*}(\boldsymbol{u}_{1},...,\boldsymbol{u}_{N}).$$
(5.149)

In order to see how good the above error estimates could be let us consider the one dimensional case with $u_j := (2j-1)/(2N)$, j = 1, ..., N. Then $\mathcal{D}^*(u_1, ..., u_N) = 1/(2N)$, and hence the estimate (5.146) leads to the error bound $V(\psi)/(2N)$. This error bound gives the correct order $O(N^{-1})$ for the error estimates (provided that ψ has bounded variation), but the involved constant $V(\psi)/2$ typically is far too large for practical calculations. Even worse, the inverse function $H^{-1}(u)$ is monotonically nondecreasing and hence its variation is given by the difference of the limits $\lim_{u\to+\infty} H^{-1}(u)$ and $\lim_{u\to-\infty} H^{-1}(u)$. Therefore, if one of this limits is infinite, i.e., the support of the corresponding random variable is unbounded, then the associated variation is infinite. Typically this variation unboundedness will carry over to the function $\psi(u) = F(H^{-1}(u))$. For example, if the function $F(\cdot)$ is monotonically nondecreasing, then

$$V(\psi) = F\left(\lim_{u \to +\infty} H^{-1}(u)\right) - F\left(\lim_{u \to -\infty} H^{-1}(u)\right)$$

This overestimation of the corresponding constant becomes even worse with increase of the dimension d.

A sequence $\{u_j\}_{j\in\mathbb{N}} \subset I^d$ is called a *low-discrepancy sequence* if $\mathcal{D}^*(u_1, ..., u_N)$ is "small" for all $N \geq 1$. We proceed now to a description of classical constructions of low-discrepancy sequences. Let us start with the one dimensional case. It is not difficult to show that $\mathcal{D}^*(u_1, ..., u_N)$ always greater than or equal to 1/(2N) and this lower bound is attained for $u_j := (2j-1)/2N$, j = 1, ..., N. While the lower bound of order $O(N^{-1})$ is attained for some N-element point sets from [0,1], there does not exist a sequence $u_1, ...,$

in [0,1] such that $\mathcal{D}^*(u_1, ..., u_N) \leq c/N$ for some c > 0 and all $N \in \mathbb{N}$. It is possible to show that a best possible rate for $\mathcal{D}^*(u_1, ..., u_N)$, for a sequence of points $u_j \in [0, 1]$, j = 1, ..., is of order $O(N^{-1} \ln N)$. We are going to construct now a sequence for which this rate is attained.

For any integer $n \ge 0$ there is a unique digit expansion

$$n = \sum_{i>0} a_i(n)b^i$$
 (5.150)

in integer base $b \ge 2$, where $a_i(n) \in \{0, 1, ..., b-1\}$, i = 0, 1, ..., and $a_i(n) = 0$ for all *i* large enough, i.e., the sum (5.150) is finite. The associated *radical-inverse function* $\phi_b(n)$, in base *b*, is defined by

$$\phi_b(n) := \sum_{i \ge 0} a_i(n) b^{-i-1}.$$
(5.151)

Note that

$$\phi_b(n) \le (b-1) \sum_{i=0}^{\infty} b^{-i-1} = 1,$$

and hence $\phi_b(n) \in [0, 1]$ for any integer $n \ge 0$.

Definition 5.28. For an integer $b \ge 2$, the van der Corput sequence in base b is the sequence $u_j := \phi_b(j), j = 0, 1, ...$

It is possible to show that to every van der Corput sequence $u_1, ...,$ in base b, corresponds constant C_b such that

$$\mathcal{D}^*(u_1, ..., u_n) \le C_b N^{-1} \ln N$$
, for all $N \in \mathbb{N}$.

A classical extension of van der Corput sequences to multidimensional settings is the following. Let $p_1 = 2, p_2 = 3, ..., p_d$ be the first d prime numbers. Then the *Halton* sequence, in the bases $p_1, ..., p_d$, is defined as

$$\boldsymbol{u}_j := (\phi_{p_1}(j), \dots, \phi_{p_d}(j)) \in I^d, \ \ j = 0, 1, \dots .$$
(5.152)

It is possible to show that for that sequence

$$\mathcal{D}^*(\boldsymbol{u}_1, ..., \boldsymbol{u}_N) \le A_d N^{-1} (\ln N)^d + O(N^{-1} (\ln N)^{d-1}), \text{ for all } N \ge 2, \qquad (5.153)$$

where $A_d = \prod_{i=1}^d \frac{p_i - 1}{2 \ln p_i}$. By the bound (5.149) of Theorem 5.27, this implies that the error of the corresponding quasi-Monte Carlo approximation is of order $O(N^{-1}(\ln N)^d)$, provided that variation $V(\psi)$ is finite. This compares favorably with the bound $O_p(N^{-1/2})$ of the Monte Carlo sampling. Note, however, that by the prime number theorem we have that $\frac{\ln A_d}{d \ln d}$ tends to 1 as $d \to \infty$. That is, the coefficient A_d , of the leading term in the right hand side of (5.153), grows superexponentially with increase of the dimension d. This makes the corresponding error bounds useless for larger values of d. It should be noticed that the above are *upper* bounds for the rates of convergence and in practice convergence rates could be much better. It seems that for low dimensional problems, say $d \leq 20$,

quasi-Monte Carlo methods are advantageous over Monte Carlo methods. With increase of the dimension d this advantage becomes less apparent. Of course, all this depends on a particular class of problems and applied quasi-Monte Carlo method. This issue requires a further investigation.

A drawback of (deterministic) quasi-Monte Carlo sequences $\{u_i\}_{i \in \mathbb{N}}$ is that there is no easy way to estimate the error of the corresponding approximations $N^{-1} \sum_{j=1}^{N} \psi(\boldsymbol{u}_j)$. In that respect bounds like (5.149) typically are too loose and impossible to calculate anyway. A way of dealing with this problem is to use a randomization of the set $\{u_1, \dots, u_N\}$, of generating points in I^d , without destroying its regular structure. Such simple randomization procedure was suggested by Cranley and Patterson [46]. That is, generate a random point u uniformly distributed over I^d , and use the randomization⁸ $ilde{u}_j := (u_j + u)$ mod 1, j = 1, ..., N. It is not difficult to show that (marginal) distribution of each random vector \tilde{u}_j is uniform on I^d . Therefore each $\psi(\tilde{u}_j)$, and hence $N^{-1} \sum_{j=1}^N \psi(\tilde{u}_j)$, is an unbiased estimator of the corresponding expectation $\mathbb{E}[\psi(U)]$. Variance of the estimator $N^{-1}\sum_{j=1}^{N}\psi(\tilde{u}_{j})$ can be significantly smaller than variance of the corresponding Monte Carlo estimator based on samples of the same size. This randomization procedure can be applied in batches. That is, it can be repeated M times for independently generated uniformly distributed vectors $u = u^i$, i = 1, ..., M, and consequently averaging the obtained replications of $N^{-1} \sum_{j=1}^{N} \psi(\tilde{u}_j)$. Simultaneously, variance of this estimator can be evalused by calculating the sample variance of the obtained M independent replications of $N^{-1} \sum_{j=1}^{N} \psi(\tilde{u}_j)$.

5.5 Variance Reduction Techniques

Consider the sample average estimators $\hat{f}_N(x)$. We have that if the sample is iid, then the variance of $\hat{f}_N(x)$ is equal to $\sigma^2(x)/N$, where $\sigma^2(x) := \operatorname{Var}[F(x,\xi)]$. In some cases it is possible to reduce the variance of generated sample averages, which in turn enhances convergence of the corresponding SAA estimators. In section 5.4 we discussed Quasi-Monte Carlo techniques for enhancing rates of convergence of sample average approximations. In this section we briefly discuss some other variance reduction techniques which seem to be useful in the SAA method.

5.5.1 Latin Hypercube Sampling

Suppose that the random data vector $\xi = \xi(\omega)$ is one dimensional with the corresponding cumulative distribution function (cdf) $H(\cdot)$. We can write then

$$\mathbb{E}[F(x,\xi)] = \int_{-\infty}^{+\infty} F(x,\xi) dH(\xi).$$
(5.154)

In order to evaluate the above integral numerically it will be much better to generate sample points evenly distributed than to use an iid sample (we already discussed this in section 5.4).

⁸For a number $a \in \mathbb{R}$ the notation " $a \mod 1$ " denotes the fractional part of a, i.e., $a \mod 1 = a - \lfloor a \rfloor$, where $\lfloor a \rfloor$ denotes the largest integer less than or equal to a. In vector case the "modulo 1" reduction is understood coordinatewise.

That is, we can generate independent random points⁹

$$U^{j} \sim U[(j-1)/N, j/N], \quad j = 1, ..., N,$$
 (5.155)

and then to construct the random sample of ξ by the inverse transformation $\xi^j := H^{-1}(U^j)$, j = 1, ..., N (compare with (5.141)).

Now suppose that j is chosen at random from the set $\{1, ..., N\}$ (with equal probability for each element of that set). Then conditional on j, the corresponding random variable U^j is uniformly distributed on the interval [(j-1)/N, j/N], and the unconditional distribution of U^j is uniform on the interval [0, 1]. Consequently, let $\{j_1, ..., j_N\}$ be a random permutation of the set $\{1, ..., N\}$. Then the random variables $\xi^{j_1}, ..., \xi^{j_N}$ have the same marginal distribution, with the same cdf $H(\cdot)$, and are negatively correlated with each other. Therefore the expected value of

$$\hat{f}_N(x) = \frac{1}{N} \sum_{i=1}^N F(x,\xi^j) = \frac{1}{N} \sum_{s=1}^N F(x,\xi^{j_s})$$
(5.156)

is f(x), while

$$\mathbb{V}\mathrm{ar}\left[\hat{f}_{N}(x)\right] = N^{-1}\sigma^{2}(x) + 2N^{-2}\sum_{s < t} \mathbb{C}\mathrm{ov}\left(F(x,\xi^{j_{s}}), F(x,\xi^{j_{t}})\right).$$
(5.157)

If the function $F(x, \cdot)$ is monotonically increasing or decreasing, than the random variables $F(x, \xi^{j_s})$ and $F(x, \xi^{j_t})$, $s \neq t$, are also negatively correlated. Therefore, the variance of $\hat{f}_N(x)$ tends to be smaller, and in some cases much smaller, than $\sigma^2(x)/N$.

Suppose now that the random vector $\xi = (\xi_1, ..., \xi_d)$ is *d*-dimensional, and that its components ξ_i , i = 1, ..., d, are distributed independently of each other. Then we can use the above procedure for each component ξ_i . That is, a random sample U^j of the form (5.155) is generated, and consequently N replications of the first component of ξ are computed by the corresponding inverse transformation applied to randomly permuted U^{j_s} . The same procedure is applied to every component of ξ with the corresponding random samples of the form (5.155) and random permutations generated independently of each other. This sampling scheme is called the *Latin Hypercube* (LH) sampling.

If the function $F(x, \cdot)$ is decomposable, i.e., $F(x, \xi) := F_1(x, \xi_1) + ... + F_d(x, \xi_d)$, then $\mathbb{E}[F(x, \xi)] = \mathbb{E}[F_1(x, \xi_1)] + ... + \mathbb{E}[F_d(x, \xi_d)]$, where each expectation is calculated with respect to a one dimensional distribution. In that case the LH sampling ensures that each expectation $\mathbb{E}[F_i(x, \xi_i)]$ is estimated in nearly optimal way. Therefore, the LH sampling works especially well in cases where the function $F(x, \cdot)$ tends to have a somewhat decomposable structure. In any case the LH sampling procedure is easy to implement and can be applied to SAA optimization procedures in a straightforward way. Since in LH sampling the random replications of $F(x, \xi)$ are correlated with each other, one cannot use variance estimates like (5.21). Therefore the LH method usually is applied in several independent batches in order to estimate variance of the corresponding estimators.

⁹For an interval $[a,b] \subset \mathbb{R}$, we denote by U[a,b] the uniform probability distribution on that interval.

5.5.2 Linear Control Random Variables Method

Suppose that we have a measurable function $A(x,\xi)$ such that $\mathbb{E}[A(x,\xi)] = 0$ for all $x \in X$. Then, for any $t \in \mathbb{R}$, the expected value of $F(x,\xi) + tA(x,\xi)$ is f(x), while

$$\operatorname{\mathbb{V}ar}\left[F(x,\xi) + tA(x,\xi)\right] = \operatorname{\mathbb{V}ar}\left[F(x,\xi)\right] + t^2 \operatorname{\mathbb{V}ar}\left[A(x,\xi)\right] + 2t \operatorname{\mathbb{C}ov}\left(F(x,\xi), A(x,\xi)\right).$$

It follows that the above variance attains its minimum, with respect to t, for

$$t^* := -\rho_{F,A}(x) \left[\frac{\mathbb{V}ar(F(x,\xi))}{\mathbb{V}ar(A(x,\xi))} \right]^{1/2},$$
(5.158)

where $\rho_{F,A}(x) := \mathbb{C}\operatorname{orr}(F(x,\xi), A(x,\xi))$, and with

$$\mathbb{V}\mathrm{ar}\big[F(x,\xi) + t^* A(x,\xi)\big] = \mathbb{V}\mathrm{ar}\,[F(x,\xi)]\,\big[1 - \rho_{F,A}(x)^2\big]. \tag{5.159}$$

For a given $x \in X$ and generated sample $\xi^1, ..., \xi^N$, one can estimate, in the standard way, the covariance and variances appearing in the right hand side of (5.158), and hence to construct an estimate \hat{t} of t^* . Then f(x) can be estimated by

$$\hat{f}_N^A(x) := \frac{1}{N} \sum_{j=1}^N \left[F(x,\xi^j) + \hat{t}A(x,\xi^j) \right].$$
(5.160)

By (5.159), the *linear control* estimator $\hat{f}_N^A(x)$ has a smaller variance than $\hat{f}_N(x)$ if $F(x,\xi)$ and $A(x,\xi)$ are highly correlated with each other.

Let us make the following observations. The estimator \hat{t} , of the optimal value t^* , depends on x and the generated sample. Therefore, it is difficult to apply linear control estimators in an SAA optimization procedure. That is, linear control estimators are mainly suitable for estimating expectations at a fixed point. Also if the same sample is used in estimating \hat{t} and $\hat{f}_N^A(x)$, then $\hat{f}_N^A(x)$ can be a slightly biased estimator of f(x).

Of course, the above Linear Control procedure can be successful only if a function $A(x,\xi)$, with mean zero and highly correlated with $F(x,\xi)$, is available. Choice of such a function is problem dependent. For instance, one can use a linear function $A(x,\xi) := \lambda(\xi)^T x$. Consider, for example, two-stage stochastic programming problems with recourse of the form (2.1)–(2.2). Suppose that the random vector $h = h(\omega)$ and matrix $T = T(\omega)$, in the second stage problem (2.2), are independently distributed, and let $\mu := \mathbb{E}[h]$. Then

$$\mathbb{E}\left[(h-\mu)^{\mathsf{T}}T\right] = \mathbb{E}\left[(h-\mu)\right]^{\mathsf{T}}\mathbb{E}\left[T\right] = 0,$$

and hence one can use $A(x,\xi) := (h - \mu)^{\mathsf{T}} T x$ as the control variable.

Let us finally remark that the above procedure can be extended in a straightforward way to a case where several functions $A_1(x,\xi), ..., A_m(x,\xi)$, each with zero mean and highly correlated with $F(x,\xi)$, are available.

5.5.3 Importance Sampling and Likelihood Ratio Methods

Suppose that ξ has a continuous distribution with probability density function (pdf) $h(\cdot)$. Let $\psi(\cdot)$ be another pdf such that the so-called *likelihood ratio* function $L(\cdot) := \frac{h(\cdot)}{\psi(\cdot)}$ is well defined. That is, if $\psi(z) = 0$ for some $z \in \mathbb{R}^d$, then h(z) = 0, and by the definition, 0/0 = 0, i.e., we do not divide a positive number by zero. Then we can write

$$f(x) = \int F(x,\xi)h(\xi)d\xi = \int F(x,\zeta)L(\zeta)\psi(\zeta)d\zeta = \mathbb{E}_{\psi}[F(x,Z)L(Z)], \quad (5.161)$$

where the integration is performed over the space \mathbb{R}^d and the notation \mathbb{E}_{ψ} emphasizes that the expectation is taken with respect to the random vector Z having pdf $\psi(\cdot)$.

Let us show that, for a fixed x, the variance of F(x, Z)L(Z) attains its minimal value for $\psi(\cdot)$ proportional to $|F(x, \cdot)h(\cdot)|$, i.e., for

$$\psi^*(\cdot) := \frac{|F(x, \cdot)h(\cdot)|}{\int |F(x, \zeta)h(\zeta)|d\zeta}.$$
(5.162)

Since $\mathbb{E}_{\psi}[F(x,Z)L(Z)] = f(x)$ and does not depend on $\psi(\cdot)$, we have that the variance of F(x,Z)L(Z) is minimized if

$$\mathbb{E}_{\psi}[F(x,Z)^{2}L(Z)^{2}] = \int \frac{F(x,\zeta)^{2}h(\zeta)^{2}}{\psi(\zeta)}d\zeta,$$
(5.163)

is minimized. Furthermore, by Cauchy inequality we have

$$\left(\int |F(x,\zeta)h(\zeta)|d\zeta\right)^2 \le \left(\int \frac{F(x,\zeta)^2 h(\zeta)^2}{\psi(\zeta)} d\zeta\right) \left(\int \psi(\zeta)d\zeta\right).$$
(5.164)

It remains to note that $\int \psi(\zeta) d\zeta = 1$ and the left hand side of (5.164) is equal to the expected value of squared F(x, Z)L(Z) for $\psi(\cdot) = \psi^*(\cdot)$.

Note that if $F(x, \cdot)$ is nonnegative valued, then $\psi^*(\cdot) = F(x, \cdot)h(\cdot)/f(x)$ and for that choice of the pdf $\psi(\cdot)$, the function $F(x, \cdot)L(\cdot)$ is identically equal to f(x). Of course, in order to achieve such absolute variance reduction to zero we need to know the expectation f(x) which was our goal in the first place. Nevertheless, it gives the idea that if we can construct a pdf $\psi(\cdot)$ roughly proportional to $|F(x, \cdot)h(\cdot)|$, then we may achieve a considerable variance reduction by generating a random sample $\zeta^1, ..., \zeta^N$ from the pdf $\psi(\cdot)$, and then estimating f(x) by

$$\tilde{f}_{N}^{\psi}(x) := \frac{1}{N} \sum_{j=1}^{N} F(x, \zeta^{j}) L(\zeta^{j}).$$
(5.165)

The estimator $\tilde{f}_N^{\psi}(x)$ is an unbiased estimator of f(x) and may have significantly smaller variance than $\hat{f}_N(x)$ depending on a successful choice of the pdf $\psi(\cdot)$.

Similar analysis can be performed in cases where ξ has a discrete distribution by replacing the integrals with the corresponding summations.

Let us remark that the above approach, called *importance sampling*, is extremely sensitive to a choice of the pdf $\psi(\cdot)$ and is notorious for its instability. This is understandable since the likelihood ratio function in the tail is the ratio of two very small numbers. For a successful choice of $\psi(\cdot)$, the method may work very well while even a small perturbation of $\psi(\cdot)$ may be disastrous. This is why a single choice of $\psi(\cdot)$ usually does not work for

different points x, and consequently cannot be used for a whole optimization procedure. Note also that $\mathbb{E}_{\psi}[L(Z)] = 1$. Therefore, $L(\zeta) - 1$ can be used as a linear control variable for the likelihood ratio estimator $\tilde{f}_{N}^{\psi}(x)$.

In some cases it is also possible to use the likelihood ratio method for estimating first and higher order derivatives of f(x). Consider, for example, the optimal value $Q(x,\xi)$ of the second stage linear program (2.2). Suppose that the vector q and matrix W are fixed, i.e., not stochastic, and for the sake of simplicity that $h = h(\omega)$ and $T = T(\omega)$ are distributed independently of each other. We have then that $Q(x,\xi) = Q(h - Tx)$, where

$$\mathcal{Q}(z) := \inf \left\{ q^{\mathsf{T}} y : W y = z, \ y \ge 0 \right\}.$$

Suppose, further, that h has a continuous distribution with pdf $\eta(\cdot)$. We have that

$$\mathbb{E}[Q(x,\xi)] = \mathbb{E}_T \left\{ \mathbb{E}_{h|T}[Q(x,\xi)] \right\},\$$

and by using the transformation z = h - Tx, since h and T are independent we obtain

$$\mathbb{E}_{h|T}[Q(x,\xi)] = \mathbb{E}_{h}[Q(x,\xi)] \\
= \int \mathcal{Q}(h-Tx)\eta(h)dh = \int \mathcal{Q}(z)\eta(z+Tx)dz \\
= \int \mathcal{Q}(\zeta)L(x,\zeta)\psi(\zeta)d\zeta = \mathbb{E}_{\psi}\left[L(x,Z)\mathcal{Q}(Z)\right],$$
(5.166)

where $\psi(\cdot)$ is a chosen pdf and $L(x,\zeta) := \eta(\zeta + Tx)/\psi(\zeta)$. If the function $\eta(\cdot)$ is smooth, then the likelihood ratio function $L(\cdot,\zeta)$ is also smooth. In that case, under mild additional conditions, first and higher order derivatives can be taken inside the expected value in the right hand side of (5.166), and consequently can be estimated by sampling. Note that the first order derivatives of $Q(\cdot,\xi)$ are piecewise constant, and hence its second order derivatives are zeros whenever defined. Therefore, second order derivatives cannot be taken inside the expectation $\mathbb{E}[Q(x,\xi)]$ even if ξ has a continuous distribution.

5.6 Validation Analysis

Suppose that we are given a feasible point $\bar{x} \in \mathcal{X}$ as a candidate for an optimal solution of the true problem. For example, \bar{x} can be an output of a run of the corresponding SAA problem. In this section we discuss ways to evaluate quality of this candidate solution. This is important, in particular, for a choice of the sample size and stopping criteria in simulation based optimization. There are basically two approaches to such validation analysis. We can either try to estimate the optimality gap $f(\bar{x}) - \vartheta^*$ between the objective value at the considered point \bar{x} and the optimal value of the true problem, or to evaluate first order (KKT) optimality conditions at \bar{x} .

Let us emphasize that the following analysis is designed for the situations where the value $f(\bar{x})$, of the true objective function at the considered point, is *finite*. In the case of two stage programming this requires, in particular, that the second stage problem, associated with first stage decision vector \bar{x} , is feasible for almost every realization of the random data.

5.6.1 Estimation of the Optimality Gap

In this section we consider the problem of estimating the optimality gap

$$\operatorname{gap}(\bar{x}) := f(\bar{x}) - \vartheta^* \tag{5.167}$$

associated with the candidate solution \bar{x} . Clearly, for any feasible $\bar{x} \in \mathcal{X}$, gap (\bar{x}) is non-negative and gap $(\bar{x}) = 0$ iff \bar{x} is an optimal solution of the true problem.

Consider the optimal value $\hat{\vartheta}_N$ of the SAA problem (5.2). We have that $\vartheta^* \geq \mathbb{E}[\hat{\vartheta}_N]$ (see the discussion following equation (5.22)). This means that $\hat{\vartheta}_N$ provides a valid statistical lower bound for the optimal value ϑ^* of the true problem. The expectation $\mathbb{E}[\hat{\vartheta}_N]$ can be estimated by averaging. That is, one can solve M times sample average approximation problems based on independently generated samples each of size N. Let $\hat{\vartheta}_N^1, ..., \hat{\vartheta}_N^M$ be the computed optimal values of these SAA problems. Then

$$\bar{v}_{N,M} := \frac{1}{M} \sum_{m=1}^{M} \hat{\vartheta}_N^m$$
(5.168)

is an unbiased estimator of $\mathbb{E}[\hat{\vartheta}_N]$. Since the samples, and hence $\hat{\vartheta}_N^1, ..., \hat{\vartheta}_N^M$, are independent and have the same distribution, we have that $\mathbb{V}ar[\bar{v}_{N,M}] = M^{-1}\mathbb{V}ar[\hat{\vartheta}_N]$, and hence we can estimate variance of $\bar{v}_{N,M}$ by

$$\hat{\sigma}_{N,M}^{2} := \frac{1}{M} \Big[\underbrace{\frac{1}{M-1} \sum_{m=1}^{M} \left(\hat{\vartheta}_{N}^{m} - \bar{v}_{N,M} \right)^{2}}_{\text{estimate of } \mathbb{V}ar[\hat{\vartheta}_{N}]} \Big].$$
(5.169)

Note that the above make sense only if the optimal value ϑ^* of the true problem is finite. Note also that the inequality $\vartheta^* \ge \mathbb{E}[\hat{\vartheta}_N]$ holds and $\hat{\vartheta}_N$ gives a valid statistical lower bound even if $f(x) = +\infty$ for some $x \in \mathcal{X}$. Note finally that the samples do not need to be iid (for example, one can use Latin Hypercube sampling), they only should be independent of each other in order to use estimate (5.169) of the corresponding variance.

In general the random variable ϑ_N , and hence its replications ϑ_N^j , does not have a normal distribution, even approximately (see Theorem 5.7 and the following afterwards discussion). However, by the CLT, the probability distribution of the average $\bar{v}_{N,M}$ becomes approximately normal as M increases. Therefore, we can use

$$L_{N,M} := \bar{v}_{N,M} - t_{\alpha,M-1}\hat{\sigma}_{N,M}$$
(5.170)

as an approximate $100(1-\alpha)\%$ confidence¹⁰ lower bound for the expectation $\mathbb{E}[\hat{\vartheta}_N]$.

We can also estimate $f(\bar{x})$ by sampling. That is, let $\hat{f}_{N'}(\bar{x})$ be the sample average estimate of $f(\bar{x})$, based on a sample of size N' generated independently of samples involved in computing \bar{x} . Let $\hat{\sigma}_{N'}^2(\bar{x})$ be an estimate of the variance of $\hat{f}_{N'}(\bar{x})$. In the case of iid sample, one can use the sample variance estimate

$$\hat{\sigma}_{N'}^2(\bar{x}) := \frac{1}{N'(N'-1)} \sum_{j=1}^{N'} \left[F(\bar{x},\xi^j) - \hat{f}_{N'}(\bar{x}) \right]^2.$$
(5.171)

¹⁰Here $t_{\alpha,\nu}$ is the α -critical value of t-distribution with ν degrees of freedom. This critical value is slightly bigger than the corresponding standard normal critical value z_{α} , and $t_{\alpha,\nu}$ quickly approaches z_{α} as ν increases.

Then

$$U_{N'}(\bar{x}) := \hat{f}_{N'}(\bar{x}) + z_{\alpha}\hat{\sigma}_{N'}(\bar{x})$$
(5.172)

gives an approximate $100(1 - \alpha)\%$ confidence upper bound for $f(\bar{x})$. Note that since N' typically is large, we use here critical value z_{α} from the standard normal distribution rather than a t-distribution.

We have that

$$\mathbb{E}\left[\hat{f}_{N'}(\bar{x}) - \bar{v}_{N,M}\right] = f(\bar{x}) - \mathbb{E}[\hat{\vartheta}_N] = \operatorname{gap}(\bar{x}) + \vartheta^* - \mathbb{E}[\hat{\vartheta}_N] \ge \operatorname{gap}(\bar{x}),$$

i.e., $f_{N'}(\bar{x}) - \bar{v}_{N,M}$ is a biased estimator of the gap (\bar{x}) . Also the variance of this estimator is equal to the sum of the variances of $\hat{f}_{N'}(\bar{x})$ and $\bar{v}_{N,M}$, and hence

$$\hat{f}_{N'}(\bar{x}) - \bar{v}_{N,M} + z_{\alpha} \sqrt{\hat{\sigma}_{N'}^2(\bar{x}) + \hat{\sigma}_{N,M}^2}$$
(5.173)

provides a conservative $100(1-\alpha)\%$ confidence upper bound for the gap (\bar{x}) . We say that this upper bound is "conservative" since in fact it gives a $100(1-\alpha)\%$ confidence upper bound for the gap $(\bar{x}) + \vartheta^* - \mathbb{E}[\hat{\vartheta}_N]$, and we have that $\vartheta^* - \mathbb{E}[\hat{\vartheta}_N] \ge 0$.

In order to calculate the estimate $f_{N'}(\bar{x})$ one needs to compute the value $F(\bar{x}, \xi^j)$ of the objective function for every generated sample realization ξ^j , j = 1, ..., N'. Typically it is much easier to compute $F(\bar{x}, \xi)$, for a given $\xi \in \Xi$, than to solve the corresponding SAA problem. Therefore, often one can use a relatively large sample size N', and hence to estimate $f(\bar{x})$ quite accurately. Evaluation of the optimal value ϑ^* by employing the estimator $\bar{v}_{N,M}$ is a more delicate problem.

There are two types of errors in using $\bar{v}_{N,M}$ as an estimator of ϑ^* , namely the bias $\vartheta^* - \mathbb{E}[\hat{\vartheta}_N]$ and variability of $\bar{v}_{N,M}$ measured by its variance. Both errors can be reduced by increasing N, and the variance can be reduced by increasing N and M. Note, however, that the computational effort in computing $\bar{v}_{N,M}$ is proportional to M, since the corresponding SAA problems should be solved M times, and to the computational time for solving a single SAA problem based on a sample of size N. Naturally one may ask what is the best way of distributing computational resources between increasing the sample size N and the number of repetitions M. This question is, of course, problem dependent. In cases where computational complexity of SAA problems grows fast with increase of the sample size N, it may be more advantageous to use a larger number of repetitions M. On the other hand it was observed empirically that the computational effort in solving SAA problems by 'good' subgradient algorithms grows only *linearly* with the sample size N. In such cases one can use a larger N and make only a few repetitions M in order to estimate the variance of $\bar{v}_{N,M}$.

The bias $\vartheta^* - \mathbb{E}[\hat{\vartheta}_N]$ does not depend on M, of course. It was shown in Proposition 5.6 that if the sample is iid, then $\mathbb{E}[\hat{\vartheta}_N] \leq \mathbb{E}[\hat{\vartheta}_{N+1}]$ for any $N \in \mathbb{N}$. It follows that the bias $\vartheta^* - \mathbb{E}[\hat{\vartheta}_N]$ decreases monotonically with increase of the sample size N. By Theorem 5.7 we have that, under mild regularity conditions,

$$\hat{\vartheta}_N = \inf_{x \in \mathcal{S}} \hat{f}_N(x) + o_p(N^{-1/2}).$$
(5.174)

Consequently, if the set S of optimal solutions of the true problem is not a singleton, then the bias $\vartheta^* - \mathbb{E}[\hat{\vartheta}_N]$ typically converges to zero, as N increases, at a rate of $O(N^{-1/2})$, and tends to be bigger for a larger set S (see equation (5.29) and the following afterwards discussion). On the other hand, in well conditioned problems, where the optimal set S is a singleton, the bias typically is of order $O(N^{-1})$ (see Theorem 5.8), and the bias tends to be of a lesser problem. Moreover, if the true problem has a sharp optimal solution x^* , then the event " $\hat{x}_N = x^*$ ", and hence the event " $\hat{\vartheta}_N = \hat{f}_N(x^*)$ ", happens with probability approaching one exponentially fast (see Theorem 5.23). Since $\mathbb{E}[\hat{f}_N(x^*)] = f(x^*)$, in such cases the bias $\vartheta^* - \mathbb{E}[\hat{\vartheta}_N] = f(x^*) - \mathbb{E}[\hat{\vartheta}_N]$ tends to be much smaller.

In the above approach the upper and lower statistical bounds were computed independently of each other. Alternatively, it is possible to use the same sample for estimating $f(\bar{x})$ and $\mathbb{E}[\hat{\vartheta}_N]$. That is, for M generated samples each of size N, the gap is estimated by

$$\widehat{\operatorname{gap}}_{N,M}(\bar{x}) := \frac{1}{M} \sum_{m=1}^{M} \left[\widehat{f}_N^m(\bar{x}) - \widehat{\vartheta}_N^m \right],$$
(5.175)

where $\hat{f}_N^m(\bar{x})$ and $\hat{\vartheta}_N^m$ are computed from the same sample m = 1, ..., M. We have that the expected value of $\widehat{\text{gap}}_{N,M}(\bar{x})$ is $f(\bar{x}) - \mathbb{E}[\hat{\vartheta}_N]$, i.e., the estimator $\widehat{\text{gap}}_{N,M}(\bar{x})$ has the same bias as $\hat{f}_N(\bar{x}) - \bar{v}_{N,M}$. On the other hand, for a problem with sharp optimal solution x^* it happens with high probability that $\hat{\vartheta}_N^m = \hat{f}_N^m(x^*)$, and as a consequence $\hat{f}_N^m(\bar{x})$ tends to be highly positively correlated with $\hat{\vartheta}_N^m$, provided that \bar{x} is close to x^* . In such cases variability of $\widehat{\text{gap}}_{N,M}(\bar{x})$ can be considerably smaller than variability of $\hat{f}_{N'}(\bar{x}) - \bar{v}_{N,M}$. This is the idea of common random number generated estimators.

Remark 19. Of course, in order to obtain a valid statistical lower bound for the optimal value ϑ^* we can use any (deterministic) lower bound for the optimal value ϑ_N of the corresponding SAA problem instead of ϑ_N itself. For example, suppose that the problem is *convex*. By convexity of $\hat{f}_N(\cdot)$ we have that for any $x' \in \mathcal{X}$ and $\gamma \in \partial \hat{f}_N(x')$ it holds that

$$\hat{f}_N(x) \ge \hat{f}_N(x') + \gamma^{\mathsf{T}}(x - x'), \quad \forall x \in \mathbb{R}^n.$$
(5.176)

Therefore, we can proceed as follows. Choose points $x_1, ..., x_r \in \mathcal{X}$, calculate subgradients $\hat{\gamma}_{iN} \in \partial \hat{f}_N(x_i)$, i = 1, ..., r, and solve the problem

$$\operatorname{Min}_{x \in \mathcal{X}} \max_{1 \le i \le r} \left\{ \widehat{f}_N(x_i) + \widehat{\gamma}_{iN}^{\mathsf{T}}(x - x_i) \right\}.$$
(5.177)

Denote by $\hat{\lambda}_N$ the optimal value of (5.177). By (5.176) we have that $\hat{\lambda}_N$ is less than or equal to the optimal value $\hat{\vartheta}_N$ of the corresponding SAA problem, and hence gives a valid statistical lower bound for ϑ^* . A possible advantage of $\hat{\lambda}_N$ over $\hat{\vartheta}_N$ is that it could be easier to solve (5.177) than the corresponding SAA problem. For instance, if the set \mathcal{X} is polyhedral, then (5.177) can be formulated as a linear programming problem.

Of course, this approach raises the question of how to choose the points $x_1, ..., x_r \in \mathcal{X}$. Suppose that the expectation function f(x) is differentiable at the points $x_1, ..., x_r$. Then for any choice of $\hat{\gamma}_{iN} \in \partial \hat{f}_N(x_i)$ we have that subgradients $\hat{\gamma}_{iN}$ converge to $\nabla f(x_i)$ w.p.1. Therefore $\hat{\lambda}_N$ converges w.p.1 to the optimal value of the problem

$$\underset{x \in \mathcal{X}}{\operatorname{Min}} \max_{1 \le i \le r} \left\{ f(x_i) + \nabla f(x_i)^{\mathsf{T}} (x - x_i) \right\},$$
(5.178)

provided that the set \mathcal{X} is bounded. Again by convexity arguments the optimal value of (5.178) is less than or equal to the optimal value ϑ^* of the true problem. If we can find such points $x_1, ..., x_r \in \mathcal{X}$ that the optimal value of (5.178) is less than ϑ^* by a 'small' amount, then it could be advantageous to use $\hat{\lambda}_N$ instead of $\hat{\vartheta}_N$. We also should keep in mind that the number r should be relatively small, otherwise we may loose the advantage of solving the 'easier' problem (5.177).

A natural approach to choosing the required points and hence to apply the above procedure is the following. By solving (once) an SAA problem, find points $x_1, ..., x_r \in \mathcal{X}$ such that the optimal value of the corresponding problem (5.177) provides with a high accuracy an estimate of the optimal value of this SAA problem. Use some (all) these points to calculate lower bound estimates $\hat{\lambda}_N^m, m = 1, ..., M$, probably with a larger sample size N. Calculate the average $\bar{\lambda}_{N,M}$ together with the corresponding sample variance and construct the associated $100(1 - \alpha)\%$ confidence lower bound similar to (5.170).

Estimation of Optimality Gap of Minimax and Expectation-Constrained Problems

Consider a minimax problem of the form (5.46). Let ϑ^* be the optimal value of this (true) minimax problem. Clearly for any $\bar{y} \in \mathcal{Y}$ we have that

$$\vartheta^* \ge \inf_{x \in \mathcal{X}} f(x, \bar{y}). \tag{5.179}$$

Now for the optimal value of the right hand side of (5.179) we can construct a valid statistical lower bound, and hence a valid statistical lower bound for ϑ^* , as before by solving the corresponding SAA problems several times and averaging calculated optimal values. Suppose, further, that the minimax problem (5.46) has a nonempty set $S_x \times S_y \subset \mathcal{X} \times \mathcal{Y}$ of saddle points, and hence its optimal value is equal to the optimal value of its dual problem (5.48). Then for any $\bar{x} \in \mathcal{X}$ we have that

$$\vartheta^* \le \sup_{y \in \mathcal{Y}} f(\bar{x}, y), \tag{5.180}$$

and the equalities in (5.179) and/or (5.180) hold iff $\bar{y} \in S_y$ and/or $\bar{x} \in S_x$. By (5.180) we can construct a valid statistical upper bound for ϑ^* by averaging optimal values of sample average approximations of the right hand side of (5.180). Of course, quality of these bounds will depend on a good choice of the points \bar{y} and \bar{x} . A natural construction for the candidate solutions \bar{y} and \bar{x} will be to use optimal solutions of a run of the corresponding minimax SAA problem (5.47).

Similar ideas can be applied to validation of stochastic problems involving constraints given as expected value functions (see equations (5.11)–(5.13)). That is, consider the following problem

$$\underset{x \in \mathcal{X}_0}{\operatorname{Min}} f(x) \text{ subject to } g_i(x) \le 0, \ i = 1, \dots, p,$$
(5.181)

where \mathcal{X}_0 is a nonempty subset of \mathbb{R}^n , $f(x) := \mathbb{E}[F(x,\xi)]$ and $g_i(x) := \mathbb{E}[G_i(x,\xi)]$, i = 1, ..., p. We have that

$$\vartheta^* = \inf_{x \in \mathcal{X}_0} \sup_{\lambda \ge 0} L(x, \lambda), \tag{5.182}$$

where ϑ^* is the optimal value and $L(x, \lambda) := f(x) + \sum_{i=1}^p \lambda_i g_i(x)$ is the Lagrangian of problem (5.181). Therefore, for any $\overline{\lambda} \ge 0$, we have that

$$\vartheta^* \ge \inf_{x \in \mathcal{X}_0} L(x, \bar{\lambda}),$$
(5.183)

and the equality in (5.183) is attained if the problem (5.179) is convex and $\overline{\lambda}$ is a Lagrange multipliers vector satisfying the corresponding first-order optimality conditions. Of course, a statistical lower bound for the optimal value of the problem in the right hand side of (5.183), is also a statistical lower bound for ϑ^* .

Unfortunately an upper bound which can be obtained by interchanging the 'inf' and 'sup' operators in (5.182) cannot be used in a straightforward way. This is because if, for a chosen $\bar{x} \in \mathcal{X}_0$, it happens that $\hat{g}_{iN}(\bar{x}) > 0$ for some $i \in \{1, ..., p\}$, then

$$\sup_{\lambda \ge 0} \left\{ \hat{f}_N(\bar{x}) + \sum_{i=1}^p \lambda_i \hat{g}_{iN}(\bar{x}) \right\} = +\infty.$$
 (5.184)

Of course, in such a case the obtained upper bound $+\infty$ is useless. This typically will be the case if \bar{x} is constructed as a solution of an SAA problem and some of the SAA constraints are active at \bar{x} . Note also that if $\hat{g}_{iN}(\bar{x}) \leq 0$ for all $i \in \{1, ..., p\}$, then the supremum in the left hand side of (5.184) is equal to $\hat{f}_N(\bar{x})$.

If we can ensure, with a high probability $1-\alpha$, that a chosen point \bar{x} is a feasible point of the true problem 5.179), then we can construct an upper bound by estimating $f(\bar{x})$ using a relatively large sample. This, in turn, can be approached by verifying, for an independent sample of size N', that $\hat{g}_{iN'}(\bar{x}) + \kappa \hat{\sigma}_{iN'}(\bar{x}) \leq 0$, i = 1, ..., p, where $\hat{\sigma}_{iN'}^2(\bar{x})$ is a sample variance of $\hat{g}_{iN'}(\bar{x})$ and κ is a positive constant chosen in such a way that the probability of $g_i(\bar{x})$ being bigger that $\hat{g}_{iN'}(\bar{x}) + \kappa \hat{\sigma}_{iN'}(\bar{x})$ is less than α/p for all $i \in \{1, ..., p\}$.

5.6.2 Statistical Testing of Optimality Conditions

Suppose that the feasible set \mathcal{X} is defined by (equality and inequality) constraints in the form

$$\mathcal{X} := \left\{ x \in \mathbb{R}^n : g_i(x) = 0, \ i = 1, ..., q; \ g_i(x) \le 0, \ i = q + 1, ..., p \right\},$$
(5.185)

where $g_i(x)$ are smooth (at least continuously differentiable) *deterministic* functions. Let $x^* \in \mathcal{X}$ be an optimal solution of the true problem and suppose that the expected value function $f(\cdot)$ is differentiable at x^* . Then, under a constraint qualification, first order (KKT) optimality conditions hold at x^* . That is, there exist Lagrange multipliers λ_i such that $\lambda_i \geq 0$, $i \in \mathcal{I}(x^*)$ and

$$\nabla f(x^*) + \sum_{i \in \mathcal{J}(x^*)} \lambda_i \nabla g_i(x^*) = 0, \qquad (5.186)$$

where $\mathcal{I}(x) := \{i : g_i(x) = 0, i = q + 1, ..., p\}$ denotes the index set of inequality constraints active at a point $x \in \mathbb{R}^n$, and $\mathcal{J}(x) := \{1, ..., q\} \cup \mathcal{I}(x)$. Note that if the constraint

functions are linear, say $g_i(x) := a_i^{\mathsf{T}} x + b_i$, then $\nabla g_i(x) = a_i$ and the above KKT conditions hold without a constraint qualification. Consider the (polyhedral) cone

$$K(x) := \left\{ z \in \mathbb{R}^n : z = \sum_{i \in \mathcal{J}(x)} \alpha_i \nabla g_i(x), \ \alpha_i \le 0, \ i \in \mathcal{I}(x) \right\}.$$
 (5.187)

Then the KKT optimality conditions can be written in the form $\nabla f(x^*) \in K(x^*)$.

Suppose now that $f(\cdot)$ is differentiable at the candidate point $\bar{x} \in \mathcal{X}$, and that the gradient $\nabla f(\bar{x})$ can be estimated by a (random) vector $\gamma_N(\bar{x})$. In particular, if $F(\cdot,\xi)$ is differentiable at \bar{x} w.p.1, then we can use the estimator

$$\gamma_N(\bar{x}) := \frac{1}{N} \sum_{j=1}^N \nabla_x F(\bar{x}, \xi^j) = \nabla \hat{f}_N(\bar{x})$$
(5.188)

associated with the generated¹¹ random sample. Note that if, moreover, the derivatives can be taken inside the expectation, that is,

$$\nabla f(\bar{x}) = \mathbb{E}[\nabla_x F(\bar{x}, \xi)], \qquad (5.189)$$

then the above estimator is unbiased, i.e., $\mathbb{E}[\gamma_N(\bar{x})] = \nabla f(\bar{x})$. In the case of two-stage linear stochastic programming with recourse, formula (5.189) typically holds if the corresponding random data have a continuous distribution. On the other hand if the random data have a discrete distribution with a finite support, then the expected value function f(x) is piecewise linear and typically is nondifferentiable at an optimal solution.

Suppose, further, that $V_N := N^{1/2} [\gamma_N(\bar{x}) - \nabla f(\bar{x})]$ converges in distribution, as N tends to infinity, to multivariate normal with zero mean vector and covariance matrix Σ , written $V_N \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma)$. For the estimator $\gamma_N(\bar{x})$ defined in (5.188), this holds by the CLT if the interchangeability formula (5.189) holds, the sample is iid, and $\nabla_x F(\bar{x}, \xi)$ has finite second order moments. Moreover, in that case the covariance matrix Σ can be estimated by the corresponding sample covariance matrix

$$\hat{\Sigma}_N := \frac{1}{N-1} \sum_{j=1}^N \left[\nabla_x F(\bar{x}, \xi^j) - \nabla \hat{f}_N(\bar{x}) \right] \left[\nabla_x F(\bar{x}, \xi^j) - \nabla \hat{f}_N(\bar{x}) \right]^{\mathsf{T}}.$$
 (5.190)

Under the above assumptions, the sample covariance matrix $\hat{\Sigma}_N$ is an unbiased and consistent estimator of Σ .

We have that if $V_N \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma)$ and the covariance matrix Σ is nonsingular, then (given a consistent estimator $\hat{\Sigma}_N$ of Σ) the following holds

$$N(\gamma_N(\bar{x}) - \nabla f(\bar{x}))^{\mathsf{T}} \hat{\Sigma}_N^{-1} (\gamma_N(\bar{x}) - \nabla f(\bar{x})) \xrightarrow{\mathcal{D}} \chi_n^2,$$
(5.191)

¹¹We emphasize that the random sample in (5.188) is generated independently of the sample used to compute the candidate point \bar{x} .

where χ_n^2 denotes chi-square distribution with *n* degrees of freedom. This allows to construct the following (approximate) $100(1-\alpha)\%$ confidence region¹² for $\nabla f(\bar{x})$:

$$\left\{z \in \mathbb{R}^n : \left(\gamma_N(\bar{x}) - z\right)\right)^{\mathsf{T}} \hat{\mathcal{L}}_N^{-1} \left(\gamma_N(\bar{x}) - z\right)\right) \le \frac{\chi^2_{\alpha,n}}{N} \right\}.$$
 (5.192)

Consider the statistic

$$T_N := N \inf_{z \in K(\bar{x})} \left(\gamma_N(\bar{x}) - z \right)^{\mathsf{T}} \hat{\Sigma}_N^{-1} \left(\gamma_N(\bar{x}) - z \right).$$
(5.193)

Note that since the cone $K(\bar{x})$ is polyhedral and $\hat{\Sigma}_N^{-1}$ is positive definite, the minimization in the right hand side of (5.193) can be formulated as a quadratic programming problem, and hence can be solved by standard quadratic programming algorithms. We have that the confidence region, defined in (5.192), does not have common points with the cone $K(\bar{x})$ iff $T_N > \chi^2_{\alpha,n}$. We can also use the statistic T_N for testing the hypothesis:

$$H_0: \nabla f(\bar{x}) \in K(\bar{x})$$
 against the alternative $H_1: \nabla f(\bar{x}) \notin K(\bar{x})$. (5.194)

The T_N statistic represents the squared distance, with respect to the norm¹³ $\|\cdot\|_{\hat{\Sigma}_N^{-1}}$, from $N^{1/2}\gamma_N(\bar{x})$ to the cone $K(\bar{x})$. Suppose for the moment that only equality constraints are present in the definition (5.185) of the feasible set, and that the gradient vectors $\nabla g_i(\bar{x})$, i = 1, ..., q, are linearly independent. Then the set $K(\bar{x})$ forms a linear subspace of \mathbb{R}^n of dimension q, and the optimal value of the right hand side of (5.193) can be written in a closed form. Consequently, it is possible to show that T_N has asymptotically noncentral chi-square distribution with n - q degrees of freedom and the noncentrality parameter¹⁴

$$\delta := N \inf_{z \in K(\bar{x})} \left(\nabla f(\bar{x}) - z \right)^{\mathsf{T}} \Sigma^{-1} \left(\nabla f(\bar{x}) - z \right).$$
(5.195)

In particular, under H_0 we have that $\delta = 0$, and hence the null distribution of T_N is asymptotically central chi-square with n - q degrees of freedom.

Consider now the general case where the feasible set is defined by equality and inequality constraints as in (5.185). Suppose that the gradient vectors $\nabla g_i(\bar{x}), i \in \mathcal{J}(\bar{x})$, are linearly independent and that the *strict complementarity* condition holds at \bar{x} , that is, the Lagrange multipliers $\lambda_i, i \in \mathcal{I}(\bar{x})$, corresponding to the active at \bar{x} inequality constraints, are positive. Then for $\gamma_N(\bar{x})$ sufficiently close to $\nabla f(\bar{x})$ the minimizer in the right hand side of (5.193) will be lying in the linear space generated by vectors $\nabla g_i(\bar{x}), i \in \mathcal{J}(\bar{x})$. Therefore, in such case the null distribution of T_N is asymptotically central chi-square with $\nu := n - |\mathcal{J}(\bar{x})|$ degrees of freedom. Consequently, for a computed value T_N^* of the statistic T_N we can calculate (approximately) the corresponding *p*-value, which is equal

¹²Here $\chi^2_{\alpha,n}$ denotes the α -critical value of chi-square distribution with n degrees of freedom. That is, if $Y \sim \chi^2_n$, then $\Pr\{Y \ge \chi^2_{\alpha,n}\} = \alpha$.

¹³For a positive definite matrix A, the norm $\|\cdot\|_A$ is defined as $\|z\|_A := (z^T A z)^{1/2}$.

¹⁴Note that under the alternative (i.e., if $\nabla f(\bar{x}) \notin K(\bar{x})$), the noncentrality parameter δ tends to infinity as $N \to \infty$. Therefore, in order to justify the above asymptotics one needs a technical assumption known as Pitman's parameter drift.

to $\Pr{\{Y \ge T_N^*\}}$, where $Y \sim \chi_{\nu}^2$. This *p*-value gives an indication of the quality of the candidate solution \bar{x} with respect to the stochastic precision.

It should be understood that by accepting (i.e., failing to reject) H_0 , we do not claim that the KKT conditions hold exactly at \bar{x} . By accepting H_0 we rather assert that we cannot separate $\nabla f(\bar{x})$ from $K(\bar{x})$, given precision of the generated sample. That is, statistical error of the estimator $\gamma_N(\bar{x})$ is bigger than the squared $\|\cdot\|_{\Sigma^{-1}}$ -norm distance between $\nabla f(\bar{x})$ and $K(\bar{x})$. Also rejecting H_0 does not necessarily mean that \bar{x} is a poor candidate for an optimal solution of the true problem. The calculated value of T_N statistic can be large, i.e., the *p*-value can be small, simply because the estimated covariance matrix $N^{-1}\Sigma_N$ of $\gamma_N(\bar{x})$ is "small". In such cases, $\gamma_N(\bar{x})$ provides an accurate estimator of $\nabla f(\bar{x})$ with the corresponding confidence region (5.192) being "small". Therefore, the above p-value should be compared with the size of the confidence region (5.192), which in turn is defined by the size of the matrix $N^{-1}\hat{\Sigma}_N$ measured, for example, by its eigenvalues. Note also that it may happen that $|\mathcal{J}(\bar{x})| = n$, and hence $\nu = 0$. Under the strict complementarity condition, this means that $\nabla f(\bar{x})$ lies in the interior of the cone $K(\bar{x})$, which in turn is equivalent to the condition that $f'(\bar{x}, d) \ge c ||d||$ for some c > 0 and all $d \in \mathbb{R}^n$. Then, by the LD principle (see (7.217) in particular), the event $\gamma_N(\bar{x}) \in K(\bar{x})$ happens with probability approaching one exponentially fast.

Let us remark again that the above testing procedure is applicable if $F(\cdot,\xi)$ is differentiable at \bar{x} w.p.1 and the interchangeability formula (5.189) holds. This typically happens in cases where the corresponding random data have a continuous distribution.

5.7 Chance Constrained Problems

Consider a chance constrained problem of the form

$$\underset{x \in \mathcal{X}}{\operatorname{Min}} f(x) \text{ subject to } p(x) \le \alpha,$$
(5.196)

where $\mathcal{X} \subset \mathbb{R}^n$ is a closed set, $f : \mathbb{R}^n \to \mathbb{R}$ is a continuous function, $\alpha \in (0, 1)$ is a given significance level and

$$p(x) := \Pr\{C(x,\xi) > 0\}$$
(5.197)

is the probability that constraint is violated at point $x \in \mathcal{X}$. We assume that ξ is a random vector, whose probability distribution P is supported on set $\Xi \subset \mathbb{R}^d$, and the function $C : \mathbb{R}^n \times \Xi \to \mathbb{R}$ is a Carathéodory function. The chance constraint " $p(x) \leq \alpha$ " can be written equivalently in the form

$$\Pr\{C(x,\xi) \le 0\} \ge 1 - \alpha. \tag{5.198}$$

Let us also remark that several chance constraints

$$\Pr\{C_i(x,\xi) \le 0, \ i = 1, ..., q\} \ge 1 - \alpha, \tag{5.199}$$

can be reduced to one chance constraint (5.198) by employing the max-function $C(x, \xi) := \max_{1 \le i \le q} C_i(x, \xi)$. Of course, in some cases this may destroy a 'nice' structure of considered functions. At this point, however, this is not important.

5.7.1 Monte Carlo Sampling Approach

We discuss now a way of solving problem (5.196) by Monte Carlo sampling. For the sake of simplicity we assume that the objective function f(x) is given explicitly and only the chance constraints should be approximated.

We can write the probability p(x) in the form of the expectation,

$$p(x) = \mathbb{E}\left[\mathbf{1}_{(0,\infty)}(C(x,\xi))\right]$$

and to estimate this probability by the corresponding SAA function (compare with (5.14)–(5.16))

$$\hat{p}_N(x) := N^{-1} \sum_{j=1}^N \mathbf{1}_{(0,\infty)} \left(C(x,\xi^j) \right).$$
(5.200)

Recall that $\mathbf{1}_{(0,\infty)}(C(x,\xi))$ is equal to 1 if $C(x,\xi) > 0$, and it is equal 0 otherwise. Therefore, $\hat{p}_N(x)$ is equal to the proportion of times that $C(x,\xi^j) > 0$, j = 1, ..., N. Consequently we can write the corresponding SAA problem as

$$\underset{x \in \mathcal{X}}{\operatorname{Min}} f(x) \text{ subject to } \hat{p}_N(x) \le \alpha.$$
(5.201)

Proposition 5.29. Let $C(x,\xi)$ be a Carathéodory function. Then the functions p(x) and $\hat{p}_N(x)$ are lower semicontinuous. Suppose, further, that the sample is iid. Then $\hat{p}_N \xrightarrow{e} p$ w.p.1. Moreover, suppose that for every $x \in \mathcal{X}$ it holds that

$$\Pr\{\xi \in \Xi : C(x,\xi) = 0\} = 0, \tag{5.202}$$

i.e., $C(x,\xi) \neq 0$ w.p.1. Then the function p(x) is continuous on \mathcal{X} and $\hat{p}_N(x)$ converges to p(x) w.p.1 uniformly on any compact subset of \mathcal{X} .

Proof. Consider function $\psi(x,\xi) := \mathbf{1}_{(0,\infty)}(C(x,\xi))$. Recall that $p(x) = \mathbb{E}_P[\psi(x,\xi)]$ and $\hat{p}_N(x) = \mathbb{E}_{P_N}[\psi(x,\xi)]$, where $P_N := N^{-1} \sum_{j=1}^N \delta(\xi^j)$ is the respective empirical measure. Since the function $\mathbf{1}_{(0,\infty)}(\cdot)$ is lower semicontinuous and $C(x,\xi)$ is a Carathéodory function, it follows that the function $\psi(x,\xi)$ is random lower semicontinuous. Lower semicontinuity of p(x) and $\hat{p}_N(x)$ follows by Fatou's lemma (see Theorem 7.47). If the sample is iid, the epiconvergence $\hat{p}_N \stackrel{e}{\to} p$ w.p.1 follows by Theorem 7.56. Note that the dominance condition, from below and from above, holds here automatically since $|\psi(x,\xi)| \leq 1$.

Suppose, further, that condition (5.202) holds. Then for every $x \in \mathcal{X}$, $\psi(\cdot, \xi)$ is continuous at x w.p.1. It follows by the Lebesgue Dominated Convergence Theorem that $p(\cdot)$ is continuous at x (see Theorem 7.48). Finally, the uniform convergence w.p.1 follows by Theorem 7.53. \Box

Since the function p(x) is lower semicontinuous and the set \mathcal{X} is closed, it follows that the feasible set of problem (5.196) is closed. If, moreover, it is nonempty and bounded, then problem (5.196) has a nonempty set \mathcal{S} of optimal solutions (recall that the objective function f(x) is assumed to be continuous here). The same applies to the corresponding SAA problem (5.201). We have here the following consistency properties of the optimal

value $\hat{\vartheta}_N$ and the set \hat{S}_N of optimal solutions of the SAA problem (5.201) (compare with Theorem 5.5).

Proposition 5.30. Suppose that the set \mathcal{X} is compact, the function f(x) is continuous, $C(x,\xi)$ is a Carathéodory function, the sample is iid and that the following condition holds: (a) there is an optimal solution \bar{x} of the true problem such that for any $\epsilon > 0$ there is $x \in \mathcal{X}$ with $||x - \bar{x}|| \leq \epsilon$ and $p(x) < \alpha$. Then $\hat{\vartheta}_N \to \vartheta^*$ and $\mathbb{D}(\hat{S}_N, \mathcal{S}) \to 0$ w.p.1 as $N \to \infty$.

Proof. By condition (a), the set S is nonempty and there is $x' \in \mathcal{X}$ such that $p(x') < \alpha$. By the LLN we have that $\hat{p}_N(x')$ converges to p(x') w.p.1. Consequently $\hat{p}_N(x') < \alpha$, and hence the SAA problem has a feasible solution, w.p.1 for N large enough. Since $\hat{p}_N(\cdot)$ is lower semicontinuous, the feasible set of SAA problem is closed and hence compact, and thus \hat{S}_N is nonempty w.p.1 for N large enough. Of course, if x' is a feasible solution of an SAA problem, then $f(x') \geq \hat{\vartheta}_N$, where $\hat{\vartheta}_N$ is the optimal value of that SAA problem.

For a given $\varepsilon > 0$ let $x' \in \mathcal{X}$ be a point sufficiently close to $\bar{x} \in \mathcal{S}$ such that $\hat{p}_N(x') < \alpha$ and $f(x') \leq f(\bar{x}) + \varepsilon$. Since $f(\cdot)$ is continuous, existence of such point is ensured by condition (a). Consequently

$$\limsup_{N \to \infty} \hat{\vartheta}_N \le f(x') \le f(\bar{x}) + \varepsilon = \vartheta^* + \varepsilon \quad \text{w.p.1.}$$
(5.203)

Since $\varepsilon > 0$ is arbitrary, it follows that

$$\limsup_{N \to \infty} \hat{\vartheta}_N \le \vartheta^* \quad \text{w.p.1.}$$
(5.204)

Now let $\hat{x}_N \in \hat{S}_N$, i.e., $\hat{x}_N \in \mathcal{X}$, $\hat{p}_N(\hat{x}_N) \leq \alpha$ and $\hat{\vartheta}_N = f(\hat{x}_N)$. Since the set \mathcal{X} is compact, we can assume by passing to a subsequence if necessary that \hat{x}_N converges to a point $\bar{x} \in \mathcal{X}$ w.p.1. Also by Proposition 5.29 we have that $\hat{p}_N \stackrel{e}{\to} p$ w.p.1, and hence

$$\liminf_{N \to \infty} \hat{p}_N(\hat{x}_N) \ge p(\bar{x}) \quad \text{w.p.1.}$$

It follows that $p(\bar{x}) \leq \alpha$ and hence \bar{x} is a feasible point of the true problem, and thus $f(\bar{x}) \geq \vartheta^*$. Also $f(\hat{x}_N) \to f(\bar{x})$ w.p.1, and hence

$$\liminf_{N \to \infty} \hat{\vartheta}_N \ge \vartheta^* \quad \text{w.p.1.} \tag{5.205}$$

It follows from (5.204) and (5.205) that $\hat{\vartheta}_N \to \vartheta^*$ w.p.1. It also follows that the point \bar{x} is an optimal solution of the true problem and consequently we obtain that $\mathbb{D}(\hat{S}_N, \mathcal{S}) \to 0$ w.p.1. \Box

The above condition (a) is essential for the consistency of $\hat{\vartheta}_N$ and \hat{S}_N . Think, for example, about a situation where the constraint $p(x) \leq \alpha$ defines just one feasible point \bar{x} such that $p(\bar{x}) = \alpha$. Then arbitrary small changes in the constraint $\hat{p}_N(x) \leq \alpha$ may result in that the feasible set of the corresponding SAA problem becomes empty. Note also that condition (a) was not used in the proof of inequality (5.205). Verification of this condition (a) can be done by ad hoc methods.

We have that, under mild regularity conditions, optimal value and optimal solutions of the SAA problem (5.201) converge w.p.1, as $N \to \infty$, to their counterparts of the true problem (5.196). There are, however, several potential problems with the SAA approach here. In order for $\hat{p}_N(x)$ to be a reasonably accurate estimate of p(x), the sample size N should be significantly bigger than α^{-1} . For small α this may result in a large sample size. Another problem is that typically the function $\hat{p}_N(x)$ is discontinuous and the SAA problem (5.201) is a combinatorial problem which could be difficult to solve. Therefore we consider the following approach.

Convex Approximation Approach

For a generated sample $\xi^1, ..., \xi^N$ consider the problem:

$$\underset{x \in \mathcal{X}}{\operatorname{Min}} f(x) \text{ subject to } C(x, \xi^j) \le 0, \ j = 1, \dots, N.$$
(5.206)

Note that for $\alpha = 0$ the SAA problem (5.201) coincides with the above problem (5.206). If the set \mathcal{X} and functions $f(\cdot)$ and $C(\cdot,\xi), \xi \in \Xi$, are convex, then (5.206) is a convex problem and could be efficiently solved provided that the involved functions are given in a closed form and the sample size N is not too large. Clearly, as $N \to \infty$ the feasible set of the above problem (5.206) will shrink to the set of $x \in X$ determined by the constraints $C(x,\xi) \leq 0$ for a.e. $\xi \in \Xi$, and hence for large N will be overly conservative for the true chance constrained problem (5.196). Nevertheless, it makes sense to ask the question for what sample size N an optimal solution of problem (5.206) is guaranteed to be a feasible point of problem (5.196).

We need the following auxiliary result.

Lemma 5.31. Suppose that the set \mathcal{X} and functions $f(\cdot)$ and $C(\cdot,\xi)$, $\xi \in \Xi$, are convex and let \bar{x}_N be an optimal solution of problem (5.206). Then there exists an index set $J \subset \{1, ..., N\}$ such that $|J| \leq n$ and \bar{x}_N is an optimal solution of the problem

$$\underset{x \in \mathcal{X}}{\operatorname{Min}} f(x) \text{ subject to } C(x, \xi^j) \le 0, \ j \in J.$$
(5.207)

Proof. Consider sets $A_0 := \{x \in \mathcal{X} : f(x) < f(\bar{x}_N)\}$ and $A_j := \{x \in \mathcal{X} : C(x, \xi^j) \leq 0\}$, j = 1, ..., N. Since \mathcal{X} , $f(\cdot)$ and $C(\cdot, \xi^j)$ are convex, these sets are convex. Now we argue by a contradiction. Suppose that the assertion of this lemma is not correct. Then the intersection of A_0 and any n sets A_j is nonempty. Since the intersection of all sets $A_j, j \in \{1, ..., N\}$, is nonempty (these sets have at least one common element \bar{x}_N), it follows that the intersection of any n + 1 sets of the family $A_j, j \in \{0, 1, ..., N\}$, is nonempty. By Helly's theorem (Theorem 7.3) this implies that the intersection of all sets $A_j, j \in \{0, 1, ..., N\}$, is nonempty. This, in turn, implies existence of a feasible point \tilde{x} of problem (5.206) such that $f(\tilde{x}) < f(\bar{x}_N)$, which contradicts optimality of \bar{x}_N . \Box

We will use the following assumptions.

(F1) For any $N \in \mathbb{N}$ and any $(\xi_1, ..., \xi_N) \in \Xi^N$, problem (5.206) attains unique optimal solution $\bar{x}_N = \bar{x}(\xi_1, ..., \xi_N)$.

Recall that sometimes we use the same notation for a random vector and its particular value (realization). In the above assumption we view $\xi_1, ..., \xi_N$ as an element of the set Ξ^N , and \bar{x}_N as a function of $\xi_1, ..., \xi_N$. Of course, if $\xi_1, ..., \xi_N$ is a random sample, then \bar{x}_N becomes a random vector.

Let $\mathcal{J} = \mathcal{J}(\xi^1, ..., \xi^N) \subset \{1, ..., N\}$ be an index set such that \bar{x}_N is an optimal solution of the problem (5.207) for $J = \mathcal{J}$. Moreover, let the index set \mathcal{J} be minimal in the sense that if any of the constraints $C(x, \xi^j) \leq 0, j \in \mathcal{J}$, is removed, then \bar{x}_N is not an optimal solution of the obtained problem. We assume that w.p.1 such minimal index set is unique. By Lemma 5.31, we have that $|\mathcal{J}| \leq n$. By P^N we denote here the product probability measure on the set Ξ^N , i.e., P^N is the probability distribution of the iid sample $\xi^1, ..., \xi^N$.

(F2) There is an integer $\mathfrak{n} \in \mathbb{N}$ such that, for any $N \ge \mathfrak{n}$, w.p.1 the minimal set $\mathcal{J} = \mathcal{J}(\xi^1, ..., \xi^N)$ is uniquely defined and has constant cardinality \mathfrak{n} , i.e., $P^N \{ |\mathcal{J}| = \mathfrak{n} \} = 1$.

By Lemma 5.31, we have that $n \leq n$.

Assumption (F1) holds, for example, if the set \mathcal{X} is compact and convex, functions $f(\cdot)$ and $C(\cdot, \xi), \xi \in \Xi$, are convex, and either $f(\cdot)$ or the feasible set of problem (5.206) is strictly convex. Assumption (F2) is more involved, it is needed in order to show an equality in the estimate (5.209) of the following theorem.

The following result is due to Campi and Garatti [34], building on work of Calafiore and Campi [33]. Denote

$$\mathfrak{b}(k;\alpha,N) := \sum_{i=0}^{k} \binom{N}{i} \alpha^{i} (1-\alpha)^{N-i}, \quad k = 0, ..., N.$$
(5.208)

That is, $\mathfrak{b}(k; \alpha, N) = \Pr(W \leq k)$, where $W \sim B(\alpha, N)$ is a random variable having binomial distribution.

Theorem 5.32. Suppose that the set \mathcal{X} and functions $f(\cdot)$ and $C(\cdot,\xi)$, $\xi \in \Xi$, are convex and conditions (F1) and (F2) hold. Then for $\alpha \in (0,1)$ and for an iid sample $\xi^1, ..., \xi^N$ of size $N \ge \mathfrak{n}$ we have that

$$\Pr\{p(\bar{x}_N) > \alpha\} = \mathfrak{b}(\mathfrak{n} - 1; \alpha, N).$$
(5.209)

Proof. Let \mathfrak{J}_n be the family of all sets $J \subset \{1, ..., N\}$ of cardinality n. We have that $|\mathfrak{J}_n| = \binom{N}{n}$. For $J \in \mathfrak{J}_n$ define the set

$$\Sigma_J := \left\{ (\xi^1, ..., \xi^N) \in \Xi^N : \mathcal{J}(\xi^1, ..., \xi^N) = J \right\},$$
(5.210)

and denote by $\hat{x}_J = \hat{x}_J(\xi^1, ..., \xi^N)$ an optimal solution of problem (5.207) for $\mathcal{J} = J$. By condition (F1), such optimal solution \hat{x}_J exists and is unique, and hence

$$\Sigma_J = \left\{ (\xi^1, ..., \xi^N) \in \Xi^N : \hat{x}_J = \bar{x}_N \right\}.$$
 (5.211)

Note that for any permutation of vectors $\xi^1, ..., \xi^N$, problem (5.206) remains the same. Therefore, any set from the family $\{\Sigma_J\}_{J\in\mathfrak{J}_n}$ can be obtained from another set of that family by an appropriate permutation of its components. Since P^N is the direct product probability measure, it follows that the probability measure of each set $\Sigma_J, J \in \mathfrak{J}_n$, is the same. The sets Σ_J are disjoint and, because of condition (F2), union of all these sets is equal to Ξ^N up to a set of P^N -measure zero. Since there are $\binom{N}{n}$ such sets, we obtain that

$$P^{N}(\Sigma_{J}) = \frac{1}{\binom{N}{\mathfrak{n}}}.$$
(5.212)

Consider the optimal solution $\bar{x}_n = \bar{x}(\xi^1, ..., \xi^n)$, for N = n, and let H(z) be the cdf of the random variable $p(\bar{x}_n)$, i.e.,

$$H(z) := P^{\mathfrak{n}} \{ p(\bar{x}_{\mathfrak{n}}) \le z \}.$$
(5.213)

Let us show that for $N \geq \mathfrak{n}$,

$$P^{N}(\Sigma_{J}) = \int_{0}^{1} (1-z)^{N-\mathfrak{n}} dH(z).$$
(5.214)

Indeed, for $z \in [0, 1]$ and $J \in \mathfrak{J}_n$ consider the sets

$$\Delta_{z} := \{ (\xi_{1}, ..., \xi_{N}) : p(\bar{x}_{N}) \in [z, z + dz] \}, \Delta_{J,z} := \{ (\xi_{1}, ..., \xi_{N}) : p(\hat{x}_{J}) \in [z, z + dz] \}.$$
(5.215)

By (5.211) we have that $\Delta_z \cap \Sigma_J = \Delta_{J,z} \cap \Sigma_J$. For $J \in \mathfrak{J}_n$ let us evaluate probability of the event $\Delta_{J,z} \cap \Sigma_J$. For the sake of notational simplicity let us take $J = \{1, ..., n\}$. Note that \hat{x}_J depends on $(\xi^1, ..., \xi^n)$ only. Therefore $\Delta_{J,z} = \Delta_z^* \times \Xi^{N-n}$, where Δ_z^* is a subset of Ξ^n corresponding to the event " $p(\hat{x}_J) \in [z, z+dz]$ ". Conditional on $(\xi^1, ..., \xi^n)$, the event $\Delta_{J,z} \cap \Sigma_J$ happens iff the point $\hat{x}_J = \hat{x}_J(\xi^1, ..., \xi^n)$ remains feasible for the remaining $N - \mathfrak{n}$ constraints, i.e., iff $C(\hat{x}_J, \xi^j) \leq 0$ for all $j = \mathfrak{n} + 1, ..., N$. If $p(\hat{x}_J) = z$, then probability of each event " $C(\hat{x}_J, \xi^j) \leq 0$ " is equal to 1 - z. Since the sample is iid, we obtain that conditional on $(\xi^1, ..., \xi^n) \in \Delta_z^*$, probability of the event $\Delta_{J,z} \cap \Sigma_J$ is equal to $(1 - z)^{N-\mathfrak{n}}$. Consequently, the unconditional probability

$$P^{N}(\Delta_{z} \cap \Sigma_{J}) = P^{N}(\Delta_{J,z} \cap \Sigma_{J}) = (1-z)^{N-\mathfrak{n}} dH(z),$$
(5.216)

and hence equation (5.214) follows.

It follows from (5.212) and (5.214) that

$$\binom{N}{\mathfrak{n}} \int_0^1 (1-z)^{N-\mathfrak{n}} dH(z) = 1, \quad N \ge \mathfrak{n}.$$
(5.217)

Let us observe that $H(z) := z^n$ satisfies equations (5.217) for all $N \ge n$. Indeed, using integration by parts, we have

$$\binom{N}{\mathfrak{n}} \int_{0}^{1} (1-z)^{N-\mathfrak{n}} dz^{\mathfrak{n}} = -\binom{N}{\mathfrak{n}} \frac{\mathfrak{n}}{N-\mathfrak{n}+1} \int_{0}^{1} z^{\mathfrak{n}-1} d(1-z)^{N-\mathfrak{n}+1} = \binom{N}{\mathfrak{n}-1} \int_{0}^{1} (1-z)^{N-\mathfrak{n}+1} dz^{\mathfrak{n}-1} = \dots = 1.$$
 (5.218)

We also have that equations (5.217) determine respective moments of random variable 1 - Z, where $Z \sim H(z)$, and hence (since random variable $p(\bar{x}_n)$ has a bounded support) by the general theory of moments these equations have unique solution. Therefore we conclude that $H(z) = z^n$, $0 \le z \le 1$, is the cdf of $p(\bar{x}_n)$.

We also have by (5.216) that

$$P^{N}\left\{p(\bar{x}_{N})\in[z,z+dz]\right\}=\sum_{J\in\mathfrak{J}_{\mathfrak{n}}}P^{N}(\Delta_{z}\cap\Sigma_{J})=\binom{N}{\mathfrak{n}}(1-z)^{N-\mathfrak{n}}dH(z).$$
 (5.219)

Therefore, since $H(z) = z^n$ and using integration by parts similar to (5.218), we can write

$$P^{N} \{ p(\bar{x}_{N}) > \alpha \}$$

$$= \binom{N}{n} \int_{\alpha}^{1} (1-z)^{N-n} dH(z) = \binom{N}{n} n \int_{\alpha}^{1} (1-z)^{N-n} z^{n-1} dz$$

$$= \binom{N}{n} \frac{n}{N-n+1} \left[-(1-z)^{N-n+1} z^{n-1} \right]_{\alpha}^{1} + \int_{\alpha}^{1} (1-z)^{N-n+1} dz^{n-1} \right]$$

$$= \binom{N}{n-1} (1-\alpha)^{N-n+1} \alpha^{n-1} + \binom{N}{n-1} \int_{\alpha}^{1} (1-z)^{N-n+1} dz^{n-1}$$

$$= \cdots = \sum_{i=0}^{n-1} \binom{N}{i} \alpha^{i} (1-\alpha)^{N-i}.$$
(5.220)

Since $\Pr\{p(\bar{x}_N) > \alpha\} = P^N\{p(\bar{x}_N) > \alpha\}$, this completes the proof. \Box

Of course, the event " $p(\bar{x}_N) > \alpha$ " means that \bar{x}_N is not a feasible point of the true problem (5.196). Recall that $\mathfrak{n} \leq n$. Therefore, given $\beta \in (0, 1)$, the inequality (5.209) implies that for sample size $N \geq n$ such that

$$\mathfrak{b}(n-1;\alpha,N) \le \beta,\tag{5.221}$$

we have with probability at least $1 - \beta$ that \bar{x}_N is a feasible solution of the true problem (5.196).

Recall that

$$\mathfrak{b}(n-1;\alpha,N) = \mathsf{Pr}(W \le n-1),$$

where $W \sim B(\alpha, N)$ is a random variable having binomial distribution. For "not too small" α and large N, good approximation of that probability is suggested by the Central Limit Theorem. That is, W has approximately normal distribution with mean $N\alpha$ and variance $N\alpha(1-\alpha)$, and hence¹⁵

$$\mathfrak{b}(n-1;\alpha,N) \approx \Phi\left(\frac{n-1-N\alpha}{\sqrt{N\alpha(1-\alpha)}}\right).$$
 (5.222)

For $N\alpha \ge n-1$, Hoeffding inequality (7.213) gives the estimate

$$\mathfrak{b}(n-1;\alpha,N) \le \exp\left\{-\frac{2(N\alpha-n+1)^2}{N}\right\},\tag{5.223}$$

and Chernoff inequality (7.215) gives

$$\mathfrak{b}(n-1;\alpha,N) \le \exp\left\{-\frac{(N\alpha-n+1)^2}{2\alpha N}\right\}.$$
(5.224)

 $^{^{15}\}text{Recall}$ that $\Phi(\cdot)$ is the cdf of standard normal distribution.

The estimates (5.221) and (5.224) show that the required sample size N should be of order $O(\alpha^{-1})$. This, of course, is not surprising since just to estimate the probability p(x), for a given x, by Monte Carlo sampling we will need a sample size of order O(1/p(x)). For example, for n = 100 and $\alpha = \beta = 0.01$, bound (5.221) suggests estimate N = 12460 for the required sample size. Normal approximation (5.222) gives practically the same estimate of N. The estimate derived from the bound (5.223) gives significantly bigger estimate of N = 40372. The estimate derived from Chernoff inequality (5.224) gives much better estimate of N = 13410.

Anyway this indicates that the "guaranteed" estimates like (5.221) could be too conservative for practical calculations. Note also that Theorem 5.32 does not make any claims about quality of \bar{x}_N as a candidate for an optimal solution of the true problem (5.196), it only guarantees its feasibility.

5.7.2 Validation of an Optimal Solution

We discuss now an approach to a practical validation of a candidate point $\bar{x} \in \mathcal{X}$ for an optimal solution of the true problem (5.196). This task is twofold, namely we need to verify feasibility and optimality of \bar{x} . Of course, if a point \bar{x} is feasible for the true problem, then $\vartheta^* \leq f(\bar{x})$, i.e., $f(\bar{x})$ gives an upper bound for the true optimal value.

Upper bounds

Let us start with verification of feasibility of the point \bar{x} . For that we need to estimate the probability $p(\bar{x}) = \Pr\{C(\bar{x},\xi) > 0\}$. We proceed by employing Monte Carlo sampling techniques. For a generated iid random sample $\xi^1, ..., \xi^N$, let m be the number of times that the constraints $C(\bar{x},\xi^j) \leq 0, j = 1, ..., N$, are violated, i.e.,

$$\mathfrak{m} := \sum_{j=1}^{N} \mathbf{1}_{(0,\infty)} \left(C(\bar{x}, \xi^j) \right).$$

Then $\hat{p}_N(\bar{x}) = \mathfrak{m}/N$ is an unbiased estimator of $p(\bar{x})$, and \mathfrak{m} has Binomial distribution $B(p(\bar{x}), N)$.

If the sample size N is significantly bigger than $1/p(\bar{x})$, then the distribution of $\hat{p}_N(\bar{x})$ can be reasonably approximated by a normal distribution with mean $p(\bar{x})$ and variance $p(\bar{x})(1-p(\bar{x}))/N$. In that case one can consider, for a given confidence level $\beta \in (0, 1/2)$, the following approximate upper bound for the probability¹⁶ $p(\bar{x})$:

$$\hat{p}_N(\bar{x}) + z_\beta \sqrt{\frac{\hat{p}_N(\bar{x})(1-\hat{p}_N(\bar{x}))}{N}}.$$
 (5.225)

Let us discuss the following, more accurate, approach for constructing an upper confidence bound for the probability $p(\bar{x})$. For a given $\beta \in (0, 1)$ consider

$$\mathfrak{U}_{\beta,N}(\bar{x}) := \sup_{\rho \in [0,1]} \left\{ \rho : \mathfrak{b}(\mathfrak{m};\rho,N) \ge \beta \right\}.$$
(5.226)

¹⁶Recall that $z_{\beta} := \Phi^{-1}(1-\beta) = -\Phi^{-1}(\beta)$, where $\Phi(\cdot)$ is the cdf of the standard normal distribution.

We have that $\mathfrak{U}_{\beta,N}(\bar{x})$ is a function of \mathfrak{m} and hence is a random variable. Note that $\mathfrak{b}(\mathfrak{m};\rho,N)$ is continuous and monotonically decreasing in $\rho \in (0,1)$. Therefore, in fact, the supremum in the right hand side of (5.226) is attained, and $\mathfrak{U}_{\beta,N}(\bar{x})$ is equal to such $\bar{\rho}$ that $\mathfrak{b}(\mathfrak{m};\bar{\rho},N) = \beta$. Denoting $V := \mathfrak{b}(\mathfrak{m};p(\bar{x}),N)$, we have that

$$\Pr\left\{p(\bar{x}) < \mathfrak{U}_{\beta,N}(\bar{x})\right\} = \Pr\left\{V > \overbrace{\mathfrak{b}(\mathfrak{m};\bar{\rho},N)}^{\beta}\right\}$$
$$= 1 - \Pr\left\{V \le \beta\right\} = 1 - \sum_{k=0}^{N} \Pr\left\{V \le \beta \middle| \mathfrak{m} = k\right\} \Pr(\mathfrak{m} = k).$$

Since

$$\Pr\left\{V \le \beta \middle| \mathfrak{m} = k\right\} = \left\{\begin{array}{ll} 1, & \text{if } \mathfrak{b}(k; p(\bar{x}), N) \le \beta, \\ 0, & \text{otherwise,} \end{array}\right.$$

and $\Pr(\mathfrak{m} = k) = \binom{N}{k} p(\bar{x})^k (1 - p(\bar{x}))^{N-k}$, it follows that

$$\sum_{k=0}^{N} \Pr\left\{V \leq \beta \middle| \mathfrak{m} = k\right\} \Pr(\mathfrak{m} = k) \leq \beta,$$

and hence

$$\Pr\left\{p(\bar{x}) < \mathfrak{U}_{\beta,N}(\bar{x})\right\} \ge 1 - \beta. \tag{5.227}$$

That is, $p(\bar{x}) < \mathfrak{U}_{\beta,N}(\bar{x})$ with probability at least $1 - \beta$. Therefore we can take $\mathfrak{U}_{\beta,N}(\bar{x})$ as an upper $(1 - \beta)$ -confidence bound for $p(\bar{x})$. In particular, if $\mathfrak{m} = 0$, then

$$\mathfrak{U}_{\beta,N}(\bar{x}) = 1 - \beta^{1/N} < N^{-1} \ln(\beta^{-1}).$$

We obtain that if $\mathfrak{U}_{\beta,N}(\bar{x}) \leq \alpha$, then \bar{x} is a feasible solution of the true problem with probability at least $1 - \beta$. In that case we can use $f(\bar{x})$ as an upper bound, with confidence $1 - \beta$, for the optimal value ϑ^* of the true problem (5.196). Since this procedure involves only calculations of $C(\bar{x}, \xi^j)$, it can be performed with a large sample size N, and hence feasibility of \bar{x} can be verified with a high accuracy provided that α is not too small.

It also could be noted that the bound given in (5.225), in a sense, is an approximation of the upper bound $\bar{\rho} = \mathfrak{U}_{\beta,N}(\bar{x})$. Indeed, by the CLT the cumulative distribution $\mathfrak{b}(k;\bar{\rho},N)$ can be approximated by $\Phi\left(\frac{k-\bar{\rho}N}{\sqrt{N\bar{\rho}(1-\bar{\rho})}}\right)$. Therefore, approximately $\bar{\rho}$ is the solution of the equation $\Phi\left(\frac{\mathfrak{m}-\rho N}{\sqrt{N\rho(1-\rho)}}\right) = \beta$, which can be written as

$$\rho = \frac{\mathfrak{m}}{N} + z_{\beta} \sqrt{\frac{\rho(1-\rho)}{N}}.$$

By approximating ρ in the right hand side of the above equation by \mathfrak{m}/N we obtain the bound (5.225).

Lower Bounds

It is more tricky to construct a valid lower statistical bound for ϑ^* . One possible approach is to apply a general methodology of the SAA method (see discussion at the end of section 5.6.1). We have that for any $\lambda \ge 0$ the following inequality holds (compare with (5.183)):

$$\vartheta^* \ge \inf_{x \in \mathcal{X}} \left[f(x) + \lambda(p(x) - \alpha) \right].$$
 (5.228)

We also have that expectation of

$$\hat{v}_N(\lambda) := \inf_{x \in \mathcal{X}} \left[f(x) + \lambda (\hat{p}_N(x) - \alpha) \right]$$
(5.229)

gives a valid lower bound for the right hand side of (5.228), and hence for ϑ^* . An unbiased estimate of $\mathbb{E}[\hat{v}_N(\lambda)]$ can be obtained by solving the right hand side problem of (5.229) several times and averaging calculated optimal values. Note, however, that there are two difficulties with applying this approach. First, recall that typically the function $\hat{p}_N(x)$ is discontinuous and hence it could be difficult to solve these optimization problems. Second, it may happen that for any choice of $\lambda \ge 0$ the optimal value of the right hand side of (5.228) is smaller than ϑ^* , i.e., there is a gap between problem (5.196) and its (Lagrangian) dual.

We discuss now an alternative approach to construction statistical lower bounds. For chosen positive integers N and M, and constant $\gamma \in [0, 1)$, let us generate M independent samples $\xi^{1,m}, ..., \xi^{N,m}$, m = 1, ..., M, each of size N, of random vector ξ . For each sample solve the associated optimization problem

$$\operatorname{Min}_{x \in \mathcal{X}} f(x) \text{ subject to } \sum_{j=1}^{N} \mathbf{1}_{(0,\infty)} \left(C(x,\xi^{j,m}) \right) \le \gamma N,$$
(5.230)

and hence calculate its optimal value $\hat{\vartheta}_{\gamma,N}^m$, m = 1, ..., M. That is, we solve M times the corresponding SAA problem at the significance level γ . In particular, for $\gamma = 0$ problem (5.230) takes the form

$$\underset{x \in \mathcal{X}}{\operatorname{Min}} f(x) \text{ subject to } C(x, \xi^{j,m}) \le 0, \, j = 1, \dots, N.$$
(5.231)

It may happen that problem (5.230) is either infeasible or unbounded from below, in which case we assign its optimal value as $+\infty$ or $-\infty$, respectively. We can view $\hat{\vartheta}_{\gamma,N}^m$, m = 1, ..., M, as an iid sample of the random variable $\hat{\vartheta}_{\gamma,N}$, where $\hat{\vartheta}_{\gamma,N}$ is the optimal value of the respective SAA problem at significance level γ . Next we rearrange the calculated optimal values in the nondecreasing order as follows $\hat{\vartheta}_{\gamma,N}^{(1)} \leq ... \leq \hat{\vartheta}_{\gamma,N}^{(M)}$, i.e., $\hat{\vartheta}_{\gamma,N}^{(1)}$ is the smallest, $\hat{\vartheta}_{\gamma,N}^{(2)}$ is the second smallest etc, among the values $\hat{\vartheta}_{\gamma,N}^m$, m = 1, ..., M. By definition, we choose an integer $L \in \{1, ..., M\}$ and use the random quantity $\hat{\vartheta}_{\gamma,N}^{(L)}$ as a lower bound of the true optimal value ϑ^* .

Let us analyze the resulting bounding procedure. Let $\tilde{x} \in X$ be a feasible point of the true problem, i.e.,

$$\Pr\{C(\tilde{x},\xi) > 0\} \le \alpha.$$
Since $\sum_{j=1}^{N} \mathbf{1}_{(0,\infty)} (C(\tilde{x}, \xi^{j,m}))$ has binomial distribution with probability of success equal to the probability of the event $\{C(\tilde{x}, \xi) > 0\}$, it follows that \tilde{x} is feasible for problem (5.230) with probability at least¹⁷

$$\sum_{i=0}^{\lfloor \gamma N \rfloor} \binom{N}{i} \alpha^{i} (1-\alpha)^{N-i} = \mathfrak{b}\left(\lfloor \gamma N \rfloor; \alpha, N\right) =: \theta_{N}$$

When \tilde{x} is feasible for (5.230), we of course have that $\hat{\vartheta}_{\gamma,N}^m \leq f(\tilde{x})$. Let $\varepsilon > 0$ be an arbitrary constant and \tilde{x} be a feasible point of the true problem such that $f(\tilde{x}) \leq \vartheta^* + \varepsilon$. Then for every $m \in \{1, ..., M\}$ we have

$$\theta := \Pr\left\{\hat{\vartheta}^m_{\gamma,N} \le \vartheta^* + \varepsilon\right\} \ge \Pr\left\{\hat{\vartheta}^m_{\gamma,N} \le f(\tilde{x})\right\} \ge \theta_N.$$

Now, in the case of $\hat{\vartheta}_{\gamma,N}^{(L)} > \vartheta^* + \varepsilon$, the corresponding realization of the random sequence $\hat{\vartheta}_{\gamma,N}^1, ..., \hat{\vartheta}_{\gamma,N}^M$ contains less than L elements which are less than or equal to $\vartheta^* + \varepsilon$. Since the elements of the sequence are independent, the probability of the latter event is $\mathfrak{b}(L-1;\theta,M)$. Since $\theta \geq \theta_N$, we have that $\mathfrak{b}(L-1;\theta,M) \leq \mathfrak{b}(L-1;\theta_N,M)$. Thus, $\Pr\left\{\hat{\vartheta}_{\gamma,N}^{(L)} > \vartheta^* + \varepsilon\right\} \leq \mathfrak{b}(L-1;\theta_N,M)$. Since the resulting inequality is valid for any $\varepsilon > 0$, we arrive at the bound

$$\Pr\left\{\hat{\vartheta}_{\gamma,N}^{(L)} > \vartheta^*\right\} \le \mathfrak{b}(L-1;\theta_N,M).$$
(5.232)

We obtain the following result.

Proposition 5.33. Given $\beta \in (0, 1)$ and $\gamma \in [0, 1)$, let us choose positive integers M, N and L in such a way that

$$\mathfrak{b}(L-1;\theta_N,M) \le \beta,\tag{5.233}$$

where $\theta_N := \mathfrak{b}(\lfloor \gamma N \rfloor; \alpha, N)$. Then

$$\Pr\left\{\hat{\vartheta}_{\gamma,N}^{(L)} > \vartheta^*\right\} \le \beta.$$
(5.234)

For given sample sizes N and M it is better to take the largest integer $L \in \{1, ..., M\}$ satisfying condition (5.233). That is, for

$$L^* := \max_{1 \le L \le M} \left\{ L : \mathfrak{b}(L-1; \theta_N, M) \le \beta \right\}$$

we have that the random quantity $\hat{\vartheta}_{\gamma,N}^{(L^*)}$ gives a lower bound for the true optimal value ϑ^* with probability at least $1 - \beta$. If no $L \in \{1, ..., M\}$ satisfying (5.233) exists, the lower bound, by definition, is $-\infty$.

The question arising in connection with the outlined bounding scheme is how to choose M, N and γ . In the convex case it is advantageous to take $\gamma = 0$, since then we

¹⁷Recall that the notation $\lfloor a \rfloor$ stands for the largest integer less than or equal to $a \in \mathbb{R}$.

need to solve convex problems (5.231), rather than combinatorial problems (5.230). Note that for $\gamma = 0$, we have that $\theta_N = (1 - \alpha)^N$ and the bound (5.233) takes the form

$$\sum_{k=0}^{L-1} \binom{M}{k} (1-\alpha)^{Nk} [1-(1-\alpha)^N]^{M-k} \le \beta.$$
(5.235)

Suppose that N and $\gamma \geq 0$ are given (fixed). Then the larger is M, the better. We can view $\hat{\vartheta}_{\gamma,N}^m$, m = 1, ..., M, as a random sample from the distribution of the random variable $\hat{\vartheta}_N$, with $\hat{\vartheta}_N$ being the optimal value of the corresponding SAA problem of the form (5.230). It follows from the definition that L^* is equal to the (lower) β -quantile of the binomial distribution $B(\theta_N, M)$. By the CLT we have that

$$\lim_{M \to \infty} \frac{L^* - \theta_N M}{\sqrt{M \theta_N (1 - \theta_N)}} = \Phi^{-1}(\beta),$$

and L^*/M tends to θ_N as $M \to \infty$. It follows that the lower bound $\hat{\vartheta}_{\gamma,N}^{(L^*)}$ converges to the θ_N -quantile of the distribution of $\hat{\vartheta}_N$ as $M \to \infty$.

In reality, however, M is bounded by the computational effort required to solve M problems of the form (5.230). Note that the effort per problem is larger the larger is the sample size N. For L = 1 (which is the smallest value of L) and $\gamma = 0$ the left hand side of (5.235) is equal to $[1-(1-\alpha)^N]^M$. Note that $(1-\alpha)^N \approx e^{-\alpha N}$ for small $\alpha > 0$. Therefore if αN is large, then one will need a very large M to make $[1 - (1 - \alpha)^N]^M$ smaller than, say, $\beta = 0.01$, and hence to get a meaningful lower bound. For example, for $\alpha N = 7$ we have that $e^{-\alpha N} = 0.0009$, and we will need M > 5000 to make $[1 - (1 - \alpha)^N]^M$ smaller than 0.01. Therefore, for $\gamma = 0$ it is recommended to take N not larger than, say, $2/\alpha$.

5.8 SAA Method Applied to Multistage Stochastic Programming

Consider a multistage stochastic programming problem, in the general form (3.1), driven by the random data process $\xi_1, \xi_2, ..., \xi_T$. The exact meaning of this formulation was discussed in section 3.1.1. In this section we discuss application of the SAA method to such multistage problems.

Consider the following sampling scheme. Generate a sample $\xi_2^1, ..., \xi_2^{N_1}$ of N_1 realizations of random vector ξ_2 . Conditional on each ξ_2^i , $i = 1, ..., N_1$, generate a random sample ξ_3^{ij} , $j = 1, ..., N_2$, of N_2 realizations of ξ_3 according to conditional distribution of ξ_3 given $\xi_2 = \xi_2^i$. Conditional on each ξ_3^{ij} generate a random sample of size N_3 of ξ_4 conditional on $\xi_3 = \xi_3^{ij}$, and so on for later stages. (Although we do not consider such possibility here, it is also possible to generate at each stage conditional samples of different sizes.) In that way we generate a scenario tree with $N = \prod_{t=1}^{T-1} N_t$ number of scenarios each taken with equal probability 1/N. We refer to this scheme as *conditional sampling*. Unless stated otherwise¹⁸, we assume that, at the first stage, the sample $\xi_2^1, ..., \xi_2^{N_1}$ is iid

¹⁸It is also possible to employ Quasi-Monte Carlo sampling in constructing conditional sampling. In some situations this may reduce variability of the corresponding SAA estimators. In the following analysis we assume independence in order to simplify statistical analysis.

and the following samples, at each stage t = 2, ..., T - 1, are conditionally iid. If, moreover, all conditional samples at each stage are independent of each other, we refer to such conditional sampling as the *independent conditional sampling*. The multistage stochastic programming problem induced by the original problem (3.1) on the scenario tree generated by conditional sampling is viewed as the sample average approximation (SAA) of the "true" problem (3.1).

It could be noted that in case of stagewise independent process $\xi_1, ..., \xi_T$, the independent conditional sampling destroys the stagewise independence structure of the original process. This is because at each stage conditional samples are independent of each other and hence are different. In the stagewise independence case an alternative approach is to use the same sample at each stage. That is, independent of each other random samples $\xi_t^1, ..., \xi_t^{N_{t-1}}$ of respective $\xi_t, t = 2, ..., T$, are generated and the corresponding scenario tree is constructed by connecting every ancestor node at stage t - 1 with the same set of children nodes $\xi_t^1, ..., \xi_t^{N_{t-1}}$. In that way stagewise independence is preserved in the scenario tree generated by conditional sampling. We refer to this sampling scheme as the *identical conditional sampling*.

5.8.1 Statistical Properties of Multistage SAA Estimators

Similar to two stage programming it makes sense to discuss convergence of the optimal value and first stage solutions of multistage SAA problems to their true counterparts as sample sizes $N_1, ..., N_{T-1}$ tend to infinity. We denote $\mathcal{N} := \{N_1, ..., N_{T-1}\}$ and by ϑ^* and $\vartheta_{\mathcal{N}}$ the optimal values of the true and the corresponding SAA multistage programs, respectively.

In order to simplify the presentation let us consider now three stage stochastic programs, i.e., T = 3. In that case conditional sampling consists of sample ξ_2^i , $i = 1, ..., N_1$, of ξ_2 and for each $i = 1, ..., N_1$ of conditional samples ξ_3^{ij} , $j = 1, ..., N_2$, of ξ_3 given $\xi_2 = \xi_2^i$. Let us write dynamic programming equations for the true problem. We have

$$Q_3(x_2,\xi_3) = \inf_{x_3 \in \mathcal{X}_3(x_2,\xi_3)} f_3(x_3,\xi_3),$$
(5.236)

$$Q_2(x_1,\xi_2) = \inf_{x_2 \in \mathcal{X}_2(x_1,\xi_2)} \left\{ f_2(x_2,\xi_2) + \mathbb{E} \left[Q_3(x_2,\xi_3) \big| \xi_2 \right] \right\},$$
(5.237)

and at the first stage we solve the problem

$$\min_{x_1 \in \mathcal{X}_1} \left\{ f_1(x_1) + \mathbb{E} \left[Q_2(x_1, \xi_2) \right] \right\}.$$
(5.238)

If we could calculate values $Q_2(x_1, \xi_2)$, we could approximate problem (5.238) by the sample average problem

$$\underset{x_1 \in \mathcal{X}_1}{\min} \left\{ \hat{f}_{N_1}(x_1) := f_1(x_1) + \frac{1}{N_1} \sum_{i=1}^{N_1} Q_2(x_1, \xi_2^i) \right\}.$$
(5.239)

However, values $Q_2(x_1,\xi_2^i)$ are not given explicitly and are approximated by

$$\hat{Q}_{2,N_2}(x_1,\xi_2^i) := \inf_{x_2 \in \mathcal{X}_2(x_1,\xi_2^i)} \left\{ f_2(x_2,\xi_2^i) + \frac{1}{N_2} \sum_{j=1}^{N_2} Q_3(x_2,\xi_3^{ij}) \right\},$$
(5.240)

 $i = 1, ..., N_1$. That is, the SAA method approximates the first stage problem (5.238) by the problem

$$\min_{x_1 \in \mathcal{X}_1} \left\{ \tilde{f}_{N_1, N_2}(x_1) := f_1(x_1) + \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{Q}_{2, N_2}(x_1, \xi_2^i) \right\}.$$
(5.241)

In order to verify consistency of the SAA estimators, obtained by solving problem (5.241), we need to show that $\tilde{f}_{N_1,N_2}(x_1)$ converges to $f_1(x_1) + \mathbb{E}[Q_2(x_1,\xi_2)]$ w.p.1 uniformly on any compact subset X of \mathcal{X}_1 (compare with analysis of section 5.1.1). That is, we need to show that

$$\lim_{N_1, N_2 \to \infty} \sup_{x_1 \in X} \left| \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{Q}_{2, N_2}(x_1, \xi_2^i) - \mathbb{E} \left[Q_2(x_1, \xi_2) \right] \right| = 0 \text{ w.p.1.}$$
(5.242)

For that it suffices to show that

$$\lim_{N_1 \to \infty} \sup_{x_1 \in X} \left| \frac{1}{N_1} \sum_{i=1}^{N_1} Q_2(x_1, \xi_2^i) - \mathbb{E} \left[Q_2(x_1, \xi_2) \right] \right| = 0 \text{ w.p.1},$$
(5.243)

and

$$\lim_{N_1, N_2 \to \infty} \sup_{x_1 \in X} \left| \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{Q}_{2, N_2}(x_1, \xi_2^i) - \frac{1}{N_1} \sum_{i=1}^{N_1} Q_2(x_1, \xi_2^i) \right| = 0 \text{ w.p.1.}$$
(5.244)

Condition (5.243) can be verified by applying a version of uniform Law of Large Numbers (see section 7.2.5). Condition (5.244) is more involved. Of course, we have that

$$\begin{aligned} \sup_{x_1 \in X} \left| \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{Q}_{2,N_2}(x_1,\xi_2^i) - \frac{1}{N_1} \sum_{i=1}^{N_1} Q_2(x_1,\xi_2^i) \right| \\ &\leq \frac{1}{N_1} \sum_{i=1}^{N_1} \sup_{x_1 \in X} \left| \hat{Q}_{2,N_2}(x_1,\xi_2^i) - Q_2(x_1,\xi_2^i) \right|, \end{aligned}$$

and hence condition (5.244) holds if $\hat{Q}_{2,N_2}(x_1,\xi_2^i)$ converges to $Q_2(x_1,\xi_2^i)$ w.p.1 as $N_2 \to \infty$ in a certain uniform way. Unfortunately an exact mathematical analysis of such condition could be quite involved. The analysis simplifies considerably if the underline random process is stagewise independent. In the present case this means that random vectors ξ_2 and ξ_3 are independent. In that case distribution of random sample ξ_3^{ij} , $j = 1, ..., N_2$, does not depend on i (in both sampling schemes whether samples ξ_3^{ij} are the same for all $i = 1, ..., N_1$, or independent of each other), and we can apply Theorem 7.53 to establish that, under mild regularity conditions, $\frac{1}{N_2} \sum_{j=1}^{N_2} Q_3(x_2, \xi_3^{ij})$ converges to $\mathbb{E}[Q_3(x_2, \xi_3)]$ w.p.1 as $N_2 \to \infty$ uniformly in x_2 on any compact subset of \mathbb{R}^{n_2} . With an additional assumptions about mapping $\mathcal{X}_2(x_1, \xi_2)$, it is possible to verify the required uniform type convergence of $\hat{Q}_{2,N_2}(x_1, \xi_2^i)$ to $Q_2(x_1, \xi_2^i)$. Again a precise mathematical analysis is quite technical and will be left out. Instead in section 5.8.2 we discuss a uniform exponential convergence of the sample average function $\tilde{f}_{N_1,N_2}(x_1)$ to the objective function $f_1(x_1) + \mathbb{E}[Q_2(x_1, \xi_2)]$ of the true problem.

Let us make the following observations. By increasing sample sizes $N_1, ..., N_{T-1}$, of conditional sampling, we eventually reconstruct the scenario tree structure of the original multistage problem. Therefore it should be expected that in the limit, as these sample sizes tend (simultaneously) to infinity, the corresponding SAA estimators of the optimal value

and first stage solutions are consistent, i.e., converge w.p.1 to their true counterparts. And, indeed, this can be shown under certain regularity conditions. However, consistency alone does not justify the SAA method since in reality sample sizes are always finite and are constrained by available computational resources. Similar to the two stage case we have here that (for minimization problems)

$$\vartheta^* \ge \mathbb{E}[\hat{\vartheta}_{\mathcal{N}}]. \tag{5.245}$$

That is, the SAA optimal value $\hat{\vartheta}_{\mathcal{N}}$ is a downwards biased estimator of the true optimal value ϑ^* .

Suppose now that the data process $\xi_1, ..., \xi_T$ is stagewise independent. As it was discussed above in that case it is possible to use two different approaches to conditional sampling, namely to use at every stage independent or the same samples for every ancestor node at the previous stage. These approaches were referred to as the independent and identical conditional samplings, respectively. Consider, for instance, the three stage stochastic programming problem (5.236)–(5.238). In the second approach of identical conditional sampling we have sample ξ_2^i , $i = 1, ..., N_1$, of ξ_2 and sample ξ_3^j , $j = 1, ..., N_2$, of ξ_3 independent of ξ_2^i . In that case formula (5.240) takes the form

$$\hat{Q}_{2,N_2}(x_1,\xi_2^i) = \inf_{x_2 \in \mathcal{X}_2(x_1,\xi_2^i)} \left\{ f_2(x_2,\xi_2^i) + \frac{1}{N_2} \sum_{j=1}^{N_2} Q_3(x_2,\xi_3^j) \right\}.$$
(5.246)

Because of independence of ξ_2 and ξ_3 we have that conditional distribution of ξ_3 given ξ_2 is the same as its unconditional distribution, and hence in both sampling approaches $\hat{Q}_{2,N_2}(x_1,\xi_2^i)$ has the same distribution independent of i. Therefore in both sampling schemes $\frac{1}{N_1} \sum_{i=1}^{N_1} \hat{Q}_{2,N_2}(x_1,\xi_2^i)$ has the same expectation, and hence we may expect that in both cases the estimator $\hat{\vartheta}_N$ has a similar bias. Variance of $\hat{\vartheta}_N$, however, could be quite different. In the case of independent conditional sampling we have that $\hat{Q}_{2,N_2}(x_1,\xi_2^i)$, $i = 1, ..., N_1$, are independent of each other, and hence

$$\mathbb{V}\operatorname{ar}\left[\frac{1}{N_{1}}\sum_{i=1}^{N_{1}}\hat{Q}_{2,N_{2}}(x_{1},\xi_{2}^{i})\right] = \frac{1}{N_{1}}\mathbb{V}\operatorname{ar}\left[\hat{Q}_{2,N_{2}}(x_{1},\xi_{2}^{i})\right].$$
(5.247)

On the other hand in the case of identical conditional sampling the right hand side of (5.246) has the same component $\frac{1}{N_2} \sum_{j=1}^{N_2} Q_3(x_2, \xi_3^j)$ for all $i = 1, ..., N_1$. Consequently $\hat{Q}_{2,N_2}(x_1, \xi_2^i)$ would tend to be positively correlated for different values of i, and as a result $\hat{\vartheta}_N$ will have a higher variance than in the case of independent conditional sampling. Therefore from a statistical point of view it is advantageous to use the independent conditional sampling.

Example 5.34 (Portfolio Selection) Consider the example of multistage portfolio selection discussed in section 1.4.2. Suppose for the moment that the problem has three stages t = 0, 1, 2. In the SAA approach we generate sample ξ_1^i , $i = 1, ..., N_0$, of returns at stage t = 1, and conditional samples ξ_2^{ij} , $j = 1, ..., N_1$, of returns at stage t = 2. The dynamic programming equations for the SAA problem can be written as follows (see equations (1.50)–(1.52)). At stage t = 1 for $i = 1, ..., N_0$, we have

$$\hat{Q}_{1,N_1}(W_1,\xi_1^i) = \sup_{x_1 \ge 0} \left\{ \frac{1}{N_1} \sum_{j=1}^{N_1} U((\xi_2^{ij})^\mathsf{T} x_1) : e^\mathsf{T} x_1 = W_1 \right\},\tag{5.248}$$

where $e \in \mathbb{R}^n$ is vector of ones, and at stage t = 0 we solve the problem

$$\max_{x_0 \ge 0} \frac{1}{N_0} \sum_{i=1}^{N_0} \hat{Q}_{1,N_1} \left((\xi_1^i)^\mathsf{T} x_0, \xi_1^i \right) \text{ s.t. } e^\mathsf{T} x_0 = W_0.$$
 (5.249)

Now let $U(W) := \ln W$ be the logarithmic utility function. Suppose that the data process is stagewise independent. Then the optimal value ϑ^* of the true problem is (see (1.58))

$$\vartheta^* = \ln W_0 + \sum_{t=0}^{T-1} \nu_t, \tag{5.250}$$

where ν_t is the optimal value of the problem

$$\max_{x_t \ge 0} \mathbb{E}\left[\ln\left(\xi_{t+1}^\mathsf{T} x_t\right)\right] \quad \text{s.t.} \quad e^\mathsf{T} x_t = 1.$$
(5.251)

Let the SAA method be applied with the identical conditional sampling, with respective sample ξ_t^j , $j = 1, ..., N_{t-1}$ of ξ_t , t = 1, ..., T. In that case the corresponding SAA problem is also stagewise independent and the optimal value of the SAA problem

$$\hat{\vartheta}_{\mathcal{N}} = \ln W_0 + \sum_{t=0}^{T-1} \hat{\nu}_{t,N_t},$$
(5.252)

where $\hat{\nu}_{t,N_t}$ is the optimal value of the problem

$$\max_{x_t \ge 0} \frac{1}{N_t} \sum_{j=1}^{N_t} \ln\left((\xi_{t+1}^j)^\mathsf{T} x_t \right) \quad \text{s.t.} \quad e^\mathsf{T} x_t = 1.$$
 (5.253)

We can view $\hat{\nu}_{t,N_t}$ as an SAA estimator of ν_t . Since here we solve a maximization rather than a minimization problem, $\hat{\nu}_{t,N_t}$ is an upwards biased estimator of ν_t , i.e., $\mathbb{E}[\hat{\nu}_{t,N_t}] \geq \nu_t$. We also have that $\mathbb{E}[\hat{\vartheta}_{\mathcal{N}}] = \ln W_0 + \sum_{t=0}^{T-1} \mathbb{E}[\hat{\nu}_{t,N_t}]$, and hence

$$\mathbb{E}[\hat{\vartheta}_{\mathcal{N}}] - \vartheta^* = \sum_{t=0}^{T-1} \left(\mathbb{E}[\hat{\nu}_{t,N_t}] - \nu_t \right).$$
(5.254)

That is, for the logarithmic utility function and identical conditional sampling, bias of the SAA estimator of the optimal value grows additively with increase of the number of stages. Also because the samples at different stages are independent of each other, we have that

$$\mathbb{V}\mathrm{ar}[\hat{\vartheta}_{\mathcal{N}}] = \sum_{t=0}^{T-1} \mathbb{V}\mathrm{ar}[\hat{\nu}_{t,N_t}].$$
(5.255)

Let now $U(W) := W^{\gamma}$, with $\gamma \in (0, 1]$, be the power utility function and suppose that the data process is stagewise independent. Then (see (1.61))

$$\vartheta^* = W_0^{\gamma} \prod_{t=0}^{T-1} \eta_t,$$
(5.256)

where η_t is the optimal value of problem

$$\underset{x_t \ge 0}{\operatorname{Max}} \mathbb{E}\left[\left(\xi_{t+1}^{\mathsf{T}} x_t\right)^{\gamma}\right] \quad \text{s.t. } e^{\mathsf{T}} x_t = 1.$$
(5.257)

For the corresponding SAA method with the identical conditional sampling, we have that

$$\hat{\vartheta}_{\mathcal{N}} = W_0^{\gamma} \prod_{t=0}^{T-1} \hat{\eta}_{t,N_t}, \qquad (5.258)$$

where $\hat{\eta}_{t,N_t}$ is the optimal value of problem

$$\max_{x_t \ge 0} \frac{1}{N_t} \sum_{j=1}^{N_t} \left((\xi_{t+1}^j)^\mathsf{T} x_t \right)^\gamma \quad \text{s.t.} \quad e^\mathsf{T} x_t = 1.$$
(5.259)

Because of the independence of the samples, and hence independence of $\hat{\eta}_{t,N_t}$, we can write $\mathbb{E}[\hat{\vartheta}_{\mathcal{N}}] = W_0^{\gamma} \prod_{t=0}^{T-1} \mathbb{E}[\hat{\eta}_{t,N_t}]$, and hence

$$\mathbb{E}[\hat{\vartheta}_{\mathcal{N}}] = \vartheta^* \prod_{t=0}^{T-1} (1 + \beta_{t,N_t}), \qquad (5.260)$$

where $\beta_{t,N_t} := \frac{\mathbb{E}[\hat{\eta}_{t,N_t}] - \eta_t}{\eta_t}$ is the relative bias of $\hat{\eta}_{t,N_t}$. That is, bias of $\hat{\vartheta}_N$ grows with increase of the number of stages in a *multiplicative* way. In particular, if the relative biases β_{t,N_t} are constant, then bias of $\hat{\vartheta}_N$ grows *exponentially* fast with increase of the number of stages.

Statistical Validation Analysis

By (5.245) we have that the optimal value $\hat{\vartheta}_{\mathcal{N}}$ of SAA problem gives a valid statistical lower bound for the optimal value ϑ^* . Therefore in order to construct a lower bound for ϑ^* one can proceed exactly in the same way as it was discussed in section 5.6.1. Unfortunately, typically the bias and variance of $\hat{\vartheta}_{\mathcal{N}}$ grow fast with increase of the number of stages, which makes the corresponding statistical lower bounds quite inaccurate already for a mild number of stages.

In order to construct an upper bound we proceed as follows. Let $x_t(\xi_{[t]})$ be a feasible policy. Recall that a policy is feasible if it satisfies the feasibility constraints (3.3). Since the multistage problem can be formulated as the minimization problem (3.4) we have that

$$\mathbb{E}\left[f_1(x_1) + f_2(\boldsymbol{x}_2(\xi_{[2]}), \xi_2) + \dots + f_T\left(\boldsymbol{x}_T(\xi_{[T]}), \xi_T\right)\right] \ge \vartheta^*,$$
(5.261)

and equality in (5.261) holds iff the policy $x_t(\xi_{[t]})$ is optimal. The expectation in the left hand side of (5.261) can be estimated in a straightforward way. That is, generate random sample $\xi_1^j, ..., \xi_T^j, j = 1, ..., N$, of N realizations (scenarios) of the random data process $\xi_1, ..., \xi_T$ and estimate this expectation by the average

$$\frac{1}{N}\sum_{j=1}^{N} \left[f_1(x_1) + f_2\left(\boldsymbol{x}_2(\boldsymbol{\xi}_{[2]}^j), \boldsymbol{\xi}_2^j\right) + \dots + f_T\left(\boldsymbol{x}_T(\boldsymbol{\xi}_{[T]}^j), \boldsymbol{\xi}_T^j\right) \right].$$
(5.262)

Note that in order to construct the above estimator we do not need to generate a scenario tree, say by conditional sampling, we only need to generate a sample of single scenarios of the data process. The above estimator (5.262) is an unbiased estimator of the expectation in the left hand side of (5.261), and hence is a valid statistical upper bound for ϑ^* . Of course, quality of this upper bound depends on a successful choice of the feasible policy, i.e., on how small is the optimality gap between the left and right hand sides of (5.261). It also depends on variability of the estimator (5.262), which unfortunately often grows fast with increase of the number of stages.

We also may address the problem of validating a given feasible first stage solution $\bar{x}_1 \in \mathcal{X}_1$. The value of the multistage problem at \bar{x}_1 is given by the optimal value of the problem

$$\begin{array}{ll} \underset{\mathbf{x}_{2},...,\mathbf{x}_{T}}{\text{Min}} & f_{1}(\bar{x}_{1}) + \mathbb{E} \left[f_{2}(\mathbf{x}_{2}(\xi_{[2]}), \xi_{2}) + ... + f_{T}\left(\mathbf{x}_{T}(\xi_{[T]}), \xi_{T}\right) \right] \\ \text{s.t.} & \mathbf{x}_{t}(\xi_{[t]}) \in \mathcal{X}_{t}(\mathbf{x}_{t-1}(\xi_{[t-1]}), \xi_{t}), \ t = 2, ..., T. \end{array}$$

$$(5.263)$$

Recall that the optimization in (5.263) is performed over feasible policies. That is, in order to validate \bar{x}_1 we basically need to solve the corresponding T - 1 stage problems. Therefore, for T > 2 validation of \bar{x}_1 can be almost as difficult as solving the original problem.

5.8.2 Complexity Estimates of Multistage Programs

In order compute value of two stage stochastic program $\min_{x \in \mathcal{X}} \mathbb{E}[F(x,\xi)]$, where $F(x,\xi)$ is the optimal value of the corresponding second stage problem, at a feasible point $\bar{x} \in \mathcal{X}$ we need to calculate the expectation $\mathbb{E}[F(\bar{x},\xi)]$. This, in turn, involves two difficulties. First, the objective value $F(\bar{x},\xi)$ is not given explicitly, its calculation requires solution of the associated second stage optimization problem. Second, the multivariate integral $\mathbb{E}[F(\bar{x},\xi)]$ cannot be evaluated with a high accuracy even for moderate values of dimension d of the random data vector ξ . Monte Carlo techniques allow to evaluate $\mathbb{E}[F(\bar{x},\xi)]$ with accuracy $\varepsilon > 0$ by employing samples of size $N = O(\varepsilon^{-2})$. The required sample size N gives, in a sense, an estimate of complexity of evaluation of $\mathbb{E}[F(\bar{x},\xi)]$ since this is how many times we will need to solve the corresponding second stage problem. It is remarkable that in order to solve the two stage stochastic program with accuracy $\varepsilon > 0$, say by the SAA method, we need a sample size basically of the same order $N = O(\varepsilon^{-2})$. These complexity estimates were analyzed in details in section 5.3. Two basic conditions required for such analysis are that the problem has relatively complete recourse and that for given x and ξ the optimal value $F(x,\xi)$ of the second stage problem can be calculated with a high accuracy.

In this section we discuss analogous estimates of complexity of the SAA method applied to multistage stochastic programming problems. From the point of view of the SAA method it is natural to evaluate complexity of a multistage stochastic program in terms of the total number of scenarios required to find a first stage solution with a given accuracy $\varepsilon > 0$.

In order to simplify the presentation we consider three stage stochastic programs, say of the form (5.236)–(5.238). Assume that for every $x_1 \in \mathcal{X}_1$ the expectation $\mathbb{E}[Q_2(x_1, \xi_2)]$ is well defined and finite valued. In particular, this assumption implies that the problem has relatively complete recourse. Let us look at the problem of computing value of the first stage problem (5.238) at a feasible point $\bar{x}_1 \in \mathcal{X}_1$. Apart from the problem of evaluating the expectation $\mathbb{E}[Q_2(\bar{x}_1, \xi_2)]$, we also face here the problem of computing $Q_2(\bar{x}_1, \xi_2)$ for different realizations of random vector ξ_2 . For that we need to solve the two stage stochastic programming problem given in the right hand side of (5.237). As it was already discussed, in order to evaluate $Q_2(\bar{x}_1, \xi_2)$ with accuracy $\varepsilon > 0$ by solving the corresponding SAA problem, given in the right hand side of (5.240), we also need a sample of size $N_2 = O(\varepsilon^{-2})$. Recall that the total number of scenarios involved in evaluation of the sample average $\tilde{f}_{N_1,N_2}(\bar{x}_1)$, defined in (5.241), is $N = N_1N_2$. Therefore we will need $N = O(\varepsilon^{-4})$ scenarios just to compute value of the first stage problem at a given feasible point with accuracy ε by the SAA method. This indicates that complexity of the SAA method, applied to multistage stochastic programs, grows exponentially with increase of the number of stages.

We discuss now in details sample size estimates of the three stage SAA program (5.239)–(5.241). For the sake of simplicity we assume that the data process is stagewise independent, i.e., random vectors ξ_2 and ξ_3 are independent. Also similar to assumptions (M1)–(M5) of section 5.3 let us make the following assumptions.

- (M'1) For every $x_1 \in \mathcal{X}_1$ the expectation $\mathbb{E}[Q_2(x_1, \xi_2)]$ is well defined and finite valued.
- (M'2) The random vectors ξ_2 and ξ_3 are independent.
- (M'3) The set \mathcal{X}_1 has finite diameter D_1 .
- (M'4) There is a constant $L_1 > 0$ such that

$$\left|Q_2(x_1',\xi_2) - Q_2(x_1,\xi_2)\right| \le L_1 \|x_1' - x_1\|$$
(5.264)

for all $x'_1, x_1 \in \mathcal{X}_1$ and a.e. ξ_2 .

(M'5) There exists a constant $\sigma_1 > 0$ such that for any $x_1 \in \mathcal{X}_1$ it holds that

$$M_{1,x_1}(t) \le \exp\left\{\sigma_1^2 t^2/2\right\}, \quad \forall t \in \mathbb{R},$$
(5.265)

where $M_{1,x_1}(t)$ is the moment generating function of $Q_2(x_1,\xi_2) - \mathbb{E}[Q_2(x_1,\xi_2)]$.

- (M'6) There is a set C of finite diameter D_2 such that for every $x_1 \in \mathcal{X}_1$ and a.e. ξ_2 , the set $\mathcal{X}_2(x_1, \xi_2)$ is contained in C.
- (M'7) There is a constant $L_2 > 0$ such that

$$\left|Q_3(x_2',\xi_3) - Q_3(x_2,\xi_3)\right| \le L_2 \|x_2' - x_2\|$$
(5.266)

for all $x'_2, x_2 \in C$ and a.e. ξ_3 .

(M'8) There exists a constant $\sigma_2 > 0$ such that for any $x_2 \in \mathcal{X}_2(x_1, \xi_2)$ and all $x_1 \in \mathcal{X}_1$ and a.e. ξ_2 it holds that

$$M_{2,x_2}(t) \le \exp\left\{\sigma_2^2 t^2/2\right\}, \quad \forall t \in \mathbb{R},$$
(5.267)

where $M_{2,x_2}(t)$ is the moment generating function of $Q_3(x_2,\xi_3) - \mathbb{E}[Q_3(x_2,\xi_3)]$.

Theorem 5.35. Under assumptions (M'1)–(M'8) and for $\varepsilon > 0$ and $\alpha \in (0, 1)$, and the sample sizes N_1 and N_2 (using either independent or identical conditional sampling schemes) satisfying

$$\left[\frac{O(1)D_1L_1}{\varepsilon}\right]^{n_1} \exp\left\{-\frac{O(1)N_1\varepsilon^2}{\sigma_1^2}\right\} + \left[\frac{O(1)D_2L_2}{\varepsilon}\right]^{n_2} \exp\left\{-\frac{O(1)N_2\varepsilon^2}{\sigma_2^2}\right\} \le \alpha, \quad (5.268)$$

we have that any $\varepsilon/2$ -optimal solution of the SAA problem (5.241) is an ε -optimal solution of the first stage (5.238) of the true problem with probability at least $1 - \alpha$.

Proof. Proof of this theorem is based on the uniform exponential bound of Theorem 7.75. Let us sketch the arguments. Assume that the conditional sampling is identical. We have that for every $x_1 \in \mathcal{X}_1$ and $i = 1, ..., N_1$,

$$\left| \hat{Q}_{2,N_2}(x_1,\xi_2^i) - Q_2(x_1,\xi_2^i) \right| \le \sup_{x_2 \in \mathcal{C}} \left| \frac{1}{N_2} \sum_{j=1}^{N_2} Q_3(x_2,\xi_3^j) - \mathbb{E}[Q_3(x_2,\xi_3)] \right|,$$

where C is the set postulated in assumption (M'6). Consequently

$$\sup_{x_{1}\in\mathcal{X}_{1}} \left| \frac{1}{N_{1}} \sum_{i=1}^{N_{1}} \hat{Q}_{2,N_{2}}(x_{1},\xi_{2}^{i}) - \frac{1}{N_{1}} \sum_{i=1}^{N_{1}} Q_{2}(x_{1},\xi_{2}^{i}) \right| \\
\leq \frac{1}{N_{1}} \sum_{i=1}^{N_{1}} \sup_{x_{1}\in\mathcal{X}_{1}} \left| \hat{Q}_{2,N_{2}}(x_{1},\xi_{2}^{i}) - Q_{2}(x_{1},\xi_{2}^{i}) \right| \\
\leq \sup_{x_{2}\in\mathcal{C}} \left| \frac{1}{N_{2}} \sum_{j=1}^{N_{2}} Q_{3}(x_{2},\xi_{3}^{j}) - \mathbb{E}[Q_{3}(x_{2},\xi_{3})] \right|.$$
(5.269)

By the uniform exponential bound (7.242) we have that

$$\Pr\left\{\sup_{x_2\in\mathcal{C}}\left|\frac{1}{N_2}\sum_{j=1}^{N_2}Q_3(x_2,\xi_3^j) - \mathbb{E}[Q_3(x_2,\xi_3)]\right| > \varepsilon/2\right\}$$

$$\leq \left[\frac{O(1)D_2L_2}{\varepsilon}\right]^{n_2}\exp\left\{-\frac{O(1)N_2\varepsilon^2}{\sigma_2^2}\right\},$$
(5.270)

and hence

$$\Pr\left\{\sup_{x_1 \in \mathcal{X}_1} \left| \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{Q}_{2,N_2}(x_1,\xi_2^i) - \frac{1}{N_1} \sum_{i=1}^{N_1} Q_2(x_1,\xi_2^i) \right| > \varepsilon/2 \right\} \leq \left[\frac{O(1)D_2L_2}{\varepsilon} \right]^{n_2} \exp\left\{ -\frac{O(1)N_2\varepsilon^2}{\sigma_2^2} \right\}.$$
(5.271)

By the uniform exponential bound (7.242) we also have that

$$\Pr\left\{\sup_{x_1\in\mathcal{X}_1} \left|\frac{1}{N_1}\sum_{i=1}^{N_1}Q_2(x_1,\xi_2^i) - \mathbb{E}[Q_2(x_1,\xi_2)]\right| > \varepsilon/2\right\}$$

$$\leq \left[\frac{O(1)D_1L_1}{\varepsilon}\right]^{n_1}\exp\left\{-\frac{O(1)N_1\varepsilon^2}{\sigma_1^2}\right\}.$$
(5.272)

Let us observe that if Z_1, Z_2 are random variables, then

$$\Pr(Z_1 + Z_2 > \varepsilon) \le \Pr(Z_1 > \varepsilon/2) + \Pr(Z_2 > \varepsilon/2).$$

Therefore it follows from (5.271) and (5.271) that

$$\Pr\left\{\sup_{x_1\in\mathcal{X}_1}\left|\tilde{f}_{N_1,N_2}(x_1) - f_1(x_1) - \mathbb{E}[Q_2(x_1,\xi_2)]\right| > \varepsilon\right\}$$

$$\leq \left[\frac{O(1)D_1L_1}{\varepsilon}\right]^{n_1} \exp\left\{-\frac{O(1)N_1\varepsilon^2}{\sigma_1^2}\right\} + \left[\frac{O(1)D_2L_2}{\varepsilon}\right]^{n_2} \exp\left\{-\frac{O(1)N_2\varepsilon^2}{\sigma_2^2}\right\},$$
(5.273)

which implies the assertion of the theorem.

In the case of the independent conditional sampling the proof can be completed in a similar way. $\hfill\square$

Remark 20. We have, of course, that

$$\left|\hat{\vartheta}_{\mathcal{N}} - \vartheta^*\right| \le \sup_{x_1 \in \mathcal{X}_1} \left|\tilde{f}_{N_1, N_2}(x_1) - f_1(x_1) - \mathbb{E}[Q_2(x_1, \xi_2)]\right|.$$
(5.274)

Therefore bound (5.273) also implies that

$$\Pr\{\left|\hat{\vartheta}_{\mathcal{N}} - \vartheta^*\right| > \varepsilon\} \leq \left[\frac{O(1)D_1L_1}{\varepsilon}\right]^{n_1} \exp\left\{-\frac{O(1)N_1\varepsilon^2}{\sigma_1^2}\right\} + \left[\frac{O(1)D_2L_2}{\varepsilon}\right]^{n_2} \exp\left\{-\frac{O(1)N_2\varepsilon^2}{\sigma_2^2}\right\}.$$
(5.275)

In particular, suppose that $N_1 = N_2$. Then for

 $n := \max\{n_1, n_2\}, \ L := \max\{L_1, L_2\}, \ D := \max\{D_1, D_2\}, \ \sigma := \max\{\sigma_1, \sigma_2\},$

the estimate (5.268) implies the following estimate of the required sample size $N_1 = N_2$:

$$\left(\frac{O(1)DL}{\varepsilon}\right)^n \exp\left\{-\frac{O(1)N_1\varepsilon^2}{\sigma^2}\right\} \le \alpha,\tag{5.276}$$

which is equivalent to

$$N_1 \ge \frac{O(1)\sigma^2}{\varepsilon^2} \left[n \ln\left(\frac{O(1)DL}{\varepsilon}\right) + \ln\left(\frac{1}{\alpha}\right) \right].$$
(5.277)

The estimate (5.277), for 3-stage programs, looks similar to the estimate (5.116), of Theorem 5.18, for two-stage programs. Recall, however, that if we use the SAA method with conditional sampling and respective sample sizes N_1 and N_2 , then the total number of scenarios is $N = N_1N_2$. Therefore, our analysis indicates that for 3-stage problems we need random samples with the total number of scenarios of order of the *square* of the corresponding sample size for two-stage problems. This analysis can be extended to *T*-stage problems with the conclusion that the total number of scenarios needed to solve the true problem with a reasonable accuracy grows *exponentially* with increase of the number of stages *T*. Some numerical experiments seem to confirm this conclusion. Of course, it should be mentioned that the above analysis does *not* prove in a *rigorous* mathematical sense that complexity of multistage programming grows exponentially with increase of the number of stages. It only indicates that if we measure computational complexity in terms of the *number of generated scenarios*, the SAA method, which showed a considerable promise for solving two stage problems, could be practically inapplicable for solving multistage problems with a large (say greater than 3) number of stages.

5.9 Stochastic Approximation Method

To an extent this section is based on Nemirovski et al [161]. Consider stochastic optimization problem (5.1). We assume that the expected value function $f(x) = \mathbb{E}[F(x,\xi)]$ is well defined, finite valued and continuous at every $x \in \mathcal{X}$, and that the set $\mathcal{X} \subset \mathbb{R}^n$ is nonempty, closed and bounded. We denote by \bar{x} an optimal solution of problem (5.1). Such an optimal solution does exist since the set \mathcal{X} is compact and f(x) is continuous. Clearly, $\vartheta^* = f(\bar{x})$ (recall that ϑ^* denotes the optimal value of problem (5.1)). We also assume throughout this section that the set \mathcal{X} is *convex* and the function $f(\cdot)$ is *convex*. Of course, if $F(\cdot, \xi)$ is convex for every $\xi \in \Xi$, then convexity of $f(\cdot)$ follows. We assume availability of the following *stochastic oracle:*

There is a mechanism which for every given x ∈ X and ξ ∈ Ξ returns value F(x, ξ) and a stochastic subgradient, a vector G(x, ξ) such that g(x) := E[G(x, ξ)] is well defined and is a subgradient of f(·) at x, i.e., g(x) ∈ ∂f(x).

Remark 21. Recall that if $F(\cdot, \xi)$ is convex for every $\xi \in \Xi$, and x is an interior point of \mathcal{X} , i.e., $f(\cdot)$ is finite valued in a neighborhood of x, then

$$\partial f(x) = \mathbb{E}\left[\partial_x F(x,\xi)\right] \tag{5.278}$$

(see Theorem 7.52). Therefore, in that case we can employ a measurable selection $G(x,\xi) \in \partial_x F(x,\xi)$ as a stochastic subgradient. Note also that for an implementation of a stochastic approximation algorithm we only need to employ stochastic subgradients, while objective values $F(x,\xi)$ are used for accuracy estimates in section 5.9.4.

We also assume that we can generate, say by Monte Carlo sampling techniques, an iid sequence ξ^j , j = 1, ..., of realizations of the random vector ξ , and hence to compute a stochastic subgradient $G(x_j, \xi^j)$ at an iterate point $x_j \in X$.

5.9.1 Classical Approach

We denote by $||x||_2 = (x^{\mathsf{T}}x)^{1/2}$ the Euclidean norm of vector $x \in \mathbb{R}^n$, and by

$$\Pi_{\mathcal{X}}(x) := \arg\min_{z \in \mathcal{X}} \|x - z\|_2 \tag{5.279}$$

the metric projection of x onto the set \mathcal{X} . Since \mathcal{X} is convex and closed, the minimizer in the right hand side of (5.279) exists and is unique. Note that $\Pi_{\mathcal{X}}$ is a nonexpanding operator, i.e.,

$$\|\Pi_{\mathcal{X}}(x') - \Pi_{\mathcal{X}}(x)\|_{2} \le \|x' - x\|_{2}, \quad \forall x', x \in \mathbb{R}^{n}.$$
(5.280)

The classical Stochastic Approximation (SA) algorithm solves problem (5.1) by mimicking a simple subgradient descent method. That is, for chosen initial point $x_1 \in \mathcal{X}$ and a sequence $\gamma_j > 0, j = 1, ...,$ of stepsizes, it generates the iterates by the formula

$$x_{j+1} = \prod_{\mathcal{X}} (x_j - \gamma_j G(x_j, \xi^j)).$$
(5.281)

The crucial question of that approach is how to choose the stepsizes γ_j . Also the set \mathcal{X} should be simple enough so that the corresponding projection can be easily calculated. We

are going now to analyze convergence of the iterates, generated by this procedure, to an optimal solution \bar{x} of problem (5.1). Note that the iterate $x_{j+1} = x_{j+1}(\xi_{[j]})$, j = 1, ..., is a function of the history $\xi_{[j]} = (\xi^1, ..., \xi^j)$ of the generated random process and hence is random, while the initial point x_1 is given (deterministic). We assume that there is number M > 0 such that

$$\mathbb{E}\left[\|G(x,\xi)\|_2^2\right] \le M^2, \quad \forall x \in \mathcal{X}.$$
(5.282)

Note that since for a random variable Z it holds that $\mathbb{E}[Z^2] \ge (\mathbb{E}|Z|)^2$, it follows from (5.282) that $\mathbb{E}||G(x,\xi)|| \le M$.

Denote

$$A_j := \frac{1}{2} \|x_j - \bar{x}\|_2^2 \quad \text{and} \quad a_j := \mathbb{E}[A_j] = \frac{1}{2} \mathbb{E}[\|x_j - \bar{x}\|_2^2]. \tag{5.283}$$

By (5.280) and since $\bar{x} \in \mathcal{X}$ and hence $\Pi_{\mathcal{X}}(\bar{x}) = \bar{x}$, we have

$$A_{j+1} = \frac{1}{2} \| \Pi_{\mathcal{X}} (x_j - \gamma_j G(x_j, \xi^j)) - \bar{x} \|_2^2$$

$$= \frac{1}{2} \| \Pi_{\mathcal{X}} (x_j - \gamma_j G(x_j, \xi^j)) - \Pi_{\mathcal{X}} (\bar{x}) \|_2^2$$

$$\leq \frac{1}{2} \| x_j - \gamma_j G(x_j, \xi^j) - \bar{x} \|_2^2$$

$$= A_j + \frac{1}{2} \gamma_j^2 \| G(x_j, \xi^j) \|_2^2 - \gamma_j (x_j - \bar{x})^{\mathsf{T}} G(x_j, \xi^j).$$
(5.284)

Since $x_j = x_j(\xi_{[j-1]})$ is independent of ξ_j , we have

$$\mathbb{E}\left[(x_j - \bar{x})^\mathsf{T} G(x_j, \xi^j) \right] = \mathbb{E}\left\{ \mathbb{E}\left[(x_j - \bar{x})^\mathsf{T} G(x_j, \xi^j) | \xi_{[j-1]} \right] \right\}$$

= $\mathbb{E}\left\{ (x_j - \bar{x})^\mathsf{T} \mathbb{E}[G(x_j, \xi^j) | \xi_{[j-1]}] \right\}$
= $\mathbb{E}\left[(x_j - \bar{x})^\mathsf{T} g(x_j) \right].$

Therefore, by taking expectation of both sides of (5.284) and since (5.282) we obtain

$$a_{j+1} \le a_j - \gamma_j \mathbb{E}\left[(x_j - \bar{x})^{\mathsf{T}} g(x_j) \right] + \frac{1}{2} \gamma_j^2 M^2.$$
 (5.285)

Suppose, further, that the expectation function f(x) is differentiable and strongly convex on \mathcal{X} with parameter c > 0, i.e.,

$$(x'-x)^{\mathsf{T}}(\nabla f(x') - \nabla f(x)) \ge c \|x'-x\|_{2}^{2}, \quad \forall x, x' \in \mathcal{X}.$$
(5.286)

Note that strong convexity of f(x) implies that the minimizer \bar{x} is unique, and that because of differentiability of f(x) it follows that $\partial f(x) = \{\nabla f(x)\}$ and hence $g(x) = \nabla f(x)$. By optimality of \bar{x} we have that

$$(x - \bar{x})^{\top} \nabla f(\bar{x}) \ge 0, \quad \forall x \in \mathcal{X},$$

$$(5.287)$$

which together with (5.286) implies that

$$\mathbb{E}\left[(x_j - \bar{x})^T \nabla f(x_j)\right] \geq \mathbb{E}\left[(x_j - \bar{x})^T (\nabla f(x_j) - \nabla f(\bar{x}))\right] \\
\geq c \mathbb{E}\left[\|x_j - \bar{x}\|_2^2\right] = 2ca_j.$$
(5.288)

Therefore it follows from (5.285) that

$$a_{j+1} \le (1 - 2c\gamma_j)a_j + \frac{1}{2}\gamma_j^2 M^2.$$
(5.289)

In the classical approach to stochastic approximation the employed stepsizes are $\gamma_j := \theta/j$ for some constant $\theta > 0$. Then by (5.289) we have

$$a_{j+1} \le (1 - 2c\theta/j)a_j + \frac{1}{2}\theta^2 M^2/j^2.$$
 (5.290)

Suppose now that $\theta > 1/(2c)$. Then it follows from (5.290) by induction that for j = 1, ...,

$$2a_j \le \frac{\max\left\{\theta^2 M^2 (2c\theta - 1)^{-1}, 2a_1\right\}}{j}.$$
(5.291)

Recall that $2a_j = \mathbb{E} \left[\|x_j - \bar{x}\|^2 \right]$ and, since x_1 is deterministic, $2a_1 = \|x_1 - \bar{x}\|_2^2$. Therefore, by (5.291) we have that

$$\mathbb{E}\left[\|x_j - \bar{x}\|_2^2\right] \le \frac{Q(\theta)}{j},\tag{5.292}$$

where

$$Q(\theta) := \max\left\{\theta^2 M^2 (2c\theta - 1)^{-1}, \|x_1 - \bar{x}\|_2^2\right\}.$$
(5.293)

The constant $Q(\theta)$ attains its optimal (minimal) value at $\theta = 1/c$.

Suppose, further, that \bar{x} is an *interior* point of \mathcal{X} and $\nabla f(x)$ is Lipschitz continuous, i.e., there is constant L > 0 such that

$$\|\nabla f(x') - \nabla f(x)\|_2 \le L \|x' - x\|_2, \quad \forall x', x \in \mathcal{X}.$$
(5.294)

Then

$$f(x) \le f(\bar{x}) + \frac{1}{2}L \|x - \bar{x}\|_2^2, \quad \forall x \in \mathcal{X},$$
(5.295)

and hence by (5.292)

$$\mathbb{E}[f(x_j) - f(\bar{x})] \le \frac{1}{2}L \mathbb{E}\left[\|x_j - \bar{x}\|_2^2\right] \le \frac{Q(\theta)L}{2j}.$$
(5.296)

We obtain that under the specified assumptions, after j iterations the expected error of the current solution in terms of the distance to the true optimal solution \bar{x} is of order $O(j^{-1/2})$, and the expected error in terms of the objective value is of order $O(j^{-1})$, provided that $\theta > 1/(2c)$. Note, however, that the classical stepsize rule $\gamma_j = \theta/j$ could be very dangerous if the parameter c of strong convexity is overestimated, i.e., if $\theta < 1/(2c)$.

Example 5.36 As a simple example consider $f(x) := \frac{1}{2}\kappa x^2$, with $\kappa > 0$ and $\mathcal{X} := [-1,1] \subset \mathbb{R}$, and assume that there is no noise, i.e., $G(x,\xi) \equiv \nabla f(x)$. Clearly $\bar{x} = 0$ is the optimal solution and zero is the optimal value of the corresponding optimization (minimization) problem. Let us take $\theta = 1$, i.e., use stepsizes $\gamma_j = 1/j$, in which case the iteration process becomes

$$x_{j+1} = x_j - f'(x_j)/j = \left(1 - \frac{\kappa}{j}\right) x_j.$$
 (5.297)

For $\kappa = 1$ the above choice of the stepsizes is optimal and the optimal solution is obtained in one iteration. Suppose now that $\kappa < 1$. Then starting with $x_1 > 0$, we have

$$x_{j+1} = x_1 \prod_{s=1}^{j} \left(1 - \frac{\kappa}{s}\right) = x_1 \exp\left\{-\sum_{s=1}^{j} \ln\left(1 + \frac{\kappa}{s - \kappa}\right)\right\} > x_1 \exp\left\{-\sum_{s=1}^{j} \frac{\kappa}{s - \kappa}\right\}$$

Moreover,

$$\sum_{s=1}^{j} \frac{\kappa}{s-\kappa} \le \frac{\kappa}{1-\kappa} + \int_{1}^{j} \frac{\kappa}{t-\kappa} dt < \frac{\kappa}{1-\kappa} + \kappa \ln j - \kappa \ln(1-\kappa).$$

It follows that

$$x_{j+1} > O(1) j^{-\kappa}$$
 and $f(x_{j+1}) > O(1) j^{-2\kappa}, \ j = 1, \dots$ (5.298)

(In the first of the above inequalities the constant $O(1) = x_1 \exp\{-\kappa/(1-\kappa)+\kappa \ln(1-\kappa)\}$, and in the second inequality the generic constant O(1) is obtained from the first one by taking square and multiplying it by $\kappa/2$.) That is, the convergence becomes extremely slow for small κ close to zero. In order to reduce the value x_j (the objective value $f(x_j)$) by factor 10, i.e., to improve the error of current solution by one digit, we will need to increase the number of iterations j by factor $10^{1/\kappa}$ (by factor $10^{1/(2\kappa)}$). For example, for $\kappa = 0.1, x_1 = 1$ and $j = 10^5$ we have that $x_j > 0.28$. In order to reduce the error of the iterate to 0.028 we will need to increase the number of iterations by factor 10^{10} , i.e., to $j = 10^{15}$.

It could be added that if f(x) loses strong convexity, i.e., the parameter c degenerates to zero, and hence no choice of $\theta > 1/(2c)$ is possible, then the stepsizes $\gamma_j = \theta/j$ may become completely unacceptable for any choice of θ .

5.9.2 Robust SA Approach

It was argued in the previous section 5.9.1 that the classical stepsizes $\gamma_j = O(j^{-1})$ can be too small to ensure a reasonable rate of convergence even in the "no noise" case. An important improvement of the SA method was developed by Polyak [188] and Polyak and Juditsky [189], where longer stepsizes were suggested with consequent averaging of the obtained iterates. Under the outlined "classical" assumptions, the resulting algorithm exhibits the same optimal $O(j^{-1})$ asymptotical convergence rate, while using an easy to implement and "robust" stepsize policy. The main ingredients of Polyak's scheme (long steps and averaging) were, in a different form, proposed already in Nemirovski and Yudin [163] for problems with general type Lipschitz continuous convex objectives and for convexconcave saddle point problems. Results of this section go back to Nemirovski and Yudin [163],[164].

Recall that $g(x) \in \partial f(x)$ and $a_j = \frac{1}{2}\mathbb{E}\left[\|x_j - \bar{x}\|_2^2\right]$, and we assume the boundedness condition (5.282). By convexity of f(x) we have that $f(x) \ge f(x_j) + (x - x_j)^{\mathsf{T}} g(x_j)$ for any $x \in \mathcal{X}$, and hence

$$\mathbb{E}\left[(x_j - \bar{x})^{\mathsf{T}} g(x_j)\right] \ge \mathbb{E}\left[f(x_j) - f(\bar{x})\right].$$
(5.299)

Together with (5.285) this implies

$$\gamma_j \mathbb{E} \big[f(x_j) - f(\bar{x}) \big] \le a_j - a_{j+1} + \frac{1}{2} \gamma_j^2 M^2.$$

It follows that whenever $1 \le i \le j$, we have

$$\sum_{t=i}^{j} \gamma_t \mathbb{E} \left[f(x_t) - f(\bar{x}) \right] \le \sum_{t=i}^{j} \left[a_t - a_{t+1} \right] + \frac{1}{2} M^2 \sum_{t=i}^{j} \gamma_t^2 \le a_i + \frac{1}{2} M^2 \sum_{t=i}^{j} \gamma_t^2.$$
(5.300)

Denote

$$\nu_t := \frac{\gamma_t}{\sum_{\tau=i}^j \gamma_\tau} \text{ and } D_{\mathcal{X}} := \max_{x \in \mathcal{X}} \|x - x_1\|_2.$$
(5.301)

Clearly $\nu_t \ge 0$ and $\sum_{t=i}^{j} \nu_t = 1$. By (5.300) we have

$$\mathbb{E}\left[\sum_{t=i}^{j} \nu_t f(x_t) - f(\bar{x})\right] \le \frac{a_i + \frac{1}{2}M^2 \sum_{t=i}^{j} \gamma_t^2}{\sum_{t=i}^{j} \gamma_t}.$$
(5.302)

Consider points

$$\tilde{x}_{i,j} := \sum_{t=i}^{j} \nu_t x_t.$$
(5.303)

Since \mathcal{X} is convex it follows that $\tilde{x}_{i,j} \in \mathcal{X}$ and by convexity of $f(\cdot)$ we have

$$f(\tilde{x}_{i,j}) \le \sum_{t=i}^{j} \nu_t f(x_t)$$

Thus, by (5.302) and in view of $a_1 \leq D_{\mathcal{X}}^2$ and $a_i \leq 4D_{\mathcal{X}}^2$, i > 1, we get

$$\mathbb{E}\left[f(\tilde{x}_{1,j}) - f(\bar{x})\right] \le \frac{D_{\mathcal{X}}^2 + M^2 \sum_{t=1}^j \gamma_t^2}{2\sum_{t=1}^j \gamma_t}, \text{ for } 1 \le j,$$
(5.304)

$$\mathbb{E}\left[f(\tilde{x}_{i,j}) - f(\bar{x})\right] \le \frac{4D_{\mathcal{X}}^2 + M^2 \sum_{t=i}^j \gamma_t^2}{2\sum_{t=i}^j \gamma_t}, \text{ for } 1 < i \le j.$$
(5.305)

Based of the above bounds on the expected accuracy of approximate solutions $\tilde{x}_{i,j}$, we can now develop "reasonable" stepsize policies along with the associated efficiency estimates.

Constant stepsizes and error estimates

Assume now that the number of iterations of the method is fixed in advance, say equal to N, and that we use the *constant* stepsize policy, i.e., $\gamma_t = \gamma$, t = 1, ..., N. It follows then from (5.304) that

$$\mathbb{E}\left[f(\tilde{x}_{1,N}) - f(\bar{x})\right] \le \frac{D_{\mathcal{X}}^2 + M^2 N \gamma^2}{2N\gamma}.$$
(5.306)

Minimizing the right hand side of (5.306) over $\gamma > 0$, we arrive at the *constant* stepsize policy

$$\gamma_t = \frac{D_X}{M\sqrt{N}}, \ t = 1, ..., N,$$
 (5.307)

along with the associated efficiency estimate

$$\mathbb{E}\left[f(\tilde{x}_{1,N}) - f(\bar{x})\right] \le \frac{D_{\mathcal{X}}M}{\sqrt{N}}.$$
(5.308)

By (5.305), with the constant stepsize policy (5.307), we also have for $1 \le K \le N$,

$$\mathbb{E}\left[f(\tilde{x}_{K,N}) - f(\bar{x})\right] \le \frac{C_{N,K} D_{\mathcal{X}} M}{\sqrt{N}},\tag{5.309}$$

where

$$C_{N,K} := \frac{2N}{N - K + 1} + \frac{1}{2}.$$

When $K/N \le 1/2$, the right hand side of (5.309) coincides, within an absolute constant factor, with the right hand side of (5.308). If we change the stepsizes (5.307) by a factor of $\theta > 0$, i.e., use the stepsizes

$$\gamma_t = \frac{\theta D_{\mathcal{X}}}{M\sqrt{N}}, \quad t = 1, \dots, N, \tag{5.310}$$

then the efficiency estimate (5.309) becomes

$$\mathbb{E}\left[f(\tilde{x}_{K,N}) - f(\bar{x})\right] \le \max\left\{\theta, \theta^{-1}\right\} \frac{C_{N,K} D_{\mathcal{X}} M}{\sqrt{N}}.$$
(5.311)

The expected error of the iterates (5.303), with constant stepsize policy (5.310), after N iterations is $O(N^{-1/2})$. Of course, this is worse than the rate $O(N^{-1})$ for the classical SA algorithm as applied to a smooth strongly convex function attaining minimum at an interior point of the set \mathcal{X} . However, the error bound (5.311) is guaranteed independently of any smoothness and/or strong convexity assumptions on $f(\cdot)$. Moreover, changing the stepsizes by factor θ results just in rescaling of the corresponding error estimate (5.311). This is in a sharp contrast with the classical approach discussed in the previous section, when such change of stepsizes can be a disaster. These observations, in particular the fact that there is no necessity in "fine tuning" the stepsizes to the objective function $f(\cdot)$, explains the adjective "robust" in the name of the method.

It can be interesting to compare sample size estimates derived from the error bounds of the (robust) SA approach with respective sample size estimates of the SAA method discussed in section 5.3.2. By Chebyshev (Markov) inequality we have that for $\varepsilon > 0$,

$$\Pr\left\{f(\tilde{x}_{1,N}) - f(\bar{x}) \ge \varepsilon\right\} \le \varepsilon^{-1} \mathbb{E}\left[f(\tilde{x}_{1,N}) - f(\bar{x})\right].$$
(5.312)

Together with (5.308) this implies that, for the constant stepsize policy (5.307),

$$\Pr\left\{f(\tilde{x}_{1,N}) - f(\bar{x}) \ge \varepsilon\right\} \le \frac{D_{\mathcal{X}}M}{\varepsilon\sqrt{N}}.$$
(5.313)

It follows that for $\alpha \in (0, 1)$ and sample size

$$N \ge \frac{D_{\mathcal{X}}^2 M^2}{\varepsilon^2 \alpha^2} \tag{5.314}$$

we are guaranteed that $\tilde{x}_{1,N}$ is an ε -optimal solution of the "true" problem (5.1) with probability at least $1 - \alpha$.

Compared with the corresponding estimate (5.126) for the sample size by the SAA method, the above estimate (5.314) is of the same order with respect to parameters $D_{\mathcal{X}}, M$ and ε . On the other hand, the dependence on the significance level α is different, in (5.126) it is of order $O(\ln(\alpha^{-1}))$, while in (5.314) it is of order $O(\alpha^{-2})$. It is possible to derive better estimates, similar to the respective estimates of the SAA method, of the required sample size by using Large Deviations theory, we will discuss this further in the next section (see Theorem 5.41 in particular).

5.9.3 Mirror Descent SA Method

The robust SA approach discussed in the previous section is tailored to Euclidean structure of the space \mathbb{R}^n . In this section we discuss a generalization of the Euclidean SA approach allowing to adjust, to some extent, the method to the geometry, not necessary Euclidean, of the problem in question. A rudimentary form of the following generalization can be found in Nemirovski and Yudin [164], from where the name "Mirror Descent" originates.

In this section we denote by $\|\cdot\|$ a *general* norm on \mathbb{R}^n . Its dual norm is defined as

$$||x||_* := \sup_{||y|| \le 1} y^\mathsf{T} x.$$

By $||x||_p := (|x_1|^p + \dots + |x_n|^p)^{1/p}$ we denote the $\ell_p, p \in [1, \infty)$, norm on \mathbb{R}^n . In particular, $||\cdot||_2$ is the Euclidean norm. Recall that the dual of $||\cdot||_p$ is the norm $||\cdot||_q$, where q > 1 is such that 1/p+1/q = 1. The dual norm of ℓ_1 norm $||x||_1 = |x_1|+\dots+|x_n|$ is the ℓ_∞ norm $||x||_\infty = \max\{|x_1|,\dots,|x_n|\}$.

Definition 5.37. We say that a function $\mathfrak{d} : \mathcal{X} \to \mathbb{R}$ is a distance generating function with modulus $\kappa > 0$ with respect to norm $\|\cdot\|$ if the following holds: $\mathfrak{d}(\cdot)$ is convex continuous on \mathcal{X} , the set

$$\mathcal{X}^{\star} := \{ x \in \mathcal{X} : \partial \mathfrak{d}(x) \neq \emptyset \}$$
(5.315)

is convex, $\mathfrak{d}(\cdot)$ is continuously differentiable on \mathcal{X}^* , and

$$(x'-x)^{\mathsf{T}}(\nabla\mathfrak{d}(x')-\nabla\mathfrak{d}(x)) \ge \kappa \|x'-x\|^2, \ \forall x, x' \in \mathcal{X}^{\star}.$$
(5.316)

Note that the set \mathcal{X}^* includes the relative interior of the set \mathcal{X} , and hence condition (5.316) implies that $\mathfrak{d}(\cdot)$ is strongly convex on \mathcal{X} with the parameter κ taken with respect to the considered norm $\|\cdot\|$.

A simple example of a distance generating function (with modulus 1 with respect to the Euclidean norm) is $\mathfrak{d}(x) := \frac{1}{2}x^{\mathsf{T}}x$. Of course, this function is continuously differentiable at every $x \in \mathbb{R}^n$. Another interesting example is the *entropy function*

$$\mathfrak{d}(x) := \sum_{i=1}^{n} x_i \ln x_i, \tag{5.317}$$

defined on the standard simplex $\mathcal{X} := \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1, x \ge 0\}$ (note that by continuity, $x \ln x = 0$ for x = 0). Here the set \mathcal{X}^* is formed by points $x \in \mathcal{X}$ having all coordinates different from zero. The set \mathcal{X}^* is the subset of \mathcal{X} of those points at which the entropy function is differentiable with $\nabla \mathfrak{d}(x) = (1 + \ln x_1, \dots, 1 + \ln x_n)$. The entropy function is strongly convex with modulus 1 on standard simplex with respect to $\|\cdot\|_1$ norm.

Indeed, it suffices to verify that $h^{\mathsf{T}} \nabla^2 \mathfrak{d}(x) h \ge ||h||_1^2$ for every $h \in \mathbb{R}^n$ and $x \in \mathcal{X}^*$. This, in turn, is verified by

$$\begin{bmatrix} \sum_{i} |h_{i}| \end{bmatrix}^{2} = \begin{bmatrix} \sum_{i} (x_{i}^{-1/2} |h_{i}|) x_{i}^{1/2} \end{bmatrix}^{2} \leq \begin{bmatrix} \sum_{i} h_{i}^{2} x_{i}^{-1} \end{bmatrix} \begin{bmatrix} \sum_{i} x_{i} \end{bmatrix}$$

= $\sum_{i} h_{i}^{2} x_{i}^{-1} = h^{T} \nabla^{2} \mathfrak{d}(x) h,$ (5.318)

where the inequality follows by Cauchy inequality.

Let us define function $V : \mathcal{X}^* \times \mathcal{X} \to \mathbb{R}_+$ as follows

$$V(x,z) := \mathfrak{d}(z) - [\mathfrak{d}(x) + \nabla \mathfrak{d}(x)^{\mathsf{T}}(z-x)].$$
(5.319)

In what follows we refer to $V(\cdot, \cdot)$ as *prox-function*¹⁹ associated with distance generating function $\mathfrak{d}(x)$. Note that $V(x, \cdot)$ is nonnegative and is strongly convex with modulus κ with respect to the norm $\|\cdot\|$. Let us define *prox-mapping* $P_x : \mathbb{R}^n \to \mathcal{X}^*$, associated with the distance generating function and a point $x \in \mathcal{X}^*$, viewed as a parameter, as follows

$$P_x(y) := \arg\min_{z \in \mathcal{X}} \left\{ y^{\mathsf{T}}(z-x) + V(x,z) \right\}.$$
 (5.320)

Observe that the minimum in the right hand side of (5.320) is attained since $\mathfrak{d}(\cdot)$ is continuous on \mathcal{X} and \mathcal{X} is compact, and a corresponding minimizer is unique since $V(x, \cdot)$ is strongly convex on \mathcal{X} . Moreover, by the definition of the set \mathcal{X}^* , all these minimizers belong to \mathcal{X}^* . Thus, the prox-mapping is well defined.

For the (Euclidean) distance generating function $\mathfrak{d}(x) := \frac{1}{2}x^{\mathsf{T}}x$, we have that $P_x(y) = \prod_{\mathcal{X}} (x - y)$. In that case the iteration formula (5.281) of the SA algorithm can be written as

$$x_{j+1} = P_{x_j}(\gamma_j G(x_j, \xi^j)), \quad x_1 \in \mathcal{X}^{\star}.$$
 (5.321)

Our goal is to demonstrate that the main properties of the recurrence (5.281) are inherited by (5.321) for any distance generating function $\mathfrak{d}(x)$.

Lemma 5.38. For every $u \in \mathcal{X}, x \in \mathcal{X}^*$ and $y \in \mathbb{R}^n$ one has

$$V(P_x(y), u) \le V(x, u) + y^{\mathsf{T}}(u - x) + (2\kappa)^{-1} ||y||_*^2.$$
(5.322)

Proof. Let $x \in \mathcal{X}^*$ and $v := P_x(y)$. Note that v is of the form $\operatorname{argmin}_{z \in \mathcal{X}} [h^{\mathsf{T}}z + \mathfrak{d}(z)]$ and thus $v \in X^*$, so that $\mathfrak{d}(\cdot)$ is differentiable at v. Since $\nabla_v V(x, v) = \nabla \mathfrak{d}(v) - \nabla \mathfrak{d}(x)$, the optimality conditions for (5.320) imply that

$$[\nabla \mathfrak{d}(v) - \nabla \mathfrak{d}(x) + y)^{\mathsf{T}}(v - u) \le 0, \quad \forall u \in \mathcal{X}.$$
 (5.323)

 $^{^{19}\}text{The}$ function $V(\cdot,\cdot)$ is also called Bregman divergence.

Therefore, for $u \in \mathcal{X}$ we have

$$\begin{aligned} V(v,u) - V(x,u) &= \left[\mathfrak{d}(u) - \nabla \mathfrak{d}(v)^{\mathsf{T}}(u-v) - \mathfrak{d}(v) \right] - \left[\mathfrak{d}(u) - \nabla \mathfrak{d}(x)^{\mathsf{T}}(u-x) - \mathfrak{d}(x) \right] \\ &= \left(\nabla \mathfrak{d}(v) - \nabla \mathfrak{d}(x) + y \right)^{\mathsf{T}}(v-u) + y^{\mathsf{T}}(u-v) - \left[\mathfrak{d}(v) - \nabla \mathfrak{d}(x)^{\mathsf{T}}(v-x) - \mathfrak{d}(x) \right] \\ &\leq y^{\mathsf{T}}(u-v) - V(x,v), \end{aligned}$$

where the last inequality follows by (5.323).

For any $a, b \in \mathbb{R}^n$ we have by the definition of the dual norm that $||a||_* ||b|| \ge a^{\mathsf{T}}b$ and hence

$$(\|a\|_*^2/\kappa + \kappa\|b\|^2)/2 \ge \|a\|_*\|b\| \ge a^{\mathsf{T}}b.$$
(5.324)

Applying this inequality with a = y and b = x - v we obtain

$$y^{\mathsf{T}}(x-v) \le \frac{\|y\|_{*}^{2}}{2\kappa} + \frac{\kappa}{2} \|x-v\|^{2}.$$

Also due to the strong convexity of $V(x, \cdot)$ and since V(x, x) = 0 we have

$$V(x,v) \geq V(x,x) + (x-v)^{\mathsf{T}} \nabla_{v} V(x,v) + \frac{1}{2} \kappa \|x-v\|^{2}$$

= $(x-v)^{\mathsf{T}} (\nabla \mathfrak{d}(v) - \nabla \mathfrak{d}(x)) + \frac{1}{2} \kappa \|x-v\|^{2}$
 $\geq \frac{1}{2} \kappa \|x-v\|^{2},$ (5.325)

where the last inequality holds by convexity of $\mathfrak{d}(\cdot)$. We get

$$\begin{split} V(v,u) - V(x,u) &\leq y^\mathsf{T}(u-v) - V(x,v) = y^\mathsf{T}(u-x) + y^\mathsf{T}(x-v) - V(x,v) \\ &\leq y^\mathsf{T}(u-x) + (2\kappa)^{-1} \|y\|_*^2, \end{split}$$

as required in (5.322). \Box

Using (5.322) with $x = x_j$, $y = \gamma_j G(x_j, \xi^j)$ and $u = \bar{x}$, and noting that by (5.321) $x_{j+1} = P_x(y)$ here, we get

$$\gamma_j (x_j - \bar{x})^{\mathsf{T}} G(x_j, \xi^j) \le V(x_j, \bar{x}) - V(x_{j+1}, \bar{x}) + \frac{\gamma_j^2}{2\kappa} \|G(x_j, \xi^j)\|_*^2.$$
(5.326)

Let us observe that for the Euclidean distance generating function $\vartheta(x) = \frac{1}{2}x^{\mathsf{T}}x$, one has $V(x, z) = \frac{1}{2}||x - z||_2^2$ and $\kappa = 1$. That is, in the Euclidean case (5.326) becomes

$$\frac{1}{2} \|x_{j+1} - \bar{x}\|_2^2 \le \frac{1}{2} \|x_j - \bar{x}\|_2^2 + \frac{1}{2}\gamma_j^2 \|G(x_j, \xi^j)\|_2^2 - \gamma_j (x_j - \bar{x})^\mathsf{T} G(x_j, \xi^j).$$
(5.327)

The above inequality is exactly the relation (5.284), which played a crucial role in the developments related to the Euclidean SA. We are about to process, in a similar way, the relation (5.326) in the case of a general distance generating function, thus arriving at the Mirror Descent SA.

Specifically, setting

$$\Delta_j := G(x_j, \xi^j) - g(x_j), \tag{5.328}$$

we can rewrite (5.326), with *j* replaced by *t*, as

$$\gamma_t(x_t - \bar{x})^{\mathsf{T}} g(x_t) \le V(x_t, \bar{x}) - V(x_{t+1}, \bar{x}) - \gamma_t \Delta_t^{\mathsf{T}}(x_t - \bar{x}) + \frac{\gamma_t^2}{2\kappa} \|G(x_t, \xi^t)\|_*^2.$$
(5.329)

Summing up over t = 1, ..., j, and taking into account that $V(x_{j+1}, u) \ge 0, u \in X$, we get

$$\sum_{t=1}^{j} \gamma_t (x_t - \bar{x})^{\mathsf{T}} g(x_t) \le V(x_1, \bar{x}) + \sum_{t=1}^{j} \frac{\gamma_t^2}{2\kappa} \|G(x_t, \xi^t)\|_*^2 - \sum_{t=1}^{j} \gamma_t \Delta_t^{\mathsf{T}} (x_t - \bar{x}).$$
(5.330)
Set $\nu_t := \frac{\gamma_t}{\sum_{\tau=1}^{j} \gamma_\tau}, t = 1, ..., j$, and

$$\tilde{x}_{1,j} := \sum_{t=1}^{j} \nu_t x_t.$$
(5.331)

By convexity of $f(\cdot)$ we have $f(x_t) - f(\bar{x}) \leq (x_t - \bar{x})^{\mathsf{T}} g(x_t)$, and hence

$$\sum_{t=1}^{j} \gamma_t (x_t - \bar{x})^{\mathsf{T}} g(x_t) \geq \sum_{t=1}^{j} \gamma_t [f(x_t) - f(\bar{x})]$$

= $\left(\sum_{t=1}^{j} \gamma_t\right) \left[\sum_{t=1}^{j} \nu_t f(x_t) - f(\bar{x})\right]$
$$\geq \left(\sum_{t=1}^{j} \gamma_t\right) [f(\tilde{x}_{1,j}) - f(\bar{x})].$$

Combining this with (5.330) we obtain

$$f(\tilde{x}_{1,j}) - f(\bar{x}) \le \frac{V(x_1, \bar{x}) + \sum_{t=1}^{j} (2\kappa)^{-1} \gamma_t^2 \|G(x_t, \xi^t)\|_*^2 - \sum_{t=1}^{j} \gamma_t \Delta_t^{\mathsf{T}}(x_t - \bar{x})}{\sum_{t=1}^{j} \gamma_t}.$$
 (5.332)

• Assume from now on that the procedure starts with the minimizer of $\mathfrak{d}(\cdot)$, that is

$$x_1 := \operatorname{argmin}_{x \in \mathcal{X}} \mathfrak{d}(x). \tag{5.333}$$

Since by the optimality of x_1 we have that $(u - x_1)^T \nabla \mathfrak{d}(x_1) \ge 0$ for any $u \in \mathcal{X}$, it follows from the definition (5.319) of the function $V(\cdot, \cdot)$ that

$$\max_{u \in \mathcal{X}} V(x_1, u) \le D^2_{\mathfrak{d}, \mathcal{X}},\tag{5.334}$$

1 /0

where

$$D_{\mathfrak{d},\mathcal{X}} := \left[\max_{u \in \mathcal{X}} \mathfrak{d}(u) - \min_{x \in \mathcal{X}} \mathfrak{d}(x) \right]^{1/2}.$$
 (5.335)

Together with (5.332) this implies

$$f(\tilde{x}_{1,j}) - f(\bar{x}) \le \frac{D_{\mathfrak{d},\mathcal{X}}^2 + \sum_{t=1}^j (2\kappa)^{-1} \gamma_t^2 \|G(x_t,\xi^t)\|_*^2 - \sum_{t=1}^j \gamma_t \Delta_t^\mathsf{T}(x_t - \bar{x})}{\sum_{t=1}^j \gamma_t}.$$
 (5.336)

We also have (see (5.325)) that $V(x_1, u) \ge \frac{1}{2}\kappa ||x_1 - u||^2$, and hence it follows from (5.334) that for all $u \in \mathcal{X}$:

$$\|x_1 - u\| \le \sqrt{\frac{2}{\kappa}} D_{\mathfrak{d}, \mathcal{X}}.$$
(5.337)

Let us assume, as in the previous section (see (5.282)), that there is a positive number M_* such that

$$\mathbb{E}\left[\|G(x,\xi)\|_*^2\right] \le M_*^2, \quad \forall x \in \mathcal{X}.$$
(5.338)

Proposition 5.39. Let $x_1 := \operatorname{argmin}_{x \in \mathcal{X}} \mathfrak{d}(x)$ and suppose that condition (5.338) holds. *Then*

$$\mathbb{E}\left[f(\tilde{x}_{1,j}) - f(\bar{x})\right] \le \frac{D^2_{\mathfrak{d},X} + (2\kappa)^{-1}M^2_* \sum_{t=1}^j \gamma_t^2}{\sum_{t=1}^j \gamma_t}.$$
(5.339)

Proof. Taking expectations of both sides of (5.336) and noting that: (i) x_t is a deterministic function of $\xi_{[t-1]} = (\xi^1, ..., \xi^{t-1})$, (ii) conditional on $\xi_{[t-1]}$, the expectation of Δ_t is 0, and (iii) the expectation of $||G(x_t, \xi^t)||_*^2$ does not exceed M_*^2 , we obtain (5.339).

Constant stepsize policy

Assume that the total number of steps N is given in advance and the constant stepsize policy $\gamma_t = \gamma$, t = 1, ..., N, is employed. Then (5.339) becomes

$$\mathbb{E}\left[f(\tilde{x}_{1,j}) - f(\bar{x})\right] \le \frac{D^2_{\mathfrak{d},\mathcal{X}} + (2\kappa)^{-1}M_*^2 N\gamma^2}{N\gamma}.$$
(5.340)

Minimizing the right hand side of (5.340) over $\gamma > 0$ we arrive at the constant stepsize policy

$$\gamma_t = \frac{\sqrt{2\kappa} D_{\mathfrak{d},\mathcal{X}}}{M_* \sqrt{N}}, \quad t = 1, \dots, N,$$
(5.341)

and the associated efficiency estimate

$$\mathbb{E}\left[f(\tilde{x}_{1,N}) - f(\bar{x})\right] \le D_{\mathfrak{d},\mathcal{X}} M_* \sqrt{\frac{2}{\kappa N}}.$$
(5.342)

This can be compared with the respective stepsize (5.307) and efficiency estimate (5.308) for the robust Euclidean SA method. Passing from the stepsizes (5.341) to the stepsizes

$$\gamma_t = \frac{\theta \sqrt{2\kappa} D_{\mathfrak{d}, \mathcal{X}}}{M_* \sqrt{N}}, \quad t = 1, ..., N,$$
(5.343)

with rescaling parameter $\theta > 0$, the efficiency estimate becomes

$$\mathbb{E}\left[f(\tilde{x}_{1,N}) - f(\bar{x})\right] \le \max\left\{\theta, \theta^{-1}\right\} D_{\mathfrak{d},\mathcal{X}} M_* \sqrt{\frac{2}{\kappa N}},\tag{5.344}$$

similar to the Euclidean case. We refer to the SA method based on (5.321),(5.331) and (5.343) as *Mirror Descent SA* algorithm with constant stepsize policy.

Comparing (5.308) to (5.342), we see that for both the Euclidean and the Mirror Descent SA algorithms, the expected inaccuracy, in terms of the objective values of the approximate solutions is $O(N^{-1/2})$. A benefit of the Mirror Descent over the Euclidean algorithm is in potential possibility to reduce the constant factor hidden in $O(\cdot)$ by adjusting the norm $\|\cdot\|$ and the distance generating function $\mathfrak{d}(\cdot)$ to the geometry of the problem.

Example 5.40 Let $\mathcal{X} := \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1, x \ge 0\}$ be the standard simplex. Consider two setups for the Mirror Descent SA. Namely, the *Euclidean setup* where the considered norm $\|\cdot\| := \|\cdot\|_2$ and $\mathfrak{d}(x) := \frac{1}{2}x^{\mathsf{T}}x$, and ℓ_1 -setup where $\|\cdot\| := \|\cdot\|_1$ and $\mathfrak{d}(\cdot)$ is the entropy function (5.317). The Euclidean setup leads to the Euclidean Robust SA which is easily implementable. Note that the Euclidean diameter of \mathcal{X} is $\sqrt{2}$ and hence is independent of n. The corresponding efficiency estimate is

$$\mathbb{E}\left[f(\tilde{x}_{1,N}) - f(\bar{x})\right] \le O(1) \max\left\{\theta, \theta^{-1}\right\} M N^{-1/2},\tag{5.345}$$

with $M^2 = \sup_{x \in \mathcal{X}} \mathbb{E} \left[\|G(x,\xi)\|_2^2 \right].$

The ℓ_1 -setup corresponds to $\mathcal{X}^{\star} = \{x \in \mathcal{X} : x > 0\}, D_{\mathfrak{d},\mathcal{X}} = \sqrt{\ln n},$

$$x_1 := \operatorname*{argmin}_{x \in \mathcal{X}} \mathfrak{d}(x) = n^{-1} (1, ..., 1)^{\mathsf{T}},$$

 $||x||_* = ||x||_{\infty}$ and $\kappa = 1$ (see (5.318) for verification that $\kappa = 1$). The associated Mirror Descent SA is easily implementable. The prox-function here is

$$V(x,z) = \sum_{i=1}^{n} z_i \ln \frac{z_i}{x_i},$$

and the prox-mapping $P_x(y)$ is given by the explicit formula

$$[P_x(y)]_i = \frac{x_i e^{-y_i}}{\sum_{k=1}^n x_k e^{-y_k}}, \ i = 1, ..., n$$

The respective efficiency estimate of the ℓ_1 -setup is

$$\mathbb{E}\left[f(\tilde{x}_{1,N}) - f(\bar{x})\right] \le O(1) \max\left\{\theta, \theta^{-1}\right\} (\ln n)^{1/2} M_* N^{-1/2}, \tag{5.346}$$

with $M_*^2 = \sup_{x \in \mathcal{X}} \mathbb{E}\left[\|G(x,\xi)\|_{\infty}^2 \right]$, provided that the constant stepsizes (5.343) are used.

To compare (5.346) and (5.345), observe that $M_* \leq M$, and the ratio M_*/M can be as small as $n^{-1/2}$. Thus, the efficiency estimate for the ℓ_1 -setup is never much worse than the estimate for the Euclidean setup, and for large n can be *far better* than the latter estimate. That is,

$$\sqrt{\frac{1}{\ln n}} \le \frac{M}{\sqrt{\ln n}M_*} \le \sqrt{\frac{n}{\ln n}},$$

with both the upper and the lower bounds being achievable. Thus, when X is a standard simplex of large dimension, we have strong reasons to prefer the ℓ_1 -setup to the usual Euclidean one.

Comparison with the SAA approach

Similar to (5.312)–(5.314), by using Chebyshev (Markov) inequality, it is possible to derive from (5.344) an estimate of the sample size N which guarantees that $\tilde{x}_{1,N}$ is an ε -optimal solution of the true problem with probability at least $1-\alpha$. It is possible, however, to obtain

much finer bounds on deviation probabilities when imposing more restrictive assumptions on the distribution of $G(x,\xi)$. Specifically, assume that there is constant $M_* > 0$ such that

$$\mathbb{E}\left[\exp\left\{\left\|G(x,\xi)\right\|_{*}^{2}/M_{*}^{2}\right\}\right] \le \exp\{1\}, \quad \forall x \in \mathcal{X}.$$
(5.347)

Note that condition (5.347) is stronger than (5.338). Indeed, if a random variable Y satisfies $\mathbb{E}[\exp\{Y/a\}] \le \exp\{1\}$ for some a > 0, then by Jensen inequality

$$\exp\{\mathbb{E}[Y/a]\} \le \mathbb{E}[\exp\{Y/a\}] \le \exp\{1\},\$$

and therefore $\mathbb{E}[Y] \leq a$. By taking $Y := \|G(x,\xi)\|_*^2$ and $a := M^2$, we obtain that (5.347) implies (5.338). Of course, condition (5.347) holds if $\|G(x,\xi)\|_* \leq M_*$ for all $(x,\xi) \in \mathcal{X} \times \Xi$.

Theorem 5.41. Suppose that condition (5.347) is fulfilled. Then for the constant stepsizes (5.343), the following holds for any $\Theta \ge 0$:

$$\Pr\left\{f(\tilde{x}_{1,N}) - f(\bar{x}) \ge \frac{C(1+\Theta)}{\sqrt{\kappa N}}\right\} \le 4\exp\{-\Theta\},\tag{5.348}$$

where $C := \left(\max\left\{ \theta, \theta^{-1} \right\} + 8\sqrt{3} \right) M_* D_{\mathfrak{d}, \mathcal{X}} / \sqrt{2}.$

Proof. By (5.336) we have

$$f(\tilde{x}_{1,N}) - f(\bar{x}) \le A_1 + A_2, \tag{5.349}$$

where

$$A_1 := \frac{D_{\mathfrak{d},\mathcal{X}}^2 + (2\kappa)^{-1} \sum_{t=1}^N \gamma_t^2 \|G(x_t,\xi^t)\|_*^2}{\sum_{t=1}^N \gamma_t} \text{ and } A_2 := \sum_{t=1}^N \nu_t \Delta_t^\mathsf{T}(\bar{x} - x_t).$$

Consider $Y_t := \gamma_t^2 \|G(x_t, \xi^t)\|_*^2$ and $c_t := M_*^2 \gamma_t^2$. Note that by (5.347),

$$\mathbb{E}\left[\exp\{Y_i/c_i\}\right] \le \exp\{1\}, \ i = 1, ..., N.$$
(5.350)

Since $\exp{\{\cdot\}}$ is a convex function we have

$$\exp\left\{\frac{\sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} c_i}\right\} = \exp\left\{\sum_{i=1}^{N} \frac{c_i(Y_i/c_i)}{\sum_{i=1}^{N} c_i}\right\} \le \sum_{i=1}^{N} \frac{c_i}{\sum_{i=1}^{N} c_i} \exp\{Y_i/c_i\}.$$

By taking expectation of both sides of the above inequality and using (5.350) we obtain

$$\mathbb{E}\left[\exp\left\{\frac{\sum_{i=1}^{N}Y_i}{\sum_{i=1}^{N}c_i}\right\}\right] \le \exp\{1\}.$$

Consequently by Chebyshev's inequality we have for any number Θ :

$$\Pr\left[\exp\left\{\frac{\sum_{i=1}^{N}Y_i}{\sum_{i=1}^{N}c_i}\right\} \ge \exp\{\Theta\}\right] \le \frac{\exp\{1\}}{\exp\{\Theta\}} = \exp\{1-\Theta\},$$

and hence

$$\Pr\left\{\sum_{i=1}^{N} Y_i \ge \Theta \sum_{i=1}^{N} c_i\right\} \le \exp\{1-\Theta\} \le 3\exp\{-\Theta\}.$$
(5.351)

That is, for any Θ :

$$\Pr\left\{\sum_{t=1}^{N} \gamma_t^2 \|G(x_t, \xi^t)\|_*^2 \ge \Theta M_*^2 \sum_{t=1}^{N} \gamma_t^2\right\} \le 3 \exp\left\{-\Theta\right\}.$$
(5.352)

For the constant stepsize policy (5.343), we obtain by (5.352) that

$$\Pr\left\{A_1 \ge \max\{\theta, \theta^{-1}\} \frac{M_* D_{\mathfrak{d}, \mathcal{X}}(1+\Theta)}{\sqrt{2\kappa N}}\right\} \le 3 \exp\left\{-\Theta\right\}.$$
(5.353)

Consider now the random variable A_2 . By (5.337) we have that

$$\|\bar{x} - x_t\| \le \|x_1 - \bar{x}\| + \|x_1 - x_t\| \le 2\sqrt{2}\kappa^{-1/2}D_{\mathfrak{d},\mathcal{X}}$$

and hence

$$\left|\Delta_{t}^{\mathsf{T}}(\bar{x}-x_{t})\right|^{2} \leq \|\Delta_{t}\|_{*}^{2}\|\bar{x}-x_{t}\|^{2} \leq 8\kappa^{-1}D_{\mathfrak{d},\mathcal{X}}^{2}\|\Delta_{t}\|_{*}^{2}$$

We also have that

$$\mathbb{E}\left[(\bar{x} - x_t)^{\mathsf{T}} \Delta_t \big| \xi_{[t-1]}\right] = (\bar{x} - x_t)^{\mathsf{T}} \mathbb{E}\left[\Delta_t \big| \xi_{[t-1]}\right] = 0 \quad \text{w.p.1},$$

and by condition (5.347) that

$$\mathbb{E}\left[\exp\left\{\left\|\Delta_{t}\right\|_{*}^{2}/(4M_{*}^{2})\right\}\left|\xi_{[t-1]}\right] \le \exp\{1\} \text{ w.p.1}$$

Consequently by applying inequality (7.219) of Proposition 7.72 with $\phi_t := \nu_t \Delta_t^{\mathsf{T}}(\bar{x} - x_t)$ and $\sigma_t^2 := 32\kappa^{-1}M_*^2 D_{\mathfrak{d},\mathcal{X}}^2 \nu_t^2$, we obtain for any $\Theta \ge 0$:

$$\Pr\left\{A_2 \ge 4\sqrt{2}\kappa^{-1/2}M_*D_{\mathfrak{d},\mathcal{X}}\Theta\sqrt{\sum_{t=1}^N\nu_t^2}\right\} \le \exp\left\{-\Theta^2/3\right\}.$$
(5.354)

Since for the constant stepsize policy we have that $\nu_t = 1/N$, t = 1, ..., N, by changing variables $\Theta^2/3$ to Θ and noting that $\Theta^{1/2} \le 1 + \Theta$ for any $\Theta \ge 0$, we obtain from (5.354) that for any $\Theta \ge 0$:

$$\Pr\left\{A_2 \ge \frac{8\sqrt{3}M_*D_{\mathfrak{d},\mathcal{X}}(1+\Theta)}{\sqrt{2\kappa N}}\right\} \le \exp\left\{-\Theta\right\}.$$
(5.355)

Finally, (5.348) follows from (5.349), (5.353) and (5.355).

By setting $\varepsilon = \frac{C(1+\Theta)}{\sqrt{\kappa N}}$, we can rewrite the estimate (5.348) in the form²⁰

$$\Pr\left\{f(\tilde{x}_{1,N}) - f(\bar{x}) > \varepsilon\right\} \le 12 \exp\left\{-\varepsilon C^{-1}\sqrt{\kappa N}\right\}.$$
(5.356)

 $^{^{20}}$ The constant 12 in the right hand side of (5.356) comes from the simple estimate $4\exp\{1\}<12.$

For $\varepsilon > 0$ this gives the following estimate of the sample size N which guarantees that $\tilde{x}_{1,N}$ is an ε -optimal solution of the true problem with probability at least $1 - \alpha$:

$$N \ge O(1)\varepsilon^{-2}\kappa^{-1}M_*^2 D_{\mathfrak{d},\mathcal{X}}^2 \ln^2(12/\alpha).$$
(5.357)

This estimate is similar to the respective estimate (5.126) of the sample size for the SAA method. However, as far as complexity of solving the problem numerically is concerned, it could be mentioned that the SAA method requires a solution of the generated optimization problem, while an SA algorithm is based on computing a single subgradient $G(x_j, \xi^j)$ at each iteration point. As a result, for the same sample size N it typically takes considerably less computation time to run an SA algorithm than to solve the corresponding SAA problem.

5.9.4 Accuracy Certificates for Mirror Descent SA Solutions

We discuss now a way to estimate lower and upper bounds for the optimal value of problem (5.1) by employing SA iterates. This will give us an accuracy certificate for obtained solutions. Assume that we run an SA procedure with respective iterates $x_1, ..., x_N$ computed according to formula (5.321). As before, set

$$\nu_t := \frac{\gamma_t}{\sum_{\tau=1}^N \gamma_\tau}, \ t = 1, ..., N, \text{ and } \tilde{x}_{1,N} := \sum_{t=1}^N \nu_t x_t$$

We assume now that the stochastic objective value $F(x,\xi)$, as well as the stochastic subgradient $G(x,\xi)$, are computable at a given point $(x,\xi) \in X \times \Xi$.

Consider

$$f_*^N := \min_{x \in X} f^N(x) \text{ and } f^{*N} := \sum_{t=1}^N \nu_t f(x_t),$$
 (5.358)

where

$$f^{N}(x) := \sum_{t=1}^{N} \nu_{t} \left[f(x_{t}) + g(x_{t})^{\mathsf{T}}(x - x_{t}) \right].$$
(5.359)

Since $\nu_t > 0$ and $\sum_{t=1}^{N} \nu_t = 1$, by convexity of f(x) we have that the function $f^N(x)$ underestimates f(x) everywhere on \mathcal{X} , and hence²¹ $f_*^N \leq \vartheta^*$. Since $\tilde{x}_{1,N} \in \mathcal{X}$ we also have that $\vartheta^* \leq f(\tilde{x}_{1,N})$, and by convexity of f that $f(\tilde{x}_{1,N}) \leq f^{*N}$. It follows that $\vartheta^* \leq f^{*N}$. That is, for any realization of the random process ξ^1 , ..., we have that

$$f_*^N \le \vartheta^* \le f^{*N}. \tag{5.360}$$

It follows, of course, that $\mathbb{E}[f_*^N] \leq \vartheta^* \leq \mathbb{E}[f^{*N}]$ as well.

Along with the "unobservable" bounds f_*^N , f^{*N} , consider their observable (computable) counterparts

$$\underline{f}^{N} := \min_{x \in X} \left\{ \sum_{t=1}^{N} \nu_{t} [F(x_{t}, \xi^{t}) + G(x_{t}, \xi^{t})^{\mathsf{T}}(x - x_{t})] \right\},$$

$$\overline{f}^{N} := \sum_{t=1}^{N} \nu_{t} F(x_{t}, \xi^{t}),$$
(5.361)

²¹Recall that ϑ^* denotes the optimal value of the true problem (5.1).

which will be referred to as *online* bounds. The bound \overline{f}^N can be easily calculated while running the SA procedure. The bound \underline{f}^N involves solving the optimization problem of minimizing a linear in x objective function over set \mathcal{X} . If the set \mathcal{X} is defined by linear constraints, this is a linear programming problem.

Since x_t is a function of $\xi_{[t-1]}$ and ξ^t is independent of $\xi_{[t-1]}$, we have that

$$\mathbb{E}\left[\overline{f}^{N}\right] = \sum_{t=1}^{N} \nu_{t} \mathbb{E}\left\{\mathbb{E}[F(x_{t},\xi^{t})|\xi_{[t-1]}]\right\} = \sum_{t=1}^{N} \nu_{t} \mathbb{E}\left[f(x_{t})\right] = \mathbb{E}[f^{*N}]$$

and

$$\mathbb{E}[\underline{f}^{N}] = \mathbb{E}\left[\mathbb{E}\left\{\min_{x\in\mathcal{X}}\left\{\sum_{t=1}^{N}\nu_{t}[F(x_{t},\xi^{t})+G(x_{t},\xi^{t})^{\mathsf{T}}(x-x_{t})]\right\}|\xi_{[t-1]}\right\}\right] \\ \leq \mathbb{E}\left[\min_{x\in\mathcal{X}}\left\{\mathbb{E}\left[\sum_{t=1}^{N}\nu_{t}[F(x_{t},\xi^{t})+G(x_{t},\xi^{t})^{\mathsf{T}}(x-x_{t})]\right]|\xi_{[t-1]}\right\}\right] \\ = \mathbb{E}\left[\min_{x\in\mathcal{X}}f^{N}(x)\right] = \mathbb{E}[f_{*}^{N}].$$

It follows that

$$\mathbb{E}[\underline{f}^N] \le \vartheta^* \le \mathbb{E}[\overline{f}^N].$$
(5.362)

That is, on average \underline{f}^N and \overline{f}^N give, respectively, a lower and an upper bound for the optimal value ϑ^* of the optimization problem (5.1).

In order to see how good are the bounds \underline{f}^N and \overline{f}^N let us estimate expectations of the corresponding errors. We will need the following result.

Lemma 5.42. Let
$$\zeta_t \in \mathbb{R}^n$$
, $v_1 \in \mathcal{X}^*$ and $v_{t+1} = P_{v_t}(\zeta_t)$, $t = 1, ..., N$. Then

$$\sum_{t=1}^N \zeta_t^\mathsf{T}(v_t - u) \le V(v_1, u) + (2\kappa)^{-1} \sum_{t=1}^N \|\zeta_t\|_*^2, \quad \forall u \in \mathcal{X}.$$
(5.363)

Proof. By the estimate (5.322) of Lemma 5.38 with $x = v_t$ and $y = \zeta_t$ we have that the following inequality holds for any $u \in \mathcal{X}$,

$$V(v_{t+1}, u) \le V(v_t, u) + \zeta_t^{\mathsf{T}}(u - v_t) + (2\kappa)^{-1} \|\zeta_t\|_*^2.$$
(5.364)

Summing this over t = 1, ..., N, we obtain

$$V(v_{N+1}, u) \le V(v_1, u) + \sum_{t=1}^{N} \zeta_t^{\mathsf{T}}(u - v_t) + (2\kappa)^{-1} \sum_{t=1}^{N} \|\zeta_t\|_*^2.$$
(5.365)

Since $V(v_{N+1}, u) \ge 0$, (5.363) follows. \Box

Consider again condition (5.338), that is

$$\mathbb{E}\left[\|G(x,\xi)\|_*^2\right] \le M_*^2, \quad \forall x \in \mathcal{X},\tag{5.366}$$

and the following condition: there is a constant Q > 0 such that

$$\mathbb{V}\mathrm{ar}[F(x,\xi)] \le Q^2, \ \forall x \in \mathcal{X}.$$
(5.367)

Note that, of course, $\mathbb{V}ar[F(x,\xi)] = \mathbb{E}\left[(F(x,\xi) - f(x))^2\right]$.

Theorem 5.43. Suppose that conditions (5.366) and (5.367) hold. Then

$$\mathbb{E}[f^{*N} - f^N_*] \le \frac{2D^2_{\mathfrak{d},\mathcal{X}} + \frac{5}{2}\kappa^{-1}M^2_*\sum_{t=1}^N \gamma^2_t}{\sum_{t=1}^N \gamma_t},$$
(5.368)

$$\mathbb{E}\left[\left|\overline{f}^{N} - f^{*N}\right|\right] \leq Q_{\sqrt{\sum_{t=1}^{N} \nu_{t}^{2}}},$$

$$\mathbb{E}\left[\left|\underline{f}^{N} - f^{N}_{*}\right|\right] \leq \left(Q + 4\sqrt{2}\kappa^{-1/2}M_{*}D_{\mathfrak{d},\mathcal{X}}\right)\sqrt{\sum_{t=1}^{N} \nu_{t}^{2}}$$

$$+ \frac{D_{\mathfrak{d},X}^{2} + 2\kappa^{-1}M_{*}^{2}\sum_{t=1}^{N} \gamma_{t}^{2}}{\sum_{t=1}^{N} \gamma_{t}}.$$
(5.369)
(5.369)

Proof. If in Lemma 5.42 we take $v_1 := x_1$ and $\zeta_t := \gamma_t G(x_t, \xi^t)$, then the corresponding iterates v_t coincide with x_t . Therefore, we have by (5.363) and since $V(x_1, u) \leq D^2_{\mathfrak{d}, \mathcal{X}}$ that

$$\sum_{t=1}^{N} \gamma_t (x_t - u)^{\mathsf{T}} G(x_t, \xi^t) \le D^2_{\mathfrak{d}, X} + (2\kappa)^{-1} \sum_{t=1}^{N} \gamma_t^2 \| G(x_t, \xi^t) \|_*^2, \quad \forall u \in \mathcal{X}.$$
(5.371)

It follows that for any $u \in X$ (compare with (5.330)),

$$\sum_{t=1}^{N} \nu_t \left[-f(x_t) + (x_t - u)^{\mathsf{T}} g(x_t) \right] + \sum_{t=1}^{N} \nu_t f(x_t)$$

$$\leq \frac{D_{\mathfrak{d},\mathcal{X}}^2 + (2\kappa)^{-1} \sum_{t=1}^{N} \gamma_t^2 \|G(x_t,\xi^t)\|_*^2}{\sum_{t=1}^{N} \gamma_t} + \sum_{t=1}^{N} \nu_t \Delta_t^{\mathsf{T}}(x_t - u),$$

where $\Delta_t := G(x_t, \xi^t) - g(x_t)$. Since

$$f^{*N} - f^N_* = \sum_{t=1}^N \nu_t f(x_t) + \max_{u \in X} \sum_{t=1}^N \nu_t \big[-f(x_t) + (x_t - u)^{\mathsf{T}} g(x_t) \big],$$

it follows that

$$f^{*N} - f^{N}_{*} \le \frac{D^{2}_{\mathfrak{d},X} + (2\kappa)^{-1} \sum_{t=1}^{N} \gamma^{2}_{t} \|G(x_{t}, \xi^{t})\|_{*}^{2}}{\sum_{t=1}^{N} \gamma_{t}} + \max_{u \in X} \sum_{t=1}^{N} \nu_{t} \Delta^{\mathsf{T}}_{t}(x_{t} - u).$$
(5.372)

Let us estimate the second term in the right hand side of (5.372). By using Lemma 5.42 with $v_1 := x_1$ and $\zeta_t := \gamma_t \Delta_t$, and the corresponding iterates $v_{t+1} = P_{v_t}(\zeta_t)$, we obtain

$$\sum_{t=1}^{N} \gamma_t \Delta_t^{\mathsf{T}}(v_t - u) \le D_{\mathfrak{d},\mathcal{X}}^2 + (2\kappa)^{-1} \sum_{t=1}^{N} \gamma_t^2 \|\Delta_t\|_*^2, \quad \forall u \in \mathcal{X}.$$
 (5.373)

Moreover,

$$\Delta_t^{\mathsf{T}}(v_t - u) = \Delta_t^{\mathsf{T}}(x_t - u) + \Delta_t^{\mathsf{T}}(v_t - x_t),$$

and hence it follows by (5.373) that

$$\max_{u \in X} \sum_{t=1}^{N} \nu_t \Delta_t^{\mathsf{T}}(x_t - u) \le \sum_{t=1}^{N} \nu_t \Delta_t^{\mathsf{T}}(x_t - v_t) + \frac{D_{\mathfrak{d}, \mathcal{X}}^2 + (2\kappa)^{-1} \sum_{t=1}^{N} \gamma_t^2 \|\Delta_t\|_*^2}{\sum_{t=1}^{N} \gamma_t}.$$
 (5.374)

Moreover, $\mathbb{E}\left[\Delta_t | \xi_{[t-1]}\right] = 0$ and v_t and x_t are functions of $\xi_{[t-1]}$, and hence

$$\mathbb{E}\left[(x_t - v_t)^{\mathsf{T}} \Delta_t\right] = \mathbb{E}\left\{(x_t - v_t)^{\mathsf{T}} \mathbb{E}[\Delta_t | \xi_{[t-1]}]\right\} = 0.$$
(5.375)

In view of condition (5.366) we have that $\mathbb{E}\left[\|\Delta_t\|_*^2\right] \leq 4M_*^2$, and hence it follows from (5.374) and (5.375) that

$$\mathbb{E}\left[\max_{u\in\mathcal{X}}\sum_{t=1}^{N}\nu_{t}\Delta_{t}^{\mathsf{T}}(x_{t}-u)\right] \leq \frac{D_{\mathfrak{d},\mathcal{X}}^{2}+2\kappa^{-1}M_{*}^{2}\sum_{t=1}^{N}\gamma_{t}^{2}}{\sum_{t=1}^{N}\gamma_{t}}.$$
(5.376)

Therefore, by taking expectation of both sides of (5.372) and using (5.366) together with (5.376) we obtain (5.368).

In order to proof (5.369) let us observe that

$$\overline{f}^N - f^{*N} = \sum_{t=1}^N \nu_t (F(x_t, \xi^t) - f(x_t)),$$

and that for $1 \leq s < t \leq N$,

$$\begin{split} & \mathbb{E} \Big[(F(x_s,\xi^s) - f(x_s))(F(x_t,\xi^t) - f(x_t)) \Big] \\ &= \mathbb{E} \Big\{ \mathbb{E} \Big[(F(x_s,\xi^s) - f(x_s))(F(x_t,\xi^t) - f(x_t)) | \xi_{[t-1]} \Big] \Big\} \\ &= \mathbb{E} \Big\{ (F(x_s,\xi_s) - f(x_s)) \mathbb{E} \Big[(F(x_t,\xi^t) - f(x_t)) | \xi_{[t-1]} \Big] \Big\} = 0. \end{split}$$

Therefore

$$\mathbb{E}\left[\left(\overline{f}^{N} - f^{*N}\right)^{2}\right] = \sum_{t=1}^{N} \nu_{t}^{2} \mathbb{E}\left[\left(F(x_{t}, \xi^{t}) - f(x_{t})\right)^{2}\right] \\
= \sum_{t=1}^{N} \nu_{t}^{2} \mathbb{E}\left\{\mathbb{E}\left[\left(F(x_{t}, \xi^{t}) - f(x_{t})\right)^{2} |\xi_{[t-1]}\right]\right\} \quad (5.377) \\
\leq Q^{2} \sum_{t=1}^{N} \nu_{t}^{2},$$

where the last inequality is implied by condition (5.367). Since for any random variable Y we have that $\sqrt{\mathbb{E}[Y^2]} \ge \mathbb{E}[|Y|]$, the inequality (5.369) follows from (5.377).

Let us now look at (5.370). Denote

$$\tilde{f}^{N}(x) := \sum_{t=1}^{N} \nu_{t} [F(x_{t}, \xi^{t}) + G(x_{t}, \xi^{t})^{\mathsf{T}}(x - x_{t})].$$

Then

$$\left|\underline{f}^{N} - f_{*}^{N}\right| = \left|\min_{x \in \mathcal{X}} \tilde{f}^{N}(x) - \min_{x \in \mathcal{X}} f^{N}(x)\right| \le \max_{x \in \mathcal{X}} \left|\tilde{f}^{N}(x) - f^{N}(x)\right|$$

and

$$\tilde{f}^N(x) - f^N(x) = \overline{f}^N - f^{*N} + \sum_{t=1}^N \nu_t \Delta_t^\mathsf{T}(x_t - x),$$

and hence

$$\left|\underline{f}^{N} - f_{*}^{N}\right| \leq \left|\overline{f}^{N} - f^{*N}\right| + \left|\max_{x \in \mathcal{X}} \sum_{t=1}^{N} \nu_{t} \Delta_{t}^{\mathsf{T}}(x_{t} - x)\right|.$$
(5.378)

For $\mathbb{E}\left[\left|\overline{f}^{N} - f^{*N}\right|\right]$ we already have the estimate (5.369). By (5.373) we have

$$\left| \max_{x \in \mathcal{X}} \sum_{t=1}^{N} \nu_t \Delta_t^{\mathsf{T}}(x_t - x) \right| \le \left| \sum_{t=1}^{N} \nu_t \Delta_t^{\mathsf{T}}(x_t - v_t) \right| + \frac{D_{\mathfrak{d},\mathcal{X}}^2 + (2\kappa)^{-1} \sum_{t=1}^{N} \gamma_t^2 \|\Delta_t\|_*^2}{\sum_{t=1}^{N} \gamma_t}.$$
(5.379)

Let us observe that for $1 \le s < t \le N$,

$$\mathbb{E}\left[(\Delta_s^\mathsf{T}(x_s - v_s))(\Delta_t^\mathsf{T}(x_t - v_t)) \right] = \mathbb{E}\left\{ \mathbb{E}\left[(\Delta_s^\mathsf{T}(x_s - v_s))(\Delta_t^\mathsf{T}(x_t - v_t))|\xi_{[t-1]} \right] \right\} \\ = \mathbb{E}\left\{ (\Delta_s^\mathsf{T}(x_s - v_s)) \mathbb{E}\left[(\Delta_t^\mathsf{T}(x_t - v_t))|\xi_{[t-1]} \right] \right\} = 0.$$

Therefore, by condition (5.366) we have

$$\mathbb{E}\left[\left(\sum_{t=1}^{N}\nu_{t}\Delta_{t}^{\mathsf{T}}(x_{t}-v_{t})\right)^{2}\right] = \sum_{t=1}^{N}\nu_{t}^{2}\mathbb{E}\left[\left|\Delta_{t}^{\mathsf{T}}(x_{t}-v_{t})\right|^{2}\right] \\
\leq \sum_{t=1}^{N}\nu_{t}^{2}\mathbb{E}\left[\left\|\Delta_{t}\right\|_{*}^{2}\left\|x_{t}-v_{t}\right\|^{2}\right] = \sum_{t=1}^{N}\nu_{t}^{2}\mathbb{E}\left[\left\|x_{t}-v_{t}\right\|^{2}\mathbb{E}\left[\left\|\Delta_{t}\right\|_{*}^{2}|\xi_{[t-1]}\right] \\
\leq 4M_{*}^{2}\sum_{t=1}^{N}\nu_{t}^{2}\mathbb{E}\left[\left\|x_{t}-v_{t}\right\|^{2}\right] \leq 32\kappa^{-1}M_{*}^{2}D_{\mathfrak{d},\mathcal{X}}^{2}\sum_{t=1}^{N}\nu_{t}^{2},$$

where the last inequality follows by (5.337). It follows that

$$\mathbb{E}\left[\left|\sum_{t=1}^{N}\nu_{t}\Delta_{t}^{\mathsf{T}}(x_{t}-v_{t})\right|\right] \leq 4\sqrt{2}\kappa^{-1/2}M_{*}D_{\mathfrak{d},\mathcal{X}}\sqrt{\sum_{t=1}^{N}\nu_{t}^{2}}.$$
(5.380)

Putting together (5.378),(5.379),(5.380) and (5.369) we obtain (5.370). □

For the constant stepsize policy (5.343), all estimates given in the right hand sides of (5.368),(5.369) and (5.370) are of order $O(N^{-1/2})$. It follows that under the specified conditions, difference between the upper \overline{f}^N and lower \underline{f}^N bounds converge on average to zero, with increase of the sample size N, at a rate of $O(N^{-1/2})$. It is also possible to derive respective large deviations rates of convergence (Lan, Nemirovski and Shapiro [140]).

Remark 22. The lower SA bound \underline{f}^N can be compared with the respective SAA bound $\hat{\vartheta}_N$ obtained by solving the corresponding SAA problem (see section 5.6.1). Suppose that the same sample $\xi^1, ..., \xi^N$ is employed for both the SA and SAA methods, that $F(\cdot, \xi)$ is convex for all $\xi \in \Xi$, and $G(x,\xi) \in \partial_x F(x,\xi)$ for all $(x,\xi) \in \mathcal{X} \times \Xi$. By convexity of $F(\cdot,\xi)$ and definition of f^N , we have

$$\hat{\vartheta}_{N} = \min_{x \in \mathcal{X}} \left\{ N^{-1} \sum_{t=1}^{N} F(x, \xi^{t}) \right\} \\
\geq \min_{x \in \mathcal{X}} \left\{ \sum_{t=1}^{N} \nu_{t} \left[F(x_{t}, \xi^{t}) + G(x_{t}, \xi^{t})^{\mathsf{T}}(x - x_{t}) \right] \right\} = \underline{f}^{N}.$$
(5.381)

Therefore, for the same sample, the SA lower bound \underline{f}^N is weaker than the SAA lower bound $\hat{\vartheta}_N$. However, it should be noted that the SA lower bound can be computed much faster than the respective SAA lower bound.

5.10 Stochastic Dual Dynamic Programming Method

The analysis of section 5.8.2 gives quite a pessimistic view of numerical complexity in terms of number of scenarios required for solving multistage stochastic programming problems. On the other hand the dynamic programming approach suffers from the famous "curse of dimensionality", the term coined by Bellman. One of the main difficulties in applying the dynamic programming is how to represent the cost-to-go functions in a computationally tractable way when the number of state variables is large. In this section we discuss an approach which can be considered as a variant of approximate dynamic programming.

In the subsequent analysis we distinguish between cutting and supporting planes (hyperplanes) of a convex function $Q : \mathbb{R}^n \to \mathbb{R}$. We say that an affine function $\ell(x) = \alpha + \beta^T x$ is a *cutting plane* of Q(x), if $Q(x) \ge \ell(x)$ for all $x \in \mathbb{R}^n$. Note that cutting plane $\ell(x)$ can be strictly smaller than Q(x) for all $x \in \mathbb{R}^n$. If, moreover, $Q(\bar{x}) = \ell(\bar{x})$ for some $\bar{x} \in \mathbb{R}^n$, it is said that $\ell(x)$ is a *supporting plane* of Q(x) at $x = \bar{x}$. This supporting plane is given by $\ell(x) = Q(\bar{x}) + g^T(x - \bar{x})$ for some subgradient $g \in \partial Q(\bar{x})$.

Consider multistage stochastic programs of the form discussed in section 3.2, which in the nested form can be written as

$$\underset{\substack{A_1x_1=b_1\\x_1\in X_1}}{\operatorname{Min}} f_1(x_1) + \mathbb{E} \Big[\min_{\substack{B_2x_1+A_2x_2=b_2\\x_2\in X_2}} f_2(x_2,\xi_2) + \dots + \mathbb{E} \Big[\min_{\substack{B_Tx_T-1+A_Tx_T=b_T\\x_T\in X_T}} f_T(x_T,\xi_T) \Big] \Big].$$
(5.382)

Here $f_t : \mathbb{R}^{n_t} \times \mathbb{R}^{d_t} \to \mathbb{R}$ and $X_t \subset \mathbb{R}^{n_t}$. The data vectors $\xi_t \in \mathbb{R}^{d_t}$ can include components of matrices B_t, A_t and vectors b_t and some additional parameters. Unless stated otherwise we make the following assumptions throughout this section.

- (i) The functions $f_t(x_t, \xi_t), t = 1, ..., T$, are random lower semicontinuous and $f_t(\cdot, \xi_t)$ are *convex* for a.e. ξ_t .
- (ii) The sets $X_t \subset \mathbb{R}^{n_t}$, t = 1, ..., T, are closed and convex.
- (iii) The random process ξ_t , t = 1, ..., T, is stagewise independent.
- (iv) The problem has relatively complete recourse.

Recall that in some cases with stagewise dependent data it is possible to reformulate the problem to make it stagewise independent by increasing the number of decision variables (see Remark 5 on page 67).

5.10.1 Approximate Dynamic Programming Approach

The dynamic programming equations for problem (5.382) can be written (compare with equations (3.52)) as

$$Q_t(x_{t-1},\xi_t) = \inf_{x_t \in X_t} \left\{ f_t(x_t,\xi_t) + \mathcal{Q}_{t+1}(x_t) : B_t x_{t-1} + A_t x_t = b_t \right\},$$
(5.383)

where

$$Q_{t+1}(x_t) := \mathbb{E}\left[Q_{t+1}(x_t, \xi_{t+1})\right], \tag{5.384}$$

t = 2, ..., T, and $Q_{T+1}(\cdot) \equiv 0$. At the first stage the following problem should be solved

$$\min_{x_1 \in X_1} f_1(x_1) + \mathcal{Q}_2(x_1) \text{ s.t. } A_1 x_1 = b_1.$$
(5.385)

Note that because of the stagewise independence assumption (ii), the cost-to-go functions $Q_{t+1}(x_t)$ do not depend on the data process (see Remark 4 on page 67).

The basic idea is to approximate the (convex) cost-to-go functions $Q_t(\cdot, \xi_t)$ and $Q_t(\cdot)$ by maxima of finite families of linear functions. First we need to discretize the data process ξ_t . We approach this by constructing a scenario tree employing the *identical* conditional sampling method. That is, a sample $\xi_t^1, ..., \xi_t^{N_t}$ of random vector $\xi_t, t = 2, ..., T$, is generated and the corresponding scenario tree is constructed by connecting every ancestor node at stage t - 1 with the same set of children nodes $\xi_t^1, ..., \xi_t^{N_t}$ (see the discussion at the beginning of section 5.8). Recall that the identical conditional sampling preserves the stagewise independence in the constructed SAA problem. The total number of scenarios of the obtained SAA problem is $\mathcal{N} = \prod_{t=2}^{T} N_t$. In order for the optimal value $\hat{\vartheta}_{\mathcal{N}}$ of the SAA problem to converge to the optimal value ϑ^* of the true problem, all sample sizes N_t should increase. Therefore the total number of scenarios \mathcal{N} quickly becomes astronomically large with increase of the number of stages T even for moderate values of sample sizes N_t . In such cases it will be hopeless to solve the SAA problem by approximating the dynamic programming equations.

For the SAA problem the corresponding dynamic programming equations take the form

$$Q_t^j(x_{t-1}) = \inf_{x_t \in X_t} \left\{ f_t(x_t, \xi_t^j) + \mathcal{Q}_{t+1}(x_t) : B_t^j x_{t-1} + A_t^j x_t = b_t^j \right\},$$
(5.386)

 $j = 1, ..., N_t$, with

$$Q_{t+1}(x_t) = \frac{1}{N_{t+1}} \sum_{j=1}^{N_{t+1}} Q_{t+1}^j(x_t).$$
(5.387)

At the last stage the cost-to-go functions $Q_T^j(x_{T-1})$, $j = 1, ..., N_T$, are given by the optimal value of the respective problems

$$\underset{x_T \in X_T}{\min} f_T(x_T, \xi_T^j) \text{ s.t. } B_T^j x_{T-1} + A_T^j x_T = b_T^j.$$
(5.388)

Under mild regularity conditions there is no duality gap between problem (5.388) and its dual, and a subgradient of $Q_T^j(\cdot)$ at a point $\bar{x}_{T-1} \in \mathbb{R}^{n_{T-1}}$ is given by $g_T^j = -B_T^j \pi_T^j$, where π_T^j is an optimal solution of the dual of problem (5.388) for $x_{T-1} = \bar{x}_{T-1}$ (see Propositions 2.21 and 2.22). In particular, this holds if the function $f_T(\cdot, \xi_T^j)$ and the set X_T are polyhedral and problem (5.388) has a finite optimal value (see Proposition 2.14). In that case problem (5.388) can be formulated as a linear programming problem.

By solving the dual of problem (5.388), and hence computing the subgradient g_T^j , for every $j = 1, ..., N_T$, we can compute the corresponding subgradient $\mathfrak{g}_T = \frac{1}{N_T} \sum_{j=1}^{N_T} g_T^j$ of

the cost-to-go function $Q_T(\cdot)$ at the point \bar{x}_{T-1} . This gives us the corresponding supporting plane

$$\ell_T(x_{T-1}) := \mathcal{Q}_T(\bar{x}_{T-1}) + \mathfrak{g}_T^{\mathsf{I}}(x_{T-1} - \bar{x}_{T-1})$$

to $\mathcal{Q}_T(\cdot)$ at \bar{x}_{T-1} . Computing several such supporting planes, at different points of $\mathbb{R}^{n_{T-1}}$, we obtain a piecewise linear approximation of the cost-to-go function $\mathcal{Q}_T(\cdot)$ by taking maximum of these supporting planes. We denote this piecewise linear approximation function by $\mathfrak{Q}_T(x_{T-1})$. Note that by the construction $\mathcal{Q}_T(x_{T-1}) \geq \mathfrak{Q}_T(x_{T-1})$ for all $x_{T-1} \in \mathbb{R}^{n_{T-1}}$.

At stage t = T - 1 the cost-to-go functions $Q_{T-1}^{j}(x_{T-2}), j = 1, ..., N_{T-1}$, are given by the optimal value of the respective problems

$$\begin{array}{ll}
& \underset{x_{T-1} \in X_{T-1}}{\operatorname{Min}} & f_{T-1}(x_{T-1}, \xi_{T-1}^{j}) + \mathcal{Q}_{T}(x_{T-1}) \\
& \text{s.t.} & B_{T-1}^{j} x_{T-2} + A_{T-1}^{j} x_{T-1} = b_{T-1}^{j}.
\end{array}$$
(5.389)

The cost-to-go function $Q_T(\cdot)$ is not known in a closed form. So we replace it by the computed approximation $\mathfrak{Q}_T(\cdot)$, and hence approximate problem (5.389) by the problem

$$\begin{array}{ll}
& \underset{x_{T-1} \in X_{T-1}}{\operatorname{Min}} & f_{T-1}(x_{T-1}, \xi_{T-1}^{j}) + \mathfrak{Q}_{T}(x_{T-1}) \\
& \text{s.t.} & B_{T-1}^{j} x_{T-2} + A_{T-1}^{j} x_{T-1} = b_{T-1}^{j}.
\end{array}$$
(5.390)

Note that by the construction the function $\mathfrak{Q}_T(\cdot)$ is polyhedral, and hence if the function $f_{T-1}(\cdot,\xi_{T-1}^j)$ and the set X_{T-1} are also polyhedral, then the objective function of problem (5.390) is polyhedral and this problem can be formulated as a linear programming problem.

Denote by $\underline{Q}_{T-1}^{j}(x_{T-2})$ the optimal value of problem (5.390) and

$$\underline{\mathcal{Q}}_{T-1}(x_{T-2}) = \frac{1}{N_{T-1}} \sum_{j=1}^{N_{T-1}} \underline{Q}_{T-1}^j(x_{T-2}).$$

Since $Q_T(x_{T-1}) \ge \mathfrak{Q}_T(x_{T-1})$, it follows that

$$Q_{T-1}^j(x_{T-2}) \ge \underline{Q}_{T-1}^j(x_{T-2}) \text{ and } \mathcal{Q}_{T-1}(x_{T-2}) \ge \underline{\mathcal{Q}}_{T-1}(x_{T-2})$$
(5.391)

for all $x_{T-2} \in \mathbb{R}^{n_{T-2}}$. By computing a solution π_{T-1}^j of the dual of problem (5.390), for a given value $x_{T-2} = \bar{x}_{T-2}$, we obtain a subgradient $g_{T-1}^j = -B_{T-1}^j \pi_{T-1}^j$ of $\underline{Q}_{T-1}^j(\cdot)$ at $\bar{x}_{T-2} \in \mathbb{R}^{n_{T-2}}$.

By computing such subgradients for all $j = 1, ..., N_{T-1}$, we obtain the corresponding subgradient $\mathfrak{g}_{T-1} = \frac{1}{N_{T-1}} \sum_{j=1}^{N_{T-1}} g_{T-1}^j$ and the supporting plane

$$\ell_{T-1}(x_{T-2}) := \underline{\mathcal{Q}}_{T-1}(\bar{x}_{T-2}) + \mathfrak{g}_{T-1}^{\mathsf{T}}(x_{T-2} - \bar{x}_{T-2})$$

of $\underline{\mathcal{Q}}_{T-1}(\cdot)$ at \bar{x}_{T-2} . By (5.391) we have that $\ell_{T-1}(\cdot)$ is a cutting plane for $\mathcal{Q}_{T-1}(\cdot)$ in the sense that $\mathcal{Q}_{T-1}(\cdot) \geq \ell_{T-1}(\cdot)$. Note, however, that $\mathcal{Q}_{T-1}(\bar{x}_{T-2})$ can be strictly bigger than $\ell_{T-1}(\bar{x}_{T-2})$. That is, $\ell_{T-1}(\cdot)$ is a supporting plane of $\underline{\mathcal{Q}}_{T-1}(\cdot)$, but could be only a cutting plane of $\mathcal{Q}_{T-1}(\cdot)$. By computing such supporting planes of $\underline{\mathcal{Q}}_{T-1}(\cdot)$, at different points of $\mathbb{R}^{n_{T-2}}$, and taking their maximum we obtain an approximation, denoted $\mathfrak{Q}_{T-1}(\cdot)$, of the function $\underline{\mathcal{Q}}_{T-1}(\cdot)$ and hence of $\mathcal{Q}_{T-1}(\cdot)$. By the construction we have that the function $\mathfrak{Q}_{T-1}(\cdot)$ is polyhedral and $\mathcal{Q}_{T-1}(\cdot) \ge \mathfrak{Q}_{T-1}(\cdot)$. Continuing this backward in time we construct the respective (approximate) dynamic programming problems

$$\min_{x_t \in X_t} f_t(x_t, \xi_t^j) + \mathfrak{Q}_{t+1}(x_t) \text{ s.t. } B_t^j x_{t-1} + A_t^j x_t = b_t^j,$$
(5.392)

 $j = 1, ..., N_t$, and eventually the approximation

$$\min_{x_1} f_1(x_1) + \mathfrak{Q}_2(x_1) \text{ s.t. } A_1 x_1 = b_1$$
(5.393)

of the first stage problem.

The computed approximations $\mathfrak{Q}_2(\cdot), ..., \mathfrak{Q}_T(\cdot)$, with $\mathfrak{Q}_{T+1}(\cdot) \equiv 0$ by definition, and a feasible first stage solution \bar{x}_1 , define a feasible policy. That is, for a given realization $\xi_2, ..., \xi_T$ of the data process, going forward in time, decisions $\bar{x}_t = \bar{x}_t(\bar{x}_{t-1}, \xi_t), t = 2, ..., T$, are computed recursively as a solution of the approximate dynamic programming problem

$$\min_{x_t \in X_t} f_t(x_t, \xi_t) + \mathfrak{Q}_{t+1}(x_t) \text{ s.t. } B_t x_{t-1} + A_t x_t = b_t,$$
(5.394)

for $x_{t-1} = \bar{x}_{t-1}$. Of course, this procedure can be performed only if problems (5.394) have optimal solutions. The problem (5.394) has an optimal solution if its feasible set is nonempty and bounded. Note that the feasible set of problem (5.394) is given by the intersection of the set X_t with the set of solutions of the equation $B_t x_{t-1} + A_t x_t = b_t$, and is the same as the feasible set of the respective SAA problem.

5.10.2 The SDDP algorithm

In order to implement the approximate dynamic programming approach, outlined in the above section 5.10.1, we face two questions. Namely, how to generate (trial) points where the corresponding cutting planes are computed, and how to estimate value of the constructed policy. The following procedure was suggested in Pereira and Pinto [178], and is known as the *Stochastic Dual Dynamic Programming* (SDDP) Method. The SDDP algorithm, applied to the SAA problem, consists of backward and forward steps. In the backward step the approximate cost-to-go functions $\mathfrak{Q}_2(\cdot), ..., \mathfrak{Q}_T(\cdot)$ and the first stage solution \bar{x}_1 are constructed in the way described in the previous section 5.10.1. These approximations are updated at successive iterations of the algorithm using trial points computed at the respective forward steps of the algorithm.

At a current iteration of the algorithm, given approximate functions $\mathfrak{Q}_2(\cdot), ..., \mathfrak{Q}_T(\cdot)$ and a feasible first stage solution \bar{x}_1 , the *forward* step of the SDDP algorithm proceeds in generating M random realizations (scenarios) $\xi_{2i}, ..., \xi_{Ti}, i = 1, ..., M$, from the scenario tree of the SAA problem. Recall that the total number of scenarios $\mathcal{N} = \prod_{t=2}^T N_t$ of the SAA problem is very large, so for moderate values of M probability of generating same scenarios is practically negligible. By computing solutions \bar{x}_{ti} of respective problems (5.394) for $x_{t-1} = \bar{x}_{t-1,i}$ and $\xi_t = \xi_{ti}, t = 2, ..., T$, the corresponding value

$$\vartheta_i := f_1(\bar{x}_1) + \sum_{t=2}^T f_t(\bar{x}_{ti}, \xi_{ti}), \ i = 1, ..., M,$$

of the current policy, for the realization $\xi_{2i}, ..., \xi_{Ti}$ of the data process, is calculated. As such, ϑ_i is an unbiased estimate of expected value of that policy, i.e.,

$$\mathbb{E}[\vartheta_i] = f_1(\bar{x}_1) + \mathbb{E}\left[\sum_{t=2}^T f_t(\bar{x}_{ti}, \xi_{ti})\right].$$

The forward step has two objectives. First, some (all) of computed solutions \bar{x}_{ti} , i = 1, ..., M, can be used as trial points in the next iteration of the backward step of the algorithm. Second, these solutions can be employed for constructing a statistical upper bound for the optimal value of the corresponding multistage program. Consider the average (sample mean) $\bar{\vartheta}_M := M^{-1} \sum_{i=1}^M \vartheta_i$ and the standard error

$$\hat{\sigma}_M := \sqrt{\frac{1}{M-1} \sum_{i=1}^M (\vartheta_i - \bar{\vartheta}_M)^2}$$
(5.395)

of the computed values ϑ_i . Since ϑ_i is an unbiased estimate of the expected value of the constructed policy, we have that $\bar{\vartheta}_M$ is also an unbiased estimate of the expected value of that policy. By invoking the Central Limit Theorem we can say that ϑ_M has an approximately normal distribution provided that M is reasonably large. This leads to the following (approximate) $(1 - \alpha)$ -confidence upper bound for the value of that policy

$$u_{\alpha,M} := \bar{\vartheta}_M + z_\alpha \frac{\hat{\sigma}_M}{\sqrt{M}}.$$
(5.396)

Here $1 - \alpha \in (0, 1)$ is a chosen confidence level and $z_{\alpha} = \Phi^{-1}(1 - \alpha)$, where $\Phi(\cdot)$ is the cdf of standard normal distribution. For example, for $\alpha=0.05$ the corresponding critical value $z_{0.05} = 1.64$. That is, with probability approximately $1 - \alpha$ the expected value of the constructed policy is less than the upper bound $u_{\alpha,M}$. Since the expected value of the constructed (feasible) policy is bigger than or equal to the optimal value of the considered multistage problem, we have that $u_{\alpha,M}$ also gives an upper bound for the optimal value of the multistage problem with confidence at least $1 - \alpha$.

Since $\mathfrak{Q}_t(\cdot)$ is the maximum of cutting planes of the cost-to-go function $\mathcal{Q}_t(\cdot)$ we have that $\mathcal{Q}_t(\cdot) \geq \mathfrak{Q}_t(\cdot), t = 2, ..., T$. Therefore the optimal value of the approximate problem computed at a backward step of the algorithm (i.e., the optimal value of problem (5.393)), gives a lower bound for the optimal value ϑ_N of the SAA problem. This lower bound is deterministic (i.e., is not based on sampling) when applied to the considered SAA problem. Since at the backward step of every iteration of the algorithm more cutting planes are added, while no cutting planes are discarded, values of the approximate functions $\mathfrak{Q}_t(\cdot)$ are monotonically nondecreasing from one iteration to the next. As a consequence the computed lower bounds are monotonically nondecreasing with increase of the number of iterations. On the other hand, the upper bound $u_{\alpha,M}$ is a function of generated scenarios and thus is stochastic even for fixed SAA problem. This upper bound may vary for different sets of random samples, in particular from one iteration to the next of the forward step of the algorithm.

Remark 23. Note that the scenarios $\xi_{2i}, ..., \xi_{Ti}, i = 1, ..., M$, can be also sampled from the original data process. Consequently the upper bound $u_{\alpha,M}$ can be used for the original



Figure 5.1. Typical behavior of the lower and upper bounds. The considered SAA problem has 8 state variables, T = 120 stages, $N_t = 100$ sample sizes per stage and $\mathcal{N} = 100^{119}$ total number of scenarios (taken from [252, section 6.2]).

(true) or the SAA problem depending on from what distribution the sampled scenarios were generated.

As far as the true problem is concerned, recall that $\vartheta^* \ge \mathbb{E}[\hat{\vartheta}_{\mathcal{N}}]$ (see (5.245)). Therefore on average the lower bound of the SAA problem is also a lower bound for the optimal value of the true problem.

Cutting plane elimination procedure

With increase of the number of iterations, as more cuts are added at the backward step of the SDDP algorithm, some of these cuts become redundant. That is, one of the cutting planes could become smaller than maximum of the other cutting planes. Clearly such cutting plane can be removed without altering the corresponding lower bound.

Let $\ell_i(x) = \alpha_i + \beta_i^T x$, i = 1, ..., m, be a collection of cutting planes of the cost-to-go function at some stage of the problem. Then, say, ℓ_1 cutting plane is redundant if

$$\ell_1(x) < \max_{2 \le i \le m} \ell_i(x), \ \forall x \in X,$$

where X is the corresponding feasible set. This means that there does not exist $x \in X$ such that $\ell_1(x) \ge \ell_i(x), i = 2, ..., m$. That is, the following system is infeasible

$$\begin{aligned} \alpha_1 + \beta_1^\mathsf{T} x &\geq \alpha_i + \beta_i^\mathsf{T} x, \ i = 2, ..., m, \\ x \in X. \end{aligned}$$
 (5.397)
Checking redundancy of the cutting plane ℓ_1 is reduced to verification of infeasibility of the system (5.397). When the set X is defined by a finite number of linear constraints, this can be formulated as a linear programming problem.

Markovian Data Process

So far we assumed that the data process is stagewise independent. Suppose now that the data process $\xi_2, ..., \xi_T$ is Markovian. That is, the conditional distribution of ξ_{t+1} , given $\xi_{[t]} = (\xi_1, ..., \xi_t)$, does not depend on $(\xi_1, ..., \xi_{t-1})$, i.e., is the same as the conditional distribution of ξ_{t+1} , given $\xi_t, t = 1, ..., T - 1$. Then the corresponding dynamic programming equations take the form (compare with (5.383)–(5.384))

$$Q_t(x_{t-1},\xi_t) = \inf_{x_t \in X_t} \left\{ f_t(x_t,\xi_t) + Q_{t+1}(x_t,\xi_t) : B_t x_{t-1} + A_t x_t = b_t \right\}, (5.398)$$
$$Q_{t+1}(x_t,\xi_t) = \mathbb{E} \left[Q_{t+1}(x_t,\xi_{t+1}) \left| \xi_t \right| \right].$$
(5.399)

Here the cost-to-go functions $Q_t(x_{t-1},\xi_t)$ and $Q_{t+1}(x_t,\xi_t)$ depend on ξ_t , but not on $(\xi_1,...,\xi_{t-1})$.

Suppose, further, that the process $\xi_2, ..., \xi_T$ can be reasonably approximated by a discretization so that it becomes a (possibly nonhomogeneous) *Markov chain*. That is, at stage t = 2, ..., T, the data process can take values $\xi_t^1, ..., \xi_t^{K_t}$, with specified probabilities of going from state $\xi_t^{j_t}$, at stage t, to state $\xi_{t+1}^{j_{t+1}}$, at stage t + 1, $j_t = 1, ..., K_t$, $j_{t+1} = 1, ..., K_{t+1}$. If the numbers K_t are reasonably small, then we still can proceed with the SDDP method by constructing cuts, and hence piecewise linear approximations, of the cost-to-go functions $Q_{t+1}(\cdot, \xi_t^{j_t}), j_t = 1, ..., K_t$. Note that in the backward step of the algorithm the cuts for $Q_{t+1}(x_t, \xi_t^{j_t})$, with respect to x_t , should be constructed *separately* for every $j_t = 1, ..., K_t$. Therefore computational complexity of backward steps of the SDDP algorithm grows more or less linearly with increase of the numbers K_t .

If a sample path is generated from the original (may be continuous) distribution of the data process (rather than its Markov chain discretization), then the computed cuts cannot be used for this sample path unless it coincides with one of the sample paths of the Markov chain. That is, in this approach the SDDP algorithm does not generate a policy for the corresponding original (true) problem, at least not in a direct way. In order to have a policy for the true problem we can proceed as follows. For a sample path of the true problem choose in some sense close to it sample path of the Markov chain. Consequently compute the policy values iteratively going forward and using the computed cuts of the cost-to-go functions $Q_{t+1}(\cdot, \xi_t^{j_t})$ associated with the corresponding sample path of the Markov chain.

5.10.3 Convergence Properties of the SDDP Algorithm

Let us first consider the SDDP algorithm applied to the two stage linear stochastic programming problem (2.1)–(2.2). Let us assume that: (i) the feasible set $\mathcal{X} := \{x : Ax = b, x \ge 0\}$, of the first stage problem, is nonempty and bounded, (ii) the matrix W and vector q are deterministic (not random), and the feasible set $\{\pi : W^T \pi \le q\}$ of the dual (2.3), of the second stage problem, is nonempty and bounded. Consider SAA problem based on a sample $\xi^1, ..., \xi^N$ of the random data vector ξ . Under the above assumption (ii), the corresponding sample average function $Q(x) = N^{-1} \sum_{j=1}^{N} Q(x, \xi^j)$ is convex, finite valued and piecewise linear. The backward step of the SDDP algorithm, applied to the SAA problem, becomes the classical Kelley's cutting plane algorithm, [125]. That is, let at *k*-th iteration, $\mathfrak{Q}^k(\cdot)$ be the corresponding approximation of $Q(\cdot)$ given by maximum of supporting planes of $Q(\cdot)$. At the next iteration the backward step solves the problem

$$\min_{x \in \mathcal{X}} c^{\mathsf{T}} x + \mathfrak{Q}^{k}(x), \tag{5.400}$$

and hence compute its optimal value v^{k+1} , an optimal solution x^{k+1} of (5.400) and a subgradient $g^{k+1} \in \partial \mathcal{Q}(x^{k+1})$. Consequently the supporting plane

$$\ell(x) := \mathcal{Q}(x^{k+1}) + (g^{k+1})^{\mathsf{T}}(x - x^{k+1})$$

is added to the collection of cutting (supporting) planes of $Q(\cdot)$. Note that for the two stage SAA problem, the backward step of the SDDP algorithm does not involve any sampling and the forward step of the algorithm is redundant.

Since the function $Q(\cdot)$ is piecewise linear, it is not difficult to show that Kelley's algorithm converges in a finite number of iterations. Arguments of that type are based on the observation that $Q(\cdot)$ is the maximum of a finite number of its supporting planes. However, the number of supporting planes can be very large and these arguments do not give an idea about rate of convergence of the algorithm. So let us look at the following proof of convergence which uses only convexity of the function $Q(\cdot)$ (cf., [222, p.160]).

Denote $f(x) := c^{\mathsf{T}} x + \mathcal{Q}(x)$ and by f^* the optimal value of the first stage problem. Let x^k , k = 1, ..., be a sequence of iteration points generated by the algorithm, $\gamma^k = c + g^k$ be the corresponding subgradients used in construction of the supporting planes and v^k be the optimal value of the problem (5.400). Note that since $x^k \in \mathcal{X}$ we have that $f(x^k) \ge f^*$, and since $\mathfrak{Q}^k(\cdot) \le \mathcal{Q}(\cdot)$ we have that $v^k \le f^*$ for all k. Note also that since $\mathcal{Q}(\cdot)$ is piecewise linear and \mathcal{X} is bounded, the subgradients of $f(\cdot)$ are bounded on \mathcal{X} , i.e., there is a constant C such that $\|\gamma\| \le C$ for all $\gamma \in \partial f(x)$ and $x \in \mathcal{X}$ (such boundedness of gradients holds for a general convex function provided the function is finite valued on a neighborhood of the set \mathcal{X}). Choose the precision level $\varepsilon > 0$ and denote

$$\mathfrak{K}_{\varepsilon} := \{k : f(x^k) - f^* \ge \varepsilon\}.$$

For k < k' we have

$$f(x^k) + (\gamma^k)^{\mathsf{T}} (x^{k'} - x^k) \le v^{k'} \le f^*,$$

and hence

$$f(x^{k}) - f^{*} \leq (\gamma^{k})^{\mathsf{T}} (x^{k} - x^{k'}) \leq \|\gamma^{k}\| \|x^{k} - x^{k'}\| \leq C \|x^{k} - x^{k'}\|.$$

This implies that for any $k, k' \in \mathfrak{K}_{\varepsilon}$ the following inequality holds

$$\|x^k - x^{k'}\| \ge \varepsilon/(2C). \tag{5.401}$$

Since the set \mathcal{X} is bounded, it follows that the set $\mathfrak{K}_{\varepsilon}$ is finite. That is, after a finite number of iterations, $f(x^k) - f^*$ becomes less than ε , i.e., the algorithm reaches the precision ε in a finite number of iterations.

For $\eta > 0$ denote by $\mathfrak{N}(\mathcal{X}, \eta)$ the maximal number of points in the set \mathcal{X} such that the distance between any two of these points is not less than η . The inequality (5.401) implies that $\mathfrak{N}(\mathcal{X}, \eta)$, with $\eta = \varepsilon/(2C)$, gives an upper bound for the number of iterations required to obtain an ε -optimal solution of the problem by Kelley's algorithm. Unfortunately, for a given η and $\mathcal{X} \subset \mathbb{R}^n$ say being a ball of fixed diameter, the number $\mathfrak{N}(\mathcal{X}, \eta)$, although is finite, grows exponentially with increase of the dimension n. Worst case analysis of Kelley's algorithm is discussed in [165, pp. 158-160], with the following example of a convex problem

$$\min_{x \in \mathbb{R}^{n+1}} f(x) \text{ s.t. } \|x\| \le 1,$$
(5.402)

where $f(x) := \max \{x_1^2 + ... + x_n^2, |x_{n+1}|\}$. It is shown there that Kelley's algorithm applied to problem (5.402) with starting point $x_0 := (0, ..., 0, 1)$, requires at least

$$\frac{\ln(\varepsilon^{-1})}{2\ln 2} \left(\frac{2}{\sqrt{3}}\right)^{n-1} \tag{5.403}$$

calls of the oracle to obtain an ε -optimal solution, i.e., the number of required iterations grows exponentially with increase of the dimension n of the problem. It was also observed empirically that Kelley's algorithm could behave quite poorly in practice.

The above analysis suggests quite a pessimistic view on Kelley's algorithm for convex problems of large dimension *n*. Unfortunately it is not clear how more efficient, bundle type algorithms, can be extended to a multistage setting. On the other hand, from the number-of-scenarios point of view complexity of multistage SAA problems grows very fast with increase of the number of stages even for problems with relatively small dimensions of the involved decision variables. It is possible to extend the above analysis to the multistage setting to show, under regularity conditions of boundedness and relatively complete recourse, that the SDDP algorithm applied to the SAA problem converges (w.p.1) with increase of the number of iterations (cf., [87]). However, the convergence can be very slow and it should be remembered that such mathematical proofs of convergence do not give guarantees that the algorithm will produce an accurate solution in a reasonable computational time. The analysis of two stage problems indicates that the SDDP method could give reasonable results for problems with a not too large number of decision (state) variables; this seems to be confirmed by numerical experiments. Of course, these considerations are problem dependent.

5.10.4 Risk Averse SDDP Method

In formulation (5.382) the expected value of the total cost (objective value) is minimized subject to the feasibility constraints. That is, the total cost is optimized (minimized) on average. Since the costs are functions of the random data process, they are random and hence are subject to random perturbations. For a particular realization of the data process these costs could be much bigger than their average (i.e., expectation) values. We will refer to the formulation (5.382) as *risk neutral* as opposed to *risk averse* approaches which we will discuss below. The goal of a risk averse approach is to avoid large values of the costs for some possible realizations of the data process at *every* stage of the considered time horizon. In the following Chapter 6 we will discuss a general theory of risk measures. The

reader is advised to read relevant parts of Chapter 6 before reading the remainder of this section.

For chosen law invariant risk measures ρ_t and their conditional analogues $\rho_{t|\xi_{[t-1]}}$, and constants γ_t , consider the following extension of the multistage problem (5.382), in the nested form, with added risk averse constraints (see section 6.8.4)

$$\underset{x_{1} \in X_{1}}{\underset{x_{1} \in X_{1}}{\underset{x_{1} \in X_{1}}{\underset{x_{1} \in X_{1}}{\underset{x_{2} \in X_{2}}{\underset{x_{2} \xrightarrow X_{2}}{\underset{x_{2}}{\underset{x_{2}}}{\underset{x_{2} \xrightarrow X_{2}}{\underset$$

A natural choice of the risk measures is

$$\varrho_t(\cdot) := \mathsf{V}@\mathsf{R}_{\alpha_t}(\cdot), \ \alpha_t \in (0,1).$$
(5.405)

That is, at stage t = 2, ..., T, the additional constraint (conditional on $\xi_{[t-1]}$)

$$\mathsf{V}@\mathsf{R}_{\alpha_t|\xi_{[t-1]}}(f_t(x_t,\xi_t)) \le \gamma_t, \tag{5.406}$$

is added to the optimization problem. The constraints (5.406) can be interpreted as (conditional) chance constraints

$$\Pr_{|\xi_{[t-1]}}(f_t(x_t,\xi_t) \le \gamma_t) \ge 1 - \alpha_t.$$
(5.407)

Note that in (5.404)–(5.407), $x_t = x_t(\xi_{[t]}), t = 2, ..., T$, are a functions of the data process.

By using risk measures of the form (5.405) we try to enforce the constraints that (conditional) probabilities of costs $f_t(x_t, \xi_t)$ being bigger than the specified upper bounds γ_t will be less than the chosen significance level α_t . Since risk measure V@R_{α} is not convex, in order to make the corresponding optimization problem computationally tractable, it is natural to replace V@R_{α} by AV@R_{α}. That is, let

$$\varrho_t(\cdot) := \mathsf{AV}@\mathsf{R}_{\alpha_t}(\cdot), \ t = 2, ..., T.$$
(5.408)

With risk measures (5.408) the dynamic programming equations for problem (5.404) can be written as follows. At the last stage t = T the cost-to-go function $Q_T(x_{T-1}, \xi_T)$ is given by the optimal value of the problem

$$\underset{x_T \in X_T}{\min} f_T(x_T, \xi_T) \text{ s.t } B_T x_{T-1} + A_T x_T = b_T.$$
(5.409)

At stage t = T - 1, ..., 2, the cost-to-go function $Q_t(x_{t-1}, \xi_{[t]})$ is given by the optimal value of problem

$$\begin{aligned}
& \underset{x_t \in X_t}{\min} \quad f_t(x_t, \xi_t) + \mathbb{E}_{|\xi_{[t]}} \left[Q_{t+1} \left(x_t, \xi_{[t+1]} \right) \right] \\
& \text{s.t.} \quad B_t x_{t-1} + A_t x_t = b_t, \\
& \quad \mathsf{AV} @ \mathsf{R}_{\alpha_{t+1} |\xi_{[t]}} \left[Q_{t+1} \left(x_t, \xi_{[t+1]} \right) \right] \leq \gamma_t.
\end{aligned} \tag{5.410}$$

Finally at the first stage the following problem should be solved

$$\begin{aligned}
& \underset{x_{1} \in X_{1}}{\min} \quad f_{1}(x_{1}) + \mathbb{E}[Q_{2}(x_{1},\xi_{2})] \\
& \text{s.t.} \quad A_{1}x_{1} = b_{1}, \\
& \mathsf{AV} @ \mathsf{R}_{\alpha_{2}} \left[Q_{2} \left(x_{1},\xi_{2} \right) \right] \leq \gamma_{1}.
\end{aligned}$$
(5.411)

It could happen that problems (5.410) become infeasible for some realizations of the random data even if the original problem (5.382) had relatively complete recourse. In order to deal with such infeasibility we move the constraints into the objective in a form of penalty. That is, let the cost-to-go functions be defined as $Q_{T+1}(\cdot, \cdot) \equiv 0$ and for t = T - 1, ..., 2,

$$Q_t(x_{t-1},\xi_{[t]}) = \inf_{x_t \in X_t} \left\{ f_t(x_t,\xi_t) + \mathcal{Q}_{t+1}(x_t,\xi_{[t]}) : B_t x_{t-1} + A_t x_t = b_t \right\}, \quad (5.412)$$

where

$$\mathcal{Q}_{t+1}(x_t, \xi_{[t]}) := \rho_{t+1|\xi_{[t]}}[Q_{t+1}(x_t, \xi_{[t+1]})]$$
(5.413)

and

$$\rho_t(\cdot) := (1 - \lambda_t) \mathbb{E}(\cdot) + \lambda_t \mathsf{AV} @\mathsf{R}_{\alpha_t}(\cdot)$$
(5.414)

for some $\lambda_t \in (0, 1)$. At the first stage the following problem should be solved

$$\min_{x_1 \in X_1} f_1(x_1) + \mathcal{Q}_2(x_1) \text{ s.t. } A_1 x_1 = b_1.$$
(5.415)

If the data process is *stagewise independent*, then the cost-to-go functions $Q_t(x_{t-1}, \xi_t)$ depend only on ξ_t , rather than $\xi_{[t]}$, the functions $Q_{t+1}(x_t)$ do not depend on the data process and the conditional expectations and risk measures become the corresponding unconditional ones.

Dynamic equations (5.412)–(5.413) correspond to the following risk averse stochastic program in the nested form (see section 6.8.4)

$$\underset{\substack{A_{1}x_{1}=b_{1}\\x_{1}\in X_{1}}}{\operatorname{Min}} \quad f_{1}(x_{1}) + \rho_{2|\xi_{1}} \Big[\inf_{B_{2}x_{1}+A_{2}x_{2}=b_{2}\\x_{2}\in X_{2}} f_{2}(x_{2},\xi_{2}) + \dots \\ + \rho_{T-1|\xi_{[T-2]}} \Big[\inf_{B_{T-1}x_{T-2}+A_{T-1}x_{T-1}=b_{T-1}\\B_{T-1}x_{T-1}\in X_{T-1}} f_{T-1}(x_{T-1},\xi_{T-1}) \\ + \rho_{T|\xi_{[T-1]}} \Big[\inf_{B_{T}x_{T-1}+A_{T}x_{T}=b_{T}\\x_{T}\in X_{T}} f_{T}(x_{T},\xi_{T})] \Big] \Big].$$
(5.416)

Recall that (see (6.23))

$$\mathsf{AV}@\mathsf{R}_{\alpha}(Z) = \inf_{u \in \mathbb{R}} \left\{ u + \alpha^{-1} \mathbb{E}[Z - u]_+ \right\}.$$
(5.417)

Using this variational representation of AV@R_{α} it is also possible to write dynamic programming equations for problem (5.416) in the following form. By (5.417) it follows that $Q_{t+1}(x_t, \xi_{[t]})$ is equal to the optimal value of the problem

$$\underset{u_{t}}{\operatorname{Min}} \mathbb{E}_{|\xi_{[t]}} \left[(1 - \lambda_{t+1}) Q_{t+1} \left(x_{t}, \xi_{[t+1]} \right) + \lambda_{t+1} \left(u_{t} + \alpha_{t+1}^{-1} [Q_{t+1} \left(x_{t}, \xi_{[t+1]} \right) - u_{t}]_{+} \right) \right].$$
(5.418)

Thus we can write the corresponding dynamic programming equations as follows. At the last stage t = T we have that $Q_T(x_{T-1}, \xi_T)$ is equal to the optimal value of problem

$$\min_{x_T \in X_T} f_T(x_T, \xi_T) \text{ s.t. } B_T x_{T-1} + A_T x_T = b_T,$$
(5.419)

and $\mathcal{Q}_T\left(x_{T-1}\xi_{[T-1]}\right)$ is equal to the optimal value of problem

$$\underset{u_{T-1}}{\operatorname{Min}} \mathbb{E}_{|\xi_{[T-1]}} \left[(1 - \lambda_T) Q_T \left(x_{T-1}, \xi_T \right) + \lambda_T u_{T-1} + \frac{\lambda_T}{\alpha_T} [Q_T \left(x_{T-1}, \xi_T \right) - u_{T-1}]_+ \right].$$
(5.420)

At stage t = T - 1 we have that $Q_{T-1}(x_{T-2}, \xi_{[T-1]})$ is equal to the optimal value of problem

$$\begin{array}{ll}
& \underset{x_{T-1}\in X_{T-1}}{\operatorname{Min}} & f_{T-1}(x_{T-1},\xi_{T-1}) + \mathcal{Q}_{T}(x_{T-1},\xi_{[T-1]}) \\
& \text{s.t.} & B_{T-1}x_{T-2} + A_{T-1}x_{T-1} = b_{T-1}.
\end{array}$$
(5.421)

By using (5.420) and (5.421) we can write that $Q_{T-1}(x_{T-2}, \xi_{[T-1]})$ is equal to the optimal value of problem

$$\begin{array}{ll} \underset{x_{T-1} \in X_{T-1} \\ u_{T-1} \in \mathbb{R} \\ \text{s.t.} \\ B_{T-1} x_{T-2} + A_{T-1} x_{T-1} = b_{T-1}, \end{array} f_{T-1} \left(x_{T-1}, \xi_{T-1} \right) + \lambda_T u_{T-1} + \mathcal{V}_T \left(x_{T-1}, u_{T-1}, \xi_{[T-1]} \right)$$
(5.422)

where $\mathcal{V}_T(x_{T-1}, u_{T-1}, \xi_{[T-1]})$ is equal to the following conditional expectation

$$\mathbb{E}_{|\xi_{[T-1]}} \Big[(1 - \lambda_T) Q_T (x_{T-1}, \xi_T) + \lambda_T \alpha_T^{-1} [Q_T (x_{T-1}, \xi_T) - u_{T-1}]_+ \Big].$$
(5.423)

By continuing this process backward we can write dynamic programming equations for t = T, ..., 2 as

$$Q_t \left(x_{t-1}, \xi_{[t]} \right) = \inf_{x_t \in X_t, u_t \in \mathbb{R}} \left\{ f_t(x_t, \xi_t) + \lambda_{t+1} u_t + \mathcal{V}_{t+1}(x_t, u_t, \xi_{[t]}) : \\ B_t x_{t-1} + A_t x_t = b_t \right\},$$
(5.424)

where $\mathcal{V}_{T+1}(x_T, u_T, \xi_{[T]}) \equiv 0, \lambda_{T+1} := 0$ and for t = T - 1, ..., 2,

$$\mathcal{V}_{t+1}\left(x_{t}, u_{t}, \xi_{[t]}\right) := \mathbb{E}_{|\xi_{[t]}} \left[(1 - \lambda_{t+1})Q_{t+1}\left(x_{t}, \xi_{[t+1]}\right) + \lambda_{t+1}\alpha_{t+1}^{-1}\left[Q_{t+1}\left(x_{t}, \xi_{[t+1]}\right) - u_{t}\right]_{+} \right].$$
(5.425)

At the first stage problem

$$\underset{x_1 \in X_1, u_1 \in \mathbb{R}}{\operatorname{Min}} f_1(x_1) + \lambda_2 u_1 + \mathcal{V}_2(x_1, u_1) \quad \text{s.t.} \ A_1 x_1 = b_1,$$
(5.426)

should be solved. Note that functions $\mathcal{V}_{t+1}(x_t, u_t, \xi_{[t]})$ are convex in (x_t, u_t) . In this formulation there are additional state variables u_t . By solving the above dynamic programming equations we find an optimal policy $\bar{x}_t = \bar{x}_t(\xi_{[t]})$ and $(1 - \alpha_{t+1})$ -quantile $\bar{u}_t = \bar{u}_t(\xi_{[t]})$ of the cost-to-go function $Q_{t+1}(\bar{x}_t, \xi_{[t+1]})$.

As it was pointed before, if the data process is stagewise independent, then the costto-go functions $Q_t(x_{t-1}, \xi_t)$ depend only on ξ_t , and the functions

$$Q_{t+1}(x_t) = \rho_{t+1}[Q_t(x_t, \xi_{t+1})]$$

and $\mathcal{V}_{t+1}(x_t, u_t)$ do not depend on the random data.

The SDDP Algorithm

In order to apply the backward step of the SDDP algorithm we need to know how to construct cuts for the cost-to-go functions. Let us start by considering the static case of two stage programming. Consider the following optimization problem

$$\min_{x \in \mathcal{X}} \left\{ \mathcal{Q}(x) := \rho[Q(x,\xi)] \right\},$$
(5.427)

where $\mathcal{X} \subset \mathbb{R}^n$ is a nonempty closed convex set, ξ is a random vector with probability distribution having finite support $\{\xi^1, ..., \xi^N\}$ with equal probabilities 1/N, $Q(x, \xi)$ is a real valued function convex in x and

$$\rho(\cdot) := (1 - \lambda)\mathbb{E}(\cdot) + \lambda \mathsf{AV}@\mathsf{R}_{\alpha}(\cdot), \ \lambda \in (0, 1).$$
(5.428)

Recall that minimum in the right hand side of (5.417) is attained at a $(1-\alpha)$ -quantile of the distribution of Z. Therefore we have that

$$Q(x) = \frac{1-\lambda}{N} \sum_{j=1}^{N} Q^{j}(x) + \lambda Q^{(\iota)}(x) + \frac{\lambda}{\alpha N} \sum_{j=1}^{N} [Q^{j}(x) - Q^{(\iota)}(x)]_{+}, \quad (5.429)$$

where $Q^j(x) := Q(x,\xi^j)$ and $Q^{(\iota)}(x)$ is the left side $(1 - \alpha)$ -quantile of the distribution of $Q(x,\xi)$. That is, numbers $Q^j(x)$, j = 1, ..., N, are arranged in the increasing order $Q^{(1)}(x) \le \cdots \le Q^{(N)}(x)$, and ι is the smallest integer such that $\iota \ge (1 - \alpha)N$. In particular, if $N < \alpha^{-1}$, then $\iota = N$ and $Q^{(\iota)}(x) = \max_{1 \le j \le N} Q^j(x)$.

Suppose that $(1 - \alpha)N$ is not an integer and the $(1 - \alpha)$ -quantile $Q^{(\iota)}(x)$ is defined uniquely, i.e., $Q^{(\iota)}(x)$ is different from the other values of $Q^j(x)$. Then the index ι remains the same for small perturbations of x and hence a subgradient of Q(x) is given by

$$\nabla \mathcal{Q}(x) = \frac{1-\lambda}{N} \sum_{j=1}^{N} \nabla Q^{j}(x) + \lambda \nabla Q^{(\iota)}(x) + \frac{\lambda}{\alpha N} \sum_{j=\iota}^{N} \left[\nabla Q^{(j)}(x) - \nabla Q^{(\iota)}(x) \right],$$
(5.430)

where $\nabla Q^{j}(x)$ is a subgradient of $Q^{j}(x)$. In fact by continuity arguments, formula (5.430) gives a subgradient of Q(x) in general.

Alternatively using (5.417) we can write problem (5.427) in the following equivalent form

$$\underset{x \in \mathcal{X}, u \in \mathbb{R}}{\operatorname{Min}} \frac{1}{N} \sum_{j=1}^{N} V^{j}(x, u),$$
(5.431)

where

$$V^{j}(x,u) := (1-\lambda)Q^{j}(x) + \lambda \left(u + \alpha^{-1}[Q^{j}(x) - u]_{+}\right).$$
(5.432)

The functions $V^{j}(x, u)$ are convex with respective subgradients

$$\nabla_x V^j(x,u) = (1-\lambda) \nabla Q^j(x) + \lambda \alpha^{-1} \mathcal{I}^j(x,u) \nabla Q^j(x),
\nabla_u V^j(x,u) = \lambda - \lambda \alpha^{-1} \mathcal{I}^j(x,u),$$
(5.433)

where $\mathcal{I}^j(x, u) = 1$ if $Q^j(x) - u > 0$, $\mathcal{I}^j(x, u) = 0$ if $Q^j(x) - u < 0$, and $\mathcal{I}^j(x, u)$ can be any number between 0 and 1 if $Q^j(x) - u = 0$.

Consider now the multistage risk averse problem (5.416). Suppose that the data process is *stagewise independent*. There are two possible approaches to proceed with the backward and forward steps of the SDDP algorithm for solving this problem. The first approach is based on the dynamic programming equations (5.412)–(5.413). Formula (5.430) can be adjusted in a straightforward way for computing subgradients of the cost-to-go functions $Q_{t+1}(x_t)$ given in (5.413). This in turn can be used for constructing cuts of the cost-togo functions in the backward steps of the SDDP algorithm in the same way as in the risk neutral case. Also the forward steps of the algorithm can be performed in the same way as in the risk neutral case. Recall that in the risk neutral case the forward steps have two functions - to produce trial points, used in the consequent backward steps, and to estimate objective value of the constructed policy. For a feasible first stage solution \bar{x}_1 and a feasible policy $\bar{x}_t = \bar{x}_t(\xi_{[t]}), t = 2, ..., T$, the corresponding objective value of problem (5.416) is given in the nested form

$$f_{1}(\bar{x}_{1}) + \rho_{2|\xi_{1}} \left[f_{2}(\bar{x}_{2},\xi_{2}) + \dots + \rho_{T-1|\xi_{[T-2]}} \left[f_{T-1}(\bar{x}_{T-1},\xi_{T-1}) + \rho_{T|\xi_{[T-1]}} f_{T}(\bar{x}_{T},\xi_{T}) \right] \right]$$
(5.434)

Unfortunately, it is not clear here how this objective value can be estimated in the forward steps of the algorithm.

The other approach is based on the dynamic equations (5.418)–(5.426). Formulas (5.433) can be adjusted for computing subgradients of the respective functions $\mathcal{V}_{t+1}(x_t, u_t)$, and hence to construct their lower piecewise linear (affine) approximations $\mathfrak{V}_{t+1}(x_t, u_t)$. In a forward step of the algorithm the values (\bar{x}_t, \bar{u}_t) of the respective policy, for a generated scenario $\xi_2, ..., \xi_T$ and feasible first stage solution \bar{x}_1 , are computed iteratively for t = 2, ..., T, as optimal solutions of problems

$$\begin{array}{ll}
& \underset{x_t \in X_t, u_t \in \mathbb{R}}{\operatorname{Min}} & f_t(x_t, \xi_t) + \lambda_{t+1} u_t + \mathfrak{V}_{t+1}(x_t, u_t) \\
& \text{s.t.} & A_t x_t = b_t - B_t \bar{x}_{t-1}.
\end{array}$$
(5.435)

The objective value of this policy is given in the nested form (5.434). It can be estimated as follows. For the sake of simplicity suppose for the moment that $\lambda_t = 1$, t = 2, ..., T, i.e., consider nested AV@R measures. Since for any $u \in \mathbb{R}$,

$$\mathsf{AV}@\mathsf{R}_{\alpha}(Z) \leq \mathbb{E}\left[u + \alpha^{-1}[Z - u]_{+}\right],$$

we have that

$$\rho_{T|\xi_{[T-1]}} \left[f_T(\bar{x}_T, \xi_T) \right] \le \mathbb{E}_{|\xi_{[T-1]}} \left[\bar{u}_T + \alpha_T^{-1} [f_T(\bar{x}_T, \xi_T) - \bar{u}_T]_+ \right].$$
(5.436)

Denote $\nu_T := \bar{u}_T + \alpha_T^{-1} [f_T(\bar{x}_T, \xi_T) - \bar{u}_T]_+$. We have then that

$$\begin{aligned} \rho_{T-1|\xi_{[T-2]}} \left[f_{T-1}(\bar{x}_{T-1},\xi_{T-1}) + \rho_{T|\xi_{[T-1]}}[f_{T}(\bar{x}_{T},\xi_{T})] \right] \\ &\leq \mathbb{E}_{|\xi_{[T-2]}} \left[\bar{u}_{T-1} + \alpha_{T-1}^{-1} \left[\mathbb{E}_{|\xi_{[T-1]}}[\nu_{T}] + f_{T-1}(\bar{x}_{T-1},\xi_{T-1}) - \bar{u}_{T-1} \right]_{+} \right] \\ &\leq \mathbb{E}_{|\xi_{[T-2]}} \left[\bar{u}_{T-1} + \alpha_{T-1}^{-1} \mathbb{E}_{|\xi_{[T-1]}} \left[\nu_{T} + f_{T-1}(\bar{x}_{T-1},\xi_{T-1}) - \bar{u}_{T-1} \right]_{+} \right] \\ &= \mathbb{E}_{|\xi_{[T-2]}} \left[\bar{u}_{T-1} + \alpha_{T-1}^{-1} \left[\nu_{T} + f_{T-1}(\bar{x}_{T-1},\xi_{T-1}) - \bar{u}_{T-1} \right]_{+} \right], \end{aligned} \tag{5.437}$$

where the last inequality follows because $\psi(\cdot) := [\cdot]_+$ is a convex function and hence by Jensen inequality $[\mathbb{E}[Z]]_+ \leq \mathbb{E}[[Z]_+]$.

This suggests the following estimator. Set $\nu_{T+1} = 0$ and compute iteratively for t = T, ..., 2,

$$\nu_t = (1 - \lambda_t) [\nu_{t+1} + f_t(\bar{x}_t, \xi_t)] + \lambda_t \big[\bar{u}_t + \alpha_t^{-1} [\nu_{t+1} + f_t(\bar{x}_t, \xi_t) - \bar{u}_t]_+ \big].$$
(5.438)

By the above analysis we have that $\mathbb{E}[\nu_2]$ gives an upper bound for the objective value of the constructed policy, and hence for the optimal value of the multistage problem (5.416). Note that since the last inequality in (5.437) could be strict (it is based on the Jensen inequality), $\mathbb{E}[\nu_2]$ could be strictly bigger than the objective value of the considered policy. The expected value $\mathbb{E}[\nu_2]$ can be estimated by averaging the computed values (realizations) of ν_2 in the forward steps of the SDDP algorithm in the same way as in the risk neutral case. This average provides an unbiased estimate of $\mathbb{E}[\nu_2]$, and can be used for construction of a statistical upper bound for the objective value of the multistage problem (5.416).

Exercises

- 5.1. Suppose that set \mathcal{X} is defined by constraints in the form (5.11) with constraint functions given as expectations as in (5.12) and the set \mathcal{X}_N defined in (5.13). Show that if sample average functions \hat{g}_{iN} converge uniformly to g_i w.p.1 on a neighborhood of x and g_i are continuous, i = 1, ..., p, then condition (a) of Theorem 5.5 holds.
- 5.2. Specify regularity conditions under which equality (5.29) follows from (5.25).
- 5.3. Let $\mathcal{X} \subset \mathbb{R}^n$ be a closed convex set. Show that the multifunction $x \mapsto \mathcal{N}_{\mathcal{X}}(x)$ is closed.
- 5.4. Prove the following extension of Theorem 5.7. Let $g : \mathbb{R}^m \to \mathbb{R}$ be continuously differentiable function, $F_i(x,\xi)$, i = 1, ..., m, be random lower semicontinuous functions, $f_i(x) := \mathbb{E}[F_i(x,\xi)]$, i = 1, ..., m, $f(x) = (f_1(x), ..., f_m(x))$, \mathcal{X} be a nonempty compact subset of \mathbb{R}^n and consider the optimization problem

$$\operatorname{Min}_{x \in \mathcal{X}} g\left(f(x)\right).$$
(5.439)

Moreover, let $\xi^1, ..., \xi^N$ be an iid random sample, $\hat{f}_{iN}(x) := N^{-1} \sum_{j=1}^N F_i(x, \xi^j)$, $i = 1, ..., m, \hat{f}_N(x) = \left(\hat{f}_{1N}(x), ..., \hat{f}_{mN}(x)\right)$ be the corresponding sample average

functions and

$$\min_{x \in \mathcal{X}} g\left(\hat{f}_N(x)\right)$$
(5.440)

be the associated SAA problem. Suppose that conditions (A1) and (A2) (used in Theorem 5.7) hold for every function $F_i(x,\xi)$, i = 1, ..., m. Let ϑ^* and $\hat{\vartheta}_N$ be the optimal values of problems (5.439) and (5.440), respectively, and S be the set of optimal solutions of problem (5.439). Show that

$$\hat{\vartheta}_N - \vartheta^* = \inf_{x \in S} \left(\sum_{i=1}^m w_i(x) \left[\hat{f}_{iN}(x) - f_i(x) \right] \right) + o_p(N^{-1/2}), \tag{5.441}$$

where

$$w_i(x) := \frac{\partial g(y_1, ..., y_m)}{\partial y_i}\Big|_{y=f(x)}, \ i = 1, ..., m.$$

Moreover, if $S = {\bar{x}}$ is a singleton, then

$$N^{1/2}\left(\hat{\vartheta}_N - \vartheta^*\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$
 (5.442)

where $\bar{w}_i := w_i(\bar{x})$ and

$$\sigma^2 = \operatorname{\mathbb{V}ar}\left[\sum_{i=1}^m \bar{w}_i F_i(\bar{x},\xi)\right].$$
(5.443)

Hint: consider function $V : C(\mathcal{X}) \times \cdots \times C(\mathcal{X}) \to \mathbb{R}$ defined as $V(\psi_1, \ldots, \psi_m) := \inf_{x \in \mathcal{X}} g(\psi_1(x), \ldots, \psi_m(x))$, and apply the functional CLT together with Delta and Danskin Theorems.

5.5. Consider matrix $\begin{bmatrix} H & A \\ A^{\mathsf{T}} & 0 \end{bmatrix}$ defined in (5.44). Assuming that matrix *H* is positive definite and matrix *A* has full column rank, verify that

$$\begin{bmatrix} H & A \\ A^{\mathsf{T}} & 0 \end{bmatrix}^{-1} = \begin{bmatrix} H^{-1} - H^{-1}A(A^{\mathsf{T}}H^{-1}A)^{-1}A^{\mathsf{T}}H^{-1} & H^{-1}A(A^{\mathsf{T}}H^{-1}A)^{-1} \\ (A^{\mathsf{T}}H^{-1}A)^{-1}A^{\mathsf{T}}H^{-1} & -(A^{\mathsf{T}}H^{-1}A)^{-1} \end{bmatrix}$$

Using this identity write the asymptotic covariance matrix of $N^{1/2}\begin{bmatrix} \hat{x}_N - \bar{x} \\ \hat{\lambda}_N - \bar{\lambda} \end{bmatrix}$, given in (5.45), explicitly.

5.6. Consider the minimax stochastic problem (5.46), the corresponding SAA problem (5.47) and let

$$\Delta_N := \sup_{x \in \mathcal{X}, y \in \mathcal{Y}} \left| \hat{f}_N(x, y) - f(x, y) \right|.$$
(5.444)

(i) Show that $|\hat{\vartheta}_N - \vartheta^*| \leq \Delta_N$, and that if \hat{x}_N is a δ -optimal solution of the SAA problem (5.47), then \hat{x}_N is a $(\delta + 2\Delta_N)$ -optimal solution of the minimax problem (5.46).

(ii) By using Theorem 7.73 conclude that, under appropriate regularity conditions, for any $\varepsilon > 0$ there exist positive constants $C = C(\varepsilon)$ and $\beta = \beta(\varepsilon)$ such that

$$\Pr\left\{\left|\hat{\vartheta}_{N}-\vartheta^{*}\right|\geq\varepsilon\right\}\leq Ce^{-N\beta}.$$
(5.445)

(iii) By using bounds (7.241) and (7.242) derive an estimate, similar to (5.116), of the sample size N which guarantees with probability at least $1 - \alpha$ that a δ -optimal solution \hat{x}_N of the SAA problem (5.47) is an ε -optimal solution of the minimax problem (5.46). Specify required regularity conditions.

- 5.7. Consider multistage SAA method based on iid conditional sampling. For corresponding sample sizes $\mathcal{N} = (N_1, ..., N_{T-1})$ and $\mathcal{N}' = (N'_1, ..., N'_{T-1})$ we say that $\mathcal{N}' \succeq \mathcal{N}$ if $N'_t \ge N_t$, t = 1, ..., T 1. Let $\hat{\vartheta}_{\mathcal{N}}$ and $\hat{\vartheta}_{\mathcal{N}'}$ be respective optimal (minimal) values of SAA problems. Show that if $\mathcal{N}' \succeq \mathcal{N}$, then $\mathbb{E}[\hat{\vartheta}_{\mathcal{N}'}] \ge \mathbb{E}[\hat{\vartheta}_{\mathcal{N}}]$.
- 5.8. Consider the following chance constrained problem

$$\underset{x \in \mathcal{X}}{\operatorname{Min}} f(x) \text{ subject to } \Pr\{T(\xi)x + h(\xi) \in C\} \ge 1 - \alpha,$$
(5.446)

where $\mathcal{X} \subset \mathbb{R}^n$ is a closed convex set, $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, $C \subset \mathbb{R}^m$ is a convex closed set, $\alpha \in (0, 1)$, matrix $T(\xi)$ and vector $h(\xi)$ are functions of random vector ξ . For example, if

$$C := \left\{ z : z = -Wy - w, \ y \in \mathbb{R}^{\ell}, \ w \in \mathbb{R}^{m}_{+} \right\},$$
(5.447)

then, for a given $x \in X$, the constraint $T(\xi)x + h(\xi) \in C$ means that the system $Wy + T(\xi)x + h(\xi) \leq 0$ has a feasible solution. Extend the results of section 5.7 to the setting of problem (5.446).

5.9. Consider the following extension of the chance constrained problem (5.196):

$$\underset{x \in \mathcal{X}}{\operatorname{Min}} f(x) \text{ subject to } p_i(x) \le \alpha_i, \ i = 1, \dots, p,$$
(5.448)

with several (individual) chance constraints. Here $\mathcal{X} \subset \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $\alpha_i \in (0, 1), i = 1, ..., p$, are given significance levels, and

$$p_i(x) = \mathsf{Pr}\{C_i(x,\xi) > 0\}, \ i = 1, ..., p,$$

with $C_i(x,\xi)$ being Carathéodory functions.

Extend the methodology of constructing lower and upper bounds, discussed in section 5.7.2, to the above problem (5.448). Use SAA problems based on *independent* samples (see Remark 9 on page 182, and equation (5.18) in particular). That is, estimate $p_i(x)$ by

$$\hat{p}_{iN_i}(x) := \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{1}_{(0,\infty)} \left(C_i(x,\xi^{ij}) \right), \ i = 1, ..., p.$$

In order to verify feasibility of a point $\bar{x} \in X$, show that

$$\Pr\{p_i(\bar{x}) < U_i(\bar{x}), \ i = 1, ..., p\} \ge \prod_{i=1}^p (1 - \beta_i),$$

where $\beta_i \in (0, 1)$ are chosen constants and

$$U_i(\bar{x}) := \sup_{\rho \in [0,1]} \left\{ \rho : \mathfrak{b}\left(\mathfrak{m}_i; \rho, N_i\right) \ge \beta_i \right\}, \ i = 1, ..., p$$

with $\mathfrak{m}_i := \hat{p}_{iN_i}(\bar{x}).$

In order to construct a lower bound, generate M independent realizations of the corresponding SAA problems, each of the same sample size $\mathcal{N} = (N_1, ..., N_p)$ and significance levels $\gamma_i \in [0, 1)$, i = 1, ..., p, and compute their optimal values $\hat{\vartheta}^1_{\gamma,\mathcal{N}}, ..., \hat{\vartheta}^M_{\gamma,\mathcal{N}}$. Arrange these values in the increasing order $\hat{\vartheta}^{(1)}_{\gamma,\mathcal{N}} \leq ... \leq \hat{\vartheta}^{(M)}_{\gamma,\mathcal{N}}$. Given significance level $\beta \in (0, 1)$, consider the following rule for choice of the corresponding integer L:

• Choose largest integer $L \in \{1, ..., M\}$ such that

$$\mathfrak{b}(L-1;\theta_{\mathcal{N}},M) \le \beta, \tag{5.449}$$

where $\theta_{\mathcal{N}} := \prod_{i=1}^{p} \mathfrak{b}(r_i; \alpha_i, N_i)$ and $r_i := \lfloor \gamma_i N_i \rfloor$.

Show that with probability at least $1 - \beta$, the random quantity $\hat{\vartheta}_{\gamma,\mathcal{N}}^{(L)}$ gives a lower bound for the true optimal value ϑ^* .

5.10. Consider the SAA problem (5.241) giving an approximation of the first stage of the corresponding three stage stochastic program. Let

$$\tilde{\vartheta}_{N_1,N_2} := \inf_{x_1 \in \mathcal{X}_1} \tilde{f}_{N_1,N_2}(x_1)$$

be the optimal value and \tilde{x}_{N_1,N_2} be an optimal solution of problem (5.241). Consider asymptotics of $\tilde{\vartheta}_{N_1,N_2}$ and \tilde{x}_{N_1,N_2} as N_1 tends to infinity while N_2 is *fixed*. Let $\vartheta_{N_2}^*$ be the optimal value and S_{N_2} be the set of optimal solutions of the problem

$$\min_{x_1 \in \mathcal{X}_1} \left\{ f_1(x_1) + \mathbb{E} \left[\hat{Q}_{2,N_2}(x_1, \xi_2^i) \right] \right\},$$
(5.450)

where the expectation is taken with respect to the distribution of the random vector $(\xi_2^i, \xi_3^{i1}, ..., \xi_3^{iN_2})$.

(i) By using results of section 5.1.1 show that $\tilde{\vartheta}_{N_1,N_2} \to \vartheta^*_{N_2}$ w.p.1 and distance from \tilde{x}_{N_1,N_2} to \mathcal{S}_{N_2} tends to 0 w.p.1 as $N_1 \to \infty$. Specify required regularity conditions.

(ii) Show that, under appropriate regularity conditions,

$$\tilde{\vartheta}_{N_1,N_2} = \inf_{x_1 \in \mathcal{S}_{N_2}} \tilde{f}_{N_1,N_2}(x_1) + o_p \left(N_1^{-1/2}\right).$$
(5.451)

Conclude that if, moreover, $S_{N_2} = {\bar{x}_1}$ is a singleton, then

$$N_1^{1/2} \left(\tilde{\vartheta}_{N_1, N_2} - \vartheta_{N_2}^* \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \sigma^2(\bar{x}_1) \right), \tag{5.452}$$

where $\sigma^2(\bar{x}_1) := \mathbb{V}ar[\hat{Q}_{2,N_2}(x_1,\xi_2^i)]$. Hint: use Theorem 5.7.

Chapter 6 Risk Averse Optimization

Andrzej Ruszczyński and Alexander Shapiro

6.1 Introduction

So far, we discussed stochastic optimization problems, in which the objective function was defined as the expected value $f(x) := \mathbb{E}[F(x,\omega)]$. The function $F : \mathbb{R}^n \times \Omega \to \mathbb{R}$ models the random outcome, for example, the random cost, and is assumed to be sufficiently regular so that the expected value function is well defined. For a feasible set $\mathcal{X} \subset \mathbb{R}^n$, the stochastic optimization model

$$\operatorname{Min}_{x \in \mathcal{X}} f(x) \tag{6.1}$$

optimizes the random outcome $F(x, \omega)$ on average. This is justified when the Law of Large Numbers can be invoked and we are interested in the long term performance, irrespective of the fluctuations of specific outcome realizations. The shortcomings of such an approach can be clearly illustrated by the example of portfolio selection discussed in section 1.4. Consider problem (1.34) of maximizing the expected return rate. Its optimal solution suggests to concentrate the investment in the assets having the highest expected return rate. This is not what we would consider reasonable, because it leaves out all considerations of the involved risk of losing all or a large portion of the capital invested. In this section, we discuss stochastic optimization from a new point of view which takes into account risk aversion.

A classical approach to risk-averse preferences is based on the expected utility theory, which has its roots in mathematical economics (we already touched upon this subject in section 1.4). In this theory, in order to compare two random outcomes we consider expected values of some scalar transformations $u : \mathbb{R} \to \mathbb{R}$ of the realization of these outcomes. In a minimization problem, a random outcome Z_1 (understood as a scalar random variable) is

preferred over a random outcome Z_2 , if

$$\mathbb{E}[u(Z_1)] < \mathbb{E}[u(Z_2)].$$

The function $u(\cdot)$, called the *disutility function*, is assumed to be nondecreasing and convex. Following this principle, instead of problem (6.1), we construct the problem

$$\min_{x \in \mathcal{X}} \mathbb{E}[u(F(x,\omega))].$$
(6.2)

Observe that it is still an expected value problem, but the function F is replaced by the composition $u \circ F$. Since $u(\cdot)$ is convex, we have by Jensen's inequality that

$$u(\mathbb{E}[F(x,\omega)]) \le \mathbb{E}[u(F(x,\omega))].$$

That is, a sure outcome of $\mathbb{E}[F(x,\omega)]$ is at least as good as the random outcome $F(x,\omega)$. In a maximization problem, we assume that $u(\cdot)$ is concave (and still nondecreasing). We call it a *utility function* in this case. Again, Jensen's inequality yields the preference in terms of expected utility:

$$u(\mathbb{E}[F(x,\omega)]) \ge \mathbb{E}[u(F(x,\omega))].$$

One of the basic difficulties in using the expected utility approach is specifying the utility or disutility function. They are very difficult to elicit; even the authors of this book cannot specify their utility functions in simple stochastic optimization problems. Moreover, using some arbitrarily selected utility functions may lead to solutions which are difficult to interpret and explain. Modern approach to modeling risk aversion in optimization problems uses the concept of risk measures. These are, generally speaking, functionals which take as their argument the entire collection of realizations $Z(\omega) = F(x, \omega), \omega \in \Omega$, understood as an object in an appropriate vector space. In the following sections we introduce this concept in a more formalized way.

6.2 Mean–risk models

6.2.1 Main ideas of mean–risk analysis

The main idea of mean-risk models is to characterize the uncertain outcome $Z_x(\omega) = F(x, \omega)$ by two scalar characteristics: the *mean* $\mathbb{E}[Z_x]$ describing the expected outcome, and the *risk (dispersion measure)* $\mathbb{D}[Z_x]$ which measures the uncertainty of the outcome. In the mean-risk approach, we select from the set of all possible solutions those that are *efficient*: for a given value of the mean they minimize the risk and for a given value of risk – minimize the mean. Such an approach has many advantages: it allows one to formulate the problem as a parametric optimization problem, and it facilitates the trade-off analysis between mean and risk.

Let us describe the mean-risk analysis on the example of the minimization problem (6.1). Suppose that the risk functional is defined as the variance $\mathbb{D}[Z] := \mathbb{V}ar[Z]$, which is well defined for $Z \in \mathcal{L}_2(\Omega, \mathcal{F}, P)$, i.e., for Z having a finite second order moment. The variance, although not the best choice, is easiest to start from. It is also important in finance. Later in this chapter, we discuss in much detail desirable properties of the risk functionals.

In the mean-risk approach, we aim at finding efficient solutions of the problem with two objectives, namely, $\mathbb{E}[Z_x]$ and $\mathbb{D}[Z_x]$, subject to the feasibility constraint $x \in \mathcal{X}$. This can be accomplished by techniques of multi-objective optimization. Most convenient, from our perspective, is the idea of *scalarization*. For a coefficient $c \ge 0$, we form a composite objective functional

$$\rho[Z] := \mathbb{E}[Z] + c\mathbb{D}[Z]. \tag{6.3}$$

The coefficient c plays the role of the price of risk. We formulate the problem

$$\min_{x \in \mathcal{X}} \mathbb{E}[Z_x] + c\mathbb{D}[Z_x].$$
(6.4)

By varying the value of the coefficient c, we can generate in this way a large ensemble of efficient solutions. We already discussed this approach for the portfolio selection problem, with $\mathbb{D}[Z] := \mathbb{V}ar[Z]$, in section 1.4.

An obvious deficiency of variance as a measure of risk is that it treats the excess over the mean equally as the shortfall. After all, in a minimization case, we are not concerned if a particular realization of Z is significantly below the mean of Z; we do not want it to be too large. Two particular classes of risk functionals, which we discuss next, play an important role in the theory of mean–risk models.

6.2.2 Semideviations

An important group of risk functionals (representing dispersion measures) are *central semide*viations. The upper semideviation of order p is defined as

$$\sigma_p^+[Z] := \left(\mathbb{E} \Big[\left(Z - \mathbb{E}[Z] \right)_+^p \Big] \right)^{1/p}, \tag{6.5}$$

where $p \in [1, \infty)$ is a fixed parameter. It is natural to assume here that the random variables (uncertain outcomes) $Z : \Omega \to \mathbb{R}$ considered belong to the space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$, i.e., that they have finite *p*-th order moments. That is, $\sigma_p^+[Z]$ is well-defined and *finite*-valued for all $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$. The corresponding mean–risk model has the general form

$$\underset{x \in \mathcal{X}}{\operatorname{Min}} \ \mathbb{E}[Z_x] + c\sigma_p^+[Z_x].$$
(6.6)

The upper semideviation measure is appropriate for minimization problems, where $Z_x(\omega) = F(x, \omega)$ represents a cost, as a function of $\omega \in \Omega$. It is aimed at penalization of an excess of Z_x over its mean. If we are dealing with a maximization problem, where Z_x represents some reward or profit, the corresponding risk functional is the *lower* semideviation

$$\sigma_p^{-}[Z] := \left(\mathbb{E}\left[\left(\mathbb{E}[Z] - Z \right)_+^p \right] \right)^{1/p}, \tag{6.7}$$

where $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$. The resulting mean-risk model has the form

$$\max_{x \in \mathcal{X}} \mathbb{E}[Z_x] - c\sigma_p^-[Z_x].$$
(6.8)

In the special case of p = 1, both left and right first order semideviations are related to the mean absolute deviation

$$\sigma_1(Z) := \mathbb{E} |Z - \mathbb{E}[Z]|. \tag{6.9}$$

Proposition 6.1. The following identity holds

$$\sigma_1^+[Z] = \sigma_1^-[Z] = \frac{1}{2}\sigma_1[Z], \quad \forall Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P).$$
(6.10)

Proof. Denote by $H(\cdot)$ the cumulative distribution function (cdf) of Z and let $\mu := \mathbb{E}[Z]$. We have

$$\sigma_1^{-}[Z] = \int_{-\infty}^{\mu} (\mu - z) \, dH(z) = \int_{-\infty}^{\infty} (\mu - z) \, dH(z) - \int_{\mu}^{\infty} (\mu - z) \, dH(z).$$

The first integral on the right hand side is equal to 0, and thus $\sigma_1^-[Z] = \sigma_1^+[Z]$. The identity (6.10) follows now from the equation $\sigma_1[Z] = \sigma_1^-[Z] + \sigma_1^+[Z]$. \Box

We conclude that using the mean absolute deviation instead of the semideviation in mean-risk models has the same effect, just the parameter c has to be halved. The identity (6.10) does not extend to semideviations of higher orders, unless the distribution of Z is symmetric.

6.2.3 Weighted mean deviations from quantiles

Let $H_Z(z) = \Pr(Z \le z)$ be the cdf of the random variable Z, and let $\alpha \in (0, 1)$. Recall that the *left-side* α -quantile of H_Z is defined as

$$H_Z^{-1}(\alpha) := \inf\{t : H_Z(t) \ge \alpha\},\tag{6.11}$$

and the *right-side* α *-quantile* as

$$\sup\{t: H_Z(t) \le \alpha\}. \tag{6.12}$$

If Z represents losses, the (left-side) quantile $H_Z^{-1}(1-\alpha)$ is also called *Value-at-Risk* and denoted V@R_{α}(Z), i.e.,

$$V@R_{\alpha}(Z) := H_{Z}^{-1}(1-\alpha) = \inf\{t : \Pr(Z \le t) \ge 1-\alpha\} \\ = \inf\{t : \Pr(Z > t) \le \alpha\}.$$
(6.13)

Its meaning is the following: losses larger than $V@R_{\alpha}(Z)$ occur with probability not exceeding α . Note that

$$\mathsf{V}@\mathsf{R}_{\alpha}(\gamma Z + \tau) = \gamma \mathsf{V}@\mathsf{R}_{\alpha}(Z) + \tau, \ \forall \gamma, \tau \in \mathbb{R}, \ \gamma \ge 0.$$
(6.14)

The weighted mean deviation from a quantile is defined as follows:

$$\mathfrak{q}_{\alpha}[Z] := \mathbb{E}\left[\max\left\{(1-\alpha)(H_Z^{-1}(\alpha)-Z),\alpha(Z-H_Z^{-1}(\alpha))\right\}\right].$$
(6.15)

The functional $\mathfrak{q}_{\alpha}[Z]$ is well defined and finite valued for all $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$. It can be easily shown that

$$\mathfrak{q}_{\alpha}[Z] = \min_{t \in \mathbb{R}} \left\{ \varphi(t) := \mathbb{E} \left[\max \left\{ (1 - \alpha)(t - Z), \alpha(Z - t) \right\} \right] \right\}.$$
(6.16)

Indeed, the right and left side derivatives of the function $\varphi(\cdot)$ are

$$\begin{aligned} \varphi'_{+}(t) &= (1-\alpha) \Pr[Z \leq t] - \alpha \Pr[Z > t], \\ \varphi'_{-}(t) &= (1-\alpha) \Pr[Z < t] - \alpha \Pr[Z \geq t]. \end{aligned}$$

At the optimal t the right derivative is nonnegative and the left derivative nonpositive, and thus

$$\Pr[Z < t] \le \alpha \le \Pr[Z \le t].$$

This means that every α -quantile is a minimizer in (6.16).

The risk functional $\mathfrak{q}_{\alpha}[\,\cdot\,]$ can be used in mean–risk models, both in the case of minimization

$$\underset{x \in \mathcal{X}}{\operatorname{Min}} \ \mathbb{E}[Z_x] + c \,\mathfrak{q}_{1-\alpha}[Z_x],$$
(6.17)

and in the case of maximization

$$\max_{x \in \mathcal{X}} \mathbb{E}[Z_x] - c \mathfrak{q}_{\alpha}[Z_x].$$
(6.18)

We use $1 - \alpha$ in the minimization problem and α in the maximization problem, because in practical applications we are interested in these quantities for small values of α .

6.2.4 Average Value-at-Risk

The mean-deviation from quantile model is closely related to the concept of Average Valueat-Risk¹. Suppose that Z represents losses and we want to satisfy the chance constraint

$$\mathsf{V}@\mathsf{R}_{\alpha}[Z_x] \le 0. \tag{6.19}$$

Recall that

$$\mathsf{V}@\mathsf{R}_{\alpha}[Z] = \inf\{t : \mathsf{Pr}(Z \le t) \ge 1 - \alpha\},\$$

and hence constraint (6.19) is equivalent to the constraint $\Pr(Z_x \leq 0) \geq 1 - \alpha$. We have that² $\Pr(Z_x > 0) = \mathbb{E}\left[\mathbf{1}_{(0,\infty)}(Z_x)\right]$, and hence constraint (6.19) can also be written as the expected value constraint:

$$\mathbb{E}\left[\mathbf{1}_{(0,\infty)}(Z_x)\right] \le \alpha. \tag{6.20}$$

The source of difficulties with probabilistic (chance) constraints is that the step function $\mathbf{1}_{(0,\infty)}(\cdot)$ is not convex and, even worse, it is discontinuous at zero. As a result, chance constraints are often nonconvex, even if the function $x \mapsto Z_x$ is convex almost surely. One possibility is to approach such problems by constructing a convex approximation of the expected value on the left of (6.20).

¹Average Value-at-Risk is often called Conditional Value-at-Risk. We adopt here the term 'Average' rather than 'Conditional' Value-at-Risk in order to avoid awkward notation and terminology while discussing later conditional risk mappings.

²Recall that $\mathbf{1}_{(0,\infty)}(z) = 0$ if $z \le 0$, and $\mathbf{1}_{(0,\infty)}(z) = 1$ if z > 0.

Let $\psi : \mathbb{R} \to \mathbb{R}$ be a nonnegative valued, nondecreasing, convex function such that $\psi(z) \ge \mathbf{1}_{(0,\infty)}(z)$ for all $z \in \mathbb{R}$. By noting that $\mathbf{1}_{(0,\infty)}(tz) = \mathbf{1}_{(0,\infty)}(z)$ for any t > 0 and $z \in \mathbb{R}$, we have that $\psi(tz) \ge \mathbf{1}_{(0,\infty)}(z)$ and hence the following inequality holds

$$\inf_{t>0} \mathbb{E}\left[\psi(tZ)\right] \ge \mathbb{E}\left[\mathbf{1}_{(0,\infty)}(Z)\right].$$

Consequently, the constraint

$$\inf_{t>0} \mathbb{E}\left[\psi(tZ_x)\right] \le \alpha \tag{6.21}$$

is a *conservative* approximation of the chance constraint (6.19) in the sense that the feasible set defined by (6.21) is contained in the feasible set defined by (6.19).

Of course, the smaller the function $\psi(\cdot)$ is the better this approximation will be. From this point of view the best choice of $\psi(\cdot)$ is to take piecewise linear function $\psi(z) := [1 + \gamma z]_+$ for some $\gamma > 0$. Since constraint (6.21) is invariant with respect to scale change of $\psi(\gamma z)$ to $\psi(z)$, we have that $\psi(z) := [1 + z]_+$ gives the best choice of such a function. For this choice of function $\psi(\cdot)$, we have that constraint (6.21) is equivalent to

$$\inf_{t>0} \{ t \mathbb{E}[t^{-1} + Z]_+ - \alpha \} \le 0,$$

or equivalently

$$\inf_{t>0} \left\{ \alpha^{-1} \mathbb{E}[Z+t^{-1}]_+ - t^{-1} \right\} \le 0.$$

Now replacing t with $-t^{-1}$ we get the form:

$$\inf_{t<0} \left\{ t + \alpha^{-1} \mathbb{E}[Z-t]_+ \right\} \le 0.$$
(6.22)

The quantity

$$\mathsf{AV}@\mathsf{R}_{\alpha}(Z) := \inf_{t \in \mathbb{R}} \left\{ t + \alpha^{-1} \mathbb{E}[Z - t]_+ \right\}$$
(6.23)

is called the Average Value-at-Risk³ of Z (at level α). Note that AV@R_{α}(Z) is well defined and finite valued for every $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$.

The function $\varphi(t) := t + \alpha^{-1} \mathbb{E}[Z - t]_+$ is convex. Its derivative at t is equal to $1 + \alpha^{-1}[H_Z(t) - 1]$, provided that the cdf $H_Z(\cdot)$ is continuous at t. If $H_Z(\cdot)$ is discontinuous at t, then the respective right and left side derivatives of $\varphi(\cdot)$ are given by the same formula with $H_Z(t)$ understood as the corresponding right and left side limits. Therefore the minimum of $\varphi(t)$, over $t \in \mathbb{R}$, is attained on the interval $[t^*, t^{**}]$, where

$$t^* := \inf\{z : H_Z(z) \ge 1 - \alpha\}$$
 and $t^{**} := \sup\{z : H_Z(z) \le 1 - \alpha\}$ (6.24)

are the respective left and right side quantiles. Recall that the left-side quantile $t^* = V@R_{\alpha}(Z)$.

Since the minimum of $\varphi(t)$ is attained at $t^* = V@R_{\alpha}(Z)$, we have that AV@R_{α}(Z) is larger than V@R_{α}(Z) by the nonnegative amount of $\alpha^{-1}\mathbb{E}[Z - t^*]_+$. Therefore

$$\inf_{t \in \mathbb{R}} \left\{ t + \alpha^{-1} \mathbb{E}[Z - t]_+ \right\} \le 0 \text{ implies that } t^* \le 0,$$

³In some publications the above concept of 'Average Value-at-Risk' is called 'Conditional Value-at-Risk' and is denoted $CV@R_{\alpha}$.

and hence constraint (6.22) is equivalent to AV@R_{α}(Z) \leq 0. Therefore, the constraint

$$\mathsf{AV}@\mathsf{R}_{\alpha}[Z_x] \le 0 \tag{6.25}$$

is equivalent to the constraint (6.22) and gives a conservative approximation⁴ of the chance constraint (6.19).

The function $\rho(Z) := AV@R_{\alpha}(Z)$, defined on a space of random variables, is *convex*, i.e., if Z and Z' are two random variables and $t \in [0, 1]$, then

$$\rho(tZ + (1-t)Z') \le t\rho(Z) + (1-t)\rho(Z').$$

This follows from the fact that the function $t + \alpha^{-1} \mathbb{E}[Z - t]_+$ is convex jointly in t and Z. Also $\rho(\cdot)$ is monotonic, i.e., if Z and Z' are two random variables such that with probability one $Z \ge Z'$, then $\rho(Z) \ge \rho(Z')$. It follows that if $G(\cdot, \xi)$ is convex for a.e. $\xi \in \Xi$, then the function $\rho[G(\cdot, \xi)]$ is also convex. Indeed, by convexity of $G(\cdot, \xi)$ and monotonicity of $\rho(\cdot)$, we have for any $t \in [0, 1]$ that

$$\rho[G(tZ + (1-t)Z'),\xi)] \le \rho[tG(Z,\xi) + (1-t)G(Z',\xi)],$$

and hence by convexity of $\rho(\cdot)$ that

$$\rho[G(tZ + (1-t)Z'),\xi)] \le t\rho[G(Z,\xi)] + (1-t)\rho[G(Z',\xi)].$$

Consequently, (6.25) is a *convex* conservative approximation of the chance constraint (6.19). Moreover, from the considered point of view, (6.25) is the best convex conservative approximation of the chance constraint (6.19).

We can now relate the concept of Average Value-at-Risk to mean deviations from quantiles. Recall that (see (6.15))

$$\mathfrak{q}_{\alpha}[Z] := \mathbb{E}\left[\max\left\{(1-\alpha)(H_Z^{-1}(\alpha)-Z),\alpha(Z-H_Z^{-1}(\alpha))\right\}\right].$$

Theorem 6.2. Let $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$ and $H(\cdot) = H_Z(\cdot)$ be its cdf. Then the following identities hold true

$$\mathsf{AV}@\mathsf{R}_{\alpha}(Z) = \frac{1}{\alpha} \int_{1-\alpha}^{1} \mathsf{V}@\mathsf{R}_{1-\tau}(Z) \, d\tau = \mathbb{E}[Z] + \frac{1}{\alpha} \mathfrak{q}_{1-\alpha}[Z]. \tag{6.27}$$

Moreover, if H(z) is continuous at $z = V@R_{\alpha}(Z)$, then

$$\mathsf{AV}@\mathsf{R}_{\alpha}(Z) = \frac{1}{\alpha} \int_{\mathsf{V}@\mathsf{R}_{\alpha}(Z)}^{+\infty} z dH(z) = \mathbb{E}\left[Z \middle| Z \ge \mathsf{V}@\mathsf{R}_{\alpha}(Z)\right].$$
(6.28)

Proof. As discussed before, the minimum in (6.23) is attained at $t^* = H^{-1}(1 - \alpha) = V@R_{\alpha}(Z)$. Therefore

$$\mathsf{AV}@\mathsf{R}_{\alpha}(Z) = t^* + \alpha^{-1}\mathbb{E}[Z - t^*]_+ = t^* + \alpha^{-1}\int_{t^*}^{+\infty} (z - t^*)dH(z).$$

$$W@R_{\alpha}(Z+\tau) = AV@R_{\alpha}(Z) + \tau.$$
(6.26)

⁴It is easy to see that for any $\tau \in \mathbb{R}$,

Consequently the constraint AV@R_{α}[Z_x] $\leq \tau$ gives a conservative approximation of the chance constraint V@R_{α}[Z_x] $\leq \tau$.

Moreover,

$$\int_{t^*}^{+\infty} dH(z) = \Pr(Z \ge t^*) = 1 - \Pr(Z \le t^*) = \alpha,$$

provided that $Pr(Z = t^*) = 0$, i.e., that H(z) is continuous at $z = V@R_{\alpha}(Z)$. This shows the first equality in (6.28), and then the second equality in (6.28) follows provided that $Pr(Z = t^*) = 0$.

The first equality in (6.27) follows from the first equality in (6.28) by the substitution $\tau = H(z)$. Finally, we have

$$AV@R_{\alpha}(Z) = t^* + \alpha^{-1}\mathbb{E}[Z - t^*]_+ = \mathbb{E}[Z] + \mathbb{E}\left\{-Z + t^* + \alpha^{-1}[Z - t^*]_+\right\}$$
$$= \mathbb{E}[Z] + \mathbb{E}\left[\max\left\{\alpha^{-1}(1 - \alpha)(Z - t^*), t^* - Z\right\}\right].$$

This proves the last equality in (6.27).

The first equation in (6.27) motivates the term *Average Value-at-Risk*. The last equation in (6.28) explains the origin of the alternative term *Conditional Value-at-Risk*.

Remark 24. By (6.27) we have that for $\alpha = 1$,

$$\mathsf{AV}@\mathsf{R}_1(Z) = \int_0^1 \mathsf{V}@\mathsf{R}_{1-\tau}(Z) \, d\tau = \int_0^1 H^{-1}(\tau) d\tau = \int_{-\infty}^{+\infty} z \, dH(z) = \mathbb{E}[Z].$$

For α tending to zero we have

$$\lim_{\alpha \downarrow 0} \mathsf{AV}@\mathsf{R}_{\alpha}(Z) = \lim_{\alpha \downarrow 0} \frac{1}{\alpha} \int_{1-\alpha}^{1} H^{-1}(\tau) \, d\tau = \mathrm{ess} \, \mathrm{sup}(Z),$$

where ess sup(Z) is the essential supremum of $Z(\omega)$. Therefore we set

$$\mathsf{AV}@\mathsf{R}_0(\cdot) := \operatorname{ess\,sup}(\cdot).$$

It follows from (6.23) that $AV@R_{\alpha}(Z)$ is monotonically decreasing in α , i.e., for $0 < \alpha < \alpha' \leq 1$ we have $AV@R_{\alpha}(Z) \geq AV@R_{\alpha'}(Z)$. In particular $AV@R_{\alpha}(Z) \geq \mathbb{E}[Z]$ for any $\alpha \in (0, 1]$. Also it follows from (6.27) that for a given $Z \in \mathcal{Z}$, $AV@R_{\alpha}(Z)$ is a continuous function of $\alpha \in (0, 1]$.

Theorem 6.2 allows us to show an important relation between the absolute semideviation $\sigma_1^+[Z]$ and the mean deviation from quantile $\mathfrak{q}_{\alpha}[Z]$.

Corollary 6.3. For every $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$ we have

$$\sigma_1^+[Z] = \max_{\alpha \in [0,1]} \mathfrak{q}_{\alpha}[Z] = \min_{t \in \mathbb{R}} \max_{\alpha \in [0,1]} \mathbb{E} \left\{ (1-\alpha)[t-Z]_+ + \alpha[Z-t]_+ \right\}.$$
(6.29)

Proof. From (6.27) we get

$$\mathfrak{q}_{1-\alpha}[Z] = \int_{1-\alpha}^{1} H_Z^{-1}(\tau) \, d\tau - \alpha \mathbb{E}[Z].$$

The right derivative of the right hand side with respect to α equals $H_Z^{-1}(1-\alpha) - \mathbb{E}[Z]$. As it is nonincreasing, the function $\alpha \mapsto \mathfrak{q}_{1-\alpha}[Z]$ is concave. Moreover, its maximum is achieved at α^* for which $\mathbb{E}[Z]$ is the $(1-\alpha^*)$ -quantile of Z. Substituting the minimizer $t^* = \mathbb{E}[Z]$ into (6.23) we conclude that

$$\mathsf{AV}@\mathsf{R}_{\alpha^*}(Z) = \mathbb{E}[Z] + \frac{1}{\alpha^*} \sigma_1^+[Z].$$

Comparison with (6.27) yields the first equality in (6.29). To prove the second equality we recall relation (6.16) and note that

$$\max((1-\alpha)(t-Z), \alpha(Z-t)) = (1-\alpha)[t-Z]_{+} + \alpha[Z-t]_{+}$$

Thus

$$\sigma_1^+[Z] = \max_{\alpha \in [0,1]} \min_{t \in \mathbb{R}} \mathbb{E} \left\{ (1-\alpha)[t-Z]_+ + \alpha [Z-t]_+ \right\}$$

As the function under the "max-min" operation is linear with respect to $\alpha \in [0, 1]$ and convex with respect to t, the "max" and "min" operations can be exchanged. This proves the second equality in (6.29).

It also follows from (6.27) that the minimization problem (6.17) can be equivalently written as follows

$$\min_{x \in \mathcal{X}} \mathbb{E}[Z_x] + c \mathfrak{q}_{1-\alpha}[Z_x] = \min_{x \in \mathcal{X}} (1 - c\alpha) \mathbb{E}[Z_x] + c\alpha \operatorname{AV}_{@} \mathsf{R}_{\alpha}[Z_x]$$

$$= \min_{x \in \mathcal{X}, t \in \mathbb{R}} \mathbb{E}\left[(1 - c\alpha) Z_x + c \left(\alpha t + [Z_x - t]_+ \right) \right].$$
(6.30)

Both x and t are variables in this problem. We conclude that for this specific mean-risk model, an equivalent expected value formulation has been found. If $c \in [0, \alpha^{-1}]$ and the function $x \mapsto Z_x$ is convex, problem (6.30) is convex.

The maximization problem (6.18) can be equivalently written as follows:

$$\max_{x \in \mathcal{X}} \mathbb{E}[Z_x] - c \, \mathfrak{q}_{\alpha}[Z_x] = -\min_{x \in \mathcal{X}} \mathbb{E}[-Z_x] + c \, \mathfrak{q}_{1-\alpha}[-Z_x]$$

$$= -\min_{x \in \mathcal{X}, t \in \mathbb{R}} \mathbb{E}\left[-(1-c\alpha)Z_x + c\left(\alpha t + [-Z_x - t]_+\right)\right]$$

$$= \max_{x \in \mathcal{X}, t \in \mathbb{R}} \mathbb{E}\left[(1-c\alpha)Z_x + c\left(\alpha t - [t-Z_x]_+\right)\right].$$
(6.31)
(6.32)

In the last problem we replaced t by-t to stress the similarity with (6.30). Again, if $c \in [0, \alpha^{-1}]$ and the function $x \mapsto Z_x(\omega)$ is convex for a.e. $\omega \in \Omega$, problem (6.31) is convex.

6.3 Coherent Risk Measures

Let (Ω, \mathcal{F}) be a sample space, equipped with the sigma algebra \mathcal{F} , on which uncertain outcomes (random functions $Z = Z(\omega)$) are defined. By a *risk measure* we understand a function⁵ $\rho(Z)$ which maps Z into the extended real line $\mathbb{R} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$.

⁵The concept of risk measure should not be confused with the concept of probability measure.

In order to make this concept precise we need to define a space \mathcal{Z} of allowable random functions $Z(\omega)$ for which $\rho(Z)$ is defined. It seems that a natural choice of \mathcal{Z} will be the space of all \mathcal{F} -measurable functions $Z: \Omega \to \mathbb{R}$. However, typically, this space is too large for development of a meaningful theory (we will discuss this later, see Proposition 6.31 in particular). Unless stated otherwise, we deal in this chapter with spaces $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, where $p \in [1, \infty)$ (see section 7.3 for an introduction to these spaces). By assuming that $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$, we assume that the random variable Z has a finite p-th order moment with respect to the reference probability measure P. Also by considering a function ρ defined on the space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$, it is implicitly assumed that actually ρ is defined on classes of functions which may differ on sets of P-measure zero, i.e., $\rho(Z) = \rho(Z')$ if $P\{\omega : Z(\omega) \neq Z'(\omega)\} = 0$. We also consider the space $\mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$ of essentially bounded functions.

We assume throughout this chapter that risk measures $\rho : \mathbb{Z} \to \overline{\mathbb{R}}$ are *proper*. That is, $\rho(Z) > -\infty$ for all $Z \in \mathbb{Z}$ and the domain

$$\operatorname{dom}(\rho) := \{ Z \in \mathcal{Z} : \rho(Z) < +\infty \}$$

is nonempty. For $Z, Z' \in \mathbb{Z}$ we denote by $Z \succeq Z'$ the pointwise partial order⁶ meaning $Z(\omega) \ge Z'(\omega)$ for *P*-almost all $\omega \in \Omega$. We also assume that the smaller the realizations of *Z*, the better; for example *Z* may represent a random cost.

We consider the following axioms associated with a risk measure ρ .

(R1) Convexity:

$$\rho(tZ + (1-t)Z') \le t\rho(Z) + (1-t)\rho(Z')$$

for all $Z, Z' \in \mathcal{Z}$ and all $t \in [0, 1]$.

- (**R2**) Monotonicity: If $Z, Z' \in \mathcal{Z}$ and $Z \succeq Z'$, then $\rho(Z) \ge \rho(Z')$.
- (R3) Translation Equivariance: If $a \in \mathbb{R}$ and $Z \in \mathcal{Z}$, then $\rho(Z + a) = \rho(Z) + a$.
- (**R4**) Positive Homogeneity: If t > 0 and $Z \in \mathcal{Z}$, then $\rho(tZ) = t\rho(Z)$.

Definition 6.4. It is said that risk measure ρ is convex if it satisfies the conditions (R1)–(R3); it is said that ρ is coherent if it satisfies the conditions (R1)–(R4).

An example of a coherent risk measure is the Average Value-at-Risk measure $\rho(Z) :=$ AV@R_{α}(Z) (further examples of risk measures will be discussed in section 6.3.2). It is natural to assume in this example that Z has a finite first order moment, i.e., to use $\mathcal{Z} := \mathcal{L}_1(\Omega, \mathcal{F}, P)$. For such space \mathcal{Z} in this example, $\rho(Z)$ is finite (real valued) for all $Z \in \mathcal{Z}$.

It follows from the convexity condition (R1) that $\rho(\frac{Z+Z'}{2}) \leq \frac{\rho(Z)+\rho(Z')}{2}$. Thus conditions (R1) and (R4) imply that

$$\rho(Z+Z') \le \rho(Z) + \rho(Z'), \quad Z, Z' \in \mathcal{Z}.$$
(6.33)

⁶This partial order corresponds to the cone $C := \mathcal{L}_p^+(\Omega, \mathcal{F}, P)$, see the discussion of section 7.3, page 495, following equation (7.274).

Conversely, the positive homogeneity condition (R4) and the subadditivity condition (6.33) imply convexity of $\rho(\cdot)$. Therefore condition (R1) in the definition of a coherent risk measure ρ can be replaced by condition (6.33).

If the random outcome represents a reward, i.e., larger realizations of Z are preferred, we can define a risk measure $\rho(Z) = \rho(-Z)$, where ρ satisfies axioms (R1)–(R4). In this case, the function ρ also satisfies (R1) and (R4). The axioms (R2) and (R3) change to:

(**R2a**) Monotonicity: If $Z, Z' \in \mathcal{Z}$ and $Z \succeq Z'$, then $\varrho(Z) \leq \varrho(Z')$.

(R3a) Translation Equivariance: If $a \in \mathbb{R}$ and $Z \in \mathcal{Z}$, then $\varrho(Z + a) = \varrho(Z) - a$.

All our considerations regarding risk measures satisfying (R1)–(R4) have their obvious counterparts for risk measures satisfying (R1)-(R2a)-(R3a)-(R4).

With each space $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$, is associated its dual space $\mathcal{Z}^* := \mathcal{L}_q(\Omega, \mathcal{F}, P)$, where $q \in (1, \infty]$ is such that 1/p + 1/q = 1.

It is convenient to consider a bilinear form $\langle \cdot, \cdot \rangle$ on the product $\mathcal{Z} \times \mathcal{Z}^*$. For $Z \in \mathcal{Z}$ and $\zeta \in \mathcal{Z}^*$ the value of the bilinear form is defined as follows:

$$\langle \zeta, Z \rangle := \int_{\Omega} \zeta(\omega) Z(\omega) dP(\omega).$$
 (6.34)

Recall that the conjugate function $\rho^* : \mathbb{Z}^* \to \overline{\mathbb{R}}$ of a risk measure ρ is defined as

$$\rho^*(\zeta) := \sup_{Z \in \mathcal{Z}} \left\{ \langle \zeta, Z \rangle - \rho(Z) \right\},\tag{6.35}$$

and the conjugate of ρ^* (the biconjugate function) as

$$\rho^{**}(Z) := \sup_{\zeta \in \mathcal{Z}^*} \left\{ \langle \zeta, Z \rangle - \rho^*(\zeta) \right\}.$$
(6.36)

By the Fenchel-Moreau Theorem (Theorem 7.82) we have that if $\rho : \mathbb{Z} \to \mathbb{R}$ is convex, proper and lower semicontinuous, then $\rho^{**} = \rho$, i.e., $\rho(\cdot)$ has the representation

$$\rho(Z) = \sup_{\zeta \in \mathcal{Z}^*} \left\{ \langle \zeta, Z \rangle - \rho^*(\zeta) \right\}, \ \forall Z \in \mathcal{Z}.$$
(6.37)

Conversely, if the representation (6.37) holds for some proper function $\rho^* : \mathbb{Z}^* \to \mathbb{R}$, then ρ is convex and lower semicontinuous. Note that if ρ is convex, proper and lower semicontinuous, then its conjugate function ρ^* is also proper. Clearly, we can write the representation (6.37) in the following equivalent form

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \left\{ \langle \zeta, Z \rangle - \rho^*(\zeta) \right\}, \quad \forall Z \in \mathcal{Z},$$
(6.38)

where $\mathfrak{A} := \operatorname{dom}(\rho^*)$ is the domain of the conjugate function ρ^* .

The following basic duality result for convex risk measures is a direct consequence of the Fenchel-Moreau Theorem.

⁷Recall that unless stated otherwise it is assumed that $p \in [1, \infty)$. For p = 1 we have that $q = \infty$, i.e, the dual space of $\mathcal{L}_1(\Omega, \mathcal{F}, P)$ is the space $\mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$.

Theorem 6.5. Suppose that $\rho : \mathbb{Z} \to \mathbb{R}$ is convex, proper and lower semicontinuous. Then the representation (6.38) holds with $\mathfrak{A} := \operatorname{dom}(\rho^*)$. Moreover, we have that: (i) Condition (R2) holds iff every $\zeta \in \mathfrak{A}$ is nonnegative, i.e., $\zeta(\omega) \ge 0$ for a.e. $\omega \in \Omega$; (ii) Condition (R3) holds iff $\int_{\Omega} \zeta dP = 1$ for every $\zeta \in \mathfrak{A}$; (iii) Condition (R4) holds iff $\rho(\cdot)$ is the support function of the set \mathfrak{A} , i.e., it can be represented in the form

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \langle \zeta, Z \rangle, \quad \forall Z \in \mathcal{Z}.$$
(6.39)

Proof. If $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is convex, proper, and lower semicontinuous, then representation (6.38) is valid by virtue of the Fenchel-Moreau Theorem (Theorem 7.82).

Now suppose that assumption (R2) holds. It follows that $\rho^*(\zeta) = +\infty$ for every $\zeta \in \mathbb{Z}^*$ which is not nonnegative. Indeed, if $\zeta \in \mathbb{Z}^*$ is not nonnegative, then there exists a set $\Delta \in \mathcal{F}$ of positive measure such that $\zeta(\omega) < 0$ for all $\omega \in \Delta$. Consequently, for $\overline{Z} := \mathbf{1}_{\Delta}$ we have that $\langle \zeta, \overline{Z} \rangle < 0$. Take any Z in the domain of ρ , i.e., such that $\rho(Z)$ is finite, and consider $Z_t := Z - t\overline{Z}$. Then for $t \ge 0$, we have that $Z \succeq Z_t$, and assumption (R2) implies that $\rho(Z) \ge \rho(Z_t)$. Consequently,

$$\rho^*(\zeta) \ge \sup_{t \in \mathbb{R}_+} \left\{ \langle \zeta, Z_t \rangle - \rho(Z_t) \right\} \ge \sup_{t \in \mathbb{R}_+} \left\{ \langle \zeta, Z \rangle - t \langle \zeta, \bar{Z} \rangle - \rho(Z) \right\} = +\infty.$$

Conversely, suppose that every $\zeta \in \mathfrak{A}$ is nonnegative. Then for every $\zeta \in \mathfrak{A}$ and $Z \succeq Z'$, we have that $\langle \zeta, Z' \rangle \ge \langle \zeta, Z \rangle$. By (6.38), this implies that if $Z \succeq Z'$, then $\rho(Z) \ge \rho(Z')$. This completes the proof of assertion (i).

Suppose that assumption (R3) holds. Then for every $Z \in dom(\rho)$ we have

$$\rho^*(\zeta) \ge \sup_{a \in \mathbb{R}} \left\{ \langle \zeta, Z + a \rangle - \rho(Z + a) \right\} = \sup_{a \in \mathbb{R}} \left\{ a \int_{\Omega} \zeta dP - a + \langle \zeta, Z \rangle - \rho(Z) \right\}.$$

It follows that $\rho^*(\zeta) = +\infty$ for any $\zeta \in \mathbb{Z}^*$ such that $\int_{\Omega} \zeta dP \neq 1$. Conversely, if $\int_{\Omega} \zeta dP = 1$, then $\langle \zeta, Z + a \rangle = \langle \zeta, Z \rangle + a$, and hence condition (R3) follows by (6.38). This completes the proof of (ii).

Clearly, if (6.39) holds, then ρ is positively homogeneous. Conversely, if ρ is positively homogeneous, then its conjugate function is the indicator function of a convex subset of \mathcal{Z}^* . Consequently, the representation (6.39) follows by (6.38).

It follows from the above theorem that if ρ is a convex risk measure (i.e., satisfies conditions (R1)–(R3)) and is proper and lower semicontinuous, then the representation (6.38) holds with \mathfrak{A} being a subset of the set of probability density functions,

$$\mathfrak{P} := \left\{ \zeta \in \mathcal{Z}^* : \int_{\Omega} \zeta(\omega) dP(\omega) = 1, \ \zeta \succeq 0 \right\}.$$
(6.40)

If, moreover, ρ is positively homogeneous (i.e., condition (R4) holds), then its conjugate ρ^* is the indicator function of a convex set $\mathfrak{A} \subset \mathcal{Z}^*$, and \mathfrak{A} is equal to the subdifferential $\partial \rho(0)$ of ρ at $0 \in \mathcal{Z}$. Furthermore, $\rho(0) = 0$ and hence by the definition of $\partial \rho(0)$ we have that

$$\mathfrak{A} = \{ \zeta \in \mathfrak{P} : \langle \zeta, Z \rangle \le \rho(Z), \ \forall Z \in \mathcal{Z} \}.$$
(6.41)

The set \mathfrak{A} is weakly^{*} closed. Recall that if the space \mathcal{Z} , and hence \mathcal{Z}^* , is reflexive, then a convex subset of \mathcal{Z}^* is closed in the weak^{*} topology of \mathcal{Z}^* iff it is closed in the strong (norm) topology of \mathcal{Z}^* . If ρ is positively homogeneous and continuous, then $\mathfrak{A} = \partial \rho(0)$ is a *bounded* (and weakly^{*} compact) subset of \mathcal{Z}^* (see Proposition 7.85). We refer to the set \mathfrak{A} as the *dual* set of the corresponding coherent risk measure ρ .

We have that if ρ is a *coherent* risk measure, then the corresponding set \mathfrak{A} is a set of probability density functions. Thus for any $\zeta \in \mathfrak{A}$ we can view $\langle \zeta, Z \rangle$ as the expectation $\mathbb{E}_Q[Z]$ taken with respect to the probability measure $dQ = \zeta dP$, defined by the density ζ , i.e., Q is absolutely continuous with respect to P and $dQ/dP = \zeta$. Consequently, the representation (6.39) can be written in the form

$$\rho(Z) = \sup_{\frac{dQ}{dP} \in \mathfrak{A}} \mathbb{E}_Q[Z], \quad \forall Z \in \mathcal{Z}.$$
(6.42)

Definition of a risk measure ρ depends on a particular choice of the corresponding space Z. In many cases there is a natural choice of Z which ensures that $\rho(Z)$ is finite valued for all $Z \in Z$. We shall see such examples in section 6.3.2. By Theorem 7.91 we have the following result which shows that for real valued convex and monotone risk measures, the assumption of lower semicontinuity in Theorem 6.5 holds automatically.

Proposition 6.6. Let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, with $p \in [1, \infty]$, and $\rho : \mathcal{Z} \to \mathbb{R}$ be a real valued risk measure satisfying conditions (R1) and (R2). Then ρ is continuous and subdifferentiable on \mathcal{Z} .

Theorem 6.5 together with Proposition 6.6 imply the following basic duality result.

Theorem 6.7. Let $\rho : \mathbb{Z} \to \mathbb{R}$, where $\mathbb{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ with $p \in [1, \infty)$. Then ρ is a real valued coherent risk measure iff there exists a convex bounded and weakly^{*} closed set $\mathfrak{A} \subset \mathfrak{P}$ such that the representation (6.39) holds. In the later case, for every $Z \in \mathbb{Z}$ the maximum in (6.39) is attained at some point of \mathfrak{A} .

Proof. If $\rho : \mathbb{Z} \to \mathbb{R}$ is a real valued coherent risk measure, then by Proposition 6.6 it is continuous, and hence by Theorem 6.5 the representation (6.39) holds with $\mathfrak{A} = \partial \rho(0)$. Moreover, the subdifferential of a convex continuous function is bounded and weakly^{*} closed (and hence is weakly^{*} compact).

Conversely, if the representation (6.39) holds with the set \mathfrak{A} being a convex subset of \mathfrak{P} and weakly* compact, then ρ is real valued and satisfies conditions (R1)–(R4). It follows by compactness of \mathfrak{A} that the maximum in (6.39) is attained for every $Z \in \mathcal{Z}$.

The following result shows that if ρ is a proper convex risk measure, then either it is finite valued and continuous on Z or it takes value $+\infty$ on a dense subset of Z. Recall that risk measure ρ is said to be *convex* if it satisfies conditions (R1),(R2) and (R3).

Proposition 6.8. Let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, with $p \in [1, \infty)$, and $\rho : \mathcal{Z} \to \mathbb{R}$ be a proper convex risk measure. Suppose that the domain of ρ has a nonempty interior. Then ρ is finite valued and continuous on \mathcal{Z} .

Proof. By Theorem 7.91 we have that ρ is continuous on the interior of its domain. Consider a point Z_0 in the interior of dom (ρ) . Since ρ is continuous at Z_0 it follows that there exists r > 0 such that $|\rho(Z) - \rho(Z_0)| < 1$ for all Z such that $||Z - Z_0|| \le r$. By changing variables $Z \mapsto Z - Z_0$, we can assume without loss of generality that $Z_0 = 0$.

Now let us observe that for any $Z \in \mathcal{Z}$ and $Z_c^+(\omega) := [Z(\omega) - c]_+, c \in \mathbb{R}$, it holds that

$$\lim_{c \to +\infty} \left\| Z_c^+ \right\|^p = \lim_{c \to +\infty} \int_{Z(\omega) \ge c} |Z(\omega) - c|^p dP(\omega) = 0.$$

Therefore there exists $c \in \mathbb{R}$ (depending on Z) such that $\left\| Z_c^+ \right\| < r$, and hence

$$\rho(Z_c^+) \le \rho(0) + 1 < +\infty.$$

Noting that $Z \leq c + Z_c^+$, we can write

$$\rho(Z) \le \rho(Z_c^+ + c) = \rho(Z_c^+) + c < +\infty$$

This shows that $\rho(Z)$ is finite valued for any $Z \in \mathcal{Z}$. Continuity of $\rho(\cdot)$ follows by Proposition 6.6. \Box

Remark 25. The analysis simplifies considerably if the space Ω is finite, say $\Omega := \{\omega_1, \ldots, \omega_n\}$ equipped with sigma algebra of all subsets of Ω and respective (positive) probabilities p_1, \ldots, p_n . Then every function $Z : \Omega \to \mathbb{R}$ is measurable and the space Z of all such functions can be identified with \mathbb{R}^n by identifying $Z \in Z$ with the vector $(Z(\omega_1), \ldots, Z(\omega_n)) \in \mathbb{R}^n$. The dual of the space \mathbb{R}^n can be identified with itself by using the standard scalar product in \mathbb{R}^n , and the set \mathfrak{P} becomes

$$\mathfrak{P} = \left\{ \zeta \in \mathbb{R}^n : \sum_{i=1}^n p_i \zeta_i = 1, \ \zeta \ge 0 \right\}.$$
(6.43)

The above set \mathfrak{P} forms a convex bounded subset of \mathbb{R}^n , and hence the set \mathfrak{A} is also bounded.

Spaces of Essentially Bounded Variables

Consider now the space $\mathcal{Z} := \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$ of essentially bounded random variables $Z : \Omega \to \mathbb{R}$. Let $\rho : \mathcal{Z} \to \mathbb{R}$ be a convex risk measure having a finite value in at least one point $\overline{Z} \in \mathcal{Z}$. For a random variable $Z \in \mathcal{Z}$ we have that $Z \preceq \overline{Z} + c$ and $\overline{Z} - c \preceq Z$ for sufficiently large $c \in \mathbb{R}$ (this follows because of the essential boundedness of Z). Hence by conditions (R2) and (R3) it follows that $\rho(Z) \le \rho(\overline{Z}) + c < +\infty$ and $-\infty < \rho(\overline{Z}) - c \le \rho(Z)$. This shows that here any convex risk measure ρ having a finite value in at least one point, is finite valued.

Thus by Proposition 6.6 it follows that ρ is continuous in the norm (strong) topology of $\mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$. In fact ρ is Lipschitz continuous. Indeed, for any $Z_1, Z_2 \in \mathcal{Z}$ we have that $Z_1 \leq Z_2 + ||Z_1 - Z_2||_{\infty}$, and hence it follows by (R2) and (R3) that $\rho(Z_1) \leq \rho(Z_2) + ||Z_1 - Z_2||_{\infty}$. Similarly $\rho(Z_2) \leq \rho(Z_1) + ||Z_2 - Z_1||_{\infty}$ and hence

$$|\rho(Z_1) - \rho(Z_2)| \le ||Z_1 - Z_2||_{\infty}.$$
(6.44)

It follows that any coherent risk measure ρ , defined on the space $\mathcal{Z} = \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$, has the dual representation (6.39) for some $\mathfrak{A} \subset \mathcal{Z}^*$. However, the dual \mathcal{Z}^* of the space $\mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$ is inconvenient to work with (the dual space \mathcal{Z}^* is formed by finitely additive measures). Therefore a general practice is to pair space $\mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$ with the space $\mathcal{L}_1(\Omega, \mathcal{F}, P)$, rather than \mathcal{Z}^* . That is, $\mathcal{L}_1(\Omega, \mathcal{F}, P)$ and $\mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$ are viewed as paired locally convex topological vector spaces with respect to the bilinear form

$$\langle \zeta, Z \rangle := \int_{\Omega} \zeta(\omega) Z(\omega) dP(\omega), \ \zeta \in \mathcal{L}_1(\Omega, \mathcal{F}, P), \ Z \in \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P).$$
(6.45)

The space $\mathcal{L}_1(\Omega, \mathcal{F}, P)$ is equipped with its strong (norm) topology and the space $\mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$ is equipped with its weak* topology (see section 7.3.2).

In order to ensure that a convex proper function $\rho : \mathbb{Z} \to \mathbb{R}$ has the dual representation of the form (6.37):

$$\rho(Z) = \sup_{\zeta \in \mathcal{L}_1(\Omega, \mathcal{F}, P)} \left\{ \langle \zeta, Z \rangle - \rho^*(\zeta) \right\}, \ \forall Z \in \mathcal{Z},$$
(6.46)

with respect to pairing of $\mathcal{Z} = \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$ with $\mathcal{L}_1(\Omega, \mathcal{F}, P)$, we need the additional condition for ρ to be lower semicontinuous with respect to the weak* topology rather than the norm (strong) topology of \mathcal{Z} . Consider the following probability space.

Definition 6.9. We say that probability space (Ω, \mathcal{F}, P) is standard uniform if $\Omega = [0, 1]$ equipped with its Borel sigma algebra and uniform probability measure P.

Of course, the standard uniform probability space is nonatomic. Consider the essential supremum risk measure $\rho(Z) := \operatorname{ess\,sup}(Z)$. This is a coherent risk measure $\rho : \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P) \to \mathbb{R}$. It has the dual representation (of the form (6.39))

$$o(Z) = \sup_{\zeta \in \mathfrak{A}} \langle \zeta, Z \rangle, \ Z \in \mathcal{Z},$$
(6.47)

with the dual set $\mathfrak{A} \subset \mathcal{L}_1(\Omega, \mathcal{F}, P)$ being the set of all density functions, i.e., $\mathfrak{A} = \mathfrak{P}$, where (as before)

$$\mathfrak{P} := \left\{ \zeta \in \mathcal{L}_1(\Omega, \mathcal{F}, P) : \int_{\Omega} \zeta dP = 1, \ \zeta \succeq 0 \right\}.$$

Thus ρ can be represented as maximum of weakly^{*} continuous (linear) functions, and hence ρ is lower semicontinuous with respect to the weak^{*} topology of $\mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$ (see Proposition 7.24). Note, however that ρ is not continuous in the weak^{*} topology of $\mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$.

Although the dual representation (6.47) holds for this ρ , the maximum in the right hand side of (6.47) may be not attained. For example, let $Z : [0,1] \to \mathbb{R}$ be a continuous function attaining its maximum at a unique point $\bar{\omega} \in [0,1]$. For such $Z \in \mathbb{Z}$ the maximum in (6.47) is not attained.

6.3.1 Differentiability Properties of Risk Measures

Recall that unless stated otherwise we assume that $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$, with $p \in [1, \infty)$. Let $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ be a convex proper lower semicontinuous risk measure. By convexity and lower semicontinuity of ρ we have that $\rho^{**} = \rho$, and hence by Proposition 7.84 that

$$\partial \rho(Z) = \operatorname*{argmax}_{\zeta \in \mathfrak{A}} \left\{ \langle \zeta, Z \rangle - \rho^*(\zeta) \right\}, \tag{6.48}$$

provided that $\rho(Z)$ is finite. If, moreover, ρ is positively homogeneous, then $\mathfrak{A} = \partial \rho(0)$ and

$$\partial \rho(Z) = \underset{\zeta \in \mathfrak{A}}{\operatorname{argmax}} \langle \zeta, Z \rangle. \tag{6.49}$$

The conditions (R1)–(R3) imply that \mathfrak{A} is a subset of the set \mathfrak{P} of probability density functions. Consequently, under conditions (R1)–(R3), $\partial \rho(Z)$ is a subset of \mathfrak{P} as well.

We can summarize this as follows.

Theorem 6.10. Let $\rho : \mathbb{Z} \to \mathbb{R}$ be a convex proper lower semicontinuous risk measure, which is finite valued and continuous at a point $Z \in \mathbb{Z}$. Then the following holds: (i) the subdifferential $\partial \rho(Z)$ is a nonempty bounded and weakly* compact subset of \mathbb{Z}^* , (ii) the subdifferential $\partial \rho(Z)$ is given by (6.48), and if furthermore ρ is positively homogeneous, then the subdifferential $\partial \rho(Z)$ is given by (6.49), (iii) ρ is Hadamard directionally differentiable at Z and

$$\rho'(Z,H) = \sup_{\zeta \in \partial \rho(Z)} \langle \zeta, H \rangle, \ \forall H \in \mathcal{Z}.$$
(6.50)

(iv) if, moreover, $\partial \rho(Z)$ is a singleton, i.e., $\partial \rho(Z) = \{\nabla \rho(Z)\}$ for some $\nabla \rho(Z) \in \mathbb{Z}^*$, then ρ is Hadamard differentiable at Z and

$$\rho'(Z, \cdot) = \langle \nabla \rho(Z), \cdot \rangle. \tag{6.51}$$

Remark 26. Let $\rho : \mathbb{Z} \to \mathbb{R}$ be a (real valued) coherent risk measure. By Proposition 6.6 we have that ρ is continuous and hence formula (6.49) applies for every $Z \in \mathbb{Z}$. It follows that $\partial \rho(Z) = \{\overline{\zeta}\}$ is a singleton iff Z exposes \mathfrak{A} at $\overline{\zeta}$. Thus, by Theorem 7.81 a dense subset \mathcal{D} of \mathbb{Z} exists such that $\partial \rho(Z)$ is a singleton, and hence ρ is Hadamard differentiable, at every $Z \in \mathcal{D}$.

We often have to deal with composite functions $\rho \circ F : \mathbb{R}^n \to \overline{\mathbb{R}}$, where $F : \mathbb{R}^n \to \mathcal{Z}$ is a mapping. We write $f(x, \omega)$, or $f_{\omega}(x)$, for $[F(x)](\omega)$, and view $f(x, \omega)$ as a random function defined on the measurable space (Ω, \mathcal{F}) . We say that the mapping F is *convex* if the function $f(\cdot, \omega)$ is convex for every $\omega \in \Omega$.

Proposition 6.11. If the mapping $F : \mathbb{R}^n \to \mathcal{Z}$ is convex and $\rho : \mathcal{Z} \to \mathbb{R}$ satisfies conditions (R1) and (R2), then the composite function $\phi(\cdot) := \rho(F(\cdot))$ is convex.

Proof. For any $x, x' \in \mathbb{R}^n$ and $t \in [0, 1]$, we have by convexity of $F(\cdot)$ and monotonicity of $\rho(\cdot)$ that

$$\rho(F(tx + (1 - t)x')) \le \rho(tF(x) + (1 - t)F(x')).$$

Hence convexity of $\rho(\cdot)$ implies that

$$\rho(F(tx + (1 - t)x')) \le t\rho(F(x)) + (1 - t)\rho(F(x')).$$

This proves convexity of $\rho(F(\cdot))$.

It should be noted that the monotonicity condition (R2) was essential in the above derivation of convexity of the composite function.

Let us discuss differentiability properties of the composite function $\phi = \rho \circ F$ at a point $\bar{x} \in \mathbb{R}^n$. As before, we assume that $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P), p \in [1, \infty)$. The mapping $F : \mathbb{R}^n \to \mathcal{Z}$ maps a point $x \in \mathbb{R}^n$ into a real valued function (or rather a class of functions which may differ on sets of *P*-measure zero) $[F(x)](\cdot)$ on Ω , also denoted $f(x, \cdot)$, which is an element of $\mathcal{L}_p(\Omega, \mathcal{F}, P)$. If *F* is convex, then $f(\cdot, \omega)$ is convex real valued and hence is continuous and has (finite valued) directional derivatives at \bar{x} , denoted $f'_{\omega}(\bar{x}, h)$. These properties are inherited by the mapping *F*.

Lemma 6.12. Let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$, and $F : \mathbb{R}^n \to \mathcal{Z}$ be a convex mapping. Then F is continuous, directionally differentiable and its directional derivative at a point $\bar{x} \in \mathbb{R}^n$ is given by

$$[F'(\bar{x},h)](\omega) = f'_{\omega}(\bar{x},h), \ \omega \in \Omega, \ h \in \mathbb{R}^n.$$
(6.52)

Proof. In order to show continuity of F we need to verify that, for an arbitrary point $\bar{x} \in \mathbb{R}^n$, $||F(x) - F(\bar{x})||_p$ tends to zero as $x \to \bar{x}$. By the Lebesgue Dominated Convergence Theorem and continuity of $f(\cdot, \omega)$ we can write that

$$\lim_{x \to \bar{x}} \int_{\Omega} \left| f(x,\omega) - f(\bar{x},\omega) \right|^p dP(\omega) = \int_{\Omega} \lim_{x \to \bar{x}} \left| f(x,\omega) - f(\bar{x},\omega) \right|^p dP(\omega) = 0, \quad (6.53)$$

provided that there exists a neighborhood $U \subset \mathbb{R}^n$ of \bar{x} such that the family $\{|f(x,\omega) - f(\bar{x},\omega)|^p\}_{x\in U}$ is dominated by a *P*-integrable function, or equivalently that $\{|f(x,\omega) - f(\bar{x},\omega)|\}_{x\in U}$ is dominated by a function from the space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$. Since $f(\bar{x}, \cdot)$ belongs to $\mathcal{L}_p(\Omega, \mathcal{F}, P)$, it suffices to verify this dominance condition for $\{|f(x,\omega)|\}_{x\in U}$.

Now let $x_1, \ldots, x_{n+1} \in \mathbb{R}^n$ be such points that the set $U := \operatorname{conv}\{x_1, \ldots, x_{n+1}\}$ forms a neighborhood of the point \bar{x} , and let $g(\omega) := \max\{f(x_1, \omega), \ldots, f(x_{n+1}, \omega)\}$. By convexity of $f(\cdot, \omega)$ we have that $f(x, \cdot) \leq g(\cdot)$ for all $x \in U$. Also since every $f(x_i, \cdot)$, $i = 1, \ldots, n+1$, is an element of $\mathcal{L}_p(\Omega, \mathcal{F}, P)$, we have that $g \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ as well. That is, $g(\omega)$ gives an upper bound for $\{|f(x, \omega)|\}_{x \in U}$. Also by convexity of $f(\cdot, \omega)$ we have that

$$f(x,\omega) \ge 2f(\bar{x},\omega) - f(2\bar{x} - x,\omega).$$

By shrinking the neighborhood U if necessary, we can assume that U is symmetrical around \bar{x} , i.e., if $x \in U$, then $2\bar{x} - x \in U$. Consequently, we have that $\tilde{g}(\omega) := 2f(\bar{x}, \omega) - g(w)$ gives a lower bound for $\{|f(x, \omega)|\}_{x \in U}$, and $\tilde{g} \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$. This shows that the required dominance condition holds and hence F is continuous at \bar{x} by (6.53).

Now for $h \in \mathbb{R}^n$ and t > 0 denote

$$R_t(\omega) := t^{-1} \left[f(\bar{x} + th, \omega) - f(\bar{x}, \omega) \right] \text{ and } Z(\omega) := f'_{\omega}(\bar{x}, h), \ \omega \in \Omega.$$

Note that $f(\bar{x} + th, \cdot)$ and $f(\bar{x}, \cdot)$ are elements of the space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$, and hence $R_t(\cdot)$ is also an element of $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ for any t > 0. Since for a.e. $\omega \in \Omega$, $f(\cdot, \omega)$ is convex

real valued, we have that $R_t(\omega)$ is monotonically nonincreasing and converges to $Z(\omega)$ as $t \downarrow 0$. Therefore we have that $R_t(\cdot) \ge Z(\cdot)$ for any t > 0. Again by convexity of $f(\cdot, \omega)$, we have that for t > 0,

$$Z(\cdot) \ge t^{-1} \left[f(\bar{x}, \cdot) - f(\bar{x} - th, \cdot) \right].$$

We obtain that $Z(\cdot)$ is bounded from above and below by functions which are elements of the space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$, and hence $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ as well.

We have that $R_t(\cdot) - Z(\cdot)$, and hence $|R_t(\cdot) - Z(\cdot)|^p$, are monotonically decreasing to zero as $t \downarrow 0$ and for any t > 0, $\mathbb{E}[|R_t - Z|^p]$ is finite. It follows by the Monotone Convergence Theorem that $\mathbb{E}[|R_t - Z|^p]$ tends to zero as $t \downarrow 0$. That is, R_t converges to Z in the norm topology of \mathcal{Z} . Since $R_t = t^{-1}[F(\bar{x} + th) - F(\bar{x})]$, this shows that F is directionally differentiable at \bar{x} and formula (6.52) follows. \Box

The following theorem can be viewed as an extension of Theorem 7.51 where a similar result is derived for $\rho(\cdot) := \mathbb{E}[\cdot]$.

Theorem 6.13. Let $Z := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$, and $F : \mathbb{R}^n \to Z$ be a convex mapping. Suppose that ρ is convex, finite valued and continuous at $\overline{Z} := F(\overline{x})$. Then the composite function $\phi = \rho \circ F$ is directionally differentiable at \overline{x} , $\phi'(\overline{x}, h)$ is finite valued for every $h \in \mathbb{R}^n$ and

$$\phi'(\bar{x},h) = \sup_{\zeta \in \partial \rho(\bar{Z})} \int_{\Omega} f'_{\omega}(\bar{x},h)\zeta(\omega)dP(\omega).$$
(6.54)

Proof. Since ρ is continuous at \overline{Z} , it follows that ρ subdifferentiable and Hadamard directionally differentiable at \overline{Z} and formula (6.50) holds. Also by Lemma 6.12, mapping F is directionally differentiable. Consequently, we can apply the chain rule (see Proposition 7.66) to conclude that $\phi(\cdot)$ is directionally differentiable at \overline{x} , $\phi'(\overline{x}, h)$ is finite valued and

$$\phi'(\bar{x},h) = \rho'(Z, F'(\bar{x},h)). \tag{6.55}$$

Together with (6.50) and (6.52), the above formula (6.55) implies (6.54).

It is also possible to write formula (6.54) in terms of the corresponding subdifferentials.

Theorem 6.14. Let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$, and $F : \mathbb{R}^n \to \mathcal{Z}$ be a convex mapping. Suppose that ρ satisfies conditions (R1) and (R2) and is finite valued and continuous at $\overline{Z} := F(\overline{x})$. Then the composite function $\phi = \rho \circ F$ is subdifferentiable at \overline{x} and

$$\partial \phi(\bar{x}) = \operatorname{cl}\left(\bigcup_{\zeta \in \partial \rho(\bar{Z})} \int_{\Omega} \partial f_{\omega}(\bar{x})\zeta(\omega)dP(\omega)\right).$$
(6.56)

Proof. Since, by Lemma 6.12, F is continuous at \bar{x} and ρ is continuous at $F(\bar{x})$, we have that ϕ is continuous at \bar{x} , and hence $\phi(x)$ is finite valued for all x in a neighborhood of \bar{x} .

Moreover, by Proposition 6.11, $\phi(\cdot)$ is convex, and hence is continuous in a neighborhood of \bar{x} and is subdifferentiable at \bar{x} . Also formula (6.54) holds. It follows that $\phi'(\bar{x}, \cdot)$ is convex, continuous, positively homogeneous and

$$\phi'(\bar{x},\cdot) = \sup_{\zeta \in \partial \rho(\bar{Z})} \eta_{\zeta}(\cdot), \tag{6.57}$$

where

$$\eta_{\zeta}(\cdot) := \int_{\Omega} f'_{\omega}(\bar{x}, \cdot)\zeta(\omega)dP(\omega).$$
(6.58)

Because of the condition (R2) we have that every $\zeta \in \partial \rho(Z)$ is nonnegative. Consequently the corresponding function η_{ζ} is convex continuous and positively homogeneous, and hence is the support function of the set $\partial \eta_{\zeta}(0)$. The supremum of these functions, given by the right hand side of (6.57), is the support function of the set $\bigcup_{\zeta \in \partial \rho(\bar{Z})} \partial \eta_{\zeta}(0)$. Applying Theorem 7.52 and using the fact that the subdifferential of $f'_{\omega}(\bar{x}, \cdot)$ at $0 \in \mathbb{R}^n$ coincides with $\partial f_{\omega}(\bar{x})$, we obtain

$$\partial \eta_{\zeta}(0) = \int_{\Omega} \partial f_{\omega}(\bar{x}) \zeta(\omega) dP(\omega).$$
(6.59)

Since $\partial \rho(\bar{Z})$ is convex, it is straightforward to verify that the set $\cup_{\zeta \in \partial \rho(\bar{Z})} \partial \eta_{\zeta}(0)$ is also convex. Consequently it follows by (6.57) and (6.59) that the subdifferential of $\phi'(\bar{x}, \cdot)$ at $0 \in \mathbb{R}^n$ is equal to the topological closure of the set $\cup_{\zeta \in \partial \rho(\bar{Z})} \partial \eta_{\zeta}(0)$, i.e., is given by the right hand side of (6.56). It remains to note that the subdifferential of $\phi'(\bar{x}, \cdot)$ at $0 \in \mathbb{R}^n$ coincides with $\partial \phi(\bar{x})$. \Box

Under the assumptions of the above theorem we have that the composite function ϕ is convex and is continuous (in fact even Lipschitz continuous) in a neighborhood of \bar{x} . It follows that ϕ is differentiable⁸ at \bar{x} iff $\partial \phi(\bar{x})$ is a singleton. This leads to the following result, where for $\zeta \succeq 0$, we say that a property holds for ζ -a.e. $\omega \in \Omega$ if the set of points $A \in \mathcal{F}$ where it does not hold has ζdP measure zero, i.e., $\int_A \zeta(\omega) dP(\omega) = 0$. Of course, if P(A) = 0, then $\int_A \zeta(\omega) dP(\omega) = 0$. That is, if a property holds for a.e. $\omega \in \Omega$ with respect to P, then it holds for ζ -a.e. $\omega \in \Omega$.

Corollary 6.15. Let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$, and $F : \mathbb{R}^n \to \mathcal{Z}$ be a convex mapping. Suppose that ρ satisfies conditions (R1) and (R2) and is finite valued and continuous at $\overline{Z} := F(\overline{x})$. Then the composite function $\phi = \rho \circ F$ is differentiable at \overline{x} if and only if the following two properties hold: (i) for every $\zeta \in \partial \rho(\overline{Z})$ the function $f(\cdot, \omega)$ is differentiable at \overline{x} for ζ -a.e. $\omega \in \Omega$, (ii) $\int_{\Omega} \nabla f_{\omega}(\overline{x}) \zeta(\omega) dP(\omega)$ is the same for every $\zeta \in \partial \rho(\overline{Z})$.

In particular, if $\partial \rho(\overline{Z}) = {\overline{\zeta}}$ is a singleton, then ϕ is differentiable at \overline{x} if and only if $f(\cdot, \omega)$ is differentiable at \overline{x} for $\overline{\zeta}$ -a.e. $\omega \in \Omega$, in which case

$$\nabla\phi(\bar{x}) = \int_{\Omega} \nabla f_{\omega}(\bar{x})\bar{\zeta}(\omega)dP(\omega).$$
(6.60)

⁸Note that since $\phi(\cdot)$ is Lipschitz continuous near \bar{x} , the notions of Gâteaux and Fréchet differentiability at \bar{x} are equivalent here.

Proof. By Theorem 6.14 we have that ϕ is differentiable at \bar{x} iff the set on the right hand side of (6.56) is a singleton. Clearly this set is a singleton iff the set $\int_{\Omega} \partial f_{\omega}(\bar{x})\zeta(\omega)dP(\omega)$ is a singleton and is the same for every $\zeta \in \partial \rho(\bar{Z})$. Since $\partial f_{\omega}(\bar{x})$ is a singleton iff $f_{\omega}(\cdot)$ is differentiable at \bar{x} , in which case $\partial f_{\omega}(\bar{x}) = \{\nabla f_{\omega}(\bar{x})\}$, we obtain that ϕ is differentiable at \bar{x} iff conditions (i) and (ii) hold. The second assertion then follows. \Box

Of course, if the set inside the parentheses on the right hand side of (6.56) is closed, then there is no need to take its topological closure. This holds true in the following case.

Corollary 6.16. Suppose that the assumptions of Theorem 6.14 are satisfied and for every $\zeta \in \partial \rho(\overline{Z})$ the function $f_{\omega}(\cdot)$ is differentiable at \overline{x} for ζ -a.e. $\omega \in \Omega$. Then

$$\partial \phi(\bar{x}) = \bigcup_{\zeta \in \partial \rho(\bar{Z})} \int_{\Omega} \nabla f_{\omega}(\bar{x}) \zeta(\omega) dP(\omega).$$
(6.61)

Proof. In view of Theorem 6.14 we only need to show that the set on the right hand side of (6.61) is closed. As ρ is continuous at \overline{Z} , the set $\partial \rho(\overline{Z})$ is weakly^{*} compact. Also the mapping $\zeta \mapsto \int_{\Omega} \nabla f_{\omega}(\overline{x}) \zeta(\omega) dP(\omega)$, from \mathcal{Z}^* to \mathbb{R}^n , is continuous with respect to the weak^{*} topology of \mathcal{Z}^* and the standard topology of \mathbb{R}^n . It follows that the image of the set $\partial \rho(\overline{Z})$ by this mapping is compact and hence is closed, i.e., the set at the right hand side of (6.61) is closed. \Box

6.3.2 Examples of Risk Measures

In this section we discuss several examples of risk measures which are commonly used in applications. In the following examples it is natural to use the space $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ for an appropriate $p \in [1, \infty)$. Note that if a random variable Z has a p-th order finite moment, then it has finite moments of any order p' smaller than p, i.e., if $1 \le p' \le p$ and $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$, then $Z \in \mathcal{L}_{p'}(\Omega, \mathcal{F}, P)$. This gives a natural embedding of $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ into $\mathcal{L}_{p'}(\Omega, \mathcal{F}, P)$ for p' < p. Unless stated otherwise, all expectations and probabilistic statements will be made with respect to the probability measure P.

Before proceeding to particular examples let us consider the following construction. Let $\rho: \mathcal{Z} \to \overline{\mathbb{R}}$ and define

$$\tilde{\rho}(Z) := \mathbb{E}[Z] + \inf_{t \in \mathbb{R}} \rho(Z - t).$$
(6.62)

Clearly we have that for any $a \in \mathbb{R}$,

$$\tilde{\rho}(Z+a) = \mathbb{E}[Z+a] + \inf_{t \in \mathbb{R}} \rho(Z+a-t) = \mathbb{E}[Z] + a + \inf_{t \in \mathbb{R}} \rho(Z-t) = \tilde{\rho}(Z) + a.$$

That is, $\tilde{\rho}$ satisfies condition (R3) irrespectively whether ρ does. It is not difficult to see that if ρ satisfies conditions (R1) and (R2), then $\tilde{\rho}$ satisfies these conditions as well. Also if ρ is

positively homogeneous, then so is $\tilde{\rho}$. Let us calculate the conjugate of $\tilde{\rho}$. We have

$$\begin{split} \tilde{\rho}^*(\zeta) &= \sup_{Z \in \mathcal{Z}} \left\{ \langle \zeta, Z \rangle - \tilde{\rho}(Z) \right\} = \sup_{Z \in \mathcal{Z}} \left\{ \langle \zeta, Z \rangle - \mathbb{E}[Z] - \inf_{t \in \mathbb{R}} \rho(Z - t) \right\} \\ &= \sup_{Z \in \mathcal{Z}, t \in \mathbb{R}} \left\{ \langle \zeta, Z \rangle - \mathbb{E}[Z] - \rho(Z - t) \right\} \\ &= \sup_{Z \in \mathcal{Z}, t \in \mathbb{R}} \left\{ \langle \zeta - 1, Z \rangle + t(\mathbb{E}[\zeta] - 1) - \rho(Z) \right\}. \end{split}$$

It follows that

$$\tilde{\rho}^*(\zeta) = \begin{cases} \rho^*(\zeta - 1), & \text{if } \mathbb{E}[\zeta] = 1, \\ +\infty, & \text{if } \mathbb{E}[\zeta] \neq 1. \end{cases}$$

The construction below can be viewed as a homogenization of a risk measure ρ : $\mathcal{Z} \to \overline{\mathbb{R}}$. Define

$$\check{\rho}(Z) := \inf_{\tau > 0} \tau \rho(\tau^{-1}Z).$$
(6.63)

For any t > 0, by making change of variables $\tau \mapsto t\tau$, we obtain that $\check{\rho}(tZ) = t\check{\rho}(Z)$. That is, $\check{\rho}$ is positively homogeneous whether ρ is or isn't. Clearly, if ρ is positively homogeneous to start with, then $\rho = \check{\rho}$.

If ρ is convex, then so is $\check{\rho}$. Indeed, observe that if ρ is convex, then function $\varphi(\tau, Z) := \tau \rho(\tau^{-1}Z)$ is convex jointly in Z and $\tau > 0$. This can be verified directly as follows. For $t \in [0, 1]$, $\tau_1, \tau_2 > 0$ and $Z_1, Z_2 \in Z$, and setting $\tau := t\tau_1 + (1 - t)\tau_2$ and $Z := tZ_1 + (1 - t)Z_2$, we have

$$t[\tau_1\rho(\tau_1^{-1}Z_1)] + (1-t)[\tau_2\rho(\tau_2^{-1}Z_2)] = \tau \left[\frac{t\tau_1}{\tau}\rho(\tau_1^{-1}Z_1) + \frac{(1-t)\tau_2}{\tau}\rho(\tau_2^{-1}Z_2)\right]$$
$$\geq \tau \rho \left(\frac{t}{\tau}Z_1 + \frac{(1-t)}{\tau}Z_2\right) = \tau \rho(\tau^{-1}Z).$$

Minimizing convex function $\varphi(\tau, Z)$ over $\tau > 0$, we obtain a convex function. It is also not difficult to see that if ρ satisfies conditions (R2) and (R3), then so is $\check{\rho}$.

Let us calculate the conjugate of $\check{\rho}$. We have

$$\begin{split} \check{\rho}^*(\zeta) &= \sup_{Z \in \mathcal{Z}} \left\{ \langle \zeta, Z \rangle - \check{\rho}(Z) \right\} = \sup_{Z \in \mathcal{Z}, \tau > 0} \left\{ \langle \zeta, Z \rangle - \tau \rho(\tau^{-1}Z) \right\} \\ &= \sup_{Z \in \mathcal{Z}, \tau > 0} \left\{ \tau [\langle \zeta, Z \rangle - \rho(Z)] \right\}. \end{split}$$

It follows that $\check{\rho}^*$ is the indicator function of the set

$$\mathfrak{A} := \{ \zeta \in \mathcal{Z}^* : \langle \zeta, Z \rangle \le \rho(Z), \ \forall Z \in \mathcal{Z} \} .$$
(6.64)

If, moreover, $\check{\rho}$ is lower semicontinuous, then $\check{\rho}$ is equal to the conjugate of $\check{\rho}^*$, and hence $\check{\rho}$ is the support function of the above set \mathfrak{A} .

Example 6.17 (Utility model) It is possible to relate the theory of convex risk measures with the utility model. Let $g : \mathbb{R} \to \overline{\mathbb{R}}$ be a proper convex nondecreasing lower semicontinuous function such that the expectation $\mathbb{E}[g(Z)]$ is well defined for all $Z \in \mathcal{Z}$ (it is allowed

here for $\mathbb{E}[g(Z)]$ to take value $+\infty$, but not $-\infty$ since the corresponding risk measure is required to be proper). We can view the function g as a *disutility* function⁹.

Proposition 6.18. Let $g : \mathbb{R} \to \overline{\mathbb{R}}$ be a proper convex nondecreasing lower semicontinuous function. Suppose that the risk measure

$$\rho(Z) := \mathbb{E}[g(Z)] \tag{6.65}$$

is well defined and proper. Then ρ is convex, lower semicontinuous, satisfies the monotonicity condition (R2) and the representation (6.37) holds with

$$\rho^*(\zeta) = \mathbb{E}[g^*(\zeta)]. \tag{6.66}$$

Moreover, if $\rho(Z)$ is finite, then

$$\partial \rho(Z) = \{ \zeta \in \mathcal{Z}^* : \zeta(\omega) \in \partial g(Z(\omega)) \text{ a.e. } \omega \in \Omega \}.$$
(6.67)

Proof. Since g is lower semicontinuous and convex, we have by Fenchel-Moreau Theorem that

$$g(z) = \sup_{\alpha \in \mathbb{R}} \left\{ \alpha z - g^*(\alpha) \right\},\,$$

where g^* is the conjugate of g. As g is proper, the conjugate function g^* is also proper. It follows that

$$\rho(Z) = \mathbb{E}\left[\sup_{\alpha \in \mathbb{R}} \left\{\alpha Z - g^*(\alpha)\right\}\right].$$
(6.68)

By the interchangeability principle (Theorem 7.92) for the space $\mathfrak{M} := \mathcal{Z}^* = \mathcal{L}_q(\Omega, \mathcal{F}, P)$, which is decomposable, we obtain

$$\rho(Z) = \sup_{\zeta \in \mathcal{Z}^*} \left\{ \langle \zeta, Z \rangle - \mathbb{E}[g^*(\zeta)] \right\}.$$
(6.69)

It follows that ρ is convex and lower semicontinuous, and representation (6.37) holds with the conjugate function given in (6.66). Moreover, since the function g is nondecreasing, it follows that ρ satisfies the monotonicity condition (R2).

Since ρ is convex proper and lower semicontinuous, and hence $\rho^{**} = \rho$, we have by Proposition 7.84 that

$$\partial \rho(Z) = \operatorname*{arg\,max}_{\zeta \in \mathfrak{A}} \left\{ \mathbb{E}[\zeta Z - g^*(\zeta)] \right\},\tag{6.70}$$

assuming that $\rho(Z)$ is finite. Together with formula (7.277) of the interchangeability principle (Theorem 7.92) this implies (6.67). \Box

The above risk measure ρ , defined in (6.65), does not satisfy condition (R3) unless $g(z) \equiv z$. We can consider the corresponding risk measure $\tilde{\rho}$, defined in (6.62), which in the present case can be written as

$$\tilde{\rho}(Z) = \inf_{t \in \mathbb{R}} \mathbb{E}[Z + g(Z - t)].$$
(6.71)

⁹We consider here minimization problems, and that is why we speak about disutility. Any disutility function g corresponds to a utility function $u : \mathbb{R} \to \mathbb{R}$ defined by u(z) = -g(-z). Note that the function u is concave and nondecreasing since the function g is convex and nondecreasing.

We have that $\tilde{\rho}$ is convex since g is convex, that $\tilde{\rho}$ is monotone (i.e., condition (R2) holds) if z + g(z) is monotonically nondecreasing, and that $\tilde{\rho}$ satisfies condition (R3). If, moreover, $\tilde{\rho}$ is lower semicontinuous, then the dual representation

$$\tilde{\rho}(Z) = \sup_{\substack{\zeta \in \mathcal{Z}^*\\\mathbb{E}[\zeta]=1}} \left\{ \langle \zeta - 1, Z \rangle - \mathbb{E}[g^*(\zeta)] \right\}$$
(6.72)

holds.

Example 6.19 (Average Value-at-Risk) The risk measure ρ associated with disutility function g, defined in (6.65), is positively homogeneous only if g is positively homogeneous. Suppose now that $g(z) := \max\{az, bz\}$, where $b \ge a$. Then $g(\cdot)$ is positively homogeneous and convex. It is natural here to use the space $\mathcal{Z} := \mathcal{L}_1(\Omega, \mathcal{F}, P)$, since $\mathbb{E}[g(Z)]$ is finite for every $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$. The conjugate function of g is the indicator function $g^* = \mathbb{I}_{[a,b]}$. Therefore it follows by Proposition 6.18 that the representation (6.39) holds with

$$\mathfrak{A} = \{ \zeta \in \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P) : \zeta(\omega) \in [a, b] \text{ a.e. } \omega \in \Omega \}$$

Note that the dual space $\mathcal{Z}^* = \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$, of the space $\mathcal{Z} := \mathcal{L}_1(\Omega, \mathcal{F}, P)$, appears naturally in the corresponding representation (6.39) since, of course, the condition that " $\zeta(\omega) \in [a, b]$ for a.e. $\omega \in \Omega$ " implies that ζ is essentially bounded.

Consider now the risk measure

$$\tilde{\rho}(Z) := \mathbb{E}[Z] + \inf_{t \in \mathbb{R}} \mathbb{E}\left\{\beta_1[t - Z]_+ + \beta_2[Z - t]_+\right\}, \quad Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P), \tag{6.73}$$

where $\beta_1 \in [0, 1]$ and $\beta_2 \ge 0$. This risk measure can be recognized as risk measure defined in (6.71), associated with function $g(z) := \beta_1[-z]_+ + \beta_2[z]_+$. For specified β_1 and β_2 , the function z + g(z) is convex and nondecreasing, and $\tilde{\rho}$ is a continuous coherent risk measure. For $\beta_1 \in (0, 1]$ and $\beta_2 > 0$, the above risk measure $\tilde{\rho}(Z)$ can be written in the form

$$\tilde{\rho}(Z) = (1 - \beta_1)\mathbb{E}[Z] + \beta_1 \mathsf{AV}@\mathsf{R}_{\alpha}(Z), \tag{6.74}$$

where $\alpha := \beta_1/(\beta_1 + \beta_2)$. Note that the right hand side of (6.73) attains its minimum at $t^* = V@R_{\alpha}(Z)$. Therefore the second term on the right hand side of (6.73) is the weighted measure of deviation from the quantile V@R_{\alpha}(Z), discussed in section 6.2.3.

The respective conjugate function is the indicator function of the set $\mathfrak{A} := \operatorname{dom}(\tilde{\rho}^*)$, and $\tilde{\rho}$ can be represented in the dual form (6.39) with

$$\mathfrak{A} = \{\zeta \in \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P) : \zeta(\omega) \in [1 - \beta_1, 1 + \beta_2] \text{ a.e. } \omega \in \Omega, \ \mathbb{E}[\zeta] = 1\}.$$
(6.75)

In particular, for $\beta_1 = 1$ we have that $\tilde{\rho}(\cdot) = AV@R_{\alpha}(\cdot)$, and hence the dual representation (6.39) of AV@R_{\alpha} holds with the set

$$\mathfrak{A} = \left\{ \zeta \in \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P) : \zeta(\omega) \in [0, \alpha^{-1}] \text{ a.e. } \omega \in \Omega, \ \mathbb{E}[\zeta] = 1 \right\}.$$
(6.76)

Since AV@R_{α}(·) is convex and continuous, it is subdifferentiable and its subdifferentials can be calculated using formula (6.49). That is,

$$\partial(\mathsf{AV}@\mathsf{R}_{\alpha})(Z) = \arg\max_{\zeta \in \mathcal{Z}^*} \left\{ \langle \zeta, Z \rangle : \zeta(\omega) \in [0, \alpha^{-1}] \text{ a.e. } \omega \in \Omega, \ \mathbb{E}[\zeta] = 1 \right\}.$$
(6.77)

Consider the maximization problem on the right hand side of (6.77). The Lagrangian of that problem is

$$L(\zeta,\lambda) = \langle \zeta, Z \rangle + \lambda(1 - \mathbb{E}[\zeta]) = \langle \zeta, Z - \lambda \rangle + \lambda,$$

and its (Lagrangian) dual is the problem:

$$\underset{\lambda \in \mathbb{R}}{\operatorname{Min}} \sup_{\zeta(\cdot) \in [0, \alpha^{-1}]} \left\{ \langle \zeta, Z - \lambda \rangle + \lambda \right\}.$$
(6.78)

We have that

$$\sup_{\zeta(\cdot)\in[0,\alpha^{-1}]}\langle\zeta,Z-\lambda\rangle=\alpha^{-1}\mathbb{E}([Z-\lambda]_+),$$

and hence the dual problem (6.78) can be written as

$$\min_{\lambda \in \mathbb{R}} \alpha^{-1} \mathbb{E}([Z - \lambda]_{+}) + \lambda.$$
(6.79)

The set of optimal solutions of problem (6.79) is the interval with the end points given by the left and right side $(1 - \alpha)$ -quantiles of the cdf $H_Z(z) = \Pr(Z \le z)$ of $Z(\omega)$. Since the set of optimal solutions of the dual problem (6.78) is a compact subset of \mathbb{R} , there is no duality gap between the maximization problem on the right hand side of (6.77) and its dual (6.78) (see Theorem 7.11). It follows that the set of optimal solutions of the right hand side of (6.77), and hence the subdifferential $\partial(AV@R_\alpha)(Z)$, is given by such feasible $\bar{\zeta}$ that $(\bar{\zeta}, \bar{\lambda})$ is a saddle point of the Lagrangian $L(\zeta, \lambda)$ for any $(1 - \alpha)$ -quantile $\bar{\lambda}$. Recall that the left-side $(1 - \alpha)$ -quantile of the cdf $H_Z(z)$ is called Value-at-Risk and denoted V@R_{\alpha}(Z). Suppose for the moment that the set of $(1 - \alpha)$ -quantiles of H_Z is a singleton, i.e., consists of one point V@R_{\alpha}(Z). Then we have

$$\partial(\mathsf{AV}@\mathsf{R}_{\alpha})(Z) = \begin{cases} \zeta(\omega) = \alpha^{-1}, & \text{if } Z(\omega) > \mathsf{V}@\mathsf{R}_{\alpha}(Z), \\ \zeta(\omega) = 0, & \text{if } Z(\omega) < \mathsf{V}@\mathsf{R}_{\alpha}(Z), \\ \zeta(\omega) \in [0, \alpha^{-1}], & \text{if } Z(\omega) = \mathsf{V}@\mathsf{R}_{\alpha}(Z). \end{cases}$$
(6.80)

If the set of $(1 - \alpha)$ -quantiles of H_Z is not a singleton, then the probability that $Z(\omega)$ belongs to that set is zero. Consequently, formula (6.80) still holds with the left-side quantile $V@R_{\alpha}(Z)$ can be replaced by any $(1 - \alpha)$ -quantile of H_Z .

It follows that $\partial(AV@R_{\alpha})(Z)$ is a singleton, and hence $AV@R_{\alpha}(\cdot)$ is Hadamard differentiable at Z, iff the following condition holds:

$$\Pr(Z < \mathsf{V}@\mathsf{R}_{\alpha}(Z)) = 1 - \alpha \text{ or } \Pr(Z > \mathsf{V}@\mathsf{R}_{\alpha}(Z)) = \alpha.$$
(6.81)

Again if the set of $(1 - \alpha)$ -quantiles is not a singleton, then the left-side quantile $V@R_{\alpha}(Z)$ in the above condition (6.81) can be replaced by any $(1 - \alpha)$ -quantile of H_Z . Note that condition (6.81) is always satisfied if the cdf $H_Z(\cdot)$ is continuous at $V@R_{\alpha}(Z)$, but may also hold even if $H_Z(\cdot)$ is discontinuous at $V@R_{\alpha}(Z)$.

Example 6.20 (Entropic risk measure) Consider utility risk measure ρ , defined in (6.65), associated with the exponential disutility function $g(z) := e^z$. That is, $\rho(Z) := \mathbb{E}[e^Z]$. A natural question is what space $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ to use here. Let us observe that unless
the sigma algebra \mathcal{F} has a finite number of elements, in which case the space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ is finite dimensional, there exist such $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ that $\mathbb{E}[e^Z] = +\infty$. In fact for any $p \in [1, \infty)$ the domain of ρ forms a dense subset of $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $\rho(\cdot)$ is discontinuous at every $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ unless $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ is finite dimensional. Nevertheless, for any $p \in [1, \infty)$ the risk measure ρ is proper and, by Proposition 6.18, is convex and lower semicontinuous. Note that if $Z : \Omega \to \mathbb{R}$ is an \mathcal{F} -measurable function such that $\mathbb{E}[e^Z]$ is finite, then $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ for any $p \ge 1$. Therefore, by formula (6.67) of Proposition 6.18, we have that if $\mathbb{E}[e^Z]$ is finite, then $\partial \rho(Z) = \{e^Z\}$ is a singleton. It could be mentioned that although $\rho(\cdot)$ is subdifferentiable at every $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ where it is finite and has unique subgradient $\nabla \rho(Z) = e^Z$, it is discontinuous and nondifferentiable at Zunless $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ is finite dimensional.

The above risk measure associated with the exponential disutility function is not positively homogeneous and does not satisfy condition (R3). Let us consider instead the following risk measure (called *entropic*)

$$\rho_{\tau}^{\text{ent}}(Z) := \tau^{-1} \ln \mathbb{E}[e^{\tau Z}], \ \tau > 0.$$
(6.82)

On the space $\mathcal{Z} := \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$ this risk measure is finite valued. As it was discussed at the end of section 6.3, it is natural to pair \mathcal{Z} with the space $\mathcal{L}_1(\Omega, \mathcal{F}, P)$. In the weak^{*} topology this risk measure is lower semicontinuous.

It can be verified that $\rho_{\tau}^{\text{ent}}(\cdot)$ is convex (this can be shown in a way similar to the proof of convexity of the logarithmic moment-generating function in section 7.2.9). Moreover, for any $a \in \mathbb{R}$,

$$\tau^{-1}\ln\mathbb{E}\left[e^{\tau(Z+a)}\right] = \tau^{-1}\ln\left(e^{\tau a}\mathbb{E}\left[e^{\tau Z}\right]\right) = \tau^{-1}\ln\mathbb{E}\left[e^{\tau Z}\right] + a,$$

i.e., ρ_{τ}^{ent} satisfies condition (R3). Clearly the monotonicity condition (R2) also holds, and hence ρ_{τ}^{ent} is a convex risk measure.

Let us calculate the conjugate of ρ_{τ}^{ent} . We have that

$$(\rho_{\tau}^{\text{ent}})^*(\zeta) = \sup_{Z \in \mathcal{Z}} \left\{ \mathbb{E}[\zeta Z] - \tau^{-1} \ln \mathbb{E}[e^{\tau Z}] \right\}, \ \zeta \in \mathcal{L}_1(\Omega, \mathcal{F}, P).$$
(6.83)

Since ρ_{τ}^{ent} satisfies conditions (R2) and (R3), it follows that dom $(\rho_{\tau}^{\text{ent}}) \subset \mathfrak{P}$, where \mathfrak{P} is the set of density functions (see (6.40)). By writing (first-order) optimality conditions for the optimization problem on the right hand side of (6.83), it is straightforward to verify that for $\zeta \in \mathfrak{P}$ such that $\zeta(\omega) > 0$ for a.e. $\omega \in \Omega$, a point \overline{Z} is an optimal solution of that problem if $\overline{Z} = \ln \zeta + a$ for some $a \in \mathbb{R}$. Substituting this into the right hand side of (6.83), and noting that the obtained expression does not depend on a, we obtain

$$(\rho_{\tau}^{\text{ent}})^{*}(\zeta) = \begin{cases} \tau^{-1} \mathbb{E}[\zeta \ln \zeta], & \text{if } \zeta \in \mathfrak{P}, \\ +\infty, & \text{if } \zeta \notin \mathfrak{P}. \end{cases}$$
(6.84)

Note that $x \ln x$ tends to zero as $x \downarrow 0$. Therefore, we set $0 \ln 0 = 0$ in the above formula (6.84).

Clearly for any $\zeta \in \mathfrak{P}$, the value $(\rho_{\tau}^{\text{ent}})^*(\zeta)$ is monotonically decreasing with increase of τ . By the dual representation of ρ_{τ}^{ent} , it follows that for any $Z \in \mathcal{Z}$ the value $\rho_{\tau}^{\text{ent}}(Z)$ is monotonically increasing with increase of τ . By l'Hôpital's rule we have

$$\lim_{\tau \downarrow 0} \frac{\ln \mathbb{E}[e^{\tau Z}]}{\tau} = \lim_{\tau \downarrow 0} \frac{d\left(\ln \mathbb{E}[e^{\tau Z}]\right)}{d\tau} = \lim_{\tau \downarrow 0} \frac{\mathbb{E}[Ze^{\tau Z}]}{\mathbb{E}[e^{\tau Z}]} = \mathbb{E}[Z],$$

that is

$$\lim_{\tau \to 0} \rho_{\tau}^{\text{ent}}(Z) = \mathbb{E}[Z]. \tag{6.85}$$

On the other hand, by (6.84) we have that

$$\lim_{\tau \to +\infty} (\rho_{\tau}^{\text{ent}})^*(\zeta) = \begin{cases} 0, & \text{if } \zeta \in \mathfrak{P}, \\ +\infty, & \text{if } \zeta \notin \mathfrak{P}. \end{cases}$$

Together with the corresponding dual representation of ρ_{τ}^{ent} this implies

$$\lim_{\tau \to +\infty} \rho_{\tau}^{\text{ent}}(Z) = \sup_{\zeta \in \mathfrak{P}} \mathbb{E}[\zeta Z].$$
(6.86)

Recall that the right hand side of (6.86) is equal to $\operatorname{ess\,sup}(Z)$ (see (6.47)). It follows that in the limit, as $\tau \to +\infty$, the entropic risk measure becomes the essential supremum risk measure

$$\lim_{\tau \to +\infty} \rho_{\tau}^{\text{ent}}(Z) = \operatorname{ess\,sup}(Z). \tag{6.87}$$

Example 6.21 (Mean-variance risk measure) Consider

$$\rho(Z) := \mathbb{E}[Z] + c \mathbb{V}ar[Z], \tag{6.88}$$

where $c \ge 0$ is a given constant. It is natural to use here the space $\mathcal{Z} := \mathcal{L}_2(\Omega, \mathcal{F}, P)$ since for any $Z \in \mathcal{L}_2(\Omega, \mathcal{F}, P)$ the expectation $\mathbb{E}[Z]$ and variance $\mathbb{V}ar[Z]$ are well defined and finite. We have here that $\mathcal{Z}^* = \mathcal{Z}$ (i.e., \mathcal{Z} is a Hilbert space) and for $Z \in \mathcal{Z}$ its norm is given by $\|Z\|_2 = \sqrt{\mathbb{E}[Z^2]}$. We also have that

$$\|Z\|_{2}^{2} = \sup_{\zeta \in \mathcal{Z}} \left\{ \langle \zeta, Z \rangle - \frac{1}{4} \|\zeta\|_{2}^{2} \right\}.$$
 (6.89)

Indeed, it is not difficult to verify that the maximum on the right hand side of (6.89) is attained at $\zeta = 2Z$.

We have that $\mathbb{V}ar[Z] = ||Z - \mathbb{E}[Z]||_2^2$, and since $||\cdot||_2^2$ is a convex and continuous function on the Hilbert space Z, it follows that $\rho(\cdot)$ is convex and continuous. Also because of (6.89), we can write

$$\mathbb{V}\mathrm{ar}[Z] = \sup_{\zeta \in \mathcal{Z}} \left\{ \langle \zeta, Z - \mathbb{E}[Z] \rangle - \frac{1}{4} \| \zeta \|_2^2 \right\}.$$

Since

$$\langle \zeta, Z - \mathbb{E}[Z] \rangle = \langle \zeta, Z \rangle - \mathbb{E}[\zeta] \mathbb{E}[Z] = \langle \zeta - \mathbb{E}[\zeta], Z \rangle, \tag{6.90}$$

we can rewrite the last expression as follows:

$$\begin{aligned} \mathbb{V}\mathrm{ar}[Z] &= \sup_{\zeta \in \mathcal{Z}} \left\{ \langle \zeta - \mathbb{E}[\zeta], Z \rangle - \frac{1}{4} \|\zeta\|_2^2 \right\} \\ &= \sup_{\zeta \in \mathcal{Z}} \left\{ \langle \zeta - \mathbb{E}[\zeta], Z \rangle - \frac{1}{4} \mathbb{V}\mathrm{ar}[\zeta] - \frac{1}{4} \left(\mathbb{E}[\zeta] \right)^2 \right\}. \end{aligned}$$

Since $\zeta - \mathbb{E}[\zeta]$ and $\mathbb{V}ar[\zeta]$ are invariant under transformations of ζ to $\zeta + a$, where $a \in \mathbb{R}$, the above maximization can be restricted to such $\zeta \in \mathcal{Z}$ that $\mathbb{E}[\zeta] = 0$. Consequently

$$\mathbb{V}\mathrm{ar}[Z] = \sup_{\substack{\zeta \in \mathcal{Z} \\ \mathbb{E}[\zeta] = 0}} \left\{ \langle \zeta, Z \rangle - \frac{1}{4} \mathbb{V}\mathrm{ar}[\zeta] \right\}.$$

Therefore the risk measure ρ , defined in (6.88), can be expressed as

$$\rho(Z) = \mathbb{E}[Z] + c \sup_{\substack{\zeta \in \mathcal{Z} \\ \mathbb{E}[\zeta] = 0}} \left\{ \langle \zeta, Z \rangle - \frac{1}{4} \mathbb{V}\mathrm{ar}\left[\zeta\right] \right\},\$$

and hence for c>0 (by making change of variables $\zeta'=c\zeta+1)$ as

$$\rho(Z) = \sup_{\substack{\zeta \in \mathcal{Z} \\ \mathbb{E}[\zeta]=1}} \left\{ \langle \zeta, Z \rangle - \frac{1}{4c} \mathbb{V}\mathrm{ar}[\zeta] \right\}.$$
(6.91)

It follows that for any c > 0 the function ρ is convex, continuous and

$$\rho^*(\zeta) = \begin{cases} \frac{1}{4c} \operatorname{Var}[\zeta], & \text{if } \mathbb{E}[\zeta] = 1, \\ +\infty, & \text{otherwise.} \end{cases}$$
(6.92)

The function ρ satisfies the translation equivariance condition (R3), e.g., because the domain of its conjugate contains only ζ such that $\mathbb{E}[\zeta] = 1$. However, for any c > 0 the function ρ is not positively homogeneous and it does not satisfy the monotonicity condition (R2), because the domain of ρ^* contains density functions which are not nonnegative.

Since $\mathbb{V}ar[Z] = \langle Z, Z \rangle - (\mathbb{E}[Z])^2$, it is straightforward to verify that $\rho(\cdot)$ is (Fréchet) differentiable and

$$\nabla \rho(Z) = 2cZ - 2c \mathbb{E}[Z] + 1. \tag{6.93}$$

Example 6.22 (Mean-deviation risk measures of order p) For $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $\mathcal{Z}^* := \mathcal{L}_q(\Omega, \mathcal{F}, P)$, with $p \in [1, \infty)$ and $c \ge 0$, consider

$$\rho(Z) := \mathbb{E}[Z] + c \left(\mathbb{E}\left[|Z - \mathbb{E}[Z]|^p \right] \right)^{1/p}.$$
(6.94)

We have that $(\mathbb{E}[|Z|^p])^{1/p} = ||Z||_p$, where $||\cdot||_p$ denotes the norm of the space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$. The function ρ is convex continuous and positively homogeneous. Also

$$||Z||_p = \sup_{||\zeta||_q \le 1} \langle \zeta, Z \rangle, \tag{6.95}$$

and hence

$$\left(\mathbb{E}\left[|Z - \mathbb{E}[Z]|^{p}\right]\right)^{1/p} = \sup_{\|\zeta\|_{q} \le 1} \langle \zeta, Z - \mathbb{E}[Z] \rangle = \sup_{\|\zeta\|_{q} \le 1} \langle \zeta - \mathbb{E}[\zeta], Z \rangle.$$
(6.96)

It follows that representation (6.39) holds with the set \mathfrak{A} given by

$$\mathfrak{A} = \left\{ \zeta' \in \mathcal{Z}^* : \zeta' = 1 + \zeta - \mathbb{E}[\zeta], \ \|\zeta\|_q \le c \right\}.$$
(6.97)

We obtain here that ρ satisfies conditions (R1), (R3) and (R4).

The monotonicity condition (R2) is more involved. Suppose that p = 1. Then $q = \infty$ and hence for any $\zeta' \in \mathfrak{A}$ and a.e. $\omega \in \Omega$ we have

$$\zeta'(\omega) = 1 + \zeta(\omega) - \mathbb{E}[\zeta] \ge 1 - |\zeta(\omega)| - \mathbb{E}[\zeta] \ge 1 - 2c.$$

It follows that if $c \in [0, 1/2]$, then $\zeta'(\omega) \ge 0$ for a.e. $\omega \in \Omega$, and hence condition (R2) follows. Conversely, take $\zeta := c(-\mathbf{1}_A + \mathbf{1}_{\Omega \setminus A})$, for some $A \in \mathcal{F}$, and $\zeta' = 1 + \zeta - \mathbb{E}[\zeta]$. We have that $\|\zeta\|_{\infty} = c$ and $\zeta'(\omega) = 1 - 2c + 2cP(A)$ for all $\omega \in A$ It follows that if c > 1/2, then $\zeta'(\omega) < 0$ for all $\omega \in A$, provided that P(A) is small enough. We obtain that for c > 1/2 the monotonicity property (R2) does not hold if the following condition is satisfied:

For any
$$\varepsilon > 0$$
 there exists $A \in \mathcal{F}$ such that $\varepsilon > P(A) > 0$. (6.98)

That is, for p = 1 the mean-deviation measure ρ satisfies (R2) if, and provided that condition (6.98) holds only if, $c \in [0, 1/2]$. (The above condition (6.98) holds, in particular, if the measure P is nonatomic.)

Suppose now that p > 1. For a set $A \in \mathcal{F}$ and $\alpha > 0$ let us take $\zeta := -\alpha \mathbf{1}_A$ and $\zeta' = 1 + \zeta - \mathbb{E}[\zeta]$. Then $\|\zeta\|_q = \alpha P(A)^{1/q}$ and $\zeta'(\omega) = 1 - \alpha + \alpha P(A)$ for all $\omega \in A$. It follows that if p > 1, then for any c > 0 the mean-deviation measure ρ does not satisfy (R2) provided that condition (6.98) holds.

Since ρ is convex continuous, it is subdifferentiable. By (6.49) and because of (6.97) and (6.90) we have here that $\partial \rho(Z)$ is formed by vectors $\zeta' = 1 + \zeta - \mathbb{E}[\zeta]$ such that $\zeta \in \arg \max_{\|\zeta\|_q \leq c} \langle \zeta, Z - \mathbb{E}[Z] \rangle$. That is,

$$\partial \rho(Z) = \left\{ \zeta' = 1 + c \,\zeta - c \,\mathbb{E}[\zeta] : \zeta \in \mathfrak{S}_Y \right\},\tag{6.99}$$

where $Y(\omega) \equiv Z(\omega) - \mathbb{E}[Z]$ and \mathfrak{S}_Y is the set of contact points of Y (see (7.251) for the definition of contact points). If $p \in (1, \infty)$, then the set \mathfrak{S}_Y is a singleton, i.e., there is unique contact point ζ_Y^* , provided that $Y(\omega)$ is not zero for a.e. $\omega \in \Omega$. In that case $\rho(\cdot)$ is Hadamard differentiable at Z and

$$\nabla \rho(Z) = 1 + c\,\zeta_Y^* - c\,\mathbb{E}[\zeta_Y^*] \tag{6.100}$$

(an explicit form of the contact point ζ_Y^* is given in equation (7.259)). If $Y(\omega)$ is zero for a.e. $\omega \in \Omega$, i.e., $Z(\omega)$ is constant w.p.1, then $\mathfrak{S}_Y = \{\zeta \in \mathbb{Z}^* : \|\zeta\|_q \le 1\}$.

For p = 1 the set \mathfrak{S}_Y is described in equation (7.260). It follows that if p = 1, and hence $q = \infty$, then the subdifferential $\partial \rho(Z)$ is a singleton iff $Z(\omega) \neq \mathbb{E}[Z]$ for a.e. $\omega \in \Omega$, in which case

$$\nabla \rho(Z) = \begin{cases} \zeta : & \zeta(\omega) = 1 + 2c \left(1 - \Pr(Z > \mathbb{E}[Z]) \right), & \text{if } Z(\omega) > \mathbb{E}[Z], \\ \zeta(\omega) = 1 - 2c \Pr(Z > \mathbb{E}[Z]), & \text{if } Z(\omega) < \mathbb{E}[Z]. \end{cases}$$
(6.101)

Example 6.23 (Mean-upper-semideviation of order p) Let $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and for $c \ge 0$ consider¹⁰

$$\rho(Z) := \mathbb{E}[Z] + c \left(\mathbb{E}\left[\left[Z - \mathbb{E}[Z] \right]_{+}^{p} \right] \right)^{1/p}.$$
(6.102)

¹⁰We denote $[a]_{+}^{p} := (\max\{0, a\})^{p}$.

For any $c \ge 0$ this function satisfies conditions (R1), (R3) and (R4), and similarly to the derivations of Example 6.22 it can be shown that representation (6.39) holds with the set \mathfrak{A} given by

$$\mathfrak{A} = \left\{ \zeta' \in \mathcal{Z}^* : \zeta' = 1 + \zeta - \mathbb{E}[\zeta], \ \|\zeta\|_q \le c, \ \zeta \succeq 0 \right\}.$$
(6.103)

Since $|\mathbb{E}[\zeta]| \leq \mathbb{E}[\zeta] \leq \|\zeta\|_q$ for any $\zeta \in \mathcal{L}_q(\Omega, \mathcal{F}, P)$, we have that every element of the above set \mathfrak{A} is nonnegative and has its expected value equal to 1. This means that the monotonicity condition (R2) holds, if and, provided that condition (6.98) holds, only if $c \in [0, 1]$. That is, ρ is a coherent risk measure if $c \in [0, 1]$.

Since ρ is convex continuous, it is subdifferentiable. Its subdifferential can be calculated in a way similar to the derivations of Example 6.22. That is, $\partial \rho(Z)$ is formed by vectors $\zeta' = 1 + \zeta - \mathbb{E}[\zeta]$ such that

$$\zeta \in \arg\max\left\{\langle \zeta, Y \rangle : \|\zeta\|_q \le c, \, \zeta \succeq 0\right\},\tag{6.104}$$

where $Y := Z - \mathbb{E}[Z]$. Suppose that $p \in (1, \infty)$. Then the set of maximizers on the right hand side of (6.104) is not changed if Y is replaced by Y_+ , where $Y_+(\cdot) := [Y(\cdot)]_+$. Consequently, if $Z(\omega)$ is not constant for a.e. $\omega \in \Omega$, and hence $Y_+ \neq 0$, then $\partial \rho(Z)$ is a singleton and

$$\nabla \rho(Z) = 1 + c \,\zeta_{Y_+}^* - c \,\mathbb{E}[\zeta_{Y_+}^*],\tag{6.105}$$

where $\zeta_{Y_+}^*$ is the contact point of Y_+ (note that the contact point of Y_+ is nonnegative since $Y_+ \succeq 0$).

Suppose now that p = 1 and hence $q = \infty$. Then the set on the right hand side of (6.104) is formed by $\zeta(\cdot)$ such that: $\zeta(\omega) = c$ if $Y(\omega) > 0$, $\zeta(\omega) = 0$, if $Y(\omega) < 0$, and $\zeta(\omega) \in [0, c]$ if $Y(\omega) = 0$. It follows that $\partial \rho(Z)$ is a singleton iff $Z(\omega) \neq \mathbb{E}[Z]$ for a.e. $\omega \in \Omega$, in which case

$$\nabla \rho(Z) = \begin{cases} \zeta(\omega) = 1 + c \left(1 - \Pr(Z > \mathbb{E}[Z]) \right), & \text{if } Z(\omega) > \mathbb{E}[Z], \\ \zeta(\omega) = 1 - c \Pr(Z > \mathbb{E}[Z]), & \text{if } Z(\omega) < \mathbb{E}[Z]. \end{cases}$$
(6.106)

It can be noted that by Lemma 6.1

$$\mathbb{E}(|Z - \mathbb{E}[Z]|) = 2\mathbb{E}([Z - \mathbb{E}[Z]]_+).$$
(6.107)

Consequently, formula (6.106) can be derived directly from (6.101).

Example 6.24 (Mean-upper-semivariance from a target) Let $\mathcal{Z} := \mathcal{L}_2(\Omega, \mathcal{F}, P)$ and for a weight $c \ge 0$ and a target $\tau \in \mathbb{R}$ consider

$$\rho(Z) := \mathbb{E}[Z] + c \mathbb{E}\left[\left[Z - \tau\right]_{+}^{2}\right].$$
(6.108)

This is a convex and continuous risk measure. We can now use (6.69) with $g(z) := z + c[z - \tau]_+^2$. Since

$$g^*(\alpha) = \begin{cases} (\alpha - 1)^2/4c + \tau(\alpha - 1), & \text{if } \alpha \ge 1, \\ +\infty, & \text{otherwise,} \end{cases}$$

we obtain that

$$\rho(Z) = \sup_{\zeta \in \mathcal{Z}, \, \zeta(\cdot) \ge 1} \left\{ \mathbb{E}[\zeta Z] - \tau \mathbb{E}[\zeta - 1] - \frac{1}{4c} \mathbb{E}[(\zeta - 1)^2] \right\}.$$
(6.109)

Consequently, representation (6.38) holds with $\mathfrak{A} = \{\zeta \in \mathcal{Z} : \zeta - 1 \succeq 0\}$ and

$$\rho^*(\zeta) = \tau \mathbb{E}[\zeta - 1] + \frac{1}{4c} \mathbb{E}[(\zeta - 1)^2], \quad \zeta \in \mathfrak{A}.$$

If c > 0 then conditions (R3) and (R4) are not satisfied by this risk measure.

Since ρ is convex continuous, it is subdifferentiable. Moreover, by using (6.67) we obtain that its subdifferentials are singletons and hence $\rho(\cdot)$ is differentiable at every $Z \in \mathcal{Z}$, and

$$\nabla \rho(Z) = \begin{cases} \zeta : & \zeta(\omega) = 1 + 2c(Z(\omega) - \tau), & \text{if } Z(\omega) \ge \tau, \\ \zeta(\omega) = 1, & \text{if } Z(\omega) < \tau. \end{cases}$$
(6.110)

The above formula can be also derived directly and it can be shown that ρ is differentiable in the sense of Fréchet.

Example 6.25 (Mean-upper-semideviation of order p from a target) Let Z be the space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$, and for $c \ge 0$ and $\tau \in \mathbb{R}$ consider

$$\rho(Z) := \mathbb{E}[Z] + c \left(\mathbb{E}\left[\left[Z - \tau \right]_{+}^{p} \right] \right)^{1/p}.$$
(6.111)

For any $c \ge 0$ and τ this risk measure satisfies conditions (R1) and (R2), but not (R3) and (R4) if c > 0. We have

$$\left(\mathbb{E}\left[\left[Z - \tau \right]_{+}^{p} \right] \right)^{1/p} = \sup_{\|\zeta\|_{q} \le 1} \mathbb{E}\left(\zeta [Z - \tau]_{+} \right) = \sup_{\|\zeta\|_{q} \le 1, \ \zeta(\cdot) \ge 0} \mathbb{E}\left(\zeta [Z - \tau]_{+} \right)$$
$$= \sup_{\|\zeta\|_{q} \le 1, \ \zeta(\cdot) \ge 0} \mathbb{E}\left(\zeta [Z - \tau] \right) = \sup_{\|\zeta\|_{q} \le 1, \ \zeta(\cdot) \ge 0} \mathbb{E}\left[\zeta Z - \tau \zeta \right].$$

We obtain that representation (6.38) holds with

1 /

$$\mathfrak{A} = \{ \zeta \in \mathcal{Z}^* : \|\zeta\|_q \le c, \ \zeta \succeq 0 \}$$

and $\rho^*(\zeta) = \tau \mathbb{E}[\zeta]$ for $\zeta \in \mathfrak{A}$.

6.3.3 Law Invariant Risk Measures

As in the previous sections, unless stated otherwise we assume here that $Z = \mathcal{L}_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$. We say that random outcomes (random variables) $Z \in Z$ and $Z' \in Z$ have the same distribution, with respect to the reference probability measure P, if $P(Z \leq z) = P(Z' \leq z)$ for all $z \in \mathbb{R}$. We also say that Z and Z' are *distributionally equivalent* and write this relation as $Z \stackrel{\mathcal{D}}{\sim} Z'$. The risk measures $\rho(Z)$ discussed in all examples considered in section 6.3.2 were dependent only on the distribution of Z. That is, each risk measure $\rho(Z)$, considered in section 6.3.2, could be formulated in terms of the cumulative

distribution function (cdf) $H_Z(t) := P(Z \le t)$ associated with $Z \in \mathcal{Z}$. Such risk measures are called *law invariant*.¹¹

Definition 6.26. A risk measure $\rho : \mathbb{Z} \to \mathbb{R}$ is said to be law invariant, with respect to the reference probability measure P, if for all $Z, Z' \in \mathbb{Z}$ the following implication holds

$$\left\{ Z \stackrel{\mathcal{D}}{\sim} Z' \right\} \Rightarrow \left\{ \rho(Z) = \rho(Z') \right\}.$$

Remark 27. Since a law invariant risk measure $\rho(Z)$ depends only on the cdf $H = H_Z$ we also write $\rho(H_Z)$ or $\rho(H)$ rather than $\rho(Z)$. It should be remembered that the class of the corresponding cumulative distribution functions depends on the considered space $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ of random variables. Therefore the notion of law invariance of risk measure $\rho : \mathcal{Z} \to \mathbb{R}$ is associated with the space \mathcal{Z} , and in particular with the choice of the reference probability space (Ω, \mathcal{F}, P) .

Let us have a careful look at what does it mean that two random variables $Z, Z' \in \mathcal{Z}$ are distributionally equivalent. Let us start with finite probability space $\Omega = \{\omega_1, \ldots, \omega_n\}$ equipped with the sigma algebra of all its subsets and probability measure P assigning equal probabilities $p_i = 1/n$ to each element of Ω . Consider a permutation π of the set Ω , i.e., $\pi : \Omega \to \Omega$ is a one-to-one and onto mapping, and denote by \mathfrak{G} the set of all these permutations. Note that $\pi \in \mathfrak{G}$ preserves the probability measure P, i.e., for any set $A \subset \Omega$ we have that $P(A) = P(\pi(A))$, where $\pi(A) = \{\pi(\omega) : \omega \in A\}$. It follows that if $Z : \Omega \to \mathbb{R}$ is a random variable and $\mathbb{I}^2 Z' = Z \circ \pi$ for some $\pi \in \mathfrak{G}$, then Z and Z'are distributionally equivalent. In fact it is not difficult to see that two random variables $Z, Z' : \Omega \to \mathbb{R}$ are distributionally equivalent iff there exists permutation $\pi \in \mathfrak{G}$ such that $Z' = Z \circ \pi$.

Now let (Ω, \mathcal{F}, P) be a general probability space. Recall that a transformation $T : \Omega \to \Omega$ is said to be *measurable* if $T^{-1}(A) \in \mathcal{F}$ for any $A \in \mathcal{F}$, where $T^{-1}(A) = \{\omega \in \Omega : T(\omega) \in A\}$.

Definition 6.27. We say that a mapping $T : \Omega \to \Omega$ is a measure-preserving transformation if T is measurable and for any $A \in \mathcal{F}$ it follows that $P(A) = P(T^{-1}(A))$, and T is one-to-one and onto. We denote by \mathfrak{G} the set of measure-preserving transformations.

This definition of measure-preserving transformations is slightly stronger than the standard one in that we assume that T is one-to-one and onto; hence T is invertable and P(A) = P(T(A)) for any $A \in \mathcal{F}$. The set \mathfrak{G} forms a group of transformations, i.e., if $T_1, T_2 \in \mathfrak{G}$, then their composition $T_1 \circ T_2 \in \mathfrak{G}$, and if $T \in \mathfrak{G}$ then its inverse $T^{-1} \in \mathfrak{G}$. In the case of finite space Ω with equal probabilities, the group of measure-preserving transformations coincides with the group of permutations of Ω . As it was pointed above, in that case random variables Z, Z' are distributionally equivalent iff there exists $\pi \in \mathfrak{G}$ such that $Z' = Z \circ \pi$. This characterization of distributional equivalence can be extended to general nonatomic spaces.

¹¹Some authors also use terms *distribution based* or *version independent*.

¹²By $Z' = Z \circ \pi$ we denote the composition $Z'(\omega) = Z(\pi(\omega))$.

Proposition 6.28. If $Z \in Z$ and $Z' = Z \circ T$ for some $T \in \mathfrak{G}$, then $Z' \in Z$ and Z and Z' are distributionally equivalent. If, moreover, the space (Ω, \mathcal{F}, P) is nonatomic, then the converse implication holds, i.e., if $Z \in Z$ and $Z' \in Z$ are distributionally equivalent, then there exists a measure-preserving transformation $T \in \mathfrak{G}$ such that ${}^{13}Z' = Z \circ T$.

Proof. Let $Z \in \mathcal{Z}$ and $Z' = Z \circ T$ for some $T \in \mathfrak{G}$. Then

$$\int |Z'(\omega)|^p dP(\omega) = \int |Z(T(\omega))|^p dP(\omega) = \int |Z(\omega)|^p dP(\omega),$$

and hence $Z' \in \mathcal{Z}$. Furthermore, for $t \in \mathbb{R}$ consider $A := \{\omega \in \Omega : Z(\omega) \leq t\}$. We have that

$$T^{-1}(A) = \{T^{-1}(\omega) : \omega \in A\} = \{T^{-1}(\omega) : Z(\omega) \le t\} = \{\omega : Z(T(\omega)) \le t\}.$$

Thus

$$\Pr(Z' \le t) = P(T^{-1}(A)) = P(A) = \Pr(Z \le t),$$

and hence Z and Z' are distributionally equivalent.

The converse implication can be proved by partitioning Ω into a finite number of subsets of equal probabilities, applying an appropriate permutation of these sets and then passing to the limit. For a formal proof we may refer, e.g., to [119, lemma A.4].

Remark 28. Suppose that the space (Ω, \mathcal{F}, P) is nonatomic. Consider measurable sets $A, B \in \mathcal{F}$, and $Z := \mathbf{1}_A$ and $Z' := \mathbf{1}_B$. Note that Z and Z' are distributionally equivalent iff P(A) = P(B), and $Z \circ T = \mathbf{1}_{T^{-1}(A)}$ for $T \in \mathfrak{G}$. Hence it follows by the above proposition that there exists a measure-preserving transformation $T \in \mathfrak{G}$ such that B = T(A) iff P(A) = P(B).

Remark 29. The assumption in the above proposition for the probability space to be nonatomic is essential for the converse implication. Consider, for example, finite space $\Omega := \{\omega_1, \omega_2, \omega_3\}$ with respective probabilities $p_1 = p_2 = 1/4$ and $p_3 = 1/2$, and random variables $Z, Z' : \Omega \to \mathbb{R}$ defined as $Z(\omega_1) = Z(\omega_2) = 1$, $Z(\omega_3) = 0$, and $Z'(\omega_1) = Z'(\omega_2) = 0$, $Z'(\omega_3) = 1$. These two random variables are distributionally equivalent, but cannot be transformed one into another by a measure-preserving transformation.

Proposition 6.29. Let $\rho : \mathbb{Z} \to \mathbb{R}$ be a proper lower semicontinuous convex risk measure. Then the following holds. If ρ is law invariant, then its conjugate function ρ^* is invariant with respect to measure-preserving transformations, i.e., for any $T \in \mathfrak{G}$ and $\zeta \in \mathbb{Z}^*$ it follows that $\rho^*(\zeta \circ T) = \rho^*(\zeta)$. Moreover, if the space (Ω, \mathcal{F}, P) is nonatomic, then ρ is law invariant iff its conjugate function ρ^* is invariant with respect to measure-preserving transformations.

¹³The equality $Z' = Z \circ T$ is understood, of course, as that $Z'(\omega) = Z(T(\omega))$ for a.e. $\omega \in \Omega$.

Proof. For any $\zeta \in \mathbb{Z}^*$ and $T \in \mathfrak{G}$ we have that $\zeta \circ T \in \mathbb{Z}^*$ and

$$\rho^*(\zeta \circ T) = \sup_{Z \in \mathcal{Z}} \left\{ \int \zeta(T(\omega)) Z(\omega) dP(\omega) - \rho(Z) \right\}
= \sup_{Z \in \mathcal{Z}} \left\{ \int \zeta(\omega) Z(T^{-1}(\omega)) dP(\omega) - \rho(Z) \right\}
= \sup_{Z' \in \mathcal{Z}} \left\{ \int \zeta(\omega) Z'(\omega) dP(\omega) - \rho(Z' \circ T) \right\},$$
(6.112)

where we made change of variables $Z' = Z \circ T^{-1}$. If ρ is law invariant and hence is invariant with respect to measure-preserving transformations, then $\rho(Z' \circ T) = \rho(Z')$ and thus it follows that $\rho^*(\zeta \circ T) = \rho^*(\zeta)$.

Conversely, suppose that the space (Ω, \mathcal{F}, P) is nonatomic and ρ^* is invariant with respect to measure-preserving transformations. By Fenchel-Moreau Theorem we have

$$\rho(Z) := \sup_{\zeta \in \mathcal{Z}^*} \left\{ \langle \zeta, Z \rangle - \rho^*(\zeta) \right\}.$$
(6.113)

Consider distributionally equivalent $Z, Z' \in \mathcal{Z}$. Then by Proposition 6.28 there exists $T \in \mathfrak{G}$ such that $Z' = Z \circ T$. Similar to (6.112) by using (6.113) and invariance of ρ^* with respect to measure-preserving transformations, we obtain that $\rho(Z) = \rho(Z')$. This shows that ρ is law invariant. \Box

If risk measure $\rho : \mathbb{Z} \to \mathbb{R}$ is coherent, then its conjugate function ρ^* is the indicator function of the set $\mathfrak{A} = \operatorname{dom}(\rho^*)$. In that case invariance of ρ^* with respect to measure-preserving transformations means that the set \mathfrak{A} is invariant with respect to measure-preserving transformations, i.e., for any $T \in \mathfrak{G}$ and $\zeta \in \mathfrak{A}$ it follows that $\zeta \circ T \in \mathfrak{A}$. Therefore we have the following consequence of Proposition 6.29.

Corollary 6.30. Let $\rho : \mathbb{Z} \to \mathbb{R}$ be a proper lower semicontinuous coherent risk measure. Then the following holds. If ρ is law invariant, then its dual set \mathfrak{A} is invariant with respect to measure-preserving transformations. Moreover, if the space (Ω, \mathcal{F}, P) is nonatomic, then ρ is law invariant iff its dual set \mathfrak{A} is invariant with respect to measure-preserving transformations.

As it was pointed above we can consider a law invariant risk measure as a function $\rho(H)$ defined on an appropriate set of cumulative distribution functions. If $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$, and the space (Ω, \mathcal{F}, P) is *nonatomic*, then the corresponding set of cumulative distribution functions $H = H_Z$, $Z \in \mathcal{Z}$, consists of the set of monotonically nondecreasing, right side continuous functions $H : \mathbb{R} \to \mathbb{R}$ such that the integral $\int_{-\infty}^{+\infty} |z|^p dH(z)$ is finite. By making change of variables t = H(z) we can write this integral condition as $\int_0^1 |H^{-1}(t)|^p dt < \infty$. We can view H^{-1} as a random variable defined on the standard uniform probability space¹⁴ and having finite *p*-th order moment. Therefore in the case of nonatomic probability space we can assume without loss of generality that (Ω, \mathcal{F}, P) is the standard uniform probability space.

Remark 30. Let us observe that a random variable $Z : \Omega \to \mathbb{R}$, defined on the standard uniform probability space, has the same probability distribution as H_Z^{-1} . Hence by Proposition 6.28 there exists a measure-preserving transformation $T \in \mathfrak{G}$ such that $H_Z^{-1} = Z \circ T$.

¹⁴Recall that the standard uniform probability space is the space $\Omega = [0, 1]$ equipped with its Borel sigma algebra and uniform probability measure P.

Thus for any random variable Z, defined on the standard uniform probability space, there exists a measure-preserving transformation $T \in \mathfrak{G}$ such that $Z \circ T$ is monotonically nondecreasing. Note that we view Z and $Z' = Z \circ T$ as random variables, i.e., Z is represented by a class of functions which may differ from each other on sets of measure zero. By saying that Z' is monotonically nondecreasing we mean that there is a monotonically nondecreasing element in the corresponding class of functions. Since a monotonically nondecreasing function has at most a countable number of discontinuity points, we can take $Z' = Z \circ T$ to be left side or right side continuous. In particular if $Z : \Omega \to \mathbb{R}$ is a monotonically nondecreasing random variable, then $Z = H_Z^{-1}$ in the sense that there is an element in the corresponding class of functions.

By Proposition 6.8 we know that if $\rho : \mathcal{L}_p(\Omega, \mathcal{F}, P) \to \mathbb{R}$, with $p \in [1, \infty)$, is a proper convex risk measure, then either $\rho(\cdot)$ is finite valued and continuous on \mathcal{Z} , or $\rho(Z) = +\infty$ on a dense set of points $Z \in \mathcal{Z}$. In all examples considered in section 6.3.2 it was possible to find an appropriate space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ on which the considered risk measure was finite valued. The following result shows that it is not possible to construct a finite valued law invariant convex risk measure on a space larger than $\mathcal{L}_1(\Omega, \mathcal{F}, P)$. This indicates that it does not make much sense to consider convex risk measures on spaces larger than $\mathcal{L}_1(\Omega, \mathcal{F}, P)$.

Proposition 6.31. Let (Ω, \mathcal{F}, P) be a nonatomic probability space and \mathcal{Z} be a linear space of random variables $Z : \Omega \to \mathbb{R}$ such that $\mathcal{L}_p(\Omega, \mathcal{F}, P) \subset \mathcal{Z}$, for some $p \in [1, \infty)$, and such that if $Z \in \mathcal{Z}$, then $[Z]_+ \in \mathcal{Z}$. Suppose that there exists a real valued law invariant convex risk measure $\rho : \mathcal{Z} \to \mathbb{R}$. Then $\mathcal{Z} \subset \mathcal{L}_1(\Omega, \mathcal{F}, P)$.

Proof. We may assume without loss of generality that (Ω, \mathcal{F}, P) is the standard uniform probability space. We argue by a contradiction. Suppose that \mathcal{Z} is not a subspace of $\mathcal{L}_1(\Omega, \mathcal{F}, P)$, i.e., there exists $\overline{Z} \in \mathcal{Z} \setminus \mathcal{L}_1(\Omega, \mathcal{F}, P)$. Then $\overline{Z} = [\overline{Z}]_+ - [-\overline{Z}]_+$ and at least one of the integrals $\int_0^1 [\overline{Z}(t)]_+ dt$ or $\int_0^1 [-\overline{Z}]_+(t)dt$ is $+\infty$. Hence we can take $\overline{Z} \in \mathcal{Z} \setminus \mathcal{L}_1(\Omega, \mathcal{F}, P)$ such that $\overline{Z} \ge 0$ and $\int_0^1 \overline{Z}(t)dt = +\infty$.

Now let $\hat{\rho} : \mathcal{L}_p(\Omega, \mathcal{F}, P) \to \mathbb{R}$ be the restriction of ρ to the space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$. Since $\hat{\rho}$ is a real valued convex risk measure, by Proposition 6.6 $\hat{\rho}$ is continuous, and thus has the dual representation (6.37). Moreover, we have that (see equation (6.126) below)

$$\hat{\rho}(Z) = \sup_{\sigma \in \mathfrak{F}_q} \left\{ \int_0^1 \sigma(t) H_Z^{-1}(t) dt - \hat{\rho}^*(\sigma) \right\}, \ Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P),$$
(6.114)

where \mathfrak{F}_q is the set of right side continuous, monotonically nondecreasing functions σ : $[0,1) \to \mathbb{R}_+$ such $\int_0^1 \sigma(t) dt = 1$ and $\int_0^1 \sigma(t)^q dt < \infty$ (see Definition 6.32). Since $\hat{\rho}$ is real valued, there is $\bar{\sigma} \in \mathfrak{F}_q$ such that $\kappa := \hat{\rho}^*(\bar{\sigma})$ is finite. It follows that

$$\hat{\rho}(Z) \ge \int_0^1 \bar{\sigma}(t) H_Z^{-1}(t) dt - \kappa, \ Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P).$$
(6.115)

Consider the sequence $Z_n(\omega) := \min\{\overline{Z}(\omega), n\}$, $n \in \mathbb{N}$. Clearly $Z_n \ge 0$ and is bounded and hence $Z_n \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$. We are going to show that $\hat{\rho}(Z_n)$ tends to $+\infty$ as $n \to \infty$. Since $\overline{Z} \ge Z_n$, and hence $\rho(\overline{Z}) \ge \rho(Z_n) = \hat{\rho}(Z_n)$, this will imply that $\rho(\overline{Z}) = +\infty$, the required contradiction.

Consider Z'_n distributionally equivalent to Z_n and monotonically nondecreasing (such Z'_n does exist, see Remark 30). Since $\int_0^1 \bar{\sigma}(t) dt = 1$ and $\bar{\sigma}(\cdot)$ is monotonically nondecreasing, it follows that there exist constants $a \in (0, 1)$ and b > 0 such that $\bar{\sigma}(t) \ge b$ for all $t \in [a, 1]$. Together with (6.115) this implies

$$\hat{\rho}(Z'_n) \ge \int_0^1 \bar{\sigma}(t) Z'_n(t) dt - \kappa \ge b \int_a^1 Z'_n(t) dt - \kappa \ge b(1-a) \int_0^1 Z'_n(t) dt - \kappa,$$

where the last inequality follows since $Z'_n(\cdot)$ is monotonically nondecreasing. Recall that $\int_0^1 \overline{Z}(t) dt = +\infty$. It follows that $\int_0^1 Z_n(t) dt$, and hence $\int_0^1 Z'_n(t) dt$, tends to $+\infty$. Thus $\hat{\rho}(Z_n)$ tends to $+\infty$ as well. This completes the proof. \Box

6.3.4 Spectral Risk Measures

Unless stated otherwise we assume in this section that (Ω, \mathcal{F}, P) is the standard uniform probability space and $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P), p \in [1, \infty)$. Consider a (real valued) law invariant coherent risk measure $\rho : \mathcal{Z} \to \mathbb{R}$. As we know it has the dual representation

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \int_0^1 \zeta(t) Z(t) dt, \quad Z \in \mathcal{Z}.$$
(6.116)

Since $Z \stackrel{\mathcal{D}}{\sim} H_Z^{-1}$, and $\rho(Z) = \rho(Z \circ T)$ for any $Z \in \mathcal{Z}$ and measure-preserving transformation $T \in \mathfrak{G}$, we can write

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}, T \in \mathfrak{G}} \int_0^1 \zeta(T(t)) H_Z^{-1}(t) dt, \ Z \in \mathcal{Z}.$$
(6.117)

Recall that here the dual set \mathfrak{A} is invariant with respect to measure-preserving transformations (see Corollary 6.30). It is also not difficult to see that since $H_Z^{-1}(\cdot)$ is monotonically nondecreasing, the maximum in the right hand side of (6.117), with respect to $T \in \mathfrak{G}$, is attained if $\zeta' = \zeta \circ T$ is monotonically nondecreasing. This motivates the following definitions.

Definition 6.32. We say that $\sigma : [0,1) \to \mathbb{R}_+$ is a spectral function if $\sigma(\cdot)$ is right side continuous, monotonically nondecreasing and such that $\int_0^1 \sigma(t)dt = 1$. We denote by \mathfrak{F} the set of spectral functions, and by $\mathfrak{F}_q := \mathfrak{F} \cap \mathcal{L}_q(\Omega, \mathcal{F}, P)$ the set of spectral functions with finite q-th moment.

Definition 6.33. We say that a set $\Upsilon \subset \mathfrak{F}_q$ is a generating set of proper lower semicontinuous law invariant coherent risk measure $\rho : \mathbb{Z} \to \overline{\mathbb{R}}$ if

$$\rho(Z) = \sup_{\sigma \in \Upsilon} \int_0^1 \sigma(t) H_Z^{-1}(t) dt, \quad Z \in \mathcal{Z}.$$
(6.118)

Recall that if $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is a lower semicontinuous coherent risk measure and $\mathfrak{C} \subset \mathcal{Z}^*$ is such that

$$\rho(Z) = \sup_{\zeta \in \mathfrak{C}} \langle \zeta, Z \rangle, \ \ Z \in \mathcal{Z},$$
(6.119)

then \mathfrak{C} is a subset of the dual set \mathfrak{A} . Therefore it follows from (6.118) that $\Upsilon \subset \mathfrak{A}$. By the above discussion such generating set Υ always exists. However, it is not defined uniquely. Indeed, let us observe that if $\sigma_1, \sigma_2 \in \mathfrak{F}_q$ are two spectral functions such that¹⁵

$$\int_{\eta}^{1} \sigma_{1}(t)dt \leq \int_{\eta}^{1} \sigma_{2}(t)dt, \ \forall \eta \in (0,1),$$
(6.120)

then

$$\int_{0}^{1} \sigma_{1}(t) H_{Z}^{-1}(t) dt \leq \int_{0}^{1} \sigma_{2}(t) H_{Z}^{-1}(t) dt, \quad Z \in \mathcal{Z}.$$
(6.121)

Therefore we always can add to the generating set Υ a spectral function which is dominated in the increasing convex order by one of the spectral functions already in Υ . In particular we can take as generating set the set of all spectral functions $\sigma \in \mathfrak{A}$. This raises the question of existence in some sense of minimal generating set. We address this question below.

Proposition 6.34. Let $\rho : \mathbb{Z} \to \mathbb{R}$ be a (real valued) law invariant coherent risk measure and $\operatorname{Exp}(\mathfrak{A})$ be the set of exposed points of its dual set \mathfrak{A} . Then $\operatorname{Exp}(\mathfrak{A})$ is invariant under measure-preserving transformations and the representation (6.119) holds with $\mathfrak{C} :=$ $\operatorname{Exp}(\mathfrak{A})$. Moreover, if the representation (6.119) holds for some weakly^{*} closed set \mathfrak{C} , then $\operatorname{Exp}(\mathfrak{A}) \subset \mathfrak{C}$.

Proof. Recall that by Corollary 6.30, for $T \in \mathfrak{G}$ we have that

$$\langle \zeta \circ T, Z \rangle = \int_0^1 \zeta(T(t)) Z(t) d\tau = \int_0^1 \zeta(t) Z(T^{-1}(t)) dt = \langle \zeta, Z \circ T^{-1} \rangle.$$

Thus $\overline{\zeta} \in \mathfrak{A}$ is a maximizer of $\langle \zeta, Z \rangle$ over $\zeta \in \mathfrak{A}$, iff $\overline{\zeta} \circ T$ is a maximizer of $\langle \zeta, T^{-1} \circ Z \rangle$ over $\zeta \in \mathfrak{A}$. Consequently we have that if $\overline{\zeta} \in \operatorname{Exp}(\mathfrak{A})$, then $\overline{\zeta} \circ T \in \operatorname{Exp}(\mathfrak{A})$. This shows that $\operatorname{Exp}(\mathfrak{A})$ is invariant under \mathfrak{G} .

Now consider the set

$$\mathcal{D} := \{ Z \in \mathcal{Z} : Z \text{ exposes } \mathfrak{A} \text{ at a point } \overline{\zeta} \}.$$

By Theorem 7.81 this set is dense in \mathcal{Z} . So for $Z \in \mathcal{Z}$ fixed, let $\{Z_n\} \subset \mathcal{D}$ be a sequence of points converging (in the norm topology) to Z. Let $\{\zeta_n\} \subset \operatorname{Exp}(\mathfrak{A})$ be a sequence of the corresponding maximizers, i.e., $\rho(Z_n) = \langle \zeta_n, Z_n \rangle$. Since \mathfrak{A} is bounded, we have that $\|\zeta_n\|^*$ is uniformly bounded. Since $\rho : \mathcal{Z} \to \mathbb{R}$ is real valued it is continuous, and thus $\rho(Z_n) \to \rho(Z)$. We also have that

$$|\rho(Z_n) - \langle \zeta_n, Z \rangle| = |\langle \zeta_n, Z_n \rangle - \langle \zeta_n, Z \rangle| \le ||\zeta_n||^* ||Z_n - Z|| \to 0.$$

¹⁵As we will discuss later, condition (6.120) means that σ_2 dominates σ_1 in the increasing convex order.

It follows that

$$\rho(Z) = \sup\{\langle \zeta_n, Z \rangle : n = 1, 2, \dots\},\$$

and hence the representation (6.119) holds with $\mathfrak{C} = \operatorname{Exp}(\mathfrak{A})$.

Let \mathfrak{C} be a weakly* closed set such that the representation (6.119) holds. It follows that \mathfrak{C} is a subset of \mathfrak{A} and is weakly* compact. Consider a point $\zeta \in \operatorname{Exp}(\mathfrak{A})$. By the definition of the set $\operatorname{Exp}(\mathfrak{A})$, there is $Z \in \mathcal{D}$ such that $\rho(Z) = \langle \zeta, Z \rangle$. Since \mathfrak{C} is weakly* compact, the maximum in (6.119) is attained and hence $\rho(Z) = \langle \zeta', Z \rangle$ for some $\zeta' \in \mathfrak{C}$. By the uniqueness of the maximizer ζ it follows that $\zeta' = \zeta$. This shows that $\operatorname{Exp}(\mathfrak{A}) \subset \mathfrak{C}$. \Box

The result of Proposition 6.34 shows that the set $\text{Exp}(\mathfrak{A}) \cap \mathfrak{F}$, of monotonically nondecreasing elements of $\text{Exp}(\mathfrak{A})$, is a minimal generating set in the following sense.

Theorem 6.35. Let $\rho : \mathbb{Z} \to \mathbb{R}$ be a (real valued) law invariant coherent risk measure and $\operatorname{Exp}(\mathfrak{A})$ be the set of exposed points of its dual set \mathfrak{A} . Then the set $\overline{\Upsilon} := \operatorname{Exp}(\mathfrak{A}) \cap \mathfrak{F}$ is a generating set of ρ , and moreover this generating set is minimal in the sense that it is contained in any weakly^{*} closed generating set.

Remark 31. Recall that $\overline{\zeta}$ is an exposed point of \mathfrak{A} iff $\partial \rho(Z) = {\overline{\zeta}}$ is a singleton for some $Z \in \mathcal{Z}$ (see Remark 26 on page 304). Therefore in order to compute $\operatorname{Exp}(\mathfrak{A})$ we can use formulas derived in section 6.3.2 for subdifferentials of coherent risk measures.

Consider $\rho := AV@R_{1-\alpha}$, $\alpha \in [0, 1)$, risk measure.¹⁶ A general formula for subdifferentials of the Average Value-at-Risk measures is given in equation (6.80). By this formula we have that the set $Exp(\mathfrak{A})$, of exposed points of the dual set of ρ , consists of functions $(1 - \alpha)^{-1} \mathbf{1}_A$ with $A \subset [0, 1]$ being set of measure $1 - \alpha$. Thus the minimal generating set $\tilde{\Upsilon} = Exp(\mathfrak{A}) \cap \mathfrak{F}$ consists of unique spectral function $(1 - \alpha)^{-1} \mathbf{1}_{[\alpha, 1]}$. We can write the corresponding spectral representation as

$$\mathsf{AV}@\mathsf{R}_{1-\alpha}(Z) = \frac{1}{1-\alpha} \int_{\alpha}^{1} H_Z^{-1}(t) dt.$$
(6.122)

Recall that $H_Z^{-1}(t) = V@R_{1-t}(Z)$. Therefore the above formula (6.122) is equivalent to (6.27).

Definition 6.36. It is said that risk measure $\rho : \mathbb{Z} \to \mathbb{R}$ is spectral, with spectral function $\sigma \in \mathfrak{F}_q$, if

$$\rho(Z) = \int_0^1 \sigma(t) H_Z^{-1}(t) dt, \quad Z \in \mathcal{Z}.$$
(6.123)

That is, a spectral risk measure $\rho : \mathbb{Z} \to \mathbb{R}$, with spectral function $\sigma \in \mathfrak{F}_q$, is a (real valued) law invariant coherent risk measure whose minimal generating set is the singleton $\{\sigma\}$. Note that since $H_Z^{-1} \in \mathbb{Z}$ and the spectral function σ is supposed to belong to the

¹⁶It will be convenient here to deal with AV@R_{1- α}, rather than AV@R_{α}, formulation of Average Value-at-Risk measures. We will comment on that later.

dual space $\mathcal{Z}^* = \mathcal{L}_q(\Omega, \mathcal{F}, P)$, the integral in the right hand side of (6.123) is well defined and finite valued.

Remark 32. Consider now a proper lower semicontinuous law invariant *convex* risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$. Since $Z \in \mathcal{Z}$ is distributionally equivalent to H_Z^{-1} and ρ is law invariant we have that $\rho(Z) = \rho(H_Z^{-1})$. Thus by invoking Fenchel-Moreau Theorem we can write

$$\rho(Z) = \sup_{\zeta \in \mathcal{Z}^*} \left\{ \int_0^1 \zeta(t) H_Z^{-1}(t) dt - \rho^*(\zeta) \right\}, \ Z \in \mathcal{Z}.$$
 (6.124)

Recall that dom(ρ^*) is a subset of the set \mathfrak{P} of density functions (see Theorem 6.5). Now for any $\zeta \in \mathcal{Z}^*$ and measure-preserving transformation $T \in \mathfrak{G}$ we have by Proposition 6.29 that $\rho^*(\zeta \circ T) = \rho^*(\zeta)$. Therefore we can write (6.124) as

$$\rho(Z) = \sup_{\zeta \in \mathfrak{P} \cap \mathcal{Z}^*, T \in \mathfrak{G}} \left\{ \int_0^1 \zeta(T(t)) H_Z^{-1}(t) dt - \rho^*(\zeta \circ T) \right\}, \ Z \in \mathcal{Z}.$$
(6.125)

As it was pointed before, since $H_Z^{-1}(\cdot)$ is monotonically nondecreasing, the maximum in the right hand side of (6.125) with respect to $T \in \mathfrak{G}$ is attained if $\zeta' = \zeta \circ T$ is monotonically nondecreasing. Note that \mathfrak{F}_q is a subset of $\mathfrak{P} \cap \mathcal{Z}^*$ consisting of monotonically nondecreasing right side continuous functions of $\mathfrak{P} \cap \mathcal{Z}^*$. Therefore it follows that

$$\rho(Z) = \sup_{\sigma \in \mathfrak{F}_q} \left\{ \int_0^1 \sigma(t) H_Z^{-1}(t) dt - \rho^*(\sigma) \right\}, \ Z \in \mathcal{Z}.$$
 (6.126)

In a similar way, by noting that for $\sigma \in \mathfrak{F}_q$ we have that $\sigma = H_{\sigma}^{-1}$ (see Remark 30 on page 321), the conjugate ρ^* can be written as

$$\rho^*(\sigma) = \sup_{Y \in \mathfrak{F}_p} \left\{ \int_0^1 Y(t)\sigma(t)dt - \rho(Y) \right\}, \ \sigma \in \mathfrak{F}_q.$$
(6.127)

Remark 33. Sometimes we will need a property slightly stronger than the monotonicity condition (R2). For two random variables $Z, Z' \in \mathcal{Z}$ we write $Z \succ Z'$ if $Z \succeq Z'$ and $Z \neq Z'$, i.e., if $Z \succeq Z'$ and Z > Z' with positive probability. Consider the following condition.

(**R***2) Strict Monotonicity: If $Z, Z' \in \mathcal{Z}$ and $Z \succ Z'$, then $\rho(Z) > \rho(Z')$.

Proposition 6.37. Let $\rho : \mathbb{Z} \to \mathbb{R}$ be a spectral risk measure with spectral function $\sigma \in \mathfrak{F}_q$. Then the strict monotonicity condition (\mathbb{R}^*2) holds iff $\sigma(t) > 0$ for all $t \in (0, 1)$.

Proof. Let $Z \succ Z'$, i.e., $Z - Z' \succeq 0$ and Z(t) - Z'(t) > 0 for $t \in A \subset (0,1)$ with P(A) > 0. Suppose that $\sigma(t) > 0$ for all $t \in (0,1)$. It follows that $\sigma(Z - Z') \succeq 0$ and $\sigma(t)[Z(t) - Z'(t)] > 0$ for $t \in A$, and hence

$$\int_0^1 \sigma(t) [Z(t) - Z'(t)] dt > 0.$$

By applying a measure preserving transformation $T \in \mathfrak{G}$ to Z and Z' if necessary, we can assume that Z' is monotonically nondecreasing, i.e., $Z' = H_{Z'}^{-1}$. Then

$$\rho(Z) = \int_0^1 \sigma(t) H_Z^{-1}(t) dt \ge \int_0^1 \sigma(t) Z(t) dt > \int_0^1 \sigma(t) Z'(t) dt = \rho(Z').$$

Conversely suppose that $\sigma(t) = 0$ for all $t \in (0, a)$ and some $a \in (0, 1)$. Consider $Z := \mathbf{1}_{[0,1]}$ and $Z' := \mathbf{1}_{[a,1]}$. Clearly $Z \succ Z'$, while $\rho(Z) = \rho(Z')$. \Box

By the above proposition we have that risk measure $\rho := AV@R_{\alpha}, \alpha \in (0, 1)$, does not satisfy the strict monotonicity condition (R*2). On the other hand risk measure $\rho(\cdot) := (1 - \beta)\mathbb{E}[\cdot] + \beta AV@R_{\alpha}(\cdot)$ does satisfy (R*2) for any $\alpha \in (0, 1)$ and $\beta \in [0, 1)$.

For a general coherent risk measure we can give the following sufficient conditions for strict monotonicity. Recall that if $\rho : \mathbb{Z} \to \mathbb{R}$ is a real valued coherent risk measure, than it is continuous, subdifferentiable and for any $Z \in \mathbb{Z}$,

$$\partial \rho(Z) = \operatorname*{argmax}_{\zeta \in \mathfrak{A}} \int_{\Omega} \zeta(\omega) Z(\omega) dP.$$
 (6.128)

Proposition 6.38. Let $\rho : \mathbb{Z} \to \mathbb{R}$ be a coherent risk measure. Suppose that for every $Z \in \mathbb{Z}$ there is $\zeta \in \partial \rho(Z)$ such that $\zeta(\omega) > 0$ for a.e. $\omega \in \Omega$. Then the strict monotonicity condition ($\mathbb{R}^{\star}2$) holds.

Proof. Let $Z \succ Z'$ and $\zeta \in \partial \rho(Z')$ be such that $\zeta(\omega) > 0$ for a.e. $\omega \in \Omega$. Then

$$\rho(Z') = \int_{\Omega} \zeta(\omega) Z'(\omega) dP < \int_{\Omega} \zeta(\omega) Z(\omega) dP \le \rho(Z).$$

This completes the proof. \Box

As an example consider mean-upper-semideviation risk measures $\rho(Z) = \mathbb{E}[Z] + c (\mathbb{E}[[Z - \mathbb{E}[Z]]_+^p])^{1/p}$. Formulas for subdifferentials $\partial \rho(Z)$ of these risk measures were derived in Example 6.23. By these formulas conditions of Proposition 6.38 hold, and hence these risk measures satisfy the strict monotonicity condition (\mathbb{R}^*2) for $p \in [1, \infty)$ and $c \in [0, 1)$.

6.3.5 Kusuoka Representations

As it was discussed in the previous section, $AV@R_{1-\alpha}$ is a spectral risk measure with spectral function $\sigma = (1 - \alpha)^{-1} \mathbf{1}_{[\alpha, 1]}$. Consider now a convex combination

$$\rho(Z) := \lambda_1 \mathsf{AV} @\mathsf{R}_{1-\alpha_1}(Z) + \dots + \lambda_m \mathsf{AV} @\mathsf{R}_{1-\alpha_m}(Z), \ Z \in \mathcal{Z},$$
(6.129)

of Average Value-at Risk measures, with $0 \le \alpha_1 < \alpha_2 < \cdots < \alpha_m < 1$ and $\lambda_i > 0$, $i = 1, \ldots, m, \sum_{i=1}^m \lambda_i = 1$, and $\mathcal{Z} = \mathcal{L}_1(\Omega, \mathcal{F}, P)$. This is also a spectral risk measure with spectral function $\sigma = \sum_{i=1}^m \lambda_i \sigma_i$, where σ_i is the spectral function of AV@R_{1- α_i}. That is,

$$\sigma = \sum_{i=1}^{m} \lambda_i (1 - \alpha_i)^{-1} \mathbf{1}_{[\alpha_i, 1]}.$$
(6.130)

We can view $\rho(Z)$, defined in (6.129), as the integral

$$\rho(Z) = \int_0^1 \mathsf{AV}@\mathsf{R}_{1-\alpha}(Z)d\mu(\alpha), \ Z \in \mathcal{Z},$$
(6.131)

with discrete probability measure¹⁷ $\mu := \sum_{i=1}^{m} \lambda_i \delta(\alpha_i)$. Such integral representation can be extended to general spectral functions. The integral in the right hand side of (6.131) can be viewed as the Lebesgue-Stieltjes integral with $\mu : \mathbb{R} \to [0, 1]$ being right side continuous monotonically nondecreasing function (distribution function) such that $\mu(\alpha) = 0$ for $\alpha < 0$ and $\mu(\alpha) = 1$ for $\alpha \ge 1$. For the discrete measure $\mu = \sum_{i=1}^{m} \lambda_i \delta(\alpha_i)$ the corresponding distribution function is $\mu(\cdot) = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{[\alpha_i,\infty)}(\cdot)$.

Consider the following transformation (linear mapping) $\mathbb{T}\mu$, from the set of probability measures (distribution functions) $\mu(\cdot)$ on the interval¹⁸ [0, 1) to the set \mathfrak{F} of spectral functions:

$$(\mathbb{T}\mu)(t) := \int_0^t (1-\alpha)^{-1} d\mu(\alpha), \ t \in [0,1).$$
(6.132)

It is not difficult to verify that indeed the function $\sigma = \mathbb{T}\mu$ is a spectral function. That is, clearly $\sigma(\cdot)$ is nonnegative valued, monotonically nondecreasing, right side continuous and

$$\int_0^1 \sigma(t)dt = \int_0^1 \int_0^t (1-\alpha)^{-1} d\mu(\alpha)dt = \int_0^1 \int_\alpha^1 (1-\alpha)^{-1} dt d\mu(\alpha) = \int_0^1 d\mu(\alpha) = 1.$$

Let us show that the inverse $\mu = \mathbb{T}^{-1}\sigma$, of the transformation defined in (6.132) can be written as

$$\mu(\alpha) = (1 - \alpha)\sigma(\alpha) + \int_0^\alpha \sigma(t)dt, \ \alpha \in [0, 1).$$
(6.133)

Indeed,¹⁹ $\mu'_+(\alpha) = (1 - \alpha)\sigma'_+(\alpha) \ge 0$ at points where $\sigma(\cdot)$ is continuous. It follows that $\mu(\alpha)$ is monotonically nondecreasing. Also $\sigma(\cdot)$ is right side continuous, and hence $\mu(\cdot)$ is right side continuous. Clearly $\mu(\alpha) \ge 0$. Since $\int_0^1 \sigma(t)dt = 1$ is finite, it follows that $\lim_{\alpha\uparrow 1}(1 - \alpha)\sigma(\alpha) = 0$, and hence $\lim_{\alpha\uparrow 1}\mu(\alpha) = \int_0^1 \sigma(t)dt = 1$. This shows that $\mu(\cdot)$ is a probability distribution function on the interval [0, 1). Note that $\mu(0) = \sigma(0)$, and thus μ has positive mass at $\alpha = 0$ if $\sigma(0) > 0$. If $\mu(0) = \sigma(0)$ is positive, we understand the integrals in (6.132) and (6.133) as taken from 0^- . Thus

$$\int_0^\alpha d\mu(\tau) = \mu(\alpha), \ \alpha \in [0,1].$$

In order to show that transformation (6.133) is inverse of (6.132), let us substitute in the right hand side of (6.133) the expression for $\sigma(\cdot)$ given in (6.132). That is

¹⁷Recall that $\delta(\alpha)$ denotes measure of mass one at α .

¹⁸By saying that measure μ is on the interval [0, 1) we emphasize that μ has mass zero at $\alpha = 1$.

¹⁹By $\mu'_{+}(\alpha) = \liminf_{h \downarrow 0} \frac{\mu(\alpha+h) - \mu(\alpha)}{h}$ we denote here the lower right side derivative.

$$(1-\alpha)\int_0^{\alpha} (1-\tau)^{-1} d\mu(\tau) + \int_0^{\alpha} \int_0^t (1-\tau)^{-1} d\mu(\tau) dt = (1-\alpha)\int_0^{\alpha} (1-\tau)^{-1} d\mu(\tau) + \int_0^{\alpha} \int_{\tau}^{\alpha} (1-\tau)^{-1} dt d\mu(\tau) = (1-\alpha)\int_0^{\alpha} (1-\tau)^{-1} d\mu(\tau) + \int_0^{\alpha} (1-\tau)^{-1} (\alpha-\tau) d\mu(\tau) = \int_0^{\alpha} d\mu(\tau) = \mu(\alpha).$$

This shows that indeed transformations (6.133) and (6.132) are inverse to each other. It follows that the mapping \mathbb{T} is one-to-one and onto.

Now let us consider spectral risk measure $\rho : \mathcal{Z} \to \mathbb{R}$ with $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$, and the corresponding spectral function $\sigma \in \mathfrak{F}_q$. The mapping \mathbb{T} suggests a one-to-one correspondence between spectral risk measures and integral representations of the form (6.131).

Theorem 6.39. *Transformation* (6.132) *gives a one-to-one correspondence between spectral risk measures and integral representations of the form* (6.131).

Proof. Consider a spectral risk measure $\rho(Z)$, of the form (6.123), with the corresponding spectral function $\sigma \in \mathfrak{F}_q$. Let $\mu := \mathbb{T}^{-1}\sigma$, and thus $\sigma = \mathbb{T}\mu$. Then

$$\rho(Z) = \int_0^1 (\mathbb{T}\mu)(t) H_Z^{-1}(t) dt = \int_0^1 \int_0^t (1-\alpha)^{-1} H_Z^{-1}(t) d\mu(\alpha) dt$$

=
$$\int_0^1 \int_\alpha^1 (1-\alpha)^{-1} H_Z^{-1}(t) dt d\mu(\alpha) = \int_0^1 \mathsf{AV} @\mathsf{R}_{1-\alpha}(Z) d\mu(\alpha),$$

where the last equality follows by (6.122).

We have that any proper lower semicontinuous law invariant coherent risk measure $\rho : \mathbb{Z} \to \mathbb{R}$ can be represented as maximum of spectral risk measures taken over spectral functions in the corresponding generating set. Together with Theorem 6.39 this gives the following result.

Theorem 6.40 (Kusuoka). Suppose that the probability space (Ω, \mathcal{F}, P) is nonatomic and let $\rho : \mathbb{Z} \to \mathbb{R}$ be a proper lower semicontinuous law invariant coherent risk measure. Then there exists a set \mathfrak{M} of probability measures on the interval [0, 1) such that

$$\rho(Z) = \sup_{\mu \in \mathfrak{M}} \int_0^1 \mathsf{AV}@\mathsf{R}_{1-\alpha}(Z) d\mu(\alpha), \ \forall Z \in \mathcal{Z}.$$
(6.134)

We refer to (6.134) as the *Kusuoka representation* of the law invariant coherent risk measure ρ . Of course, by making change of variables $\alpha \mapsto 1 - \alpha$ we can write representation (6.134) as

$$\rho(Z) = \sup_{\mu \in \mathfrak{M}'} \int_0^1 \mathsf{AV}@\mathsf{R}_{\alpha}(Z) d\mu(\alpha), \quad \forall Z \in \mathcal{Z},$$
(6.135)

with \mathfrak{M}' being a set of probability measures on the interval (0, 1].

Remark 34. As it was mentioned before, the generating set is not defined uniquely. Nevertheless, by Theorem 6.35 we have that for a real valued law invariant coherent risk measure $\rho : \mathbb{Z} \to \mathbb{R}$, the set $\overline{\Upsilon} := \operatorname{Exp}(\mathfrak{A}) \cap \mathfrak{F}$ is in a sense a minimal generating set of ρ . By taking the set of probability measures $\overline{\mathfrak{M}} := \mathbb{T}^{-1}(\overline{\Upsilon})$ on [0, 1), we obtain, in a sense, a minimal (Kusuoka) representation of the form (6.134). Recall that the transformation \mathbb{T} and its inverse \mathbb{T}^{-1} are defined in (6.132) and (6.133), respectively. In particular, if $\mu = \sum_{i=1}^{m} \lambda_i \delta(\alpha_i)$, where λ_i are positive numbers such that $\sum_{i=1}^{m} \lambda_i = 1$, and $0 \le \alpha_1 < \alpha_2 < \cdots < \alpha_m < 1$, then $\sigma = \mathbb{T}\mu$ is given in (6.130).

Example 6.41 Consider the absolute semideviation risk measure

$$\rho(Z) := \mathbb{E}[Z] + c \mathbb{E}\left\{\left[Z - \mathbb{E}[Z]\right]_{+}\right\},\tag{6.136}$$

where $c \in [0, 1]$ and $\mathcal{Z} = \mathcal{L}_1(\Omega, \mathcal{F}, P)$. Recall that $\overline{\zeta}$ is an exposed point of \mathfrak{A} iff $\partial \rho(Z) = {\overline{\zeta}}$ is a singleton for some $Z \in \mathcal{Z}$. The subdifferential of $\rho(Z)$ was computed in Example 6.23. It follows from formula (6.106) that if $Z \neq \mathbb{E}[Z]$ w.p.1, then Z exposes \mathfrak{A} at the point

$$\bar{\zeta}(\omega) = \begin{cases} 1 - c\kappa, & \text{if } Z(\omega) < \mathbb{E}[Z], \\ 1 + c(1 - \kappa), & \text{if } Z(\omega) > \mathbb{E}[Z], \end{cases}$$
(6.137)

where $\kappa := \Pr(Z > \mathbb{E}[Z])$. Consequently the set $\operatorname{Exp}(\mathfrak{A}) \cap \mathfrak{F}$, of exposed spectral points, is given by $\{\sigma_{\kappa} : \kappa \in (0, 1)\}$, where

$$\sigma_{\kappa}(t) = \begin{cases} 1 - c\kappa, & \text{if } 0 \le t < 1 - \kappa, \\ 1 + c(1 - \kappa), & \text{if } 1 - \kappa \le t \le 1. \end{cases}$$
(6.138)

Each σ_{κ} is a step function, hence by (6.130) each $\mu_{\kappa} = \mathbb{T}^{-1} \sigma_{\kappa}$ can be written as

$$\mu_{\kappa} = (1 - c\kappa)\delta(0) + c\kappa\delta(1 - \kappa).$$

Consequently the set \mathfrak{M} in Kusuoka representation (6.134) is

$$\mathfrak{M} = \bigcup_{\kappa \in (0,1)} \left\{ (1 - c\kappa)\delta(0) + c\kappa\delta(1 - \kappa) \right\}.$$

As it was pointed above, in a sense this set is minimal. The corresponding (minimal) Kusuoka representation is given by

$$\rho(Z) = \sup_{\kappa \in (0,1)} \left\{ (1 - c\kappa) \mathsf{AV} @ \mathsf{R}_1(Z) + c\kappa \mathsf{AV} @ \mathsf{R}_\kappa(Z) \right\}.$$
(6.139)

Recalling that $AV@R_1(\cdot) = \mathbb{E}(\cdot)$, and hence using definition (6.23) of $AV@R_\alpha$ we can write (6.139) as

$$\rho(Z) = \sup_{\substack{\kappa \in [0,1] \ t \in \mathbb{R} \\ t \in \mathbb{R}, \ \kappa \in [0,1]}} \inf_{t \in \mathbb{R}} \mathbb{E} \left\{ Z + c\kappa(t-Z) + c[Z-t]_+ \right\}$$

$$= \inf_{\substack{t \in \mathbb{R}, \ \kappa \in [0,1]}} \mathbb{E} \left\{ Z + c\kappa(t-Z) + c[Z-t]_+ \right\}.$$
(6.140)

The above formula is equivalent to formula (6.29), derived in Corollary 6.3, for $\sigma_1^+[Z]$.

Remark 35. So far we dealt with spaces $\mathcal{L}_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$. Consider now space $\mathcal{Z} = \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$ paired with space $\mathcal{L}_1(\Omega, \mathcal{F}, P)$ and a law invariant coherent risk measure $\rho : \mathcal{Z} \to \mathbb{R}$. Suppose that ρ is lower semicontinuous with respect to the weak^{*} topology, and hence has the dual representation

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \int_0^1 \zeta(t) Z(t) dt, \tag{6.141}$$

for some set $\mathfrak{A} \subset \mathcal{L}_1(\Omega, \mathcal{F}, P)$. We still can write such dual representation in the form (6.118) for some generating set $\Upsilon \subset \mathfrak{A} \cap \mathfrak{F}$. Consequently by using the transformation $\mu = \mathbb{T}^{-1}\mu$, given in (6.133), we obtain Kusuoka representation (6.134) with the set \mathfrak{M} being a set of probability measures on the interval [0, 1).

For example, consider the essential supremum risk measure $\rho(Z) := \operatorname{ess\,sup}(Z)$. It has Kusuoka representation (6.134) with the set \mathfrak{M} consisting, for example, from measures $\delta(\alpha_n)$ such that $\alpha_n \uparrow 1$. Of course if we define AV@R₀ := ess sup (see Remark 24 on page 296), then we can use the singleton set $\mathfrak{M} = \{\delta(1)\}$.

Comonotonicity

Let X and Y be two random variables with respective cdfs H_X and H_Y , and $H(x, y) := \Pr(X \le x, Y \le y)$ be their joint cdf. It is said that X and Y are *comonotonic* if $(X, Y) \stackrel{\mathcal{D}}{\sim} (H_X^{-1}(U), H_Y^{-1}(U))$, where U is a random variable uniformly distributed on the interval [0, 1].

Definition 6.42. It is said that a risk measure $\rho : \mathbb{Z} \to \mathbb{R}$ is comonotonic if for any two comonotonic random variables $X, Y \in \mathbb{Z}$ it follows that $\rho(X + Y) = \rho(X) + \rho(Y)$.

Let us observe that V@R_{α}, $\alpha \in (0, 1)$, risk measure is comonotonic. Indeed, let X and Y be comonotonic random variables. We can assume that $X = H_X^{-1}(U)$ and $Y = H_Y^{-1}(U)$, with U being a random variable uniformly distributed on the interval [0, 1]. Consider function $g(\cdot) := H_X^{-1}(\cdot) + H_Y^{-1}(\cdot)$. Note that $g : (0, 1) \to \mathbb{R}$ is a monotonically nondecreasing left side continuous function. Then

$$V@R_{\alpha}(X+Y) = \inf \{t : \Pr(g(U) \le t) \ge 1 - \alpha \} \\ = \inf \{t : \Pr(U \le g^{-1}(t)) \ge 1 - \alpha \} \\ = \inf \{t : g^{-1}(t) \ge 1 - \alpha \} \\ = g(1-\alpha) = H_X^{-1}(1-\alpha) + H_Y^{-1}(1-\alpha) \\ = V@R_{\alpha}(X) + V@R_{\alpha}(Y).$$

By (6.27) it follows that if X and Y are comonotonic, then

$$\begin{aligned} \mathsf{AV}@\mathsf{R}_{\alpha}(X+Y) &= \frac{1}{\alpha} \int_{1-\alpha}^{1} \mathsf{V}@\mathsf{R}_{1-\tau}(X+Y) \, d\tau \\ &= \frac{1}{\alpha} \int_{1-\alpha}^{1} \left[\mathsf{V}@\mathsf{R}_{1-\tau}(X) + \mathsf{V}@\mathsf{R}_{1-\tau}(Y) \right] d\tau \\ &= \mathsf{AV}@\mathsf{R}_{\alpha}(X) + \mathsf{AV}@\mathsf{R}_{\alpha}(Y). \end{aligned}$$

That is, AV@R_{α} is comonotonic for all $\alpha \in (0, 1]$. Consequently if μ is a probability measure on the interval [0, 1), then the risk measure $\int_0^1 AV@R_{1-\alpha}(Z)d\mu(\alpha)$ is coherent

law invariant and comonotonic. By Theorem 6.39 it follows that a spectral risk measure is comonotonic. We show now that the converse is also true.

Theorem 6.43. Suppose that the probability space (Ω, \mathcal{F}, P) is nonatomic and let $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$, and $\rho : \mathcal{Z} \to \mathbb{R}$ be a law invariant, coherent risk measure. Then the following properties are equivalent: (i) Risk measure ρ is spectral. (ii) Risk measure ρ is comonotonic. (iii) There exists a (uniquely defined) probability measure μ on the interval [0, 1) such that

$$\rho(Z) = \int_0^1 \mathsf{AV}@\mathsf{R}_{1-\alpha}(Z)d\mu(\alpha), \ Z \in \mathcal{Z}.$$
(6.142)

Proof. Let $X, Y \in \mathcal{Z}$ be comonotonic random variables. We can assume that $(X, Y) \sim (H_X^{-1}(U), H_Y^{-1}(U))$, where U is a random variable uniformly distributed on the interval [0, 1]. It follows that $X + Y \sim H_X^{-1}(U) + H_Y^{-1}(U)$. Suppose that the risk measure ρ is spectral with spectral function $\sigma \in \mathfrak{F}_q$. Then

$$\begin{array}{lll} \rho(X+Y) &=& \int_0^1 \sigma(t) (H_X^{-1}(t) + H_Y^{-1}(t)) dt \\ &=& \int_0^1 \sigma(t) H_X^{-1}(t) dt + \int_0^1 \sigma(t) H_Y^{-1}(t) dt \\ &=& \rho(X) + \rho(Y). \end{array}$$

This proves the implication (i) \Rightarrow (ii).

Conversely, suppose that ρ is comonotonic. We can assume that the space (Ω, \mathcal{F}, P) is standard uniform. Consider the dual representation of ρ :

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \int_0^1 \zeta(t) Z(t) dt, \ Z \in \mathcal{Z}.$$
(6.143)

Recall that by Corollary 6.30 the dual set \mathfrak{A} is invariant with respect to measure-preserving transformations. Therefore ρ is spectral, with spectral function σ , iff for any $Z \in \mathcal{Z}$ the maximum in the right hand side of (6.143) is attained at an element of \mathfrak{A} distributionally equivalent to σ .

Now let us argue by a contradiction. Suppose that there exist $X, Y \in \mathbb{Z}$ such that the maximum in the right hand side of (6.143) for Z = X and Z = Y is attained at respective points of \mathfrak{A} which are not distributionally equivalent. Note that the set of left side continuous monotonically nondecreasing random variables in \mathbb{Z} can be represented by random variables $H_Z^{-1}(U), Z \in \mathbb{Z}$, with U being a random variable uniformly distributed on the interval [0, 1]. By applying appropriate measure-preserving transformations to $X, Y : [0, 1] \to \mathbb{R}$, we can assume that these functions are left side continuous monotonically nondecreasing, and hence are comonotonic. It follows that there exist two comonotonic random variables $X, Y \in \mathbb{Z}$ such that the maximum in the right hand side of (6.143) for Z = X and Z = Y is not attained at the same $\zeta \in \mathfrak{A}$. On the other hand for Z = X + Ythe maximum in the right hand side of (6.143) is attained at some $\overline{\zeta} \in \mathfrak{A}$, and hence

$$\rho(X+Y) = \int_0^1 \bar{\zeta}(t)(X(t)+Y(t))dt = \int_0^1 \bar{\zeta}(t)X(t)dt + \int_0^1 \bar{\zeta}(t)Y(t)dt.$$

Since $\bar{\zeta} \in \mathfrak{A}$ we have that

$$\int_0^1 \bar{\zeta}(t) X(t) dt \le \rho(X) \text{ and } \int_0^1 \bar{\zeta}(t) Y(t) dt \le \rho(Y),$$

with at least one of these inequalities being strict (this is because the maximum in the right hand side of (6.143) for Z = X and Z = Y is not attained at a same point of \mathfrak{A}). Thus $\rho(X + Y) < \rho(X) + \rho(Y)$, and hence ρ is not comonotonic. This proves the implication (ii) \Rightarrow (i).

The equivalence of (i) and (iii) was shown in Theorem 6.39. \Box

Convex Law Invariant Risk Measures

Consider a proper lower semicontinuous law invariant convex risk measure $\rho : \mathbb{Z} \to \mathbb{R}$. As it was discussed in Remark 32, $\rho(Z)$ has a dual representation of the form (6.126) with the corresponding conjugate $\rho^*(\sigma)$ of the form (6.127). By proceeding in the same way as in the derivation of Theorem 6.40 we obtain the following result.

Theorem 6.44. Suppose that the probability space (Ω, \mathcal{F}, P) is nonatomic and let $\rho : \mathbb{Z} \to \mathbb{R}$ be a proper lower semicontinuous law invariant convex risk measure. Then there exists a set \mathfrak{M} of probability measures on the interval [0, 1) such that

$$\rho(Z) = \sup_{\mu \in \mathfrak{M}} \left\{ \int_0^1 \mathsf{AV} @\mathsf{R}_{1-\alpha}(Z) d\mu(\alpha) - \beta(\mu) \right\}, \ Z \in \mathcal{Z},$$
(6.144)

where

$$\beta(\mu) := \sup_{Y \in \mathfrak{F}_p} \left\{ \int_0^1 \mathsf{AV}@\mathsf{R}_{1-\alpha}(Y) d\mu(\alpha) - \rho(Y) \right\}.$$
(6.145)

Proof. It was shown in Remark 32, on page 326, that $\rho(Z)$ has representation (6.126) with $\rho^*(\sigma)$ given in (6.127). Similar to the proof of Theorem 6.39, we can write for $\sigma = \mathbb{T}\mu$, where \mathbb{T} is defined in (6.132),

$$\int_{0}^{1} \sigma(t) H_{Z}^{-1}(t) dt - \rho^{*}(\sigma) = \int_{0}^{1} (\mathbb{T}\mu)(t) H_{Z}^{-1}(t) dt - \rho^{*}(\mathbb{T}\mu) \\
= \int_{0}^{1} \int_{0}^{1} (1 - \alpha)^{-1} H_{Z}^{-1}(t) d\mu(\alpha) dt - \rho^{*}(\mathbb{T}\mu) \\
= \int_{0}^{1} \int_{\alpha}^{1} (1 - \alpha)^{-1} H_{Z}^{-1}(t) dt d\mu(\alpha) - \rho^{*}(\mathbb{T}\mu) \\
= \int_{0}^{1} A \mathsf{V}@R_{1-\alpha}(Z) d\mu(\alpha) - \rho^{*}(\mathbb{T}\mu).$$
(6.146)

Together with (6.126) this implies that

$$\rho(Z) = \sup_{\mu \in \mathfrak{M}} \left\{ \int_0^1 \mathsf{AV} @\mathsf{R}_{1-\alpha}(Z) d\mu(\alpha) - \rho^*(\mathbb{T}\mu) \right\}, \tag{6.147}$$

where $\mathfrak{M} := \mathbb{T}^{-1}(\mathfrak{F}_q)$. Similar to (6.146) we have for $Y \in \mathfrak{F}_q$ that

$$\int_0^1 Y(t)(\mathbb{T}\mu)(t)dt - \rho(Y) = \int_0^1 \mathsf{AV} @\mathsf{R}_{1-\alpha}(Y)d\mu(\alpha) - \rho(Y),$$

and hence by (6.127)

$$\rho^*(\mathbb{T}\mu) = \sup_{Y \in \mathfrak{F}_q} \left\{ \int_0^1 \mathsf{AV}@\mathsf{R}_{1-\alpha}(Y) d\mu(\alpha) - \rho(Y) \right\}.$$
(6.148)

Combining (6.147) with (6.148) completes the proof.

6.3.6 Probability Spaces with Atoms

So far we considered cases when the reference probability space is nonatomic. Of course this rules out, for example, discrete probability spaces. So what can be said about law invariant risk measures defined on probability spaces with atoms? Let us consider the following construction. Suppose that the reference probability space can be embedded into the standard uniform probability space. That is, let (Ω, \mathcal{F}, P) be the standard uniform probability space and \mathcal{G} be a sigma subalgebra of \mathcal{F} such that the considered reference probability space is equivalent (isomorphic) to (Ω, \mathcal{G}, P) . So we can view (Ω, \mathcal{G}, P) as the reference probability space. For example, let the reference probability space be discrete with a countable (finite) number of elementary events and respective probabilities p_1, p_2, \ldots . Let us partition $\Omega = [0, 1]$ into intervals A_1, A_2, \ldots , of respective lengths p_1, p_2, \ldots , and consider sigma algebra $\mathcal{G} \subset \mathcal{F}$ generated by these intervals. This will give the required embedding of the discrete probability space into the standard uniform probability space.

Consider space $\hat{\mathcal{Z}} := \mathcal{L}_p(\Omega, \mathcal{G}, P), p \in [1, \infty)$, and a proper lower semicontinuous law invariant risk measure $\varrho : \hat{\mathcal{Z}} \to \overline{\mathbb{R}}$. Note that ϱ is supposed to be law invariant with respect to the *reference* probability space. That is, if $Z_1, Z_2 \in \hat{\mathcal{Z}}$ are two \mathcal{G} -measurable distributionally equivalent random variables, then $\varrho(Z_1) = \varrho(Z_2)$. Note also that the space $\hat{\mathcal{Z}}$ is a subspace of the space $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$. Therefore a relevant question is whether it is possible to extend ϱ to a law invariant risk measure on the space \mathcal{Z} . For a risk measure $\rho : \mathcal{Z} \to \mathbb{R}$ we denote by $\rho|_{\hat{\mathcal{Z}}}$ the restriction of ρ to the space $\hat{\mathcal{Z}}$, i.e., $\hat{\rho} = \rho|_{\hat{\mathcal{Z}}}$ is defined on the space $\hat{\mathcal{Z}}$ and $\hat{\rho}(Z) = \rho(Z)$ for $Z \in \hat{\mathcal{Z}}$.

Definition 6.45. We say that a proper lower semicontinuous law invariant coherent (convex) risk measure $\varrho : \hat{Z} \to \overline{\mathbb{R}}$ is regular if there exists a proper lower semicontinuous law invariant coherent (convex) risk measure $\rho : Z \to \overline{\mathbb{R}}$ such that $\rho|_{\hat{Z}} = \varrho$.

For example, the Average Value-at-Risk and mean-semideviation measures are regular risk measures irrespective of the reference probability space. Recall that \mathfrak{F} denotes the set of spectral functions $\sigma : [0,1) \to \mathbb{R}$ and $\mathfrak{F}_q \subset \mathfrak{F}$ denotes the set of spectral functions with finite *q*-th order moment. As before we denote by \mathfrak{G} the set of measure-preserving transformations of the *reference* space (Ω, \mathcal{G}, P) . Note that \mathfrak{G} is a subgroup of the group of measure-preserving transformations of the standard uniform probability space (Ω, \mathcal{F}, P) .

Remark 36. Let $\varrho : \hat{Z} \to \overline{\mathbb{R}}$ be a proper lower semicontinuous law invariant coherent risk measure. It follows from the above definition and derivations of section 6.3.4 that ϱ is

regular iff there exists a set $\Upsilon \subset \mathfrak{F}_q$ such that

$$\varrho(Z) = \sup_{\sigma \in \Upsilon} \int_0^1 \sigma(t) H_Z^{-1}(t) dt, \quad Z \in \hat{\mathcal{Z}}.$$
(6.149)

It also follows by Theorem 6.40 that ρ is regular iff it has the Kusuoka representation (6.134). Therefore existence of Kusuoka representation for the risk measure ρ can be reduced to verification of regularity of ρ . Note that \mathcal{G} -measurability of $Z \in \hat{\mathcal{Z}}$ does not necessarily imply \mathcal{G} -measurability of H_Z^{-1} .

Proposition 6.46. Suppose that the following condition holds:

(♣) For any *G*-measurable random variable *Z* : [0,1] → ℝ there exists measure preserving transformation *T* ∈ 𝔅, of the reference space (Ω, *G*, *P*), such that *Z* ∘ *T* is monotonically nondecreasing.

Then every proper lower semicontinuous law invariant coherent risk measure $\varrho : \hat{Z} \to \overline{\mathbb{R}}$ is regular.

Proof. Let $\varrho : \hat{Z} \to \overline{\mathbb{R}}$ be a proper lower semicontinuous law invariant coherent risk measure. It has the dual representation

$$\varrho(Z) = \sup_{\zeta \in \hat{\mathfrak{A}}} \int_0^1 \zeta(t) Z(t) dt, \ Z \in \hat{\mathcal{Z}},$$
(6.150)

where $\hat{\mathfrak{A}} \subset \hat{\mathcal{Z}}^*$ is the corresponding dual set of density functions. Since ρ is law invariant, the set $\hat{\mathfrak{A}}$ is invariant with respect to measure-preserving transformations $T \in \mathfrak{G}$ (see Corollary 6.30).

Suppose that condition (\clubsuit) holds. Consider an element $Y \in \hat{Z}$. By condition (\clubsuit) there exists $T \in \mathfrak{G}$ such that $Z = Y \circ T$ is monotonically nondecreasing, and we can take Z to be left side continuous. It follows that $Z \in \hat{Z}$ and $Z \stackrel{\mathcal{D}}{\sim} Y$, and hence $\varrho(Z) = \varrho(Y)$, and that $Z = H_Y^{-1}$. So let $Z \in \hat{Z}$ be monotonically nondecreasing and consider an element $\zeta \in \hat{\mathfrak{A}}$. By condition (\clubsuit) there exists $T \in \mathfrak{G}$ such that $\eta = \zeta \circ T$ is a monotonically nondecreasing function. Since $\hat{\mathfrak{A}}$ is invariant with respect to transformations of the group \mathfrak{G} , it follows that $\eta \in \hat{\mathfrak{A}}$. Also by monotonicity of Z we have that $\int_0^1 \zeta(t)Z(t)dt \leq \int_0^1 \eta(t)Z(t)dt$, and hence it suffices to take the maximum in (6.150) with respect to $\zeta \in \hat{\Upsilon}$, where $\hat{\Upsilon}$ is the subset of $\hat{\mathfrak{A}}$ formed by monotonically nondecreasing $\eta \in \hat{\mathfrak{A}}$. Define now

$$\rho(Z) = \sup_{\eta \in \hat{\Upsilon}} \int_0^1 \eta(t) H_Z^{-1}(t) dt, \ Z \in \mathcal{Z}.$$
(6.151)

We have that $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is a proper lower semicontinuous law invariant coherent risk measure and $\rho|_{\hat{\mathcal{Z}}} = \varrho$. This shows that ϱ is regular. \Box

Suppose now that the reference probability space $\{\omega_1, \ldots, \omega_n\}$ is finite. As it was discussed above it can be embedded into the standard uniform probability space. The

corresponding space \hat{Z} consists of all functions $Z : \{\omega_1, \ldots, \omega_n\} \to \mathbb{R}$. If all respective probabilities p_1, \ldots, p_n are equal to each other, then the associated group \mathfrak{G} of measure preserving transformations is given by the set of permutations of $\{\omega_1, \ldots, \omega_n\}$. It follows that condition (\clubsuit) of Proposition 6.46 holds and hence we have the following result.

Corollary 6.47. Let the reference probability space $\{\omega_1, \ldots, \omega_n\}$ be finite, equipped with equal probabilities $p_i = 1/n, i = 1, \ldots, n$. Then every law invariant coherent risk measure $\varrho: \hat{Z} \to \mathbb{R}$ is regular, and hence has a Kusuoka representation.

We give now an example of law invariant coherent risk measures which are not regular.

Example 6.48 Consider the above framework where the reference probability space is embedded into the standard uniform probability space, and $\hat{\mathcal{Z}} = \mathcal{L}_p(\Omega, \mathcal{G}, P)$. Suppose that the reference probability space (Ω, \mathcal{G}, P) has atoms. Thus there exists r > 0 such that the set $A^r := \{\omega \in \Omega : P(\{\omega\}) = r\}$, of all atoms having probability r, is nonempty. Clearly the set A^r is finite, let $N := |A^r|$ be the cardinality of A^r (the set A^r could be a singleton, i.e., it could be that N = 1). The set $A^r = \bigcup_{i=1}^N A_i$, where $A_i \subset \Omega$, $i = 1, \ldots, N$, are intervals of length r. Note that since $Z \in \hat{Z}$ is \mathcal{G} -measurable, $Z(\cdot)$ is constant on each A_i ; we denote by $Z(\omega)$ this constant when $\omega \in A_i$.

Assume that the reference probability space is not a finite set equipped with equal probabilities, so that $P(\Omega \setminus A^r) > 0$. Define

$$\varrho(Z) := \frac{1}{N} \sum_{\omega \in A^r} Z(\omega), \ Z \in \hat{\mathcal{Z}}.$$
(6.152)

Clearly $\rho : \hat{\mathcal{Z}} \to \mathbb{R}$ is a coherent risk measure. Suppose further that the following condition holds.

(\bigstar) If $Z, Z' \in \hat{Z}$ are distributionally equivalent, then there exists a permutation π of the set A^r such that $Z'(\omega) = Z(\pi(\omega))$ for any $\omega \in A^r$.

Under this condition, the risk measure ρ is law invariant as well. Let us show that it is not regular.

Indeed, let us argue by a contradiction. Suppose that $\rho = \hat{\rho}$ for some proper lower semicontinuous law invariant coherent risk measure $\rho : \mathbb{Z} \to \mathbb{R}$ and $\hat{\rho} := \rho|_{\hat{\mathbb{Z}}}$. Note that $\rho(Z)$ depends only on values of $Z(\cdot)$ on the set A^r , i.e., for any $Z, Z' \in \hat{\mathbb{Z}}$ we have that $\rho(Z) = \rho(Z')$ if $Z(\omega) = Z'(\omega)$ for all $\omega \in A^r$. In particular, if $Z(\omega) = 0$ for all $\omega \in A^r$, then $\rho(Z) = 0$. It follows that $\rho(\mathbf{1}_B) = 0$ for any \mathcal{G} -measurable set $B \subset \Omega \setminus A^r$. Note that the set $\Omega \setminus A^r$ has a nonempty interior since the set A^r is a union of a finite number of intervals and $P(\Omega \setminus A^r) > 0$, and hence we can take $B \subset \Omega \setminus A^r$ to be an open interval. Since ρ is law invariant it follows that $\rho(\mathbf{1}_{T(B)}) = 0$ for any measurepreserving transformation $T : \Omega \to \Omega$ of the standard uniform probability space. Hence there is a family of sets $B_i \in \mathcal{F}$, $i = 1, \ldots, m$, such that $\rho(\mathbf{1}_{B_i}) = 0$, $i = 1, \ldots, m$, and $[0,1] = \bigcup_{i=1}^m B_i$. It follows that for any bounded $Z \in \mathcal{Z}$ there are $c_i \in \mathbb{R}$, $i = 1, \ldots, m$, such that $Z \leq \sum_{i=1}^m c_i \mathbf{1}_{B_i}$. Consequently

$$\rho(Z) \le \rho\left(\sum_{i=1}^{m} c_i \mathbf{1}_{B_i}\right) \le \sum_{i=1}^{m} c_i \rho(\mathbf{1}_{B_i}) = 0,$$

this clearly is a contradiction.

It follows that the risk measure ϱ , defined in (6.152), does not have a Kusuoka representation. Note that it was only essential in the above construction that the restriction of the risk measure $\varrho : \hat{Z} \to \mathbb{R}$ to the set A^r is a law invariant coherent risk measure. So, for example, under the above assumptions the risk measure $\varrho(Z) := \max\{Z(\omega) : \omega \in A^r\}$ is also not regular.

As far as condition (\blacklozenge) is concerned, suppose for example that the reference space is finite, equipped with respective probabilities p_1, \ldots, p_n . For some $r \in \{p_1, \ldots, p_n\}$ let $\mathcal{I}_r := \{i : p_i = r, i = 1, \ldots, n\}$. Suppose that

$$\sum_{i \in I} p_i \neq \sum_{j \in J} p_j, \quad \forall I \subset \mathcal{I}_r, \ \forall J \subset \{1, \dots, n\} \setminus \mathcal{I}_r.$$
(6.153)

Then condition (\blacklozenge) holds for the set $A^r := \{\omega_i : i \in \mathcal{I}_r\}$. That is, condition (6.153) ensures existence of a nonregular risk measure ϱ .

6.3.7 Stochastic Orders

For law invariant risk measures it makes sense to discuss their monotonicity properties with respect to various stochastic orders defined for (real valued) random variables. Many stochastic orders can be characterized by a class \mathcal{U} of functions $u : \mathbb{R} \to \mathbb{R}$ as follows. For (real valued) random variables Z_1 and Z_2 it is said that Z_2 dominates Z_1 , denoted $Z_2 \succeq_{\mathcal{U}} Z_1$, if $\mathbb{E}[u(Z_2)] \ge \mathbb{E}[u(Z_1)]$ for all $u \in \mathcal{U}$ for which the corresponding expectations do exist. This stochastic order is called the *integral stochastic order* with *generator* \mathcal{U} . In particular, the *usual stochastic order*, written $Z_2 \succeq_{(1)} Z_1$, corresponds to the generator \mathcal{U} formed by all nondecreasing functions $u : \mathbb{R} \to \mathbb{R}$. Equivalently, $Z_2 \succeq_{(1)} Z_1$ iff $H_{Z_2}(t) \le H_{Z_1}(t)$ for all $t \in \mathbb{R}$. The relation $\succeq_{(1)}$ is also frequently called the *first order stochastic dominance* (see Definition 4.3). We say that the integral stochastic order is an example of increasing integral stochastic order.

Definition 6.49. A law invariant risk measure $\rho : \mathbb{Z} \to \mathbb{R}$ is consistent (monotone) with the integral stochastic order \succeq_{u} if for all $Z_1, Z_2 \in \mathbb{Z}$ we have the implication

$$\{Z_2 \succeq_{\mathcal{U}} Z_1\} \Rightarrow \{\rho(Z_2) \ge \rho(Z_1)\}.$$

For an increasing integral stochastic order we have that if $Z_2(\omega) \ge Z_1(\omega)$ for a.e. $\omega \in \Omega$, then $u(Z_2(\omega)) \ge u(Z_1(\omega))$ for any $u \in \mathcal{U}$ and a.e. $\omega \in \Omega$, and hence $\mathbb{E}[u(Z_2)] \ge \mathbb{E}[u(Z_1)]$. That is, if $Z_2 \succeq Z_1$ in the almost sure sense, then $Z_2 \succeq_{\mathcal{U}} Z_1$. It follows that if ρ is law invariant and consistent with respect to an increasing integral stochastic order, then it satisfies the monotonicity condition (R2). In other words if ρ does not satisfy condition (R2), then it cannot be consistent with any increasing integral stochastic order. In particular, for c > 1 the mean-semideviation risk measure, defined in Example 6.23, is not consistent with any increasing integral stochastic ofdes.

A general way of proving consistency of law invariant risk measures with stochastic orders can be obtained via the following construction. For a given pair of random variables Z_1 and Z_2 in Z, consider another pair of random variables, \hat{Z}_1 and \hat{Z}_2 , which have distributions identical to the original pair, i.e., $\hat{Z}_1 \stackrel{\mathcal{D}}{\sim} Z_1$ and $\hat{Z}_2 \stackrel{\mathcal{D}}{\sim} Z_2$. The construction is such that the postulated consistency result becomes evident. For this method to be applicable, it is convenient to assume that the probability space (Ω, \mathcal{F}, P) is nonatomic. Then there exists a measurable function $U : \Omega \to \mathbb{R}$ (uniform random variable) such that $P(U \le t) = t$ for all $t \in [0, 1]$.

Theorem 6.50. Suppose that the probability space (Ω, \mathcal{F}, P) is nonatomic. Then the following holds: if a risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is law invariant, then it is consistent with the usual stochastic order iff it satisfies the monotonicity condition (R2).

Proof. By the discussion preceding the theorem, it is sufficient to prove that (R2) implies consistency with the usual stochastic order.

For a uniform random variable $U(\omega)$ consider the random variables $\hat{Z}_1 := H_{Z_1}^{-1}(U)$ and $\hat{Z}_2 := H_{Z_2}^{-1}(U)$. We obtain that if $Z_2 \succeq_{(1)} Z_1$, then $\hat{Z}_2(\omega) \ge \hat{Z}_1(\omega)$ for all $\omega \in \Omega$, and hence by virtue of (R2), $\rho(\hat{Z}_2) \ge \rho(\hat{Z}_1)$. By construction, $\hat{Z}_1 \stackrel{\mathcal{D}}{\sim} Z_1$ and $\hat{Z}_2 \stackrel{\mathcal{D}}{\sim} Z_2$. Since the risk measure is law invariant, we conclude that $\rho(Z_2) \ge \rho(Z_1)$. Consequently, the risk measure ρ is consistent with the usual stochastic order. \Box

It is said that Z_1 is smaller than Z_2 in the *increasing convex order*, written $Z_1 \preceq_{icx} Z_2$, if $\mathbb{E}[u(Z_1)] \leq \mathbb{E}[u(Z_2)]$ for all increasing convex functions $u : \mathbb{R} \to \mathbb{R}$ such that the expectations exist. Clearly this is an integral stochastic order with the corresponding generator given by the set of increasing convex functions. It is equivalent to the *second order stochastic dominance* relation for the negative variables: $-Z_1 \succeq_{(2)} -Z_2$ (recall that we are dealing here with minimization rather than maximization problems). Indeed, applying Definition 4.4 to $-Z_1$ and $-Z_2$ for k = 2 and using identity (4.7) we see that

$$\mathbb{E}\left\{[Z_1 - \eta]_+\right\} \le \mathbb{E}\left\{[Z_2 - \eta]_+\right\}, \quad \forall \ \eta \in \mathbb{R}.$$
(6.154)

Since any convex nondecreasing function u(z) can be arbitrarily close approximated by a positive combination of functions $u_k(z) = \beta_k + [z - \eta_k]_+$, inequality (6.154) implies that $\mathbb{E}[u(Z_1)] \leq \mathbb{E}[u(Z_2)]$, as claimed (compare with the statement (4.8)). Note also that for standard uniform probability space (Ω, \mathcal{F}, P) , condition (6.154) is equivalent to condition (6.120).

Theorem 6.51. Suppose that the probability space (Ω, \mathcal{F}, P) is nonatomic. Then any proper lower semicontinuous law invariant convex risk measure $\rho : \mathbb{Z} \to \overline{\mathbb{R}}$ is consistent with the increasing convex order.

Proof. By using definition (6.23) of AV@R_{α} and the property that $Z_1 \leq_{icx} Z_2$ iff condition (6.154) holds, it is straightforward to verify that AV@R_{α} is consistent with the increasing convex order. Now by using the Kusuoka representation (6.144), for convex risk measures (Theorem 6.44), and noting that the operations of taking convex combinations and maximum preserve consistency with the increasing convex order, we can complete the proof.

Corollary 6.52. Suppose that the probability space (Ω, \mathcal{F}, P) is nonatomic. Let $\rho : \mathbb{Z} \to \mathbb{R}$ be a proper lower semicontinuous law invariant convex risk measure and \mathcal{G} be a sigma subalgebra of the sigma algebra \mathcal{F} . Then

$$\rho\left(\mathbb{E}\left[Z|\mathcal{G}\right]\right) \le \rho(Z), \quad \forall Z \in \mathcal{Z},\tag{6.155}$$

and

$$\mathbb{E}\left[Z\right] \le \rho(Z), \quad \forall Z \in \mathcal{Z}. \tag{6.156}$$

Proof. Consider $Z \in \mathcal{Z}$ and $Z' := \mathbb{E}[Z|\mathcal{G}]$. For every convex function $u : \mathbb{R} \to \mathbb{R}$ we have

$$\mathbb{E}[u(Z')] = \mathbb{E}\left[u\left(\mathbb{E}[Z|\mathcal{G}]\right)\right] \le \mathbb{E}[\mathbb{E}(u(Z)|\mathcal{G})] = \mathbb{E}\left[u(Z)\right],$$

where the inequality is implied by Jensen's inequality. This shows that $Z' \leq_{icx} Z$, and hence (6.155) follows by Theorem 6.51.

In particular, for $\mathcal{G} := \{\Omega, \emptyset\}$, it follows by (6.155) that $\rho(Z) \ge \rho(\mathbb{E}[Z])$, and since $\rho(\mathbb{E}[Z]) = \mathbb{E}[Z]$ this completes the proof. \Box

An intuitive interpretation of property (6.155) is that if we reduce variability of a random variable Z by employing conditional averaging $Z' = \mathbb{E}[Z|\mathcal{G}]$, then the risk measure $\rho(Z')$ becomes smaller, while $\mathbb{E}[Z'] = \mathbb{E}[Z]$.

Remark 37. It was only essential in the proof of Theorem 6.51, that the risk measure ρ has the Kusuoka representation. Therefore the assumption for the space (Ω, \mathcal{F}, P) to be nonatomic in Theorem 6.51 can be replaced by the assumption that the risk measure ρ is regular (see the discussion of section 6.3.6, and specifically Remark 36). The same remark applies to Corollary 6.52. In particular, for the mean-semideviation measures, defined in Example 6.23, and for the Average Value-at-Risk measures, consistency with the increasing convex order follows and the inequality (6.155) holds irrespective of the associated reference probability space.

In general, the inequalities of Corollary 6.52, and hence the implication of Theorem 6.51, do not hold for arbitrary reference probability spaces. For example, let $\Omega = \{\omega_1, \omega_2\}$ with respective probabilities $p_1 = 1/3$ and $p_2 = 2/3$, and consider $\varrho(Z) := Z(\omega_1)$. As it was discussed in Example 6.48, ϱ is a law invariant coherent risk measure, although it is not regular. For random variable $Z : \Omega \to \mathbb{R}$ defined as $Z(\omega_1) = 0$, $Z(\omega_2) = 1$, we have that $\mathbb{E}[Z] = 2/3$, while $\varrho(Z) = 0$. That is, the inequality (6.156) does not hold.

6.4 Ambiguous Chance Constraints

Owing to the dual representation (6.38), measures of risk are related to robust and ambiguous models. Consider a chance constraint of the form

$$P\{C(x,\omega) \le 0\} \ge 1 - \alpha.$$
(6.157)

Here P is a probability measure on a measurable space (Ω, \mathcal{F}) and $C : \mathbb{R}^n \times \Omega \to \mathbb{R}$ is a random function. It is assumed in this formulation of chance constraint that the probability

measure (distribution), with respect to which the corresponding probabilities are calculated, is known.

Suppose now that the underlying probability distribution is not known exactly, but rather is assumed to belong to a specified family of probability distributions. Problems involving such constrains are called *ambiguous chance constrained* problems. For a specified uncertainty set \mathfrak{M} of probability measures on (Ω, \mathcal{F}) , the corresponding ambiguous chance constraint defines a feasible set $\mathcal{X} \subset \mathbb{R}^n$ which can be written as

$$\mathcal{X} := \left\{ x : Q\{C(x,\omega) \le 0\} \ge 1 - \alpha, \ \forall Q \in \mathfrak{M} \right\}.$$
(6.158)

The set \mathcal{X} can be written in the following equivalent form

$$\mathcal{X} = \left\{ x \in \mathbb{R}^n : \sup_{Q \in \mathfrak{M}} \mathbb{E}_Q \left[\mathbf{1}_{A_x} \right] \le \alpha \right\},$$
(6.159)

where

$$A_x := \{ \omega \in \Omega : C(x, \omega) > 0 \}.$$
(6.160)

We consider the uncertainty sets \mathfrak{M} of the following form. Let (Ω, \mathcal{F}, P) be a *nonatomic* probability space (reference space) and \mathfrak{A} be a nonempty set of probability density functions $h : \Omega \to \mathbb{R}_+$. Assume that \mathfrak{A} is *invariant* with respect to measure-preserving transformations (see Definition 6.27). That is, if $h \in \mathfrak{A}$ and $T : \Omega \to \Omega$ is a measure-preserving transformation, then $h \circ T \in \mathfrak{A}$. Define \mathfrak{M} to be a set of probability measures, absolutely continuous with respect to P, of the form

$$\mathfrak{M} := \{Q : dQ/dP \in \mathfrak{A}\}.$$
(6.161)

That is, the set \mathfrak{M} is formed by probability measures absolutely continuous with respect to P and \mathfrak{M} is invariant with respect to measure-preserving transformations. (Note that if Q is an absolutely continuous with respect to P probability measure, with h := dQ/dP, and $Q'(A) := Q(T(A)), A \in \mathcal{F}$, for some measure-preserving transformation $T : \Omega \to \Omega$, then Q' is also absolutely continuous with respect to P probability measure and $dQ'/dP = h \circ T$.)

For an \mathcal{F} -measurable set $A \subset \Omega$ consider

$$g(A) := \sup_{Q \in \mathfrak{M}} \mathbb{E}_Q \left[\mathbf{1}_A \right] = \sup_{h \in \mathfrak{A}} \int_A h(\omega) dP(\omega).$$
(6.162)

We can view $g(\cdot)$ as a function $g: \mathcal{F} \to \mathbb{R}$. Because of the invariance of \mathfrak{A} with respect to measure-preserving transformations, we have that g(A) = g(B) if P(A) = P(B). (Recall that P(A) = P(B), for some $A, B \in \mathcal{F}$, iff there exists a measure-preserving transformation T such that B = T(A), see Remark 28 on page 320.) That is, g(A) is a function of P(A) and can be written as $g(A) = \mathfrak{p}(P(A))$ for some function $\mathfrak{p}: [0,1] \to \mathbb{R}$. Since the reference space is nonatomic, it follows that the set $\{P(A) : A \in \mathcal{F}\}$ coincides with the interval [0, 1], and hence the function \mathfrak{p} is uniquely defined on [0,1].

Proposition 6.53. The function $\mathfrak{p} : [0,1] \to \mathbb{R}$ has the following properties: (i) $\mathfrak{p}(0) = 0$ and $\mathfrak{p}(1) = 1$, (ii) $\mathfrak{p}(\cdot)$ is monotonically nondecreasing on the interval [0,1], (iii) $\mathfrak{p}(\cdot)$ is monotonically increasing on the interval $[0,\bar{\tau}]$, where

$$\bar{\tau} := \inf\{t \in [0,1] : \mathfrak{p}(t) = 1\},$$
(6.163)

(iv) if $\mathfrak{M} = \{P\}$, then $\mathfrak{p}(t) = t$ for all $t \in [0, 1]$; and if $\mathfrak{M} \neq \{P\}$, then $\mathfrak{p}(t) > t$ for all $t \in (0, 1)$, (v) $\mathfrak{p}(\cdot)$ is continuous on the interval (0, 1].

Proof. Clearly if $A, B \in \mathcal{F}$ and $A \subset B$ and hence $P(A) \leq P(B)$, then $g(A) \leq g(B)$. It follows that function \mathfrak{p} is monotonically nondecreasing. If P(A) = 0, then by the right hand side of (6.162) we have that g(A) = 0. It follows that $\mathfrak{p}(0) = 0$. Similarly $g(\Omega) = 1$ and hence $\mathfrak{p}(1) = 1$. This proves properties (i) and (ii).

We can assume without loss of generality that the reference probability space is standard uniform. We have that for any $h \in \mathfrak{A}$ there exists a measure-preserving transformation T such that $h \circ T$ is monotonically nondecreasing (see Remark 30 on page 321). Since \mathfrak{A} is invariant with respect to measure-preserving transformations, $h \circ T \in \mathfrak{A}$. Denote by \mathfrak{A}^* the subset of \mathfrak{A} formed by monotonically nondecreasing (density) functions $h \in \mathfrak{A}$. It follows that

$$\mathfrak{p}(t) = \sup_{h \in \mathfrak{A}^*} \psi_h(t), \ t \in [0, 1], \tag{6.164}$$

where $\psi_h(t) := \int_{1-t}^1 h(\omega) d\omega$. It is straightforward to verify that for every $h \in \mathfrak{A}^*$ the function $\psi_h(\cdot)$ is concave continuous monotonically increasing on [0,1] with $\psi_h(0) = 0$ and $\psi_h(1) = 1$.

To prove (iii) we argue by a contradiction. Suppose that $\mathfrak{p}(a) = \mathfrak{p}(b)$ for some $0 \le a < b \le \overline{\tau}$. For $\varepsilon > 0$ let $h \in \mathfrak{A}^*$ be such that $\mathfrak{p}(a) \le \psi_h(a) + \varepsilon$. By concavity of $\psi_h(\cdot)$ we have that

$$\psi_h(1) \le \psi_h(a) + \frac{\psi_h(b) - \psi_h(a)}{b - a}(1 - a)$$

We also have that $\mathfrak{p}(a) - \varepsilon \leq \psi_h(a) \leq \psi_h(b) \leq \mathfrak{p}(b) = \mathfrak{p}(a)$. It follows that $1 \leq \mathfrak{p}(a) + \varepsilon(1-a)/(b-a)$, and hence since $\varepsilon > 0$ is arbitrary that $1 \leq \mathfrak{p}(a)$. On the other hand $\mathfrak{p}(a) < 1$ since $a < \overline{\tau}$. We obtain a contradiction. This proves (iii).

If $\mathfrak{M} = \{P\}$, then $\mathfrak{A} = \{\mathbf{1}_{\Omega}\}$ and hence $\mathfrak{p}(t) = t$ for all $t \in [0, 1]$. If $\mathfrak{M} \neq \{P\}$, then the set \mathfrak{A}^* contains a nonconstant function h^* , and hence

$$\mathbf{\mathfrak{p}}(t) \ge \int_{1-t}^{1} h^*(\omega) d\omega > t, \ t \in (0,1).$$

This proves (iv).

Since $\mathfrak{p}(\cdot)$ is a maximum of a family of continuous functions, it is lower semicontinuous (see Proposition 7.24). In the present case this means that $\lim_{t\uparrow t_0} \mathfrak{p}(t) = \mathfrak{p}(t_0)$ for any $t_0 \in (0, 1]$, i.e., $\mathfrak{p}(\cdot)$ is continuous from the left. Let us show that it is also continuous from the right on the interval (0, 1]. Consider a point $t_0 \in (0, 1)$ and $t_0 < t < 1$. For $h \in \mathfrak{A}^*$ we have

$$0 \le \psi_h(t) - \psi_h(t_0) = \int_{1-t}^{1-t_0} h(\omega) d\omega.$$
 (6.165)

We also have that

$$1 \geq \int_{1-t}^{1} h(\omega)d\omega = \int_{1-t}^{1-t_0} h(\omega)d\omega + \int_{1-t_0}^{1} h(\omega)d\omega$$

$$\geq \int_{1-t}^{1-t_0} h(\omega)d\omega + \frac{t_0}{t-t_0} \int_{1-t}^{1-t_0} h(\omega)d\omega$$

$$= \frac{t}{t-t_0} \int_{1-t}^{1-t_0} h(\omega)d\omega,$$

where the last inequality follows because $h(\cdot)$ is monotonically nondecreasing. Hence the right hand side of (6.165) can be bounded by $(t - t_0)/t$ uniformly in $h \in \mathfrak{A}^*$. It follows that

$$0 \le \mathbf{p}(t) - \mathbf{p}(t_0) \le (t - t_0)/t.$$
 (6.166)

Since $t_0 > 0$, we have that $(t - t_0)/t$ converges to 0 as $t \downarrow t_0$. Together with (6.166) this implies that $\lim_{t\downarrow t_0} \mathfrak{p}(t) = \mathfrak{p}(t_0)$. This proves that $\mathfrak{p}(t)$ is continuous on (0, 1). Continuity at t = 1 follows from the inequalities $t \leq \mathfrak{p}(t) \leq 1$. This completes the proof of (v). \Box

As we will see below function $\mathfrak{p}(t)$ can be discontinuous at t = 0. That is,

$$\alpha_0 := \lim_{t \downarrow 0} \mathfrak{p}(t) \tag{6.167}$$

can be strictly bigger than $\mathfrak{p}(0) = 0$. The set \mathcal{X} of the form (6.159) can be written as

$$\mathcal{X} = \left\{ x \in \mathbb{R}^n : \mathfrak{p}(P(A_x)) \le \alpha \right\},\tag{6.168}$$

and hence as

$$\mathcal{X} = \left\{ x \in \mathbb{R}^n : P(A_x) \le \alpha^* \right\},\tag{6.169}$$

where $\alpha^* := \mathfrak{p}^{-1}(\alpha)$. That is,

$$\mathcal{X} = \{ x : P\{C(x,\omega) \le 0\} \ge 1 - \alpha^* \},$$
(6.170)

i.e., the set \mathcal{X} can be defined by a chance constraint with respect to the reference distribution P and with the respective probability level $1 - \alpha^*$.

By the above Proposition 6.53, for $\alpha \in (\alpha_0, 1)$ the corresponding value $\alpha^* = \mathfrak{p}^{-1}(\alpha)$ is given as the unique solution of the equation $\mathfrak{p}(\alpha^*) = \alpha$; and $\mathfrak{p}^{-1}(\alpha) = 0$ for $\alpha \in [0, \alpha_0]$. The function $\mathfrak{p}^{-1}(\cdot)$ is continuous on [0, 1). Also unless \mathfrak{M} is the singleton $\{P\}$, it holds that $\mathfrak{p}^{-1}(\alpha) < \alpha$ for $\alpha \in (0, 1)$.

Remark 38. Consider Value-at-Risk

$$\mathsf{V}@\mathsf{R}^{Q}_{\alpha}(Z) = \inf \left\{ t : Q\{Z > t\} \le \alpha \right\},\$$

associated with probability measure $Q \in \mathfrak{M}$. We have that for any $c \in \mathbb{R}$, $V@R^Q_\alpha(Z) \leq c$ is equivalent to $Q\{Z > c\} \leq \alpha$. It follows that

$$\sup_{Q \in \mathfrak{M}} \mathsf{V}_{\alpha}^{Q}(Z) \le c \quad \Leftrightarrow \quad \sup_{Q \in \mathfrak{M}} Q\{Z > c\} \le \alpha.$$
(6.171)

By the above analysis we also have that the right hand side of (6.171) is equivalent to $P\{Z > c\} \le \alpha^*$, and hence to $V@R^P_{\alpha^*}(Z) \le c$, where $\alpha^* = \mathfrak{p}^{-1}(\alpha)$. It follows that

$$\sup_{Q \in \mathfrak{M}} \mathsf{V}_{\alpha}\mathsf{R}_{\alpha}^{Q}(Z) = \mathsf{V}_{\alpha}\mathsf{R}_{\alpha^{*}}^{P}(Z).$$
(6.172)

In particular, for α_0 defined in (6.167) and $\alpha \in [0, \alpha_0]$ we have that

$$\sup_{Q \in \mathfrak{M}} \mathsf{V}_{\alpha}^{Q}(Z) = \operatorname{ess\,sup}(Z). \tag{6.173}$$

Consider now coherent risk measure

$$\rho(Z) := \sup_{\substack{\frac{dQ}{dP} \in \mathfrak{A}}} \mathbb{E}_Q[Z]$$
(6.174)

defined on the space $\mathcal{Z} = \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$. Since the set \mathfrak{A} is invariant with respect to measure-preserving transformations, this risk measure is law invariant. Using this risk measure we can write the set \mathcal{X} in the form

$$\mathcal{X} = \left\{ x \in \mathbb{R}^n : \rho(\mathbf{1}_{A_x}) \le \alpha \right\}.$$
(6.175)

In some cases the corresponding function $\mathfrak{p} = \mathfrak{p}_{\rho}$ can be computed in a closed form.

Consider the Average Value-at-Risk measure $\rho(\cdot) := AV@R_{\gamma}(\cdot), \gamma \in (0,1]$. By direct calculations it is straightforward to verify that for any $A \in \mathcal{F}$,

$$\mathsf{AV}@\mathsf{R}_{\gamma}(\mathbf{1}_{A}) = \begin{cases} \gamma^{-1}P(A), & \text{if } P(A) \leq \gamma, \\ 1, & \text{if } P(A) > \gamma. \end{cases}$$

Consequently the corresponding function \mathfrak{p}_{ρ} can be written as

$$\mathbf{\mathfrak{p}}_{\rho}(t) = \begin{cases} \gamma^{-1}t, & \text{if } t \in [0,\gamma], \\ 1, & \text{if } t \in (\gamma,1]. \end{cases}$$
(6.176)

In particular, for $\rho(\cdot) := AV@R_1(\cdot)$, i.e., for $\rho := \mathbb{E}(\cdot)$, we have that $\mathfrak{p}_{\rho}(t) = t$ for $t \in [0, 1]$. For $\rho(\cdot) := AV@R_0(\cdot)$, i.e., for $\rho := \operatorname{ess\,sup}(\cdot)$, we have that $\mathfrak{p}_{\rho}(t) = 1$ for $t \in (0, 1]$, and $\mathfrak{p}_{\rho}(0) = 0$. That is, in that case the function $\mathfrak{p}_{\rho}(\cdot)$ is discontinuous at 0.

Now let $\rho := \sum_{i=1}^{m} \lambda_i \rho_i$ be a convex combination of law invariant coherent risk measures ρ_i , i = 1, ..., m. For $A \in \mathcal{F}$ we have that $\rho(\mathbf{1}_A) = \sum_{i=1}^{m} \lambda_i \rho_i(\mathbf{1}_A)$ and hence $\mathfrak{p}_{\rho} = \sum_{i=1}^{m} \lambda_i \mathfrak{p}_{\rho_i}$. By taking $\rho_i := \mathsf{AV}@R_{\gamma_i}$, with $\gamma_i \in (0, 1]$, i = 1, ..., m, and using (6.176), we obtain that $\mathfrak{p}_{\rho} : [0, 1] \to [0, 1]$ is a piecewise linear nondecreasing concave function with $\mathfrak{p}_{\rho}(0) = 0$ and $\mathfrak{p}_{\rho}(1) = 1$. In particular, let

$$\rho := \beta \mathsf{AV}@\mathsf{R}_{\gamma} + (1 - \beta)\mathsf{AV}@\mathsf{R}_{1}, \tag{6.177}$$

where $\beta, \gamma \in (0, 1)$. Then

$$\mathfrak{p}_{\rho}(t) = \begin{cases} (1-\beta+\gamma^{-1}\beta)t, & \text{if } t \in [0,\gamma], \\ \beta+(1-\beta)t, & \text{if } t \in (\gamma,1]. \end{cases}$$
(6.178)

That is, for an appropriate choice of β , $\gamma \in (0, 1)$, the corresponding function \mathfrak{p}_{ρ} can be any piecewise linear nondecreasing concave function consisting of two linear pieces and such that $\mathfrak{p}_{\rho}(0) = 0$ and $\mathfrak{p}_{\rho}(1) = 1$. More generally, let μ be a probability measure on [0, 1] and $\rho := \int_{0}^{1} \mathsf{AV}@R_{\gamma}d\mu(\gamma)$. In that case the corresponding function $\mathfrak{p}_{\rho} : [0, 1] \to \mathbb{R}$ becomes a nondecreasing concave function with $\mathfrak{p}_{\rho}(0) = 0$ and $\mathfrak{p}_{\rho}(1) = 1$ (it is discontinuous at t = 0 iff μ has a positive mass at t = 0).

Let $\rho(\cdot)$ be the convex combination of AV@R_{γ}(\cdot) and AV@R₁(\cdot) = $\mathbb{E}(\cdot)$ defined in (6.177). Then by (6.178) it follows that for this risk measure and for $\alpha < \beta + (1 - \beta)\gamma$,

$$\alpha^* = \frac{\alpha}{1 + \beta(\gamma^{-1} - 1)}.$$
(6.179)

In particular, for $\beta = 1$, i.e., for $\rho = AV@R_{\gamma}$, we have that $\alpha^* = \gamma \alpha$ for $\alpha \in [0, 1)$.

As another example consider the mean-upper-semideviation risk measure of order $p \in [1, \infty)$. That is

$$\rho(Z) := \mathbb{E}[Z] + c \left(\mathbb{E}\left[\left[Z - \mathbb{E}[Z] \right]_+^p \right] \right)^{1/p}.$$

We have here that $\rho(\mathbf{1}_A) = P(A) + c[P(A)(1 - P(A))^p]^{1/p}$, and hence

$$\mathbf{p}_{\rho}(t) = t + c t^{1/p} (1 - t), \ t \in [0, 1].$$
 (6.180)

In particular, for p = 1 we have that $\mathbf{p}_{a}(t) = (1 + c)t - ct^{2}$, and hence

$$\alpha^* = \frac{1 + c - \sqrt{(1 + c)^2 - 4\alpha c}}{2c}.$$
(6.181)

Note that for c > 1 the above function $\mathfrak{p}_{\rho}(\cdot)$ is not monotonically nondecreasing on the interval [0, 1]. This should be not surprising since for c > 1 and nonatomic P, the corresponding mean-upper-semideviation risk measure is not monotone.

Example 6.54 Let (Ω, \mathcal{F}, P) be a nonatomic probability space. For a probability measure Q on (Ω, \mathcal{F}) , absolutely continuous with respect to P, the Kullback-Leibler divergence from Q to P is defined as

$$D_{KL}(Q||P) := \int_{\Omega} \ln\left(\frac{dQ}{dP}\right) dQ = \int_{\Omega} \ln\left(\frac{dQ}{dP}\right) \frac{dQ}{dP} dP.$$
(6.182)

By Jensen inequality, using convexity of function $-\ln(\cdot)$, we have

$$\int_{\Omega} \ln\left(\frac{dQ}{dP}\right) dQ = -\int_{\Omega} \ln\left(\frac{dP}{dQ}\right) dQ \ge -\ln\left(\int_{\Omega} \frac{dP}{dQ} dQ\right) = -\ln 1 = 0.$$

That is, $D_{KL}(Q||P) \ge 0$. Since $-\ln(\cdot)$ is strictly convex, it follows that $D_{KL}(Q||P) = 0$ iff Q = P.

Consider the uncertainty set \mathfrak{M} of the form

$$\mathfrak{M} := \{Q : D_{KL}(Q \| P) \le c\}$$
(6.183)

for some constant c > 0. This set can be written in the form (6.161) with the corresponding set

$$\mathfrak{A} = \left\{ h \in \mathfrak{P} : \int_{\Omega} h(\omega) \ln(h(\omega)) dP(\omega) \le c \right\},$$
(6.184)

where \mathfrak{P} denotes the set of probability density functions on the space (Ω, \mathcal{F}, P) . It is convenient to write \mathfrak{A} in the following equivalent way

$$\mathfrak{A} = \left\{ h \in \mathfrak{P} : \int_{\Omega} \psi(h(\omega)) dP(\omega) \le c \right\},\tag{6.185}$$

where $\psi(x) := x \ln x - x + 1$, $x \ge 0$. The function $\psi : \mathbb{R}_+ \to \mathbb{R}$ has the following properties: (i) $\psi(x) \ge 0$ for all $x \ge 0$, (ii) $\psi(1) = 0$ and $\psi(x) > 0$ for all x > 1, (iii) $\psi(\cdot)$ is convex continuous. In the following derivations we consider sets \mathfrak{A} of the form (6.185) with respective function $\psi(\cdot)$ satisfying the above conditions (i)-(iii).

For any measure-preserving transformation $T: \Omega \to \Omega$ we have that

$$\int_{\Omega} \psi(h(T(\omega))) dP(\omega) = \int_{\Omega} \psi(h(\omega)) dP(\omega),$$

and hence the set \mathfrak{A} is invariant with respect to measure-preserving transformations. The corresponding function $\mathfrak{p}(\cdot)$ can be written as

$$\mathfrak{p}(t) = \sup_{h \in \mathfrak{A}} \int_{A} h(\omega) dP(\omega), \qquad (6.186)$$

where P(A) = t and the set \mathfrak{A} consists of measurable functions $h : \Omega \to \mathbb{R}_+$ such that $\int_{\Omega} h(\omega) dP(\omega) = 1$ and $\int_{\Omega} \psi(h(\omega)) dP(\omega) \leq c$.

Let us observe that if $h \in \mathfrak{A}$ then for any $B \in \mathcal{F}$, $P(B) \neq 0$, the function

$$\hat{h}(\omega) := \left\{ \begin{array}{ll} h(\omega), & \text{if} \quad \omega \in \Omega \setminus B, \\ \frac{1}{P(B)} \int_B h dP, & \text{if} \quad \omega \in B, \end{array} \right.$$

also belongs to A. Indeed

$$\begin{split} \int_{\Omega} \psi(\hat{h}) dP &= \int_{\Omega \setminus B} \psi(\hat{h}) dP + \int_{B} \psi(\hat{h}) dP \\ &= \int_{\Omega \setminus B} \psi(h) dP + P(B) \psi\left(\frac{1}{P(B)} \int_{B} h dP\right) \\ &\leq \int_{\Omega \setminus B} \psi(h) dP + \int_{B} \psi(h) dP = \int_{\Omega} \psi(h) dP, \end{split}$$

where the inequality holds by Jensen inequality since function $\psi(\cdot)$ is convex.

Also for B = A we have that $\int_A h dP = \int_A \hat{h} dP$. It follows that it suffices to perform optimization (maximization) in (6.186) over feasible functions of the form $h = x_1 \mathbf{1}_{\Omega \setminus A} + x_2 \mathbf{1}_A$, $x_1 \ge 0, x_2 \ge 0$. Hence for $t \in (0, 1]$ the function $\mathfrak{p}(t)$ is equal to the optimal value of the problem

$$\begin{array}{ll}
& \underset{x_1 \ge 0, x_2 \ge 0}{\text{Max}} & x_2 t \\
& \underset{x_1(1-t) + x_2 t = 1, \\
& \psi(x_1)(1-t) + \psi(x_2) t \le c.
\end{array}$$
(6.187)

By making change of variables $y = x_2 t$ we can write problem (6.187) in the following equivalent form

$$\underset{0 \le y \le 1}{\operatorname{Max}} y \text{ s.t. } \theta_t(y) \le c, \tag{6.188}$$

where

$$\theta_t(y) := (1-t)\psi\left(\frac{1-y}{1-t}\right) + t\psi\left(\frac{y}{t}\right).$$
(6.189)

The following properties of function $\theta_t(\cdot)$ follow from the properties (i)–(iii) of function $\psi(\cdot)$. For every $t \in (0, 1)$ the function $\theta_t : [0, 1] \to \mathbb{R}_+$ is nonnegative valued convex continuous, attains its minimal value of 0 at y = t, and is monotonically increasing on the interval [t, 1]. By Proposition 6.53 we have that $\mathfrak{p}(t)$ is continuous at every $t \in (0, 1]$ (this also can be proved here directly by using Theorem 7.23), and monotonically increasing on the interval $[0, \overline{\tau}]$. Also $\mathfrak{p}^{-1}(\alpha) < \alpha$ for $\alpha \in (0, 1)$ and $\mathfrak{p}^{-1}(\cdot)$ is continuous on [0, 1). As we will see below $\mathfrak{p}(t)$ can be discontinuous at t = 0.

For $\theta_t(1) > c$ the equation $\theta_t(y) = c$ has a solution $\overline{y}(t) \in [t, 1]$. Since $\theta_t(\cdot)$ is monotonically increasing on [t, 1] this solution is unique. Moreover, since $\theta_t(1) = 1 - t + t\psi(1/t)$, we have for $t \in (0, 1)$,

$$\mathfrak{p}(t) = \begin{cases} \bar{y}(t) & \text{if } 1 - t + t\psi(1/t) > c, \\ 1 & \text{if } 1 - t + t\psi(1/t) \le c. \end{cases}$$
(6.190)

That is, here $\bar{\tau} = \inf\{t \in [0,1] : 1 - t + t\psi(1/t) \le c\}$. It follows that for $\alpha \in (0,1)$ the corresponding value α^* is the less than α root of the equation

$$(1 - \alpha^*)\psi\left(\frac{1 - \alpha}{1 - \alpha^*}\right) + \alpha^*\psi\left(\frac{\alpha}{\alpha^*}\right) = c,$$
(6.191)

if this root is nonnegative, and $\alpha^* = 0$ otherwise.

In case of $\psi(x) = x \ln x - x + 1$, i.e., in case of the Kullback-Leibler divergence, $\theta_t(y) = (1 - y) \ln \left(\frac{1 - y}{1 - t}\right) + y \ln \left(\frac{y}{t}\right)$, and hence the corresponding value $\alpha^* = \mathfrak{p}^{-1}(\alpha)$ is the less than α root of the equation

$$(1-\alpha)\ln\left(\frac{1-\alpha}{1-\alpha^*}\right) + \alpha\ln\left(\frac{\alpha}{\alpha^*}\right) = c.$$
(6.192)

We have here that $\theta_t(1) = \ln(1/t)$. Thus the condition $\theta_t(1) \le c$ is equivalent to $t \ge e^{-c}$, and hence $\mathfrak{p}(t) = 1$ for $t \in [e^{-c}, 1]$. It follows that $\mathfrak{p}^{-1}(\alpha) < e^{-c}$ for any $\alpha \in (0, 1)$. Here the function $\psi(\cdot)$ satisfies the property $\lim_{x\to+\infty} \psi(x)/x = +\infty$. This property implies that $\lim_{t\downarrow 0} \theta_t(y) = +\infty$ for $y \in (0, 1)$. This in turn implies that $\mathfrak{p}(t)$ tends to zero as $t \downarrow 0$. Thus $\mathfrak{p}(t)$ is continuous at t = 0, and hence is continuous on [0,1]. Consequently $\mathfrak{p}^{-1}(\alpha) > 0$ for any $\alpha \in (0, 1]$.

Consider now $\psi(x) := |x - 1|$. Then for $t \in (0, 1)$ and $y \in [0, 1]$ we have that $\theta_t(y) = 2|y - t|$. Consequently $\mathfrak{p}(t) = \min\{t + c/2, 1\}$ for $t \in (0, 1]$ and $\mathfrak{p}(0) = 0$. It follows that $\mathfrak{p}^{-1}(\alpha) = [\alpha - c/2]_+$. The function $\mathfrak{p}(t)$ is discontinuous at t = 0. For $\alpha \in [0, c/2]$ we have that $\mathfrak{p}^{-1}(\alpha) = 0$, and hence the corresponding set \mathcal{X} is given by

$$\mathcal{X} = \{ x \in \mathbb{R}^n : P\{C(x,\omega) \le 0\} = 1 \}.$$
(6.193)

That is for $\alpha \in [0, c/2]$ the set \mathcal{X} is defined by the constraints $C(x, \omega) \leq 0$ for *P*-a.e. $\omega \in \Omega$.

6.5 Optimization of Risk Measures

As before, we use the spaces $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $\mathcal{Z}^* = \mathcal{L}_q(\Omega, \mathcal{F}, P)$. Consider the composite function $\phi(\cdot) := \rho(F(\cdot))$, also denoted $\phi = \rho \circ F$, associated with a mapping $F : \mathbb{R}^n \to \mathcal{Z}$ and a risk measure $\rho : \mathcal{Z} \to \mathbb{R}$. We already studied properties of such composite functions in section 6.3.1. Again we write $f(x, \omega)$ or $f_{\omega}(x)$ for $[F(x)](\omega)$, and view $f(x, \omega)$ as a random function defined on the measurable space (Ω, \mathcal{F}) . Note that F(x) is an element of space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ and hence $f(x, \cdot)$ is \mathcal{F} -measurable and finite valued. If, moreover, $f(\cdot, \omega)$ is continuous for a.e. $\omega \in \Omega$, then $f(x, \omega)$ is a Carathéodory function, and hence is random lower semicontinuous.

In this section we discuss optimization problems of the form

$$\operatorname{Min}_{x \in \mathcal{X}} \left\{ \phi(x) := \rho(F(x)) \right\}.$$
(6.194)

Unless stated otherwise, we assume that the feasible set \mathcal{X} is a nonempty convex closed subset of \mathbb{R}^n . Of course, if we use $\rho(\cdot) := \mathbb{E}[\cdot]$, then problem (6.194) becomes a standard stochastic problem of optimizing (minimizing) the expected value of the random function $f(x, \omega)$. In that case we can view the corresponding optimization problem as *risk neutral*. However, a particular realization of $f(x, \omega)$ could be quite different from its expectation $\mathbb{E}[f(x, \omega)]$. This motivates an introduction, in the corresponding optimization procedure, of some type of risk control. In the analysis of Portfolio Selection (see section 1.4) we discussed an approach of using variance as a measure of risk. There is, however, a problem with such approach since the corresponding mean-variance risk measure is not monotone (see Example 6.21). We shall discuss this later.

Unless stated otherwise we assume that the risk measure ρ is proper, lower semicontinuous and satisfies conditions (R1)–(R2). By Theorem 6.5 we can use representation (6.38) to write problem (6.194) in the form

$$\underset{x \in \mathcal{X}}{\operatorname{Min}} \sup_{\zeta \in \mathfrak{A}} \Phi(x, \zeta), \tag{6.195}$$

where $\mathfrak{A} := \operatorname{dom}(\rho^*)$ and the function $\Phi : \mathbb{R}^n \times \mathcal{Z}^* \to \overline{\mathbb{R}}$ is defined by

$$\Phi(x,\zeta) := \int_{\Omega} f(x,\omega)\zeta(\omega)dP(\omega) - \rho^*(\zeta).$$
(6.196)

If, moreover, ρ is positively homogeneous, then ρ^* is the indicator function of the set \mathfrak{A} and hence $\rho^*(\cdot)$ is identically zero on \mathfrak{A} . That is, if ρ is a proper lower semicontinuous coherent risk measure, then problem (6.194) can be written as the minimax problem

$$\underset{x \in \mathcal{X}}{\operatorname{Min}} \sup_{\zeta \in \mathfrak{A}} \mathbb{E}_{\zeta}[f(x,\omega)], \tag{6.197}$$

where

$$\mathbb{E}_{\zeta}[f(x,\omega)] := \int_{\Omega} f(x,\omega) \zeta(\omega) dP(\omega)$$

denotes the expectation with respect to ζdP . Note that, by the definition, $F(x) \in \mathbb{Z}$ and $\zeta \in \mathbb{Z}^*$, and hence

$$\mathbb{E}_{\zeta}[f(x,\omega)] = \langle F(x), \zeta \rangle$$

is finite valued.

Suppose that the mapping $F : \mathbb{R}^n \to \mathbb{Z}$ is convex, i.e., for a.e. $\omega \in \Omega$ the function $f(\cdot, \omega)$ is convex. This implies that for every $\zeta \succeq 0$ the function $\Phi(\cdot, \zeta)$ is convex and if, moreover, $\zeta \in \mathfrak{A}$, then $\Phi(\cdot, \zeta)$ is real valued and hence continuous. We also have that $\langle F(x), \zeta \rangle$ is linear and $\rho^*(\zeta)$ is convex in $\zeta \in \mathbb{Z}^*$, and hence for every $x \in \mathcal{X}$ the function $\Phi(x, \cdot)$ is concave. Therefore, under various regularity conditions, there is no duality gap between problem (6.194) and its dual

$$\operatorname{Max}_{\zeta \in \mathfrak{A}} \inf_{x \in \mathcal{X}} \left\{ \int_{\Omega} f(x,\omega)\zeta(\omega)dP(\omega) - \rho^*(\zeta) \right\},\tag{6.198}$$

which is obtained by interchanging the "min" and "max" operators in (6.195) (recall that the set \mathcal{X} is assumed to be nonempty closed and convex). In particular, if there exists a saddle point $(\bar{x}, \bar{\zeta}) \in \mathcal{X} \times \mathfrak{A}$ of the minimax problem (6.195), then there is no duality gap between problems (6.195) and (6.198), and \bar{x} and $\bar{\zeta}$ are optimal solutions of (6.195) and (6.198), respectively.

Proposition 6.55. Suppose that mapping $F : \mathbb{R}^n \to \mathcal{Z}$ is convex and risk measure $\rho : \mathcal{Z} \to \mathbb{R}$ is proper, lower semicontinuous and satisfies conditions (R1)–(R2). Then $(\bar{x}, \bar{\zeta}) \in \mathcal{X} \times \mathfrak{A}$ is a saddle point of $\Phi(x, \zeta)$ if and only if $\bar{\zeta} \in \partial \rho(\bar{Z})$ and

$$0 \in \mathcal{N}_{\mathcal{X}}(\bar{x}) + \mathbb{E}_{\bar{\zeta}}[\partial f_{\omega}(\bar{x})], \tag{6.199}$$

where $\overline{Z} := F(\overline{x})$.

Proof. By the definition, $(\bar{x}, \bar{\zeta})$ is a saddle point of $\Phi(x, \zeta)$ iff

$$\bar{x} \in \arg\min_{x \in \mathcal{X}} \Phi(x, \bar{\zeta}) \text{ and } \bar{\zeta} \in \arg\max_{\zeta \in \mathfrak{A}} \Phi(\bar{x}, \zeta).$$
 (6.200)

The first of the above conditions means that $\bar{x} \in \arg \min_{x \in \mathcal{X}} \psi(x)$, where

$$\psi(x) := \int_{\Omega} f(x,\omega) \bar{\zeta}(\omega) dP(\omega).$$

Since \mathcal{X} is convex and $\psi(\cdot)$ is convex real valued, by the standard optimality conditions this holds iff $0 \in \mathcal{N}_{\mathcal{X}}(\bar{x}) + \partial \psi(\bar{x})$. Moreover, by Theorem 7.52 we have $\partial \psi(\bar{x}) = \mathbb{E}_{\bar{\zeta}}[\partial f_{\omega}(\bar{x})]$. Therefore, condition (6.199) and the first condition in (6.200) are equivalent. The second condition (6.200) and the condition $\bar{\zeta} \in \partial \rho(\bar{Z})$ are equivalent by equation (6.48). \Box

Under the assumptions of Proposition 6.55, existence of $\bar{\zeta} \in \partial \rho(\bar{Z})$ in equation (6.199) can be viewed as an optimality condition for problem (6.194). Sufficiency of that condition follows directly from the fact that it implies that $(\bar{x}, \bar{\zeta})$ is a saddle point of the min-max problem (6.195). In order for that condition to be necessary we need to verify existence of a saddle point for problem (6.195).

Proposition 6.56. Let \bar{x} be an optimal solution of the problem (6.194). Suppose that the mapping $F : \mathbb{R}^n \to \mathcal{Z}$ is convex and risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is proper, lower
semicontinuous and satisfies conditions (R1)–(R2), and is continuous at $\overline{Z} := F(\overline{x})$. Then there exists $\overline{\zeta} \in \partial \rho(\overline{Z})$ such that $(\overline{x}, \overline{\zeta})$ is a saddle point of $\Phi(x, \zeta)$.

Proof. By monotonicity of ρ (condition (R2)) it follows from the optimality of \bar{x} that (\bar{x}, \bar{Z}) is an optimal solution of the problem

$$\min_{\substack{x,Z\}\in S}}\rho(Z),\tag{6.201}$$

where $S := \{(x, Z) \in \mathcal{X} \times \mathcal{Z} : F(x) \succeq Z\}$. Since F is convex, the set S is convex, and since F is continuous (see Lemma 6.12) the set S is closed. Also because ρ is convex and continuous at \overline{Z} , the following (first order) optimality condition holds at $(\overline{x}, \overline{Z})$ (see Remark 59, page 493):

$$0 \in \partial \rho(\bar{Z}) \times \{0\} + \mathcal{N}_S(\bar{x}, \bar{Z}). \tag{6.202}$$

This means that there exists $\bar{\zeta} \in \partial \rho(\bar{Z})$ such that $(-\bar{\zeta}, 0) \in \mathcal{N}_S(\bar{x}, \bar{Z})$. This in turn implies that

$$\langle \bar{\zeta}, Z - \bar{Z} \rangle \ge 0, \ \forall (x, Z) \in S.$$
 (6.203)

Setting Z := F(x) we obtain that

$$\langle \bar{\zeta}, F(x) - F(\bar{x}) \rangle \ge 0, \quad \forall x \in \mathcal{X}.$$
 (6.204)

It follows that \bar{x} is a minimizer of $\langle \bar{\zeta}, F(x) \rangle$ over $x \in \mathcal{X}$, and hence \bar{x} is a minimizer of $\Phi(x, \bar{\zeta})$ over $x \in \mathcal{X}$. That is, \bar{x} satisfies first of the two conditions in (6.200). Moreover, as it was shown in the proof of Proposition 6.55, this implies condition (6.199), and hence $(\bar{x}, \bar{\zeta})$ is a saddle point by Proposition 6.55. \Box

Corollary 6.57. Suppose that problem (6.194) has optimal solution \bar{x} , the mapping $F : \mathbb{R}^n \to \mathcal{Z}$ is convex and the risk measure $\rho : \mathcal{Z} \to \mathbb{R}$ is proper, lower semicontinuous and satisfies conditions (R1)–(R2), and is continuous at $\bar{Z} := F(\bar{x})$. Then there is no duality gap between problems (6.195) and (6.198), and problem (6.198) has an optimal solution.

Propositions 6.55 and 6.56 imply the following optimality conditions.

Theorem 6.58. Suppose that mapping $F : \mathbb{R}^n \to Z$ is convex and risk measure $\rho : Z \to \mathbb{R}$ is proper, lower semicontinuous and satisfies conditions (R1)–(R2). Consider a point $\bar{x} \in \mathcal{X}$ and let $\bar{Z} := F(\bar{x})$. Then a sufficient condition for \bar{x} to be an optimal solution of the problem (6.194) is existence of $\zeta \in \partial \rho(\bar{Z})$ such that equation (6.199) holds. This condition is also necessary if ρ is continuous at \bar{Z} .

It could be noted that if $\rho(\cdot) := \mathbb{E}[\cdot]$, then its subdifferential consists of unique subgradient $\overline{\zeta}(\cdot) \equiv 1$. In that case condition (6.199) takes the form

$$0 \in \mathcal{N}_{\mathcal{X}}(\bar{x}) + \mathbb{E}[\partial f_{\omega}(\bar{x})]. \tag{6.205}$$

Note that since it is assumed that $F(x) \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$, the expectation $\mathbb{E}[f_{\omega}(x)]$ is well defined and finite valued for all x, and hence $\partial \mathbb{E}[f_{\omega}(x)] = \mathbb{E}[\partial f_{\omega}(x)]$ (see Theorem 7.52).

6.5.1 Dualization of Nonanticipativity Constraints

We assume again that $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $\mathcal{Z}^* = \mathcal{L}_q(\Omega, \mathcal{F}, P)$, that $F : \mathbb{R}^n \to \mathcal{Z}$ is convex and $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is proper lower semicontinuous and satisfies conditions (R1) and (R2). A way of representing problem (6.194) is to consider the decision vector x as a function of the elementary event $\omega \in \Omega$ and then to impose an appropriate nonaniticipativity constraint. That is, let \mathfrak{M} be a linear space of \mathcal{F} -measurable mappings $\chi : \Omega \to \mathbb{R}^n$. Define $F_{\chi}(\omega) := f(\chi(\omega), \omega)$ and

$$\mathfrak{M}_{\mathcal{X}} := \{ \chi \in \mathfrak{M} : \chi(\omega) \in \mathcal{X}, \text{ a.e. } \omega \in \Omega \}.$$
(6.206)

We assume that the space \mathfrak{M} is chosen in such a way that $F_{\chi} \in \mathcal{Z}$ for every $\chi \in \mathfrak{M}$ and for every $x \in \mathbb{R}^n$ the constant mapping $\chi(\omega) \equiv x$ belongs to \mathfrak{M} . Then we can write problem (6.194) in the following equivalent form

$$\underset{(\chi,x)\in\mathfrak{M}_{\mathcal{X}}\times\mathbb{R}^{n}}{\operatorname{Min}}\rho(F_{\chi}) \text{ subject to } \chi(\omega) = x, \text{ a.e. } \omega \in \Omega.$$
(6.207)

Formulation (6.207) allows developing a duality framework associated with the *nonanticipativity* constraint $\chi(\cdot) = x$. In order to formulate such duality we need to specify the space \mathfrak{M} and its dual. It looks natural to use $\mathfrak{M} := \mathcal{L}_{p'}(\Omega, \mathcal{F}, P; \mathbb{R}^n)$, for some $p' \in [1, +\infty)$, and its dual $\mathfrak{M}^* := \mathcal{L}_{q'}(\Omega, \mathcal{F}, P; \mathbb{R}^n)$, $q' \in (1, +\infty]$. It is also possible to employ $\mathfrak{M} := \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P; \mathbb{R}^n)$. Unfortunately this Banach space is not reflexive. Nevertheless, it can be paired with the space $\mathcal{L}_1(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ by defining the corresponding scalar product in the usual way. As long as the risk measure is lower semicontinuous and subdifferentiable in the corresponding weak topology, we can use this setting as well.

The (Lagrangian) dual of problem (6.207) can be written in the form

$$\operatorname{Max}_{\lambda \in \mathfrak{M}^*} \left\{ \inf_{(\chi, x) \in \mathfrak{M}_{\mathcal{X}} \times \mathbb{R}^n} L(\chi, x, \lambda) \right\},$$
(6.208)

where

$$L(\chi, x, \lambda) := \rho(F_{\chi}) + \mathbb{E}[\lambda^{\mathsf{T}}(\chi - x)], \ (\chi, x, \lambda) \in \mathfrak{M} \times \mathbb{R}^{n} \times \mathfrak{M}^{*}.$$
(6.209)

Note that

$$\inf_{\in \mathbb{R}^n} L(\chi, x, \lambda) = \begin{cases} L(\chi, 0, \lambda), & \text{if } \mathbb{E}[\lambda] = 0, \\ -\infty, & \text{if } \mathbb{E}[\lambda] \neq 0. \end{cases}$$

Therefore the dual problem (6.209) can be rewritten in the form

$$\operatorname{Max}_{\lambda \in \mathfrak{M}^*} \left\{ \inf_{\chi \in \mathfrak{M}_{\mathcal{X}}} L_0(\chi, \lambda) \right\} \text{ subject to } \mathbb{E}[\lambda] = 0,$$
(6.210)

where $L_0(\chi, \lambda) := L(\chi, 0, \lambda) = \rho(F_{\chi}) + \mathbb{E}[\lambda^{\mathsf{T}}\chi].$

We have that the optimal value of problem (6.207) (which is the same as the optimal value of problem (6.194)) is greater than or equal to the optimal value of its dual (6.210). Moreover, under some regularity conditions, their optimal values are equal to each other. In particular, if Lagrangian $L(\chi, x, \lambda)$ has a saddle point $((\bar{\chi}, \bar{x}), \bar{\lambda})$, then there is no duality gap between problems (6.207) and (6.210), and $(\bar{\chi}, \bar{x})$ and $\bar{\lambda}$ are optimal solutions of

problems (6.207) and (6.210), respectively. Noting that $L(\chi, 0, \lambda)$ is linear in x and in λ , we have that $((\bar{\chi}, \bar{x}), \bar{\lambda})$ is a saddle point of $L(\chi, x, \lambda)$ iff the following conditions hold:

$$\bar{\chi}(\omega) = \bar{x}, \text{ a.e. } \omega \in \Omega, \text{ and } \mathbb{E}[\lambda] = 0,$$

$$\bar{\chi} \in \arg\min_{\chi \in \mathfrak{M}_{\mathcal{X}}} L_0(\chi, \bar{\lambda}).$$
(6.211)

Unfortunately, it may be not be easy to verify existence of such saddle point.

We can approach the duality analysis by conjugate duality techniques. For a perturbation vector $y \in \mathfrak{M}$ consider the problem

$$\underset{(\chi,x)\in\mathfrak{M}_{\mathcal{X}}\times\mathbb{R}^{n}}{\operatorname{Min}}\rho(F_{\chi}) \text{ subject to } \chi(\omega) = x + y(\omega),$$
(6.212)

and let $\vartheta(y)$ be its optimal value. Note that a perturbation in the vector x, in the constraints of problem (6.207), can be absorbed into $y(\omega)$. Clearly for y = 0, problem (6.212) coincides with the unperturbed problem (6.207), and $\vartheta(0)$ is the optimal value of the unperturbed problem (6.207). Assume that $\vartheta(0)$ is finite. Then there is no duality gap between problem (6.207) and its dual (6.208) iff $\vartheta(y)$ is lower semicontinuous at y = 0. Again it may be not easy to verify lower semicontinuity of the optimal value function $\vartheta : \mathfrak{M} \to \mathbb{R}$. By the general theory of conjugate duality we have the following result.

Proposition 6.59. Suppose that $F : \mathbb{R}^n \to \mathbb{Z}$ is convex, $\rho : \mathbb{Z} \to \mathbb{R}$ satisfies conditions (R1)–(R2) and the function $\rho(F_{\chi})$, from \mathfrak{M} to \mathbb{R} , is lower semicontinuous. Suppose, further, that $\vartheta(0)$ is finite and $\vartheta(y) < +\infty$ for all y in a neighborhood (in the norm topology) of $0 \in \mathfrak{M}$. Then there is no duality gap between problems (6.207) (6.208), and the dual problem (6.208) has an optimal solution.

Proof. Since ρ satisfies conditions (R1) and (R2) and F is convex, we have that the function $\rho(F_{\chi})$ is convex, and by the assumption it is lower semicontinuous. The assertion then follows by a general result of conjugate duality for Banach spaces (see Theorem 7.88).

In order to apply the above result we need to verify lower semicontinuity of the function $\rho(F_{\chi})$. This function is lower semicontinuous if $\rho(\cdot)$ is lower semicontinuous and the mapping $\chi \mapsto F_{\chi}$, from \mathfrak{M} to \mathcal{Z} , is continuous. If the set Ω is finite, and hence the spaces \mathcal{Z} and \mathfrak{M} are finite dimensional, then continuity of $\chi \mapsto F_{\chi}$ follows from the continuity of F. In the infinite dimensional setting this should be verified by specialized methods. The assumption that $\vartheta(0)$ is finite means that the optimal value of the problem (6.207) is finite, and the assumption that $\vartheta(y) < +\infty$ means that the corresponding problem (6.212) has a feasible solution.

6.5.2 Interchangeability Principle for Risk Measures

By removing the nonanticipativity constraint $\chi(\cdot) = x$ we obtain the following relaxation of the problem (6.207):

$$\underset{\chi \in \mathfrak{M}_{\mathcal{X}}}{\operatorname{Min}} \rho(F_{\chi}),$$
(6.213)

where \mathfrak{M}_X is defined in (6.206). Similarly to the interchangeability principle for the expectation operator (Theorem 7.92), we have the following result for monotone risk measures. By $\inf_{x \in \mathcal{X}} F(x)$ we denote the pointwise minimum, i.e.,

$$\left[\inf_{x\in\mathcal{X}}F(x)\right](\omega) := \inf_{x\in\mathcal{X}}f(x,\omega), \ \omega\in\Omega.$$
(6.214)

Proposition 6.60. Let $Z := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $\mathfrak{M} := \mathcal{L}_{p'}(\Omega, \mathcal{F}, P; \mathbb{R}^n)$, where $p, p' \in [1, \infty]$, $\mathfrak{M}_{\mathcal{X}}$ be defined in (6.206), $\rho : Z \to \overline{\mathbb{R}}$ be a proper risk measure satisfying monotonicity condition (R2), and $F : \mathbb{R}^n \to Z$ be such that $\inf_{x \in \mathcal{X}} F(x) \in Z$. Suppose that ρ is continuous at $\Psi := \inf_{x \in \mathcal{X}} F(x)$. Then

$$\inf_{\chi \in \mathfrak{M}_{\mathcal{X}}} \rho(F_{\chi}) = \rho\left(\inf_{x \in \mathcal{X}} F(x)\right).$$
(6.215)

Proof. For any $\chi \in \mathfrak{M}_{\mathcal{X}}$ we have that $\chi(\cdot) \in \mathcal{X}$, and hence the following inequality holds

$$\left[\inf_{x\in\mathcal{X}}F(x)\right](\omega)\leq F_{\chi}(\omega), \text{ a.e. } \omega\in\Omega.$$

By monotonicity of ρ this implies that $\rho(\Psi) \leq \rho(F_{\chi})$, and hence

$$\rho\left(\Psi\right) \le \inf_{\chi \in \mathfrak{M}_{\chi}} \rho(F_{\chi}). \tag{6.216}$$

Since ρ is proper we have that $\rho(\Psi) > -\infty$. If $\rho(\Psi) = +\infty$, then by (6.216) the left hand side of (6.215) is also $+\infty$ and hence (6.215) holds. Therefore we can assume that $\rho(\Psi)$ is finite.

Let us derive now the converse of (6.216) inequality. Since it is assumed that $\Psi \in \mathcal{Z}$, we have that $\Psi(\omega)$ is finite valued for a.e. $\omega \in \Omega$ and measurable. Therefore for a sequence $\varepsilon_k \downarrow 0$ and a.e. $\omega \in \Omega$ and all $k \in \mathbb{N}$, we can choose $\chi_k(\omega) \in \mathcal{X}$ such that $|f(\chi_k(\omega), \omega) - \Psi(\omega)| \leq \varepsilon_k$ and $\chi_k(\cdot)$ are measurable. We also can truncate $\chi_k(\cdot)$, if necessary, in such a way that each χ_k belongs to $\mathfrak{M}_{\mathcal{X}}$, and $f(\chi_k(\omega), \omega)$ monotonically converges to $\Psi(\omega)$ for a.e. $\omega \in \Omega$. We have then that $f(\chi_k(\cdot), \cdot) - \Psi(\cdot)$ is nonnegative valued and is dominated by a function from the space \mathcal{Z} . It follows by the Lebesgue Dominated Convergence Theorem that F_{χ_k} converges to Ψ in the norm topology of \mathcal{Z} . Since ρ is continuous at Ψ , it follows that $\rho(F_{\chi_k})$ tends to $\rho(\Psi)$. Also $\inf_{\chi \in \mathfrak{M}_{\mathcal{X}}} \rho(F_{\chi}) \leq \rho(F_{\chi_k})$, and hence the required converse inequality

$$\inf_{\chi \in \mathfrak{M}_{\chi}} \rho(F_{\chi}) \le \rho(\Psi) \tag{6.217}$$

follows.

Remark 39. It follows from (6.215) that if

$$\bar{\chi} \in \arg\min_{\chi \in \mathfrak{M}_{\mathcal{X}}} \rho(F_{\chi}), \tag{6.218}$$

then

$$\bar{\chi}(\omega) \in \arg\min_{x \in \mathcal{X}} f(x, \omega), \text{ a.e. } \omega \in \Omega.$$
 (6.219)

Conversely, suppose that the function $f(x, \omega)$ is random lower semicontinuous. Then the multifunction $\omega \mapsto \arg \min_{x \in X} f(x, \omega)$ is measurable. Therefore $\bar{\chi}(\omega)$ in the left hand side of (6.219) can be chosen to be measurable. If, moreover, $\bar{\chi} \in \mathfrak{M}$ (this holds, in particular, if the set \mathcal{X} is bounded and hence $\bar{\chi}(\cdot)$ is bounded), then the inclusion (6.218) follows.

Consider now a setting of two stage programming. That is, suppose that the function $[F(x)](\omega) = f(x, \omega)$, of the first stage problem

$$\min_{x \in \mathcal{X}} \rho(F(x)),$$
(6.220)

is given by the optimal value of the following second stage problem

$$\underset{y \in \mathcal{G}(x,\omega)}{\operatorname{Min}} g(x, y, \omega),$$
(6.221)

where $g : \mathbb{R}^n \times \mathbb{R}^m \times \Omega \to \mathbb{R}$ and $\mathcal{G} : \mathbb{R}^n \times \Omega \rightrightarrows \mathbb{R}^m$. Under appropriate regularity conditions, from which the most important is the monotonicity condition (R2), we can apply the interchangeability principle to the optimization problem (6.221) to obtain

$$\rho(F(x)) = \inf_{y(\cdot) \in \mathcal{G}(x, \cdot)} \rho(g(x, y(\omega), \omega)), \tag{6.222}$$

where now $y(\cdot)$ is an element of an appropriate functional space and the notation $y(\cdot) \in \mathcal{G}(x, \cdot)$ means that $y(\omega) \in \mathcal{G}(x, \omega)$ w.p.1. If the interchangeability principle (6.222) holds, then the two stage problem (6.220)–(6.221) can be written as one large optimization problem:

$$\min_{x \in \mathcal{X}, y(\cdot) \in \mathcal{G}(x, \cdot)} \rho(g(x, y(\omega), \omega)).$$
(6.223)

In particular, suppose that the set Ω is finite, say $\Omega = \{\omega_1, \ldots, \omega_K\}$, i.e., there is a finite number K of scenarios. In that case we can view function $Z : \Omega \to \mathbb{R}$ as vector $(Z(\omega_1), \ldots, Z(\omega_K)) \in \mathbb{R}^K$, and hence to identify the space Z with \mathbb{R}^K . Then problem (6.223) takes the form

$$\underset{x \in \mathcal{X}, y_k \in \mathcal{G}(x, \omega_k), k=1, \dots, K}{\operatorname{Min}} \rho \left[(g(x, y_1, \omega_1), \dots, g(x, y_K, \omega_K)) \right].$$
(6.224)

Moreover, consider the linear case where $\mathcal{X} := \{x : Ax = b, x \ge 0\}, g(x, y, \omega) := c^{\mathsf{T}}x + q(\omega)^{\mathsf{T}}y$ and

$$\mathcal{G}(x,\omega) := \{ y : T(\omega)x + W(\omega)y = h(\omega), \ y \ge 0 \}.$$

Assume that ρ satisfies conditions (R1)–(R3) and the set $\Omega = \{\omega_1, \ldots, \omega_K\}$ is finite. Then problem (6.224) takes the form:

$$\underset{x,y_{1},\dots,y_{K}}{\min} c^{\mathsf{T}}x + \rho \left[\left(q_{1}^{\mathsf{T}}y_{1},\dots,q_{K}^{\mathsf{T}}y_{K} \right) \right] \\$$
s.t. $Ax = b, \ x \ge 0, \ T_{k}x + W_{k}y_{k} = h_{k}, \ y_{k} \ge 0, \ k = 1,\dots,K,$
(6.225)

where $(q_k, T_k, W_k, h_k) := (q(\omega_k), T(\omega_k), W(\omega_k), h(\omega_k)), k = 1, ..., K.$

6.5.3 Examples

Let $\mathcal{Z} := \mathcal{L}_1(\Omega, \mathcal{F}, P)$ and consider

$$\rho(Z) := \mathbb{E}[Z] + \inf_{t \in \mathbb{R}} \mathbb{E}\{\beta_1[t - Z]_+ + \beta_2[Z - t]_+\}, \ Z \in \mathcal{Z},$$
(6.226)

where $\beta_1 \in [0, 1]$ and $\beta_2 \ge 0$ are some constants. Properties of this risk measure were studied in Example 6.19 (see equations (6.73) and (6.74) in particular). We can write the corresponding optimization problem (6.194) in the following equivalent form

$$\underset{(x,t)\in\mathcal{X}\times\mathbb{R}}{\operatorname{Min}} \mathbb{E}\left\{f_{\omega}(x) + \beta_{1}[t - f_{\omega}(x)]_{+} + \beta_{2}[f_{\omega}(x) - t]_{+}\right\}.$$
(6.227)

That is, by adding one extra variable we can formulate the corresponding optimization problem as an expectation minimization problem.

Risk Averse Optimization of an Inventory Model

Let us consider again the inventory model analyzed in section 1.2. Recall that the objective of that model is to minimize the total cost

$$F(x,d) = cx + b[d-x]_{+} + h[x-d]_{+},$$

where c, b and h are nonnegative constants representing costs of ordering, back ordering and holding, respectively. Again we assume that b > c > 0, i.e., the back order cost is *bigger* than the ordering cost. A risk averse extension of the corresponding (expected value) problem (1.4) can be formulated in the form

$$\min_{x \ge 0} \{ f(x) := \rho[F(x, D)] \},$$
(6.228)

where ρ is a specified risk measure.

Assume that the risk measure ρ is coherent, i.e., satisfies conditions (R1)–(R4), and that demand $D = D(\omega)$ belongs to an appropriate space $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$. Assume, further, that $\rho : \mathcal{Z} \to \mathbb{R}$ is *real valued*. It follows that there exists a convex set $\mathfrak{A} \subset \mathfrak{P}$, where $\mathfrak{P} \subset \mathcal{Z}^*$ is the set of probability density functions, such that

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \int_{\Omega} Z(\omega) \zeta(\omega) dP(\omega), \ Z \in \mathcal{Z}$$

Consequently we have that

$$\rho[F(x,D)] = \sup_{\zeta \in \mathfrak{A}} \int_{\Omega} F(x,D(\omega))\zeta(\omega)dP(\omega).$$
(6.229)

To each $\zeta \in \mathfrak{P}$ corresponds the cumulative distribution function H of D with respect to the measure $Q := \zeta dP$, that is,

$$H(z) = Q(D \le z) = \mathbb{E}_{\zeta}[\mathbf{1}_{D \le z}] = \int_{\{\omega: D(\omega) \le z\}} \zeta(\omega) dP(\omega).$$
(6.230)

We have then that

$$\int_{\Omega} F(x, D(\omega))\zeta(\omega)dP(\omega) = \int F(x, z)dH(z).$$

Denote by \mathfrak{M} the set of cumulative distribution functions H associated with densities $\zeta \in \mathfrak{A}$. \mathfrak{A} . The correspondence between $\zeta \in \mathfrak{A}$ and $H \in \mathfrak{M}$ is given by formula (6.230), and depends on $D(\cdot)$ and the reference probability measure P. Then we can rewrite (6.229) in the form

$$\rho[F(x,D)] = \sup_{H \in \mathfrak{M}} \int F(x,z) dH(z) = \sup_{H \in \mathfrak{M}} \mathbb{E}_H[F(x,D)].$$
(6.231)

This leads to the following minimax formulation of the risk averse optimization problem (6.228):

$$\underset{x \ge 0}{\operatorname{Min}} \sup_{H \in \mathfrak{M}} \mathbb{E}_H[F(x, D)].$$
(6.232)

Note that we also have that $\rho(D) = \sup_{H \in \mathfrak{M}} \mathbb{E}_H[D]$.

In the subsequent analysis we deal with the minimax formulation (6.232), rather than the risk averse formulation (6.228), viewing \mathfrak{M} as a given set of cumulative distribution functions. We show next that the minimax problem (6.232), and hence the risk averse problem (6.228), structurally is similar to the corresponding (expected value) problem (1.4). We assume that every $H \in \mathfrak{M}$ is such that H(z) = 0 for any z < 0 (recall that the demand cannot be negative). We also assume that $\sup_{H \in \mathfrak{M}} \mathbb{E}_H[D] < +\infty$, which follows from the assumption that $\rho(\cdot)$ is real valued.

Proposition 6.61. Let \mathfrak{M} be a set of cumulative distribution functions such that H(z) = 0for any $H \in \mathfrak{M}$ and z < 0, and $\sup_{H \in \mathfrak{M}} \mathbb{E}_H[D] < +\infty$. Consider function $f(x) := \sup_{H \in \mathfrak{M}} \mathbb{E}_H[F(x, D)]$. Then there exists a cdf \overline{H} , depending on the set \mathfrak{M} and $\eta := b/(b+h)$, such that $\overline{H}(z) = 0$ for any z < 0, and the function f(x) can be written in the form

$$f(x) = b \sup_{H \in \mathfrak{M}} \mathbb{E}_H[D] + (c-b)x + (b+h) \int_{-\infty}^{-} \bar{H}(z) dz.$$
(6.233)

Proof. We have (see formula (1.5)) that for $H \in \mathfrak{M}$,

$$\mathbb{E}_H[F(x,D)] = b \mathbb{E}_H[D] + (c-b)x + (b+h) \int_0^x H(z)dz.$$

Therefore we can write $f(x) = (c - b)x + (b + h)\mathfrak{g}(x)$, where

$$\mathfrak{g}(x) := \sup_{H \in \mathfrak{M}} \left\{ \eta \mathbb{E}_H[D] + \int_{-\infty}^x H(z) dz \right\}.$$
(6.234)

Since every $H \in \mathfrak{M}$ is a monotonically nondecreasing function, we have that $x \mapsto \int_{-\infty}^{x} H(z) dz$ is a convex function. It follows that the function $\mathfrak{g}(x)$ is given by the maximum of convex functions and hence is convex. Moreover, $\mathfrak{g}(x) \ge 0$ and

$$\mathfrak{g}(x) \le \eta \sup_{H \in \mathfrak{M}} \mathbb{E}_H[D] + [x]_+.$$
(6.235)

and hence $\mathfrak{g}(x)$ is finite valued for any $x \in \mathbb{R}$. Also for any $H \in \mathfrak{M}$ and z < 0 we have that H(z) = 0, and hence $\mathfrak{g}(x) = \eta \sup_{H \in \mathfrak{M}} \mathbb{E}_H[D]$ for any x < 0.

Consider the right hand side derivative of $\mathfrak{g}(x)$:

$$\mathfrak{g}^+(x) := \lim_{t \downarrow 0} \frac{\mathfrak{g}(x+t) - \mathfrak{g}(x)}{t},$$

and define $\overline{H}(\cdot) := \mathfrak{g}^+(\cdot)$. Since $\mathfrak{g}(x)$ is real valued convex, its right hand side derivative $\mathfrak{g}^+(x)$ exists, is finite and for any $x \ge 0$ and a < 0,

$$\mathfrak{g}(x) = \mathfrak{g}(a) + \int_{a}^{x} \mathfrak{g}^{+}(z)dz = \eta \sup_{H \in \mathfrak{M}} \mathbb{E}_{H}[D] + \int_{-\infty}^{x} \bar{H}(z)dz.$$
(6.236)

Note that definition of the function $\mathfrak{g}(\cdot)$, and hence $\overline{H}(\cdot)$, involves the constant η and set \mathfrak{M} only. Let us also observe that the right hand side derivative $\mathfrak{g}^+(x)$, of a real valued convex function, is monotonically nondecreasing and right side continuous. Moreover, $\mathfrak{g}^+(x) = 0$ for x < 0 since $\mathfrak{g}(x)$ is constant for x < 0. We also have that $\mathfrak{g}^+(x)$ tends to one as $x \to +\infty$. Indeed, since $\mathfrak{g}^+(x)$ is monotonically nondecreasing it tends to a limit, denoted r, as $x \to +\infty$. We have then that $\mathfrak{g}(x)/x \to r$ as $x \to +\infty$. It follows from (6.235) that $r \leq 1$, and by (6.234) that for any $H \in \mathfrak{M}$,

$$\liminf_{x \to +\infty} \frac{\mathfrak{g}(x)}{x} \ge \liminf_{x \to +\infty} \frac{1}{x} \int_{-\infty}^{x} H(z) dz \ge 1,$$

and hence $r \ge 1$. It follows that r = 1.

We obtain that $\overline{H}(\cdot) = \mathfrak{g}^+(\cdot)$ is a cumulative distribution function of some probability distribution and the representation (6.233) holds. \Box

It follows from the representation (6.233) that the set of optimal solutions of the risk averse problem (6.228) is an interval given by the set of κ -quantiles of the cdf $\bar{H}(\cdot)$, where $\kappa := \frac{b-c}{b+h}$ (compare with Remark 1, page 3).

In some specific cases it is possible to calculate the corresponding cdf \overline{H} in a closed form. Consider the risk measure ρ defined in (6.226):

$$\rho(Z) := \mathbb{E}[Z] + \inf_{t \in \mathbb{R}} \mathbb{E}\left\{\beta_1[t-Z]_+ + \beta_2[Z-t]_+\right\},\$$

where the expectations are taken with respect to some reference cdf $H^*(\cdot)$. The corresponding set \mathfrak{M} is formed by cumulative distribution functions $H(\cdot)$ such that

$$(1 - \beta_1) \int_S dH^* \le \int_S dH \le (1 + \beta_2) \int_S dH^*$$
(6.237)

for any Borel set $S \subset \mathbb{R}$ (compare with formula (6.75)). Recall that for $\beta_1 = 1$ this risk measure is $\rho(Z) = AV@R_{\alpha}(Z)$ with $\alpha = 1/(1 + \beta_2)$. Suppose that the reference distribution of the demand is uniform on the interval [0, 1], i.e., $H^*(z) = z$ for $z \in [0, 1]$. It follows that any $H \in \mathfrak{M}$ is continuous, H(0) = 0 and H(1) = 1, and

$$\mathbb{E}_{H}[D] = \int_{0}^{1} z dH(z) = zH(z) \big|_{0}^{1} - \int_{0}^{1} H(z) dz = 1 - \int_{0}^{1} H(z) dz.$$

Consequently we can write function g(x), defined in (6.234), for $x \in [0, 1]$ in the form

$$\mathfrak{g}(x) = \eta + \sup_{H \in \mathfrak{M}} \left\{ (1 - \eta) \int_0^x H(z) dz - \eta \int_x^1 H(z) dz \right\}.$$
 (6.238)

Suppose, further, that h = 0 (i.e., there are no holding costs) and hence $\eta = 1$. In that case

$$\mathfrak{g}(x) = 1 - \inf_{H \in \mathfrak{M}} \int_{x}^{1} H(z) dz, \text{ for } x \in [0, 1].$$
(6.239)

By using the first inequality of (6.237) with S := [0, z] we obtain that $H(z) \ge (1 - \beta_1)z$ for any $H \in \mathfrak{M}$ and $z \in [0, 1]$. Similarly, by the second inequality of (6.237) with S := [z, 1] we have that $H(z) \ge 1 + (1 + \beta_2)(z - 1)$ for any $H \in \mathfrak{M}$ and $z \in [0, 1]$. Consequently, the cdf

$$\bar{H}(z) := \max\{(1 - \beta_1)z, (1 + \beta_2)z - \beta_2\}, \ z \in [0, 1],$$
(6.240)

is dominated by any other cdf $H \in \mathfrak{M}$, and it can be verified that $\overline{H} \in \mathfrak{M}$. Therefore the minimum on the right hand side of (6.239) is attained at \overline{H} for any $x \in [0, 1]$, and hence this cdf \overline{H} fulfils equation (6.233).

Note that for any $\beta_1 \in (0,1)$ and $\beta_2 > 0$, the cdf $\overline{H}(\cdot)$ defined in (6.240) is strictly less than the reference cdf $H^*(\cdot)$ on the interval (0,1). Consequently the corresponding risk averse optimal solution $\overline{H}^{-1}(\kappa)$ is bigger than the risk neutral optimal solution $H^{*-1}(\kappa)$. It should be not surprising that in the absence of holding costs it will be safer to order a larger quantity of the product.

Risk Averse Portfolio Selection

Consider the portfolio selection problem introduced in section 1.4. A risk averse formulation of the corresponding optimization problem can be written in the form

$$\min_{x \in \mathcal{X}} \rho\Big(-\sum_{i=1}^{n} \xi_i x_i\Big),\tag{6.241}$$

where ρ is a chosen risk measure and $\mathcal{X} := \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = W_0, x \ge 0\}$. We use the negative of the return as an argument of the risk measure, because we developed our theory for the minimization, rather than maximization framework. An example below shows what could be a problem of using risk measures with dispersions measured by variance or standard deviation.

Example 6.62 Let n = 2, $W_0 = 1$ and the risk measure ρ be of the form

$$\rho(Z) := \mathbb{E}[Z] + c \mathbb{D}[Z], \tag{6.242}$$

where c > 0 and $\mathbb{D}[\cdot]$ is a dispersion measure. Let the dispersion measure be either $\mathbb{D}[Z] := \sqrt{\mathbb{V}ar[Z]}$ or $\mathbb{D}[Z] := \mathbb{V}ar[Z]$. Suppose, further, that the space $\Omega := \{\omega_1, \omega_2\}$ consists of two points with associated probabilities p and 1 - p, for some $p \in (0, 1)$. Define (random) return rates $\xi_1, \xi_2 : \Omega \to \mathbb{R}$ as follows: $\xi_1(\omega_1) = a$ and $\xi_1(\omega_2) = 0$, where a is some positive number, and $\xi_2(\omega_1) = \xi_2(\omega_2) = 0$. Obviously, is better to invest

in asset 1 than asset 2. Now, for $\mathbb{D}[Z] := \sqrt{\mathbb{Var}[Z]}$, we have that $\rho(-\xi_2) = 0$ and $\rho(-\xi_1) = -pa + ca\sqrt{p(1-p)}$. It follows that $\rho(-\xi_1) > \rho(-\xi_2)$ for any c > 0 and $p < (1+c^{-2})^{-1}$. Similarly, for $\mathbb{D}[Z] := \mathbb{Var}[Z]$ we have that $\rho(-\xi_1) = -pa + ca^2p(1-p)$, $\rho(-\xi_2) = 0$, and hence $\rho(-\xi_1) > \rho(-\xi_2)$ again, provided $p < 1 - (ca)^{-1}$. That is, although ξ_2 dominates ξ_1 in the sense that $\xi_1(\omega) \ge \xi_2(\omega)$ for every possible realization of $(\xi_1(\omega), \xi_2(\omega))$, we have that $\rho(-\xi_1) > \rho(-\xi_2)$.

Here $[F(x)](\omega) := -\xi_1(\omega)x_1 - \xi_2(\omega)x_2$. Let $\bar{x} := (1, 0)$ and $x^* := (0, 1)$. Note that the feasible set X is formed by vectors $t\bar{x} + (1-t)x^*$, $t \in [0, 1]$. We have that $[F(x)](\omega) = -\xi_1(\omega)x_1$, and hence $[F(\bar{x})](\omega)$ is dominated by $[F(x)](\omega)$ for any $x \in X$ and $\omega \in \Omega$. And yet, under the specified conditions, we have that $\rho[F(\bar{x})] = \rho(-\xi_1)$ is greater than $\rho[F(x^*)] = \rho(-\xi_2)$, and hence \bar{x} is *not* an optimal solution of the corresponding optimization (minimization) problem. This should be not surprising, because the chosen risk measure is not monotone, i.e., it does not satisfy the condition (R2), for c > 0 (see Examples 6.21 and 6.22).

Suppose now that ρ is a real valued coherent risk measure. We can write then problem (6.241) in the corresponding min-max form (6.197), that is

$$\operatorname{Min}_{x \in \mathcal{X}} \sup_{\zeta \in \mathfrak{A}} \sum_{i=1}^{n} \left(-\mathbb{E}_{\zeta}[\xi_i] \right) x_i.$$

Equivalently,

$$\operatorname{Max}_{x \in \mathcal{X}} \inf_{\zeta \in \mathfrak{A}} \sum_{i=1}^{n} \left(\mathbb{E}_{\zeta}[\xi_i] \right) x_i.$$
(6.243)

Since the feasible set \mathcal{X} is compact, problem (6.241) always has an optimal solution \bar{x} . Also (see Proposition 6.56) the min-max problem (6.243) has a saddle point, and $(\bar{x}, \bar{\zeta})$ is a saddle point iff

$$\bar{\zeta} \in \partial \rho(\bar{Z}) \text{ and } \bar{x} \in \arg \max_{x \in \mathcal{X}} \sum_{i=1}^{n} \bar{\mu}_{i} x_{i},$$
(6.244)

where $\bar{Z}(\omega) := -\sum_{i=1}^{n} \xi_i(\omega) \bar{x}_i$ and $\bar{\mu}_i := \mathbb{E}_{\bar{\zeta}}[\xi_i]$.

An interesting insight into the risk averse solution is provided by its game-theoretical interpretation. For $W_0 = 1$ the portfolio allocations x can be interpreted as a *mixed strategy* of the investor (for another W_0 the fractions x_i/W_0 are the mixed strategy). The measure ζ represents the mixed strategy of the opponent (the market). It is not chosen from the set of all possible mixed strategies, but rather from the set \mathfrak{A} . The risk averse solution (6.244) corresponds to the equilibrium of the game.

It is not difficult to see that the set $\arg \max_{x \in X} \sum_{i=1}^{n} \overline{\mu}_i x_i$ is formed by all convex combinations of vectors $W_0 e_i$, $i \in \mathcal{I}$, where $e_i \in \mathbb{R}^n$ denotes *i*-th coordinate vector (with zero entries except the *i*-th entry equal 1), and

$$\mathcal{I} := \{ i' : \bar{\mu}_{i'} = \max_{1 < i < n} \bar{\mu}_i, \ i' = 1, \dots, n \}.$$

Also $\partial \rho(Z) \subset \mathfrak{A}$ (see formula (6.49) for the subdifferential $\partial \rho(Z)$).

6.6 Statistical Properties of Risk Measures

All examples of risk measures discussed in section 6.3.2 were constructed with respect to a reference probability measure (distribution) P. Suppose now that the "true" probability distribution P is estimated by an empirical measure (distribution) P_N based on a sample of size N. In this section we discuss statistical properties of the respective estimates of the "true values" of the corresponding risk measures.

6.6.1 Average Value-at-Risk

Recall that the Average Value-at-Risk, $AV@R_{1-\alpha}(Z)$ at level $1 - \alpha \in (0, 1)$, of a random variable Z, is given by the optimal value of the minimization problem

$$\min_{t \in \mathbb{R}} f(t),$$
(6.245)

where

$$f(t) := \mathbb{E}[F_{\alpha}(t, Z)] \text{ with } F_{\alpha}(t, z) := t + (1 - \alpha)^{-1}[z - t]_{+},$$
(6.246)

and the expectation is taken with respect to the probability distribution P of Z. We assume that $\mathbb{E}|Z| < +\infty$, which implies that $AV@R_{1-\alpha}(Z)$ is finite. Suppose now that we have an iid random sample Z^1, \ldots, Z^N of N realizations of Z. Then we can estimate $\theta^* := AV@R_{1-\alpha}(Z)$ by replacing distribution P with its empirical estimate²⁰ $P_N := N^{-1} \sum_{j=1}^N \delta(Z^j)$. Since $\mathbb{E}_{P_N}[Z-t]_+ = N^{-1} \sum_{j=1}^N [Z^j - t]_+$, this leads to the sample estimate $\hat{\theta}_N$, of θ^* , given by the optimal value of the following problem

$$\min_{t \in \mathbb{R}} \hat{f}_N(t),$$
(6.247)

where

$$\hat{f}_N(t) := \frac{1}{N} \sum_{j=1}^N F_\alpha(t, Z^j) = t + \frac{1}{(1-\alpha)N} \sum_{j=1}^N \left[Z^j - t \right]_+$$

Let us observe that problem (6.245) can be viewed as a stochastic programming problem and problem (6.247) as its sample average approximation. That is,

$$\theta^* = \inf_{t \in \mathbb{R}} f(t) \text{ and } \hat{\theta}_N = \inf_{t \in \mathbb{R}} \hat{f}_N(t).$$
(6.248)

Therefore, results of section 5.1 can be applied here in a straightforward way. Recall that the set of optimal solutions of problem (6.245) is the interval $[t^*, t^{**}]$, where

$$t^* = \inf\{z : H_Z(z) \ge \alpha\} = \mathsf{V}@\mathsf{R}_{1-\alpha}(Z) \text{ and } t^{**} = \sup\{z : H_Z(z) \le \alpha\}$$
 (6.249)

are the respective left and right side α -quantiles of the distribution of Z (see page 294). Since for any $\alpha \in (0,1)$ the interval $[t^*, t^{**}]$ is finite and problem (6.245) is convex, we have by Theorem 5.4 that

$$\theta_N \to \theta^* \quad \text{w.p.1 as } N \to \infty.$$
 (6.250)

²⁰Recall that $\delta(z)$ denotes measure of mass one at point z.

That is, $\hat{\theta}_N$ is a consistent estimator of $\theta^* = AV@R_{1-\alpha}(Z)$. This also follows from a general result of Theorem 7.58.

Assume now that $\mathbb{E}[Z^2] < +\infty$. Then the assumptions (A1) and (A2) of Theorem 5.7 hold, and hence

$$\hat{\theta}_N = \inf_{t \in [t^*, t^{**}]} \hat{f}_N(t) + o_p(N^{-1/2}).$$
(6.251)

Moreover, if $t^* = t^{**} = V@R_{1-\alpha}(Z)$, i.e., the left and right side α -quantiles of the distribution of Z are the same, then

$$N^{1/2}\left(\hat{\theta}_N - \theta^*\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$
 (6.252)

where $\sigma^2 = (1 - \alpha)^{-2} \mathbb{V} \operatorname{ar} ([Z - t^*]_+).$

The estimator $\hat{\theta}_N$ has a negative bias, i.e., $\mathbb{E}[\hat{\theta}_N] - \theta^* \leq 0$, and (see Proposition 5.6)

$$\mathbb{E}[\hat{\theta}_N] \le \mathbb{E}[\hat{\theta}_{N+1}], \quad N = 1, \dots, \tag{6.253}$$

i.e., the bias is monotonically decreasing with increase of the sample size N. If $t^* = t^{**}$, then this bias is of order $O(N^{-1})$ and can be estimated using results of section 5.1.3. The first and second order derivatives of the expectation function f(t) are

$$f'(t) = 1 + (1 - \alpha)^{-1} (H_Z(t) - 1),$$

provided that the cumulative distribution function $H_Z(\cdot)$ is continuous at t, and $f''(t) = (1 - \alpha)^{-1}h_Z(t)$, provided that the density $h_Z(t) = \partial H_Z(t)/\partial t$ exists. We obtain (see Theorem 5.8 and the discussion on page 188), under appropriate regularity conditions, in particular if $t^* = t^{**} = V@R_{1-\alpha}(Z)$ and the density $h_Z(t^*) = \partial H_Z(t^*)/\partial t$ exists and $h_Z(t^*) \neq 0$, that

$$\hat{\theta}_N - \hat{f}_N(t^*) = N^{-1} \inf_{\tau \in \mathbb{R}} \left\{ \tau V + \frac{1}{2} \tau^2 f''(t^*) \right\} + o_p(N^{-1})$$
$$= -\frac{(1-\alpha)V^2}{2Nh_Z(t^*)} + o_p(N^{-1}), \tag{6.254}$$

where $V \sim \mathcal{N}(0,\gamma^2)$ with

$$\gamma^{2} = \mathbb{V}\mathrm{ar}\left((1-\alpha)^{-1}\frac{\partial[Z-t^{*}]_{+}}{\partial t}\right) = \frac{H_{Z}(t^{*})(1-H_{Z}(t^{*}))}{(1-\alpha)^{2}} = \frac{\alpha}{1-\alpha}.$$

Consequently, under appropriate regularity conditions,

$$N\left[\hat{\theta}_N - \hat{f}_N(t^*)\right] \xrightarrow{\mathcal{D}} - \left[\frac{\alpha}{2h_Z(t^*)}\right] \chi_1^2$$
(6.255)

and (see Remark 57 on page 470)

$$\mathbb{E}[\hat{\theta}_N] - \theta^* = -\frac{\alpha}{2Nh_Z(t^*)} + o(N^{-1}).$$
(6.256)

Remark 40. The set of optimal solutions of the sample optimization problem (6.247) is given by the set of α -quantiles of the empirical cdf of the sample Z^1, \ldots, Z^N . That is,

such an optimal solution is the (left side) quantile $Z^{(k)}$, where $Z^{(1)} \leq \cdots \leq Z^{(N)}$ are sample values arranged in the increasing order and²¹ $k := \lceil \alpha N \rceil$. For small $1 - \alpha > 0$ the sample qauntile $Z^{(k)}$ can be very unstable especially if Z has a distribution with unbounded support. In particular if $N < (1 - \alpha)^{-1}$, then $\lceil \alpha N \rceil = N$, i.e., the corresponding sample estimator is given by the maximum $\max_{1 \leq j \leq N} Z^j$. Therefore the asymptotics (6.251)– (6.252) and (6.254)–(6.256) could give a reasonable approximation when the sample size N is significantly bigger than $(1 - \alpha)^{-1}$.

It is also possible to give large deviations type bounds for convergence of $\hat{\theta}_N$ to θ^* . Let (ℓ, u) be an interval containing the interval $[t^*, t^{**}]$, i.e., $\ell < t^*$ and $u > t^{**}$, where t^* and t^{**} are defined in (6.249). Recall that $[t^*, t^{**}]$ is the set of optimal solutions of the problem (6.245). By convexity arguments we have that the set of optimal solutions of the corresponding SAA problem (6.247) is contained in the interval (ℓ, u) if $|\hat{f}_N(\ell) - f(\ell)| < \kappa$, $|\hat{f}_N(u) - f(u)| < \kappa$ and $|\hat{f}_N(t^*) - f(t^*)| < \kappa$, where

$$\kappa := \frac{1}{2} \min \left\{ f(\ell) - f(t^*), f(u) - f(t^*) \right\}.$$
(6.257)

Indeed, κ is a positive number and the right side derivative of $\hat{f}_N(t)$ at $t = \ell$ is

$$\inf_{\delta>0} \frac{\hat{f}_N(\ell+\delta) - \hat{f}_N(\ell)}{\delta} \le \frac{f(t^*) + \kappa - (f(\ell) - \kappa)}{t^* - \kappa} < 0.$$

And similarly the left side derivative at t = u is positive.

We can now proceed as in section 5.3.2. Consider random variable

$$Y_t := [Z - t]_+ - \mathbb{E}[Z - t]_+,$$

and its moment generating function $M_t(\tau) = \mathbb{E}[\exp(\tau Y_t)]$. Let us make the following assumption (compare with condition (M4) of section 5.3.2).

(\mathfrak{E}) There are constants $\sigma > 0$ and a > 0 such that for any $t \in [\ell, u]$ the following inequality holds

$$M_t(\tau) \le \exp(\sigma^2 \tau^2/2), \ \ \tau \in [-a, a].$$
 (6.258)

Note that $F_{\alpha}(t, Z) - f(t) = (1 - \alpha)^{-1}Y_t$, and the moment generating function of this random variable is

$$M_t^*(\tau) = \mathbb{E}\left[\exp\{\tau(1-\alpha)^{-1}Y_t\}\right] = M_t\left((1-\alpha)^{-1}\tau\right).$$

Hence condition (6.258) is the same as

$$M_t^*(\tau) \le \exp\{\bar{\sigma}^2 \tau^2/2\}, \ \tau \in [-a, a],$$
 (6.259)

where $\bar{\sigma}^2 := (1 - \alpha)^{-2} \sigma^2$. Condition (6.259) implies the following estimate for the corresponding LD rate function

$$\begin{split} I(z) &:= \sup_{\tau \in \mathbb{R}} \{\tau z - \log M_t^*(\tau)\} \ge \sup_{-a \le \tau \le a} \{\tau z - \bar{\sigma}^2 \tau^2 / 2\} \\ &= \begin{cases} z^2 / (2\bar{\sigma}^2), & \text{if } |z| \le a\bar{\sigma}^2, \\ a|z| - a^2 \bar{\sigma}^2 / 2, & \text{if } |z| \ge a\bar{\sigma}^2. \end{cases} \end{split}$$

²¹Here [a] denotes the smallest integer greater than or equal to $a \in \mathbb{R}$.

We also have that

$$\left|F_{\alpha}(t',Z) - F_{\alpha}(t,Z)\right| = (1-\alpha)^{-1} \left|[Z-t']_{+} - [Z-t]_{+}\right| \le (1-\alpha)^{-1} |t'-t|.$$

Consequently for $0 < \varepsilon \leq a\bar{\sigma}^2$ we have the following estimate (see Theorem 7.75 and Remark 16 on page 208)

$$\Pr\left(\sup_{\ell \le t \le u} \left| \hat{f}_N(t) - f(t) \right| \ge \varepsilon\right) \le \frac{8(u-\ell)}{(1-\alpha)\varepsilon} \exp\left[-\frac{N\varepsilon^2 (1-\alpha)^2}{32\sigma^2} \right].$$
 (6.260)

We obtain the following result.

Proposition 6.63. Suppose that condition (\mathfrak{E}) is satisfied, and let $\ell < t^*$, $u > t^{**}$ and κ be the constant defined in (6.257). Then for $0 < \varepsilon < \min{\{\kappa, (1 - \alpha)^{-2}a\sigma^2\}}$ it follows that

$$\Pr\left(\left|\hat{\theta}_N - \theta^*\right| \ge \varepsilon\right) \le \frac{8(u-\ell)}{(1-\alpha)\varepsilon} \exp\left[-\frac{N\varepsilon^2(1-\alpha)^2}{32\sigma^2}\right].$$
(6.261)

For confidence level $\gamma \in (0, 1)$ and sample size

$$N \ge \ln\left[\frac{8(u-\ell)}{(1-\alpha)\varepsilon\gamma}\right]\frac{32\sigma^2}{(1-\alpha)^2\varepsilon^2},\tag{6.262}$$

it follows from (6.261) that $\Pr\left(\left|\hat{\theta}_N - \theta^*\right| \ge \varepsilon\right) \le \gamma$. Of course, the above estimate (6.262) of the sample size is very conservative. It also indicates that N should be significantly bigger than $(1 - \alpha)^{-1}$ for the sample estimate $\hat{\theta}_N$ to give a reasonably accurate estimate of θ^* .

Example 6.64 Suppose that α is close to one, and that $0 \le Z \le 1$ w.p.1. Then for $\theta^* = AV@R_{1-\alpha}(Z)$ we have that $0 \le \theta^* \le 1$. This follows immediately from the representation (6.27). Suppose that $N < (1 - \alpha)^{-1}$, in which case $\hat{\theta}_N = \max\{Z_1, \ldots, Z_N\}$. Suppose further that $\varepsilon \ge 1 - \theta^*$. Then

$$\Pr\left(\left|\hat{\theta}_{N}-\theta^{*}\right| \geq \varepsilon\right) = \Pr\left(\hat{\theta}_{N} \leq \theta^{*}-\varepsilon\right) = \Pr\left(Z_{i} \leq \theta^{*}-\varepsilon, \ i=1,\ldots,N\right)$$
$$= [H(\theta^{*}-\varepsilon)]^{N} \leq [H(1-\varepsilon)]^{N},$$
(6.263)

where $H(\cdot)$ is the cdf of Z. That is, in this case in order for $\hat{\theta}_N$ to estimate θ^* with accuracy $\varepsilon > 0$ and confidence $1 - \gamma$ it suffices to use sample of size

$$N \ge \frac{\ln(\gamma^{-1})}{\ln[(H(1-\varepsilon))^{-1}]},$$
(6.264)

provided that $H(1 - \varepsilon) < 1$. That is, if for some c > 0 and small $\varepsilon > 0$ it holds that $H(1 - \varepsilon) \le 1 - c\varepsilon$, then for α sufficiently close to one and small $\varepsilon > 0$, the required sample size is of order $O(\varepsilon^{-1})$.

Example 6.65 Consider the exponential distribution $H(z) = 1 - e^{-z}$, $z \ge 0$. For t = 0 the moment generating function of $Y_0 = Z - \mathbb{E}[Z]$ is

$$M_0(\tau) = \begin{cases} (1-\tau)^{-1} - \tau, & \text{if } \tau < 1, \\ +\infty, & \text{if } \tau \ge 1. \end{cases}$$

We have here that $H^{-1}(\alpha) = \ln[(1-\alpha)^{-1}]$ and $\theta^* = 1 + \ln[(1-\alpha)^{-1}]$. Therefore for $N < (1-\alpha)^{-1}$ we have

$$\Pr\left(\left|\hat{\theta}_{N} - \theta^{*}\right| \geq \varepsilon\right) \geq \Pr\left(\hat{\theta}_{N} \geq 1 + \ln[(1 - \alpha)^{-1}] + \varepsilon\right) = \Pr\left(\max\{Z_{1}, \dots, Z_{N}\} \geq 1 + \ln[(1 - \alpha)^{-1}] + \varepsilon\right) = 1 - \Pr\left(Z_{i} \leq 1 + \ln[(1 - \alpha)^{-1}] + \varepsilon, \ i = 1, \dots, N\right) = 1 - [1 - \exp\{-1 + \ln(1 - \alpha) - \varepsilon\}]^{N} = 1 - [1 - e^{-1 - \varepsilon}(1 - \alpha)]^{N}.$$
(6.265)

Suppose now that $N = \lfloor (1 - \alpha)^{-1} \rfloor$. Since

$$\lim_{\alpha \uparrow 1} \left[1 - e^{-1 - \varepsilon} (1 - \alpha) \right]^{(1 - \alpha)^{-1}} = e^{-e^{-1 - \varepsilon}},$$

we have then that

$$\liminf_{\alpha\uparrow 1} \Pr\left(\hat{\theta}_N \ge \theta^* + \varepsilon\right) \ge 1 - e^{-e^{-1-\varepsilon}}.$$

That is, as α tends to one, and hence $N = \lfloor (1-\alpha)^{-1} \rfloor$ tends to ∞ , probability of the event " $|\hat{\theta}_N - \theta^*| \ge \varepsilon''$ does not tend to zero.

Similar analysis can be performed for a convex combination of Average Value-at-Risk measures. That is, let

$$\theta^* := \sum_{i=1}^m \lambda_i \mathsf{AV} @\mathsf{R}_{1-\alpha_i}(Z), \tag{6.266}$$

where $0 \le \alpha_1 < \cdots < \alpha_m < 1$ and $\lambda_i > 0$ are positive weights summing up to one. Let $\hat{\theta}_N$ be the sample estimate of θ^* based on an iid sample of size N. Then the representation of the form (6.248) holds with

$$f(t) := \sum_{i=1}^{m} \left\{ \lambda_i t_i + \frac{\lambda_i}{1-\alpha_i} \mathbb{E}[Z-t_i]_+ \right\},$$

$$\hat{f}_N(t) := \sum_{i=1}^{m} \left\{ \lambda_i t_i + \frac{\lambda_i}{(1-\alpha_i)N} \sum_{j=1}^{N} [Z^j - t_i]_+ \right\},$$

and the corresponding minimization over $t \in \mathbb{R}^m$. If $\mathbb{E}[Z^2] < +\infty$ and the left and right side α_i -quantiles, i = 1, ..., m, of the distribution of Z are the same, then the asymptotic convergence result (6.252) holds with

$$\sigma^2 = \operatorname{Var}\left\{\sum_{i=1}^m \frac{\lambda_i}{1-\alpha_i} \left[Z - t_i^*\right]_+\right\},\tag{6.267}$$

where $t_i^* := V @R_{1-\alpha_i}(Z) = H_Z^{-1}(\alpha_i).$

We can view the right hand side of (6.266) as the Kusuoka representation (6.131) of the corresponding spectral risk measure. That is, we can write equation (6.266) as

$$\theta^* = \int_0^1 \mathsf{AV} @\mathsf{R}_{1-\alpha}(Z) d\mu(\alpha), \tag{6.268}$$

where $\mu := \sum_{i=1}^{m} \lambda_i \delta(\alpha_i)$. For a general probability measure μ , on the interval [0, 1), an analogue of formula (6.267), for the corresponding asymptotic variance of $N^{1/2}(\hat{\theta}_N - \theta^*)$ will be

$$\sigma^{2} = \operatorname{Var}\left\{\int_{0}^{1} \frac{1}{1-\alpha} \left[Z - t^{*}(\alpha)\right]_{+} d\mu(\alpha)\right\},$$
(6.269)

where $t^*(\alpha) := V@R_{1-\alpha}(Z) = H_Z^{-1}(\alpha)$.

Of course, the above formula (6.269) should be rigorously justified. As it was pointed in Remark 40, sample estimate of AV@R_{1- α}(Z) could become very unstable for small values of 1 - α . Similarly the right hand side of (6.269) could become large if the distribution of Z has an unbounded support and the distribution $\mu(\alpha)$ has a heavy tail as $\alpha \to 1$. We will discuss this further at the end of section 6.6.3.

6.6.2 Absolute Semideviation Risk Measure

Consider the mean absolute semideviation risk measure

$$\rho_c(Z) := \mathbb{E} \{ Z + c[Z - \mathbb{E}(Z)]_+ \}, \qquad (6.270)$$

where $c \in [0,1]$ and the expectation is taken with respect to the probability distribution P of Z. We assume that $\mathbb{E}|Z| < +\infty$, and hence $\rho_c(Z)$ is finite. For a random sample Z^1, \ldots, Z^N of Z, the corresponding estimator of $\theta^* := \rho_c(Z)$ is

$$\hat{\theta}_N = N^{-1} \sum_{j=1}^N \left(Z^j + c [Z^j - \bar{Z}]_+ \right), \tag{6.271}$$

where $\bar{Z} = N^{-1} \sum_{j=1}^{N} Z^{j}$.

We have that $\rho_c(Z)$ is equal to the optimal value of the following convex-concave minimax problem

$$\min_{t \in \mathbb{R}} \max_{\gamma \in [0,1]} \mathbb{E}\left[F(t,\gamma,Z)\right],$$
(6.272)

where

$$F(t, \gamma, z) := z + c\gamma[z - t]_{+} + c(1 - \gamma)[t - z]_{+}$$

= z + c[z - t]_{+} + c(1 - \gamma)(t - z). (6.273)

This follows by virtue of Corollary 6.3. More directly we can argue as follows. Denote $\mu := \mathbb{E}[Z]$. We have that

$$\sup_{\gamma \in [0,1]} \mathbb{E} \{ Z + c\gamma [Z - t]_+ + c(1 - \gamma)[t - Z]_+ \}$$

= $\mathbb{E}[Z] + c \max \{ \mathbb{E}([Z - t]_+), \mathbb{E}([t - Z]_+) \}.$

Moreover, $\mathbb{E}([Z - t]_+) = \mathbb{E}([t - Z]_+)$ if $t = \mu$, and either $\mathbb{E}([Z - t]_+)$ or $\mathbb{E}([t - Z]_+)$ is bigger than $\mathbb{E}([Z - \mu]_+)$ if $t \neq \mu$. This implies the assertion and also shows that the

minimum in (6.272) is attained at unique point $t^* = \mu$. It also follows that the set of saddle points of the minimax problem (6.272) is given by $\{\mu\} \times [\gamma^*, \gamma^{**}]$, where

$$\gamma^* = \Pr(Z < \mu) \text{ and } \gamma^{**} = \Pr(Z \le \mu) = H_Z(\mu).$$
 (6.274)

In particular, if the cdf $H_Z(\cdot)$ is continuous at $\mu = \mathbb{E}[Z]$, then there is unique saddle point $(\mu, H_Z(\mu))$.

Consequently, $\hat{\theta}_N$ is equal to the optimal value of the corresponding SAA problem

$$\min_{t \in \mathbb{R}} \max_{\gamma \in [0,1]} N^{-1} \sum_{j=1}^{N} F(t, \gamma, Z^j).$$
(6.275)

Therefore we can apply results of section 5.1.4 in a straightforward way. We obtain that $\hat{\theta}_N$ converges w.p.1 to θ^* as $N \to \infty$. Moreover, assuming that $\mathbb{E}[Z^2] < +\infty$ we have by Theorem 5.10 that

$$\hat{\theta}_{N} = \max_{\gamma \in [\gamma^{*}, \gamma^{**}]} N^{-1} \sum_{j=1}^{N} F(\mu, \gamma, Z^{j}) + o_{p}(N^{-1/2})
= \bar{Z} + cN^{-1} \sum_{j=1}^{N} [Z^{j} - \mu]_{+} + c\Psi(\bar{Z} - \mu) + o_{p}(N^{-1/2}),$$
(6.276)

where $\bar{Z} = N^{-1} \sum_{j=1}^{N} Z^{j}$ and function $\Psi(\cdot)$ is defined as

$$\Psi(z) := \begin{cases} (1 - \gamma^*)z, & \text{if } z > 0, \\ (1 - \gamma^{**})z, & \text{if } z \le 0. \end{cases}$$

If, moreover, the cdf $H_Z(\cdot)$ is continuous at μ , and hence $\gamma^* = \gamma^{**} = H_Z(\mu)$, then

$$N^{1/2}(\hat{\theta}_N - \theta^*) \xrightarrow{\mathcal{D}} N(0, \sigma^2), \tag{6.277}$$

where $\sigma^2 = \mathbb{V}ar[F(\mu, H_Z(\mu), Z)].$

This analysis can be extended to risk averse optimization problems of the form (6.194). That is, consider problem

$$\min_{x \in \mathcal{X}} \Big\{ \rho_c[G(x,\xi)] = \mathbb{E} \Big\{ G(x,\xi) + c[G(x,\xi) - \mathbb{E}(G(x,\xi))]_+ \Big\} \Big\},$$
(6.278)

where $\mathcal{X} \subset \mathbb{R}^n$ and $G : \mathcal{X} \times \Xi \to \mathbb{R}$. Its sample average approximation (SAA) is obtained by replacing the true distribution of the random vector ξ with the empirical distribution associated with a random sample ξ^1, \ldots, ξ^N , that is

$$\min_{x \in \mathcal{X}} \frac{1}{N} \sum_{j=1}^{N} \left\{ G(x,\xi^j) + c \left[G(x,\xi^j) - \frac{1}{N} \sum_{j=1}^{N} G(x,\xi^j) \right]_+ \right\}.$$
(6.279)

Assume that the set \mathcal{X} is convex compact and function $G(\cdot, \xi)$ is convex for a.e. ξ . Then, for $c \in [0, 1]$, problems (6.278) and (6.279) are convex. By using the min-max representation (6.272), problem (6.278) can be written as the minimax problem

$$\underset{(x,t)\in\mathcal{X}\times\mathbb{R}}{\operatorname{Min}} \max_{\gamma\in[0,1]} \mathbb{E}\left[F(t,\gamma,G(x,\xi))\right],$$
(6.280)

where function $F(t, \gamma, z)$ is defined in (6.273). The function $F(t, \gamma, z)$ is convex and monotonically increasing in z. Therefore, by convexity of $G(\cdot, \xi)$, the function $F(t, \gamma, G(x, \xi))$ is convex in $x \in X$, and hence (6.280) is a convex-concave minimax problem. Consequently results of section 5.1.4 can be applied.

Let ϑ^* and ϑ_N be the optimal values of the true problem (6.278) and the SAA problem (6.279), respectively, and S be the set of optimal solutions of the true problem (6.278). By Theorem 5.10 and the above analysis we obtain, assuming that conditions specified in Theorem 5.10 are satisfied, that

$$\hat{\vartheta}_{N} = N^{-1} \inf_{\substack{x \in S \\ t = \mathbb{Z}[G(x,\xi)]}} \max_{\gamma \in [\gamma^{*}, \gamma^{**}]} \left\{ \sum_{j=1}^{N} F\left(t, \gamma, G(x,\xi^{j})\right) \right\} + o_{p}(N^{-1/2}), \quad (6.281)$$

where

 $\gamma^* := \mathsf{Pr}\big\{G(x,\xi) < \mathbb{E}[G(x,\xi)]\big\} \ \text{ and } \ \gamma^{**} := \mathsf{Pr}\big\{G(x,\xi) \leq \mathbb{E}[G(x,\xi)]\big\}, \ x \in S.$

Note that the points $((x, \mathbb{E}[G(x,\xi)]), \gamma)$, where $x \in S$ and $\gamma \in [\gamma^*, \gamma^{**}]$, form the set of saddle points of the convex-concave minimax problem (6.280), and hence the interval $[\gamma^*, \gamma^{**}]$ is the same for any $x \in S$.

Moreover, assume that $S = \{\bar{x}\}$ is a singleton, i.e., problem (6.278) has unique optimal solution \bar{x} , and the cdf of the random variable $Z = G(\bar{x},\xi)$ is continuous at $\mu := \mathbb{E}[G(\bar{x},\xi)]$, and hence $\gamma^* = \gamma^{**}$. Then it follows that $N^{1/2}(\hat{\vartheta}_N - \vartheta^*)$ converges in distribution to normal with zero mean and variance

$$\mathbb{V}ar\{G(\bar{x},\xi) + c[G(\bar{x},\xi) - \mu]_{+} + c(1 - \gamma^{*})(G(\bar{x},\xi) - \mu)\}$$

6.6.3 Von Mises Statistical Functionals

In the two examples, of AV@R_{α} and absolute semideviation, of risk measures considered in the above sections it was possible to use their variational representations in order to apply results and methods developed in section 5.1. A possible approach to deriving large sample asymptotics of law invariant coherent risk measures is to use the Kusuoka representation described in Theorem 6.40 (such approach was developed in [180]). In this section we discuss an alternative approach of *Von Mises statistical functionals* borrowed from Statistics. We view now a (law invariant) risk measure $\rho(Z)$ as a function $\theta(P)$ of the corresponding probability measure P. For example, with the (upper) semideviation risk measure $\sigma_p^+[Z]$, defined in (6.5), we associate the functional

$$\theta(P) := \left(\mathbb{E}_P \left[\left(Z - \mathbb{E}_P[Z] \right)_+^p \right] \right)^{1/p}.$$
(6.282)

The sample estimate of $\theta(P)$ is obtained by replacing probability measure P with the empirical measure P_N . That is, we estimate $\theta^* = \theta(P)$ by $\hat{\theta}_N = \theta(P_N)$.

Let Q be an arbitrary probability measure, defined on the same probability space as P, and consider the convex combination $(1 - \tau)P + \tau Q = P + \tau(Q - P)$, with $\tau \in [0, 1]$, of P and Q. Suppose that the following limit exists

$$\theta'(P,Q-P) := \lim_{\tau \downarrow 0} \frac{\theta(P + \tau(Q-P)) - \theta(P)}{\tau}.$$
(6.283)

The above limit is just the directional derivative of $\theta(\cdot)$ at P in the direction Q - P. If, moreover, the directional derivative $\theta'(P, \cdot)$ is linear, then $\theta(\cdot)$ is Gâteaux differentiable at P. Consider now the approximation

$$\theta(P_N) - \theta(P) \approx \theta'(P, P_N - P).$$
 (6.284)

By this approximation

$$N^{1/2}(\hat{\theta}_N - \theta^*) \approx \theta'(P, N^{1/2}(P_N - P)),$$
 (6.285)

and we can use $\theta'(P, N^{1/2}(P_N - P))$ to derive asymptotics of $N^{1/2}(\hat{\theta}_N - \theta^*)$.

Suppose, further, that $\theta'(P, \cdot)$ is linear, i.e., $\theta(\cdot)$ is Gâteaux differentiable at P. Then, since $P_N = N^{-1} \sum_{j=1}^N \delta(Z^j)$, we have that

$$\theta'(P, P_N - P) = \frac{1}{N} \sum_{j=1}^{N} IF_{\theta}(Z^j),$$
(6.286)

where

$$IF_{\theta}(z) := \theta'(P, \delta(z) - P) \tag{6.287}$$

is the so-called *influence function* (also called influence curve) of $\theta(\cdot)$.

It follows from the linearity of $\theta'(P, \cdot)$ that $\mathbb{E}_P[IF_{\theta}(Z)] = 0$. Indeed, linearity of $\theta'(P, \cdot)$ means that it is a linear functional and hence can be represented as

$$\theta'(P,Q-P) = \int g \, d(Q-P) = \int g \, dQ - \mathbb{E}_P[g(Z)]$$

for some function g in an appropriate functional space. Consequently $IF_{\theta}(z) = g(z) - \mathbb{E}_{P}[g(Z)]$, and hence

$$\mathbb{E}_P[IF_\theta(Z)] = \mathbb{E}_P\{g(Z) - \mathbb{E}_P[g(Z)]\} = 0$$

Then by the Central Limit Theorem we have that $N^{-1/2} \sum_{j=1}^{N} IF_{\theta}(Z^j)$ converges in distribution to normal with zero mean and variance $\mathbb{E}_P[IF_{\theta}(Z)^2]$. This suggests the following asymptotics

$$N^{1/2}(\hat{\theta}_N - \theta^*) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathbb{E}_P[IF_{\theta}(Z)^2]\right).$$
(6.288)

It should be mentioned at this point that the above derivations *do not* prove in a rigorous way validity of the asymptotics (6.288). The main technical difficulty is to give a rigorous justification for the approximation (6.285) leading to the corresponding convergence in distribution. This can be compared with the Delta method, discussed in section 7.2.8 and applied in section 5.1, where first (and second) order approximations were derived in functional spaces rather than spaces of measures. Anyway, formula (6.288) usually gives correct asymptotics and is routinely used in statistical applications.

Let us consider, for example, the statical functional

$$\theta(P) := \mathbb{E}_P \big[Z - \mathbb{E}_P [Z] \big]_+, \tag{6.289}$$

associated with $\sigma_1^+[Z]$. Denote $\mu := \mathbb{E}_P[Z]$. Then

$$\theta(P + \tau(Q - P)) = \tau \left(\mathbb{E}_Q [Z - \mu]_+ - \mathbb{E}_P [Z - \mu]_+ \right) \\ + \mathbb{E}_P [Z - \mu - \tau(\mathbb{E}_Q [Z] - \mu)]_+ + o(\tau).$$

Moreover, the right side derivative at $\tau = 0$ of the second term in the right hand side of the above equation is $(1 - H_Z(\mu))(\mathbb{E}_Q[Z] - \mu)$, provided that the cdf $H_Z(z)$ is continuous at $z = \mu$. It follows that if the cdf $H_Z(z)$ is continuous at $z = \mu$, then

$$\theta'(P,Q-P) = \mathbb{E}_Q[Z-\mu]_+ - \mathbb{E}_P[Z-\mu]_+ + (1-H_Z(\mu))(\mathbb{E}_Q[Z]-\mu),$$

and hence

$$IF_{\theta}(z) = [z - \mu]_{+} - \mathbb{E}_{P} [Z - \mu]_{+} + (1 - H_{Z}(\mu))(z - \mu).$$
(6.290)

It can be seen now that $\mathbb{E}_P[IF_{\theta}(Z)] = 0$ and

$$\mathbb{E}_{P}[IF_{\theta}(Z)^{2}] = \mathbb{V}\mathrm{ar}\{[Z-\mu]_{+} + (1-H_{Z}(\mu))(Z-\mu)\}$$

That is, the asymptotics (6.288) here are exactly the same as the ones derived in the previous section 6.6.2 (compare with (6.277)).

In a similar way it is possible to compute the influence function of the statistical functional defined in (6.282), associated with $\sigma_p^+[Z]$, for p > 1. For example, for p = 2 the corresponding influence function can be computed, provided that the cdf $H_Z(z)$ is continuous at $z = \mu$, as

$$IF_{\theta}(z) = \frac{1}{2\theta^*} \Big([z-\mu]_+^2 - \theta^{*2} + 2\kappa(1-H_Z(\mu))(z-\mu) \Big), \tag{6.291}$$

where $\theta^* := \theta(P) = (\mathbb{E}_P[Z - \mu]_+^2)^{1/2}$ and $\kappa := \mathbb{E}_P[Z - \mu]_+ = \frac{1}{2}\mathbb{E}_P|Z - \mu|.$

Now let us compute the influence function of a spectral risk measure, given in the form of Kusuoka representation,

$$\theta(P) := \int_0^1 \mathsf{AV} @\mathsf{R}_{1-\alpha}(P) d\mu(\alpha), \tag{6.292}$$

where μ is a probability measure on the interval [0,1) such that $\theta(P)$ is finite valued. Given a probability measure Q having finite first order moment, consider function

$$\psi_{\alpha}(\tau) := \mathsf{AV}@\mathsf{R}_{1-\alpha}(P + \tau(Q - P)).$$

We have that

$$\psi_{\alpha}(\tau) = \inf_{t \in \mathbb{R}} \left\{ t + \frac{1-\tau}{1-\alpha} \mathbb{E}_P[Z-t]_+ + \frac{\tau}{1-\alpha} \mathbb{E}_Q[Z-t]_+ \right\},\tag{6.293}$$

and for $\tau = 0$ the set of minimizers of the right hand side of (6.293) is given by the interval $[t^*(\alpha), t^{**}(\alpha)]$ with $t^*(\alpha)$ and $t^{**}(\alpha)$ being the respective left and right side α -quantiles

of the probability distribution P. It follows by Danskin Theorem (Theorem 7.25) that for $\alpha \in (0, 1)$, the right side derivative of $\psi_{\alpha}(\tau)$ at $\tau = 0$ is

$$G(\alpha) := \inf_{t \in [t^*(\alpha), t^{**}(\alpha)]} (1 - \alpha)^{-1} \left(\mathbb{E}_Q[Z - t]_+ - \mathbb{E}_P[Z - t]_+ \right), \tag{6.294}$$

Consequently, $\theta(\cdot)$ is directionally differentiable at P in the direction Q - P and

$$\theta'(P, Q - P) = \lim_{\tau \downarrow 0} \int_0^1 \frac{\psi_\alpha(\tau) - \psi_\alpha(0)}{\tau} \, d\mu(\alpha) = \int_0^1 G(\alpha) d\mu(\alpha), \tag{6.295}$$

provided the limit and integral operators can be interchanged. By the Lebesgue Dominated Convergence Theorem this interchangeability holds if there exists an μ -integrable function $c(\alpha)$ and $\eta > 0$ such that

$$|\psi_{\alpha}(\tau) - \psi_{\alpha}(0)| \le c(\alpha)\tau, \ \tau \in [0,\eta].$$
(6.296)

We obtain that

$$IF_{\theta}(z) = \int_{0}^{1} (1-\alpha)^{-1} [z-t^{*}(\alpha)]_{+} d\mu(\alpha) - \int_{0}^{1} (1-\alpha)^{-1} \mathbb{E}_{P} [Z-t^{*}(\alpha)]_{+} d\mu(\alpha),$$
(6.297)

provided that the following conditions are satisfied.

- (i) For μ-almost every α ∈ [0,1) the distribution P has unique α-quantile t*(α) = V@R_{1-α}(P), i.e., t*(α) = t**(α).
- (ii) The dominance condition (6.296) holds for $Q := P_N$.

Then it follows that $\mathbb{E}_P[IF_{\theta}(Z)] = 0$, $\mathbb{V}ar(IF_{\theta}(Z)) = \mathbb{E}_P[IF_{\theta}(Z)^2]$ and

$$\mathbb{E}_{P}[IF_{\theta}(Z)^{2}] = \mathbb{V}\mathrm{ar}\left\{\int_{0}^{1} (1-\alpha)^{-1} \left[Z - t^{*}(\alpha)\right]_{+} d\mu(\alpha)\right\}.$$
(6.298)

Note that (6.298) gives the same expression as formula (6.269) for the asymptotic variance of the corresponding sample estimate $\hat{\theta}_N$. The above derivations indicate necessity of the assumption (i) for asymptotic normality of the estimator $\hat{\theta}_N$.

6.7 The Problem of Moments

Due to the duality representation (6.39) of a coherent risk measure, the corresponding risk averse optimization problem (6.194) can be written as the minimax problem (6.197). So far, risk measures were defined on an appropriate functional space which in turn was dependent on a reference probability distribution. One can take an opposite point of view by defining a min-max problem of the form

$$\underset{x \in \mathcal{X}}{\min} \sup_{P \in \mathfrak{M}} \mathbb{E}_{P}[f(x,\omega)]$$
(6.299)

in a direct way, for a specified set \mathfrak{M} of probability measures on a measurable space (Ω, \mathcal{F}) . Note that we do not assume in this section existence of a reference measure P and do not work in a functional space of corresponding density functions. In fact, it will be essential here to consider discrete measures on the space (Ω, \mathcal{F}) . We denote by $\bar{\mathfrak{P}}$ the set of probability measures²² on (Ω, \mathcal{F}) . For a probability measure $P \in \bar{\mathfrak{P}}$, the expectation $\mathbb{E}_P[f(x, \omega)]$

 $^{^{22}} The set <math display="inline">\bar{\mathfrak{P}}$ of probability measures should be distinguished from the set \mathfrak{P} of probability density functions used before.

is given by the integral

$$\mathbb{E}_{P}[f(x,\omega)] = \int_{\Omega} f(x,\omega) dP(\omega).$$

The set \mathfrak{M} can be viewed as an *uncertainty set* for the underlying probability distribution. Minimax problems of the form (6.299) are referred to as *distributionally robust* stochastic programming problems. Of course, there are various ways to define the uncertainty set \mathfrak{M} . In some situations, it is reasonable to assume that we have a knowledge about certain moments of the corresponding probability distribution. That is, suppose that the set \mathfrak{M} is defined by moment constraints as follows

$$\int_{\Omega} \Psi(\omega) dP(\omega) \preceq_C b, \tag{6.300}$$

where $b \in \mathbb{R}^q$, $C \subset \mathbb{R}^q$ is a closed convex cone, $\Psi : \Omega \to \mathbb{R}^q$ is a measurable mapping and the integral (expectation) of $\Psi(\omega) = (\psi_1(\omega), ..., \psi_q(\omega))$ is taken componentwise. That is, the set \mathfrak{M} consists of probability measures P on (Ω, \mathcal{F}) such that the integrals $\int_{\Omega} \psi_i dP$, i = 1, ..., q, are well defined and the constraint (6.300) holds. Note that the condition that P is a *probability* measure, can be formulated explicitly as the constraint²³ $\int_{\Omega} dP = 1$ and $P \succeq 0$.

Recall that $a \leq_C b$ denotes partial order with respect to the cone C, i.e., $a \leq_C b$ means that $b - a \in C$ or equivalently $a \in b - C$. In particular if $C = \{0_q\}$ consists of the null vector $0_q \in \mathbb{R}^q$, then $a \leq_C b$ means that a = b. If $C = \{0_p\} \times \mathbb{R}^{q-p}_+$, then $a \leq_C b$ means that $a_i = b_i$, i = 1, ..., p, and $a_i \leq b_i$, i = p + 1, ..., q. Another interesting example is when Ψ maps Ω into the space of $m \times m$ symmetric matrices and C is given by the cone of $m \times m$ positive semidefinite matrices.

We assume that every finite subset of Ω is \mathcal{F} -measurable. This is a mild assumption. For example, if Ω is a metric space equipped with its Borel sigma algebra, then this is certainly holds true. We denote by $\bar{\mathfrak{P}}_m^*$ the set of probability measures on (Ω, \mathcal{F}) having a finite support of at most m points. That is, every measure $P \in \bar{\mathfrak{P}}_m^*$ can be represented in the form $P = \sum_{i=1}^m \alpha_i \delta(\omega_i)$, where α_i are nonnegative numbers summing up to one and $\delta(\omega)$ denotes measure of mass one at the point $\omega \in \Omega$. Similarly, we denote $\mathfrak{M}_m^* := \mathfrak{M} \cap \bar{\mathfrak{P}}_m^*$, i.e., \mathfrak{M}_m^* is the subset of \mathfrak{M} consisting of probability measures having finite support of at most m points. Note that the set \mathfrak{M} is convex while, for a fixed m, the set \mathfrak{M}_m^* is not necessarily convex. By Theorem 7.37, to any $P \in \mathfrak{M}$ corresponds a probability measure $Q \in \bar{\mathfrak{P}}$ with a finite support of at most q + 1 points such that $\mathbb{E}_P[\psi_i(\omega)] = \mathbb{E}_Q[\psi_i(\omega)]$, $i = 1, \ldots, q$. That is, if the set \mathfrak{M} is nonempty, then its subset \mathfrak{M}_{q+1}^* is also nonempty.

For a given $x \in \mathcal{X}$ and $\psi_0(\omega) := f(x, \omega)$, consider the corresponding problem of moments:

$$\max_{P \in \mathfrak{M}} \mathbb{E}_{P}[\psi_{0}(\omega)]. \tag{6.301}$$

Proposition 6.66. Problem (6.301) is equivalent to the problem

$$\max_{P \in \mathfrak{M}_{q+1}^*} \mathbb{E}_P[\psi_0(\omega)].$$
(6.302)

²³Recall that the notation " $P \succeq 0$ " means that P is a nonnegative (not necessarily probability) measure on (Ω, \mathcal{F}) .

The equivalence is in the sense that both problems have the same optimal value, and if problem (6.301) has an optimal solution, then it also has an optimal solution $P^* \in \mathfrak{M}_{q+1}^*$. In particular if q = 0, i.e., the set \mathfrak{M} consists of all probability measures on (Ω, \mathcal{F}) , then problem (6.301) is equivalent to the problem of maximization of $\psi_0(\omega)$ over $\omega \in \Omega$.

Proof. If the set \mathfrak{M} is empty, then its subset \mathfrak{M}_{q+1}^* is also empty, and hence both problems (6.301) and (6.302) have optimal value $+\infty$. So suppose that \mathfrak{M} is nonempty. Let $P \in \mathfrak{M}$, by Theorem 7.37 there exists $Q \in \mathfrak{M}_{q+2}^*$ such that $\mathbb{E}_P[\psi_0(\omega)] = \mathbb{E}_Q[\psi_0(\omega)]$. It follows that problem (6.301) is equivalent to the problem of maximization of $\mathbb{E}_P[\psi_0(\omega)]$ over $P \in \mathfrak{M}_{q+2}^*$, and if problem (6.301) has an optimal solution, then it has an optimal solution in the set \mathfrak{M}_{q+2}^* . The problem of maximization of $\mathbb{E}_P[\psi_0(\omega)]$ over $P \in \mathfrak{M}_{q+2}^*$ can be written as

$$\begin{aligned}
& \underset{\alpha \in \mathbb{R}^{m}_{+}}{\operatorname{Max}} \sum_{j=1}^{m} \alpha_{j} \psi_{0}(\omega_{j}) \\
& \text{subject to } \sum_{j=1}^{m} \alpha_{j} \psi_{i}(\omega_{j}) = s_{i}, \ i = 1, \dots, q, \\
& \sum_{j=1}^{m} \alpha_{j} = 1, \\
& s \in b - C,
\end{aligned}$$
(6.303)

where m := q+2. For fixed $\omega_1, \ldots, \omega_m \in \Omega$ and $s \in b-C$, the above problem (6.303) is a linear programming problem in α . Its feasible set is bounded, and provided that its feasible set is nonempty, its optimum is attained at an extreme point of its feasible set which has at most q + 1 nonzero components of α . Therefore it suffices to take the maximum over $P \in \mathfrak{M}^*_{q+1}$. \Box

The Lagrangian of problem (6.301) is

$$L(P,\lambda_0,\lambda) = \int_{\Omega} \psi_0(\omega) dP(\omega) + \lambda_0 \left(1 - \int_{\Omega} dP(\omega)\right) + \lambda^{\mathsf{T}} \left(\int_{\Omega} \Psi(\omega) dP(\omega) - b\right) \\ = \int_{\Omega} \left(\psi_0(\omega) - \lambda_0 + \lambda^{\mathsf{T}} \Psi(\omega)\right) dP(\omega) + \lambda_0 - \lambda^{\mathsf{T}} b.$$

We have that

$$\inf_{\lambda_0 \in \mathbb{R}, \lambda \in C^*} L(P, \lambda_0, \lambda) = \begin{cases} \int_{\Omega} \psi_0(\omega) dP(\omega), & \text{if } \int_{\Omega} dP = 1, \ \int_{\Omega} \Psi(\omega) dP(\omega) \in b - C, \\ -\infty, & \text{otherwise,} \end{cases}$$

where C^* is the polar of the cone C. Therefore problem (6.301) can be written in the following minimax form

$$\max_{P \succeq 0} \inf_{\lambda_0 \in \mathbb{R}, \lambda \in C^*} L(P, \lambda_0, \lambda).$$
(6.304)

The (Lagrangian) dual of the problem (6.301) is obtained by interchanging the *max* and *min* operators in (6.304). Now

$$\max_{P \succeq 0} L(P, \lambda_0, \lambda) = \begin{cases} \lambda_0 - \lambda^{\mathsf{T}} b & \text{if } \psi_0(\omega) - \lambda_0 + \lambda^{\mathsf{T}} \Psi(\omega) \le 0, \ \omega \in \Omega, \\ +\infty, & \text{otherwise.} \end{cases}$$

This can be verified by considering atomic measures $P = \alpha \delta(\omega), \alpha > 0, \omega \in \Omega$. Therefore the Lagrangian dual of problem (6.301) can be written as (we change here λ to $-\lambda$)

$$\begin{array}{ll}
& \underset{\lambda_0 \in \mathbb{R}, \ \lambda \in -C^*}{\operatorname{Min}} & \lambda_0 + \lambda^{\mathsf{T}} b \\
& \text{s.t.} & \psi_0(\omega) - \lambda^{\mathsf{T}} \Psi(\omega) \leq \lambda_0, \ \omega \in \Omega.
\end{array}$$
(6.305)

It is also possible to write the above problem (6.305) as the following min-max problem

$$\underset{\lambda \in -C^*}{\operatorname{Min}} \sup_{\omega \in \Omega} \left\{ \lambda^{\mathsf{T}} b + \psi_0(\omega) - \lambda^{\mathsf{T}} \Psi(\omega) \right\}.$$
(6.306)

In particular, if $C = \{0_p\} \times \mathbb{R}^{q-p}_+$, then $-C^* = \mathbb{R}^p \times \mathbb{R}^{q-p}_+$, and hence the dual problem takes the form

$$\begin{array}{ll}
& \underset{\lambda_{0} \in \mathbb{R}, \, \lambda \in \mathbb{R}^{p} \times \mathbb{R}^{q-p}_{+}}{\operatorname{Min}} & \lambda_{0} + \sum_{i=1}^{q} b_{i} \lambda_{i} \\
& \underset{s.t.}{\text{s.t.}} & \psi_{0}(\omega) - \sum_{i=1}^{q} \lambda_{i} \psi_{i}(\omega) \leq \lambda_{0}, \, \omega \in \Omega.
\end{array}$$
(6.307)

If moreover the set Ω is finite, then problem (6.301) and its dual (6.307) are linear programming problems. In that case there is no duality gap between these problems unless both are infeasible.

If the set Ω is infinite, then the dual problem (6.305) involves an infinite number of constraints and becomes a semi-infinite programming problem. In that case one needs to verify some regularity conditions in order to ensure the "no duality gap" property. Note that optimal value of the dual problem (6.305) is always greater than or equal to the optimal value of the corresponding problem of moments (6.301). Therefore if the dual problem (6.305) is unbounded from below, i.e., its optimal value is $-\infty$, then the optimal value of the problem (6.301) is also $-\infty$, i.e., the feasible set \mathfrak{M} of problem (6.301) is empty.

By the theory of conjugate duality (see Example 7.10 and Theorem 7.8 of section 7.1.3), we have the following result.

Proposition 6.67. *Suppose that dual problem* (6.305) *has a nonempty and bounded set of optimal solutions. Then there is no duality gap between problems* (6.301) *and* (6.305).

Another regularity condition ensuring the "no duality gap" property is the following.

Proposition 6.68. Let Ω be a compact metric space equipped with its Borel sigma algebra. Suppose that the set \mathfrak{M} is nonempty and functions $\psi_i(\cdot)$, $i = 1, \ldots, q$, and $\psi_0(\cdot)$ are continuous on Ω . Then there is no duality gap between problems (6.301) and (6.305), and problem (6.301) has an optimal solution.

Proof. We briefly outline the proof. By Proposition 6.66 it suffices to give a proof in the framework of the problem (6.303), which can be written in the form

$$\begin{array}{ll}
\underset{(\omega_1,\ldots,\omega_m,\alpha)\in\Upsilon}{\max} & \sum_{j=1}^m \alpha_j \psi_0(\omega_j) \\
\text{s.t.} & \sum_{j=1}^m \alpha_j \Psi(\omega_j) \in b - C,
\end{array}$$
(6.308)

where $\Upsilon := \Omega \times \cdots \times \Omega \times \Delta_m$ with $\Delta_m := \{\alpha \in \mathbb{R}^m_+ : \sum_{j=1}^m \alpha_j = 1\}$. The set Υ is compact and the feasible set of problem (6.308) is nonempty. It follows that the feasible set of problem (6.308) is compact, and hence this problem has an optimal solution. Moreover, the optimal value of problem (6.308), considered as a function of vector *b*, is lower semicontinuous. This can be shown in the same way as in the proof of Theorem 7.23. Then by Theorem 7.8 we obtain that there is no duality gap between problems (6.301) and (6.305). \Box

If for every $x \in \mathcal{X}$ and $\psi_0(\cdot) = f(x, \cdot)$ there is no duality gap between problems (6.301) and (6.305), then the corresponding min-max problem (6.299) is equivalent to the following semi-infinite programming problem

$$\begin{array}{ll}
& \underset{x \in \mathcal{X}, \, \lambda_0 \in \mathbb{R}, \, \lambda \in -C^*}{\operatorname{Min}} & \lambda_0 + \lambda^{\mathsf{T}} b \\ & \text{s.t.} & f(x, \omega) - \lambda^{\mathsf{T}} \Psi(\omega) \leq \lambda_0, \, \omega \in \Omega. \end{array}$$
(6.309)

We also can write problem (6.309) as the following min-max problem

$$\underset{x \in \mathcal{X}, \lambda \in -C^*}{\operatorname{Min}} \sup_{\omega \in \Omega} \left\{ \lambda^{\mathsf{T}} b + f(x, \omega) - \lambda^{\mathsf{T}} \Psi(\omega) \right\}.$$
(6.310)

Remark 41. Let \mathfrak{M} be the set of *all* probability measures on (Ω, \mathcal{F}) . Then by the above analysis we have that it suffices in problem (6.299) to take the maximum over measures of mass one, and hence problem (6.299) is equivalent to the following (deterministic) minimax problem

$$\underset{x \in \mathcal{X}}{\operatorname{Min}} \sup_{\omega \in \Omega} f(x, \omega). \tag{6.311}$$

Suppose now that Ω is a (nonempty) convex compact subset of \mathbb{R}^d equipped with its Borel sigma algebra. Then by Minkowski Theorem, the set Ω is equal to the convex hull of its extreme points. Recall that a point $e \in \Omega$ is said to be an *extreme* point of Ω if there do not exist points $e_1, e_2 \in \Omega$, different from e, such that e belongs to the interval $[e_1, e_2]$. In other words, e is an extreme point of Ω if whenever $e = te_1 + (1-t)e_2$ for some $e_1, e_2 \in \Omega$ and $t \in (0, 1)$, then $e_1 = e_2 = e$. We denote by $\text{Ext}(\Omega)$ the set of extreme points of Ω . If the functions $\psi_i : \mathbb{R}^d \to \mathbb{R}, i = 1, \ldots, q$, are affine and the function $\psi_0 : \Omega \to \mathbb{R}$ is convex, then for any $\lambda \in -C^*$ the function $\psi_0(\cdot) - \lambda^T \Psi(\cdot)$ is convex, and hence maximum in the problem (6.306) is attained at extreme points of Ω . For the corresponding problem (6.301) this suggests the following result, for which we give a direct proof.

Theorem 6.69. Suppose that the set $\Omega \subset \mathbb{R}^d$ is nonempty convex compact, the set \mathfrak{M} is nonempty, the functions $\psi_i : \mathbb{R}^d \to \mathbb{R}$, i = 1, ..., q, are affine, and the function $\psi_0 : \Omega \to \mathbb{R}$ is convex continuous. Then maximum in problem of moments (6.301) is attained at a probability measure of the form $P^* = \sum_{i=1}^{q+1} \alpha_i \delta(e_i)$, where $e_i \in \text{Ext}(\Omega)$ and $\alpha_i \in [0, 1]$, i = 1, ..., q + 1, with $\sum_{i=1}^{q+1} \alpha_i = 1$.

Proof. By adding a constant to function $\psi_i(\cdot)$, i = 1, ..., q, we can make it linear, while absorbing this constant in the right hand side of the corresponding constraint in (6.300).

Therefore without loss of generality we can assume that all functions $\psi_i(\cdot)$, $i = 1, \ldots, q$, are linear.

By Proposition 6.66 it suffices to perform the maximization over $P \in \mathfrak{M}_k^*$, where k = q + 1. Since the set \mathfrak{M} is nonempty, the set \mathfrak{M}_k^* is also nonempty. Moreover, since the set Ω is compact, the functions ψ_i , $i = 1, \ldots, q$, are linear and ψ_0 is continuous, it follows by compactness arguments that problem (6.301) has an optimal solution of the form $P^* = \sum_{i=1}^k \alpha_i \delta(\omega_i)$ for some $\omega_i \in \Omega$ and $\alpha_i \in [0, 1]$ such that $\sum_{i=1}^k \alpha_i = 1$. We need to show that the points ω_i can be chosen to be extreme points of the set Ω .

Suppose that one of the points ω_i , say ω_1 , is not an extreme point of Ω . Since (by Minkowski Theorem) Ω is equal to the convex hull of $\text{Ext}(\Omega)$, there exist points $e_1, \ldots, e_m \in \text{Ext}(\Omega)$ and $t_j \in (0, 1)$, with $\sum_{j=1}^m t_j = 1$, such that $\omega_1 = \sum_{j=1}^m t_j e_j$. Consider the probability measure

$$P' := \alpha_1 \sum_{j=1}^m t_j \delta(e_j) + \sum_{i=2}^k \alpha_i \delta(\omega_i).$$

Then using linearity of ψ_{ℓ} , $\ell = 1, \ldots, q$, we have that

$$\mathbb{E}_{P'}[\psi_{\ell}(\omega)] = \alpha_1 \sum_{j=1}^m t_j \psi_{\ell}(e_j) + \sum_{i=2}^k \alpha_i \psi_{\ell}(\omega_i)$$

= $\alpha_1 \psi_{\ell} \Big(\sum_{j=1}^m t_j e_j \Big) + \sum_{i=2}^k \alpha_i \psi_{\ell}(\omega_i)$
= $\alpha_1 \psi_{\ell}(\omega_1) + \sum_{i=2}^k \alpha_i \psi_{\ell}(\omega_i) = \mathbb{E}_{P^*}[\psi_{\ell}(\omega)].$

Therefore P' satisfies the feasibility constraints of problem (6.301) as well as P^* . Now by convexity of ψ_0 we can write

$$\mathbb{E}_{P'}[\psi_0(\omega)] = \alpha_1 \sum_{j=1}^m t_j \psi_0(e_j) + \sum_{i=2}^k \alpha_i \psi_0(\omega_i)$$

$$\geq \alpha_1 \psi_0 \left(\sum_{j=1}^m t_j e_j \right) + \sum_{i=2}^k \alpha_i \psi_0(\omega_i)$$

$$= \alpha_1 \psi_0(\omega_1) + \sum_{i=2}^k \alpha_i \psi_0(\omega_i) = \mathbb{E}_{P^*}[\psi_0(\omega)].$$

Thus P' is also an optimal solution of problem (6.301). It follows that there exists an optimal solution with a finite support of a set of extreme points. It remains to note that by Theorem 7.37 this support can be chosen to have no more than q + 1 points. \Box

6.8 Multistage Risk Averse Optimization

In this section we discuss an extension of risk averse optimization to a multistage setting. In order to simplify the presentation, we start our analysis with a process on a finite probability space, where evolution of the state of the system is represented by a scenario tree.

6.8.1 Scenario Tree Formulation

Consider a scenario tree representation of evolution of the corresponding data process (see section 3.1.3). The basic idea of multistage stochastic programming is that if we are currently at a state of the system at stage t, represented by a node of the scenario tree, then our decision at that node is based on our knowledge about the distribution of the next possible realizations of the data process, which are represented by its children nodes at stage t + 1.

1

In the risk neutral approach, we optimize the corresponding conditional expectation of the objective function. This allows to write the associated dynamic programming equations. This idea can be extended to optimization of a risk measure, conditional on a current state of the system. We are going to discuss now such construction in detail.

As in section 3.1.4, we denote by Ω_t the set of all nodes at stage $t = 1, \ldots, T$, by $K_t := |\Omega_t|$ the cardinality of Ω_t and by C_a the set of children nodes of a node a of the tree. Note that $\{C_a\}_{a \in \Omega_t}$ forms a partition of the set Ω_{t+1} , i.e., $C_a \cap C_{a'} = \emptyset$ if $a \neq a'$ and $\Omega_{t+1} = \bigcup_{a \in \Omega_t} C_a$, $t = 1, \ldots, T-1$. With the set Ω_T we associate the algebra \mathcal{F}_T of all its subsets. Let \mathcal{F}_{T-1} be the subalgebra of \mathcal{F}_T generated by sets C_a , $a \in \Omega_{T-1}$, i.e., these sets form the set of elementary events of \mathcal{F}_{T-1} (recall that $\{C_a\}_{a \in \Omega_{T-1}}$ forms a partition of Ω_T). By this construction there is a one-to-one correspondence between elementary events of \mathcal{F}_{T-1} and the set Ω_{T-1} of nodes at stage T-1. By continuing this process we construct a sequence of sigma algebras $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T$ (called *filtration*). Note that \mathcal{F}_1 corresponds to the unique root node and hence $\mathcal{F}_1 = \{\emptyset, \Omega_T\}$. In this construction there is a one-to-one correspondence between of \mathcal{F}_t , and hence we can identify every node $a \in \Omega_t$ with an elementary event of \mathcal{F}_t . By taking all children of every node of C_a at later stages, we eventually can identify with C_a a subset of Ω_T .

Suppose, further, that there is a probability distribution defined on the scenario tree. As it was discussed in section 3.1.3, such probability distribution can be defined by introducing conditional probabilities of going from a node of the tree to its children nodes. That is, with a node $a \in \Omega_t$ is associated a probability vector²⁴ $p^a \in \mathbb{R}^{|C_a|}$ of conditional probabilities of moving from a to nodes in C_a . Equipped with probability vector p^a , the set C_a becomes a probability space, with the corresponding sigma algebra of all subsets of C_a , and any function $Z : C_a \to \mathbb{R}$ can be viewed as a random variable. Since the space of functions $Z : C_a \to \mathbb{R}$ can be identified with the space $\mathbb{R}^{|C_a|}$, we identify such random variable Z with an element of the vector space $\mathbb{R}^{|C_a|}$. With every $Z \in \mathbb{R}^{|C_a|}$ is associated the expectation $\mathbb{E}_{p^a}[Z]$, which can be considered as a conditional expectation given that we are currently at node a.

Now with every node a at stage t = 1, ..., T - 1 we associate a risk measure $\rho^a(Z)$ defined on the space of functions $Z : C_a \to \mathbb{R}$, that is we choose a family of risk measures

$$\rho^a : \mathbb{R}^{|C_a|} \to \mathbb{R}, \ a \in \Omega_t, \ t = 1, \dots, T - 1.$$
(6.312)

Of course, there are many ways how such risk measures can be defined. For instance, for a given probability distribution on the scenario tree, we can use conditional expectations

$$p^{a}(Z) := \mathbb{E}_{p^{a}}[Z], \ a \in \Omega_{t}, \ t = 1, \dots, T - 1.$$
 (6.313)

Such choice of risk measures ρ^a leads to the risk neutral formulation of a corresponding multistage stochastic program. For a risk averse approach we can use any class of coherent risk measures discussed in section 6.3.2, as, for example,

$$\rho^{a}[Z] := \inf_{t \in \mathbb{R}} \left\{ t + \lambda_{a}^{-1} \mathbb{E}_{p^{a}} \left[Z - t \right]_{+} \right\}, \ \lambda_{a} \in (0, 1),$$
(6.314)

²⁴A vector $p = (p_1, \ldots, p_n) \in \mathbb{R}^n$ is said to be a *probability vector* if all its components p_i are nonnegative and $\sum_{i=1}^n p_i = 1$. If $Z = (Z_1, \ldots, Z_n) \in \mathbb{R}^n$ is viewed as a random variable, then its expectation with respect to p is $\mathbb{E}_p[Z] = \sum_{i=1}^n p_i Z_i$.

corresponding to AV@R risk measure, or

$$\rho^{a}[Z] := \mathbb{E}_{p^{a}}[Z] + c_{a}\mathbb{E}_{p^{a}}[Z - \mathbb{E}_{p^{a}}[Z]]_{+}, \ c_{a} \in [0, 1],$$
(6.315)

corresponding to the absolute semideviation risk measure. Another interesting example is the max-risk measure

$$\rho^a[Z] := \max_{\omega \in C_a} Z(\omega). \tag{6.316}$$

This risk measure corresponds to $AV@R_0$ (see Remark 24 on page 296).

Since Ω_{t+1} is the union of the disjoint sets C_a , $a \in \Omega_t$, we can write $\mathbb{R}^{K_{t+1}}$ as the Cartesian product of the spaces $\mathbb{R}^{|C_a|}$, $a \in \Omega_t$. That is, $\mathbb{R}^{K_{t+1}} = \mathbb{R}^{|C_{a_1}|} \times \cdots \times \mathbb{R}^{|C_{a_{K_t}}|}$, where $\{a_1, \ldots, a_{K_t}\} = \Omega_t$. Define the mappings

$$\rho_{t+1} := (\rho^{a_1}, \dots, \rho^{a_{K_t}}) : \mathbb{R}^{K_{t+1}} \to \mathbb{R}^{K_t}, \ t = 1, \dots, T-1,$$
(6.317)

associated with risk measures ρ^a . Recall that the set Ω_{t+1} of nodes at stage t+1 is identified with the set of elementary events of sigma algebra \mathcal{F}_{t+1} , and its sigma subalgebra \mathcal{F}_t is generated by sets C_a , $a \in \Omega_t$.

We denote by Z_T the space of all functions $Z : \Omega_T \to \mathbb{R}$. As it was mentioned before we can identify every such function with a vector of the space \mathbb{R}^{K_T} , i.e., the space Z_T can be identified with the space \mathbb{R}^{K_T} . We have that a function $Z : \Omega_T \to \mathbb{R}$ is \mathcal{F}_{T-1} -measurable iff it is constant on every set C_a , $a \in \Omega_{T-1}$. We denote by Z_{T-1} the subspace of Z_T formed by \mathcal{F}_{T-1} -measurable functions. The space Z_{T-1} can be identified with $\mathbb{R}^{K_{T-1}}$. And so on, we can construct a sequence Z_t , $t = 1, \ldots, T$, of spaces of \mathcal{F}_t measurable functions $Z : \Omega_T \to \mathbb{R}$ such that $Z_1 \subset \cdots \subset Z_T$ and each Z_t can be identified with the space \mathbb{R}^{K_t} . Recall that $K_1 = 1$, and hence Z_1 can be identified with \mathbb{R} . We view the mapping ρ_{t+1} , defined in (6.317), as a mapping from the space Z_{t+1} into the space Z_t . Conversely, with any mapping $\rho_{t+1} : Z_{t+1} \to Z_t$ we can associate a family of risk measures of the form (6.312).

For a mapping $\rho_{t+1} : \mathbb{Z}_{t+1} \to \mathbb{Z}_t$ consider the following conditions²⁵:

 $(\mathbf{R}'\mathbf{1})$ Convexity:

$$\rho_{t+1}(\alpha Z + (1-\alpha)Z') \preceq \alpha \rho_{t+1}(Z) + (1-\alpha)\rho_{t+1}(Z')$$

for any $Z, Z' \in \mathcal{Z}_{t+1}$ and $\alpha \in [0, 1]$.

- (**R'2**) Monotonicity: If $Z, Z' \in \mathcal{Z}_{t+1}$ and $Z \succeq Z'$, then $\rho_{t+1}(Z) \succeq \rho_{t+1}(Z')$.
- (**R'3**) Translation Equivariance: If $Y \in \mathcal{Z}_t$ and $Z \in \mathcal{Z}_{t+1}$, then $\rho_{t+1}(Z+Y) = \rho_{t+1}(Z) + Y$.
- (**R'4**) Positive homogeneity: If $\alpha \ge 0$ and $Z \in \mathcal{Z}_{t+1}$, then $\rho_{t+1}(\alpha Z) = \alpha \rho_{t+1}(Z)$.

We say that ρ_{t+1} is a *coherent (convex) conditional risk mapping* if it satisfies conditions (R'1)–(R'4) (conditions (R'1)–(R'3)). It is straightforward to see that conditions (R'1), (R'2) and (R'4) hold iff the corresponding conditions (R1), (R2) and (R4), defined in section 6.3, hold for every risk measure ρ^a associated with ρ_{t+1} . Also by construction of ρ_{t+1} ,

²⁵For $Z_1, Z_2 \in \mathcal{Z}_t$ the inequality $Z_2 \succeq Z_1$ is understood componentwise.

we have that condition (R'3) holds iff condition (R3) holds for all ρ^a . That is, ρ_{t+1} is a *coherent (convex) conditional risk mapping* iff every corresponding risk measure ρ^a is a *coherent (convex) risk measure*. In the sequel we will deal mainly with coherent conditional risk mappings. Therefore, unless stated otherwise, we simply say that ρ_{t+1} is a *conditional risk mapping* if it satisfies conditions (R'1)–(R'4).

By Theorem 6.5 with each coherent risk measure ρ^a , $a \in \Omega_t$, is associated a set $\mathfrak{A}(a)$ of probability measures (vectors) such that

$$\rho^a(Z) = \max_{p \in \mathfrak{A}(a)} \mathbb{E}_p[Z].$$
(6.318)

Here $Z \in \mathbb{R}^{K_{t+1}}$ is a vector corresponding to function $Z : \Omega_{t+1} \to \mathbb{R}$, and $\mathfrak{A}(a) = \mathfrak{A}_{t+1}(a)$ is a closed convex set of probability vectors $p \in \mathbb{R}^{K_{t+1}}$ such that $p_k = 0$ if $k \in \Omega_{t+1} \setminus C_a$, i.e., all probability measures of $\mathfrak{A}_{t+1}(a)$ are supported on the set C_a . We can now represent the corresponding conditional risk mapping ρ_{t+1} as a maximum of conditional expectations as follows. Let $\nu = (\nu_a)_{a \in \Omega_t}$ be a probability distribution on Ω_t , assigning *positive* probability ν_a to every $a \in \Omega_t$, and define

$$\mathfrak{V}_{t+1} := \left\{ \mu = \sum_{a \in \Omega_t} \nu_a p^a : p^a \in \mathfrak{A}_{t+1}(a) \right\}.$$
(6.319)

It is not difficult to see that $\mathfrak{V}_{t+1} \subset \mathbb{R}^{K_{t+1}}$ is a convex set of probability vectors. Moreover, since each $\mathfrak{A}_{t+1}(a)$ is compact, the set \mathfrak{V}_{t+1} is also compact and hence is closed. Consider a probability distribution (measure) $\mu = \sum_{a \in \Omega_t} \nu_a p^a \in \mathfrak{V}_{t+1}$. We have that, for $a \in \Omega_t$, the corresponding conditional distribution given the event C_a is p^a , and²⁶

$$\mathbb{E}_{\mu}\left[Z|\mathcal{F}_{t}\right](a) = \mathbb{E}_{p^{a}}[Z], \quad Z \in \mathcal{Z}_{t+1}.$$
(6.320)

It follows then by (6.318) that

$$\rho_{t+1}(Z) = \max_{\mu \in \mathfrak{V}_{t+1}} \mathbb{E}_{\mu}\left[Z|\mathcal{F}_t\right],\tag{6.321}$$

where the maximum on the right hand side of (6.321) is taken pointwise in $a \in \Omega_t$. That is, formula (6.321) means that

$$[\rho_{t+1}(Z)](a) = \max_{p \in \mathfrak{A}_{t+1}(a)} \mathbb{E}_p[Z], \quad Z \in \mathcal{Z}_{t+1}, \ a \in \Omega_t.$$
(6.322)

Note that, in this construction, choice of the distribution ν is arbitrary and any distribution of \mathfrak{V}_{t+1} agrees with the distribution ν on Ω_t (see Proposition 6.74 of section 6.8.2 for a further discussion of such max-representations of conditional risk mappings).

We are ready now to formulate a risk averse multistage programming problem. For a sequence $\rho_{t+1} : \mathcal{Z}_{t+1} \to \mathcal{Z}_t, t = 1, ..., T - 1$, of conditional risk mappings consider the following risk averse formulation analogous to the nested risk neutral formulation (3.1):

$$\underset{x_{1}\in\mathcal{X}_{1}}{\min} f_{1}(x_{1}) + \rho_{2} \Big| \inf_{x_{2}\in\mathcal{X}_{2}(x_{1})} f_{2}(x_{2}) + \cdots \\
+ \rho_{T-1} \Big[\inf_{x_{T-1}\in\mathcal{X}_{T}(x_{T-2})} f_{T-1}(x_{T-1}) + \rho_{T} [\inf_{x_{T}\in\mathcal{X}_{T}(x_{T-1})} f_{T}(x_{T})] \Big] \Big].$$
(6.323)

²⁶Recall that the conditional expectation $\mathbb{E}_{\mu}[\cdot | \mathcal{F}_t]$ is a mapping from \mathcal{Z}_{t+1} into \mathcal{Z}_t .

Here $\Omega := \Omega_T$, the objective functions $f_t : \mathbb{R}^{n_{t-1}} \times \Omega \to \mathbb{R}$ are real valued random functions, and $\mathcal{X}_t : \mathbb{R}^{n_{t-1}} \times \Omega \rightrightarrows \mathbb{R}^{n_t}$, t = 2, ..., T, are multifunctions such that $f_t(x_t, \cdot)$ and $\mathcal{X}_t(x_{t-1}, \cdot)$ are \mathcal{F}_t -measurable for all x_t and x_{t-1} . We use here the notation

$$\inf_{x_t \in \mathcal{X}_t(x_{t-1})} f_t(x_t) := \inf_{x_t(\omega) \in \mathcal{X}_t(x_{t-1}(\omega), \omega)} f_t(x_t(\omega), \omega), \ \omega \in \Omega,$$
(6.324)

to denote the corresponding random variables.

Note that if the corresponding risk measures ρ^a are defined as conditional expectations (6.313), then the above multistage problem (6.323) coincides with the risk neutral multistage problem (3.1).

There are several ways how the above nested formulation (6.323) can be formalized. Similarly to (3.4), we can write problem (6.323) in the form

$$\underset{x_{1},\boldsymbol{x}_{2},...,\boldsymbol{x}_{T}}{\min} f_{1}(x_{1}) + \rho_{2} \Big[f_{2}(\boldsymbol{x}_{2}) + \dots + \rho_{T-1} \big[f_{T-1} \left(\boldsymbol{x}_{T-1} \right) + \rho_{T} [f_{T} \left(\boldsymbol{x}_{T} \right)] \big] \Big]$$
s.t. $x_{1} \in \mathcal{X}_{1}, \ \boldsymbol{x}_{t}(\omega) \in \mathcal{X}_{t}(\boldsymbol{x}_{t-1}(\omega), \omega), \ t = 2, \dots, T.$

$$(6.325)$$

Optimization in (6.325) is performed over functions $x_t : \Omega \to \mathbb{R}$, t = 1, ..., T, satisfying the corresponding constraints, which imply that each x_t is \mathcal{F}_t -measurable and hence each $f_t(x_t)$ is \mathcal{F}_t -measurable. The requirement for x_t to be \mathcal{F}_t -measurable is another way of formulating the *nonanticipativity constraints*. Therefore, it can be viewed that the optimization in (6.325) is performed over feasible policies.

Consider the function $\varrho: \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_T \to \mathbb{R}$ defined as

$$\varrho(Z_1, \dots, Z_T) := Z_1 + \rho_2 \Big[Z_2 + \dots + \rho_{T-1} \big[Z_{T-1} + \rho_T [Z_T] \big] \Big].$$
(6.326)

By condition (R'3) we have that $Z_{T-1} + \rho_T[Z_T] = \rho_T[Z_{T-1} + Z_T]$ and hence

$$\rho_{T-1} [Z_{T-1} + \rho_T [Z_T]] = \rho_{T-1} \circ \rho_T [Z_{T-1} + Z_T].$$

By continuing this process we obtain that

$$\varrho(Z_1,\ldots,Z_T) = \bar{\rho}(Z_1 + \cdots + Z_T), \tag{6.327}$$

where $\bar{\rho} := \rho_2 \circ \cdots \circ \rho_T$. We refer to $\bar{\rho}$ as the *composite risk measure*. That is,

$$\bar{\rho}(Z_1 + \dots + Z_T) = Z_1 + \rho_2 \Big[Z_2 + \dots + \rho_{T-1} \big[Z_{T-1} + \rho_T [Z_T] \big] \Big], \tag{6.328}$$

defined for $Z_t \in \mathcal{Z}_t$, t = 1, ..., T. Recall that \mathcal{Z}_1 is identified with \mathbb{R} , and hence Z_1 is a real number and $\bar{\rho} : \mathcal{Z}_T \to \mathbb{R}$ is a real valued function. Conditions (R'1)–(R'4) imply that $\bar{\rho}$ is a coherent risk measure.

As above we have that since $f_{T-1}(\boldsymbol{x}_{T-1}(\omega), \omega)$ is \mathcal{F}_{T-1} -measurable, it follows by condition (R'3) that

$$f_{T-1}(\boldsymbol{x}_{T-1}(\omega),\omega) + \rho_T [f_T(\boldsymbol{x}_T)](\omega) = \rho_T [f_{T-1}(\boldsymbol{x}_{T-1}) + f_T(\boldsymbol{x}_T)](\omega).$$

Continuing this process backwards we obtain that the objective function of (6.325) can be formulated using the composite risk measure. That is, problem (6.325) can be written in the form

$$\begin{array}{ll} \underset{x_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T}{\operatorname{Min}} & \bar{\rho} \Big[f_1(x_1) + f_2(\boldsymbol{x}_2) + \dots + f_T(\boldsymbol{x}_T) \Big] \\ \text{s.t.} & x_1 \in \mathcal{X}_1, \ \boldsymbol{x}_t(\omega) \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\omega), \omega), \ t = 2, \dots, T. \end{array}$$
(6.329)

Remark 42. If the conditional risk mappings ρ^a are defined as the conditional expectations (6.313), then the composite risk measure $\bar{\rho}$ becomes the corresponding expectation operator, and (6.329) coincides with the risk neutral multistage program written in the form (3.4). Unfortunately it is not easy to write the composite risk measure $\bar{\rho}$ in a closed form even for relatively simple conditional risk mappings. Apart from conditional expectations, another example where the composite risk measure can be written explicitly is the max-risk measure (6.316). In that case the composite risk measure $\bar{\rho}$ also has the form of max-risk measure:

$$\bar{\varrho}(Z) = \max_{\omega \in \Omega_T} Z(\omega), \ Z \in \mathcal{Z}_T.$$
(6.330)

It turns out that the expectation and max-risk measures are the only examples, among coherent risk measures, for which the corresponding composite risk measure has a tractable representation. We will discuss this later.

An alternative approach to formalize the nested formulation (6.323) is to write dynamic programming equations. That is, for the last period T we have

$$Q_T(x_{T-1},\omega) := \inf_{x_T \in \mathcal{X}_T(x_{T-1},\omega)} f_T(x_T,\omega),$$
(6.331)

and we define

$$Q_T(x_{T-1}) := \rho_T[Q_T(x_{T-1})]. \tag{6.332}$$

Observe that $Q_T(x_{T-1})$ is random; it depends on the node in Ω_{T_1} . For $t = T - 1, \ldots, 2$, we recursively apply the conditional risk measures

$$Q_t(x_{t-1}) := \rho_t \left[Q_t(x_{t-1}) \right], \tag{6.333}$$

where

$$Q_t(x_{t-1},\omega) := \inf_{x_t \in \mathcal{X}_t(x_{t-1},\omega)} \Big\{ f_t(x_t,\omega) + \mathcal{Q}_{t+1}(x_t,\omega) \Big\}.$$
 (6.334)

Equations (6.333) and (6.334) can be combined into one equation:

$$Q_t(x_{t-1},\omega) = \inf_{x_t \in \mathcal{X}_t(x_{t-1},\omega)} \Big\{ f_t(x_t,\omega) + \rho_{t+1} \left[Q_{t+1}(x_t) \right](\omega) \Big\}.$$
 (6.335)

Finally, at the first stage we solve the problem

$$\min_{x_1 \in \mathcal{X}_1} f_1(x_1) + \rho_2[Q_2(x_1)].$$
(6.336)

A policy $\bar{x}_1, \ldots, \bar{x}_T$ is optimal iff \bar{x}_1 is an optimal solution of the first stage problem (6.336) and for $t = 2, \ldots, T$, it satisfies the optimality conditions (compare with (3.10)):

$$\bar{\boldsymbol{x}}_{t}(\omega) \in \operatorname*{argmin}_{x_{t} \in \mathcal{X}_{t}(\bar{\boldsymbol{x}}_{t-1}(\omega),\omega)} \left\{ f_{t}(x_{t},\omega) + \rho_{t+1}\left[Q_{t+1}(x_{t})\right](\omega) \right\}, \text{ w.p.1.}$$
(6.337)

It is important to emphasize that conditional risk mappings $\rho_t(Z)$ are defined on real valued functions $Z(\omega)$. Therefore it is implicitly assumed in the above equations that the cost-to-go (value) functions $Q_t(x_{t-1}, \omega)$ are real valued. In particular, this implies that the

considered problem should have relatively complete recourse. Also, in the development of the dynamic programming equations, the monotonicity condition (R'2) plays a crucial role, because only then we can use the interchangeability property of risk measures (Proposition 6.60) to move the optimization under the risk operation.

Remark 43. By using representation (6.321), we can write the dynamic programming equations (6.335) in the form

$$Q_t(x_{t-1},\omega) = \inf_{x_t \in \mathcal{X}_t(x_{t-1},\omega)} \Big\{ f_t(x_t,\omega) + \sup_{\mu \in \mathfrak{V}_{t+1}} \mathbb{E}_{\mu} \left[Q_{t+1}(x_t) \big| \mathcal{F}_t \right](\omega) \Big\}.$$
(6.338)

Note that the left and right hand side functions in (6.338) are \mathcal{F}_t -measurable, and hence this equation can be written in terms of $a \in \Omega_t$ instead of $\omega \in \Omega$. Recall that every $\mu \in \mathfrak{V}_{t+1}$ is representable in the form $\mu = \sum_{a \in \Omega_t} \nu_a p^a$ (see (6.319)) and that

$$\mathbb{E}_{\mu}\left[Q_{t+1}(x_t)\middle|\mathcal{F}_t\right](a) = \mathbb{E}_{p^a}[Q_{t+1}(x_t)], \ a \in \Omega_t.$$
(6.339)

We say that the problem is *convex* if the functions $f_t(\cdot, \omega)$, $Q_t(\cdot, \omega)$ and the sets $\mathcal{X}_t(x_{t-1}, \omega)$ are convex for every $\omega \in \Omega$ and $t = 1, \ldots, T$. If the problem is convex, then (since the set \mathfrak{V}_{t+1} is convex compact) the 'inf' and 'sup' operators on the right hand side of (6.338) can be interchanged to obtain a dual problem, and for a given x_{t-1} and every $a \in \Omega_t$ the dual problem has an optimal solution $\bar{p}^a \in \mathfrak{A}_{t+1}(a)$. Consequently, for $\bar{\mu}_{t+1} := \sum_{a \in \Omega_t} \nu_a \bar{p}^a$ an optimal solution of the original problem and the corresponding cost-to-go functions satisfy the following dynamic programming equations:

$$Q_t(x_{t-1},\omega) = \inf_{x_t \in \mathcal{X}_t(x_{t-1},\omega)} \Big\{ f_t(x_t,\omega) + \mathbb{E}_{\bar{\mu}_{t+1}} \big[Q_{t+1}(x_t) | \mathcal{F}_t \big](\omega) \Big\}.$$
 (6.340)

Moreover, it is possible to choose the "worst case" distributions $\bar{\mu}_{t+1}$ in consistent way, such that each $\bar{\mu}_{t+1}$ coincides with $\bar{\mu}_t$ on \mathcal{F}_t . Indeed, consider the first-stage problem (6.336). We have that (recall that at the first stage there is only one node, $\mathcal{F}_1 = \{\emptyset, \Omega\}$ and $\mathfrak{V}_2 = \mathfrak{A}_2$)

$$\rho_2[Q_2(x_1)] = \sup_{\mu \in \mathfrak{V}_2} \mathbb{E}_{\mu}[Q_2(x_1)|\mathcal{F}_1] = \sup_{\mu \in \mathfrak{V}_2} \mathbb{E}_{\mu}[Q_2(x_1)].$$
(6.341)

By convexity and compactness of \mathfrak{V}_2 and convexity of $Q_2(\cdot)$, a measure $\bar{\mu}_2 \in \mathfrak{V}_2$ exists (an optimal solution of the dual problem) such that the optimal value of the first stage problem is equal to the optimal value of the problem

$$\min_{x_1 \in \mathcal{X}_1} \mathbb{E}_{\bar{\mu}_2}[Q_2(x_1)].$$
(6.342)

Moreover, the set of optimal solutions of the first stage problem is contained in the set of optimal solutions of problem (6.342).

Let \bar{x}_1 be an optimal solution of the first stage problem. Then we can choose $\bar{\mu}_3 \in \mathfrak{V}_3$, of the form $\bar{\mu}_3 := \sum_{a \in \Omega_2} \nu_a \bar{p}^a$, such that equation (6.340) holds with t = 2 and $x_1 = \bar{x}_1$. Moreover, we can take the probability measure $\nu = (\nu_a)_{a \in \Omega_2}$ to be the same as $\bar{\mu}_2$, and hence to ensure that $\bar{\mu}_3$ coincides with $\bar{\mu}_2$ on \mathcal{F}_2 . Next, for every node $a \in \Omega_2$ choose a corresponding (second-stage) optimal solution and repeat the construction to produce an appropriate $\bar{\mu}_4 \in \mathfrak{V}_4$, and so on for later stages.

Proceeding in this way and assuming existence of optimal solutions, we can construct a probability distribution $\bar{\mu}_2, \ldots, \bar{\mu}_T$ on the scenario tree such that the obtained multistage problem of the standard form (3.1), has the same cost-to-go (value) functions as the original problem (6.323) and has an optimal solution which also is an optimal solution of the problem (6.323). In this sense, the expected-value multistage problem, with dynamic programming equations (6.340), is "almost equivalent" to the original problem.

Remark 44. Let us define, for every node $a \in \Omega_t$, t = 1, ..., T - 1, the corresponding set $\mathfrak{A}(a) = \mathfrak{A}_{t+1}(a)$ to be the set of *all* probability measures (vectors) on the set C_a (recall that $C_a \subset \Omega_{t+1}$ is the set of children nodes of a, and that all probability measures of $\mathfrak{A}_{t+1}(a)$ are supported on C_a). Then the maximum on the right hand side of (6.318) is attained at a measure of mass one at a point of the set C_a . Consequently, by (6.339) for such choice of the sets $\mathfrak{A}_{t+1}(a)$, the conditional risk measures ρ^a become the max-risk measures (6.316), and the dynamic programming equations (6.338) can be written as

$$Q_t(x_{t-1}, a) = \inf_{x_t \in \mathcal{X}_t(x_{t-1}, a)} \Big\{ f_t(x_t, a) + \max_{\omega \in C_a} Q_{t+1}(x_t, \omega) \Big\}, \ a \in \Omega_t.$$
(6.343)

It is interesting to note (see the above Remark 43) that if the problem is convex, then it is possible to construct a probability distribution (on the considered scenario tree), defined by a sequence $\bar{\mu}_t$, t = 2, ..., T, of consistent probability distributions, such that the obtained (risk neutral) multistage program is "almost equivalent" to the "min-max" formulation (6.343).

6.8.2 Conditional Risk Mappings

In this section we discuss a general concept of conditional risk mappings which can be applied to a risk averse formulation of multistage programs. The material of this section can be considered as an extension to an infinite dimensional setting of the developments presented in the previous section. Similarly to the presentation of coherent risk measures, given in section 6.3, we use the framework of \mathcal{L}_p spaces, $p \in [1, \infty)$. That is, let Ω be a sample space equipped with sigma algebras $\mathcal{F}_1 \subset \mathcal{F}_2$ (i.e., \mathcal{F}_1 is subalgebra of \mathcal{F}_2) and a probability measure P on (Ω, \mathcal{F}_2) . Consider the spaces $\mathcal{Z}_1 := \mathcal{L}_p(\Omega, \mathcal{F}_1, P)$ and $\mathcal{Z}_2 := \mathcal{L}_p(\Omega, \mathcal{F}_2, P)$. Since \mathcal{F}_1 is a subalgebra of \mathcal{F}_2 , it follows that $\mathcal{Z}_1 \subset \mathcal{Z}_2$.

For a mapping $\rho : \mathcal{Z}_2 \to \mathcal{Z}_1$ consider the following conditions (axioms):

 $(\mathbf{R}'\mathbf{1})$ Convexity:

$$\rho(\alpha Z + (1 - \alpha)Z') \preceq \alpha \rho(Z) + (1 - \alpha)\rho(Z'),$$

for any $Z, Z' \in \mathbb{Z}_2$ and $\alpha \in [0, 1]$.

- (**R'2**) Monotonicity: If $Z, Z' \in \mathcal{Z}_2$ and $Z \succeq Z'$, then $\rho(Z) \succeq \rho(Z')$.
- (**R'3**) Translation Equivariance: If $Y \in \mathcal{Z}_1$ and $Z \in \mathcal{Z}_2$, then $\rho(Z+Y) = \rho(Z) + Y$.
- (**R'4**) Positive homogeneity: If $\alpha \ge 0$ and $Z \in \mathbb{Z}_2$, then $\rho(\alpha Z) = \alpha \rho(Z)$.

The above conditions coincide with the respective conditions of the previous section which were defined in a finite dimensional setting. If the sigma algebra \mathcal{F}_1 is trivial, i.e., $\mathcal{F}_1 = \{\emptyset, \Omega\}$, then the space \mathcal{Z}_1 can be identified with \mathbb{R} , and conditions $(\mathbf{R}'1)-(\mathbf{R}'4)$ define a coherent risk measure. As in the previous section we refer to mapping $\rho : \mathcal{Z}_2 \to \mathcal{Z}_1$ satisfying conditions $(\mathbf{R}'1)-(\mathbf{R}'3)$ as a *convex* conditional risk mapping, while we refer to ρ as a *coherent* conditional risk mapping, or simply *conditional risk mapping*, if it satisfies conditions $(\mathbf{R}'1)-(\mathbf{R}'4)$.

We also will need stronger condition of strict monotonicity (compare with Remark 33 on page 326):

(**R**^{*}'2) Strict Monotonicity: If $Z, Z' \in \mathbb{Z}_2$ and $Z \succ Z'$, then $\rho(Z) \succ \rho(Z')$.

Remark 45. Consider a convex conditional risk mapping $\rho : \mathbb{Z}_2 \to \mathbb{Z}_1$. For a density function $\eta \in \mathbb{Z}_1^*$ (i.e., $\eta \succeq 0$ and $\int \eta \, dP = 1$), consider the corresponding functional

$$\varrho_{\eta}(Z) := \int_{\Omega} \eta \, \rho(Z) \, dP, \quad Z \in \mathcal{Z}_2.$$
(6.344)

Conditions (R'1)–(R'3) for ρ imply that the corresponding conditions (R1)–(R3) hold for ρ_{η} , and hence $\rho_{\eta} : \mathbb{Z}_2 \to \mathbb{R}$ is a convex risk measure. By Proposition 6.6, ρ_{η} is continuous. That is, if $Z_k \in \mathbb{Z}_2$ in a sequence converging to $Z \in \mathbb{Z}_2$ and $\eta \in \mathbb{Z}_1^*$, $\eta \succeq 0$, then

$$\lim_{k \to \infty} \int_{\Omega} \eta \,\rho(Z_k) \,dP = \int_{\Omega} \eta \,\rho(Z) \,dP.$$
(6.345)

For a general $\eta \in \mathbb{Z}_1^*$ we have the same conclusion (6.345) by applying the above arguments to the positive and negative parts of η . We obtain that ρ is *continuous* with respect to the strong (norm) topology of \mathbb{Z}_2 and weak topology of \mathbb{Z}_1 .

Theorem 6.70. Let $\rho : \mathbb{Z}_2 \to \mathbb{Z}_1$ be a conditional risk mapping, and $Y \in \mathbb{Z}_1$ and $Z \in \mathbb{Z}_2$ be such that $Y \succeq 0$ and $YZ \in \mathbb{Z}_2$. Then

$$\rho(YZ) = Y\rho(Z). \tag{6.346}$$

Proof. Recall that conditional risk mapping $\rho : \mathbb{Z}_2 \to \mathbb{Z}_1$ satisfies conditions (R'1)–(R'4). Let us show first that

$$\rho(\mathbf{1}_A Z) = \mathbf{1}_A \rho(Z), \quad \text{for all} \quad A \in \mathcal{F}_1, \ Z \in \mathcal{Z}_2.$$
(6.347)

Consider a density function $\eta \in \mathcal{Z}_1^*$ and the associated functional ϱ_η defined in (6.344). Conditions (R'1)–(R'4) for ρ imply that the corresponding conditions (R1)–(R4) hold for ϱ_η , and hence ϱ_η is a coherent risk measure defined on the space $\mathcal{Z}_2 = \mathcal{L}_p(\Omega, \mathcal{F}_2, P)$. Moreover, for any $B \in \mathcal{F}_1$ we have by (R'3) that for any $Z \in \mathcal{Z}_2$ and $\alpha \in \mathbb{R}$,

$$\varrho_{\eta}(Z + \alpha \mathbf{1}_{B}) = \int_{\Omega} \eta \rho(Z + \alpha \mathbf{1}_{B}) dP = \int_{\Omega} \eta [\rho(Z) + \alpha \mathbf{1}_{B}] dP
= \varrho_{\eta}(Z) + \alpha \int_{B} \eta dP.$$
(6.348)

As ρ_{η} is a coherent risk measure, by virtue of Theorem 6.5 it can be represented in the following form:

$$\varrho_{\eta}(Z) = \sup_{\zeta \in \mathfrak{A}(\eta)} \int_{\Omega} \zeta(\omega) Z(\omega) \, dP(\omega), \tag{6.349}$$

for some set $\mathfrak{A}(\eta) \subset \mathcal{L}_q(\Omega, \mathcal{F}_2, P)$ of probability density functions. Let us denote by A the support set of η . We can make the following observation.

• Each density $\zeta \in \mathfrak{A}(\eta)$ is supported on the set A.

Indeed, we argue by a contradiction. For any $B \in \mathcal{F}_1$, such that $P(B \cap A) = 0$, and any $\alpha \in \mathbb{R}$, we have by (6.348) that $\varrho_\eta(Z + \alpha \mathbf{1}_B) = \varrho_\eta(Z)$. On the other hand, if there exists $\bar{\zeta} \in \mathfrak{A}(\eta)$ such that $\int_B \bar{\zeta} dP > 0$, then it follows from (6.349) that

$$\varrho_{\eta}(Z + \alpha \mathbf{1}_B) \ge \int_{\Omega} \bar{\zeta}(\omega) Z(\omega) \, dP(\omega) + \alpha \int_{B} \bar{\zeta}(\omega) \, dP(\omega).$$

The right hand side of the above inequality tends to $+\infty$ as $\alpha \to +\infty$. Therefore by taking α large enough we obtain the required contradiction.

As all densities $\zeta \in \mathfrak{A}(\eta)$ are supported on A, we conclude from (6.349) that

$$\varrho_{\eta}(Z) = \sup_{\zeta \in \mathfrak{A}(\eta)} \int_{A} \zeta(\omega) Z(\omega) \, dP(\omega) = \varrho_{\eta}(\mathbf{1}_{A}Z).$$

Comparison with (6.344) leads to the conclusion that

$$\int_{A} \eta \,\rho(Z) \,dP = \int_{A} \eta \,\rho(\mathbf{1}_{A}Z) \,dP, \tag{6.350}$$

for every $Z \in \mathcal{Z}_2$, every set $A \in \mathcal{F}_1$ of positive measure P, and for every density function $\eta \in \mathcal{L}_q(\Omega, \mathcal{F}_1, P)$ supported on A. This means that we must have

$$\mathbf{1}_A \rho(Z) = \mathbf{1}_A \rho(\mathbf{1}_A Z), \quad \text{a.s.}, \tag{6.351}$$

for every $Z \in Z_2$ and every set $A \in \mathcal{F}_1$. Replacing A with $A^c = \Omega \setminus A$ and Z with $\mathbf{1}_A Z$ in (6.351) we also see that

$$\mathbf{1}_{A^c}\rho(\mathbf{1}_A Z) = \mathbf{1}_{A^c}\rho(\mathbf{1}_{A^c}\mathbf{1}_A Z) = 0.$$

We can just add $\mathbf{1}_{A^c}\rho(\mathbf{1}_A Z)$ to the right hand side of (6.351) to conclude that

$$\mathbf{1}_A \rho(Z) = \mathbf{1}_A \rho(\mathbf{1}_A Z) + \mathbf{1}_{A^c} \rho(\mathbf{1}_A Z) = \rho(\mathbf{1}_A Z)$$

This proves (6.347).

It follows from (6.347) that for any $A \in \mathcal{F}_1, Z \in \mathcal{Z}_2$ and $\omega \in A$ we have

$$[\rho(\mathbf{1}_A Z)](\omega) = [\rho(Z)\mathbf{1}_A](\omega) = [\rho(Z)](\omega).$$
(6.352)

Consider now an \mathcal{F}_1 -measurable step function $Y \succeq 0$, i.e., $Y = \sum_{i=1}^m \alpha_i \mathbf{1}_{B_i}$ where $\alpha_i \ge 0$ and B_i are disjoint \mathcal{F}_1 -measurable sets such that $\bigcup_{i=1}^m B_i = \Omega$. For $j \in \{1, ..., m\}$ and $\omega \in B_j$ we can write

$$\begin{aligned} \rho(YZ)](\omega) &= [\rho(YZ)\mathbf{1}_{B_j}](\omega) = [\rho(\mathbf{1}_{B_j}YZ)](\omega) \\ &= [\rho(\alpha_j\mathbf{1}_{B_j}Z)](\omega) = \alpha_j[\rho(\mathbf{1}_{B_j}Z)](\omega) \\ &= \alpha_j[\rho(Z)](\omega) = [\rho(Z)Y](\omega), \end{aligned}$$

where for $\omega \in B_j$ we used properties: $\mathbf{1}_{B_j}(\omega) = 1$, property (6.347), $\mathbf{1}_{B_j}Y = \alpha_j\mathbf{1}_{B_j}$, positive homogeneity of ρ , property (6.352) and $Y(\omega) = \alpha_j$. This proves (6.346) for step functions.

For a general \mathcal{F}_1 -measurable $Y \succeq 0$ we can complete the proof by taking a sequence of \mathcal{F}_1 -measurable step functions Y_k converging to Y (in the norm topology of \mathcal{Z}_1) and going to the limit (see Remark 45 on page 382). \Box

Examples of coherent risk measures, discussed in section 6.3.2, have conditional risk mapping analogues. Such conditional analogues can be constructed as follows. Consider a regular law invariant risk measure ρ (recall that by "regular" we mean that it can be defined on the standard uniform probability space, see section 6.3.6). Then $\rho(H)$ can be viewed as a function of cdf H. The conditional analogue of ρ is obtained by replacing cdf (distribution) H with the corresponding conditional distribution $H|\mathcal{F}_1$. Often, this can be done by replacing the expectation operator with the corresponding conditional expectation $\mathbb{E}[\cdot|\mathcal{F}_1]$ operator. Let us look at some examples.

Example 6.71 (Conditional expectation) In itself, the conditional expectation mapping $\mathbb{E}[\cdot |\mathcal{F}_1] : \mathcal{Z}_2 \to \mathcal{Z}_1$ is a conditional risk mapping. Indeed, for any $p \ge 1$ and $Z \in \mathcal{L}_p(\Omega, \mathcal{F}_2, P)$ we have by Jensen inequality that $\mathbb{E}[|Z|^p |\mathcal{F}_1] \succeq |\mathbb{E}[Z|\mathcal{F}_1]|^p$, and hence

$$\int_{\Omega} \left| \mathbb{E}[Z|\mathcal{F}_1] \right|^p dP \le \int_{\Omega} \mathbb{E}\left[|Z|^p |\mathcal{F}_1 \right] dP = \mathbb{E}\left[|Z|^p \right] < +\infty.$$
(6.353)

This shows that, indeed, $\mathbb{E}[\cdot |\mathcal{F}_1]$ maps \mathcal{Z}_2 into \mathcal{Z}_1 . The conditional expectation is a linear operator, and hence conditions (R'1) and (R'4) follow. The monotonicity condition (R'2) also clearly holds. Condition (R'3) is a property of conditional expectation. The strict monotonicity condition (R*2) also holds.

Example 6.72 (Conditional AV@R) Using definition (6.23) of AV@R, its conditional analogue can be defined as follows. Let $\mathcal{Z}_i := \mathcal{L}_1(\Omega, \mathcal{F}_i, P), i = 1, 2$. For $\alpha \in (0, 1]$ define mapping AV@R_{α}($\cdot | \mathcal{F}_1$) : $\mathcal{Z}_2 \to \mathcal{Z}_1$ as

$$[\mathsf{AV}@\mathsf{R}_{\alpha}(Z|\mathcal{F}_{1})](\omega) := \inf_{Y \in \mathcal{Z}_{1}} \left\{ Y(\omega) + \alpha^{-1} \mathbb{E}\left[[Z - Y]_{+} \big| \mathcal{F}_{1} \right](\omega) \right\}, \ \omega \in \Omega.$$
(6.354)

The minimum in the right hand side of (6.354) is attained at $\overline{Y} = V@R_{\alpha}(Z|\mathcal{F}_1)$ (see (6.24)), where $V@R_{\alpha}(Z|\mathcal{F}_1)$ is the (left side) $(1-\alpha)$ -quantile of the conditional distribution of Z. Therefore we can write

$$\mathsf{AV}@\mathsf{R}_{\alpha}(Z|\mathcal{F}_{1}) = \mathsf{V}@\mathsf{R}_{\alpha}(Z|\mathcal{F}_{1}) + \alpha^{-1}\mathbb{E}\big[[Z - \mathsf{V}@\mathsf{R}_{\alpha}(Z|\mathcal{F}_{1})]_{+}\big|\mathcal{F}_{1}\big].$$
(6.355)

It follows that AV@R_{α}($Z|\mathcal{F}_1$) is \mathcal{F}_1 -measurable. Furthermore, it is not difficult to verify that this mapping satisfies conditions (R'1)–(R'4). Alternatively by using equation (6.27), the conditional AV@R can be written as follows

$$AV@R_{\alpha}(Z|\mathcal{F}_{1}) = \frac{1}{\alpha} \int_{1-\alpha}^{1} V@R_{1-\tau}(Z|\mathcal{F}_{1}) d\tau.$$
(6.356)

Similarly to (6.74), for $\beta \in [0, 1]$ and $\alpha \in (0, 1)$, we can also consider the following conditional risk mapping
$$\rho_{\alpha,\beta|\mathcal{F}_1}(Z) := (1-\beta)\mathbb{E}[Z|\mathcal{F}_1] + \beta \mathsf{AV}@\mathsf{R}_{\alpha}(Z|\mathcal{F}_1).$$
(6.357)

Of course, the conditional risk mapping $\rho_{\alpha,\beta|\mathcal{F}_1}$ corresponds to the coherent risk measure $\rho_{\alpha,\beta}(Z) = (1 - \beta)\mathbb{E}[Z] + \beta AV@R_{\alpha}(Z)$. For $\alpha \in [0,1)$ and $\beta \in [0,1)$ the strict monotonicity condition (R*'2) also holds (compare with Remark 33 on page 326).

Example 6.73 (Conditional mean-upper-semideviation) An analogue of the mean-uppersemideviation risk measure (of order p) can be constructed as follows. Let $\mathcal{Z}_i := \mathcal{L}_p(\Omega, \mathcal{F}_i, P)$, i = 1, 2. For $c \in [0, 1]$ define

$$\rho_{c|\mathcal{F}_1}(Z) := \mathbb{E}[Z|\mathcal{F}_1] + c \left(\mathbb{E}\Big[\left[Z - \mathbb{E}[Z|\mathcal{F}_1] \right]_+^p | \mathcal{F}_1 \Big] \right)^{1/p}.$$
(6.358)

In particular, for p = 1 this gives an analogue of the absolute semideviation risk measure. For $p \in [1, \infty)$ and $c \in [0, 1)$ the strict monotonicity condition ($\mathbb{R}^{\star'2}$) also holds (compare with Remark 33 on page 326).

In the discrete case of scenario tree formulation (discussed in the previous section) the above examples correspond to taking the same respective risk measure at every node of the considered tree at stage t = 1, ..., T.

Similarly to (6.321), we show now that a conditional risk mapping can be represented as a maximum of a family of conditional expectations. To simplify the derivations, we assume that the subalgebra \mathcal{F}_1 has a countable number of elementary events. That is, there is a (countable) partition $\{A_i\}_{i\in\mathbb{N}}$, of the sample space Ω which generates \mathcal{F}_1 , i.e., $\cup_{i\in\mathbb{N}}A_i = \Omega$, the sets $A_i, i \in \mathbb{N}$, are disjoint and form the family of elementary events of sigma algebra \mathcal{F}_1 . Since \mathcal{F}_1 is a subalgebra of \mathcal{F}_2 , we have of course that $A_i \in \mathcal{F}_2, i \in \mathbb{N}$. We also have that a function $Z : \Omega \to \mathbb{R}$ is \mathcal{F}_1 -measurable iff it is constant on every set A_i , $i \in \mathbb{N}$.

Consider a conditional risk mapping $\rho : \mathbb{Z}_2 \to \mathbb{Z}_1$. Let

$$\mathfrak{N} := \{ i \in \mathbb{N} : P(A_i) \neq 0 \}.$$

For $i \in \mathfrak{N}$ consider probability density function $\eta_i := \frac{1}{P(A_i)} \mathbf{1}_{A_i}$ and risk measure

$$\varrho_i(Z) := \int_{\Omega} \eta_i \rho(Z) dP = \frac{1}{P(A_i)} \int_{A_i} \rho(Z) dP, \quad Z \in \mathcal{Z}_2.$$

As in (6.349) with every ϱ_i , $i \in \mathfrak{N}$, is associated a set $\mathfrak{A}_i \subset \mathbb{Z}_1^*$ of probability density functions, supported on the set A_i , such that

$$\varrho_i(Z) = \sup_{\zeta \in \mathfrak{A}_i} \int_{\Omega} \zeta(\omega) Z(\omega) \, dP(\omega).$$
(6.359)

Now let $\nu = (\nu_i)_{i \in \mathbb{N}}$ be a probability distribution (measure) on (Ω, \mathcal{F}_1) , assigning probability ν_i to the event $A_i, i \in \mathbb{N}$. Assume that ν is such that $\nu(A_i) = 0$ iff $P(A_i) = 0$ (i.e., ν is absolutely continuous with respect to P and P is absolutely continuous with respect to ν on (Ω, \mathcal{F}_1)), otherwise the probability measure ν is arbitrary. Define the following family of probability measures on (Ω, \mathcal{F}_2) :

$$\mathfrak{V} := \left\{ \mu = \sum_{i \in \mathfrak{N}} \nu_i \mu_i : d\mu_i = \zeta_i \, dP, \ \zeta_i \in \mathfrak{A}_i, \ i \in \mathfrak{N} \right\}.$$
(6.360)

Note that since $\nu_i \ge 0$, $i \in \mathfrak{N}$, and $\sum_{i \in \mathfrak{N}} \nu_i = 1$, every $\mu \in \mathfrak{V}$ is a probability measure. Since ζ_i is supported on A_i , we have $\mu_i(A_i) = \int_{A_i} \zeta_i dP = 1$, and

$$\mathbb{E}_{\mu_i}[Z|\mathcal{F}_1](\omega) = \begin{cases} \int_{A_i} Z\zeta_i \, dP, & \text{if } \omega \in A_i, \\ 0, & \text{otherwise.} \end{cases}$$
(6.361)

It follows that for $Z \in \mathbb{Z}_2$, $\omega \in A_i$ and $\mu = \sum_{i \in \mathfrak{N}} \nu_i \mu_i \in \mathfrak{V}$,

$$\mathbb{E}_{\mu}[Z|\mathcal{F}_1](\omega) = \frac{1}{\mu(A_i)} \int_{A_i} Zd\mu = \frac{1}{\nu_i} \int_{A_i} \nu_i \zeta_i ZdP = \int_{A_i} Z\zeta_i \, dP.$$

Consequently by the max-representations (6.359) it follows that for $Z \in \mathcal{Z}_2$ and $\omega \in A_i$,

$$\sup_{\mu \in \mathfrak{V}} \mathbb{E}_{\mu}[Z|\mathcal{F}_1](\omega) = \sup_{\zeta_i \in \mathfrak{A}_i} \int_{A_i} Z\zeta_i \, dP = \varrho_i(Z).$$
(6.362)

Also since $[\rho(Z)](\cdot)$ is \mathcal{F}_1 -measurable, and hence is constant on every set A_i , we have that for any $i \in \mathfrak{N}$, $[\rho(Z)](\omega) = \varrho_i(Z)$ for every $\omega \in A_i$. We obtain the following result.

Proposition 6.74. Let $Z_i := \mathcal{L}_p(\Omega, \mathcal{F}_i, P)$, i = 1, 2, with $\mathcal{F}_1 \subset \mathcal{F}_2$, and let $\rho : Z_2 \to Z_1$ be a conditional risk mapping. Suppose that \mathcal{F}_1 has a countable number of elementary events. Then

$$\rho(Z) = \sup_{\mu \in \mathfrak{V}} \mathbb{E}_{\mu}[Z|\mathcal{F}_1], \quad \forall Z \in \mathcal{Z}_2,$$
(6.363)

where \mathfrak{V} is a family of probability measures on (Ω, \mathcal{F}_2) , specified in (6.360), corresponding to a probability distribution ν on (Ω, \mathcal{F}_1) .

6.8.3 Dynamic Risk Measures

In our introductory presentation in section 6.8.1, we constructed a risk-averse multistage optimization problem (6.323), by postulating that the risk of the random cost sequence $Z_1 = f_1(x_1), Z_2 = f_2(x_2), \dots, Z_T = f_T(x_T)$ is evaluated as follows:

$$\varrho_{1,T}(Z_1, Z_2, \dots, Z_T) = Z_1 + \rho_2 \Big(Z_2 + \dots + \rho_{T-1} \big(Z_{T-1} + \rho_T(Z_T) \big) \Big), \quad (6.364)$$

where $\rho_{t+1} : Z_{t+1} \to Z_t, t = 1, ..., T - 1$, are conditional risk mappings. Our objective in this section is to provide a justification for this nested structure.

Consider a probability space (Ω, \mathcal{F}, P) , a filtration $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T \subset \mathcal{F}$, and an adapted sequence of random variables $Z_t, t = 1, \ldots, T$. We assume that $\mathcal{F}_1 = \{\Omega, \emptyset\}$, and thus Z_1 is in fact deterministic. In all our considerations we interpret the variables Z_t as

stage-wise costs. Define the spaces $\mathcal{Z}_t := \mathcal{L}_p(\Omega, \mathcal{F}_t, P), p \in [1, \infty), t = 1, \dots, T$, and let $\mathcal{Z}_{t,T} := \mathcal{Z}_t \times \cdots \times \mathcal{Z}_T.$

As we are planning to use our risk evaluation of the sequence Z_1, Z_2, \ldots, Z_T for the purpose of making decisions at stages $1, 2, \ldots, T$, we need to evaluate the risk of the tail subsequences Z_t, \ldots, Z_T from the perspective of each stage t. This motivates the following definition.

Definition 6.75. A dynamic risk measure is a sequence of mappings $\varrho_{t,T} : \mathcal{Z}_{t,T} \to \mathcal{Z}_t$, $t = 1, \ldots, T$, satisfying for all $t = 1, \ldots, T$ and all (Z_t, \ldots, Z_T) and (W_t, \ldots, W_T) in $\mathcal{Z}_{t,T}$ the following monotonicity property:

if
$$(Z_t, \ldots, Z_T) \preceq (W_t, \ldots, W_T)$$
 then $\varrho_{t,T}(Z_t, \ldots, Z_T) \preceq \varrho_{t,T}(W_t, \ldots, W_T)$. (6.365)

The value of the "tail risk measure" $\varrho_{t,T}(Z_t, \ldots, Z_T)$ can be interpreted as a fair onetime \mathcal{F}_t -measurable charge we would be willing to incur at time t, instead of the sequence of random future costs Z_t, \ldots, Z_T . The key issue associated with dynamic preferences is the question of their consistency over time.

Definition 6.76. A dynamic risk measure $\{\varrho_{t,T}\}_{t=1}^{T}$ is called time consistent if for all $1 \leq \tau < \theta \leq T$ and all sequences $Z, W \in \mathcal{Z}_{\tau,T}$ the conditions

$$Z_k = W_k, \ k = \tau, \dots, \theta - 1 \quad and \quad \varrho_{\theta,T}(Z_\theta, \dots, Z_T) \preceq \varrho_{\theta,T}(W_\theta, \dots, W_T) \quad (6.366)$$

imply that

$$\varrho_{\tau,T}(Z_{\tau},\ldots,Z_T) \preceq \varrho_{\tau,T}(W_{\tau},\ldots,W_T).$$
(6.367)

In words, if Z will be at least as good as W from the perspective of some future time θ , and they are identical between now time τ and future time θ , then Z should not be worse than W from today's perspective. Note that time consistency implies that if in (6.366) actually the equality $\varrho_{\theta,T}(Z_{\theta},\ldots,Z_T) = \varrho_{\theta,T}(W_{\theta},\ldots,W_T)$ holds, then the equality $\varrho_{\tau,T}(Z_{\tau},\ldots,Z_T) = \varrho_{\tau,T}(W_{\tau},\ldots,W_T)$ follows. For a dynamic risk measure $\{\varrho_{t,T}\}_{t=1}^T$ we can define a broader family of conditional

risk measures, by setting

$$\varrho_{\tau,\theta}(Z_{\tau},\ldots,Z_{\theta}) := \varrho_{\tau,T}(Z_{\tau},\ldots,Z_{\theta},0,\ldots,0), \quad 1 \le \tau \le \theta \le T.$$
(6.368)

We can derive the following structure of a time consistent dynamic risk measure.

Theorem 6.77. Suppose a dynamic risk measure $\{\varrho_{t,T}\}_{t=1}^T$ satisfies for all $t = 1, \ldots, T$ and all $Z_t \in \mathcal{Z}_t$ the condition:

$$\varrho_{t,T}(Z_t, 0, \dots, 0) = Z_t. \tag{6.369}$$

Then it is time consistent if and only if for all $1 \le \tau \le \theta \le T$ and all $Z \in \mathcal{Z}_{1,T}$ the following identity holds:

$$\varrho_{\tau,T}(Z_{\tau},\ldots,Z_{\theta},\ldots,Z_{T}) = \varrho_{\tau,\theta}(Z_{\tau},\ldots,Z_{\theta-1},\varrho_{\theta,T}(Z_{\theta},\ldots,Z_{T})).$$
(6.370)

Proof. Consider two sequences:

$$Z = (Z_{\tau}, \dots, Z_{\theta-1}, Z_{\theta}, Z_{\theta+1}, \dots, Z_T),$$

$$W = (Z_{\tau}, \dots, Z_{\theta-1}, \rho_{\theta,T}(Z_{\theta}, \dots, Z_T), 0, \dots, 0).$$

Suppose the measure $\{\varrho_{t,T}\}_{t=1}^{T}$ is time consistent. Then, by (6.369),

$$\varrho_{\theta,T}(W_{\theta},\ldots,W_T) = \varrho_{\theta,T}(\varrho_{\theta,T}(Z_{\theta},\ldots,Z_T),0,\ldots,0) = \varrho_{\theta,T}(Z_{\theta},\ldots,Z_T).$$

Using Definition 6.76 we get $\rho_{\tau,T}(Z) = \rho_{\tau,T}(W)$. Equation (6.368) then yields (6.370).

Conversely, suppose the identity (6.370). Consider Z and W satisfying conditions (6.366). Then, by Definition 6.75, we have

$$\varrho_{\tau,T}(Z_{\tau},\ldots,Z_{\theta-1},\varrho_{\theta,T}(Z_{\theta},\ldots,Z_{T}),0,\ldots,0)) \\
\leq \varrho_{\tau,T}(W_{\tau},\ldots,W_{\theta-1},\varrho_{\theta,T}(W_{\theta},\ldots,W_{T}),0,\ldots,0).$$

Using equation (6.370) we obtain (6.367).

In order to justify the nested structure (6.364), we employ a form of translation condition for dynamic risk measures:

$$\varrho_{t,T}(Z_t, Z_{t+1}, \dots, Z_T) = Z_t + \varrho_{t,T}(0, Z_{t+1}, \dots, Z_T).$$
(6.371)

Together with the normalization $\rho_{t,T}(0, ..., 0) = 0$, it implies (6.369). If the risk measure is time consistent and satisfies (6.371), then we obtain the chain of equations:

$$\varrho_{t,T}(Z_t,\ldots,Z_{T-1},Z_T) = \varrho_{t,T}(Z_t,\ldots,\varrho_{T-1,T}(Z_{T-1},Z_T),0) = \varrho_{t,T-1}(Z_t,\ldots,\varrho_{T-1,T}(Z_{T-1},Z_T)) = \varrho_{t,T-1}(Z_t,\ldots,Z_{T-1}+\varrho_{T-1,T}(0,Z_T)).$$

In the first equation we used the identity (6.370), in the second one – equation (6.368), and in the third one – condition (6.371). Define one-step conditional risk measures ρ_{t+1} : $\mathcal{Z}_{t+1} \rightarrow \mathcal{Z}_t$, $t = 1, \ldots, T - 1$ as follows:

$$\rho_{t+1}(Z_{t+1}) = \varrho_{t,t+1}(0, Z_{t+1}).$$

Proceeding in this way, we obtain for all t = 1, ..., T the following recursive relation:

$$\varrho_{t,T}(Z_t, \dots, Z_T) = Z_t + \rho_{t+1} \bigg(Z_{t+1} + \rho_{t+2} \Big(Z_{t+2} + \dots + \rho_{T-1} \big(Z_{T-1} + \rho_T (Z_T) \big) \Big) \bigg).$$
(6.372)

It follows that a time consistent dynamic risk measure is completely defined by one-step conditional risk measures ρ_{t+1} , $t = 1, \ldots, T-1$. For t = 1 formula (6.372) becomes formally identical with (6.364) and defines a risk measure of the *entire* sequence $Z \in \mathcal{Z}_{1,T}$ (with a deterministic Z_1). In order to obtain complete equivalence, we need to restrict the class of conditional risk measures ρ_{t+1} , $t = 1, \ldots, T-1$, to the mappings satisfying

conditions (R'1)–(R'4). In this case, we can use condition (R'3) recursively, to obtain the representation

$$\varrho_{1,T}(Z_1, Z_2, \dots, Z_T) = \rho_2 \Big(\rho_3 \Big(\cdots \big(\rho_T (Z_1 + Z_2 + \dots + Z_T) \big) \Big) \Big).$$
(6.373)

We conclude that under the specified conditions

$$\varrho_{1,T}(Z_1, Z_2, \dots, Z_T) = \bar{\rho}(Z_1 + Z_2 + \dots + Z_T), \tag{6.374}$$

with

$$\bar{\rho} = \rho_{2|\mathcal{F}_1} \circ \cdots \rho_{T|\mathcal{F}_{T-1}} \tag{6.375}$$

for some sequence $\rho_{t+1|\mathcal{F}_t} : \mathcal{Z}_{t+1} \to \mathcal{Z}_t, t = 1, ..., T-1$, of coherent conditional risk mappings.

Let us focus on the structure (6.374). A question arises whether a coherent (convex) risk measure $\bar{\rho} : \mathcal{Z}_T \to \mathbb{R}$ is representable in the composite form (6.375). In this respect we have the following observation.

Remark 46. Consider the following construction. Let the space $\Omega := \Omega_1 \times \Omega_2$, where $\Omega_1 = \{\omega_1, \ldots, \omega_n\}$ and $\Omega_2 = \{\omega'_1, \ldots, \omega'_n\}$ are finite spaces with *n* elementary events equipped with equal probabilities 1/n. That is, Ω is a finite space with n^2 elementary events each having probability $1/n^2$. Consider filtration $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3$, where $\mathcal{F}_1 = \{\emptyset, \Omega\}$, \mathcal{F}_2 is generated by the elementary events $\{\omega_i\} \times \Omega_2, i = 1, \ldots, n, \text{ and } \mathcal{F}_3$ consists of all subsets of Ω . This corresponds to a scenario tree with T = 3 stages, *n* children nodes of the root node with probability 1/n of moving to each node of stage 2, and *n* children nodes of each node of the second stage with the corresponding probability 1/n of moving to the last stage 3. Let \mathcal{Z}_t be the spaces of \mathcal{F}_t -measurable random variables, t = 1, 2, 3. Recall that here two random variables $Z, Z' : \Omega \to \mathbb{R}$ are distributionally equivalent iff $Z' = Z \circ \pi$ for some permutation π of the set Ω .

Consider a law invariant risk measure $\bar{\rho} : \mathcal{Z}_3 \to \mathbb{R}$. If $\bar{\rho}(\cdot) = \mathbb{E}[\cdot]$, then it can be represented as a composition of (conditional) expectation operators. Another example of decomposable risk measure is the max-risk measure $\bar{\rho}(\cdot) = \max_{\omega \in \Omega} Z(\omega)$, which can be represented as the composition of (conditional) max-risk mappings. It turns out that these are only examples of *law invariant* risk measures representable as the composition of law invariant *coherent* conditional risk mappings. Proof of the following result can be found in [246].

Theorem 6.78. In the above construction let $\bar{\rho} = \rho_{2|\mathcal{F}_1} \circ \rho_{3|\mathcal{F}_2}$ be a composite risk measure, where $\rho_{2|\mathcal{F}_1} : \mathcal{Z}_2 \to \mathcal{Z}_1$ and $\rho_{3|\mathcal{F}_2} : \mathcal{Z}_3 \to \mathcal{Z}_2$ are law invariant coherent conditional²⁷ risk mappings. Then $\bar{\rho} : \mathcal{Z}_3 \to \mathbb{R}$ is law invariant if and only if either both $\rho_{2|\mathcal{F}_1}$ and $\rho_{3|\mathcal{F}_2}$ are conditional expectations or both $\rho_{2|\mathcal{F}_1}$ and $\rho_{3|\mathcal{F}_2}$ are conditional max-mappings.

It should be stressed, however, that the concept of law invariance employed in this setting requires that the risk evaluation of the sequence Z_1, Z_2, \ldots, Z_T depends only on the marginal distribution of the sum $Z_1 + Z_2 + \cdots + Z_T$, but is not allowed to depend on

²⁷Since $\mathcal{Z}_1 = \mathbb{R}$, $\rho_{2|\mathcal{F}_1}(\cdot)$ actually is real valued.

conditional distributions at intermediate times 1 < t < T. This is the source of the paradox discussed above. If we choose any law invariant coherent conditional risk mappings $\rho_{t+1|\mathcal{F}_t} : \mathcal{Z}_{t+1} \to \mathcal{Z}_t, t = 1, ..., T - 1$, their composition (6.375) will not necessarily be law invariant in the sense employed here.

Remark 47. Let $\rho : \mathcal{Z}_T \to \mathbb{R}$ be a law invariant coherent risk measure and $\rho_{|\mathcal{F}_t} : \mathcal{Z}_T \to \mathcal{Z}_t$, $t = 1, \ldots, T$, be its conditional analogues. Consider the corresponding sequence of multiperiod mappings:

$$\varrho_{t,T}(Z_t, \dots, Z_T) := \rho_{|\mathcal{F}_t}(Z_t + \dots + Z_T), \ t = 1, \dots, T.$$
(6.376)

This sequence satisfies condition (6.371). If it was time consistent, then by the above derivation of (6.372) it would follow that

$$\rho_{|\mathcal{F}_t}(Z_t + \dots + Z_T) = Z_t + \rho_{|\mathcal{F}_{t+1}} [Z_{t+1} + \dots + \rho_{|\mathcal{F}_T}[Z_T]].$$
(6.377)

However, in general the decomposition (6.377) holds only for the expectation and max risk measures (see Theorem 6.78). Therefore multiperiod mappings (6.376) are time consistent only for the expectation and max risk measures.

6.8.4 Risk Averse Multistage Stochastic Programming

There are several ways how risk averse stochastic programming can be formulated in a multistage setting. We discuss now a nested formulation similar to the derivations of section 6.8.1 and 6.8.3. Let (Ω, \mathcal{F}, P) be a probability space and $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T$ be a sequence of nested sigma algebras with $\mathcal{F}_1 = \{\emptyset, \Omega\}$ being trivial sigma algebra and $\mathcal{F}_T = \mathcal{F}$ (such sequence of sigma algebras is called a *filtration*). For $p \in [1, \infty)$ let $\mathcal{Z}_t := \mathcal{L}_p(\Omega, \mathcal{F}_t, P)$, $t = 1, \ldots, T$, be the corresponding sequence of spaces of \mathcal{F}_t -measurable and p-integrable functions, and $\rho_{t+1|\mathcal{F}_t} : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$, $t = 1, \ldots, T - 1$, be a selected family of conditional risk mappings. It is straightforward to verify that the composition

$$\rho_{t|\mathcal{F}_{t-1}} \circ \cdots \circ \rho_{T|\mathcal{F}_{T-1}} : \mathcal{Z}_T \xrightarrow{\rho_{T|\mathcal{F}_{T-1}}} \mathcal{Z}_{T-1} \xrightarrow{\rho_{T-1|\mathcal{F}_{T-2}}} \cdots \xrightarrow{\rho_{t|\mathcal{F}_{t-1}}} \mathcal{Z}_{t-1}, \qquad (6.378)$$

t = 2, ..., T, of such conditional risk mappings is also a conditional risk mapping. In particular, the space Z_1 can be identified with \mathbb{R} and hence the composition $\rho_{2|\mathcal{F}_1} \circ \cdots \circ \rho_{T|\mathcal{F}_{T-1}} : Z_T \to \mathbb{R}$ is a real valued coherent risk measure.

Similarly to (6.323), we consider the following nested risk averse formulation of multistage programs:

$$\underset{x_{1} \in \mathcal{X}_{1}}{\min} f_{1}(x_{1}) + \rho_{2|\mathcal{F}_{1}} \left[\inf_{x_{2} \in \mathcal{X}_{2}(x_{1})} f_{2}(x_{2}) + \dots \\ \dots + \rho_{T-1|\mathcal{F}_{T-2}} \left[\inf_{x_{T-1} \in \mathcal{X}_{T}(x_{T-2})} f_{T-1}(x_{T-1}) \\ + \rho_{T|\mathcal{F}_{T-1}} \left[\inf_{x_{T} \in \mathcal{X}_{T}(x_{T-1})} f_{T}(x_{T}) \right] \right]. \quad (6.379)$$

Here $f_t : \mathbb{R}^{n_{t-1}} \times \Omega \to \mathbb{R}$ and $\mathcal{X}_t : \mathbb{R}^{n_{t-1}} \times \Omega \rightrightarrows \mathbb{R}^{n_t}$, $t = 2, \ldots, T$, are such that $f_t(x_t, \cdot) \in \mathcal{Z}_t$ and $\mathcal{X}_t(x_{t-1}, \cdot)$ are \mathcal{F}_t -measurable for all x_t and x_{t-1} .

As it was discussed in section 6.8.1, the above nested formulation (6.379) has two equivalent interpretations. Namely, it can be formulated as

$$\underset{x_{1}, \boldsymbol{x}_{2}, \dots, \boldsymbol{x}_{T}}{\min} f_{1}(x_{1}) + \rho_{2|\mathcal{F}_{1}} \left[f_{2}(\boldsymbol{x}_{2}) + \dots + \rho_{T-1|\mathcal{F}_{T-2}} \left[f_{T-1}(\boldsymbol{x}_{T-1}) + \rho_{T|\mathcal{F}_{T-1}} \left[f_{T}(\boldsymbol{x}_{T}) \right] \right] \right]$$
s.t. $x_{1} \in \mathcal{X}_{1}, \ \boldsymbol{x}_{t}(\omega) \in \mathcal{X}_{t}(\boldsymbol{x}_{t-1}(\omega), \omega), \ t = 2, \dots, T,$

$$(6.380)$$

where the optimization is performed over \mathcal{F}_t -measurable $\boldsymbol{x}_t : \Omega \to \mathbb{R}, t = 1, ..., T$, satisfying the corresponding constraints, and such that $f_t(\boldsymbol{x}_t(\cdot), \cdot) \in \mathcal{Z}_t$. Recall that the nonanticipativity is enforced here by the \mathcal{F}_t -measurability of $\boldsymbol{x}_t(\cdot)$. By using the *composite risk measure* $\bar{\rho} := \rho_{2|\mathcal{F}_1} \circ \cdots \circ \rho_{T|\mathcal{F}_{T-1}}$, we also can write (6.380) in the form

$$\begin{array}{ll} \underset{x_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T}{\text{Min}} & \bar{\rho} \Big[f_1(x_1) + f_2(\boldsymbol{x}_2) + \dots + f_T(\boldsymbol{x}_T) \Big] \\ \text{s.t.} & x_1 \in \mathcal{X}_1, \ \boldsymbol{x}_t(\omega) \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\omega), \omega), \ t = 2, \dots, T. \end{array}$$
(6.381)

Recall that for $Z_t \in \mathcal{Z}_t, t = 1, ..., T$,

$$\bar{\rho}(Z_1 + \dots + Z_T) = Z_1 + \rho_{2|\mathcal{F}_1} \Big[Z_2 + \dots + \rho_{T-1|\mathcal{F}_{T-2}} \big[Z_{T-1} + \rho_{T|\mathcal{F}_{T-1}} [Z_T] \big] \Big],$$
(6.382)

and that conditions $(\mathbf{R}'1)$ – $(\mathbf{R}'4)$ imply that $\bar{\rho}: \mathbb{Z}_T \to \mathbb{R}$ is a coherent risk measure.

Alternatively we can write the corresponding dynamic programming equations (compare with (6.331)–(6.336)):

$$Q_T(x_{T-1},\omega) = \inf_{x_T \in \mathcal{X}_T(x_{T-1},\omega)} f_T(x_T,\omega),$$
(6.383)

$$Q_t(x_{t-1},\omega) = \inf_{x_t \in \mathcal{X}_t(x_{t-1},\omega)} \left\{ f_t(x_t,\omega) + \mathcal{Q}_{t+1}(x_t,\omega) \right\}, \ t = T - 1, \dots, 2, \quad (6.384)$$

where

$$Q_t(x_{t-1},\omega) = \rho_{t|\mathcal{F}_{t-1}} \left[Q_t(x_{t-1}) \right](\omega), \ t = T, \dots, 2.$$
(6.385)

Finally, at the first stage we solve the problem

$$\min_{x_1 \in \mathcal{X}_1} f_1(x_1) + \rho_{2|\mathcal{F}_1}[Q_2(x_1)].$$
(6.386)

A policy $\bar{x}_1, \ldots, \bar{x}_T$ is optimal iff \bar{x}_1 is an optimal solution of the first stage problem (6.386) and for $t = 2, \ldots, T$, it satisfies the optimality conditions:

$$\bar{\boldsymbol{x}}_{t}(\omega) \in \operatorname*{argmin}_{x_{t} \in \mathcal{X}_{t}(\bar{\boldsymbol{x}}_{t-1}(\omega),\omega)} \left\{ f_{t}(x_{t},\omega) + \rho_{t+1}\left[Q_{t+1}(x_{t})\right](\omega) \right\}, \text{ w.p.1.}$$
(6.387)

It should be remarked that we need to ensure here that the cost-to-go functions are *p*-integrable, i.e., $Q_t(x_{t-1}, \cdot) \in \mathbb{Z}_t$ for $t = 1, \ldots, T-1$ and all feasible x_{t-1} .

Conditional Risk Mappings of a Data Process

In applications we are often deal with a data process represented by a sequence of random vectors ξ_1, \ldots, ξ_T , say defined on a probability space (Ω, \mathcal{F}, P) . We can associate with this data process the filtration²⁸ $\mathcal{F}_t := \sigma(\xi_1, \ldots, \xi_t), t = 1, \ldots, T$, and hence to proceed with handling conditional risk mapping defined with respect to the obtained filtration $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T$.

However, often it is more convenient to deal with conditional risk mappings defined directly in terms of the data process rather that the respective sequence of sigma algebras. For example, consider

$$\rho_{t|\xi_{[t-1]}}(Z) := (1 - \beta_t) \mathbb{E}\left[Z|\xi_{[t-1]}\right] + \beta_t \mathsf{AV} @\mathsf{R}_{\alpha_t}(Z|\xi_{[t-1]}), \quad t = 2, \dots, T, \quad (6.388)$$

where

$$\mathsf{AV}@\mathsf{R}_{\alpha_t}(Z|\xi_{[t-1]}) := \inf_{Y \in \mathcal{Z}_{t-1}} \left\{ Y + \alpha_t^{-1} \mathbb{E}\big[[Z - Y]_+ \big| \xi_{[t-1]} \big] \right\}.$$
(6.389)

Here $\beta_t \in [0, 1]$ and $\alpha_t \in (0, 1)$ are chosen constants, $Z_t := \mathcal{L}_1(\Omega, \mathcal{F}_t, P)$, where $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T$ is the filtration associated with the process ξ_t , and the minimum on the right hand side of (6.389) is taken pointwise in $\omega \in \Omega$. Compared with (6.354), the conditional AV@R is defined in (6.389) in terms of the conditional expectation with respect to the history $\xi_{[t-1]}$ of the data process rather than the corresponding sigma algebra \mathcal{F}_{t-1} . We can also consider conditional mean-upper-semideviation risk mappings of the form:

$$\rho_{t|\xi_{[t-1]}}(Z) := \mathbb{E}[Z|\xi_{[t-1]}] + c_t \left(\mathbb{E}\Big[\left[Z - \mathbb{E}[Z|\xi_{[t-1]}] \right]_+^p \big| \xi_{[t-1]} \Big] \right)^{1/p}, \tag{6.390}$$

defined in terms of the data process.

Let us also assume that the objective functions $f_t(x_t, \xi_t)$ and feasible sets $\mathcal{X}_t(x_{t-1}, \xi_t)$ are given in terms of the data process. Then formulation (6.380) takes the form

$$\underset{x_{1}, \boldsymbol{x}_{2}, \dots, \boldsymbol{x}_{T}}{\min} f_{1}(x_{1}) + \rho_{2|\xi_{[1]}} \left[f_{2}(\boldsymbol{x}_{2}(\xi_{[2]}), \xi_{2}) + \dots + \rho_{T-1|\xi_{[T-2]}} \left[f_{T-1} \left(\boldsymbol{x}_{T-1}(\xi_{[T-1]}), \xi_{T-1} \right) + \rho_{T|\xi_{[T-1]}} \left[f_{T} \left(\boldsymbol{x}_{T}(\xi_{[T]}), \xi_{T} \right) \right] \right] \right] \\$$
s.t. $x_{1} \in \mathcal{X}_{1}, \ \boldsymbol{x}_{t}(\xi_{[t]}) \in \mathcal{X}_{t}(\boldsymbol{x}_{t-1}(\xi_{[t-1]}), \xi_{t}), \ t = 2, \dots, T,$

$$(6.391)$$

where the optimization is performed over feasible policies. The corresponding dynamic programming equations (6.384)–(6.385) take the form:

$$Q_t(x_{t-1},\xi_{[t]}) = \inf_{x_t \in \mathcal{X}_t(x_{t-1},\xi_t)} \Big\{ f_t(x_t,\xi_t) + \mathcal{Q}_{t+1}(x_t,\xi_{[t]}) \Big\},$$
(6.392)

where

$$\mathcal{Q}_{t+1}(x_t,\xi_{[t]}) = \rho_{t+1|\xi_{[t]}} \big[Q_{t+1}(x_t,\xi_{[t+1]}) \big].$$
(6.393)

²⁸Recall that $\sigma(\xi_1, \ldots, \xi_t)$ denotes the smallest sigma algebra with respect to which $\xi_{[t]} = (\xi_1, \ldots, \xi_t)$ is measurable.

Of course, if we set $\rho_{t|\xi_{[t-1]}}(\cdot) := \mathbb{E}\left[\cdot|\xi_{[t-1]}\right]$, then the above equations (6.392)– (6.393) coincide with the corresponding risk neutral dynamic programming equations. Also in that case the composite measure $\bar{\rho}$ becomes the corresponding expectation operator and hence formulation (6.381) coincides with the respective risk neutral formulation (3.4). Unfortunately, in the general case it is quite difficult to write the composite measure $\bar{\rho}$ in an explicit form (see Remark 42 on page 379).

Remark 48. Note that with $\rho_{t|\xi_{[t-1]}}$, defined in (6.388) or (6.390), is associated coherent risk measure ρ_t which is obtained by replacing the conditional expectations with respective (unconditional) expectations. Note also that if random variable $Z \in \mathcal{Z}_t$ is independent of $\xi_{[t-1]}$, then the conditional expectations on the right hand sides of (6.388)–(6.390) coincide with the respective unconditional expectations, and hence $\rho_{t|\xi_{[t-1]}}(Z)$ does not depend on $\xi_{[t-1]}$ and coincides with $\rho_t(Z)$.

If the process ξ_t is *stagewise independent*, then the dynamic programming equations (6.392)–(6.393) take the form:

$$Q_t(x_{t-1},\xi_t) = \inf_{x_t \in \mathcal{X}_t(x_{t-1},\xi_t)} \Big\{ f_t(x_t,\xi_t) + \mathcal{Q}_{t+1}(x_t) \Big\},$$
(6.394)

where

$$Q_{t+1}(x_t) = \rho_{t+1} [Q_{t+1}(x_t, \xi_{t+1})].$$
(6.395)

Here the cost-to-go functions $Q_{t+1}(x_t, \xi_{[t]}) = Q_{t+1}(x_t)$ do not depend on $\xi_{[t]}$, and the cost-to-go functions $Q_t(x_{t-1}, \xi_t)$ depend only on ξ_t rather than $\xi_{[t]}$. This can be shown by induction, going backward in time, in the same way as in the risk neutral case (see Remark 4 on page 67).

6.8.5 Time Consistency of Multiperiod Problems

Consider the setting of section 6.8.4. For $Z_{1,T} = Z_1 \times \cdots \times Z_T$ and a multiperiod risk measure $\varrho : Z_{1,T} \to \mathbb{R}$ consider the following risk averse optimization problem

$$\begin{array}{ll}
& \underset{x_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T}{\operatorname{Min}} & \varrho \left(f_1(x_1), f_2(\boldsymbol{x}_2), \dots, f_T(\boldsymbol{x}_T) \right) \\
& \text{s.t.} & x_1 \in \mathcal{X}_1, \ \boldsymbol{x}_t(\omega) \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\omega), \omega), \ t = 2, \dots, T,
\end{array}$$
(6.396)

where optimization is performed over policies $x_1, x_2, ..., x_T$, i.e., over \mathcal{F}_t -measurable $x_t : \Omega \to \mathbb{R}, t = 1, ..., T$, such that $f_t(x_t(\cdot), \cdot) \in \mathbb{Z}_t$. The feasibility constraints in (6.396) should be satisfied w.p.1. The nonanticipativity is enforced here by the \mathcal{F}_t -measurability of $x_t(\omega), t = 1, ..., T$.

For the multiperiod risk measure

$$\varrho(Z) := \mathbb{E}[Z_1 + \dots + Z_T], \ Z \in \mathcal{Z}, \tag{6.397}$$

the optimization problem (6.396) becomes the risk neutral multistage problem, which can be written in the form (3.4) or (3.33). In the risk neutral case, formulation (3.4) is equivalent to the corresponding nested formulation (3.1), which in turn can be represented by the dynamic programming equations (3.8)–(3.9). This equivalence is based on the tower property of the expectation operator:

$$\mathbb{E}[Z_1 + \ldots + Z_T] = Z_1 + \mathbb{E}_{|\mathcal{F}_1} \Big[Z_2 + \cdots + \mathbb{E}_{|\mathcal{F}_{T-2}} \big[Z_{T-1} + \mathbb{E}_{|\mathcal{F}_{T-1}} [Z_T] \big] \Big], \quad (6.398)$$

for $Z = (Z_1, ..., Z_T) \in \mathcal{Z}$. As it was discussed in sections 6.8.3 and 6.8.4, in the risk averse setting similar equivalence holds if ρ is given in the nested form

$$\varrho(Z) := Z_1 + \rho_{2|\mathcal{F}_1} \Big[Z_2 + \dots + \rho_{T-1|\mathcal{F}_{T-2}} \big[Z_{T-1} + \rho_{T|\mathcal{F}_{T-1}} [Z_T] \big] \Big], \ Z \in \mathcal{Z}.$$
(6.399)

Let us make the following observation. If we are currently at a certain stage of the system, then we know the past and hence it is reasonable to require that our decisions should be based on that information alone and should not involve unknown data. This is the nonanticipativity constraint, which was discussed in the previous sections. However, if we believe in the considered model, we also have an idea what can and what cannot happen in the future. Think, for example, about a scenario tree representing evolution of the data process. If we are currently at a certain node of that tree, representing current state of the system, we already know that only scenarios passing through this node can happen in the future. Consequently it is natural to require for an optimal policy to be optimal at every stage of the process conditional on the current state of the system looking into the future. Therefore, apart form the nonanticipativity constraint, it is also reasonable to think about the following concept.

Principle of conditional optimality. At every state of the system, optimality of our decisions should not depend on scenarios which we already know cannot happen in the future.

As we will see this principle may not hold even for seemingly natural formulations of risk averse multistage problems (see examples 6.84 and 6.86 below). However, if the nested formulation (6.399) is employed and the conditional risk measures are convex, i.e., satisfy the conditions (R'1)–(R'3), then this principle will be satisfied.

A well known quotation of Bellman [15], coming from his pioneering work on dynamic programming, asserts that: "An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision." In order to make this precise we have to define what do we mean by saying that an optimal policy remains optimal at every stage of the process conditional on an observed realization of the data process.

As far as the multiperiod optimization problem (6.396) is concerned, a (feasible) policy $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_T$, is optimal if it minimizes the objective function of this problem; there is no mentioning of what optimality means from a dynamical point of view. In the risk neutral case, when $\varrho(Z)$ is given as the expectation (6.397), the decomposable property (6.398) of the expectation operator gives a natural answer to this question - at stage t of the process the optimality is understood in terms of the conditional expectation $\mathbb{E}_{|\mathcal{F}_t}$. That is, the corresponding policy $\bar{x}_t, \ldots, \bar{x}_T$ is optimal for the problem (compare with (3.33))

$$\underset{\boldsymbol{x}_{t},\ldots,\boldsymbol{x}_{T}}{\underset{\text{s.t.}}{\underset{\boldsymbol{x}_{\tau}(\omega) \in \mathcal{X}_{\tau}(\boldsymbol{x}_{\tau-1}(\omega),\omega), \tau = t,\ldots,T,}}} \mathbb{E}_{|\mathcal{F}_{t}[f_{t}(\boldsymbol{x}_{t}) + \ldots + f_{T}(\boldsymbol{x}_{T})]}$$
(6.400)

A policy which is an optimal solution of the problem (3.33), remains optimal for the problem (6.400), t = 2, ..., T, *conditional* on \mathcal{F}_t and x_{t-1} . In that sense we say that the risk neutral formulation (3.33) is *time consistent*. For the risk averse formulations the situation is more delicate. We will discuss this below.

Problem (6.400) can be also written in the nested form

$$\underset{x_t \in \mathcal{X}_t(x_{t-1})}{\min} f_t(x_t) + \mathbb{E}_{|\mathcal{F}_t|} \Big[\inf_{x_{t+1} \in \mathcal{X}_{t+1}(x_t)} f_{t+1}(x_{t+1}) + \cdots + \mathbb{E}_{|\mathcal{F}_{[T-1]}|} \Big[\inf_{x_T \in \mathcal{X}_T(x_{T-1})} f_T(x_T) \Big] \Big],$$
(6.401)

conditional on \mathcal{F}_t and x_{t-1} (for meaning of the notation $\inf_{x_{t+1} \in \mathcal{X}_{t+1}(x_t)} f_{t+1}(x_{t+1})$ see (6.324)).

Remark 49. If we use the multistage formulation (3.4), in terms of the data process ξ_1, \ldots, ξ_T , then the corresponding optimization problem at stage t can be written as

$$\begin{array}{ll}
& \underset{\boldsymbol{x}_{t},...,\boldsymbol{x}_{T}}{\text{Min}} & \mathbb{E}_{|\xi_{[t]}} \left[f_{t}(\boldsymbol{x}_{t}(\xi_{[t]}),\xi_{t}) + \ldots + f_{T} \left(\boldsymbol{x}_{T}(\xi_{[T]}),\xi_{T} \right) \right] \\
& \text{s.t.} & \boldsymbol{x}_{\tau}(\xi_{[\tau]}) \in \mathcal{X}_{\tau}(\boldsymbol{x}_{\tau-1}(\xi_{[\tau-1]}),\xi_{\tau}), \ \tau = t, \ldots, T,
\end{array}$$
(6.402)

conditional on $\xi_{[t]} = (\xi_1, \dots, \xi_t)$ and x_{t-1} , for $t = 2, \dots, T$. Similar to (3.11), problem (6.402) can be written in the nested form

$$\underset{x_{t}\in\mathcal{X}_{t}(x_{t-1},\xi_{t})}{\min} f_{t}(x_{t},\xi_{t}) + \mathbb{E}_{|\xi_{[t]}} \Big[\inf_{x_{t+1}\in\mathcal{X}_{t+1}(x_{t},\xi_{t+1})} f_{t+1}(x_{t+1},\xi_{t+1}) + \dots + \mathbb{E}_{|\xi_{[T-1]}} \Big[\inf_{x_{T}\in\mathcal{X}_{T}(x_{T-1},\xi_{T})} f_{T}(x_{T},\xi_{T}) \Big] \Big],$$
(6.403)

conditional on ξ_1, \ldots, ξ_t and x_{t-1} . Again in that sense the risk neutral formulation (3.4) is time consistent.

Similar arguments can be applied if the multiperiod risk measure ρ is given in the nested form (6.364) for a chosen sequence $\rho_{t+1|\mathcal{F}_t} : \mathcal{Z}_{t+1} \to \mathcal{Z}_t, t = 1, ..., T - 1$, of conditional risk mappings. In that case $\rho(Z) = \overline{\rho}(Z_1 + \cdots + Z_T)$, where $\overline{\rho} = \rho_{2|\mathcal{F}_1} \circ \cdots \circ \rho_{T|\mathcal{F}_{T-1}}$ is the composite risk measure, and the corresponding optimization problem can be written as in (6.381). Because of the decomposable form (6.382) of the composite risk measure $\overline{\rho}$, problem (6.381) can be also written in the nested form (6.380). Therefore an optimal policy for problem (6.381) is also optimal, at stage $t = 2, \ldots, T$, for the problem

$$\begin{array}{ll}
& \underset{\boldsymbol{x}_{t},\ldots,\boldsymbol{x}_{T}}{\text{Min}} & \bar{\rho}_{t,T} \left[f_{t}(\boldsymbol{x}_{t}) + \cdots + f_{T}\left(\boldsymbol{x}_{T}\right) \right] \\
& \text{s.t.} & \boldsymbol{x}_{\tau}(\omega) \in \mathcal{X}_{\tau}(\boldsymbol{x}_{\tau-1}(\omega), \omega), \ \tau = t, \ldots, T,
\end{array}$$
(6.404)

conditional on \mathcal{F}_t and x_{t-1} . (This follows from the optimality conditions (6.387) associated with the corresponding dynamic programming equations.) Here $\bar{\rho}_{t,T} := \rho_{t+1|\mathcal{F}_t} \circ \cdots \circ \rho_{T|\mathcal{F}_{T-1}} : \mathcal{Z}_T \to \mathcal{Z}_t$. In that sense the nested formulation of risk averse problems is time consistent. Unfortunately, as it was discussed in section 6.8.3, the composite risk measure $\bar{\rho}$, and the composite mappings $\bar{\rho}_{t,T}$, are not given in a closed form. Similar to (6.379), problem (6.404) can be written in the following nested form

$$\underset{x_{t}\in\mathcal{X}_{t}(x_{t-1})}{\min} f_{t}(x_{t}) + \rho_{t+1|\mathcal{F}_{t}} \bigg[\inf_{x_{t+1}\in\mathcal{X}_{t+1}(x_{t})} f_{t+1}(x_{t+1}) + \dots + \rho_{T|\mathcal{F}_{T-1}} \bigg[\inf_{x_{T}\in\mathcal{X}_{T}(x_{T-1})} f_{T}(x_{T}) \bigg] \bigg].$$
(6.405)

Now let us give a formal definition of time consistency for the risk averse multiperiod problem (6.396). We assume that with the multiperiod problem (6.396) is associated a dynamic measure of risk (see Definition 6.75), that is, a sequence of mappings $\rho_{t,T}$: $\mathcal{Z}_t \times \cdots \times \mathcal{Z}_T \to \mathcal{Z}_t, t = 1, \ldots, T$, satisfying (6.365).

In addition to the condition of time consistency of Definition 6.76, we also consider the following condition of *strict* time consistency.

Definition 6.79. A time consistent dynamic risk measure $\{\varrho_{t,T}\}_{t=1}^{T}$ is called strictly time consistent if for all $1 \le \tau < \theta \le T$ and all sequences $Z, W \in \mathcal{Z}_{\tau,T}$ the conditions

$$Z_k = W_k, \ k = \tau, \dots, \theta - 1 \quad and \quad \varrho_{\theta,T}(Z_\theta, \dots, Z_T) \prec \varrho_{\theta,T}(W_\theta, \dots, W_T) \quad (6.406)$$

imply that

$$\varrho_{\tau,T}(Z_{\tau},\ldots,Z_T) \prec \varrho_{\tau,T}(W_{\tau},\ldots,W_T).$$
(6.407)

Remark 50. The condition of time consistency of Definition 6.76 holds for multiperiod mappings of the nested form

$$\varrho_{t,T}(Z_t, \dots, Z_T) := \bar{\rho}_{t,T}(Z_t + \dots + Z_T), \tag{6.408}$$

where

$$\bar{\rho}_{t,T}(Z_t + \dots + Z_T) := \rho_{t+1|\mathcal{F}_t} \circ \dots \circ \rho_{T|\mathcal{F}_{T-1}}(Z_t + \dots + Z_T) \\
= Z_t + \rho_{t+1|\mathcal{F}_t} \left[Z_{t+1} + \dots + \rho_{T|\mathcal{F}_{T-1}}[Z_T] \right],$$
(6.409)

and $\rho_{t+1|\mathcal{F}_t} : \mathcal{Z}_{t+1} \to \mathcal{Z}_t, t = 1, ..., T - 1$, is a sequence of conditional risk mappings. (Compare this with Remark 47 on page 390.)

Indeed, let $Z, W \in \mathbb{Z}_{\tau} \times \cdots \times \mathbb{Z}_{T}$ satisfy condition (6.366). Then

$$\varrho_{\tau,T}(Z_{\tau},\ldots,Z_{T}) = Z_{\tau} + \rho_{\tau+1|\mathcal{F}_{t}} \left[Z_{\tau+1} + \cdots + \rho_{T|\mathcal{F}_{T-1}}[Z_{T}] \right] \\
= \bar{\rho}_{\tau,T}(Z_{\tau} + \cdots + Z_{\theta-1} + \bar{\rho}_{\theta,T}(Z_{\theta} + \cdots + Z_{T})) \\
\leq \bar{\rho}_{\tau,T}(Z_{\tau} + \cdots + Z_{\theta-1} + \bar{\rho}_{\theta,T}(W_{\theta} + \cdots + W_{T})) \\
= \bar{\rho}_{\tau,T}(W_{\tau} + \cdots + W_{\theta-1} + \bar{\rho}_{\theta,T}(W_{\theta} + \cdots + W_{T})) \\
= \varrho_{\tau,T}(W_{\tau},\ldots,W_{T}),$$

where the inequality is implied by the monotonicity condition (R'2).

In order to have the stronger conclusion (6.407) of Definition 6.79, we need the strict monotonicity condition $(\mathbb{R}^{*'2})$: $\{Z \succ Z'\} \Rightarrow \{\rho(Z) \succ \rho(Z')\}$ (see Remark 33 on page 326 for a discussion of strict monotonicity). That is, strict time consistency holds under time consistency with the strict monotonicity condition $(\mathbb{R}^{*'2})$.

Consider the following optimization problem

$$\begin{array}{ll}
& \underset{x_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T}{\text{Min}} & \varrho_{1,T} \left(f_1(x_1), f_2(\boldsymbol{x}_2), \dots, f_T(\boldsymbol{x}_T) \right) \\
& \text{s.t.} & x_1 \in \mathcal{X}_1, \ \boldsymbol{x}_t(\omega) \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\omega), \omega), \ t = 2, \dots, T.
\end{array}$$
(6.410)

As before the optimization is performed over policies x_t , t = 1, ..., T, adapted to the filtration $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T$.

Proposition 6.80. Let $\varrho_{t,T} : \mathcal{Z}_t \times \cdots \times \mathcal{Z}_T \to \mathcal{Z}_t$, $t = 1, \ldots, T$, be a dynamic risk measure, and $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_T$, be an optimal policy of the problem (6.410). Assume that at least one of the following conditions holds:

- (i) the dynamic risk measure $\{\varrho_{t,T}\}$ is strictly time consistent,
- (ii) the dynamic risk measure $\{\varrho_{t,T}\}$ is time consistent and the policy $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_T$ is the unique optimal solution of the problem (6.410).

Then for t = 2, ..., T, conditional on \mathcal{F}_t and \bar{x}_{t-1} , the policy $\bar{x}_t, ..., \bar{x}_T$, is optimal for the problem

$$\begin{array}{ll} \underset{\boldsymbol{x}_{t},\dots,\boldsymbol{x}_{T}}{\operatorname{Min}} & \varrho_{t,T}\left(f_{t}(\boldsymbol{x}_{t}),\dots,f_{T}\left(\boldsymbol{x}_{T}\right)\right) \\ \text{s.t.} & \boldsymbol{x}_{\tau}(\omega) \in \mathcal{X}_{\tau}(\boldsymbol{x}_{\tau-1}(\omega),\omega), \ \tau = t,\dots,T. \end{array}$$
(6.411)

Proof. Suppose condition (i) holds and problem (6.411) has a feasible policy $\tilde{x}_t, \ldots, \tilde{x}_T$, such that

$$\varrho_{t,T}(f_t(\tilde{\boldsymbol{x}}_t),\ldots,f_T(\tilde{\boldsymbol{x}}_T)) \prec \varrho_{t,T}(f_t(\bar{\boldsymbol{x}}_t),\ldots,f_T(\bar{\boldsymbol{x}}_T)).$$
(6.412)

By strict time consistency, this implies that

$$\varrho_{1,T}(f_1(\bar{x}_1),\ldots,f_{t-1}(\bar{x}_{t-1}),f_t(\tilde{x}_t),\ldots,f_T(\tilde{x}_T)))
< \varrho_{1,T}(f_1(\bar{x}_1),\ldots,f_{t-1}(\bar{x}_{t-1}),f_t(\bar{x}_t),\ldots,f_T(\bar{x}_T)),$$
(6.413)

which contradicts optimality of the policy $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_T$. This shows that $\bar{x}_t, \ldots, \bar{x}_T$ is optimal for the problem (6.411).

Now suppose condition (ii). If we replace the strict inequality sign " \prec " by " \preceq " in the inequality (6.412), then owing to time consistency we can conclude that (6.413) holds with "<" replaced by " \leq ". That is, we obtain that $\bar{x}_1, \ldots, \bar{x}_{t-1}, \tilde{x}_t, \ldots, \tilde{x}_T$ is an optimal solution of problem (6.410). By uniqueness of the optimal policy $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_T$ we obtain the required conclusion. \Box

Remark 51. It might be tempting to impose the following "forward" condition on the dynamic risk measure $\{\varrho_{t,T}\}$:

(F) For all $1 \le \tau < \theta \le T$ and $Z, Z' \in \mathcal{Z}_{\tau} \times \cdots \times \mathcal{Z}_{T}$, the conditions

$$Z_t = Z'_t, \ t = \tau, \dots, \theta - 1, \text{ and } \varrho_{\tau,T}(Z_\tau, \dots, Z_T) \preceq \varrho_{\tau,T}(Z'_\tau, \dots, Z'_T) \quad (6.414)$$

imply that

$$\varrho_{\theta,T}(Z_{\theta},\ldots,Z_T) \preceq \varrho_{\theta,T}(Z'_{\theta},\ldots,Z'_T).$$
(6.415)

Note that for an optimal solution $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_T$, of the problem (6.410) and a feasible solution $\tilde{x}_t, \dots, \tilde{x}_T$, of the problem (6.411) we have that

$$\varrho_{1,T}\left(f_1(\bar{x}_1),\ldots,f_{t-1}(\bar{x}_{t-1}),f_t(\tilde{x}_t),\ldots,f_T(\tilde{x}_T)\right) \\
\geq \varrho_{1,T}\left(f_1(\bar{x}_1),\ldots,f_{t-1}(\bar{x}_{t-1}),f_t(\bar{x}_t),\ldots,f_T(\bar{x}_T)\right).$$

Under condition (F) this would imply that

$$\varrho_{t,T}(f_t(\tilde{\boldsymbol{x}}_t),\ldots,f_T(\tilde{\boldsymbol{x}}_T)) \geq \varrho_{t,T}(f_t(\bar{\boldsymbol{x}}_t),\ldots,f_T(\bar{\boldsymbol{x}}_T)),$$

and hence that $\bar{x}_t, \ldots, \bar{x}_T$, is an optimal solution of the problem (6.411). That is, condition (F) implies the desired property of problem (6.411).

Unfortunately, condition (F) does not hold even for a sequence of conditional expectation mappings

$$\varrho_{t,T}(Z_t,\ldots,Z_T) := \mathbb{E}_{|\mathcal{F}_t}[Z_t + \cdots + Z_T].$$
(6.416)

Indeed, for $\tau = 1$ condition (6.414) becomes

$$\mathbb{E}[Z_1 + \dots + Z_{\theta-1} + Z_\theta + \dots + Z_T] \le \mathbb{E}[Z_1 + \dots + Z_{\theta-1} + Z_\theta' + \dots + Z_T'],$$

or equivalently

$$\mathbb{E}[Z_{\theta} + \dots + Z_T] \le \mathbb{E}[Z'_{\theta} + \dots + Z'_T].$$

However, this inequality does not imply the required *pointwise* inequality

$$\mathbb{E}_{|\mathcal{F}_{\theta}}[Z_{\theta} + \dots + Z_T] \preceq \mathbb{E}_{|\mathcal{F}_{\theta}}[Z'_{\theta} + \dots + Z'_T]$$

of condition (6.415).

Definition 6.81. We say that an optimal policy $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_T$, of the multiperiod problem (6.396) is time consistent if for $t = 1, \ldots, T$, the policy $\bar{x}_t, \ldots, \bar{x}_T$ is optimal for the problem (6.411), conditional on \mathcal{F}_t and \bar{x}_{t-1} for $t = 2, \ldots, T$.

That is, let $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_T$, be an optimal policy of the multiperiod problem (6.396). It is said that this policy is time consistent if it is also optimal for the problem (6.410) and moreover for stages $t = 2, \ldots, T$, the policy $\bar{x}_t, \ldots, \bar{x}_T$, is optimal for the problem (6.411) *conditional* on \mathcal{F}_t and \bar{x}_{t-1} . In this formulation the principle of conditional optimality holds automatically.

Definition 6.82. We say that the multiperiod problem (6.396) is time consistent if its every optimal policy is time consistent. We say that the multiperiod problem (6.396) is weakly time consistent if at least one of its optimal policies is time consistent.

Note that definition 6.82 implicitly assumes that the multiperiod problem (6.396) possesses optimal solutions (possibly more than one), otherwise by the definition problem (6.396) is time consistent with respect to any choice of the multiperiod mappings. If the multiperiod problem (6.396) has only one optimal solution, then of course the concepts of time consistency and weak time consistency do coincide. Note also that by Proposition

6.80, the property of strict time consistency of the dynamic risk measure is *sufficient* for time consistency of problem (6.410) (i.e., for $\rho = \rho_{1,T}$).

The property of time consistency of a dynamic risk measure is sufficient for weak time consistency of problem (6.396).

Proposition 6.83. Let $\varrho_{t,T} : Z_t \times \cdots \times Z_T \to Z_t$, $t = 1, \dots, T$, be a time consistent dynamic risk measure. If for every optimal policy $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_T$ of problem (6.410) and for every $t = 2, \dots, T$ the problems (6.411) conditioned on $x_{t-1} = \bar{x}_{t-1}$, have an optimal solution, then problem (6.396) is weakly time consistent.

Proof. Let $\bar{x}_1^1, \bar{x}_2^1, \ldots, \bar{x}_T^1$ be optimal policy of problem (6.410). Consider problem (6.411) for t = 2, conditioned on $x_1 = \bar{x}_1^1$. Let $\bar{x}_2^2, \ldots, \bar{x}_T^2$ denote its optimal solution. By optimality,

$$\varrho_{2,T}(f_2(\bar{\boldsymbol{x}}_2^2),\ldots,f_T(\bar{\boldsymbol{x}}_T^2)) \preceq \varrho_{2,T}(f_2(\bar{\boldsymbol{x}}_2^1),\ldots,f_T(\bar{\boldsymbol{x}}_T^1)).$$

Owing to time consistency of the dynamic risk measure,

$$\varrho_{1,T}(f_1(\bar{\boldsymbol{x}}_1^1), f_2(\bar{\boldsymbol{x}}_2^2), \dots, f_T(\bar{\boldsymbol{x}}_T^2)) \preceq \varrho_{1,T}(f_1(\bar{\boldsymbol{x}}_1^1), f_2(\bar{\boldsymbol{x}}_2^1), \dots, f_T(\bar{\boldsymbol{x}}_T^1)).$$

It follows that $\bar{x}_1^1, \bar{x}_2^2, \ldots, \bar{x}_T^2$ is also an optimal solution of problem (6.410). Consider now problem (6.411) for t = 3, conditioned on $x_2 = \bar{x}_2^2$. Let $\bar{x}_3^3, \ldots, \bar{x}_T^3$ denote its optimal solution. Using time consistency of the risk measure again, we observe that

$$\varrho_{1,T}\left(f_1(\bar{\boldsymbol{x}}_1^1), f_2(\bar{\boldsymbol{x}}_2^2), f_3(\bar{\boldsymbol{x}}_3^3), \dots, f_T(\bar{\boldsymbol{x}}_T^3)\right) \leq \varrho_{1,T}\left(f_1(\bar{\boldsymbol{x}}_1^1), f_2(\bar{\boldsymbol{x}}_2^2), f_3(\bar{\boldsymbol{x}}_3^2), \dots, f_T(\bar{\boldsymbol{x}}_T^2)\right).$$

It follows that $\bar{x}_1^1, \bar{x}_2^2, \bar{x}_3^3, \dots, \bar{x}_T^3$ is also an optimal solution of problem (6.410). Continuing in this way, we construct an optimal time-consistent policy $\bar{x}_1^1, \bar{x}_2^2, \bar{x}_3^3, \dots, \bar{x}_T^T$.

Let us emphasize that the concept of time consistency of a problem depends on the choice of the multiperiod mappings $\rho_{t,T}$, t = 1, ..., T. In the risk-neutral case, with ρ of the form (6.397), time consistency follows with the natural choice of conditional multiperiod mappings:

$$\varrho_{t,T}(Z_t,\ldots,Z_T) := \mathbb{E}_{|\mathcal{F}_t}(Z_t+\cdots+Z_T).$$
(6.417)

In the case of the nested risk averse formulation (6.380), the corresponding multiperiod problem (6.381) is time consistent with respect to the multiperiod mappings

$$\varrho_{t,T}(Z_t, \dots, Z_T) := \bar{\rho}_{t,T}(Z_t + \dots + Z_T) \\
= Z_t + \rho_{t+1}|_{\mathcal{F}_t} \left[Z_{t+1} + \dots + \rho_T|_{\mathcal{F}_{T-1}}[Z_T] \right],$$
(6.418)

where $\bar{\rho}_{t,T} := \rho_{t+1|\mathcal{F}_t} \circ \cdots \circ \rho_{T|\mathcal{F}_{T-1}}$. For a general multiperiod problem, the choice of "natural" multiperiod mappings may not be obvious. We discuss this later (see section 6.8.6, in particular).

Example 6.84 Consider a multiperiod risk measure ρ of the form

$$\varrho(Z_1, \dots, Z_T) := Z_1 + \sum_{t=2}^T \rho_t(Z_t), \ (Z_1, \dots, Z_T) \in \mathcal{Z},$$
(6.419)

with $\rho_t := \mathsf{AV}@\mathsf{R}_{\alpha}, \alpha \in (0, 1), t = 2, \dots, T$. That is,

$$\varrho(Z_1,\ldots,Z_T) := Z_1 + \sum_{t=2}^T \mathsf{AV} @\mathsf{R}_{\alpha}(Z_t).$$

By using definition (6.23) of AV@R $_{\alpha}$ we can write

$$\varrho(Z_1, \dots, Z_T) := \inf_{r_2, \dots, r_T} \mathbb{E}\left\{ Z_1 + \sum_{t=2}^T \left(r_t + \alpha^{-1} [Z_t - r_t]_+ \right) \right\}.$$
 (6.420)

Thus the corresponding optimization problem (6.396) can be formulated as

$$\underset{r,x_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_T}{\min} \quad f_1(x_1) + \sum_{t=2}^T r_t + \mathbb{E}\left\{\sum_{t=2}^T \alpha^{-1}[f_t(\boldsymbol{x}_T(\omega),\omega) - r_t]_+\right\} \quad (6.421)$$
s.t.
$$x_1 \in \mathcal{X}_1, \ \boldsymbol{x}_t(\omega) \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\omega),\omega), \ t = 2,\ldots,T,$$

with additional variables $r = (r_2, ..., r_T)$. Problem (6.421) can be viewed as a standard multistage stochastic program with (r, x_1) being first stage decision variables.

The corresponding dynamic programming equations can be written with the cost-togo function $Q_t(x_{t-1}, r_t, \ldots, r_T, \omega)$, for $t = 2, \ldots, T$, given by the optimal value of the problem

$$\begin{aligned}
& \underset{x_t}{\operatorname{Min}} \quad \alpha^{-1}[f_t(x_t,\omega) - r_t]_+ + \mathbb{E}_{|\mathcal{F}_t}\left[Q_{t+1}(x_t, r_t, \dots, r_T, \omega)\right] \\
& \text{s.t.} \quad x_t \in \mathcal{X}_t(x_{t-1}, \omega),
\end{aligned}$$
(6.422)

where $Q_{T+1}(\cdot, \cdot) \equiv 0$. At the first stage the following problem should be solved

$$\min_{x_1 \in \mathcal{X}_1, r \in \mathbb{R}^{T-1}} r_2 + \dots + r_T + f(x_1) + \mathbb{E}[Q_2(x_1, r, \omega)].$$
(6.423)

Although it was possible to write dynamic programming equations for problem (6.421), note that decision variables r_2, \ldots, r_T are decided at the first stage and their optimal values depend on all scenarios starting at the root node at stage t = 1. Consequently, the optimal decisions at later stages depend on scenarios other than following a considered node. That is, here *the principle of conditional optimality does not hold.*

6.8.6 Minimax Approach to Risk-Averse Multistage Programming

Again, we use the framework of the previous sections 6.8.4–6.8.5. Consider a (nonempty) set \mathfrak{M} of probability measures P, defined on the sample space (Ω, \mathcal{F}) , and the functional

$$\hat{\rho}(Z_T) := \sup_{P \in \mathfrak{M}} \mathbb{E}_P[Z_T]$$
(6.424)

defined on a linear space Z_T of random variables $Z_T : \Omega \to \mathbb{R}$. If every $P \in \mathfrak{M}$ is absolutely continuous with respect to the reference probability measure P_0 , then $\hat{\rho} : Z_T \to \mathbb{R}$ becomes a coherent risk measure.

With the functional $\hat{\rho}$ is associated the following minimax (distributionally robust) stochastic programming problem

$$\begin{array}{ll}
\underset{x_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T}{\operatorname{Min}} & \sup_{P \in \mathfrak{M}} \mathbb{E}_P \left[f_1(x_1) + f_2(\boldsymbol{x}_2) + \dots + f_T(\boldsymbol{x}_T) \right] \\
\text{s.t.} & x_1 \in \mathcal{X}_1, \ \boldsymbol{x}_t(\omega) \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\omega), \omega), \ t = 2, \dots, T.
\end{array}$$
(6.425)

Of course, if the set \mathfrak{M} is a singleton, then (6.425) becomes the risk-neutral formulation (3.33).

In order to construct a nested counterpart of the minimax problem (6.425) we proceed as follows. We can write

$$\mathbb{E}_{P}[Z_{T}] = \mathbb{E}_{P|\mathcal{F}_{1}}\Big[\mathbb{E}_{P|\mathcal{F}_{2}}\Big[\cdots\mathbb{E}_{P|\mathcal{F}_{T-1}}[Z_{T}]\Big]\Big],\tag{6.426}$$

and hence

$$\hat{\rho}(Z_T) = \sup_{P \in \mathfrak{M}} \mathbb{E}_{P|\mathcal{F}_1} \Big[\mathbb{E}_{P|\mathcal{F}_2} \big[\cdots \mathbb{E}_{P|\mathcal{F}_{T-1}}[Z_T] \big] \Big] \\ \leq \sup_{P \in \mathfrak{M}} \mathbb{E}_{P|\mathcal{F}_1} \Big[\sup_{P \in \mathfrak{M}} \mathbb{E}_{P|\mathcal{F}_2} \big[\cdots \sup_{P \in \mathfrak{M}} \mathbb{E}_{P|\mathcal{F}_{T-1}}[Z_T] \big] \Big].$$
(6.427)

That is,

$$\hat{\rho}(Z_T) \le \rho_{2|\mathcal{F}_1} \Big[\rho_{3|\mathcal{F}_2} \big[\cdots \rho_{T|\mathcal{F}_{T-1}} [Z_T] \big] \Big], \tag{6.428}$$

where for $t = 2, \ldots, T$,

$$\rho_{t|\mathcal{F}_{t-1}}[Z_T] := \sup_{P \in \mathfrak{M}} \mathbb{E}_{P|\mathcal{F}_{t-1}}[Z_t].$$
(6.429)

For $(Z_1, \ldots, Z_T) \in \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_T$, with $Z_t, t = 1, \ldots, T$, being appropriate linear spaces of \mathcal{F}_t -measurable random variables, we can write (6.428) as

$$\hat{\rho}(Z_1 + \dots + Z_T) \le Z_1 + \rho_{2|\mathcal{F}_1} \Big[Z_2 + \rho_{3|\mathcal{F}_2} \big[Z_3 + \dots + \rho_{T|\mathcal{F}_{T-1}} [Z_T] \big] \Big].$$
(6.430)

This suggests the corresponding nested counterpart of the minimax problem (6.425) of the form (6.380), that is

$$\underset{x_{1},\ldots,\boldsymbol{x}_{T}}{\operatorname{Min}} \quad f_{1}(x_{1}) + \underset{P\in\mathfrak{M}}{\sup} \mathbb{E}_{P|\mathcal{F}_{1}} \left[f_{2}(\boldsymbol{x}_{2}(\omega),\omega) + \underset{P\in\mathfrak{M}}{\sup} \mathbb{E}_{P|\mathcal{F}_{2}} \left[f_{3}(\boldsymbol{x}_{3}(\omega),\omega) + \underset{P\in\mathfrak{M}}{\sup} \mathbb{E}_{P|\mathcal{F}_{T-1}} \left[f_{T}(\boldsymbol{x}_{T}(\omega),\omega) \right] \right] \right]$$
s.t.
$$x_{1} \in \mathcal{X}_{1}, \ \boldsymbol{x}_{t}(\omega) \in \mathcal{X}_{t}(\boldsymbol{x}_{t-1}(\omega),\omega), \ t = 2,\ldots,T.$$

$$(6.431)$$

It follows from (6.430) that the optimal value of the nested problem (6.431) is grater than or equal to the optimal value of the minimax problem (6.425).

In accordance with Definition 6.82 we say that the minimax problem (6.425) is (weakly) *time consistent* if its every (at least one) optimal solution policy is also optimal for the nested problem (6.431). The dynamic programming equations, of the form (6.383)–(6.385), for the nested formulation (6.431) can be written as

$$Q_t(x_{t-1},\omega) = \inf_{x_t \in \mathcal{X}_t(x_{t-1},\omega)} \left\{ f_t(x_t,\omega) + \sup_{P \in \mathfrak{M}} \mathbb{E}_{P|\mathcal{F}_t} \left[Q_{t+1}(x_t) \right](\omega) \right\}, \quad (6.432)$$

with $Q_{T+1}(\cdot, \cdot) \equiv 0$. That is, problem (6.425) is time consistent if its every optimal policy $\bar{x}_1, \ldots, \bar{x}_T$ satisfies the optimality conditions: \bar{x}_1 is an optimal solution of the corresponding first stage problem and for $t = 2, \ldots, T$,

$$\bar{\boldsymbol{x}}_{t}(\omega) \in \operatorname*{argmin}_{x_{t}\in\mathcal{X}_{t}(\bar{\boldsymbol{x}}_{t-1}(\omega),\omega)} \left\{ f_{t}(x_{t},\omega) + \sup_{P\in\mathfrak{M}} \mathbb{E}_{P|\mathcal{F}_{t}}\left[Q_{t+1}(x_{t})\right](\omega) \right\}, \text{ w.p.1.} \quad (6.433)$$

The nested formulation and the corresponding dynamic programming equations are simplified in the stagewise independent case. That is, let $\xi_t = \xi_t(\omega)$, t = 1, ..., T, be a data process and $\mathcal{F}_t := \sigma(\xi_1, ..., \xi_t)$, t = 1, ..., T, be the corresponding filtration. Recall that ξ_1 is deterministic and $\mathcal{F}_1 = \{\emptyset, \Omega\}$. Suppose that

$$\mathfrak{M} := \{ P = P_1 \times \dots \times P_T : P_t \in \mathcal{M}_t, \ t = 1, \dots, T \},$$
(6.434)

where \mathcal{M}_t is a set of probability distributions (measures) of random vector $\xi_t = \xi_t(\omega)$, $t = 1, \ldots, T$. That is, the set \mathfrak{M} consists of probability distributions of (ξ_1, \ldots, ξ_T) with mutually independent components.

In that case the dynamic programming equations for the corresponding nested formulation take the form (compare with (3.18)-(3.19) and (6.394)-(6.395))

$$Q_t(x_{t-1},\xi_t) = \inf_{x_t \in \mathcal{X}_t(x_{t-1},\xi_t)} \left\{ f_t(x_t,\xi_t) + \sup_{P_{t+1} \in \mathcal{M}_{t+1}} \mathbb{E}_{P_{t+1}} \left[Q_{t+1}(x_t,\xi_{t+1}) \right] \right\},$$
(6.435)

with the optimality conditions for an optimal policy $\bar{x}_t = \bar{x}_t(\xi_{[t]}), t = 1, ..., T$,

$$\bar{\boldsymbol{x}}_{t} \in \operatorname*{argmin}_{x_{t} \in \mathcal{X}_{t}(\bar{\boldsymbol{x}}_{t-1},\xi_{t})} \Big\{ f_{t}(x_{t},\xi_{t}) + \underset{P_{t+1} \in \mathcal{M}_{t+1}}{\sup} \mathbb{E}_{P_{t+1}} \left[Q_{t+1}(x_{t},\xi_{t+1}) \right] \Big\}, \text{ w.p.1.}$$
(6.436)

Example 6.85 (Distributionally robust inventory model) Consider the inventory model discussed in section 1.2.3. Suppose that the demand process D_1, \ldots, D_T is stagewise independent and let \mathcal{M}_t be a specified set of probability distributions of the demand D_t , $t = 1, \ldots, T$. The minimax (distributionally robust) counterpart of the inventory problem (1.17) becomes

$$\begin{array}{ll}
& \min_{x_t \ge y_t} & \sup_{P \in \mathfrak{M}} \mathbb{E}_P \Big\{ \sum_{t=1}^T c_t (x_t - y_t) + \psi_t (x_t, D_t) \Big\} \\
& \text{s.t.} & y_{t+1} = x_t - D_t, \ t = 1, \dots, T - 1, \end{array}$$
(6.437)

where $\mathfrak{M} := \{ P = P_1 \times \cdots \times P_T : P_t \in \mathcal{M}_t \}$ and

$$\psi_t(x_t, d_t) := b_t [d_t - x_t]_+ + h_t [x_t - d_t]_+.$$

For t = T, ..., 2, the cost-to-go function $Q_t(y_t)$, of the dynamic programming equations of the nested counterpart of the minimax problem (6.437), is given by the optimal value of the problem (compare with (1.20))

$$\min_{x_t \ge y_t} \left\{ c_t(x_t - y_t) + \sup_{P_t \in \mathcal{M}_t} \mathbb{E}_{P_t} \Big[\psi_t(x_t, D_t) + Q_{t+1} \left(x_t - D_t \right) \Big] \right\},$$
(6.438)

where $Q_{T+1}(\cdot) \equiv 0$. At the first stage we need to solve problem

$$\underset{x_1 \ge y_1}{\operatorname{Min}} c_1(x_1 - y_1) + \underset{P_1 \in \mathcal{M}_1}{\sup} \mathbb{E}_{P_1} \left\{ \psi_1(x_1, D_1) + Q_2 \left(x_1 - D_1 \right) \right\}.$$
(6.439)

Suppose now that for the distribution of the demand D_t we only specify its range (support) given by a (finite) interval $[\alpha_t, \beta_t] \subset \mathbb{R}_+$ and its mean $\mu_t \in [\alpha_t, \beta_t]$. That is,

$$\mathcal{M}_t := \{ P_t : \text{supp}(P_t) = [\alpha_t, \beta_t], \ \mathbb{E}_{P_t}[D_t] = \mu_t \}, \ t = 1, \dots, T.$$
(6.440)

Since the functions $\psi_t(\cdot, d_t)$ and $Q_{t+1}(\cdot)$ are convex, the maximum over $P_t \in \mathcal{M}_t$ in problem (6.438) is attained at the probability measure

$$\bar{P}_t = p_t \delta(\alpha_t) + (1 - p_t) \delta(\beta_t)$$
, where $p_t := \frac{\beta_t - \mu_t}{\beta_t - \alpha_t}$

supported on the end points of the interval $[\alpha_t, \beta_t]$ (see Theorem 6.69). This measure \bar{P}_t does not depend on x_t . Therefore the nested problem is equivalent to the risk neutral inventory problem

$$\begin{array}{ll}
& \min_{x_t \ge y_t} & \mathbb{E}_{\bar{P}} \Big\{ \sum_{t=1}^T c_t(x_t - y_t) + \psi_t(x_t, D_t) \Big\} \\
& \text{s.t.} & y_{t+1} = x_t - D_t, \ t = 1, \dots, T - 1, \end{array}$$
(6.441)

corresponding to the probability measure $\overline{P} := \overline{P}_1 \times \cdots \times \overline{P}_T$ of the demand (D_1, \ldots, D_T) . It could be noted that problem (6.441) can be represented by a scenario tree with each node at every stage (except the last stage) having two children nodes.

Since $P \in \mathfrak{M}$ it follows that the optimal value of the distributionally robust problem (6.437) is greater than or equal to the optimal value the problem (6.441). On the other hand, as it was pointed above, the optimal value of the nested problem is always greater than or equal to the optimal value of the corresponding distributionally robust problem. It follows that the distributionally robust problem (6.437), its nested counterpart problem and the risk neutral problem (6.441) are equivalent in the sense that their optimal values are equal to each other and they have the same optimal solutions. As a consequence we obtain that in the present case the minimax (distributionally robust) problem (6.437) is *time consistent*.

The situation changes dramatically if in the definition of the sets \mathcal{M}_t we also specify second order moments (variances) of the respective demands D_t . In that case it is possible to construct examples where the corresponding minimax problem (6.437) is time consistent, only weakly time consistent or not time consistent at all. We refer to [274] for a discussion of this case.

6.8.7 Portfolio Selection and Inventory Model Examples

Example 6.86 (Risk Averse Multistage Portfolio Selection) Consider the multistage portfolio selection problem discussed in section 1.4.2. A nested formulation of risk averse multistage portfolio selection can be written as

$$\operatorname{Min}\left\{\bar{\rho}(-W_T) := \rho_1 \left[\cdots \rho_{T-1|W_{T-2}} \left[\rho_{T|W_{T-1}} \left[-W_T \right] \right] \right] \right\}$$

s.t. $W_{t+1} = \sum_{i=1}^n \xi_{i,t+1} x_{it}, \sum_{i=1}^n x_{it} = W_t, \ x_t \ge 0, \ t = 0, \dots, T-1.$ (6.442)

We use here conditional risk mappings formulated in terms of the respective conditional expectations, like the conditional AV@R (see (6.388)) and conditional mean semideviations (see (6.390)), and the notation $\rho_{t|W_{t-1}}$ stands for a conditional risk mapping defined in terms of the respective conditional expectations given W_{t-1} . By $\rho_t(\cdot)$ we denote the corresponding (unconditional) risk measures. For example, to the conditional AV@R_{α}($\cdot|\xi_{[t-1]}$) corresponds the respective (unconditional) AV@R_{α}(\cdot). If we set $\rho_{t|W_{t-1}} := \mathbb{E}_{|W_{t-1}}$, $t = 1, \ldots, T$, then since

$$\mathbb{E}\Big[\cdots\mathbb{E}\Big[\mathbb{E}\left[-W_T|W_{T-1}\right]\Big|W_{T-2}\Big]\Big]=\mathbb{E}\left[-W_T\right],$$

we obtain the risk neutral formulation. Note also that in order to formulate this as a minimization, rather than a maximization, problem we changed the sign of ξ_{it} .

Suppose that the random process ξ_t is *stagewise independent*. Let us write dynamic programming equations. At the last stage we have to solve problem

$$\underset{x_{T-1} \ge 0, W_T}{\min} \rho_{T|W_{T-1}} \left[-W_T \right]
s.t. W_T = \sum_{i=1}^n \xi_{iT} x_{i,T-1}, \sum_{i=1}^n x_{i,T-1} = W_{T-1}.$$
(6.443)

Since W_{T-1} is a function of $\xi_{[T-1]}$, by the stagewise independence we have that ξ_T , and hence W_T , are independent of W_{T-1} . It follows by positive homogeneity of ρ_T that the optimal value of (6.443) is $Q_{T-1}(W_{T-1}) = W_{T-1}\nu_{T-1}$, where ν_{T-1} is the optimal value of

$$\underset{x_{T-1} \ge 0, W_T}{\min} \rho_T[-W_T]$$
s.t. $W_T = \sum_{i=1}^n \xi_{iT} x_{i,T-1}, \sum_{i=1}^n x_{i,T-1} = 1,$
(6.444)

and an optimal solution of (6.443) is $\bar{x}_{T-1}(W_{T-1}) = W_{T-1}x_{T-1}^*$, where x_{T-1}^* is an optimal solution of (6.444). And so on we obtain that the optimal policy $\bar{x}_t(W_t)$ here is *myopic*. That is, $\bar{x}_t(W_t) = W_t x_t^*$, where x_t^* is an optimal solution of

$$\underset{x_t \ge 0, W_{t+1}}{\min} \rho_{t+1}[-W_{t+1}]$$
s.t. $W_{t+1} = \sum_{i=1}^{n} \xi_{i,t+1} x_{it}, \ \sum_{i=1}^{n} x_{it} = 1$
(6.445)

(compare with Remark 1.4.3 on page 21). Note that the composite risk measure $\bar{\rho}$ can be quite complicated here.

An alternative, multiperiod risk averse approach can be formulated as

$$\operatorname{Min} \rho[-W_T]$$

s.t. $W_{t+1} = \sum_{i=1}^n \xi_{i,t+1} x_{it}, \ \sum_{i=1}^n x_{it} = W_t, \ x_t \ge 0, \ t = 0, \dots, T-1,$ (6.446)

for an explicitly defined risk measure ρ . Let, for example,

$$\rho(\cdot) := (1 - \beta)\mathbb{E}[\cdot] + \beta \mathsf{AV}@\mathsf{R}_{\alpha}(\cdot), \quad \beta \in [0, 1], \quad \alpha \in (0, 1).$$
(6.447)

Then problem (6.446) becomes

$$\operatorname{Min} (1-\beta)\mathbb{E}[-W_T] + \beta \left(-r + \alpha^{-1}\mathbb{E}[r - W_T]_+\right)$$

s.t. $W_{t+1} = \sum_{i=1}^n \xi_{i,t+1} x_{it}, \ \sum_{i=1}^n x_{it} = W_t, \ x_t \ge 0, \ t = 0, \dots, T-1,$ (6.448)

where $r \in \mathbb{R}$ is the (additional) first stage decision variable. After r is decided, at the first stage, the problem comes to minimizing $\mathbb{E}[U(W_T)]$ at the last stage, where $U(W) := (1 - \beta)W + \beta \alpha^{-1}[W - r]_+$ can be viewed as a disutility function.

The respective dynamic programming equations become as follows. The last stage value function $Q_{T-1}(W_{T-1}, r)$ is given by the optimal value of the problem

$$\underset{x_{T-1} \ge 0, W_T}{\min} \mathbb{E} \Big[-(1-\beta)W_T + \beta \alpha^{-1} [r - W_T]_+ \Big]
s.t. W_T = \sum_{i=1}^n \xi_{iT} x_{i,T-1}, \sum_{i=1}^n x_{i,T-1} = W_{T-1}.$$
(6.449)

Proceeding in this way, at stages t = T - 2, ..., 1 we consider the problems

$$\underset{x_t \ge 0, W_{t+1}}{\min} \mathbb{E} \left\{ Q_{t+1}(W_{t+1}, r) \right\}$$
s.t. $W_{t+1} = \sum_{i=1}^{n} \xi_{i,t+1} x_{it}, \sum_{i=1}^{n} x_{it} = W_t,$
(6.450)

whose optimal value is denoted $Q_t(W_t, r)$. Finally, at stage t = 0 we solve the problem

$$\begin{aligned}
& \underset{x_0 \ge 0, r, W_1}{\min} - \beta r + \mathbb{E}[Q_1(W_1, r)] \\
& \text{s.t. } W_1 = \sum_{i=1}^n \xi_{i1} x_{i0}, \ \sum_{i=1}^n x_{i0} = W_0.
\end{aligned}$$
(6.451)

In the above multiperiod risk averse approach the optimal policy is not myopic and the principle of conditional optimality is not satisfied.

Example 6.87 (Risk Averse Multistage Inventory Model) Consider the multistage inventory problem (1.17). The nested risk averse formulation of that problem can be written as

$$\begin{aligned}
& \underset{x_t \ge y_t}{\min} c_1(x_1 - y_1) + \rho_1 \Big[\psi_1(x_1, D_1) + c_2(x_2 - y_2) + \rho_{2|D_{[1]}} \Big[\psi_2(x_2, D_2) + \cdots \\
& + c_{T-1}(x_{T-1} - y_{T-1}) + \rho_{T-1|D_{[T-2]}} \Big[\psi_{T-1}(x_{T-1}, D_{T-1}) \\
& + c_T(x_T - y_T) + \rho_{T|D_{[T-1]}} [\psi_T(x_T, D_T)] \Big] \Big] \\
& \text{s.t. } y_{t+1} = x_t - D_t, \ t = 1, \dots, T - 1,
\end{aligned}$$
(6.452)

where y_1 is a given initial inventory level,

$$\psi_t(x_t, d_t) := b_t [d_t - x_t]_+ + h_t [x_t - d_t]_+$$

and $\rho_{t|D_{[t-1]}}(\cdot)$, $t = 2, \ldots, T$, are chosen conditional risk mappings. Recall that the notation $\rho_{t|D_{[t-1]}}(\cdot)$ stands for a conditional risk mapping obtained by using conditional expectations, conditional on $D_{[t-1]}$, and note that $\rho_1(\cdot)$ is real valued and is a coherent risk measure.

As it was discussed before, there are two equivalent interpretations of problem (6.452). We can write it as an optimization problem with respect to feasible policies $x_t(d_{[t-1]})$ (compare with (6.391)):

F

$$\underset{x_{1}, \boldsymbol{x}_{2}, \dots, \boldsymbol{x}_{T}}{\min} c_{1}(x_{1} - y_{1}) + \rho_{1} \left[\psi_{1}(x_{1}, D_{1}) + c_{2}(\boldsymbol{x}_{2}(D_{1}) - x_{1} + D_{1}) \right. \\ \left. + \rho_{2|D_{1}} \left[\psi_{2}(\boldsymbol{x}_{2}(D_{1}), D_{2}) + \cdots \right. \\ \left. + c_{T-1}(\boldsymbol{x}_{T-1}(D_{[T-2]}) - \boldsymbol{x}_{T-2}(D_{[T-3]}) + D_{T-2}) \right. \\ \left. + \rho_{T-1|D_{[T-2]}} \left[\psi_{T-1}(\boldsymbol{x}_{T-1}(D_{[T-2]}), D_{T-1}) \right. \\ \left. + c_{T}(\boldsymbol{x}_{T}(D_{[T-1]}) - \boldsymbol{x}_{T-1}(D_{[T-2]}) + D_{T-1}) \right. \\ \left. + \rho_{T|D_{[T-1]}} \left[\psi_{T}(\boldsymbol{x}_{T}(D_{[T-1]}), D_{T}) \right] \right] \right]$$
(6.453)

s.t.
$$x_1 \ge y_1, \ x_2(D_1) \ge x_1 - D_1,$$

 $x_t(D_{[t-1]}) \ge x_{t-1}(D_{[t-2]}) - D_{t-1}, \ t = 3, \dots, T.$

Alternatively we can write dynamic programming equations. At the last stage t = T, for observed inventory level y_T , we need to solve the problem:

$$\lim_{x_T \ge y_T} c_T(x_T - y_T) + \rho_{T|D_{[T-1]}} [\psi_T(x_T, D_T)].$$
(6.454)

The optimal value of problem (6.454) is denoted $Q_T(y_T, D_{[T-1]})$. Continuing in this way, we write for t = T - 1, ..., 2 the following dynamic programming equations

$$Q_t(y_t, D_{[t-1]}) = \min_{x_t \ge y_t} c_t(x_t - y_t) + \rho_{t|D_{[t-1]}} \left[\psi(x_t, D_t) + Q_{t+1} \left(x_t - D_t, D_{[t]} \right) \right].$$
(6.455)

Finally, at the first stage we need to solve problem

$$\underset{x_1 \ge y_1}{\operatorname{Min}} c_1(x_1 - y_1) + \rho_1 \big[\psi(x_1, D_1) + Q_2 \left(x_1 - D_1, D_1 \right) \big].$$
(6.456)

Suppose now that the process D_t is stagewise independent. Then, by exactly the same argument as in section 1.2.3, the cost-to-go (value) function $Q_t(y_t, d_{[t-1]}) = Q_t(y_t)$, $t = 2, \ldots, T$, is independent of $d_{[t-1]}$, and by convexity arguments the optimal policy $\bar{x}_t = \bar{x}_t(d_{[t-1]})$ is a basestock policy. That is, $\bar{x}_t = \max\{y_t, x_t^*\}$, where x_t^* is an optimal solution of

$$\underset{x_{t}}{\operatorname{Min}} c_{t} x_{t} + \rho_{t} \big[\psi(x_{t}, D_{t}) + Q_{t+1} \left(x_{t} - D_{t} \right) \big].$$
(6.457)

Recall that ρ_t denotes the coherent risk measure corresponding to the conditional risk mapping $\rho_{t|D_{[t-1]}}$.

Exercises

6.1. Let $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$ be a random variable with cdf $H(z) := P\{Z \leq z\}$. Note that $\lim_{z \downarrow t} H(z) = H(t)$ and denote $H^-(t) := \lim_{z \uparrow t} H(z)$. Consider functions $\phi_1(t) := \mathbb{E}[t - Z]_+, \phi_2(t) := \mathbb{E}[Z - t]_+$ and $\phi(t) := \beta_1 \phi_1(t) + \beta_2 \phi_2(t)$, where β_1, β_2 are positive constants. Show that ϕ_1, ϕ_2 and ϕ are real valued convex functions with subdifferentials

$$\partial \phi_1(t) = [H^-(t), H(t)]$$
 and $\partial \phi_2(t) = [-1 + H^-(t), -1 + H(t)],$
 $\partial \phi(t) = [(\beta_1 + \beta_2)H^-(t) - \beta_2, (\beta_1 + \beta_2)H(t) - \beta_2].$

Conclude that the set of minimizers of $\phi(t)$ over $t \in \mathbb{R}$ is the (closed) interval of $[\beta_2/(\beta_1 + \beta_2)]$ -quantiles of $H(\cdot)$.

6.2. Show that for any $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$,

$$\lim_{\alpha \downarrow 0} \alpha \operatorname{AV} @ \mathsf{R}_{\alpha}(Z) = 0.$$

6.3. For $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$ and $\alpha \in (0, 1)$ consider

$$\mathsf{AV}@\mathsf{R}^*_{\alpha}(Z) := \sup_{t \in \mathbb{R}} \left\{ t + \alpha^{-1} \mathbb{E}[Z - t]_{-} \right\}, \tag{6.458}$$

where $[a]_{-} = \min\{a, 0\} = a - [a]_{+}$. Show that maximum in the right hand side of (6.458) is attained at any α -quantile of the distribution of Z, and

$$\mathsf{AV}@\mathsf{R}^*_{\alpha}(Z) = \frac{1}{\alpha}\mathbb{E}[Z] - \frac{1-\alpha}{\alpha}\mathsf{AV}@\mathsf{R}_{1-\alpha}(Z). \tag{6.459}$$

Derive properties of AV@ R^*_{α} analogous to the respective properties of the Average Value-at-Risk.

- 6.4. Show that $\rho := V@R_{\alpha}, \alpha \in (0, 1)$, satisfies conditions (R2)–(R4), defined on page 298.
- 6.5. (i) Let $Y \sim \mathcal{N}(\mu, \sigma^2)$. Show that

$$\mathsf{V}@\mathsf{R}_{\alpha}(Y) = \mu + z_{\alpha}\sigma, \tag{6.460}$$

where $z_{\alpha} := \Phi^{-1}(1-\alpha)$, and

$$\mathsf{AV}@\mathsf{R}_{\alpha}(Y) = \mu + \frac{\sigma}{\alpha\sqrt{2\pi}}e^{-z_{\alpha}^{2}/2}.$$
(6.461)

(ii) Let Y^1, \ldots, Y^N be an iid sample of $Y \sim \mathcal{N}(\mu, \sigma^2)$. Compute the asymptotic variance and asymptotic bias of the sample estimator $\hat{\theta}_N$, of $\theta^* = \mathsf{AV}@\mathsf{R}_{\alpha}(Y)$, discussed in section 6.6.1.

6.6. Consider the chance constraint

$$\Pr\left\{\sum_{i=1}^{n}\xi_{i}x_{i} \ge b\right\} \ge 1 - \alpha, \tag{6.462}$$

where $\xi \sim \mathcal{N}(\mu, \Sigma)$ (see problem (1.43)). Note that this constraint can be written as

$$\mathsf{V}@\mathsf{R}_{\alpha}\left(b-\sum_{i=1}^{n}\xi_{i}x_{i}\right) \leq 0.$$
(6.463)

Consider the following constraint

$$\mathsf{AV}@\mathsf{R}_{\gamma}\left(b-\sum_{i=1}^{n}\xi_{i}x_{i}\right) \leq 0.$$
(6.464)

Show that constraints (6.462) and (6.464) are equivalent if $z_{\alpha} = \frac{1}{\gamma\sqrt{2\pi}}e^{-z_{\gamma}^2/2}$.

- 6.7. Consider the inverse function $\alpha^* = \mathfrak{p}^{-1}(\alpha)$, defined in section 6.4. Compute this function for $\psi(x) = (x-1)^2$ in example 6.54.
- 6.8. Consider the function $\phi(x) := AV@R_{\alpha}(F_x)$, where $F_x = F_x(\omega) = F(x,\omega)$ is a real valued random variable, on a probability space (Ω, \mathcal{F}, P) , depending on $x \in \mathbb{R}^n$. Assume that: (i) for a.e. $\omega \in \Omega$ the function $F(\cdot, \omega)$ is continuously differentiable on a neighborhood V of a point $x_0 \in \mathbb{R}^n$, (ii) the families $|F(x, \omega)|$, $x \in V$, and $||\nabla_x F(x, \omega)||$, $x \in V$, are dominated by a P-integrable function, (iii) the random variable F_x has continuous distribution for all $x \in V$. Show that, under these conditions, $\phi(x)$ is directionally differentiable at x_0 and

$$\phi'(x_0, d) = \alpha^{-1} \inf_{t \in [a, b]} \mathbb{E}\left\{ d^{\mathsf{T}} \nabla_x ([F(x_0, \omega) - t]_+) \right\},$$
(6.465)

where a and b are the respective left and right side $(1 - \alpha)$ -quantiles of the cdf of the random variable F_{x_0} . Conclude that if, moreover, $a = b = V@R_{\alpha}(F_{x_0})$, then $\phi(\cdot)$ is differentiable at x_0 and

$$\nabla \phi(x_0) = \alpha^{-1} \mathbb{E} \left[\mathbf{1}_{\{F_{x_0} > a\}}(\omega) \nabla_x F(x_0, \omega) \right].$$
(6.466)

Hint: use Theorem 7.49 together with Danskin Theorem 7.25.

- 6.9. Show that the set of saddle points of the minimax problem (6.272) is given by $\{\mu\} \times [\gamma^*, \gamma^{**}]$, where γ^* and γ^{**} are defined in (6.274).
- 6.10. Consider the absolute semideviation risk measure

$$\rho_c(Z) := \mathbb{E}\left\{Z + c[Z - \mathbb{E}(Z)]_+\right\}, \ Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P),$$

where $c \in [0, 1]$, and the following risk averse optimization problem:

$$\underbrace{\operatorname{Min}_{x \in X} \underbrace{\mathbb{E}\left\{G(x,\xi) + c[G(x,\xi) - \mathbb{E}(G(x,\xi))]_{+}\right\}}_{\rho_{c}[G(x,\xi)]}.$$
(6.467)

Viewing the optimal value of problem (6.467) as Von Mises statistical functional of the probability measure P, compute its influence function.

Hint: use derivations of section 6.6.3 together with Danskin Theorem.

6.11. Consider the risk averse optimization problem (6.228) related to the inventory model. Let the corresponding risk measure be of the form ρ_λ(Z) = E[Z] + λD(Z), where D(Z) is a measure of variability of Z = Z(ω) and λ is a nonnegative trade off coefficient between expectation and variability. Higher values of λ reflects a higher degree of risk aversion. Suppose that ρ_λ is a coherent risk measure for all λ ∈ [0, 1] and let S_λ be the set of optimal solutions of the corresponding risk averse problem. Suppose that the sets S₀ and S₁ are nonempty.

Show that if $S_0 \cap S_1 = \emptyset$, then S_λ is monotonically nonincreasing or monotonically nondecreasing in $\lambda \in [0, 1]$ depending upon whether $S_0 > S_1$ or $S_0 < S_1$. If $S_0 \cap S_1 \neq \emptyset$, then $S_\lambda = S_0 \cap S_1$ for any $\lambda \in (0, 1)$.

6.12. Consider the newsvendor problem with cost function

$$F(x,d) = cx + b[d-x]_{+} + h[x-d]_{+}, \text{ where } b > c \ge 0, h > 0,$$

and the following minimax problem

$$\underset{x \ge 0}{\operatorname{Min}} \sup_{H \in \mathfrak{M}} \mathbb{E}_{H}[F(x, D)],$$
(6.468)

where \mathfrak{M} is the set of cumulative distribution functions (probability measures) supported on (final) interval $[l, u] \subset \mathbb{R}_+$ and having a given mean $\overline{d} \in [l, u]$. Show that for any $x \in [l, u]$ the maximum of $\mathbb{E}_H[F(x, D)]$ over $H \in \mathfrak{M}$ is attained at the probability measure $\overline{H} = p\delta(l) + (1-p)\delta(u)$, where $p = (u - \overline{d})/(u - l)$, i.e., the cdf $\overline{H}(\cdot)$ is the step function

$$\bar{H}(z) = \begin{cases} 0, & \text{if } z < l, \\ p, & \text{if } l \le z < u, \\ 1, & \text{if } u \le z. \end{cases}$$

Conclude that \overline{H} is the cdf specified in Proposition 6.61, and that $\overline{x} = \overline{H}^{-1}(\kappa)$, where $\kappa = (b - c)/(b + h)$, is the optimal solution of problem (6.468). That is, $\overline{x} = l$ if $\kappa < p$, and $\overline{x} = u$ if $\kappa > p$, where $\kappa = \frac{b-c}{b+h}$.

6.13. Consider the following version of the newsvendor problem. A newsvendor has to decide about quantity x of a product to purchase at the cost of c per unit. He can sell this product at the price s per unit and unsold products can be returned to the vendor at the price of r per unit. It is assumed that $0 \le r < c < s$. If the demand D turns out to be greater than or equal to the order quantity x, then he makes the profit sx - cx = (s - c)x, while if D is less than x, his profit is sD + r(x - D) - cx. Thus the profit is a function of x and D and is given by

$$F(x,D) = \begin{cases} (s-c)x, & \text{if } x \le D, \\ (r-c)x + (s-r)D, & \text{if } x > D. \end{cases}$$
(6.469)

(a) Assuming that demand $D \ge 0$ is a random variable with cdf $H(\cdot)$, show that the expectation function $f(x) := \mathbb{E}_H[F(x, D)]$ can be represented in the form

$$f(x) = (s-c)x - (s-r)\int_0^x H(z)dz.$$
 (6.470)

Conclude that the set of optimal solutions of the problem

$$\operatorname{Max}_{x \ge 0} \left\{ f(x) := \mathbb{E}_H[F(x, D)] \right\}$$
(6.471)

is an interval given by the set of κ -quantiles of the cdf $H(\cdot)$ with $\kappa := (s - c)/(s - r)$.

(b) Consider the following risk averse version of the Newsvendor problem

$$\min_{x \ge 0} \{ \phi(x) := \rho[-F(x, D)] \}.$$
(6.472)

Here ρ is a real valued coherent risk measure representable in the form (6.231) and H^* is the corresponding reference cdf.

Show that: (i) The function $\phi(x)=\rho[-F(x,D)]$ can be represented in the form

$$\phi(x) = (c-s)x + (s-r)\int_0^x \bar{H}(z)dz$$
(6.473)

for some cdf \overline{H} .

(ii) If $\rho(\cdot) := AV@R_{\alpha}(\cdot)$, then $\overline{H}(z) = \min \{\alpha^{-1}H^*(z), 1\}$. Conclude that in that case optimal solutions of the risk averse problem (6.472) are smaller than the risk neutral problem (6.471).

- 6.14. Consider Theorem 6.69. Suppose that the cone $C := \{0_p\} \times \mathbb{R}^{q-p}_+$. Show that in that case the conclusion of Theorem 6.69 holds if instead of assuming that the functions ψ_i , i = 1, ..., q, are affine, it is assumed that ψ_i , i = 1, ..., p, are affine and ψ_i , i = p + 1, ..., q, are concave and continuous on Ω .
- 6.15. Consider the following risk averse approach to multistage portfolio selection. Let ξ_1, \ldots, ξ_T be the respective data process (of random returns) and consider the following chance constrained nested formulation

$$\text{Max } \mathbb{E}[W_T]$$
s.t. $W_{t+1} = \sum_{i=1}^n \xi_{i,t+1} x_{it}, \sum_{i=1}^n x_{it} = W_t, \ x_t \ge 0,$

$$\text{Pr} \left\{ W_{t+1} \ge \kappa W_t \, \big| \, \xi_{[t]} \right\} \ge 1 - \alpha, \ t = 0, \dots, T - 1,$$

$$(6.474)$$

where $\kappa \in (0, 1)$ and $\alpha \in (0, 1)$ are given constants. Dynamic programming equations for this problem can be written as follows. At the last stage t = T - 1 the cost-to-go function $Q_{T-1}(W_{T-1}, \xi_{[T-1]})$ is given by the optimal value of the problem

$$\begin{aligned}
& \max_{x_{T-1} \ge 0, W_T} \mathbb{E} \Big[W_T \, \big| \, \xi_{[T-1]} \Big] \\
& \text{s.t. } W_T = \sum_{i=1}^n \xi_{iT} x_{i,T-1}, \, \sum_{i=1}^n x_{i,T-1} = W_{T-1}, \\
& \mathsf{Pr} \, \big\{ W_T \ge \kappa W_{T-1} \, \big| \, \xi_{[T-1]} \big\} \,,
\end{aligned}$$
(6.475)

and at stage t = T - 2, ..., 1, the cost-to-go function $Q_t(W_t, \xi_{[t]})$ is given by the optimal value of the problem

$$\max_{\substack{x_t \ge 0, W_{t+1}}} \mathbb{E} \Big[Q_{t+1}(W_{t+1}, \xi_{[t+1]}) \, \big| \, \xi_{[t]} \Big]$$

s.t. $W_{t+1} = \sum_{i=1}^{n} \xi_{i,t+1} x_{i,t}, \ \sum_{i=1}^{n} x_{i,t} = W_t,$ (6.476)
$$\Pr \left\{ W_{t+1} \ge \kappa W_t \, \big| \, \xi_{[t]} \right\}.$$

Assuming that the process ξ_t is stagewise independent show that the optimal policy is myopic and is given by $\bar{x}_t(W_t) = W_t x_t^*$, where x_t^* is an optimal solution of the problem

$$\max_{x_t \ge 0} \sum_{i=1}^{n} \mathbb{E}\left[\xi_{i,t+1}\right] x_{i,t}
s.t. \sum_{i=1}^{n} x_{i,t} = 1, \ \Pr\left\{\sum_{i=1}^{n} \xi_{i,t+1} x_{i,t} \ge \kappa\right\} \ge 1 - \alpha.$$
(6.477)

6.16. In settings of section 6.8.1, with finite space Ω and $\mathcal{Z} = \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_T$, consider functional $\varrho : \mathcal{Z} \to \mathbb{R}$ defined as

$$\varrho(Z_1,\ldots,Z_T) := \max\left\{Z_1, \max_{\omega\in\Omega} Z_2(\omega), \ldots, \max_{\omega\in\Omega} Z_T(\omega)\right\}.$$
 (6.478)

Compared with the discussion of section 6.8.3, this functional ρ satisfies the respective conditions (R1),(R2) and (R4), but not the condition (R*3). For the corresponding multiperiod optimization problem, of the form (6.396), derive the dynamic programming equations²⁹:

$$Q_t(x_{t-1},\omega) = \inf_{x_t \in \mathcal{X}_t(x_{t-1},\omega)} f_t(x_t,\omega) \lor \rho_{t+1} \left[Q_{t+1}(x_t,\omega) \right],$$
(6.479)

where $\rho_{t+1} : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$ is the conditional max-mapping, $t = 1, \ldots, T-1$, and $\rho_{T+1}[Q_{T+1}(\cdot, \cdot)] \equiv 0$. Discuss time consistency of this optimization problem.

²⁹Recall that $a \lor b = \max\{a, b\}$ for $a, b \in \mathbb{R}$.

Chapter 7 Background Material

Alexander Shapiro

In this chapter we discuss some concepts and results from convex analysis, probability theory, functional analysis and optimization theories, needed for a development of the material in this book. We give or outline proofs of some results while others are referred to the literature. Of course, a careful derivation of the required material goes far beyond the scope of this book. So this chapter should be considered as a source of references rather than a systematic presentation of the material.

We denote by \mathbb{R}^n the standard *n*-dimensional vector space, of (column) vectors $x = (x_1, ..., x_n)^T$, equipped with the scalar product $x^T y = \sum_{i=1}^n x_i y_i$. Unless stated otherwise we denote by $\|\cdot\|$ the Euclidean norm $\|x\| = \sqrt{x^T x}$. The notation A^T stands for the transpose of matrix (vector) A, and " := " stands for "equal by definition" to distinguish it from the usual equality sign. By $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$ we denote the set of extended real numbers. The domain of an extended real valued function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is defined as

$$\operatorname{dom} f := \{ x \in \mathbb{R}^n : f(x) < +\infty \}.$$

It is said that f is proper if $f(x) > -\infty$ for all $x \in \mathbb{R}^n$ and its domain, dom f, is nonempty. The function f is said to be *lower semicontinuous* (lsc) at a point $x_0 \in \mathbb{R}^n$ if $f(x_0) \leq \lim \inf_{x \to x_0} f(x)$. It is said that f is lower semicontinuous if it is lsc at every point of \mathbb{R}^n . The largest lower semicontinuous function which is less than or equal to f is denoted lsc f. It is not difficult to show that f is lower semicontinuous if and only if (iff) its epigraph

$$\operatorname{epi} f := \{ (x, \alpha) \in \mathbb{R}^{n+1} : f(x) \le \alpha \},\$$

is a closed subset of \mathbb{R}^{n+1} . We often have to deal with polyhedral functions.

Definition 7.1. An extended real valued function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called polyhedral if it is proper convex and lower semicontinuous, its domain is a convex closed polyhedron and $f(\cdot)$ is piecewise linear on its domain.

By $\mathbf{1}_A(\cdot)$ we denote the *characteristic*¹ function

$$\mathbf{1}_{A}(x) := \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A, \end{cases}$$
(7.1)

and by $\mathbb{I}_A(\cdot)$ the *indicator* function

$$\mathbb{I}_A(x) := \begin{cases} 0, & \text{if } x \in A, \\ +\infty, & \text{if } x \notin A, \end{cases}$$
(7.2)

of set A.

By cl(A) we denote the topological closure of set $A \subset \mathbb{R}^n$. For sets $A, B \subset \mathbb{R}^n$ we denote by

$$dist(x, A) := \inf_{x' \in A} \|x - x'\|$$
(7.3)

the distance from $x \in \mathbb{R}^n$ to A, and by

$$\mathbb{D}(A,B) := \sup_{x \in A} \operatorname{dist}(x,B) \text{ and } \mathbb{H}(A,B) := \max\left\{\mathbb{D}(A,B), \mathbb{D}(B,A)\right\}$$
(7.4)

the *deviation* of the set A from the set B and the *Hausdorff distance* between the sets A and B, respectively. By the definition, $dist(x, A) = +\infty$ if A is empty, and $\mathbb{H}(A, B) = +\infty$ if A or B is empty.

7.1 Optimization and Convex Analysis

7.1.1 Directional Differentiability

Consider a mapping $g : \mathbb{R}^n \to \mathbb{R}^m$. It is said that g is directionally differentiable at a point $x_0 \in \mathbb{R}^n$ in a direction $h \in \mathbb{R}^n$ if the limit

$$g'(x_0, h) := \lim_{t \downarrow 0} \frac{g(x_0 + th) - g(x_0)}{t}$$
(7.5)

exists, in which case $g'(x_0, h)$ is called the *directional derivative* of g(x) at x_0 in the direction h. If g is directionally differentiable at x_0 in every direction $h \in \mathbb{R}^n$, then it is said that g is *directionally differentiable* at x_0 . Note that whenever exists, $g'(x_0, h)$ is positively homogeneous in h, i.e., $g'(x_0, th) = tg'(x_0, h)$ for any $t \ge 0$. If g(x) is directionally differentiable at x_0 and $g'(x_0, h)$ is *linear* in h, then it is said that g(x) is Gâteaux differentiable at x_0 . Equation (7.5) can be also written in the form

$$g(x_0 + h) = g(x_0) + g'(x_0, h) + r(h),$$
(7.6)

where the remainder term r(h) is such that $r(th)/t \to 0$, as $t \downarrow 0$, for any fixed $h \in \mathbb{R}^n$. If, moreover, $g'(x_0, h)$ is linear in h and the remainder term r(h) is "uniformly small" in the sense that $r(h)/||h|| \to 0$ as $h \to 0$, i.e., r(h) = o(h), then it is said that g(x) is differentiable at x_0 in the sense of Fréchet, or simply differentiable at x_0 .

¹Function $\mathbf{1}_{A}(\cdot)$ is often also called the indicator function of the set A. We call it here characteristic function in order to distinguish it from the indicator function $\mathbb{I}_{A}(\cdot)$.

Clearly Fréchet differentiability implies Gâteaux differentiability. The converse of that is not necessarily true. However, the following theorem shows that for locally Lipschitz continuous mappings both concepts do coincide. Recall that a mapping (function) g: $\mathbb{R}^n \to \mathbb{R}^m$ is said to be *Lipschitz continuous* on a set $X \subset \mathbb{R}^n$, if there is a constant $c \ge 0$ such that

$$||g(x_1) - g(x_2)|| \le c ||x_1 - x_2||$$
, for all $x_1, x_2 \in X$

If g is Lipschitz continuous on a neighborhood of every point of X (probably with different Lipschitz constants), then it is said that g is *locally Lipschitz continuous* on X.

Theorem 7.2. Suppose that mapping $g : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous in a neighborhood of a point $x_0 \in \mathbb{R}^n$ and directionally differentiable at x_0 . Then $g'(x_0, \cdot)$ is Lipschitz continuous on \mathbb{R}^n and

$$\lim_{h \to 0} \frac{g(x_0 + h) - g(x_0) - g'(x_0, h)}{\|h\|} = 0.$$
(7.7)

Proof. For $h_1, h_2 \in \mathbb{R}^n$ we have

$$||g'(x_0, h_1) - g'(x_0, h_2)|| = \lim_{t \downarrow 0} \frac{||g(x_0 + th_1) - g(x_0 + th_2)||}{t}.$$

Also, since g is Lipschitz continuous near x_0 say with Lipschitz constant c, we have that for t > 0 small enough

$$|g(x_0 + th_1) - g(x_0 + th_2)|| \le ct ||h_1 - h_1||.$$

It follows that $||g'(x_0, h_1) - g'(x_0, h_2)|| \le c||h_1 - h_1||$ for any $h_1, h_2 \in \mathbb{R}^n$, i.e., $g'(x_0, \cdot)$ is Lipschitz continuous on \mathbb{R}^n .

Consider now a sequence $t_k \downarrow 0$ and a sequence $\{h_k\}$ converging to a point $h \in \mathbb{R}^n$. We have that

$$g(x_0 + t_k h_k) - g(x_0) = \left(g(x_0 + t_k h) - g(x_0)\right) + \left(g(x_0 + t_k h_k) - g(x_0 + t_k h)\right)$$

and

$$||g(x_0 + t_k h_k) - g(x_0 + t_k h)|| \le ct_k ||h_k - h||$$

for all k large enough. It follows that

$$g'(x_0, h) = \lim_{k \to \infty} \frac{g(x_0 + t_k h_k) - g(x_0)}{t_k}.$$
(7.8)

The proof of (7.7) can be completed now by arguing by a contradiction and using the fact that every bounded sequence in \mathbb{R}^n has a convergent subsequence.

We have that g is differentiable at a point $x \in \mathbb{R}^n$, iff

$$g(x+h) - g(x) = [\nabla g(x)]h + o(h), \tag{7.9}$$

where $\nabla g(x)$ is the so-called $m \times n$ Jacobian matrix of partial derivatives $[\partial g_i(x)/\partial x_j]$, i = 1, ..., m, j = 1, ..., n. If m = 1, i.e., g(x) is real valued, we call $\nabla g(x)$ the gradient of g at x. In that case equation (7.9) takes the form

$$g(x+h) - g(x) = h^{\mathsf{T}} \nabla g(x) + o(h).$$
 (7.10)

Note that when $g(\cdot)$ is real valued, we write its gradient $\nabla g(x)$ as a *column* vector. This is why there is a slight discrepancy between the notation of (7.10) and notation of (7.9) where the Jacobian matrix is of order $m \times n$. If g(x, y) is a function (mapping) of two vector variables x and y and we consider derivatives of $g(\cdot, y)$ while keeping y constant, we write the corresponding gradient (Jacobian matrix) as $\nabla_x g(x, y)$.

Clarke Generalized Gradient

Consider now a *locally Lipschitz continuous* function $f : U \to \mathbb{R}$ defined on an open set $U \subset \mathbb{R}^n$. By Rademacher's theorem we have that f(x) is differentiable on U almost everywhere. That is, the subset of U where f is not differentiable has Lebesgue measure zero. At a point $\bar{x} \in U$ consider the set of all limits of the form $\lim_{k\to\infty} \nabla f(x_k)$ such that $x_k \to \bar{x}$ and f is differentiable at x_k . This set is nonempty and compact, its convex hull is called *Clarke generalized gradient* of f at \bar{x} and denoted $\partial^{\circ} f(\bar{x})$. The generalized directional derivative of f at \bar{x} is defined as

$$f^{\circ}(\bar{x}, d) := \limsup_{\substack{x \to \bar{x} \\ t \downarrow 0}} \frac{f(x + td) - f(x)}{t}.$$
(7.11)

It is possible to show that $f^{\circ}(\bar{x}, \cdot)$ is the support function of the set $\partial^{\circ} f(\bar{x})$. That is

$$f^{\circ}(\bar{x},d) = \sup_{z \in \partial^{\circ} f(\bar{x})} z^{\mathsf{T}} d, \quad \forall d \in \mathbb{R}^{n}.$$
(7.12)

Function f is called regular in the sense of Clarke, or *Clarke regular*, at $\bar{x} \in \mathbb{R}^n$ if $f(\cdot)$ is directionally differentiable at \bar{x} and $f'(\bar{x}, \cdot) = f^{\circ}(\bar{x}, \cdot)$. Any convex function f is Clarke regular and its Clarke generalized gradient $\partial^{\circ} f(\bar{x})$ coincides with the respective subdifferential in the sense of convex analysis. For a concave function f, the function -f is Clarke regular, and we shall call it Clarke regular with the understanding that we modify the regularity requirement above to apply to -f. In this case we have also $\partial^{\circ}(-f)(\bar{x}) = -\partial^{\circ} f(\bar{x})$.

We say that f is *continuously differentiable* at a point $\bar{x} \in U$ if $\partial^{\circ} f(\bar{x})$ is a singleton. In other words, f is continuously differentiable at \bar{x} if f is differentiable at \bar{x} and $\nabla f(x)$ is continuous at \bar{x} on the set where f is differentiable. Note that continuous differentiability of f at a point \bar{x} does not imply differentiability of f at every point of any neighborhood of the point \bar{x} .

Consider a composite real valued function f(x) := g(h(x)) with $h : \mathbb{R}^m \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}$, and assume that g and h are locally Lipschitz continuous. Then

$$\partial^{\circ} f(x) \subset \operatorname{cl}\left\{\operatorname{conv}\left(\sum_{i=1}^{n} \alpha_{i} v_{i} : \alpha \in \partial^{\circ} g(y), \ v_{i} \in \partial^{\circ} h_{i}(x), \ i = 1, ..., n\right)\right\},$$
(7.13)

where $\alpha = (\alpha_1, ..., \alpha_n)$, y = h(x) and $h_1, ..., h_n$ are components of h. The equality in (7.13) holds true, if any one of the following conditions is satisfied: (i) g and h_i , i = 1, ..., n, are Clarke-regular and every element in $\partial^{\circ} g(y)$ has non-negative components, (ii) g is differentiable and n = 1, (iii) g is Clarke-regular and h is differentiable.

7.1.2 Elements of Convex Analysis

Let C be a subset of \mathbb{R}^n . It is said that $x \in \mathbb{R}^n$ is an *interior point* of C if there is a neighborhood N of x such that $N \subset C$. The set of interior points of C is denoted int(C). The *convex hull* of C, denoted conv(C), is the smallest convex set including C. It is said that C is a *cone* if for any $x \in C$ and $t \ge 0$ it follows that $tx \in C$. The *polar cone* of a cone $C \subset \mathbb{R}^n$ is defined as

$$C^* := \left\{ z \in \mathbb{R}^n : z^\mathsf{T} x \le 0, \ \forall \, x \in C \right\}.$$

$$(7.14)$$

We have that the polar of the polar cone $C^{**} = (C^*)^*$ is equal to the topological closure of the convex hull of C, and that $C^{**} = C$ iff the cone C is convex and closed.

Let C be a nonempty *convex* subset of \mathbb{R}^n . The affine space generated by C is the space of points in \mathbb{R}^n of the form tx + (1 - t)y, where $x, y \in C$ and $t \in \mathbb{R}$. It is said that a point $x \in \mathbb{R}^n$ belongs to the *relative interior* of the set C if x is an interior point of C relative to the affine space generated by C, i.e., there exists a neighborhood of x such that its intersection with the affine space generated by C is included in C. The relative interior set of C is denoted $\operatorname{ri}(C)$. Note that if the interior of C is nonempty, then the affine space generated by $C \subset \mathbb{R}^n$ is nonempty, then the affine space generated by C coincides with \mathbb{R}^n , and hence in that case $\operatorname{ri}(C) = \operatorname{int}(C)$. Note also that the relative interior of any convex set $C \subset \mathbb{R}^n$ is nonempty. The *recession cone* of the set C is formed by vectors $h \in \mathbb{R}^n$ such that for any $x \in C$ and any t > 0 it follows that $x + th \in C$. The recession cone of the convex set C is convex and is closed if the set C is closed. Also the convex set C is bounded iff its recession cone is $\{0\}$.

Theorem 7.3 (Helly). Let A_i , $i \in \mathcal{I}$, be a family of convex subsets of \mathbb{R}^n . Suppose that the intersection of any n + 1 sets of this family is nonempty and either the index set \mathcal{I} is finite or the sets A_i , $i \in \mathcal{I}$, are closed and there exists no common nonzero recession direction to the sets A_i , $i \in \mathcal{I}$. Then the intersection of all sets A_i , $i \in \mathcal{I}$, is nonempty.

The support function $S(\cdot) = S_C(\cdot)$ of a (nonempty) set $C \subset \mathbb{R}^n$ is defined as

$$\mathbf{S}(h) := \sup_{z \in C} z^{\mathsf{T}} h. \tag{7.15}$$

The support function $S(\cdot)$ is convex, positively homogeneous and lower semicontinuous. The support function of a set C coincides with the support function of the set cl(conv C). If $S_1(\cdot)$ and $S_2(\cdot)$ are support functions of *convex closed* sets C_1 and C_2 , respectively, then $S_1(\cdot) \leq S_2(\cdot)$ iff $C_1 \subset C_2$, and $S_1(\cdot) = S_2(\cdot)$ iff $C_1 = C_2$.

Let $C \subset \mathbb{R}^n$ be a convex closed set. The *normal cone* to C at a point $x_0 \in C$ is defined as

$$\mathcal{N}_C(x_0) := \left\{ z : z^{\mathsf{T}}(x - x_0) \le 0, \ \forall x \in C \right\}.$$
(7.16)

By definition $\mathcal{N}_C(x_0) := \emptyset$ if $x_0 \notin C$. The topological closure of the *radial cone* $\mathcal{R}_C(x_0) := \bigcup_{t>0} \{t(C-x_0)\}$ is called the *tangent cone* to C at $x_0 \in C$, and denoted $\mathcal{T}_C(x_0)$. Both cones $\mathcal{T}_C(x_0)$ and $\mathcal{N}_C(x_0)$ are closed and convex, and each one is the polar cone of the other.

Consider an extended real valued function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$. It is not difficult to show that f is convex iff its epigraph epif is a convex subset of \mathbb{R}^{n+1} . Suppose that f is a *convex* function and $x_0 \in \mathbb{R}^n$ is a point such that $f(x_0)$ is *finite*. Then f(x) is directionally differentiable at x_0 , its directional derivative $f'(x_0, \cdot)$ is an extended real valued convex positively homogeneous function and can be written in the form

$$f'(x_0, h) = \inf_{t>0} \frac{f(x_0 + th) - f(x_0)}{t}.$$
(7.17)

Moreover, if x_0 is in the interior of the domain of $f(\cdot)$, then f(x) is Lipschitz continuous in a neighborhood of x_0 , the directional derivative $f'(x_0, h)$ is finite valued for any $h \in \mathbb{R}^n$, and f(x) is differentiable at x_0 iff $f'(x_0, h)$ is linear in h.

It is said that a vector $z \in \mathbb{R}^n$ is a *subgradient* of f(x) at x_0 if

$$f(x) - f(x_0) \ge z^{\mathsf{T}}(x - x_0), \quad \forall x \in \mathbb{R}^n.$$
 (7.18)

The set of all subgradients of f(x), at x_0 , is called the *subdifferential* and denoted $\partial f(x_0)$. The subdifferential $\partial f(x_0)$ is a closed convex subset of \mathbb{R}^n . It is said that f is *subdifferentiable* at x_0 if $\partial f(x_0)$ is nonempty. If f is subdifferentiable at x_0 , then the normal cone $\mathcal{N}_{\text{dom } f}(x_0)$, to the domain of f at x_0 , forms the recession cone of the set $\partial f(x_0)$. It is also clear that if f is subdifferentiable at x_0 , then $f(x) > -\infty$ for any x and hence f is proper.

By duality theory of convex analysis we have that if the directional derivative $f'(x_0, \cdot)$ is lower semicontinuous, then

$$f'(x_0, h) = \sup_{z \in \partial f(x_0)} z^{\mathsf{T}} h, \ \forall h \in \mathbb{R}^n,$$
(7.19)

i.e., $f'(x_0, \cdot)$ is the support function of the set $\partial f(x_0)$. In particular, if x_0 is an interior point of the domain of f(x), then $f'(x_0, \cdot)$ is continuous, $\partial f(x_0)$ is nonempty and compact and (7.19) holds. Conversely, if $\partial f(x_0)$ is nonempty and compact, then x_0 is an interior point of the domain of f(x). Also f(x) is differentiable at x_0 iff $\partial f(x_0)$ is a singleton, i.e., contains only one element, which then coincides with the gradient $\nabla f(x_0)$.

Theorem 7.4 (Moreau–Rockafellar). Let $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$, i = 1, ..., m, be proper convex functions, $f(\cdot) := f_1(\cdot) + ... + f_m(\cdot)$ and x_0 be a point such that $f_i(x_0)$ are finite, i.e., $x_0 \in \bigcap_{i=1}^m \text{dom } f_i$. Then

$$\partial f_1(x_0) + \dots + \partial f_m(x_0) \subset \partial f(x_0). \tag{7.20}$$

Moreover,

$$\partial f_1(x_0) + \dots + \partial f_m(x_0) = \partial f(x_0). \tag{7.21}$$

if any one of the following conditions holds: (i) the set $\bigcap_{i=1}^{m} \operatorname{ri}(\operatorname{dom} f_i)$ is nonempty, (ii) the functions $f_1, ..., f_k, k \leq m$, are polyhedral and the intersection of the sets $\bigcap_{i=1}^{k} \operatorname{dom} f_i$ and $\bigcap_{i=k+1}^{m} \operatorname{ri}(\operatorname{dom} f_i)$ is nonempty, (iii) there exists a point $\bar{x} \in \operatorname{dom} f_m$ such that $\bar{x} \in \operatorname{int}(\operatorname{dom} f_i)$, i = 1, ..., m - 1.

In particular, if all functions $f_1, ..., f_m$ in the above theorem are polyhedral, then the equation (7.21) holds without an additional regularity condition.

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be an extended real valued function. The *conjugate function* of f is

$$f^*(z) := \sup_{x \in \mathbb{R}^n} \{ z^\mathsf{T} x - f(x) \}.$$
(7.22)

The conjugate function $f^* : \mathbb{R}^n \to \overline{\mathbb{R}}$ is always convex and lower semicontinuous. The conjugate of f^* is denoted f^{**} . Note that if $f(x) = -\infty$ at some $x \in \mathbb{R}^n$, then $f^*(\cdot) \equiv +\infty$ and $f^{**}(\cdot) \equiv -\infty$.

Theorem 7.5 (Fenchel-Moreau). Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper extended real valued convex function. Then

$$f^{**} = \operatorname{lsc} f. \tag{7.23}$$

It follows from (7.23) that if f is proper and convex, then $f^{**} = f$ iff f is lower semicontinuous. Also it immediately follows from the definitions that

$$z \in \partial f(x)$$
 iff $f^*(z) + f(x) = z^{\mathsf{T}} x$.

By applying that to the function f^{**} , instead of f, we obtain that $z \in \partial f^{**}(x)$ iff $f^{***}(z) + f^{**}(x) = z^{\mathsf{T}}x$. Now by the Fenchel–Moreau Theorem we have that $f^{***} = f^*$, and hence $z \in \partial f^{**}(x)$ iff $f^*(z) + f^{**}(x) = z^{\mathsf{T}}x$. Consequently we obtain

$$\partial f^{**}(x) = \operatorname*{arg\,max}_{z \in \mathbb{R}^n} \left\{ z^{\mathsf{T}} x - f^*(z) \right\},\tag{7.24}$$

and if $f^{**}(x) = f(x)$ and is finite, then $\partial f^{**}(x) = \partial f(x)$.

Strong convexity. Let $X \subset \mathbb{R}^n$ be a nonempty closed convex set. It is said that a function $f: X \to \mathbb{R}$ is *strongly convex*, with parameter c > 0, if²

$$tf(x') + (1-t)f(x) \ge f(tx' + (1-t)x) + \frac{1}{2}ct(1-t)\|x' - x\|^2,$$
(7.25)

for all $x, x' \in X$ and $t \in [0, 1]$. It is not difficult to verify that f is strongly convex iff the function $\psi(x) := f(x) - \frac{1}{2}c||x||^2$ is convex on X.

Indeed, convexity of ψ means that the inequality

$$tf(x') - \frac{1}{2}ct\|x'\|^2 + (1-t)f(x) - \frac{1}{2}c(1-t)\|x\|^2$$

$$\geq f(tx' + (1-t)x) - \frac{1}{2}c\|tx' + (1-t)x\|^2,$$

holds for all $t \in [0, 1]$ and $x, x' \in X$. By the identity

$$t||x'||^{2} + (1-t)||x||^{2} - ||tx' + (1-t)x||^{2} = t(1-t)||x' - x||^{2},$$

this is equivalent to (7.25).

If the set X has a nonempty interior and $f : X \to \mathbb{R}$ is continuous and differentiable at every point $x \in int(X)$, then f is strongly convex iff

$$f(x') \ge f(x) + (x' - x)^{\mathsf{T}} \nabla f(x) + \frac{1}{2} c \|x' - x\|^2, \quad \forall x, x' \in \operatorname{int}(X),$$
(7.26)

or equivalently iff

$$(x' - x)^{\mathsf{T}}(\nabla f(x') - \nabla f(x)) \ge c \|x' - x\|^2, \ \forall x, x' \in \text{int}(X).$$
(7.27)

²Unless stated otherwise we denote by $\|\cdot\|$ the Euclidean norm on \mathbb{R}^n .

7.1.3 Optimization and Duality

Consider a real valued function $L: X \times Y \to \mathbb{R}$, where X and Y are arbitrary sets. We can associate with the function L(x, y) the following two optimization problems:

$$\operatorname{Min}_{x \in X} \left\{ f(x) := \sup_{y \in Y} L(x, y) \right\},$$
(7.28)

$$\operatorname{Max}_{y \in Y} \{ g(y) := \inf_{x \in X} L(x, y) \},$$
(7.29)

viewed as dual to each other. We have that for any $x \in X$ and $y \in Y$,

$$g(y) = \inf_{x' \in X} L(x', y) \le L(x, y) \le \sup_{y' \in Y} L(x, y') = f(x),$$

and hence the optimal value of problem (7.28) is greater than or equal to the optimal value of problem (7.29). It is said that a point $(\bar{x}, \bar{y}) \in X \times Y$ is a *saddle point* of L(x, y) if

$$L(\bar{x}, y) \le L(\bar{x}, \bar{y}) \le L(x, \bar{y}), \quad \forall (x, y) \in X \times Y.$$
(7.30)

Theorem 7.6. The following holds: (i) The optimal value of problem (7.28) is greater than or equal to the optimal value of problem (7.29). (ii) Problems (7.28) and (7.29) have the same optimal value and each has an optimal solution if and only if there exists a saddle point (\bar{x}, \bar{y}) . In that case \bar{x} and \bar{y} are optimal solutions of problems (7.28) and (7.29), respectively. (iii) If problems (7.28) and (7.29) have the same optimal value, then the set of saddle points coincides with the Cartesian product of the sets of optimal solutions of (7.28) and (7.29).

Suppose that there is no duality gap between problems (7.28) and (7.29), i.e., their optimal values are equal to each other, and let \bar{y} be an optimal solution of problem (7.29). By the above we have that the set of optimal solutions of problem (7.28) is contained in the set of optimal solutions of the problem

$$\min_{x \in X} L(x, \bar{y}),$$
(7.31)

and the common optimal value of problems (7.28) and (7.29) is equal to the optimal value of (7.31). In applications of the above results to optimization problems with constraints, the function L(x, y) usually is the Lagrangian and y is a vector of Lagrange multipliers. The inclusion of the set of optimal solutions of (7.28) into the set of optimal solutions of (7.31) can be strict (see the following example).

Example 7.7 Consider the linear problem

$$\underset{x \in \mathbb{R}}{\operatorname{Min}} x \text{ subject to } x \ge 0.$$
(7.32)

This problem has unique optimal solution $\bar{x} = 0$ and can be written in the minimax form (7.28) with L(x, y) := x - yx, $Y := \mathbb{R}_+$ and $X := \mathbb{R}$. The objective function g(y) of its dual (of the form (7.29)) is equal to $-\infty$ for all y except y = 1 for which g(y) = 0. There is no duality gap here between the primal and dual problems and the dual problem has unique feasible point $\bar{y} = 1$, which is also its optimal solution. The corresponding problem (7.31) takes here the form of minimizing $L(x, 1) \equiv 0$ over $x \in \mathbb{R}$, with the set of optimal solutions equal to \mathbb{R} . That is, in this example the set of optimal solutions of (7.28) is a strict subset of the set of optimal solutions of (7.31).
Conjugate Duality

An alternative approach to duality, referred to as *conjugate duality*, is the following. Consider an extended real valued function $\psi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, where \mathcal{X} is an abstract vector (linear) space and \mathcal{Y} is a *finite dimensional* vector space. We assume that with the space \mathcal{X} is associated a space \mathcal{X}^* of linear functionals $x^* : \mathcal{X} \to \mathbb{R}$ and the corresponding scalar product $\langle x^*, x \rangle := x^*(x), x^* \in \mathcal{X}^*, x \in \mathcal{X}$. We also assume that there is a scalar product $\langle y^*, y \rangle, y^*, y \in \mathcal{Y}$, defined on the space \mathcal{Y} . For example if $\mathcal{Y} := \mathbb{R}^n$, then we can use the standard scalar product $\langle y^*, y \rangle := (y^*)^{\mathsf{T}}y$, and if \mathcal{Y} is the linear space of $m \times m$ symmetric matrices, we can use³ $\langle Y^*, Y \rangle := \mathrm{Tr}(Y^*Y)$. In this section we assume that the space \mathcal{Y} , we will discuss this further in section 7.3.1.

Let $\vartheta(y)$ be the optimal value of the parameterized problem

$$\min_{x \in \mathcal{X}} \psi(x, y), \tag{7.33}$$

i.e., $\vartheta(y) := \inf_{x \in \mathcal{X}} \psi(x, y)$. Note that implicitly the optimization in the above problem is performed over the domain of the function $\psi(\cdot, y)$, i.e., dom $\psi(\cdot, y)$ can be viewed as the feasible set of problem (7.33). If the set dom $\psi(\cdot, y)$ is empty, then by the definition $\vartheta(y) = +\infty$.

The conjugate of the function $\vartheta(y)$ can be expressed in terms of the conjugate of $\psi(x,y)$. That is, the conjugate of ψ is

$$\psi^*(x^*, y^*) := \sup_{(x,y)\in\mathcal{X} imes\mathcal{Y}} \left\{ \langle x^*, x
angle + \langle y^*, y
angle - \psi(x,y)
ight\},$$

and hence the conjugate of ϑ can be written as

$$\begin{aligned} \vartheta^*(y^*) &:= \sup_{y \in \mathcal{Y}} \left\{ \langle y^*, y \rangle - \vartheta(y) \right\} = \sup_{y \in \mathcal{Y}} \left\{ \langle y^*, y \rangle - \inf_{x \in \mathcal{X}} \psi(x, y) \right\} \\ &= \sup_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \left\{ \langle y^*, y \rangle - \psi(x, y) \right\} = \psi^*(0, y^*). \end{aligned}$$

Consequently, the conjugate of ϑ^* is

$$\vartheta^{**}(y) = \sup_{y^* \in \mathcal{Y}} \left\{ \langle y^*, y \rangle - \psi^*(0, y^*) \right\}.$$
(7.34)

This leads to the following dual of (7.33):

$$\max_{y^* \in \mathcal{Y}} \left\{ \langle y^*, y \rangle - \psi^*(0, y^*) \right\}.$$
(7.35)

We assume further that the function $\psi : \mathcal{X} \times \mathcal{Y} \to \overline{\mathbb{R}}$ is *convex*. Then it is straightforward to verify that the optimal value function $\vartheta : \mathcal{Y} \to \overline{\mathbb{R}}$ is also convex. In the above formulation of problem (7.33) and its (conjugate) dual (7.35) we have that $\vartheta(y)$ and $\vartheta^{**}(y)$ are optimal values of (7.33) and (7.35), respectively. By the Fenchel–Moreau Theorem we have that either $\vartheta^{**}(\cdot)$ is identically $-\infty$, or

$$\vartheta^{**}(y) = (\operatorname{lsc} \vartheta)(y), \quad \forall \, y \in \mathcal{Y}.$$
(7.36)

³By Tr(A) we denote trace of an $m \times m$ matrix.

It follows that $\vartheta^{**}(y) \leq \vartheta(y)$ for any $y \in \mathcal{Y}$. It is said that there is no duality gap between (7.33) and its dual (7.35) if $\vartheta^{**}(y) = \vartheta(y)$.

It is said that the problem (7.33) is *subconsistent*, for a given value of y, if $\operatorname{lsc} \vartheta(y) < +\infty$. If problem (7.33) is feasible, i.e., dom $\psi(\cdot, y)$ is nonempty, then $\vartheta(y) < +\infty$, and hence (7.33) is subconsistent.

Theorem 7.8. Suppose that the function $\psi(\cdot, \cdot)$ is convex. Then the following holds: (i) The optimal value function $\vartheta(\cdot)$ is convex. (ii) If problem (7.33) is subconsistent, then $\vartheta^{**}(y) = \vartheta(y)$ if and only if the optimal value function $\vartheta(\cdot)$ is lower semicontinuous at y. (iii) If $\vartheta^{**}(y)$ is finite, then the set of optimal solutions of the dual problem (7.35) coincides with $\partial \vartheta^{**}(y)$. (iv) The set of optimal solutions of the dual problem (7.35) is nonempty and bounded if and only if $\vartheta(y)$ is finite and $\vartheta(\cdot)$ is continuous at y.

Proof. It is straightforward to verify that convexity of $\vartheta(\cdot)$ follows from convexity of $\psi(\cdot, \cdot)$. Assertion (ii) follows by the Fenchel–Moreau Theorem. Assertion (iii) follows from formula (7.24). If $\vartheta(\cdot)$ is continuous at y, then it is lower semicontinuous at y, and hence $\vartheta^{**}(y) = \vartheta(y)$. Moreover, in that case $\partial \vartheta^{**}(y) = \partial \vartheta(y)$ and is nonempty and bounded provided that $\vartheta(y)$ is finite. It follows then that the set of optimal solutions of the dual problem (7.35) is nonempty and bounded. Conversely, if the set of optimal solutions of (7.35) is nonempty and bounded, then, by (iii), $\partial \vartheta^{**}(y)$ is nonempty and bounded, and hence by convex analysis $\vartheta(\cdot)$ is continuous at y. Note also that if $\partial \vartheta(y)$ is nonempty, then $\vartheta^{**}(y) = \vartheta(y)$ and $\partial \vartheta^{**}(y) = \partial \vartheta(y)$. \Box

The above analysis can be also used in order to describe differentiability properties of the optimal value function $\vartheta(\cdot)$ in terms of its subdifferentials.

Theorem 7.9. Suppose that the function $\psi(\cdot, \cdot)$ is convex and let $y \in \mathcal{Y}$ be a given point. Then the following holds: (i) The optimal value function $\vartheta(\cdot)$ is subdifferentiable at y if and only if $\vartheta(\cdot)$ is lower semicontinuous at y and the dual problem (7.35) possesses an optimal solution. (ii) The subdifferential $\vartheta\vartheta(y)$ is nonempty and bounded if and only if $\vartheta(y)$ is finite and the set of optimal solutions of the dual problem (7.35) is nonempty and bounded. (iii) In both above cases $\vartheta\vartheta(y)$ coincides with the set of optimal solutions of the dual problem (7.35).

Since $\vartheta(\cdot)$ is convex, we also have that $\partial\vartheta(y)$ is nonempty and bounded iff $\vartheta(y)$ is finite and $y \in \operatorname{int}(\operatorname{dom} \vartheta)$. The condition $y \in \operatorname{int}(\operatorname{dom} \vartheta)$ means the following: there exists a neighborhood N of y such that for any $y' \in N$ the domain of $\psi(\cdot, y')$ is nonempty.

Example 7.10 Consider the following problem

$$\begin{array}{ll}
& \underset{x \in X}{\operatorname{Min}} & f(x) \\
& \text{s.t.} & G(x) \preceq_C y,
\end{array}$$
(7.37)

where X is a subset of \mathcal{X} , $f : \mathcal{X} \to \mathbb{R}$, $G : \mathcal{X} \to \mathcal{Y}$, $y \in \mathcal{Y}$, $C \subset \mathcal{Y}$ is a closed convex cone and $a \preceq_C b$ denotes the partial order with respect to the cone C, i.e., $a \preceq_C b$ means that $b - a \in C$. We can formulate this problem in the form (7.33) by defining

$$\psi(x,y) := \bar{f}(x) + F(y - G(x)),$$

where $\overline{f}(\cdot) := f(\cdot) + \mathbb{I}_X(\cdot)$ and $F(\cdot) := \mathbb{I}_C(\cdot)$.

Suppose that the problem (7.37) is convex, that is, the set X and the function f(x) are convex and the mapping G is convex with respect to the cone C. (In particular, if $C := \mathbb{R}^m_+$, then convexity of $G(\cdot) = (g_1(\cdot), ..., g_m(\cdot))$ with respect to C means that the functions $g_1(\cdot), ..., g_m(\cdot)$ are convex.) Then it is straightforward to verify that the function $\psi(x, y)$ is also convex. Let us calculate the conjugate of the function $\psi(x, y)$,

$$\begin{split} \psi^*(x^*, y^*) &= \sup_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \left\{ \langle x^*, x \rangle + \langle y^*, y \rangle - \bar{f}(x) - F(y - G(x)) \right\} = \\ \sup_{x \in \mathcal{X}} \left\{ \langle x^*, x \rangle - \bar{f}(x) + \langle y^*, G(x) \rangle + \sup_{y \in \mathcal{Y}} \left[\langle y^*, (y - G(x)) - F(y - G(x)) \right] \right\}. \end{split}$$

By change of variables z = y - G(x) we obtain that

$$\sup_{y\in\mathcal{Y}} \left[\langle y^*, (y-G(x)) \rangle - F(y-G(x)) \right] = \sup_{z\in\mathcal{Y}} \left[\langle y^*, z \rangle - F(z) \right] = \mathbb{I}_{C^*}(y^*),$$

where C^* is the polar of the cone C, and hence

$$\psi^*(x^*, y^*) = \sup_{x \in X} \left\{ \langle x^*, x \rangle - f(x) + \langle y^*, G(x) \rangle \right\} + \mathbb{I}_{C^*}(y^*),$$

Consequently the dual of the problem (7.37) can be written in the form

$$\operatorname{Max}_{\lambda \in C^*} \left\{ \langle \lambda, y \rangle + \inf_{x \in X} L(x, \lambda) \right\},$$
(7.38)

where $L(x, \lambda) := f(x) - \langle \lambda, G(x) \rangle$, is the corresponding Lagrangian. Note that we changed the notation from y^* to λ in order to emphasize that the above problem (7.38) is the standard Lagrangian dual of (7.37) with λ being vector of Lagrange multipliers. The results of Propositions 7.8 and 7.9 can be applied to problem (7.37) and its dual (7.38) in a straightforward way. In particular, if $C := \mathbb{R}^m_+$, then the dual problem (7.38) becomes

$$\max_{\lambda \le 0} \left\{ \lambda^{\mathsf{T}} y + \inf_{x \in X} L(x, \lambda) \right\},\tag{7.39}$$

where $L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x)$.

As another example consider a function $L : \mathbb{R}^n \times Y \to \overline{\mathbb{R}}$, where Y is a vector space (not necessarily finite dimensional), and the corresponding pair of dual problems (7.28) and (7.29). Define

$$\varphi(y,z) := \sup_{x \in \mathbb{R}^n} \left\{ z^\mathsf{T} x - L(x,y) \right\}, \quad (y,z) \in Y \times \mathbb{R}^n.$$
(7.40)

Note that the problem

$$\max_{y \in Y} \{-\varphi(y,0)\}\tag{7.41}$$

⁴Recall that \mathbb{I}_X denotes the indicator function of the set X.

coincides with the problem (7.29). Note also that for every $y \in Y$ the function $\varphi(y, \cdot)$ is the conjugate of $L(\cdot, y)$. Suppose that for every $y \in Y$ the function $L(\cdot, y)$ is convex and lsc. Then by the Fenchel-Moreau Theorem we have that the conjugate of the conjugate of $L(\cdot, y)$ coincides with $L(\cdot, y)$. Consequently the dual of (7.41), of the form (7.35), coincides with the problem (7.28). This leads to the following result.

Theorem 7.11. Let Y be an abstract vector space and $L : \mathbb{R}^n \times Y \to \overline{\mathbb{R}}$. Suppose that: (i) for every $x \in \mathbb{R}^n$ the function $L(x, \cdot)$ is concave, (ii) for every $y \in Y$ the function $L(\cdot, y)$ is convex and lower semicontinuous, (iii) problem (7.28) has a nonempty and bounded set of optimal solutions. Then the optimal values of problems (7.28) and (7.29) are equal to each other.

Proof. Consider function $\varphi(y, z)$, defined in (7.40), and the corresponding optimal value function

$$\vartheta(z) := \inf_{y \in Y} \varphi(y, z). \tag{7.42}$$

Since $\varphi(y, z)$ is given by maximum of convex in (y, z) functions, it is convex, and hence $\vartheta(z)$ is also convex. We have that $-\vartheta(0)$ is equal to the optimal value of the problem (7.29) and $-\vartheta^{**}(0)$ is equal to the optimal value of (7.28). We also have that

$$\vartheta^*(z^*) = \sup_{y \in Y} L(z^*, y),$$

and (see (7.24))

$$\partial \vartheta^{**}(0) = -\arg\min_{z^* \in \mathbb{R}^n} \vartheta^*(z^*) = -\arg\min_{z^* \in \mathbb{R}^n} \left\{ \sup_{y \in Y} L(z^*, y) \right\}.$$

That is, $-\partial \vartheta^{**}(0)$ coincides with the set of optimal solutions of the problem (7.28). It follows by assumption (iii) that $\partial \vartheta^{**}(0)$ is nonempty and bounded. Since $\vartheta^{**} : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a convex function, this in turn implies that $\vartheta^{**}(\cdot)$ is continuous in a neighborhood of $0 \in \mathbb{R}^n$. It follows that $\vartheta(\cdot)$ is also continuous in a neighborhood of $0 \in \mathbb{R}^n$, and hence $\vartheta^{**}(0) = \vartheta(0)$. This completes the proof. \Box

Remark 52. Note that it follows from the lower semicontinuity of $L(\cdot, y)$ that the maxfunction $f(x) = \sup_{y \in Y} L(x, y)$ is also lsc. Indeed, the epigraph of $f(\cdot)$ is given by the intersection of the epigraphs of $L(\cdot, y)$, $y \in Y$, and hence is closed. Therefore, if in addition, the set $X \subset \mathbb{R}^n$ is compact and problem (7.28) has a finite optimal value, then the set of optimal solutions of (7.28) is nonempty and compact, and hence bounded.

Hoffman's Lemma

The following result about Lipschitz continuity of linear systems is known as Hoffman's lemma. For a vector $a = (a_1, ..., a_m)^{\mathsf{T}} \in \mathbb{R}^m$, we use notation $(a)_+$ componentwise, i.e., $(a)_+ := ([a_1]_+, ..., [a_m]_+)^{\mathsf{T}}$, where $[a_i]_+ := \max\{0, a_i\}$.

Theorem 7.12 (Hoffman). Consider the multifunction $\mathcal{M}(b) := \{x \in \mathbb{R}^n : Ax \leq b\}$, where A is a given $m \times n$ matrix. Then there exists a positive constant κ , depending on A,

such that for any $x \in \mathbb{R}^n$ and any $b \in \text{dom } \mathcal{M}$,

$$\operatorname{dist}(x, \mathcal{M}(b)) \le \kappa \| (Ax - b)_+ \|.$$
(7.43)

Proof. Suppose that $b \in \text{dom } \mathcal{M}$, i.e., the system $Ax \leq b$ has a feasible solution. Note that for any $a \in \mathbb{R}^n$ we have that $||a|| = \sup_{||z||^* \leq 1} z^{\mathsf{T}}a$, where $|| \cdot ||^*$ is the dual of the norm $|| \cdot ||$. Then we have

$$\operatorname{dist}(x, \mathcal{M}(b)) = \inf_{x' \in \mathcal{M}(b)} \|x - x'\| = \inf_{Ax' \le b} \sup_{\|z\|^* \le 1} z^{\mathsf{T}}(x - x') = \sup_{\|z\|^* \le 1} \inf_{Ax' \le b} z^{\mathsf{T}}(x - x'),$$

where the interchange of the 'min' and 'max' operators can be justified, for example, by applying theorem 7.11 (see Remark 52 on page 424). By making change of variables y = x - x' and using linear programming duality we obtain

$$\inf_{Ax' \le b} z^{\mathsf{T}}(x - x') = \inf_{Ay \ge Ax - b} z^{\mathsf{T}}y = \sup_{\lambda \ge 0, A^{\mathsf{T}}\lambda = z} \lambda^{\mathsf{T}}(Ax - b).$$

It follows that

$$\operatorname{dist}(x, \mathcal{M}(b)) = \sup_{\lambda \ge 0, \, \|A^{\mathsf{T}}\lambda\|^* \le 1} \lambda^{\mathsf{T}} (Ax - b).$$
(7.44)

Since any two norms on \mathbb{R}^n are equivalent, we can assume without loss of generality that $\|\cdot\|$ is the ℓ_1 norm, and hence its dual is the ℓ_{∞} norm. For such choice of a polyhedral norm, we have that the set $S := \{\lambda : \lambda \ge 0, \|A^{\mathsf{T}}\lambda\|^* \le 1\}$ is polyhedral. We obtain that the right hand side of (7.44) is given by a maximization of a linear function over the polyhedral set S and has a finite optimal value (since the left hand side of (7.44) is finite), and hence has an optimal solution $\overline{\lambda}$. It follows that

$$\operatorname{dist}(x,\mathcal{M}(b)) = \bar{\lambda}^{\mathsf{T}}(Ax-b) \leq \bar{\lambda}^{\mathsf{T}}(Ax-b)_{+} \leq \|\bar{\lambda}\|^{*} \|(Ax-b)_{+}\|.$$

It remains to note that the polyhedral set S depends only on A, and can be represented as the direct sum $S = S_0 + C$ of a bounded polyhedral set S_0 and a polyhedral cone C, and that optimal solution $\overline{\lambda}$ can be taken to be an extreme point of the polyhedral set S_0 . Consequently, (7.43) follows with $\kappa := \max_{\lambda \in S_0} \|\lambda\|^*$. \Box

The term $||(Ax - b)_+||$, in the right hand side of (7.43), measures the infeasibility of the point x.

Consider now the following linear programming problem

$$\underset{x \in \mathbb{R}^n}{\operatorname{Min}} c^{\mathsf{T}} x \text{ subject to } Ax \le b.$$
(7.45)

A slight variation of the proof of Hoffman's lemma leads to the following result.

Theorem 7.13. Let S(b) be the set of optimal solutions of problem (7.45). Then there exists a positive constant γ , depending only on A, such that for any $b, b' \in \text{dom } S$ and any $x \in S(b)$,

$$\operatorname{dist}(x, \mathcal{S}(b')) \le \gamma \|b - b'\|. \tag{7.46}$$

Proof. Problem (7.45) can be written in the following equivalent form:

$$\underset{t \in \mathbb{R}}{\operatorname{Min}} t \text{ subject to } Ax \le b, \ c^{\mathsf{T}}x - t \le 0.$$
(7.47)

Denote by $\mathcal{M}(b)$ the set of feasible points of problem (7.47), i.e.,

$$\mathcal{M}(b) := \left\{ (x,t) : Ax \le b, \ c^{\mathsf{T}}x - t \le 0 \right\}.$$

Let $b, b' \in \operatorname{dom} S$ and consider a point $(x, t) \in \mathcal{M}(b)$. Proceeding as in the proof of Theorem 7.12 we can write

dist
$$((x,t), \mathcal{M}(b')) = \sup_{\|(z,a)\|^* \le 1} \inf_{Ax' \le b', \ c^{\mathsf{T}}x' \le t'} z^{\mathsf{T}}(x-x') + a(t-t').$$

By changing variables y = x - x' and s = t - t' and using linear programming duality, we have

$$\inf_{Ax' \le b', c^{\mathsf{T}}x' \le t'} z^{\mathsf{T}}(x-x') + a(t-t') = \sup_{\lambda \ge 0, A^{\mathsf{T}}\lambda + ac = z} \lambda^{\mathsf{T}}(Ax-b') + a(c^{\mathsf{T}}x-t)$$

for $a \ge 0$, and for a < 0 the above minimum is $-\infty$. By using ℓ_1 norm $\|\cdot\|$, and hence ℓ_{∞} norm $\|\cdot\|^*$, we obtain that

dist
$$((x,t), \mathcal{M}(b')) = \overline{\lambda}^{\mathsf{T}}(Ax - b') + \overline{a}(c^{\mathsf{T}}x - t),$$

where $(\bar{\lambda}, \bar{a})$ is an optimal solution of the problem

$$\max_{\lambda \ge 0, a \ge 0} \lambda^{\mathsf{T}} (Ax - b') + a(c^{\mathsf{T}}x - t) \text{ subject to } \|A^{\mathsf{T}}\lambda + ac\|^* \le 1, \ a \le 1.$$
(7.48)

By normalizing c we can assume without loss of generality that $||c||^* \leq 1$. Then by replacing the constraint $||A^T\lambda + ac||^* \leq 1$ with the constraint $||A^T\lambda||^* \leq 2$ we increase the feasible set of problem (7.48), and hence increase its optimal value. Let $(\hat{\lambda}, \hat{a})$ be an optimal solution of the obtained problem. Note that $\hat{\lambda}$ can be taken to be an extreme point of the polyhedral set $S := \{\lambda : ||A^T\lambda||^* \leq 2\}$. The polyhedral set S depends only on A and has a finite number of extreme points. Therefore $||\hat{\lambda}||^*$ can be bounded by a constant γ which depends only on A. Since $(x, t) \in \mathcal{M}(b)$, and hence $Ax - b \leq 0$ and $c^Tx - t \leq 0$, we have

$$\hat{\lambda}^{\mathsf{T}}(Ax-b') = \hat{\lambda}^{\mathsf{T}}(Ax-b) + \hat{\lambda}^{\mathsf{T}}(b-b') \le \hat{\lambda}^{\mathsf{T}}(b-b') \le \|\hat{\lambda}\|^* \|b-b'\|$$

and $\hat{a}(c^{\mathsf{T}}x - t) \leq 0$, and hence

dist
$$((x,t), \mathcal{M}(b')) \le \|\hat{\lambda}\|^* \|b - b'\| \le \gamma \|b - b'\|.$$
 (7.49)

The above inequality implies (7.46).

7.1.4 Optimality Conditions

Consider optimization problem

$$\min_{x \in X} f(x),$$
(7.50)

where $X \subset \mathbb{R}^n$ and $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is an extended real valued function.

First Order Optimality Conditions

Convex case. Suppose that the function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *convex*. It follows immediately from the definition of the subdifferential that if $f(\bar{x})$ is finite for some point $\bar{x} \in \mathbb{R}^n$, then $f(x) \ge f(\bar{x})$ for all $x \in \mathbb{R}^n$ iff

$$0 \in \partial f(\bar{x}). \tag{7.51}$$

That is, condition (7.51) is necessary and sufficient for the point \bar{x} to be a (global) minimizer of f(x) over $x \in \mathbb{R}^n$.

Suppose, further, that the set $X \subset \mathbb{R}^n$ is convex and closed and the function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper and convex, and consider a point $\overline{x} \in X \cap \text{dom} f$. It follows that the function $\overline{f}(x) := f(x) + \mathbb{I}_X(x)$ is convex, and of course the point \overline{x} is an optimal solution of the problem (7.50) iff \overline{x} is a (global) minimizer of $\overline{f}(x)$. Suppose that

$$\operatorname{ri}(X) \cap \operatorname{ri}(\operatorname{dom} f) \neq \emptyset. \tag{7.52}$$

Then by the Moreau-Rockafellar Theorem we have that $\partial \bar{f}(\bar{x}) = \partial f(\bar{x}) + \partial \mathbb{I}_X(\bar{x})$. Recalling that $\partial \mathbb{I}_X(\bar{x}) = \mathcal{N}_X(\bar{x})$, we obtain that \bar{x} is an optimal solution of problem (7.50) iff

$$0 \in \partial f(\bar{x}) + \mathcal{N}_X(\bar{x}), \tag{7.53}$$

provided that the regularity condition (7.52) holds. Note that (7.52) holds, in particular, if $\bar{x} \in int(dom f)$.

Nonconvex case. Assume that the function $f : \mathbb{R}^n \to \mathbb{R}$ is real valued continuously differentiable and the set X is closed (not necessarily convex).

Definition 7.14. The contingent (Bouligand) cone to X at $x \in X$, denoted $\mathcal{T}_X(x)$, is formed by vectors $h \in \mathbb{R}^n$ such that there exist sequences $h_k \to h$ and $t_k \downarrow 0$ such that $x + t_k h_k \in X$.

Note that $\mathcal{T}_X(x)$ is nonempty only if $x \in X$. If the set X is convex, then the contingent cone $\mathcal{T}_X(x)$ coincides with the corresponding tangent cone. We have the following simple necessary condition for a point $\bar{x} \in X$ to be a locally optimal solution of problem (7.50).

Proposition 7.15. Let $\bar{x} \in X$ be a locally optimal solution of problem (7.50). Then

$$h^{\mathsf{T}} \nabla f(\bar{x}) \ge 0, \quad \forall h \in \mathcal{T}_X(\bar{x}).$$
 (7.54)

Proof. Consider $h \in \mathcal{T}_X(\bar{x})$ and let $h_k \to h$ and $t_k \downarrow 0$ be sequences such that $x_k := \bar{x} + t_k h_k \in X$. Since $\bar{x} \in X$ is a local minimizer of f(x) over $x \in X$, we have that $f(x_k) - f(\bar{x}) \ge 0$. We also have that

$$f(x_k) - f(\bar{x}) = t_k h^{\mathsf{T}} \nabla f(\bar{x}) + o(t_k),$$

and hence (7.54) follows.

Condition (7.54) means that $\nabla f(\bar{x}) \in -[\mathcal{T}_X(\bar{x})]^*$. If the set X is convex, then the polar $[\mathcal{T}_X(\bar{x})]^*$ of the tangent cone $\mathcal{T}_X(\bar{x})$ coincides with the normal cone $\mathcal{N}_X(\bar{x})$. Therefore, if $f(\cdot)$ is convex and differentiable and X is convex, then optimality conditions (7.53) and (7.54) are equivalent.

Suppose now that the set X is given in the following form

$$X := \{ x \in \mathbb{R}^n : G(x) \in K \},$$
(7.55)

where $G(\cdot) = (g_1(\cdot), ..., g_m(\cdot)) : \mathbb{R}^n \to \mathbb{R}^m$ is a continuously differentiable mapping and $K \subset \mathbb{R}^m$ is a closed convex cone. In particular, if $K := \{0_q\} \times \mathbb{R}^{m-q}$, where $0_q \in \mathbb{R}^q$ is the null vector and $\mathbb{R}^{m-q}_{-} = \{y \in \mathbb{R}^{m-q} : y \leq 0\}$, then formulation (7.55) becomes

$$X = \{x \in \mathbb{R}^n : g_i(x) = 0, \ i = 1, ..., q, \ g_i(x) \le 0, \ i = q + 1, ..., m\}.$$
(7.56)

Under some regularity conditions (called *constraint qualifications*) we have the following formula for the contingent cone $\mathcal{T}_X(\bar{x})$ at a feasible point $\bar{x} \in X$:

$$\mathcal{T}_X(\bar{x}) = \left\{ h \in \mathbb{R}^n : [\nabla G(\bar{x})] h \in \mathcal{T}_K(G(\bar{x})) \right\},\tag{7.57}$$

where $\nabla G(\bar{x}) = [\nabla g_1(\bar{x}), ..., \nabla g_m(\bar{x})]^T$ is the corresponding $m \times n$ Jacobian matrix. The following condition is called *Robinson constraint qualification*

$$\nabla G(\bar{x})]\mathbb{R}^n + \mathcal{T}_K(G(\bar{x})) = \mathbb{R}^m.$$
(7.58)

If the cone K has a nonempty interior, Robinson constraint qualification is equivalent to the following condition

$$\exists h: \ G(\bar{x}) + [\nabla G(\bar{x})]h \in \operatorname{int}(K).$$
(7.59)

In case X is given in the form (7.56), Robinson constraint qualification is equivalent to the *Mangasarian-Fromovitz constraint qualification:*

$$\nabla g_i(\bar{x}), \quad i = 1, ..., q, \text{ are linearly independent,} \exists h : \quad h^{\mathsf{T}} \nabla g_i(\bar{x}) = 0, \quad i = 1, ..., q, \\ \quad h^{\mathsf{T}} \nabla g_i(\bar{x}) < 0, \quad i \in \mathcal{I}(\bar{x}), \end{cases}$$
(7.60)

where $\mathcal{I}(\bar{x}) := \{i \in \{q+1, ..., m\} : g_i(\bar{x}) = 0\}$ denotes the set of active at \bar{x} inequality constraints.

Consider the Lagrangian

$$L(x,\lambda) := f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

associated with problem (7.50) and the constraint mapping G(x). Under a constraint qualification ensuring validity of formula (7.57), first order necessary optimality condition (7.54) can be written in the following dual form: there exists a vector $\lambda \in \mathbb{R}^m$ of Lagrange multipliers such that

$$\nabla_x L(\bar{x}, \lambda) = 0, \quad G(\bar{x}) \in K, \quad \lambda \in K^*, \quad \lambda^{\top} G(\bar{x}) = 0.$$
(7.61)

Denote by $\Lambda(\bar{x})$ the set of Lagrange multipliers vectors λ satisfying (7.61).

Theorem 7.16. Let \bar{x} be a locally optimal solution of problem (7.50). Then the set $\Lambda(\bar{x})$ Lagrange multipliers is nonempty and bounded iff Robinson constraint qualification holds.

In particular, if $\Lambda(\bar{x})$ is a singleton (i.e., there exists unique Lagrange multiplier vector), then Robinson constraint qualification holds. If the set X is defined by a finite number of constraints in the form (7.56), then optimality conditions (7.61) are often referred to as the Karush-Kuhn-Tucker (KKT) necessary optimality conditions.

Second Order Optimality Conditions

We assume in this section that the function f(x) is real valued twice continuously differentiable and denote by $\nabla^2 f(x)$ the Hessian matrix of second order partial derivatives of fat x. Let \bar{x} be a locally optimal solution of problem (7.50). Consider the set (cone)

$$C(\bar{x}) := \left\{ h \in \mathcal{T}_X(\bar{x}) : h^\mathsf{T} \nabla f(\bar{x}) = 0 \right\}.$$

$$(7.62)$$

The cone $C(\bar{x})$ represents those feasible directions along which the first order approximation of f(x) at \bar{x} is zero, and is called the *critical cone*. The set

$$\mathcal{T}_X^2(x,h) := \left\{ z \in \mathbb{R}^n : \text{dist}\left(x + th + \frac{1}{2}t^2z, X\right) = o(t^2), \ t \ge 0 \right\}$$
(7.63)

is called the (inner) second order tangent set to X at the point $x \in X$ in the direction h. That is, the set $\mathcal{T}_X^2(x,h)$ is formed by vectors z such that $x + th + \frac{1}{2}t^2z + r(t) \in X$ for some $r(t) = o(t^2), t \ge 0$. Note that this implies that $x + th + o(t) \in X$, and hence $\mathcal{T}_X^2(x,h)$ can be nonempty only if $h \in \mathcal{T}_X(x)$.

Proposition 7.17. Let \bar{x} be a locally optimal solution of problem (7.50). Then⁵

$$h^{\mathsf{T}} \nabla^2 f(\bar{x}) h - \mathsf{S}\left(-\nabla f(\bar{x}), \mathcal{T}_X^2(\bar{x}, h)\right) \ge 0, \quad \forall h \in C(\bar{x}).$$

$$(7.64)$$

Proof. For some $h \in C(\bar{x})$ and $z \in \mathcal{T}_X^2(\bar{x}, h)$ consider the (parabolic) curve $x(t) := \bar{x} + th + \frac{1}{2}t^2z$. By the definition of the second order tangent set, we have that there exists $r(t) = o(t^2)$ such that $x(t) + r(t) \in X$, $t \ge 0$. It follows by local optimality of \bar{x} that $f(x(t) + r(t)) - f(\bar{x}) \ge 0$ for all $t \ge 0$ small enough. Since $r(t) = o(t^2)$, by the second order Taylor expansion we have

$$f(x(t) + r(t)) - f(\bar{x}) = th^{\mathsf{T}} \nabla f(\bar{x}) + \frac{1}{2}t^{2} \left[z^{\mathsf{T}} \nabla f(\bar{x}) + h^{\mathsf{T}} \nabla^{2} f(\bar{x}) h \right] + o(t^{2}).$$

Since $h \in C(\bar{x})$ the first term in the right hand side of the above equation vanishes. It follows that

$$z^{\mathsf{T}} \nabla f(\bar{x}) + h^{\mathsf{T}} \nabla^2 f(\bar{x}) h \ge 0, \quad \forall h \in C(\bar{x}), \; \forall z \in \mathcal{T}_X^2(\bar{x}, h).$$
(7.65)

⁵Recall that $S(v, A) = \sup_{z \in A} z^{\mathsf{T}} v$ denotes the support function of set A.

Condition (7.65) can be written in the form:

$$\inf_{z \in \mathcal{T}_X^2(\bar{x},h)} \left\{ z^\mathsf{T} \nabla f(\bar{x}) + h^\mathsf{T} \nabla^2 f(\bar{x})h \right\} \ge 0, \quad \forall h \in C(\bar{x}).$$
(7.66)

Since

$$\inf_{z \in \mathcal{T}_X^2(\bar{x},h)} z^{\mathsf{T}} \nabla f(\bar{x}) = -\sup_{z \in \mathcal{T}_X^2(\bar{x},h)} z^{\mathsf{T}}(-\nabla f(\bar{x})) = -\mathsf{s}\left(-\nabla f(\bar{x}), \mathcal{T}_X^2(\bar{x},h)\right),$$

the second order necessary conditions (7.65) can be written in the form (7.64).

If the set X is polyhedral, then for $\bar{x} \in X$ and $h \in \mathcal{T}_X(\bar{x})$ the second order tangent set $\mathcal{T}_X^2(\bar{x}, h)$ is equal to the sum of $\mathcal{T}_X(\bar{x})$ and the linear space generated by vector h. Since for $h \in C(\bar{x})$ we have that $h^{\mathsf{T}} \nabla f(\bar{x}) = 0$ and because of the first order optimality conditions (7.54), it follows that if the set X is polyhedral, then the term $\mathsf{S}\left(-\nabla f(\bar{x}), \mathcal{T}_X^2(\bar{x}, h)\right)$ in (7.64) vanishes. In general, this term is nonpositive and corresponds to a curvature of the set X at \bar{x} .

If the set X is given in the form (7.55) with the mapping G(x) being twice continuously differentiable, then the second order optimality conditions (7.64) can be written in the following dual form.

Theorem 7.18. Let \bar{x} be a locally optimal solution of problem (7.50). Suppose that Robinson constraint qualification (7.58) is fulfilled. Then the following second order necessary conditions hold

$$\sup_{\lambda \in \Lambda(\bar{x})} \left\{ h^{\mathsf{T}} \nabla_{xx}^2 L(\bar{x}, \lambda) h - \mathsf{s}(\lambda, \mathfrak{T}(h)) \right\} \ge 0, \quad \forall h \in C(\bar{x}),$$
(7.67)

where $\mathfrak{T}(h) := \mathcal{T}_K^2(G(\bar{x}), [\nabla G(\bar{x})]h).$

Note that if the cone K is polyhedral, then the curvature term $S(\lambda, \mathfrak{T}(h))$ in (7.67) vanishes. In general, $S(\lambda, \mathfrak{T}(h)) \leq 0$ and the second order necessary conditions (7.67) are stronger than the "standard" second order conditions:

$$\sup_{\lambda \in \Lambda(\bar{x})} h^{\mathsf{T}} \nabla_{xx}^2 L(\bar{x}, \lambda) h \ge 0, \quad \forall h \in C(\bar{x}).$$
(7.68)

Second order sufficient conditions. Consider condition

$$h^{\mathsf{T}} \nabla^2 f(\bar{x}) h - \mathsf{S}\left(-\nabla f(\bar{x}), \mathcal{T}_X^2(\bar{x}, h)\right) > 0, \quad \forall h \in C(\bar{x}), \ h \neq 0.$$
 (7.69)

This condition is obtained from the second order necessary condition (7.64) by replacing " ≥ 0 " sign with the strict inequality sign "> 0". Necessity of second order conditions (7.64) was derived by verifying optimality of \bar{x} along parabolic curves. There is no reason *a priori* that verification of (local) optimality along parabolic curves is sufficient to ensure local optimality of \bar{x} . Therefore in order to verify sufficiency of condition (7.69) we need an additional condition.

Definition 7.19. It is said that the set X is second order regular at $\bar{x} \in X$ if for any sequence $x_k \in X$ of the form $x_k = \bar{x} + t_k h + \frac{1}{2} t_k^2 r_k$, where $t_k \downarrow 0$ and $t_k r_k \to 0$, it follows that

$$\lim_{k \to \infty} \operatorname{dist} \left(r_k, \mathcal{T}_X^2(\bar{x}, h) \right) = 0.$$
(7.70)

Note that in the above definition the term $\frac{1}{2}t_k^2r_k = o(t_k)$, and hence such a sequence $x_k \in X$ can exist only if $h \in \mathcal{T}_X(\bar{x})$. It turns out that second order regularity can be verified in many interesting cases. In particular, any polyhedral set is second order regular, the cone of positive semidefinite symmetric matrices is second order regular, etc. We refer to [26, section 3.3] for a discussion of this concept.

Recall that it is said that the *quadratic growth condition* holds at $\bar{x} \in X$ if there exist constant c > 0 and a neighborhood N of \bar{x} such that

$$f(x) \ge f(\bar{x}) + c \|x - \bar{x}\|^2, \ \forall x \in X \cap N.$$
(7.71)

Of course, the quadratic growth condition implies that \bar{x} is a locally optimal solution of problem (7.50).

Proposition 7.20. Let $\bar{x} \in X$ be a feasible point of problem (7.50) satisfying first order necessary conditions (7.54). Suppose that X is second order regular at \bar{x} . Then the second order conditions (7.69) are necessary and sufficient for the quadratic growth at \bar{x} to hold.

Proof. Suppose that conditions (7.69) hold. In order to verify the quadratic growth condition we argue by a contradiction, so suppose that it does not hold. Then there exists a sequence $x_k \in X \setminus {\bar{x}}$ converging to \bar{x} and a sequence $c_k \downarrow 0$ such that

$$f(x_k) - f(\bar{x}) \le c_k \|x_k - \bar{x}\|^2.$$
(7.72)

Denote $t_k := ||x_k - \bar{x}||$ and $h_k := t_k^{-1}(x_k - \bar{x})$. By passing to a subsequence if necessary we can assume that h_k converges to a vector h. Clearly $h \neq 0$ and by the definition of $\mathcal{T}_X(\bar{x})$ it follows that $h \in \mathcal{T}_X(\bar{x})$. Moreover, by (7.72) we have

$$c_k t_k^2 \ge f(x_k) - f(\bar{x}) = t_k h^\mathsf{T} \nabla f(\bar{x}) + o(t_k),$$

and hence $h^{\mathsf{T}} \nabla f(\bar{x}) \leq 0$. Because of the first order necessary conditions it follows that $h^{\mathsf{T}} \nabla f(\bar{x}) = 0$, and hence $h \in C(\bar{x})$.

Denote $r_k := 2t_k^{-1}(h_k - h)$. We have that $x_k = \bar{x} + t_k h + \frac{1}{2}t_k^2 r_k \in X$ and $t_k r_k \to 0$. Consequently it follows by the second order regularity that there exists a sequence $z_k \in \mathcal{T}_X^2(\bar{x}, h)$ such that $r_k - z_k \to 0$. Since $h^\mathsf{T} \nabla f(\bar{x}) = 0$, by the second order Taylor expansion we have

$$f(x_k) = f(\bar{x} + t_k h + \frac{1}{2}t_k^2 r_k) = f(\bar{x}) + \frac{1}{2}t_k^2 \left[z_k^{\mathsf{T}} \nabla f(\bar{x}) + h^{\mathsf{T}} \nabla^2 f(\bar{x}) h \right] + o(t_k^2).$$

Moreover, since $z_k \in \mathcal{T}^2_X(\bar{x}, h)$ we have that

$$z_k^{\mathsf{T}} \nabla f(\bar{x}) + h^{\mathsf{T}} \nabla^2 f(\bar{x}) h \ge c,$$

where c is equal to the left hand side of (7.69), which by the assumption is positive. It follows that

$$f(x_k) \ge f(\bar{x}) + \frac{1}{2}c \|x_k - \bar{x}\|^2 + o(\|x_k - \bar{x}\|^2),$$

a contradiction with (7.72).

Conversely, suppose that the quadratic growth condition holds at \bar{x} . It follows that the function $\phi(x) := f(x) - \frac{1}{2}c||x - \bar{x}||^2$ also attains its local minimum over X at \bar{x} . Note that $\nabla \phi(\bar{x}) = \nabla f(\bar{x})$ and $h^T \nabla^2 \phi(\bar{x})h = h^T \nabla^2 f(\bar{x})h - c||h||^2$. Therefore, by the second order necessary conditions (7.64), applied to the function ϕ , it follows that the left hand side of (7.69) is greater than or equal to $c||h||^2$. This completes the proof. \Box

If the set X is given in the form (7.55), then similar to Theorem 7.18 it is possible to formulate second order sufficient conditions (7.69) in the following dual form.

Theorem 7.21. Let $\bar{x} \in X$ be a feasible point of problem (7.50) satisfying first order necessary conditions (7.61). Suppose that Robinson constraint qualification (7.58) is fulfilled and the set (cone) K is second order regular at $G(\bar{x})$. Then the following conditions are necessary and sufficient for the quadratic growth at \bar{x} to hold:

$$\sup_{\lambda \in \Lambda(\bar{x})} \left\{ h^{\mathsf{T}} \nabla_{xx}^2 L(\bar{x}, \lambda) h - \mathsf{S}(\lambda, \mathfrak{T}(h)) \right\} > 0, \quad \forall h \in C(\bar{x}), \ h \neq 0,$$
(7.73)

where $\mathfrak{T}(h) := \mathcal{T}_K^2(G(\bar{x}), [\nabla G(\bar{x})]h).$

Note again that if the cone K is polyhedral, then K is second order regular and the curvature term $S(\lambda, \mathfrak{T}(h))$ in (7.73) vanishes.

7.1.5 Perturbation Analysis

Continuity Properties of Optimal Value Functions

Consider optimization problem

$$\underset{x \in \Theta(u)}{\operatorname{Min}} g(x, u),$$
(7.74)

depending on parameter $u \in U$. Here U is a finite dimensional vector space or even more generally a metric space, $g : \mathbb{R}^n \times U \to \mathbb{R}$ and $\Theta : U \rightrightarrows \mathbb{R}^n$ is a multifunction (point-to-set mapping) assigning to $u \in U$ a set $\Theta(u) \subset \mathbb{R}^n$. Typical example of the multifunction Θ is

$$\Theta(u) := \{ x \in \mathbb{R}^n : G(x, u) \in K \},$$
(7.75)

where $K \subset \mathbb{R}^m$ is a closed convex cone and $G : \mathbb{R}^n \times U \to \mathbb{R}^m$.

With problem (7.74) we associate the optimal value function

$$v(u) := \inf_{x \in \Theta(u)} g(x, u), \tag{7.76}$$

and the set of optimal solutions

$$\mathcal{S}(u) := \operatorname*{argmin}_{x \in \Theta(u)} g(x, u). \tag{7.77}$$

Definition 7.22. It is said that the inf-compactness condition holds if there exists a bounded set $C \subset \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ such that for all u in a neighborhood of u_0 the level set

$$\operatorname{lev}_{\alpha}g(\cdot, u) := \{x \in \Theta(u) : g(x, u) \le \alpha\}$$

is nonempty and contained in C.

Recall that it is said that the multifunction $\Theta(\cdot)$ is *closed* if $u_k \to u, x_k \in \Theta(u_k)$ and $x_k \to x$, then $x \in \Theta(u)$. If the multifunction $\Theta(\cdot)$ is closed, then it is closed valued, i.e., the set $\Theta(u) \subset \mathbb{R}^n$ is closed for every $u \in U$. For example, if $\Theta(\cdot)$ is given in the form (7.75) and the mapping $G(\cdot, \cdot)$ is continuous, then $\Theta(\cdot)$ is closed.

Theorem 7.23. Let u_0 be a point in U. Suppose that: (i) the function g(x, u) is continuous on $\mathbb{R}^n \times U$, (ii) the multifunction $\Theta(\cdot)$ is closed, (iii) the inf-compactness condition holds, (iv) for any neighborhood V of the set $S(u_0)$ there exists a neighborhood U of u_0 such that $V \cap \Theta(u) \neq \emptyset$ for all $u \in U$. Then the optimal value function v(u) is continuous at $u = u_0$ and for any $u_k \to u_0$ and $x_k \in S(u_k)$ it follows that $\operatorname{dist}(x_k, S(u_0))$ tends to zero.

Proof. Consider a sequence $\{u_k\}$ converging to the point u_0 . By the inf-compactness condition we have that there exists a bounded set $C \subset \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ such that the level sets $\operatorname{lev}_{\alpha}g(\cdot, u_k)$ are nonempty and contained in C for all k large enough. Since $g(\cdot, u)$ is continuous and the sets $\Theta(u_k)$ are closed, it follows that the sets $\mathcal{S}(u_0)$ and $\mathcal{S}(u_k)$ are nonempty, compact and contained in the set C for all k large enough. We argue now by a contradiction. Suppose that $\operatorname{dist}(x_k, \mathcal{S}(u_0))$ does not tend to zero for some sequence $x_k \in \mathcal{S}(u_k)$. By passing to a subsequence if necessary we can assume that the sequence x_k converges to a point $\bar{x} \in C$. It follows that $\bar{x} \notin \mathcal{S}(u_0)$. Since the multifunction $\Theta(\cdot)$ is closed, we have that $\bar{x} \in \Theta(u_0)$. Also since $v(u_k) = g(x_k, u_k)$ tends to $g(\bar{x}, u_0)$ and $g(\bar{x}, u_0) > v(u_0)$, we obtain that $\liminf_{k\to\infty} v(u_k) > v(u_0)$. On the other hand by assumption (iv) there is a sequence $x'_k \in \Theta(u_k)$ such that $\operatorname{dist}(x'_k, \mathcal{S}(u_0))$ tends to zero. By passing to a subsequence if necessary we can assume that x'_k converges to a point $x^* \in \mathcal{S}(u_0)$. Since $v(u_k) \leq g(x'_k, u_k)$ and $g(x'_k, u_k)$ tends to $g(x^*, u_0) = v(u_0)$, it follows that $\lim_{k \to \infty} v(u_k) > v(u_0)$. On the other hand by assumption (iv) there is a sequence $x'_k \in \Theta(u_k)$ such that $\operatorname{dist}(x'_k, \mathcal{S}(u_0))$ tends to zero. By passing to a subsequence if necessary we can assume that x'_k converges to a point $x^* \in \mathcal{S}(u_0)$. Since $v(u_k) \leq g(x'_k, u_k)$ and $g(x'_k, u_k)$ tends to $g(x^*, u_0) = v(u_0)$, it follows that $\lim_{k \to \infty} v(u_k) \leq v(u_0)$. This gives the required contradiction. \Box

Some remarks about assumptions of the above theorem are now in order. Assumption (iv) can be ensured by some type of constraint qualification. In particular, if $\Theta(\cdot)$ is given in the form (7.75) with the mapping G(x, u) being continuous on $\mathbb{R}^n \times U$ and convex in x with respect to -K, then the assumption (iv) holds provided that the Slater condition is satisfied for $u = u_0$, i.e., there is a point $\bar{x} \in \mathbb{R}^n$ such that $G(\bar{x}, u_0) \in \text{int}(K)$. If G(x, u)is differentiable in x and $\nabla_x G(x, u)$ is continuous, then Robinson constraint qualification can be applied to $G(\cdot, u_0)$.

If $\Theta(u) \equiv \Theta$ does not depend on $u \in U$, then assumption (ii) means that the set $\Theta \subset \mathbb{R}^n$ is closed. Also in that case assumption (iv) holds automatically provided the set $S(u_0)$ is nonempty, which in turn is ensured by the assumptions (i)–(iii).

Differentiability Properties of Optimal value Functions

We often have to deal with optimal value functions (max-functions) of the form

$$\phi(x) := \sup_{\theta \in \Theta} g(x, \theta), \tag{7.78}$$

where $g : \mathbb{R}^n \times \Theta \to \mathbb{R}$. In applications the set Θ usually is a subset of a finite dimensional vector space. At this point, however, this is not important and we can assume that Θ is an abstract metric (or even topological) space. Denote

$$\bar{\Theta}(x) := \arg \max_{\theta \in \Theta} g(x, \theta).$$

Let us first mention the following simple result.

Proposition 7.24. Suppose that for every $\theta \in \Theta$ the function $g(\cdot, \theta)$ is lower semicontinuous. Then the max-function $\phi(\cdot)$ is lower semicontinuous.

Proof. Recall that a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is lower semicontinuous iff its epigraph is a closed subset of \mathbb{R}^{n+1} . Note also that epigraph of the max-function $\phi(\cdot)$ is equal to the intersection of the epigraphs $\operatorname{epi}\{g(\cdot, \theta)\}, \theta \in \Theta$. Since intersection of a family of closed sets is a closed set, the result follows. \Box

The following result about directional differentiability of the max-function is often called Danskin Theorem.

Theorem 7.25 (Danskin). Let Θ be a nonempty, compact topological space and $g : \mathbb{R}^n \times \Theta \to \mathbb{R}$ be such that $g(\cdot, \theta)$ is differentiable for every $\theta \in \Theta$ and $\nabla_x g(x, \theta)$ is continuous on $\mathbb{R}^n \times \Theta$. Then the corresponding max-function $\phi(x)$ is locally Lipschitz continuous, directionally differentiable and

$$\phi'(x,h) = \sup_{\theta \in \bar{\Theta}(x)} h^{\mathsf{T}} \nabla_x g(x,\theta).$$
(7.79)

In particular, if for some $x \in \mathbb{R}^n$ the set $\overline{\Theta}(x) = {\overline{\theta}}$ is a singleton, then the max-function is differentiable at x and

$$\nabla \phi(x) = \nabla_x g(x, \theta). \tag{7.80}$$

In the convex case we have the following result giving a description of subdifferentials of max-functions.

Theorem 7.26 (Levin-Valadier). Let Θ be a nonempty compact topological space and $g : \mathbb{R}^n \times \Theta \to \mathbb{R}$ be a real valued function. Suppose that: (i) for every $\theta \in \Theta$ the function $g_{\theta}(\cdot) = g(\cdot, \theta)$ is convex on \mathbb{R}^n , (ii) for every $x \in \mathbb{R}^n$ the function $g(x, \cdot)$ is upper semicontinuous on Θ . Then the max-function $\phi(x)$ is convex real valued and

$$\partial \phi(x) = \operatorname{cl}\left\{\operatorname{conv}\left(\cup_{\theta \in \bar{\Theta}(x)} \partial g_{\theta}(x)\right)\right\}.$$
(7.81)

Let us make the following observations regarding the above theorem. Since Θ is compact and by the assumption (ii), we have that the set $\overline{\Theta}(x)$ is nonempty and compact. Since the function $\phi(\cdot)$ is convex real valued, it is subdifferentiable at every $x \in \mathbb{R}^n$ and its subdifferential $\partial \phi(x)$ is a convex, closed bounded subset of \mathbb{R}^n . It follows then from (7.81) that the set $A := \bigcup_{\theta \in \overline{\Theta}(x)} \partial g_{\theta}(x)$ is bounded. Suppose further that:

(iii) For every $x \in \mathbb{R}^n$ the function $g(x, \cdot)$ is continuous on Θ .

Then the set A is closed, and hence is compact. Indeed, consider a sequence $z_k \in A$. Then, by the definition of the set $A, z_k \in \partial g_{\theta_k}(x)$ for some sequence $\theta_k \in \overline{\Theta}(x)$. Since $\overline{\Theta}(x)$ is compact and A is bounded, by passing to a subsequence if necessary, we can assume that θ_k converges to a point $\overline{\theta} \in \overline{\Theta}(x)$ and z_k converges to a point $\overline{z} \in \mathbb{R}^n$. By the definition of subgradients z_k we have that for any $x' \in \mathbb{R}^n$ the following inequality holds

$$g_{\theta_k}(x') - g_{\theta_k}(x) \ge z_k^{\mathsf{I}}(x'-x).$$

By passing to the limit in the above inequality as $k \to \infty$, we obtain that $\bar{z} \in \partial g_{\bar{\theta}}(x)$. It follows that $\bar{z} \in A$, and hence A is closed. Now since convex hull of a compact subset of \mathbb{R}^n is also compact, and hence is closed, we obtain that if the assumption (ii) in the above theorem is strengthen to the assumption (iii), then the set inside the parentheses in (7.81) is closed, and hence formula (7.81) takes the form

$$\partial \phi(x) = \operatorname{conv}\left(\cup_{\theta \in \bar{\Theta}(x)} \partial g_{\theta}(x)\right).$$
 (7.82)

Second Order Perturbation Analysis

Consider the following parameterization of problem (7.50):

$$\min_{x \in X} f(x) + t\eta_t(x),$$
(7.83)

depending on parameter $t \in \mathbb{R}_+$. We assume that the set $X \subset \mathbb{R}^n$ is nonempty and compact and consider a convex compact set $U \subset \mathbb{R}^n$ such that $X \subset int(U)$. It follows, of course, that the set U has a nonempty interior. Consider the space $W^{1,\infty}(U)$ of Lipschitz continuous functions $\psi : U \to \mathbb{R}$ equipped with the norm

$$\|\psi\|_{1,U} := \sup_{x \in U} |\psi(x)| + \sup_{x \in U'} \|\nabla\psi(x)\|,$$
(7.84)

with $U' \subset \operatorname{int}(U)$ being the set of points where $\psi(\cdot)$ is differentiable. Recall that by Rademacher Theorem, a function $\psi(\cdot) \in W^{1,\infty}(U)$ is differentiable at almost every point of U. We assume that the functions $f(\cdot)$ and $\eta_t(\cdot)$, $t \in \mathbb{R}_+$, are Lipschitz continuous on U, i.e., $f, \eta_t \in W^{1,\infty}(U)$. We also assume that η_t converges (in the norm topology) to a function $\delta \in W^{1,\infty}(U)$, that is, $\|\eta_t - \delta\|_{1,U} \to 0$ as $t \downarrow 0$.

Denote by v(t) the optimal value and by $\tilde{x}(t)$ an optimal solution of (7.83), i.e.,

$$v(t):=\inf_{x\in X}\big\{f(x)+t\eta_t(x)\big\} \ \text{and} \ \ \tilde{x}(t)\in \arg\min_{x\in X}\big\{f(x)+t\eta_t(x)\big\}.$$

We will be interested in second order differentiability properties of v(t) and first order differentiability properties of $\tilde{x}(t)$ at t = 0. We assume that f(x) has unique minimizer \bar{x} over $x \in X$, i.e., the set of optimal solutions of the unperturbed problem (7.50) is the singleton $\{\bar{x}\}$. Moreover, we assume that $\delta(\cdot)$ is differentiable at \bar{x} and f(x) is twice continuously differentiable at \bar{x} . Since X is compact and the objective function of problem (7.83) is continuous, it has an optimal solution for any t.

The following result is taken from [26, section 4.10.3].

Theorem 7.27. Let \bar{x} be unique optimal solution of problem (7.50). Suppose that: (i) the set X is compact and second order regular at \bar{x} , (ii) η_t converges (in the norm topology) to $\delta \in W^{1,\infty}(U)$ as $t \downarrow 0$, (iii) $\delta(x)$ is differentiable at \bar{x} and f(x) is twice continuously differentiable at \bar{x} , (iv) the quadratic growth condition (7.71) holds. Then

$$v(t) = v(0) + t\eta_t(\bar{x}) + \frac{1}{2}t^2\mathfrak{V}_f(\delta) + o(t^2), \quad t \ge 0,$$
(7.85)

where $\mathfrak{V}_{f}(\delta)$ is the optimal value of the auxiliary problem

$$\underset{h \in C(\bar{x})}{\operatorname{Min}} \left\{ 2h^{\mathsf{T}} \nabla \delta(\bar{x}) + h^{\mathsf{T}} \nabla^2 f(\bar{x}) h - \mathsf{s} \left(-\nabla f(\bar{x}), \mathcal{T}_X^2(\bar{x}, h) \right) \right\}.$$
(7.86)

Moreover, if (7.86) has unique optimal solution \bar{h} , then

$$\tilde{x}(t) = \bar{x} + th + o(t), \ t \ge 0.$$
 (7.87)

Proof. Since the minimizer \bar{x} is unique and the set X is compact, it is not difficult to show that, under the specified assumptions, $\tilde{x}(t)$ tends to \bar{x} as $t \downarrow 0$. Moreover, we have that $\|\tilde{x}(t) - \bar{x}\| = O(t), t > 0$. Indeed, by the quadratic growth condition, for t > 0 small enough and some c > 0 it follows that

$$v(t) = f(\tilde{x}(t)) + t\eta_t(\tilde{x}(t)) \ge f(\bar{x}) + c \|\tilde{x}(t) - \bar{x}\|^2 + t\eta_t(\tilde{x}(t)).$$

Since $\bar{x} \in X$ we also have that $v(t) \leq f(\bar{x}) + t\eta_t(\bar{x})$. Consequently,

$$t |\eta_t(\tilde{x}(t)) - \eta_t(\bar{x})| \ge c \|\tilde{x}(t) - \bar{x}\|^2.$$

Moreover, $|\eta_t(\tilde{x}(t)) - \eta_t(\bar{x})| = O(||\tilde{x}(t) - \bar{x}||)$, and hence $||\tilde{x}(t) - \bar{x}|| = O(t)$.

Let $h \in C(\bar{x})$ and $w \in \mathcal{T}_X^2(\bar{x}, h)$. By the definition of the second order tangent set it follows that there is a path $x(t) \in X$ of the form $x(t) = \bar{x} + th + \frac{1}{2}t^2w + o(t^2)$. Since $x(t) \in X$ we have that $v(t) \leq f(x(t)) + t\eta_t(x(t))$. Moreover, by using the second order Taylor expansion of f(x) at $x = \bar{x}$ we have

$$f(x(t)) = f(\bar{x}) + th^{\mathsf{T}} \nabla f(\bar{x}) + \frac{1}{2} t^2 w^{\mathsf{T}} \nabla f(\bar{x}) + \frac{1}{2} h^{\mathsf{T}} \nabla^2 f(\bar{x}) h + o(t^2),$$

and since $h \in C(\bar{x})$ we have that $h^{\mathsf{T}} \nabla f(\bar{x}) = 0$. Also since $\|\eta_t - \delta\|_{1,\infty} \to 0$, we have by the Mean Value Theorem that

$$\eta_t(x(t)) - \delta(x(t)) = \eta_t(\bar{x}) - \delta(\bar{x}) + o(t),$$

and since $\delta(x)$ is differentiable at \bar{x} that

$$\delta(x(t)) = \delta(\bar{x}) + th^{\mathsf{T}} \nabla \delta(\bar{x}) + o(t).$$

Putting this all together and noting that $f(\bar{x}) = v(0)$ we obtain that

 $f(x(t)) + t\eta_t(x(t)) = v(0) + t\eta_t(\bar{x}) + t^2 h^{\mathsf{T}} \nabla \delta(\bar{x}) + \frac{1}{2} t^2 h^{\mathsf{T}} \nabla^2 f(\bar{x}) h + \frac{1}{2} t^2 w^{\mathsf{T}} \nabla f(\bar{x}) + o(t^2).$

Consequently

$$\limsup_{t \downarrow} \frac{v(t) - v(0) - t\eta_t(\bar{x})}{\frac{1}{2}t^2} \le 2h^{\mathsf{T}} \nabla \delta(\bar{x}) + h^{\mathsf{T}} \nabla^2 f(\bar{x})h + w^{\mathsf{T}} \nabla f(\bar{x}).$$
(7.88)

Since the above inequality (7.88) holds for any $w \in \mathcal{T}^2_X(\bar{x}, h)$, by taking minimum (with respect to w) in the right hand side of (7.88) we obtain for any $h \in C(\bar{x})$:

$$\limsup_{t\downarrow 0} \frac{v(t) - v(0) - t\eta_t(\bar{x})}{\frac{1}{2}t^2} \le 2h^{\mathsf{T}} \nabla \delta(\bar{x}) + h^{\mathsf{T}} \nabla^2 f(\bar{x})h - \mathsf{s}\big(-\nabla f(\bar{x}), \mathcal{T}_X^2(\bar{x}, h)\big).$$

In order to show the converse estimate we argue as follows. Consider a sequence $t_k \downarrow 0$ and $x_k := \tilde{x}(t_k)$. Since $\|\tilde{x}(t) - \bar{x}\| = O(t)$, we have that $(x_k - \bar{x})/t_k$ is bounded, and hence by passing to a subsequence if necessary we can assume that $(x_k - \bar{x})/t_k$ converges to a vector h. Since $x_k \in X$, it follows that $h \in \mathcal{T}_X(\bar{x})$. Moreover,

$$v(t_k) = f(x_k) + t_k \eta_{t_k}(x_k) = f(\bar{x}) + t_k h^{\mathsf{T}} \nabla f(\bar{x}) + t_k \delta(\bar{x}) + o(t_k),$$

and by Danskin Theorem $v'(0) = \delta(\bar{x})$. It follows that $h^{\mathsf{T}} \nabla f(\bar{x}) = 0$, and hence $h \in C(\bar{x})$. Consider $r_k := 2(x_k - \bar{x} - t_k h)/t_k^2$, i.e., r_k are such that $x_k = \bar{x} + t_k h + \frac{1}{2}t_k^2 r_k$. Note that $t_k r_k \to 0$ and $x_k \in X$ and hence, by the second order regularity of X, there exists $w_k \in \mathcal{T}^2_X(\bar{x},h)$ such that $||r_k - w_k|| \to 0$. Finally,

$$\begin{split} v(t_k) &= f(x_k) + t_k \eta_{t_k}(x_k) \\ &= f(\bar{x}) + t_k \eta_{t_k}(\bar{x}) + t_k^2 h^{\mathsf{T}} \nabla \delta(\bar{x}) + \frac{1}{2} t_k^2 h^{\mathsf{T}} \nabla^2 f(\bar{x}) h + \frac{1}{2} t_k^2 w_k^{\mathsf{T}} \nabla f(\bar{x}) + o(t_k^2) \\ &\geq v(0) + t_k \eta_{t_k}(\bar{x}) + t_k^2 h^{\mathsf{T}} \nabla \delta(\bar{x}) + \frac{1}{2} t_k^2 h^{\mathsf{T}} \nabla^2 f(\bar{x}) h \\ &+ \frac{1}{2} t_k^2 \inf_{w \in \mathcal{T}_X^2(\bar{x}, h)} w^{\mathsf{T}} \nabla f(\bar{x}) + o(t_k^2). \end{split}$$

It follows that

$$\liminf_{t\downarrow 0} \frac{v(t) - v(0) - t\eta_t(\bar{x})}{\frac{1}{2}t^2} \ge 2h^\mathsf{T} \nabla \delta(\bar{x}) + h^\mathsf{T} \nabla^2 f(\bar{x})h - \mathsf{s} \Big(-\nabla f(\bar{x}), \mathcal{T}_X^2(\bar{x}, h) \Big).$$

This completes the proof of (7.85).

Also by the above analysis we have that any accumulation point of $(\tilde{x}(t) - \bar{x})/t$, as $t \downarrow 0$, is an optimal solution of problem (7.86). Since $(\tilde{x}(t) - \bar{x})/t$ is bounded, the assertion (7.87) follows by compactness arguments. Π

As in the case of second order optimality conditions, we have here that if the set Xis polyhedral, then the curvature term $S(-\nabla f(\bar{x}), \mathcal{T}^2_X(\bar{x}, h))$ in (7.86) vanishes.

Suppose now that the set X is given in the form (7.55) with the mapping G(x) being twice continuously differentiable. Suppose further that Robinson constraint qualification (7.58), for the unperturbed problem, holds. Then the optimal value of problem (7.86) can be written in the following dual form

$$\mathfrak{V}_{f}(\delta) = \inf_{h \in C(\bar{x})} \sup_{\lambda \in \Lambda(\bar{x})} \left\{ 2h^{\mathsf{T}} \nabla \delta(\bar{x}) + h^{\mathsf{T}} \nabla_{xx}^{2} L(\bar{x}, \lambda) h - \mathsf{s}(\lambda, \mathfrak{T}(h)) \right\},$$
(7.89)

where $\mathfrak{T}(h) := \mathcal{T}_K^2(G(\bar{x}), [\nabla G(\bar{x})]h)$. Note again that if the set K is polyhedral, then the curvature term $S(\lambda, \mathfrak{T}(h))$ in (7.89) vanishes.

Minimax Problems

In this section we consider the following minimax problem

$$\operatorname{Min}_{x \in X} \left\{ \phi(x) := \sup_{y \in Y} f(x, y) \right\},$$
(7.90)

and its dual

$$\operatorname{Max}_{y \in Y} \left\{ \iota(y) := \inf_{x \in X} f(x, y) \right\}.$$
(7.91)

We assume that the sets $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ are convex and compact, and the function $f: X \times Y \to \mathbb{R}$ is continuous⁶, i.e., $f \in C(X, Y)$. Moreover, assume that f(x, y) is convex in $x \in X$ and concave in $y \in Y$. Under these conditions there is no duality gap between problems (7.90) and (7.91), i.e., the optimal values of these problems are equal to each other. Moreover, the max-function $\phi(x)$ is continuous on X and problem (7.90) has a nonempty set of optimal solutions, denoted X^* , the min-function $\iota(y)$ is continuous on Y and problem (7.91) has a nonempty set of optimal solutions, denoted Y^* , and $X^* \times Y^*$ forms the set of saddle points of the minimax problems (7.90) and (7.91).

Consider the following perturbation of the minimax problem (7.90):

$$\min_{x \in X} \sup_{y \in Y} \{ f(x, y) + t\eta_t(x, y) \},$$
(7.92)

where $\eta_t \in C(X, Y)$, $t \ge 0$. Denote by v(t) the optimal value of the above parameterized problem (7.92). Clearly v(0) is the optimal value of the unperturbed problem (7.90). We assume that η_t converges uniformly (i.e., in the sup-norm) as $t \downarrow 0$ to a function $\gamma \in C(X, Y)$, that is

$$\lim_{t \downarrow 0} \sup_{x \in X, y \in Y} \left| \eta_t(x, y) - \gamma(x, y) \right| = 0.$$

Theorem 7.28. Suppose that: (i) the sets $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ are convex and compact, (ii) for all $t \ge 0$ the function $\zeta_t := f + t\eta_t$ is continuous on $X \times Y$, convex in $x \in X$ and concave in $y \in Y$, (iii) η_t converges uniformly as $t \downarrow 0$ to a function $\gamma \in C(X, Y)$. Then

(0)

(.)

$$\lim_{t \downarrow 0} \frac{v(t) - v(0)}{t} = \inf_{x \in X^*} \sup_{y \in Y^*} \gamma(x, y).$$
(7.93)

Proof. Consider a sequence $t_k \downarrow 0$. Denote $\eta_k := \eta_{t_k}$ and $\zeta_k := \zeta_{t_k} = f + t_k \eta_k$. By the assumption (ii) we have that functions $\zeta_k(x, y)$ are continuous and convex-concave on $X \times Y$. Also by the definition

$$v(t_k) = \inf_{x \in X} \sup_{y \in Y} \zeta_k(x, y).$$

For a point $x^* \in X^*$ we can write

$$v(0) = \sup_{y \in Y} f(x^*, y) \text{ and } v(t_k) \le \sup_{y \in Y} \zeta_k(x^*, y).$$

⁶Recall that C(X, Y) denotes the space of continuous functions $\psi : X \times Y \to \mathbb{R}$ equipped with the sup-norm $\|\psi\| = \sup_{(x,y) \in X \times Y} |\psi(x,y)|.$

Since the set Y is compact and function $\zeta_k(x^*, \cdot)$ is continuous, we have that the set $\arg \max_{y \in Y} \zeta_k(x^*, y)$ is nonempty. Let $y_k \in \arg \max_{y \in Y} \zeta_k(x^*, y)$. We have that

$$\arg\max_{y\in Y} f(x^*, y) = Y$$

and, since ζ_k tends (uniformly) to f, we have that y_k tends in distance to Y^* (i.e., the distance from y_k to Y^* tends to zero as $k \to \infty$). By passing to a subsequence if necessary we can assume that y_k converges to a point $y^* \in Y$ as $k \to \infty$. It follows that $y^* \in Y^*$, and of course we have that

$$\sup_{y \in Y} f(x^*, y) \ge f(x^*, y_k).$$

Also since η_k tends uniformly to γ , it follows that $\eta_k(x^*, y_k) \to \gamma(x^*, y^*)$. Consequently

$$v(t_k) - v(0) \le \zeta_k(x^*, y_k) - f(x^*, y_k) = t_k \eta_k(x^*, y_k) = t_k \gamma(x^*, y^*) + o(t_k).$$

We obtain that for any $x^* \in X^*$ there exists $y^* \in Y^*$ such that

$$\limsup_{k \to \infty} \frac{v(t_k) - v(0)}{t_k} \le \gamma(x^*, y^*)$$

It follows that

$$\limsup_{k \to \infty} \frac{v(t_k) - v(0)}{t_k} \le \inf_{x \in X^*} \sup_{y \in Y^*} \gamma(x, y).$$
(7.94)

In order to prove the converse inequality we proceed as follows. Consider a sequence $x_k \in \arg\min_{x \in X} \theta_k(x)$, where $\theta_k(x) := \sup_{y \in Y} \zeta_k(x, y)$. We have that $\theta_k : X \to \mathbb{R}$ are continuous functions converging uniformly in $x \in X$ to the max-function $\phi(x) = \sup_{y \in Y} f(x, y)$. Consequently x_k converges in distance to the set $\arg\min_{x \in X} \phi(x)$, which is equal to X^* . By passing to a subsequence if necessary we can assume that x_k converges to a point $x^* \in X^*$. For any $y \in Y^*$ we have $v(0) \leq f(x_k, y)$. Since $\zeta_k(x, y)$ is convexconcave, it has a nonempty set of saddle points $X_k^* \times Y_k^*$. We have that $x_k \in X_k^*$, and hence $v(t_k) \geq \zeta_k(x_k, y)$ for any $y \in Y$. It follows that for any $y \in Y^*$ the following holds

$$v(t_k) - v(0) \ge \zeta_k(x_k, y) - f(x_k, y) = t_k \gamma_k(x^*, y) + o(t_k),$$

and hence

$$\liminf_{k \to \infty} \frac{v(t_k) - v(0)}{t_k} \ge \gamma(x^*, y)$$

Since y was an arbitrary element of Y^* , we obtain that

$$\liminf_{k \to \infty} \frac{v(t_k) - v(0)}{t_k} \ge \sup_{y \in Y^*} \gamma(x^*, y),$$

and hence

$$\liminf_{k \to \infty} \frac{v(t_k) - v(0)}{t_k} \ge \inf_{x \in X^*} \sup_{y \in Y^*} \gamma(x, y).$$
(7.95)

The assertion of the theorem follows from (7.94) and (7.95).

7.1.6 Epiconvergence

Consider a sequence $f_k : \mathbb{R}^n \to \overline{\mathbb{R}}$, k = 1, ..., of extended real valued functions. It is said that the functions f_k epiconverge to a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, written $f_k \stackrel{\text{e}}{\to} f$, if the epigraphs of the functions f_k converge, in a certain set-valued sense, to the epigraph of f. It is also possible to define the epiconvergence in the following equivalent way.

Definition 7.29. It is said that f_k epiconverge to f if for any point $x \in \mathbb{R}^n$ the following two conditions hold: (i) for any sequence x_k converging to x one has

$$\liminf_{k \to \infty} f_k(x_k) \ge f(x), \tag{7.96}$$

(ii) there exists a sequence x_k converging to x such that⁷

$$\limsup_{k \to \infty} f_k(x_k) \le f(x). \tag{7.97}$$

Epiconvergence $f_k \xrightarrow{e} f$ implies that the function f is lower semicontinuous.

For $\varepsilon \ge 0$ we say that a point $\bar{x} \in \mathbb{R}^n$ is an ε -minimizer⁸ of f if $f(\bar{x}) \le \inf f(x) + \varepsilon$ (we write here ' $\inf f(x)$ ' for ' $\inf_{x \in \mathbb{R}^n} f(x)$ '). Clearly, for $\varepsilon = 0$ the set of ε -minimizers of f coincides with the set $\arg \min f$ (of minimizers of f).

Proposition 7.30. Suppose that $f_k \xrightarrow{e} f$. Then

$$\limsup_{k \to \infty} \left[\inf f_k(x) \right] \le \inf f(x). \tag{7.98}$$

Suppose, further, that: (i) for some $\varepsilon_k \downarrow 0$ there exists an ε_k -minimizer x_k of $f_k(\cdot)$ such that the sequence x_k converges to a point \bar{x} . Then $\bar{x} \in \arg \min f$ and

$$\lim_{k \to \infty} \left[\inf f_k(x) \right] = \inf f(x). \tag{7.99}$$

Proof. Consider a point $\bar{x} \in \mathbb{R}^n$ and let x_k be a sequence converging to \bar{x} such that the inequality (7.97) holds. Clearly $f_k(x_k) \ge \inf f_k(x)$ for all k. Together with (7.97) this implies that

$$f(\bar{x}) \ge \limsup_{k \to \infty} f_k(x_k) \ge \limsup_{k \to \infty} [\inf f_k(x)].$$

Since the above holds for any \bar{x} , the inequality (7.98) follows.

Now let x_k be a sequence of ε_k -minimizers of f_k converging to a point \bar{x} . We have then that $f_k(x_k) \leq \inf f_k(x) + \varepsilon_k$, and hence by (7.98) we obtain

$$\liminf_{k \to \infty} \left[\inf f_k(x) \right] = \liminf_{k \to \infty} \left[\inf f_k(x) + \varepsilon_k \right] \ge \liminf_{k \to \infty} f_k(x_k) \ge f(\bar{x}) \ge \inf f(x).$$

⁷Note that here some (all) points x_k can be equal to x.

⁸For the sake of convenience we allow in this section for a minimizer, or ε -minimizer, \bar{x} to be such that $f(\bar{x})$ is not finite, i.e., can be equal to $+\infty$ or $-\infty$.

Together with (7.98) this implies (7.99) and $f(\bar{x}) = \inf f(x)$. This completes the proof.

Assumption (i) in the above proposition can be ensured by various boundedness conditions. Proof of the following theorem can be found in [217, Theorem 7.17].

Theorem 7.31. Let $f_k : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a sequence of convex functions and $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex lower semicontinuous function such that dom f has a nonempty interior. Then the following are equivalent: (i) $f_k \stackrel{e}{\to} f$, (ii) there exists a dense subset D of \mathbb{R}^n such that $f_k(x) \to f(x)$ for all $x \in D$, (iii) $f_k(\cdot)$ converges uniformly to $f(\cdot)$ on every compact set C that does not contain a boundary point of dom f.

7.2 Probability

7.2.1 Probability Spaces and Random Variables

Let Ω be an abstract set. It is said that a set \mathcal{F} of subsets of Ω is a sigma algebra (also called sigma field) if: (i) it is closed under standard set theoretic operations (i.e., if $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}, A \cup B \in \mathcal{F}$ and $A \setminus B \in \mathcal{F}$), (ii) the set Ω belongs to \mathcal{F} , and (iii) if $A_i \in \mathcal{F}$, $i \in \mathbb{N}$, then $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$. The set Ω equipped with a sigma algebra \mathcal{F} is called a *sample* or *measurable* space and denoted (Ω, \mathcal{F}) . A set $A \subset \Omega$ is said to be \mathcal{F} -measurable if $A \in \mathcal{F}$. It is said that the sigma algebra \mathcal{F} is *generated* by its subset \mathcal{G} if any \mathcal{F} -measurable set can be obtained from sets belonging to \mathcal{G} by set theoretic operations and by taking the union of a countable family of sets from \mathcal{G} . That is, \mathcal{F} is generated by \mathcal{G} if \mathcal{F} is the smallest sigma algebra containing \mathcal{G} . If we have two sigma algebras \mathcal{F}_1 and \mathcal{F}_2 defined on the same set Ω , then it is said that \mathcal{F}_1 is a subalgebra of \mathcal{F}_2 if $\mathcal{F}_1 \subset \mathcal{F}_2$. The smallest possible sigma algebra on Ω consists of two elements Ω and the empty set \emptyset . Such sigma algebra is called *trivial.* An \mathcal{F} -measurable set A is said to be elementary if any \mathcal{F} -measurable subset of A is either the empty set or the set A. If the sigma algebra \mathcal{F} is finite, then it is generated by a family $A_i \subset \Omega$, i = 1, ..., n, of disjoint elementary sets and has 2^n elements. The sigma algebra generated by the set of open (or closed) subsets of a finite dimensional space \mathbb{R}^m is called its Borel sigma algebra. An element of this sigma algebra is called a Borel set. For a considered set $\Xi \subset \mathbb{R}^m$ we denote by \mathcal{B} the sigma algebra of all Borel subsets of Ξ .

A function $P : \mathcal{F} \to \mathbb{R}_+$ is called a (sigma-additive) *measure* on (Ω, \mathcal{F}) if for every collection $A_i \in \mathcal{F}, i \in \mathbb{N}$, such that $A_i \cap A_j = \emptyset$ for all $i \neq j$, we have

$$P(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} P(A_i).$$
(7.100)

In this definition it is assumed that for every $A \in \mathcal{F}$, and in particular for $A = \Omega$, P(A) is finite. Sometimes such measures are called *finite*. An important example of a measure which is not finite is the Lebesgue measure on \mathbb{R}^m . Unless stated otherwise we assume that a considered measure is finite. A measure P is said to be a *probability measure* if $P(\Omega) = 1$. A sample space (Ω, \mathcal{F}) equipped with a probability measure P is called a *probability space* and denoted (Ω, \mathcal{F}, P) . Recall that \mathcal{F} is said to be P-complete if $A \subset B$, $B \in \mathcal{F}$ and P(B) = 0, implies that $A \in \mathcal{F}$, and hence P(A) = 0. Since it is always possible to enlarge

⁹By \mathbb{N} we denote the set of positive integers.

the sigma algebra and extend the measure in such a way as to get complete space, we can assume without loss of generality that considered probability measures are *complete*. It is said that an event $A \in \mathcal{F}$ happens *P*-almost surely (a.s.) or almost everywhere (a.e.) if P(A) = 1, or equivalently $P(\Omega \setminus A) = 0$. We also sometimes say that such an event happens with probability one (w.p.1).

Let *P* an *Q* be two measures on a measurable space (Ω, \mathcal{F}) . It is said that *Q* is *absolutely continuous* with respect to *P* if $A \in \mathcal{F}$ and P(A) = 0 implies that Q(A) = 0. If the measure *Q* is finite, this is equivalent to condition: for every $\varepsilon > 0$ there exists $\delta > 0$ such that if $P(A) < \delta$, then $Q(A) < \varepsilon$.

Theorem 7.32 (Radon-Nikodym). If P and Q are measures on (Ω, \mathcal{F}) , then Q is absolutely continuous with respect to P if and only if there exists a measurable function $f: \Omega \to \mathbb{R}_+$ such that $Q(A) = \int_A f dP$ for every $A \in \mathcal{F}$.

The function f in the representation $Q(A) = \int_A f dP$ is called *density* of measure Q with respect to measure P. If the measure Q is a probability measure, then f is called the *probability density function* (pdf). The Radon-Nikodym Theorem says that measure Q has a density with respect to P iff Q is absolutely continuous with respect to P. We write this as f = dQ/dP or dQ = f dP.

A mapping $V : \Omega \to \mathbb{R}^m$ is said to be *measurable* if for any Borel set $A \in \mathcal{B}$, its inverse image $V^{-1}(A) := \{\omega \in \Omega : V(\omega) \in A\}$ is \mathcal{F} -measurable.¹⁰ A measurable mapping¹¹ $V(\omega)$ from probability space (Ω, \mathcal{F}, P) into \mathbb{R}^m is called a *random vector*. Note that the mapping V generates the probability measure¹² (also called the probability distribution) $P(A) := P(V^{-1}(A))$ on $(\mathbb{R}^m, \mathcal{B})$. The smallest closed set $\Xi \subset \mathbb{R}^m$ such that $P(\Xi) = 1$ is called the support of measure P. We can view the space (Ξ, \mathcal{B}) equipped with probability measure P as a probability space (Ξ, \mathcal{B}, P) . This probability space provides all relevant probabilistic information about the considered random vector. In that case we write $\Pr(A)$ for the probability of the event $A \in \mathcal{B}$. We often denote by ξ data vector of a considered problem. Sometimes we view ξ as a random vector $\xi : \Omega \to \mathbb{R}^m$ supported on a set $\Xi \subset \mathbb{R}^m$, and sometimes as an element $\xi \in \Xi$, i.e., as a particular realization of the random data vector. Usually, the meaning of such notation will be clear from the context and will not cause any confusion. If in doubt, in order to emphasize that we view ξ as a random vector, we sometimes write $\xi = \xi(\omega)$.

A measurable mapping (function) $Z : \Omega \to \mathbb{R}$ is called a *random variable*. Its probability distribution is completely defined by the cumulative distribution function (cdf) $H_Z(z) := \Pr\{Z \le z\}$. Note that since the Borel sigma algebra of \mathbb{R} is generated by the family of half line intervals $(-\infty, a]$, in order to verify measurability of $Z(\omega)$ it suffices to verify measurability of sets $\{\omega \in \Omega : Z(\omega) \le z\}$ for all $z \in \mathbb{R}$. We denote random vectors (variables) by capital letters, like V, Z etc., or $\xi(\omega)$, and often suppress their explicit

¹⁰In fact it suffices to verify \mathcal{F} -measurability of $V^{-1}(A)$ for any family of sets generating the Borel sigma algebra of \mathbb{R}^m .

¹¹We can view the notation $V(\omega)$ to denote a (measurable) mapping $V : \Omega \to \mathbb{R}^m$, i.e., to denote the corresponding random vector; or as its particular value at $\omega \in \Omega$, i.e., its particular realization. Which one of these two meanings is used usually is clear from the context.

¹²With some abuse of notation we also denote here by P the probability distribution induced by the probability measure P on (Ω, \mathcal{F}) .

dependence on $\omega \in \Omega$. The coordinate functions $V_1(\omega), ..., V_m(\omega)$ of the *m*-dimensional random vector $V(\omega)$ are called its components. While considering a random vector V we often talk about its probability distribution as the joint distribution of its components (random variables) $V_1, ..., V_m$.

Since we often deal with random variables which are given as optimal values of optimization problems we need to consider random variables $Z(\omega)$ which can also take values $+\infty$ or $-\infty$, i.e., functions $Z: \Omega \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}}$ denotes the set of extended real numbers. Such functions $Z: \Omega \to \overline{\mathbb{R}}$ are referred to as *extended real valued* functions. Operations between real numbers and symbols $\pm\infty$ are clear except for such operations as adding $+\infty$ and $-\infty$ which should be avoided. Measurability of an extended real valued function $Z(\omega)$ is defined in the standard way, i.e., $Z(\omega)$ is measurable if the set { $\omega \in \Omega : Z(\omega) \leq z$ } is \mathcal{F} -measurable for any $z \in \mathbb{R}$. A measurable extended real valued function is called an (extended) random variable. Note that here

$$\lim_{z \to +\infty} H_Z(z) = \lim_{z \to +\infty} \Pr\{Z \le z\}$$

is equal to the probability of the event $\{\omega \in \Omega : Z(\omega) < +\infty\}$ and can be less than one if the event $\{\omega \in \Omega : Z(\omega) = +\infty\}$ has a positive probability.

The *expected value* or *expectation* of an (extended) random variable $Z : \Omega \to \overline{\mathbb{R}}$ is defined by the integral

$$\mathbb{E}_P[Z] := \int_{\Omega} Z(\omega) dP(\omega).$$
(7.101)

When there is no ambiguity as to what probability measure is considered, we omit the subscript P and simply write $\mathbb{E}[Z]$. For a nonnegative valued measurable function $Z(\omega)$ such that the event $\Upsilon := \{\omega \in \Omega : Z(\omega) = +\infty\}$ has zero probability the above integral is defined in the usual way and can take value $+\infty$. If probability of the event Υ is positive, then, by definition, $\mathbb{E}[Z] = +\infty$. For a general (not necessarily nonnegative valued) random variable we would like to define¹³ $\mathbb{E}[Z] := \mathbb{E}[Z_+] - \mathbb{E}[(-Z)_+]$. In order to do that we have to ensure that we do not add $+\infty$ and $-\infty$. We say that the expected value $\mathbb{E}[Z]$ of an (extended real valued) random variable $Z(\omega)$ is well defined if it does not happen that both $\mathbb{E}[Z_+]$ and $\mathbb{E}[(-Z)_+]$ are $+\infty$, in which case $\mathbb{E}[Z] = \mathbb{E}[Z_+] - \mathbb{E}[(-Z)_+]$. That is, in order to verify that the expected value of $Z(\omega)$ is well defined one has to check that $Z(\omega)$ is measurable and either $\mathbb{E}[Z_+] < +\infty$ or $\mathbb{E}[(-Z)_+] < +\infty$. Note that if $Z(\omega)$ and $Z'(\omega)$ are two (extended) random variables such that their expectations are well defined and $Z(\omega) = Z'(\omega)$ for all $\omega \in \Omega$ except possibly on a set of measure zero, then $\mathbb{E}[Z] = \mathbb{E}[Z']$. It is said that $Z(\omega)$ is P-integrable if the expected value $\mathbb{E}[Z]$ is well defined and finite. The expected value of a random vector is defined componentwise.

If the random variable $Z(\omega)$ can take only a countable (finite) number of different values, say $z_1, z_2, ...$, then it is said that $Z(\omega)$ has a *discrete* distribution (*discrete* distribution with a *finite support*). In such cases all relevant probabilistic information is contained in the probabilities $p_i := \Pr\{Z = z_i\}$. In that case $\mathbb{E}[Z] = \sum_i p_i z_i$.

Let $f_n(\omega)$ be a sequence of real valued measurable functions on a probability space (Ω, \mathcal{F}, P) . By $f_n \uparrow f$ a.e. we mean that for almost every $\omega \in \Omega$ the sequence $f_n(\omega)$ is monotonically nondecreasing and hence converges to a limit denoted $f(\omega)$, where $f(\omega)$ can be equal to $+\infty$. We have the following classical results about convergence of integrals.

¹³Recall that $Z_{+} := \max\{0, Z\}.$

Theorem 7.33 (Monotone Convergence Theorem). Suppose that $f_n \uparrow f$ a.e. and there exists a *P*-integrable function $g(\omega)$ such that $f_n(\cdot) \ge g(\cdot)$. Then $\int_{\Omega} f dP$ is well defined and $\int_{\Omega} f_n dP \uparrow \int_{\Omega} f dP$.

Theorem 7.34 (Fatou's lemma). Suppose that there exists a *P*-integrable function $g(\omega)$ such that $f_n(\cdot) \ge g(\cdot)$. Then

$$\liminf_{n \to \infty} \int_{\Omega} f_n \, dP \ge \int_{\Omega} \liminf_{n \to \infty} f_n \, dP. \tag{7.102}$$

Theorem 7.35 (Lebesgue Dominated Convergence Theorem). Suppose that there exists a *P*-integrable function $g(\omega)$ such that $|f_n| \leq g$ a.e., and that $f_n(\omega)$ converges to $f(\omega)$ for almost every $\omega \in \Omega$. Then f is *P*-integrable and

$$\int_{\Omega} f_n dP \to \int_{\Omega} f dP. \tag{7.103}$$

It is said that the sequence f_n is *uniformly integrable* if

$$\lim_{c \to \infty} \sup_{n \in \mathbb{N}} \int_{\{|f_n| \ge c\}} |f_n| dP = 0.$$
(7.104)

If $|f_n|$ are dominated by a *P*-integrable function *g*, i.e., $|f_n| \le g$ a.e., then f_n are uniformly integrable. Therefore the following theorem can be viewed as a slight generalization of the Lebesgue Dominated Convergence Theorem.

Theorem 7.36. Suppose that $f_n(\omega)$ converges to $f(\omega)$ for almost every $\omega \in \Omega$. Then the following holds. (i) If f_n are uniformly integrable, then f is P-integrable and (7.103) follows. (ii) If f and f_n are nonnegative valued and P-integrable, then (7.103) implies that f_n are uniformly integrable.

We also have the following useful result. Unless stated otherwise we always assume that considered measures are finite and nonnegative, i.e., $\mu(A)$ is a finite nonnegative number for every $A \in \mathcal{F}$.

Theorem 7.37 (Richter-Rogosinski). Let (Ω, \mathcal{F}) be a measurable space, $f_1, ..., f_m$ be measurable on (Ω, \mathcal{F}) real valued functions, and μ be a measure on (Ω, \mathcal{F}) such that $f_1, ..., f_m$ are μ -integrable. Suppose that every finite subset of Ω is \mathcal{F} -measurable. Then there exists a measure η on (Ω, \mathcal{F}) with a finite support of at most m points such that

$$\int_{\Omega} f_i d\mu = \int_{\Omega} f_i d\eta, \ i = 1, ..., m.$$
(7.105)

Proof. First we show that there exists a measure η with finite support such that (7.105) holds. The proof proceeds by induction on m. It can be easily shown that the assertion holds for m = 1. Consider the set $S \subset \mathbb{R}^m$ generated by vectors of the form

 $(\int f_1 d\mu', ..., \int f_m d\mu')$ with μ' being a measure on Ω with a finite support. We have to show that vector $a := (\int f_1 d\mu, ..., \int f_m d\mu)$ belongs to S. Note that the set S is a convex cone. Suppose that $a \notin S$. Then, by the separation theorem, there exists $c \in \mathbb{R}^m \setminus \{0\}$ such that $c^{\mathsf{T}}a \leq c^{\mathsf{T}}x$, for all $x \in S$. Since S is a cone, it follows that $c^{\mathsf{T}}a \leq 0$. This implies that for $f := \sum_{i=1}^n c_i f_i$ we have that $\int f d\mu \leq 0$ and $\int f d\mu \leq \int f d\mu'$ for any measure μ' with a finite support. In particular, by taking measures of the form¹⁴ $\mu' = \gamma \delta(\omega)$, with $\gamma > 0$ and $\omega \in \Omega$, we obtain from the second inequality that $\int f d\mu \leq \gamma f(\omega)$. This implies that $f(\omega) \geq 0$ for all $\omega \in \Omega$. Indeed, if $f(\omega) < 0$, then we can make $\gamma f(\omega)$ arbitrary small by taking γ large enough and hence eventually contradict the inequality $\int f d\mu \leq \gamma f(\omega)$. Together with the first inequality this implies that $\int f d\mu = 0$.

Consider the set $\Omega' := \{\omega \in \Omega : f(\omega) = 0\}$. Note that the function f is measurable and hence $\Omega' \in \mathcal{F}$. Since $\int f d\mu = 0$ and $f(\cdot)$ is nonnegative valued, it follows that Ω' is a support of μ , i.e., $\mu(\Omega') = \mu(\Omega)$. If $\mu(\Omega) = 0$, then the assertion clearly holds. Therefore suppose that $\mu(\Omega) > 0$. Then $\mu(\Omega') > 0$, and hence Ω' is nonempty. It also follows that $\int_{\Omega} f_i d\mu = \int_{\Omega'} f_i d\mu$ for all i = 1, ..., m.

We have that $\sum_{i=1}^{n} c_i f_i(\omega) = 0$ for all $\omega \in \Omega'$. Since $c \neq 0$, there is $\ell \in \{1, ..., n\}$ such that $c_\ell \neq 0$. By the induction assumption there exists a measure η with a finite support such that $\int_{\Omega} f_i d\mu = \int_{\Omega} f_i d\eta$ for all $i \neq \ell$. Also $f_\ell(\omega) = \sum_{i \neq \ell} \alpha_i f_i(\omega), \omega \in \Omega'$, for $\alpha_i := -c_i/c_\ell$, and hence $\int_{\Omega} f_\ell d\mu = \int_{\Omega} f_\ell d\eta$. This proves existence of a measure η with finite support such that (7.105) holds.

So let $\eta = \sum_{j=1}^{k} \bar{x}_j \delta(\omega_j)$ be a measure with $\bar{x}_j \ge 0$, j = 1, ..., k, having finite support $\{\omega_1, ..., \omega_k\}$, such that (7.105) holds. We can interpret (7.105) as that $\bar{x} = (\bar{x}_1, ..., \bar{x}_k)^T$ is a solution of the linear system $Ax = b, x \ge 0$, where b is $m \times 1$ vector with components $b_i = \int_{\Omega} f_i d\mu$ and A is $m \times k$ matrix with elements $a_{ij} = f_i(\omega_j)$. Since \bar{x} is a solution of this system, this system is feasible. Hence this system has a solution with only at most m nonzero components (this fact is well known in linear programming). This completes the proof. \Box

Let us remark that if the measure μ is a probability measure, i.e., $\mu(\Omega) = 1$, then by adding the constraint $\int_{\Omega} d\eta = 1$, we obtain in the above theorem that there exists a probability measure η on (Ω, \mathcal{F}) with a finite support of at most m + 1 points such that $\int_{\Omega} f_i d\mu = \int_{\Omega} f_i d\eta$ for all i = 1, ..., m.

Also let us recall two famous inequalities. *Chebyshev inequality*¹⁵ says that if $Z : \Omega \to \mathbb{R}_+$ is a nonnegative valued random variable, then

$$\Pr(Z \ge \alpha) \le \alpha^{-1} \mathbb{E}[Z], \quad \forall \alpha > 0. \tag{7.106}$$

Proof of (7.106) is rather simple, we have

$$\Pr(Z \ge \alpha) = \mathbb{E}\big[\mathbf{1}_{[\alpha, +\infty)}(Z)\big] \le \mathbb{E}\big[\alpha^{-1}Z\big] = \alpha^{-1}\mathbb{E}[Z].$$

Jensen inequality says that if $V : \Omega \to \mathbb{R}^m$ is a random vector, $\nu := \mathbb{E}[V]$ and $f : \mathbb{R}^m \to \mathbb{R}$ is a convex function, then

$$\mathbb{E}\left[f(V)\right] \ge f(\nu),\tag{7.107}$$

¹⁴We denote by $\delta(\omega)$ measure of mass one at the point ω , and refer to such measures as *Dirac* measures.

¹⁵Sometimes (7.106) is called Markov inequality, while Chebyshev inequality is referred to the inequality (7.106) applied to the function $(Z - \mathbb{E}[Z])^2$.

provided the above expectations are finite. Indeed, for a subgradient $g \in \partial f(\nu)$ we have that

$$f(V) \ge f(\nu) + g^{\top}(V - \nu).$$
 (7.108)

By taking expectation of the both sides of (7.108) we obtain (7.107).

Finally, let us mention the following simple inequality. Let $Y_1, Y_2 : \Omega \to \mathbb{R}$ be random variables and a_1, a_2 be numbers. Then the intersection of the events $\{\omega : Y_1(\omega) < a_1\}$ and $\{\omega : Y_2(\omega) < a_2\}$ is included in the event $\{\omega : Y_1(\omega) + Y_2(\omega) < a_1 + a_2\}$, or equivalently the event $\{\omega : Y_1(\omega) + Y_2(\omega) \ge a_1 + a_2\}$ is included in the union of the events $\{\omega : Y_1(\omega) \ge a_1\}$ and $\{\omega : Y_2(\omega) \ge a_2\}$. It follows that

$$\Pr\{Y_1 + Y_2 \ge a_1 + a_2\} \le \Pr\{Y_1 \ge a_1\} + \Pr\{Y_2 \ge a_2\}.$$
(7.109)

7.2.2 Conditional Probability and Conditional Expectation

For two events A and B the conditional probability of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$
 (7.110)

provided that $P(B) \neq 0$. Now let X and Y be discrete random variables with joint mass function p(x, y) := P(X = x, Y = y). Of course, since X and Y are discrete, p(x, y)is nonzero only for a finite or countable number of values of x and y. The marginal mass functions of X and Y are $p_X(x) := P(X = x) = \sum_y p(x, y)$ and $p_Y(y) := P(Y = y) = \sum_x p(x, y)$, respectively. It is natural to define conditional mass function of X given that Y = y as

$$p_{X|Y}(x|y) := P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p(x, y)}{p_Y(y)}$$
(7.111)

for all values of y such that $p_Y(y) > 0$. We have that X is independent of Y iff $p(x, y) = p_X(x)p_Y(y)$ holds for all x and y, which is equivalent to that $p_{X|Y}(x|y) = p_X(x)$ for all y such that $p_Y(y) > 0$.

If X and Y have continuous distribution with a joint pdf f(x, y), then the *conditional* pdf of X, given that Y = y, is defined in a way similar to (7.111) for all values of y such that $f_Y(y) > 0$ as

$$f_{X|Y}(x|y) := \frac{f(x,y)}{f_Y(y)}.$$
(7.112)

Here $f_Y(y) := \int_{-\infty}^{+\infty} f(x, y) dx$ is the marginal pdf of Y. In the continuous case the *conditional expectation* of X, given that Y = y, is defined for all values of y such that $f_Y(y) > 0$ as

$$\mathbb{E}[X|Y=y] := \int_{-\infty}^{+\infty} x f_{X|Y}(x|y) dx.$$
 (7.113)

In the discrete case it is defined in a similar way.

Note that $\mathbb{E}[X|Y = y]$ is a function of y, say $h(y) := \mathbb{E}[X|Y = y]$. Let us denote by $\mathbb{E}[X|Y]$ that function of random variable Y, i.e., $\mathbb{E}[X|Y] := h(Y)$. We have then the following important formula

$$\mathbb{E}[X] = \mathbb{E}\left|\mathbb{E}[X|Y]\right|. \tag{7.114}$$

In the continuous case, for example, we have

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y) dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f_{X|Y}(x|y) dx f_Y(y) dy,$$

and hence

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} \mathbb{E}[X|Y=y] f_Y(y) dy.$$
(7.115)

The above definitions can be extended to the case where X and Y are two random *vectors* in a straightforward way.

It is also useful to define conditional expectation in the following abstract form. Let X be a nonnegative valued integrable random variable on a probability space (Ω, \mathcal{F}, P) , and let \mathcal{G} be a subalgebra of \mathcal{F} . Define a measure on \mathcal{G} by $\nu(G) := \int_G X dP$ for any $G \in \mathcal{G}$. This measure is finite because X is integrable, and is absolutely continuous with respect to P. Hence by the Radon-Nikodym Theorem there is a \mathcal{G} -measurable function $h(\omega)$ such that $\nu(G) = \int_G h dP$. This function $h(\omega)$, viewed as a random variable has the following properties: (i) $h(\omega)$ is \mathcal{G} -measurable and integrable, (ii) it satisfies the equation $\int_G h dP = \int_G X dP$ for any $G \in \mathcal{G}$. By definition we say that a random variable, denoted $\mathbb{E}[X|\mathcal{G}]$, is said to be the *conditional expected value* of X given \mathcal{G} , if it satisfies the following two properties:

- (i) $\mathbb{E}[X|\mathcal{G}]$ is \mathcal{G} -measurable and integrable,
- (ii) $\mathbb{E}[X|\mathcal{G}]$ satisfies the functional equation

$$\int_{G} \mathbb{E}[X|\mathcal{G}]dP = \int_{G} XdP, \ \forall G \in \mathcal{G}.$$
(7.116)

The above construction showed existence of such random variable for nonnegative X. If X is not necessarily nonnegative, apply the same construction to the positive and negative part of X.

There will be many random variable satisfying the above properties (i) and (ii). Any one of them is called a *version* of the conditional expected value. We sometimes write it as $\mathbb{E}[X|\mathcal{G}](\omega)$ or $\mathbb{E}[X|\mathcal{G}]_{\omega}$ to emphasize that this a random variable. Any two versions of $\mathbb{E}[X|\mathcal{G}]$ are equal to each other with probability one. Note that, in particular, for $G = \Omega$ it follows from (ii) that

$$\mathbb{E}[X] = \int_{\Omega} \mathbb{E}[X|\mathcal{G}]dP = \mathbb{E}\big[\mathbb{E}[X|\mathcal{G}]\big].$$
(7.117)

Note also that if the sigma algebra \mathcal{G} is trivial, i.e., $\mathcal{G} = \{\emptyset, \Omega\}$, then $\mathbb{E}[X|\mathcal{G}]$ is constant equal to $\mathbb{E}[X]$.

Conditional probability $P(A|\mathcal{G})$ of event $A \in \mathcal{F}$ can be defined as $P(A|\mathcal{G}) = \mathbb{E}[\mathbf{1}_A|\mathcal{G}]$. In that case the corresponding properties (i) and (ii) take the form:

- (i') $P(A|\mathcal{G})$ is \mathcal{G} -measurable and integrable,
- (ii') $P(A|\mathcal{G})$ satisfies the functional equation

$$\int_{G} P(A|\mathcal{G})dP = P(A \cap G), \ \forall G \in \mathcal{G}.$$
(7.118)

7.2.3 Measurable Multifunctions and Random Functions

Let \mathcal{G} be a mapping from Ω into the set of subsets of \mathbb{R}^n , i.e., \mathcal{G} assigns to each $\omega \in \Omega$ a subset (possibly empty) $\mathcal{G}(\omega)$ of \mathbb{R}^n . We refer to \mathcal{G} as a *multifunction* and write $\mathcal{G} : \Omega \rightrightarrows$ \mathbb{R}^n . It is said that \mathcal{G} is *closed valued* if $\mathcal{G}(\omega)$ is a closed subset of \mathbb{R}^n for every $\omega \in \Omega$. A closed valued multifunction \mathcal{G} is said to be *measurable*, if for every closed set $A \subset \mathbb{R}^n$ one has that the inverse image $\mathcal{G}^{-1}(A) := \{\omega \in \Omega : \mathcal{G}(\omega) \cap A \neq \emptyset\}$ is \mathcal{F} -measurable. Note that measurability of \mathcal{G} implies that the domain

dom
$$\mathcal{G} := \{ \omega \in \Omega : \mathcal{G}(\omega) \neq \emptyset \} = \mathcal{G}^{-1}(\mathbb{R}^n)$$

of \mathcal{G} is an \mathcal{F} -measurable subset of Ω .

Proposition 7.38. A closed valued multifunction $\mathcal{G} : \Omega \rightrightarrows \mathbb{R}^n$ is measurable iff the (extended real valued) function $d(\omega) := \operatorname{dist}(x, \mathcal{G}(\omega))$ is measurable for any $x \in \mathbb{R}^n$.

Proof. Recall that by the definition $\operatorname{dist}(x, \mathcal{G}(\omega)) = +\infty$ if $\mathcal{G}(\omega) = \emptyset$. Note also that $\operatorname{dist}(x, \mathcal{G}(\omega)) = ||x - y||$ for some $y \in \mathcal{G}(\omega)$, because of closedness of set $\mathcal{G}(\omega)$. Therefore, for any $t \ge 0$ and $x \in \mathbb{R}^n$ we have that

$$\{\omega \in \Omega : \operatorname{dist}(x, \mathcal{G}(\omega)) \le t\} = \mathcal{G}^{-1}(x + tB),$$

where $B := \{x \in \mathbb{R}^n : ||x|| \leq 1\}$. It remains to note that it suffices to verify the measurability of $\mathcal{G}^{-1}(A)$ for closed sets of the form A = x + tB, $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^n$. \Box

Remark 53. Suppose now that Ω is a Borel subset of \mathbb{R}^m equipped with its Borel sigma algebra. Suppose, further, that the multifunction $\mathcal{G} : \Omega \rightrightarrows \mathbb{R}^n$ is *closed*. That is, if $\omega_k \to \omega$, $x_k \in \mathcal{G}(\omega_k)$ and $x_k \to x$, then $x \in \mathcal{G}(\omega)$. Of course, any closed multifunction is closed valued. It follows that for any $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^n$ the level set $\{\omega \in \Omega : \operatorname{dist}(x, \mathcal{G}(\omega)) \le t\}$ is closed, and hence the function $d(\omega) := \operatorname{dist}(x, \mathcal{G}(\omega))$ is measurable. Consequently we obtain that any closed multifunction $\mathcal{G} : \Omega \rightrightarrows \mathbb{R}^n$ is measurable.

It is said that a mapping $G : \operatorname{dom} \mathcal{G} \to \mathbb{R}^n$ is a selection of \mathcal{G} if $G(\omega) \in \mathcal{G}(\omega)$ for all $\omega \in \operatorname{dom} \mathcal{G}$. If, in addition, the mapping G is measurable, it is said that G is a *measurable selection* of \mathcal{G} .

Theorem 7.39 (Measurable selection theorem). A closed valued multifunction \mathcal{G} : $\Omega \rightrightarrows \mathbb{R}^n$ is measurable iff its domain is an \mathcal{F} -measurable subset of Ω and there exists a

countable family $\{G_i\}_{i \in \mathbb{N}}$, of measurable selections of \mathcal{G} , such that for every $\omega \in \Omega$, the set $\{G_i(\omega) : i \in \mathbb{N}\}$ is dense in $\mathcal{G}(\omega)$.

In particular, we have by the above theorem that if $\mathcal{G} : \Omega \rightrightarrows \mathbb{R}^n$ is a closed valued measurable multifunction, then there exists at least one measurable selection of \mathcal{G} . In [217, Theorem 14.5] the result of the above theorem is called *Castaing representation*.

Consider a function $F : \mathbb{R}^n \times \Omega \to \mathbb{R}$. We say that F is a *random function* if for every fixed $x \in \mathbb{R}^n$, the function $F(x, \cdot)$ is \mathcal{F} -measurable. For a random function $F(x, \omega)$ we can define the corresponding expected value function

$$f(x) := \mathbb{E}[F(x,\omega)] = \int_{\Omega} F(x,\omega) dP(\omega).$$

We say that f(x) is well defined if the expectation $\mathbb{E}[F(x,\omega)]$ is well defined for every $x \in \mathbb{R}^n$. Also for every $\omega \in \Omega$ we can view $F(\cdot, \omega)$ as an extended real valued function.

Definition 7.40. It is said that the function $F(x, \omega)$ is *random lower semicontinuous* if the associated epigraphical multifunction $\omega \mapsto \text{epi} F(\cdot, \omega)$ is closed valued and measurable.

In some publications random lower semicontinuous functions are called *normal integrands*. It follows from the above definitions that if $F(x, \omega)$ is random lsc, then the multifunction $\omega \mapsto \operatorname{dom} F(\cdot, \omega)$ is measurable, and $F(x, \cdot)$ is measurable for every fixed $x \in \mathbb{R}^n$. Close valuedness of the epigraphical multifunction means that for every $\omega \in \Omega$, the epigraph epi $F(\cdot, \omega)$ is a closed subset of \mathbb{R}^{n+1} , i.e., $F(\cdot, \omega)$ is lower semicontinuous. Note, however, that the lower semicontinuity in x and measurability in ω does not imply measurability of the corresponding epigraphical multifunction, and hence random lower semicontinuous is given by the so-called *Carathéodory* functions, i.e., real valued functions $F : \mathbb{R}^n \times \Omega \to \mathbb{R}$ such that $F(x, \cdot)$ is \mathcal{F} -measurable for every $x \in \mathbb{R}^n$ and $F(\cdot, \omega)$ continuous for a.e. $\omega \in \Omega$.

Theorem 7.41. Suppose that the sigma algebra \mathcal{F} is *P*-complete. Then an extended real valued function $F : \mathbb{R}^n \times \Omega \to \overline{\mathbb{R}}$ is random lsc iff the following two properties hold: (i) for every $\omega \in \Omega$, the function $F(\cdot, \omega)$ is lsc, (ii) the function $F(\cdot, \cdot)$ is measurable with respect to the sigma algebra of $\mathbb{R}^n \times \Omega$ given by the product of the sigma algebras \mathcal{B} and \mathcal{F} .

With a random function $F(x, \omega)$ we associate its *optimal value function* $\vartheta(\omega) := \inf_{x \in \mathbb{R}^n} F(x, \omega)$ and the *optimal solution multifunction* $\mathcal{X}^*(\omega) := \arg \min_{x \in \mathbb{R}^n} F(x, \omega)$.

Theorem 7.42. Let $F : \mathbb{R}^n \times \Omega \to \overline{\mathbb{R}}$ be a random lsc function. Then the optimal value function $\vartheta(\omega)$ and the optimal solution multifunction $\mathcal{X}^*(\omega)$ are both measurable.

Note that it follows from lower semicontinuity of $F(\cdot, \omega)$ that the optimal solution multifunction $\mathcal{X}^*(\omega)$ is closed valued. Note also that if $F(x, \omega)$ is random lsc and $\mathcal{G} : \Omega \Longrightarrow \mathbb{R}^n$ is a closed valued measurable multifunction, then the function

$$\bar{F}(x,\omega) := \begin{cases} F(x,\omega), & \text{if } x \in \mathcal{G}(\omega), \\ +\infty, & \text{if } x \notin \mathcal{G}(\omega), \end{cases}$$

is also random lsc. Consequently the corresponding optimal value $\omega \mapsto \inf_{x \in \mathcal{G}(\omega)} F(x, \omega)$ and the optimal solution multifunction $\omega \mapsto \arg \min_{x \in \mathcal{G}(\omega)} F(x, \omega)$ are both measurable, and hence by the measurable selection theorem, there exists a measurable selection $\bar{x}(\omega) \in$ $\arg \min_{x \in \mathcal{G}(\omega)} F(x, \omega)$.

Theorem 7.43. Let $F : \mathbb{R}^{n+m} \times \Omega \to \overline{\mathbb{R}}$ be a random lsc function and

$$\vartheta(x,\omega) := \inf_{y \in \mathbb{R}^m} F(x, y, \omega) \tag{7.119}$$

be the associated optimal value function. Suppose that there exists a bounded set $S \subset \mathbb{R}^m$ such that dom $F(x, \cdot, \omega) \subset S$ for all $(x, \omega) \in \mathbb{R}^n \times \Omega$. Then the optimal value function $\vartheta(x, \omega)$ is random lsc.

Let us observe that the above framework of random lsc functions is aimed at minimization problems. Of course, the problem of maximization of $\mathbb{E}[F(x,\omega)]$ is equivalent to minimization of $\mathbb{E}[-F(x,\omega)]$. Therefore, for maximization problems one would need the comparable concept of random upper semicontinuous functions.

Consider a multifunction $\mathcal{G}: \Omega \rightrightarrows \mathbb{R}^n$. Denote

$$\|\mathcal{G}(\omega)\| := \sup\{\|G(\omega)\| : G(\omega) \in \mathcal{G}(\omega)\},\$$

and by conv $\mathcal{G}(\omega)$ the convex hull of set $\mathcal{G}(\omega)$. If the set $\Omega = \{\omega_1, ..., \omega_K\}$ is finite and equipped with respective probabilities p_k , k = 1, ..., K, then it is natural to define the integral

$$\int_{\Omega} \mathcal{G}(\omega) dP(\omega) := \sum_{k=1}^{K} p_k \mathcal{G}(\omega_k), \qquad (7.120)$$

where the sum of two sets $A, B \subset \mathbb{R}^n$ and multiplication by a scalar $\gamma \in \mathbb{R}$ are defined in the natural way, $A+B := \{a+b : a \in A, b \in B\}$ and $\gamma A := \{\gamma a : a \in A\}$. For a general measure P on a sample space (Ω, \mathcal{F}) the corresponding integral is defined as follows.

Definition 7.44. The integral $\int_{\Omega} \mathcal{G}(\omega) dP(\omega)$ is defined as the set of all points of the form $\int_{\Omega} G(\omega) dP(\omega)$, where $G(\omega)$ is a *P*-integrable selection of $\mathcal{G}(\omega)$, i.e., $G(\omega) \in \mathcal{G}(\omega)$ for *a.e.* $\omega \in \Omega$, $G(\omega)$ is measurable and $\int_{\Omega} ||G(\omega)|| dP(\omega)$ is finite.

If the multifunction $\mathcal{G}(\omega)$ is convex valued, i.e., the set $\mathcal{G}(\omega)$ is convex for a.e. $\omega \in \Omega$, then $\int_{\Omega} \mathcal{G}dP$ is a convex set. It turns out that $\int_{\Omega} \mathcal{G}dP$ is always convex (even if $\mathcal{G}(\omega)$ is not convex valued) if the measure P does not have atoms, i.e., is nonatomic¹⁶. The following theorem often is referred to Aumann (1965).

Theorem 7.45 (Aumann). Suppose that the measure P is nonatomic and let $\mathcal{G} : \Omega \Rightarrow \mathbb{R}^n$ be a multifunction. Then the set $\int_{\Omega} \mathcal{G}dP$ is convex. Suppose, further, that $\mathcal{G}(\omega)$ is closed valued and measurable, and there exists a P-integrable function $g(\omega)$ such that $\|\mathcal{G}(\omega)\| \leq C$

¹⁶It is said that measure P, and the space (Ω, \mathcal{F}, P) , is *nonatomic* if any set $A \in \mathcal{F}$, such that P(A) > 0, contains a subset $B \in \mathcal{F}$ such that P(A) > P(B) > 0.

 $g(\omega)$ for a.e. $\omega \in \Omega$. Then

$$\int_{\Omega} \mathcal{G}(\omega) dP(\omega) = \int_{\Omega} \left(\operatorname{conv} \mathcal{G}(\omega) \right) dP(\omega).$$
(7.121)

The above theorem is a consequence of a theorem due to A.A. Lyapunov (1940).

Theorem 7.46 (Lyapunov). Let $\mu_1, ..., \mu_n$ be a finite collection of nonatomic measures on a measurable space (Ω, \mathcal{F}) . Then the set $\{(\mu_1(S), ..., \mu_n(S)) : S \in \mathcal{F}\}$ is a closed and convex subset of \mathbb{R}^n .

7.2.4 Expectation Functions

Consider a random function $F : \mathbb{R}^n \times \Omega \to \overline{\mathbb{R}}$ and the corresponding expected value (or simply expectation) function $f(x) = \mathbb{E}[F(x,\omega)]$. Recall that by assuming that $F(x,\omega)$ is a random function we assume that $F(x, \cdot)$ is measurable for every $x \in \mathbb{R}^n$. We have that the function f(x) is well defined on a set $X \subset \mathbb{R}^n$ if for every $x \in X$ either $\mathbb{E}[F(x,\omega)_+] < +\infty$ or $\mathbb{E}[(-F(x,\omega))_+] < +\infty$. The expectation function inherits various properties of the functions $F(\cdot, \omega), \omega \in \Omega$. As it is shown in the next proposition the lower semicontinuity of the expected value function follows from the lower semicontinuity of $F(\cdot, \omega)$.

Theorem 7.47. Suppose that for P-almost every $\omega \in \Omega$ the function $F(\cdot, \omega)$ is lsc at a point x_0 and there exists P-integrable function $Z(\omega)$ such that $F(x, \omega) \ge Z(\omega)$ for P-almost all $\omega \in \Omega$ and all x in a neighborhood of x_0 . Then for all x in a neighborhood of x_0 the expected value function $f(x) := \mathbb{E}[F(x, \omega)]$ is well defined and lsc at x_0 .

Proof. It follows from the assumption that $F(x, \omega)$ is bounded from below by a *P*-integrable function, that $f(\cdot)$ is well defined in a neighborhood of x_0 . Moreover, by Fatou's lemma we have

$$\liminf_{x \to x_0} \int_{\Omega} F(x,\omega) \, dP(\omega) \ge \int_{\Omega} \liminf_{x \to x_0} F(x,\omega) \, dP(\omega). \tag{7.122}$$

Together with lsc of $F(\cdot, \omega)$ this implies lower semicontinuity of f at x_0 . \Box

With a little stronger assumptions we can show that the expectation function is continuous.

Theorem 7.48. Suppose that for *P*-almost every $\omega \in \Omega$ the function $F(\cdot, \omega)$ is continuous at x_0 and there exists *P*-integrable function $Z(\omega)$ such that $|F(x, \omega)| \leq Z(\omega)$ for *P*-almost every $\omega \in \Omega$ and all x in a neighborhood of x_0 . Then for all x in a neighborhood of x_0 the expected value function f(x) is well defined and continuous at x_0 .

Proof. It follows from the assumption that $|F(x, \omega)|$ is dominated by a *P*-integrable function, that f(x) is well defined and finite valued for all x in a neighborhood of x_0 . Moreover, by the Lebesgue Dominated Convergence Theorem we can take the limit inside the integral,

which together with the continuity assumption implies

$$\lim_{x \to x_0} \int_{\Omega} F(x,\omega) dP(\omega) = \int_{\Omega} \lim_{x \to x_0} F(x,\omega) dP(\omega) = \int_{\Omega} F(x_0,\omega) dP(\omega).$$
(7.123)

This shows the continuity of f(x) at x_0 . \Box

Consider, for example, the characteristic function $F(x,\omega) := \mathbf{1}_{(-\infty,x]}(\xi(\omega))$, with $x \in \mathbb{R}$ and $\xi = \xi(\omega)$ being a real valued random variable. We have then that $f(x) = \Pr(\xi \leq x)$, i.e., that $f(\cdot)$ is the cumulative distribution function of ξ . It follows that in this example the expected value function is continuous at a point x_0 iff the probability of the event $\{\xi = x_0\}$ is zero. Note that $x = \xi(\omega)$ is the only point at which the function $F(\cdot, \omega)$ is discontinuous.

We say that random function $F(x, \omega)$ is *convex*, if the function $F(\cdot, \omega)$ is convex for a.e. $\omega \in \Omega$. Convexity of $F(\cdot, \omega)$ implies convexity of the expectation function f(x). Indeed, if $F(x, \omega)$ is convex and the measure P is discrete, then f(x) is a weighted sum, with positive coefficients, of convex functions and hence is convex. For general measures convexity of the expectation function follows by passing to the limit. Recall that if f(x)is convex, then it is continuous on the interior of its domain. In particular, if f(x) is real valued for all $x \in \mathbb{R}^n$, then it is continuous on \mathbb{R}^n .

We discuss now differentiability properties of the expected value function f(x). We sometimes write $F_{\omega}(\cdot)$ for the function $F(\cdot, \omega)$ and denote by $F'_{\omega}(x_0, h)$ the directional derivative of $F_{\omega}(\cdot)$ at the point x_0 in the direction h. Definitions and basic properties of directional derivatives are given in section 7.1.1. Consider the following conditions:

- (A1) The expectation $f(x_0)$ is well defined and finite valued at a given point $x_0 \in \mathbb{R}^n$.
- (A2) There exists a positive valued random variable $C(\omega)$ such that $\mathbb{E}[C(\omega)] < +\infty$, and for all x_1, x_2 in a neighborhood of x_0 and almost every $\omega \in \Omega$ the following inequality holds

$$|F(x_1,\omega) - F(x_2,\omega)| \le C(\omega) ||x_1 - x_2||.$$
(7.124)

- (A3) For almost every ω the function $F_{\omega}(\cdot)$ is directionally differentiable at x_0 .
- (A4) For almost every ω the function $F_{\omega}(\cdot)$ is differentiable at x_0 .

Theorem 7.49. We have the following: (a) If conditions (A1) and (A2) hold, then the expected value function f(x) is Lipschitz continuous in a neighborhood of x_0 . (b) If conditions (A1)–(A3) hold, then the expected value function f(x) is directionally differentiable at x_0 , and

$$f'(x_0, h) = \mathbb{E}\left[F'_{\omega}(x_0, h)\right] \quad \text{for all } h. \tag{7.125}$$

(c) If conditions (A1), (A2) and (A4) hold, then f(x) is differentiable at x_0 and

$$\nabla f(x_0) = \mathbb{E}\left[\nabla_x F(x_0, \omega)\right]. \tag{7.126}$$

Proof. It follows from (7.124) that for any x_1, x_2 in a neighborhood of x_0 ,

$$|f(x_1) - f(x_2)| \le \int_{\Omega} |F(x_1, \omega) - F(x_2, \omega)| \, dP(\omega) \le c ||x_1 - x_2||,$$

where $c := \mathbb{E}[C(\omega)]$. Together with assumption (A1) this implies that f(x) is well defined, finite valued and Lipschitz continuous in a neighborhood of x_0 .

Suppose now that assumptions (A1)–(A3) hold. For $t \neq 0$ consider the ratio

$$R_t(\omega) := t^{-1} \left[F(x_0 + th, \omega) - F(x_0, \omega) \right].$$

By assumption (A2) we have that $|R_t(\omega)| \leq C(\omega) ||h||$, and by assumption (A3) that

$$\lim_{t \downarrow 0} R_t(\omega) = F'_{\omega}(x_0, h) \quad \text{w.p.1.}$$

Therefore, it follows by the Lebesgue Dominated Convergence Theorem that

$$\lim_{t\downarrow 0} \int_{\Omega} R_t(\omega) \, dP(\omega) = \int_{\Omega} \lim_{t\downarrow 0} R_t(\omega) \, dP(\omega).$$

Together with assumption (A3) this implies formula (7.125). This proves the assertion (b).

Finally, if $F'_{\omega}(x_0, h)$ is linear in h for almost every ω , i.e., the function $F_{\omega}(\cdot)$ is differentiable at x_0 w.p.1, then (7.125) implies that $f'(x_0, h)$ is linear in h, and hence (7.126) follows. Note that since f(x) is locally Lipschitz continuous, we only need to verify linearity of $f'(x_0, \cdot)$ in order to establish (Fréchet) differentiability of f(x) at x_0 (see theorem 7.2). This completes proof of (c).

The above analysis shows that two basic conditions for interchangeability of the expectation and differentiation operators, i.e., for the validity of formula (7.126), are the above conditions (A2) and (A4). The following lemma shows that if, in addition to the assumptions (A1)–(A3), the directional derivative $F'_{\omega}(x_0, h)$ is *convex* in h w.p.1, then f(x) is differentiable at x_0 if and only if $F(\cdot, \omega)$ is differentiable at x_0 w.p.1.

Lemma 7.50. Let $\psi : \mathbb{R}^n \times \Omega \to \mathbb{R}$ be a random function such that for almost every $\omega \in \Omega$ the function $\psi(\cdot, \omega)$ is convex and positively homogeneous, and the expected value function $\phi(h) := \mathbb{E}[\psi(h, \omega)]$ is well defined and finite valued. Then the expected value function $\phi(\cdot)$ is linear if and only if the function $\psi(\cdot, \omega)$ is linear w.p.1.

Proof. We have here that the expected value function $\phi(\cdot)$ is convex and positively homogeneous. Moreover, it immediately follows from the linearity properties of the expectation operator that if the function $\psi(\cdot, \omega)$ is linear w.p.1, then $\phi(\cdot)$ is also linear.

Conversely, let $e_1, ..., e_n$ be a basis of the space \mathbb{R}^n . Since $\phi(\cdot)$ is convex and positively homogeneous, it follows that $\phi(e_i) + \phi(-e_i) \ge \phi(0) = 0$, i = 1, ..., n. Furthermore, since $\phi(\cdot)$ is finite valued, it is the support function of a convex compact set. This convex set is a singleton iff

$$\phi(e_i) + \phi(-e_i) = 0, \quad i = 1, ..., n.$$
(7.127)

Therefore, $\phi(\cdot)$ is linear iff condition (7.127) holds. Consider the sets

$$A_i := \left\{ \omega \in \Omega : \psi(e_i, \omega) + \psi(-e_i, \omega) > 0 \right\}$$

Thus the set of $\omega \in \Omega$ such that $\psi(\cdot, \omega)$ is not linear coincides with the set $\bigcup_{i=1}^{n} A_i$. If $P(\bigcup_{i=1}^{n} A_i) > 0$, then at least one of the sets A_i has a positive measure. Let, for example,

 $P(A_1)$ be positive. Then $\phi(e_1) + \phi(-e_1) > 0$, and hence $\phi(\cdot)$ is not linear. This completes the proof. \Box

Regularity conditions which are required for formula (7.125) to hold are simplified further if the random function $F(x, \omega)$ is convex. In that case, by using the Monotone Convergence Theorem instead of the Lebesgue Dominated Convergence Theorem, it is possible to prove the following result.

Theorem 7.51. Suppose that the random function $F(x, \omega)$ is convex and the expected value function f(x) is well defined and finite valued in a neighborhood of a point x_0 . Then f(x) is convex, directionally differentiable at x_0 and formula (7.125) holds. Moreover, f(x) is differentiable at x_0 if and only if $F_{\omega}(x)$ is differentiable at x_0 w.p.1, in which case formula (7.126) holds.

Proof. The convexity of f(x) follows from convexity of $F_{\omega}(\cdot)$. Since f(x) is convex and finite valued near x_0 it follows that f(x) is directionally differentiable at x_0 with finite directional derivative $f'(x_0, h)$ for every $h \in \mathbb{R}^n$. Consider a direction $h \in \mathbb{R}^n$. Since f(x) is finite valued near x_0 , we have that $f(x_0)$ and, for some $t_0 > 0$, $f(x_0 + t_0 h)$ are finite. It follows from the convexity of $F_{\omega}(\cdot)$ that the ratio

$$R_t(\omega) := t^{-1} \left[F(x_0 + th, \omega) - F(x_0, \omega) \right]$$

is monotonically decreasing to $F'_{\omega}(x_0, h)$ as $t \downarrow 0$. Also we have that

$$\mathbb{E}\left|R_{t_0}(\omega)\right| \le t_0^{-1} \left(\mathbb{E}\left|F(x_0 + t_0 h, \omega)\right| + \mathbb{E}\left|F(x_0, \omega)\right|\right) < +\infty.$$

Then it follows by the Monotone Convergence Theorem that

$$\lim_{t \downarrow 0} \mathbb{E}[R_t(\omega)] = \mathbb{E}\left[\lim_{t \downarrow 0} R_t(\omega)\right] = \mathbb{E}\left[F'_{\omega}(x_0, h)\right].$$
(7.128)

Since $\mathbb{E}[R_t(\omega)] = t^{-1}[f(x_0 + th) - f(x_0)]$, we have that the left hand side of (7.128) is equal to $f'(x_0, h)$, and hence formula (7.125) follows.

The last assertion follows then from Lemma 7.50. \Box

Remark 54. It is possible to give a version of the above result for a particular direction $h \in \mathbb{R}^n$. That is, suppose that: (i) the expected value function f(x) is well defined in a neighborhood of a point x_0 , (ii) $f(x_0)$ is finite, (iii) for almost every $\omega \in \Omega$ the function $F_{\omega}(\cdot) := F(\cdot, \omega)$ is convex, (iv) $\mathbb{E}[F(x_0 + t_0h, \omega)] < +\infty$ for some $t_0 > 0$. Then $f'(x_0, h) < +\infty$ and formula (7.125) holds. Note also that if the above assumptions (i)–(iii) are satisfied and $\mathbb{E}[F(x_0+th, \omega)] = +\infty$ for any t > 0, then clearly $f'(x_0, h) = +\infty$.

Often the expectation operator smoothes the integrand $F(x, \omega)$. Consider, for example, $F(x, \omega) := |x - \xi(\omega)|$ with $x \in \mathbb{R}$ and $\xi(\omega)$ being a real valued random variable. Suppose that $f(x) = \mathbb{E}[F(x, \omega)]$ is finite valued. We have here that $F(\cdot, \omega)$ is convex and $F(\cdot, \omega)$ is differentiable everywhere except $x = \xi(\omega)$. The corresponding derivative is given by $\partial F(x,\omega)/\partial x = 1$ if $x > \xi(\omega)$ and $\partial F(x,\omega)/\partial x = -1$ if $x < \xi(\omega)$. Therefore, f(x) is differentiable at x_0 iff the event $\{\xi(\omega) = x_0\}$ has zero probability, in which case

$$df(x_0)/dx = \mathbb{E}\left[\partial F(x_0,\omega)/\partial x\right] = \Pr(\xi < x_0) - \Pr(\xi > x_0).$$
(7.129)

If the event $\{\xi(\omega) = x_0\}$ has positive probability, then the directional derivatives $f'(x_0, h)$ exist but are not linear in h, that is,

$$f'(x_0, -1) + f'(x_0, 1) = 2 \Pr(\xi = x_0) > 0.$$
(7.130)

We can also investigate differentiability properties of the expectation function by studying the subdifferentiability of the integrand. Suppose for the moment that the set Ω is finite, say $\Omega := \{\omega_1, ..., \omega_K\}$ with $P\{\omega = \omega_k\} = p_k > 0$, and that the functions $F(\cdot, \omega), \omega \in \Omega$, are proper. Then $f(x) = \sum_{k=1}^{K} p_k F(x, \omega_k)$ and dom $f = \bigcap_{k=1}^{K} \text{dom } F_k$, where $F_k(\cdot) := F(\cdot, \omega_k)$. The Moreau–Rockafellar Theorem (Theorem 7.4) allows us to express the subdifferential of f(x) as the sum of subdifferentials of $p_k F(x, \omega_k)$. That is, suppose that: (i) the set $\Omega = \{\omega_1, ..., \omega_K\}$ is finite, (ii) for every $\omega_k \in \Omega$ the function $F_k(\cdot) := F(\cdot, \omega_k)$ is proper and convex, (iii) the sets $ri(\text{dom } F_k), k = 1, ..., K$, have a common point. Then for any $x_0 \in \text{dom } f$,

$$\partial f(x_0) = \sum_{k=1}^{K} p_k \partial F(x_0, \omega_k).$$
(7.131)

Note that the above regularity assumption (iii) holds, in particular, if the interior of dom f is nonempty.

The subdifferentials at the right hand side of (7.131) are taken with respect to x. Note that $\partial F(x_0, \omega_k)$, and hence $\partial f(x_0)$, in (7.131) can be unbounded or empty. Suppose that all probabilities p_k are positive. It follows then from (7.131) that $\partial f(x_0)$ is a singleton iff all subdifferentials $\partial F(x_0, \omega_k)$, k = 1, ..., K, are singletons. That is, $f(\cdot)$ is differentiable at a point $x_0 \in \text{dom } f$ iff all $F(\cdot, \omega_k)$ are differentiable at x_0 .

Remark 55. In the case of a finite set Ω we didn't have to worry about the measurability of the multifunction $\omega \mapsto \partial F(x, \omega)$. Consider now a general case where the measurable space does not need to be finite. Suppose that the function $F(x, \omega)$ is random lower semicontinuous and for a.e. $\omega \in \Omega$ the function $F(\cdot, \omega)$ is convex and proper. Then for any $x \in \mathbb{R}^n$, the multifunction $\omega \mapsto \partial F(x, \omega)$ is measurable. Indeed, consider the conjugate

$$F^*(z,\omega) := \sup_{x \in \mathbb{R}^n} \left\{ z^\mathsf{T} x - F(x,\omega) \right\}$$

of the function $F(\cdot, \omega)$. It is possible to show that the function $F^*(z, \omega)$ is also random lower semicontinuous. Moreover, by the Fenchel–Moreau Theorem, $F^{**} = F$ and by convex analysis (see (7.24))

$$\partial F(x,\omega) = \operatorname*{arg\,max}_{z \in \mathbb{R}^n} \{ z^\mathsf{T} x - F^*(z,\omega) \}.$$

Then it follows by Theorem 7.42 that the multifunction $\omega \mapsto \partial F(x, \omega)$ is measurable.

In general we have the following extension of formula (7.131).

Theorem 7.52. Suppose that: (i) the function $F(x, \omega)$ is random lower semicontinuous, (ii) for a.e. $\omega \in \Omega$ the function $F(\cdot, \omega)$ is convex, (iii) the expectation function f is proper, (iv) the domain of f has a nonempty interior. Then for any $x_0 \in \text{dom } f$,

$$\partial f(x_0) = \int_{\Omega} \partial F(x_0, \omega) \, dP(\omega) + \mathcal{N}_{\operatorname{dom} f}(x_0). \tag{7.132}$$

Proof. Consider a point $z \in \int_{\Omega} \partial F(x_0, \omega) dP(\omega)$. By the definition of that integral we have then that there exists a *P*-integrable selection $G(\omega) \in \partial F(x_0, \omega)$ such that $z = \int_{\Omega} G(\omega) dP(\omega)$. Consequently, for a.e. $\omega \in \Omega$ the following holds

$$F(x,\omega) - F(x_0,\omega) \ge G(\omega)^{\mathsf{T}}(x-x_0), \ \forall x \in \mathbb{R}^n.$$

By taking the integral of the both sides of the above inequality we obtain that z is a subgradient of f at x_0 . This shows that

$$\int_{\Omega} \partial F(x_0, \omega) \, dP(\omega) \subset \partial f(x_0). \tag{7.133}$$

In particular, it follows from (7.133) that if $\partial f(x_0)$ is empty, then the set at the right hand side of (7.132) is also empty. If $\partial f(x_0)$ is nonempty, i.e., f is subdifferentiable at x_0 , then $\mathcal{N}_{\text{dom } f}(x_0)$ forms the recession cone of $\partial f(x_0)$. In any case it follows from (7.133) that

$$\int_{\Omega} \partial F(x_0, \omega) \, dP(\omega) + \mathcal{N}_{\operatorname{dom} f}(x_0) \subset \partial f(x_0). \tag{7.134}$$

Note that inclusion (7.134) holds irrespective of assumption (iv).

Proving the converse of inclusion (7.134) is a more delicate problem. Let us outline main steps of such a proof based on the interchangeability property of the directional derivative and integral operators. We can assume that both sets at the left and right hand sides of (7.133) are nonempty. Since the subdifferentials $\partial F(x_0, \omega)$ are convex, it is quite easy to show that the set $\int_{\Omega} \partial F(x_0, \omega) dP(\omega)$ is convex. With some additional effort it is possible to show that this set is closed. Let us denote by $S_1(\cdot)$ and $S_2(\cdot)$ the support functions of the sets at the left and right hand sides of (7.134), respectively. By virtue of inclusion (7.133), $\mathcal{N}_{\text{dom } f}(x_0)$ forms the recession cone of the set at the left hand side of (7.134) as well. Since the tangent cone $\mathcal{T}_{\text{dom } f}(x_0)$ is the polar of $\mathcal{N}_{\text{dom } f}(x_0)$, it follows that $S_1(h) = S_2(h) = +\infty$ for any $h \notin \mathcal{T}_{\text{dom } f}(x_0)$. Suppose now that (7.132) does not hold, i.e., inclusion (7.134) is strict. Then $S_1(h) < S_2(h)$ for some $h \in \mathcal{T}_{\text{dom } f}(x_0)$. Moreover, by assumption (iv), the tangent cone $\mathcal{T}_{\text{dom } f}(x_0)$ has a nonempty interior and there exists \bar{h} in the interior of $\mathcal{T}_{\text{dom } f}(x_0)$ such that $S_1(\bar{h}) < S_2(\bar{h})$. For such \bar{h} the directional derivative $f'(x_0, h)$ is finite for all h in a neighborhood of \bar{h} , $f'(x_0, \bar{h}) = S_2(\bar{h})$ and (see Remark 54 on page 454)

$$f'(x_0,\bar{h}) = \int_{\Omega} F'_{\omega}(x_0,\bar{h}) \, dP(\omega).$$
Also, $F'_{\omega}(x_0, h)$ is finite for a.e. ω and for all h in a neighborhood of \bar{h} , and hence $F'_{\omega}(x_0, \bar{h}) = \bar{h}^{\mathsf{T}} G(\omega)$ for some $G(\omega) \in \partial F(x_0, \omega)$. Moreover, since the multifunction $\omega \mapsto \partial F(x_0, \omega)$ is measurable, we can choose a measurable $G(\omega)$ here. Consequently,

$$\int_{\Omega} F'_{\omega}(x_0, \bar{h}) \, dP(\omega) = \bar{h}^{\mathsf{T}} \int_{\Omega} G(\omega) \, dP(\omega)$$

Since $\int_{\Omega} G(\omega) dP(\omega)$ is a point of the set at the left hand side of (7.133), we obtain that $s_1(\bar{h}) \ge f'(x_0, \bar{h}) = s_2(\bar{h})$, a contradiction. \Box

In particular, if x_0 is an *interior* point of the domain of f, then under the assumptions of the above theorem we have that

$$\partial f(x_0) = \int_{\Omega} \partial F(x_0, \omega) \, dP(\omega). \tag{7.135}$$

Also it follows from formula (7.132) that $f(\cdot)$ is differentiable at x_0 iff x_0 is an interior point of the domain of f and $\partial F(x_0, \omega)$ is a singleton for a.e. $\omega \in \Omega$, i.e., $F(\cdot, \omega)$ is differentiable at x_0 w.p.1.

7.2.5 Uniform Laws of Large Numbers

Consider a sequence $\xi^i = \xi^i(\omega)$, $i \in \mathbb{N}$, of *d*-dimensional random vectors defined on a probability space (Ω, \mathcal{F}, P) . As it was discussed in section 7.2.1 we can view ξ^i as random vectors supported on a (closed) set $\Xi \subset \mathbb{R}^d$, equipped with its Borel sigma algebra \mathcal{B} . We say that ξ^i , $i \in \mathbb{N}$, are *identically distributed* if each ξ^i has the same probability distribution on (Ξ, \mathcal{B}) . If, moreover, ξ^i , $i \in \mathbb{N}$, are independent, we say that they are *independent identically distributed* (iid). Consider a measurable function $F : \Xi \to \mathbb{R}$ and the sequence $F(\xi^i)$, $i \in \mathbb{N}$, of random variables. If ξ^i are identically distributed, then $F(\xi^i)$, $i \in \mathbb{N}$, are also identically distributed and hence their expectations $\mathbb{E}[F(\xi^i)]$ are constant, i.e., $\mathbb{E}[F(\xi^i)] = \mathbb{E}[F(\xi^1)]$ for all $i \in \mathbb{N}$. The Law of Large Numbers (LLN) says that if ξ^i are identically distributed and the expectation $\mathbb{E}[F(\xi^1)]$ is well defined, then, under some regularity conditions¹⁷,

$$N^{-1}\sum_{i=1}^{N} F(\xi^{i}) \to \mathbb{E}\left[F(\xi^{1})\right] \text{ w.p.1 as } N \to \infty.$$
(7.136)

In particular, the classical LLN states that the convergence (7.136) holds if the sequence ξ^i is iid.

Consider now a random function $F: X \times \Xi \to \mathbb{R}$, where X is a nonempty subset of \mathbb{R}^n and $\xi = \xi(\omega)$ is a random vector supported on the set Ξ . Suppose that the corresponding expected value function $f(x) := \mathbb{E}[F(x,\xi)]$ is well defined and finite valued for every $x \in X$. Let $\xi^i = \xi^i(\omega), i \in \mathbb{N}$, be an iid sequence of random vectors having the same

¹⁷Sometimes (7.136) is referred to as the *strong* LLN to distinguish it from the *weak* LLN where the convergence is ensured 'in probability' instead of 'with probability one'. Unless stated otherwise we deal with the strong LLN.

distribution as the random vector ξ , and

$$\hat{f}_N(x) := N^{-1} \sum_{i=1}^N F(x, \xi^i)$$
(7.137)

be the so-called *sample average* functions. Note that the sample average function $\hat{f}_N(x)$ depends on the random sequence $\xi^1, ..., \xi^N$, and hence is a random function. Since we assumed that all $\xi^i = \xi^i(\omega)$ are defined on the same probability space, we can view $\hat{f}_N(x) = \hat{f}_N(x, \omega)$ as a sequence of functions of $x \in X$ and $\omega \in \Omega$.

We have that for every fixed $x \in X$ the LLN holds, i.e.,

$$f_N(x) \to f(x) \text{ w.p.1 as } N \to \infty.$$
 (7.138)

This means that for a.e. $\omega \in \Omega$, the sequence $\hat{f}_N(x, \omega)$ converges to f(x). That is, for any $\varepsilon > 0$ and a.e. $\omega \in \Omega$ there exists $N^* = N^*(\varepsilon, \omega, x)$ such that $|\hat{f}_N(x) - f(x)| < \varepsilon$ for any $N \ge N^*$. It should be emphasized that N^* depends on ε and ω , and also on $x \in X$. We may refer to (7.138) as a *pointwise* LLN. In some applications we will need a stronger form of LLN where the number N^* can be chosen independent of $x \in X$. That is, we say that $\hat{f}_N(x)$ converges to f(x) w.p.1 uniformly on X if

$$\sup_{x \in X} \left| \hat{f}_N(x) - f(x) \right| \to 0 \quad \text{w.p.1} \quad \text{as } N \to \infty, \tag{7.139}$$

and refer to this as the uniform LLN.

We have the following basic result. It is said that $F(x,\xi)$, $x \in X$, is *dominated* by an integrable function if there exists a nonnegative valued measurable function $g(\xi)$ such that $\mathbb{E}[g(\xi)] < +\infty$ and for every $x \in X$ the inequality $|F(x,\xi)| \leq g(\xi)$ holds w.p.1.

Theorem 7.53. Let X be a nonempty compact subset of \mathbb{R}^n and suppose that: (i) for any $x \in X$ the function $F(\cdot, \xi)$ is continuous at x for almost every $\xi \in \Xi$, (ii) $F(x, \xi)$, $x \in X$, is dominated by an integrable function, (iii) the sample is iid. Then the expected value function f(x) is finite valued and continuous on X, and $\hat{f}_N(x)$ converges to f(x) w.p.1 uniformly on X.

Proof. It follows from the assumption (ii) that $|f(x)| \leq \mathbb{E}[g(\xi)]$, and consequently $|f(x)| < +\infty$ for all $x \in X$. Consider a point $x \in X$ and let x_k be a sequence of points in X converging to x. By the Lebesgue Dominated Convergence Theorem, assumption (ii) implies that

$$\lim_{k \to \infty} \mathbb{E}\left[F(x_k, \xi)\right] = \mathbb{E}\left[\lim_{k \to \infty} F(x_k, \xi)\right].$$

Since, by (i), $F(x_k, \xi) \to F(x, \xi)$ w.p.1, it follows that $f(x_k) \to f(x)$, and hence f(x) is continuous.

Choose now a point $\bar{x} \in X$, a sequence γ_k of positive numbers converging to zero, and define $V_k := \{x \in X : ||x - \bar{x}|| \le \gamma_k\}$ and

$$\Delta_k(\xi) := \sup_{x \in V_k} \left| F(x,\xi) - F(\bar{x},\xi) \right|.$$
(7.140)

By the assumption (i) we have that for a.e. $\xi \in \Xi$, $\Delta_k(\xi)$ tends to zero as $k \to \infty$. Moreover, by the assumption (ii) we have that $\Delta_k(\xi)$, $k \in \mathbb{N}$, are dominated by an integrable function, and hence by the Lebesgue Dominated Convergence Theorem we have that

$$\lim_{k \to \infty} \mathbb{E}\left[\Delta_k(\xi)\right] = \mathbb{E}\left[\lim_{k \to \infty} \Delta_k(\xi)\right] = 0.$$
(7.141)

We also have that

$$\left|\hat{f}_{N}(x) - \hat{f}_{N}(\bar{x})\right| \leq \frac{1}{N} \sum_{i=1}^{N} \left|F(x,\xi^{i}) - F(\bar{x},\xi^{i})\right|,$$

and hence

$$\sup_{x \in V_k} \left| \hat{f}_N(x) - \hat{f}_N(\bar{x}) \right| \le \frac{1}{N} \sum_{i=1}^N \Delta_k(\xi^i).$$
(7.142)

Since the sequence ξ^i is iid, it follows by the LLN that the right hand side of (7.142) converges w.p.1 to $\mathbb{E}[\Delta_k(\xi)]$ as $N \to \infty$. Together with (7.141) this implies that for any given $\varepsilon > 0$ there exists a neighborhood W of \bar{x} and $\bar{N} = \bar{N}_W(\omega)$ such that w.p.1 for all $N \ge \bar{N}$ it holds that

$$\sup_{x \in W \cap X} \left| \hat{f}_N(x) - \hat{f}_N(\bar{x}) \right| < \varepsilon.$$

Since X is compact, there exists a finite number of points $x_1, ..., x_m \in X$ and corresponding neighborhoods $W_1, ..., W_m$ covering X such that w.p.1 for $N \ge \overline{N}(\omega) := \max\{\overline{N}_{W_1}(\omega), ..., \overline{N}_{W_m}(\omega)\}$ the following holds

$$\sup_{x \in W_j \cap X} \left| \hat{f}_N(x) - \hat{f}_N(x_j) \right| < \varepsilon, \quad j = 1, ..., m.$$
(7.143)

Furthermore, since f(x) is continuous on X, these neighborhoods can be chosen in such a way that

$$\sup_{x \in W_j \cap X} \left| f(x) - f(x_j) \right| < \varepsilon, \quad j = 1, ..., m.$$
(7.144)

Again by the LLN we have that $\hat{f}_N(x)$ converges pointwise to f(x) w.p.1. Therefore,

$$\left|\hat{f}_{N}(x_{j}) - f(x_{j})\right| < \varepsilon, \quad j = 1, ..., m,$$
(7.145)

w.p.1 for sufficiently large $N \ge N^*(\omega)$. It follows from (7.143)–(7.145) that w.p.1 for N larger than $\max\{\bar{N}(\omega), N^*(\omega)\}$ it holds that

$$\sup_{x \in X} \left| \hat{f}_N(x) - f(x) \right| < 3\varepsilon.$$
(7.146)

Since $\varepsilon>0$ was arbitrary, we obtain that (7.139) follows and hence the proof is complete. \Box

Remark 56. It could be noted that the assumption (i) in the above theorem means that $F(\cdot, \xi)$ is continuous at any given point $x \in X$ w.p.1. This does not mean, however, that

 $F(\cdot,\xi)$ is continuous on X w.p.1. Take, for example, $F(x,\xi) := \mathbf{1}_{\mathbb{R}_+}(x-\xi), x, \xi \in \mathbb{R}$, i.e., $F(x,\xi) = 1$ if $x \ge \xi$ and $F(x,\xi) = 0$ otherwise. We have here that $F(\cdot,\xi)$ is always discontinuous at $x = \xi$, and that the expectation $\mathbb{E}[F(x,\xi)]$ is equal to the probability $\Pr(\xi \le x)$, i.e., $f(x) = \mathbb{E}[F(x,\xi)]$ is the cumulative distribution function (cdf) of ξ . The assumption (i) means here that for any given x, probability of the event " $x = \xi$ " is zero, i.e., that the cdf of ξ is continuous at x. In this example, the sample average function $\widehat{f}_N(\cdot)$ is just the empirical cdf of the considered random sample. The fact that the empirical cdf converges to its true counterpart uniformly on \mathbb{R} w.p.1 is known as the Glivenko-Cantelli Theorem. In fact the Glivenko-Cantelli Theorem states that the uniform convergence holds even if the corresponding cdf is discontinuous.

The analysis simplifies further if for a.e. $\xi \in \Xi$ the function $F(\cdot,\xi)$ is convex, i.e., the random function $F(x,\xi)$ is convex. We can view $\hat{f}_N(x) = \hat{f}_N(x,\omega)$ as a sequence of random functions defined on a common probability space (Ω, \mathcal{F}, P) . Recall definition 7.29 of epiconvergence of extended real valued functions. We say that functions \hat{f}_N epiconverge to f w.p.1, written $\hat{f}_N \stackrel{e}{\to} f$ w.p.1, if for a.e. $\omega \in \Omega$ the functions $\hat{f}_N(\cdot,\omega)$ epiconverge to $f(\cdot)$. In the following theorem we assume that function $F(x,\xi) : \mathbb{R}^n \times \Xi \to \overline{\mathbb{R}}$ is an extended real valued function, i.e., can take values $\pm \infty$.

Theorem 7.54. Suppose that for almost every $\xi \in \Xi$ the function $F(\cdot, \xi)$ is an extended real valued convex function, the expected value function $f(\cdot)$ is lower semicontinuous and its domain, dom f, has a nonempty interior, and the pointwise LLN holds. Then $\hat{f}_N \stackrel{\text{e}}{\to} f$ w.p.1.

Proof. It follows from the assumed convexity of $F(\cdot,\xi)$ that the function $f(\cdot)$ is convex and that w.p.1 the functions $\hat{f}_N(\cdot)$ are convex. Let us choose a countable and dense subset D of \mathbb{R}^n . By the pointwise LLN we have that for any $x \in D$, $\hat{f}_N(x)$ converges to f(x)w.p.1 as $N \to \infty$. This means that there exists a set $\Upsilon_x \subset \Omega$ of P-measure zero such that for any $\omega \in \Omega \setminus \Upsilon_x$, $\hat{f}_N(x, \omega)$ tends to f(x) as $N \to \infty$. Consider the set $\Upsilon := \bigcup_{x \in D} \Upsilon_x$. Since the set D is countable and $P(\Upsilon_x) = 0$ for every $x \in D$, we have that $P(\Upsilon) = 0$. We also have that for any $\omega \in \Omega \setminus \Upsilon$, $\hat{f}_N(x, \omega)$ converges to f(x), as $N \to \infty$, pointwise on D. It follows then by Theorem 7.31 that $\hat{f}_N(\cdot, \omega) \stackrel{\text{e}}{\to} f(\cdot)$ for any $\omega \in \Omega \setminus \Upsilon$. That is, $\hat{f}_N(\cdot) \stackrel{\text{e}}{\to} f(\cdot)$ w.p.1. \Box

We also have the following result. It can be proved in a way similar to the proof of the above theorem by using Theorem 7.31.

Theorem 7.55. Suppose that the random function $F(x, \xi)$ is convex and let X be a compact subset of \mathbb{R}^n . Suppose that the expectation function f(x) is finite valued on a neighborhood of X and that the pointwise LLN holds for every x in a neighborhood of X. Then $\hat{f}_N(x)$ converges to f(x) w.p.1 uniformly on X.

It is worthwhile to note that in some cases the pointwise LLN can be verified by ad hoc methods, and hence the above epi-convergence and uniform LLN for *convex* random functions can be applied, without the assumption of independence.

For iid random samples we have the following version of epi-convergence LLN.

Theorem 7.56 (Artstein-Wets). Suppose that: (i) the function $F(x,\xi)$ is random lower semicontinuous, (ii) for every $\bar{x} \in \mathbb{R}^n$ there exists a neighborhood \mathcal{V} of \bar{x} and P-integrable function $h : \Xi \to \mathbb{R}$ such that $F(x,\xi) \ge h(\xi)$ for all $x \in \mathcal{V}$ and a.e. $\xi \in \Xi$, (iii) the sample is iid. Then $\hat{f}_N \stackrel{\text{e}}{\to} f$ w.p.1.

We will give a proof for a somewhat more general result in Theorem 7.60 below.

Uniform LLN for Derivatives

Let us discuss now uniform LLN for derivatives of the sample average function. By Theorem 7.49 we have that, under the corresponding assumptions (A1),(A2) and (A4), the expectation function is differentiable at the point x_0 and the derivatives can be taken inside the expectation, i.e., formula (7.126) holds. Now if we assume that the expectation function is well defined and finite valued, $\nabla_x F(\cdot, \xi)$ is continuous on X for a.e. $\xi \in \Xi$, and $\|\nabla_x F(x,\xi)\|$, $x \in X$, is dominated by an integrable function, then the assumptions (A1),(A2) and (A4) hold and by Theorem 7.53 we obtain that f(x) is continuously differentiable on X and $\nabla \hat{f}_N(x)$ converges to $\nabla f(x)$ w.p.1 uniformly on X. However, in many interesting applications the function $F(\cdot,\xi)$ is not everywhere differentiable for any $\xi \in \Xi$, and yet the expectation function is smooth. Such simple example of $F(x,\xi) := |x - \xi|$ was discussed after Remark 54 on page 454.

Theorem 7.57. Let $U \subset \mathbb{R}^n$ be an open set, X be a nonempty compact subset of U and $F: U \times \Xi \to \mathbb{R}$ be a random function. Suppose that: (i) $\{F(x,\xi)\}_{x \in X}$ is dominated by an integrable function, (ii) there exists an integrable function $C(\xi)$ such that

$$|F(x',\xi) - F(x,\xi)| \le C(\xi) ||x' - x|| \text{ a.e. } \xi \in \Xi, \ \forall x, x' \in U.$$
(7.147)

(iii) for every $x \in X$ the function $F(\cdot, \xi)$ is continuously differentiable at x w.p.1. Then the following holds: (a) the expectation function f(x) is finite valued and continuously differentiable on X, (b) for all $x \in X$ the corresponding derivatives can be taken inside the integral, i.e.,

$$\nabla f(x) = \mathbb{E}\left[\nabla_x F(x,\xi)\right],\tag{7.148}$$

(c) Clarke generalized gradient $\partial^{\circ} \hat{f}_N(x)$ converges to $\nabla f(x)$ w.p.1 uniformly on X, i.e.,

$$\lim_{N \to \infty} \sup_{x \in X} \mathbb{D}\left(\partial^{\circ} \hat{f}_N(x), \{\nabla f(x)\}\right) = 0 \quad \text{w.p.1.}$$
(7.149)

Proof. Assumptions (i) and (ii) imply that the expectation function f(x) is finite valued for all $x \in U$. Note that assumption (ii) is basically the same as assumption (A2) and, of course, assumption (iii) implies assumption (A4) of Theorem 7.49. Consequently, it follows by Theorem 7.49 that $f(\cdot)$ is differentiable at every point $x \in X$ and the interchangeability formula (7.148) holds. Moreover, it follows from (7.147) that $\|\nabla_x F(x,\xi)\| \leq C(\xi)$

for a.e. ξ and all $x \in U$ where $\nabla_x F(x,\xi)$ is defined. Hence by assumption (iii) and the Lebesgue Dominated Convergence Theorem, we have that for any sequence x_k in U converging to a point $x \in X$ it follows that

$$\lim_{k \to \infty} \nabla f(x_k) = \mathbb{E} \left[\lim_{k \to \infty} \nabla_x F(x_k, \xi) \right] = \mathbb{E} \left[\nabla_x F(x, \xi) \right] = \nabla f(x).$$

We obtain that $f(\cdot)$ is continuously differentiable on X.

The assertion (c) can be proved by following the same steps as in the proof of Theorem 7.53. That is, consider a point $\bar{x} \in X$, a sequence V_k of shrinking neighborhoods of \bar{x} and

$$\Delta_k(\xi) := \sup_{x \in V_k^*(\xi)} \|\nabla_x F(x,\xi) - \nabla_x F(\bar{x},\xi)\|.$$

Here $V_k^*(\xi)$ denotes the set of points of V_k where $F(\cdot, \xi)$ is differentiable. By assumption (iii) we have that $\Delta_k(\xi) \to 0$ for a.e. ξ . Also

$$\Delta_k(\xi) \le \|\nabla_x F(\bar{x},\xi)\| + \sup_{x \in V_k^*(\xi)} \|\nabla_x F(x,\xi)\| \le 2C(\xi),$$

and hence $\Delta_k(\xi)$, $k \in \mathbb{N}$, are dominated by the integrable function $2C(\xi)$. Consequently

$$\lim_{k \to \infty} \mathbb{E}\left[\Delta_k(\xi)\right] = \mathbb{E}\left[\lim_{k \to \infty} \delta_k(\xi)\right] = 0,$$

and the remainder of the proof can be completed in the same way as the proof of Theorem 7.53 using compactness arguments. \Box

7.2.6 Law of Large Numbers for Risk Measures

Consider settings of section 6.3.3. That is, let (Ω, \mathcal{F}, P) be a nonatomic probability space, $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$, with $p \in [1, \infty)$, and $\rho : \mathcal{Z} \to \mathbb{R}$ be a real valued law invariant risk measure. As it was pointed in Remark 27 on page 319, $\rho(H_Z)$ can be considered as a function of cdf H_Z of $Z \in \mathcal{Z}$. Let $Z^1, ..., Z^N$ be a random sample of $Z \in \mathcal{Z}$, and

$$\hat{H}_N(z) = N^{-1} \sum_{i=1}^N \mathbf{1}_{(-\infty,z]}(Z^i)$$

be the corresponding empirical cdf. Then a sample (empirical) estimate of $\rho(Z)$ is given by $\rho(\hat{H}_N)$. We have the following LLN for the sample estimates $\rho(\hat{H}_N)$.

Theorem 7.58. Let $\rho : \mathbb{Z} \to \mathbb{R}$ be a real valued law invariant convex risk measure and \hat{H}_N be the empirical cdf associated with an iid sample of $Z \in \mathbb{Z}$. Then $\rho(\hat{H}_N)$ converges to $\rho(Z)$ w.p.1 as $N \to \infty$.

Proof. We can view H_Z^{-1} as an element of the space $\hat{\mathcal{Z}} := \mathcal{L}_p(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$, where $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$ is the standard uniform probability space, and by law invariance of ρ we have

that $\rho(Z) = \rho(H_Z^{-1})$. The function \hat{H}_N^{-1} is piecewise constant and hence $\hat{H}_N^{-1} \in \hat{Z}$. Consider the set $\mathcal{C} \subset [0, 1]$ of points where function $H^{-1} = H_Z^{-1}$ is discontinuous. Since H^{-1} is monotonically nondecreasing, the set \mathcal{C} is countable and hence has measure zero. By Glivenko-Cantelli theorem we have that w.p.1 \hat{H}_N converges to H uniformly on \mathbb{R} . It follows that w.p.1 \hat{H}_N^{-1} converges pointwise to H^{-1} on the set $[0,1] \setminus \mathcal{C}$, and hence w.p.1

$$\lim_{N \to \infty} \int_0^1 |H^{-1}(t) - \hat{H}_N^{-1}(t)|^p dt = \int_0^1 \lim_{N \to \infty} |H^{-1}(t) - \hat{H}_N^{-1}(t)|^p dt = 0, \quad (7.150)$$

provided the limit and integral operators can be interchanged. This interchangeability is justified if w.p.1 the sequence $\psi_N(\cdot) := |H^{-1}(\cdot) - \hat{H}_N^{-1}(\cdot)|^p$ is uniformly integrable (see Theorem 7.36).

Let us show that the uniform integrability indeed holds. We have (triangle inequality) that

$$\left(\int_0^1 \psi_N(\cdot)dt\right)^{1/p} \le \left(\int_0^1 |H^{-1}(t)|^p dt\right)^{1/p} + \left(\int_0^1 |\hat{H}_N^{-1}(t)|^p dt\right)^{1/p}.$$
(7.151)

The term

$$\left(\int_0^1 |H^{-1}(t)|^p dt\right)^{1/p} = \left(\int_0^\infty |z|^p dH(z)\right)^{1/p} = ||Z||_p, \tag{7.152}$$

with the right hand side of (7.152) being a finite constant. Therefore it is sufficient to verify uniform integrability of $|\hat{H}_N^{-1}(\cdot)|^p$. We have that

$$\int_{0}^{1} |\hat{H}_{N}^{-1}(t)|^{p} dt = \int_{0}^{\infty} |z|^{p} d\hat{H}_{N}(z) = \frac{1}{N} \sum_{i=1}^{N} |Z^{i}|^{p},$$
(7.153)

and by the Law of Large Numbers $N^{-1} \sum_{i=1}^{N} |Z^i|^p$ converges w.p.1 to $\mathbb{E}|Z|^p$. Since $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ we have that $\mathbb{E}|Z|^p$ is finite. It follows that $\int_0^1 |\hat{H}_N^{-1}(t)|^p dt$ converges w.p.1 to a finite limit, which implies that w.p.1 $|\hat{H}_N^{-1}(\cdot)|^p$ are uniformly integrable (see Theorem 7.36).

By (7.150) this shows that \hat{H}_N^{-1} converges to H^{-1} w.p.1 in the norm topology of the space $\mathcal{L}_p(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$. Since the risk measure ρ is real valued, it is continuous in the norm topology (Proposition 6.6). It follows that $\rho(\hat{H}_N^{-1})$ converges to $\rho(H_Z^{-1})$ w.p.1.

For example, the LLN holds for the sample estimate of AV@R_{α}(Z), $\alpha \in (0, 1]$. This was shown in section 6.6.1 by different arguments. It is interesting to note that for the Value-at-Risk measure V@R_{α}(Z) = $H_Z^{-1}(1 - \alpha)$, the LLN does not hold if $H_Z^{-1}(\cdot)$ is discontinuous at $1 - \alpha$.

Now consider a function $F : \mathbb{R}^n \times \Xi \to \mathbb{R}$ and random vector $\xi = \xi(\omega)$ supported on the set $\Xi \subset \mathbb{R}^d$. With some abuse of the notation we write $F(x, \omega)$ for the random function $F(x, \xi(\omega))$. We assume the following.

($\mathfrak{R}1$) For every $x \in \mathbb{R}^n$ the random variable $F_x(\omega) = F(x, \omega)$ belongs to the space \mathcal{Z} .

Then we can consider the composite function $\phi : \mathbb{R}^n \to \mathbb{R}$ defined as $\phi(x) := \rho(F_x)$.

Let $\xi^i = \xi^i(\omega)$, i = 1, ..., N, be an iid sample of the random vector ξ , and \hat{H}_{xN} be the corresponding empirical cdf associated with the sample $F(x, \xi^1), ..., F(x, \xi^N)$. Then $\hat{\phi}_N(x) := \rho(\hat{H}_{xN})$ gives the corresponding sample estimate of the composite function $\phi(x)$. By Theorem 7.58 we have that if the risk measure ρ is law invariant and convex, then for any given $x \in \mathbb{R}^n$, $\hat{\phi}_N(x)$ converges to $\phi(x)$ w.p.1. Recall that if $F(\cdot, \omega)$ is convex for a.e. $\omega \in \Omega$ and $\rho : \mathcal{Z} \to \mathbb{R}$ is a convex risk measure, then the composite function ϕ is convex. Therefore the following result can be proved in the same way as Theorem 7.54.

Theorem 7.59. Suppose that: (i) condition (\Re 1) holds, (ii) for a.e. $\omega \in \Omega$ the function $F(\cdot, \omega)$ is convex, (iii) the risk measure $\rho : \mathbb{Z} \to \mathbb{R}$ is law invariant and convex. Then the functions $\phi : \mathbb{R}^n \to \mathbb{R}$ and $\hat{\phi}_N : \mathbb{R}^n \to \mathbb{R}$ are convex and $\hat{\phi}_N \stackrel{e}{\to} \phi$ w.p.1.

The question of epiconvergence of ϕ_N to ϕ is considerably more delicate in the nonconvex case. Consider the following conditions.

- ($\Re 2$) The function $F(x, \omega)$ is random lower semicontinuous.
- (\Re 3) For every $\bar{x} \in \mathbb{R}^n$ there is a neighborhood $\mathcal{V}_{\bar{x}}$ of \bar{x} and a function $h \in \mathcal{Z}$ such that $F(x, \cdot) \geq h(\cdot)$ for all $x \in \mathcal{V}_{\bar{x}}$.

Theorem 7.60. Suppose that conditions $(\Re 1)$ – $(\Re 3)$ hold and the risk measure $\rho : \mathbb{Z} \to \mathbb{R}$ is law invariant and convex. Then the functions ϕ is lower semicontinuous and $\hat{\phi}_N \xrightarrow{e} \phi$ w.p.1.

Proof. Since the risk measure ρ is convex, it has the dual representation (6.38). Consider a point $\bar{x} \in \mathbb{R}^n$. By condition ($\mathfrak{R}3$) we have for any $\zeta \in \mathfrak{A}$ and $x \in \mathcal{V}_{\bar{x}}$ that $\zeta(\cdot)F(x, \cdot) \geq \zeta(\cdot)h(\cdot)$. Moreover, since $h \in \mathbb{Z}$ and $\zeta \in \mathbb{Z}^*$ we have that $\int_{\Omega} \zeta(\omega)h(\omega)dP(\omega)$ is finite. Hence by Fatou's lemma it follows that for a sequence $x_k \to \bar{x}$,

$$\liminf_{k \to \infty} \int_{\Omega} \zeta(\omega) F(x_k, \omega) dP(\omega) \geq \int_{\Omega} \liminf_{k \to \infty} \zeta(\omega) F(x_k, \omega) dP(\omega) \\ \geq \int_{\Omega} \zeta(\omega) F(\bar{x}, \omega) dP(\omega),$$

where the last inequality follows by lower semicontinuity of $F(\cdot, \omega)$ (condition (\Re 2)). That is, function $x \mapsto \int_{\Omega} \zeta(\omega) F(x, \omega) dP(\omega) - \rho^*(\zeta)$ is lower semicontinuous. Since $\phi(x) = \rho(F_x)$ is given by maximum of such functions, it follows that $\phi(x)$ is also lower semicontinuous (see Proposition 7.24).

In order to show that $\hat{\phi}_N \stackrel{e}{\to} \phi$ w.p.1 we need to verify the respective conditions (7.96) and (7.97) of Definition 7.29. That is, to verify the condition (7.96) we have to show that there exists a set $\Delta \subset \Omega$ of measure zero such that for any point $\bar{x} \in \mathbb{R}^n$ and any sequence x_N converging to \bar{x} it holds that

$$\liminf_{N \to \infty} \hat{\phi}_N(x_N, \omega) \ge \phi(\bar{x}), \quad \forall \omega \in \Omega \setminus \Delta.$$
(7.154)

Note that the set Δ should not depend on \bar{x} . Similarly for the condition (7.97).

To verify condition (7.154) we proceed as follows. For some sequence $\gamma_k \downarrow 0$ of positive numbers consider $V_k := \{x \in \mathbb{R}^n : ||x - \bar{x}|| < \gamma_k\}$, and let

$$\vartheta_k(\omega) := \inf_{x \in V_k} F(x, \omega), \ k \in \mathbb{N}.$$

Since $F(x, \omega)$ is random lower semicontinuous (condition ($\Re 2$)), we have that $\vartheta_k(\omega)$ is measurable (Theorem 7.42). By condition ($\Re 3$) we have $\vartheta_k(\cdot) \ge h(\cdot)$ for all k large enough (such that $V_k \subset \mathcal{V}_{\bar{x}}$). Of course, we also have that $F(x, \cdot) \ge \vartheta_k(\cdot)$ for any $x \in V_k$. It follows that $\vartheta_k \in \mathbb{Z}$. Consider the dual representation (6.38) of ρ and let

$$\bar{\zeta} \in \arg\max_{\zeta \in \mathfrak{A}} \left\{ \int_{\Omega} \zeta(\omega) F(\bar{x}, \omega) dP(\omega) - \rho^*(\zeta) \right\}$$

(recall that such maximizer exists). That is, $\bar{\zeta} \in \mathfrak{A}$ and

$$\phi(\bar{x}) = \int_{\Omega} \bar{\zeta}(\omega) F(\bar{x}, \omega) dP(\omega) - \rho^*(\bar{\zeta}).$$

Since $\vartheta_k \in \mathcal{Z}$ we have by (6.38) that

$$\rho(\vartheta_k) \ge \int_{\Omega} \bar{\zeta}(\omega) \vartheta_k(\omega) dP(\omega) - \rho^*(\bar{\zeta}).$$

By condition (\Re 3), $\bar{\zeta}(\cdot)\vartheta_k(\cdot)$ is bounded from below, on a neighborhood of \bar{x} , by integrable function $\bar{\zeta}(\cdot)h(\cdot)$, and hence applying Fatou's lemma we have

$$\liminf_{k \to \infty} \int_{\Omega} \bar{\zeta}(\omega) \vartheta_k(\omega) dP(\omega) \ge \int_{\Omega} \liminf_{k \to \infty} \bar{\zeta}(\omega) \vartheta_k(\omega) dP(\omega), \tag{7.155}$$

and by lower semicontinuity of $F(\cdot, \omega)$

$$\int_{\Omega} \liminf_{k \to \infty} \bar{\zeta}(\omega) \vartheta_k(\omega) dP(\omega) - \rho^*(\bar{\zeta}) \ge \int_{\Omega} \bar{\zeta}(\omega) F(\bar{x}, \omega) dP(\omega) - \rho^*(\bar{\zeta}) = \phi(\bar{x}).$$
(7.156)

We obtain that

$$\liminf_{k \to \infty} \rho(\vartheta_k) \ge \phi(\bar{x}). \tag{7.157}$$

Now let us choose $\varepsilon > 0$. By (7.157) there exists $\bar{k} = \bar{k}(\varepsilon)$ such that

$$\rho(\vartheta_{\bar{k}}) \ge \phi(\bar{x}) - \varepsilon. \tag{7.158}$$

Let $\hat{\rho}_N(\vartheta_{\bar{k}})$ be the empirical estimate of $\rho(\vartheta_{\bar{k}})$ based on the same sample as the sample used for the estimate $\hat{\phi}_N(\cdot)$, i.e., $\hat{\rho}_N(\vartheta_{\bar{k}}) = \rho(\tilde{H}_N)$, where \tilde{H}_N is the empirical cdf of the sample $Y^i = \inf_{x \in V_{\bar{k}}} F(x, \xi^i)$, i = 1, ..., N. By Theorem 7.58 we have that $\hat{\rho}_N(\vartheta_{\bar{k}}) \to \rho(\vartheta_{\bar{k}})$ w.p.1 as $N \to \infty$. Hence for a.e. $\omega \in \Omega$ there is $\bar{N}_{\bar{x}}(\omega)$ such that $\hat{\rho}_N(\vartheta_{\bar{k}}) \ge \rho(\vartheta_{\bar{k}}) - \varepsilon$, for all $N \ge \bar{N}_{\bar{x}}(\omega)$. Together with (7.158) this implies that

$$\hat{\rho}_N(\vartheta_{\bar{k}}) \ge \phi(\bar{x}) - 2\varepsilon \tag{7.159}$$

for a.e. $\omega \in \Omega$ and $N \geq \overline{N}_{\overline{x}}(\omega)$.

For any $x \in V_{\bar{k}}$ we have (for the same random sample) that the empirical cdf of $F(x, \cdot)$ dominates the empirical cdf of $\vartheta_{\bar{k}}(\cdot)$, and hence (see Theorem 6.50)

$$\hat{\phi}_N(x) \ge \hat{\rho}_N(\vartheta_{\bar{k}}). \tag{7.160}$$

By (7.159) it follows that

$$\inf_{x \in V_k} \hat{\phi}_N(x) \ge \phi(\bar{x}) - 2\varepsilon \tag{7.161}$$

for a.e. $\omega \in \Omega$ and $N \ge \overline{N}_{\overline{x}}(\omega)$. That is, there exist $\overline{N}(\omega)$, a set $\Upsilon \subset \Omega$ of measure zero and a neighborhood V (both depending on \overline{x} and ε) such that

$$\phi_N(x) \ge \phi(\bar{x}) - 2\varepsilon,$$
(7.162)

for all $N \ge \overline{N}(\omega)$, $x \in V$ and $\omega \in \Omega \setminus \Upsilon$. It follows that there exists a countable number of points $x_1, ..., \text{ in } \mathbb{R}^n$, with the corresponding neighborhoods $V_1, ...,$ covering \mathbb{R}^n . Let $\Upsilon_1, ...,$ be the corresponding sets of measure zero and $\overline{\Upsilon} := \bigcup_{i=1}^{\infty} \Upsilon_i$. Note that the set $\overline{\Upsilon}$ has measure zero, and that $\overline{\Upsilon}$ depends on ε , but not on a particular point of \mathbb{R}^n . It follows that for any $\overline{x} \in \mathbb{R}^n$ there is a neighborhood W and $N^*(\omega)$ such that (7.162) holds for all $x \in W, N \ge N^*(\omega)$ and $\omega \in \Omega \setminus \overline{\Upsilon}$.

Consequently for any point \bar{x} and a sequence x_N converging to \bar{x} we have that

$$\liminf_{N \to \infty} \hat{\phi}_N(x_N, \omega) \ge \phi(\bar{x}) - 2\varepsilon \tag{7.163}$$

for all $\omega \in \Omega \setminus \Upsilon$. Choose now a sequence of positive numbers $\varepsilon_i \downarrow 0$ and let Υ_i be the corresponding sets of measure zero. Set $\Delta := \bigcup_{i=1}^{\infty} \tilde{\Upsilon}_i$. Then (7.163) implies that (7.154) holds for any $\bar{x} \in \mathbb{R}^n$ and any sequence x_N converging to \bar{x} . This proves that condition (7.154) holds for a.e. $\omega \in \Omega$.

In order to verify the respective condition (7.97) of Definition 7.29 we need the following result. There exists a countable set $\mathcal{D} \subset \mathbb{R}^n$ such that for any point $\bar{x} \in \mathbb{R}^n$ there exists a sequence $x_k \in \mathcal{D}$ converging to \bar{x} and

$$\limsup_{k \to \infty} \phi(x_k) \le \phi(\bar{x}). \tag{7.164}$$

Such set \mathcal{D} can be constructed as follows. Let $A \subset \mathbb{R}$ be a countable and dense subset of \mathbb{R} . Consider level sets $\mathcal{L}_a := \{x \in \mathbb{R}^n : \phi(x) \leq a\}$, and let $D_a, a \in A$, be a countable and dense subset of \mathcal{L}_a . (Of course, some of the sets \mathcal{L}_a and D_a can be empty.) Define $\mathcal{D} := \bigcup_{a \in A} D_a$. Clearly the set \mathcal{D} is countable. The condition (7.164) also holds. Indeed, let $a_k \in A$ be a monotonically decreasing sequence converging to $\phi(\bar{x})$. Note that $\bar{x} \in \mathcal{L}_{a_k}$ for all k. Therefore there exists a point $x_k \in D_{a_k}$ such that $||x_k - \bar{x}|| \leq 1/k$. We have then that $x_k \to \bar{x}$ and $\phi(x_k) \leq a_k$, and hence (7.164) follows.

Let \mathcal{D} be such a set. Consider a point $\bar{x} \in \mathbb{R}^n$ and let $x_k \in \mathcal{D}$ be a sequence of points converging to \bar{x} such that (7.164) holds. For a given $x \in \mathbb{R}^n$ we have by Theorem 7.58 that $\hat{\phi}_N(x)$ converges to $\phi(x)$ w.p.1. That is, there exists a set Υ_x of measure zero such that $\hat{\phi}_N(x,\omega)$ converges to $\phi(x)$ for every $\omega \in \Omega \setminus \Upsilon_x$. Consider the set $\tilde{\Upsilon} := \bigcup_{x \in \mathcal{D}} \Upsilon_x$. Since the set \mathcal{D} is countable, the set $\tilde{\Upsilon}$ has measure zero. We have that $\hat{\phi}_N(x,\omega)$ converges to $\phi(x)$ for every $x \in \mathcal{D}$ and $\omega \in \Omega \setminus \Upsilon$. Hence there is a sequence $N_k = N_k(\omega)$ of positive integers such that for all k,

$$|\hat{\phi}_{N_k}(x_k,\omega) - \phi(x_k)| < 1/k, \quad \forall \omega \in \Omega \setminus \hat{\Upsilon}.$$
 (7.165)

Now let x'_N be a sequence of points such that $x'_{N_k} = x_k$ for all k. We have then that $x'_{N_k} \to \bar{x}$ and by (7.164) and (7.165) that

$$\limsup_{k \to \infty} \hat{\phi}_{N_k}(x'_{N_k}, \omega) \le \phi(\bar{x}), \quad \forall \omega \in \Omega \setminus \tilde{\Upsilon}.$$
(7.166)

This shows that condition (7.97) holds w.p.1.

7.2.7 Law of Large Numbers for Random Sets and Subdifferentials

Consider a measurable multifunction $\mathcal{A} : \Omega \Rightarrow \mathbb{R}^n$. Assume that \mathcal{A} is compact valued, i.e., $\mathcal{A}(\omega)$ is a nonempty compact subset of \mathbb{R}^n for every $\omega \in \Omega$. Let us denote by \mathfrak{C}_n the space of nonempty compact subsets of \mathbb{R}^n . Equipped with the Hausdorff distance between two sets $A, B \in \mathfrak{C}_n$, the space \mathfrak{C}_n becomes a metric space. We equip \mathfrak{C}_n with the sigma algebra \mathcal{B} of its Borel subsets (generated by the family of closed subsets of \mathfrak{C}_n). This makes $(\mathfrak{C}_n, \mathcal{B})$ a sample (measurable) space. Of course, we can view the multifunction \mathcal{A} as a mapping from Ω into \mathfrak{C}_n . We have that the multifunction $\mathcal{A} : \Omega \Rightarrow \mathbb{R}^n$ is measurable iff the corresponding mapping $\mathcal{A} : \Omega \to \mathfrak{C}_n$ is measurable.

We say $A_i : \Omega \to \mathfrak{C}_n$, $i \in \mathbb{N}$, is an iid sequence of realizations of \mathcal{A} if each $A_i = A_i(\omega)$ has the same probability distribution on $(\mathfrak{C}_n, \mathcal{B})$ as $\mathcal{A}(\omega)$, and A_i , $i \in \mathbb{N}$, are independent. We have the following (strong) LLN for an iid sequence of random sets.

Theorem 7.61 (Artstein-Vitale). Let A_i , $i \in \mathbb{N}$, be an iid sequence of realizations of a measurable mapping $\mathcal{A} : \Omega \to \mathfrak{C}_n$ such that $\mathbb{E}[||\mathcal{A}(\omega)||] < \infty$. Then

$$N^{-1}(A_1 + \dots + A_N) \to \mathbb{E}[\operatorname{conv}(\mathcal{A})] \quad \text{w.p.1 as } N \to \infty,$$
 (7.167)

where the convergence is understood in the sense of the Hausdorff metric.

In order to understand the above result let us make the following observations. There is a one-to-one correspondence between *convex* sets $A \in \mathfrak{C}_n$ and finite valued convex positively homogeneous functions on \mathbb{R}^n , defined by $A \mapsto \mathsf{s}_A$, where $\mathsf{s}_A(h) := \sup_{z \in A} z^\mathsf{T} h$ is the support function of A. Note that for any two convex sets $A, B \in \mathfrak{C}_n$ we have that $\mathsf{s}_{A+B}(\cdot) = \mathsf{s}_A(\cdot) + \mathsf{s}_B(\cdot)$, and $A \subset B$ iff $\mathsf{s}_A(\cdot) \leq \mathsf{s}_B(\cdot)$. Consequently, for convex sets $A_1, A_2 \in \mathfrak{C}_n$ and $B_r := \{x : ||x|| \leq r\}, r \geq 0$, we have

$$\mathbb{D}(A_1, A_2) = \inf \left\{ r \ge 0 : A_1 \subset A_2 + B_r \right\}$$
(7.168)

and

$$\inf \left\{ r \ge 0 : A_1 \subset A_2 + B_r \right\} = \inf \left\{ r \ge 0 : \mathbf{S}_{A_1}(\cdot) \le \mathbf{S}_{A_2}(\cdot) + \mathbf{S}_{B_r}(\cdot) \right\}.$$
(7.169)

Moreover, $S_{B_r}(h) = \sup_{\|z\| \le r} z^T h = r \|h\|^*$, where $\|\cdot\|^*$ is the dual of the norm $\|\cdot\|$. We obtain that

$$\mathbb{H}(A_1, A_2) = \sup_{\|h\|^* \le 1} |\mathbf{s}_{A_1}(h) - \mathbf{s}_{A_2}(h)|.$$
(7.170)

It follows that if the multifunction $\mathcal{A}(\omega)$ is compact and *convex* valued, then the convergence assertion (7.167) is equivalent to:

$$\sup_{\|h\|^* \le 1} \left| N^{-1} \sum_{i=1}^N \mathsf{s}_{A_i}(h) - \mathbb{E}\bigl[\mathsf{s}_{\mathcal{A}}(h)\bigr] \right| \to 0 \quad \text{w.p.1 as } N \to \infty.$$
(7.171)

Therefore, for compact and convex valued multifunction $\mathcal{A}(\omega)$, Theorem 7.61 is a direct consequence of Theorem 7.55. For general compact valued multifunctions, the averaging operation (in the left hand side of (7.167)) makes a "convexifation" of the limiting set.

Consider now a random lsc convex function $F : \mathbb{R}^n \times \Xi \to \overline{\mathbb{R}}$ and the corresponding sample average function $\hat{f}_N(x)$ based on an iid sequence $\xi^i = \xi^i(\omega), i \in \mathbb{N}$ (see (7.137)). Recall that for any $x \in \mathbb{R}^n$, the multifunction $\xi \mapsto \partial F(x,\xi)$ is measurable (see Remark 55 on page 455). In a sense the following result can be viewed as a particular case of Theorem 7.61 for compact convex valued multifunctions.

Theorem 7.62. Let $F : \mathbb{R}^n \times \Xi \to \overline{\mathbb{R}}$ be a random lsc convex function and $\hat{f}_N(x)$ be the corresponding sample average functions based on an iid sequence ξ^i . Suppose that the expectation function f(x) is well defined and finite valued in a neighborhood of a point $\overline{x} \in \mathbb{R}^n$. Then

$$\mathbb{H}(\partial \hat{f}_N(\bar{x}), \partial f(\bar{x})) \to 0 \text{ w.p.1 as } N \to \infty.$$
(7.172)

Proof. By Theorem 7.51 we have that f(x) is directionally differentiable at \bar{x} and

$$f'(\bar{x},h) = \mathbb{E} \left| F'_{\xi}(\bar{x},h) \right|. \tag{7.173}$$

Note that since $f(\cdot)$ is finite valued near \bar{x} , the directional derivative $f'(\bar{x}, \cdot)$ is finite valued as well. We also have that

$$\hat{f}'_N(\bar{x},h) = N^{-1} \sum_{i=1}^N F'_{\xi^i}(\bar{x},h).$$
(7.174)

Therefore, by the LLN it follows that $\hat{f}'_N(\bar{x},\cdot)$ converges to $f'(\bar{x},\cdot)$ pointwise w.p.1 as $N \to \infty$. Consequently, by Theorem 7.55 we obtain that $\hat{f}'_N(\bar{x},\cdot)$ converges to $f'(\bar{x},\cdot)$ w.p.1 uniformly on the set $\{h : \|h\|^* \leq 1\}$. Since $\hat{f}'_N(\bar{x},\cdot)$ is the support function of the set $\partial \hat{f}_N(\bar{x})$, it follows by (7.170) that $\partial \hat{f}_N(\bar{x})$ converges (in the Hausdorff metric) w.p.1 to $\mathbb{E}[\partial F(\bar{x},\xi)]$. It remains to note that by Theorem 7.52 we have $\mathbb{E}[\partial F(\bar{x},\xi)] = \partial f(\bar{x})$. \Box

The problem in trying to extend the pointwise convergence (7.172) to a uniform type of convergence is that the multifunction $x \mapsto \partial f(x)$ is not continuous even if f(x) is convex real valued.¹⁸

¹⁸This multifunction is upper semicontinuous in the sense that if the function $f(\cdot)$ is convex and continuous at \bar{x} , then $\lim_{x\to\bar{x}} \mathbb{D}(\partial f(x), \partial f(\bar{x})) = 0$.

Let us consider now the ε -subdifferential, $\varepsilon \ge 0$, of a convex real valued function $f: \mathbb{R}^n \to \mathbb{R}$, defined as

$$\partial_{\varepsilon} f(\bar{x}) := \left\{ z \in \mathbb{R}^n : f(x) - f(\bar{x}) \ge z^{\mathsf{T}}(x - \bar{x}) - \varepsilon, \ \forall x \in \mathbb{R}^n \right\}.$$
(7.175)

Clearly for $\varepsilon = 0$, the ε -subdifferential coincides with the usual subdifferential (at the respective point). It is possible to show that for $\varepsilon > 0$ the multifunction $x \mapsto \partial_{\varepsilon} f(x)$ is continuous (in the Hausdorff metric) on \mathbb{R}^n .

Theorem 7.63. Let $g_k : \mathbb{R}^n \to \mathbb{R}$, $k \in \mathbb{N}$, be a sequence of convex real valued (deterministic) functions. Suppose that for every $x \in \mathbb{R}^n$ the sequence $g_k(x)$, $k \in \mathbb{N}$, converges to a finite limit g(x), i.e., functions $g_k(\cdot)$ converge pointwise to the function $g(\cdot)$. Then the function g(x) is convex, and for any $\varepsilon > 0$ the ε -subdifferentials $\partial_{\varepsilon}g_k(\cdot)$ converge uniformly to $\partial_{\varepsilon}g(\cdot)$ on any nonempty compact set $X \subset \mathbb{R}^n$, i.e.,

$$\lim_{k \to \infty} \sup_{x \in X} \mathbb{H} \left(\partial_{\varepsilon} g_k(x), \partial_{\varepsilon} g(x) \right) = 0.$$
(7.176)

Proof. Convexity of $g(\cdot)$ means that

$$g(tx_1 + (1-t)x_2) \le tg(x_1) + (1-t)g(x_2), \ \forall x_1, x_2 \in \mathbb{R}^n, \ \forall t \in [0,1].$$

This follows from convexity of functions $g_k(\cdot)$ by passing to the limit.

By continuity and compactness arguments we have that in order to prove (7.176) it suffices to show that if x_k is a sequence of points converging to a point \bar{x} , then the Hausdorff distance $\mathbb{H}(\partial_{\varepsilon}g_k(x_k), \partial_{\varepsilon}g(\bar{x}))$ tends to zero as $k \to \infty$. Consider the ε -directional derivative of g at x:

$$g'_{\varepsilon}(x,h) := \inf_{t>0} \frac{g(x+th) - g(x) + \varepsilon}{t}.$$
(7.177)

It is known that $g'_{\varepsilon}(x, \cdot)$ is the support function of the set $\partial_{\varepsilon}g(x)$. Therefore, since convergence of a sequence of nonempty convex compact sets in the Hausdorff metric is equivalent to the pointwise convergence of the corresponding support functions, it suffices to show that for any given $h \in \mathbb{R}^n$:

$$\lim_{k \to \infty} g'_{k\varepsilon}(x_k, h) = g'_{\varepsilon}(\bar{x}, h).$$

Let us fix t > 0. Then

$$\limsup_{k \to \infty} g'_{k\varepsilon}(x_k, h) \le \limsup_{k \to \infty} \frac{g_k(x_k + th) - g_k(x_k) + \varepsilon}{t} = \frac{g(\bar{x} + th) - g(\bar{x}) + \varepsilon}{t}.$$

Since t > 0 was arbitrary this implies that

$$\limsup_{k \to \infty} g'_{k\varepsilon}(x_k, h) \le g'_{\varepsilon}(\bar{x}, h).$$

Now let us suppose for a moment that the minimum of $t^{-1}[g(\bar{x}+th)-g(\bar{x})+\varepsilon]$, over t > 0, is attained on a bounded set $T_{\varepsilon} \subset \mathbb{R}_+$. It follows then by convexity that for k large enough, $t^{-1}[g_k(x_k + th) - g_k(x_k) + \varepsilon]$ attains its minimum over t > 0, say at a point t_k , and $dist(t_k, T_{\varepsilon}) \to 0$. Note that $\inf T_{\varepsilon} > 0$. Consequently

$$\liminf_{k \to \infty} g'_{k\varepsilon}(x_k, h) = \liminf_{k \to \infty} \frac{g_k(x_k + t_k h) - g_k(x_k) + \varepsilon}{t_k} \ge g'_{\varepsilon}(\bar{x}, h).$$

In the general case the proof can be completed by adding the term $\alpha ||x - \bar{x}||^2$, $\alpha > 0$, to the functions $g_k(x)$ and g(x) and passing to the limit $\alpha \downarrow 0$. \Box

The above result is deterministic. It can be easily translated into the stochastic framework as follows.

Theorem 7.64. Suppose that the random function $F(x,\xi)$ is convex, and for every $x \in \mathbb{R}^n$ the expectation f(x) is well defined and finite and the sample average $\hat{f}_N(x)$ converges to f(x) w.p.1. Then for any $\varepsilon > 0$ the ε -subdifferentials $\partial_{\varepsilon} \hat{f}_N(x)$ converge uniformly to $\partial_{\varepsilon} f(x)$ w.p.1 on any nonempty compact set $X \subset \mathbb{R}^n$, i.e.,

$$\sup_{x \in X} \mathbb{H}\left(\partial_{\varepsilon} \hat{f}_N(x), \partial_{\varepsilon} f(x)\right) \to 0 \quad \text{w.p.1 as } N \to \infty.$$
(7.178)

Proof. In a way similar to the proof of Theorem 7.55 it can be shown that for a.e. $\omega \in \Omega$, $\hat{f}_N(x)$ converges pointwise to f(x) on a countable and dense subset of \mathbb{R}^n . By the convexity arguments it follows that w.p.1, $\hat{f}_N(x)$ converges pointwise to f(x) on \mathbb{R}^n (see Theorem 7.31), and hence the proof can be completed by applying Theorem 7.63. \Box

Note that the assumption that the expectation function $f(\cdot)$ is finite valued on \mathbb{R}^n implies that $F(\cdot,\xi)$ is finite valued for a.e. ξ , and since $F(\cdot,\xi)$ is convex it follows that $F(\cdot,\xi)$ is continuous. Consequently, it follows that $F(x,\xi)$ is a Carathéodory function, and hence is random lower semicontinuous. Note also that the equality $\partial_{\varepsilon} \hat{f}_N(x) =$ $N^{-1} \sum_{i=1}^N \partial_{\varepsilon} F(x,\xi^i)$ holds for $\varepsilon = 0$ (by the Moreau-Rockafellar Theorem), but does *not* hold for $\varepsilon > 0$ and N > 1.

7.2.8 Delta Method

In this section we discuss the so-called Delta method approach to asymptotic analysis of stochastic problems. Let Z_k , $k \in \mathbb{N}$, be a sequence of random variables converging in distribution to a random variable Z, denoted $Z_k \xrightarrow{\mathcal{D}} Z$.

Remark 57. It can be noted that convergence in distribution does not imply convergence of the expected values $\mathbb{E}[Z_k]$ to $\mathbb{E}[Z]$, as $k \to \infty$, even if all these expected values are finite. This implication holds under the additional condition that Z_k are *uniformly integrable*, that is

$$\lim_{c \to \infty} \sup_{k \in \mathbb{N}} \mathbb{E}\left[Z_k(c)\right] = 0, \tag{7.179}$$

where $Z_k(c) := |Z_k|$ if $|Z_k| \ge c$, and $Z_k(c) := 0$ otherwise. A simple sufficient condition ensuring uniform integrability, and hence the implication that $Z_k \xrightarrow{\mathcal{D}} Z$ implies $\mathbb{E}[Z_k] \to$ $\mathbb{E}[Z]$, is the following: there exists $\varepsilon > 0$ such that $\sup_{k \in \mathbb{N}} \mathbb{E}\left[|Z_k|^{1+\varepsilon}\right] < \infty$. Indeed, for c > 0 we have

$$\mathbb{E}\left[Z_k(c)\right] \le c^{-\varepsilon} \mathbb{E}\left[Z_k(c)^{1+\varepsilon}\right] \le c^{-\varepsilon} \mathbb{E}\left[|Z_k|^{1+\varepsilon}\right],$$

from which the assertion follows.

Remark 58 (stochastic order notation). The notation $O_p(\cdot)$ and $o_p(\cdot)$ stands for a probabilistic analogue of the usual order notation $O(\cdot)$ and $o(\cdot)$, respectively. That is, let X_k and Z_k be sequences of random variables. It is written that $Z_k = O_p(X_k)$ if for any $\varepsilon > 0$ there exists c > 0 such that $\Pr(|Z_k/X_k| > c) \le \varepsilon$ for all $k \in \mathbb{N}$. It is written that $Z_k = o_p(X_k)$ if for any $\varepsilon > 0$ it holds that $\lim_{k\to\infty} \Pr(|Z_k/X_k| > \varepsilon) = 0$. Usually this is used with the sequence X_k being deterministic. In particular, the notation $Z_k = O_p(1)$ asserts that the sequence Z_k is bounded in probability, and $Z_k = o_p(1)$ means that the sequence Z_k converges in probability to zero.

First Order Delta Method

In order to investigate asymptotic properties of sample estimators it will be convenient to use the Delta method, which we are going to discuss now. Let $Y_N \in \mathbb{R}^d$ be a sequence of random vectors, converging in probability to a vector $\mu \in \mathbb{R}^d$. Suppose that there exists a sequence τ_N of positive numbers, tending to infinity, such that $\tau_N(Y_N - \mu)$ converges in distribution to a random vector Y, i.e., $\tau_N(Y_N - \mu) \xrightarrow{\mathcal{D}} Y$. Let $G : \mathbb{R}^d \to \mathbb{R}^m$ be a vector valued function, differentiable at μ . That is

$$G(y) - G(\mu) = J(y - \mu) + r(y), \qquad (7.180)$$

where $J := \nabla G(\mu)$ is the $m \times d$ Jacobian matrix of G at μ , and the remainder r(y) is of order $o(||y - \mu||)$, i.e., $r(y)/||y - \mu|| \to 0$ as $y \to \mu$. It follows from (7.180) that

$$\tau_N \left[G(Y_N) - G(\mu) \right] = J \left[\tau_N (Y_N - \mu) \right] + \tau_N r(Y_N).$$
(7.181)

Since $\tau_N(Y_N - \mu)$ converges in distribution, it is bounded in probability, and hence $||Y_N - \mu||$ is of stochastic order $O_p(\tau_N^{-1})$. It follows that

$$r(Y_N) = o(||Y_N - \mu||) = o_p(\tau_N^{-1}),$$

and hence $\tau_N r(Y_N)$ converges in probability to zero. Consequently we obtain by (7.181) that

$$\tau_N \left[G(Y_N) - G(\mu) \right] \xrightarrow{\mathcal{D}} JY. \tag{7.182}$$

This formula is routinely employed in multivariate analysis and is known as the (finite dimensional) Delta Theorem. In particular, suppose that $N^{1/2}(Y_N - \mu)$ converges in distribution to a (multivariate) normal distribution with zero mean vector and covariance matrix Σ , written $N^{1/2}(Y_N - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma)$. Often, this can be ensured by an application of the Central Limit Theorem. Then it follows by (7.182) that

$$N^{1/2}\left[G(Y_N) - G(\mu)\right] \xrightarrow{\mathcal{D}} \mathcal{N}(0, J\Sigma J^{\mathsf{T}}).$$
(7.183)

We need to extend this method in several directions. The random functions $f_N(\cdot)$ can be viewed as random elements in an appropriate functional space. This motivates us to extend formula (7.182) to a Banach space setting. Let B_1 and B_2 be two Banach spaces, and $G : B_1 \rightarrow B_2$ be a mapping. Suppose that G is directionally differentiable at a considered point $\mu \in B_1$, i.e., the limit

$$G'_{\mu}(d) := \lim_{t \downarrow 0} \frac{G(\mu + td) - G(\mu)}{t}$$
(7.184)

exists for all $d \in B_1$. If, in addition, the directional derivative $G'_{\mu}: B_1 \to B_2$ is linear and continuous, then it is said that G is Gâteaux differentiable at μ . Note that, in any case, the directional derivative $G'_{\mu}(\cdot)$ is positively homogeneous, that is $G'_{\mu}(\alpha d) = \alpha G'_{\mu}(d)$ for any $\alpha \ge 0$ and $d \in B_1$.

It follows from (7.184) that

$$G(\mu + d) - G(\mu) = G'_{\mu}(d) + r(d)$$

with the remainder r(d) being "small" along any fixed direction d, i.e., $r(td)/t \to 0$ as $t \downarrow 0$. This property is not sufficient, however, to neglect the remainder term in the corresponding asymptotic expansion and we need a stronger notion of directional differentiability. It is said that G is directionally differentiable at μ in the sense of Hadamard if the directional derivative $G'_{\mu}(d)$ exists for all $d \in B_1$ and, moreover,

$$G'_{\mu}(d) = \lim_{\substack{t \downarrow 0 \\ d' \to d}} \frac{G(\mu + td') - G(\mu)}{t}.$$
(7.185)

Proposition 7.65. Let B_1 and B_2 be Banach spaces, $G : B_1 \to B_2$ and $\mu \in B_1$. Then the following holds. (i) If $G(\cdot)$ is Hadamard directionally differentiable at μ , then the directional derivative $G'_{\mu}(\cdot)$ is continuous. (ii) If $G(\cdot)$ is Lipschitz continuous in a neighborhood of μ and directionally differentiable at μ , then $G(\cdot)$ is Hadamard directionally differentiable at μ .

The above properties can a be proved in a way similar to the proof of Theorem 7.2. We also have the following *chain rule*.

Proposition 7.66 (Chain Rule). Let B_1 , B_2 and B_3 be Banach spaces, and $G : B_1 \to B_2$ and $F : B_2 \to B_3$ be mappings. Suppose that G is directionally differentiable at a point $\mu \in B_1$ and F is Hadamard directionally differentiable at $\eta := G(\mu)$. Then the composite mapping $F \circ G : B_1 \to B_3$ is directionally differentiable at μ and

$$(F \circ G)'(\mu, d) = F'(\eta, G'(\mu, d)), \quad \forall d \in B_1.$$
 (7.186)

Proof. Since G is directionally differentiable at μ , we have for $t \ge 0$ and $d \in B_1$ that

$$G(\mu + td) = G(\mu) + tG'(\mu, d) + o(t).$$

Since F is Hadamard directionally differentiable at $\eta := G(\mu)$, it follows that

$$F(G(\mu + td)) = F(G(\mu) + tG'(\mu, d) + o(t)) = F(\eta) + tF'(\eta, G'(\mu, d)) + o(t).$$

This implies that $F \circ G$ is directionally differentiable at μ and formula (7.186) holds.

Now let B_1 and B_2 be equipped with their Borel σ -algebras \mathcal{B}_1 and \mathcal{B}_2 , respectively. An \mathcal{F} -measurable mapping from a probability space (Ω, \mathcal{F}, P) into B_1 is called a random element of B_1 . Consider a sequence X_N of random elements of B_1 . It is said that X_N converges in distribution (weakly) to a random element Y of B_1 , and denoted $X_N \xrightarrow{\mathcal{D}} Y$, if the expected values $\mathbb{E}[f(X_N)]$ converge to $\mathbb{E}[f(Y)]$, as $N \to \infty$, for any bounded and continuous function $f : B_1 \to \mathbb{R}$. Let us formulate now the first version of the Delta Theorem. Recall that a Banach space is said to be *separable* if it has a countable dense subset.

Theorem 7.67 (Delta Theorem). Let B_1 and B_2 be Banach spaces, equipped with their Borel σ -algebras, Y_N be a sequence of random elements of B_1 , $G : B_1 \to B_2$ be a mapping, and τ_N be a sequence of positive numbers tending to infinity as $N \to \infty$. Suppose that the space B_1 is separable, the mapping G is Hadamard directionally differentiable at a point $\mu \in B_1$, and the sequence $X_N := \tau_N(Y_N - \mu)$ converges in distribution to a random element Y of B_1 . Then

$$\tau_N \left[G(Y_N) - G(\mu) \right] \xrightarrow{\mathcal{D}} G'_{\mu}(Y), \tag{7.187}$$

and

$$\tau_N \left[G(Y_N) - G(\mu) \right] = G'_\mu(X_N) + o_p(1). \tag{7.188}$$

Note that, because of the Hadamard directional differentiability of G, the mapping $G'_{\mu}: B_1 \to B_2$ is continuous, and hence is measurable with respect to the Borel σ -algebras of B_1 and B_2 . The above infinite dimensional version of the Delta Theorem can be proved easily by using the following Skorohod-Dudley almost sure representation theorem.

Theorem 7.68 (Representation Theorem). Suppose that a sequence of random elements X_N , of a separable Banach space B, converges in distribution to a random element Y. Then there exists a sequence X'_N , Y', defined on a single probability space, such that $X'_N \stackrel{\mathcal{D}}{\sim} X_N$, for all $N, Y' \stackrel{\mathcal{D}}{\sim} Y$ and $X'_N \to Y'$ with probability one.

Here $Y' \stackrel{\mathcal{D}}{\sim} Y$ means that the probability measures induced by Y' and Y coincide.

Proof. [of Theorem 7.67] Consider the sequence $X_N := \tau_N(Y_N - \mu)$ of random elements of B_1 . By the Representation Theorem, there exists a sequence X'_N, Y' , defined on a single probability space, such that $X'_N \stackrel{\mathcal{D}}{\sim} X_N, Y' \stackrel{\mathcal{D}}{\sim} Y$ and $X'_N \to Y'$ w.p.1. Consequently for $Y'_N := \mu + \tau_N^{-1}X'_N$, we have $Y'_N \stackrel{\mathcal{D}}{\sim} Y_N$. It follows then from Hadamard directional differentiability of G that

$$\tau_N [G(Y'_N) - G(\mu)] \to G'_\mu(Y')$$
 w.p.1. (7.189)

Since convergence with probability one implies convergence in distribution and the terms in (7.189) have the same distributions as the corresponding terms in (7.187), the asymptotic result (7.187) follows.

Now since $G'_{\mu}(\cdot)$ is continuous and $X'_{N} \to Y'$ w.p.1, we have that

$$G'_{\mu}(X'_N) \to G'_{\mu}(Y')$$
 w.p.1. (7.190)

Together with (7.189) this implies that the difference between $G'_{\mu}(X'_N)$ and the left hand side of (7.189) tends w.p.1, and hence in probability, to zero. We obtain that

$$\tau_N \left[G(Y'_N) - G(\mu) \right] = G'_\mu \left[\tau_N (Y'_N - \mu) \right] + o_p(1),$$

which implies (7.188).

Let us now formulate the second version of the Delta Theorem where the mapping G is restricted to a subset K of the space B_1 . We say that G is Hadamard directionally differentiable at a point μ tangentially to the set K if for any sequence d_N of the form $d_N := (y_N - \mu)/t_N$, where $y_N \in K$ and $t_N \downarrow 0$, and such that $d_N \to d$, the following limit exists

$$G'_{\mu}(d) = \lim_{N \to \infty} \frac{G(\mu + t_N d_N) - G(\mu)}{t_N}.$$
(7.191)

Equivalently the above condition (7.191) can be written in the form

$$G'_{\mu}(d) = \lim_{\substack{t \downarrow 0 \\ d' \to_{K} d}} \frac{G(\mu + td') - G(\mu)}{t},$$
(7.192)

where the notation $d' \rightarrow_{K} d$ means that $d' \rightarrow d$ and $\mu + td' \in K$.

Since $y_N \in K$, and hence $\mu + t_N d_N \in K$, the mapping G needs only to be defined on the set K. Recall that the *contingent (Bouligand) cone* to K at μ , denoted $\mathcal{T}_K(\mu)$, is formed by vectors $d \in B$ such that there exist sequences $d_N \to d$ and $t_N \downarrow 0$ such that $\mu + t_N d_N \in K$. Note that $\mathcal{T}_K(\mu)$ is nonempty only if μ belongs to the topological closure of the set K. If the set K is convex, then the contingent cone $\mathcal{T}_K(\mu)$ coincides with the corresponding tangent cone. By the above definitions we have that $G'_{\mu}(\cdot)$ is defined on the set $\mathcal{T}_K(\mu)$. The following "tangential" version of the Delta Theorem can be easily proved in a way similar to the proof of Theorem 7.67.

Theorem 7.69 (Delta Theorem). Let B_1 and B_2 be Banach spaces, K be a subset of B_1 , $G : K \to B_2$ be a mapping, and Y_N be a sequence of random elements of B_1 . Suppose that: (i) the space B_1 is separable, (ii) the mapping G is Hadamard directionally differentiable at a point μ tangentially to the set K, (iii) for some sequence τ_N of positive numbers tending to infinity, the sequence $X_N := \tau_N(Y_N - \mu)$ converges in distribution to a random element Y, (iv) $Y_N \in K$, with probability one, for all N large enough. Then

$$\tau_N \left[G(Y_N) - G(\mu) \right] \xrightarrow{\mathcal{D}} G'_{\mu}(Y). \tag{7.193}$$

Moreover, if the set K is convex, then equation (7.188) holds.

Note that it follows from the assumptions (iii) and (iv) that the distribution of Y is concentrated on the contingent cone $\mathcal{T}_K(\mu)$, and hence the distribution of $G'_{\mu}(Y)$ is well defined.

Second Order Delta Theorem

Our third variant of the Delta Theorem deals with a second order expansion of the mapping G. That is, suppose that G is directionally differentiable at μ and define

$$G_{\mu}^{\prime\prime}(d) := \lim_{\substack{t\downarrow 0\\ d'\to d}} \frac{G(\mu + td') - G(\mu) - tG_{\mu}^{\prime}(d')}{\frac{1}{2}t^2}.$$
(7.194)

If the mapping G is twice continuously differentiable, then this second order directional derivative $G''_{\mu}(d)$ coincides with the second order term in the Taylor expansion of $G(\mu + d)$. The above definition of $G''_{\mu}(d)$ makes sense for directionally differentiable mappings. However, in interesting applications, where it is possible to calculate $G''_{\mu}(d)$, the mapping G is actually (Gâteaux) differentiable. We say that G is second order Hadamard directionally differentiable at μ if the second order directional derivative $G''_{\mu}(d)$, defined in (7.194), exists for all $d \in B_1$. We say that G is second order Hadamard directionally differentiable at μ tangentially to a set $K \subset B_1$ if for all $d \in \mathcal{T}_K(\mu)$ the limit

$$G_{\mu}^{\prime\prime}(d) = \lim_{\substack{t \downarrow 0 \\ d' \to \kappa^d}} \frac{G(\mu + td') - G(\mu) - tG_{\mu}^{\prime}(d')}{\frac{1}{2}t^2}$$
(7.195)

exists.

Note that if G is first and second order Hadamard directionally differentiable at μ tangentially to K, then $G'_{\mu}(\cdot)$ and $G''_{\mu}(\cdot)$ are continuous on $\mathcal{T}_{K}(\mu)$, and that $G''_{\mu}(\alpha d) = \alpha^{2}G''_{\mu}(d)$ for any $\alpha \geq 0$ and $d \in \mathcal{T}_{K}(\mu)$.

Theorem 7.70 (second-order Delta Theorem). Let B_1 and B_2 be Banach spaces, K be a convex subset of B_1 , Y_N be a sequence of random elements of B_1 , $G : K \to B_2$ be a mapping, and τ_N be a sequence of positive numbers tending to infinity as $N \to \infty$. Suppose that: (i) the space B_1 is separable, (ii) G is first and second order Hadamard directionally differentiable at μ tangentially to the set K, (iii) the sequence $X_N := \tau_N(Y_N - \mu)$ converges in distribution to a random element Y of B_1 , (iv) $Y_N \in K$ w.p.1 for N large enough. Then

$$\tau_N^2 \left[G(Y_N) - G(\mu) - G'_{\mu}(Y_N - \mu) \right] \xrightarrow{\mathcal{D}} \frac{1}{2} G''_{\mu}(Y),$$
 (7.196)

and

$$G(Y_N) = G(\mu) + G'_{\mu}(Y_N - \mu) + \frac{1}{2}G''_{\mu}(Y_N - \mu) + o_p(\tau_N^{-2}).$$
(7.197)

Proof. Let X'_N , Y' and Y'_N be elements as in the proof of Theorem 7.67. Recall that their existence is guaranteed by the Representation Theorem. Then by the definition of G''_{μ} we have

$$au_N^2 \left[G(Y'_N) - G(\mu) - au_N^{-1} G'_\mu(X'_N) \right] \to \frac{1}{2} G''_\mu(Y') \quad \text{w.p.1.}$$

Note that $G'_{\mu}(\cdot)$ is defined on $\mathcal{T}_{K}(\mu)$ and, since K is convex, $X'_{N} = \tau_{N}(Y'_{N} - \mu) \in \mathcal{T}_{K}(\mu)$. Therefore the expression in the left hand side of the above limit is well defined. Since convergence w.p.1 implies convergence in distribution, formula (7.196) follows. Since $G''_{\mu}(\cdot)$ is continuous on $\mathcal{T}_{K}(\mu)$, and, by convexity of K, $Y'_{N} - \mu \in \mathcal{T}_{K}(\mu)$ w.p.1, we have that $\tau_{N}^{2}G''_{\mu}(Y'_{N} - \mu) \rightarrow G''_{\mu}(Y')$ w.p.1. Since convergence w.p.1 implies convergence in probability, formula (7.197) then follows. \Box

7.2.9 Exponential Bounds of the Large Deviations Theory

Consider an iid sequence Y_1, \ldots, Y_N of replications of a real valued random variable Y, and let $Z_N := N^{-1} \sum_{i=1}^N Y_i$ be the corresponding sample average. Then for any real numbers a and t > 0 we have that $\Pr(Z_N \ge a) = \Pr(e^{tZ_N} \ge e^{ta})$, and hence, by Chebyshev's inequality

$$\Pr(Z_N \ge a) \le e^{-ta} \mathbb{E}\left[e^{tZ_N}\right] = e^{-ta} [M(t/N)]^N,$$

where $M(t) := \mathbb{E}[e^{tY}]$ is the moment generating function of Y. Suppose that Y has finite mean $\mu := \mathbb{E}[Y]$ and let $a \ge \mu$. By taking the logarithm of both sides of the above inequality, changing variables t' = t/N and minimizing over t' > 0, we obtain

$$\frac{1}{N}\ln\left[\Pr(Z_N \ge a)\right] \le -I(a),\tag{7.198}$$

where

$$I(z) := \sup_{t \in \mathbb{R}} \left\{ tz - \Lambda(t) \right\}$$
(7.199)

is the conjugate of the logarithmic moment generating function $\Lambda(t) := \ln M(t)$. In the LD theory, I(z) is called the *(large deviations) rate* function, and the inequality (7.198) corresponds to the upper bound of Cramér's LD Theorem.

Note that the moment generating function $M(\cdot)$ is convex, positive valued, M(0) = 1and its domain dom M is a subinterval of \mathbb{R} containing zero. It follows by Theorem 7.49 that $M(\cdot)$ is infinitely differentiable at every interior point of its domain. Moreover, if $a := \inf(\operatorname{dom} M)$ is finite, then $M(\cdot)$ is right side continuous at a, and similarly for the $b := \sup(\operatorname{dom} M)$. It follows that $M(\cdot)$, and hence $\Lambda(\cdot)$, are proper lower semicontinuous functions. The logarithmic moment generating function $\Lambda(\cdot)$ is also convex. Indeed, $\operatorname{dom} \Lambda = \operatorname{dom} M$ and at an interior point t of $\operatorname{dom} \Lambda$,

$$\Lambda''(t) = \frac{\mathbb{E}\left[Y^2 e^{tY}\right] \mathbb{E}\left[e^{tY}\right] - \mathbb{E}\left[Y e^{tY}\right]^2}{M(t)^2}.$$
(7.200)

Moreover, the matrix $\begin{bmatrix} Y^2 e^{tY} & Y e^{tY} \\ Y e^{tY} & e^{tY} \end{bmatrix}$ is positive semidefinite, and hence its expectation is also a positive semidefinite matrix. Consequently the determinant of the later matrix is nonnegative, i.e.,

$$\mathbb{E}\left[Y^2 e^{tY}\right] \mathbb{E}\left[e^{tY}\right] - \mathbb{E}\left[Y e^{tY}\right]^2 \ge 0.$$

We obtain that $\Lambda''(\cdot)$ is nonnegative at every point of the interior of dom Λ , and hence $\Lambda(\cdot)$ is convex.

Note that the constraint t > 0 is removed in the above definition of the rate function $I(\cdot)$. This is because of the following. Consider the function $\psi(t) := ta - \Lambda(t)$. The

function $\Lambda(t)$ is convex, and hence $\psi(t)$ is concave. Suppose that the moment generating function $M(\cdot)$ is finite valued at some $\overline{t} > 0$. Then M(t) is finite for all $t \in [0, \overline{t}]$ and right side differentiable at t = 0. Moreover, the right side derivative of M(t) at t = 0 is μ , and hence the right side derivative of $\psi(t)$ at t = 0 is positive if $a > \mu$. Consequently in that case $\psi(t) > \psi(0)$ for all t > 0 small enough, and hence I(a) > 0 and the supremum in (7.199) is not changed if the constraint t > 0 is removed. If $a = \mu$, then the supremum in (7.199) is attained at t = 0 and hence I(a) = 0. In that case the inequality (7.198) trivially holds. Now if $M(t) = +\infty$ for all t > 0, then I(a) = 0 for any $a \ge \mu$ and the inequality (7.198) trivially holds.

For $a \leq \mu$ the upper bound (7.198) takes the form

$$\frac{1}{N}\ln\left[\Pr(Z_N \le a)\right] \le -I(a),\tag{7.201}$$

which of course can be written as

$$\Pr(Z_N \le a) \le e^{-I(a)N} \tag{7.202}$$

The rate function I(z) is convex and has the following properties. Suppose that the random variable Y has finite mean $\mu := \mathbb{E}[Y]$. Then $\Lambda'(0) = \mu$ and hence the maximum in the right hand side of (7.199) is attained at $t^* = 0$. It follows that $I(\mu) = 0$ and

$$I'(\mu) = t^* \mu - \Lambda'(t^*) = -\Lambda(0) = 0,$$

and hence I(z) attains its minimum at $z = \mu$. Suppose, further, that the moment generating function M(t) is finite valued for all t in a neighborhood of t = 0. Then $\Lambda(t)$ is infinitely differentiable at t = 0, and $\Lambda'(0) = \mu$ and $\Lambda''(0) = \sigma^2$, where $\sigma^2 := \operatorname{Var}[Y]$. It follows by the above discussion that in that case I(a) > 0 for any $a \neq \mu$. We also have then that $I'(\mu) = 0$ and $I''(\mu) = \sigma^{-2}$, and hence by Taylor's expansion,

$$I(a) = \frac{(a-\mu)^2}{2\sigma^2} + o(|a-\mu|^2).$$
(7.203)

If Y has normal distribution $N(\mu, \sigma^2)$, then its logarithmic moment generating function is $\Lambda(t) = \mu t + \sigma^2 t^2/2$. In that case

$$I(a) = \frac{(a-\mu)^2}{2\sigma^2}.$$
(7.204)

The constant I(a) in (7.198) gives, in a sense, the best possible exponential rate at which the probability $Pr(Z_N \ge a)$ converges to zero. This follows from the lower bound

$$\liminf_{N \to \infty} \frac{1}{N} \ln \left[\Pr(Z_N \ge a) \right] \ge -I(a) \tag{7.205}$$

of Cramér's LD Theorem, which holds for $a \ge \mu$.

Another, closely related, exponential type inequalities can be derived for bounded random variables.

Proposition 7.71. Let Y be a random variable such that $a \leq Y \leq b$, for some $a, b \in \mathbb{R}$, and $\mathbb{E}[Y] = 0$. Then

$$\mathbb{E}[e^{tY}] \le e^{t^2(b-a)^2/8}, \ \forall t \ge 0.$$
(7.206)

Proof. If Y is identically zero, then (7.206) obviously holds. Therefore we can assume that Y is not identically zero. Since $\mathbb{E}[Y] = 0$, it follows that a < 0 and b > 0.

Any $Y \in [a, b]$ can be represented as convex combination $Y = \tau a + (1 - \tau)b$, where $\tau = (b - Y)/(b - a)$. Since e^y is a convex function, it follows that

$$e^{Y} \le \frac{b-Y}{b-a}e^{a} + \frac{Y-a}{b-a}e^{b}.$$
 (7.207)

Taking expectation from both sides of (7.207) and using $\mathbb{E}[Y] = 0$, we obtain

$$\mathbb{E}\left[e^{Y}\right] \le \frac{b}{b-a}e^{a} - \frac{a}{b-a}e^{b}.$$
(7.208)

The right hand side of (7.207) can be written as $e^{g(u)}$, where u := b - a, $g(x) := -\alpha x + \ln(1 - \alpha + \alpha e^x)$ and $\alpha := -a/(b - a)$. Note that $\alpha > 0$ and $1 - \alpha > 0$.

Let us observe that g(0) = g'(0) = 0 and

$$g''(x) = \frac{\alpha(1-\alpha)}{(1-\alpha)^2 e^{-x} + \alpha^2 e^x + 2\alpha(1-\alpha)}.$$
(7.209)

Moreover, $(1-\alpha)^2 e^{-x} + \alpha^2 e^x \ge 2\alpha(1-\alpha)$, and hence $g''(x) \le 1/4$ for any x. By Taylor expansion of $g(\cdot)$ at zero, we have $g(u) = u^2 g''(\tilde{u})/2$ for some $\tilde{u} \in (0, u)$. It follows that $g(u) \le u^2/8 = (b-a)^2/8$, and hence

$$\mathbb{E}[e^Y] \le e^{(b-a)^2/8}.$$
(7.210)

Finally, (7.206) follows from (7.210) by rescaling Y to tY for $t \ge 0$.

In particular, if $|Y| \le b$ and $\mathbb{E}[Y] = 0$, then $|-Y| \le b$ and $\mathbb{E}[-Y] = 0$ as well, and hence by (7.206) we have

$$\mathbb{E}[e^{tY}] \le e^{t^2 b^2/2}, \quad \forall t \in \mathbb{R}.$$
(7.211)

Let Y be a (real valued) random variable supported on a bounded interval $[a, b] \subset \mathbb{R}$, and $\mu := \mathbb{E}[Y]$. Then it follows from (7.206) that the rate function of $Y - \mu$ satisfies

$$I(z) \ge \sup_{t \in \mathbb{R}} \left\{ tz - t^2 (b-a)^2 / 8 \right\} = 2z^2 / (b-a)^2$$

Together with (7.202) this implies the following. Let $Y_1, ..., Y_N$ be an iid sequence of realizations of Y and Z_N be the corresponding average. Then for $\tau > 0$ it holds that

$$\Pr\left(Z_N \ge \mu + \tau\right) \le e^{-2\tau^2 N/(b-a)^2}.$$
(7.212)

The bound (7.212) is often referred to as *Hoeffding inequality*.

In particular, let $W \sim B(p, n)$ be a random variable having Binomial distribution, i.e., $\Pr(W = k) = \binom{n}{k} p^k (1-p)^{n-k}$, k = 0, ..., n. Recall that W can be represented as $W = Y_1 + ... + Y_n$, where $Y_1, ..., Y_n$ is an iid sequence of Bernoulli random variables with $\Pr(Y_i = 1) = p$ and $\Pr(Y_i = 0) = 1 - p$. It follows from Hoeffding's inequality that for a nonnegative integer $k \le np$,

$$\Pr\left(W \le k\right) \le \exp\left\{-\frac{2(np-k)^2}{n}\right\}.$$
(7.213)

For small p it is possible to improve the above estimate as follows. For $Y \sim Bernoulli(p)$ we have

$$\mathbb{E}[e^{tY}] = pe^t + 1 - p = 1 - p(1 - e^t).$$

By using the inequality $e^{-x} \ge 1 - x$ with $x := p(1 - e^t)$, we obtain

$$\mathbb{E}[e^{tY}] \le \exp[p(e^t - 1)],$$

and hence for z > 0,

$$I(z) := \sup_{t \in \mathbb{R}} \left\{ tz - \ln \mathbb{E}[e^{tY}] \right\} \ge \sup_{t \in \mathbb{R}} \left\{ tz - p(e^t - 1) \right\} = z \ln \frac{z}{p} - z + p.$$

Moreover, since $\ln(1+x) \ge x - x^2/2$ for $x \ge 0$, we obtain

$$I(z) \ge \frac{(z-p)^2}{2p} \text{ for } z \ge p.$$

By (7.198) it follows that

$$\Pr\left(n^{-1}W \ge z\right) \le \exp\left\{-n(z-p)^2/(2p)\right\}, \text{ for } z \ge p.$$
(7.214)

Alternatively, this can be written as

$$\Pr\left(W \le k\right) \le \exp\left\{-\frac{(np-k)^2}{2pn}\right\},\tag{7.215}$$

for a nonnegative integer $k \le np$. The above inequality (7.215) is often called *Chernoff inequality*. For small p it can be significantly better than Hoeffding inequality (7.213).

The above, one dimensional, LD results can be extended to multivariate and even infinite dimensional settings, and also to non iid random sequences. In particular, suppose that Y is a d-dimensional random vector and let $\mu := \mathbb{E}[Y]$ be its mean vector. We can associate with Y its moment generating function M(t), of $t \in \mathbb{R}^d$, and the rate function I(z) defined in the same way as in (7.199) with the supremum taken over $t \in \mathbb{R}^d$ and tzdenoting the standard scalar product of vectors $t, z \in \mathbb{R}^d$. Consider a (Borel) measurable set $A \subset \mathbb{R}^d$. Then, under certain regularity conditions, the following Large Deviations Principle holds:

$$-\inf_{z \in \operatorname{int}(A)} I(z) \leq \liminf_{N \to \infty} N^{-1} \ln [\Pr(Z_N \in A)]$$

$$\leq \limsup_{N \to \infty} N^{-1} \ln [\Pr(Z_N \in A)]$$

$$\leq -\inf_{z \in \operatorname{cl}(A)} I(z),$$
(7.216)

where int(A) and cl(A) denote the interior and topological closure, respectively, of the set A. In the above one dimensional setting, the LD principle (7.216) was derived for sets $A := [a, +\infty)$.

We have that if $\mu \in int(A)$ and the moment generating function M(t) is finite valued for all t in a neighborhood of $0 \in \mathbb{R}^d$, then $\inf_{z \in \mathbb{R}^d \setminus (intA)} I(z)$ is positive. Moreover, if the sequence is iid, then

$$\limsup_{N \to \infty} N^{-1} \ln \left[\Pr(Z_N \notin A) \right] < 0, \tag{7.217}$$

i.e., the probability $\Pr(Z_N \in A) = 1 - \Pr(Z_N \notin A)$ approaches one exponentially fast as N tends to infinity.

Finally, let us derive the following useful result.

Proposition 7.72. Let $\xi^1, \xi^2, ...$ be a sequence of iid random variables (vectors), $\sigma_t > 0$, t = 1, ..., be a sequence of deterministic numbers and $\phi_t = \phi_t(\xi_{[t]})$ be (measurable) functions of $\xi_{[t]} = (\xi^1, ..., \xi^t)$ such that

$$\mathbb{E}\left[\phi_t | \xi_{[t-1]}\right] = 0 \text{ and } \mathbb{E}\left[\exp\{\phi_t^2 / \sigma_t^2\} | \xi_{[t-1]}\right] \le \exp\{1\} \text{ w.p.1.}$$
(7.218)

Then for any $\Theta \ge 0$:

$$\Pr\left\{\sum_{t=1}^{N}\phi_t \ge \Theta\sqrt{\sum_{t=1}^{N}\sigma_t^2}\right\} \le \exp\{-\Theta^2/3\}.$$
(7.219)

Proof. Let us set $\tilde{\phi}_t := \phi_t / \sigma_t$. By condition (7.218) we have that $\mathbb{E}[\tilde{\phi}_t | \xi_{[t-1]}] = 0$ and $\mathbb{E}[\exp{\{\tilde{\phi}_t^2\}} | \xi_{[t-1]}] \le \exp\{1\}$ w.p.1. By Jensen inequality it follows that for any $a \ge 1$:

$$\mathbb{E}\left[\exp\{a\tilde{\phi}_t^2\}|\xi_{[t-1]}\right] = \mathbb{E}\left[(\exp\{\tilde{\phi}_t^2\})^a|\xi_{[t-1]}\right] \le \left(\mathbb{E}\left[\exp\{\tilde{\phi}_t^2\}|\xi_{[t-1]}\right]\right)^a \le \exp\{a\}.$$

We also have that $\exp\{x\} \le x + \exp\{9x^2/16\}$ for all x (this inequality can be verified by direct calculations), and hence for any $\lambda \in [0, 4/3]$:

$$\mathbb{E}\left[\exp\{\lambda\tilde{\phi}_t\}|\xi_{[t-1]}\right] \le \mathbb{E}\left[\exp\{(9\lambda^2/16)\tilde{\phi}_t^2\}|\xi_{[t-1]}\right] \le \exp\{9\lambda^2/16\}.$$
(7.220)

Moreover, we have that $\lambda x \leq \frac{3}{8}\lambda^2 + \frac{2}{3}x^2$ for any λ and x, and hence

$$\mathbb{E}\left[\exp\{\lambda\tilde{\phi}_t\}|\xi_{[t-1]}\right] \le \exp\{3\lambda^2/8\}\mathbb{E}\left[\exp\{2\tilde{\phi}_t^2/3\}|\xi_{[t-1]}\right] \le \exp\{2/3 + 3\lambda^2/8\}.$$

Combining the latter inequality with (7.220), we get

$$\mathbb{E}\left[\exp\{\lambda\tilde{\phi}_t\}|\xi_{[t-1]}\right] \le \exp\{3\lambda^2/4\}, \ \forall \lambda \ge 0.$$

Going back to ϕ_t , the above inequality reads

$$\mathbb{E}\left[\exp\{\gamma\phi_t\}|\xi_{[t-1]}\right] \le \exp\{3\gamma^2\sigma_t^2/4\}, \quad \forall \gamma \ge 0.$$
(7.221)

Now, since ϕ_{τ} is a deterministic function of $\xi_{[\tau]}$ and using (7.221), we obtain for any $\gamma \ge 0$:

$$\mathbb{E}\left[\exp\left\{\gamma\sum_{\tau=1}^{t}\phi_{\tau}\right\}\right] = \mathbb{E}\left[\exp\left\{\gamma\sum_{\tau=1}^{t-1}\phi_{\tau}\right\}\mathbb{E}\left(\exp\{\gamma\phi_{t}\}|\xi_{[t-1]}\right)\right] \\ \leq \exp\{3\gamma^{2}\sigma_{t}^{2}/4\}\mathbb{E}\left[\exp\{\gamma\sum_{\tau=1}^{t-1}\phi_{\tau}\}\right],$$

and hence

$$\mathbb{E}\left[\exp\left\{\gamma\sum_{t=1}^{N}\phi_{t}\right\}\right] \le \exp\left\{3\gamma^{2}\sum_{t=1}^{N}\sigma_{t}^{2}/4\right\}.$$
(7.222)

By Chebyshev's inequality, we have for $\gamma > 0$ and Θ :

$$\Pr\left\{ \sum_{t=1}^{N} \phi_t \ge \Theta \sqrt{\sum_{t=1}^{N} \sigma_t^2} \right\} = \Pr\left\{ \exp\left[\gamma \sum_{t=1}^{N} \phi_t\right] \ge \exp\left[\gamma \Theta \sqrt{\sum_{t=1}^{N} \sigma_t^2}\right] \right\}$$
$$\le \exp\left[-\gamma \Theta \sqrt{\sum_{t=1}^{N} \sigma_t^2}\right] \mathbb{E}\left\{ \exp\left[\gamma \sum_{t=1}^{N} \phi_t\right] \right\}.$$

Together with (7.222) this implies for $\Theta \ge 0$:

$$\Pr\left\{ \sum_{t=1}^{N} \phi_t \ge \Theta \sqrt{\sum_{t=1}^{N} \sigma_t^2} \right\} \le \inf_{\gamma > 0} \exp\left\{ \frac{3}{4} \gamma^2 \sum_{t=1}^{N} \sigma_t^2 - \gamma \Theta \sqrt{\sum_{t=1}^{N} \sigma_t^2} \right\}$$
$$= \exp\left\{ -\Theta^2/3 \right\}.$$

This completes the proof. \Box

7.2.10 Uniform Exponential Bounds

Consider the setting of section 7.2.5 with a sequence ξ^i , $i \in \mathbb{N}$, of random realizations of an d-dimensional random vector $\xi = \xi(\omega)$, a function $F : X \times \Xi \to \mathbb{R}$ and the corresponding sample average function $\hat{f}_N(x)$. We assume here that the sequence ξ^i , $i \in \mathbb{N}$, is iid, the set $X \subset \mathbb{R}^n$ is nonempty and compact and the expectation function $f(x) = \mathbb{E}[F(x,\xi)]$ is well defined and finite valued for all $x \in X$. We discuss now uniform exponential rates of convergence of $\hat{f}_N(x)$ to f(x). Denote by

$$M_x(t) := \mathbb{E}\left[e^{t(F(x,\xi) - f(x))}\right]$$

the moment generating function of the random variable $F(x,\xi) - f(x)$. Let us make the following assumptions.

- (C1) For every $x \in X$ the moment generating function $M_x(t)$ is finite valued for all t in a neighborhood of zero.
- (C2) There exists a (measurable) function $\kappa : \Xi \to \mathbb{R}_+$ such that

$$|F(x',\xi) - F(x,\xi)| \le \kappa(\xi) ||x' - x||$$
(7.223)

for all $\xi \in \Xi$ and all $x', x \in X$.

(C3) The moment generating function $M_{\kappa}(t) := \mathbb{E}\left[e^{t\kappa(\xi)}\right]$ of $\kappa(\xi)$ is finite valued for all t in a neighborhood of zero.

Theorem 7.73. Suppose that conditions (C1)–(C3) hold and the set X is compact. Then for any $\varepsilon > 0$ there exist positive constants C and $\beta = \beta(\varepsilon)$, independent of N, such that

$$\Pr\left\{\sup_{x\in X}\left|\hat{f}_N(x) - f(x)\right| \ge \varepsilon\right\} \le Ce^{-N\beta}.$$
(7.224)

Proof. By the upper bound (7.198) of Cramér's Large Deviation Theorem we have that for any $x \in X$ and $\varepsilon > 0$ it holds that

$$\Pr\{\hat{f}_N(x) - f(x) \ge \varepsilon\} \le \exp\{-NI_x(\varepsilon)\},\tag{7.225}$$

where

$$I_x(z) := \sup_{t \in \mathbb{R}} \left\{ zt - \ln M_x(t) \right\}$$
(7.226)

is the LD rate function of random variable $F(x,\xi) - f(x)$. Similarly

$$\Pr\{\hat{f}_N(x) - f(x) \le -\varepsilon\} \le \exp\{-NI_x(-\varepsilon)\},\$$

and hence

$$\Pr\left\{\left|\hat{f}_{N}(x) - f(x)\right| \ge \varepsilon\right\} \le \exp\left\{-NI_{x}(\varepsilon)\right\} + \exp\left\{-NI_{x}(-\varepsilon)\right\}.$$
(7.227)

By assumption (C1) we have that both $I_x(\varepsilon)$ and $I_x(-\varepsilon)$ are positive for every $x \in X$.

For a $\nu > 0$, let $\bar{x}_1, ..., \bar{x}_K \in X$ be such that for every $x \in X$ there exists \bar{x}_i , $i \in \{1, ..., K\}$, such that $||x - \bar{x}_i|| \le \nu$, i.e., $\{\bar{x}_1, ..., \bar{x}_K\}$ is a ν -net in X. We can choose this net in such a way that

$$K \le \left[\varrho D/\nu\right]^n,\tag{7.228}$$

where

$$D := \sup_{x', x \in X} \|x' - x\|$$

is the diameter of X and ρ is a constant depending on the chosen norm $\|\cdot\|$. By (7.223) we have that

$$|f(x') - f(x)| \le L ||x' - x||, \tag{7.229}$$

where $L := \mathbb{E}[\kappa(\xi)]$ is finite by assumption (C3). Moreover,

$$\left|\hat{f}_{N}(x') - \hat{f}_{N}(x)\right| \le \hat{\kappa}_{N} \|x' - x\|,$$
(7.230)

where $\hat{\kappa}_N := N^{-1} \sum_{j=1}^N \kappa(\xi^j)$. Again, because of condition (C3), by Cramér's LD Theorem we have that for any L' > L there is a constant $\ell > 0$ such that

$$\Pr\left\{\hat{\kappa}_N \ge L'\right\} \le \exp\{-N\ell\}. \tag{7.231}$$

Consider

$$Z_i := \hat{f}_N(\bar{x}_i) - f(\bar{x}_i), \ i = 1, ..., K.$$

We have that the event $\{\max_{1 \le i \le K} |Z_i| \ge \varepsilon\}$ is equal to the union of the events $\{|Z_i| \ge \varepsilon\}$, i = 1, ..., K, and hence

$$\Pr\left\{\max_{1\leq i\leq K} |Z_i| \geq \varepsilon\right\} \leq \sum_{i=1}^{K} \Pr\left(|Z_i| \geq \varepsilon\right).$$

Together with (7.227) this implies that

$$\Pr\left\{\max_{1\leq i\leq K} \left| \hat{f}_N(\bar{x}_i) - f(\bar{x}_i) \right| \geq \varepsilon\right\} \leq 2\sum_{i=1}^K \exp\left\{ -N[I_{\bar{x}_i}(\varepsilon) \wedge I_{\bar{x}_i}(-\varepsilon)] \right\}.$$
(7.232)

For an $x \in X$ let $i(x) \in \arg \min_{1 \le i \le K} ||x - \bar{x}_i||$. By construction of the ν -net we have that $||x - \bar{x}_{i(x)}|| \le \nu$ for every $x \in X$. Then

$$\begin{aligned} \left| \hat{f}_N(x) - f(x) \right| &\leq \left| \hat{f}_N(x) - \hat{f}_N(\bar{x}_{i(x)}) \right| + \left| \hat{f}_N(\bar{x}_{i(x)}) - f(\bar{x}_{i(x)}) \right| + \left| f(\bar{x}_{i(x)}) - f(x) \right| \\ &\leq \hat{\kappa}_N \nu + \left| \hat{f}_N(\bar{x}_{i(x)}) - f(\bar{x}_{i(x)}) \right| + L\nu. \end{aligned}$$

Let us take now a ν -net with such ν that $L\nu = \varepsilon/4$, i.e., $\nu := \varepsilon/(4L)$. Then

$$\Pr\left\{\sup_{x\in X} \left|\hat{f}_N(x) - f(x)\right| \ge \varepsilon\right\} \le \Pr\left\{\hat{\kappa}_N \nu + \max_{1\le i\le K} \left|\hat{f}_N(\bar{x}_i) - f(\bar{x}_i)\right| \ge 3\varepsilon/4\right\}.$$

Moreover, we have that

$$\Pr\left\{\hat{\kappa}_N\nu\geq\varepsilon/2\right\}\leq\exp\{-N\ell\}$$

where ℓ is a positive constant specified in (7.231) for L' := 2L. Consequently

$$\Pr\left\{\sup_{x\in X} \left|\hat{f}_{N}(x) - f(x)\right| \geq \varepsilon\right\}$$

$$\leq \exp\{-N\ell\} + \Pr\left\{\max_{1\leq i\leq K} \left|\hat{f}_{N}(\bar{x}_{i}) - f(\bar{x}_{i})\right| \geq \varepsilon/4\right\}$$

$$\leq \exp\{-N\ell\} + 2\sum_{i=1}^{K} \exp\left\{-N\left[I_{\bar{x}_{i}}(\varepsilon/4) \land I_{\bar{x}_{i}}(-\varepsilon/4)\right]\right\}.$$
(7.233)

Since the above choice of the ν -net does not depend on the sample (although it depends on ε), and both $I_{\bar{x}_i}(\varepsilon/4)$ and $I_{\bar{x}_i}(-\varepsilon/4)$ are positive, i = 1, ..., K, we obtain that (7.233) implies (7.224), and hence completes the proof. \Box

In the convex case the (Lipschitz continuity) condition (C2) holds, in a sense, automatically. That is, we have the following result.

Theorem 7.74. Let $U \subset \mathbb{R}^n$ be a convex open set. Suppose that: (i) for a.e. $\xi \in \Xi$ the function $F(\cdot,\xi) : U \to \mathbb{R}$ is convex, and (ii) for every $x \in U$ the moment generating function $M_x(t)$ is finite valued for all t in a neighborhood of zero. Then for every compact set $X \subset U$ and $\varepsilon > 0$ there exist positive constants C and $\beta = \beta(\varepsilon)$, independent of N, such that

$$\Pr\left\{\sup_{x\in X}\left|\hat{f}_{N}(x)-f(x)\right|\geq\varepsilon\right\}\leq Ce^{-N\beta}.$$
(7.234)

Proof. We have here that the expectation function f(x) is convex and finite valued for all $x \in U$. Let X be a (nonempty) compact subset of U. For $\gamma \ge 0$ consider the set

$$X_{\gamma} := \{ x \in \mathbb{R}^n : \operatorname{dist}(x, X) \le \gamma \}.$$

Since the set U is open, we can choose $\gamma > 0$ such that $X_{\gamma} \subset U$. The set X_{γ} is compact and by convexity of $f(\cdot)$ we have that $f(\cdot)$ is continuous and hence is bounded on X_{γ} . That is, there is constant c > 0 such that $|f(x)| \le c$ for all $x \in X_{\gamma}$. Also by convexity of $f(\cdot)$ we have for any $\tau \in [0, 1]$, $x \in X$ and $y \in \mathbb{R}^n$ such that $x + y \in U$ and $x - y/\tau \in U$:

$$f(x) = f\left(\frac{1}{1+\tau}(x+y) + \frac{\tau}{1+\tau}(x-y/\tau)\right) \le \frac{1}{1+\tau}f(x+y) + \frac{\tau}{1+\tau}f(x-y/\tau).$$

It follows that if $||y|| \le \tau \gamma, \tau \in [0, 1]$, and hence $x + y \in X_{\gamma}$ and $x - y/\tau \in X_{\gamma}$, then

$$f(x+y) \ge (1+\tau)f(x) - \tau f(x-y/\tau) \ge f(x) - 2\tau c.$$
(7.235)

Now we proceed similar to the proof of Theorem 7.73. Let $\varepsilon > 0$ and $\nu > 0$, and $\bar{x}_1, ..., \bar{x}_K \in U$ be a ν -net for X_{γ} . As in the proof of Theorem 7.73, this ν -net will be dependent on ε , but not on the random sample $\xi^1, ..., \xi^N$. Consider the event

$$A_N := \left\{ \max_{1 \le i \le K} \left| \hat{f}_N(\bar{x}_i) - f(\bar{x}_i) \right| \le \varepsilon \right\}.$$

By (7.225) and (7.227) we have similar to (7.232) that $Pr(A_N) \ge 1 - \alpha_N$, where

$$\alpha_N := 2 \sum_{i=1}^K \exp\left\{-N[I_{\bar{x}_i}(\varepsilon) \wedge I_{\bar{x}_i}(-\varepsilon)]\right\}$$

Consider a point $x \in X$ and let $\mathcal{I} \subset \{1, ..., K\}$ be such index set that x is a convex combination of points $\bar{x}_i, i \in \mathcal{I}$, i.e., $x = \sum_{i \in \mathcal{I}} t_i \bar{x}_i$, for some positive numbers t_i summing up to one. Moreover, let \mathcal{I} be such that $||x - \bar{x}_i|| \leq a\nu$, for all $i \in \mathcal{I}$, where a > 0 is a constant independent of x and the net. By convexity of $\hat{f}_N(\cdot)$ we have that $\hat{f}_N(x) \leq \sum_{i \in \mathcal{I}} t_i \hat{f}_N(\bar{x}_i)$. It follows that the event A_N is included in the event $\{\hat{f}_N(x) \leq \sum_{i \in \mathcal{I}} t_i f(\bar{x}_i) + \varepsilon\}$. By (7.235) we also have that

$$f(x) \ge f(\bar{x}_i) - 2\tau c, \quad \forall i \in \mathcal{I},$$

provided that $a\nu \leq \tau\gamma$. Setting $\tau := \varepsilon/(2c)$, we obtain that the event A_N is included in the event $B_x := \left\{ \hat{f}_N(x) \leq f(x) + 2\varepsilon \right\}$, provided that¹⁹ $\nu \leq O(1)\varepsilon$. It follows that the event A_N is included in the event $\cap_{x \in X} B_x$, and hence

$$\Pr\left\{\sup_{x\in X} \left(\hat{f}_N(x) - f(x)\right) \le 2\varepsilon\right\} = \Pr\left\{\bigcap_{x\in X} B_x\right\} \ge \Pr\left\{A_N\right\} \ge 1 - \alpha_N, \quad (7.236)$$

provided that $\nu \leq O(1)\varepsilon$.

In order to derive the converse to (7.236) estimate let us observe that by convexity of $\hat{f}_N(\cdot)$ we have with probability at least $1 - \alpha_N$ that $\sup_{x \in X_\gamma} \hat{f}_N(x) \le c + \varepsilon$. Also by using (7.235) we have with probability at least $1 - \alpha_N$ that $\inf_{x \in X_\gamma} \hat{f}_N(x) \ge -(c + \varepsilon)$, provided that $\nu \le O(1)\varepsilon$. That is, with probability at least $1 - 2\alpha_N$ we have that

$$\sup_{x \in X_{\gamma}} \left| \hat{f}_N(x) \right| \le c + \varepsilon,$$

¹⁹Recall that O(1) denotes a generic constant, here $O(1) = \gamma/(2ca)$.

$$\Pr\left\{\sup_{x\in X} \left(f(x) - \hat{f}_N(x)\right) \le 2\varepsilon\right\} \ge 1 - 3\alpha_N.$$
(7.237)

Since, by the condition (ii), $I_{\bar{x}_i}(\varepsilon)$ and $I_{\bar{x}_i}(-\varepsilon)$ are positive, this completes the proof.

Now let us strengthen condition (C1) to the following condition:

(C4) There exists constant $\sigma > 0$ such that for any $x \in X$, the following inequality holds:

$$M_x(t) \le \exp\left\{\sigma^2 t^2/2\right\}, \quad \forall t \in \mathbb{R}.$$
(7.238)

It follows from condition (7.238) that $\ln M_x(t) \leq \sigma^2 t^2/2$, and hence²⁰

$$I_x(z) \ge \frac{z^2}{2\sigma^2}, \quad \forall z \in \mathbb{R}.$$
(7.239)

Consequently, inequality (7.233) implies

$$\Pr\left\{\sup_{x\in X}\left|\hat{f}_{N}(x) - f(x)\right| \ge \varepsilon\right\} \le \exp\{-N\ell\} + 2K\exp\left\{-\frac{N\varepsilon^{2}}{32\sigma^{2}}\right\},$$
(7.240)

where ℓ is a constant specified in (7.231) with L' := 2L, $K = [\rho D/\nu]^n$, $\nu = \varepsilon/(4L)$ and hence

$$K = \left[4\rho DL/\varepsilon\right]^n. \tag{7.241}$$

If we assume further that the Lipschitz constant in (7.223) does not depend on ξ , i.e., $\kappa(\xi) \equiv L$, then the first term in the right hand side of (7.240) can be omitted. Therefore we obtain the following result.

Theorem 7.75. Suppose that conditions (C2)– (C4) hold and that the set X has finite diameter D. Then

$$\Pr\left\{\sup_{x\in X} \left|\hat{f}_N(x) - f(x)\right| \ge \varepsilon\right\} \le \exp\{-N\ell\} + 2\left[\frac{4\varrho DL}{\varepsilon}\right]^n \exp\left\{-\frac{N\varepsilon^2}{32\sigma^2}\right\}.$$
 (7.242)

Moreover, if $\kappa(\xi) \equiv L$ *in condition* (C2), *then condition* (C3) *holds automatically and the term* $\exp\{-N\ell\}$ *in the right hand side of* (7.242) *can be omitted.*

As it was shown in the proof of Theorem 7.74, in the convex case estimates of the form (7.242), with different constants, can be obtained without assuming the (Lipschitz continuity) condition (C2).

²⁰Recall that if random variable $F(x,\xi) - f(x)$ has normal distribution with variance σ^2 , then its moment generating function is equal to the right hand side of (7.238), and hence the inequalities (7.238) and (7.239) hold as equalities.

Exponential Convergence of Generalized Gradients

The above results can be also applied to establishing rates of convergence of directional derivatives and generalized gradients (subdifferentials) of $\hat{f}_N(x)$ at a given point $\bar{x} \in X$. Consider the following condition.

(C5) For a.e. $\xi \in \Xi$, the function $F_{\xi}(\cdot) = F(\cdot, \xi)$ is directionally differentiable at a point $\bar{x} \in X$.

Consider the expected value function $f(x) = \mathbb{E}[F(x,\xi)] = \int_{\Xi} F(x,\xi) dP(\xi)$. Suppose that $f(\bar{x})$ is finite and condition (C2) holds with the respective Lipschitz constant $\kappa(\xi)$ being *P*-integrable, i.e., $\mathbb{E}[\kappa(\xi)] < +\infty$. Then it follows that f(x) is finite valued and Lipschitz continuous on *X* with Lipschitz constant $\mathbb{E}[\kappa(\xi)]$. Moreover, the following result for Clarke generalized gradient of f(x) holds (cf., [45, Theorem 2.7.2]).

Theorem 7.76. Suppose that condition (C2) holds with $\mathbb{E}[\kappa(\xi)] < +\infty$, and let \bar{x} be an interior point of the set X such that $f(\bar{x})$ is finite. If, moreover, $F(\cdot,\xi)$ is Clarke-regular at \bar{x} for a.e. $\xi \in \Xi$, then f is Clarke-regular at \bar{x} and

$$\partial^{\circ} f(\bar{x}) = \int_{\Xi} \partial^{\circ} F(\bar{x},\xi) dP(\xi), \qquad (7.243)$$

where Clarke generalized gradient $\partial^{\circ} F(\bar{x}, \xi)$ is taken with respect to x.

The above result can be extended to an infinite dimensional setting with the set X being a subset of a separable Banach space \mathcal{X} . Formula (7.243) can be interpreted in the following way. For every $\gamma \in \partial^{\circ} f(\bar{x})$, there exists a measurable selection $\Gamma(\xi) \in \partial^{\circ} F(\bar{x}, \xi)$ such that for every $v \in \mathcal{X}^*$, the function $\langle v, \Gamma(\cdot) \rangle$ is integrable and

$$\langle v, \gamma \rangle = \int_{\Xi} \langle v, \Gamma(\xi) \rangle dP(\xi).$$

In this way, γ can be considered as an integral of a measurable selection from $\partial^{\circ} F(\bar{x}, \cdot)$.

Theorem 7.77. Let \bar{x} be an interior point of the set X. Suppose that $f(\bar{x})$ is finite and conditions (C2)–(C3) and (C5) hold. Then for any $\varepsilon > 0$ there exist positive constants C and $\beta = \beta(\varepsilon)$, independent of N, such that²¹

$$\Pr\left\{\sup_{d\in S^{n-1}}\left|\hat{f}'_N(\bar{x},d) - f'(\bar{x},d)\right| > \varepsilon\right\} \le Ce^{-N\beta}.$$
(7.244)

Moreover, suppose that for a.e. $\xi \in \Xi$ the function $F(\cdot, \xi)$ is Clarke-regular at \bar{x} . Then

$$\Pr\left\{\mathbb{H}\left(\partial^{\circ}\hat{f}_{N}(\bar{x}),\partial^{\circ}f(\bar{x})\right)>\varepsilon\right\}\leq Ce^{-N\beta}.$$
(7.245)

Furthermore, if in condition (C2) $\kappa(\xi) \equiv L$ is constant, then

$$\Pr\left\{\mathbb{H}\left(\partial^{\circ}\hat{f}_{N}(\bar{x}),\partial^{\circ}f(\bar{x})\right) > \varepsilon\right\} \le 2\left[\frac{4\varrho L}{\varepsilon}\right]^{n}\exp\left\{-\frac{N\varepsilon^{2}}{128L^{2}}\right\}.$$
(7.246)

 $^{{}^{21}\}text{By }S^{n-1}:=\{d\in\mathbb{R}^n:\|d\|=1\}\text{ we denote the unit sphere taken with respect to a norm }\|\cdot\|\text{ on }\mathbb{R}^n.$

Proof. Since $f(\bar{x})$ is finite, conditions (C2)-(C3) and (C5) imply that $f(\cdot)$ is finite valued and Lipschitz continuous in a neighborhood of \bar{x} , $f(\cdot)$ is directionally differentiable at \bar{x} , its directional derivative $f'(\bar{x}, \cdot)$ is Lipschitz continuous, and $f'(\bar{x}, \cdot) = \mathbb{E}[\eta(\cdot, \xi)]$, where $\eta(\cdot, \xi) := F'_{\xi}(\bar{x}, \cdot)$ (see Theorem 7.49). We also have here that $\hat{f}'_N(\bar{x}, \cdot) = \hat{\eta}_N(\cdot)$, where

$$\hat{\eta}_N(d) := \frac{1}{N} \sum_{i=1}^N \eta(d, \xi^i), \ d \in \mathbb{R}^n,$$
(7.247)

and $\mathbb{E}[\hat{\eta}_N(d)] = f'(\bar{x}, d)$, for all $d \in \mathbb{R}^n$. Moreover, conditions (C2) and (C5) imply that $\eta(\cdot, \xi)$ is Lipschitz continuous on \mathbb{R}^n , with Lipschitz constant $\kappa(\xi)$, and in particular that $|\eta(d,\xi)| \leq \kappa(\xi) ||d||$ for any $d \in \mathbb{R}^n$ and $\xi \in \Xi$. Hence together with condition (C3) this implies that, for every $d \in \mathbb{R}^n$, the moment generating function of $\eta(d,\xi)$ is finite valued in a neighborhood of zero.

Consequently, the estimate (7.244) follows directly from Theorem 7.73. If $F_{\xi}(\cdot)$ is Clarke-regular for a.e. $\xi \in \Xi$, then $\hat{f}_N(\cdot)$ is also Clarke-regular and

$$\partial^{\circ} \hat{f}_N(\bar{x}) = N^{-1} \sum_{i=1}^N \partial^{\circ} F_{\xi^i}(\bar{x}).$$

By applying (7.244) together with equation (7.170) for sets $A_1 := \partial^{\circ} \hat{f}_N(\bar{x})$ and $A_2 := \partial^{\circ} f(\bar{x})$, we obtain (7.245).

Now if $\kappa(\xi) \equiv L$ is constant, then $\eta(\cdot,\xi)$ is Lipschitz continuous on \mathbb{R}^n , with Lipschitz constant L, and $|\eta(d,\xi)| \leq L$ for every $d \in S^{n-1}$ and $\xi \in \Xi$. Consequently, for any $d \in S^{n-1}$ and $\xi \in \Xi$ we have that $|\eta(d,\xi) - \mathbb{E}[\eta(d,\xi)]| \leq 2L$, and hence for every $d \in S^{n-1}$ the moment generating function $M_d(t)$ of $\eta(d,\xi) - \mathbb{E}[\eta(d,\xi)]$ is bounded $M_d(t) \leq \exp\{2t^2L^2\}$, for all $t \in \mathbb{R}$ (see (7.211)). It follows by Theorem 7.75 that

$$\Pr\left\{\sup_{d\in S^{n-1}}\left|\hat{f}_{N}'(\bar{x},d) - f'(\bar{x},d)\right| > \varepsilon\right\} \le 2\left[\frac{4\varrho L}{\varepsilon}\right]^{n} \exp\left\{-\frac{N\varepsilon^{2}}{128L^{2}}\right\},$$
(7.248)

and hence (7.246) follows.

7.3 Elements of Functional Analysis

Recall that given a linear (vector) space \mathcal{V} , a function $\phi : \mathcal{V} \to \mathbb{R}$ is said to be *sublinear* if $\phi(\alpha x) = \alpha \phi(x)$ for all $x \in \mathcal{V}$ and $\alpha \ge 0$ (positive homogeneity), and $\phi(x + y) \le \phi(x) + \phi(y)$ for all $x, y \in \mathcal{V}$ (subadditivity).

Theorem 7.78 (Hahn-Banach). Let \mathcal{V} be a linear space, $\phi : \mathcal{V} \to \mathbb{R}$ be a sublinear function and $\ell : \mathcal{U} \to \mathbb{R}$ be a linear functional on a linear subspace $\mathcal{U} \subset \mathcal{V}$ which is dominated by ϕ on \mathcal{U} , i.e., $\ell(x) \leq \phi(x)$ for all $x \in \mathcal{U}$. Then there exists a linear functional $\hat{\ell} : \mathcal{V} \to \mathbb{R}$ such that $\hat{\ell}(x) = \ell(x)$ for all $x \in \mathcal{U}$, and $\hat{\ell}(x) \leq \phi(x)$ for all $x \in \mathcal{V}$.

Note that since the function ϕ is sublinear, it follows that $\phi(0) = 0$. Therefore by taking the space $\mathcal{U} := \{0\}$ and defining $\ell(0) = 0$, Hahn-Banach Theorem implies existence of a linear functional $\hat{\ell} : \mathcal{V} \to \mathbb{R}$ dominated by ϕ on \mathcal{V} .

A linear space \mathcal{Z} equipped with a norm $\|\cdot\|$ is said to be a *Banach space* if it is complete, i.e., every Cauchy sequence in \mathcal{Z} has a limit. Let \mathcal{Z} be a Banach space. Unless stated otherwise, all topological statements (like convergence, continuity, lower semicontinuity, etc) will be made with respect to the norm topology of \mathcal{Z} . It is said that Banach space \mathcal{Z} is *separable* if it has a dense countable subset.

The space of all linear continuous functionals $\zeta : \mathbb{Z} \to \mathbb{R}$ forms the dual of space \mathbb{Z} and is denoted \mathbb{Z}^* . For $\zeta \in \mathbb{Z}^*$ and $z \in \mathbb{Z}$ we denote $\langle \zeta, z \rangle := \zeta(z)$ and view it as a scalar product on $\mathbb{Z}^* \times \mathbb{Z}$. The space \mathbb{Z}^* , equipped with the dual norm

$$\|\zeta\|_* := \sup_{\|z\| \le 1} \langle \zeta, z \rangle, \tag{7.249}$$

is also a Banach space. Consider the dual \mathcal{Z}^{**} of the space \mathcal{Z}^* . There is a natural embedding of \mathcal{Z} into \mathcal{Z}^{**} given by identifying $z \in \mathcal{Z}$ with linear functional $\langle \cdot, z \rangle$ on \mathcal{Z}^* . In that sense, \mathcal{Z} can be considered as a subspace of \mathcal{Z}^{**} . It is said that Banach space \mathcal{Z} is *reflexive* if \mathcal{Z} coincides with \mathcal{Z}^{**} . It follows from the definition of the dual norm that

$$\langle \zeta, z \rangle | \le \| \zeta \|_* \| z \|, \quad z \in \mathcal{Z}, \, \zeta \in \mathcal{Z}^*.$$
(7.250)

Together with the strong (norm) topology of Z we sometimes need to consider its weak topology, which is the weakest topology in which all linear functionals $\langle \zeta, \cdot \rangle, \zeta \in Z^*$, are continuous. The dual space Z^* can be also equipped with its weak* topology, which is the weakest topology in which all linear functionals $\langle \cdot, z \rangle, z \in Z$, are continuous. If the space Z is reflexive, then Z^* is also reflexive and its weak* and weak topologies do coincide. Note also that a convex subset of Z is closed in the strong topology iff it is closed in the weak topology of Z.

Theorem 7.79 (Banach-Alaoglu). Let \mathcal{Z} be Banach space. The closed unit ball $B_{\mathcal{Z}^*} := \{\zeta \in \mathcal{Z}^* : \|\zeta\|_* \leq 1\}$ is compact in the weak* topology of \mathcal{Z}^* .

It follows that any bounded (in the dual norm $\|\cdot\|_*$) and weakly* closed subset of \mathcal{Z}^* is weakly* compact. Therefore if K is a nonempty bounded and weakly* closed subset of \mathcal{Z}^* , then for any $z \in \mathcal{Z}$ the set

$$\mathfrak{S}_{K}(z) := \arg\max\left\{\langle \zeta, z \rangle : \zeta \in K\right\}$$
(7.251)

is nonempty. For the unit ball $B_{\mathcal{Z}^*}$ we refer to the corresponding set $\mathfrak{S}(z) = \mathfrak{S}_{B_{\mathcal{Z}^*}}(z)$ as the set of *contact* points of z, i.e., $\overline{\zeta}$ is a contact point of z if

$$\zeta \in \operatorname*{argmax}_{\|\zeta\|_* \le 1} \langle \zeta, z \rangle.$$

Definition 7.80. Let $K \subset \mathbb{Z}^*$ be a nonempty weakly^{*} compact set. It is said that $\overline{\zeta} \in K$ is a weak^{*} exposed point of K if there exists $z \in \mathbb{Z}$ such that $\langle \zeta, z \rangle$ attains its maximum over $\zeta \in K$ at the unique point $\overline{\zeta}$, i.e., the set $\mathfrak{S}_K(z) = \{\overline{\zeta}\}$ is a singleton. In that case it is said that z exposes K at $\overline{\zeta}$. We denote by $\operatorname{Exp}(K)$ the set of exposed points of K.

The following result is going back to Mazur [154]. We give its proof for the sake of completeness (this proof was suggested to us by A. Nemirovski).

Theorem 7.81. If Z is a separable Banach space and K is a nonempty weakly^{*} compact subset of Z^* , then the set of points $z \in Z$ which expose K at some point $\overline{\zeta} \in K$ is a dense subset of Z

Proof. Let $z_1, z_2, ...$ be a sequence, forming a dense subset of the unit sphere $\{z \in \mathbb{Z} : \|z\| = 1\}$. Since the space \mathbb{Z} is separable, such sequence exists. Let $\lambda_1, \lambda_2, ...$ be a sequence of positive numbers such that $\sum_{i=1}^{\infty} \lambda_i \leq \varepsilon$ for some $\varepsilon > 0$. Given $z \in \mathbb{Z}$ consider functional $f_{\varepsilon} : \mathbb{Z}^* \to \mathbb{R}$ defined as

$$f_{\varepsilon}(\zeta) := \langle \zeta, z \rangle + \frac{1}{2} \sum_{i=1}^{\infty} \lambda_i \langle \zeta, z_i \rangle^2.$$

Note that since $||z_i|| = 1$,

$$\sum_{i=1}^{\infty} \lambda_i \langle \zeta, z_i \rangle^2 \le \sum_{i=1}^{\infty} \lambda_i \|\zeta\|_*^2 \le \varepsilon \|\zeta\|_*^2,$$

and hence $f_{\varepsilon}(\zeta)$ is well defined. The functional $f_{\varepsilon}(\cdot)$ is weakly^{*} continuous, and hence attains its maximum over K at a point $\overline{\zeta} \in K$. Moreover, $f_{\varepsilon}(\cdot)$ is (Gâteaux) differentiable, with derivative $\nabla f_{\varepsilon}(\zeta) \in \mathcal{Z}$ given by

$$\nabla f_{\varepsilon}(\zeta) = z + \sum_{i=1}^{\infty} \lambda_i \langle \zeta, z_i \rangle z_i.$$
(7.252)

For $\zeta_1 \neq \zeta_2$ we have that

$$\frac{f_{\varepsilon}(\zeta_1) + f_{\varepsilon}(\zeta_2)}{2} - f_{\varepsilon}\left(\frac{\zeta_1 + \zeta_2}{2}\right) = \frac{1}{4}\sum_{i=1}^{\infty}\lambda_i\langle\zeta_1 - \zeta_2, z_i\rangle^2 > 0,$$

where the inequality follows since the sequence z_i is dense on the unit sphere. Hence $f_{\varepsilon}(\cdot)$ is *strictly* convex, and thus

$$f_{\varepsilon}(\zeta) > f_{\varepsilon}(\bar{\zeta}) + \langle \zeta - \bar{\zeta}, \nabla f_{\varepsilon}(\bar{\zeta}) \rangle, \ \zeta \neq \bar{\zeta}.$$

Since $\bar{\zeta}$ is a corresponding maximizer, i.e., $f_{\varepsilon}(\bar{\zeta}) \ge f_{\varepsilon}(\zeta)$ for all $\zeta \in K$, it follows that

$$f_{\varepsilon}(\bar{\zeta}) > f_{\varepsilon}(\bar{\zeta}) + \langle \zeta - \bar{\zeta}, \nabla f_{\varepsilon}(\bar{\zeta}) \rangle, \ \zeta \in K, \ \zeta \neq \bar{\zeta}.$$

Thus

$$\langle \zeta - \bar{\zeta}, \nabla f_{\varepsilon}(\bar{\zeta}) \rangle < 0, \ \zeta \in K, \ \zeta \neq \bar{\zeta},$$

and hence $\bar{\zeta}$ is the unique maximizer of $\langle \zeta, \nabla f_{\varepsilon}(\bar{\zeta}) \rangle$ over $\zeta \in K$. That is, $\nabla f_{\varepsilon}(\bar{\zeta})$ exposes K at $\bar{\zeta}$.

Let \mathcal{D} be the set of points of the form $\nabla f_{\varepsilon}(\overline{\zeta})$ for all $z \in \mathbb{Z}$ and $\varepsilon > 0$. By the above we have that \mathcal{D} is a set of points in \mathbb{Z} which expose K. Also by (7.252) we have

$$||z - \nabla f_{\varepsilon}(\bar{\zeta})|| \le \varepsilon ||\bar{\zeta}||^* \le \varepsilon C,$$

where $C := \sup_{\zeta \in K} \|\zeta\|_*$ is a finite constant since the set K is bounded. Hence any $z \in \mathbb{Z}$ can be approximated with an arbitrary precision by a point $z' \in \mathcal{D}$, i.e., \mathcal{D} is a dense subset of \mathbb{Z} . This completes the proof. \Box

An important class of Banach spaces are $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ spaces, where (Ω, \mathcal{F}, P) is a probability space and $p \in [1, \infty)$. The space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ consists of all \mathcal{F} -measurable functions $\phi : \Omega \to \mathbb{R}$ such that $\int_{\Omega} |\phi(\omega)|^p dP(\omega) < +\infty$. More precisely, an element of $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ is a class of such functions $\phi(\omega)$ which may differ from each other on sets of P-measure zero. Equipped with the norm

$$\|\phi\|_p := \left(\int_{\Omega} |\phi(\omega)|^p \, dP(\omega)\right)^{1/p},\tag{7.253}$$

 $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ becomes a Banach space.

We also use the space $\mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$ of functions (or rather classes of functions which may differ on sets of *P*-measure zero) $\phi : \Omega \to \mathbb{R}$ which are \mathcal{F} -measurable and essentially bounded. A function ϕ is said to be essentially bounded if its sup-norm

$$\|\phi\|_{\infty} := \operatorname{ess\,sup}_{\omega \in \Omega} |\phi(\omega)| \tag{7.254}$$

is finite, where

$$\operatorname{ess\,sup}_{\omega\in\Omega} |\phi(\omega)| := \inf\left\{ \sup_{\omega\in\Omega} |\psi(\omega)| : \phi(\omega) = \psi(\omega) \text{ a.e. } \omega \in \Omega \right\}.$$
(7.255)

Note that in case of infinite probability space (Ω, \mathcal{F}, P) , the space $\mathcal{Z} = \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$ is not separable. Indeed, we have that $\|\mathbf{1}_{A} - \mathbf{1}_{A'}\|_{\infty} = 1$ for any sets $A, A' \in \mathcal{F}$ such that $P(A \neq A') > 0$. Since the set of all \mathcal{F} -measurable subsets of Ω uncountable, it is not possible to construct a countable dense subset of \mathcal{Z} .

In particular, suppose that the set $\Omega := \{\omega_1, ..., \omega_n\}$ is finite, and let \mathcal{F} be the sigma algebra of all subsets of Ω and $p_1, ..., p_n$ be (positive) probabilities of the corresponding elementary events. In that case every element $z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ can be viewed as a finite dimensional vector $(z(\omega_1), ..., z(\omega_n))$, and $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ can be identified with the space \mathbb{R}^n equipped with the corresponding norm

$$||z||_p := \left(\sum_{i=1}^n p_i |z(\omega_i)|^p\right)^{1/p}.$$
(7.256)

We also use spaces $\mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^m)$, with $p \in [1, \infty]$. For $p \in [1, \infty)$ this space is formed by all \mathcal{F} -measurable functions (mappings) $\psi : \Omega \to \mathbb{R}^m$ such that $\int_{\Omega} \|\psi(\omega)\|^p dP(\omega) < +\infty$, with the corresponding norm $\|\cdot\|$ on \mathbb{R}^m being, for example, the Euclidean norm. For $p = \infty$, the corresponding space consists of all essentially bounded functions $\psi : \Omega \to \mathbb{R}^m$.

For $p \in (1, \infty)$ the dual of $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ is the space $\mathcal{L}_q(\Omega, \mathcal{F}, P)$, where $q \in (1, \infty)$ is such that 1/p+1/q = 1, and these spaces are reflexive. This duality is derived by Hölder inequality

$$\int_{\Omega} |\zeta(\omega)z(\omega)| dP(\omega) \le \left(\int_{\Omega} |\zeta(\omega)|^q dP(\omega)\right)^{1/q} \left(\int_{\Omega} |z(\omega)|^p dP(\omega)\right)^{1/p}.$$
 (7.257)

For points $z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and $\zeta \in \mathcal{L}_q(\Omega, \mathcal{F}, P)$, their scalar product is defined as

$$\langle \zeta, z \rangle := \int_{\Omega} \zeta(\omega) z(\omega) dP(\omega).$$
 (7.258)

The dual of $\mathcal{L}_1(\Omega, \mathcal{F}, P)$ is the space $\mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$, and these spaces are not reflexive.

If $z(\omega)$ is not zero for a.e. $\omega \in \Omega$, then the equality in (7.257) holds iff $\zeta(\omega)$ is proportional²² to $\operatorname{sign}(z(\omega))|z(\omega)|^{p/q}$. It follows that for $p \in (1, \infty)$, with every nonzero $z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ is associated unique maximizer of $\langle \zeta, z \rangle$ over the unit ball $B_{\mathbb{Z}^*}$ of the dual space $\mathcal{L}_q(\Omega, \mathcal{F}, P)$. That is, $z \neq 0$ exposes $B_{\mathbb{Z}^*}$ at the contact point, denoted ζ_z^* , which can be written in the form

$$\zeta_z^*(\omega) = \frac{\text{sign}(z(\omega))|z(\omega)|^{p/q}}{\|z\|_p^{p/q}}.$$
(7.259)

In particular, for p = 2 and q = 2, $\zeta_z^* = ||z||_2^{-1} z$.

For p=1 and $z\in \mathcal{L}_1(\Omega,\mathcal{F},P)$ the corresponding set of contact points can be described as follows

$$\mathfrak{S}(z) = \begin{cases} \zeta(\omega) = 1, & \text{if } z(\omega) > 0, \\ \zeta \in \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P) : \zeta(\omega) = -1, & \text{if } z(\omega) < 0, \\ \zeta(\omega) \in [-1, 1], & \text{if } z(\omega) = 0. \end{cases}$$
(7.260)

It follows that $\mathfrak{S}(z)$ is a singleton iff $z(\omega) \neq 0$ for a.e. $\omega \in \Omega$, in which case $\mathfrak{S}(z) = \{\operatorname{sign}(z)\}$.

7.3.1 Conjugate Duality and Differentiability

Let \mathcal{Z} be a Banach space, \mathcal{Z}^* be its dual space and $f : \mathcal{Z} \to \overline{\mathbb{R}}$ be an extended real valued function. Similar to the final dimensional case we define the *conjugate function* of f as

$$f^*(\zeta) := \sup_{z \in \mathcal{Z}} \left\{ \langle \zeta, z \rangle - f(z) \right\}.$$
(7.261)

The conjugate function $f^*: \mathcal{Z}^* \to \overline{\mathbb{R}}$ is always convex and lower semicontinuous. The biconjugate function $f^{**}: \mathcal{Z} \to \overline{\mathbb{R}}$, i.e., the conjugate of f^* , is

$$f^{**}(z) := \sup_{\zeta \in \mathcal{Z}^*} \left\{ \langle \zeta, z \rangle - f^*(\zeta) \right\}.$$
(7.262)

The basic duality theorem still holds in the considered infinite dimensional framework.

Theorem 7.82 (Fenchel-Moreau). Let \mathcal{Z} be a Banach space and $f : \mathcal{Z} \to \overline{\mathbb{R}}$ be a proper extended real valued convex function. Then

$$f^{**} = \operatorname{lsc} f. \tag{7.263}$$

²²For $a \in \mathbb{R}$, sign(a) is equal to 1 if a > 0, to -1 if a < 0, and to 0 if a = 0.

It follows from (7.263) that if f is proper and convex, then $f^{**} = f$ iff f is lower semicontinuous. A basic difference between finite and infinite dimensional frameworks is that in the infinite dimensional case a proper convex function can be discontinuous at an interior point of its domain. As the following result shows, for a convex proper function continuity and lower semicontinuity properties on the interior of its domain are the same.

Proposition 7.83. Let Z be a Banach space and $f : Z \to \mathbb{R}$ be a convex lower semicontinuous function having a finite value in at least one point. Then f is proper and is continuous on int(dom f).

In particular, it follows from the above proposition that if $f : \mathbb{Z} \to \mathbb{R}$ is real valued convex and lower semicontinuous, then f is continuous on \mathbb{Z} .

The subdifferential of a function $f : \mathbb{Z} \to \mathbb{R}$, at a point z_0 such that $f(z_0)$ is finite, is defined in a way similar to the finite dimensional case. That is,

$$\partial f(z_0) := \left\{ \zeta \in \mathcal{Z}^* : f(z) - f(z_0) \ge \langle \zeta, z - z_0 \rangle, \ \forall z \in \mathcal{Z} \right\}.$$
(7.264)

It is said that f is subdifferentiable at z_0 if $\partial f(z_0)$ is nonempty. Clearly, if f is subdifferentiable at some point $z_0 \in \mathbb{Z}$, then f is proper and lower semicontinuous at z_0 . Similar to the finite dimensional case we have the following.

Proposition 7.84. Let \mathcal{Z} be a Banach space and $f : \mathcal{Z} \to \overline{\mathbb{R}}$ be a convex function and $z \in \mathcal{Z}$ be such that $f^{**}(z)$ is finite. Then

$$\partial f^{**}(z) = \arg\max_{\zeta \in \mathcal{Z}^*} \left\{ \langle \zeta, z \rangle - f^*(\zeta) \right\}.$$
(7.265)

Moreover, if $f^{**}(z) = f(z)$, then $\partial f^{**}(z) = \partial f(z)$.

Proposition 7.85. Let Z be a Banach space, $f : Z \to \mathbb{R}$ be a convex function. Suppose that f is finite valued and continuous at a point $z_0 \in Z$. Then f is subdifferentiable at z_0 , $\partial f(z_0)$ is nonempty, convex, bounded and weakly^{*} compact subset of Z^* , f is Hadamard directionally differentiable at z_0 and

$$f'(z_0, h) = \sup_{\zeta \in \partial f(z_0)} \langle \zeta, h \rangle.$$
(7.266)

Note that by the definition, every element of the subdifferential $\partial f(z_0)$ (called *sub-gradient*) is a continuous linear functional on \mathcal{Z} . A linear (not necessarily continuous) functional $\ell : \mathcal{Z} \to \mathbb{R}$ is called an *algebraic subgradient* of f at z_0 if

$$f(z_0 + h) - f(z_0) \ge \ell(h), \ \forall h \in \mathcal{Z}.$$
(7.267)

Of course, if the algebraic subgradient ℓ is also continuous, then $\ell \in \partial f(z_0)$.

Proposition 7.86. Let \mathcal{Z} be a Banach space and $f : \mathcal{Z} \to \mathbb{R}$ be a proper convex function. Then the set of algebraic subgradients at any point $z_0 \in int(dom f)$ is nonempty.
Proof. Consider the directional derivative function $\delta(h) := f'(z_0, h)$. The directional derivative is defined here in the same way as in section 7.1.1. Since f is convex we have that

$$f'(z_0, h) = \inf_{t>0} \frac{f(z_0 + th) - f(z_0)}{t},$$
(7.268)

and $\delta(\cdot)$ is convex, positively homogeneous. Moreover, since $z_0 \in \operatorname{int}(\operatorname{dom} f)$ and hence f(z) is finite valued for all z in a neighborhood of z_0 , it follows by (7.268) that $\delta(h)$ is finite valued for all $h \in \mathbb{Z}$. That is, $\delta(\cdot)$ is a real valued subadditive and positively homogeneous function. Consequently, by Hahn-Banach Theorem we have that there exists a linear functional $\ell : \mathbb{Z} \to \mathbb{R}$ such that $\delta(h) \ge \ell(h)$ for all $h \in \mathbb{Z}$. Since $f(z_0 + h) \ge f(z_0) + \delta(h)$ for any $h \in \mathbb{Z}$, it follows that ℓ is an algebraic subgradient of f at z_0 . \Box

There is also the following version of Moreau-Rockafellar Theorem in the infinite dimensional setting.

Theorem 7.87 (Moreau-Rockafellar). Let $f_1, f_2 : \mathbb{Z} \to \overline{\mathbb{R}}$ be convex proper lower semicontinuous functions, $f := f_1 + f_2$ and $\overline{z} \in \text{dom}(f_1) \cap \text{dom}(f_2)$. Then

$$\partial f(\bar{z}) = \partial f_1(\bar{z}) + \partial f_2(\bar{z}), \tag{7.269}$$

provided that the following regularity condition holds

$$0 \in \inf \left\{ \operatorname{dom}(f_1) - \operatorname{dom}(f_2) \right\}.$$
(7.270)

In particular, equation (7.269) holds if f_1 is continuous at \overline{z} .

Remark 59. It is possible to derive the following (first order) necessary optimality condition from the above theorem. Let S be a convex closed subset of Z and $f : Z \to \mathbb{R}$ be a convex proper lower semicontinuous function. We have that a point $z_0 \in S$ is a minimizer of f(z) over $z \in S$ iff z_0 is a minimizer of $\psi(z) := f(z) + \mathbb{I}_S(z)$ over $z \in Z$. The last condition is equivalent to the condition that $0 \in \partial \psi(z_0)$. Since S is convex and closed, the indicator function $\mathbb{I}_S(\cdot)$ is convex lower semicontinuous, and $\partial \mathbb{I}_S(z_0) = \mathcal{N}_S(z_0)$. Therefore, we have the following.

If $z_0 \in S \cap \text{dom}(f)$ is a minimizer of f(z) over $z \in S$, then

$$0 \in \partial f(z_0) + \mathcal{N}_S(z_0), \tag{7.271}$$

provided that $0 \in int \{ dom(f) - S \}$. In particular, (7.271) holds, if f is continuous at z_0 .

It is also possible to apply the conjugate duality theory to dual problems of the form (7.33) and (7.35) in an infinite dimensional setting. That is, let \mathcal{X} and \mathcal{Y} be Banach spaces, $\psi : \mathcal{X} \times \mathcal{Y} \to \overline{\mathbb{R}}$ and $\vartheta(y) := \inf_{x \in \mathcal{X}} \psi(x, y)$.

Theorem 7.88. Let \mathcal{X} and \mathcal{Y} be Banach spaces. Suppose that the function $\psi(x, y)$ is proper convex and lower semicontinuous, and that $\vartheta(\bar{y})$ is finite. Then $\vartheta(y)$ is continuous at \bar{y} iff for every y in a neighborhood of \bar{y} , $\vartheta(y) < +\infty$, i.e., $\bar{y} \in \operatorname{int}(\operatorname{dom} \vartheta)$.

If $\vartheta(y)$ is continuous at \bar{y} , then there is no duality gap between the corresponding primal and dual problems and the set of optimal solutions of the dual problem coincides with $\partial \vartheta(\bar{y})$ and is nonempty and weakly^{*} compact.

7.3.2 Paired Locally Convex Topological Vector Spaces

When Banach space Z is not reflexive, the dual Z^{**} of its dual Z^* is larger than Z. This leads to a certain asymmetry between duality frameworks for spaces Z and Z^* discussed in the previous sections. This can be remedied by endowing Z^* with its weak^{*}, rather than strong (norm), topology. In that respect let us consider the following concept of paired spaces.

Definition 7.89. Let \mathcal{X} and \mathcal{Y} be locally convex topological vector spaces and let $\langle \cdot, \cdot \rangle$ be a bilinear form (scalar product) on $\mathcal{X} \times \mathcal{Y}$, i.e., $\langle \cdot, y \rangle$ is a linear functional on \mathcal{X} for every $y \in \mathcal{Y}$ and $\langle x, \cdot \rangle$ is a linear functional on \mathcal{Y} for every $x \in \mathcal{X}$. It is said that \mathcal{X} and \mathcal{Y} have topologies compatible with the pairing $\langle \cdot, \cdot \rangle$, or in short that \mathcal{X} and \mathcal{Y} are paired spaces, if the set of linear continuous functionals on \mathcal{X} coincides with the set $\{\langle \cdot, y \rangle : y \in \mathcal{Y}\}$ and the set of linear continuous functionals on \mathcal{Y} coincides with the set $\{\langle x, \cdot \rangle : x \in \mathcal{X}\}$.

Rather than giving a formal definition of locally convex topological vector spaces, let us note that a Banach space Z endowed with its strong or weak topology is a locally convex topological vector space. Also its dual space Z^* endowed with its weak* topology becomes a locally convex topological vector space. If the space Z is reflexive, then Zand Z^* endowed with their strong or weak topologies and scalar product $\langle \zeta, z \rangle := \zeta(z)$, $\zeta \in Z^*, z \in Z$, become paired spaces. On the other hand, if Z is not reflexive, then Z can be paired with Z^* by endowing Z with its strong topology and Z^* with its weak* topology.

For example, consider the space $\mathcal{Z} = \mathcal{L}_1(\Omega, \mathcal{F}, P)$. Recall that its dual is the space $\mathcal{Z}^* = \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$, and these spaces are not reflexive unless they are finite dimensional. The dual \mathcal{Z}^{**} of $\mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$ is rather complicated (it is formed from finitely additive probability measures). Therefore a common practice is to pair $\mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$ with $\mathcal{L}_1(\Omega, \mathcal{F}, P)$, rather than with its dual \mathcal{Z}^{**} , by using the scalar product

$$\langle \zeta, Z \rangle := \int \zeta(\omega) Z(\omega) dP(\omega), \ \zeta \in \mathcal{L}_1(\Omega, \mathcal{F}, P), \ Z \in \mathcal{L}_\infty(\Omega, \mathcal{F}, P).$$

Now let \mathcal{X} and \mathcal{Y} be paired spaces. For a function $f : \mathcal{X} \to \overline{\mathbb{R}}$ we can define its conjugate

$$f^*(y) := \sup_{x \in \mathcal{X}} \left\{ \langle x, y \rangle - f(x) \right\},\tag{7.272}$$

and its biconjugate

$$f^{**}(x) := \sup_{y \in \mathcal{Y}} \{ \langle x, y \rangle - f^{*}(y) \}.$$
 (7.273)

Then the Fenchel-Moreau Theorem still holds. That is, if f is convex, then $f^{**} = \operatorname{lsc} f$. It should be remembered, of course, that lower semicontinuity here should be taken with respect to the corresponding compatible topologies. It also could be noted that an analogue of Proposition 7.84 holds.

7.3.3 Lattice Structure

Let $C \subset Z$ be a closed convex pointed²³ cone. It defines an order relation on the space Z. That is, $z_1 \succeq z_2$ if $z_1 - z_2 \in C$. It is not difficult to verify that this order relation defines a *partial order* on Z, i.e., the following conditions hold for any $z, z', z'' \in Z$: (i) $z \succeq z$, (ii) if $z \succeq z'$ and $z' \succeq z''$, then $z \succeq z''$ (transitivity), (iii) if $z \succeq z'$ and $z' \succeq z$, then z = z'. This partial order relation is also compatible with the algebraic operations, i.e., the following conditions hold: (iv) if $z \succeq z'$ and $t \ge 0$, then $tz \succeq tz'$, (v) if $z' \succeq z''$ and $z \in Z$, then $z' + z \succeq z'' + z$.

It is said that $u \in \mathbb{Z}$ is the *least upper bound* (or *supremum*) of $z, z' \in \mathbb{Z}$, written $u = z \lor z'$, if $u \succeq z$ and $u \succeq z'$ and, moreover, if $u' \succeq z$ and $u' \succeq z'$ for some $u' \in \mathbb{Z}$, then $u' \succeq u$. By the above property (iii) we have that if the least upper bound $z \lor z'$ exists, then it is unique. It is said that the considered partial order induces a *lattice structure* on \mathbb{Z} if the least upper bound $z \lor z'$ exists for any $z, z' \in \mathbb{Z}$. Denote $z_+ := z \lor 0, z_- := (-z) \lor 0$, and $|z| := z_+ \lor z_- = z \lor (-z)$. It is said that Banach space \mathbb{Z} with lattice structure is a *Banach lattice* if $z, z' \in \mathbb{Z}$ and $|z| \succeq |z'|$ implies $||z|| \ge ||z'||$.

For $p \in [1, \infty]$, consider Banach space $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ and cone $\mathcal{C} := \mathcal{L}_p^+(\Omega, \mathcal{F}, P)$, where

$$\mathcal{L}_{p}^{+}(\Omega, \mathcal{F}, P) := \left\{ z \in \mathcal{L}_{p}(\Omega, \mathcal{F}, P) : z(\omega) \ge 0 \text{ for a.e. } \omega \in \Omega \right\}.$$
(7.274)

This cone C is closed, convex and pointed. The corresponding partial order means that $z \succeq z'$ iff $z(\omega) \ge z'(\omega)$ for a.e. $\omega \in \Omega$. It has a lattice structure with

$$(z \lor z')(\omega) = \max\{z(\omega), z'(\omega)\},\$$

and $|z|(\omega) = |z(\omega)|$. Also the property: "if $z \succeq z' \succeq 0$, then $||z|| \ge ||z'||$ ", clearly holds. It follows that space $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ with cone $\mathcal{L}_p^+(\Omega, \mathcal{F}, P)$ forms a Banach lattice.

Theorem 7.90 (Klee-Nachbin-Namioka). Let Z be a Banach lattice and $\ell : Z \to \mathbb{R}$ be a linear functional. Suppose that ℓ is positive, i.e., $\ell(z) \ge 0$ for any $z \succeq 0$. Then ℓ is continuous.

Proof. We have that linear functional ℓ is continuous iff it is bounded on the unit ball of \mathcal{Z} , i.e, iff there exists positive constant c such that $|\ell(z)| \leq c||z||$ for all $z \in \mathcal{Z}$. First, let us show that there exists c > 0 such that $\ell(z) \leq c||z||$ for all $z \succeq 0$. Recall that $z \succeq 0$ iff $z \in \mathcal{C}$. We argue by a contradiction. Suppose that this is incorrect. Then there exists a sequence $z_k \in \mathcal{C}$ such that $||z_k|| = 1$ and $\ell(z_k) \geq 2^k$ for all $k \in \mathbb{N}$. Consider $\overline{z} := \sum_{k=1}^{\infty} 2^{-k} z_k$. Note that $\sum_{k=1}^{n} 2^{-k} z_k$ forms a Cauchy sequence in \mathcal{Z} and hence is convergent, i.e., the point \overline{z} is well defined. Note also that since \mathcal{C} is closed, it follows that $\sum_{k=m}^{\infty} 2^{-k} z_k \in \mathcal{C}$, and hence it follows by positivity of ℓ that $\ell(\sum_{k=m}^{\infty} 2^{-k} z_k) \geq 0$ for any $m \in \mathbb{N}$. Therefore we have

$$\ell(\bar{z}) = \ell\left(\sum_{k=1}^{n} 2^{-k} z_k\right) + \ell\left(\sum_{k=n+1}^{\infty} 2^{-k} z_k\right) \ge \ell\left(\sum_{k=1}^{n} 2^{-k} z_k\right)$$

= $\sum_{k=1}^{n} 2^{-k} \ell(z_k) \ge n,$

for any $n \in \mathbb{N}$. This gives a contradiction.

²³Recall that cone C is said to be *pointed* if $z \in C$ and $-z \in C$ implies that z = 0.

Now for any $z \in \mathcal{Z}$ we have

$$|z| = z_+ \lor z_- \succeq z_+ = |z_+|.$$

It follows that for v = |z| we have that $||v|| \ge ||z_+||$, and similarly $||v|| \ge ||z_-||$. Since $z = z_+ - z_-$ and $\ell(z_-) \ge 0$ by positivity of ℓ , it follows that

$$\ell(z) = \ell(z_+) - \ell(z_-) \le \ell(z_+) \le c ||z_+|| \le c ||z||,$$

and similarly

$$-\ell(z) = -\ell(z_{+}) + \ell(z_{-}) \le \ell(z_{-}) \le c ||z_{-}|| \le c ||z||.$$

It follows that $|\ell(z)| \leq c ||z||$, which completes the proof. \Box

Suppose that Banach space Z has a lattice structure. It is said that a function $f : Z \to \overline{\mathbb{R}}$ is *monotone* if $z \succeq z'$ implies that $f(z) \ge f(z')$.

Theorem 7.91. Let \mathcal{Z} be a Banach lattice and $f : \mathcal{Z} \to \overline{\mathbb{R}}$ be proper convex and monotone. Then $f(\cdot)$ is continuous and subdifferentiable on the interior of its domain.

Proof. Let $z_0 \in \text{int}(\text{dom} f)$. By Proposition 7.86, function f possesses an algebraic subgradient ℓ at z_0 . It follows from monotonicity of f that ℓ is positive. Indeed, if $\ell(h) < 0$ for some $h \in C$, then it follows by (7.267) that

$$f(z_0 - h) \ge f(z_0) - \ell(h) > f(z_0),$$

which contradicts monotonicity of f. It follows by Theorem 7.90 that ℓ is continuous, and hence $\ell \in \partial f(z_0)$. This shows that f is subdifferentiable at every point of int(dom f). This, in turn, implies that f is lower semicontinuous on int(dom f), and hence by Proposition 7.83 is continuous on int(dom f). \Box

The above result can be applied to any space $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P), p \in [1, \infty]$, equipped with the lattice structure induced by the cone $\mathcal{C} := \mathcal{L}_p^+(\Omega, \mathcal{F}, P)$.

7.3.4 Interchangeability Principle

Let (Ω, \mathcal{F}, P) be a probability space and $f : \mathbb{R}^m \times \Omega \to \overline{\mathbb{R}}$ be an extended real valued function. Consider function $\vartheta(\omega) := \inf_{x \in \mathbb{R}^n} f(x, \omega)$. Suppose for the moment that for a.e. $\omega \in \Omega$ the minimum is attained at $\overline{x} = \overline{x}(\omega)$, i.e., $\overline{x}(\omega) \in \arg \min_{x \in \mathbb{R}^n} f(x, \omega)$. Then $\vartheta(\omega) = f(\overline{x}(\omega), \omega)$ and hence $\vartheta(\omega)$ is finite valued. If we assume further that the function $f(x, \omega)$ is random lower semicontinuous, then $\vartheta(\omega)$ is measurable (see Theorem 7.42) and hence ϑ can be viewed as a random variable. Also $\overline{x}(\omega)$ can be selected to be measurable. It follows that

$$\mathbb{E}[\vartheta(\omega)] = \mathbb{E}[f(\bar{x}(\omega), \omega)] \ge \inf_{x(\cdot)} \mathbb{E}[f(x(\omega), \omega)],$$
(7.275)

where the minimization in the right hand side of (7.275) is performed over an appropriate space of measurable mappings $x(\cdot) : \Omega \to \mathbb{R}^n$. Conversely, we have that $f(x(\omega), \omega) \ge$ $\vartheta(\omega)$, and hence $\mathbb{E}[\vartheta(\omega)] = \inf_{x(\cdot)} \mathbb{E}[f(x(\omega), \omega)]$. That is, in the above sense the expectation and minimization operators can be interchanged. More accurately we have the following result (taken from [217, Theorem 14.60]).

It is said that a linear space \mathfrak{M} of \mathcal{F} -measurable functions (mappings) $\psi : \Omega \to \mathbb{R}^m$ is *decomposable* if for every $\psi \in \mathfrak{M}$ and $A \in \mathcal{F}$, and every bounded and \mathcal{F} -measurable function $\gamma : \Omega \to \mathbb{R}^m$, the space \mathfrak{M} also contains the function $\eta(\cdot) := \mathbf{1}_{\Omega \setminus A}(\cdot)\psi(\cdot) + \mathbf{1}_A(\cdot)\gamma(\cdot)$. For example, spaces $\mathfrak{M} := \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^m)$, with $p \in [1, \infty]$, are decomposable.

Theorem 7.92. Let \mathfrak{M} be a decomposable space and $f : \mathbb{R}^m \times \Omega \to \overline{\mathbb{R}}$ be a random lower semicontinuous function. Then

$$\mathbb{E}\left[\inf_{x\in\mathbb{R}^m} f(x,\omega)\right] = \inf_{\chi\in\mathfrak{M}} \mathbb{E}[F_{\chi}],\tag{7.276}$$

where $F_{\chi}(\omega) := f(\chi(\omega), \omega)$, provided that the right hand side of (7.276) is less than $+\infty$. Moreover, if the common value of both sides in (7.276) is not $-\infty$, then

$$\bar{\chi} \in \underset{\chi \in \mathfrak{M}}{\operatorname{argmin}} \mathbb{E}[F_{\chi}] \quad \text{iff} \quad \bar{\chi}(\omega) \in \underset{x \in \mathbb{R}^m}{\operatorname{argmin}} f(x, \omega) \text{ for a.e. } \omega \in \Omega, \text{ and } \bar{\chi} \in \mathfrak{M}.$$
(7.277)

Clearly the above interchangeability principle can be applied to a maximization, rather than minimization, procedure simply by replacing function $f(x, \omega)$ with $-f(x, \omega)$. For an extension of this interchangeability principle to risk measures see Proposition 6.60.

Exercises

- 7.1. Show that function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is lsc iff its epigraph epif is a closed subset of \mathbb{R}^{n+1} .
- 7.2. Show that a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is polyhedral iff its epigraph is a convex closed polyhedron, and f(x) is finite for at least one x.
- 7.3. Give an example of a function $f : \mathbb{R}^2 \to \mathbb{R}$ which is Gâteaux but not Fréchet differentiable.
- 7.4. Show that if $g : \mathbb{R}^n \to \mathbb{R}^m$ is Hadamard directionally differentiable at $x_0 \in \mathbb{R}^n$, then $g'(x_0, \cdot)$ is continuous and g is Fréchet directionally differentiable at x_0 . Conversely, if g is Fréchet directionally differentiable at x_0 and $g'(x_0, \cdot)$ is continuous, then g is Hadamard directionally differentiable at x_0 .
- 7.5. Show that if $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a convex function, finite valued at a point $x_0 \in \mathbb{R}^n$, then formula (7.17) holds and $f'(x_0, \cdot)$ is convex. If, moreover, $f(\cdot)$ is finite valued in a neighborhood of x_0 , then $f'(x_0, h)$ is finite valued for all $h \in \mathbb{R}^n$.
- 7.6. Let S(·) be the support function of a nonempty set C ⊂ ℝⁿ. Show that the conjugate of S(·) is the indicator function of the set cl(conv(C)).
- 7.7. Let $C \subset \mathbb{R}^n$ be a closed convex set and $x \in C$. Show that the normal cone $\mathcal{N}_C(x)$ is equal to the subdifferential of the indicator function $\mathbb{I}_C(\cdot)$ at x.

- 7.8. Let \mathcal{X} and \mathcal{Y} be vector spaces and $\psi : \mathcal{X} \times \mathcal{Y} \to \overline{\mathbb{R}}$ be a convex function. Show that the function $\vartheta(y) := \inf_{x \in \mathcal{X}} \psi(x, y)$ is convex.
- 7.9. Show that if multifunction $\mathcal{G} : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is closed valued and upper semicontinuous, then it is closed. Conversely, if \mathcal{G} is closed and the set dom \mathcal{G} is compact, then \mathcal{G} is upper semicontinuous.
- 7.10. Give an example of a sequence $f_k : [0,1] \to \mathbb{R}$ of continuous functions epiconverging to a continuous function $f : [0,1] \to \mathbb{R}$ and such that $\sup_{x \in [0,1]} |f_k(x) - f(x)| =$ 1, i.e., f_k does not converge to f uniformly on the interval [0,1].
- 7.11. Consider function $F(x, \omega)$ used in Theorem 7.49. Show that if $F(\cdot, \omega)$ is differentiable for a.e. ω , then condition (A2) of that theorem is equivalent to the condition: there exists a neighborhood V of x_0 such that

$$\mathbb{E}\left[\sup_{x\in V} \|\nabla_x F(x,\omega)\|\right] < \infty.$$
(7.278)

- 7.12. Show that if $f(x) := \mathbb{E}|x \xi|$, then formula (7.129) holds. Conclude that $f(\cdot)$ is differentiable at $x_0 \in \mathbb{R}$ iff $\Pr(\xi = x_0) = 0$.
- 7.13. Verify equalities (7.168) and (7.169), and hence conclude (7.170).
- 7.14. Show that the estimate (7.224) of Theorem 7.73 still holds if the bound (7.223) in condition (C2) is replaced by:

$$|F(x',\xi) - F(x,\xi)| \le \kappa(\xi) ||x' - x||^{\gamma}$$
(7.279)

for some constant $\gamma > 0$. Show how the estimate (7.242) of Theorem 7.75 should be corrected in that case.

Chapter 8 Bibliographical Remarks

Chapter 1

The Newsvendor (sometimes called Newsboy) problem, portfolio selection and supply chain models are classical with numerous papers written on each subject. It will be far beyond this monograph to give a complete review of all relevant literature. Our main purpose in discussing these models is to introduce such basic concepts as a recourse action, probabilistic (chance) constraints, here-and-now and wait-and-see solutions, the nonanticipativity principle, dynamic programming equations etc. We give below just a few basic references.

The Newsvendor Problem is a classical model used in Inventory Management. Its origin is going back to the paper by Edgeworth [73]. In the stochastic setting studying of the Newsvendor Problem started with the classical paper by Arrow, Harris and Marchak [5]. The optimality of the basestock policy for the multi-stage inventory model was first proved in Clark and Scarf [44]. The worst case distribution approach to the Newsvendor Problem was initiated by Scarf [231], where the case when only the mean and variance of the distribution of the demand are known was analyzed. For a thorough discussion and relevant references for single and multi-stage inventory models we can refer to Zipkin [276].

Modern portfolio theory was introduced by Markowitz [151, 152]. The concept of utility function has a long history. Its origins are going back as far as to work of Daniel Bernoulli (1738). Axiomatic approach to the expected utility theory was introduced by von Neumann and Morgenstern [266].

For an introduction to Supply Chain Network Design see, e.g., Nagurney [160]. The material of section 1.5 is based on Santoso et al [229].

For a thorough discussion of robust optimization we refer to the book by Ben-Tal, El Ghaoui and Nemirovski [16].

Chapters 2 and 3

Stochastic programming with recourse originated in the works of Beale [14], Dantzig [48] and Tintner [260].

Properties of the optimal value $Q(x, \xi)$ of the second stage linear programming problem and of its expectation $\mathbb{E}[Q(x,\xi)]$ were first studied by Kall [120, 121], Walkup and Wets [268, 269], and Wets [271, 272]. Example 2.5 is discussed in Birge and Louveaux [22]. Polyhedral and convex two-stage problems, discussed in sections 2.2 and 2.3, are natural extensions of the linear two-stage problems. Many additional examples and analysis of particular models can be found in Birge and Louveaux [22], Kall and Wallace [123], Wallace and Ziemba [270]. For a thorough analysis of simple recourse models, see Kall and Mayer [122].

Duality analysis of stochastic problems, and in particular dualization of the nonanticipativity constraints, was developed by Eisner and Olsen [75], Wets [273] and Rockafellar and Wets [215, 216] (see also Rockafellar [212] and the references therein).

Expected value of perfect information is a classical concept in decision theory (see, Raiffa and Schlaifer [201], Raiffa [200]). In stochastic programming this and related concepts were analyzed by Madansky [148], Spivey [255], Avriel and Williams [11], Dempster [53], Huang, Vertinsky and Ziemba [113].

Numerical methods for solving two- and multi-stage stochastic programming problems are extensively discussed in Birge and Louveaux [22], Ruszczyński [222], Kall and Mayer [122], where the Reader can also find detailed references to original contributions.

There is also an extensive literature on constructing scenario trees for multistage models, encompassing various techniques using probability metrics, pseudo-random sequences, lower and upper bounding trees, and moment matching. The Reader is referred to Kall and Mayer [122], Heitsch and Römisch [98], Hochreiter and Pflug [109], Casey and Sen [35], Pennanen [176], Dupačova, Growe-Kuska and Römisch [71], and the references therein.

Linear (affine) decision rules were already discussed quite sometime ago (see Garstka and Wets [83] and references therein). For a more recent discussion, in the framework of stochastic programming, we can refer to Shapiro and Nemirovski [249], and Georghiou, Wiesemann and Kuhn [84].

An extensive stochastic programming bibliography can be found at the web site: http://mally.eco.rug.nl/spbib.html, maintained by Maarten van der Vlerk.

Chapter 4

Models involving probabilistic (chance) constraints were introduced by Charnes, Cooper and Symonds [36], Miller and Wagner [156], and Prékopa [190]. Problems with integrated chance constraints are considered in [95]. Models with stochastic dominance constraints are introduced and analyzed by Dentcheva and Ruszczyński in [62, 64, 65]. The notion of stochastic ordering or stochastic dominance of first order has been introduced in statistics in Mann and Whitney [150], and Lehmann [142], and further applied and developed in economics (see Quirk and Saposnik [199], Fishburn [77] and Hadar and Russell [94].)

An essential contribution to the theory and solutions of problems with chance constraints was the theory of α -concave measures and functions. In Prékopa [191, 192] the concept of logarithmic concave measures was introduced and studied. This notion was generalized to α -concave measures and functions in Borell [27, 28], Brascamp and Lieb [30], and Rinott [204], and further analyzed in Tamm [259], and Norkin [171]. Approximations of the probability function by Steklov-Sobolev transformation was suggested

by Norkin in [169]. Differentiability properties of probability functions are studied in Uryasev [261, 262], Kibzun and Tretyakov [126], Kibzun and Uryasev [127], and Raik [202]. The first definition of α -concave discrete multivariate distributions was introduced in Barndorff–Nielsen [13]. The generalized definition of α -concave functions on a set, which we have adopted here, was introduced in Dentcheva, Prekopa, and Ruszczyński [58]. It facilitates the development of optimality and duality theory of probabilistic optimization. Its consequences for probabilistic optimization are explored in Dentcheva, Prékopa, and Ruszczyński [60]. The notion of *p*-efficient points was first introduced in Prékopa [193]. Similar concept is used in Sen [233]. The concept was studied and applied in the context of discrete distributions and linear problems in the papers Dentcheva, Prékopa, and Ruszczyński [58, 60] and Prékopa and B. Vízvári and T. Badics [196]. Optimization problems with probabilistic set-covering constraint are investigated in Beraldi and Ruszczyński [17, 18], where efficient enumeration procedures of p-efficient points of 0-1 variable are employed. Optimality conditions based on the concept of *p*-efficient points for linear problems with discrete distributions are developed in are developed in Dentcheva, Prékopa, and Ruszczyński [58]. Optimality conditions for non-linear convex problems with chance constraints involving discrete distributions are developed in Dentcheva, Prékopa, and Ruszczyński [59] and further generalized for arbitrary distributions in Dentcheva, Lai, and Ruszczyński [55]. The duality result is presented first in Dentcheva and Martinez [57], where numerical methods with regularization for solving problems of this type are developed. The description of the tangent and the normal cone of the probabilistically constrained sets given in section 4.3.4 is a slightly modified version of the ones given in Dentcheva and Martinez [56], where also the example 4.65 is included. The optimality conditions under differentiability assumptions presented in section 4.3.4 are based on the conditions developed in Dentcheva and Martinez [56]. For further information related to differentiability of probability functions in the context of chance constraints, we refer to Henrion and Möller [102], where formulae for the gradient of the probability function of the multivariate normal distribution for events described by linear constraints are developed. Additionally, lower estimates for the norm of gradients of Gaussian distribution functions are given in Henrion [101]. It is shown how the precision of computing gradients in probabilistically constrained optimization problems can be controlled by the precision of function values for Gaussian distribution functions. There is a wealth of research on estimating probabilities of events. We refer to Boros and Prékopa [29], Bukszár [31], Bukszár and Prékopa [32], Dentcheva, Prékopa, Ruszczynski [60], Prékopa [194], and Szántai [257], where probability bounds are used in the context of chance constraints.

Statistical approximations of probabilistically constrained problems were analyzed in Salinetti [228], Kankova [124], Deák [50], and Gröwe [90]. Stability of models with probabilistic constraints is addressed in Dentcheva [54], Henrion [100, 99], and Henrion and Römisch [219, 103].

Many applied models in engineering, where reliability is frequently a central issue (e.g., in telecommunication, transportation, hydrological network design and operation, engineering structure design, electronic manufacturing problems, etc.) include optimization under probabilistic constraints. We do not list these applied works here. In finance, the concept of Value-at-Risk enjoys great popularity (see, e.g., Dowd [68], Pflug [181], Pflug and Römisch [182]). The concept of stochastic dominance plays a fundamental role in economics and statistics. We refer to Mosler and Scarsini [159], Shaked and Shanthikumar [234], and Szekli [258] for more information and a general overview on stochastic orders.

Chapter 5

The concept of SAA estimators is closely related to the Maximum Likelihood (ML) method and M-estimators developed in Statistics literature. However, the motivation and scope of applications are quite different. In Statistics the involved constraints typically are of a simple nature and do not play such an essential role as in stochastic programming. Also in applications of Monte Carlo sampling techniques to stochastic programming the respective sample is generated in the computer and its size can be controlled, while in statistical applications the data are typically given and cannot be easily changed.

Starting with a pioneering work of Wald [267], consistency properties of the Maximum Likelihood and M-estimators were studied in numerous publications. Epi-convergence approach to studying consistency of statistical estimators was discussed in Dupačová and Wets [70]. In the context of stochastic programming, consistency of SAA estimators was also investigated by tools of epi-convergence analysis in King and Wets [130] and Robinson [209].

Proposition 5.6 appeared in Norkin, Pflug and Ruszczyński [170] and Mak, Morton and Wood [149]. Theorems 5.7, 5.11 and 5.10 are taken from Shapiro [236] and [242], respectively. The approach to second order asymptotics, discussed in section 5.1.3, is based on Dentcheva and Römisch [61] and Shapiro [237]. Starting with the classical asymptotic theory of the ML method, asymptotics of statistical estimators were investigated in numerous publications. Asymptotic normality of *M*-estimators was proved, under quite weak differentiability assumptions, in Huber [114]. An extension of the SAA method to stochastic generalized equations is a natural one. Stochastic variational inequalities were discussed by Gürkan, Özge and Robinson [93]. Proposition 5.14 and Theorem 5.15 are similar to the results obtained in [93, Theorems 1 and 2]. Asymptotics of SAA estimators of optimal solutions of stochastic programs are discussed in King and Rockafellar [129] and Shapiro [235].

The idea of using Monte Carlo sampling for solving stochastic optimization problems of the form (5.1) certainly is not new. There is a variety of sampling based optimization techniques which were suggested in the literature. It will be beyond the scope of this chapter to give a comprehensive survey of these methods, we mention a few approaches related to the material of this chapter. One approach uses the Infinitesimal Perturbation Analysis (IPA) techniques to estimate the gradients of $f(\cdot)$, which consequently are employed in the Stochastic Approximation (SA) method. For a discussion of the IPA and SA methods we refer to Ho and Cao [108], Glasserman [88], and Kushner and Clark [138], Nevelson and Hasminskii [166], respectively. For an application of this approach to optimization of queueing systems see Chong and Ramadge [43] and L'Ecuyer and Glynn [141], for example. Closely related to this approach is the Stochastic Quasi-Gradient method (see Ermoliev [76]).

Another class of methods uses sample average estimates of the values of the objective function, and may be its gradients (subgradients), in an "interior" fashion. Such methods are aimed at solving the true problem (5.1) by employing sampling estimates of $f(\cdot)$ and $\nabla f(\cdot)$ blended into a particular optimization algorithm. Typically, the sample is updated or a different sample is used each time function or gradient (subgradient) estimates are

required at a current iteration point. In this respect we can mention, in particular, the statistical L-shaped method of Infanger [115] and the stochastic decomposition method of Higle and Sen [106].

In this chapter we mainly discussed an "exterior" approach, in which a sample is generated outside of an optimization procedure and consequently the constructed sample average approximation (SAA) problem is solved by an appropriate deterministic optimization algorithm. There are several advantages in such approach. The method separates sampling procedures and optimization techniques. This makes it easy to implement and, in a sense, universal. From the optimization point of view, given a sample $\xi^1, ..., \xi^N$, the obtained optimization problem can be considered as a stochastic program with the associated scenarios $\xi^1, ..., \xi^N$, each taken with equal probability N^{-1} . Therefore any optimization algorithm which is developed for a considered class of stochastic programs can be applied to the constructed SAA problem in a straightforward way. Also the method is ideally suited for a parallel implementation. From the theoretical point of view there is available a quite well developed statistical inference of the SAA method. This, in turn, gives a possibility of error estimation, validation analysis and hence stopping rules. Finally, various variance reduction techniques can be conveniently combined with the SAA method.

It is difficult to point out an exact origin of the SAA method. The idea is simple indeed and it was used by various authors under different names. Variants of this approach are known as the stochastic counterpart method (Rubinstein and Shapiro [220],[221]) and sample-path optimization (Plambeck, Fu, Robinson and Suri [187] and Robinson [209]), for example. Also similar ideas were used in statistics for computing maximum likelihood estimators by Monte Carlo techniques based on Gibbs sampling (see, e.g., Geyer and Thompson [85] and references therein). Numerical experiments with the SAA approach, applied to linear and discrete (integer) stochastic programming problems, can be also found in more recent publications [3],[146],[149],[265].

The complexity analysis of the SAA method, discussed in section 5.3, is motivated by the following observations. Suppose for the moment that components ξ_i , i = 1, ..., d, of the random data vector $\xi \in \mathbb{R}^d$ are independently distributed. Suppose, further, that we use r points for discretization of the (marginal) probability distribution of each component ξ_i . Then the resulting number of scenarios is $K = r^d$, i.e., it grows exponentially with increase of the number of random parameters. Already with, say, r = 4 and d = 20we will have an astronomically large number of scenarios $4^{20} \approx 10^{12}$. In such situations it seems hopeless just to calculate with a high accuracy the value $f(x) = \mathbb{E}[F(x,\xi)]$ of the objective function at a given point $x \in \mathcal{X}$, much less to solve the corresponding optimization problem²⁴. And, indeed, it is shown in Dyer and Stougie [72] that, under the assumption that the stochastic parameters are independently distributed, two-stage linear stochastic programming problems are #P-hard. This indicates that, in general, two-stage stochastic programming problems cannot be solved with a high accuracy, as say with accuracy of order 10^{-3} or 10^{-4} , as it is common in deterministic optimization. On the other hand, quite often in applications it does not make much sense to try to solve the corresponding stochastic problem with a high accuracy since the involved inaccuracies resulting from

²⁴Of course, in some very specific situations it is possible to calculate $\mathbb{E}[F(x,\xi)]$ in a closed form. Also if $F(x,\xi)$ is decomposable into the sum $\sum_{i=1}^{d} F_i(x,\xi_i)$, then $\mathbb{E}[F(x,\xi)] = \sum_{i=1}^{d} \mathbb{E}[F_i(x,\xi_i)]$ and hence the problem is reduced to calculations of one dimensional integrals. This happens in the case of the so-called simple recourse.

inexact modeling, distribution approximations etc, could be far bigger. In some situations the randomization approach based on Monte Carlo sampling techniques allows to solve stochastic programs with a reasonable accuracy and a reasonable computational effort.

The material of section 5.3.1 is based on Kleywegt, Shapiro and Homem-De-Mello [131]. The extension of that analysis to general feasible sets, given in section 5.3.2, is discussed in Shapiro [238, 240, 243] and Shapiro and Nemirovski [249]. The material of section 5.3.3 is based on Shapiro and Homem-de-Mello [251], where proof of Theorem 5.24 can be found.

In practical applications, in order to speed up the convergence, it is often advantageous to use Quasi-Monte Carlo techniques. Theoretical bounds for the error of numerical integration by Quasi-Monte Carlo methods are proportional to $(\log N)^d N^{-1}$, i.e., are of order $O((\log N)^d N^{-1})$, with the respective proportionality constant A_d depending on d. For small d it is "almost" the same as of order $O(N^{-1})$, which of course is better than $O_p(N^{-1/2})$. However, the theoretical constant A_d grows superexponentially with increase of d. Therefore, for larger values of d one often needs a very large sample size N for Quasi-Monte Carlo methods to become advantageous. It is beyond the scope of this chapter to give a thorough discussion of Quasi-Monte Carlo methods. A brief discussion of Quasi-Monte Carlo techniques is given in section 5.4. For a further readings on that topic we may refer to Niederreiter [167]. For applications of Quasi-Monte Carlo techniques to stochastic programming see, e.g., Koivu [133], Homem-de-Mello [112], Pennanen and Koivu [177], Heitsch, Leövey and Römisch [97].

For a discussion of variance reduction techniques in Monte Carlo sampling we refer to Fishman [78] and a survey paper by Avramidis and Wilson [10], for example. In the context of stochastic programming, variance reduction techniques were discussed in Rubinstein and Shapiro [221], Dantzig and Infanger [49], Higle [104] and Bailey, Jensen and Morton [12], for example.

The statistical bounds of section 5.6.1 were suggested in Norkin, Pflug and Ruszczyński [170], and developed in Mak, Morton and Wood [149]. The "common random numbers" estimator $\widehat{gap}_{N,M}(\overline{x})$ of the optimality gap was introduced in [149]. The KKT statistical test, discussed in section 5.6.2, was developed in Shapiro and Homem-de-Mello [250], so that the material of that section is based on [250]. See also Higle and Sen [105].

The estimate of the sample size derived in Theorem 5.32 is due to Campi and Garatti [34]. This result builds on a previous work of Calafiore and Campi [33], and from the considered point of view gives a tightest possible estimate of the required sample size. Construction of upper and lower statistical bounds for chance constrained problems, discussed in section 5.7, is based on Nemirovski and Shapiro [162]. For some numerical experiments with these bounds see Luedtke and Ahmed [147].

The extension of the SAA method to multistage stochastic programming, discussed in section 5.8 and referred to as conditional sampling, is a natural one. A discussion of consistency of conditional sampling estimators is given, e.g., in Shapiro [239]. Discussion of the portfolio selection (Example 5.34) is based on Blomvall and Shapiro [24]. Complexity of the SAA approach to multistage programming is discussed in Shapiro and Nemirovski [249] and Shapiro [241].

Section 5.9 is based on Nemirovski et al [161]. The origins of the Stochastic Approximation algorithms are going back to the pioneering paper by Robbins and Monro [205]. For a thorough discussion of the asymptotic theory of the SA method we can refer to Kushner and Clark [138] and Nevelson and Hasminskii [166]. The robust SA approach was developed in Polyak [188] and Polyak and Juditsky [189]. The main ingredients of Polyak's scheme (long steps and averaging) were, in a different form, proposed in Nemirovski and Yudin [163].

The SDDP algorithm, discussed in section 5.10, was introduced by Pereira and Pinto [178] and became a popular method for scheduling of hydro-thermal electricity systems. Almost sure convergence of the SDDP algorithm is discussed in Philpott and Guan [185] and Girardeau, Leclere and Philpott [87]. Statistical analysis of the SDDP method is presented in [245]. For risk measures given by convex combinations of the expectation and Average Value-at-Risk, it was shown in Shapiro [245] how to incorporate this approach into the SDDP algorithm with a little additional effort. This was implemented in an extensive numerical study in Philpott and de Matos [184]. For additional developments and extensive numerical studies we refer to [252], [253]. For a recent study of the SDDP method we can also refer to Guigues [92] and references therein.

Chapter 6

Foundations of the expected utility theory were developed in von Neumann and Morgenstern [266]. The dual utility theory is developed in Quiggin [197, 198] and Yaari [275].

The mean–variance model was introduced and analyzed in Markowitz [151, 152, 153]. Deviations and semideviations in mean–risk analysis were analyzed in Kijima and Ohnishi [128], Konno [134], Ogryczak and Ruszczyński [173, 174], Ruszczyński and Vanderbei [227]. Weighted deviations from quantiles, relations to stochastic dominance and Lorenz curves are discussed in Ogryczak and Ruszczyński [175]. For Conditional (Average) Value at Risk see Acerbi and Tasche [1], Rockafellar and Uryasev [213], Pflug [181]. A general class of convex approximations of chance constraints is developed in Nemirovski and Shapiro [162].

The theory of coherent measures of risk was initiated in Artzner, Delbaen, Eber and Heath [8] and further developed, *inter alia*, by Delbaen [51], Föllmer and Schied [80], Leitner [143], Rockafellar, Uryasev and Zabarankin [214]. Our presentation is based on Ruszczyński and Shapiro [224, 226]. Kusuoka representation of law invariant coherent risk measures (Theorem 6.40) was derived in [139] for $\mathcal{L}_{\infty}(\Omega, \mathcal{F}, P)$ spaces. For an extension to $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ spaces see, e.g., Pflug and Römisch [182]. Our presentation is based on Shapiro [248] and Pichler and Shapiro [186]. Theory of consistency with stochastic orders was initiated in [173] and developed in [174, 175]. In section 6.3.6 we follow [186]. Similar definition of regularity of risk measures was given in Noyan and Rudolf [172]. Example 6.54 of section 6.4 is motivated by Jiang and Guan [118].

An alternative approach to asymptotic analysis of law invariant coherent risk measures (see section 6.6.3), was developed in Pflug and Wozabal [180] based on Kusuoka representation. Application to portfolio optimization is discussed in Miller and Ruszczyński [157].

The theory of conditional risk mappings was initiated by Scandolo [230] and further developed by Riedel [203], Detlefsen and Scandolo [67], Ruszczyński and Shapiro [224, 225]. The theory of dynamic measures of risk in discrete time was developed in Artzner, Delbaen, Eber, Heath and Ku [9], Cheridito, Delbaen and Kupper [37], Detlefsen and Scandolo [67], Eichhorn and Römisch [74], Pflug and Römisch [182], Pflug and Ruszczyński [179].

Much work on time-consistency of dynamic measures of risk focused on the case when we have just one final cost and we are evaluating it from the perspective of earlier stages. In the finance literature time consistency is usually defined via the composition property of dynamic risk measures; see, Artzner, Delbaen, Eber, Heath and Ku [9], Bion-Nadal [20, 21], Cheridito, Delbaen and Kupper [38], Cheridito and Kupper [39], Föllmer and Penner [79], Frittelli and Scandolo [81], Frittelli and Rosazza Gianin [82], Klöppel and Schweizer [132]. Our presentation in section 6.8.3 is based on Ruszczyński [223], where the composition property is derived from simpler assumptions. It is worth noting that time consistency of utility models was already discussed by Koopmans [135] and Kreps and Porteus [136, 137].

The principle of conditional optimality in 6.8.5 is discussed in Shapiro [244], Example 6.84 is taken from that paper. Two-stage risk-averse optimization is discussed in Miller and Ruszczyński [158]. An approach is to define time-consistency of optimization models directly through the dynamic programming principle was initiated by Boda and Filar [25] for a portfolio problem. Most of the presentation in section 6.8.5 is original. The minimax approach of section 6.8.6 is related to robust dynamic programming discussed in Iyengar [117] and Nilim and El Ghaoui [168].

Our inventory example is based on Ahmed, Cakmak and Shapiro [2]. For related risk-averse inventory models, see Choi and Ruszczyński [41] and Choi, Ruszczyński and Zhao [42].

Chapter 7

There are many monographs where concepts of directional differentiability are discussed in detail, see, e.g., [26]. A thorough discussion of Clarke generalized gradient and regularity in the sense of Clarke can be found in Clarke [45]. Classical references on (finite dimensional) convex analysis are books by Rockafellar [210] and Hiriart-Urruty and Lemaréchal [107]. For a proof of Fenchel-Moreau Theorem (in an infinite dimensional setting) see, e.g., [211]. For a development of conjugate duality (in an infinite dimensional setting) we refer to Rockafellar [211]. Theorem 7.12 (Hoffman's Lemma) appeared in [111].

Theorem 7.25 appeared in Danskin [47]. Theorem 7.26 is going back to Levin [144] and Valadier [263] (see also Ioffe and Tihomirov [116, p. 213]). For a general discussion of second order optimality conditions and perturbation analysis of optimization problems we can refer to Bonnans and Shapiro [26] and references therein. Theorem 7.28 is an adaptation of a result going back to Gol'shtein [89]. For a thorough discussion of epiconvergence we refer to Rockafellar and Wets [217]. Theorem 7.31 is taken from [217, Theorem 7.17].

There are many books on probability theory. Of course, it will be beyond the scope of this monograph to give a thorough development of that theory. In that respect we can mention the excellent book by Billingsley [19]. Theorem 7.37 appeared in Rogosinski [218]. A thorough discussion of measurable multifunctions and random lower semicontinuous functions can be found in Rockafellar and Wets [217, Chapter 14], to which the interested reader is referred for a further reading. For a proof of Aumann and Lyapunov Theorems (Theorems 7.45 and 7.46) see, e.g., [116, section 8.2].

Theorem 7.52 originated in Strassen [256], where the interchangeability of the subdifferential and integral operators was shown in the case when the expectation function is continuous. The present formulation of Theorem 7.52 is taken from [116, Theorem 4, p. 351].

Uniform Laws of Large Numbers take their origin in the Glivenko-Cantelli Theorem. For a further discussion of the uniform LLN we refer to van der Vaart and Welner [264]. Epi-convergence Law of Large Numbers (Theorem 7.56) is due to Artstein and Wets [7]. The uniform convergence w.p.1 of Clarke generalized gradients, specified in part (c) of Theorem 7.57, was obtained in [235]. The material of section 7.2.6, and in particular Theorem 7.60, is taken from Shapiro [247]. Law of Large Numbers for random sets (Theorem 7.61) appeared in Artstein and Vitale [6]. The uniform convergence of ε -subdifferentials (Theorem 7.63) was derived in [254].

Finite dimensional Delta Method is well known and routinely used in theoretical Statistics. The infinite dimensional version (Theorem 7.67) is going back to Grübel [91], Gill [86] and King [129]. The tangential version (Theorem 7.69) appeared in [236].

There is a large literature on Large Deviations Theory (see, e.g., a book by Dembo and Zeitouni [52]). Hoeffding inequality appeared in [110] and Chernoff inequality in [40]. Theorem 7.76, about interchangeability of Clarke generalized gradient and integral operators, can be derived by using the interchangeability formula (7.125) for directional derivatives, Strassen's Theorem 7.52 and the fact that in the Clarke regular case the directional derivative is the support function of the corresponding Clarke generalized gradient (see [45] for details).

A classical reference for functional analysis is Dunford and Schwartz [69]. The concept of algebraic subgradient and Theorem 7.90 are taken from Levin [145] (unfortunately this excellent book was not translated from Russian). Theorem 7.91 is from Ruszczyński and Shapiro [226]. The interchangeability principle (Theorem 7.92) is taken from [217, Theorem 14.60]. Similar results can be found in [116, Proposition 2, p.340] and [145, Theorem 0.9].

Bibliography

- C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance*, 26:14911507, 2002.
- [2] S. Ahmed, U. Cakmak, and A. Shapiro. Coherent risk measures in inventory problems. *European Journal of Operational Research*, 182:226–238, 2007.
- [3] S. Ahmed and A. Shapiro. The sample average approximation method for stochastic programs with integer recourse. *E-print available at* http://www.optimization-online.org, 2002.
- [4] A. Araujo and E. Giné. *The Central Limit Theorem for Real and Banach Valued Random Variables*. Wiley, New York, 1980.
- [5] K. Arrow, T. Harris, and J. Marshack. Optimal inventory policy. *Econometrica*, 19:250–272, 1951.
- [6] Z. Artstein and R.A. Vitale. A strong law of large numbers for random compact sets. *The Annals of Probability*, 3:879–882, 1975.
- [7] Z. Artstein and R.J.B. Wets. Consistency of minimizers and the slln for stochastic programs. *Journal of Convex Analysis*, 2:1–17, 1996.
- [8] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.
- [9] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, and Ku H. Coherent multiperiod risk adjusted values and bellmans principle. *Annals of Operations Research*, 152:5–22, 2007.
- [10] A.N. Avramidis and J.R. Wilson. Integrated variance reduction strategies for simulation. Operations Research, 44:327–346, 1996.
- [11] M. Avriel and A. Williams. The value of information and stochastic programming. Operations Research, 18:947–954, 1970.
- [12] T.G. Bailey, P. Jensen, and D.P. Morton. Response surface analysis of two-stage stochastic linear programming with recourse. *Naval Research Logistics*, 46:753– 778, 1999.

- [13] O. Barndorff-Nielsen. Unimodality and exponential families. *Comm. Statistics*, 1:189–216, 1973.
- [14] E. M. L. Beale. On minimizing a convex function subject to linear inequalities. *Journal of the Royal Statistical Society, Series B*, 17:173–184, 1955.
- [15] R. E. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.
- [16] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.
- [17] P. Beraldi and A. Ruszczyński. The probabilistic set covering problem. Operations Research, 50:956–967, 1999.
- [18] P. Beraldi and A. Ruszczyński. A branch and bound method for stochastic integer problems under probabilistic constraints. *Optimization Methods and Software*, 17:359–382, 2002.
- [19] P. Billingsley. Probability and Measure. John Wiley & Sons, New York, 1995.
- [20] J. Bion-Nadal. Dynamic risk measures: time consistency and risk measures from BMO martingales. *Finance Stoch.*, 12(2):219–244, 2008.
- [21] J. Bion-Nadal. Time consistent dynamic risk processes. *Stochastic Process. Appl.*, 119(2):633–654, 2009.
- [22] J.R. Birge and F. Louveaux. Introduction to Stochastic Programming. Springer-Verlag, New York, NY, 1997.
- [23] G. Birkhoff. Tres obsevaciones sobre el algebra lineal. Univ. Nac. Tucumán Rev. Ser. A, 5:147–151, 1946.
- [24] J. Blomvall and A. Shapiro. Solving multistage asset investment problems by Monte Carlo based optimization. *Mathematical Programming*, 108:571–595, 2007.
- [25] K. Boda and J. A. Filar. Time consistent dynamic risk measures. *Math. Methods Oper. Res.*, 63(1):169–186, 2006.
- [26] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer-Verlag, New York, NY, 2000.
- [27] C. Borell. Convex measures on locally convex spaces. Ark. Mat., 12:239–252, 1974.
- [28] C. Borell. Convex set functions in d-space. *Periodica Mathematica Hungarica*, 6:111–136, 1975.
- [29] E. Boros and A. Prékopa. Close form two-sided bounds for probabilities that exactly *r* and at least *r* out of *n* events occur. *Math. Opns. Res.*, 14:317–342, 1989.

- [30] H. J. Brascamp and E. H. Lieb. On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems including inequalities for log concave functions, and with an application to the diffusion equations. *Journal of Functional Analysis*, 22:366–389, 1976.
- [31] J. Bukszár. Probability bounds with multitrees. Adv. Appl. Prob., 33:437–452, 2001.
- [32] J. Bukszár and A. Prékopa. Probability bounds with cherry trees. *Mathematics of Operations Research*, 26:174–192, 2001.
- [33] G. Calafiore and M.C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51:742–753, 2006.
- [34] M.C. Campi and S. Garatti. The exact feasibility of randomized solutions of robust convex programs. SIAM J. Optimization, 19:1211–1230, 2008.
- [35] M.S. Casey and S. Sen. The scenario generation algorithm for multistage stochastic linear programming. *Mathematics Of Operations Research*, 30:615–631, 2005.
- [36] A. Charnes, W. W. Cooper, and G. H. Symonds. Cost horizons and certainty equivalents; an approach to stochastic programming of heating oil. *Management Science*, 4:235–263, 1958.
- [37] P. Cheridito, F. Delbaen, and M. Kupper. Coherent and convex risk measures for bounded càdlàg processes. *Stochastic Processes and their Applications*, 112:1–22, 2004.
- [38] P. Cheridito, F. Delbaen, and M. Kupper. Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability*, 11:57–106, 2006.
- [39] P. Cheridito and M. Kupper. Composition of time-consistent dynamic monetary risk measures in discrete time. *Int. J. Theor. Appl. Finance*, 14(1):137–162, 2011.
- [40] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum observations. *Annals Math. Statistics*, 23:493–507, 1952.
- [41] S. Choi and A. Ruszczyński. A risk-averse newsvendor with law invariant coherent measures of risk. Oper. Res. Lett., 36(1):77–82, 2008.
- [42] S. Choi, A. Ruszczyński, and Y. Zhao. A multiproduct risk-averse newsvendor with law-invariant coherent measures of risk. *Oper. Res.*, 59(2):346–364, 2011.
- [43] E.K.P. Chong and P.J. Ramadge. Optimization of queues using an infinitesimal perturbation analysis-based stochastic algorithm with general update times. *SIAM J. Control and Optimization*, 31:698–732, 1993.
- [44] A. Clark and H. Scarf. Optimal policies for a multi-echelon inventory problem. *Management Science*, 6:475–490, 1960.

- [45] F.H. Clarke. Optimization and nonsmooth analysis. Canadian Mathematical Society Series of Monographs and Advanced Texts. A Wiley-Interscience Publication. John Wiley & Sons Inc., New York, 1983.
- [46] R. Cranley and T. N. L. Patterson. Randomization of number theoretic methods for multiple integration. SIAM J. Numer. Anal., 13:904–914, 1976.
- [47] J.M. Danskin. The Theory of Max-Min and Its Applications to Weapons Allocation Problems. Springer-Verlag, New York, 1967.
- [48] G.B. Dantzig. Linear programming under uncertainty. *Management Science*, 1:197–206, 1955.
- [49] G.B. Dantzig and G. Infanger. Large-scale stochastic linear programs—importance sampling and Benders decomposition. In *Computational and Applied Mathematics I* (*Dublin, 1991*), pages 111–120. North-Holland, Amsterdam, 1992.
- [50] I. Deák. Linear regression estimators for multinormal distributions in optimization of stochastic programming problems. *European Journal of Operational Research*, 111:555–568, 1998.
- [51] P. Delbaen. Coherent risk measures on general probability spaces. In *Essays in Honour of Dieter Sondermann*. Springer-Verlag, Berlin, Germany, 2002.
- [52] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer-Verlag, New York, NY, 1998.
- [53] M. A. H. Dempster. The expected value of perfect information in the optimal evolution of stochastic problems. In M. Arato, D. Vermes, and A. V. Balakrishnan, editors, *Stochastic Differential Systems*, volume 36 of *Lecture Notes in Information* and Control, pages 25–40. Springer-Verlag, Berkeley, CA, 1981.
- [54] D. Dentcheva. Regular Castaing representations with application to stochastic programming. SIAM Journal on Optimization, 10:732–749, 2000.
- [55] D. Dentcheva, B. Lai, and A Ruszczyński. Dual methods for probabilistic optimization. *Mathematical Methods of Operations Research*, 60:331–346, 2004.
- [56] D. Dentcheva and G. Martinez. Augmented lagrangian method for probabilistic optimization. *Annals of operations research*, 200:109–130, 2012.
- [57] D. Dentcheva and G. Martinez. Regularization methods for optimization problems with probabilistic constraints. *Mathemathical Programming, Ser. A*, 138:223–251, 2013.
- [58] D. Dentcheva, A. Prékopa, and A. Ruszczyński. Concavity and efficient points of discrete distributions in probabilistic programming. *Mathematical Programming*, 89:55–77, 2000.

- [59] D. Dentcheva, A. Prekopa, and A. Ruszczyński. On convex probabilistic programs with discrete distributions. *Nonlinear Analysis: Theory, Methods & Applications*, 47:1997–2009, 2001.
- [60] D. Dentcheva, A. Prékopa, and A. Ruszczyński. Bounds for probabilistic integer programming problems. *Discrete Applied Mathematics*, 124:55–65, 2002.
- [61] D. Dentcheva and W. Römisch. Differential stability of two-stage stochastic programs. SIAM Journal on Optimization, 11:87–112, 2001.
- [62] D. Dentcheva and A. Ruszczyński. Optimization with stochastic dominance constraints. SIAM Journal on Optimization, 14:548–566, 2003.
- [63] D. Dentcheva and A Ruszczyński. Convexification of stochastic ordering. *Comptes Rendus de l'Academie Bulgare des Sciences*, 57:11–16, 2004.
- [64] D. Dentcheva and A. Ruszczyński. Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints. *Mathematical Pro*gramming, 99:329–350, 2004.
- [65] D. Dentcheva and A Ruszczyński. Semi-infinite probabilistic optimization: First order stochastic dominance constraints. *Optimization*, 53:583–601, 2004.
- [66] D. Dentcheva and A. Ruszczyński. Portfolio optimization under stochastic dominance constraints. *Journal of Banking and Finance*, 30:433–451, 2006.
- [67] K. Detlefsen and G. Scandolo. Conditional and dynamic convex risk measures. *Finance Stoch.*, 9(4):539–561, 2005.
- [68] K. Dowd. Beyond Value at Risk. The Science of Risk Management. Wiley & Sons, New York, 1997.
- [69] N. Dunford and J. Schwartz. Linear Operators, Vol I. Interscience, New York, 1958.
- [70] J. Dupačová and R.J.B. Wets. Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *The Annals of Statistics*, 16:1517–1549, 1988.
- [71] J. Dupačová, N. Growe-Kuska, and W. Römisch. Scenario reduction in stochastic programming - An approach using probability metrics. *Mathematical Programming*, 95:493–511, 2003.
- [72] M. Dyer and L. Stougie. Computational complexity of stochastic programming problems. *Mathematical Programming*, 106:423–432, 2006.
- [73] F. Edgeworth. The mathematical theory of banking. *Royal Statistical Society*, 51:113–127, 1888.
- [74] A. Eichhorn and W. Römisch. Polyhedral risk measures in stochastic programming. SIAM Journal on Optimization, 16:69–95, 2005.

- [75] M.J. Eisner and P. Olsen. Duality for stochastic programming interpreted as L.P. in L_p space. SIAM J. Appl. Math., 28:779–792, 1975.
- [76] Y. Ermoliev. Stochastic quasi-gradient methods and their application to systems optimization. *Stochastics*, 4:1–37, 1983.
- [77] P.C. Fishburn. Utility Theory for Decision Making. John Wiley & Sons, New York, 1970.
- [78] G.S. Fishman. *Monte Carlo, Concepts, Algorithms and Applications*. Springer-Verlag, New York, 1999.
- [79] H. Föllmer and I. Penner. Convex risk measures and the dynamics of their penalty functions. *Statist. Decisions*, 24(1):61–96, 2006.
- [80] H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6:429–447, 2002.
- [81] M. Fritelli and G. Scandolo. Risk measures and capital requirements for processes. *Mathematical Finance*, 16:589–612, 2006.
- [82] M. Frittelli and E. Rosazza Gianin. Dynamic convex risk measures. In G. Szegö, editor, *Risk Measures for the 21st Century*, pages 227–248. John Wiley & Sons, Chichester, 2005.
- [83] S. Garstka and R.J.B. Wets. On decision rules in stochastic programming. *Mathe-matical Programming*, 7:117143, 1974.
- [84] A. Georghiou, W. Wiesemann, and D. Kuhn. Generalized decision rule approximations for stochastic programming via liftings. www.optimization-online.org/DB_HTML/2010/08/2701.html, 2013.
- [85] C.J. Geyer and E.A. Thompson. Constrained Monte Carlo maximum likelihood for dependent data, (with discussion). J. Roy. Statist. Soc. Ser. B, 54:657–699, 1992.
- [86] R.D. Gill. Non-and-semiparametric maximum likelihood estimators and the von mises method (part i). *Scandinavian Journal of Statistics*, 16:97–124, 1989.
- [87] P. Girardeau, V. Leclere, and A. B. Philpott. On the convergence of decomposition methods for multi-stage stochastic convex programs. *Optimization Online*, 2013.
- [88] P. Glasserman. Gradient Estimation via Perturbation Analysis. Kluwer Academic Publishers, Norwell, MA, 1991.
- [89] E.G. Gol'shtein. *Theory of Convex Programming*. Translations of Mathematical Monographs, American Mathematical Society, Providence, 1972.
- [90] N. Gröwe. Estimated stochastic programs with chance constraint. *European Journal* of Operations Research, 101:285–305, 1997.

- [91] R. Grübel. The length of the short. Annals of Statistics, 16:619–628, 1988.
- [92] V. Guigues. SDDP for some interstage dependent risk-averse problems and application to hydro-thermal planning. *Computational Optimization and Applications*, to appear.
- [93] G. Gurkan, A.Y. Ozge, and S.M. Robinson. Sample-path solution of stochastic variational inequalities. *Mathematical Programming*, 21:313–333, 1999.
- [94] J. Hadar and W. Russell. Rules for ordering uncertain prospects. *The American Economic Review*, 59:25–34, 1969.
- [95] W. K. Klein Haneveld. Duality in Stochastic Linear and Dynamic Programming, volume 274 of Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, New York, 1986.
- [96] G.H. Hardy, J.E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, Cambridge, MA, 1934.
- [97] H. Heitsch, H. Leövey, and W. Römisch. Are quasi-monte carlo algorithms efficient for two-stage stochastic programs? *Stochastic Programming E-Print Series*, 5-2012, 2012.
- [98] H. Heitsch and W. Römisch. Scenario tree modeling for multistage stochastic programs. *Mathematical Programming*, 118:371–406, 2009.
- [99] R. Henrion. Perturbation analysis of chance-constrained programs under variation of all constraint data. In K. Marti et al., editor, *Dynamic Stochastic Optimization*, Lecture Notes in Economics and Mathematical Systems, pages 257–274. Springer-Verlag, Heidelberg.
- [100] R. Henrion. On the connectedness of probabilistic constraint sets. *Journal of Opti*mization Theory and Applications, 112:657–663, 2002.
- [101] R. Henrion. Gradient estimates for gaussian distribution functions: application to probabilistically constrained optimization problems. *Numerical Algebra, Control* and Optimization, 2:655–668, 2012.
- [102] R. Henrion and A. Möller. A gradient formula for linear chance constraints under gaussian distribution. *Mathematics of Operations Research*, 37:475–488, 2012.
- [103] R. Henrion and W. Römisch. Metric regularity and quantitative stability in stochastic programs with probabilistic constraints. *Mathematical Programming*, 84:55–88, 1998.
- [104] J.L. Higle. Variance reduction and objective function evaluation in stochastic linear programs. *INFORMS Journal on Computing*, 10(2):236–247, 1998.
- [105] J.L. Higle and S. Sen. Duality and statistical tests of optimality for two stage stochastic programs. *Math. Programming (Ser. B)*, 75(2):257–275, 1996.

- [106] J.L. Higle and S. Sen. Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming. Kluwer Academic Publishers, Dordrecht, 1996.
- [107] J.-B. Hiriart-Urruty and C. Lemaréchal. Convex Analysis and Minimization Algorithms I and II. Springer-Verlag, New York, 1993.
- [108] Y.C. Ho and X.R. Cao. Perturbation Analysis of Discrete Event Dynamic Systems. Kluwer Academic Publishers, Norwell, MA, 1991.
- [109] R. Hochreiter and G. Ch. Pflug. Financial scenario generation for stochastic multistage decision processes as facility location problems. *Annals Of Operations Research*, 152:257–272, 2007.
- [110] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.
- [111] A. Hoffman. On approximate solutions of systems of linear inequalities. *Journal* of Research of the National Bureau of Standards, Section B, Mathematical Sciences, 49:263–265, 1952.
- [112] T. Homem-de-Mello. On rates of convergence for stochastic optimization problems under non-independent and identically distributed sampling. *SIAM Journal on Optimization*, 19:524–551, 2008.
- [113] C. C. Huang, I. Vertinsky, and W. T. Ziemba. Sharp bounds on the value of perfect information. *Operations Research*, 25:128–139, 1977.
- [114] P.J. Huber. The behavior of maximum likelihood estimates under nonstandard conditions, volume 1. Univ. California Press., 1967.
- [115] G. Infanger. Planning Under Uncertainty: Solving Large Scale Stochastic Linear Programs. Boyd and Fraser Publishing Co., Danvers, MA, 1994.
- [116] A.D. Ioffe and V.M. Tihomirov. *Theory of Extremal Problems*. North-Holland Publishing Company, Amsterdam, 1979.
- [117] G. N. Iyengar. Robust dynamic programming. *Mathematics of Operations Research*, 30:257–280, 2005.
- [118] R. Jiang and Y. Guan. Data-driven chance constrained stochastic program. www.optimization-online.org/DB_HTML/2013/09/4044.html, 2013.
- [119] E. Jouini, W. Schachermayer, and N. Touzi. Law invariant risk measures have the fatou property. *Advances in Mathematical Economics*, 9:49–71, 2006.
- [120] P. Kall. Qualitative aussagen zu einigen problemen der stochastischen programmierung. Z. Warscheinlichkeitstheorie u. Vervandte Gebiete, 6:246–272, 1966.
- [121] P. Kall. Stochastic Linear Programming. Springer-Verlag, Berlin, 1976.

- [122] P. Kall and J. Mayer. Stochastic Linear Programming. Springer, New York, 2005.
- [123] P. Kall and S. W. Wallace. *Stochastic Programming*. John Wiley & Sons, Chichester, England, 1994.
- [124] V. Kankova. On the convergence rate of empirical estimates in chance constrained stochastic programming. *Kybernetika (Prague)*, 26:449–461, 1990.
- [125] J.E. Kelley. The cutting-plane method for solving convex programs. Journal of the Society for Industrial and Applied Mathematics, 8:703–712, 1960.
- [126] A.I. Kibzun and G.L. Tretyakov. Differentiability of the probability function. *Doklady Akademii Nauk*, 354:159–161, 1997. (in Russian).
- [127] A.I. Kibzun and S. Uryasev. Differentiability of probability function. *Stochastic Analysis and Applications*, 16:1101–1128, 1998.
- [128] M. Kijima and M. Ohnishi. Mean-risk analysis of risk aversion and wealth effects on optimal portfolios with multiple investment opportunities. *Ann. Oper. Res.*, 45:147– 163, 1993.
- [129] A.J. King and R.T. Rockafellar. Asymptotic theory for solutions in statistical estimation and stochastic programming. *Mathematics of Operations Research*, 18:148– 162, 1993.
- [130] A.J. King and R.J.-B. Wets. Epi-consistency of convex stochastic programs. Stochastics Stochastics Rep., 34(1-2):83–92, 1991.
- [131] A.J. Kleywegt, A. Shapiro, and T. Homem-De-Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12:479–502, 2001.
- [132] S. Klöppel and M. Schweizer. Dynamic indifference valuation via convex risk measures. *Math. Finance*, 17(4):599–627, 2007.
- [133] M. Koivu. Variance reduction in sample approximations of stochastic programs. *Mathematical Programming*, 103:463–485, 2005.
- [134] H. Konno and H. Yamazaki. Mean–absolute deviation portfolio optimization model and its application to tokyo stock market. *Management Science*, 37:519–531, 1991.
- [135] T.C. Koopmans. Stationary ordinal utility and impatience. *Econometrica*, 28:287– 309, 1960.
- [136] D. M. Kreps and E. L. Porteus. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica*, 46(1):185–200, 1978.
- [137] D. M. Kreps and E.L. Porteus. Temporal von Neumann-Morgenstern and induced preferences. J. Econom. Theory, 20(1):81–109, 1979.
- [138] H.J. Kushner and D.S. Clark. Stochastic Approximation Methods for Constrained and Unconstrained Systems. Springer-Verlag, Berlin, Germany, 1978.

- [139] S. Kusuoka. On law-invariant coherent risk measures. In Kusuoka S. and Maruyama T., editors, *Advances in Mathematical Economics, Vol. 3*, pages 83–95. Springer, Tokyo, 2001.
- [140] G. Lan, A. Nemirovski, and A. Shapiro. Validation analysis of robust stochastic approximation method. *Mathematical Programming*, 134:425–458, 2012.
- [141] P. L'Ecuyer and P.W. Glynn. Stochastic optimization by simulation: Convergence proofs for the GI/G/1 queue in steady-state. *Management Science*, 11:1562–1578, 1994.
- [142] E. Lehmann. Ordered families of distributions. Annals of Mathematical Statistics, 26:399–419, 1955.
- [143] J. Leitner. A short note on second-order stochastic dominance preserving coherent risk measures. *Mathematical Finance*, 15:649–651, 2005.
- [144] V.L. Levin. Application of a theorem of E. Helly in convex programming, problems of best approximation and related topics. *Mat. Sbornik*, 79:250–263, 1969. (in Russian).
- [145] V.L. Levin. Convex Analysis in Spaces of Measurable Functions and Its Applications in Economics. Nauka, Moskva, 1985. (in Russian).
- [146] J. Linderoth, A. Shapiro, and S. Wright. The empirical behavior of sampling methods for stochastic programming. *Annals of Operations Research*, 142:215–241, 2006.
- [147] J. Luedtke and S. Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19:674–699, 2008.
- [148] A. Madansky. Inequalities for stochastic linear programming problems. *Management Science*, 6:197–204, 1960.
- [149] W.K. Mak, D.P. Morton, and R.K. Wood. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24:47–56, 1999.
- [150] H.B. Mann and D.R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statistics*, 18:50–60, 1947.
- [151] H. M. Markowitz. Portfolio selection. Journal of Finance, 7:77–91, 1952.
- [152] H. M. Markowitz. Portfolio Selection. Wiley, New York, 1959.
- [153] H. M. Markowitz. Mean–Variance Analysis in Portfolio Choice and Capital Markets. Blackwell, Oxford, 1987.
- [154] S. Mazur. Uber konvexe Mengen in linearen normierten Räumen. Studia Math., 4:70–84, 1933.

- [155] M. Meyer and S. Reisner. Characterizations of affinely-rotation-invariant logconcave meaqsures by section-centroid location. In *Geometric Aspects of Functional Analysis*, volume 1469 of *Lecture Notes in Mathematics*, pages 145–152. Springer-Verlag, Berlin, 1989–90.
- [156] L.B. Miller and H. Wagner. Chance-constrained programming with joint constraints. Operations Research, 13:930–945, 1965.
- [157] N. Miller and A. Ruszczyński. Risk-adjusted probability measures in portfolio optimization with coherent measures of risk. *European Journal of Operational Research*, 191:193–206, 2008.
- [158] N. Miller and A. Ruszczyński. Risk-averse two-stage stochastic linear programming: modeling and decomposition. *Operations Research*, 59(1):125–132, 2011.
- [159] K. Mosler and M. Scarsini. Stochastic Orders and Decision Under Risk. Institute of Mathematical Statistics, Hayward, California, 1991.
- [160] A. Nagurney. Supply Chain Network Economics: Dynamics of Prices, Flows, and Profits. Edward Elgar Publishing, 2006.
- [161] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19:1574– 1609, 2009.
- [162] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. SIAM Journal on Optimization, 17:969–996, 2006.
- [163] A. Nemirovski and D. Yudin. On Cezaro's convergence of the steepest descent method for approximating saddle point of convex-concave functions. *Soviet Math. Dokl.*, 19, 1978.
- [164] A. Nemirovski and D. Yudin. Problem Complexity and Method Efficiency in Optimization. John Wiley, New York, 1983.
- [165] Yu. Nesterov. Introductory Lectures on Convex Optimization. Kluwer, Boston, 2004.
- [166] M.B. Nevelson and R.Z. Hasminskii. Stochastic Approximation and Recursive Estimation. American Mathematical Society Translations of Mathematical Monographs, Vol. 47, Rhode Island, 1976.
- [167] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia, 1992.
- [168] A. Nilim and L. El Ghaoui. Robust control of Markov decision processes with uncertain transition matrices. *Operations Research*, 53:780–798, 2005.
- [169] V.I. Norkin. The analysis and optimization of probability functions. IIASA Working Paper, WP-93-6, 1993.

- [170] V.I. Norkin, G.Ch. Pflug, and A. Ruszczyński. A branch and bound method for stochastic global optimization. *Mathematical Programming*, 83:425–450, 1998.
- [171] V.I. Norkin and N.V. Roenko. α-Concave functions and measures and their applications. *Kibernet. Sistem. Anal.*, 189:77–88, 1991. Russian. Translation in: *Cybernet. Systems Anal.* 27 (1991) 860–869.
- [172] N. Noyan and G. Rudolf. Optimization with multivariate conditional value-at-risk constraints. www.optimization-online.org/DB_HTML/2012/04/3444.html, 2012.
- [173] W. Ogryczak and A. Ruszczyński. From stochastic dominance to mean-risk models: semideviations as risk measures. *European Journal of Operational Research*, 116:33–50, 1999.
- [174] W. Ogryczak and A. Ruszczyński. On consistency of stochastic dominance and mean-semideviation models. *Mathematical Programming*, 89:217–232, 2001.
- [175] W. Ogryczak and A. Ruszczyński. Dual stochastic dominance and related mean risk models. SIAM Journal on Optimization, 13:60–78, 2002.
- [176] T. Pennanen. Epi-convergent discretizations of multistage stochastic programs. *Mathematics of Operations Research*, 30:245–256, 2005.
- [177] T. Pennanen and M. Koivu. Epi-convergent discretizations of stochastic programs via integration quadratures. *Numerische Mathematik*, 100:141–163, 2005.
- [178] M.V.F. Pereira and L.M.V.G. Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52:359–375, 1991.
- [179] G. Pflug and A. Ruszczyński. A risk measure for income processes. In G. Szegö, editor, *Risk Measures for the 21st Century*, pages 249–269. John Wiley & Sons, Chichester, 2004.
- [180] G. Ch. Pflug and N. Wozabal. Asymptotic distribution of law-invariant risk functionals. *Finance and Stochastics*, 14:397–418, 2010.
- [181] G.Ch. Pflug. Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic Constrained Optimization - Methodology and Applications*, pages 272–281. Kluwer Academic Publishers, Dordrecht, 2000.
- [182] G.Ch. Pflug and W. Römisch. *Modeling, Measuring and Managing Risk*. World Scientific, Singapore, 2007.
- [183] R.R. Phelps. *Convex functions, monotone operators, and differentiability*, volume 1364 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1989.
- [184] A.B. Philpott and V.L. de Matos. Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. *European Journal of Operational Research*, 218:470–483, 2012.

- [185] A.B. Philpott and Z. Guan. On the convergence of stochastic dual dynamic programming and related methods. *Operations Research Letters*, 36:450455, 2008.
- [186] A. Pichler and A. Shapiro. Uniqueness of Kusuoka representations. www.optimization-online.org/DB_HTML/2012/10/3660.html, 2013.
- [187] E.L. Plambeck, B.R. Fu, S.M. Robinson, and R. Suri. Sample-path optimization of convex stochastic performance functions. *Mathematical Programming, Series B*, 75:137–176, 1996.
- [188] B.T. Polyak. New stochastic approximation type procedures. Automat. i Telemekh., 7:98–107, 1990.
- [189] B.T. Polyak and A.B. Juditsky. Acceleration of stochastic approximation by averaging. SIAM J. Control and Optimization, 30:838–855, 1992.
- [190] A. Prékopa. On probabilistic constrained programming. In *Proceedings of the Princeton Symposium on Mathematical Programming*, pages 113–138. Princeton University Press, Princeton, New Jersey, 1970.
- [191] A. Prékopa. Logarithmic concave measures with applications to stochastic programming. Acta Scientiarium Mathematicarum (Szeged), 32:301–316, 1971.
- [192] A. Prékopa. On logarithmic concave measures and functions. Acta Scientiarium Mathematicarum (Szeged), 34:335–343, 1973.
- [193] A. Prékopa. Dual method for the solution of a one-stage stochastic programming problem with random rhs obeying a discrete probality distribution. ZOR-Methods and Models of Operations Research, 34:441–461, 1990.
- [194] A. Prékopa. Sharp bound on probabilities using linear programming. Operations Research, 38:227–239, 1990.
- [195] A. Prékopa. Stochastic Programming. Kluwer Acad. Publ., Dordrecht, Boston, 1995.
- [196] A. Prékopa, B. Vízvári, and T. Badics. Programming under probabilistic constraint with discrete random variable. In L. Grandinetti et al., editor, *New Trends in Mathematical Programming*, pages 235–255. Kluwer, Boston, 2003.
- [197] J. Quiggin. A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3:225–243, 1982.
- [198] J. Quiggin. Generalized Expected Utility Theory The Rank-Dependent Expected Utility Model. Kluwer, Dordrecht, 1993.
- [199] J.P Quirk and R. Saposnik. Admissibility and measurable utility functions. *Review of Economic Studies*, 29:140–146, 1962.
- [200] H. Raiffa. Decision Analysis. Addison-Wesley, Reading, MA, 1968.

- [201] H. Raiffa and R. Schlaifer. Applied Statistical Decision Theory. Studies in Managerial Economics. Harvard University, Boston, MA, 1961.
- [202] E. Raik. The differentiability in the parameter of the probability function and optimization of the probability function via the stochastic pseudogradient method. *Eesti NSV Teaduste Akdeemia Toimetised. Füüsika-Matemaatika*, 24:860–869, 1975. (in Russian).
- [203] F. Riedel. Dynamic coherent risk measures. Stochastic Processes and Their Applications, 112:185–200, 2004.
- [204] Y. Rinott. On convexity of measures. Annals of Probability, 4:1020–1026, 1976.
- [205] H. Robbins and S. Monro. A stochastic approximation method. Annals of Math. Stat., 22:400–407, 1951.
- [206] S.M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5:43–62, 1980.
- [207] S.M. Robinson. Generalized equations and their solutions, Part II: Applications to nonlinear programming. *Mathematical Programming Study*, 19:200–221, 1982.
- [208] S.M. Robinson. Normal maps induced by linear transformations. *Mathematics of Operations Research*, 17:691–714, 1992.
- [209] S.M. Robinson. Analysis of sample-path optimization. *Mathematics of Operations Research*, 21:513–528, 1996.
- [210] R.T. Rockafellar. Convex Analysis. Princeton University Press, Princeton, 1970.
- [211] R.T. Rockafellar. *Conjugate Duality and Optimization*. Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1974.
- [212] R.T. Rockafellar. Duality and optimality in multistage stochastic programming. Ann. Oper. Res., 85:1–19, 1999. Stochastic Programming. State of the Art, 1998 (Vancouver, BC).
- [213] R.T. Rockafellar and S. Uryasev. Optimization of conditional value at risk. *The Journal of Risk*, 2:21–41, 2000.
- [214] R.T. Rockafellar, S. Uryasev, and M. Zabarankin. Generalized deviations in risk analysis. *Finance and Stochastics*, 10:51–74, 2006.
- [215] R.T. Rockafellar and R.J.-B. Wets. Stochastic convex programming: basic duality. *Pacific J. Math.*, 62(1):173–195, 1976.
- [216] R.T. Rockafellar and R.J.-B. Wets. Stochastic convex programming: singular multipliers and extended duality singular multipliers and duality. *Pacific J. Math.*, 62(2):507–522, 1976.
- [217] R.T. Rockafellar and R.J.-B. Wets. Variational Analysis. Springer, Berlin, 1998.

- [218] W.W. Rogosinski. Moments of non-negative mass. Proc. Roy. Soc. London Ser. A, 245:1–27, 1958.
- [219] W. Römisch. Stability of stochastic programming problems. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 483–554. Elsevier, Amsterdam, 2003.
- [220] R.Y. Rubinstein and A. Shapiro. Optimization of static simulation models by the score function method. *Mathematics and Computers in Simulation*, 32:373–392, 1990.
- [221] R.Y. Rubinstein and A. Shapiro. Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method. John Wiley & Sons, Chichester, England, 1993.
- [222] A. Ruszczyński. Decompositon methods. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, Handbooks in Operations Research and Management Science, pages 141–211. Elsevier, Amsterdam, 2003.
- [223] A. Ruszczyński. Risk-averse dynamic programming for Markov decision processes. *Math. Program.*, 125(2, Ser. B):235–261, 2010.
- [224] A. Ruszczyński and A. Shapiro. Optimization of risk measures. In G. Calafiore and F. Dabbene, editors, *Probabilistic and Randomized Methods for Design under Uncertainty*, pages 117–158. Springer-Verlag, London, 2005.
- [225] A. Ruszczyński and A. Shapiro. Conditional risk mappings. *Mathematics of Oper*ations Research, 31:544–561, 2006.
- [226] A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Mathematics of Operations Research*, 31:433–452, 2006.
- [227] A. Ruszczyński and R. Vanderbei. Frontiers of stochastically nondominated portfolios. *Econometrica*, 71:1287–1297, 2003.
- [228] G. Salinetti. Approximations for chance constrained programming problems. Stochastics, 10:157–169, 1983.
- [229] T. Santoso, S. Ahmed, M. Goetschalckx, and A. Shapiro. A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research*, 167:95–115, 2005.
- [230] G. Scandolo. *Risk Measures in a Dynamic Setting*. PhD thesis, Università degli Studi di Milano, Milan, Italy, 2003.
- [231] H. Scarf. A min-max solution of an inventory problem. In *Studies in The Mathematical Theory of Inventory and Production*, Stanford University Press, pages 201–209. Stanford University Press, California, 1958.

- [232] L. Schwartz. Analyse Mathematique, volume I and II. Mir/Hermann, Moscow/Paris, 1972/1967.
- [233] S. Sen. Relaxations for the probabilistically constrained programs with discrete random variables. *Operations Research Letters*, 11:81–86, 1992.
- [234] M. Shaked and J. G. Shanthikumar. Stochastic Orders and Their Applications. Academic Press, Boston, 1994.
- [235] A. Shapiro. Asymptotic properties of statistical estimators in stochastic programming. Annals of Statistics, 17:841–858, 1989.
- [236] A. Shapiro. Asymptotic analysis of stochastic programs. Annals of Operations Research, 30:169–186, 1991.
- [237] A. Shapiro. Statistical inference of stochastic optimization problems. In S. Uryasev, editor, *Probabilistic Constrained Optimization: Methodology and Applications*, pages 282–304. Kluwer Academic Publishers, 2000.
- [238] A. Shapiro. Monte Carlo approach to stochastic programming. In B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, editors, *Proceedings of the 2001 Winter Simulation Conference*, pages 428–431. 2001.
- [239] A. Shapiro. Inference of statistical bounds for multistage stochastic programming problems. *Mathematical Methods of Operations Research*, 58:57–68, 2003.
- [240] A. Shapiro. Monte Carlo sampling methods. In A. Ruszczyński and A. Shapiro, editors, *Handbook in OR & MS, Vol. 10*, pages 353–425. North-Holland Publishing Company, 2003.
- [241] A. Shapiro. On complexity of multistage stochastic programs. Operations Research Letters, 34:1–8, 2006.
- [242] A. Shapiro. Asymptotics of minimax stochastic programs. *Statistics and Probability Letters*, 78:150–157, 2008.
- [243] A. Shapiro. Stochastic programming approach to optimization under uncertainty. *Mathematical Programming, Series B*, 112:183–220, 2008.
- [244] A. Shapiro. On a time consistency concept in risk averse multi-stage stochastic programming. *Operations Research Letters*, 37:143 – 147, 2009.
- [245] A. Shapiro. Analysis of stochastic dual dynamic programming method. European Journal of Operational Research, 209:63–72, 2011.
- [246] A. Shapiro. Time consistency of dynamic risk measures. Operations Research Letters, 40:436–439, 2012.
- [247] A. Shapiro. Consistency of sample estimates of risk averse stochastic programs. *Journal of Applied Probability*, 50:533–541, 2013.

- [248] A. Shapiro. On Kusuoka representation of law invariant risk measures. *Mathematics of Operations Research*, 38:142–152, 2013.
- [249] A. Shapiro and A. Nemirovski. On complexity of stochastic programming problems. In V. Jeyakumar and A.M. Rubinov, editors, *Continuous Optimization: Current Trends and Applications*, pages 111–144. Springer-Verlag, 2005.
- [250] A. Shapiro and T. Homem-de-Mello. A simulation-based approach to two-stage stochastic programming with recourse. *Mathematical Programming*, 81:301–325, 1998.
- [251] A. Shapiro and T. Homem-de-Mello. On rate of convergence of Monte Carlo approximations of stochastic programs. *SIAM Journal on Optimization*, 11:70–86, 2000.
- [252] A. Shapiro, W. Tekaya, J. Paulo da Costa, and M. Pereira Soares. Final report for technical cooperation between georgia institute of technology and ONS Operador Nacional do Sistema Elétrico. www2.isye.gatech.edu/people/faculty/Alex_Shapiro/ONS-FR.pdf 2012.
- [253] A. Shapiro, W. Tekaya, J. Paulo da Costa, and M. Pereira Soares. Risk neutral and risk averse stochastic dual dynamic programming method. *European Journal of Operational Research*, 224:375–391, 2013.
- [254] A. Shapiro and Y. Wardi. Convergence analysis of stochastic algorithms. *Mathe-matics of Operations Research*, 21:615–628, 1996.
- [255] W. A. Spivey. Decision making and probabilistic programming. *Industrial Manage*ment Review, 9:57–67, 1968.
- [256] V. Strassen. The existence of probability measures with given marginals. Annals of Mathematical Statistics, 38:423–439, 1965.
- [257] T. Szántai. Improved bounds and simulation procedures on the value of the multivariate normal probability distribution function. *Annals of Oper. Res.*, 100:85–101, 2000.
- [258] R. Szekli. Stochastic Ordering and Dependence in Applied Probability. Springer-Verlag, New York, 1995.
- [259] E. Tamm. On g-concave functions and probability measures. *Eesti NSV Teaduste Akademia Toimetised (News of the Estonian Academy of Sciences) Füüs. Mat.*, 26:376–379, 1977.
- [260] G. Tintner. Stochastic linear programming with applications to agricultural economics. In H. A. Antosiewicz, editor, *Proc. 2nd Symp. Linear Programming*, pages 197–228. National Bureau of Standards, Washington, D.C., 1955.
- [261] S. Uryasev. A differentiation formula for integrals over sets given by inclusion. *Numerical Functional Analysis and Optimization*, 10:827–841, 1989.

- [262] S. Uryasev. Derivatives of probability and integral functions. In P. M. Pardalos and C. M. Floudas, editors, *Encyclopedia of Optimization*, pages 267–352. Kluwer Acad. Publ., 2001.
- [263] M. Valadier. Sous-différentiels d'une borne supérieure et d'une somme continue de fonctions convexes. *Comptes Rendus de l'Académie des Sciences de Paris Série A*, 268:39–42, 1969.
- [264] A.W. van der Vaart and A. Wellner. Weak Convergence and Empirical Processes. Springer-Verlag, New York, 1996.
- [265] B. Verweij, S. Ahmed, A.J. Kleywegt, G. Nemhauser, and A. Shapiro. The sample average approximation method applied to stochastic routing problems: A computational study. *Computational Optimization and Applications*, 24:289–333, 2003.
- [266] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, 1944.
- [267] A. Wald. Note on the consistency of the maximum likelihood estimates. Annals Mathematical Statistics, 20:595–601, 1949.
- [268] D.W. Walkup and R.J.-B. Wets. Some practical regularity conditions for nonlinear programms. SIAM J. Control, 7:430–436, 1969.
- [269] D.W. Walkup and R.J.B. Wets. Stochastic programs with recourse: special forms. In Proceedings of the Princeton Symposium on Mathematical Programming (Princeton Univ., 1967), pages 139–161, Princeton, N.J., 1970. Princeton Univ. Press.
- [270] S. Wallace and W. T. Ziemba, editors. Applications of Stochastic Programming. SIAM, New York, 2005.
- [271] R.J.B. Wets. Programming under uncertainty: the equivalent convex program. SIAM Journal on Applied Mathematics, 14:89–105, 1966.
- [272] R.J.B. Wets. Stochastic programs with fixed recourse: the equivalent deterministic program. SIAM Review, 16:309–339, 1974.
- [273] R.J.B. Wets. Duality relations in stochastic programming. In Symposia Mathematica, Vol. XIX (Convegno sulla Programmazione Matematica e sue Applicazioni, INDAM, Rome, pages 341–355, London, 1976. Academic Press.
- [274] L. Xin, D.A. Goldberg, and A. Shapiro. Distributionally robust multistage inventory models with moment constraints. *arXiv:1304.3074*, 2013.
- [275] M. E. Yaari. The dual theory of choice under risk. *Econometrica*, 55:95–115, 1987.
- [276] P.H. Zipkin. Foundations of Inventory Management. McGraw-Hil, 2000.

Index

:= equal by definition, 413 A^{T} transpose of matrix (vector) A, 413 $C(\mathcal{X})$ space of continuous functions, 185 C^* polar of cone C, 417 $C^1(\mathcal{V},\mathbb{R}^n)$ space of continuously differentiable mappings, 197 IF_{θ} influence function, 367 L^{\perp} orthogonal of (linear) space L, 41 O(1) generic constant, 209 $O_p(\cdot)$ term, 471 $V_d(A)$ Lebesgue measure of set $A \subset \mathbb{R}^d$, 215 $W^{1,\infty}(U)$ space of Lipschitz continuous functions, 186, 435 $[a]_{+} = \max\{a, 0\}, 2$ $\mathbb{I}_A(\cdot)$ indicator function of set A, 414 $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ space, 490 $\Lambda(\bar{x})$ set of Lagrange multipliers vectors, 429 $\mathcal{N}(\mu, \Sigma)$ normal distribution, 16 \mathcal{N}_C normal cone to set C, 417 $\Phi(z)$ cdf of standard normal distribution, 16 $\Pi_{\mathcal{X}}$ metric projection onto set \mathcal{X} , 252 $\xrightarrow{\mathcal{D}}$ convergence in distribution, 183 $\mathcal{T}_{X}^{2}(x,h)$ second order tangent set, 429 AV@R Average Value-at-Risk, 294 $\hat{\mathfrak{P}}$ set of probability measures, 369 $\mathbb{D}(A, B)$ deviation of set A from set B, 414 $\mathbb{D}[Z_x]$ dispersion measure of random variable Z_x , 290 \mathbb{E} expectation, 443 $\mathbb{H}(A, B)$ Hausdorff distance between sets A and B, 414 \mathbb{N} set of positive integers, 441

 \mathbb{R}^n *n*-dimensional space, 413 A domain of the conjugate of risk measure ρ , 299 \mathfrak{C}_n space of nonempty compact subsets of \mathbb{R}^n , 467 & group of measure-preserving transformations, 319 \mathfrak{P} set of probability density functions, 300, 303 $\mathfrak{b}(k;\alpha,N)$ cdf of binomial distribution, 235 d distance generating function, 258 $\mathfrak{g}^+(x)$ right hand side derivative, 356 cl(A) topological closure of set A, 414 $\operatorname{conv}(C)$ convex hull of set C, 417 \mathbb{C} orr(X, Y) correlation of X and Y, 221 $\mathbb{C}ov(X, Y)$ covariance of X and Y, 201 q_{α} weighted mean deviation, 292 $S_C(\cdot)$ support function of set C, 417 $\delta(\omega)$ measure of mass one at the point ω , 445 dist(x, A) distance from point x to set A, 414 $\operatorname{dom} f$ domain of function f, 413 dom \mathcal{G} domain of multifunction \mathcal{G} , 448 \mathbb{R} set of extended real numbers, 413 epif epigraph of function f, 413 $\stackrel{e}{\rightarrow}$ epiconvergence, 460 Exp(K) set of exposed point of the set K, 488 $Ext(\Omega)$ set of extreme points of the set Ω, 373 \hat{S}_N the set of optimal solutions of the SAA problem, 176 $\hat{\mathcal{S}}_{N}^{\varepsilon}$ the set of ε -optimal solutions of the SAA problem, 201

 $\hat{\vartheta}_N$ optimal value of the SAA problem, 176 $f_N(x)$ sample average function, 175 $\mathbf{1}_A(\cdot)$ characteristic function of set A, 414 int(C) interior of set C, 417 |a| integer part of $a \in \mathbb{R}$, 241 lsc f lower semicontinuous hull of function f, 413 \mathcal{R}_C radial cone to set C, 417 \mathcal{T}_C tangent cone to set C, 417 $\nabla^2 f(x)$ Hessian matrix of second order partial derivatives, 199 ∂ subdifferential, 418 ∂° Clarke generalized gradient, 416 ∂_{ε} epsilon subdifferential, 469 pos W positive hull of matrix W, 29 \leq_C partial order defined by cone C, 50 Pr(A) probability of event A, 442 ri relative interior, 417 $\mathcal{S}^{\varepsilon}$ the set of ε -optimal solutions of the true problem, 201 $\sigma(\xi_1, ..., \xi_t)$ subalgebra generated by $(\xi_1, ..., \xi_t)$, 78 σ_p^+ upper semideviation, 291 σ_p^- lower semideviation, 291 F set of spectral functions, 323 Tr(A) trace of a square matrix, 421 $V@R_{\alpha}$ Value-at-Risk, 292 $\mathbb{V}ar[X]$ variance of X, 15 ϑ^* optimal value of the true problem, 176 $\xi_{[t]} = (\xi_1, \dots, \xi_t)$ history of the process, 63 $a \lor b = \max\{a, b\}, 207$ f^* conjugate of function f, 418 $f^{\circ}(x,d)$ generalized directional derivative, 416 g'(x,h) directional derivative, 414 $o_p(\cdot)$ term, 471 p-efficient point, 126 iid independently identically distributed, 176 approximation conservative, 294

conservative, 294 Average Value-at-Risk, 293, 294, 296, 311 dual representation, 311 Banach lattice, 495 Banach space, 488 reflexive, 488 separable, 488 Borel set, 441 bounded in probability, 471 Bregman divergence, 259 Capacity Expansion, 31, 42, 59 chain rule, 472 chance constrained problem ambiguous, 340 disjunctive semi-infinite formulation, 127 chance constraints, 6, 11, 16, 231 Clarke generalized gradient, 416 CLT Central Limit Theorem, 163 common random number generation method, 201comonotonic random variables, 331 complexity of multistage programs, 248 of two-stage programs, 201, 208 conditional expectation, 65, 446 conditional probability, 446 conditional risk mapping, 377, 382 conditional sampling, 242 identical, 243 independent, 243 Conditional Value-at-Risk, 293, 294, 296 cone contingent, 427, 474 critical, 198, 429 normal, 417 pointed, 495 polar, 29 recession, 29 tangent, 417 confidence interval, 183 conjugate duality, 421, 493 constraint nonanticipativity, 54, 350 constraint qualification linear independence, 189, 199 Mangasarian-Fromovitz, 428 Robinson, 428
Slater, 182 contact point, 488, 491 contingent cone, 427, 474 convergence in distribution, 183, 470 in probability, 471 weak, 473 with probability one, 457 convex hull, 417 cumulative distribution function, 2 of random vector, 11 cutting plane, 271 decision rule, 21, 64 Delta Theorem, 473, 474 finite dimensional, 471 second-order, 475 deviation of a set, 414 diameter of a set. 207 differential uniform dominance condition, 166 directional derivative, 414 ε -directional derivative, 469 generalized, 416 Hadamard, 472 second order, 475 tangentially to a set, 475 distribution asymptotically normal, 183 Binomial, 479 conditional, 446 Dirichlet, 107 discrete, 443 discrete with a finite support, 443 empirical, 176 Gamma, 111 log-concave, 105 log-normal, 116 multivariate normal, 16, 104 multivariate Student, 171 normal, 183 Pareto, 172 uniform, 104 Wishart, 111 domain

of multifunction, 448 of a function, 413 dual feasibility condition, 145 duality gap, 420, 422 dynamic programming equations, 7, 65, 379 empirical cdf, 3 empirical distribution, 176 entropy function, 258 epiconvergence, 440 with probability one, 460 epigraph of a function, 413 epsilon subdifferential, 469 estimator common random number, 226 consistent, 177 linear control, 221 unbiased, 176 expected value, 443 well defined, 443 expected value of perfect information, 61 Fatou's lemma, 444 filtration, 74, 78, 375, 390 floating body of a probability measure, 113 Fréchet differentiability, 414 function α -concave, 102 α -concave on a set, 114 affine, 82 biconjugate, 299, 491 Carathéodory, 176, 190, 449 characteristic, 414 Clarke-regular, 112, 416 composite, 304 conjugate, 299, 418, 491 continuously differentiable, 416 cost-to-go, 65, 69, 379 cumulative distribution (cdf), 2, 442 distance generating, 258 disutility, 290, 310 essentially bounded, 490 extended real valued, 443 indicator, 29, 414

530

influence, 367 integrable, 443 likelihood ratio, 221 log-concave, 103 logarithmically concave, 103 lower semicontinuous, 413 moment generating, 476 monotone, 496 optimal value, 449 polyhedral, 28, 42, 413, 497 proper, 413 quasi-concave, 104 radical-inverse, 218 random, 449 random lower semicontinuous, 449 random polyhedral, 43 sample average, 458 spectral, 323 strongly convex, 419 subdifferentiable, 418, 492 sublinear, 487 support, 417 utility, 290, 310 well defined, 451 Gâteaux differentiability, 414, 472 generalized equation sample average approximation, 196 generic constant O(1), 209 gradient, 416 Hadamard differentiability, 472 Hausdorff distance, 414 here-and-now solution, 10 Hessian matrix, 429 higher order distribution functions, 99 Hoffman's lemma, 424 identically distributed, 457 importance sampling, 222 independent identically distributed, 457 inequality Chebyshev, 445

Chebyshev, 445 Chernoff, 479 Hölder, 490 Hoeffding, 478

Jensen, 445 Markov, 445 Minkowski for matrices, 109 inf-compactness condition, 178, 433 interchangeability principle, 496 for risk measures, 352 for two-stage programming, 49 interior of a set, 417 inventory model, 1, 354 Jacobian matrix, 416 Kelley's cutting plane algorithm, 278 Kullback-Leibler divergence, 344 Kusuoka representation, 329 Lagrange multiplier, 429 large deviations rate function, 476 lattice, 495 Law of Large Numbers, 2, 457 for random sets, 467 pointwise, 458 strong, 457 uniform, 458 weak, 457 least upper bound, 495 Lebesgue-Stieltjes integral, 328 Lindeberg condition, 163 Lipschitz continuous, 415 lower bound statistical, 224 Lyapunov condition, 163 mapping convex, 50 measurable, 442 Markov chain, 73 Markovian process, 63 mean absolute deviation, 291 measurable multifunction, 448 measurable selection, 448 measurable transformation, 319 measure α -concave, 105 absolutely continuous, 442

complete, 441

Dirac, 445 finite, 441 Lebesgue, 441 nonatomic, 450 sigma-additive, 441 metric projection, 252 Mirror Descent SA, 262 model state equations, 71 model state variables, 71 moment generating function, 476 multifunction, 448 closed, 195, 433, 448 closed valued, 195, 448 convex, 50 convex valued, 50, 450 measurable, 448 optimal solution, 449 upper semicontinuous, 468 newsvendor problem, 1, 409 node ancestor, 72 children, 72 root, 72 nonanticipativity, 7, 53, 63 nonanticipativity constraints, 76, 378 nonatomic probability space, 450 norm dual, 258, 488 normal cone, 417 normal integrands, 449 optimality conditions first order, 228, 427 Karush-Kuhn-Tucker (KKT), 195, 228, 429 second order, 199, 429 paired spaces, 494 partial order, 495 permutation, 319 point contact, 488, 491 exposed, 488 extreme, 373 saddle, 420

polar cone, 417 policy basestock, 8, 406 feasible, 8, 17, 64 fixed mix, 21 implementable, 8, 17, 64 myopic, 20, 404 optimal, 8, 65, 70 piecewise affine, 82 piecewise linear, 82 time consistent, 398 portfolio selection, 13, 357 positive hull, 29 positively homogeneous, 198 Preface, iii principle of conditional optimality, 394 probabilistic constraints, 6, 11, 95, 182 individual, 98 joint, 98 probabilistic liquidity constraint, 102 probability density function, 442 probability distribution, 442 probability measure, 441 probability space, 441 nonatomic, 450 standard uniform, 303 probability vector, 375 problem chance constrained, 95, 231 distributionally robust, 370, 400 first stage, 10 of moments, 369 piecewise linear, 213 risk averse, 279 risk neutral, 279 second stage, 10 semi-infinite programming, 372 subconsistent, 422 time consistent, 398 two stage, 10 weakly time consistent, 398 process autoregressive, 67 stagewise independent, 63 stochastic, 63 prox-function, 259

prox-mapping, 259 quadratic growth condition, 211, 431 quantile, 16 left-side, 3, 292 right-side, 3, 292 radial cone, 417 random function convex, 452 random variable, 442 random vector, 442 recession cone, 417 recourse complete, 33 fixed, 33, 45 relatively complete, 10, 33 simple, 33 recourse action, 2 relative interior, 417 risk mapping coherent conditional, 376, 382 convex conditional, 376, 382 risk measure, 297 absolute semideviation, 364, 408 coherent, 298 comonotonic, 331 composite, 378, 391 consistent with stochastic orders, 337 convex, 298 distribution based, 319 dynamic, 387 entropic, 313 essential supremum, 296, 303, 331 law invariant, 319 mean-deviation, 315 mean-upper-semideviation, 316 mean-upper-semideviation from a target, 317 mean-variance, 314 proper, 298 regular, 334 spectral, 325 strictly time consistent, 396 time consistent, 387 version independent, 319

robust optimization, 12 saddle point, 420 sample independently identically distributed (iid), 176 random, 175 sample average approximation (SAA), 175 multistage, 243 sample covariance matrix, 229 sampling Latin Hypercube, 219 Monte Carlo, 200 scenario tree, 72 scenarios, 3, 30 SDDP algorithm, 274 second order regularity, 431 second order tangent set, 429 semi-infinite probabilistic problem, 164 semideviation lower. 291 upper, 291 separable space, 473 sequence Halton, 218 log-concave, 115 low-discrepancy, 217 van der Corput, 218 set dual, 301 elementary, 441 generating, 323 sigma algebra, 441 Borel, 441 trivial, 441 significance level, 5 simplex, 259 Slater condition, 182 solution ε -optimal, 201 sharp, 211, 212 space Banach, 488 decomposable, 497 dual, 488 Hilbert, 314

measurable, 441 paired, 494 probability, 441 reflexive, 488 sample, 441 stagewise independence, 8, 63, 73 star discrepancy, 215 stationary point of α -concave function, 113 Stochastic Approximation, 252 stochastic dominance kth order, 99 first order, 98, 337 higher order, 99 second order, 338 stochastic dominance constraint, 100 stochastic generalized equations, 194 stochastic order, 98, 337 increasing convex, 324, 338 usual, 337 stochastic programming nested risk averse multistage, 377, 390 stochastic programming problem minimax, 190 multiperiod, 68 multistage, 64 multistage linear, 70 two-stage convex, 50 two-stage linear, 27 two-stage polyhedral, 42 stochastic-order constraint, 99 strict complementarity condition, 200, 230 strongly regular solution of a generalized equation, 196 subdifferential, 418, 492 subgradient, 418, 492 algebraic, 492 stochastic, 252 supply chain model, 22 support of a measure, 36, 442 of a set, 417 support function, 28, 417, 418 supporting plane, 271

tangent cone, 417 theorem Artstein-Vitale, 467 Artstein-Wets, 461 Aumann, 450 Banach-Alaoglu, 488 Birkhoff, 121 Central Limit, 163 Cramér's Large Deviations, 476 Danskin, 434 Fenchel-Moreau, 299, 419, 491 functional CLT, 184 Glivenko-Cantelli, 460, 463 Hahn-Banach, 487 Helly, 417 Hlawka, 217 Klee-Nachbin-Namioka, 495 Koksma, 216 Kusuoka, 329 Lebesgue dominated convergence, 444 Levin-Valadier, 434 Lyapunov, 451 Mazur, 488 measurable selection, 448 Minkowski, 373 monotone convergence, 444 Moreau-Rockafellar, 418, 493 Rademacher, 416, 435 Radon-Nikodym, 442 Richter-Rogosinski, 444 Skorohod-Dudley almost sure representation, 473 time consistency, 401 of problem formulation, 395 policy, 398 risk measure, 387 strict, 396 topology strong (norm), 488 weak, 488 weak*, 488 transformation measurable, 319 measure-preserving, 319 uncertainty set, 12, 370

uniformly integrable, 444, 470 upper bound consevative, 225 statistical, 225 utility model, 309 value function, 7 Value-at-Risk, 16, 292, 312 constraint, 16 variation of a function, 216 variational inequality, 194 stochastic, 194 Von Mises statistical functional, 366

wait-and-see solution, 10, 60 weighted mean deviation, 292

534