

# LECTURES ON STOCHASTIC PROGRAMMING

## *MODELING AND THEORY*

**Alexander Shapiro**

Georgia Institute of Technology  
Atlanta, Georgia

**Darinka Dentcheva**

Stevens Institute of Technology  
Hoboken, New Jersey

**Andrzej Ruszczyński**

Rutgers University  
New Brunswick, New Jersey



Society for Industrial  
and Applied Mathematics  
Philadelphia



Mathematical Programming Society  
Philadelphia

Copyright © 2009 by the Society for Industrial and Applied Mathematics and the Mathematical Programming Society

10 9 8 7 6 5 4 3 2 1

All rights reserved. Printed in the United States of America. No part of this book may be reproduced, stored, or transmitted in any manner without the written permission of the publisher. For information, write to the Society for Industrial and Applied Mathematics, 3600 Market Street, 6th Floor, Philadelphia, PA 19104-2688 USA.

Trademarked names may be used in this book without the inclusion of a trademark symbol. These names are used in an editorial context only; no infringement of trademark is intended.

Cover image appears courtesy of Julia Shapiro.

### **Library of Congress Cataloging-in-Publication Data**

Shapiro, Alexander, 1949-

Lectures on stochastic programming : modeling and theory / Alexander Shapiro, Darinka Dentcheva, Andrzej Ruszczyński.

p. cm. -- (MPS-SIAM series on optimization ; 9)

Includes bibliographical references and index.


ISBN 978-0-898716-87-0

1. Stochastic programming. I. Dentcheva, Darinka. II. Ruszczyński, Andrzej P. III. Title.

T57.79.S54 2009

519.7--dc22

2009022942

 is a registered trademark.



is a registered trademark.

To Julia, Benjamin, Daniel, Natan, and Yael;

to Tsonka, Konstatin, and Marek;

and to the memory of Feliks, Maria, and Dentcho



# Contents

<b>List of Notations</b>	<b>xi</b>
<b>Preface</b>	<b>xiii</b>
<b>1 Stochastic Programming Models</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Inventory . . . . .	1
1.2.1 The News Vendor Problem . . . . .	1
1.2.2 Chance Constraints . . . . .	5
1.2.3 Multistage Models . . . . .	6
1.3 Multiproduct Assembly . . . . .	9
1.3.1 Two-Stage Model . . . . .	9
1.3.2 Chance Constrained Model . . . . .	10
1.3.3 Multistage Model . . . . .	12
1.4 Portfolio Selection . . . . .	13
1.4.1 Static Model . . . . .	13
1.4.2 Multistage Portfolio Selection . . . . .	16
1.4.3 Decision Rules . . . . .	21
1.5 Supply Chain Network Design . . . . .	22
Exercises . . . . .	25
<b>2 Two-Stage Problems</b>	<b>27</b>
2.1 Linear Two-Stage Problems . . . . .	27
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 . . . . .	32
2.1.4 Optimality Conditions . . . . .	38
2.2 Polyhedral 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 . . . . .	49
2.4 Nonanticipativity . . . . .	52

2.4.1	Scenario Formulation . . . . .	52
2.4.2	Dualization of Nonanticipativity Constraints . . . . .	54
2.4.3	Nonanticipativity Duality for General Distributions . . . . .	56
2.4.4	Value of Perfect Information . . . . .	59
	Exercises . . . . .	60
<b>3</b>	<b>Multistage Problems</b>	<b>63</b>
3.1	Problem Formulation . . . . .	63
3.1.1	The General Setting . . . . .	63
3.1.2	The Linear Case . . . . .	65
3.1.3	Scenario Trees . . . . .	69
3.1.4	Algebraic Formulation of Nonanticipativity Constraints . . . . .	71
3.2	Duality . . . . .	76
3.2.1	Convex Multistage Problems . . . . .	76
3.2.2	Optimality Conditions . . . . .	77
3.2.3	Dualization of Feasibility Constraints . . . . .	80
3.2.4	Dualization of Nonanticipativity Constraints . . . . .	82
	Exercises . . . . .	84
<b>4</b>	<b>Optimization Models with Probabilistic Constraints</b>	<b>87</b>
4.1	Introduction . . . . .	87
4.2	Convexity in Probabilistic Optimization . . . . .	94
4.2.1	Generalized Concavity of Functions and Measures . . . . .	94
4.2.2	Convexity of Probabilistically Constrained Sets . . . . .	106
4.2.3	Connectedness of Probabilistically Constrained Sets . . . . .	113
4.3	Separable Probabilistic Constraints . . . . .	114
4.3.1	Continuity and Differentiability Properties of Distribution Functions . . . . .	114
4.3.2	$p$ -Efficient Points . . . . .	115
4.3.3	Optimality Conditions and Duality Theory . . . . .	122
4.4	Optimization Problems with Nonseparable Probabilistic Constraints . . . . .	132
4.4.1	Differentiability of Probability Functions and Optimality Conditions . . . . .	133
4.4.2	Approximations of Nonseparable Probabilistic Constraints . . . . .	136
4.5	Semi-infinite Probabilistic Problems . . . . .	144
	Exercises . . . . .	150
<b>5</b>	<b>Statistical Inference</b>	<b>155</b>
5.1	Statistical Properties of Sample Average Approximation Estimators . . . . .	155
5.1.1	Consistency of SAA Estimators . . . . .	157
5.1.2	Asymptotics of the SAA Optimal Value . . . . .	163
5.1.3	Second Order Asymptotics . . . . .	166
5.1.4	Minimax Stochastic Programs . . . . .	170
5.2	Stochastic Generalized Equations . . . . .	174
5.2.1	Consistency of Solutions of the SAA Generalized Equations . . . . .	175

5.2.2	Asymptotics of SAA Generalized Equations Estimators . . . . .	177
5.3	Monte Carlo Sampling Methods . . . . .	180
5.3.1	Exponential Rates of Convergence and Sample Size Estimates in the Case of a Finite Feasible Set . . . . .	181
5.3.2	Sample Size Estimates in the General Case . . . . .	185
5.3.3	Finite Exponential Convergence . . . . .	191
5.4	Quasi-Monte Carlo Methods . . . . .	193
5.5	Variance-Reduction Techniques . . . . .	198
5.5.1	Latin Hypercube Sampling . . . . .	198
5.5.2	Linear Control Random Variables Method . . . . .	200
5.5.3	Importance Sampling and Likelihood Ratio Methods . . . . .	200
5.6	Validation Analysis . . . . .	202
5.6.1	Estimation of the Optimality Gap . . . . .	202
5.6.2	Statistical Testing of Optimality Conditions . . . . .	207
5.7	Chance Constrained Problems . . . . .	210
5.7.1	Monte Carlo Sampling Approach . . . . .	210
5.7.2	Validation of an Optimal Solution . . . . .	216
5.8	SAA Method Applied to Multistage Stochastic Programming . . . . .	220
5.8.1	Statistical Properties of Multistage SAA Estimators . . . . .	221
5.8.2	Complexity Estimates of Multistage Programs . . . . .	226
5.9	Stochastic Approximation Method . . . . .	230
5.9.1	Classical Approach . . . . .	230
5.9.2	Robust SA Approach . . . . .	233
5.9.3	Mirror Descent SA Method . . . . .	236
5.9.4	Accuracy Certificates for Mirror Descent SA Solutions . . . . .	244
	Exercises . . . . .	249
<b>6</b>	<b>Risk Averse Optimization</b> . . . . .	<b>253</b>
6.1	Introduction . . . . .	253
6.2	Mean-Risk Models . . . . .	254
6.2.1	Main Ideas of Mean-Risk Analysis . . . . .	254
6.2.2	Semideviations . . . . .	255
6.2.3	Weighted Mean Deviations from Quantiles . . . . .	256
6.2.4	Average Value-at-Risk . . . . .	257
6.3	Coherent Risk Measures . . . . .	261
6.3.1	Differentiability Properties of Risk Measures . . . . .	265
6.3.2	Examples of Risk Measures . . . . .	269
6.3.3	Law Invariant Risk Measures and Stochastic Orders . . . . .	279
6.3.4	Relation to Ambiguous Chance Constraints . . . . .	285
6.4	Optimization of Risk Measures . . . . .	288
6.4.1	Dualization of Nonanticipativity Constraints . . . . .	291
6.4.2	Examples . . . . .	295
6.5	Statistical Properties of Risk Measures . . . . .	300
6.5.1	Average Value-at-Risk . . . . .	300
6.5.2	Absolute Semideviation Risk Measure . . . . .	301
6.5.3	Von Mises Statistical Functionals . . . . .	304
6.6	The Problem of Moments . . . . .	306

6.7	Multistage Risk Averse Optimization . . . . .	308
6.7.1	Scenario Tree Formulation . . . . .	308
6.7.2	Conditional Risk Mappings . . . . .	315
6.7.3	Risk Averse Multistage Stochastic Programming . . . . .	318
	Exercises . . . . .	328
<b>7</b>	<b>Background Material</b>	<b>333</b>
7.1	Optimization and Convex Analysis . . . . .	334
7.1.1	Directional Differentiability . . . . .	334
7.1.2	Elements of Convex Analysis . . . . .	336
7.1.3	Optimization and Duality . . . . .	339
7.1.4	Optimality Conditions . . . . .	346
7.1.5	Perturbation Analysis . . . . .	351
7.1.6	Epiconvergence . . . . .	357
7.2	Probability . . . . .	359
7.2.1	Probability Spaces and Random Variables . . . . .	359
7.2.2	Conditional Probability and Conditional Expectation . . . . .	363
7.2.3	Measurable Multifunctions and Random Functions . . . . .	365
7.2.4	Expectation Functions . . . . .	368
7.2.5	Uniform Laws of Large Numbers . . . . .	374
7.2.6	Law of Large Numbers for Random Sets and Subdifferentials . . . . .	379
7.2.7	Delta Method . . . . .	382
7.2.8	Exponential Bounds of the Large Deviations Theory . . . . .	387
7.2.9	Uniform Exponential Bounds . . . . .	393
7.3	Elements of Functional Analysis . . . . .	399
7.3.1	Conjugate Duality and Differentiability . . . . .	401
7.3.2	Lattice Structure . . . . .	403
	Exercises . . . . .	405
<b>8</b>	<b>Bibliographical Remarks</b>	<b>407</b>
	<b>Bibliography</b>	<b>415</b>
	<b>Index</b>	<b>431</b>

# List of Notations

- $:=$ , equal by definition, 333  
 $A^T$ , transpose of matrix (vector)  $A$ , 333  
 $C(X)$ , space of continuous functions, 165  
 $C^*$ , polar of cone  $C$ , 337  
 $C^1(\mathcal{V}, \mathbb{R}^n)$ , space of continuously differentiable mappings, 176  
 $IF_{\mathcal{F}}$ , influence function, 304  
 $L^\perp$ , orthogonal of (linear) space  $L$ , 41  
 $O(1)$ , generic constant, 188  
 $O_p(\cdot)$ , term, 382  
 $S^\varepsilon$ , the set of  $\varepsilon$ -optimal solutions of the true problem, 181  
 $V_d(A)$ , Lebesgue measure of set  $A \subset \mathbb{R}^d$ , 195  
 $W^{1,\infty}(U)$ , space of Lipschitz continuous functions, 166, 353  
 $[a]_+ = \max\{a, 0\}$ , 2  
 $\mathbb{I}_A(\cdot)$ , indicator function of set  $A$ , 334  
 $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ , space, 399  
 $\Lambda(\bar{x})$ , set of Lagrange multipliers vectors, 348  
 $\mathcal{N}(\mu, \Sigma)$ , normal distribution, 16  
 $\mathcal{N}_C$ , normal cone to set  $C$ , 337  
 $\Phi(z)$ , cdf of standard normal distribution, 16  
 $\Pi_X$ , metric projection onto set  $X$ , 231  
 $\xrightarrow{\mathcal{D}}$ , convergence in distribution, 163  
 $\mathcal{T}_X^2(x, h)$ , second order tangent set, 348  
 $\text{AV@R}$ , Average Value-at-Risk, 258  
 $\tilde{\mathfrak{P}}$ , set of probability measures, 306  
 $\mathbb{D}(A, B)$ , deviation of set  $A$  from set  $B$ , 334  
 $\mathbb{D}[Z]$ , dispersion measure of random variable  $Z$ , 254  
 $\mathbb{E}$ , expectation, 361  
 $\mathbb{H}(A, B)$ , Hausdorff distance between sets  $A$  and  $B$ , 334  
 $\mathbb{N}$ , set of positive integers, 359  
 $\mathbb{R}^n$ ,  $n$ -dimensional space, 333  
 $\mathfrak{A}$ , domain of the conjugate of risk measure  $\rho$ , 262  
 $\mathfrak{C}_n$ , the space of nonempty compact subsets of  $\mathbb{R}^n$ , 379  
 $\mathfrak{P}$ , set of probability density functions, 263  
 $\mathfrak{S}_z$ , set of contact points, 399  
 $\mathfrak{b}(k; \alpha, N)$ , cdf of binomial distribution, 214  
 $\mathfrak{d}$ , distance generating function, 236  
 $\mathfrak{g}^+(x)$ , right-hand-side derivative, 297  
 $\text{cl}(A)$ , topological closure of set  $A$ , 334  
 $\text{conv}(C)$ , convex hull of set  $C$ , 337  
 $\text{Corr}(X, Y)$ , correlation of  $X$  and  $Y$ , 200  
 $\text{Cov}(X, Y)$ , covariance of  $X$  and  $Y$ , 180  
 $\mathfrak{q}_\alpha$ , weighted mean deviation, 256  
 $\mathfrak{s}_C(\cdot)$ , support function of set  $C$ , 337  
 $\text{dist}(x, A)$ , distance from point  $x$  to set  $A$ , 334  
 $\text{dom } f$ , domain of function  $f$ , 333  
 $\text{dom } \mathfrak{g}$ , domain of multifunction  $\mathfrak{g}$ , 365  
 $\overline{\mathbb{R}}$ , set of extended real numbers, 333  
 $\text{epi } f$ , epigraph of function  $f$ , 333  
 $\xrightarrow{\varepsilon}$ , epiconvergence, 377  
 $\hat{S}_N$ , the set of optimal solutions of the SAA problem, 156  
 $\hat{S}_N^\varepsilon$ , the set of  $\varepsilon$ -optimal solutions of the SAA problem, 181  
 $\hat{\vartheta}_N$ , optimal value of the SAA problem, 156  
 $\hat{f}_N(x)$ , sample average function, 155  
 $\mathbf{1}_A(\cdot)$ , characteristic function of set  $A$ , 334  
 $\text{int}(C)$ , interior of set  $C$ , 336  
 $[a]$ , integer part of  $a \in \mathbb{R}$ , 219  
 $\text{lsc } f$ , lower semicontinuous hull of function  $f$ , 333



- $\mathcal{R}_C$ , radial cone to set  $C$ , 337  
 $\mathcal{T}_C$ , tangent cone to set  $C$ , 337  
 $\nabla^2 f(x)$ , Hessian matrix of second order  
     partial derivatives, 179  
 $\partial$ , subdifferential, 338  
 $\partial^\circ$ , Clarke generalized gradient, 336  
 $\partial_\varepsilon$ , epsilon subdifferential, 380  
 $\text{pos } W$ , positive hull of matrix  $W$ , 29  
 $\Pr(A)$ , probability of event  $A$ , 360  
 $\text{ri}$ , relative interior, 337  
 $\sigma_p^+$ , upper semideviation, 255  
 $\sigma_p^-$ , lower semideviation, 255  
 $\text{V@R}_\alpha$ , Value-at-Risk, 256  
 $\mathbb{V}\text{ar}[X]$ , variance of  $X$ , 14  
 $\vartheta^*$ , optimal value of the true problem, 156  
 $\xi_{[t]} = (\xi_1, \dots, \xi_t)$ , history of the process,  
     63  
 $a \vee b = \max\{a, b\}$ , 186  
 $f^*$ , conjugate of function  $f$ , 338  
 $f^\circ(x, d)$ , generalized directional deriva-  
     tive, 336  
 $g'(x, h)$ , directional derivative, 334  
 $o_p(\cdot)$ , term, 382  
 $p$ -efficient point, 116  
 $\text{iid}$ , independently identically distributed,  
     156

# Preface

The main topic of this book is 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, from telecommunication and medicine to finance. 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 into seven 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-stage and multistage stochastic programming problems. We analyze properties of the models and develop optimality conditions and duality theory in a rather general setting. Our analysis covers general distributions of uncertain parameters and provides special results for discrete distributions, which are relevant for numerical methods. Due to specific properties of two- and multistage 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 in 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 constrained only by the computational resources needed to solve the constructed SAA problem. One of the byproducts of this theory is the complexity analysis of two-stage and multistage stochastic programming. Already in the case of two-stage stochastic programming, the number of scenarios (discretization points) grows exponentially with an increase in the number of random parameters. Furthermore, for multistage problems, the computational complexity also 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 refer to the literature. The subject index can help the reader quickly find 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, except in section 5.9, where the stochastic approximation method and its relation to complexity estimates are 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 a 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 the 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*

## Chapter 1

# Stochastic Programming Models

*Andrzej Ruszczyński and Alexander Shapiro*

## 1.1 Introduction

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. We can mention linear programming, quadratic programming, convex optimization, and nonlinear optimization. *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 a wealth of different problem settings, reflecting different aspects of the applied problem at hand. This chapter illustrates 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 help us to convey our main message.

## 1.2 Inventory

### 1.2.1 The News Vendor 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  $b \geq 0$ . The cost of this is equal to  $b(d - x)$  if  $d > x$  and is 0 otherwise. On the other hand, if  $d < x$ , then a holding cost of

$h(x - d) \geq 0$  is incurred. The total cost is then equal to<sup>1</sup>

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

We assume that  $b > c$ , i.e., the backorder 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 as

$$\text{Min}_{x \geq 0} F(x, d). \quad (1.2)$$

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

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

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 a 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

$$\text{Min}_{x \geq 0} \{ f(x) := \mathbb{E}[F(x, D)] \}. \quad (1.4)$$

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)  $H(x) := \Pr(D \leq x)$  of the random variable  $D$ . Note that  $H(x) = 0$  for all  $x < 0$ , because the demand cannot be negative. The expectation  $\mathbb{E}[F(x, D)]$  can be written in the following form:

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

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

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  $x \geq 0$  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  $D \geq 0$ , we have that  $f(0) = b\mathbb{E}[D]$ . Also, we have that

$$\begin{aligned} f'(z) &= c + \mathbb{E} \left[ \frac{\partial}{\partial z} (b[D - z]_+ + h[z - D]_+) \right] \\ &= c - b \Pr(D \geq z) + h \Pr(D \leq z) \\ &= c - b(1 - H(z)) + hH(z) \\ &= c - b + (b + h)H(z). \end{aligned}$$

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**

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

**Remark 1.** Recall that for  $\kappa \in (0, 1)$  the left-side  $\kappa$ -quantile of the cdf  $H(\cdot)$  is defined as  $H^{-1}(\kappa) := \inf\{t : H(t) \geq \kappa\}$ . In a similar way, the right-side  $\kappa$ -quantile is defined as  $\sup\{t : H(t) \leq \kappa\}$ . If the left and right  $\kappa$ -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  $\kappa$ -quantiles.

Suppose for the moment that the random variable  $D$  has **a finitely supported distribution**, i.e., it takes values  $d_1, \dots, d_K$  (called *scenarios*) with respective probabilities  $p_1, \dots, p_K$ . In that case, its cdf  $H(\cdot)$  is a step function with jumps of size  $p_k$  at each  $d_k$ ,  $k = 1, \dots, K$ . Formula (1.6) for an optimal solution still holds with the corresponding left-side (right-side)  $\kappa$ -quantile, coinciding with one of the points  $d_k$ ,  $k = 1, \dots, K$ . For example, the scenarios may represent historical data collected over a period of time. In such a case, the corresponding cdf is viewed as the **empirical cdf, giving an approximation (estimation) of the true cdf**, and the associated  $\kappa$ -quantile is viewed as the sample estimate of the  $\kappa$ -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 mentioned earlier, the optimal solution of such (deterministic) problem is  $\bar{d}$ . The mean  $\bar{d}$  can be very different from the  $\kappa$ -quantile  $\bar{x} = H^{-1}(\kappa)$ . It is also worth mentioning that **sample quantiles typically are 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 1. By using the representation (1.3), we can write problem (1.2) as the linear programming problem

$$\begin{aligned} \text{Min } & v \\ \text{s.t. } & v \geq (c - b)x + bd, \\ & v \geq (c + h)x - hd. \end{aligned} \quad (1.7)$$

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:

$$\begin{aligned} \text{Min } & \sum_{k=1}^K p_k v_k \\ \text{s.t. } & v_k \geq (c - b)x + bd_k, \quad k = 1, \dots, K, \\ & v_k \geq (c + h)x - hd_k, \quad k = 1, \dots, K. \end{aligned} \quad (1.8)$$

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 a 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

$$\text{Min } \max_{x \geq 0, d \in [l, u]} F(x, d). \quad (1.9)$$

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

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

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:

$$\text{Min}_{x \geq 0} \sup_{H \in \mathfrak{M}} \mathbb{E}_H[F(x, D)], \quad (1.10)$$

where  $\mathfrak{M}$  denotes the set of probability measures supported on the interval  $[l, u]$  and having mean  $\bar{d}$ , 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 6.6 (see also problem 6.8 on p. 330).

### 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  $\tau$  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:

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

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  $\tau/c$ , 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  $\tau$  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

$$\Pr\{F(x, D) > \tau\} \leq \alpha \quad (1.12)$$

or equivalently,

$$\Pr\{F(x, D) \leq \tau\} \geq 1 - \alpha. \quad (1.13)$$

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  $\tau$ ) is small (i.e., less than  $\alpha$ ). We have that

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

For  $x \leq \tau/c$ , the inequalities on the right-hand side of (1.14) are consistent, and hence for such  $x$ ,

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

The chance constraint (1.13) becomes

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

Even for small (but positive) values of  $\alpha$ , 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  $t = 1, \dots, T$ . At the beginning, at  $t = 1$ , there is (known) inventory level  $y_1$ . At each period  $t = 1, \dots, T$ , 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.,  $x_t \geq y_t$ . After the inventory is replenished, demand  $d_t$  is realized,<sup>2</sup> 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  $h_t \geq 0$ ,  $t = 1, \dots, T$ . 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:

$$\begin{aligned} \text{Min}_{x_t \geq y_t} \quad & \sum_{t=1}^T \mathbb{E}\{c_t(x_t - y_t) + b_t[D_t - x_t]_+ + h_t[x_t - D_t]_+\} \\ \text{s.t.} \quad & y_{t+1} = x_t - D_t, \quad t = 1, \dots, T-1. \end{aligned} \quad (1.17)$$

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.

<sup>2</sup>As before, we denote by  $d_t$  a particular realization of the random variable  $D_t$ .

## 1.2. Inventory

7

Consider the demand process  $D_t, t = 1, \dots, T$ . We denote by  $D_{[t]} := (D_1, \dots, D_t)$  the history of the demand process up to time  $t$ , and by  $d_{[t]} := (d_1, \dots, d_t)$  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

$$\text{Min}_{x_T \geq y_T} c_T(x_T - y_T) + \mathbb{E} \{ b_T[D_T - x_T]_+ + h_T[x_T - D_T]_+ | D_{[T-1]} = d_{[T-1]} \}. \quad (1.18)$$

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

$$\begin{aligned} \text{Min}_{x_{T-1} \geq y_{T-1}} & c_{T-1}(x_{T-1} - y_{T-1}) \\ & + \mathbb{E} \{ b_{T-1}[D_{T-1} - x_{T-1}]_+ + h_{T-1}[x_{T-1} - D_{T-1}]_+ \\ & + Q_T(x_{T-1} - D_{T-1}, D_{[T-1]}) | D_{[T-2]} = d_{[T-2]} \}. \end{aligned} \quad (1.19)$$

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

$$\begin{aligned} Q_t(y_t, d_{[t-1]}) = \min_{x_t \geq y_t} & c_t(x_t - y_t) + \mathbb{E} \{ b_t[D_t - x_t]_+ \\ & + h_t[x_t - D_t]_+ + Q_{t+1}(x_t - D_t, D_{[t]}) | D_{[t-1]} = d_{[t-1]} \}, \end{aligned} \quad (1.20)$$

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

$$\text{Min}_{x_1 \geq y_1} c_1(x_1 - y_1) + \mathbb{E} \{ b_1[D_1 - x_1]_+ + h_1[x_1 - D_1]_+ + Q_2(x_1 - D_1, D_1) \}. \quad (1.21)$$

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 backward 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  $D_{[t-1]}$ ,  $t = 2, \dots, T$ . 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  $t = T - 1, \dots, 2$ , 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  $t = 2, \dots, T$ , 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  $d_{[t-1]} = (d_1, \dots, d_{t-1})$  and decisions  $(x_1, \dots, x_{t-1})$ . Therefore, we may think about a sequence of possible decisions  $x_t = x_t(d_{[t-1]})$ ,  $t = 1, \dots, T$ , 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  $x_t \geq y_t$ ,  $t = 1, \dots, T$ , 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

characterization

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

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  $x_t \rightarrow +\infty$ . Similarly, as  $b_t > c_t$ , it also tends to  $+\infty$  if  $x_t \rightarrow -\infty$ . 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 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 backward in time. The situation starts to change dramatically with an 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.

algorithm

## 1.3 Multiproduct 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  $i = 1, \dots, n$  and  $j = 1, \dots, m$ . 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  $D = (D_1, \dots, D_n)$ . Before the demand is known, the manufacturer may preorder the parts from outside suppliers at a 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 not used are assessed salvage values  $s_j < c_j$ . The unsatisfied demand is lost.

Suppose the numbers of parts ordered are equal to  $x_j$ ,  $j = 1, \dots, m$ . 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  $z_i$ ,  $i = 1, \dots, n$ , and the numbers of parts left in inventory by  $y_j$ ,  $j = 1, \dots, m$ . For an observed value (a realization)  $d = (d_1, \dots, d_n)$  of the random demand vector  $D$ , we can find the best production plan by solving the following linear programming problem:

$$\begin{aligned} \text{Min}_{z,y} \quad & \sum_{i=1}^n (l_i - q_i) z_i - \sum_{j=1}^m s_j y_j \\ \text{s.t.} \quad & y_j = x_j - \sum_{i=1}^n a_{ij} z_i, \quad j = 1, \dots, m, \\ & 0 \leq z_i \leq d_i, \quad i = 1, \dots, n, \quad y_j \geq 0, \quad j = 1, \dots, m. \end{aligned}$$

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

$$\begin{aligned} \text{Min}_{z,y} \quad & (l - q)^T z - s^T y \\ \text{s.t.} \quad & y = x - A^T z, \\ & 0 \leq z \leq d, \quad y \geq 0. \end{aligned} \tag{1.23}$$

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 optimization problem

$$\text{Min}_{x \geq 0} \quad c^T x + \mathbb{E}[Q(x, D)], \tag{1.24}$$

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*, where (1.23) is called the *second-stage problem* and (1.24) is 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  $d^1, \dots, d^K$  occurring with positive probabilities  $p_1, \dots, p_K$ , 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:

$$\begin{aligned} \text{Min } c^T x + \sum_{k=1}^K p_k [(l - q)^T z^k - s^T y^k] \\ \text{s.t. } y^k = x - A^T z^k, \quad k = 1, \dots, K, \\ 0 \leq z^k \leq d^k, \quad y^k \geq 0, \quad k = 1, \dots, K, \\ x \geq 0, \end{aligned} \quad (1.25)$$

where the minimization is performed over vector variables  $x$  and  $z^k, y^k, k = 1, \dots, K$ . 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 worth noticing the following. There are three types of decision variables here: 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  $D = (D_1, \dots, D_n)$ , we need to have  $x \geq A^T D$ . If we have the parts needed, there is no need for the production planning stage, as in problem (1.23). We simply produce  $z_i = D_i, i = 1, \dots, n$ , 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

### 1.3. Multiproduct Assembly

11

formulation of the corresponding problem:

$$\begin{aligned} & \text{Min}_{x \geq 0} c^\top x \\ & \text{s.t. } \Pr \{A^\top D \leq x\} \geq 1 - \alpha. \end{aligned} \quad (1.26)$$

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

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

$$H_W(x) \geq 1 - \alpha, \quad (1.27)$$

where  $H_W(x) := \Pr(W \leq x)$  is the cumulative distribution function of the  $n$ -dimensional random vector  $W = A^\top D$ . 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

$$\{x \in \mathbb{R}_+^m : \Pr(A^\top D \leq x) \geq 1 - \alpha\}$$

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

$$\{x \in \mathbb{R}_+^m : A^\top d \leq x, \ d \in \mathfrak{D}, \ \Pr(\mathfrak{D}) \geq 1 - \alpha\}. \quad (1.28)$$

In the 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  $\mathfrak{D}_\alpha$  in such a way that  $\Pr(\mathfrak{D}_\alpha) \geq 1 - \alpha$ . In that way we obtain a simplified version of problem (1.26):

$$\begin{aligned} & \text{Min}_{x \geq 0} c^\top x \\ & \text{s.t. } A^\top d \leq x, \quad \forall d \in \mathfrak{D}_\alpha. \end{aligned} \quad (1.29)$$

The set  $\mathfrak{D}_\alpha$  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  $\mathfrak{D}_\alpha$  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  $D_t$ ,  $t = 1, \dots, T$ , where each  $D_t = (D_{t1}, \dots, D_{tn})$  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  $D_{[t]} := (D_1, \dots, D_t)$  to denote the history of the demand process in periods  $1, \dots, t$ . 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  $x_{t-1} = (x_{t-1,1}, \dots, x_{t-1,n})$  the vector of quantities ordered at the beginning of stage  $t$ , *before* the demand vector  $D_t$  becomes known. The numbers of units 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  $t = 1, \dots, T$ . 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:

$$\begin{aligned} \text{Min}_{z_T, y_T} \quad & (l - q)^T z_T - s^T y_T \\ \text{s.t.} \quad & y_T = y_{T-1} + x_{T-1} - A^T z_T, \\ & 0 \leq z_T \leq d_T, \quad y_T \geq 0, \end{aligned} \tag{1.30}$$

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 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

$$\begin{aligned} \text{Min}_{z_{T-1}, y_{T-1}, x_{T-1}} \quad & (l - q)^T z_{T-1} + h^T y_{T-1} + c^T x_{T-1} + \mathcal{Q}_T(x_{T-1}, y_{T-1}, d_{[T-1]}) \\ \text{s.t.} \quad & y_{T-1} = y_{T-2} + x_{T-2} - A^T z_{T-1}, \\ & 0 \leq z_{T-1} \leq d_{T-1}, \quad y_{T-1} \geq 0. \end{aligned} \tag{1.31}$$



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

$$\begin{aligned} \text{Min}_{z_t, y_t, x_t} \quad & (l - q)^\top z_t + h^\top y_t + c^\top x_t + Q_{t+1}(x_t, y_t, d_{[t]}) \\ \text{s.t.} \quad & y_t = y_{t-1} + x_{t-1} - A^\top z_t, \\ & 0 \leq z_t \leq d_t, \quad y_t \geq 0, \end{aligned} \quad (1.32)$$

with

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

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

$$\text{Min}_{x_0 \geq 0} c^\top x_0 + \mathbb{E}[Q_1(x_0, D_1)]. \quad (1.33)$$

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  $i = 1, \dots, n$ . 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  $\sum_{i=1}^n x_i \leq W_0$ . Suppose, further, that one possible investment 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:

$$\text{Max}_{x \geq 0} \mathbb{E}[W_1] \quad \text{s.t.} \quad \sum_{i=1}^n x_i = W_0. \quad (1.34)$$

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  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ . 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  $\mu^* := \max_{1 \leq i \leq n} \mu_i$ . Of course, from the practical point of view, such a solution is not very appealing. Putting everything into one asset can be very dangerous, because if its realized return rate is bad, 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:

$$\text{Max}_{x \geq 0} \mathbb{E}[U(W_1)] \quad \text{s.t.} \quad \sum_{i=1}^n x_i = W_0. \quad (1.35)$$

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

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

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:

$$\text{Max}_{x \geq 0} \mathbb{E}[Q(x, \xi)] \quad \text{s.t.} \quad \sum_{i=1}^n x_i = W_0, \quad (1.37)$$

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

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

We can view the above problem (1.38) as the second-stage program. Given a realization  $\xi = (\xi_1, \dots, \xi_n)$  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 in which the concept of risk can be formalized. For instance, we can evaluate risk by variability of  $W$  measured by its *variance*  $\text{Var}[W] = \mathbb{E}[W^2] - (\mathbb{E}[W])^2$ . Since  $W_1$  is a linear function of the random variables  $\xi_i$ , we have that

$$\text{Var}[W_1] = x^\top \Sigma x = \sum_{i,j=1}^n \sigma_{ij} x_i x_j,$$

where  $\Sigma = [\sigma_{ij}]$  is the covariance matrix of the random vector  $\xi$ . (Note that the covariance matrices of the random vectors  $\xi = (\xi_1, \dots, \xi_n)$  and  $R = (R_1, \dots, R_n)$  are identical.) This leads to the optimization problem of maximizing the expected return subject to the

#### 1.4. Portfolio Selection

15

additional constraint  $\text{Var}[W_1] \leq \nu$ , where  $\nu > 0$  is a specified constant. This problem can be written as

$$\text{Max}_{x \geq 0} \sum_{i=1}^n \mu_i x_i \quad \text{s.t.} \quad \sum_{i=1}^n x_i = W_0, \quad x^T \Sigma x \leq \nu. \quad (1.39)$$

Since the covariance matrix  $\Sigma$  is positive semidefinite, the constraint  $x^T \Sigma x \leq \nu$  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  $\text{Var}[W_1] = 0$ , 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:

$$\text{Min}_{\lambda \geq 0} \text{Max}_{\substack{\sum_{i=1}^n x_i = W_0 \\ x \geq 0}} \left\{ \sum_{i=1}^n \mu_i x_i - \lambda (x^T \Sigma x - \nu) \right\}. \quad (1.40)$$

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

$$\text{Max}_{x \geq 0} \sum_{i=1}^n \mu_i x_i - \bar{\lambda} x^T \Sigma x \quad \text{s.t.} \quad \sum_{i=1}^n x_i = W_0. \quad (1.41)$$

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  $\bar{\lambda} \nu$  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  $\text{Var}[W_1]$ , keeping the expected return  $\mathbb{E}[W_1]$  above a specified value  $\tau$ . That is,

$$\text{Min}_{x \geq 0} x^T \Sigma x \quad \text{s.t.} \quad \sum_{i=1}^n x_i = W_0, \quad \sum_{i=1}^n \mu_i x_i \geq \tau. \quad (1.42)$$

For appropriately chosen constants  $\nu$ ,  $\bar{\lambda}$ , and  $\tau$ , 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  $\xi$  and do not require complete knowledge of the probability distribution of  $\xi$ .

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

$$\text{Max}_{x \geq 0} \sum_{i=1}^n \mu_i x_i \quad \text{s.t.} \quad \sum_{i=1}^n x_i = W_0, \quad \Pr \left\{ \sum_{i=1}^n \xi_i x_i \geq b \right\} \geq 1 - \alpha. \quad (1.43)$$

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  $\xi$  has a

multivariate normal distribution with mean vector  $\mu$  and covariance matrix  $\Sigma$ , 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  $x^\top \Sigma x$ , and

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

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 form<sup>3</sup>

$$b - \sum_{i=1}^n \mu_i x_i + z_\alpha \sqrt{x^\top \Sigma x} \leq 0, \quad (1.45)$$

where  $z_\alpha := \Phi^{-1}(1 - \alpha)$  is the  $(1 - \alpha)$ -quantile of the standard normal distribution. Note that since matrix  $\Sigma$  is positive semidefinite,  $\sqrt{x^\top \Sigma x}$  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

$$\text{Max}_{x \geq 0} \sum_{i=1}^n \mu_i x_i - \eta \sqrt{x^\top \Sigma x} \quad \text{s.t.} \quad \sum_{i=1}^n x_i = W_0, \quad (1.46)$$

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.,

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

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

$$\text{V@R}_\alpha\left(b - \sum_{i=1}^n \xi_i x_i\right) \leq 0. \quad (1.48)$$

It is possible to write a 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  $\xi_i$  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  $t = 1, \dots, T - 1$  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  $x_0 = (x_{10}, \dots, x_{n0})$  be initial

<sup>3</sup>Note that if  $x^\top \Sigma x = 0$ , i.e.,  $\text{Var}(W_1) = 0$ , then the chance constraint of problem (1.43) holds iff  $\sum_{i=1}^n \mu_i x_i \geq b$ . In that case equivalence to the constraint (1.45) obviously holds.

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  $R_{1t}, \dots, R_{nt}$ , at periods  $t = 1, \dots, T$ , form a random process with a known distribution. Actually, we will work with the (vector valued) random process  $\xi_1, \dots, \xi_T$ , where  $\xi_t = (\xi_{1t}, \dots, \xi_{nt})$  and  $\xi_{it} := 1 + R_{it}$ ,  $i = 1, \dots, n$ ,  $t = 1, \dots, T$ . At time period  $t = 1$  we can rebalance the portfolio by specifying the amounts  $x_1 = (x_{11}, \dots, x_{n1})$  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  $\xi_1$ , i.e.,  $x_1 = x_1(\xi_1)$ . Similarly, at time  $t$  our decision  $x_t = (x_{1t}, \dots, x_{nt})$  is a function  $x_t = x_t(\xi_{[t]})$  of the available information given by realization  $\xi_{[t]} = (\xi_1, \dots, \xi_t)$  of the data process up to time  $t$ . A sequence of specific functions  $x_t = x_t(\xi_{[t]})$ ,  $t = 0, 1, \dots, T - 1$ , 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  $x_{it}(\xi_{[t]}) \geq 0$ ,  $i = 1, \dots, n$ ,  $t = 0, \dots, T - 1$ , and the balance of wealth constraints

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

At period  $t = 1, \dots, T$ , 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

$$\text{Max } \mathbb{E}[U(W_T)]. \quad (1.49)$$

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  $R_1, \dots, R_T$ . 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, independent of other assets, then the total number of scenarios is  $2^{nT}$ . It also should be ensured that  $1 + R_{it} \geq 0$ ,  $i = 1, \dots, n$ ,  $t = 1, \dots, T$ , for all possible realizations of the random data.

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

$$\begin{aligned} & \text{Max}_{x_{T-1} \geq 0, W_T} \mathbb{E}\{U[W_T] | \xi_{[T-1]}\} \\ & \text{s.t. } W_T = \sum_{i=1}^n \xi_{iT} x_{i,T-1}, \quad \sum_{i=1}^n x_{i,T-1} = W_{T-1}, \end{aligned} \quad (1.50)$$

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  $t = T - 2, \dots, 1$ , we consider the problem

$$\begin{aligned} \text{Max}_{x_t \geq 0, W_{t+1}} \quad & \mathbb{E}\{Q_{t+1}(W_{t+1}, \xi_{[t+1]})|\xi_{[t]}\} \\ \text{s.t. } W_{t+1} = & \sum_{i=1}^n \xi_{i,t+1}x_{i,t}, \quad \sum_{i=1}^n x_{i,t} = W_t, \end{aligned} \quad (1.51)$$

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

$$\begin{aligned} \text{Max}_{x_0 \geq 0, W_1} \quad & \mathbb{E}[Q_1(W_1, \xi_1)] \\ \text{s.t. } W_1 = & \sum_{i=1}^n \xi_{i1}x_{i0}, \quad \sum_{i=1}^n x_{i0} = W_0. \end{aligned} \quad (1.52)$$

For a general distribution of the data process  $\xi_t$ , it may be hard to solve these dynamic programming equations. The situation simplifies dramatically if the process  $\xi_t$  is stagewise independent, i.e.,  $\xi_t$  is (stochastically) independent of  $\xi_1, \dots, \xi_{t-1}$  for  $t = 2, \dots, T$ . 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  $Q_t(W_t)$ ,  $t = 1, \dots, T - 1$ , does not depend on  $\xi_{[t]}$ . That is,  $Q_{T-1}(W_{T-1})$  is the optimal value of the problem

$$\begin{aligned} \text{Max}_{x_{T-1} \geq 0, W_T} \quad & \mathbb{E}\{U[W_T]\} \\ \text{s.t. } W_T = & \sum_{i=1}^n \xi_{iT}x_{i,T-1}, \quad \sum_{i=1}^n x_{i,T-1} = W_{T-1}, \end{aligned}$$

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

$$\begin{aligned} \text{Max}_{x_t \geq 0, W_{t+1}} \quad & \mathbb{E}\{Q_{t+1}(W_{t+1})\} \\ \text{s.t. } W_{t+1} = & \sum_{i=1}^n \xi_{i,t+1}x_{i,t}, \quad \sum_{i=1}^n x_{i,t} = W_t \end{aligned}$$

for  $t = T - 2, \dots, 1$ .

The other relevant question is what utility function to use. Let us consider the *logarithmic* 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:

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

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

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

By (1.53) its optimal value is

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

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

$$\begin{aligned} \text{Max}_{x_{T-2} \geq 0} \mathbb{E} \left\{ v_{T-1}(\xi_{[T-1]}) + \ln \left( \sum_{i=1}^n \xi_{i,T-1} x_{i,T-2} \right) \middle| \xi_{[T-2]} \right\} \\ \text{s.t.} \quad \sum_{i=1}^n x_{i,T-2} = W_{T-2}. \end{aligned} \quad (1.55)$$

Of course, we have that

$$\begin{aligned} \mathbb{E} \left\{ v_{T-1}(\xi_{[T-1]}) + \ln \left( \sum_{i=1}^n \xi_{i,T-1} x_{i,T-2} \right) \middle| \xi_{[T-2]} \right\} \\ = \mathbb{E} \left\{ v_{T-1}(\xi_{[T-1]}) \middle| \xi_{[T-2]} \right\} + \mathbb{E} \left\{ \ln \left( \sum_{i=1}^n \xi_{i,T-1} x_{i,T-2} \right) \middle| \xi_{[T-2]} \right\}, \end{aligned}$$

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

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

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

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

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

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

in a completely myopic fashion.

By definition, we set  $\xi_0$  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, the first-stage optimal solution  $\bar{x}_0$  is given by an optimal solution of the problem

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

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

$$\vartheta^* = \ln W_0 + v_0 + \sum_{t=1}^{T-1} \mathbb{E} [v_t(\xi_{[t]})], \quad (1.58)$$

where  $v_t(\xi_{[t]})$  is the optimal value of problem (1.56) for  $W_t = 1$ . Note that  $v_0 + \ln W_0$  is the optimal value of problem (1.57) with  $v_0$  being the (deterministic) optimal value of (1.57) for  $W_0 = 1$ .

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

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

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

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

defines the optimal policy.

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

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

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

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

The optimal value  $\vartheta^*$  of the corresponding multistage problem (1.49) is

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

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

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

As 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 nonoptimal. However, it is readily implementable, once the current wealth  $W_t$  is observed. As mentioned, rules for calculating decisions as functions of the observations gathered up to time  $t$ , similar 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

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

Since the random process  $\xi_1, \dots, \xi_T$  is stagewise independent, by independence of  $\xi_{t+1}$  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,

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

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

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

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

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

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

It follows from (1.64) and (1.66) that

$$\frac{\mathbb{V}\text{ar}[W_{t+1}]}{(\mathbb{E}[W_{t+1}])^2} = \frac{x_t^{*\top} \Sigma_{t+1} x_t^*}{(x_t^{*\top} \mu_{t+1})^2} + \frac{\mathbb{V}\text{ar}[W_t]}{(\mathbb{E}[W_t])^2} \quad (1.67)$$

and hence

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

This shows that if the terms  $x_{\tau-1}^{*\top} \Sigma_{\tau} x_{\tau-1}^* / (x_{\tau-1}^{*\top} \mu_{\tau})^2$  are of the same order for  $\tau = 1, \dots, T$ , then the ratio of the standard deviation  $\sqrt{\text{Var}[W_T]}$  to the expected wealth  $\mathbb{E}[W_T]$  is of order  $O(\sqrt{T})$  with an increase in 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. We first describe a deterministic mathematical formulation for the supply chain design problem.

Denote by  $\mathcal{S}$ ,  $\mathcal{P}$ , and  $\mathcal{C}$  the respective (finite) sets of suppliers, processing facilities, and customers. The union  $\mathcal{N} := \mathcal{S} \cup \mathcal{P} \cup \mathcal{C}$  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:

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

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

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

## 1.5. Supply Chain Network Design

23

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

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

$$x \in \mathcal{X}, \quad y \geq 0. \quad (1.74)$$

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 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 be procured only 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:

$$\text{Min}_{x \in \mathcal{X}, y \geq 0} c^\top x + q^\top y \quad (1.75)$$

$$\text{s.t.} \quad Ny = 0, \quad (1.76)$$

$$Cy \geq d, \quad (1.77)$$

$$Sy \leq s, \quad (1.78)$$

$$Ry \leq Mx, \quad (1.79)$$

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 the time at which a decision about vector  $x \in \mathcal{X}$  should be made, i.e., which facilities to build 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:

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

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

$$\text{Min}_{y \geq 0} q^T y \quad (1.81)$$

$$\text{s.t. } Ny = 0, \quad (1.82)$$

$$Cy \geq d, \quad (1.83)$$

$$Sy \leq s, \quad (1.84)$$

$$Ry \leq Mx \quad (1.85)$$

with  $\xi = (q, d, s, R, M)$  being the vector of the involved parameters. Of course, the above optimization problem depends on the data vector  $\xi$ . 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  $\xi$  as random variables with a specified joint probability distribution. Then we can formulate the stochastic programming problem

$$\text{Min}_{x \in \mathcal{X}} c^T x + \mathbb{E}[Q(x, \xi)], \quad (1.86)$$

where the expectation is taken with respect to the probability distribution of the random vector  $\xi$ . 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  $\xi$ , 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 modeled by writing the second-stage problem in the form

$$\text{Min}_{y \geq 0, z \geq 0} q^T y + h^T z \quad (1.87)$$

$$\text{s.t. } Ny = 0, \quad (1.88)$$

$$Cy + z \geq d, \quad (1.89)$$

$$Sy \leq s, \quad (1.90)$$

$$Ry \leq Mx, \quad (1.91)$$

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  $z \geq d$  clearly satisfy the constraints of this problem.

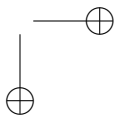
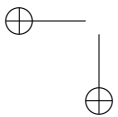
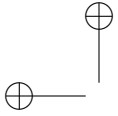
## Exercises

- 1.1. Consider the expected value function  $f(x) := \mathbb{E}[F(x, D)]$ , where function  $F(x, d)$  is defined in (1.1). (i) Show that function  $F(x, d)$  is convex in  $x$  and hence that  $f(x)$  is also convex. (ii) Show that  $f(\cdot)$  is differentiable at a point  $x > 0$  iff the cdf  $H(\cdot)$  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  $H^{-1}(\kappa) = \inf\{t : H(t) \geq \kappa\}$  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  $\tau/c$ . (ii) Verify equation (1.15). (iii) Assume that the probability distribution of the demand  $D$  is supported on an interval  $[l, u]$  with  $0 \leq l \leq u < +\infty$ . Show that if the significance level  $\alpha = 0$ , then the constraint (1.16) becomes

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

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

- 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 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  $W_1 = \xi^\top x$  is given by formula  $\mathbb{V}\text{ar}[W_1] = x^\top \Sigma x$ , where  $\Sigma = \mathbb{E}[(\xi - \mu)(\xi - \mu)^\top]$  is the covariance matrix of the random vector  $\xi$  and  $\mu = \mathbb{E}[\xi]$ .
- 1.10. Show that the optimal value function  $Q_t(W_t, \xi_{[t]})$ , defined in (1.51), is convex in  $W_t$ .
- 1.11. Let  $D$  be a random variable with cdf  $H(t) = \Pr(D \leq t)$  and  $D^1, \dots, D^N$  be an iid random sample of  $D$  with the corresponding empirical cdf  $\hat{H}_N(\cdot)$ . Let  $a = H^{-1}(\kappa)$  and  $b = \sup\{t : H(t) \leq \kappa\}$  be respective left- and right-side  $\kappa$ -quantiles of  $H(\cdot)$ . Show that  $\min\{|\hat{H}_N^{-1}(\kappa) - a|, |\hat{H}_N^{-1}(\kappa) - b|\}$  tends w.p. 1 to 0 as  $N \rightarrow \infty$ .



## 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

$$\begin{aligned} \text{Min}_{x \in \mathbb{R}^n} \quad & c^\top x + \mathbb{E}[Q(x, \xi)] \\ \text{s.t.} \quad & Ax = b, \quad x \geq 0, \end{aligned} \tag{2.1}$$

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

$$\begin{aligned} \text{Min}_{y \in \mathbb{R}^m} \quad & q^\top y \\ \text{s.t.} \quad & Tx + Wy = h, \quad y \geq 0. \end{aligned} \tag{2.2}$$

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

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

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

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

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

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

$$\Pi(q) := \{\pi : W^T\pi \leq q\} \quad (2.5)$$

is nonempty, then

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

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  $\xi$ , the function  $Q(\cdot, \xi)$  is convex. Moreover, if the set  $\{\pi : W^T\pi \leq q\}$  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  $s_q(\cdot)$ .  $\square$

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*

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

where

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

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



## 2.1. Linear Two-Stage Problems

29

**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))

$$\partial s_q(\chi_0) = \arg \max_{\pi} \{ \pi^\top \chi_0 - \mathbb{I}_q(\pi) \} = \arg \max_{\pi \in \Pi(q)} \pi^\top \chi_0. \quad (2.8)$$

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.  $\square$

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

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

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

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

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

$$\Pi_0 := \Pi(0) = \{ \pi : W^\top \pi \leq 0 \}. \quad (2.10)$$

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

$$\Pi_0^* = \text{pos } W. \quad (2.11)$$

If  $\chi_0 \in \text{int}(\text{pos } W)$ , 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  $\pi_0^\top \chi_0 = 0$ . By perturbing  $\chi_0$  a little to some  $\chi$ , we would be able to keep  $\chi$  within  $\text{pos } W$  and get  $\pi_0^\top \chi > 0$ , which is a contradiction, because  $\text{pos } W$  is the polar of  $\Pi_0$ . Therefore the set of maximizers in (2.6) is the convex hull of the vertices  $v$  of  $\Pi(q)$  for which  $v^\top \chi = s_q(\chi)$ . Note that  $\Pi(q)$  must have vertices in this case, because otherwise the polar to  $\Pi_0$  would have no interior.

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

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

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

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

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

$$\begin{aligned} \text{Min}_{y_1, \dots, y_K} \quad & \sum_{k=1}^K p_k q_k^\top y_k \\ \text{s.t.} \quad & T_k x + W_k y_k = h_k, \\ & y_k \geq 0, \quad k = 1, \dots, K. \end{aligned} \quad (2.14)$$

If for at least one  $k \in \{1, \dots, K\}$  the system  $T_k x + W_k y_k = h_k, y_k \geq 0$ , 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:

$$\begin{aligned} \text{Min}_{x, y_1, \dots, y_K} \quad & c^\top x + \sum_{k=1}^K p_k q_k^\top y_k \\ \text{s.t.} \quad & T_k x + W_k y_k = h_k, \quad k = 1, \dots, K, \\ & Ax = b, \\ & x \geq 0, \quad y_k \geq 0, \quad k = 1, \dots, K. \end{aligned} \quad (2.15)$$

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

**Proposition 2.3.** *Suppose that the probability distribution of  $\xi$  has finite support  $\Xi = \{\xi_1, \dots, \xi_K\}$  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  $x_0 \in \text{dom } \phi$ ,*

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

## 2.1. Linear Two-Stage Problems

31

**Proof.** Since  $\phi(\bar{x})$  is finite, all values  $Q(\bar{x}, \xi_k)$ ,  $k = 1, \dots, K$ , 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  $\text{dom } \phi = \bigcap_{k=1}^K \text{dom } Q_k$ , 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

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

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.  $\square$

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

$$\partial Q(x_0, \xi_k) = -T_k^\top \arg \max \{ \pi^\top (h_k - T_k x_0) : W_k^\top \pi \leq q_k \}. \quad (2.18)$$

It follows that the expectation function  $\phi$  is differentiable at  $x_0$  iff for every  $\xi = \xi_k$ ,  $k = 1, \dots, K$ , 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  $x \geq x^{\min}$ , where  $x^{\min}$  are the existing capacities.

At each node  $n$  of the graph, we have a random demand  $\xi_n$  for shipments to  $n$ . (If  $\xi_n$  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  $\xi_n$  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  $n$ . The second-stage problem is the network flow problem

$$\text{Min } \sum_{a \in \mathcal{A}} q_a y_a \quad (2.19)$$

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

$$0 \leq y_a \leq x_a, \quad a \in \mathcal{A}. \quad (2.21)$$

This problem depends on the random demand vector  $\xi$  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  $\mu_n$ ,  $n \in \mathcal{N}$ , the optimal Lagrange multipliers (node potentials) associated with the node balance equations (2.20), and denote by  $\pi_a$ ,  $a \in \mathcal{A}$ , the (nonnegative) Lagrange multipliers associated with the constraints (2.21). The dual problem has the form

$$\begin{aligned} \text{Max} \quad & - \sum_{n \in \mathcal{N}} \xi_n \mu_n - \sum_{(i,j) \in \mathcal{A}} x_{ij} \pi_{ij} \\ \text{s.t.} \quad & -\pi_{ij} + \mu_i - \mu_j \leq q_{ij}, \quad (i, j) \in \mathcal{A}, \\ & \pi \geq 0. \end{aligned}$$

As  $\sum_{n \in \mathcal{N}} \xi_n = 0$ , the values of  $\mu_n$  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  $\pi_{ij}$  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  $y_{ij} \leq x_{ij}$  becomes relevant. The dual problem becomes equivalent to

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

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  $T^T = [0 \ -I]$  in this case, formula (2.18) provides the description of the subdifferential of  $Q(\cdot, \xi)$  at  $x_0$ :

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

The first-stage problem has the form

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

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

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

### 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 semi-continuous function (see section 7.2.3). Therefore, it follows by Theorem 7.37 that  $Q(\cdot, \cdot)$

## 2.1. Linear Two-Stage Problems

33

is measurable with respect to the Borel sigma algebra of  $\mathbb{R}^n \times \mathbb{R}^d$ . Also for every  $\xi$  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  $\mathbb{E}[(-Q(x, \xi))_+]$  is 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  $y \geq 0$  has a solution for every  $\chi$ . 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  $W^T \pi \leq 0$ .

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

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

the feasible set of the second-stage problem (2.2) is nonempty for almost everywhere (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,  $Q(x, \xi(\omega)) < +\infty$  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:

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

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  $\xi$  has a finite support, or
- (ii) the recourse is fixed.

Indeed, sufficiency is clear. If  $\xi$  has a finite support, i.e., the set  $\Xi$  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  $x \in X$ , then  $Q(x, \xi) < +\infty$  for all  $\xi$  in  $\Xi$ , except possibly for a subset of  $\Xi$  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  $\text{pos } W$  is convex and closed and thus  $\Xi_0(x)$  is convex and closed as well. By assumption,  $P[\Xi_0(x)] = 1$  for every  $x \in X$ . Thus  $\bigcap_{x \in X} \Xi_0(x)$  is convex, closed, and has probability 1. The support of  $\xi$  must be its subset.

**Example 2.5.** Consider

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

with  $x \in [0, 1]$  and  $\xi$  being a random variable whose probability density function is  $p(z) := 2z$ ,  $0 \leq z \leq 1$ . 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  $\xi$  may change that. Take, for example, some discretization of the distribution of  $\xi$  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 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 7.11), there exists a constant  $\kappa$ , 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:

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

where  $B := \{\pi : \|\pi\| \leq 1\}$  and  $\|\cdot\|$  denotes the Euclidean norm. This inclusion allows us 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

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

Consider  $q_0 = 0$ . The support function  $s_0(\cdot)$  of the cone  $\Pi_0$  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  $s_q(\chi) \leq \kappa \|q\| \|\chi\|$  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  $\text{pos } W$ , and

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

**Proposition 2.6.** *Suppose that the recourse is fixed and*

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

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

$$h - Tx \in \text{pos } W. \quad (2.29)$$

**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  $\kappa$  such that for any  $\chi$ ,

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

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

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

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

**Remark 2.** If  $q$  and  $(h, T)$  are independent and have finite first moments,<sup>4</sup> then

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

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  $\xi$ , 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, and (iii) condition (2.28) holds.*

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

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

**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

<sup>4</sup>We 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.

because  $\Pi(q)$  is closed. By Hoffman's lemma (see (2.25)) there is a constant  $\kappa$  such that  $\|\pi(q)\| \leq \kappa \|q\|$ . Then for every  $x$  the following holds w.p. 1:

$$s_q(h - Tx) \geq \pi(q)^T(h - Tx) \geq -\kappa \|q\|(\|h\| + \|T\| \|x\|). \quad (2.31)$$

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  $\text{dom } \phi$  follow from the convexity and lower semicontinuity of  $\phi$ . 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:

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

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

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

It follows that

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

With condition (2.28) this implies the Lipschitz continuity of  $\phi$  on its domain.  $\square$

Denote by  $\Sigma$  the support<sup>5</sup> of the probability distribution (measure) of  $(h, T)$ . Formula (2.30) means that a point  $x$  belongs to  $\text{dom } \phi$  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  $\text{dom } \phi$  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

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

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  $\text{pos } W$  is a convex polyhedral cone and hence can be represented as the intersection of a finite number of half spaces  $A_i = \{\chi : a_i^T \chi \leq 0\}$ ,  $i = 1, \dots, \ell$ . 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 that the set  $\bigcap_{h \in \mathcal{H}} (-h + \text{pos } W)$  can be represented as the intersection of half spaces of the form  $b_i + A_i$ ,  $i = 1, \dots, \ell$ , 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 7.47 and formula (2.7) of Proposition 2.2 we have the following result.

<sup>5</sup>Recall that the support of the probability measure is the smallest closed set such that the probability (measure) of its complement is zero.



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

$$\partial\phi(x_0) = -\mathbb{E}[T^\top \mathcal{D}(x_0, \xi)] + \mathcal{N}_{\text{dom } \phi}(x_0), \quad (2.34)$$

where

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

Moreover,  $\phi$  is differentiable at  $x_0$  iff  $x_0$  belongs to the interior of  $\text{dom } \phi$  and the set  $\mathcal{D}(x_0, \xi)$  is a singleton w.p. 1.

As discussed earlier, when the distribution of  $\xi$  has a finite support (i.e., there is a finite number of scenarios), the expectation function  $\phi$  is piecewise linear on its domain and is differentiable everywhere only in the trivial case if it is linear.<sup>6</sup> In the case of a continuous distribution of  $\xi$ , the expectation operator smoothes the piecewise linear function  $Q(\cdot, \xi)$ .

**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  $\phi$  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  $\text{dom } \phi$ . For fixed  $T$  and  $q$ , consider the multifunction

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

Conditional on  $(T, q)$ , the set  $\mathcal{D}(x, \xi)$  coincides with  $\mathfrak{Z}(h)$ . Since  $x \in \text{dom } \phi$ , relation (2.30) implies that  $h - Tx \in \text{pos } W$  w.p. 1. For every  $h - Tx \in \text{pos } W$ , 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  $h - Tx \in (\text{pos } W) \setminus A$ , 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 event “ $\mathcal{D}(x, \xi)$  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  $x \in \text{int}(\text{dom } \phi)$  the gradient  $\nabla\phi(x)$  coincides with the (unique) subgradient of  $\phi$  at  $x$  and that  $\nabla\phi(\cdot)$  is continuous at  $x$ .  $\square$

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.

<sup>6</sup>By linear, we mean here that it is of the form  $a^\top x + b$ . It is more accurate to call such a function affine.

### 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

$$\begin{aligned} \text{Min}_x \quad & c^\top x + \sum_{k=1}^K p_k Q(x, \xi_k) \\ \text{s.t.} \quad & Ax = b, \quad x \geq 0, \end{aligned} \quad (2.35)$$

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  $x_0 \in \text{dom } \phi$ ,

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

where

$$\mathfrak{D}(x_0, \xi_k) := \arg \max \{ \pi^\top (h_k - T_k x_0) : W_k^\top \pi \leq q_k \}.$$

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

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

$$\begin{aligned} \sum_{k=1}^K p_k T_k^\top \pi_k + A^\top \mu &\leq c, \\ \bar{x}^\top \left( c - \sum_{k=1}^K p_k T_k^\top \pi_k - A^\top \mu \right) &= 0. \end{aligned} \quad (2.37)$$

**Proof.** Necessary and sufficient optimality conditions for minimizing  $c^\top x + \phi(x)$  over  $x \in X$  can be written as

$$0 \in c + \partial\phi(\bar{x}) + \mathcal{N}_X(\bar{x}), \quad (2.38)$$

where  $\mathcal{N}_X(\bar{x})$  is the normal cone to the feasible set  $X$ . Note that condition (2.38) implies that the sets  $\mathcal{N}_X(\bar{x})$  and  $\partial\phi(\bar{x})$  are nonempty and hence  $\bar{x} \in X$  and  $\phi(\bar{x})$  is finite. Note also that there is no need here for additional regularity conditions since  $\phi(\cdot)$  and  $X$  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

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

## 2.1. Linear Two-Stage Problems

39

Observe that

$$\mathcal{N}_X(\bar{x}) = \{A^\top \mu - h : h \geq 0, h^\top \bar{x} = 0\}. \quad (2.39)$$

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

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

$$\begin{aligned} \text{Min}_{x, y_1, \dots, y_K} \quad & c^\top x + \sum_{k=1}^K p_k q_k^\top y_k \\ \text{s.t.} \quad & T_k x + W_k y_k = h_k, \quad k = 1, \dots, K, \\ & Ax = b, \\ & x \geq 0, \\ & y_k \geq 0, \quad k = 1, \dots, K. \end{aligned} \quad (2.40)$$

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

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

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

$$\begin{aligned} \text{Max}_{\mu, \pi_1, \dots, \pi_K} \quad & b^\top \mu + \sum_{k=1}^K p_k h_k^\top \pi_k \\ \text{s.t.} \quad & c - A^\top \mu - \sum_{k=1}^K p_k T_k^\top \pi_k \geq 0, \\ & q_k - W_k^\top \pi_k \geq 0, \quad k = 1, \dots, K. \end{aligned}$$

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

$$\begin{aligned} & \sum_{k=1}^K p_k T_k^\top \pi_k + A^\top \mu \leq c, \\ & \bar{x}^\top \left( c - \sum_{k=1}^K p_k T_k^\top \pi_k - A^\top \mu \right) = 0, \\ & q_k - W_k^\top \pi_k \geq 0, \quad k = 1, \dots, K, \\ & \bar{y}_k^\top (q_k - W_k^\top \pi_k) = 0, \quad k = 1, \dots, K. \end{aligned}$$

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

If we deal with general distributions of the 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,  $\text{int}(\text{dom } \phi) \cap X$  is nonempty, and  $\mathcal{N}_{\text{dom } \phi}(\bar{x}) \subset \mathcal{N}_X(\bar{x})$ . Then  $\bar{x}$  is an optimal solution of problem (2.1)–(2.2) iff there exist a measurable function  $\pi(\omega) \in \mathfrak{D}(x, \xi(\omega))$ ,  $\omega \in \Omega$ , and a vector  $\mu \in \mathbb{R}^m$  such that*

$$\begin{aligned}\mathbb{E}[T^\top \pi] + A^\top \mu &\leq c, \\ \bar{x}^\top (c - \mathbb{E}[T^\top \pi] - A^\top \mu) &= 0.\end{aligned}$$

**Proof.** Since  $\text{int}(\text{dom } \phi) \cap X$  is nonempty, we have by the Moreau–Rockafellar theorem that

$$\partial (c^\top \bar{x} + \phi(\bar{x}) + \mathbb{I}_X(\bar{x})) = c + \partial \phi(\bar{x}) + \partial \mathbb{I}_X(\bar{x}).$$

Also,  $\partial \mathbb{I}_X(\bar{x}) = \mathcal{N}_X(\bar{x})$ . Therefore, we have here that (2.38) is necessary and sufficient optimality conditions for minimizing  $c^\top x + \phi(x)$  over  $x \in X$ . Using the characterization of the subdifferential of  $\phi(\cdot)$  given in (2.8), we conclude that (2.38) is equivalent to existence of a measurable function  $\pi(\omega) \in \mathfrak{D}(x_0, \xi(\omega))$  such that

$$0 \in c - \mathbb{E}[T^\top \pi] + \mathcal{N}_{\text{dom } \phi}(\bar{x}) + \mathcal{N}_X(\bar{x}). \quad (2.41)$$

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

The additional technical condition  $\mathcal{N}_{\text{dom } \phi}(\bar{x}) \subset \mathcal{N}_X(\bar{x})$  was needed in the above derivations in order to eliminate the term  $\mathcal{N}_{\text{dom } \phi}(\bar{x})$  in (2.41). In particular, this condition holds if  $\bar{x} \in \text{int}(\text{dom } \phi)$ , in which case  $\mathcal{N}_{\text{dom } \phi}(\bar{x}) = \{0\}$ , or in the case of relatively complete recourse, i.e., when  $X \subset \text{dom } \phi$ . 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,  $\text{int}(\text{dom } \phi) \cap X$  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  $\pi(\omega) \in \mathfrak{D}(x, \xi(\omega))$ ,  $\omega \in \Omega$ , and a vector  $\mu \in \mathbb{R}^m$  such that*

$$\begin{aligned}T^\top \mathbb{E}[\pi] + A^\top \mu &\leq c, \\ \bar{x}^\top (c - T^\top \mathbb{E}[\pi] - A^\top \mu) &= 0.\end{aligned}$$

## 2.1. Linear Two-Stage Problems

41

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

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

Now we need to calculate the cone  $\mathcal{N}_{\text{dom } \phi}(\bar{x})$ . 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

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

where  $\Sigma$  is the support of the distribution of the random vector  $h$ . The tangent cone to  $\text{dom } \phi$  at  $\bar{x}$  has the form

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

Defining the linear subspace

$$L := \bigcap_{h \in \Sigma} \text{lin}(-h + T\bar{x}),$$

we can write the tangent cone as

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

Therefore the normal cone equals

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

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

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

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

$$\begin{aligned} -T^\top \mathbb{E}[\mathcal{D}(\bar{x}, \xi)] + \mathcal{N}_{\text{dom } \phi}(\bar{x}) &= -T^\top \mathbb{E}[\mathcal{D}(\bar{x}, \xi)] - T^\top (\Pi_0 \cap L^\perp) \\ &= -T^\top \mathbb{E}[\mathcal{D}(\bar{x}, \xi)], \end{aligned}$$

and the result follows.  $\square$

**Example 2.13 (Capacity Expansion, continued).** Let us return to Example 2.13 and suppose the support  $\Xi$  of the random demand vector  $\xi$  is compact. Only the right-hand side  $\xi$  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:

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

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

In particular, for every  $(i, j) \in \mathcal{A}$  such that  $\bar{x}_{ij} > x_{ij}^{\min}$  we have equality in (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,

$$\text{Min}_x f_1(x) + \mathbb{E}[Q(x, \omega)], \quad (2.44)$$

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

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

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  $\alpha_j$ ,  $j = 1, \dots, J_1$ , vectors  $a_k$  and scalars  $b_k$ ,  $k = 1, \dots, K_1$ , such that  $f_1(x)$  can be represented as follows:

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

and its domain  $\text{dom } f_1 = \{x : a_k^\top x \leq b_k, k = 1, \dots, K_1\}$  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  $\gamma_j = \gamma_j(\omega)$ ,  $j = 1, \dots, J_2$ , random vectors  $d_k = d_k(\omega)$ , and

## 2.2. Polyhedral Two-Stage Problems

43

random scalars  $r_k = r_k(\omega)$ ,  $k = 1, \dots, K_2$ , such that  $f_2(y, \omega)$  can be represented as follows:

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

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

Note that (linear) constraints of the second-stage problem which are independent of  $x$ , for example,  $y \geq 0$ , 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 reformulated as a linear two-stage model. For example, the second-stage problem (2.45) can be written as follows:

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

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 redefined in an appropriate way. In order to avoid all these manipulations and unnecessary notational complications that come 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):

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

We have

$$\begin{aligned} \inf_y L(y, \pi; x, \omega) &= \pi^\top (h(\omega) - T(\omega)x) + \inf_y [f_2(y, \omega) - \pi^\top W(\omega)y] \\ &= \pi^\top (h(\omega) - T(\omega)x) - f_2^*(W(\omega)^\top \pi, \omega), \end{aligned}$$

where  $f_2^*(\cdot, \omega)$  is the conjugate<sup>7</sup> of  $f_2(\cdot, \omega)$ . We obtain that the dual of problem (2.45) can be written as

$$\text{Max}_{\pi} [\pi^\top (h(\omega) - T(\omega)x) - f_2^*(W(\omega)^\top \pi, \omega)]. \quad (2.46)$$

By the duality theory of linear programming, if, for some  $(x, \omega)$ , 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.

<sup>7</sup>Note that since  $f_2(\cdot, \omega)$  is polyhedral, so is  $f_2^*(\cdot, \omega)$ .

**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*

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

**Proof.** Let us define the function  $\psi(\pi) := f_2^*(W^\top \pi)$ . (For simplicity we suppress the argument  $\omega$ .) 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.  $\square$

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  $\omega_k$  with respective (positive) probabilities  $p_k, k = 1, \dots, K$ . Then

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

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

$$\begin{aligned} \text{Min}_{y_1, \dots, y_K} \quad & \sum_{k=1}^K p_k f_2(y_k, \omega_k) \\ \text{s.t.} \quad & T_k x + W_k y_k = h_k, \quad k = 1, \dots, K, \end{aligned} \quad (2.48)$$

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  $k \in \{1, \dots, K\}$  the set

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

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  $x \in \mathbb{R}^n$ .*



## 2.2. Polyhedral Two-Stage Problems

45

Then the function  $\phi(\cdot)$  is polyhedral, and for any  $x_0 \in \text{dom } \phi$ ,

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

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

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

It follows that the expectation function  $\phi$  is differentiable at  $x_0$  iff for every  $\omega_k$ ,  $k = 1, \dots, K$ , 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 set<sup>8</sup>  $\mathcal{Y} := \text{dom } f_2(\cdot, \omega)$  are fixed, i.e., do not depend on  $\omega$ . In that case,

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

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

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

This means that for a.e.  $\omega$  the system

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

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  $\kappa$  such that for almost all  $\omega$  we can find a solution  $\bar{y}(\omega)$  of the system (2.52) with

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

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

$$\begin{aligned} Q(x, \omega) &\leq \max_{1 \leq j \leq J_2} \{ \gamma_j(\omega) + q_j(\omega)^\top \bar{y}(\omega) \} \\ &\leq Q(x, \omega_0) + \sum_{j=1}^{J_2} |\gamma_j(\omega) - \gamma_j(\omega_0)| \\ &\quad + \kappa \sum_{j=1}^{J_2} \|q_j(\omega)\| (\|h(\omega) - h(\omega_0)\| + \|x\| \|T(\omega) - T(\omega_0)\|). \end{aligned} \quad (2.53)$$

<sup>8</sup>Note that since it is assumed that  $f_2(\cdot, \omega)$  is polyhedral, it follows that the set  $\mathcal{Y}$  is nonempty and polyhedral.

**Proposition 2.16.** *Suppose that the recourse is fixed and*

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

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

**Proof.** The proof uses (2.53), similar to the proof of Proposition 2.6.  $\square$

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

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

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

$$\begin{aligned} & \text{Max}_{y, v} v \\ & \text{s.t. } z^\top y - \gamma_j(\omega) - q_j(\omega)^\top y \geq v, \quad j = 1, \dots, J_2, \quad y \in \mathcal{Y}. \end{aligned}$$

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  $\pi_j \geq 0, j = 1, \dots, J_2$ , such that  $\sum_{j=1}^{J_2} \pi_j = 1$  and

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

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

$$\Pi(\omega) := \{\pi : W^\top \pi \in \text{conv}\{q_j(\omega), j = 1, \dots, J_2\} + \mathcal{C}^*\}.$$

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, and (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,  $\phi$  is convex, lower semicontinuous and Lipschitz continuous on  $\text{dom } \phi$ , its domain  $\text{dom } \phi$  is a convex closed subset of  $\mathbb{R}^n$ , and*

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

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

$$\partial \phi(x_0) = -\mathbb{E}[T^\top \mathcal{D}(x_0, \omega)] + \mathcal{N}_{\text{dom } \phi}(x_0), \quad (2.57)$$

**Proof.** Note that the dual problem (2.46) is feasible iff  $W^T \pi \in \text{dom } f_2^*(\cdot, \omega)$ . 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.  $\omega$ . The remainder of the proof is similar to the linear case (Propositions 2.7 and 2.8).  $\square$

### 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  $\omega_k$  occurring with probabilities  $p_k$ ,  $k = 1, \dots, K$ .

**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  $\pi_k \in \mathfrak{D}(\bar{x}, \omega_k)$ ,  $k = 1, \dots, K$ , such that*

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

**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

$$0 \in \partial[f_1(\bar{x}) + \phi(\bar{x})]. \quad (2.59)$$

In particular, the above condition requires  $f_1(\bar{x})$  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  $\phi$ . The proof can be completed now by using the formula for  $\partial \phi(\bar{x})$  given in Proposition 2.15.  $\square$

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, and (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*

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

**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  $\phi(\cdot)$ .  $\square$

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:

$$\text{Min}_{x \in X} \{f(x) := \mathbb{E}[F(x, \omega)]\}, \quad (2.61)$$

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

$$\text{Min}_{y \in \mathcal{G}(x, \omega)} g(x, y, \omega). \quad (2.62)$$

Here  $X \subset \mathbb{R}^n$ ,  $g : \mathbb{R}^n \times \mathbb{R}^m \times \Omega \rightarrow \mathbb{R}$ , 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  $g(x, y, \omega) := c^\top x + q(\omega)^\top y$  and

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

We also use the 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) also can be written in the following equivalent form:

$$\text{Min}_{y \in \mathbb{R}^m} \bar{g}(x, y, \omega), \quad (2.63)$$

where

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

We assume that the function  $\bar{g}(x, y, \omega)$  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 lower semicontinuous. Random lower semicontinuity of  $\bar{g}(x, y, \omega)$  implies that the optimal value function  $F(x, \cdot)$  is measurable (see Theorem 7.37). 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 lower semicontinuous. The indicator function  $\mathbb{I}_{\mathcal{G}_\omega(x)}(y)$  is random lower semicontinuous 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 7.36). Of course, if  $g(x, y, \omega)$  and  $\mathbb{I}_{\mathcal{G}_\omega(x)}(y)$  are random lower semicontinuous, then their sum  $\bar{g}(x, y, \omega)$  is also random lower semicontinuous.

### 2.3. General Two-Stage Problems

49

Now let  $\mathfrak{Y}$  be a linear *decomposable* space of measurable mappings from  $\Omega$  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

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

provided that the right-hand side of (2.65) is less than  $+\infty$  (see Theorem 7.80). 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:*

$$\begin{aligned} \text{Min}_{x \in \mathbb{R}^n, y \in \mathfrak{Y}} \quad & \mathbb{E} [g(x, y(\omega), \omega)] \\ \text{s.t. } & x \in X, \quad y(\omega) \in \mathcal{G}(x, \omega) \text{ a.e. } \omega \in \Omega. \end{aligned} \quad (2.66)$$

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 that 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 (2.66).

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

$$\begin{aligned} \text{Min}_{x, y_1, \dots, y_K} \quad & \sum_{k=1}^K p_k g(x, y_k, \omega_k) \\ \text{s.t. } & x \in X, \quad y_k \in \mathcal{G}(x, \omega_k), \quad k = 1, \dots, K. \end{aligned} \quad (2.67)$$

#### 2.3.2 Convex Two-Stage Problems

We say that the two-stage problem (2.61)–(2.62) is *convex* if the set  $X$  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  $t \in [0, 1]$ :

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

Equivalently this condition can be written as

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

The multifunction  $\mathcal{G}_\omega$  satisfying the above condition (2.69) is called *convex*. By taking  $x = x'$  we obtain that if the multifunction  $\mathcal{G}_\omega$  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 form

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

where  $Y$  is a nonempty convex closed subset of  $\mathbb{R}^m$  and  $T = (t_1, \dots, t_\ell) : \mathbb{R}^n \times \Omega \rightarrow \mathbb{R}^\ell$ ,  $W = (w_1, \dots, w_\ell) : \mathbb{R}^m \times \Omega \rightarrow \mathbb{R}^\ell$ , and  $C \subset \mathbb{R}^\ell$  is a closed convex cone. Cone  $C$  defines a partial order, denoted " $\preceq_C$ ", on the space  $\mathbb{R}^\ell$ . 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  $t_i(x, \omega) + w_i(y, \omega) \leq 0, i = 1, \dots, \ell$ . We assume that  $t_i(x, \omega)$  and  $w_i(y, \omega), i = 1, \dots, \ell$ , 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  $G : \mathbb{R}^n \rightarrow \mathbb{R}^\ell$  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

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

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

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

$$\begin{aligned} & \text{Min}_{y \in Y} q_\omega(y) \\ & \text{s.t. } W_\omega(y) + \chi_\omega \preceq_C 0 \end{aligned} \quad (2.71)$$

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

$$\text{Max}_{\pi \succeq_C 0} \{ \pi^\top \chi_\omega + \inf_{y \in Y} L_\omega(y, \pi) \}, \quad (2.72)$$

where

$$L_\omega(y, \pi) := q_\omega(y) + \pi^\top W_\omega(y)$$

### 2.3. General Two-Stage Problems

51

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  $\chi_\omega$  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) iff the optimal value function  $\vartheta(\cdot, \omega)$  is lower semicontinuous at  $\chi_\omega$ .*
- (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 iff the optimal value function  $\vartheta(\cdot, \omega)$  is subdifferentiable at  $\chi_\omega$ .*
- (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 iff  $\chi_\omega \in \text{int}(\text{dom } \vartheta(\cdot, \omega))$ .*

The regularity condition  $\chi_\omega \in \text{int}(\text{dom } \vartheta(\cdot, \omega))$  means that for all small perturbations of  $\chi_\omega$  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*

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

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

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

Then

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

**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  $\chi$  iff  $\mathfrak{D}(\chi, \omega)$  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  $\chi$  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

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

holds true.  $\square$

Let us focus on the expectation function  $f(x) := \mathbb{E}[F(x, \omega)]$ . If the set  $\Omega$  is finite, say,  $\Omega = \{\omega_1, \dots, \omega_K\}$  with corresponding probabilities  $p_k, k = 1, \dots, K$ , 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  $\Omega$  is not assumed to be finite. By combining Proposition 2.22 and Theorem 7.47 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

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

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

$$\nabla f(\bar{x}) = \nabla c(\bar{x}) + \mathbb{E}[\nabla T_\omega(\bar{x})^\top \pi(\omega)]. \quad (2.78)$$

We obtain the following conditions for optimality.

**Proposition 2.24.** *Let  $\bar{x} \in X \cap \text{int}(\text{dom } f)$  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  $\pi(\omega) \in \mathfrak{D}(T_\omega(\bar{x}), \omega)$  such that*

$$-c(\bar{x}) - \mathbb{E}[\nabla T_\omega(\bar{x})^\top \pi(\omega)] \in \mathcal{N}_X(\bar{x}). \quad (2.79)$$

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

## 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.,  $\Omega = \{\omega_1, \dots, \omega_K\}$  with respective (positive) probabilities  $p_1, \dots, p_K$ . Let us relax the first-stage problem by replacing vector



## 2.4. Nonanticipativity

53

$x$  with  $K$  vectors  $x_1, x_2, \dots, x_K$ , one for each scenario. We obtain the following relaxation of problem (2.61):

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

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

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

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),  $k = 1, \dots, K$ . For example, in the case of the two-stage linear program (2.15), relaxation of the form (2.80) leads to solving  $K$  smaller problems,

$$\begin{aligned} &\text{Min}_{x_k \geq 0, y_k \geq 0} c^\top x_k + q_k^\top y_k \\ &\text{s.t. } Ax_k = b, \quad T_k x_k + W_k y_k = h_k. \end{aligned}$$

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

$$(x_1, \dots, x_K) \in \mathcal{L}, \quad (2.82)$$

where  $\mathcal{L} := \{x = (x_1, \dots, x_K) : x_1 = \dots = x_K\}$  is a linear subspace of the  $nK$ -dimensional vector space  $\mathcal{X} := \mathbb{R}^n \times \dots \times \mathbb{R}^n$ . 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

$$\begin{aligned} &\text{Min}_{x_1, \dots, x_K} \sum_{k=1}^K p_k F(x_k, \omega_k) \\ &\text{s.t. } x_1 = \dots = x_K, \quad x_k \in X, \quad k = 1, \dots, K. \end{aligned} \quad (2.83)$$

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

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

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  $\mathcal{X}$  equipped with the scalar product

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

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

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

Constraint (2.84) can be compactly written as

$$\mathbf{x} = \mathbf{P}\mathbf{x}.$$

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

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

The range space of  $\mathbf{P}$ , 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

$$\begin{aligned} x_1 &= x_2, \\ x_2 &= x_3, \\ &\vdots \\ x_{K-1} &= x_K. \end{aligned} \quad (2.87)$$

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  $\lambda_k \in \mathbb{R}^n$ ,  $k = 1, \dots, K$ , we can write the Lagrangian

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

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

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

## 2.4. Nonanticipativity

55

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

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

Let us observe that shifting the multipliers  $\lambda_k, k = 1, \dots, K$ , 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  $\mathbf{P}\boldsymbol{\lambda} = 0$ . Dualization of problem (2.80) with respect to the nonanticipativity constraints takes the form of the following problem:

$$\text{Max}_{\boldsymbol{\lambda}} \left\{ D(\boldsymbol{\lambda}) := \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) \right\} \text{ s.t. } \mathbf{P}\boldsymbol{\lambda} = 0. \quad (2.88)$$

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  $\mathbf{P}\boldsymbol{\lambda} = 0$ , the Lagrangian can be written simply as

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

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

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

where  $L_k(x_k, \lambda_k) := F(x_k, \omega_k) + \lambda_k^{\top} x_k$ . 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

$$\begin{aligned} & \text{Min}_{x_k, y_k} (c + \lambda_k)^\top x_k + q_k^\top y_k \\ & \text{s.t. } Ax_k = b, \\ & \quad T_k x_k + W_k y_k = h_k, \\ & \quad x_k \geq 0, \quad y_k \geq 0. \end{aligned}$$

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

$$\text{Min}_{x_k \in X} \sum_{k=1}^K p_k L_k(x_k, \bar{\lambda}_k) \quad (2.89)$$

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  $\bar{x} = (\bar{x}_1, \dots, \bar{x}_K)$ , then  $\bar{x} \in \mathcal{L}$ , i.e.,  $\bar{x}_1 = \dots = \bar{x}_K$ , and this is also the optimal solution of problem (2.61) with  $\bar{x}$  being equal to the common value of  $\bar{x}_1, \dots, \bar{x}_K$ . Note also that the above problem (2.89) is separable, i.e.,  $\bar{x}$  is an optimal solution of (2.89) iff for every  $k = 1, \dots, K$ ,  $\bar{x}_k$  is an optimal solution of the problem

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

### 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 the general (not necessarily finite-scenarios) case. For the sake of convenience we write problem (2.61) in the form

$$\text{Min}_{x \in \mathbb{R}^n} \{ \bar{f}(x) := \mathbb{E}[\bar{F}(x, \omega)] \}, \quad (2.90)$$

where  $\bar{F}(x, \omega) := F(x, \omega) + \mathbb{I}_X(x)$ , i.e.,  $\bar{F}(x, \omega) = F(x, \omega)$  if  $x \in X$  and  $\bar{F}(x, \omega) = +\infty$  otherwise. Let  $\mathfrak{X}$  be a linear *decomposable* space of measurable mappings from  $\Omega$  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  $x \in \mathfrak{X}$  the expectation  $\mathbb{E}[\bar{F}(x(\omega), \omega)]$  is well defined. Then we can write problem (2.90) in the equivalent form

$$\text{Min}_{x \in \mathcal{L}} \mathbb{E}[\bar{F}(x(\omega), \omega)], \quad (2.91)$$

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

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

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

## 2.4. Nonanticipativity

57

Consider the dual<sup>9</sup>  $\mathfrak{X}^* := \mathcal{L}_q(\Omega, \mathcal{F}, P; \mathbb{R}^n)$  of the space  $\mathfrak{X}$  and define the scalar product (bilinear form)

$$\langle \lambda, x \rangle := \mathbb{E}[\lambda^\top x] = \int_{\Omega} \lambda(\omega)^\top x(\omega) dP(\omega), \quad \lambda \in \mathfrak{X}^*, \quad x \in \mathfrak{X}.$$

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

$$\langle \lambda, Px \rangle = \mathbb{E}[\lambda]^\top \mathbb{E}[x] = \langle P^* \lambda, x \rangle,$$

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

With problem (2.91) is associated the following Lagrangian:

$$L(x, \lambda) := \mathbb{E}[\bar{F}(x(\omega), \omega)] + \mathbb{E}[\lambda^\top (x - \mathbb{E}[x])].$$

Note that

$$\mathbb{E}[\lambda^\top (x - \mathbb{E}[x])] = \langle \lambda, x - Px \rangle = \langle \lambda - P^* \lambda, x \rangle,$$

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

$$L(x, \lambda) = \mathbb{E}[\bar{F}(x(\omega), \omega) + \lambda(\omega)^\top x(\omega)] \quad \text{for } \mathbb{E}[\lambda] = 0. \quad (2.92)$$

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

$$\text{Max}_{\lambda \in \mathfrak{X}^*} \left\{ D(\lambda) := \inf_{x \in \mathfrak{X}} L(x, \lambda) \right\} \quad \text{s.t. } \mathbb{E}[\lambda] = 0. \quad (2.93)$$

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

By the interchangeability principle (Theorem 7.80) we have

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

Consequently,

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

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

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

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.

<sup>9</sup>Recall 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  $x \in \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ , then its expectation  $\mathbb{E}[x] = \int_{\Omega} x(\omega) dP(\omega)$  is well defined and is an element of vector space  $\mathbb{R}^n$ .

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  $(\bar{x}, \bar{\lambda}) \in \mathcal{X} \times \mathcal{X}^*$  is a saddle point of the Lagrangian iff

$$\bar{x} \in \arg \min_{x \in \mathcal{X}} L(x, \bar{\lambda}) \text{ and } \bar{\lambda} \in \arg \max_{\lambda: \mathbb{E}[\lambda] = 0} L(\bar{x}, \lambda). \quad (2.95)$$

By the interchangeability principle (see (7.247) of Theorem 7.80), we have that the first condition in (2.95) can be written in the following equivalent form:

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

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  $X$  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

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

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 of (2.93).*

Recall that the inclusion  $\mathbb{E} [\partial \bar{F}_\omega(\bar{x})] \subset \partial \bar{f}(\bar{x})$  always holds (see (7.125) in the proof of Theorem 7.47). 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  $\mathbb{E} [\partial \bar{F}_\omega(\bar{x})] = \partial \bar{f}(\bar{x})$ , then (2.97) follows. Therefore, Theorems 2.25 and 7.47 imply the following result.

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

$$\text{Min}_{x \in \mathcal{X}} \mathbb{E} [\bar{F}(x(\omega), \omega) + \bar{\lambda}(\omega)^\top x(\omega)].$$

**Proof.** Since  $\bar{x}$  is an optimal solution of problem (2.90), we have that  $\bar{x} \in X$  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 7.47 that  $\mathbb{E}[\partial F_\omega(\bar{x})] = \partial f(\bar{x})$ . Furthermore, since  $\bar{x} \in \text{int}(\text{dom} f)$ , we have that  $\partial \bar{f}(\bar{x}) = \partial f(\bar{x}) + \mathcal{N}_X(\bar{x})$ , and hence  $\mathbb{E}[\partial \bar{F}_\omega(\bar{x})] = \partial \bar{f}(\bar{x})$ . By optimality of  $\bar{x}$ , we also have that  $0 \in \partial \bar{f}(\bar{x})$ . Consequently,  $0 \in \mathbb{E}[\partial \bar{F}_\omega(\bar{x})]$ , and hence the proof can be completed by applying Theorem 2.25.  $\square$

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

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

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 an optimal solution of the scenario problem, because the first term of  $\lambda_{ij}$  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.  $\blacksquare$

#### 2.4.4 Value of Perfect Information

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

$$\text{Min}_{x \in \mathfrak{X}} \mathbb{E}[\bar{F}(x(\omega), \omega)]. \quad (2.98)$$

This relaxation is obtained by removing the nonanticipativity constraint from the formulation (2.91) of the first-stage problem. By the interchangeability principle (Theorem 7.80) we have that the optimal value of the above problem (2.98) is equal to  $\mathbb{E}[\inf_{x \in \mathbb{R}^n} \bar{F}(x, \omega)]$ . The value  $\inf_{x \in \mathbb{R}^n} \bar{F}(x, \omega)$  is equal to the optimal value of the problem

$$\text{Min}_{x \in X, y \in \mathcal{G}(x, \omega)} g(x, y, \omega). \quad (2.99)$$

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 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

$$\text{Min}_{x \geq 0, y \geq 0} c^\top x + q^\top y \quad \text{s.t.} \quad Ax = b, \quad Tx + Wy = h.$$

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

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

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  $\bar{x} \in \arg \min_{x \in X} \mathbb{E}[F(x, \omega)]$ . As  $F(\bar{x}, \omega) - \inf_{x \in X} F(x, \omega) \geq 0$  for all  $\omega \in \Omega$ , we conclude that

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

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

$$F(\bar{x}, \omega) = \inf_{x \in X} F(x, \omega) \quad \text{for a.e. } \omega \in \Omega. \quad (2.102)$$

In particular, this happens if  $\bar{F}_\omega(x)$  has a minimizer independent of  $\omega \in \Omega$ . This, of course, may happen only in rather specific situations.

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

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

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

## Exercises

2.1. Consider the assembly problem discussed in section 1.3.1 in two cases:

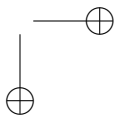
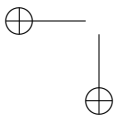
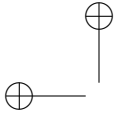
- (i) The demand which is not satisfied from the preordered quantities of parts is lost.



- (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  $s_i$ ,  $i = 1, \dots, n$ . The company considers repositioning its 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 a unique optimal solution  $\bar{\pi}$ , in which case  $\nabla_x Q(x, \xi) = -T^T \bar{\pi}$ .
- 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  $q$ , and (iii) the system  $W^T \pi \leq 0$  has only one solution  $\pi = 0$ .
- 2.6. Show that if random vector  $\xi$  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  $\bar{g}(x, y, \omega)$  is convex in  $(x, y)$ . Show that the indicator function  $\mathbb{I}_{\mathfrak{g}_\omega(x)}(y)$  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  $\langle x - Px, y \rangle = 0$  for any  $x \in \mathfrak{X}$  and  $y \in \mathfrak{L}$ , i.e., that  $P$  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 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  $\xi_1, \dots, \xi_T$  is revealed gradually over time, in  $T$  periods, and our decisions should be adapted to this process. The decision process has the form

$$\begin{aligned} &\text{decision } (x_1) \rightsquigarrow \text{observation } (\xi_2) \rightsquigarrow \text{decision } (x_2) \rightsquigarrow \\ &\dots \rightsquigarrow \text{observation } (\xi_T) \rightsquigarrow \text{decision } (x_T). \end{aligned}$$

We view the sequence  $\xi_t \in \mathbb{R}^{d_t}, t = 1, \dots, T$ , of data vectors as a *stochastic process*, i.e., as a sequence of random variables with a specified probability distribution. We use notation  $\xi_{[t]} := (\xi_1, \dots, \xi_t)$  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  $\xi_t$  is stochastically independent of  $\xi_{[t-1]}$ ,  $t = 2, \dots, T$ . It is said that the process is *Markovian* if for every  $t = 2, \dots, T$ , the conditional distribution of  $\xi_t$  given  $\xi_{[t-1]}$  is the same as the conditional distribution of  $\xi_t$  given  $\xi_{t-1}$ . Of course, if the process is stagewise independent, then it is Markovian. As before, we often use the same notation  $\xi_t$  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.

In a generic form a  $T$ -stage stochastic programming problem can be written in the nested formulation

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

driven by the random data process  $\xi_1, \xi_2, \dots, \xi_T$ . Here  $x_t \in \mathbb{R}^{n_t}$ ,  $t = 1, \dots, T$ , are decision variables,  $f_t : \mathbb{R}^{n_t} \times \mathbb{R}^{d_t} \rightarrow \mathbb{R}$  are continuous functions and  $\mathcal{X}_t : \mathbb{R}^{n_{t-1}} \times \mathbb{R}^{d_t} \rightrightarrows \mathbb{R}^{n_t}$ ,  $t = 2, \dots, T$ , are measurable closed valued multifunctions. The first-stage data, i.e., the vector  $\xi_1$ , the function  $f_1 : \mathbb{R}^{n_1} \rightarrow \mathbb{R}$ , 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. In a typical formulation,

$$\begin{aligned} f_t(x_t, \xi_t) &:= c_t^\top x_t, \quad \mathcal{X}_1 := \{x_1 : A_1 x_1 = b_1, x_1 \geq 0\}, \\ \mathcal{X}_t(x_{t-1}, \xi_t) &:= \{x_t : B_t x_{t-1} + A_t x_t = b_t, x_t \geq 0\}, \quad t = 2, \dots, T. \end{aligned}$$

Here,  $\xi_1 := (c_1, A_1, b_1)$  is known at the first-stage (and hence is nonrandom), and  $\xi_t := (c_t, B_t, A_t, b_t) \in \mathbb{R}^{d_t}$ ,  $t = 2, \dots, T$ , are data vectors,<sup>10</sup> some (or all) elements of which can be random.

There are several equivalent ways to make this formulation precise. One approach is to consider decision variables  $x_t = \mathbf{x}_t(\xi_{[t]})$ ,  $t = 1, \dots, T$ , as functions of the data process  $\xi_{[t]}$  up to time  $t$ . Such a sequence of (measurable) mappings  $\mathbf{x}_t : \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_t} \rightarrow \mathbb{R}^{n_t}$ ,  $t = 1, \dots, T$ , is called an *implementable policy* (or simply a *policy*) (recall that  $\xi_1$  is deterministic). An implementable policy is said to be *feasible* if it satisfies the feasibility constraints, i.e.,

$$\mathbf{x}_t(\xi_{[t]}) \in \mathcal{X}_t(\mathbf{x}_{t-1}(\xi_{[t-1]}), \xi_t), \quad t = 2, \dots, T, \quad \text{w.p. } 1. \quad (3.2)$$

We can formulate the multistage problem (3.1) in the form

$$\begin{aligned} \text{Min}_{x_1, \mathbf{x}_2, \dots, \mathbf{x}_T} \quad & \mathbb{E}[f_1(x_1) + f_2(\mathbf{x}_2(\xi_{[2]}), \xi_2) + \cdots + f_T(\mathbf{x}_T(\xi_{[T]}), \xi_T)] \\ \text{s.t.} \quad & x_1 \in \mathcal{X}_1, \quad \mathbf{x}_t(\xi_{[t]}) \in \mathcal{X}_t(\mathbf{x}_{t-1}(\xi_{[t-1]}), \xi_t), \quad t = 2, \dots, T. \end{aligned} \quad (3.3)$$

Note that optimization in (3.3) is performed over implementable and feasible policies and that policies  $\mathbf{x}_2, \dots, \mathbf{x}_T$  are *functions* of the data process, and hence are elements of appropriate functional spaces, while  $x_1 \in \mathbb{R}^{n_1}$  is a deterministic vector. Therefore, unless the data process  $\xi_1, \dots, \xi_T$  has a finite number of realizations, formulation (3.3) leads to an infinite dimensional optimization problem. This is a natural extension of the formulation (2.66) of the two-stage problem.

Another possible way is to write the corresponding *dynamic programming* equations. That is, consider the last-stage problem

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

<sup>10</sup>If data involves matrices, then their elements can be stacked columnwise to make it a vector.

### 3.1. Problem Formulation

65

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

$$\begin{aligned} \text{Min}_{x_t} \quad & f_t(x_t, \xi_t) + \mathbb{E} \{ Q_{t+1}(x_t, \xi_{[t+1]}) \mid \xi_{[t]} \} \\ \text{s.t.} \quad & x_t \in \mathcal{X}_t(x_{t-1}, \xi_t), \end{aligned}$$

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

$$\text{Min}_{x_1 \in \mathcal{X}_1} f_1(x_1) + \mathbb{E}[Q_2(x_1, \xi_2)].$$

The corresponding dynamic programming equations are

$$Q_t(x_{t-1}, \xi_{[t]}) = \inf_{x_t \in \mathcal{X}_t(x_{t-1}, \xi_t)} \{ f_t(x_t, \xi_t) + Q_{t+1}(x_t, \xi_{[t+1]}) \}, \quad (3.4)$$

where

$$Q_{t+1}(x_t, \xi_{[t]}) := \mathbb{E} \{ Q_{t+1}(x_t, \xi_{[t+1]}) \mid \xi_{[t]} \}.$$

An implementable policy  $\bar{x}_t(\xi_{[t]})$  is *optimal* iff for  $t = 1, \dots, T$ ,

$$\bar{x}_t(\xi_{[t]}) \in \arg \min_{x_t \in \mathcal{X}_t(\bar{x}_{t-1}(\xi_{[t-1]}), \xi_t)} \{ f_t(x_t, \xi_t) + Q_{t+1}(x_t, \xi_{[t+1]}) \}, \quad \text{w.p. 1}, \quad (3.5)$$

where for  $t = T$  the term  $Q_{T+1}$  is omitted and for  $t = 1$  the set  $X_1$  depends only on  $\xi_1$ . 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.

If the process  $\xi_1, \dots, \xi_T$  is Markovian, then conditional distributions in the above equations, given  $\xi_{[t]}$ , are the same as the respective conditional distributions given  $\xi_t$ . In that case each cost-to-go function  $Q_t$  depends on  $\xi_t$  rather than the whole  $\xi_{[t]}$  and we can write it as  $Q_t(x_{t-1}, \xi_t)$ . 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 simply as  $Q_t(x_{t-1})$ .

#### 3.1.2 The Linear Case

We discuss linear multistage problems in more detail. Let  $x_1, \dots, x_T$  be decision vectors corresponding to time periods (stages)  $1, \dots, T$ . Consider the following linear programming problem:

$$\begin{aligned} \text{Min} \quad & c_1^\top x_1 + c_2^\top x_2 + c_3^\top x_3 + \dots + c_T^\top x_T \\ \text{s.t.} \quad & A_1 x_1 = b_1, \\ & B_2 x_1 + A_2 x_2 = b_2, \\ & B_3 x_2 + A_3 x_3 = b_3, \\ & \dots \\ & B_T x_{T-1} + A_T x_T = b_T, \\ & x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad \dots \quad x_T \geq 0. \end{aligned} \quad (3.6)$$

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  $b_t$ ,  $t = 2, \dots, T$ , 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:

$$\begin{aligned} & \text{decision } (x_1) \\ & \text{observation } \xi_2 := (c_2, B_2, A_2, b_2) \\ & \text{decision } (x_2) \\ & \vdots \\ & \text{observation } \xi_T := (c_T, B_T, A_T, b_T) \\ & \text{decision } (x_T). \end{aligned}$$

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  $t = 1, \dots, T$ .

Let us denote by  $\xi_t$  the data vector, realization of which becomes known at time period  $t$ . In the setting of the multiperiod problem (3.6),  $\xi_t$  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 (3.6) 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 several possible ways to formulate multistage stochastic programs in a precise mathematical form. In one such formulation  $x_t = x_t(\xi_{[t]})$ ,  $t = 2, \dots, T$ , is viewed as a function of  $\xi_{[t]}$ , and the minimization in (3.6) 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 3.1.4. Another possible approach is to write dynamic programming equations, which we discuss 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,  $x_1, \dots, x_{T-1}$ , have been chosen. Our problem is, therefore, a simple linear programming problem

$$\begin{aligned} & \text{Min}_{x_T} c_T^\top x_T \\ & \text{s.t. } B_T x_{T-1} + A_T x_T = b_T, \\ & \quad x_T \geq 0. \end{aligned}$$

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  $Q_T(x_{T-1}, \xi_T)$ .

### 3.1. Problem Formulation

67

At stage  $T - 1$  we know  $x_{T-2}$  and  $\xi_{[T-1]}$ . We face, therefore, the following stochastic programming problem:

$$\begin{aligned} \text{Min}_{x_{T-1}} \quad & c_{T-1}^\top x_{T-1} + \mathbb{E}[Q_T(x_{T-1}, \xi_T) \mid \xi_{[T-1]}] \\ \text{s.t.} \quad & B_{T-1}x_{T-2} + A_{T-1}x_{T-1} = b_{T-1}, \\ & x_{T-1} \geq 0. \end{aligned}$$

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  $t = 2, \dots, T - 1$ , we have the problem

$$\begin{aligned} \text{Min}_{x_t} \quad & c_t^\top x_t + \mathbb{E}[Q_{t+1}(x_t, \xi_{[t+1]}) \mid \xi_{[t]}] \\ \text{s.t.} \quad & B_t x_{t-1} + A_t x_t = b_t, \\ & x_t \geq 0. \end{aligned} \tag{3.7}$$

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,  $x_1 \in \mathbb{R}^{n_1}$ ,

$$\begin{aligned} \text{Min}_{x_1} \quad & c_1^\top x_1 + \mathbb{E}[Q_2(x_1, \xi_2)] \\ \text{s.t.} \quad & A_1 x_1 = b_1, \\ & x_1 \geq 0. \end{aligned} \tag{3.8}$$

Note that all subsequent stages  $t = 2, \dots, T$  are absorbed in the above problem into the function  $Q_2(x_1, \xi_2)$  through the corresponding expected values. Note also that since  $\xi_1$  is not random, the optimal value  $Q_2(x_1, \xi_2)$  does not depend on  $\xi_1$ . In particular, if  $T = 2$ , then (3.8) coincides with the formulation (2.1) of a two-stage linear problem.

The dynamic programming equations here take the form (compare with (3.4))

$$Q_t(x_{t-1}, \xi_{[t]}) = \inf_{x_t} \{c_t^\top x_t + \mathcal{Q}_{t+1}(x_t, \xi_{[t+1]}) : B_t x_{t-1} + A_t x_t = b_t, x_t \geq 0\},$$

where

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

Also an implementable policy  $\bar{x}_t(\xi_{[t]})$  is *optimal* if for  $t = 1, \dots, T$  the condition

$$\bar{x}_t(\xi_{[t]}) \in \arg \min_{x_t} \{c_t^\top x_t + \mathcal{Q}_{t+1}(x_t, \xi_{[t+1]}) : A_t x_t = b_t - B_t \bar{x}_{t-1}(\xi_{[t-1]}), x_t \geq 0\}$$

holds for almost every realization of the random process. (For  $t = T$  the term  $\mathcal{Q}_{T+1}$  is omitted and for  $t = 1$  the term  $B_t \bar{x}_{t-1}$  is omitted.) If the process  $\xi_t$  is Markovian, then each cost-to-go function depends on  $\xi_t$  rather than  $\xi_{[t]}$ , and we can simply write  $Q_t(x_{t-1}, \xi_t)$ ,  $t = 2, \dots, T$ . If, moreover, the stagewise independence condition holds, then each expectation function  $\mathcal{Q}_t$  does not depend on realizations of the random process, and we can write it as  $\mathcal{Q}_t(x_{t-1})$ ,  $t = 2, \dots, T$ .

The *nested formulation* of the linear multistage problem can be written as follows (compare with (3.1)):

$$\text{Min}_{\substack{A_1 x_1 = b_1 \\ x_1 \geq 0}} c_1^\top x_1 + \mathbb{E} \left[ \min_{\substack{B_2 x_1 + A_2 x_2 = b_2 \\ x_2 \geq 0}} c_2^\top x_2 + \mathbb{E} \left[ \dots + \mathbb{E} \left[ \min_{\substack{B_T x_{T-1} + A_T x_T = b_T \\ x_T \geq 0}} c_T^\top x_T \right] \right] \right]. \tag{3.9}$$

Suppose now that we deal with an underlying model with a full lower block triangular constraint matrix:

$$\begin{array}{ll}
 \text{Min} & c_1^\top x_1 + c_2^\top x_2 + c_3^\top x_3 + \cdots + c_T^\top x_T \\
 \text{s.t.} & A_{11}x_1 = b_1, \\
 & A_{21}x_1 + A_{22}x_2 = b_2, \\
 & A_{31}x_1 + A_{32}x_2 + A_{33}x_3 = b_3, \\
 & \dots \\
 & A_{T1}x_1 + A_{T2}x_2 + \cdots + A_{T,T-1}x_{T-1} + A_{TT}x_T = b_T, \\
 & x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad \dots \quad x_T \geq 0.
 \end{array} \tag{3.10}$$

In the constraint matrix of (3.6), the respective blocks  $A_{t1}, \dots, A_{t,t-2}$  were assumed to be zeros. This allowed us to express there the optimal value  $Q_t$  of (3.7) as a function of the immediately preceding decision,  $x_{t-1}$ , rather than all earlier decisions  $x_1, \dots, x_{t-1}$ . In the case of problem (3.10), each respective subproblem of the form (3.7) depends on the entire history of our decisions,  $x_{[t-1]} := (x_1, \dots, x_{t-1})$ . It takes on the form

$$\begin{array}{ll}
 \text{Min}_{x_t} & c_t^\top x_t + \mathbb{E}[Q_{t+1}(x_{[t]}, \xi_{[t+1]}) \mid \xi_{[t]}] \\
 \text{s.t.} & A_{t1}x_1 + \cdots + A_{t,t-1}x_{t-1} + A_{t,t}x_t = b_t, \\
 & x_t \geq 0.
 \end{array} \tag{3.11}$$

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 (3.10), the vectors  $x_{[t]} = (x_1, \dots, x_t)$  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 (3.11) 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

$$\begin{array}{ll}
 \text{Min} & c_1^\top x_1 + c_2^\top x_2 + c_3^\top x_3 + \cdots + c_T^\top x_T \\
 \text{s.t.} & A_{11}x_1 = b_1, \\
 & B_1x_1 + A_{22}x_2 = b_2, \\
 & B_1x_1 + B_2x_2 + A_{33}x_3 = b_3, \\
 & \dots \\
 & B_1x_1 + B_2x_2 + \cdots + B_{T-1}x_{T-1} + A_{TT}x_T = b_T, \\
 & x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad \dots \quad x_T \geq 0,
 \end{array}$$



### 3.1. Problem Formulation

69

in which all blocks  $A_{it}$ ,  $i = t + 1, \dots, T$ , are identical and observed at time  $t$ . Then we can define the state variables  $r_t$ ,  $t = 1, \dots, T$ , recursively by the state equation  $r_t = r_{t-1} + B_t x_t$ ,  $t = 1, \dots, T - 1$ , where  $r_0 = 0$ . Subproblem (3.11) simplifies substantially:

$$\begin{aligned} \text{Min}_{x_t, r_t} \quad & c_t^T x_t + \mathbb{E} [Q_{t+1}(r_t, \xi_{[t+1]}) \mid \xi_{[t]}] \\ \text{s.t.} \quad & r_{t-1} + A_{t,t} x_t = b_t, \\ & r_t = r_{t-1} + B_t x_t, \\ & x_t \geq 0. \end{aligned}$$

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 sign constraints  $x_t \geq 0$  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  $\xi_1, \dots, \xi_T$  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  $1, \dots, T$ . At level  $t = 1$  we have only one *root node*, and we associate with it the value of  $\xi_1$  (which is known at stage  $t = 1$ ). At level  $t = 2$  we have as many nodes as many different realizations of  $\xi_2$  may occur. Each of them is connected with the root node by an arc. For each node  $\iota$  of level  $t = 2$  (which corresponds to a particular realization  $\xi_2^\iota$  of  $\xi_2$ ) we create at least as many nodes at level 3 as different values of  $\xi_3$  may follow  $\xi_2^\iota$ , and we connect them with the node  $\iota$ , 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  $\xi_t^\iota$  are vectors, and it may happen that some of the values  $\xi_t^\iota$ , 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  $\Omega_t$  the set of all nodes at stage  $t = 1, \dots, T$ . In particular,  $\Omega_1$  consists of a unique root node,  $\Omega_2$  has as many elements as many different realizations of  $\xi_2$  may occur, etc. For a node  $\iota \in \Omega_t$  we denote by  $C_\iota \subset \Omega_{t+1}$ ,  $t = 1, \dots, T - 1$ , the set of all children nodes of  $\iota$ , and by  $a(\iota) \in \Omega_{t-1}$ ,  $t = 2, \dots, T$ , the ancestor node of  $\iota$ . We have that  $\Omega_{t+1} = \cup_{\iota \in \Omega_t} C_\iota$  and the sets  $C_\iota$  are disjoint, i.e.,  $C_\iota \cap C_{\iota'} = \emptyset$  if  $\iota \neq \iota'$ . Note again that with different nodes at stage  $t \geq 3$  may be associated the same numerical values (realizations) of the corresponding data process  $\xi_t$ . *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  $\xi_1, \dots, \xi_T$ . By construction, there is one-to-one correspondence between scenarios and the set  $\Omega_T$ , and hence the total number  $K$  of scenarios is equal to the cardinality<sup>11</sup> of the set  $\Omega_T$ , i.e.,  $K = |\Omega_T|$ .

<sup>11</sup>We denote by  $|\Omega|$  the number of elements in a (finite) set  $\Omega$ .

Next we should 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  $\xi_{t+1}$  given  $\xi_{[t]}$ ,  $t = 1, \dots, T-1$ . That is, if we are currently at a node  $\iota \in \Omega_t$ , we need to specify probability of moving from  $\iota$  to a node  $\eta \in C_\iota$ . Let us denote this probability by  $\rho_{\iota\eta}$ . Note that  $\rho_{\iota\eta} \geq 0$  and  $\sum_{\eta \in C_\iota} \rho_{\iota\eta} = 1$ , and that probabilities  $\rho_{\iota\eta}$  are in one-to-one correspondence with arcs of the scenario tree. Probabilities  $\rho_{\iota\eta}$ ,  $\eta \in C_\iota$ , represent conditional distribution of  $\xi_{t+1}$  given that the path of the process  $\xi_1, \dots, \xi_t$  ended at the node  $\iota$ .

Every scenario can be defined by its nodes  $\iota_1, \dots, \iota_T$ , arranged in the chronological order, i.e., node  $\iota_2$  (at level  $t = 2$ ) is connected to the root node,  $\iota_3$  is connected to the node  $\iota_2$ , 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, a set of conditional probabilities defines a probability distribution on the set of scenarios. Conversely, it is possible to derive these conditional probabilities from scenario probabilities  $p_k$ ,  $k = 1, \dots, K$ , as follows. Let us denote by  $\mathcal{S}^{(\iota)}$  the set of scenarios passing through node  $\iota$  (at level  $t$ ) of the scenario tree, and let  $p^{(\iota)} := \Pr[\mathcal{S}^{(\iota)}]$ , i.e.,  $p^{(\iota)}$  is the sum of probabilities of all scenarios passing through node  $\iota$ . If  $\iota_1, \iota_2, \dots, \iota_t$ , with  $\iota_1$  being the root node and  $\iota_t = \iota$ , is the history of the process up to node  $\iota$ , then the probability  $p^{(\iota)}$  is given by 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  $\iota$ . 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  $\iota$ , then  $\mathcal{S}^{(\iota)} \subset \mathcal{S}^{(a)}$  and hence  $p^{(\iota)} \leq p^{(a)}$ . Consequently, if  $p^{(a)} > 0$ , then  $\rho_{a\iota} = p^{(\iota)}/p^{(a)}$ . Otherwise  $\mathcal{S}^{(a)}$  is empty, i.e., no scenario is passing through the node  $a$ , and hence no scenario is passing through the node  $\iota$ .

If the process  $\xi_1, \dots, \xi_T$  is stagewise independent, then the conditional distribution of  $\xi_{t+1}$  given  $\xi_{[t]}$  is the same as the unconditional distribution of  $\xi_{t+1}$ ,  $t = 1, \dots, T-1$ . In that case at every stage  $t = 1, \dots, T-1$ , 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  $Z_t$ ,  $t = 1, 2, \dots$ , that can take a finite number  $\{z_1, \dots, z_m\}$  of different values is a *Markov chain* if

$$\Pr\{Z_{t+1} = z_j \mid Z_t = z_i, Z_{t-1} = z_{i-1}, \dots, Z_1 = z_{i_1}\} = \Pr\{Z_{t+1} = z_j \mid Z_t = z_i\}$$

for all states  $z_{i-1}, \dots, z_{i_1}, z_i, z_j$  and all  $t = 1, 2, \dots$ . Denote

$$p_{ij} := \Pr\{Z_{t+1} = z_j \mid Z_t = z_i\}, \quad i, j = 1, \dots, m.$$

In some situations, it is natural to model the data process as a Markov chain with the corresponding state space<sup>12</sup>  $\{\zeta^1, \dots, \zeta^m\}$  and probabilities  $p_{ij}$  of moving from state  $\zeta^i$  to state  $\zeta^j$ ,  $i, j = 1, \dots, m$ . 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,  $\zeta^i$ . At stage  $t = 2$  there are  $m$  nodes to which are assigned values  $\zeta^1, \dots, \zeta^m$  with the

<sup>12</sup>In our model, values  $\zeta^1, \dots, \zeta^m$  can be numbers or vectors.

corresponding probabilities  $p_{i1}, \dots, p_{im}$ . At stage  $t = 3$  there are  $m^2$  nodes, such that each node at stage  $t = 2$ , associated with a state  $\zeta^a$ ,  $a = 1, \dots, m$ , is the ancestor of  $m$  nodes at stage  $t = 3$  to which are assigned values  $\zeta^1, \dots, \zeta^m$  with the corresponding conditional probabilities  $p_{a1}, \dots, p_{am}$ . 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)  $\xi_t$  can take values  $\zeta^1, \dots, \zeta^m$  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)  $\xi_1, \dots, \xi_T$  are independently distributed iff  $p_{ij} = p_{i'j}$  for any  $i, i', j = 1, \dots, m$ , i.e., the conditional probability  $p_{ij}$  of moving from state  $\zeta^i$  to state  $\zeta^j$  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  $Q_t(x_{t-1}, \xi_t)$ , going backward, 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)  $\xi_t = \zeta^i$ ,  $i = 1, \dots, m$ , of the process. On the other hand, if we want to write the corresponding optimization problem (in the case of a finite number of scenarios) 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.

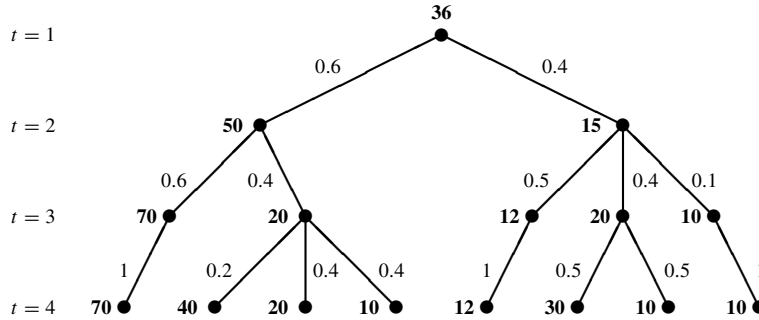
An important concept associated with the data process is the corresponding filtration. We associate with the set  $\Omega_T$  the sigma algebra  $\mathcal{F}_T$  of all its subsets. The set  $\Omega_T$  can be represented as the union of disjoint sets  $C_\iota$ ,  $\iota \in \Omega_{T-1}$ . Let  $\mathcal{F}_{T-1}$  be the subalgebra of  $\mathcal{F}_T$  generated by the sets  $C_\iota$ ,  $\iota \in \Omega_{T-1}$ . As they are disjoint, they are the elementary events of  $\mathcal{F}_{T-1}$ . By this construction, there is one-to-one correspondence between elementary events of  $\mathcal{F}_{T-1}$  and the set  $\Omega_{T-1}$  of nodes at stage  $T-1$ . By continuing in this way we construct a sequence of sigma algebras  $\mathcal{F}_1 \subset \dots \subset \mathcal{F}_T$ , called *filtration*. In this construction, elementary events of sigma algebra  $\mathcal{F}_t$  are subsets of  $\Omega_T$  which are in one-to-one correspondence with the nodes  $\iota \in \Omega_t$ . Of course, the cardinality  $|\mathcal{F}_t| = 2^{|\Omega_t|}$ . In particular,  $\mathcal{F}_1$  corresponds to the unique root at stage  $t = 1$  and hence  $\mathcal{F}_1 = \{\emptyset, \Omega_T\}$ .

### 3.1.4 Algebraic Formulation of Nonanticipativity Constraints

Suppose that in our basic problem (3.6) 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 decisions<sup>13</sup>  $\mathbf{x}^k = (x_1^k, x_2^k, \dots, x_T^k)$ . That is, with each possible scenario  $k = 1, \dots, K$  (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

<sup>13</sup>To avoid ugly collisions of subscripts, we change our notation a little and put the index of the scenario,  $k$ , as a superscript.





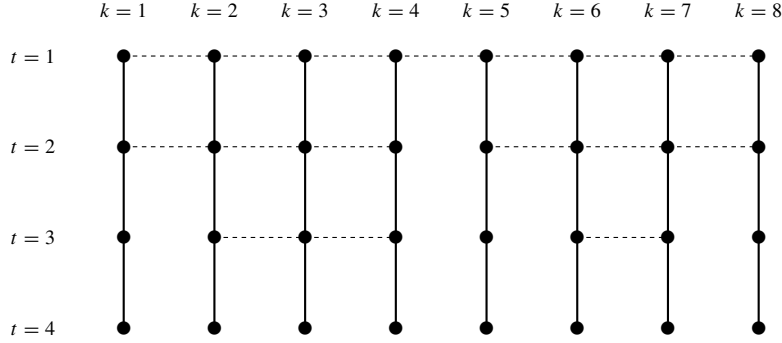
**Figure 3.1.** *Scenario tree. Nodes represent information states. Paths from the root to leaves represent scenarios. Numbers along the arcs represent conditional probabilities of moving to the next node. Bold numbers represent numerical values of the process.*

Since the two-stage problem obtained is a relaxation of the multistage problem (3.6), its optimal value gives a lower bound for the optimal value of problem (3.6) 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  $x_2, \dots, x_{K-1}$ .

**Example 3.1 (Scenario Tree).** As 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 a 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  $\xi_t = (c_t, B_t, A_t, b_t)$ ,  $t = 1, \dots, T$ , 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  $b_1, \dots, b_T$  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 right to left): 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 right to left, 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  $\xi_3$ , 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  $\xi_4$ . 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.



**Figure 3.2.** Sequences of decisions for scenarios from Figure 3.1. Horizontal dotted lines represent the equations of nonanticipativity.

The process  $b_t$  (and hence the process  $\xi_t$ ) 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

$$\begin{aligned} \Pr \{b_4 = 10 \mid b_3 = 20\} &= \frac{\Pr \{b_4 = 10, b_3 = 20\}}{\Pr \{b_3 = 20\}} \\ &= \frac{0.5 \cdot 0.4 \cdot 0.4 + 0.4 \cdot 0.4 \cdot 0.6}{0.4 \cdot 0.4 + 0.4 \cdot 0.6} = 0.44. \end{aligned}$$

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 (3.12), where the problem was separated into  $K$  different problems, one for each scenario  $\omega_k = (\xi_1^k, \dots, \xi_T^k)$ ,  $k = 1, \dots, K$ . The corresponding nonanticipativity constraints can be written in several way. One possible way is to write them, similarly to (2.84) for two-stage models, as

$$x_t = \mathbb{E} [x_t \mid \xi_{[t]}], \quad t = 1, \dots, T. \quad (3.15)$$

Another way is to use *filtration* associated with the data process. Let  $\mathcal{F}_t$  be the sigma algebra generated by  $\xi_{[t]}$ ,  $t = 1, \dots, T$ . That is,  $\mathcal{F}_t$  is the minimal subalgebra of the sigma algebra  $\mathcal{F}$  such that  $\xi_1(\omega), \dots, \xi_t(\omega)$  are  $\mathcal{F}_t$ -measurable. Since  $\xi_1$  is not random,  $\mathcal{F}_1$  contains only two sets:  $\emptyset$  and  $\Omega$ . We have that  $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}_T \subset \mathcal{F}$ . In the case of finitely many scenarios, we discussed construction of such a filtration at the end of section 3.1.3. We can write (3.15) in the following equivalent form

$$x_t = \mathbb{E} [x_t \mid \mathcal{F}_t], \quad t = 1, \dots, T. \quad (3.16)$$

(See section 7.2.2 for a definition of conditional expectation with respect to a sigma subalgebra.) Condition (3.16) holds iff  $x_t(\omega)$  is measurable with respect to  $\mathcal{F}_t$ ,  $t = 1, \dots, T$ . 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  $(x_1^k, \dots, x_T^k)$  of decision vectors which can be considered as an element of a vector space of dimension  $n_1 + \dots + n_T$ . The space of all such sequences  $(x_1^k, \dots, x_T^k)$ ,  $k = 1, \dots, K$ , is a vector space, denoted  $\mathfrak{X}$ , of dimension  $(n_1 + \dots + n_T)K$ . The nonanticipativity constraints (3.14) define a linear subspace of  $\mathfrak{X}$ , denoted  $\mathfrak{L}$ . Define the scalar product on the space  $\mathfrak{X}$ ,

$$\langle x, y \rangle := \sum_{k=1}^K \sum_{t=1}^T p_k(x_t^k)^\top y_t^k, \quad (3.17)$$

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 (3.14) can be derived by using the following construction, which extends to the multistage case the system (2.87).

Let  $\Omega_t$  be the set of nodes at level  $t$ . For a node  $\iota \in \Omega_t$  we denote by  $\mathcal{J}^{(\iota)}$  the set of scenarios that pass through node  $\iota$  and are, therefore, indistinguishable on the basis of the information available up to time  $t$ . As explained before, the sets  $\mathcal{J}^{(\iota)}$  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  $\mathcal{J}_t^1, \dots, \mathcal{J}_t^{\gamma_t}$ .

Let us assume that all scenarios  $1, \dots, K$  are ordered in such a way that each set  $\mathcal{J}_t^v$  is a set of consecutive numbers  $l_t^v, l_t^v + 1, \dots, r_t^v$ . Then nonanticipativity can be expressed by the system of equations

$$x_t^s - x_t^{s+1} = 0, \quad s = l_t^v, \dots, r_t^v - 1, \quad t = 1, \dots, T-1, \quad v = 1, \dots, \gamma_t. \quad (3.18)$$

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 (3.18) 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:

$$M^1 x^1 + \dots + M^K x^K = 0. \quad (3.19)$$

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.

**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

$$\begin{bmatrix} I & -I & -I & -I & -I & -I & -I & -I \\ & I & -I & -I & I & -I & -I & -I \\ & & I & -I & -I & I & -I & -I \\ & & & I & -I & -I & I & -I \\ & & & & I & -I & -I & I \\ & & & & & I & -I & -I \\ & & & & & & I & -I \\ & & & & & & & I \end{bmatrix}.$$

**Figure 3.3.** The nonanticipativity constraint matrix  $M$  corresponding to the scenario tree from Figure 3.1. The subdivision corresponds to the scenario submatrices  $M^1, \dots, M^8$ .

the form

$$\begin{aligned} x_1^1 - x_1^2 &= 0, & x_1^2 - x_1^3 &= 0, & \dots, & x_1^7 - x_1^8 &= 0, \\ x_2^1 - x_2^2 &= 0, & x_2^2 - x_2^3 &= 0, & x_2^3 - x_2^4 &= 0, \\ x_2^5 - x_2^6 &= 0, & x_2^6 - x_2^7 &= 0, & x_2^7 - x_2^8 &= 0, \\ x_3^2 - x_3^3 &= 0, & x_3^3 - x_3^4 &= 0, & x_3^6 - x_3^7 &= 0. \end{aligned}$$

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. ■

## 3.2 Duality

### 3.2.1 Convex Multistage Problems

In this section we consider multistage problems of the form (3.1) with

$$\mathcal{X}_t(x_{t-1}, \xi_t) := \{x_t : B_t x_{t-1} + A_t x_t = b_t\}, \quad t = 2, \dots, T, \quad (3.20)$$

$\mathcal{X}_1 := \{x_1 : A_1 x_1 = b_1\}$  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.  $\xi_t$ . In particular, if

$$f_t(x_t, \xi_t) := \begin{cases} c_t^\top x_t & \text{if } x_t \geq 0, \\ +\infty & \text{otherwise,} \end{cases} \quad (3.21)$$

then the problem becomes the linear multistage problem given in the nested formulation (3.9). 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  $(A_t, B_t, b_t) = (A_t(\xi_t), B_t(\xi_t), b_t(\xi_t))$ ,  $t = 1, \dots, T$ , form a random process.

Dynamic programming equations take here the form

$$Q_t(x_{t-1}, \xi_{[t]}) = \inf_{x_t} \{f_t(x_t, \xi_t) + Q_{t+1}(x_t, \xi_{[t+1]}) : B_t x_{t-1} + A_t x_t = b_t\}, \quad (3.22)$$

where

$$Q_{t+1}(x_t, \xi_{[t]}) := \mathbb{E} \{Q_{t+1}(x_t, \xi_{[t+1]}) \mid \xi_{[t]}\}.$$

For every  $t = 1, \dots, T$  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  $t = T, \dots, 1$ . 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 (3.22). With the optimization problem on the right-hand side of (3.22) is associated the following Lagrangian:

$$L_t(x_t, \pi_t) := f_t(x_t, \xi_t) + Q_{t+1}(x_t, \xi_{[t+1]}) + \pi_t^\top (b_t - B_t x_{t-1} - A_t x_t).$$

This Lagrangian also depends on  $\xi_{[t]}$  and  $x_{t-1}$ , which we omit for brevity of the notation. Denote

$$\psi_t(x_t, \xi_{[t]}) := f_t(x_t, \xi_t) + Q_{t+1}(x_t, \xi_{[t+1]}).$$

The dual functional is

$$\begin{aligned} D_t(\pi_t) &:= \inf_{x_t} L_t(x_t, \pi_t) \\ &= -\sup_{x_t} \{\pi_t^\top A_t x_t - \psi_t(x_t, \xi_{[t]})\} + \pi_t^\top (b_t - B_t x_{t-1}) \\ &= -\psi_t^*(A_t^\top \pi_t, \xi_{[t]}) + \pi_t^\top (b_t - B_t x_{t-1}), \end{aligned}$$

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 (3.22) as follows:

$$\text{Max}_{\pi_t} \{-\psi_t^*(A_t^\top \pi_t, \xi_{[t]}) + \pi_t^\top (b_t - B_t x_{t-1})\}. \quad (3.23)$$

Both optimization problems, (3.22) and its dual (3.23), are convex. Under various regularity conditions there is no duality gap between problems (3.22) and (3.23). In particular, we can formulate the following two conditions.

- (D1) The functions  $f_t(x_t, \xi_t)$ ,  $t = 1, \dots, T$ , 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 (3.22) 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 (3.23). All subdifferentials in the subsequent formulas are taken with respect to  $x_t$  for an appropriate  $t = 1, \dots, T$ .

**Proposition 3.3.** *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 (3.22) and (3.23), i.e.,*

$$Q_t(x_{t-1}, \xi_{[t]}) = \sup_{\pi_t} \{ -\psi_t^*(A_t^\top \pi_t, \xi_{[t]}) + \pi_t^\top (b_t - B_t x_{t-1}) \}, \quad (3.24)$$

- (ii)  *$\bar{x}_t$  is an optimal solution of (3.22) iff there exists  $\bar{\pi}_t = \bar{\pi}_t(\xi_{[t]})$  such that  $\bar{\pi}_t \in \mathfrak{D}_t(x_{t-1}, \xi_{[t]})$  and*

$$0 \in \partial L_t(\bar{x}_t, \bar{\pi}_t), \quad (3.25)$$

- (iii) *the function  $Q_t(\cdot, \xi_{[t]})$  is subdifferentiable at  $x_{t-1}$  and*

$$\partial Q_t(x_{t-1}, \xi_{[t]}) = -B_t^\top \mathfrak{D}_t(x_{t-1}, \xi_{[t]}). \quad (3.26)$$

**Proof.** Consider the optimal value function

$$\vartheta(y) := \inf_{x_t} \{ \psi_t(x_t, \xi_{[t]}) : A_t x_t = y \}.$$

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 (3.26) then follows by the chain rule. Condition (3.25) 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.  $\square$

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 (3.5), which in the present case can be written as follows:

$$\bar{x}_t(\xi_{[t]}) \in \arg \min_{x_t} \{ f_t(x_t, \xi_t) + Q_{t+1}(x_t, \xi_{[t]}) : A_t x_t = b_t - B_t \bar{x}_{t-1}(\xi_{[t-1]}) \}. \quad (3.27)$$

### 3.2. Duality

79

Since the optimization problem on the right-hand side of (3.27) is convex, subject to linear constraints, we have that a feasible policy is optimal iff it satisfies the following optimality conditions: for  $t = 1, \dots, T$  and a.e.  $\xi_{[t]}$  there exists  $\bar{\pi}_t(\xi_{[t]})$  such that the following condition holds:

$$0 \in \partial [f_t(\bar{x}_t(\xi_{[t]}), \xi_t) + \mathcal{Q}_{t+1}(\bar{x}_t(\xi_{[t]}), \xi_{[t]})] - A_t^\top \bar{\pi}_t(\xi_{[t]}). \quad (3.28)$$

Recall that all subdifferentials are taken with respect to  $x_t$ , and for  $t = T$  the term  $\mathcal{Q}_{T+1}$  is omitted.

We shall use the following regularity condition:

(D3) For  $t = 2, \dots, T$  and a.e.  $\xi_{[t]}$  the function  $\mathcal{Q}_t(\cdot, \xi_{[t-1]})$  is finite valued.

The above condition implies, of course, that  $\mathcal{Q}_t(\cdot, \xi_{[t]})$  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  $\mathcal{Q}_t(\cdot, \xi_{[t]})$  is required to be finite for all small perturbations of  $b_t$ .

**Proposition 3.4.** *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  $\bar{\pi}_t(\xi_{[t]})$ ,  $t = 1, \dots, T$ , such that the condition*

$$0 \in \partial f_t(\bar{x}_t(\xi_{[t]}), \xi_t) - A_t^\top \bar{\pi}_t(\xi_{[t]}) + \mathbb{E} [\partial \mathcal{Q}_{t+1}(\bar{x}_t(\xi_{[t]}), \xi_{[t+1]}) | \xi_{[t]}] \quad (3.29)$$

*holds true for a.e.  $\xi_{[t]}$  and  $t = 1, \dots, T$ . Moreover, multipliers  $\bar{\pi}_t(\xi_{[t]})$  satisfy (3.29) iff for a.e.  $\xi_{[t]}$  it holds that*

$$\bar{\pi}_t(\xi_{[t]}) \in \mathfrak{D}(\bar{x}_{t-1}(\xi_{[t-1]}), \xi_{[t]}). \quad (3.30)$$

**Proof.** Suppose that condition (D3) holds. Then by the Moreau–Rockafellar theorem (Theorem 7.4) we have that at  $\bar{x}_t = \bar{x}_t(\xi_{[t]})$ ,

$$\partial [f_t(\bar{x}_t, \xi_t) + \mathcal{Q}_{t+1}(\bar{x}_t, \xi_{[t]})] = \partial f_t(\bar{x}_t, \xi_t) + \partial \mathcal{Q}_{t+1}(\bar{x}_t, \xi_{[t]}).$$

Also by Theorem 7.47 the subdifferential of  $\mathcal{Q}_{t+1}(\bar{x}_t, \xi_{[t]})$  can be taken inside the expectation to obtain the last term in the right-hand side of (3.29). Note that conditional on  $\xi_{[t]}$  the term  $\bar{x}_t = \bar{x}_t(\xi_{[t]})$  is fixed. Optimality conditions (3.29) then follow from (3.28). Suppose, further, that condition (D2) holds. Then there is no duality gap between problems (3.22) and (3.23), and the second assertion follows by (3.27) and Proposition 3.3(ii).

If condition (D1) holds, then functions  $f_t(x_t, \xi_t)$  and  $\mathcal{Q}_{t+1}(x_t, \xi_{[t]})$  are random polyhedral, and hence the same arguments can be applied without additional regularity conditions.  $\square$

Formula (3.26) makes it possible to write optimality conditions (3.29) in the following form.

**Theorem 3.5.** *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  $\bar{\pi}_t(\xi_{[t]})$ ,  $t = 1, \dots, T$ , such that*

$$0 \in \partial f_t(\bar{x}_t(\xi_{[t]}), \xi_t) - A_t^\top \bar{\pi}_t(\xi_{[t]}) - \mathbb{E} [B_{t+1}^\top \bar{\pi}_{t+1}(\xi_{[t+1]}) | \xi_{[t]}] \quad (3.31)$$

*for a.e.  $\xi_{[t]}$  and  $t = 1, \dots, T$ , where for  $t = T$  the corresponding term  $T + 1$  is omitted.*

**Proof.** By Proposition 3.4 we have that a feasible policy  $\bar{x}_t(\xi_{[t]})$  is optimal iff conditions (3.29) and (3.30) hold true. For  $t = 1$  this means the existence of  $\bar{\pi}_1 \in \mathfrak{D}_1$  such that

$$0 \in \partial f_1(\bar{x}_1) - A_1^\top \bar{\pi}_1 + \mathbb{E}[\partial Q_2(\bar{x}_1, \xi_2)]. \quad (3.32)$$

Recall that  $\xi_1$  is known, and hence the set  $\mathfrak{D}_1$  is fixed. By (3.26) we have

$$\partial Q_2(\bar{x}_1, \xi_2) = -B_2^\top \mathfrak{D}_2(\bar{x}_1, \xi_2). \quad (3.33)$$

Formulas (3.32) and (3.33) mean that there exists a measurable selection

$$\bar{\pi}_2(\xi_2) \in \mathfrak{D}_2(\bar{x}_1, \xi_2)$$

such that (3.31) holds for  $t = 1$ . By the second assertion of Proposition 3.4, the same selection  $\bar{\pi}_2(\xi_2)$  can be used in (3.29) for  $t = 2$ . Proceeding in that way we obtain existence of measurable selections

$$\bar{\pi}_t(\xi_t) \in \mathfrak{D}_t(\bar{x}_{t-1}(\xi_{[t-1]}), \xi_{[t]})$$

satisfying (3.31).  $\square$

In particular, consider the multistage linear problem given in the nested formulation (3.9). That is, functions  $f_t(x_t, \xi_t)$  are defined in the form (3.21), which can be written as

$$f_t(x_t, \xi_t) = c_t^\top x_t + \mathbb{I}_{\mathbb{R}_+^{n_t}}(x_t).$$

Then  $\partial f_t(x_t, \xi_t) = \{c_t + \mathcal{N}_{\mathbb{R}_+^{n_t}}(x_t)\}$  at every point  $x_t \geq 0$ , and hence optimality conditions (3.31) take the form

$$0 \in \mathcal{N}_{\mathbb{R}_+^{n_t}}(\bar{x}_t(\xi_{[t]})) + c_t - A_t^\top \bar{\pi}_t(\xi_{[t]}) - \mathbb{E}[B_{t+1}^\top \bar{\pi}_{t+1}(\xi_{[t+1]}) | \xi_{[t]}].$$

### 3.2.3 Dualization of Feasibility Constraints

Consider the linear multistage program given in the nested formulation (3.9). In this section we discuss dualization of that problem with respect to the feasibility constraints. As 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  $\xi_t$  of the data process of that problem is formed from some (or all) elements of  $(c_t, B_t, A_t, b_t)$ ,  $t = 1, \dots, T$ . 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 (3.9) we associate the Lagrangian

$$\begin{aligned} L(x, \pi) &:= \mathbb{E} \left\{ \sum_{t=1}^T [c_t^\top x_t + \pi_t^\top (b_t - B_t x_{t-1} - A_t x_t)] \right\} \\ &= \mathbb{E} \left\{ \sum_{t=1}^T [c_t^\top x_t + \pi_t^\top b_t - \pi_t^\top A_t x_t - \pi_{t+1}^\top B_{t+1} x_t] \right\} \\ &= \mathbb{E} \left\{ \sum_{t=1}^T [b_t^\top \pi_t + (c_t - A_t^\top \pi_t - B_{t+1}^\top \pi_{t+1})^\top x_t] \right\} \end{aligned}$$

### 3.2. Duality

81

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  $x_t = x_t(\xi_{[t]})$ , are functions of the data process up to time  $t$ .

The dual functional is defined as

$$D(\pi) := \inf_{x \geq 0} L(x, \pi),$$

where the minimization is performed over variables  $x_t = x_t(\xi_{[t]})$ ,  $t = 1, \dots, T$ , in an appropriate functional space subject to the nonnegativity constraints. The Lagrangian dual of (3.9) is the problem

$$\text{Max}_{\pi} D(\pi), \quad (3.34)$$

where  $\pi$  lives in an appropriate functional space. Since, for a given  $\pi$ , the Lagrangian  $L(\cdot, \pi)$  is separable in  $x_t = x_t(\cdot)$ , by the interchangeability principle (Theorem 7.80) we can move the operation of minimization with respect to  $x_t$  inside the conditional expectation  $\mathbb{E}[\cdot | \xi_{[t]}]$ . Therefore, we obtain

$$D(\pi) = \mathbb{E} \left\{ \sum_{t=1}^T \left[ b_t^\top \pi_t + \inf_{x_t \in \mathbb{R}_+^{n_t}} (c_t - A_t^\top \pi_t - \mathbb{E}[B_{t+1}^\top \pi_{t+1} | \xi_{[t]}])^\top x_t \right] \right\}.$$

Clearly we have that  $\inf_{x_t \in \mathbb{R}_+^{n_t}} (c_t - A_t^\top \pi_t - \mathbb{E}[B_{t+1}^\top \pi_{t+1} | \xi_{[t]}])^\top x_t$  is equal to zero if  $A_t^\top \pi_t + \mathbb{E}[B_{t+1}^\top \pi_{t+1} | \xi_{[t]}] \leq c_t$ , and to  $-\infty$  otherwise. It follows that in the present case the dual problem (3.34) can be written as

$$\begin{aligned} \text{Max}_{\pi} \mathbb{E} \left[ \sum_{t=1}^T b_t^\top \pi_t \right] \\ \text{s.t. } A_t^\top \pi_t + \mathbb{E}[B_{t+1}^\top \pi_{t+1} | \xi_{[t]}] \leq c_t, \quad t = 1, \dots, T, \end{aligned} \quad (3.35)$$

where for the uniformity of notation we set all  $T + 1$  terms equal to zero. Each multiplier vector  $\pi_t = \pi_t(\xi_{[t]})$ ,  $t = 1, \dots, T$ , of problem (3.35) is a function of  $\xi_{[t]}$ . In that sense, these multipliers form a dual implementable policy. Optimization in (3.35) 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  $\pi_t(\cdot)$ ,  $t = 1, \dots, T$ , 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 3.6.** *Suppose that the data process has a finite number of possible realizations (scenarios). Then the optimal values of problems (3.9) and (3.35) 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 (3.9) and (3.35).

### 3.2.4 Dualization of Nonanticipativity Constraints

In this section we deal with a problem which is slightly more general than linear problem (3.12). Let  $f_t(x_t, \xi_t)$ ,  $t = 1, \dots, T$ , be *random polyhedral* functions, and consider the problem

$$\begin{aligned}
 \text{Min} \quad & \sum_{k=1}^K p_k \left[ f_1(x_1^k) + f_2^k(x_2^k) + f_3^k(x_3^k) + \dots + f_T^k(x_T^k) \right] \\
 \text{s.t.} \quad & \begin{aligned}
 & A_1 x_1^k & & & & & & = b_1, \\
 & B_2^k x_1^k & + & A_2^k x_2^k & & & & = b_2^k, \\
 & & & B_3^k x_2^k & + & A_3^k x_3^k & & = b_3^k, \\
 & \dots & & \dots & & \dots & & \dots \\
 & & & & & B_T^k x_{T-1}^k & + & A_T^k x_T^k & = b_T^k, \\
 & x_1^k \geq 0, & x_2^k \geq 0, & x_3^k \geq 0, & \dots & x_T^k \geq 0, \\
 & & & & & & & k = 1, \dots, K.
 \end{aligned}
 \end{aligned}$$

Here  $\xi_1^k, \dots, \xi_T^k$ ,  $k = 1, \dots, K$ , 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  $(B_t^k, A_t^k, b_t^k) := (B_t(\xi_t^k), A_t(\xi_t^k), b_t(\xi_t^k))$ ,  $t = 2, \dots, T$ . This problem can be formulated as a multistage stochastic programming problem by enforcing the corresponding nonanticipativity constraints.

As discussed in section 3.1.4, there are many ways to write nonanticipativity constraints. For example, let  $\mathfrak{X}$  be the linear space of all sequences  $(x_1^k, \dots, x_T^k)$ ,  $k = 1, \dots, K$ , and  $\mathfrak{L}$  be the linear subspace of  $\mathfrak{X}$  defined by the nonanticipativity constraints. (These spaces were defined above (3.17).) We can write the corresponding multistage problem in the following lucid form:

$$\text{Min}_{\mathbf{x} \in \mathfrak{X}} f(\mathbf{x}) := \sum_{k=1}^K \sum_{t=1}^T p_k f_t^k(x_t^k) \quad \text{s.t.} \quad \mathbf{x} \in \mathfrak{L}. \quad (3.36)$$

Clearly,  $f(\cdot)$  is a polyhedral function, so if problem (3.36) 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 (3.36),

$$L(\mathbf{x}, \boldsymbol{\lambda}) := f(\mathbf{x}) + \langle \boldsymbol{\lambda}, \mathbf{x} \rangle,$$

with the scalar product  $\langle \cdot, \cdot \rangle$  defined in (3.17). By the definition of the subspace  $\mathfrak{L}$ , every point  $\mathbf{x} \in \mathfrak{L}$  can be viewed as an implementable policy. By  $\mathfrak{L}^\perp := \{\mathbf{y} \in \mathfrak{X} : \langle \mathbf{y}, \mathbf{x} \rangle = 0, \forall \mathbf{x} \in \mathfrak{L}\}$  we denote the orthogonal subspace to the subspace  $\mathfrak{L}$ .

**Theorem 3.7.** *A policy  $\bar{\mathbf{x}} \in \mathfrak{L}$  is an optimal solution of (3.36) iff there exists a multiplier vector  $\bar{\boldsymbol{\lambda}} \in \mathfrak{L}^\perp$  such that*

$$\bar{\mathbf{x}} \in \arg \min_{\mathbf{x} \in \mathfrak{X}} L(\mathbf{x}, \bar{\boldsymbol{\lambda}}). \quad (3.37)$$

**Proof.** Let  $\bar{\boldsymbol{\lambda}} \in \mathfrak{L}^\perp$  and  $\bar{\mathbf{x}} \in \mathfrak{L}$  be a minimizer of  $L(\cdot, \bar{\boldsymbol{\lambda}})$  over  $\mathfrak{X}$ . Then by the first-order optimality conditions we have that  $0 \in \partial_{\mathbf{x}} L(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}})$ . Note that there is no need here for a constraint qualification since the problem is polyhedral. Now  $\partial_{\mathbf{x}} L(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}}) = \partial f(\bar{\mathbf{x}}) + \bar{\boldsymbol{\lambda}}$ .

### 3.2. Duality

83

Since  $\mathcal{N}_{\mathcal{L}}(\bar{\mathbf{x}}) = \mathcal{L}^\perp$ , it follows that  $0 \in \partial f(\bar{\mathbf{x}}) + \mathcal{N}_{\mathcal{L}}(\bar{\mathbf{x}})$ , which is a sufficient condition for  $\bar{\mathbf{x}}$  to be an optimal solution of (3.36). Conversely, if  $\bar{\mathbf{x}}$  is an optimal solution of (3.36), then necessarily  $0 \in \partial f(\bar{\mathbf{x}}) + \mathcal{N}_{\mathcal{L}}(\bar{\mathbf{x}})$ . This implies existence of  $\bar{\boldsymbol{\lambda}} \in \mathcal{L}^\perp$  such that  $0 \in \partial_{\bar{\mathbf{x}}} L(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}})$ . This, in turn, implies that  $\bar{\mathbf{x}} \in \mathcal{L}$  is a minimizer of  $L(\cdot, \bar{\boldsymbol{\lambda}})$  over  $\mathcal{X}$ .  $\square$

Also, we can define the dual function

$$D(\boldsymbol{\lambda}) := \inf_{\mathbf{x} \in \mathcal{X}} L(\mathbf{x}, \boldsymbol{\lambda}),$$

and the dual problem

$$\text{Max}_{\boldsymbol{\lambda} \in \mathcal{L}^\perp} D(\boldsymbol{\lambda}). \quad (3.38)$$

Since the problem considered is polyhedral, we have by the standard theory of linear programming the following results.

**Theorem 3.8.** *The optimal values of problems (3.36) and (3.38) 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  $\boldsymbol{\lambda} \in \mathcal{L}^\perp$ . Let us decipher this condition. For  $\boldsymbol{\lambda} = (\lambda_t^k)_{t=1, \dots, T, k=1, \dots, K}$ , the condition  $\boldsymbol{\lambda} \in \mathcal{L}^\perp$  is equivalent to

$$\sum_{t=1}^T \sum_{k=1}^K p_k \langle \lambda_t^k, x_t^k \rangle = 0, \quad \forall \mathbf{x} \in \mathcal{L}.$$

We can write this in a more abstract form as

$$\mathbb{E} \left[ \sum_{t=1}^T \langle \lambda_t, x_t \rangle \right] = 0, \quad \forall \mathbf{x} \in \mathcal{L}. \quad (3.39)$$

Since<sup>14</sup>  $\mathbb{E}_{|t} x_t = x_t$  for all  $\mathbf{x} \in \mathcal{L}$ , and  $\langle \lambda_t, \mathbb{E}_{|t} x_t \rangle = \langle \mathbb{E}_{|t} \lambda_t, x_t \rangle$ , we obtain from (3.39) that

$$\mathbb{E} \left[ \sum_{t=1}^T \langle \mathbb{E}_{|t} \lambda_t, x_t \rangle \right] = 0, \quad \forall \mathbf{x} \in \mathcal{L},$$

which is equivalent to

$$\mathbb{E}_{|t} [\lambda_t] = 0, \quad t = 1, \dots, T. \quad (3.40)$$

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

$$\text{Max}_{\boldsymbol{\lambda} \in \mathcal{L}^\perp} D(\boldsymbol{\lambda}) \quad \text{s.t.} \quad \mathbb{E}_{|t} [\lambda_t] = 0, \quad t = 1, \dots, T. \quad (3.41)$$

<sup>14</sup>In order to simplify notation, we denote in the remainder of this section by  $\mathbb{E}_{|t}$  the conditional expectation, conditional on  $\xi_{[t]}$ .

**Corollary 3.9.** *A policy  $\bar{x} \in \mathcal{L}$  is an optimal solution of (3.36) iff there exist multipliers vector  $\bar{\lambda}$  satisfying (3.40) such that*

$$\bar{x} \in \arg \min_{x \in \mathcal{X}} L(x, \bar{\lambda}).$$

*Moreover, problem (3.36) has an optimal solution iff problem (3.41) 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 (3.18) 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 (3.20).
  - (b) Transform the problem to an equivalent linear multistage stochastic programming problem.
- 3.2. Consider the cost-to-go function  $Q_t(x_{t-1}, \xi_{[t]}), t = 2, \dots, T$ , of the linear multistage problem defined as the optimal value of problem (3.7). 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 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 stagewise 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  $j \neq i$  on day  $t$ , where  $t = 1, 2, \dots, T$ , is modeled as a random variable  $D_{ij}(t)$ . The total capacity of vehicles currently available at depot  $i$  is denoted  $s_i, i = 1, \dots, n$ . 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  $s_i, i = 1, \dots, n$ , 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 stagewise 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 (3.12) with nonanticipativity constraints in the form (3.18).
- 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  $t = 1, \dots, T$ . 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 price are available. There are  $n$  put options with strike prices  $p_1, \dots, p_n$  and corresponding costs  $c_1, \dots, c_n$ . For example, if you buy one put option  $i$ , at any time  $t = 1, \dots, T$  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  $\pi_1, \dots, \pi_m$  and corresponding costs  $q_1, \dots, q_m$ . For example, if you buy one call option  $j$ , at any time  $t = 1, \dots, T$  you may exercise it and cash  $S_t - \pi_j$  (this makes sense only when  $\pi_j < S_t$ ). 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:

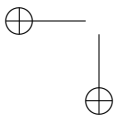
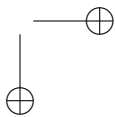
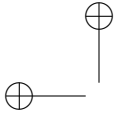
- (i) The expected value of your holdings at the end of period  $T$ .
- (ii) The expected value of a piecewise linear utility function evaluated at the value of your final holdings. Its form is

$$u(C_T) = \begin{cases} C_T & \text{if } C_T \geq 0, \\ (1 + R)C_T & \text{if } C_T < 0, \end{cases}$$

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

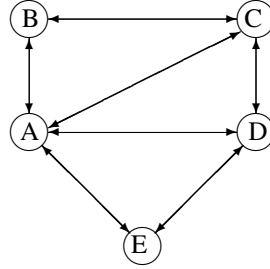
In this chapter, we discuss stochastic optimization problems with *probabilistic* (also called *chance*) constraints of the form

$$\begin{aligned} \text{Min } & c(x) \\ \text{s.t. } & \Pr\{g_j(x, Z) \leq 0, j \in \mathcal{J}\} \geq p, \\ & x \in \mathcal{X}. \end{aligned} \tag{4.1}$$

Here  $\mathcal{X} \subset \mathbb{R}^n$  is a nonempty set,  $c : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g_j : \mathbb{R}^n \times \mathbb{R}^s \rightarrow \mathbb{R}$ ,  $j \in \mathcal{J}$ , where  $\mathcal{J}$  is an index set,  $Z$  is an  $s$ -dimensional random vector, and  $p$  is a modeling parameter. We denote by  $P_Z$  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 reflects the point of view that for a given decision  $x$  we do not reject the statistical hypothesis that the constraints  $g_j(x, Z) \leq 0$ ,  $j \in \mathcal{J}$ , are satisfied. We discussed examples and a motivation for such problems in Chapter 1 in the contexts of inventory, multiproduct, and portfolio selection models. We emphasize that imposing constraints on probability of events is particularly appropriate whenever high uncertainty is involved and reliability is a central issue. In such cases, constraints on the expected value may not be sufficient to reflect our attitude to undesirable outcomes.

We also note that the objective function  $c(x)$  can represent an expected value function, i.e.,  $c(x) = \mathbb{E}[f(x, Z)]$ ; however, we focus on the analysis of the probabilistic constraints at the moment.



**Figure 4.1.** *Vehicle routing network*

We can write the probability  $\Pr\{A(x)\}$  as the expected value of the characteristic function of the event  $A(x)$ , i.e.,  $\Pr\{A(x)\} = \mathbb{E}[\mathbf{1}_{A(x)}]$ . The discontinuity of the characteristic function and the complexity of the event  $A(x)$  make such problems qualitatively different from the expectation models. 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 to satisfy transportation demand. Figure 4.1 depicts a small network, and the table in Figure 4.2 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  $j$ ,  $j = 1, \dots, n$ . The demand for transportation on each arc is given by the random variables  $Z_i, i = 1, \dots, m$ . We set  $Z = (Z_1, \dots, Z_m)^T$ . A cost  $c_j$  is associated with operating a vehicle on route  $j$ . Setting  $c = (c_1, \dots, c_n)^T$ , the model can be formulated as follows:<sup>15</sup>

$$\text{Min}_x \quad c^T x \tag{4.2}$$

$$\text{s.t. } \Pr\{Tx \geq Z\} \geq p, \tag{4.3}$$

$$x \in \mathbb{Z}_+^n. \tag{4.4}$$

In practical applications, we may have a 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.

<sup>15</sup>The notation  $\mathbb{Z}_+$  is used to denote the set of nonnegative integer numbers.

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

**Example 4.2 (Portfolio Optimization with Value-at-Risk Constraint).** We consider  $n$  investment opportunities with random return rates  $R_1, \dots, R_n$  in the next year. We have certain initial capital and our aim is to invest it in such a way that the expected value of our investment after a year is maximized, under the condition that the chance of losing no more than a given fraction of this amount is at least  $p$ , where  $p \in (0, 1)$ . Such a requirement is called a *Value-at-Risk* (V@R) constraint (already discussed in Chapter 1).

Let  $x_1, \dots, x_n$  be the fractions of our capital invested in the  $n$  assets. After a year, our investment changes in value according to a rate that can be expressed as

$$g(x, R) = \sum_{i=1}^n R_i x_i.$$

We formulate the following stochastic optimization problem with a probabilistic constraint:

$$\begin{aligned}
 & \text{Max} \quad \sum_{i=1}^n \mathbb{E}[R_i] x_i \\
 & \text{s.t.} \quad \Pr \left\{ \sum_{i=1}^n R_i x_i \geq \eta \right\} \geq p, \\
 & \quad \sum_{i=1}^n x_i = 1, \\
 & \quad x \geq 0.
 \end{aligned} \tag{4.5}$$

For example,  $\eta = -0.1$  may be chosen if we aim at protecting against losses larger than 10%. ■

The constraint

$$\Pr\{g_j(x, Z) \leq 0, j \in \mathcal{J}\} \geq p$$

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:

$$\Pr\{T^i x \geq Z_i\} \geq p_i, \quad i = 1, \dots, m,$$

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

$$F_X(\eta) \leq F_Y(\eta), \quad \forall \eta \in \mathbb{R}.$$

The left-continuous inverse  $F_X^{(-1)}$  of the cumulative distribution function of a random variable  $X$  is defined as follows:

$$F_X^{(-1)}(p) = \inf \{ \eta : F_1(X; \eta) \geq p \}, \quad p \in (0, 1).$$

Given  $p \in (0, 1)$ , the number  $q = q(X; p)$  is called a  $p$ -quantile of the random variable  $X$  if

$$\Pr\{X < q\} \leq p \leq \Pr\{X \leq q\}.$$

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

$$X \succeq_{(1)} Y \Leftrightarrow F_X^{(-1)}(p) \geq F_Y^{(-1)}(p), \quad \forall p \in (0, 1). \quad (4.6)$$

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:

$$F_X^{(k)}(\eta) = \int_{-\infty}^{\eta} F_X^{(k-1)}(t) dt \quad \text{for } k = 2, 3, 4, \dots$$

We can express the integrated distribution function  $F_X^{(2)}$  as the expected shortfall function. Integrating by parts, for each value  $\eta$ , we have the following formula:<sup>16</sup>

$$F_X^{(2)}(\eta) = \int_{-\infty}^{\eta} F_X(\alpha) d\alpha = \mathbb{E}[(\eta - X)_+]. \quad (4.7)$$

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), the second order dominance relation can be expressed in an equivalent way as follows:

$$X \succeq_{(2)} Y \text{ iff } \mathbb{E}[(\eta - X)_+] \leq \mathbb{E}[(\eta - Y)_+], \quad \forall \eta \in \mathbb{R}. \quad (4.8)$$

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  $k$ th order if

$$F_X^{(k)}(\eta) \leq F_Y^{(k)}(\eta), \quad \forall \eta \in \mathbb{R}.$$

We denote this relation by  $X \succeq_{(k)} Y$ .

We call the following semi-infinite (probabilistic) problem a *stochastic optimization problem with a stochastic ordering constraint*:

$$\begin{aligned} & \text{Min}_x \quad c(x) \\ & \text{s.t.} \quad \Pr \{g(x, Z) \leq \eta\} \leq F_Y(\eta), \quad \eta \in [a, b], \\ & \quad \quad x \in \mathcal{X}. \end{aligned} \quad (4.9)$$

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  $k \geq 2$ :

$$\begin{aligned} & \text{Min}_x \quad c(x) \\ & \text{s.t.} \quad F_{g(x, Z)}^{(k)}(\eta) \leq F_Y^{(k)}(\eta), \quad \eta \in [a, b], \\ & \quad \quad x \in \mathcal{X}. \end{aligned} \quad (4.10)$$

<sup>16</sup>Recall that  $[a]_+ = \max\{a, 0\}$ .

**Example 4.5 (Portfolio Selection Problem with Stochastic Ordering 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:

$$\begin{aligned} \text{Max} \quad & \sum_{i=1}^n \mathbb{E}[R_i]x_i \\ \text{s.t.} \quad & \Pr \left\{ \sum_{i=1}^n R_i x_i \leq \eta \right\} \leq p_\eta, \quad \forall \eta \in \mathbb{R}, \\ & \sum_{i=1}^n x_i = 1, \\ & x \geq 0. \end{aligned} \tag{4.11}$$

Using higher order stochastic dominance relations, we formulate a portfolio optimization model of form

$$\begin{aligned} \text{Max} \quad & \sum_{i=1}^n \mathbb{E}[R_i]x_i \\ \text{s.t.} \quad & \sum_{i=1}^n R_i x_i \succeq_{(k)} Y, \\ & \sum_{i=1}^n x_i = 1, \\ & x \geq 0. \end{aligned} \tag{4.12}$$

A second order dominance constraint on the portfolio return rate represents a constraint on the shortfall function:

$$\sum_{i=1}^n R_i x_i \succeq_{(2)} Y \iff \mathbb{E} \left[ \left( \eta - \sum_{i=1}^n R_i x_i \right)_+ \right] \leq \mathbb{E}[(\eta - Y)_+], \quad \forall \eta \in \mathbb{R}.$$

The second order dominance constraint can also be viewed as a continuum of *Average Value-at-Risk*<sup>17</sup> (AV@R) constraints. For more information on this connection, see Dentcheva and Ruszczyński [56]. ■

We stress that 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  $g_i$ ,  $i = 1, \dots, m$ , that is, functions of form  $g_i(x, z) = \hat{g}_i(x) + h_i(z)$ . The probabilistic constraint becomes

$$\Pr\{\hat{g}_i(x) \geq -h_i(Z), \ i = 1, \dots, m\} \geq p.$$

<sup>17</sup>Average Value-at-Risk is also called Conditional Value-at-Risk.



We can view the inequalities under the probability as a deterministic vector function  $\hat{g} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $\hat{g} = [\hat{g}_1, \dots, \hat{g}_m]^T$  constrained from below by a random vector  $Y$  with  $Y_i = -h_i(Z)$ ,  $i = 1, \dots, m$ . The problem can be formulated as

$$\begin{aligned} \text{Min}_x \quad & c(x) \\ \text{s.t.} \quad & \Pr\{\hat{g}(x) \geq Y\} \geq p, \\ & x \in \mathcal{X}, \end{aligned} \tag{4.13}$$

where the inequality  $a \leq b$  for two vectors  $a, b \in \mathbb{R}^n$  is understood componentwise.

We note again that the objective function can have a more specific form:

$$c(x) = \mathbb{E}[f(x, Z)].$$

By virtue of Theorem 7.43, we have that if the function  $f(\cdot, Z)$  is continuous at  $x_0$  w.p. 1 and there exists an integrable random variable  $\hat{Z}$  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  $f(\cdot, Z)$  for a.e.  $Z$  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.

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 example 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  $t = 1, \dots, T$ . 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  $1, \dots, T$  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:

$$c_t = c_{t-1} + \sum_{i=1}^n a_{it}x_i, \quad t = 1, \dots, T.$$

Using cumulative cash flows and cumulative liabilities permits the carryover of capital from one stage to the next, 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  $c = (c_1, \dots, c_T)^T$ . Let us assume that we quantify our preferences by using concave utility function  $U : \mathbb{R} \rightarrow \mathbb{R}$ . We would like to maximize the final capital at hand in a risk-averse manner. The problem takes on the form

$$\begin{aligned} & \text{Max}_{x,c} \mathbb{E}[U(c_T - Z_T)] \\ & \text{s.t. } \Pr\{c_t \geq Z_t, t = 1, \dots, T\} \geq p, \\ & \quad c_t = c_{t-1} + \sum_{i=1}^n a_{it}x_i, \quad t = 1, \dots, T, \\ & \quad x \geq 0. \end{aligned}$$

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  $f : \Omega \rightarrow \mathbb{R}_+$  defined on a convex set  $\Omega \subset \mathbb{R}^n$ .

**Definition 4.7.** A nonnegative function  $f(x)$  defined on a convex set  $\Omega \subset \mathbb{R}^n$  is said to be  $\alpha$ -concave, where  $\alpha \in [-\infty, +\infty]$ , if for all  $x, y \in \Omega$  and all  $\lambda \in [0, 1]$  the following inequality holds true:

$$f(\lambda x + (1 - \lambda)y) \geq m_\alpha(f(x), f(y), \lambda),$$

where  $m_\alpha : \mathbb{R}_+ \times \mathbb{R}_+ \times [0, 1] \rightarrow \mathbb{R}$  is defined as follows:

$$m_\alpha(a, b, \lambda) = 0 \quad \text{if } ab = 0,$$

and if  $a > 0, b > 0, 0 \leq \lambda \leq 1$ , then

$$m_\alpha(a, b, \lambda) = \begin{cases} a^\lambda b^{1-\lambda} & \text{if } \alpha = 0, \\ \max\{a, b\} & \text{if } \alpha = \infty, \\ \min\{a, b\} & \text{if } \alpha = -\infty, \\ (\lambda a^\alpha + (1 - \lambda)b^\alpha)^{1/\alpha} & \text{otherwise.} \end{cases}$$

## 4.2. Convexity in Probabilistic Optimization

95

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  $\alpha$  and all  $\lambda \in (0, 1)$ . Furthermore,  $m_\alpha(a, b, \lambda)$  has the following important 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:

$$\begin{aligned} \ln m_\alpha(a, b, \lambda) &= \ln(\lambda a^\alpha + (1 - \lambda)b^\alpha)^{1/\alpha} = \frac{1}{\alpha} \ln(\lambda e^{\alpha \ln a} + (1 - \lambda)e^{\alpha \ln b}) \\ &= \frac{1}{\alpha} \ln(1 + \alpha(\lambda \ln a + (1 - \lambda) \ln b) + o(\alpha^2)). \end{aligned}$$

Applying the l'Hôpital rule to the right-hand side in order to calculate its limit when  $\alpha \rightarrow 0$ , we obtain

$$\begin{aligned} \lim_{\alpha \rightarrow 0} \ln m_\alpha(a, b, \lambda) &= \lim_{\alpha \rightarrow 0} \frac{\lambda \ln a + (1 - \lambda) \ln b + o(\alpha)}{1 + \alpha(\lambda \ln a + (1 - \lambda) \ln b) + o(\alpha^2)} \\ &= \lim_{\alpha \rightarrow 0} \frac{\ln(a^\lambda b^{(1-\lambda)}) + o(\alpha)}{1 + \alpha \ln(a^\lambda b^{(1-\lambda)}) + o(\alpha^2)} = \ln(a^\lambda b^{(1-\lambda)}). \end{aligned}$$

Now we turn to the monotonicity of the mapping. First, let us consider the case of  $0 < \alpha < \beta$ . We set

$$h(\alpha) = m_\alpha(a, b, \lambda) = \exp\left(\frac{1}{\alpha} \ln[\lambda a^\alpha + (1 - \lambda)b^\alpha]\right).$$

Calculating its derivative, we obtain

$$h'(\alpha) = h(\alpha) \left( \frac{1}{\alpha} \cdot \frac{\lambda a^\alpha \ln a + (1 - \lambda)b^\alpha \ln b}{\lambda a^\alpha + (1 - \lambda)b^\alpha} - \frac{1}{\alpha^2} \ln[\lambda a^\alpha + (1 - \lambda)b^\alpha] \right).$$

We have to demonstrate that the expression on the right-hand side is nonnegative. Substituting  $x = a^\alpha$  and  $y = b^\alpha$ , we obtain

$$h'(\alpha) = \frac{1}{\alpha^2} h(\alpha) \left( \frac{\lambda x \ln x + (1 - \lambda)y \ln y}{\lambda x + (1 - \lambda)y} - \ln[\lambda x + (1 - \lambda)y] \right).$$

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

$$\frac{\lambda x \ln x + (1 - \lambda)y \ln y}{\lambda x + (1 - \lambda)y} - \ln[\lambda x + (1 - \lambda)y] \geq 0.$$

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:

$$m_\alpha(a, b, \lambda) = \left[ m_{-\alpha}\left(\frac{1}{a}, \frac{1}{b}, \lambda\right) \right]^{-1} \leq \left[ m_{-\beta}\left(\frac{1}{a}, \frac{1}{b}, \lambda\right) \right]^{-1} = m_\beta(a, b, \lambda).$$

In the case of  $0 = \alpha < \beta$ , we can select a sequence  $\{\alpha_k\}$  such that  $\alpha_k > 0$  and  $\lim_{k \rightarrow \infty} \alpha_k = 0$ . 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  $\alpha \rightarrow \infty$  or  $\alpha \rightarrow -\infty$ . We observe that

$$\max\{\lambda^{1/\alpha}a, (1-\lambda)^{1/\alpha}b\} \leq m_\alpha(a, b, \lambda) \leq \max\{a, b\}.$$

Passing to the limit, we obtain that

$$\lim_{\alpha \rightarrow \infty} m_\alpha(a, b, \lambda) = \max\{a, b\}.$$

We also conclude that

$$\lim_{\alpha \rightarrow -\infty} m_\alpha(a, b, \lambda) = \lim_{\alpha \rightarrow -\infty} [m_{-\alpha}(1/a, 1/b, \lambda)]^{-1} = [\max\{1/a, 1/b\}]^{-1} = \min\{a, b\}.$$

This completes the proof.  $\square$

This statement has the very important implication that  $\alpha$ -concavity entails  $\beta$ -concavity for all  $\beta \leq \alpha$ . Therefore, all  $\alpha$ -concave functions are  $(-\infty)$ -concave, that is, *quasi-concave*.

**Example 4.9.** Consider the density function of a nondegenerate multivariate normal distribution on  $\mathbb{R}^s$ :

$$\theta(x) = \frac{1}{\sqrt{(2\pi)^s \det(\Sigma)}} \exp \left\{ -\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right\},$$

where  $\Sigma$  is a positive definite symmetric matrix of dimension  $s \times s$ ,  $\det(\Sigma)$  denotes the determinant of the matrix  $\Sigma$ , and  $\mu \in \mathbb{R}^s$ . We observe that

$$\ln \theta(x) = -\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) - \ln \left( \sqrt{(2\pi)^s \det(\Sigma)} \right)$$

is a concave function. Therefore, we conclude that  $\theta$  is 0-concave, or log-concave.  $\blacksquare$

**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 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  $\Omega$ . The function  $\theta(x)$  is quasi-concave on  $\mathbb{R}^s$  and  $+\infty$ -concave on  $\Omega$ .  $\blacksquare$

## 4.2. Convexity in Probabilistic Optimization

97

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  $\mathbb{R}^s$ .

**Definition 4.11.** A probability measure  $P$  defined on the Lebesgue measurable subsets of a convex set  $\Omega \subset \mathbb{R}^s$  is said to be  $\alpha$ -concave if for any Borel measurable sets  $A, B \subset \Omega$  and for all  $\lambda \in [0, 1]$  we have the inequality

$$P(\lambda A + (1 - \lambda)B) \geq m_\alpha(P(A), P(B), \lambda),$$

where  $\lambda A + (1 - \lambda)B = \{\lambda x + (1 - \lambda)y : x \in A, y \in B\}$ .

We say that a random vector  $Z$  with values in  $\mathbb{R}^n$  has an  $\alpha$ -concave distribution if the probability measure  $P_Z$  induced by  $Z$  on  $\mathbb{R}^n$  is  $\alpha$ -concave.

**Lemma 4.12.** If a random vector  $Z$  induces an  $\alpha$ -concave probability measure on  $\mathbb{R}^s$ , then its cumulative distribution function  $F_Z$  is an  $\alpha$ -concave function.

**Proof.** Indeed, for given points  $z^1, z^2 \in \mathbb{R}^s$  and  $\lambda \in [0, 1]$ , we define

$$A = \{z \in \mathbb{R}^s : z_i \leq z_i^1, i = 1, \dots, s\} \quad \text{and} \quad B = \{z \in \mathbb{R}^s : z_i \leq z_i^2, i = 1, \dots, s\}.$$

Then the inequality for  $F_Z$  follows from the inequality in Definition 4.11.  $\square$

**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  $A, B \subset \mathbb{R}^s$  and  $\lambda \in (0, 1)$ , we define the set  $C = \lambda A + (1 - \lambda)B$ . Denote the projections of  $A, B$  and  $C$  on the coordinate axis by  $A_i, B_i$  and  $C_i, i = 1, \dots, s$ , respectively. For any number  $r \in C_i$  there is  $c \in C$  such that  $c_i = r$ , which implies that we have  $a \in A$  and  $b \in B$  with  $\lambda a + (1 - \lambda)b = c$  and  $r = \lambda a_i + (1 - \lambda)b_i$ . In other words,  $r \in \lambda A_i + (1 - \lambda)B_i$ , and we conclude that  $C_i \subset \lambda A_i + (1 - \lambda)B_i$ . On the other hand, if  $r \in \lambda A_i + (1 - \lambda)B_i$ , then we have  $a \in A$  and  $b \in B$  such that  $r = \lambda a_i + (1 - \lambda)b_i$ . Setting  $c = \lambda a + (1 - \lambda)b$ , we conclude that  $r \in C_i$ . We obtain

$$\begin{aligned} \ln[P_Z(C)] &= \sum_{i=1}^s \ln[P_{Z_i}(C_i)] = \sum_{i=1}^s \ln[P_{Z_i}(\lambda A_i + (1 - \lambda)B_i)] \\ &\geq \sum_{i=1}^s (\lambda \ln[P_{Z_i}(A_i)] + (1 - \lambda) \ln[P_{Z_i}(B_i)]) \\ &= \lambda \ln[P_Z(A)] + (1 - \lambda) \ln[P_Z(B)]. \quad \square \end{aligned}$$

As usually, concavity properties of a function imply a certain continuity of the function. We formulate without proof two theorems addressing this issue.

**Theorem 4.14 (Borell [24]).** 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.

We can relate the  $\alpha$ -concavity property of a measure to generalized concavity of its density. (See Brascamp and Lieb [26], Prékopa [159], Rinott [168], and the references therein.)

**Theorem 4.15.** *Let  $\Omega$  be a convex subset of  $\mathbb{R}^s$  and let  $m > 0$  be the dimension of the smallest affine subspace  $L$  containing  $\Omega$ . The probability measure  $P$  on  $\Omega$  is  $\gamma$ -concave with  $\gamma \in [-\infty, 1/m]$  iff its probability density function with respect to the Lebesgue measure on  $L$  is  $\alpha$ -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 define and positive on a nondegenerate convex set  $\Omega \subset \mathbb{R}^s$ . Denote  $c = \int_{\Omega} \theta(x) dx$ . If  $\theta(x)$  is  $\alpha$ -concave with  $\alpha \in [-1/s, \infty]$  and positive on the interior of  $\Omega$ , then the measure  $P$  on  $\Omega$  defined by setting that*

$$P(A) = \frac{1}{c} \int_A \theta(x) dx, \quad A \subset \Omega,$$

*is  $\gamma$ -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.** We observed in Example 4.10 that the density of the uniform distribution on a convex body  $\Omega$  is a  $\infty$ -concave function. Hence, it generates a  $1/s$ -concave measure on  $\Omega$ . 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  $\alpha_1, \dots, \alpha_s$  and the simplex

$$S = \left\{ x \in \mathbb{R}^s : \sum_{i=1}^s x_i \leq 1, x_i \geq 0, i = 1, \dots, s \right\}.$$

The density function of the *Dirichlet distribution* with parameters  $\alpha_1, \dots, \alpha_s$  is defined as follows:

$$\theta(x) = \begin{cases} \frac{\Gamma(\alpha_1 + \dots + \alpha_s)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_s)} x_1^{\alpha_1-1} x_2^{\alpha_2-1} \dots x_s^{\alpha_s-1} & \text{if } x \in \text{int } S, \\ 0 & \text{otherwise.} \end{cases}$$

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

$$\ln \theta(x) = \sum_{i=1}^s (\alpha_i - 1) \ln x_i + \ln \Gamma(\alpha_1 + \dots + \alpha_s) - \sum_{i=1}^s \ln \Gamma(\alpha_i).$$

## 4.2. Convexity in Probabilistic Optimization

99

If  $\alpha_i \geq 1$  for all  $i = 1, \dots, s$ , then  $\ln \theta(\cdot)$  is a concave function on the interior of  $S$  and, therefore,  $\theta(x)$  is log-concave on  $\text{cl } S$ . If all parameters satisfy  $\alpha_i \leq 1$ , then  $\theta(x)$  is log-convex on  $\text{cl } (S)$ . For other sets of parameters, this density function does not have any generalized concavity properties. ■

The next results provide calculus rules for  $\alpha$ -concave functions.

**Theorem 4.19.** *If the function  $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$  is  $\alpha$ -concave and the function  $g : \mathbb{R}^n \rightarrow \mathbb{R}_+$  is  $\beta$ -concave, where  $\alpha, \beta \geq 1$ , then the function  $h : \mathbb{R}^n \rightarrow \mathbb{R}$ , defined as  $h(x) = f(x) + g(x)$  is  $\gamma$ -concave with  $\gamma = \min\{\alpha, \beta\}$ .*

**Proof.** Given points  $x_1, x_2 \in \mathbb{R}^n$  and a scalar  $\lambda \in (0, 1)$ , we set  $x_\lambda = \lambda x_1 + (1 - \lambda)x_2$ . Both functions  $f$  and  $g$  are  $\gamma$ -concave by virtue of Lemma 4.8. Using the Minkowski inequality, which holds true for  $\gamma \geq 1$ , we obtain

$$\begin{aligned} f(x_\lambda) + g(x_\lambda) &\geq [\lambda(f(x_1))^\gamma + (1 - \lambda)(f(x_2))^\gamma]^\frac{1}{\gamma} + [\lambda(g(x_1))^\gamma + (1 - \lambda)(g(x_2))^\gamma]^\frac{1}{\gamma} \\ &\geq [\lambda(f(x_1) + g(x_1))^\gamma + (1 - \lambda)(f(x_2) + g(x_2))^\gamma]^\frac{1}{\gamma}. \end{aligned}$$

This completes the proof. □

**Theorem 4.20.** *Let  $f$  be a concave function defined on a convex set  $C \subset \mathbb{R}^s$  and  $g : \mathbb{R} \rightarrow \mathbb{R}$  be a nonnegative nondecreasing  $\alpha$ -concave function,  $\alpha \in [-\infty, \infty]$ . Then the function  $g \circ f$  is  $\alpha$ -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  $f(z) \geq \lambda f(x) + (1 - \lambda)f(y)$ . By monotonicity and  $\alpha$ -concavity of  $g$ , we obtain the following chain of inequalities:

$$[g \circ f](z) \geq g(\lambda f(x) + (1 - \lambda)f(y)) \geq m_\alpha(g(f(x)), g(f(y)), \lambda).$$

This proves the assertion. □

**Theorem 4.21.** *Let the function  $f : \mathbb{R}^m \times \mathbb{R}^s \rightarrow \mathbb{R}_+$  be such that for all  $y \in Y \subset \mathbb{R}^s$  the function  $f(\cdot, y)$  is  $\alpha$ -concave ( $\alpha \in [-\infty, \infty]$ ) on the convex set  $X \subset \mathbb{R}^m$ . Then the function  $\varphi(x) = \inf_{y \in Y} f(x, y)$  is  $\alpha$ -concave on  $X$ .*

**Proof.** Let  $x_1, x_2 \in X$  and a scalar  $\lambda \in (0, 1)$  be given. We set  $z = \lambda x_1 + (1 - \lambda)x_2$ . We can find a sequence of points  $y_k \in Y$  such that

$$\varphi(z) = \inf_{y \in Y} f(z, y) = \lim_{k \rightarrow \infty} f(z, y_k).$$

Using the  $\alpha$ -concavity of the function  $f(\cdot, y)$ , we conclude that

$$f(z, y_k) \geq m_\alpha(f(x_1, y_k), f(x_2, y_k), \lambda).$$

The mapping  $(a, b) \mapsto m_\alpha(a, b, \lambda)$  is monotone for nonnegative  $a$  and  $b$  and  $\lambda \in (0, 1)$ . Therefore, we have that

$$f(z, y_k) \geq m_\alpha(\varphi(x_1), \varphi(x_2), \lambda).$$

Passing to the limit, we obtain the assertion.  $\square$

**Lemma 4.22.** *If  $\alpha_i > 0$ ,  $i = 1, \dots, m$ , and  $\sum_{i=1}^m \alpha_i = 1$ , then the function  $f : \mathbb{R}_+^m \rightarrow \mathbb{R}$ , 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:

$$\begin{aligned} f(\lambda x + (1 - \lambda)y) &= \left( a_1^{\frac{1}{\alpha_1}} + a_2^{\frac{1}{\alpha_1}} \right)^{\alpha_1} \left( b_1^{\frac{1}{\alpha_2}} + b_2^{\frac{1}{\alpha_2}} \right)^{\alpha_2} \\ &\geq a_1 b_1 + a_2 b_2 = \lambda x_1^{\alpha_1} x_2^{\alpha_2} + (1 - \lambda) y_1^{\alpha_1} y_2^{\alpha_2}. \end{aligned}$$

The assertion in the general case follows by induction.  $\square$

**Theorem 4.23.** *If the functions  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}_+$ ,  $i = 1, \dots, m$ , are  $\alpha_i$ -concave and  $\alpha_i$  are such that  $\sum_{i=1}^m \alpha_i^{-1} > 0$ , then the function  $g : \mathbb{R}^{nm} \rightarrow \mathbb{R}_+$ , defined as  $g(x) = \prod_{i=1}^m f_i(x_i)$  is  $\gamma$ -concave with  $\gamma = (\sum_{i=1}^m \alpha_i^{-1})^{-1}$ .*

**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  $f_i$ ,  $i = 1, \dots, m$ , we have the following inequality:

$$\prod_{i=1}^m f_i(x_\lambda) \geq \prod_{i=1}^m \left( \lambda f_i(x_1)^{\alpha_i} + (1 - \lambda) f_i(x_2)^{\alpha_i} \right)^{1/\alpha_i}.$$

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  $\gamma$ , we obtain

$$\left( \prod_{i=1}^m f_i(x_\lambda) \right)^\gamma \geq \prod_{i=1}^m (\lambda y_{i1} + (1 - \lambda) y_{i2})^{\gamma/\alpha_i}.$$

We continue the chain of inequalities using Lemma 4.22:

$$\prod_{i=1}^m (\lambda y_{i1} + (1 - \lambda) y_{i2})^{\gamma/\alpha_i} \geq \lambda \prod_{i=1}^m [y_{i1}]^{\gamma/\alpha_i} + (1 - \lambda) \prod_{i=1}^m [y_{i2}]^{\gamma/\alpha_i}.$$

Putting the inequalities together and using the substitutions at the right-hand side of the last inequality, we conclude that

$$\prod_{i=1}^m [f_i(x_\lambda)]^\gamma \geq \lambda \prod_{i=1}^m [f_i(x_1)]^\gamma + (1 - \lambda) \prod_{i=1}^m [f_i(x_2)]^\gamma,$$

as required.  $\square$



In the special case, when the functions  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $i = 1, \dots, k$ , are concave, we can apply Theorem 4.23 consecutively to conclude that  $f_1 f_2$  is  $\frac{1}{2}$ -concave and  $f_1 \cdots f_k$  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  $\lambda$  of  $A$ ,  $\gamma\lambda$  is an eigenvalue of  $\gamma A$ , and, hence,  $\det(\gamma A) = \gamma^n \det(A)$ . We could apply the Minkowski inequality for matrices,

$$[\det(A + B)]^{\frac{1}{n}} \geq [\det(A)]^{\frac{1}{n}} + [\det(B)]^{\frac{1}{n}}, \quad (4.14)$$

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

$$\begin{aligned} \det(A + B) &= \det(A^{1/2} A^{-1/2} (A + B) A^{-1/2} A^{1/2}) \\ &= \det(A^{-1/2} (A + B) A^{-1/2}) \det(A) \\ &= \det(I + A^{-1/2} B A^{-1/2}) \det(A). \end{aligned} \quad (4.15)$$

Notice that  $A^{-1/2} B A^{-1/2}$  is symmetric positive definite and, therefore, we can choose an  $n \times n$  orthogonal matrix  $R$ , which diagonalizes it. We obtain

$$\begin{aligned} \det(I + A^{-1/2} B A^{-1/2}) &= \det(R^T (I + A^{-1/2} B A^{-1/2}) R) \\ &= \det(I + R^T A^{-1/2} B A^{-1/2} R). \end{aligned}$$

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

$$\begin{aligned} [\det(I + A^{-1/2} B A^{-1/2})]^{\frac{1}{n}} &= [\det(I + R^T A^{-1/2} B A^{-1/2} R)]^{\frac{1}{n}} \\ &\geq 1 + [\det(R^T A^{-1/2} B A^{-1/2} R)]^{\frac{1}{n}} \\ &= 1 + [\det(B)]^{\frac{1}{n}} [\det(A)]^{-\frac{1}{n}}. \end{aligned}$$

Combining this inequality with (4.15), we obtain (4.14) in the general case.  $\square$

**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  $\gamma$ -concave with  $\gamma = (\alpha_1 + \cdots + \alpha_m - s)^{-1}$  provided that  $\alpha_i > 1$ ,  $i = 1, \dots, m$ . Due to Corollary 4.16, the Dirichlet distribution is a  $(\alpha_1 + \cdots + \alpha_m)^{-1}$ -concave probability measure.  $\blacksquare$

**Theorem 4.26.** *If the  $s$ -dimensional random vector  $Z$  has an  $\alpha$ -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  $\alpha$ -concave probability distribution.*

**Proof.** Let  $A \subset \mathbb{R}^m$  and  $B \subset \mathbb{R}^m$  be two Borel sets. We define

$$A_1 = \{z \in \mathbb{R}^s : Tz \in A\} \quad \text{and} \quad B_1 = \{z \in \mathbb{R}^s : Tz \in B\}.$$

The sets  $A_1$  and  $A_2$  are Borel sets as well due to the continuity of the linear mapping  $z \mapsto Tz$ . Furthermore, for  $\lambda \in [0, 1]$  we have the relation

$$\lambda A_1 + (1 - \lambda)B_1 \subset \{z \in \mathbb{R}^s : Tz \in \lambda A + (1 - \lambda)B\}.$$

Denoting  $P_Z$  and  $P_Y$  the probability measure of  $Z$  and  $Y$  respectively, we obtain

$$\begin{aligned} P_Y\{\lambda A + (1 - \lambda)B\} &\geq P_Z\{\lambda A_1 + (1 - \lambda)B_1\} \\ &\geq m_\alpha(P_Z\{A_1\}, P_Z\{B_1\}, \lambda) \\ &= m_\alpha(P_Y\{A\}, P_Y\{B\}, \lambda). \end{aligned}$$

This completes the proof.  $\square$

**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  $\vartheta Y$  has the standard gamma distribution. It is not difficult to check that this density function is log-concave, provided  $\vartheta \geq 1$ .

A *multivariate gamma distribution* can be defined by a certain linear transformation of  $m$  independent random variables  $Z_1, \dots, Z_m$  ( $1 \leq m \leq 2^s - 1$ ) that have the standard gamma distribution. Let an  $s \times m$  matrix  $A$  with 0–1 elements be given. Setting  $Z = (Z_1, \dots, Z_{2^s-1})$ , 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.  $\blacksquare$

**Example 4.28.** The *Wishart distribution* arises in estimation of covariance matrices and can be considered as a multidimensional version of the  $\chi^2$ -distribution. More precisely, let us assume that  $Z$  is an  $s$ -dimensional random vector having multivariate normal distribution with a nonsingular covariance matrix  $\Sigma$  and expectation  $\mu$ . Given an iid sample  $Z^1, \dots, Z^N$  from this distribution, we consider the matrix

$$\sum_{i=1}^N (Z^i - \bar{Z})(Z^i - \bar{Z})^\top,$$

#### 4.2. Convexity in Probabilistic Optimization

103

where  $\bar{Z}$  is the sample mean. This matrix has Wishart distribution with  $N - 1$  degrees of freedom. We denote the trace of a matrix  $A$  by  $\text{tr}(A)$ .

If  $N > s$ , the Wishart distribution is a continuous distribution on the space of symmetric square matrices with probability density function defined by

$$f(A) = \begin{cases} \frac{\det(A)^{\frac{N-s-2}{2}} \exp(-\frac{1}{2}\text{tr}(\Sigma^{-1}A))}{2^{\frac{N-1}{2}s} \pi^{\frac{s(s-1)}{4}} \det(\Sigma)^{\frac{N-1}{2}} \prod_{i=1}^s \Gamma(\frac{N-i}{2})} & \text{for } A \text{ positive definite,} \\ 0 & \text{otherwise.} \end{cases}$$

If  $s = 1$  and  $\Sigma = 1$ , this density becomes the  $\chi^2$ -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  $N \geq s + 2$ , 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  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is called *regular in the sense of Clarke* or *Clarke-regular*, at a point  $x$ , if the directional derivative  $f'(x; d)$  exists and

$$f'(x; d) = \lim_{y \rightarrow x, t \downarrow 0} \frac{f(y + td) - f(y)}{t}, \quad \forall d \in \mathbb{R}^n.$$

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 Clarke generalized gradient of  $f$  at the point  $x$ . For convex functions  $\partial^\circ f(x) = \partial f(x)$ .

**Theorem 4.29.** *If  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is  $\alpha$ -concave ( $\alpha \in \mathbb{R}$ ) on some open set  $U \subset \mathbb{R}^n$  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  $\alpha$ -concave function, then an appropriate transformation of  $f$  is a concave function on  $U$ . We define

$$\tilde{f}(x) = \begin{cases} (f(x))^\alpha & \text{if } \alpha \neq 0, \\ \ln f(x) & \text{if } \alpha = 0. \end{cases}$$

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,  $\tilde{f}(x)$  is subdifferentiable at any  $x \in U$ . Further, we represent  $f$  as follows:

$$f(x) = \begin{cases} (\tilde{f}(x))^{1/\alpha} & \text{if } \alpha \neq 0, \\ \exp(\tilde{f}(x)) & \text{if } \alpha = 0. \end{cases}$$

In this representation,  $f$  is a composition of a continuously differentiable function and a concave function. By virtue of Clarke [38, 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

$$\partial^\circ f(x) = \begin{cases} \frac{1}{\alpha} (\bar{f}(x))^{1/\alpha-1} \partial \bar{f}(x) & \text{if } \alpha \neq 0, \\ \exp(\bar{f}(x)) \partial \bar{f}(x) & \text{if } \alpha = 0. \end{cases}$$

Substituting the definition of  $\bar{f}$  yields the result.  $\square$

For a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , we consider the set of points at which it takes positive values. It is denoted by  $\text{dompos } f$ , i.e.,

$$\text{dompos } f = \{x \in \mathbb{R}^n : f(x) > 0\}.$$

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  $\alpha$ -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  $X \subset \text{dompos } f$ , we call  $\hat{x} \in X$  a stationary point of  $f$  on  $X$  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  $\alpha$ -concave function  $f$  and the set  $X \subset \text{dompos } f$  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

$$0 \in f(\hat{x}) \partial(\ln f(\hat{x})) + \mathcal{N}_X(\hat{x}). \quad (4.16)$$

Using that  $f(\hat{x}) > 0$ , we obtain

$$0 \in \partial(\ln f(\hat{x})) + \mathcal{N}_X(\hat{x}). \quad (4.17)$$

As the function  $\bar{f}(x) = \ln f(x)$  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  $\bar{f}(x) = [f(x)]^\alpha$ .  $\square$

Another important property of  $\alpha$ -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  $z \in \mathbb{R}^s$ ,

$$P\{x \in \mathbb{R}^s : z^\top x \geq s_{C_p}(z)\} = 1 - p,$$

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 [128] without proof.

**Theorem 4.33.** Any nondegenerate 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  $\alpha$ -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  $\alpha$ -concave on the set  $\mathcal{A} \subset \mathbb{R}^s$  with  $\alpha \in [-\infty, \infty]$  if

$$F(z) \geq m_\alpha(F(x), F(y), \lambda)$$

for all  $z, x, y \in \mathcal{A}$ , and  $\lambda \in (0, 1)$  such that  $z \geq \lambda x + (1 - \lambda)y$ .

Observe that if  $\mathcal{A} = \mathbb{R}^s$ , then this definition coincides with the usual definition of  $\alpha$ -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 4.** If the distribution function of a random vector  $Z$  is  $\alpha$ -concave on  $\mathbb{R}^s$  then the distribution function of  $Y = \lceil Z \rceil$  is  $\alpha$ -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  $\alpha$ -concave on  $\mathbb{Z}^s$  for all  $\alpha \in [-\infty, \infty]$ .

Indeed, let  $x$  and  $y$  be binary vectors,  $\lambda \in (0, 1)$ , and  $z \geq \lambda x + (1 - \lambda)y$ . As  $z$  is integer and  $x$  and  $y$  binary, then  $z \geq x$  and  $z \geq y$ . Hence,  $F(z) \geq \max\{F(x), F(y)\}$  by the monotonicity of the cumulative distribution function. Consequently,  $F$  is  $\infty$ -concave. Using Lemma 4.8 we conclude that  $F_Z$  is  $\alpha$ -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  $\mathcal{A}$  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 subvectors  $Z^l$ ,  $l = 1, \dots, L$ , are  $s_l$ -dimensional and  $\sum_{l=1}^L s_l = s$ . Assume that  $Z^l$ ,  $l = 1, \dots, L$ , are independent and that their marginal distribution functions  $F_{Z^l} : \mathbb{R}^{s_l} \rightarrow [0, 1]$  are  $\alpha_l$ -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  $\alpha$ -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  $z \geq \lambda x + (1 - \lambda)y$ . 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:

$$\begin{aligned} \ln[F_Z(z)] &\geq \ln[F_Z(\lambda x + (1 - \lambda)y)] = \sum_{l=1}^L \ln[F_{Z^l}(\lambda x^l + (1 - \lambda)y^l)] \\ &\geq \sum_{l=1}^L [\lambda \ln[F_{Z^l}(x^l)] + (1 - \lambda) \ln[F_{Z^l}(y^l)]] \\ &\geq \lambda \sum_{l=1}^L \ln[F_{Z^l}(x^l)] + (1 - \lambda) \sum_{l=1}^L \ln[F_{Z^l}(y^l)] \\ &= \lambda[F_Z(x)] + (1 - \lambda)[F_Z(y)]. \end{aligned}$$

This concludes the proof.  $\square$

For integer random variables our definition of  $\alpha$ -concavity is related to log-concavity of sequences.

**Definition 4.37.** A sequence  $p_k$ ,  $k \in \mathbb{Z}$ , is called log-concave if

$$p_k^2 \geq p_{k-1}p_{k+1}, \quad \forall k \in \mathbb{Z}.$$

We have the following property. (See Prékopa [159, Theorem 4.7.2].)

**Theorem 4.38.** Suppose that for an integer random variable  $Y$  the probabilities  $p_k = \Pr\{Y = k\}$ ,  $k \in \mathbb{Z}$ , form a log-concave sequence. Then the distribution function of  $Y$  is  $\alpha$ -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  $\alpha$ -concave probability distribution, then the function*

$$G(x) = P\{g_j(x, Z) \geq 0, j \in \mathcal{J}\} \quad (4.18)$$

*is  $\alpha$ -concave on the set*

$$D = \{x \in \mathbb{R}^n : \exists z \in \mathbb{R}^s \text{ such that } g_j(x, z) \geq 0, j \in \mathcal{J}\}.$$

**Proof.** Given the points  $x_1, x_2 \in D$  and  $\lambda \in (0, 1)$ , we define the sets

$$A_i = \{z \in \mathbb{R}^s : g_j(x_i, z) \geq 0, j \in \mathcal{J}\}, \quad i = 1, 2,$$

and  $B = \lambda A_1 + (1 - \lambda)A_2$ . We consider

$$G(\lambda x_1 + (1 - \lambda)x_2) = P\{g_j(\lambda x_1 + (1 - \lambda)x_2, Z) \geq 0, j \in \mathcal{J}\}.$$

If  $z \in B$ , then there exist points  $z_i \in A_i$  such that  $z = \lambda z_1 + (1 - \lambda)z_2$ . By virtue of the quasi concavity of  $g_j$  we obtain that

$$g_j(\lambda x_1 + (1 - \lambda)x_2, \lambda z_1 + (1 - \lambda)z_2) \geq \min\{g_j(x_1, z_1), g_j(x_2, z_2)\} \geq 0, \quad \forall j \in \mathcal{J}.$$

This implies that  $z \in \{z \in \mathbb{R}^s : g_j(\lambda x_1 + (1 - \lambda)x_2, z) \geq 0, j \in \mathcal{J}\}$ , which entails that  $\lambda x_1 + (1 - \lambda)x_2 \in D$  and that

$$G(\lambda x_1 + (1 - \lambda)x_2) \geq P\{B\}.$$

Using the  $\alpha$ -concavity of the measure, we conclude that

$$G(\lambda x_1 + (1 - \lambda)x_2) \geq P\{B\} \geq m_\alpha\{P\{A_1\}, P\{A_2\}, \lambda\} = m_\alpha\{G(x_1), G(x_2), \lambda\},$$

as desired.  $\square$

**Example 4.40 (The Log-Normal Distribution).** The probability density function of the one-dimensional log-normal distribution with parameters  $\mu$  and  $\sigma$  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 demonstrate it for the multidimensional case.

The  $m$ -dimensional random vector  $Z$  has the log-normal distribution if the vector  $Y = (\ln Z_1, \dots, \ln Z_m)^\top$  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

$$F_Z(z) = \Pr\{Z_1 \leq z_1, \dots, Z_m \leq z_m\} = \Pr\{z_1 - e_1^Y \geq 0, \dots, z_m - e_m^Y \geq 0\}.$$

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.  $\blacksquare$

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  $\alpha$ -concave probability distribution. Then the following set is convex and closed:*

$$X_0 = \{x \in \mathbb{R}^n : \Pr\{g_i(x, Z) \geq 0, i = 1, \dots, m\} \geq p\}. \quad (4.19)$$

**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

$$G(\lambda x_1 + (1 - \lambda)x_2) \geq m_\alpha\{G(x_1), G(x_2), \lambda\} \geq \min\{G(x_1), G(x_2)\} \geq p.$$

The closedness of the set follows from the continuity of  $\alpha$ -concave functions.  $\square$

We consider the case of a separable mapping  $g$  when the random quantities appear only on the right-hand side of the inequalities.

**Theorem 4.42.** *Let the mapping  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be such that each component  $g_i$  is a concave function. Furthermore, assume that the random vector  $Z$  has independent components and the one-dimensional marginal distribution functions  $F_{Z_i}$ ,  $i = 1, \dots, m$ , are  $\alpha_i$ -concave. Furthermore, let  $\sum_{i=1}^m \alpha_i^{-1} > 0$ . Then the set*

$$X_0 = \{x \in \mathbb{R}^n : \Pr\{g(x) \geq Z\} \geq p\}$$

*is convex.*

**Proof.** Indeed, the probability function appearing in the definition of the set  $X_0$  can be described as follows:

$$G(x) = P\{g_i(x) \geq Z_i, i = 1, \dots, m\} = \prod_{i=1}^m F_{Z_i}(g_i(x_i)).$$

Due to Theorem 4.20, the functions  $F_{Z_i} \circ g_i$  are  $\alpha_i$ -concave. Using Theorem 4.23, we conclude that  $G(\cdot)$  is  $\gamma$ -concave with  $\gamma = (\sum_{i=1}^m \alpha_i^{-1})^{-1}$ . The convexity of  $X_0$  follows the same argument as in Corollary 4.41.  $\square$

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 4.43.** *Assume that  $g(\cdot, \cdot)$  is a quasi-concave function jointly in both arguments, and that  $Z$  has an  $\alpha$ -concave distribution. Then the following sets are convex and closed:*

$$X_d = \{x \in \mathbb{R}^n : g(x, Z) \succeq_{(1)} Y\},$$

$$X_c = \{x \in \mathbb{R}^n : \Pr\{g(x, Z) \geq \eta\} \geq \Pr\{Y \geq \eta\}, \quad \forall \eta \in [a, b]\}.$$

**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:

$$\Pr\{g(x, Z) \geq \eta\} \geq \Pr\{Y \geq \eta\}, \quad \forall \eta \in \mathbb{R}.$$



## 4.2. Convexity in Probabilistic Optimization

109

Therefore, the first set can be defined as follows:

$$X_d = \{x \in \mathbb{R}^n : \Pr\{g(x, Z) - \eta \geq 0\} \geq \Pr\{Y \geq \eta\} \forall \eta \in \mathbb{R}\}.$$

For any  $\eta \in \mathbb{R}$ , we define the set

$$X(\eta) = \{x \in \mathbb{R}^n : \Pr\{g(x, Z) - \eta \geq 0\} \geq \Pr\{Y \geq \eta\}\}.$$

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  $X_c = \bigcap_{\eta \in [a, b]} X(\eta)$ .  $\square$

Let us observe that affine in each argument functions  $g_i(x, z) = z^\top x + b_i$  are not necessarily quasi-concave in both arguments  $(x, z)$ . We can apply Theorem 4.39 to conclude that the set

$$X_l = \{x \in \mathbb{R}^n : \Pr\{x^\top a_i \leq b_i(Z), i = 1, \dots, m\} \geq p\} \quad (4.20)$$

is convex if  $a_i, i = 1, \dots, m$  are deterministic vectors. We have the following.

**Corollary 4.44.** *The set  $X_l$  is convex whenever  $b_i(\cdot)$  are quasi-concave functions and  $Z$  has a quasi-concave probability distribution function.*

**Example 4.45 (Vehicle Routing Continued).** We return to Example 4.1. The probabilistic constraint (4.3) has the form

$$\Pr\{TX \geq Z\} \geq p_\eta.$$

If the vector  $Z$  of a random demand has an  $\alpha$ -concave distribution, 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.  $\blacksquare$

If the functions  $g_i$  are not separable, we can invoke Theorem 4.33.

**Theorem 4.46.** *Let  $p_i \in (\frac{1}{2}, 1)$  for all  $i = 1, \dots, n$ . The set*

$$X_p = \{x \in \mathbb{R}^n : P_{Z_i}\{x^\top Z_i \leq b_i\} \geq p_i, i = 1, \dots, m\} \quad (4.21)$$

*is convex whenever  $Z_i$  has a nondegenerate log-concave probability distribution, which is symmetric around some point  $\mu_i \in \mathbb{R}^n$ .*

**Proof.** If the random vector  $Z_i$  has a nondegenerate 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 nondegenerate log-concave distribution.

Given points  $x_1, x_2 \in X_p$  and a number  $\lambda \in [0, 1]$ , we define

$$K_i(x) = \{a \in \mathbb{R}^n : a^\top x \leq b_i\}, \quad i = 1, \dots, n.$$

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  $p_i$ :

$$P_{Y_i}\{y \in \mathbb{R}^n : y^\top x \leq s_{C_{p_i}}(x)\} = p_i \quad \forall x \in \mathbb{R}^n.$$

Thus,

$$P_{Z_i} \{z \in \mathbb{R}^n : z^\top x \leq s_{c_{p_i}}(x) + \mu_i^\top x\} = p_i \quad \forall x \in \mathbb{R}^n. \quad (4.22)$$

Since  $P_{Z_i} \{K_i(x_1)\} \geq p_i$  and  $P_{Z_i} \{K_i(x_2)\} \geq p_i$  by assumption, then

$$\begin{aligned} K_i(x_j) &\subset \{z \in \mathbb{R}^n : z^\top x \leq s_{c_{p_i}}(x) + \mu_i^\top x\}, \quad j = 1, 2, \\ b_i &\geq s_{c_{p_i}}(x_1) + \mu_i^\top x_j, \quad j = 1, 2. \end{aligned}$$

The properties of the support function entail that

$$\begin{aligned} b_i &\geq \lambda[s_{c_{p_i}}(x_1) + \mu_i^\top x_1] + (1 - \lambda)[s_{c_{p_i}}(x_2) + \mu_i^\top x_2] \\ &= s_{c_{p_i}}(\lambda x_1) + s_{c_{p_i}}((1 - \lambda)x_2) + \mu_i^\top \lambda x_1 + (1 - \lambda)x_2 \\ &\geq s_{c_{p_i}}(\lambda x_1 + (1 - \lambda)x_2) + \mu_i^\top \lambda x_1 + (1 - \lambda)x_2. \end{aligned}$$

Consequently, the set  $K_i(x_\lambda)$  with  $x_\lambda = \lambda x_1 + (1 - \lambda)x_2$  contains the set

$$\{z \in \mathbb{R}^n : z^\top x_\lambda \leq s_{c_{p_i}}(x_\lambda) + \mu_i^\top x_\lambda\},$$

and, therefore, using (4.22) we obtain that

$$P_{Z_i} \{K_i(\lambda x_1 + (1 - \lambda)x_2)\} \geq p_i.$$

Since  $i$  was arbitrary, we obtain that  $\lambda x_1 + (1 - \lambda)x_2 \in X_p$ .  $\square$

**Example 4.47 (Portfolio Optimization Continued).** Let us consider the Portfolio Example 4.2. The probabilistic constraint has the form

$$\Pr \left\{ \sum_{i=1}^n R_i x_i \leq \eta \right\} \leq p_\eta.$$

If the random vector  $R = (R_1, \dots, R_n)^\top$  has a multidimensional normal distribution or a uniform distribution, then the feasible set in this example is convex by virtue of the last corollary since both distributions are symmetric and log-concave.  $\blacksquare$

There is an important relation between the sets constrained by first and second order stochastic dominance relation to a benchmark random variable (see Dentcheva and Ruszczyński [53]). We denote the space of integrable random variables by  $\mathcal{L}_1(\Omega, \mathcal{F}, P)$  and set

$$\begin{aligned} A_1(Y) &= \{X \in \mathcal{L}_1(\Omega, \mathcal{F}, P) : X \succeq_{(1)} Y\}, \\ A_2(Y) &= \{X \in \mathcal{L}_1(\Omega, \mathcal{F}, P) : X \succeq_{(2)} Y\}. \end{aligned}$$

**Proposition 4.48.** 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

$$F_X^{(2)}(\eta) = \mathbb{E}[(\eta - X)_+]. \quad (4.23)$$

## 4.2. Convexity in Probabilistic Optimization

111

Therefore, an equivalent representation of the second order stochastic dominance relation is given by the relation

$$\mathbb{E}[(\eta - X)_+] \leq \mathbb{E}[(\eta - Y)_+], \quad \forall \eta \in \mathbb{R}. \quad (4.24)$$

For every  $\eta \in \mathbb{R}$  the functional  $X \rightarrow \mathbb{E}[(\eta - X)_+]$  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.  $\square$

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 4.49.** 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.  $\blacksquare$

Directly from the definition we see that first order dominance relation implies the second order dominance. Hence,  $A_1(Y) \subset A_2(Y)$ . We have demonstrated that the set  $A_2(Y)$  is convex; therefore, we also have

$$\text{conv}(A_1(Y)) \subset A_2(Y). \quad (4.25)$$

We find sufficient conditions for the opposite inclusion.

**Theorem 4.50.** Assume that  $\Omega = \{\omega_1, \dots, \omega_N\}$ ,  $\mathcal{F}$  contains all subsets of  $\Omega$ , and  $P\{\omega_k\} = 1/N$ ,  $k = 1, \dots, N$ . If  $Y : (\Omega, \mathcal{F}, P) \rightarrow \mathbb{R}$  is a random variable, then

$$\text{conv}(A_1(Y)) = A_2(Y).$$

**Proof.** To prove the inverse inclusion to (4.25), suppose that  $X \in A_2(Y)$ . Under the assumptions of the theorem, we can identify  $X$  and  $Y$  with vectors  $x = (x_1, \dots, x_N)$  and  $y = (y_1, \dots, y_N)$  such that  $x_i = X(i)$  and  $y_i = Y(i)$ ,  $i = 1, \dots, N$ . As the probabilities of all elementary events are equal, the second order stochastic dominance relation coincides with the concept of *weak majorization*, which is characterized by the following system of inequalities:

$$\sum_{k=1}^l x_{[k]} \geq \sum_{k=1}^l y_{[k]}, \quad l = 1, \dots, N,$$

where  $x_{[k]}$  denotes the  $k$ th smallest component of  $x$ .

As established by Hardy, Littlewood, and Polya [73], weak majorization is equivalent to the existence of a doubly stochastic matrix  $\Pi$  such that

$$x \geq \Pi y.$$

By Birkhoff’s theorem [20], we can find permutation matrices  $Q^1, \dots, Q^M$  and nonnegative reals  $\alpha_1, \dots, \alpha_M$  totaling 1, such that

$$\Pi = \sum_{j=1}^M \alpha_j Q^j.$$

Setting  $z^j = Q^j y$ , we conclude that

$$x \geq \sum_{j=1}^M \alpha_j z^j.$$

Identifying random variables  $Z^j$  on  $(\Omega, \mathcal{F}, P)$  with the vectors  $z^j$ , we also see that

$$X(\omega) \geq \sum_{j=1}^M \alpha_j Z^j(\omega)$$

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 with 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  $\hat{Z}^j \in A_1(Y)$ ,  $j = 1, \dots, M$ , and

$$X(\omega) = \sum_{j=1}^M \alpha_j \hat{Z}^j(\omega),$$

as required.  $\square$

This result does not extend to general probability spaces, as the following example illustrates.

**Example 4.51.** 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  $X \succeq_{(1)} Y$  iff  $X(\omega_1) \geq -1$  and  $X(\omega_2) \geq 1$ . 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  $\text{conv } A_1(Y) = A_1(Y)$ .  $\blacksquare$

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  $\omega_2$  is split in two equally likely events  $\omega_{21}, \omega_{22}$ , then we could use Theorem 4.50 to obtain the equality  $\text{conv } A_1(Y) = A_2(Y)$ . 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 4.52.** Assume that the probability space  $(\Omega, \mathcal{F}, P)$  is nonatomic. Then

$$A_2(Y) = \text{cl}\{\text{conv}(A_1(Y))\}.$$

**Proof.** If the space  $(\Omega, \mathcal{F}, P)$  is nonatomic, we can partition  $\Omega$  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 considered in Theorem 4.50. Passing to the limit with  $N \rightarrow \infty$ , we obtain the desired result. We refer the interested reader to Dentcheva and Ruszczyński [55] for technical details of the proof.  $\square$

### 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:

$$X = \{x \in \mathcal{X} : \Pr[g_j(x, Z) \geq 0, j \in \mathcal{J}] \geq p\},$$

where  $\mathcal{J}$  is an arbitrary index set. The functions  $g_j : \mathbb{R}^n \times \mathbb{R}^s \rightarrow \mathbb{R}$  are continuous,  $Z$  is an  $s$ -dimensional random vector, and  $p \in (0, 1)$  is a prescribed probability. It will be demonstrated later (Lemma 4.61) that the probabilistically constrained set  $X$  with separable functions  $g_j$  is a union of cones intersected by  $\mathcal{X}$ . Thus,  $X$  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 [84].

**Theorem 4.53.** *Assume that the functions  $g_j(\cdot, Z)$ ,  $j \in \mathcal{J}$  are quasi-concave and that they satisfy the following condition: for all  $x^1, x^2 \in \mathbb{R}^n$  there exists a point  $x^* \in \mathcal{X}$  such that*

$$g_j(x^*, z) \geq \min\{g_j(x^1, z), g_j(x^2, z)\}, \quad \forall z \in \mathbb{R}^s, \forall j \in \mathcal{J}.$$

*Then the set  $X$  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

$$\pi(t) = \begin{cases} (1-2t)x^1 + 2tx^* & \text{for } 0 \leq t \leq 1/2, \\ 2(1-t)x^* + (2t-1)x^2 & \text{for } 1/2 < t \leq 1. \end{cases}$$

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  $0 \leq t \leq 1/2$  the following inequality:

$$g_j((1-2t)x^1 + 2tx^*, z) \geq \min\{g_j(x^1, z), g_j(x^*, z)\} = g_j(x^1, z).$$

Therefore,

$$\Pr\{g_j(\pi(t), Z) \geq 0, j \in \mathcal{J}\} \geq \Pr\{g_j(x^1) \geq 0, j \in \mathcal{J}\} \geq p \quad \text{for } 0 \leq t \leq 1/2.$$

A similar argument applies for  $1/2 < t \leq 1$ . Consequently,  $\pi(t) \in X$ , and this proves the assertion.  $\square$

### 4.3 Separable Probabilistic Constraints

We focus our attention on problems with separable probabilistic constraints. The problem that we analyze in this section is

$$\begin{aligned} & \underset{x}{\text{Min}} \ c(x) \\ & \text{s.t. } \Pr\{g(x) \geq Z\} \geq p, \\ & \quad x \in \mathcal{X}. \end{aligned} \tag{4.26}$$

We assume that  $c : \mathbb{R}^n \rightarrow \mathbb{R}$  is a convex function and  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is such that each component  $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$  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 (4.26), we can express it as a constraint on a distribution function:

$$\Pr\{g(x) \geq Z\} \geq p \iff F_Z(g(x)) \geq p.$$

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  $\alpha$ -concave distribution function with  $\alpha \in \mathbb{R}$  and that the support of it,  $\text{supp } P_Z$ , has nonempty interior in  $\mathbb{R}^s$ . Then  $F_Z(\cdot)$  is locally Lipschitz continuous on  $\text{int supp } P_Z$  by virtue of Theorem 4.29.

**Example 4.54.** 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

$$F(z) = \begin{cases} 0 & \text{for } z \leq 0, \\ \sqrt{z} & \text{for } z \in (0, 1), \\ 1 & \text{for } z \geq 1. \end{cases}$$

The density  $\theta$  is unbounded. We observe that  $F$  is continuous but it is not Lipschitz continuous at  $z = 0$ . The density  $\theta$  is also not  $(-1)$ -concave and that means that the corresponding probability distribution is not quasi-concave. ■

**Theorem 4.55.** 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.

### 4.3. Separable Probabilistic Constraints

115

**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.  $\square$

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 4.56.** *Assume that  $P_Z$  has a continuous density  $\theta(\cdot)$  and that all one-dimensional marginal distribution functions are continuous as well. Then the distribution function  $F_Z$  is continuously differentiable.*

**Proof.** In order to simplify the notation, we demonstrate the statement for  $s = 2$ . It will be clear how to extend the proof for  $s > 2$ . We have that

$$F_Z(z_1, z_2) = \Pr(Z_1 \leq z_1, Z_2 \leq z_2) = \int_{-\infty}^{z_1} \int_{-\infty}^{z_2} \theta(t_1, t_2) dt_2 dt_1 = \int_{-\infty}^{z_1} \psi(t_1, z_2) dt_1,$$

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 the Newton–Leibnitz theorem we have that

$$\frac{\partial F_Z}{\partial z_1}(z_1, z_2) = \psi(z_1, z_2) = \int_{-\infty}^{z_2} \theta(z_1, t_2) dt_2.$$

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  $\lim_{k \rightarrow \infty} y^k = z$ , we have

$$\begin{aligned} \left| \frac{\partial F_Z}{\partial z_1}(z) - \frac{\partial F_Z}{\partial z_1}(y^k) \right| &= \left| \int_{-\infty}^{z_2} \theta(z_1, t) dt - \int_{-\infty}^{y_2^k} \theta(y_1^k, t) dt \right| \\ &\leq \left| \int_{z_2}^{y_2^k} \theta(y_1^k, t) dt \right| + \left| \int_{-\infty}^{z_2} [\theta(z_1, t) - \theta(y_1^k, t)] dt \right|. \end{aligned}$$

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,  $|\int_{z_2}^{y_2^k} \theta(y_1^k, t) dt| \rightarrow 0$  whenever  $k \rightarrow \infty$ . Furthermore,  $|\int_{-\infty}^{z_2} [\theta(z_1, t) - \theta(y_1^k, t)] dt| \rightarrow 0$  as well, due to the continuity of the one-dimensional marginal function  $F_{Z_1}$ . Moreover, by the same reason, the convergence is uniform about  $z_1$ . 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.  $\square$

#### 4.3.2 $p$ -Efficient Points

We concentrate on deriving an equivalent algebraic description for the feasible set of problem (4.26).

The  $p$ -level set of the distribution function  $F_Z(z) = \Pr\{Z \leq z\}$  of  $Z$  is defined as follows:

$$\mathcal{Z}_p = \{z \in \mathbb{R}^m : F_Z(z) \geq p\}. \quad (4.27)$$

Clearly, problem (4.26) can be compactly rewritten as

$$\begin{aligned} & \underset{x}{\text{Min}} \ c(x) \\ & \text{s.t. } g(x) \in \mathcal{Z}_p, \\ & \quad x \in \mathcal{X}. \end{aligned} \quad (4.28)$$

**Lemma 4.57.** *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.  $\square$

We introduce the key concept of a  $p$ -efficient point.

**Definition 4.58.** *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  $F(v) \geq p$  and there is no  $z \leq v$ ,  $z \neq v$  such that  $F(z) \geq p$ .*

The  $p$ -efficient points are minimal points of the level set  $\mathcal{Z}_p$  with respect to the partial order in  $\mathbb{R}^m$  generated by the nonnegative cone  $\mathbb{R}_+^m$ .

Clearly, for a scalar random variable  $Z$  and for every  $p \in (0, 1)$  there is exactly one  $p$ -efficient point, which is the smallest  $v$  such that  $F_Z(v) \geq p$ , i.e.,  $F_Z^{(-1)}(p)$ .

**Lemma 4.59.** *Let  $p \in (0, 1)$  and let*

$$l = (F_{Z_1}^{(-1)}(p), \dots, F_{Z_m}^{(-1)}(p)). \quad (4.29)$$

*Then every  $v \in \mathbb{R}^m$  such that  $F_Z(v) \geq p$  must satisfy the inequality  $v \geq l$ .*

**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  $i = 1, \dots, m$ , and, therefore, we obtain that the set of  $p$ -efficient points is bounded from below.  $\square$

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 [150, Lemma 3.12] about the existence of conical support points, but we can easily prove it directly.

**Theorem 4.60.** *It holds that  $\mathcal{Z}_p = \bigcup_{j \in \mathcal{E}} K_j$ .*



### 4.3. Separable Probabilistic Constraints

117

**Proof.** If  $y \in \mathcal{Z}_p$ , then either  $y$  is  $p$ -efficient or there exists a vector  $w$  such that  $w \leq y$ ,  $w \neq y$ ,  $w \in \mathcal{Z}_p$ . By Lemma 4.59, one must have  $l \leq w \leq y$ . The set  $Z_1 = \{z \in \mathcal{Z}_p : l \leq z \leq y\}$  is compact because the set  $\mathcal{Z}_p$  is closed by virtue of Lemma 4.57. Thus, there exists  $w^1 \in Z_1$  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  $Z_2 = \{z \in \mathcal{Z}_p : l \leq z \leq w^1\}$  and choose a point  $w^2 \in Z_2$  with the minimal second coordinate. Proceeding in the same way, we shall find the minimal element  $w^m$  in the set  $\mathcal{Z}_p$  with  $w^m \leq w^{m-1} \leq \dots \leq y$ . Therefore,  $y \in w^m + \mathbb{R}_+^m$ , and this completes the proof.  $\square$

By virtue of Theorem 4.60 we obtain (for  $0 < p < 1$ ) the following *disjunctive semi-infinite* formulation of problem (4.28):

$$\begin{aligned} & \text{Min}_x c(x) \\ & \text{s.t. } g(x) \in \bigcup_{j \in \mathcal{E}} K_j, \\ & x \in \mathcal{X}. \end{aligned} \tag{4.30}$$

This formulation provides insight into the structure of the feasible set and the nature of its nonconvexity. The main difficulty here is the implicit character of the disjunctive constraint.

Let  $S$  stand for the simplex in  $\mathbb{R}^{m+1}$ ,

$$S = \left\{ \alpha \in \mathbb{R}^{m+1} : \sum_{i=1}^{m+1} \alpha_i = 1, \alpha_i \geq 0 \right\}.$$

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  $\mathcal{Z}_p$ .

**Lemma 4.61.** *It holds that*

$$\text{conv}(\mathcal{Z}_p) = E + \mathbb{R}_+^m.$$

**Proof.** By Theorem 4.60, every point  $y \in \text{conv} \mathcal{Z}$  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,  $y \in \sum_{i=1}^{m+1} \alpha_i v^{j_i} + \mathbb{R}_+^m$ .  $\square$

We also have the representation  $E = \left\{ \sum_{i=1}^{m+1} \alpha_i v^{j_i} : \alpha \in S, j_i \in \mathcal{E} \right\}$ .

**Theorem 4.62.** *For every  $p \in (0, 1)$  the set  $\text{conv} \mathcal{Z}_p$  is closed.*

**Proof.** Consider a sequence  $\{z^k\}$  of points of  $\text{conv} \mathcal{Z}_p$  which is convergent to a point  $\bar{z}$ . Using Carathéodory's theorem again, we have

$$z^k = \sum_{i=1}^{m+1} \alpha_i^k y_i^k$$

with  $y_i^k \in \mathcal{Z}_p$ ,  $\alpha_i^k \geq 0$ , and  $\sum_{i=1}^{m+1} \alpha_i^k = 1$ . By passing to a subsequence, if necessary, we can assume that the limits

$$\bar{\alpha}_i = \lim_{k \rightarrow \infty} \alpha_i^k$$

exist for all  $i = 1, \dots, m+1$ . By Lemma 4.59, 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

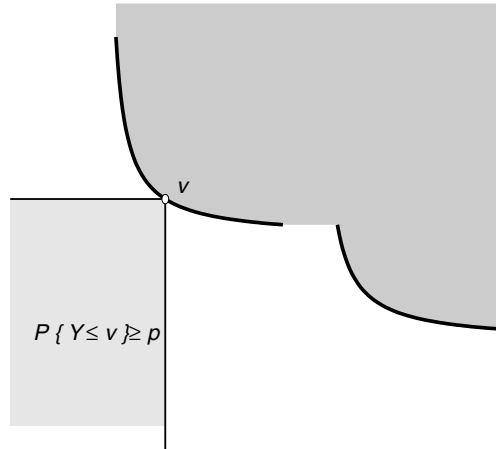
$$z^k \geq \sum_{i \in I} \alpha_i^k y_i^k. \quad (4.31)$$

We observe that  $0 \leq \alpha_i^k y_i^k \leq z^k$  for all  $i \in I$  and all  $k$ . Since  $\{z^k\}$  is convergent and  $\alpha_i^k \rightarrow \bar{\alpha}_i > 0$ , 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 4.57,  $\bar{y}_i \in \mathcal{Z}_p$ . Passing to the limit in inequality (4.31), we obtain

$$\bar{z} \geq \sum_{i \in I} \bar{\alpha}_i \bar{y}_i \in \text{conv} \mathcal{Z}.$$

Due to Lemma 4.61, we conclude that  $\bar{z} \in \text{conv} \mathcal{Z}_p$ .  $\square$

For a general random vector, the set of  $p$ -efficient points may be unbounded and not closed, as illustrated in Figure 4.3.



**Figure 4.3.** Example of a set  $\mathcal{Z}_p$  with  $p$ -efficient points  $v$ .

### 4.3. Separable Probabilistic Constraints

119

We encounter also a relation between the  $p$ -efficient points and the extreme points of the convex hull of  $\mathcal{Z}_p$ .

**Theorem 4.63.** *For every  $p \in (0, 1)$ , the set of extreme points of  $\text{conv } \mathcal{Z}_p$  is nonempty and it is contained in the set of  $p$ -efficient points.*

**Proof.** Consider the lower bound  $l$  defined in (4.29). The set  $\text{conv } \mathcal{Z}_p$  is included in  $l + \mathbb{R}_+^m$ , by virtue of Lemmas 4.59 and 4.61. Therefore, it does not contain any line. Since  $\text{conv } \mathcal{Z}_p$  is closed by Theorem 4.62, it has at least one extreme point.

Let  $w$  be an extreme point of  $\text{conv } \mathcal{Z}_p$ . Suppose that  $w$  is not a  $p$ -efficient point. Then Theorem 4.60 implies that there exists a  $p$ -efficient point  $v \leq w$ ,  $v \neq w$ . Since  $w + \mathbb{R}_+^m \subset \text{conv } \mathcal{Z}_p$ , the point  $w$  is a convex combination of  $v$  and  $w + (w - v)$ . Consequently,  $w$  cannot be extreme.  $\square$

The representation becomes very handy when the vector  $Z$  has a discrete distribution on  $\mathbb{Z}^m$ , in particular, if the problem is of form (4.57). 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 nonuniform 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 4.64.** *For each  $p \in (0, 1)$  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  $F_Z(y) \geq p$ . By Lemma 4.59, 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  $l \leq v \leq y$  must be  $p$ -efficient.

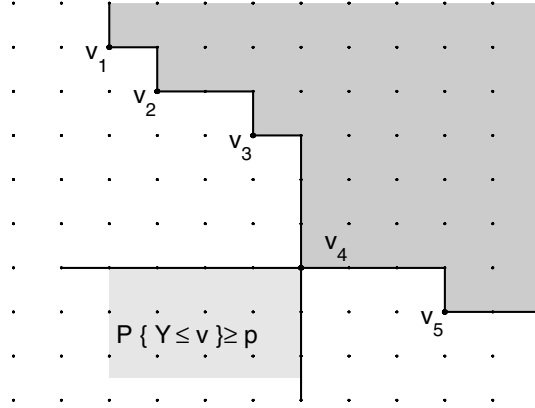
Now we prove the finiteness of the set of  $p$ -efficient points. Suppose that there exists an infinite sequence of different  $p$ -efficient points  $v^j$ ,  $j = 1, 2, \dots$ . Since they are integer, and the first coordinate  $v_1^j$  is bounded from below by  $l_1$ , with no loss of generality we may select a subsequence which is nondecreasing in the first coordinate. By a similar token, we can select further subsequences which are nondecreasing in the first  $k$  coordinates ( $k = 1, \dots, m$ ). Since the dimension  $m$  is finite, we obtain a subsequence of different  $p$ -efficient points which is nondecreasing in all coordinates. This contradicts the definition of a  $p$ -efficient point.  $\square$

Note the crucial role of Lemma 4.59 in this proof. In conclusion, we have obtained that the disjunctive formulation (4.30) of problem (4.28) has a *finite* index set  $\mathcal{E}$ .

Figure 4.4 illustrates the structure of the probabilistically constrained set for a discrete random variable.

The concept of  $\alpha$ -concavity on a set can be used at this moment to find an equivalent representation of the set  $\mathcal{Z}_p$  for a random vector with a discrete distribution.

**Theorem 4.65.** *Let  $\mathcal{A}$  be the set of all possible values of an integer random vector  $Z$ . If the distribution function  $F_Z$  of  $Z$  is  $\alpha$ -concave on  $\mathcal{A} + \mathbb{Z}_+^m$  for some  $\alpha \in [-\infty, \infty]$ , then*



**Figure 4.4.** Example of a discrete set  $\mathcal{Z}_p$  with  $p$ -efficient points  $v_1, \dots, v_5$ .

for every  $p \in (0, 1)$  one has

$$\mathcal{Z}_p = \left\{ y \in \mathbb{R}^m : y \geq z \geq \sum_{j \in \mathcal{E}} \lambda_j v^j, \sum_{j \in \mathcal{E}} \lambda_j = 1, \lambda_j \geq 0, z \in \mathbb{Z}^m \right\},$$

where  $v^j, j \in \mathcal{E}$ , are the  $p$ -efficient points of  $F$ .

**Proof.** The representation (4.30) implies that

$$\mathcal{Z}_p \subset \left\{ y \in \mathbb{R}^m : y \geq z \geq \sum_{j \in \mathcal{E}} \lambda_j v^j, \sum_{j \in \mathcal{E}} \lambda_j = 1, \lambda_j \geq 0, z \in \mathbb{Z}^m \right\}.$$

We have to show that every point  $y$  from the set at the right-hand side belongs to  $\mathcal{Z}$ . By the monotonicity of the distribution function  $F_Z$ , we have  $F_Z(y) \geq F_Z(z)$  whenever  $y \geq z$ . Therefore, it is sufficient to show that  $\Pr\{Z \leq z\} \geq p$  for all  $z \in \mathbb{Z}^m$  such that  $z \geq \sum_{j \in \mathcal{E}} \lambda_j v^j$  with  $\lambda_j \geq 0, \sum_{j \in \mathcal{E}} \lambda_j = 1$ . We consider five cases with respect to  $\alpha$ .

*Case 1:*  $\alpha = \infty$ . It follows from the definition of  $\alpha$ -concavity that

$$F_Z(z) \geq \max\{F_Z(v^j), j \in \mathcal{E} : \lambda_j \neq 0\} \geq p.$$

*Case 2:*  $\alpha = -\infty$ . Since  $F_Z(v^j) \geq p$  for each index  $j \in \mathcal{E}$  such that  $\lambda_j \neq 0$ , the assertion follows as in Case 1.

### 4.3. Separable Probabilistic Constraints

121

*Case 3:*  $\alpha = 0$ . By the definition of  $\alpha$ -concavity, we have the following inequalities:

$$F_Z(z) \geq \prod_{j \in \mathcal{E}} [F_Z(v^j)]^{\lambda_j} \geq \prod_{j \in \mathcal{E}} p^{\lambda_j} = p.$$

*Case 4:*  $\alpha \in (-\infty, 0)$ . By the definition of  $\alpha$ -concavity,

$$[F_Z(z)]^\alpha \leq \sum_{j \in \mathcal{E}} \lambda_j [F_Z(v^j)]^\alpha \leq \sum_{j \in \mathcal{E}} \lambda_j p^\alpha = p^\alpha.$$

Since  $\alpha < 0$ , we obtain  $F_Z(z) \geq p$ .

*Case 5:*  $\alpha \in (0, \infty)$ . By the definition of  $\alpha$ -concavity,

$$[F_Z(z)]^\alpha \geq \sum_{j \in \mathcal{E}} \lambda_j [F_Z(v^j)]^\alpha \geq \sum_{j \in \mathcal{E}} \lambda_j p^\alpha = p^\alpha,$$

concluding that  $z \in \mathcal{Z}$ , as desired.  $\square$

The consequence of this theorem is that under the  $\alpha$ -concavity assumption, all integer points contained in  $\text{conv} \mathcal{Z}_p = E + \mathbb{R}_+^m$  satisfy the probabilistic constraint. This demonstrates the importance of the notion of  $\alpha$ -concavity for discrete distribution functions as introduced in Definition 4.34. For example, the set  $\mathcal{Z}_p$  illustrated in Figure 4.4 does not correspond to any  $\alpha$ -concave distribution function, because its convex hull contains integer points which do not belong to  $\mathcal{Z}_p$ . These are the points (3,6), (4,5), and (6,2).

Under the conditions of Theorem 4.65, problem (4.28) can be formulated in the following equivalent way:

$$\text{Min}_{x, z, \lambda} c(x) \tag{4.32}$$

$$\text{s.t. } g(x) \geq z, \tag{4.33}$$

$$z \geq \sum_{j \in \mathcal{E}} \lambda_j v^j, \tag{4.34}$$

$$z \in \mathbb{Z}^m, \tag{4.35}$$

$$\sum_{j \in \mathcal{E}} \lambda_j = 1, \tag{4.36}$$

$$\lambda_j \geq 0, \quad j \in \mathcal{E}, \tag{4.37}$$

$$x \in \mathcal{X}. \tag{4.38}$$

In this way, we have replaced the probabilistic constraint by algebraic equations and inequalities, together with the integrality requirement (4.35). This condition cannot be dropped, in general. However, if other conditions of the problem imply that  $g(x)$  is integer, then we may remove  $z$  entirely from the problem formulation. In this case, we replace constraints (4.33)–(4.35) with

$$g(x) \geq \sum_{j \in \mathcal{E}} \lambda_j v^j.$$

For example, if the definition of  $\mathcal{X}$  contains the constraint  $x \in \mathbb{Z}^n$ , and, in addition,  $g(x) = Tx$ , where  $T$  is a matrix with integer elements, then we can dispose of the variable  $z$ .

If  $Z$  takes values on a nonuniform grid, condition (4.35) should be replaced by the requirement that  $z$  is a grid point.

**Corollary 4.66.** *If the distribution function  $F_Z$  of an integer random vector  $Z$  is  $\alpha$ -concave on the set  $\mathbb{Z}_+^m$  for some  $\alpha \in [-\infty, \infty]$ , then for every  $p \in (0, 1)$  one has*

$$\mathcal{Z}_p \cap \mathbb{Z}_+^m = \text{conv} \mathcal{Z}_p \cap \mathbb{Z}_+^m.$$

### 4.3.3 Optimality Conditions and Duality Theory

In this section, we return to problem formulation (4.28). We assume that  $c : \mathbb{R}^n \rightarrow \mathbb{R}$  is a convex function. The mapping  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  has concave components  $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ . The set  $\mathcal{X} \subset \mathbb{R}^n$  is closed and convex; the random vector  $Z$  takes values in  $\mathbb{R}^m$ . The set  $\mathcal{Z}_p$  is defined as in (4.27). We split variables and consider the following formulation of the problem:

$$\begin{aligned} & \text{Min}_{x,z} c(x) \\ & \text{s.t. } g(x) \geq z, \\ & \quad x \in \mathcal{X}, \\ & \quad z \in \mathcal{Z}_p. \end{aligned} \tag{4.39}$$

Associating a Lagrange multiplier  $u \in \mathbb{R}_+^m$  with the constraint  $g(x) \geq z$ , we obtain the Lagrangian function:

$$L(x, z, u) = c(x) + u^\top (z - g(x)).$$

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

$$h(u) = \inf \{c(x) - u^\top g(x) : x \in \mathcal{X}\}, \tag{4.40}$$

$$d(u) = \inf \{u^\top z : z \in \mathcal{Z}_p\}. \tag{4.41}$$

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  $\Psi^*$  of the problem:

$$\sup_{u \geq 0} \Psi(u). \tag{4.42}$$

We call (4.42) the dual problem to problem (4.39). For  $u \not\geq 0$  one has  $d(u) = -\infty$ , because the set  $\mathcal{Z}_p$  contains a translation of  $\mathbb{R}_+^m$ . The function  $d(\cdot)$  is concave. Note that  $d(u) = -s_{\mathcal{Z}_p}(-u)$ , where  $s_{\mathcal{Z}_p}(\cdot)$  is the support function of the set  $\mathcal{Z}_p$ . By virtue of Theorem 4.62 and Hiriart-Urruty and Lemaréchal [89, Chapter V, Proposition 2.2.1], we have

$$d(u) = \inf \{u^\top z : z \in \text{conv} \mathcal{Z}_p\}. \tag{4.43}$$

### 4.3. Separable Probabilistic Constraints

123

Let us consider the *convex hull problem*:

$$\begin{aligned} & \text{Min}_{x,z} c(x) \\ & \text{s.t. } g(x) \geq z, \\ & \quad x \in \mathcal{X}, \\ & \quad z \in \text{conv} \mathcal{Z}_p. \end{aligned} \tag{4.44}$$

We impose the following constraint qualification condition:

$$\text{There exist points } x^0 \in \mathcal{X} \text{ and } z^0 \in \text{conv} \mathcal{Z}_p \text{ such that } g(x^0) > z^0. \tag{4.45}$$

If this constraint qualification condition is satisfied, then the duality theory in convex programming Rockafellar [174, Corollary 28.2.1] implies that there exists  $\hat{u} \geq 0$  at which the minimum in (4.42) is attained, and  $\Psi^* = \Psi(\hat{u})$  is the optimal value of the convex hull problem (4.44).

We now study in detail the structure of the dual functional  $\Psi$ . We shall characterize the solution sets of the two subproblems (4.40) and (4.41), which provide the values of the dual functional. Observe that the normal cone to the positive orthant at a point  $u \geq 0$  is the following:

$$\mathcal{N}_{\mathbb{R}_+^m}(u) = \{d \in \mathbb{R}_-^m : d_i = 0 \text{ if } u_i > 0, i = 1, \dots, m\}. \tag{4.46}$$

We define the set

$$V(u) = \{v \in \mathbb{R}^m : u^\top v = d(u) \text{ and } v \text{ is a } p\text{-efficient point}\}. \tag{4.47}$$

**Lemma 4.67.** *For every  $u > 0$  the solution set of (4.41) is nonempty. For every  $u \geq 0$  it has the following form:  $\hat{Z}(u) = V(u) - \mathcal{N}_{\mathbb{R}_+^m}(u)$ .*

**Proof.** First we consider the case  $u > 0$ . Then every recession direction  $q$  of  $\mathcal{Z}_p$  satisfies  $u^\top q > 0$ . Since  $\mathcal{Z}_p$  is closed, a solution to (4.41) must exist. Suppose that a solution  $z$  to (4.41) is not a  $p$ -efficient point. By virtue of Theorem 4.60, there is a  $p$ -efficient  $v \in \mathcal{Z}_p$  such that  $v \leq z$ , and  $v \neq z$ . Thus,  $u^\top v < u^\top z$ , which is a contradiction. Therefore, we conclude that there is a  $p$ -efficient point  $v$ , which solves problem (4.41).

Consider the general case  $u \geq 0$  and assume that the solution set of problem (4.41) 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  $u^\top v \leq u^\top z$  holds by the nonnegativity of  $u$ . Consequently,  $u^\top v = u^\top z$  for all  $p$ -efficient  $v \leq z$ , which is equivalent to  $z \in \{v\} - \mathcal{N}_{\mathbb{R}_+^m}(u)$ , as required.

If the solution set of (4.41) is empty, then  $V(u) = \emptyset$  by definition and the assertion is true as well.  $\square$

The last result allows us to calculate the subdifferential of the function  $d$  in a closed form.

**Lemma 4.68.** *For every  $u \geq 0$  one has  $\partial d(u) = \text{conv}(V(u)) - \mathcal{N}_{\mathbb{R}_+^m}(u)$ . If  $u > 0$ , then  $\partial d(u)$  is nonempty.*

**Proof.** From (4.41) we obtain  $d(u) = -s_{Z_p}(-u)$ , where  $s_{Z_p}(\cdot)$  is the support function of  $Z_p$  and, consequently, of  $\text{conv } Z_p$ . Consider the indicator function  $\mathbb{I}_{\text{conv } Z_p}(\cdot)$  of the set  $\text{conv } Z_p$ . By virtue of Corollary 16.5.1 in Rockafellar [174], we have

$$s_{Z_p}(u) = \mathbb{I}_{\text{conv } Z_p}^*(u),$$

where the latter function is the conjugate of the indicator function  $\mathbb{I}_{\text{conv } Z_p}(\cdot)$ . Thus,

$$\partial d(u) = -\partial \mathbb{I}_{\text{conv } Z_p}^*(-u).$$

Recall that  $\text{conv } Z_p$  is closed, by Theorem 4.62. Using Rockafellar [174, Theorem 23.5], we observe that  $y \in \partial \mathbb{I}_{\text{conv } Z_p}^*(-u)$  iff  $\mathbb{I}_{\text{conv } Z_p}^*(-u) + \mathbb{I}_{\text{conv } Z_p}(y) = -y^\top u$ . It follows that  $y \in \text{conv } Z_p$  and  $\mathbb{I}_{\text{conv } Z_p}^*(-u) = -y^\top u$ . Consequently,

$$y^\top u = d(u). \quad (4.48)$$

Since  $y \in \text{conv } Z_p$  we can represent it as follows:

$$y = \sum_{j=1}^{m+1} \alpha_j e^j + w,$$

where  $e^j$ ,  $j = 1, \dots, m+1$ , are extreme points of  $\text{conv } Z_p$  and  $w \geq 0$ . Using Theorem 4.63 we conclude that  $e^j$  are  $p$ -efficient points. Moreover, applying  $u$ , we obtain

$$y^\top u = \sum_{j=1}^{m+1} \alpha_j u^\top e^j + u^\top w \geq d(u), \quad (4.49)$$

because  $u^\top e^j \geq d(u)$  and  $u^\top w \geq 0$ . Combining (4.48) and (4.49) we conclude that  $u^\top e^j = d(u)$  for all  $j$ , and  $u^\top w = 0$ . Thus  $y \in \text{conv } V(u) - \mathcal{N}_{\mathbb{R}_+^m}(u)$ .

Conversely, if  $y \in \text{conv } V(u) - \mathcal{N}_{\mathbb{R}_+^m}(u)$ , then (4.48) 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 4.67.  $\square$

Now, we analyze the function  $h(\cdot)$ . Define the set of minimizers in (4.40),

$$X(u) = \{x \in \mathcal{X} : c(x) - u^\top g(x) = h(u)\}.$$

Since the set  $\mathcal{X}$  is convex and the objective function of problem (4.40) is convex for all  $u \geq 0$ , we conclude that the solution set  $X(u)$  is convex for all  $u \geq 0$ .

**Lemma 4.69.** 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:

$$\partial h(u) = \text{conv} \{-g(x) : x \in X(u)\}.$$

**Proof.** The function  $h$  is concave on  $\mathbb{R}^m$ . Since the set  $\mathcal{X}$  is compact,  $c$  is convex, and  $g_i$ ,  $i = 1, \dots, m$ , are concave, the set  $X(u)$  is compact. Therefore, the subdifferential of



### 4.3. Separable Probabilistic Constraints

125

$h(u)$  for every  $u \in \mathbb{R}^m$  is the closure of  $\text{conv} \{-g(x) : x \in X(u)\}$ . (See Hiriart-Urruty and Lemaréchal [89, 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  $h(u)$ .  $\square$

This analysis provides the basis for the following necessary and sufficient optimality conditions for problem (4.42).

**Theorem 4.70.** *Assume that the constraint qualification condition (4.45) is satisfied and that the set  $\mathcal{X}$  is compact. A vector  $u \geq 0$  is an optimal solution of (4.42) iff there exists a point  $x \in X(u)$ , points  $v^1, \dots, v^{m+1} \in V(u)$  and scalars  $\beta_1, \dots, \beta_{m+1} \geq 0$  with  $\sum_{j=1}^{m+1} \beta_j = 1$  such that*

$$\sum_{j=1}^{m+1} \beta_j v^j - g(x) \in \mathcal{N}_{\mathbb{R}_+^m}(u). \quad (4.50)$$

**Proof.** Using Rockafellar [174, Theorem 27.4], the necessary and sufficient optimality condition for (4.42) has the form

$$0 \in -\partial\Psi(u) + \mathcal{N}_{\mathbb{R}_+^m}(u). \quad (4.51)$$

Since  $\text{int dom } d \neq \emptyset$  and  $\text{dom } h = \mathbb{R}^m$ , we have  $\partial\Psi(u) = \partial h(u) + \partial d(u)$ . Using Lemma 4.68 and Lemma 4.69, we conclude that there exist

$$\begin{aligned} & p\text{-efficient points } v^j \in V(u), \quad j = 1, \dots, m+1, \\ & \beta^j \geq 0, \quad j = 1, \dots, m+1, \quad \sum_{j=1}^{m+1} \beta_j = 1, \\ & x^j \in X(u), \quad j = 1, \dots, m+1, \\ & \alpha^j \geq 0, \quad j = 1, \dots, m+1, \quad \sum_{j=1}^{m+1} \alpha_j = 1, \end{aligned} \quad (4.52)$$

such that

$$\sum_{j=1}^{m+1} \alpha_j g(x^j) - \sum_{j=1}^{m+1} \beta_j v^j \in -\mathcal{N}_{\mathbb{R}_+^m}(u). \quad (4.53)$$

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 (4.50). 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,

$$c(x) - \sum_{i=1}^m u_i g_i(x) = h(u) = c(x^j) - \sum_{i=1}^m u_i g_i(x^j), \quad j = 1, \dots, m+1. \quad (4.54)$$

Multiplying the last equation by  $\alpha_j$  and adding, we obtain

$$c(x) - \sum_{i=1}^m u_i g_i(x) = \sum_{j=1}^{m+1} \alpha_j \left[ c(x^j) - \sum_{i=1}^m u_i g_i(x^j) \right] \geq c(x) - \sum_{i=1}^m u_i \sum_{j=1}^{m+1} \alpha_j g_i(x^j).$$

The last inequality follows from the convexity of  $c$ . We have the following inequality:

$$\sum_{i=1}^m u_i \left[ g_i(x) - \sum_{j=1}^{m+1} \alpha_j g_i(x^j) \right] \leq 0.$$

Since the functions  $g_i$  are concave, we have  $g_i(x) \geq \sum_{j=1}^{m+1} \alpha_j g_i(x^j)$ . 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 (4.53) and obtain (4.50), as required.

Now, we prove the converse implication. Assume that we have  $x \in X(u)$ , points  $v^1, \dots, v^{m+1} \in V(u)$ , and scalars  $\beta_1, \dots, \beta_{m+1} \geq 0$  with  $\sum_{j=1}^{m+1} \beta_j = 1$  such that (4.50) holds true. By Lemma 4.68 and Lemma 4.69 we have

$$-g(x) + \sum_{j=1}^{m+1} \beta_j v^j \in \partial\Psi(u).$$

Thus (4.50) implies (4.51), which is a necessary and sufficient optimality condition for problem (4.42).  $\square$

Using these optimality conditions we obtain the following duality result.

**Theorem 4.71.** *Assume that the constraint qualification condition (4.45) for problem (4.39) is satisfied, the probability distribution of the vector  $Z$  is  $\alpha$ -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 (4.39), then there exists a vector  $\hat{u} \geq 0$ , which is an optimal solution of (4.42) and the optimal values of both problems are equal. If  $\hat{u}$  is an optimal solution of problem (4.42), then there exist a point  $\hat{x}$  such that  $(\hat{x}, g(\hat{x}))$  is a solution of problem (4.39), and the optimal values of both problems are equal.*

**Proof.** The  $\alpha$ -concavity assumption implies that problems (4.39) and (4.44) are the same. If  $\hat{u}$  is optimal solution of problem (4.42), we obtain the existence of points  $\hat{x} \in X(\hat{u})$ ,  $v^1, \dots, v^{m+1} \in V(u)$  and scalars  $\beta_1, \dots, \beta_{m+1} \geq 0$  with  $\sum_{j=1}^{m+1} \beta_j = 1$  such that the optimality conditions in Theorem 4.70 are satisfied. Setting  $\hat{z} = g(\hat{x})$  we have to show that  $(\hat{x}, \hat{z})$  is an optimal solution of problem (4.39) 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

### 4.3. Separable Probabilistic Constraints

127

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

$$\begin{aligned} h(\hat{u}) &= c(\hat{x}) - \hat{u}^\top g(\hat{x}) = c(\hat{x}) - \hat{u}^\top \left( \sum_{j=1}^{m+1} \beta_j v^j + y \right) \\ &= c(\hat{x}) - \sum_{j=1}^{m+1} \beta_j d(\hat{u}) - \hat{u}^\top y = c(\hat{x}) - d(\hat{u}). \end{aligned}$$

Thus,

$$c(\hat{x}) = h(\hat{u}) + d(\hat{u}) = \Psi^* \geq c^*,$$

which proves that  $(\hat{x}, \hat{z})$  is an optimal solution of problem (4.39) and  $\Psi^* = c^*$ .

If  $(\hat{x}, \hat{z})$  is a solution of (4.39), then by Rockafellar [174, Theorem 28.4] there is a vector  $\hat{u} \geq 0$  such that  $\hat{u}_i(\hat{z}_i - g_i(\hat{x})) = 0$  and

$$0 \in \partial c(\hat{x}) + \partial \hat{u}^\top g(\hat{x}) - \hat{z} + \mathcal{N}_{\mathcal{X} \times \mathcal{Z}}(\hat{x}, \hat{z}).$$

This means that

$$0 \in \partial c(\hat{x}) - \partial u^\top g(\hat{x}) + \mathcal{N}_{\mathcal{X}}(\hat{x}) \quad (4.55)$$

and

$$0 \in \hat{u} + \mathcal{N}_{\mathcal{Z}}(\hat{z}). \quad (4.56)$$

The first inclusion (4.55) is optimality condition for problem (4.40), and thus  $x \in X(\hat{u})$ . By virtue of Rockafellar [174, Theorem 23.5] the inclusion (4.56) is equivalent to  $\hat{z} \in \partial \mathbb{I}_{\mathcal{Z}_p}^*(\hat{u})$ . Using Lemma 4.68 we obtain that

$$\hat{z} \in \partial d(\hat{u}) = \text{conv} V(\hat{u}) - \mathcal{N}_{\mathbb{R}_+^m}(\hat{u}).$$

Thus, there exists points  $v^1, \dots, v^{m+1} \in V(\hat{u})$  and scalars  $\beta_1, \dots, \beta_{m+1} \geq 0$  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 4.70 are satisfied. Thus,  $\hat{u}$  is an optimal solution of problem (4.42).  $\square$

For the special case of discrete distribution and linear constraints 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  $c(x) = c^\top x$  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:

$$\begin{aligned} \text{Min}_x \quad & c^\top x \\ \text{s.t.} \quad & \Pr\{Tx \geq Z\} \geq p, \\ & Ax \geq b, \\ & x \geq 0. \end{aligned} \tag{4.57}$$

Here  $A$  is an  $s \times n$  matrix and  $b \in \mathbb{R}^s$ .

**Definition 4.72.** Problem (4.57) satisfies the dual feasibility condition if

$$\Lambda = \{(u, w) \in \mathbb{R}_+^{m+s} : A^\top w + T^\top u \leq c\} \neq \emptyset.$$

**Theorem 4.73.** Assume that the feasible set of (4.57) is nonempty and that  $Z$  has a discrete distribution on  $\mathbb{Z}^m$ . Then (4.57) has an optimal solution iff it satisfies the LQ condition, defined in (4.72).

**Proof.** If (4.57) has an optimal solution, then for some  $j \in \mathcal{E}$  the linear optimization problem

$$\begin{aligned} \text{Min}_x \quad & c^\top x \\ \text{s.t.} \quad & Tx \geq v^j, \\ & Ax \geq b, \\ & x \geq 0, \end{aligned} \tag{4.58}$$

has an optimal solution. By duality in linear programming, its dual problem

$$\begin{aligned} \text{Max}_{u,w} \quad & u^\top v^j + b^\top w \\ \text{s.t.} \quad & T^\top u + A^\top w \leq c, \\ & u, w \geq 0, \end{aligned} \tag{4.59}$$

has an optimal solution and the optimal values of both programs are equal. Thus, the dual feasibility condition (4.72) must be satisfied. On the other hand, if the dual feasibility condition is satisfied, all dual programs (4.59) for  $j \in \mathcal{E}$  have nonempty feasible sets, so the objective values of all primal problems (4.58) are bounded from below. Since at least one of them has a nonempty feasible set by assumption, an optimal solution must exist.  $\square$

**Example 4.74 (Vehicle Routing Continued).** We return to the vehicle routing Example 4.1, introduced at the beginning of the chapter. The convex hull problem reads

$$\begin{aligned} \text{Min}_{x,\lambda} \quad & c^\top x \\ \text{s.t.} \quad & \sum_{i=1}^n t_{il} x_i \geq \sum_{j \in \mathcal{E}} \lambda_j v^j, \end{aligned} \tag{4.60}$$

$$\begin{aligned} & \sum_{j \in \mathcal{E}} \lambda_j = 1, \\ & x \geq 0, \lambda \geq 0. \end{aligned} \tag{4.61}$$

### 4.3. Separable Probabilistic Constraints

129

We assign a Lagrange multiplier  $u$  to constraint (4.60) and a multiplier  $\mu$  to constraint (4.61). The dual problem has the form

$$\begin{aligned} & \text{Max}_{u, \mu} \mu \\ & \text{s.t. } \sum_{l=1}^m t_{il} u_l \leq c_i, \quad i = 1, 2, \dots, n, \\ & \quad \mu \leq u^\top v^j, \quad j \in \mathcal{E}, \\ & \quad u \geq 0. \end{aligned}$$

We see that  $u_l$  provides the increase of routing cost if the demand on arc  $l$  increases by one unit,  $\mu$  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  $\text{Min}_{z \in \mathcal{Z}} u^\top z$  identifies  $p$ -efficient points, which represent critical demand levels. The optimal value of this problem provides the minimum total cost of a critical demand. ■

Our duality theory finds interesting interpretation in the context of the cash matching problem in Example 4.6.

**Example 4.75 (Cash Matching Continued).** Recall the problem formulation

$$\begin{aligned} & \text{Max}_{x, c} \mathbb{E}[U(c_T - Z_T)] \\ & \text{s.t. } \Pr\{c_t \geq Z_t, \quad t = 1, \dots, T\} \geq p, \\ & \quad c_t = c_{t-1} + \sum_{i=1}^n a_{it} x_i, \quad t = 1, \dots, T, \\ & \quad x \geq 0. \end{aligned}$$

If the vector  $Z$  has a quasi-concave distribution (e.g., joint normal distribution), the resulting problem is convex.

The convex hull problem (4.44) can be written as follows:

$$\text{Max}_{x, \lambda, c} \mathbb{E}[U(c_T - Z_T)] \tag{4.62}$$

$$\text{s.t. } c_t = c_{t-1} + \sum_{i=1}^n a_{it} x_i, \quad t = 1, \dots, T, \tag{4.63}$$

$$c_t \geq \sum_{j=1}^{T+1} \lambda_j v_t^j, \quad t = 1, \dots, T, \tag{4.64}$$

$$\sum_{j=1}^{T+1} \lambda_j = 1, \tag{4.65}$$

$$\lambda \geq 0, \quad x \geq 0. \tag{4.66}$$

In constraint (4.64) the vectors  $v^j = (v_1^j, \dots, v_T^j)$  for  $j = 1, \dots, T + 1$  are  $p$ -efficient trajectories of the cumulative liabilities  $Z = (Z_1, \dots, Z_T)$ . Constraints (4.64)–(4.66)

require that the cumulative cash flows are greater than or equal to some convex combination of  $p$ -efficient trajectories. Recall that by Lemma 4.61, no more than  $T + 1$   $p$ -efficient trajectories are needed. Unfortunately, we do not know the optimal collection of these trajectories.

Let us assign nonnegative Lagrange multipliers  $u = (u_1, \dots, u_T)$  to the constraint (4.64), multipliers  $w = (w_1, \dots, w_T)$  to the constraints (4.63) and a multiplier  $\rho \in \mathbb{R}$  to the constraint (4.65). To simplify notation, we define the function  $\bar{U} : \mathbb{R} \rightarrow \mathbb{R}$  as follows:

$$\bar{U}(y) = \mathbb{E}[U(y - Z_T)].$$

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:

$$\begin{aligned} D(w, u, \rho) &= \min_{x \geq 0, \lambda \geq 0, c} \left\{ -\bar{U}(c_T) + \sum_{t=1}^T w_t \left( c_t - c_{t-1} - \sum_{i=1}^n a_{it} x_i \right) \right. \\ &\quad \left. + \sum_{t=1}^T u_t \left( \sum_{j=1}^{T+1} \lambda_j v_t^j - c_t \right) + \rho \left( 1 - \sum_{j=1}^{T+1} \lambda_j \right) \right\} \\ &= -\max_{x \geq 0} \sum_{i=1}^n \sum_{t=1}^T a_{it} w_t x_i + \min_{\lambda \geq 0} \sum_{j=1}^{T+1} \left( \sum_{t=1}^T v_t^j u_t - \rho \right) \lambda_j + \rho \\ &\quad + \min_c \left\{ \sum_{t=1}^{T-1} c_t (w_t - u_t - w_{t+1}) - w_1 c_0 + c_T (w_T - u_T) - \bar{U}(c_T) \right\} \\ &= \rho - w_1 c_0 + \bar{U}^*(w_T - u_T). \end{aligned}$$

The dual problem becomes

$$\text{Min}_{u, w, \rho} -\bar{U}^*(w_T - u_T) + w_1 c_0 - \rho \quad (4.67)$$

$$\text{s.t. } w_t = w_{t+1} + u_t, \quad t = T-1, \dots, 1, \quad (4.68)$$

$$\sum_{t=1}^T w_t a_{it} \leq 0, \quad i = 1, \dots, n, \quad (4.69)$$

$$\rho \leq \sum_{t=1}^T u_t v_t^j, \quad j = 1, \dots, T+1. \quad (4.70)$$

$$u \geq 0. \quad (4.71)$$

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 (4.68) 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  $T$ .

### 4.3. Separable Probabilistic Constraints

131

The constraints (4.69) represent the *nonarbitrage 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  $w_t$ . At the end of the planning horizon, we collect all loans and pay off the debt. The profit from this operation should be nonpositive for each bond in order to comply with the no-free-lunch condition, which is expressed via (4.69).

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

$$\begin{aligned} L_t^j &= v_t^j - v_{t-1}^j, \quad t = 2, \dots, T, \\ L_1^j &= v_1^j. \end{aligned}$$

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 (4.70) 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

$$\begin{aligned} D(w, u, \rho) &= - \max_{x \geq 0} \sum_{i=1}^n \sum_{t=1}^T a_{it} w_t x_i + \min_{\lambda \geq 0} \sum_{j=1}^{T+1} \left( \sum_{t=1}^T v_t^j u_t - \rho \right) \lambda_j + \rho \\ &\quad + \min_c \left\{ \sum_{t=1}^{T-1} c_t (w_t - u_t - w_{t+1}) - w_1 c_0 + c_T (w_T - u_T - 1) \right\} \\ &= \rho - w_1 c_0. \end{aligned}$$

The objective function of the dual problem takes on the form

$$\text{Min}_{u, w, \rho} w_1 c_0 - \rho,$$

and the constraints (4.68) 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 (4.70) is the difference between the cost of covering the  $j$ th  $p$ -efficient liability sequence by borrowed money and the total liability.

The variable  $\rho$  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. ■

## 4.4 Optimization Problems with Nonseparable Probabilistic Constraints

In this section, we concentrate on the following problem:

$$\begin{aligned} \text{Min}_x \quad & c(x) \\ \text{s.t.} \quad & \Pr\{g(x, Z) \geq 0\} \geq p, \\ & x \in \mathcal{X}. \end{aligned} \tag{4.72}$$

The parameter  $p \in (0, 1)$  denotes some probability level. We assume that the functions  $c : \mathbb{R}^n \times \mathbb{R}^s \rightarrow \mathbb{R}$  and  $g : \mathbb{R}^n \times \mathbb{R}^s \rightarrow \mathbb{R}^m$  are continuous and the set  $\mathcal{X} \subset \mathbb{R}^n$  is a closed convex set. We define the constraint function as follows:

$$G(x) = \Pr\{g(x, Z) \geq 0\}.$$

Recall that if  $G(\cdot)$  is  $\alpha$ -concave function,  $\alpha \in \mathbb{R}$ , then a transformation of it is a concave function. In this case, we define

$$\bar{G}(x) = \begin{cases} \ln p - \ln[G(x)] & \text{if } \alpha = 0, \\ p^\alpha - [G(x)]^\alpha & \text{if } \alpha > 0, \\ [G(x)]^\alpha - p^\alpha & \text{if } \alpha < 0. \end{cases} \tag{4.73}$$

We obtain the following equivalent formulation of problem (4.72):

$$\begin{aligned} \text{Min}_x \quad & c(x) \\ \text{s.t.} \quad & \bar{G}(x) \leq 0, \\ & x \in \mathcal{X}. \end{aligned} \tag{4.74}$$

Assuming that  $c(\cdot)$  is convex, we have a convex problem.

Recall that Slater's condition is satisfied for problem (4.72) if there is a point  $x^s \in \text{int } \mathcal{X}$  such that  $\bar{G}(x^s) > 0$ . Using optimality conditions for convex optimization problems, we can infer the following conditions for problem (4.72).

**Theorem 4.76.** *Assume that  $c(\cdot)$  is a continuous convex function, the functions  $g : \mathbb{R}^n \times \mathbb{R}^s \rightarrow \mathbb{R}^m$  are quasi-concave,  $Z$  has an  $\alpha$ -concave distribution, and the set  $\mathcal{X} \subset \mathbb{R}^n$  is closed and convex. Furthermore, let Slater's condition be satisfied and  $\text{int dom } G \neq \emptyset$ .*



#### 4.4. Optimization Problems with Nonseparable Probabilistic Constraints 133

A point  $\hat{x} \in \mathcal{X}$  is an optimal solution of problem (4.72) iff there is a number  $\lambda \in \mathbb{R}_+$  such that  $\lambda[G(\hat{x}) - p] = 0$  and

$$0 \in \partial c(\hat{x}) + \lambda \frac{1}{\alpha} G(\hat{x})^{1-\alpha} \partial G(\hat{x})^\alpha + \mathcal{N}_{\mathcal{X}}(\hat{x}) \quad \text{if } \alpha \neq 0,$$

or

$$0 \in \partial c(\hat{x}) + \lambda G(\hat{x}) \partial (\ln G(\hat{x})) + \mathcal{N}_{\mathcal{X}}(\hat{x}) \quad \text{if } \alpha = 0.$$

**Proof.** Under the assumptions of the theorem, problem (4.72) can be reformulated in form (4.74), 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 nonempty interior, and therefore the assumptions of Theorem 4.29 are satisfied.  $\square$

##### 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 piecewise smooth boundary such that  $\text{supp } P_Z = \text{cl}\{\text{int}(\text{supp } P_Z)\}$ . 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:

$$H(x) = \{z \in \mathbb{R}^s : g_i(x, z) \geq 0, i = 1, \dots, m\}.$$

We denote the boundary of a set  $H(x)$  by  $\text{bd}H(x)$ . For an open set  $U \subset \mathbb{R}^n$  containing the origin, we set

$$H_U = \text{cl} \left( \bigcup_{x \in U} H(x) \right) \quad \text{and} \quad \Delta H_U = \text{cl} \left( \bigcup_{x \in U} \text{bd}H(x) \right),$$

$$V_U = \text{cl } U \times H_U \quad \text{and} \quad \Delta V_U = \text{cl } U \times \Delta H_U.$$

For any of these sets, we indicate with upper subscript  $r$  its restriction to the  $\text{supp } P_Z$ , e.g.,  $H_U^r = H_U \cap \text{supp } P_Z$ . Let

$$S_i(x) = \{z \in \text{supp } P_Z : g_i(x, z) = 0, g_j(x, z) \geq 0, j \neq i\}, \quad i = 1, \dots, m.$$

We use the notation

$$S(x) = \bigcup_{i=1}^M S_i(x), \quad \Delta H_i = \text{int} \left( \bigcup_{x \in U} (\partial \{g_i(x, z) \geq 0\} \cap H^r(x)) \right).$$

The  $(m - 1)$ -dimensional Lebesgue measure is denoted by  $P_{m-1}$ . We assume that the functions  $g_i(x, z)$ ,  $i = 1, \dots, m$ , are continuously differentiable and such that  $\text{bd}H(x) =$

$S(x)$  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,  $\Delta H_U$  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 4.77.** *Assume that*

- (i) *the vector functions  $\nabla_x g(x, z)$  and  $\nabla_z g(x, z)$  are continuous on  $\Delta V_U^r$ ;*
- (ii) *the vector functions  $\nabla_z g(x, z) > 0$  (componentwise) on the set  $\Delta V_U^r$ ;*
- (iii) *the function  $\|\nabla_x g(x, z)\| > 0$  on  $\Delta V_U^r$ .*

*Then the probability function  $G(x) = \Pr\{g(x, Z) \geq 0\}$  has partial derivatives for almost all  $x \in U$  that can be represented as a surface integral,*

$$\left( \frac{\partial G(x)}{\partial x_i} \right)_{i=1}^n = \int_{\text{bd } H(x) \cap \text{supp } P_Z} \frac{\theta(z)}{\|\nabla_z g(x, z)\|} \nabla_x g(x, z) dS.$$

**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:

$$\begin{aligned} G(x) - G(y) &= \int_{H(x)} \theta(z) dz - \int_{H(y)} \theta(z) dz \\ &= \int_{H^r(x) \setminus H^r(y)} \theta(z) dz - \int_{H^r(y) \setminus H^r(x)} \theta(z) dz. \end{aligned} \quad (4.75)$$

By the implicit function theorem, the equation  $g(x, z) = 0$  determines a differentiable function  $x(z)$  such that

$$g(x(z), z) = 0 \quad \text{and} \quad \nabla_z x(z) = - \frac{\nabla_x g(x, z)}{\nabla_z g(x, z)} \Big|_{x=x(z)}.$$

Moreover, the constraint  $g(x, z) \geq 0$  is equivalent to  $x \geq x(z)$  for all  $(x, z) \in U \times \Delta H_U^r$ , 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 (4.75) as follows:

$$G(x) - G(y) = - \int_{\{z \in \mathbb{R}^s : y \geq x(z) > x\}} \theta(z) dz.$$

Now, we apply Schwarz [194, Vol. 1, Theorem 108] and obtain

$$\begin{aligned} G(x) - G(y) &= - \int_x^y \int_{\{z \in \mathbb{R}^s : x(z)=t\}} \frac{\theta(z)}{\|\nabla_z x(z)\|} dS dt \\ &= \int_y^x \int_{\text{bd } H(x)^r} \frac{|\nabla_x g(t, z)| \theta(z)}{\|\nabla_z g(x, z)\|} dS dt. \end{aligned}$$

#### 4.4. Optimization Problems with Nonseparable Probabilistic Constraints 135

By Fubini's theorem [194, Vol. 1, Theorem 77], the inner integral converges almost everywhere with respect to the Lebesgue measure. Therefore, we can apply Schwarz [194, 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

$$\frac{\partial}{\partial x} G(x) = \int_{\text{bd} H^r(x)} \frac{\nabla_x g(x, z) \theta(z)}{\|\nabla_z g(x, z)\|} dS.$$

We have used assumption (ii) to set  $|\nabla_x g(x, z)| = \nabla_x g(x, z)$ .  $\square$

Obviously, the statement remains valid if assumption (ii) is replaced by the opposite strict inequality, so that the function  $g(x, z)$  would be strictly decreasing on  $U \times \Delta H_U^r$ .

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 is elaborated in Ermoliev [65] and Uryasev [216].

The conditions of this theorem can be slightly modified so that the result and the formula for the derivative are valid for piecewise smooth function.

**Theorem 4.78 (Raik [166]).** *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  $\Delta H_i$  for each  $i = 1, \dots, m$ ;*
- (ii) *the vector functions  $\nabla_z g_i(x, z)$  and  $\nabla_x g_i(x, z)$  are continuous and bounded on the set  $U \times \Delta H_i$  for each  $i = 1, \dots, m$ ;*
- (iii) *the function  $\|\nabla_x g_i(x, z)\| \geq \delta > 0$  on the set  $U \times \Delta H_i$  for each  $i = 1, \dots, m$ ;*
- (iv) *the following conditions are satisfied for all  $i = 1, \dots, m$  and all  $x \in U$ :*

$$P_{m-1}\{S_i(x) \cap S_j(x)\} = 0, \quad i \neq j, \quad P_{m-1}\{\text{bd}(\text{supp } P_Z \cap S_i(x))\} = 0.$$

*Then the probability function  $G(x)$  is differentiable on  $U$  and*

$$\nabla G(x) = \sum_{i=1}^m \int_{S_i(x)} \frac{\theta(z)}{\|\nabla_z g_i(x, z)\|} \nabla_x g_i(x, z) dS. \quad (4.76)$$

The precise proof of this theorem is omitted. We refer to Kibzun and Tretyakov [104] and Kibzun and Uryasev [105] for more information on this topic.

For example, if  $g(x, Z) = x^\top Z$ ,  $m = 1$ , and  $Z$  has a nondegenerate multivariate normal distribution  $\mathcal{N}(\bar{z}, \Sigma)$ , then  $g(x, Z) \sim \mathcal{N}(x^\top \bar{z}, x^\top \Sigma x)$ , and hence the probability function  $G(x) = \Pr\{g(x, Z) \geq 0\}$  can be written in the form

$$G(x) = \Phi\left(\frac{x^\top \bar{z}}{\sqrt{x^\top \Sigma x}}\right),$$

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 (4.72), 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

$$\sum_{i=1}^m \int_{S_i(x)} \frac{\theta(z)}{\|\nabla_z g_i(\hat{x}, z)\|} (x^r - \hat{x})^\top \nabla_x g_i(\hat{x}, z) dS < 0. \quad (4.77)$$

This condition implies Robinson's condition. We obtain the following necessary optimality conditions.

**Theorem 4.79.** *Under the assumption of Theorem 4.78, let the constraint qualification (4.77) be satisfied, let the function  $c(\cdot)$  be continuously differentiable, and let  $\hat{x} \in \mathcal{X}$  be an optimal solution of problem (4.72). Then there is a multiplier  $\lambda \geq 0$  such that*

$$0 \in \nabla c(\hat{x}) - \lambda \sum_{i=1}^m \int_{S_i(x)} \frac{\theta(z)}{\|\nabla_z g_i(x, z)\|} \nabla_x g_i(x, z) dS + \mathcal{N}_{\mathcal{X}}(\hat{x}), \quad (4.78)$$

$$\lambda[G(\hat{x}) - p] = 0. \quad (4.79)$$

**Proof.** The statement follows from the necessary optimality conditions for smooth optimization problems and formula (4.76).  $\square$

## 4.4.2 Approximations of Nonseparable Probabilistic Constraints

### Smoothing Approximation via Steklov Transformation

In order to apply the optimality conditions formulated in Theorem 4.76, we need to calculate the subdifferential of the probability function  $\bar{G}$  defined by the formula (4.73). The calculation involves the subdifferential of the probability function and the characteristic function of the event

$$\{g_i(x, z) \geq 0, i = 1, \dots, m\}.$$

The latter function may be discontinuous. To alleviate this difficulty, we shall approximate the function  $G(x)$  by smooth functions.

Let  $k : \mathbb{R} \rightarrow \mathbb{R}$  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  $\varepsilon > 0$ :

$$F_K^\varepsilon(\tau) = \int_{-\infty}^{+\infty} \mathbf{1}_{[0, \infty)}(\tau + \varepsilon t) k(t) dt = \frac{1}{\varepsilon} \int_{-\infty}^{+\infty} \mathbf{1}_{[0, \infty)}(t) k\left(\frac{t - \tau}{\varepsilon}\right) dt. \quad (4.80)$$

We see that by the definition of  $F_K^\varepsilon$  and  $\mathbf{1}_{[0, \infty)}$ , and by the symmetry of  $k(\cdot)$  we have

$$\begin{aligned} F_K^\varepsilon(\tau) &= \int_{-\infty}^{+\infty} \mathbf{1}_{[0, \infty)}(\tau + \varepsilon t) k(t) dt = \int_{-\tau/\varepsilon}^{+\infty} k(t) dt \\ &= \int_{-\infty}^{\tau/\varepsilon} k(-t) dt = \int_{-\infty}^{\tau/\varepsilon} k(t) dt \\ &= F_K\left(\frac{\tau}{\varepsilon}\right). \end{aligned} \quad (4.81)$$

#### 4.4. Optimization Problems with Nonseparable Probabilistic Constraints 137

Setting

$$g_M(x, z) = \min_{1 \leq i \leq m} g_i(x, z),$$

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 (4.81), we can approximate the constraint function  $G(\cdot)$  by the function

$$\begin{aligned} G_\varepsilon(x) &= \int_{\mathbb{R}^s} F_K^\varepsilon(g_M(x, z) - c) dP_z \\ &= \int_{\mathbb{R}^s} F_K\left(\frac{g_M(x, z) - c}{\varepsilon}\right) dP_z \\ &= \frac{1}{\varepsilon} \int_{\mathbb{R}^s} \int_{-\infty}^{-c} k\left(\frac{t + g_M(x, z)}{\varepsilon}\right) dt dP_z. \end{aligned} \quad (4.82)$$

Now, we show that the functions  $G_\varepsilon(\cdot)$  uniformly converge to  $G(\cdot)$  when  $\varepsilon$  converges to zero.

**Theorem 4.80.** *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*

$$\Pr\{z \in \mathbb{R}^s : g_M(x, z) = c\} = 0.$$

*Then for any compact set  $\mathbf{C} \subset \mathcal{X}$  the functions  $G_\varepsilon$  uniformly converge on  $\mathbf{C}$  to  $G$  when  $\varepsilon \rightarrow 0$ , i.e.,*

$$\lim_{\varepsilon \downarrow 0} \max_{x \in \mathbf{C}} |G_\varepsilon(x) - G(x)| = 0.$$

**Proof.** Defining  $\delta(\varepsilon) = \varepsilon^{1-\beta}$  with  $\beta \in (0, 1)$ , we have

$$\lim_{\varepsilon \rightarrow 0} \delta(\varepsilon) = 0 \quad \text{and} \quad \lim_{\varepsilon \rightarrow 0} F_K\left(\frac{\delta(\varepsilon)}{\varepsilon}\right) = 1, \quad \lim_{\varepsilon \rightarrow 0} F_K\left(\frac{-\delta(\varepsilon)}{\varepsilon}\right) = 0. \quad (4.83)$$

Define for any  $\delta > 0$  the sets

$$\begin{aligned} A(x, \delta) &= \{z \in \mathbb{R}^s : g_M(x, z) - c \leq -\delta\}, \\ B(x, \delta) &= \{z \in \mathbb{R}^s : g_M(x, z) - c \geq \delta\}, \\ C(x, \delta) &= \{z \in \mathbb{R}^s : |g_M(x, z) - c| \leq \delta\}. \end{aligned}$$

On the set  $A(x, \delta(\varepsilon))$  we have  $\mathbf{1}_{[0, \infty)}(g_M(x, z) - c) = 0$  and, using (4.81), we obtain

$$F_K^\varepsilon(g_M(x, z) - c) = F_K\left(\frac{g_M(x, z) - c}{\varepsilon}\right) \leq F_K\left(\frac{-\delta(\varepsilon)}{\varepsilon}\right).$$

On the set  $B(x, \delta(\varepsilon))$  we have  $\mathbf{1}_{[0, \infty)}(g_M(x, z) - c) = 1$  and

$$F_K^\varepsilon(g_M(x, z) - c) = F_K\left(\frac{g_M(x, z) - c}{\varepsilon}\right) \geq F_K\left(\frac{\delta(\varepsilon)}{\varepsilon}\right).$$

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:

$$\begin{aligned} & |G(x) - G_\varepsilon(x)| \\ & \leq \int_{\mathbb{R}^s} |\mathbf{1}_{[0, \infty)}(g_M(x, z) - c) - F_K^\varepsilon(g_M(x, z) - c)| dP_z \\ & \leq F_K\left(\frac{-\delta(\varepsilon)}{\varepsilon}\right) \int_{A(x, \delta(\varepsilon))} dP_Z + \left(1 - F_K\left(\frac{\delta(\varepsilon)}{\varepsilon}\right)\right) \int_{B(x, \delta(\varepsilon))} dP_Z + 2 \int_{C(x, \delta(\varepsilon))} dP_Z \\ & \leq F_K\left(\frac{-\delta(\varepsilon)}{\varepsilon}\right) + \left(1 - F_K\left(\frac{\delta(\varepsilon)}{\varepsilon}\right)\right) + 2P_Z(C(x, \delta(\varepsilon))). \end{aligned}$$

The first two terms on the right-hand side of the inequality converge to zero when  $\varepsilon \rightarrow 0$  by the virtue of (4.83). It remains to show that  $\lim_{\varepsilon \rightarrow 0} P_Z\{C(x, \delta(\varepsilon))\} = 0$  uniformly with respect to  $x \in \mathbf{C}$ . The function  $(x, z, \delta) \mapsto |g_M(x, z) - c| - \delta$  is continuous in  $(x, \delta)$  and measurable in  $z$ . Thus, it is uniformly continuous with respect to  $(x, \delta)$  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

$$\Theta(x, \delta) = P\{|g_M(x, z) - c| - \delta \leq 0\} = P\{\cap_{\beta > \delta} C(x, \beta)\}$$

is uniformly continuous with respect to  $(x, \delta)$  on  $\mathbf{C} \times [-\delta_0, \delta_0]$ . 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,

$$\lim_{\varepsilon \rightarrow 0} P_Z\{z \in \mathbb{R}^s : |g_M(x, z) - c| \leq \delta(\varepsilon)\} = \lim_{\delta \rightarrow 0} \Theta(x, \delta) = 0.$$

As  $\Theta(\cdot, \delta)$  is continuous, the convergence is uniform on compact sets with respect to the first argument.  $\square$

Now, we derive a formula for the Clarke generalized gradients of the approximation  $G_\varepsilon$ . We define the index set

$$I(x, z) = \{i : g_i(x, z) = g_M(x, z), 1 \leq i \leq m\}.$$

**Theorem 4.81.** *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:*

$$\sup\{s \in \partial g_i(y, z), \|y - x\| \leq \delta\} \leq l_\delta(x, z), \delta > 0, \quad \forall i = 1, \dots, m,$$

where  $l_\delta(x, z)$  is an integrable function of  $z$  for all  $x \in \mathbf{X}$ . Then  $G_\varepsilon(\cdot)$  is Lipschitz continuous and Clarke-regular, and its Clarke generalized gradient set is given by

$$\partial^\circ G_\varepsilon(x) = \frac{1}{\varepsilon} \int_{\mathbb{R}^s} k\left(\frac{g_M(x, z) - c}{\varepsilon}\right) \text{conv}\{\partial g_i(x, z) : i \in I(x, z)\} dP_Z.$$

#### 4.4. Optimization Problems with Nonseparable Probabilistic Constraints 139

**Proof.** Under the assumptions of the theorem, the function  $F_K(\cdot)$  is monotone 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

$$\partial g_M(y, z) = \text{conv}\{s_i \in \partial g_i(y, z) : g_i(y, z) = g_M(y, z)\}.$$

Thus the subgradients of  $g_M$  are uniformly bounded:

$$\sup\{s \in \partial g_M(y, z), \|y - x\| \leq \delta\} \leq l_\delta(x, z), \delta > 0.$$

Therefore, the composite function  $F_K\left(\frac{g_M(x, z) - c}{\varepsilon}\right)$  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 [38, 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.  $\square$

Now we show that if we choose  $K$  to have an  $\alpha$ -concave distribution, and all assumptions of Theorem 4.39 are satisfied, the generalized concavity property of the approximated probability function is preserved.

**Theorem 4.82.** *If the density function  $k$  is  $\alpha$ -concave ( $\alpha \geq 0$ ),  $Z$  has  $\gamma$ -concave distribution ( $\gamma \geq 0$ ), the functions  $g_i(\cdot, z)$ ,  $i = 1, \dots, m$ , are quasi-concave, then the approximate probability function  $G_\varepsilon$  has a  $\beta$ -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  $\alpha$ -concave ( $\alpha \geq 0$ ), then  $K$  has a  $\gamma$ -concave distribution with  $\gamma = \alpha/(1 + s\alpha)$ . If  $Z$  has  $\gamma'$ -concave distribution ( $\gamma' \geq 0$ ), then the random vector  $(Z, K)^\top$  has a  $\beta$ -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 (4.82), we can write

$$\begin{aligned} G_\varepsilon(x) &= \int_{\mathbb{R}^s} F_K^\varepsilon\left(\frac{g_M(x, z) - c}{\varepsilon}\right) dP_z = \int_{\mathbb{R}^s} \int_{-\infty}^{(g_M(x, z) - c)/\varepsilon} k(t) dt dP_z \\ &= \int_{\mathbb{R}^s} \int_{-\infty}^{\infty} \mathbf{1}_{\{(g_M(x, z) - c)/\varepsilon > t\}} dP_K dP_z = \int_{\mathbb{R}^s} \int_{H_\varepsilon(x)} dP_K dP_z, \end{aligned} \quad (4.84)$$

where

$$H_\varepsilon(x) = \{(z, t) \in \mathbb{R}^{s+1} : g_M(x, z) - \varepsilon t \geq c\}.$$

Since  $g_M(\cdot, z)$  is quasi-concave, the set  $H_\varepsilon(x)$  is convex. Representation (4.84) of  $G_\varepsilon$  and the  $\beta$ -concavity of  $(Z, K)$  imply the assumptions of Theorem 4.39, and, thus, the function  $G_\varepsilon$  is  $\beta$ -concave.  $\square$

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 4.83.** *In addition to the assumptions of Theorems 4.80, 4.81, and 4.82. Then on the set  $\{x \in \mathbb{R}^n : G(x) > 0\}$ , the function  $G_\varepsilon$  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  $\varepsilon \downarrow 0$ ,  $x^\varepsilon \rightarrow x$  and  $s^\varepsilon \in \partial^\circ G_\varepsilon(x^\varepsilon)$  such that  $s^\varepsilon \rightarrow s$ , then  $s \in \partial^\circ G(x)$ .*

**Proof.** Consider a point  $x$  such that  $G(x) > 0$  and points  $x^\varepsilon \rightarrow x$  as  $\varepsilon \downarrow 0$ . All points  $x^\varepsilon$  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 4.80, this neighborhood can be chosen small enough, so that  $G_\varepsilon(y) > 0$  for all  $\varepsilon$  small enough, as well. The functions  $G_\varepsilon$  are generalized concave by virtue of Theorem 4.82. It follows that  $G_\varepsilon$  are locally Lipschitz continuous, directionally differentiable, and Clarke-regular due to Theorem 4.29. Using the uniform convergence of  $G_\varepsilon$  on compact sets and the definition of Clarke generalized gradient, we can pass to the limit with  $\varepsilon \downarrow 0$  in the inequality

$$\lim_{t \downarrow 0, y \rightarrow x^\varepsilon} \frac{1}{t} [G_\varepsilon(y + td) - G_\varepsilon(y)] \geq d^\top s^\varepsilon \quad \text{for any } d \in \mathbb{R}^n.$$

Consequently,  $s \in \partial^\circ G(x)$ .  $\square$

Using the approximate probability function we can solve the following approximation of problem (4.72):

$$\begin{aligned} & \text{Min}_x \quad c(x) \\ & \text{s.t.} \quad G_\varepsilon(x) \geq p, \\ & \quad \quad x \in \mathcal{X}. \end{aligned} \tag{4.85}$$

Under the conditions of Theorem 4.83 the function  $G_\varepsilon$  is  $\beta$ -concave for some  $\beta \geq 0$ . We can specify the necessary and sufficient optimality conditions for the approximate problem.

**Theorem 4.84.** *In addition to the assumptions of Theorem 4.83, assume that  $c(\cdot)$  is a convex function, the Slater condition for problem (4.85) is satisfied, and  $\text{int} G_\varepsilon \neq \emptyset$ . A point  $\hat{x} \in \mathcal{X}$*



#### 4.4. Optimization Problems with Nonseparable Probabilistic Constraints 141

is an optimal solution of problem (4.85) iff a nonpositive number  $\lambda$  exists such that

$$0 \in \partial c(\hat{x}) + s\lambda \int_{\mathbb{R}^s} k\left(\frac{g_M(\hat{x}, z) - c}{\varepsilon}\right) \text{conv}\{\partial g_i(\hat{x}, z) : i \in I(\hat{x}, z)\} dP_Z + \mathcal{N}_{\mathcal{X}}(\hat{x}),$$

$$\lambda[G_\varepsilon(\hat{x}) - p] = 0.$$

Here

$$s = \begin{cases} \alpha\varepsilon^{-1}[G_\varepsilon(\hat{x})]^{\alpha-1} & \text{if } \beta \neq 0, \\ [\varepsilon G_\varepsilon(\hat{x})]^{-1} & \text{if } \beta = 0. \end{cases}$$

**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  $\bar{G}_\varepsilon$ , and formulate the problem

$$\begin{aligned} & \text{Min}_x c(x) \\ & \text{s.t. } \ln p - \bar{G}_\varepsilon(x) \leq 0, \\ & x \in \mathcal{X}. \end{aligned} \tag{4.86}$$

Clearly,  $\hat{x}$  is a solution of the problem (4.86) iff it is a solution of problem (4.85). Problem (4.86) 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 iff a number  $\lambda_0 > 0$  exists such that

$$0 \in \partial c(x) + \lambda_0 \partial[-\bar{G}_\varepsilon(\hat{x})] + \mathcal{N}_{\mathcal{X}}(\hat{x}), \tag{4.87}$$

$$\lambda_0[G_\varepsilon(\hat{x}) - p] = 0. \tag{4.88}$$

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  $G_\varepsilon$ , 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 (4.87), we obtain the result.  $\square$

#### 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

$$\begin{aligned} & \text{Max } c(x) \\ & \text{s.t. } \Pr\{x^\top Z \geq \eta\} \geq p, \\ & x \in \mathcal{X}. \end{aligned} \tag{4.89}$$

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 formulated a portfolio optimization problem under a Value-at-Risk constraint.

$$\begin{aligned}
 & \text{Max } \sum_{i=1}^n \mathbb{E}[R_i]x_i \\
 & \text{s.t. } \Pr\left\{\sum_{i=1}^n R_i x_i \geq -\eta\right\} \geq p, \\
 & \sum_{i=1}^n x_i \leq 1, \\
 & x \geq 0.
 \end{aligned} \tag{4.90}$$

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  $R_1, \dots, R_n$  have a joint normal probability distribution. Recall that the normal distribution is log-concave and the probabilistic constraint in problem (4.90) 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  $\bar{r}_i = \mathbb{E}[R_i]$ ,  $\bar{r} = (\bar{r}_1, \dots, \bar{r}_n)^\top$ , and assume that  $\bar{r}$  is not the zero-vector. Further, let  $\Sigma$  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  $\mathbb{E}[G(x, R)] = \bar{r}^\top x$  and variance  $\text{Var}[G(x, R)] = x^\top \Sigma x$ . Assuming that  $\Sigma$  is positive definite, the probabilistic constraint

$$\Pr\{G(x, R) \geq -\eta\} \geq p$$

can be written in the form (see the discussion on page 16)

$$z_p \sqrt{x^\top \Sigma x} - \bar{r}^\top x \leq \eta.$$

Hence problem (4.90) can be written in the following form:

$$\begin{aligned}
 & \text{Max } \bar{r}^\top x \\
 & \text{s.t. } z_p \sqrt{x^\top \Sigma x} - \bar{r}^\top x \leq \eta, \\
 & \sum_{i=1}^n x_i \leq 1, \\
 & x \geq 0.
 \end{aligned} \tag{4.91}$$

Note that  $\sqrt{x^\top \Sigma x}$  is a convex function, of  $x$ , and  $z_p = \Phi^{-1}(p)$  is positive for  $p > 1/2$ , and hence (4.91) is a convex programming problem.

#### 4.4. Optimization Problems with Nonseparable Probabilistic Constraints 143

Now, we consider the general optimization problem (4.89). 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  $x^\top Z$  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 nonzero components, so that the central limit theorem would be applicable to  $x^\top Z$ . 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  $Z_i, i = 1, 2, \dots$ , 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  $\sigma_i^2 = \text{Var}[Z_i]$ . Setting

$$s_n^2 = \sum_{i=1}^n \sigma_i^2 \quad \text{and} \quad r_n^3 = \sum_{i=1}^n \mathbb{E}(|Z_i - \mu_i|^3),$$

we assume that  $r_n^3$  is finite for every  $n$  and that

$$\lim_{n \rightarrow \infty} \frac{r_n}{s_n} = 0. \quad (4.92)$$

Then the distribution of the random variable

$$\frac{\sum_{i=1}^n (Z_i - \mu_i)}{s_n} \quad (4.93)$$

converges toward the standard normal distribution as  $n \rightarrow \infty$ .

The condition (4.92) 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

$$Y_{in} = \begin{cases} (Z_i - \mu_i)^2 / s_n^2 & \text{if } |Z_i - \mu_i| > \varepsilon s_n, \\ 0 & \text{otherwise.} \end{cases}$$

The Lindeberg's condition reads

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{E}(Y_{in}) = 0.$$

Let us denote  $\bar{z} = (\mu_1, \dots, \mu_n)^\top$ . Under the conditions of the CLT, the distribution of our random variable  $x^\top Z$  is close to the normal distribution with expected value  $x^\top \bar{z}$  and variance  $\sum_{i=1}^n \sigma_i^2 x_i^2$  for problems of large dimensions. Our probabilistic constraint takes on the form

$$\frac{\bar{z}^\top x - \eta}{\sqrt{\sum_{i=1}^n \sigma_i^2 x_i^2}} \geq z_p.$$

Define  $\mathcal{X} = \{x \in \mathbb{R}_+^n : \sum_{i=1}^n x_i \leq 1\}$ . Denoting the matrix with diagonal elements  $\sigma_1, \dots, \sigma_n$  by  $D$ , problem (4.89) can be replaced by the following approximate problem:

$$\begin{aligned} \text{Min}_x \quad & c(x) \\ \text{s.t.} \quad & z_p \|Dx\| \leq \bar{z}^\top x - \eta, \\ & x \in \mathcal{X}. \end{aligned}$$

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  $D$  has a simple diagonal form. There are versions of the CLT 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:

$$\begin{aligned} \text{Min}_x \quad & c(x) \\ \text{s.t.} \quad & \Pr\{g(x, Z) \geq \eta\} \geq \Pr\{Y \geq \eta\}, \quad \eta \in [a, b], \\ & x \in \mathcal{X}. \end{aligned}$$

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  $G(x, \eta) = P\{z : g(x, z) \geq \eta\}$ . 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  $x$  as 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  $\text{bd } H(z, \eta)$  is the surface of the set  $H(z, \eta) = \{z : g(x, z) \geq \eta\}$  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:

$$\begin{aligned} u(\cdot) &\text{ is nondecreasing and right continuous;} \\ u(t) &= 0, \quad \forall t \leq a; \\ u(t) &= u(b), \quad \forall t \geq b. \end{aligned}$$

It is evident that  $\mathcal{U}([a, b])$  is a convex cone.

First we derive a useful formula.

**Lemma 4.85.** *For any real random variable  $Z$  and any measure  $\mu \in \mathcal{M}([a, b])$  we have*

$$\int_a^b \Pr\{Z \geq \eta\} d\mu(\eta) = \mathbb{E}[u(Z)], \quad (4.94)$$

where  $u(z) = \mu([a, z])$ .

**Proof.** We extend the measure  $\mu$  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

$$\begin{aligned} \int_a^b \Pr\{Z \geq \eta\} d\mu(\eta) &= \int_a^\infty \Pr\{Z \geq \eta\} d\mu(\eta) = \int_a^\infty \int_\eta^\infty dP_Z(z) d\mu(\eta) \\ &= \int_a^\infty \int_a^z d\mu(\eta) dP_Z(z) = \int_a^\infty \mu([a, z]) dP_Z(z) = \mathbb{E}[\mu([a, Z])]. \end{aligned}$$

We define  $u(z) = \mu([a, z])$  and obtain the stated result.  $\square$

Let us observe that if the measure  $\mu$  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

$$u(z_1) = \mu([a, z_1]) = \mu([a, z_2]) + \mu((z_2, z_1]) \geq \mu([a, z_2]) = u(z_2).$$

Furthermore,  $u(z) = \mu([a, z]) = \mu([a, b]) = u(b)$  for  $z \geq b$ .

We introduce the functional  $L : \mathbb{R}^n \times \mathcal{U} \rightarrow \mathbb{R}$  associated with problem (4.9):

$$L(x, u) = c(x) + \mathbb{E}[u(g(x, Z)) - u(Y)].$$

We shall see that the functional  $L$  plays the role of a Lagrangian of the problem.

We also set  $v(\eta) = \Pr\{Y \geq \eta\}$ .

**Definition 4.86.** *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*

$$\min_{a \leq \eta \leq b} \left[ G(\hat{x}, \eta) + \nabla_x G(\hat{x}, \eta)(x^0 - \hat{x}) - v(\eta) \right] > 0.$$

**Theorem 4.87.** *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

$$-\nabla_x L(\hat{x}, \hat{u}) \in \mathcal{N}_{\mathcal{X}}(\hat{x}), \quad (4.95)$$

$$\mathbb{E}[\hat{u}(g(\hat{x}, Z))] = \mathbb{E}[\hat{u}(Y)]. \quad (4.96)$$

**Proof.** We consider the mapping  $\Gamma : \mathcal{X} \rightarrow \mathcal{C}([a, b])$  defined as follows:

$$\Gamma(x)(\eta) = \Pr\{g(x, Z) \geq \eta\} - v(\eta), \quad \eta \in [a, b]. \quad (4.97)$$

We define  $K$  as the cone of nonnegative functions in  $\mathcal{C}([a, b])$ . Problem (4.9) can be formulated as follows:

$$\begin{aligned} & \text{Min}_x \quad c(x) \\ & \text{s.t.} \quad \Gamma(x) \in K, \\ & \quad \quad x \in \mathcal{X}. \end{aligned} \quad (4.98)$$

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. Second, the differential uniform dominance condition is equivalent to Robinson's constraint qualification condition:

$$0 \in \text{int}\left\{\Gamma(\hat{x}) + \nabla_x \Gamma(\hat{x})(\mathcal{X} - \hat{x}) - K\right\}. \quad (4.99)$$

Indeed, it is easy to see that the uniform dominance condition implies Robinson's condition. On the other hand, if Robinson's condition holds true, then there exists  $\varepsilon > 0$  such that the function identically equal to  $\varepsilon$  is an element of the set at the right-hand side of (4.99). Then we can find  $x^0$  such that

$$\Gamma(\hat{x})(\eta) + [\nabla_x \Gamma(\hat{x})(\eta)](x^0 - \hat{x}) \geq \varepsilon, \quad \forall \eta \in [a, b].$$

Consequently, the uniform dominance condition is satisfied.

By the Riesz representation theorem, the space dual to  $\mathcal{C}([a, b])$  is the space  $\mathcal{M}([a, b])$  of regular countably additive measures on  $[a, b]$  having finite variation. The Lagrangian  $\Lambda : \mathbb{R}^n \times \mathcal{M}([a, b]) \rightarrow \mathbb{R}$  for problem (4.98) is defined as follows:

$$\Lambda(x, \mu) = c(x) + \int_a^b \Gamma(x)(\eta) d\mu(\eta). \quad (4.100)$$

The necessary optimality conditions for problem (4.98) have the form, *There exists a measure  $\hat{\mu} \in \mathcal{M}_+([a, b])$  such that*

$$-\nabla_x \Lambda(\hat{x}, \hat{\mu}) \in \mathcal{N}_{\mathcal{X}}(\hat{x}), \quad (4.101)$$

$$\int_a^b \Gamma(\hat{x})(\eta) d\hat{\mu}(\eta) = 0. \quad (4.102)$$

Using Lemma 4.85, we obtain the equation for all  $x$ :

$$\begin{aligned} \int_a^b \Gamma(x)(\eta) d\hat{\mu}(\eta) &= \int_a^b \left( \Pr\{g(x, z) \geq \eta\} - \Pr\{Y \geq \eta\} \right) d\hat{\mu}(\eta) \\ &= \mathbb{E}[\hat{u}(g(x, Z))] - \mathbb{E}[\hat{u}(Y)], \end{aligned}$$

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 (4.102) is equivalent to (4.96). Moreover,

$$\Lambda(x, \mu) = L(x, u),$$

and, therefore, (4.101) is equivalent to (4.95).  $\square$

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. 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 4.88.** *Under the assumptions of Theorem 4.87 there exist piecewise constant utility function  $w(\cdot) \in \mathcal{U}$  satisfying the necessary optimality conditions (4.95)–(4.96). Moreover, the function  $w(\cdot)$  has at most  $n + 2$  jump points: there exist numbers  $\eta_i \in [a, b]$ ,  $i = 1, \dots, k$ , such that the function  $w(\cdot)$  is constant on the intervals  $(-\infty, \eta_1]$ ,  $(\eta_1, \eta_2], \dots, (\eta_k, \infty)$ , and  $0 \leq k \leq n + 2$ .*

**Proof.** Consider the mapping  $\Gamma$  defined by (4.97). 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 (4.101) can be rewritten as follows:

$$-\nabla_x c(\hat{x}) - \int_a^b \nabla_x \Gamma(\hat{x})(\eta) d\hat{\mu}(\eta) \in \mathcal{N}_X(\hat{x}).$$

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

$$\int_a^b \nabla_x \Gamma(\hat{x})(\eta) d\mu(\eta) = g. \quad (4.103)$$

At the optimal solution  $\hat{x}$  we have  $\Gamma(\hat{x})(\cdot) \leq 0$  and  $\hat{\mu} \geq 0$ . Therefore, the complementarity condition (4.102) can be equivalently expressed as the equation

$$\int_a^b \Gamma(\hat{x})(\eta) d\mu(\eta) = 0. \quad (4.104)$$

Every nonnegative solution  $\mu$  of (4.103)–(4.104) can be used as the Lagrange multiplier satisfying conditions (4.101)–(4.102) at  $\hat{x}$ . Define

$$a = \int_a^b d\hat{\mu}(\eta).$$

We can add to (4.103)–(4.104) the condition

$$\int_a^b d\mu(\eta) = a. \quad (4.105)$$

The system of three equations (4.103)–(4.105) 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 (4.103)–(4.104) are equivalent to (4.101)–(4.102).

Now, consider the case of  $\hat{\mu} \not\equiv 0$ . In this case, we have  $a > 0$ . Normalizing by  $a$ , we notice that (4.103)–(4.105) are equivalent to the following inclusion:

$$\begin{bmatrix} g/a \\ 0 \end{bmatrix} \in \text{conv} \left\{ \begin{bmatrix} \nabla_x \Gamma(\hat{x})(\eta) \\ \Gamma(\hat{x})(\eta) \end{bmatrix} : \eta \in [a, b] \right\} \subset \mathbb{R}^{n+1}.$$

By Carathéodory's theorem, there exist numbers  $\eta_i \in [a, b]$ , and  $\alpha_i \geq 0$ ,  $i = 1, \dots, k$ , such that

$$\begin{aligned} \begin{bmatrix} g/a \\ 0 \end{bmatrix} &= \sum_{i=1}^k \alpha_i \begin{bmatrix} \nabla_x \Gamma(\hat{x})(\eta_i) \\ \Gamma(\hat{x})(\eta_i) \end{bmatrix}, \\ \sum_{i=1}^k \alpha_i &= 1, \end{aligned}$$

and

$$1 \leq k \leq n + 2.$$

We define atomic measure  $\nu$  having atoms of mass  $c\alpha_i$  at points  $\eta_i$ ,  $i = 1, \dots, k$ . It satisfies (4.103)–(4.104):

$$\begin{aligned} \int_a^b \nabla_x \Gamma(\hat{x})(\eta) d\nu(\eta) &= \sum_{i=1}^k \nabla_x \Gamma(\hat{x})(\eta_i) c\alpha_i = g, \\ \int_a^b \Gamma(\hat{x})(\eta) d\nu(\eta) &= \sum_{i=1}^k \Gamma(\hat{x})(\eta_i) c\alpha_i = 0. \end{aligned}$$

Recall that (4.103)–(4.104) are equivalent to (4.101)–(4.102). Now, applying Lemma 4.85, 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 true.  $\square$



It follows from Theorem 4.88 that if the dominance constraint is active, then there exist at least one and at most  $n + 2$  *target values*  $\eta_i$  and *target probabilities*  $v_i = \Pr\{Y \geq \eta_i\}$ ,  $i = 1, \dots, k$ , which are critical for problem (4.9). They define a relaxation of (4.9) involving finitely many probabilistic constraints:

$$\begin{aligned} \text{Min}_x \quad & c(x) \\ \text{s.t.} \quad & \Pr\{g(x, Z) \geq \eta_i\} \geq v_i, \quad i = 1, \dots, k, \\ & x \in \mathcal{X}. \end{aligned}$$

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

$$\eta_1 < \eta_2 < \dots < \eta_k$$

and the corresponding probabilities by  $p_i$ ,  $i = 1, \dots, k$ . Then the dominance constraint is equivalent to

$$\Pr\{g(x, Z) \geq \eta_i\} \geq \sum_{j=i}^k p_j, \quad i = 1, \dots, k.$$

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  $\alpha$ -concave probability distribution.

**Theorem 4.89.** *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 (4.95)–(4.96) are satisfied. If the function  $c$  is convex, the function  $g$  satisfies the concavity assumptions above and the variable  $Z$  has an  $\alpha$ -concave probability distribution, then  $\hat{x}$  is an optimal solution of problem (4.9).*

**Proof.** By virtue of Theorem 4.43, the feasible set of problem (4.98) is convex and closed.

Let the operator  $\Gamma$  and the cone  $K$  be defined as in the proof of Theorem 4.87. Using Lemma 4.85, we observe that optimality conditions (4.101)–(4.102) for problem (4.98) 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  $\Gamma$  is differentiable, we have

$$\frac{1}{\tau} [\Gamma(\hat{x} + \tau d) - \Gamma(\hat{x})] \rightarrow \nabla_x \Gamma(\hat{x})(d) \quad \text{whenever } \tau \downarrow 0.$$

This implies that

$$\nabla_x \Gamma(\hat{x})(d) \in \mathcal{T}_K(\Gamma(\hat{x})),$$

where  $\mathcal{T}_K(\gamma)$  denotes the tangent cone to  $K$  at  $\gamma$ . Since

$$\mathcal{T}_K(\gamma) = K + \{t\gamma : t \in \mathbb{R}\},$$

there exists  $t \in \mathbb{R}$  such that

$$\nabla_x \Gamma(\hat{x})(d) + t\Gamma(\hat{x}) \in K. \quad (4.106)$$

Condition (4.101) 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

$$\nabla_x c(\hat{x})(d) + \int_a^b (\nabla_x \Gamma(\hat{x})(\eta))(d) d\mu(\eta) \geq 0. \quad (4.107)$$

Condition (4.102), relation (4.106), and the nonnegativity of  $\mu$  imply that

$$\int_a^b (\nabla_x \Gamma(\hat{x})(\eta))(d) d\mu(\eta) = \int_a^b \left[ (\nabla_x \Gamma(\hat{x})(\eta))(d) + t(\Gamma(\hat{x}))(\eta) \right] d\mu(\eta) \leq 0.$$

Substituting into (4.107) we conclude that

$$d^\top \nabla_x c(\hat{x}) \geq 0$$

for every feasible direction  $d$  at  $\hat{x}$ . By the convexity of  $c$ , for every feasible point  $x$  we obtain the inequality

$$c(x) \geq c(\hat{x}) + d^\top \nabla_x c(\hat{x}) \geq c(\hat{x}),$$

as stated.  $\square$

## Exercises

4.1. Are the following density functions  $\alpha$ -concave and do they define a  $\gamma$ -concave probability measure? What are  $\alpha$  and  $\gamma$ ?

- (a) If the  $m$ -dimensional random vector  $Z$  has the normal distribution with expected value  $\mu = 0$  and covariance matrix  $\Sigma$ , the random variable  $Y$  is independent of  $Z$  and has the  $\chi_k^2$  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:

$$\theta_m(x) = \frac{\Gamma(\frac{m+k}{2})}{\Gamma(\frac{k}{2})\sqrt{(2\pi)^m \det(\Sigma)}} \left(1 + \frac{1}{k} x^\top \Sigma^{\frac{1}{2}} x\right)^{-(m+k)/2}.$$

If  $m = k = 1$ , then this function reduces to the well-known univariate Cauchy density

$$\theta_1(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad -\infty < x < \infty.$$

- (b) The density function of the  $m$ -dimensional  $F$ -distribution with parameters  $n_0, \dots, n_m$ , and  $n = \sum_{i=1}^m n_i$ , is defined as follows:

$$\theta(x) = c \prod_{i=1}^m x_i^{n_i/2-1} \left( n_0 + \sum_{i=1}^m n_i x_i \right)^{-n/2}, \quad x_i \geq 0, \quad i = 1, \dots, m,$$

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  $\Sigma$ . 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  $s_1 \geq q$  and  $s_2 \geq q$ . 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  $x_i > \Theta_i$ ,  $i = 1, \dots, s$ , and  $\theta(x) = 0$  otherwise. Here  $\Theta_i$ ,  $i = 1, \dots, s$  are positive constants.

- 4.2. Assume that  $P$  is an  $\alpha$ -concave probability distribution and  $A \subset \mathbb{R}^n$  is a convex set. Prove that the function  $f(x) = P(A+x)$  is  $\alpha$ -concave.

- 4.3. Prove that if  $\theta : \mathbb{R} \rightarrow \mathbb{R}$  is a log-concave probability density function, then the functions

$$F(x) = \int_{t \leq x} \theta(t) dt \quad \text{and} \quad \bar{F}(x) = 1 - F(x)$$

are log-concave as well.

- 4.4. Check that the binomial, the Poisson, the geometric, and the hypergeometric one-dimensional 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  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$ , respectively. We define  $Y_1 = \min\{Z_1, Z_3\}$  and  $Y_2 = \min\{Z_2, Z_3\}$ . Describe  $G(\eta_1, \eta_2) = P(Y_1 \geq \eta_1, Y_2 \geq \eta_2)$  for nonnegative scalars  $\eta_1$  and  $\eta_2$  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  $\chi^2$ -random variable with one degree of freedom, and  $A$  be an  $n \times n$  positive definite matrix. Is the set

$$\left\{x \in \mathbb{R}^n : \Pr(Z - \sqrt{(x^T A x) W} \geq 0) \geq 0.9\right\}$$

convex?

- 4.7. If  $Y$  is an  $m$ -dimensional random vector with a log-normal distribution, and  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is such that each component  $g_i$  is a concave function, show that the set

$$C = \left\{x \in \mathbb{R}^n : \Pr(g(x) \geq Y) \geq 0.9\right\}$$

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:

$$\begin{aligned} & \text{Min}_x c^T x \\ & \text{s.t. } \Pr\{g_i(x) \geq Y_i, i = 1, 2\} \geq 0.9, \\ & \quad x \geq 0. \end{aligned}$$

Here  $c \in \mathbb{R}^n$ ,  $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $i = 1, 2$ , is a concave function, and  $Y_1$  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:

$$\left\{x \in \mathbb{R}^3 : \Pr(x_1^2 + x_2^2 + Yx_2 + x_2x_3 + Yx_3 \leq Z) \geq 0.9\right\}.$$

- 4.10. Assume that the random variable  $Z$  is uniformly distributed in the interval  $[-1, 1]$  and  $e = (1, \dots, 1)^T$ . Prove that the following set is convex:

$$\left\{x \in \mathbb{R}^n : \Pr(\exp(x^T y) \geq (e^T y)Z, \quad \forall y \in \mathbb{R}^n : \|y\| \leq 1) \geq 0.95\right\}.$$

- 4.11. Let  $Z$  be a two-dimensional random vector with Dirichlet distribution. Show that the following set is convex:

$$\left\{x \in \mathbb{R}^2 : \Pr(\min(x_1 + 2x_2 + Z_1, x_1 Z_2 - x_1^2 - Z_2^2) \geq y) \geq e^{-y} \quad \forall y \in [\frac{1}{4}, 4]\right\}.$$

- 4.12. Let  $Z$  be an  $n$ -dimensional random vector uniformly distributed on a set  $A$ . Check whether the set

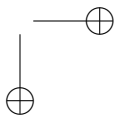
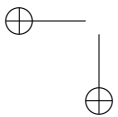
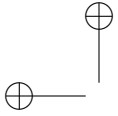
$$\left\{x \in \mathbb{R}^n : \Pr(x^T Z \leq 1) \geq 0.95\right\}$$

is convex for the following cases:

- (a)  $A = \{z \in \mathbb{R}^n : \|z\| \leq 1\}$ .
- (b)  $A = \{z \in \mathbb{R}^n : 0 \leq z_i \leq i, i = 1, \dots, m\}$ .
- (c)  $A = \{z \in \mathbb{R}^n : Tz \leq 0, -1 \leq z_i \leq 1, i = 1, \dots, m\}$ , where  $T$  is an  $(n-1) \times n$  matrix of form

$$T = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{pmatrix}.$$

- 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  $p$ -efficient points of  $F_Z$  for  $p = 0.8$ .



## Chapter 5

# Statistical Inference

*Alexander Shapiro*

## 5.1 Statistical Properties of Sample Average Approximation Estimators

Consider the following stochastic programming problem:

$$\text{Min}_{x \in X} \{f(x) := \mathbb{E}[F(x, \xi)]\}. \quad (5.1)$$

Here  $X$  is a nonempty closed subset of  $\mathbb{R}^n$ ,  $\xi$  is a random vector whose probability distribution  $P$  is supported on a set  $\Xi \subset \mathbb{R}^d$ , and  $F : X \times \Xi \rightarrow \mathbb{R}$ . 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  $x \in X$ . This implies, of course, that for every  $x \in X$  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  $\xi^1, \dots, \xi^N$  of  $N$  realizations of the random vector  $\xi$ . This random sample can be viewed as historical data of  $N$  observations of  $\xi$ , or it can be generated in the computer by Monte Carlo sampling techniques. For any  $x \in X$  we can estimate the expected value  $f(x)$  by averaging values  $F(x, \xi^j)$ ,  $j = 1, \dots, N$ . This leads to the so-called sample average approximation (SAA)

$$\text{Min}_{x \in X} \left\{ \hat{f}_N(x) := \frac{1}{N} \sum_{j=1}^N F(x, \xi^j) \right\} \quad (5.2)$$

of the “true” problem (5.1). Let us observe that we can write the sample average function as the expectation

$$\hat{f}_N(x) = \mathbb{E}_{P_N}[F(x, \xi)] \quad (5.3)$$

taken with respect to the *empirical distribution*<sup>18</sup> (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  $\xi^1, \dots, \xi^N$ , each taken with probability  $1/N$ .

As with data vector  $\xi$ , the sample  $\xi^1, \dots, \xi^N$  can be considered from two points of view: as a sequence of random vectors or as a particular realization of that sequence. Which 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  $\xi^1, \dots, \xi^N$  each taken with probability  $1/N$ . We always assume that each random vector  $\xi^j$  in the sample has the same (marginal) distribution  $P$  as the data vector  $\xi$ . If, moreover, each  $\xi^j$ ,  $j = 1, \dots, N$ , 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,  $\hat{f}_N(x)$  converges pointwise w.p. 1 to  $f(x)$  as  $N \rightarrow \infty$ . 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  $N \rightarrow \infty$ . We denote by  $\vartheta^*$  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  $\hat{f}_N(x)$  as defined on a common probability space  $(\Omega, \mathcal{F}, P)$ . For example, in the case of the iid sample, a standard construction is to consider the set  $\Omega := \Xi^\infty$  of sequences  $\{(\xi_1, \dots)\}_{\xi_i \in \Xi, i \in \mathbb{N}}$ , 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  $\xi$ . Then  $\hat{f}_N(x) = \hat{f}_N(x, \omega)$  is also a Carathéodory function and hence is a random lower semicontinuous function. It follows (see section 7.2.3 and Theorem 7.37 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 7.34). This takes care of the measurability questions.

Next we discuss statistical properties of the SAA estimators  $\hat{\vartheta}_N$  and  $\hat{S}_N$ . Let us make the following useful observation.

**Proposition 5.1.** *Let  $f : X \rightarrow \mathbb{R}$  and  $f_N : X \rightarrow \mathbb{R}$  be a sequence of (deterministic) real valued functions. Then the following two properties are equivalent: (i) for any  $\bar{x} \in X$  and any sequence  $\{x_N\} \subset X$  converging to  $\bar{x}$  it follows that  $f_N(x_N)$  converges to  $f(\bar{x})$ , and (ii) the function  $f(\cdot)$  is continuous on  $X$  and  $f_N(\cdot)$  converges to  $f(\cdot)$  uniformly on any compact subset of  $X$ .*

<sup>18</sup>Recall that  $\Delta(\xi)$  denotes measure of mass one at the point  $\xi$ .



**Proof.** Suppose that property (i) holds. Consider a point  $\bar{x} \in X$ , a sequence  $\{x_N\} \subset X$  converging to  $\bar{x}$  and a number  $\varepsilon > 0$ . By taking a sequence with each element equal  $x_1$ , we have by (i) that  $f_N(x_1) \rightarrow f(x_1)$ . Therefore, there exists  $N_1$  such that  $|f_{N_1}(x_1) - f(x_1)| < \varepsilon/2$ . 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:  $x'_i = x_1, i = 1, \dots, N_1, x'_i = x_2, i = N_1 + 1, \dots, N_2$ , 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  $f(x_k) \rightarrow f(\bar{x})$  and hence  $f(\cdot)$  is continuous at  $\bar{x}$ .

Now let  $C$  be a compact subset of  $X$ . Arguing by 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  $|f_N(x_N) - f(x_N)| \geq \varepsilon$  for all  $N$ . Since  $C$  is compact, we can assume that  $\{x_N\}$  converges to a point  $\bar{x} \in C$ . We have

$$|f_N(x_N) - f(x_N)| \leq |f_N(x_N) - f(\bar{x})| + |f(x_N) - f(\bar{x})|. \quad (5.4)$$

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 than  $\varepsilon/2$  for  $N$  large enough. This gives a designed contradiction.

Conversely, suppose that property (ii) holds. Consider a sequence  $\{x_N\} \subset X$  converging to a point  $\bar{x} \in X$ . We can assume that this sequence is contained in a compact subset of  $X$ . By employing the inequality

$$|f_N(x_N) - f(\bar{x})| \leq |f_N(x_N) - f(x_N)| + |f(x_N) - f(\bar{x})| \quad (5.5)$$

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.  $\square$

### 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  $\theta$  is *consistent* if  $\hat{\theta}_N$  converges w.p. 1 to  $\theta$  as  $N \rightarrow \infty$ . Let us consider first consistency of the SAA estimator of the optimal value. We have that for any fixed  $x \in X$ ,  $\hat{\vartheta}_N \leq \hat{f}_N(x)$ , and hence if the pointwise LLN holds, then

$$\limsup_{N \rightarrow \infty} \hat{\vartheta}_N \leq \lim_{N \rightarrow \infty} \hat{f}_N(x) = f(x) \text{ w.p. 1.}$$

It follows that if the pointwise LLN holds, then

$$\limsup_{N \rightarrow \infty} \hat{\vartheta}_N \leq \vartheta^* \text{ w.p. 1.} \quad (5.6)$$

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  $N \rightarrow \infty$ , uniformly on  $X$ . Then  $\hat{\vartheta}_N$  converges to  $\vartheta^*$  w.p. 1 as  $N \rightarrow \infty$ .*

**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  $N \geq N^*$ :

$$\sup_{x \in X} |\hat{f}_N(x, \omega) - f(x)| \leq \varepsilon. \quad (5.7)$$

It follows then that  $|\hat{\vartheta}_N(\omega) - \vartheta^*| \leq \varepsilon$  for all  $N \geq N^*$ , which completes the proof.  $\square$

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 (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  $N \rightarrow \infty$ , uniformly in  $x \in C$ , and (iv) w.p. 1 for  $N$  large enough the set  $\hat{S}_N$  is nonempty and  $\hat{S}_N \subset C$ . Then  $\hat{\vartheta}_N \rightarrow \vartheta^*$  and  $\mathbb{D}(\hat{S}_N, S) \rightarrow 0$  w.p. 1 as  $N \rightarrow \infty$ .*

**Proof.** Assumptions (i) and (iv) imply that both the true and the SAA problem can be restricted to the set  $X \cap C$ . Therefore we can assume without loss of generality that the set  $X$  is compact. The assertion that  $\hat{\vartheta}_N \rightarrow \vartheta^*$  w.p. 1 follows by Proposition 5.2. It suffices to show now that  $\mathbb{D}(\hat{S}_N(\omega), S) \rightarrow 0$  for every  $\omega \in \Omega$  such that  $\hat{\vartheta}_N(\omega) \rightarrow \vartheta^*$  and assumptions (iii) and (iv) hold. This is basically a deterministic result; therefore, we omit  $\omega$  for the sake of notational convenience.

We argue now by a contradiction. Suppose that  $\mathbb{D}(\hat{S}_N, S) \not\rightarrow 0$ . Since  $X$  is compact, by passing to a subsequence if necessary, we can assume that there exists  $\hat{x}_N \in \hat{S}_N$  such that  $\text{dist}(\hat{x}_N, S) \geq \varepsilon$  for some  $\varepsilon > 0$  and that  $\hat{x}_N$  tends to a point  $x^* \in X$ . It follows that  $x^* \notin S$  and hence  $f(x^*) > \vartheta^*$ . Moreover,  $\hat{\vartheta}_N = \hat{f}_N(\hat{x}_N)$  and

$$\hat{f}_N(\hat{x}_N) - f(x^*) = [\hat{f}_N(\hat{x}_N) - f(\hat{x}_N)] + [f(\hat{x}_N) - f(x^*)]. \quad (5.8)$$

The first term in the right-hand side of (5.8) tends to zero by 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.  $\square$

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  $\hat{f}_N(x_N) \rightarrow f(\bar{x})$  w.p. 1. Assumption (iv) in the above theorem holds, in particular, if the feasible set  $X$  is closed, the functions  $\hat{f}_N(x)$  are lower semicontinuous, and for some  $\alpha > \vartheta^*$  the level sets  $\{x \in X : \hat{f}_N(x) \leq \alpha\}$  are uniformly bounded w.p. 1. This condition is often referred to as the *inf-compactness condition*. Conditions ensuring the uniform convergence of  $\hat{f}_N(x)$  to  $f(x)$  (assumption (iii)) are given in Theorems 7.48 and 7.50, for example.

The assertion that  $\mathbb{D}(\hat{S}_N, S) \rightarrow 0$  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  $\text{dist}(\hat{x}_N, S) \rightarrow 0$  w.p. 1. If, moreover,  $S = \{\bar{x}\}$  is a singleton, i.e., the true problem has unique optimal solution  $\bar{x}$ ,

### 5.1. Statistical Properties of Sample Average Approximation Estimators

159

then this means that  $\hat{x}_N \rightarrow \bar{x}$  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

$$\bar{F}(x, \xi) := F(x, \xi) + \mathbb{I}_X(x), \quad \bar{f}(x) := f(x) + \mathbb{I}_X(x), \quad \bar{f}_N(x) := \hat{f}_N(x) + \mathbb{I}_X(x), \quad (5.9)$$

i.e.,  $\bar{f}(x) = f(x)$  if  $x \in X$  and  $\bar{f}(x) = +\infty$  if  $x \notin X$ , and similarly for functions  $F(\cdot, \xi)$  and  $\hat{f}_N(\cdot)$ . Clearly  $\bar{f}(x) = \mathbb{E}[\bar{F}(x, \xi)]$  and  $\bar{f}_N(x) = N^{-1} \sum_{j=1}^N \bar{F}(x, \xi^j)$ . Note that if the set  $X$  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  $X$  is closed and convex, (iv) the expected value function  $f$  is lower semicontinuous and there exists a point  $\bar{x} \in X$  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, and (vi) the LLN holds pointwise. Then  $\hat{\vartheta}_N \rightarrow \vartheta^*$  and  $\mathbb{D}(\hat{S}_N, S) \rightarrow 0$  w.p. 1 as  $N \rightarrow \infty$ .*

**Proof.** Clearly we can restrict both the true and the SAA problem to the affine space generated by the convex set  $X$ . Relative to that affine space, the set  $X$  has a nonempty interior. Therefore, without loss of generality we can assume that the set  $X$  has a nonempty interior. Since it is assumed that  $f(x)$  possesses an optimal solution, we have that  $\vartheta^*$  is finite and hence  $f(x) \geq \vartheta^* > -\infty$  for all  $x \in X$ . Since  $f(x)$  is convex and is greater than  $-\infty$  on an open set (e.g., interior of  $X$ ), it follows that  $f(\cdot)$  is subdifferentiable at any point  $x \in \text{int}(X)$  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  $\bar{F}(x, \xi)$ . Since  $X$  is convex and closed, it follows that  $\bar{f}$  is convex and lower semicontinuous. Moreover, because of the assumption (iv) and since the interior of  $X$  is nonempty, we have that  $\text{dom } \bar{f}$  has a nonempty interior. By Theorem 7.49 it follows then that  $\bar{f}_N \xrightarrow{e} \bar{f}$  w.p. 1. Consider a compact set  $K$  with a nonempty interior and such that it does not contain a boundary point of  $\text{dom } \bar{f}$ , and  $\bar{f}(x)$  is finite valued on  $K$ . Since  $\text{dom } \bar{f}$  has a nonempty interior, such a set exists. Then it follows from  $\bar{f}_N \xrightarrow{e} \bar{f}$  that  $\bar{f}_N(\cdot)$  converge to  $\bar{f}(\cdot)$  uniformly on  $K$ , all w.p. 1 (see Theorem 7.27). It follows that w.p. 1 for  $N$  large enough the functions  $\bar{f}_N(x)$  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  $\tilde{S}_N$  of minimizers of  $\bar{f}_N(x)$  over  $C$ . Since  $C$  is nonempty and compact and  $\bar{f}_N(x)$  is lower semicontinuous and proper for  $N$  large enough, and because by the pointwise LLN we have that for any  $x \in S$ ,  $\bar{f}_N(x)$  is finite w.p. 1 for  $N$  large enough, the set  $\tilde{S}_N$  is nonempty w.p. 1 for  $N$  large enough. Let us show that  $\mathbb{D}(\tilde{S}_N, S) \rightarrow 0$  w.p. 1. Let  $\omega \in \Omega$  be such that  $\bar{f}_N(\cdot, \omega) \xrightarrow{e} \bar{f}(\cdot)$ . 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  $\bar{f}_N(x, \omega)$  over  $C$  such that  $\text{dist}(\tilde{x}_N, S) \geq \varepsilon$  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 7.26 that  $x^* \in \arg \min_{x \in C} \bar{f}(x)$ . Since  $\arg \min_{x \in C} \bar{f}(x) = S$ , we obtain a contradiction.

Now because of the convexity assumptions, any minimizer of  $\tilde{f}_N(x)$  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.  $\square$

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 7.42 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,  $f(x)$  from taking value  $+\infty$  at some points  $x \in X$ . 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  $f(x)$  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. Of course, it can be both. For example, in the case of two-stage programming it may happen that for some  $x \in X$  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  $\xi^j$  the corresponding second-stage problem will be infeasible, and hence  $\hat{f}_N(x) = +\infty$ . 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  $X$  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

$$\text{Min}_{x \in X_N} \hat{f}_N(x), \quad (5.10)$$

where  $X_N$  is a subset of  $\mathbb{R}^n$  depending on the sample and therefore is random. As before we denote by  $\hat{\vartheta}_N$  and  $\hat{S}_N$  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  $x_N \in X_N$  and  $x_N$  converges w.p. 1 to a point  $x$ , then  $x \in X$ .*
- (b) *For some point  $x \in S$  there exists a sequence  $x_N \in X_N$  such that  $x_N \rightarrow x$  w.p. 1.*

*Then  $\hat{\vartheta}_N \rightarrow \vartheta^*$  and  $\mathbb{D}(\hat{S}_N, S) \rightarrow 0$  w.p. 1 as  $N \rightarrow \infty$ .*

**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  $\hat{S}_N \subset X_N$ , we have that  $\hat{x}_N \in X_N$ , and hence it follows by condition (a) that  $x^* \in X$ . 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  $\liminf_{N \rightarrow \infty} \hat{\vartheta}_N \geq \vartheta^*$  w.p. 1. On the other hand, by condition (b), there exists a sequence  $x_N \in X_N$  converging to a point  $x \in S$  w.p. 1. Consequently,  $\hat{\vartheta}_N \leq \hat{f}_N(x_N) \rightarrow f(x) = \vartheta^*$  w.p. 1, and hence  $\limsup_{N \rightarrow \infty} \hat{\vartheta}_N \leq \vartheta^*$ . It follows that

## 5.1. Statistical Properties of Sample Average Approximation Estimators 161

$\hat{\vartheta}_N \rightarrow \vartheta^*$  w.p. 1. The remainder of the proof can be completed by the same arguments as in the proof of Theorem 5.3.  $\square$

The SAA problem (5.10) is convex if the functions  $\hat{f}_N(\cdot)$  and the sets  $X_N$  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  $X_N$ .

Suppose, for example, that the set  $X$  is defined by the constraints

$$X := \{x \in X_0 : g_i(x) \leq 0, i = 1, \dots, p\}, \quad (5.11)$$

where  $X_0$  is a nonempty closed subset of  $\mathbb{R}^n$  and the constraint functions are given as the expected value functions

$$g_i(x) := \mathbb{E}[G_i(x, \xi)], \quad i = 1, \dots, p, \quad (5.12)$$

with  $G_i(x, \xi)$ ,  $i = 1, \dots, p$ , being random lower semicontinuous functions. Then the set  $X$  can be estimated by

$$X_N := \{x \in X_0 : \hat{g}_{iN}(x) \leq 0, i = 1, \dots, p\}, \quad (5.13)$$

where

$$\hat{g}_{iN}(x) := \frac{1}{N} \sum_{j=1}^N G_i(x, \xi^j).$$

If for a given point  $x \in X_0$ , 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 5.** Let us note that the samples used in construction of the SAA functions  $\hat{f}_N$  and  $\hat{g}_{iN}$ ,  $i = 1, \dots, p$ , can be the same or can be different, independent of each other. That is, for random samples  $\xi^{i1}, \dots, \xi^{iN_i}$ , possibly of different sample sizes  $N_i$ ,  $i = 1, \dots, p$ , and independent of each other and of the random sample used in  $\hat{f}_N$ , the corresponding SAA functions are

$$\hat{g}_{iN_i}(x) := \frac{1}{N_i} \sum_{j=1}^{N_i} G_i(x, \xi^{ij}), \quad i = 1, \dots, p.$$

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 9 on page 173).

In order to ensure condition (b) of Theorem 5.5, one needs to impose a constraint qualification (on the true problem). Consider, for example,  $X := \{x \in \mathbb{R} : g(x) \leq 0\}$  with  $g(x) := x^2$ . Clearly  $X = \{0\}$ , while an arbitrary small perturbation of the function  $g(\cdot)$  can result in the corresponding set  $X_N$  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  $X_0$  is convex and for every  $\xi \in \Xi$  the functions  $G_i(\cdot, \xi)$  are convex, and hence the corresponding expected value functions  $g_i(\cdot)$ ,  $i = 1, \dots, p$ , 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  $x^* \in X_0$  such that  $g_i(x^*) < 0$ ,  $i = 1, \dots, p$ .

As another example, suppose that the feasible set is given by probabilistic (chance) constraints in the form

$$X = \{x \in \mathbb{R}^n : \Pr(C_i(x, \xi) \leq 0) \geq 1 - \alpha_i, i = 1, \dots, p\}, \quad (5.14)$$

where  $\alpha_i \in (0, 1)$  and  $C_i : \mathbb{R}^n \times \Xi \rightarrow \mathbb{R}$ ,  $i = 1, \dots, p$ , are Carathéodory functions. Of course, we have that<sup>19</sup>

$$\Pr(C_i(x, \xi) \leq 0) = \mathbb{E}[\mathbf{1}_{(-\infty, 0]}(C_i(x, \xi))]. \quad (5.15)$$

Consequently, we can write the above set  $X$  in the form (5.11)–(5.12) with  $X_0 := \mathbb{R}^n$  and

$$G_i(x, \xi) := 1 - \alpha_i - \mathbf{1}_{(-\infty, 0]}(C_i(x, \xi)). \quad (5.16)$$

The corresponding set  $X_N$  can be written as

$$X_N = \left\{x \in \mathbb{R}^n : \frac{1}{N} \sum_{j=1}^N \mathbf{1}_{(-\infty, 0]}(C_i(x, \xi^j)) \geq 1 - \alpha_i, i = 1, \dots, p\right\}. \quad (5.17)$$

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 “ $C_i(x, \xi^j) \leq 0$ ”,  $j = 1, \dots, N$ , 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  $X$  and  $i = 1, \dots, p$ , the event “ $C_i(x, \xi) = 0$ ” happens with probability zero, and hence  $G_i(\cdot, \xi)$  is continuous at  $x$  for a.e.  $\xi$ . By Theorem 7.48 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 6.** As pointed out in Remark 5, it is possible to use different, independent of each other, random samples  $\xi^{i1}, \dots, \xi^{iN_i}$ , possibly of different sample sizes  $N_i$ ,  $i = 1, \dots, p$ , for constructing the corresponding SAA functions. That is, constraints  $\Pr(C_i(x, \xi) > 0) \leq \alpha_i$  are approximated by

$$\frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{1}_{(0, \infty)}(C_i(x, \xi^{ij})) \leq \alpha_i, i = 1, \dots, p. \quad (5.18)$$

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.

<sup>19</sup>Recall that  $\mathbf{1}_{(-\infty, 0]}(t) = 1$  if  $t \leq 0$  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  $x \in X$ . 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  $\sigma^2(x) := \mathbb{V}\text{ar}[F(x, \xi)]$  is supposed to be finite. Moreover, by the CLT we have that

$$N^{1/2} [\hat{f}_N(x) - f(x)] \xrightarrow{\mathcal{D}} Y_x, \quad (5.19)$$

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  $f(x)$ :

$$\left[ \hat{f}_N(x) - \frac{z_{\alpha/2} \hat{\sigma}(x)}{\sqrt{N}}, \hat{f}_N(x) + \frac{z_{\alpha/2} \hat{\sigma}(x)}{\sqrt{N}} \right], \quad (5.20)$$

where  $z_{\alpha/2} := \Phi^{-1}(1 - \alpha/2)$  and<sup>20</sup>

$$\hat{\sigma}^2(x) := \frac{1}{N-1} \sum_{j=1}^N [F(x, \xi^j) - \hat{f}_N(x)]^2 \quad (5.21)$$

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  $x' \in X$  the inequality  $\hat{f}_N(x') \geq \inf_{x \in X} \hat{f}_N(x)$  holds. By taking the expected value of both sides of this inequality and minimizing the left-hand side over all  $x' \in X$ , we obtain

$$\inf_{x \in X} \mathbb{E} [\hat{f}_N(x)] \geq \mathbb{E} \left[ \inf_{x \in X} \hat{f}_N(x) \right]. \quad (5.22)$$

Note that the inequality (5.22) holds even if  $f(x) = +\infty$  or  $f(x) = -\infty$  for some  $x \in X$ . Since  $\mathbb{E}[\hat{f}_N(x)] = f(x)$ , it follows that  $\vartheta^* \geq \mathbb{E}[\hat{\vartheta}_N]$ . In fact, typically,  $\mathbb{E}[\hat{\vartheta}_N]$  is strictly less than  $\vartheta^*$ , i.e.,  $\hat{\vartheta}_N$  is a downward *biased* estimator of  $\vartheta^*$ . 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  $N \in \mathbb{N}$ .*

<sup>20</sup>Here  $\Phi(\cdot)$  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

$$\begin{aligned} \mathbb{E}[\hat{\vartheta}_{N+1}] &= \mathbb{E} \left[ \inf_{x \in X} \hat{f}_{N+1}(x) \right] \\ &= \mathbb{E} \left[ \inf_{x \in X} \frac{1}{N+1} \sum_{i=1}^{N+1} \left( \frac{1}{N} \sum_{j \neq i} F(x, \xi^j) \right) \right] \\ &\geq \mathbb{E} \left[ \frac{1}{N+1} \sum_{i=1}^{N+1} \left( \inf_{x \in X} \frac{1}{N} \sum_{j \neq i} F(x, \xi^j) \right) \right] \\ &= \frac{1}{N+1} \sum_{i=1}^{N+1} \mathbb{E} \left[ \inf_{x \in X} \frac{1}{N} \sum_{j \neq i} F(x, \xi^j) \right] \\ &= \frac{1}{N+1} \sum_{i=1}^{N+1} \mathbb{E}[\hat{\vartheta}_N] = \mathbb{E}[\hat{\vartheta}_N], \end{aligned}$$

which completes the proof.  $\square$

### First Order Asymptotics of the SAA Optimal Value

We use the following assumptions about the integrand  $F$ :

**(A1)** For some point  $\tilde{x} \in X$  the expectation  $\mathbb{E}[F(\tilde{x}, \xi)^2]$  is finite.

**(A2)** There exists a measurable function  $C : \Xi \rightarrow \mathbb{R}_+$  such that  $\mathbb{E}[C(\xi)^2]$  is finite and

$$|F(x, \xi) - F(x', \xi)| \leq C(\xi) \|x - x'\| \quad (5.23)$$

for all  $x, x' \in X$  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  $x \in X$ . Moreover, it follows from (5.23) that

$$|f(x) - f(x')| \leq \kappa \|x - x'\|, \quad \forall x, x' \in X,$$

where  $\kappa := \mathbb{E}[C(\xi)]$ , and hence  $f(x)$  is Lipschitz continuous on  $X$ . If  $X$  is compact, we have then that the set  $S$ , of minimizers of  $f(x)$  over  $X$ , is nonempty.

Let  $Y_x$  be random variables defined in (5.19). These variables depend on  $x \in X$  and we also use notation  $Y(x) = Y_x$ . By the (multivariate) CLT we have that for any finite set  $\{x_1, \dots, x_m\} \subset X$ , the random vector  $(Y(x_1), \dots, Y(x_m))$  has a multivariate normal distribution with zero mean and the same covariance matrix as the covariance matrix of  $(F(x_1, \xi), \dots, F(x_m, \xi))$ . Moreover, by assumptions (A1) and (A2), compactness of  $X$ , 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*<sup>21</sup> of  $C(X)$ . This is a so-called functional CLT (see, e.g., Araujo and Giné [4, Corollary 7.17]).

<sup>21</sup>Recall that  $C(X)$  denotes the space of continuous functions equipped with the sup-norm. A random element of  $C(X)$  is a mapping  $Y : \Omega \rightarrow C(X)$  from a probability space  $(\Omega, \mathcal{F}, P)$  into  $C(X)$  which is measurable with respect to the Borel sigma algebra of  $C(X)$ , 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  $X$  is compact, and assumptions (A1) and (A2) are satisfied. Then the following holds:*

$$\hat{\vartheta}_N = \inf_{x \in S} \hat{f}_N(x) + o_p(N^{-1/2}), \quad (5.24)$$

$$N^{1/2} \left( \hat{\vartheta}_N - \vartheta^* \right) \xrightarrow{\mathcal{D}} \inf_{x \in S} Y(x). \quad (5.25)$$

If, moreover,  $S = \{\bar{x}\}$  is a singleton, then

$$N^{1/2} \left( \hat{\vartheta}_N - \vartheta^* \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(\bar{x})). \quad (5.26)$$

**Proof.** Proof is based on the functional CLT and the Delta theorem (Theorem 7.59). Consider Banach space  $C(X)$  of continuous functions  $\psi : X \rightarrow \mathbb{R}$  equipped with the sup-norm  $\|\psi\| := \sup_{x \in X} |\psi(x)|$ . Define the min-value function  $V(\psi) := \inf_{x \in X} \psi(x)$ . Since  $X$  is compact, the function  $V : C(X) \rightarrow \mathbb{R}$  is real valued and measurable (with respect to the Borel sigma algebra of  $C(X)$ ). Moreover, it is not difficult to see that  $|V(\psi_1) - V(\psi_2)| \leq \|\psi_1 - \psi_2\|$  for any  $\psi_1, \psi_2 \in C(X)$ , i.e.,  $V(\cdot)$  is Lipschitz continuous with Lipschitz constant one. By the Danskin theorem (Theorem 7.21),  $V(\cdot)$  is directionally differentiable at any  $\mu \in C(X)$  and

$$V'_\mu(\delta) = \inf_{x \in \bar{X}(\mu)} \delta(x), \quad \forall \delta \in C(X), \quad (5.27)$$

where  $\bar{X}(\mu) := \arg \min_{x \in X} \mu(x)$ . Since  $V(\cdot)$  is Lipschitz continuous, directional differentiability in the Hadamard sense follows (see Proposition 7.57). As 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  $C(X)$ . Noting that  $\hat{\vartheta}_N = V(\hat{f}_N)$ ,  $\vartheta^* = V(f)$ , and  $\bar{X}(f) = S$ , and by applying the Delta theorem to the min-function  $V(\cdot)$  at  $\mu := f$  and using (5.27), we obtain (5.25) and that

$$\hat{\vartheta}_N - \vartheta^* = \inf_{x \in S} [\hat{f}_N(x) - f(x)] + o_p(N^{-1/2}). \quad (5.28)$$

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).  $\square$

Under mild additional conditions (see Remark 32 on page 382), 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  $N \rightarrow \infty$ , that is,

$$\mathbb{E}[\hat{\vartheta}_N] - \vartheta^* = N^{-1/2}\mathbb{E}\left[\inf_{x \in S} Y(x)\right] + o(N^{-1/2}). \quad (5.29)$$

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 7 on page 168.

### 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, Theorem 7.62, 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  $X \subset \text{int}(U)$ , and we work with the space  $W^{1,\infty}(U)$  of Lipschitz continuous functions  $\psi : U \rightarrow \mathbb{R}$  equipped with the norm

$$\|\psi\|_{1,U} := \sup_{x \in U} |\psi(x)| + \sup_{x \in U'} \|\nabla \psi(x)\|, \quad (5.30)$$

where  $U' \subset \text{int}(U)$  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  $x \in X$ , and is twice continuously differentiable at  $\bar{x}$ .
- (S2) The set  $X$  is second order regular at  $\bar{x}$ .
- (S3) The quadratic growth condition (7.70) 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  $V : W^{1,\infty}(U) \rightarrow \mathbb{R}$  defined as

$$V(\psi) := \inf_{x \in X} \psi(x), \quad \psi \in W^{1,\infty}(U).$$

By Theorem 7.23, under 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  $\delta \in \mathcal{K}$ :

$$V_f''(\delta) = \inf_{h \in C(\bar{x})} \{2h^T \nabla \delta(\bar{x}) + h^T \nabla^2 f(\bar{x}) h - s(-\nabla f(\bar{x}), \mathcal{T}_X^2(\bar{x}, h))\}. \quad (5.31)$$

Here  $C(\bar{x})$  is the critical cone of the true problem,  $\mathcal{T}_X^2(\bar{x}, h)$  is the second order tangent set to  $X$  at  $\bar{x}$  and  $s(\cdot, A)$  denotes the support function of set  $A$ . (See page 386 for the definition of second order directional derivatives.)

Moreover, suppose that the set  $X$  is given in the form

$$X := \{x \in \mathbb{R}^n : G(x) \in K\}, \quad (5.32)$$

where  $G : \mathbb{R}^n \rightarrow \mathbb{R}^m$  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 (7.84)), which results in the following formula for the second order directional derivative in a direction  $\delta \in \mathcal{K}$ :

$$V_f''(\delta) = \inf_{h \in C(\bar{x})} \sup_{\lambda \in \Lambda(\bar{x})} \{2h^\top \nabla \delta(\bar{x}) + h^\top \nabla_{xx}^2 L(\bar{x}, \lambda)h - s(\lambda, \mathfrak{T}(h))\}. \quad (5.33)$$

Here

$$\mathfrak{T}(h) := \mathcal{T}_K^2(G(\bar{x}), [\nabla G(\bar{x})]h), \quad (5.34)$$

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*

$$\hat{\vartheta}_N = \hat{f}_N(\bar{x}) + \frac{1}{2} V_f''(\hat{f}_N - f) + o_p(N^{-1}), \quad (5.35)$$

and

$$N[\hat{\vartheta}_N - \hat{f}_N(\bar{x})] \xrightarrow{\mathcal{D}} \frac{1}{2} V_f''(Y). \quad (5.36)$$

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

$$N^{1/2}(\hat{x}_N - \bar{x}) \xrightarrow{\mathcal{D}} \bar{h}(Y). \quad (5.37)$$

**Proof.** By the second order Delta theorem, Theorem 7.62, 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

$$N[\hat{\vartheta}_N - \vartheta^* - V_f'(\hat{f}_N - f)] \xrightarrow{\mathcal{D}} \frac{1}{2} V_f''(Y).$$

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  $\mathfrak{x} : W^{1,\infty}(U) \rightarrow \mathbb{R}^n$  such that

$$\mathfrak{x}(\psi) \in \arg \min_{x \in X} \psi(x), \quad \psi \in W^{1,\infty}(U).$$

We have that  $\mathfrak{x}(f) = \bar{x}$ , and by (7.82) of Theorem 7.23 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 the Delta theorem, Theorem 7.61, this completes the proof of (5.37).  $\square$

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  $W^{1,\infty}(X)$ . 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

$$\Sigma = \mathbb{E}[(\nabla F(\bar{x}, \xi) - \nabla f(\bar{x}))(\nabla F(\bar{x}, \xi) - \nabla f(\bar{x}))^\top], \quad (5.38)$$

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 7.44).

Let  $Z$  be a random vector having normal distribution,  $Z \sim \mathcal{N}(0, \Sigma)$ , with covariance matrix  $\Sigma$  defined in (5.38), and let the set  $X$  be given in the form (5.32). Then by the above discussion and formula (5.33), we have that under appropriate regularity conditions,

$$N[\hat{\vartheta}_N - \hat{f}_N(\bar{x})] \xrightarrow{\mathcal{D}} \frac{1}{2}\mathfrak{v}(Z), \quad (5.39)$$

where  $\mathfrak{v}(Z)$  is the optimal value of the problem

$$\text{Min}_{h \in C(\bar{x})} \sup_{\lambda \in \Lambda(\bar{x})} \{2h^\top Z + h^\top \nabla_{xx}^2 L(\bar{x}, \lambda)h - s(\lambda, \mathfrak{T}(h))\}, \quad (5.40)$$

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  $\hat{h} = \mathfrak{h}(Z)$ , then

$$N^{1/2}(\hat{x}_N - \bar{x}) \xrightarrow{\mathcal{D}} \mathfrak{h}(Z). \quad (5.41)$$

Recall also that if the cone  $K$  is polyhedral, then the curvature term  $s(\lambda, \mathfrak{T}(h))$  vanishes.

**Remark 7.** 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; to see this, just take  $h = 0$  in (5.40).

As an example, consider the case where the set  $X$  is defined by a finite number of constraints:

$$X := \{x \in \mathbb{R}^n : g_i(x) = 0, i = 1, \dots, q, g_i(x) \leq 0, i = q + 1, \dots, p\} \quad (5.42)$$

with the functions  $g_i(x)$ ,  $i = 1, \dots, p$ , being twice continuously differentiable. This is a particular form of (5.32) with  $G(x) := (g_1(x), \dots, g_p(x))$  and  $K := \{0_q\} \times \mathbb{R}_-^{p-q}$ . Denote

$$\mathcal{I}(\bar{x}) := \{i : g_i(\bar{x}) = 0, i = q + 1, \dots, p\}$$

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  $\nabla g_i(\bar{x})$ ,  $i \in \{1, \dots, q\} \cup \mathcal{I}(\bar{x})$ , are linearly independent. Then the corresponding set of Lagrange multipliers is a singleton,  $\Lambda(\bar{x}) = \{\bar{\lambda}\}$ . In that case

$$C(\bar{x}) = \{h : h^\top \nabla g_i(\bar{x}) = 0, i \in \{1, \dots, q\} \cup \mathcal{I}_+(\bar{\lambda}), h^\top \nabla g_i(\bar{x}) \leq 0, i \in \mathcal{I}_0(\bar{\lambda})\},$$

where

$$\mathcal{I}_0(\bar{\lambda}) := \{i \in \mathcal{I}(\bar{x}) : \bar{\lambda}_i = 0\} \quad \text{and} \quad \mathcal{I}_+(\bar{\lambda}) := \{i \in \mathcal{I}(\bar{x}) : \bar{\lambda}_i > 0\}.$$

Consequently problem (5.40) takes the form

$$\begin{aligned} \text{Min}_{h \in \mathbb{R}^n} \quad & 2h^\top Z + h^\top \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})h \\ \text{s.t.} \quad & h^\top \nabla g_i(\bar{x}) = 0, i \in \{1, \dots, q\} \cup \mathcal{I}_+(\bar{\lambda}), h^\top \nabla g_i(\bar{x}) \leq 0, i \in \mathcal{I}_0(\bar{\lambda}). \end{aligned} \quad (5.43)$$

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

$$\begin{bmatrix} H & A \\ A^\top & 0 \end{bmatrix} \begin{bmatrix} h \\ \alpha \end{bmatrix} = \begin{bmatrix} Z \\ 0 \end{bmatrix}. \quad (5.44)$$

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  $\nabla g_i(\bar{x})$ ,  $i \in \{1, \dots, q\} \cup \mathcal{I}(\bar{x})$ . Then

$$N^{1/2} \begin{bmatrix} \hat{x}_N - \bar{x} \\ \hat{\lambda}_N - \bar{\lambda} \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, J^{-1} \Upsilon J^{-1}), \quad (5.45)$$

where

$$J := \begin{bmatrix} H & A \\ A^\top & 0 \end{bmatrix} \quad \text{and} \quad \Upsilon := \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix},$$

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  $\{h : A^\top h = 0\}$ . 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 8.** As 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  $X$  is given in the form (5.42). This curvature term can also be written in an explicit form for some nonpolyhedral cones, in particular for the cone of positive semidefinite matrices (see [22, section 5.3.6]).

### 5.1.4 Minimax Stochastic Programs

Sometimes it is worthwhile to consider minimax stochastic programs of the form

$$\text{Min sup}_{x \in X, y \in Y} \{ f(x, y) := \mathbb{E}[F(x, y, \xi)] \}, \quad (5.46)$$

where  $X \subset \mathbb{R}^n$  and  $Y \subset \mathbb{R}^m$  are closed sets,  $F : X \times Y \times \Xi \rightarrow \mathbb{R}$  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,

$$\text{Min sup}_{x \in X, y \in Y} \left\{ \hat{f}_N(x, y) := \frac{1}{N} \sum_{j=1}^N F(x, y, \xi^j) \right\}. \quad (5.47)$$

As before, denote by,  $\vartheta^*$  and  $\hat{\vartheta}_N$  the optimal values of (5.46) and (5.47), respectively, and by  $S_x \subset X$  and  $\hat{S}_{x,N} \subset X$  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  $X$  and  $Y$  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  $X \times Y$  and an integrable, with respect to the probability distribution of the random vector  $\xi$ , 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 7.43 it follows that the expected value function  $f(x, y)$  is continuous on  $X \times Y$ . Since  $Y$  is compact, this implies that the max-function

$$\phi(x) := \sup_{y \in Y} f(x, y)$$

is continuous on  $X$ . 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

$$\hat{\phi}_N(x, \omega) := \sup_{y \in Y} \hat{f}_N(x, y, \omega)$$

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  $\hat{\vartheta}_N \rightarrow \vartheta^*$  and  $\mathbb{D}(\hat{S}_{x,N}, S_x) \rightarrow 0$  w.p. 1 as  $N \rightarrow \infty$ .*

**Proof.** By Theorem 7.48 we have that under the specified assumptions,  $\hat{f}_N(x, y)$  converges to  $f(x, y)$  w.p. 1 uniformly on  $X \times Y$ . That is,  $\Delta_N \rightarrow 0$  w.p. 1 as  $N \rightarrow \infty$ , where

$$\Delta_N := \sup_{(x,y) \in X \times Y} \left| \hat{f}_N(x, y) - f(x, y) \right|.$$

### 5.1. Statistical Properties of Sample Average Approximation Estimators

171

Consider  $\hat{\phi}_N(x) := \sup_{y \in Y} \hat{f}_N(x, y)$  and  $\phi(x) := \sup_{y \in Y} f(x, y)$ . We have that

$$\sup_{x \in X} |\hat{\phi}_N(x) - \phi(x)| \leq \Delta_N,$$

and hence  $|\hat{\vartheta}_N - \vartheta^*| \leq \Delta_N$ . It follows that  $\hat{\vartheta}_N \rightarrow \vartheta^*$  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  $\mathbb{D}(\hat{S}_{x,N}, S_x) \rightarrow 0$  w.p. 1, one can proceed exactly in the same way as in the proof of Theorem 5.3.  $\square$

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 or a.e.  $\xi \in \Xi$  the function  $F(\cdot, \cdot, \xi)$  is convex–concave on  $X \times Y$ , i.e., the function  $F(\cdot, y, \xi)$  is convex on  $X$  for every  $y \in Y$ , and the function  $F(x, \cdot, \xi)$  is concave on  $Y$  for every  $x \in X$ .

It follows that the expected value function  $f(x, y)$  is convex concave and continuous on  $X \times Y$ . Consequently, problem (5.46) and its dual

$$\text{Max}_{y \in Y} \inf_{x \in X} f(x, y) \quad (5.48)$$

have nonempty and bounded sets of optimal solutions  $S_x \subset X$  and  $S_y \subset Y$ , 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  $(x, y) \in X \times Y$ , the expectation  $\mathbb{E}[F(x, y, \xi)^2]$  is finite, and there exists a measurable function  $C : \Xi \rightarrow \mathbb{R}_+$  such that  $\mathbb{E}[C(\xi)^2]$  is finite and the inequality

$$|F(x, y, \xi) - F(x', y', \xi)| \leq C(\xi)(\|x - x'\| + \|y - y'\|) \quad (5.49)$$

holds for all  $(x, y), (x', y') \in X \times Y$  and a.e.  $\xi \in \Xi$ .

The above assumption implies that  $f(x, y)$  is Lipschitz continuous on  $X \times Y$  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*

$$\hat{\vartheta}_N = \inf_{x \in S_x} \sup_{y \in S_y} \hat{f}_N(x, y) + o_p(N^{-1/2}). \quad (5.50)$$

*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  $\sigma^2 = \mathbb{V}\text{ar}[F(\bar{x}, \bar{y}, \xi)]$ .*

**Proof.** Consider the space  $C(X, Y)$  of continuous functions  $\psi : X \times Y \rightarrow \mathbb{R}$  equipped with the sup-norm  $\|\psi\| = \sup_{x \in X, y \in Y} |\psi(x, y)|$ , and set  $\mathcal{K} \subset C(X, Y)$  formed by convex–concave on  $X \times Y$  functions. It is not difficult to see that the set  $\mathcal{K}$  is a closed (in the

norm topology of  $C(X, Y)$ ) and convex cone. Consider the optimal value function  $V : C(X, Y) \rightarrow \mathbb{R}$  defined as

$$V(\psi) := \inf_{x \in X} \sup_{y \in Y} \psi(x, y) \quad \text{for } \psi \in C(X, Y). \quad (5.51)$$

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  $\gamma \in \mathcal{T}_{\mathcal{K}}(f)$ :

$$V'_f(\gamma) := \lim_{\substack{t \downarrow 0, \eta \rightarrow \gamma \\ f+t\eta \in \mathcal{K}}} \frac{V(f+t\eta) - V(f)}{t}. \quad (5.52)$$

By Theorem 7.24 we have that the optimal value function  $V(\cdot)$  is Hadamard directionally differentiable at  $f$  tangentially to the set  $\mathcal{K}$  and

$$V'_f(\gamma) = \inf_{x \in S_x} \sup_{y \in S_y} \gamma(x, y) \quad (5.53)$$

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  $C(X, Y)$ . 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 7.61, we can complete the proof.  $\square$

Suppose now that the feasible set  $X$  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  $X_0$  is convex and for all  $\xi \in \Xi$  the functions  $F(\cdot, \xi)$  and  $G_i(\cdot, \xi)$ ,  $i = 1, \dots, p$ , 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  $\Lambda$  of Lagrange multipliers vectors  $\lambda = (\lambda_1, \dots, \lambda_p)$  satisfying the optimality conditions

$$\bar{x} \in \arg \min_{x \in X_0} L(x, \lambda), \quad \lambda_i \geq 0 \quad \text{and} \quad \lambda_i g_i(\bar{x}) = 0, \quad i = 1, \dots, p. \quad (5.54)$$

The set  $\Lambda$  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  $X_N$  given in the form (5.13). That is,  $\hat{\vartheta}_N$  is the optimal value of the problem

$$\text{Min}_{x \in X_0} \hat{f}_N(x) \quad \text{subject to} \quad \hat{g}_{iN}(x) \leq 0, \quad i = 1, \dots, p, \quad (5.55)$$

with  $\hat{f}_N(x)$  and  $\hat{g}_{iN}(x)$  being the SAA functions of the respective integrands  $F(x, \xi)$  and  $G_i(x, \xi)$ ,  $i = 1, \dots, p$ . Assume that conditions (A1) and (A2), formulated on page 164, are satisfied for the integrands  $F$  and  $G_i$ ,  $i = 1, \dots, p$ , 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  $G_i(x, \xi)$ ,  $i = 1, \dots, p$ .

**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, and (iv) the assumptions (A1) and (A2) hold for the integrands  $F$  and  $G_i$ ,  $i = 1, \dots, p$ . Then*

$$N^{1/2} \left( \hat{\vartheta}_N - \vartheta^* \right) \xrightarrow{\mathcal{D}} \inf_{x \in S} \sup_{\lambda \in \Lambda} \left[ Y(x) + \sum_{i=1}^p \lambda_i Y_i(x) \right]. \quad (5.56)$$

If, moreover,  $S = \{\bar{x}\}$  and  $\Lambda = \{\bar{\lambda}\}$  are singletons, then

$$N^{1/2} \left( \hat{\vartheta}_N - \vartheta^* \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2) \quad (5.57)$$

with

$$\sigma^2 := \mathbb{V}\text{ar} \left[ F(\bar{x}, \xi) + \sum_{i=1}^p \bar{\lambda}_i G_i(\bar{x}, \xi) \right]. \quad (5.58)$$

**Proof.** Since the problem is convex and the Slater condition (for the true problem) holds, we have that  $\vartheta^*$  is equal to the optimal value of the (Lagrangian) dual

$$\text{Max}_{\lambda \geq 0} \inf_{x \in X_0} L(x, \lambda), \quad (5.59)$$

and the set of optimal solutions of (5.59) is nonempty and compact and coincides with the set of Lagrange multipliers  $\Lambda$ . 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 now be completed by applying Theorem 5.10.  $\square$

**Remark 9.** 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  $\xi^1, \dots, \xi^N$  for estimating the functions  $f(x)$  and  $g_i(x)$ ,  $i = 1, \dots, p$ , by their SAA counterparts. The other is to use independent samples, possibly of different sizes, for each of these functions (see Remark 5 on page 161). The asymptotic results of Theorem 5.11 are for the case of the same sample. The (asymptotic) variance  $\sigma^2$ , given in (5.58), is equal to the sum of the variances of  $F(\bar{x}, \xi)$  and  $\bar{\lambda}_i G_i(\bar{x}, \xi)$ ,  $i = 1, \dots, p$ , 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  $\xi$  whose distribution is supported on a set  $\Xi \subset \mathbb{R}^d$ , a mapping  $\Phi : \mathbb{R}^n \times \Xi \rightarrow \mathbb{R}^n$ , 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

$$\phi(x) \in \Gamma(x) \quad (5.60)$$

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  $\Gamma(\cdot) := \mathcal{N}_X(\cdot)$ , where  $X$  is a nonempty closed convex subset of  $\mathbb{R}^n$  and  $\mathcal{N}_X(x)$  denotes the (outward) normal cone to  $X$  at  $x$ . Recall that, by the definition,  $\mathcal{N}_X(x) = \emptyset$  if  $x \notin X$ . In that case  $\bar{x}$  is a solution of (5.60) iff  $\bar{x} \in X$  and the following so-called variational inequality holds:

$$(x - \bar{x})^\top \phi(\bar{x}) \leq 0, \quad \forall x \in X. \quad (5.61)$$

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  $X = \mathbb{R}^n$ , then  $\mathcal{N}_X(x) = \{0\}$  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  $X$  of the optimization problem (5.1) is defined by constraints in the form

$$X := \{x \in \mathbb{R}^n : g_i(x) = 0, i = 1, \dots, q, g_i(x) \leq 0, i = q + 1, \dots, p\} \quad (5.62)$$

with  $g_i(x) := \mathbb{E}[G_i(x, \xi)]$ ,  $i = 1, \dots, p$ , then the corresponding first-order Karush–Kuhn–Tucker (KKT) optimality conditions can be written in a form of variational inequality. That is, let  $z := (x, \lambda) \in \mathbb{R}^{n+p}$  and

$$\begin{aligned} L(z, \xi) &:= F(x, \xi) + \sum_{i=1}^p \lambda_i G_i(x, \xi), \\ \ell(z) &:= \mathbb{E}[L(z, \xi)] = f(x) + \sum_{i=1}^p \lambda_i g_i(x) \end{aligned}$$

be the corresponding Lagrangians. Define

$$\Phi(z, \xi) := \begin{bmatrix} \nabla_x L(z, \xi) \\ G_1(x, \xi) \\ \vdots \\ G_p(x, \xi) \end{bmatrix} \text{ and } \Gamma(z) := \mathcal{N}_K(z), \quad (5.63)$$

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

$$\mathcal{N}_K(z) = \left\{ (v, \gamma) \in \mathbb{R}^{n+p} : \begin{array}{l} v = 0 \text{ and } \gamma_i = 0, i = 1, \dots, q, \\ \gamma_i = 0, i \in \mathcal{I}_+(\lambda), \gamma_i \leq 0, i \in \mathcal{I}_0(\lambda) \end{array} \right\}, \quad (5.64)$$

where

$$\begin{aligned}\mathcal{I}_0(\lambda) &:= \{i : \lambda_i = 0, i = q + 1, \dots, p\}, \\ \mathcal{I}_+(\lambda) &:= \{i : \lambda_i > 0, i = q + 1, \dots, p\},\end{aligned}\quad (5.65)$$

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

$$\phi(z) := \mathbb{E}[\Phi(z, \xi)] = \begin{bmatrix} \nabla_x \ell(z) \\ g_1(x) \\ \dots \\ g_p(x) \end{bmatrix}, \quad (5.66)$$

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  $x_k \rightarrow x$ ,  $y_k \in \Gamma(x_k)$  and  $y_k \rightarrow y$ , 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  $x \mapsto \mathcal{N}_K(x)$  is closed.

Now let  $\xi^1, \dots, \xi^N$  be a random sample of  $N$  realizations of the random vector  $\xi$  and let  $\hat{\phi}_N(x) := N^{-1} \sum_{j=1}^N \Phi(x, \xi^j)$  be the corresponding sample average estimate of  $\phi(x)$ . We refer to

$$\hat{\phi}_N(x) \in \Gamma(x) \quad (5.67)$$

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  $S$  and  $\hat{S}_N$  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  $\hat{S}_N \subset C$ , and (iv)  $\hat{\phi}_N(x)$  converges to  $\phi(x)$  w.p. 1 uniformly on  $C$  as  $N \rightarrow \infty$ . Then  $\mathbb{D}(\hat{S}_N, S) \rightarrow 0$  w.p. 1 as  $N \rightarrow \infty$ .*

**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.  $\omega$ . Therefore we omit saying “w.p. 1.” Consider a sequence  $\hat{x}_N \in \hat{S}_N$ . Because of assumption (iii), by passing to a subsequence if necessary, we need to show only 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  $\hat{\phi}_N(\hat{x}_N) \rightarrow \phi(x^*)$  (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.  $\square$

A few remarks about the assumptions involved in the above consistency result are now in order. By Theorem 7.48 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  $\xi \in \Xi$  the function  $\Phi(\cdot, \xi)$  is continuous on  $C$  and  $\|\Phi(x, \xi)\|_{x \in C}$  is dominated by an integrable function.

There are two parts to 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 often can be verified by ad hoc methods. The second assumption is more subtle. We will discuss it next. The following concept of strong regularity is due to Robinson [170].

**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 (linearized) generalized equation

$$\delta + \phi(\bar{x}) + \nabla \phi(\bar{x})(x - \bar{x}) \in \Gamma(x) \quad (5.68)$$

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 the 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  $\psi : \mathcal{V} \rightarrow \mathbb{R}^n$  equipped with the norm

$$\|\psi\|_{1, \mathcal{V}} := \sup_{x \in \mathcal{V}} \|\psi(x)\| + \sup_{x \in \mathcal{V}} \|\nabla \psi(x)\|.$$

The following (deterministic) result is essentially due to Robinson [171].

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

$$\hat{x}(u) = \bar{x}(u(\bar{x}) - \phi(\bar{x})) + o(\|u - \phi\|_{1, \mathcal{V}}). \quad (5.69)$$

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  $\|\hat{\phi}_N - \phi\|_{1,\mathcal{V}} \rightarrow 0$  w.p. 1 as  $N \rightarrow \infty$ . 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  $\hat{x}_N \rightarrow \bar{x}$  w.p. 1 as  $N \rightarrow \infty$ .*

The assumption that  $\|\hat{\phi}_N - \phi\|_{1,\mathcal{V}} \rightarrow 0$  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 7.48, in the case of iid sampling this is ensured by the following assumption:

**(E3)** For a.e.  $\xi$  the mapping  $\Phi(\cdot, \xi)$  is continuously differentiable on  $\mathcal{V}$ , and  $\|\Phi(x, \xi)\|_{x \in \mathcal{V}}$  and  $\|\nabla_x \Phi(x, \xi)\|_{x \in \mathcal{V}}$  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 let  $\bar{x}$  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.  $\xi$  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, and (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  $N \rightarrow \infty$ .*

Note again that if the sample is iid, then assumption (iv) in the above theorem is implied by 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

$$\tilde{x}(\delta) = \bar{x} - J^{-1}\delta. \quad (5.70)$$

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

$$N^{1/2}(\hat{x}_N - \bar{x}) = -J^{-1}Y_N + o_p(1), \quad (5.71)$$

where  $Y_N := N^{1/2}[\hat{\phi}_N(\bar{x}) - \phi(\bar{x})]$ . Moreover, in the case of iid sample, we have by the CLT that  $Y_N \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma)$ , where  $\Sigma$  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  $\Gamma(\cdot) := \mathcal{N}_X(\cdot)$  with the set  $X$  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

$$\mathcal{C}_X(\bar{x}) := \{y \in \mathcal{T}_X(\bar{x}) : y^\top \phi(\bar{x}) = 0\}, \quad (5.72)$$

called the *critical cone*, and the Jacobian matrix  $J := \nabla \phi(\bar{x})$ . Then for all  $\delta$  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

$$\delta + Jd \in \mathcal{N}_{\mathcal{C}_X(\bar{x})}(d). \quad (5.73)$$

Note that the mapping  $\tilde{d}(\cdot)$  is positively homogeneous, i.e., for any  $\delta \in \mathbb{R}^n$  and  $t \geq 0$ , 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

$$N^{1/2}(\hat{x}_N - \bar{x}) \xrightarrow{\mathcal{D}} \tilde{d}(Y), \quad (5.74)$$

where  $Y \sim \mathcal{N}(0, \Sigma)$  and  $\Sigma$  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  $\mathcal{C}_X(\bar{x})$  is a linear space.

In the case  $\Gamma(\cdot) := \mathcal{N}_X(\cdot)$ , with the set  $X$  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  $\mathcal{C}_X(\bar{x})$ . The interested reader is referred to [172], [79] for a discussion of this topic. Let us just remark that if  $\mathcal{C}_X(\bar{x})$  is a linear subspace of  $\mathbb{R}^n$ , then the variational inequality (5.73) can be written in the form

$$P\delta + PJd = 0, \quad (5.75)$$

where  $P$  denotes the orthogonal projection matrix onto the linear space  $\mathcal{C}_X(\bar{x})$ . Then  $\bar{x}$  is strongly regular iff the matrix (mapping)  $PJ$  restricted to the linear space  $\mathcal{C}_X(\bar{x})$  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 (7.191)) 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  $\Gamma(\cdot) := \mathcal{N}_X(\cdot)$  and  $\bar{x} \in S$ , then  $\phi(\bar{x}) \in \text{int } \Gamma(\bar{x})$  iff the critical cone  $\mathcal{C}_X(\bar{x})$  is equal to  $\{0\}$ . In that case, the variational inequality (5.73) has solution  $\tilde{d} = 0$  for any  $\delta$ , 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

## 5.2. Stochastic Generalized Equations

179

feasible set  $X$  given in the form (5.62). (We assume that the expectation functions  $f(x)$  and  $g_i(x)$ ,  $i = 1, \dots, p$ , are continuously differentiable.) Let  $\bar{x}$  be an optimal solution of the (expected value) problem (5.1). It is said that the LICQ holds at the point  $\bar{x}$  if the gradient vectors  $\nabla g_i(\bar{x})$ ,  $i \in \{i : g_i(\bar{x}) = 0, i = 1, \dots, p\}$ , (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  $\mathcal{I}_0(\bar{\lambda})$  and  $\mathcal{I}_+(\bar{\lambda})$  be the index sets defined in (5.65). Then

$$\mathcal{T}_K(\bar{z}) = \mathbb{R}^n \times \mathbb{R}^q \times \{\gamma \in \mathbb{R}^{p-q} : \gamma_i \geq 0, i \in \mathcal{I}_0(\bar{\lambda})\}. \quad (5.76)$$

In order to simplify notation, let us assume that *all* constraints are active at  $\bar{x}$ , i.e.,  $g_i(\bar{x}) = 0$ ,  $i = 1, \dots, p$ . 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  $\mathcal{C}_K(\bar{z}) = \mathcal{T}_K(\bar{z})$ .

Assuming, further, that  $f(x)$  and  $g_i(x)$ ,  $i = 1, \dots, p$ , are twice continuously differentiable, we have that the following second order necessary conditions hold at  $\bar{x}$ :

$$h^\top \nabla_{xx}^2 \ell(\bar{z}) h \geq 0, \quad \forall h \in C_X(\bar{x}), \quad (5.77)$$

where

$$C_X(\bar{x}) := \{h : h^\top \nabla g_i(\bar{x}) = 0, i \in \{1, \dots, q\} \cup \mathcal{I}_+(\bar{\lambda}), h^\top \nabla g_i(\bar{x}) \leq 0, i \in \mathcal{I}_0(\bar{\lambda})\}.$$

The corresponding second order sufficient conditions are

$$h^\top \nabla_{xx}^2 \ell(\bar{z}) h > 0, \quad \forall h \in C_X(\bar{x}) \setminus \{0\}. \quad (5.78)$$

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:

$$h^\top \nabla_{xx}^2 \ell(\bar{z}) h > 0, \quad \forall h \in \text{lin}(C_X(\bar{x})) \setminus \{0\}, \quad (5.79)$$

where

$$\text{lin}(C_X(\bar{x})) := \{h : h^\top \nabla g_i(\bar{x}) = 0, i \in \{1, \dots, q\} \cup \mathcal{I}_+(\bar{\lambda})\}. \quad (5.80)$$

Under the LICQ, the set defined in the right-hand side of (5.80) is, indeed, the linear space generated by the cone  $C_X(\bar{x})$ . We also have here

$$J := \nabla \phi(\bar{z}) = \begin{bmatrix} H & A \\ A^\top & 0 \end{bmatrix}, \quad (5.81)$$

where  $H := \nabla_{xx}^2 \ell(\bar{z})$  and  $A := [\nabla g_1(\bar{x}), \dots, \nabla g_p(\bar{x})]$ .

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  $\hat{z}_N = [\hat{x}_N, \hat{\lambda}_N]$  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)):

$$N^{1/2}(\hat{z}_N - \bar{z}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, J^{-1} \Sigma J^{-1}), \quad (5.82)$$

where  $\Sigma$  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  $\xi^1, \dots, \xi^N$  of  $N$  realizations of the random vector  $\xi$  can be generated in the computer. In the Monte Carlo sampling method this is accomplished by generating a sequence  $U^1, U^2, \dots$  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  $\omega := \{U^1, U^2, \dots\}$  as an element of the probability space equipped with the corresponding product probability measure, and the sample  $\xi^j = \xi^j(\omega)$ ,  $j = 1, 2, \dots$ , as a function of  $\omega$ . 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  $\hat{f}_N(x)$  can be computed in two somewhat different ways. The generated sample  $\xi^1, \dots, \xi^N$  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  $x_1, x_2 \in X$ . 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

$$\mathbb{V}\text{ar}[\hat{f}_N(x_1) - \hat{f}_N(x_2)] = \mathbb{V}\text{ar}[\hat{f}_N(x_1)] + \mathbb{V}\text{ar}[\hat{f}_N(x_2)]. \quad (5.83)$$

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

$$\mathbb{V}\text{ar}[\hat{f}_N(x_1) - \hat{f}_N(x_2)] = \mathbb{V}\text{ar}[\hat{f}_N(x_1)] + \mathbb{V}\text{ar}[\hat{f}_N(x_2)] - 2\text{Cov}(\hat{f}_N(x_1), \hat{f}_N(x_2)). \quad (5.84)$$

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 us 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 insight into the *complexity* of solving the true (expected value) problem.

Unless stated otherwise, we assume in this section that the random sample  $\xi^1, \dots, \xi^N$  is iid, and make the following assumption:

**(M1)** The expectation function  $f(x)$  is well defined and finite valued for all  $x \in X$ .

For  $\varepsilon \geq 0$  we denote by

$$S^\varepsilon := \{x \in X : f(x) \leq \vartheta^* + \varepsilon\} \quad \text{and} \quad \hat{S}_N^\varepsilon := \{x \in X : \hat{f}_N(x) \leq \hat{\vartheta}_N + \varepsilon\}$$

the sets of  $\varepsilon$ -optimal solutions of the true and the SAA problems, respectively.

### 5.3.1 Exponential Rates of Convergence and Sample Size Estimates in the Case of a Finite Feasible Set

In this section we assume that the feasible set  $X$  is finite, although its cardinality  $|X|$  can be very large. Since  $X$  is finite, the sets  $S^\varepsilon$  and  $\hat{S}_N^\varepsilon$  are nonempty and finite. For parameters  $\varepsilon \geq 0$  and  $\delta \in [0, \varepsilon]$ , consider the event  $\{\hat{S}_N^\delta \subset S^\varepsilon\}$ . This event means that any  $\delta$ -optimal solution of the SAA problem is an  $\varepsilon$ -optimal solution of the true problem. We estimate now the probability of that event.

We can write

$$\{\hat{S}_N^\delta \not\subset S^\varepsilon\} = \bigcup_{x \in X \setminus S^\varepsilon} \bigcap_{y \in X} \{\hat{f}_N(x) \leq \hat{f}_N(y) + \delta\}, \quad (5.85)$$

and hence

$$\Pr(\hat{S}_N^\delta \not\subset S^\varepsilon) \leq \sum_{x \in X \setminus S^\varepsilon} \Pr\left(\bigcap_{y \in X} \{\hat{f}_N(x) \leq \hat{f}_N(y) + \delta\}\right). \quad (5.86)$$

Consider a mapping  $u : X \setminus S^\varepsilon \rightarrow X$ . If the set  $X \setminus S^\varepsilon$  is empty, then any feasible point  $x \in X$  is an  $\varepsilon$ -optimal solution of the true problem. Therefore we assume that this set is nonempty. It follows from (5.86) that

$$\Pr(\hat{S}_N^\delta \not\subset S^\varepsilon) \leq \sum_{x \in X \setminus S^\varepsilon} \Pr\{\hat{f}_N(x) - \hat{f}_N(u(x)) \leq \delta\}. \quad (5.87)$$

We assume that the mapping  $u(\cdot)$  is chosen in such a way that

$$f(u(x)) \leq f(x) - \varepsilon^*, \quad \forall x \in X \setminus S^\varepsilon, \quad (5.88)$$

and for some  $\varepsilon^* \geq \varepsilon$ . Note that such a mapping always exists. For example, if we use a mapping  $u : X \setminus S^\varepsilon \rightarrow S$ , then (5.88) holds with

$$\varepsilon^* := \min_{x \in X \setminus S^\varepsilon} f(x) - \vartheta^* \quad (5.89)$$

and that  $\varepsilon^* > \varepsilon$  since the set  $X$  is finite. Different choices of  $u(\cdot)$  give a certain flexibility to the following derivations.

For each  $x \in X \setminus S^\varepsilon$ , define

$$Y(x, \xi) := F(u(x), \xi) - F(x, \xi). \quad (5.90)$$

Note that  $\mathbb{E}[Y(x, \xi)] = f(u(x)) - f(x)$ , and hence  $\mathbb{E}[Y(x, \xi)] \leq -\varepsilon^*$  for all  $x \in X \setminus S^\varepsilon$ . 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

$$\Pr(\hat{S}_N^\delta \not\subset S^\varepsilon) \leq \sum_{x \in X \setminus S^\varepsilon} \Pr\{\hat{Y}_N(x) \geq -\delta\}. \quad (5.91)$$

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 (7.173) implies

$$1 - \Pr(\hat{S}_N^\delta \subset S^\varepsilon) \leq \sum_{x \in X \setminus S^\varepsilon} e^{-N I_x(-\delta)}. \quad (5.92)$$

Note that inequality (5.92) is valid for any random sample of size  $N$ . Let us make the following assumption:

**(M2)** For every  $x \in X \setminus S^\varepsilon$ , the moment-generating function  $\mathbb{E}[e^{tY(x, \xi)}]$  of the random variable  $Y(x, \xi) = F(u(x), \xi) - F(x, \xi)$  is finite valued in a neighborhood of  $t = 0$ .

Assumption (M2) holds, for example, if the support  $\Xi$  of  $\xi$  is a bounded subset of  $\mathbb{R}^d$ , or if  $Y(x, \cdot)$  grows at most linearly and  $\xi$  has a distribution from an exponential family.

**Theorem 5.16.** *Let  $\varepsilon$  and  $\delta$  be nonnegative numbers. Then*

$$1 - \Pr(\hat{S}_N^\delta \subset S^\varepsilon) \leq |X| e^{-N \eta(\delta, \varepsilon)}, \quad (5.93)$$

where

$$\eta(\delta, \varepsilon) := \min_{x \in X \setminus S^\varepsilon} I_x(-\delta). \quad (5.94)$$

Moreover, if  $\delta < \varepsilon^*$  and assumption (M2) holds, then  $\eta(\delta, \varepsilon) > 0$ .

**Proof.** Inequality (5.93) is an immediate consequence of inequality (5.92). If  $\delta < \varepsilon^*$ , then  $-\delta > -\varepsilon^* \geq \mathbb{E}[Y(x, \xi)]$ , and hence it follows by assumption (M2) that  $I_x(-\delta) > 0$  for every  $x \in X \setminus S^\varepsilon$ . (See the discussion above equation (7.178).) This implies that  $\eta(\delta, \varepsilon) > 0$ .  $\square$

The following asymptotic result is an immediate consequence of inequality (5.93):

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \ln [1 - \Pr(\hat{S}_N^\delta \subset S^\varepsilon)] \leq -\eta(\delta, \varepsilon). \quad (5.95)$$

It means that the probability of the event that any  $\delta$ -optimal solution of the SAA problem provides an  $\varepsilon$ -optimal solution of the true problem approaches one *exponentially fast* as  $N \rightarrow \infty$ . Note that since it is possible to employ a mapping  $u : X \setminus S^\varepsilon \rightarrow S$  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  $\delta$  close to  $-\mathbb{E}[Y(x, \xi)]$ , we can write by (7.178) that

$$I_x(-\delta) \approx \frac{(-\delta - \mathbb{E}[Y(x, \xi)])^2}{2\sigma_x^2} \geq \frac{(\varepsilon^* - \delta)^2}{2\sigma_x^2}, \quad (5.96)$$

where

$$\sigma_x^2 := \text{Var}[Y(x, \xi)] = \text{Var}[F(u(x), \xi) - F(x, \xi)]. \quad (5.97)$$

Let us make now the following assumption:

**(M3)** There is a constant  $\sigma > 0$  such that for any  $x \in X \setminus S^\varepsilon$  the moment-generating function  $M_x(t)$  of the random variable  $Y(x, \xi) - \mathbb{E}[Y(x, \xi)]$  satisfies

$$M_x(t) \leq \exp(\sigma^2 t^2 / 2), \quad \forall t \in \mathbb{R}. \quad (5.98)$$

It follows from assumption (M3) that

$$\ln \mathbb{E}[e^{tY(x, \xi)}] - t\mathbb{E}[Y(x, \xi)] = \ln M_x(t) \leq \sigma^2 t^2 / 2, \quad (5.99)$$

and hence the rate function  $I_x(\cdot)$ , of  $Y(x, \xi)$ , satisfies

$$I_x(z) \geq \sup_{t \in \mathbb{R}} \{t(z - \mathbb{E}[Y(x, \xi)]) - \sigma^2 t^2 / 2\} = \frac{(z - \mathbb{E}[Y(x, \xi)])^2}{2\sigma^2}, \quad \forall z \in \mathbb{R}. \quad (5.100)$$

In particular, it follows that

$$I_x(-\delta) \geq \frac{(-\delta - \mathbb{E}[Y(x, \xi)])^2}{2\sigma^2} \geq \frac{(\varepsilon^* - \delta)^2}{2\sigma^2} \geq \frac{(\varepsilon - \delta)^2}{2\sigma^2}. \quad (5.101)$$

Consequently the constant  $\eta(\delta, \varepsilon)$  satisfies

$$\eta(\delta, \varepsilon) \geq \frac{(\varepsilon - \delta)^2}{2\sigma^2}, \quad (5.102)$$

and hence the bound (5.93) of Theorem 5.16 takes the form

$$1 - \Pr(\hat{S}_N^\delta \subset S^\varepsilon) \leq |X| e^{-N(\varepsilon - \delta)^2 / (2\sigma^2)}. \quad (5.103)$$

This leads to the following result giving an estimate of the sample size which guarantees that any  $\delta$ -optimal solution of the SAA problem is an  $\varepsilon$ -optimal solution of the true problem with probability at least  $1 - \alpha$ .

**Theorem 5.17.** *Suppose that assumptions (M1) and (M3) hold. Then for  $\varepsilon > 0$ ,  $0 \leq \delta < \varepsilon$ , and  $\alpha \in (0, 1)$ , and for the sample size  $N$  satisfying*

$$N \geq \frac{2\sigma^2}{(\varepsilon - \delta)^2} \ln \left( \frac{|X|}{\alpha} \right), \quad (5.104)$$

it follows that

$$\Pr(\hat{S}_N^\delta \subset S^\varepsilon) \geq 1 - \alpha. \quad (5.105)$$

**Proof.** By setting the right-hand side of the estimate (5.103) to  $\leq \alpha$  and solving the obtained inequality, we obtain (5.104).  $\square$

**Remark 10.** 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  $X$  and on the tolerance probability (significance level)  $\alpha$ . The constant  $\sigma$ , postulated in assumption (M3), measures, in a sense, variability of a considered problem. If, for some  $x \in X$ , the random variable  $Y(x, \xi)$  has a normal distribution with mean  $\mu_x$  and variance  $\sigma_x^2$ , then its moment-generating function is equal to  $\exp(\mu_x t + \sigma_x^2 t^2/2)$ , and hence the moment-generating function  $M_x(t)$ , specified in assumption (M3), is equal to  $\exp(\sigma_x^2 t^2/2)$ . In that case,  $\sigma^2 := \max_{x \in X \setminus S^\varepsilon} \sigma_x^2$  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

$$|Y(x, \xi) - \mathbb{E}[Y(x, \xi)]| \leq b, \quad \forall x \in X \text{ and a.e. } \xi \in \Xi,$$

then by Hoeffding inequality (see Proposition 7.63 and estimate (7.186)) 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 11.** 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  $\sigma$  and  $\varepsilon$  seems to be unavoidable. In order to see that, consider a simple case when the feasible set  $X$  consists of just two elements, i.e.,  $X = \{x_1, x_2\}$  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  $\varepsilon$ -optimal solution if  $\hat{f}_N(x_2) - \hat{f}_N(x_1) > 0$ . 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  $\sigma^2$ , 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  $\alpha$  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  $\sigma^2/\varepsilon^2$  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 access to a random sample, then solving the corresponding SAA problem is in a sense a best way to proceed.

**Remark 12.** Condition (5.98) of assumption (M3) can be replaced by a more general condition,

$$M_x(t) \leq \exp(\psi(t)), \quad \forall t \in \mathbb{R}, \quad (5.106)$$

where  $\psi(t)$  is a convex even function with  $\psi(0) = 0$ . Then, similar to (5.100), we have

$$I_x(z) \geq \sup_{t \in \mathbb{R}} \{t(z - \mathbb{E}[Y(x, \xi)]) - \psi(t)\} = \psi^*(z - \mathbb{E}[Y(x, \xi)]), \quad \forall z \in \mathbb{R}, \quad (5.107)$$

where  $\psi^*$  is the conjugate of function  $\psi$ . Consequently, the estimate (5.93) takes the form

$$1 - \Pr(\hat{S}_N^\delta \subset S^\varepsilon) \leq |X| e^{-N\psi^*(\varepsilon - \delta)}, \quad (5.108)$$

and hence the estimate (5.104) takes the form

$$N \geq \frac{1}{\psi^*(\varepsilon - \delta)} \ln \left( \frac{|X|}{\alpha} \right). \quad (5.109)$$

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  $|t| \leq a$  and  $\psi(t) := +\infty$  otherwise. In that case  $\psi^*(z) = z^2/(2\sigma^2)$  for  $|z| \leq a\sigma^2$  and  $\psi^*(z) = a|z| - a^2\sigma^2$  for  $|z| > a\sigma^2$ . Consequently, the estimate (5.104) of Theorem 5.17 still holds provided that  $0 < \varepsilon - \delta \leq a\sigma^2$ .

### 5.3.2 Sample Size Estimates in the General Case

Suppose now that  $X$  is a bounded, not necessarily finite, subset of  $\mathbb{R}^n$ , and that  $f(x)$  is finite valued for all  $x \in X$ . Then we can proceed in a way similar to the derivations of section 7.2.9. Let us make the following assumptions:

**(M4)** For any  $x', x \in X$  there exists constant  $\sigma_{x',x} > 0$  such that the moment-generating function  $M_{x',x}(t) = \mathbb{E}[e^{tY_{x',x}}]$  of random variable  $Y_{x',x} := [F(x', \xi) - f(x')] - [F(x, \xi) - f(x)]$  satisfies

$$M_{x',x}(t) \leq \exp(\sigma_{x',x}^2 t^2/2), \quad \forall t \in \mathbb{R}. \quad (5.110)$$

**(M5)** There exists a (measurable) function  $\kappa : \Xi \rightarrow \mathbb{R}_+$  such that its moment-generating function  $M_\kappa(t)$  is finite valued for all  $t$  in a neighborhood of zero and

$$|F(x', \xi) - F(x, \xi)| \leq \kappa(\xi) \|x' - x\| \quad (5.111)$$

for a.e.  $\xi \in \Xi$  and all  $x', x \in X$ .

Of course, it follows from (5.110) that

$$M_{x',x}(t) \leq \exp(\sigma^2 t^2/2), \quad \forall x', x \in X, \quad \forall t \in \mathbb{R}, \quad (5.112)$$

where

$$\sigma^2 := \sup_{x',x \in X} \sigma_{x',x}^2. \quad (5.113)$$

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  $\sigma_{x',x}^2 := \text{Var}[Y_{x',x}]$ .

The assumption (M5) implies that the expectation  $\mathbb{E}[\kappa(\xi)]$  is finite and the function  $f(x)$  is Lipschitz continuous on  $X$  with Lipschitz constant  $L = \mathbb{E}[\kappa(\xi)]$ . It follows that the optimal value  $\vartheta^*$  of the true problem is finite, provided the set  $X$  is bounded. (Recall that it was assumed that  $X$  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

$$\Pr(\hat{\kappa}_N > L') \leq \exp(-N\beta), \quad (5.114)$$

where  $\hat{\kappa}_N := N^{-1} \sum_{j=1}^N \kappa(\xi^j)$ . Note that it follows from (5.111) that w.p. 1

$$|\hat{f}_N(x') - \hat{f}_N(x)| \leq \hat{\kappa}_N \|x' - x\|, \quad \forall x', x \in X, \quad (5.115)$$

i.e.,  $\hat{f}_N(\cdot)$  is Lipschitz continuous on  $X$  with Lipschitz constant  $\hat{\kappa}_N$ .

By  $D := \sup_{x,x' \in X} \|x' - x\|$  we denote the diameter of the set  $X$ . Of course, the set  $X$  is bounded iff its diameter is finite. We also use notation  $a \vee b := \max\{a, b\}$  for numbers  $a, b \in \mathbb{R}$ .

**Theorem 5.18.** *Suppose that assumptions (M1) and (M4)–(M5) hold, with the corresponding constant  $\sigma^2$  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*

$$N \geq \frac{8\sigma^2}{(\varepsilon - \delta)^2} \left[ n \ln \left( \frac{8\varrho L' D}{\varepsilon - \delta} \right) + \ln \left( \frac{2}{\alpha} \right) \right] \vee \left[ \beta^{-1} \ln \left( \frac{2}{\alpha} \right) \right], \quad (5.116)$$

it follows that

$$\Pr(\hat{S}_N^\delta \subset S^\varepsilon) \geq 1 - \alpha. \quad (5.117)$$

**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  $\bar{x}_1, \dots, \bar{x}_M \in X$  be such that for every  $x \in X$  there exists  $\bar{x}_i$ ,  $i \in \{1, \dots, M\}$ , such that  $\|x - \bar{x}_i\| \leq \nu$ , i.e., the set  $X' := \{\bar{x}_1, \dots, \bar{x}_M\}$  forms a  $\nu$ -net in  $X$ . We can choose this net in such a way that

$$M \leq (\varrho D/\nu)^n \quad (5.118)$$

for a constant  $\varrho > 0$ . If the  $X' \setminus S^{\varepsilon'}$  is empty, then any point of  $X'$  is an  $\varepsilon'$ -optimal solution of the true problem. Otherwise, choose a mapping  $u : X' \setminus S^{\varepsilon'} \rightarrow S$  and consider the sets  $\tilde{S} := \cup_{x \in X'} \{u(x)\}$  and  $\tilde{X} := X' \cup \tilde{S}$ . Note that  $\tilde{X} \subset X$  and  $|\tilde{X}| \leq (2\varrho D/\nu)^n$ . Now let

us replace the set  $X$  by its subset  $\tilde{X}$ . 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 solution 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  $\delta'$ -optimal solution of the reduced SAA problem is an  $\varepsilon'$ -optimal solutions of the reduced (and hence unreduced) true problem provided that

$$N \geq \frac{8\sigma^2}{(\varepsilon - \delta)^2} \left[ n \ln \left( \frac{8\rho L' D}{\varepsilon - \delta} \right) + \ln \left( \frac{2}{\alpha} \right) \right]. \quad (5.119)$$

(Note that the right-hand side of (5.119) is greater than or equal to the estimate

$$\frac{2\sigma^2}{(\varepsilon' - \delta')^2} \ln \left( \frac{2|\tilde{X}|}{\alpha} \right)$$

required by Theorem 5.17.) We also have by (5.114) that for

$$N \geq \beta^{-1} \ln \left( \frac{2}{\alpha} \right), \quad (5.120)$$

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  $\delta$ -optimal solution of the (unreduced) SAA problem. Then there is a point  $x' \in \tilde{X}$  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  $\delta'$ -optimal solution of the reduced SAA problem, provided that  $\hat{\kappa}_N \leq L'$ . Consequently, we have that  $x'$  is an  $\varepsilon'$ -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

$$f(\hat{x}) \leq f(x') + L\nu \leq f(x') + L'\nu \leq \vartheta^* + \varepsilon' + L'\nu = \vartheta^* + \varepsilon.$$

We obtain that if  $N$  satisfies both inequalities (5.119) and (5.120), then with probability at least  $1 - \alpha$ , any  $\delta$ -optimal solution of the SAA problem is an  $\varepsilon$ -optimal solution of the true problem. The required estimate (5.116) follows.  $\square$

It is also possible to derive sample size estimates of the form (5.116) directly from the uniform exponential bounds derived in section 7.2.9; see Theorem 7.67 in particular.

**Remark 13.** 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  $0 < \varepsilon - \delta \leq a\sigma^2$ . (See Remark 12 on page 185.)

In a sense, the above estimate (5.116) of the sample size gives an estimate of *complexity* 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  $\sigma^2/\varepsilon^2$ , which suggests complexity of order  $\sigma^2/\varepsilon^2$  with respect to the desirable accuracy. This is in a sharp contrast to deterministic (convex) optimization, where complexity usually is bounded in terms of  $\ln(\varepsilon^{-1})$ . It seems that such dependence on  $\sigma$  and  $\varepsilon$  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 the probability distribution of the random variable  $F(x, \xi) - F(x', \xi)$  to have sufficiently light tails. In a sense, the constant  $\sigma^2$  can be viewed as a bound reflecting variability of the random variables  $F(x, \xi) - F(x', \xi)$  for  $x, x' \in X$ . Naturally, larger variability of the data should result in more difficulty in solving the problem. (See Remark 11 on page 184.)

This suggests that by using Monte Carlo sampling techniques one can solve two-stage 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*. 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 us 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.

Next we discuss 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  $x', x \in X$  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

$$M_{x',x}(t) \leq \exp(\lambda^2 \|x' - x\|^2 t^2 / 2), \quad \forall t \in \mathbb{R}. \quad (5.121)$$

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  $X$  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*

$$N \geq \frac{O(1)\lambda^2 D^2}{(\varepsilon - \delta)^2} \left[ n \ln \left( \frac{O(1)LD}{\varepsilon - \delta} \right) + \ln \left( \frac{1}{\alpha} \right) \right], \quad (5.122)$$

*it follows that*

$$\Pr(\hat{S}_N^\delta \subset S^\varepsilon) \geq 1 - \alpha. \quad (5.123)$$



For example, suppose that the Lipschitz constant  $\kappa(\xi)$  in assumption (M5) can be taken independent of  $\xi$ . That is, there exists a constant  $L > 0$  such that

$$|F(x', \xi) - F(x, \xi)| \leq L\|x' - x\| \quad (5.124)$$

for a.e.  $\xi \in \Xi$  and all  $x', x \in X$ . It follows that the expectation function  $f(x)$  is also Lipschitz continuous on  $X$  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 (7.186)) that

$$M_{x',x}(t) \leq \exp(2L^2\|x' - x\|^2 t^2), \quad \forall t \in \mathbb{R}. \quad (5.125)$$

Consequently, we can take  $\lambda = 2L$  in (5.121) and the estimate (5.122) takes the form

$$N \geq \left( \frac{O(1)LD}{\varepsilon - \delta} \right)^2 \left[ n \ln \left( \frac{O(1)LD}{\varepsilon - \delta} \right) + \ln \left( \frac{1}{\alpha} \right) \right]. \quad (5.126)$$

**Remark 14.** It was assumed in Theorem 5.18 that the set  $X$  has a finite diameter, i.e., that  $X$  is bounded. For convex problems, this assumption can be relaxed. Assume that the problem is convex, the optimal value  $\vartheta^*$  of the true problem is finite, and for some  $a > \varepsilon$  the set  $S^a$  has a finite diameter  $D_a^*$ . (Recall that  $S^a := \{x \in X : f(x) \leq \vartheta^* + a\}$ .) We refer here to the respective true and SAA problems, obtained by replacing the feasible set  $X$  by its subset  $S^a$ , as reduced problems. Note that the set  $S^\varepsilon$ , of  $\varepsilon$ -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  $\delta$ -optimal solutions of the reduced SAA problem are  $\varepsilon$ -optimal solutions of the true problem. Let us observe now that in this case the set of  $\delta$ -optimal solutions of the reduced SAA problem coincides with the set of  $\delta$ -optimal solutions of the original SAA problem. Indeed, suppose that the original SAA problem has a  $\delta$ -optimal solution  $x^* \in X \setminus S^a$ . Let  $\bar{x} \in \arg \min_{x \in S^a} \hat{f}_N(x)$ , 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  $\delta$ -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  $\delta$ -optimal solution of the SAA problem is an  $\varepsilon$ -optimal solution of the true problem. Also, assumptions (M4) and (M5) should be verified for  $x, x'$  in the set  $S^a$  only.

**Remark 15.** 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 assumption (M4) to verify condition (5.110) only for every  $x \in X \setminus S^{\varepsilon'}$  and  $x' := u(x)$ , where  $u : X \setminus S^{\varepsilon'} \rightarrow S$  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  $x \in X \setminus S^{\varepsilon'}$  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, assumption (M6) holds, then for such  $x$  and  $x'$  we have  $\sigma_{x',x}^2 \leq \lambda^2 \bar{D}^2$ , where  $\bar{D} := \sup_{x \in X \setminus S^{\varepsilon'}} \text{dist}(x, S)$ . Suppose, further, that the problem is convex. Then (see Remark 14) for any  $a > \varepsilon$ , we can use  $S^a$

instead of  $X$ . Therefore, if the problem is convex and the assumption (M6) holds, we can write the following estimate of the required sample size:

$$N \geq \frac{O(1)\lambda^2 \bar{D}_{a,\varepsilon}^2}{\varepsilon - \delta} \left[ n \ln \left( \frac{O(1)LD_a^*}{\varepsilon - \delta} \right) + \ln \left( \frac{1}{\alpha} \right) \right], \quad (5.127)$$

where  $D_a^*$  is the diameter of  $S^a$  and  $\bar{D}_{a,\varepsilon} := \sup_{x \in S^a \setminus S^{\varepsilon'}} \text{dist}(x, S)$ .

**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  $\gamma \geq 1$ ,  $c > 0$ , and  $r > 0$ , the following growth condition holds:*

$$f(x) \geq \vartheta^* + c [\text{dist}(x, S)]^\gamma, \quad \forall x \in S^r. \quad (5.128)$$

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

$$N \geq \frac{O(1)\lambda^2}{c^{2/\gamma} \varepsilon^{2(\gamma-1)/\gamma}} \left[ n \ln \left( \frac{O(1)LD_a^*}{\varepsilon} \right) + \ln \left( \frac{1}{\alpha} \right) \right], \quad (5.129)$$

it follows that

$$\Pr(\hat{S}_N^\delta \subset S^\varepsilon) \geq 1 - \alpha. \quad (5.130)$$

**Proof.** It follows from (5.128) that for any  $a \leq r$  and  $x \in S^a$ , the inequality  $\text{dist}(x, S) \leq (a/c)^{1/\gamma}$  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).  $\square$

Note that since  $a = \min\{2\varepsilon, r\} \leq r$ , we have that  $S^a \subset S^r$ , and if  $S = \{x^*\}$  is a singleton, then it follows from (5.128) that  $D_a^* \leq 2(a/c)^{1/\gamma}$ . 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:

$$N \geq O(1)c^{-2}\lambda^2 \left[ n \ln (O(1)c^{-1}L) + \ln (\alpha^{-1}) \right], \quad (5.131)$$

which does not depend on  $\varepsilon$ . 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  $X := \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ . Suppose, further, that random vector  $\xi$  has normal distribution  $\mathcal{N}(0, \sigma^2 I_n)$ , where  $\sigma^2$  is a positive constant and  $I_n$  is the  $n \times n$  identity matrix, i.e., components  $\xi_i$  of  $\xi$  are independent and  $\xi_i \sim \mathcal{N}(0, \sigma^2)$ ,  $i = 1, \dots, n$ . It follows that  $f(x) = \|x\|^{2m}$ , and hence for  $\varepsilon \in [0, 1]$  the set of  $\varepsilon$ -optimal solutions of the true problem is given by  $\{x : \|x\|^{2m} \leq \varepsilon\}$ . Now let  $\xi^1, \dots, \xi^N$

be an iid random sample of  $\xi$  and  $\bar{\xi}_N := (\xi^1 + \dots + \xi^N)/N$ . The corresponding sample average function is

$$\hat{f}_N(x) = \|x\|^{2m} - 2m \bar{\xi}_N^\top x, \quad (5.132)$$

and the optimal solution  $\hat{x}_N$  of the SAA problem is  $\hat{x}_N = \|\bar{\xi}_N\|^{-b} \bar{\xi}_N$ , where

$$b := \begin{cases} \frac{2m-2}{2m-1} & \text{if } \|\bar{\xi}_N\| \leq 1, \\ 1 & \text{if } \|\bar{\xi}_N\| > 1. \end{cases}$$

It follows that for  $\varepsilon \in (0, 1)$ , the optimal solution of the corresponding SAA problem is an  $\varepsilon$ -optimal solution of the true problem iff  $\|\bar{\xi}_N\|^v \leq \varepsilon$ , where  $v := \frac{2m}{2m-1}$ . 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  $\|\bar{\xi}_N\|^v > \varepsilon$  is equal to the probability  $\Pr(\chi_n^2 > N \varepsilon^{2/v} / \sigma^2)$ . 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

$$N > \frac{n\sigma^2}{\varepsilon^{2/v}} \quad (5.133)$$

in order to have the property, “with probability  $1 - \alpha$  an (exact) optimal solution of the SAA problem is an  $\varepsilon$ -optimal solution of the true problem.” Compared with (5.116), the lower bound (5.133) also grows linearly in  $n$  and is proportional to  $\sigma^2 / \varepsilon^{2/v}$ . It remains to note that the constant  $v$  decreases to 1 as  $m$  increases.

Note that in this example the growth condition (5.128) holds with  $\gamma = 2m$  and that the power constant of  $\varepsilon$  in the estimate (5.133) is in accordance with the estimate (5.129). Note also that here

$$[F(x', \xi) - f(x')] - [F(x, \xi) - f(x)] = 2m \xi^\top (x - x')$$

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  $N$  satisfying the estimate (5.133) in order to produce an  $\varepsilon$ -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  $x^* \in X$  is a sharp (optimal) solution of the true problem (5.1) if there exists constant  $c > 0$  such that

$$f(x) \geq f(x^*) + c\|x - x^*\|, \quad \forall x \in X. \quad (5.134)$$

Condition (5.134) corresponds to 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  $X$  we have that the tangent cone  $\mathcal{T}_X(x^*)$ , to  $X$  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

$$f'(x^*, h) \geq c\|h\|, \quad \forall h \in \mathcal{T}_X(x^*). \quad (5.135)$$

Since condition (5.135) is local, we have that it actually suffices to verify (5.134) for all  $x \in X$  in a neighborhood of  $x^*$ .

**Theorem 5.23.** *Suppose that the problem is convex and assumption (M1) holds, and let  $x^* \in X$  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*

$$1 - \Pr(\hat{S}_N = \{x^*\}) \leq Ce^{-N\beta}; \quad (5.136)$$

*i.e., the probability of the event that “ $x^*$  is the unique optimal solution of the SAA problem” converges to 1 exponentially fast with the 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 7.54). It follows w.p. 1 for  $N$  large enough that

$$\hat{f}'_N(x^*, h) \geq (c/2)\|h\|, \quad \forall h \in \mathcal{T}_X(x^*), \quad (5.137)$$

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  $\hat{f}'_N(x^*, \cdot)$  converges to  $f'(x^*, \cdot)$  exponentially fast on the unit sphere. (See inequality (7.219) of Theorem 7.69.) By taking  $\varepsilon := c/2$  in (7.219), we can conclude that (5.136) follows.  $\square$

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:

$$f(x) \geq \vartheta^* + c[\text{dist}(x, S)], \quad \forall x \in X. \quad (5.138)$$

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  $X$  is a convex closed polyhedron.
- (P2) The support set  $\Xi = \{\xi_1, \dots, \xi_K\}$  is finite.
- (P3) For every  $\xi \in \Xi$  the function  $F(\cdot, \xi)$  is polyhedral.

Conditions (P1)–(P3) hold in the 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  $\Xi$ , it is possible to show the following result (cf. [208]).

**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*

$$1 - \Pr(\hat{S}_N \neq \emptyset \text{ and } \hat{S}_N \text{ is a face of } S) \leq Ce^{-N\beta}; \quad (5.139)$$

*i.e., the probability of the event that “ $\hat{S}_N$  is nonempty and forms a face of the set  $S$ ” converges to 1 exponentially fast with the 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  $\xi$ ) is small, the Monte Carlo approach may not be 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 [138], 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  $\xi$  be a real valued random variable having cdf  $H(z) = \Pr(\xi \leq z)$ . Suppose that we want to evaluate the expectation

$$\mathbb{E}[F(\xi)] = \int_{-\infty}^{+\infty} F(z) dH(z), \quad (5.140)$$

where  $F : \mathbb{R} \rightarrow \mathbb{R}$  is a measurable function. Let  $U \sim U[0, 1]$ , i.e.,  $U$  is a random variable uniformly distributed on  $[0, 1]$ . Then random variable<sup>22</sup>  $H^{-1}(U)$  has cdf  $H(\cdot)$ . Therefore, by making a change of variables we can write the expectation (5.140) as

$$\mathbb{E}[\psi(U)] = \int_0^1 \psi(u) du, \quad (5.141)$$

where  $\psi(u) := F(H^{-1}(u))$ .

Evaluation of the above expectation by the Monte Carlo method is based on generating an iid sample  $U^1, \dots, U^N$  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

$$\int_0^1 \psi(u) du \approx \frac{1}{N} \sum_{j=1}^N \psi(u_j) \quad (5.142)$$

by using some points  $u_j \in [(j-1)/N, j/N]$ ,  $j = 1, \dots, N$ , e.g., taking midpoints  $u_j := (2j-1)/(2N)$  of equally spaced partition intervals  $[(j-1)/N, j/N]$ ,  $j = 1, \dots, N$ .

<sup>22</sup>Recall that  $H^{-1}(u) := \inf\{z : H(z) \geq u\}$ .

If the function  $\psi(u)$  is Lipschitz continuous on  $[0,1]$ , then the error of the Riemann sum approximation<sup>23</sup> 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  $U^1, \dots, U^N$  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 in the possibility of estimating the approximation error by calculating the sample variance,

$$s^2 := (N-1)^{-1} \sum_{j=1}^N [\psi(U^j) - \bar{\psi}_N]^2,$$

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, 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 an 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  $\psi : I^d \rightarrow \mathbb{R}$  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  $N = M^d$  subintervals<sup>24</sup> 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  $\text{Var}[\psi(U)]$ . Yet the problem of uneven covering of  $I^d$  by an iid sample  $U^j, j = 1, \dots, N$ , remains persistent.

Quasi-Monte Carlo methods employ the approximation

$$\mathbb{E}[\psi(U)] \approx \frac{1}{N} \sum_{j=1}^N \psi(u_j) \quad (5.143)$$

for a carefully chosen (deterministic) sequence of points  $u_1, \dots, u_N \in I^d$ . From the numerical point of view, it is important to be able to generate such a sequence iteratively as an infinite sequence of points  $u_j, j = 1, \dots$ , in  $I^d$ . In that way, one does not need to recalculate already calculated function values  $\psi(u_j)$  with the increase of  $N$ . A basic requirement for this sequence is that the right-hand side of (5.143) converges to  $\mathbb{E}[\psi(U)]$

<sup>23</sup>If  $\psi(u)$  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  $\xi$  is supported on the whole real line, e.g., is normal, then  $H^{-1}(u)$  tends to  $\infty$  as  $u$  tends to 0 or 1. In that case,  $\psi(u)$  typically will be discontinuous at  $u = 0$  and  $u = 1$ .

<sup>24</sup>A set  $A \subset \mathbb{R}^d$  is said to be a ( $d$ -dimensional) interval if  $A = [a_1, b_1] \times \dots \times [a_d, b_d]$ .

as  $N \rightarrow \infty$ . It is not difficult to show that this holds (for any Riemann-integrable function  $\psi(\mathbf{u})$ ) if

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \mathbf{1}_A(\mathbf{u}_j) = V_d(A) \quad (5.144)$$

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  $\{\mathbf{u}_1, \dots, \mathbf{u}_N\} \subset I^d$  is defined by

$$\mathcal{D}^*(\mathbf{u}_1, \dots, \mathbf{u}_N) := \sup_{A \in \mathcal{I}} \left| \frac{1}{N} \sum_{j=1}^N \mathbf{1}_A(\mathbf{u}_j) - V_d(A) \right|, \quad (5.145)$$

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  $\mathbf{u}_j \in I^d$ ,  $j = 1, \dots$ , condition (5.144) holds iff  $\lim_{N \rightarrow \infty} \mathcal{D}^*(\mathbf{u}_1, \dots, \mathbf{u}_N) = 0$ . A more important property of the star discrepancy is that it is possible to give error bounds in terms of  $\mathcal{D}^*(\mathbf{u}_1, \dots, \mathbf{u}_N)$  for quasi-Monte Carlo approximations. Let us start with the one-dimensional case. Recall that variation of a function  $\psi : [0, 1] \rightarrow \mathbb{R}$  is the sup  $\sum_{i=1}^m |\psi(t_i) - \psi(t_{i-1})|$ , where the supremum is taken over all partitions  $0 = t_0 < t_1 < \dots < t_m = 1$  of the interval  $[0, 1]$ . It is said that  $\psi$  has bounded variation if its variation is finite.

**Theorem 5.26 (Koksma).** If  $\psi : [0, 1] \rightarrow \mathbb{R}$  has bounded variation  $V(\psi)$ , then for any  $u_1, \dots, u_N \in [0, 1]$  we have

$$\left| \frac{1}{N} \sum_{j=1}^N \psi(u_j) - \int_0^1 \psi(u) du \right| \leq V(\psi) \mathcal{D}^*(u_1, \dots, u_N). \quad (5.146)$$

**Proof.** We can assume that the sequence  $u_1, \dots, u_N$  is arranged in increasing order, and we set  $u_0 = 0$  and  $u_{N+1} = 1$ . That is,  $0 = u_0 \leq u_1 \leq \dots \leq u_{N+1} = 1$ . 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 using summation by parts we have

$$\frac{1}{N} \sum_{j=1}^N \psi(u_j) = \psi(u_{N+1}) - \sum_{j=0}^N \frac{j}{N} [\psi(u_{j+1}) - \psi(u_j)];$$

we can write

$$\begin{aligned} \frac{1}{N} \sum_{j=1}^N \psi(u_j) - \int_0^1 \psi(u) du &= - \sum_{j=0}^N \frac{j}{N} [\psi(u_{j+1}) - \psi(u_j)] + \int_0^1 u d\psi(u) \\ &= \sum_{j=0}^N \int_{u_j}^{u_{j+1}} \left(u - \frac{j}{N}\right) d\psi(u). \end{aligned}$$

Also for any  $u \in [u_j, u_{j+1}]$ ,  $j = 0, \dots, N$ , we have

$$\left| u - \frac{j}{N} \right| \leq \mathcal{D}^*(u_1, \dots, u_N).$$

It follows that

$$\begin{aligned} \left| \frac{1}{N} \sum_{j=1}^N \psi(u_j) - \int_0^1 \psi(u) du \right| &\leq \sum_{j=0}^N \int_{u_j}^{u_{j+1}} \left| u - \frac{j}{N} \right| d\psi(u) \\ &\leq \mathcal{D}^*(u_1, \dots, u_N) \sum_{j=0}^N |\psi(u_{j+1}) - \psi(u_j)|, \end{aligned}$$

and, of course,  $\sum_{j=0}^N |\psi(u_{j+1}) - \psi(u_j)| \leq V(\psi)$ . This completes the proof.  $\square$

This can be extended to a multidimensional setting as follows. Consider a function  $\psi : I^d \rightarrow \mathbb{R}$ . The variation of  $\psi$ , in the sense of Vitali, is defined as

$$V^{(d)}(\psi) := \sup_{\mathcal{P} \in \mathcal{J}} \sum_{A \in \mathcal{P}} |\Delta_\psi(A)|, \quad (5.147)$$

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  $\psi$  at the vertices of  $A$  (i.e., function values at adjacent vertices have opposite signs). The variation of  $\psi$ , in the sense of Hardy and Krause, is defined as

$$V(\psi) := \sum_{k=1}^d \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq d} V^{(k)}(\psi; i_1, \dots, i_k), \quad (5.148)$$

where  $V^{(k)}(\psi; i_1, \dots, i_k)$  is the variation in the sense of Vitali of restriction of  $\psi$  to the  $k$ -dimensional face of  $I^d$  defined by  $u_j = 1$  for  $j \notin \{i_1, \dots, i_k\}$ .

**Theorem 5.27 (Hlawka).** *If  $\psi : I^d \rightarrow \mathbb{R}$  has bounded variation  $V(\psi)$  on  $I^d$  in the sense of Hardy and Krause, then for any  $\mathbf{u}_1, \dots, \mathbf{u}_N \in I^d$  we have*

$$\left| \frac{1}{N} \sum_{j=1}^N \psi(\mathbf{u}_j) - \int_{I^d} \psi(\mathbf{u}) d\mathbf{u} \right| \leq V(\psi) \mathcal{D}^*(\mathbf{u}_1, \dots, \mathbf{u}_N). \quad (5.149)$$

In order to see how good the above error estimates could be, let us consider the one-dimensional case with  $u_j := (2j-1)/(2N)$ ,  $j = 1, \dots, N$ . Then  $\mathcal{D}^*(u_1, \dots, u_N) = 1/(2N)$ , 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  $\psi$  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  $\lim_{u \rightarrow +\infty} H^{-1}(u)$  and  $\lim_{u \rightarrow -\infty} H^{-1}(u)$ . Therefore, if one of these 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

$$V(\psi) = F\left(\lim_{u \rightarrow +\infty} H^{-1}(u)\right) - F\left(\lim_{u \rightarrow -\infty} H^{-1}(u)\right).$$



This overestimation of the corresponding constant becomes even worse with an increase in the dimension  $d$ .

A sequence  $\{\mathbf{u}_j\}_{j \in \mathbb{N}} \subset I^d$  is called a *low-discrepancy sequence* if  $\mathcal{D}^*(u_1, \dots, u_N)$  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  $\mathcal{D}^*(u_1, \dots, u_N)$  always greater than or equal to  $1/(2N)$  and this lower bound is attained for  $u_j := (2j - 1)/2N$ ,  $j = 1, \dots, N$ . 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  $u_1, \dots$  in  $[0, 1]$  such that  $\mathcal{D}^*(u_1, \dots, u_N) \leq c/N$  for some  $c > 0$  and all  $N \in \mathbb{N}$ . It is possible to show that a best possible for  $\mathcal{D}^*(u_1, \dots, u_N)$ , for a sequence of points  $u_j \in [0, 1]$ ,  $j = 1, \dots$ , is of order  $O(N^{-1} \ln N)$ . We are now going to construct a sequence for which this rate is attained.

For any integer  $n \geq 0$  there is a unique digit expansion

$$n = \sum_{i \geq 0} a_i(n) b^i \quad (5.150)$$

in integer base  $b \geq 2$ , where  $a_i(n) \in \{0, 1, \dots, b-1\}$ ,  $i = 0, 1, \dots$ , 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

$$\phi_b(n) := \sum_{i \geq 0} a_i(n) b^{-i-1}. \quad (5.151)$$

Note that

$$\phi_b(n) \leq (b-1) \sum_{i=0}^{\infty} b^{-i-1} = 1,$$

and hence  $\phi_b(n) \in [0, 1]$  for any integer  $n \geq 0$ .

**Definition 5.28.** For an integer  $b \geq 2$ , the van der Corput sequence in base  $b$  is the sequence  $u_j := \phi_b(j)$ ,  $j = 0, 1, \dots$

It is possible to show that to every van der Corput sequence  $u_1, \dots$ , in base  $b$ , corresponds constant  $C_b$  such that

$$\mathcal{D}^*(u_1, \dots, u_n) \leq C_b N^{-1} \ln N \quad \forall N \in \mathbb{N}.$$

A classical extension of van der Corput sequences to multidimensional settings is the following. Let  $p_1 = 2, p_2 = 3, \dots, p_d$  be the first  $d$  prime numbers. Then the *Halton sequence*, in the bases  $p_1, \dots, p_d$ , is defined as

$$\mathbf{u}_j := (\phi_{p_1}(j), \dots, \phi_{p_d}(j)) \in I^d, \quad j = 0, 1, \dots \quad (5.152)$$

It is possible to show that for that sequence,

$$\mathcal{D}^*(\mathbf{u}_1, \dots, \mathbf{u}_N) \leq A_d N^{-1} (\ln N)^d + O(N^{-1} (\ln N)^{d-1}) \quad \forall N \geq 2, \quad (5.153)$$

where  $A_d = \prod_{i=1}^d \frac{p_i-1}{2 \ln p_i}$ . By 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  $d \rightarrow \infty$ . 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  $\{\mathbf{u}_j\}_{j \in \mathbb{N}}$  is that there is no easy way to estimate the error of the corresponding approximations  $N^{-1} \sum_{j=1}^N \psi(\mathbf{u}_j)$ . 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  $\{\mathbf{u}_1, \dots, \mathbf{u}_N\}$ , of generating points in  $I^d$  without destroying its regular structure. Such a simple randomization procedure was suggested by Cranley and Patterson [39]. That is, generate a random point  $\mathbf{u}$  uniformly distributed over  $I^d$ , and use the randomization<sup>25</sup>  $\tilde{\mathbf{u}}_j := (\mathbf{u}_j + \mathbf{u}) \bmod 1$ ,  $j = 1, \dots, N$ . It is not difficult to show that (marginal) distribution of each random vector  $\tilde{\mathbf{u}}_j$  is uniform on  $I^d$ . Therefore, each  $\psi(\tilde{\mathbf{u}}_j)$ , and hence  $N^{-1} \sum_{j=1}^N \psi(\tilde{\mathbf{u}}_j)$ , is an unbiased estimator of the corresponding expectation  $\mathbb{E}[\psi(\mathbf{U})]$ . Variance of the estimator  $N^{-1} \sum_{j=1}^N \psi(\tilde{\mathbf{u}}_j)$  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  $\mathbf{u} = \mathbf{u}^i$ ,  $i = 1, \dots, M$ , and consequently averaging the obtained replications of  $N^{-1} \sum_{j=1}^N \psi(\tilde{\mathbf{u}}_j)$ . Simultaneously, variance of this estimator can be evaluated by calculating the sample variance of the obtained  $M$  independent replications of  $N^{-1} \sum_{j=1}^N \psi(\tilde{\mathbf{u}}_j)$ .

## 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  $\sigma^2(x) := \text{Var}[F(x, \xi)]$ . 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 then write

$$\mathbb{E}[F(x, \xi)] = \int_{-\infty}^{+\infty} F(x, \xi) dH(\xi). \quad (5.154)$$

<sup>25</sup>For a number  $a \in \mathbb{R}$  the notation “ $a \bmod 1$ ” denotes the fractional part of  $a$ , i.e.,  $a \bmod 1 = a - [a]$ , where  $[a]$  denotes the largest integer less than or equal to  $a$ . In the vector case, the “modulo 1” reduction is understood coordinatewise.

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. (This was already discussed in section 5.4.) That is, we can generate independent random points<sup>26</sup>

$$U^j \sim U[(j-1)/N, j/N], \quad j = 1, \dots, N, \quad (5.155)$$

and then construct the random sample of  $\xi$  by the inverse transformation  $\xi^j := H^{-1}(U^j)$ ,  $j = 1, \dots, N$  (compare with (5.141)).

Now suppose that  $j$  is chosen at random from the set  $\{1, \dots, N\}$  (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  $\{j_1, \dots, j_N\}$  be a random permutation of the set  $\{1, \dots, N\}$ . Then the random variables  $\xi^{j_1}, \dots, \xi^{j_N}$  have the same marginal distribution, with the same cdf  $H(\cdot)$ , and are negatively correlated with each other. Therefore, the expected value of

$$\hat{f}_N(x) = \frac{1}{N} \sum_{i=1}^N F(x, \xi^i) = \frac{1}{N} \sum_{s=1}^N F(x, \xi^{j_s}) \quad (5.156)$$

is  $f(x)$ , while

$$\mathbb{V}\text{ar}[\hat{f}_N(x)] = N^{-1}\sigma^2(x) + 2N^{-2} \sum_{s < t} \mathbb{C}\text{ov}(F(x, \xi^{j_s}), F(x, \xi^{j_t})). \quad (5.157)$$

If the function  $F(x, \cdot)$  is monotonically increasing or decreasing, then 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  $\xi = (\xi_1, \dots, \xi_d)$  is  $d$ -dimensional and that its components  $\xi_i$ ,  $i = 1, \dots, d$ , are distributed independently of each other. Then we can use the above procedure for each component  $\xi_i$ . That is, a random sample  $U^j$  of the form (5.155) is generated, and consequently  $N$  replications of the first component of  $\xi$  are computed by the corresponding inverse transformation applied to randomly permuted  $U^{j_s}$ . The same procedure is applied to every component of  $\xi$  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.,  $F(x, \xi) := F_1(x, \xi_1) + \dots + F_d(x, \xi_d)$ , then  $\mathbb{E}[F(x, \xi)] = \mathbb{E}[F_1(x, \xi_1)] + \dots + \mathbb{E}[F_d(x, \xi_d)]$ , 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 a 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.

<sup>26</sup>For 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

$$\mathbb{V}\text{ar}[F(x, \xi) + tA(x, \xi)] = \mathbb{V}\text{ar}[F(x, \xi)] + t^2 \mathbb{V}\text{ar}[A(x, \xi)] + 2t \mathbb{C}\text{ov}(F(x, \xi), A(x, \xi)).$$

It follows that the above variance attains its minimum, with respect to  $t$ , for

$$t^* := -\rho_{F,A}(x) \left[ \frac{\mathbb{V}\text{ar}(F(x, \xi))}{\mathbb{V}\text{ar}(A(x, \xi))} \right]^{1/2}, \quad (5.158)$$

where  $\rho_{F,A}(x) := \mathbb{C}\text{orr}(F(x, \xi), A(x, \xi))$ , and with

$$\mathbb{V}\text{ar}[F(x, \xi) + t^* A(x, \xi)] = \mathbb{V}\text{ar}[F(x, \xi)] [1 - \rho_{F,A}(x)^2]. \quad (5.159)$$

For a given  $x \in X$  and generated sample  $\xi^1, \dots, \xi^N$ , one can estimate, in the standard way, the covariance and variances appearing in the right-hand side of (5.158), and hence construct an estimate  $\hat{t}^*$  of  $t^*$ . Then  $f(x)$  can be estimated by

$$\hat{f}_N^A(x) := \frac{1}{N} \sum_{j=1}^N [F(x, \xi^j) + \hat{t}^* A(x, \xi^j)]. \quad (5.160)$$

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

$$\mathbb{E}[(h - \mu)^T T] = \mathbb{E}[(h - \mu)^T] \mathbb{E}[T] = 0,$$

and hence one can use  $A(x, \xi) := (h - \mu)^T T x$  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  $A_1(x, \xi), \dots, A_m(x, \xi)$ , each with zero mean and highly correlated with  $F(x, \xi)$ , are available.

### 5.5.3 Importance Sampling and Likelihood Ratio Methods

Suppose that  $\xi$  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  $\mathbb{E}_\psi$  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

$$\psi^*(\cdot) := \frac{|F(x, \cdot)h(\cdot)|}{\int |F(x, \zeta)h(\zeta)|d\zeta}. \quad (5.162)$$

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

$$\mathbb{E}_\psi[F(x, Z)^2L(Z)^2] = \int \frac{F(x, \zeta)^2h(\zeta)^2}{\psi(\zeta)}d\zeta \quad (5.163)$$

is minimized. Furthermore, by the Cauchy inequality we have

$$\left( \int |F(x, \zeta)h(\zeta)|d\zeta \right)^2 \leq \left( \int \frac{F(x, \zeta)^2h(\zeta)^2}{\psi(\zeta)}d\zeta \right) \left( \int \psi(\zeta)d\zeta \right). \quad (5.164)$$

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  $\zeta^1, \dots, \zeta^N$  from the pdf  $\psi(\cdot)$ , and then estimating  $f(x)$  by

$$\tilde{f}_N^\psi(x) := \frac{1}{N} \sum_{j=1}^N F(x, \zeta^j)L(\zeta^j). \quad (5.165)$$

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  $\xi$  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

$$Q(z) := \inf \{q^\top y : Wy = z, y \geq 0\}.$$

Suppose, further, that  $h$  has a continuous distribution with pdf  $\eta(\cdot)$ . We have that

$$\mathbb{E}[Q(x, \xi)] = \mathbb{E}_T \{ \mathbb{E}_{h|T}[Q(x, \xi)] \},$$

and by using the transformation  $z = h - Tx$ , since  $h$  and  $T$  are independent we obtain

$$\begin{aligned} \mathbb{E}_{h|T}[Q(x, \xi)] &= \mathbb{E}_h[Q(x, \xi)] \\ &= \int Q(h - Tx) \eta(h) dh = \int Q(z) \eta(z + Tx) dz \\ &= \int Q(\zeta) L(x, \zeta) \psi(\zeta) d\zeta = \mathbb{E}_\psi [L(x, Z) Q(Z)], \end{aligned} \quad (5.166)$$

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  $\xi$  has a continuous distribution.

## 5.6 Validation Analysis

Suppose that we are given a feasible point  $\bar{x} \in X$  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

$$\text{gap}(\bar{x}) := f(\bar{x}) - \vartheta^* \quad (5.167)$$

associated with the candidate solution  $\bar{x}$ . Clearly, for any feasible  $\bar{x} \in X$ ,  $\text{gap}(\bar{x})$  is non-negative and  $\text{gap}(\bar{x}) = 0$  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 (5.22).) This means that  $\hat{\vartheta}_N$  provides a valid *statistical lower bound* for the optimal value  $\vartheta^*$  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  $\hat{\vartheta}_N^1, \dots, \hat{\vartheta}_N^M$  be the computed optimal values of these SAA problems. Then

$$\bar{v}_{N,M} := \frac{1}{M} \sum_{m=1}^M \hat{\vartheta}_N^m \quad (5.168)$$

is an unbiased estimator of  $\mathbb{E}[\hat{\vartheta}_N]$ . Since the samples, and hence  $\hat{\vartheta}_N^1, \dots, \hat{\vartheta}_N^M$ , are independent and have the same distribution, we have that  $\mathbb{V}\text{ar}[\bar{v}_{N,M}] = M^{-1} \mathbb{V}\text{ar}[\hat{\vartheta}_N]$ , and hence we can estimate variance of  $\bar{v}_{N,M}$  by

$$\hat{\sigma}_{N,M}^2 := \frac{1}{M} \underbrace{\left[ \frac{1}{M-1} \sum_{m=1}^M \left( \hat{\vartheta}_N^m - \bar{v}_{N,M} \right)^2 \right]}_{\text{estimate of } \mathbb{V}\text{ar}[\hat{\vartheta}_N]}. \quad (5.169)$$

Note that the above make sense only if the optimal value  $\vartheta^*$  of the true problem is finite. Note also that the inequality  $\vartheta^* \geq \mathbb{E}[\hat{\vartheta}_N]$  holds and  $\hat{\vartheta}_N$  gives a valid statistical lower bound even if  $f(x) = +\infty$  for some  $x \in X$ . Note finally that the samples do not need to be iid (for example, one can use LH 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  $\hat{\vartheta}_N$ , and hence its replications  $\hat{\vartheta}_N^j$ , does not have a normal distribution, even approximately. (See Theorem 5.7 and the discussion that follows.) However, by the CLT, the probability distribution of the average  $\bar{v}_{N,M}$  becomes approximately normal as  $M$  increases. Therefore, we can use

$$L_{N,M} := \bar{v}_{N,M} - t_{\alpha, M-1} \hat{\sigma}_{N,M} \quad (5.170)$$

as an approximate  $100(1 - \alpha)\%$  confidence<sup>27</sup> 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 the iid sample, one can use the sample variance estimate

$$\hat{\sigma}_{N'}^2(\bar{x}) := \frac{1}{N'(N'-1)} \sum_{j=1}^{N'} \left[ F(\bar{x}, \xi^j) - \hat{f}_{N'}(\bar{x}) \right]^2. \quad (5.171)$$

Then

$$U_{N'}(\bar{x}) := \hat{f}_{N'}(\bar{x}) + z_{\alpha} \hat{\sigma}_{N'}(\bar{x}) \quad (5.172)$$

<sup>27</sup>Here  $t_{\alpha, \nu}$  is the  $\alpha$ -critical value of  $t$ -distribution with  $\nu$  degrees of freedom. This critical value is slightly bigger than the corresponding standard normal critical value  $z_{\alpha}$ , and  $t_{\alpha, \nu}$  quickly approaches  $z_{\alpha}$  as  $\nu$  increases.

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  $z_\alpha$  from the standard normal distribution rather than a  $t$ -distribution.

We have that

$$\mathbb{E}[\hat{f}_{N'}(\bar{x}) - \bar{v}_{N,M}] = f(\bar{x}) - \mathbb{E}[\hat{\vartheta}_N] = \text{gap}(\bar{x}) + \vartheta^* - \mathbb{E}[\hat{\vartheta}_N] \geq \text{gap}(\bar{x}),$$

i.e.,  $\hat{f}_{N'}(\bar{x}) - \bar{v}_{N,M}$  is a biased estimator of the  $\text{gap}(\bar{x})$ . 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

$$\hat{f}_{N'}(\bar{x}) - \bar{v}_{N,M} + z_\alpha \sqrt{\hat{\sigma}_{N'}^2(\bar{x}) + \hat{\sigma}_{N,M}^2} \quad (5.173)$$

provides a conservative  $100(1 - \alpha)\%$  confidence upper bound for the  $\text{gap}(\bar{x})$ . We say that this upper bound is “conservative” since in fact it gives a  $100(1 - \alpha)\%$  confidence upper bound for the  $\text{gap}(\bar{x}) + \vartheta^* - \mathbb{E}[\hat{\vartheta}_N]$ , and we have that  $\vartheta^* - \mathbb{E}[\hat{\vartheta}_N] \geq 0$ .

In order to calculate the estimate  $\hat{f}_{N'}(\bar{x})$ , one needs to compute the value  $F(\bar{x}, \xi^j)$  of the objective function for every generated sample realization  $\xi^j$ ,  $j = 1, \dots, N'$ . 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 estimate  $f(\bar{x})$  quite accurately. Evaluation of the optimal value  $\vartheta^*$  by employing the estimator  $\bar{v}_{N,M}$  is a more delicate problem.

There are two types of error in using  $\bar{v}_{N,M}$  as an estimator of  $\vartheta^*$ , 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 an increase of the sample size  $N$ . By Theorem 5.7 we have that, under mild regularity conditions,

$$\hat{\vartheta}_N = \inf_{x \in S} \hat{f}_N(x) + o_p(N^{-1/2}). \quad (5.174)$$

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 (5.29) and the following 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

$$\widehat{\text{gap}}_{N,M}(\bar{x}) := \frac{1}{M} \sum_{m=1}^M [\hat{f}_N^m(\bar{x}) - \hat{\vartheta}_N^m], \quad (5.175)$$

where  $\hat{f}_N^m(\bar{x})$  and  $\hat{\vartheta}_N^m$  are computed from the *same* sample  $m = 1, \dots, M$ . 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 16.** Of course, in order to obtain a valid statistical lower bound for the optimal value  $\vartheta^*$  we can use any (deterministic) lower bound for the optimal value  $\hat{\vartheta}_N$  of the corresponding SAA problem instead of  $\hat{\vartheta}_N$  itself. For example, suppose that the problem is *convex*. By convexity of  $\hat{f}_N(\cdot)$  we have that for any  $x' \in X$  and  $\gamma \in \partial \hat{f}_N(x')$  it holds that

$$\hat{f}_N(x) \geq \hat{f}_N(x') + \gamma^\top (x - x'), \quad \forall x \in \mathbb{R}^n. \quad (5.176)$$

Therefore, we can proceed as follows. Choose points  $x_1, \dots, x_r \in X$ , calculate subgradients  $\hat{\gamma}_{iN} \in \partial \hat{f}_N(x_i)$ ,  $i = 1, \dots, r$ , and solve the problem

$$\text{Min} \max_{x \in X} \left\{ \hat{f}_N(x) + \hat{\gamma}_{iN}^\top (x - x_i) \right\}. \quad (5.177)$$

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  $\vartheta^*$ . 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  $X$  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  $x_1, \dots, x_r \in X$ . Suppose that the expectation function  $f(x)$  is differentiable at the points  $x_1, \dots, x_r$ . 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

$$\text{Min} \max_{x \in X} \left\{ f(x) + \nabla f(x_i)^\top (x - x_i) \right\}, \quad (5.178)$$

provided that the set  $X$  is bounded. Again by convexity arguments, the optimal value of (5.178) is less than or equal to the optimal value  $\vartheta^*$  of the true problem. If we can find such

points  $x_1, \dots, x_r \in X$  that the optimal value of (5.178) is less than  $\vartheta^*$  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 lose the advantage of solving the easier problem (5.177).

A natural approach to choosing the required points and hence to applying the above procedure is the following. By solving (once) an SAA problem, find points  $x_1, \dots, x_r \in X$  such that the optimal value of the corresponding problem (5.177) provides us with high accuracy an estimate of the optimal value of this SAA problem. Use some (all) of these points to calculate lower bound estimates  $\hat{\lambda}_N^m, m = 1, \dots, M$ , 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  $\vartheta^*$  be the optimal value of this (true) minimax problem. Clearly for any  $\bar{y} \in Y$  we have that

$$\vartheta^* \geq \inf_{x \in X} f(x, \bar{y}). \quad (5.179)$$

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  $\vartheta^*$ , 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  $S_x \times S_y \subset X \times Y$  of saddle points, and hence its optimal value is equal to the optimal value of its dual problem (5.48). Then for any  $\bar{x} \in X$  we have that

$$\vartheta^* \leq \sup_{y \in Y} f(\bar{x}, y), \quad (5.180)$$

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  $\vartheta^*$  by averaging optimal values of sample average approximations of the right-hand side of (5.180). Of course, the 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 (5.11)–(5.13)). That is, consider the problem

$$\text{Min}_{x \in X_0} f(x) \text{ s.t. } g_i(x) \leq 0, \quad i = 1, \dots, p, \quad (5.181)$$

where  $X_0$  is a nonempty subset of  $\mathbb{R}^n$ ,  $f(x) := \mathbb{E}[F(x, \xi)]$ , and  $g_i(x) := \mathbb{E}[G_i(x, \xi)]$ ,  $i = 1, \dots, p$ . We have that

$$\vartheta^* = \inf_{x \in X_0} \sup_{\lambda \geq 0} L(x, \lambda), \quad (5.182)$$

where  $\vartheta^*$  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  $\bar{\lambda} \geq 0$ , we have that

$$\vartheta^* \geq \inf_{x \in X_0} L(x, \bar{\lambda}), \quad (5.183)$$

and the equality in (5.183) is attained if the problem (5.179) is convex and  $\bar{\lambda}$  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  $\vartheta^*$ .

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  $\bar{x} \in X_0$ , it happens that  $\hat{g}_{iN}(\bar{x}) > 0$  for some  $i \in \{1, \dots, p\}$ , then

$$\sup_{\lambda \geq 0} \left\{ \hat{f}_N(\bar{x}) + \sum_{i=1}^p \lambda_i \hat{g}_{iN}(\bar{x}) \right\} = +\infty. \quad (5.184)$$

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  $i \in \{1, \dots, p\}$ , 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  $\hat{g}_{iN'}(\bar{x}) + \kappa \hat{\sigma}_{iN'}^2(\bar{x}) \leq 0$ ,  $i = 1, \dots, p$ , where  $\hat{\sigma}_{iN'}^2(\bar{x})$  is a sample variance of  $\hat{g}_{iN'}(\bar{x})$  and  $\kappa$  is a positive constant chosen in such a way that the probability of  $g_i(\bar{x})$  being bigger than  $\hat{g}_{iN'}(\bar{x}) + \kappa \hat{\sigma}_{iN'}^2(\bar{x})$  is less than  $\alpha/p$  for all  $i \in \{1, \dots, p\}$ .

### 5.6.2 Statistical Testing of Optimality Conditions

Suppose that the feasible set  $X$  is defined by (equality and inequality) constraints in the form

$$X := \{x \in \mathbb{R}^n : g_i(x) = 0, \ i = 1, \dots, q; \ g_i(x) \leq 0, \ i = q + 1, \dots, p\}, \quad (5.185)$$

where  $g_i(x)$  are smooth (at least continuously differentiable) *deterministic* functions. Let  $x^* \in X$  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  $\lambda_i$  such that  $\lambda_i \geq 0$ ,  $i \in \mathcal{I}(x^*)$  and

$$\nabla f(x^*) + \sum_{i \in \mathcal{J}(x^*)} \lambda_i \nabla g_i(x^*) = 0, \quad (5.186)$$

where  $\mathcal{I}(x) := \{i : g_i(x) = 0, \ i = q + 1, \dots, p\}$  denotes the index set of inequality constraints active at a point  $x \in \mathbb{R}^n$ , and  $\mathcal{J}(x) := \{1, \dots, q\} \cup \mathcal{I}(x)$ . Note that if the constraint functions are linear, say,  $g_i(x) := a_i^\top x + b_i$ , then  $\nabla g_i(x) = a_i$  and the above KKT conditions hold without a constraint qualification. Consider the (polyhedral) cone

$$K(x) := \left\{ z \in \mathbb{R}^n : z = \sum_{i \in \mathcal{J}(x)} \alpha_i \nabla g_i(x), \ \alpha_i \leq 0, \ i \in \mathcal{I}(x) \right\}. \quad (5.187)$$

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  $\bar{x} \in X$  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

$$\gamma_N(\bar{x}) := \frac{1}{N} \sum_{j=1}^N \nabla_x F(\bar{x}, \xi^j) = \nabla \hat{f}_N(\bar{x}) \quad (5.188)$$

associated with the generated<sup>28</sup> random sample. Note that if, moreover, the derivatives can be taken inside the expectation, that is,

$$\nabla f(\bar{x}) = \mathbb{E}[\nabla_x F(\bar{x}, \xi)], \quad (5.189)$$

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  $\Sigma$ , 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  $\Sigma$  can be estimated by the corresponding sample covariance matrix

$$\hat{\Sigma}_N := \frac{1}{N-1} \sum_{j=1}^N \left[ \nabla_x F(\bar{x}, \xi^j) - \nabla \hat{f}_N(\bar{x}) \right] \left[ \nabla_x F(\bar{x}, \xi^j) - \nabla \hat{f}_N(\bar{x}) \right]^\top. \quad (5.190)$$

Under the above assumptions, the sample covariance matrix  $\hat{\Sigma}_N$  is an unbiased and consistent estimator of  $\Sigma$ .

We have that if  $V_N \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma)$  and the covariance matrix  $\Sigma$  is nonsingular, then (given a consistent estimator  $\hat{\Sigma}_N$  of  $\Sigma$ ) the following holds:

$$N(\gamma_N(\bar{x}) - \nabla f(\bar{x}))^\top \hat{\Sigma}_N^{-1} (\gamma_N(\bar{x}) - \nabla f(\bar{x})) \xrightarrow{\mathcal{D}} \chi_n^2, \quad (5.191)$$

where  $\chi_n^2$  denotes chi-square distribution with  $n$  degrees of freedom. This allows us to construct the following (approximate)  $100(1 - \alpha)\%$  confidence region<sup>29</sup> for  $\nabla f(\bar{x})$ :

$$\left\{ z \in \mathbb{R}^n : (\gamma_N(\bar{x}) - z)^\top \hat{\Sigma}_N^{-1} (\gamma_N(\bar{x}) - z) \leq \frac{\chi_{\alpha,n}^2}{N} \right\}. \quad (5.192)$$

Consider the statistic

$$T_N := N \inf_{z \in K(\bar{x})} (\gamma_N(\bar{x}) - z)^\top \hat{\Sigma}_N^{-1} (\gamma_N(\bar{x}) - z). \quad (5.193)$$

<sup>28</sup>We emphasize that the random sample in (5.188) is generated independently of the sample used to compute the candidate point  $\bar{x}$ .

<sup>29</sup>Here  $\chi_{\alpha,n}^2$  denotes the  $\alpha$ -critical value of chi-square distribution with  $n$  degrees of freedom. That is, if  $Y \sim \chi_n^2$ , then  $\Pr\{Y \geq \chi_{\alpha,n}^2\} = \alpha$ .

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  $T_N > \chi_{\alpha,n}^2$ . We can also use the statistic  $T_N$  for testing the hypothesis:

$$H_0 : \nabla f(\bar{x}) \in K(\bar{x}) \text{ against the alternative } H_1 : \nabla f(\bar{x}) \notin K(\bar{x}). \quad (5.194)$$

The  $T_N$  statistic represents the squared distance, with respect to the norm<sup>30</sup>  $\|\cdot\|_{\hat{\Sigma}_N^{-1}}$ , 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  $\nabla g_i(\bar{x})$ ,  $i = 1, \dots, q$ , 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 parameter<sup>31</sup>

$$\delta := N \inf_{z \in K(\bar{x})} (\nabla f(\bar{x}) - z)^\top \Sigma^{-1} (\nabla f(\bar{x}) - z). \quad (5.195)$$

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 to  $\Pr\{Y \geq T_N^*\}$ , 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 the  $T_N$  statistic can be large, i.e., the  $p$ -value can be small, simply because the estimated covariance matrix  $N^{-1}\hat{\Sigma}_N$  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

<sup>30</sup>For a positive definite matrix  $A$ , the norm  $\|\cdot\|_A$  is defined as  $\|z\|_A := (z^\top A z)^{1/2}$ .

<sup>31</sup>Note that under the alternative (i.e., if  $\nabla f(\bar{x}) \notin K(\bar{x})$ ), the noncentrality parameter  $\delta$  tends to infinity as  $N \rightarrow \infty$ . Therefore, in order to justify the above asymptotics, one needs a technical assumption known as Pitman’s parameter drift.

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  $\bar{f}'(\bar{x}, d) \geq c\|d\|$  for some  $c > 0$  and all  $d \in \mathbb{R}^n$ . Then, by the LD principle (see (7.192) 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

$$\text{Min}_{x \in X} f(x) \quad \text{s.t.} \quad p(x) \leq \alpha, \quad (5.196)$$

where  $X \subset \mathbb{R}^n$  is a closed set,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is a continuous function,  $\alpha \in (0, 1)$  is a given significance level, and

$$p(x) := \Pr\{C(x, \xi) > 0\} \quad (5.197)$$

is the probability that constraint is violated at point  $x \in X$ . We assume that  $\xi$  is a random vector, whose probability distribution  $P$  is supported on set  $\Xi \subset \mathbb{R}^d$ , and the function  $C : \mathbb{R}^n \times \Xi \rightarrow \mathbb{R}$  is a Carathéodory function. The chance constraint  $p(x) \leq \alpha$  can be written equivalently in the form

$$\Pr\{C(x, \xi) \leq 0\} \geq 1 - \alpha. \quad (5.198)$$

Let us also remark that several chance constraints

$$\Pr\{C_i(x, \xi) \leq 0, \quad i = 1, \dots, q\} \geq 1 - \alpha \quad (5.199)$$

can be reduced to one chance constraint (5.198) by employing the max-function  $C(x, \xi) := \max_{1 \leq i \leq q} C_i(x, \xi)$ . 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,

$$p(x) = \mathbb{E}[\mathbf{1}_{(0, \infty)}(C(x, \xi))],$$

and estimate this probability by the corresponding SAA function (compare with (5.14)–(5.16))

$$\hat{p}_N(x) := N^{-1} \sum_{j=1}^N \mathbf{1}_{(0, \infty)}(C(x, \xi^j)). \quad (5.200)$$

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  $C(x, \xi^j) > 0$ ,  $j = 1, \dots, N$ . Consequently we can write the corresponding SAA problem as

$$\text{Min}_{x \in X} f(x) \text{ s.t. } \hat{p}_N(x) \leq \alpha. \quad (5.201)$$

**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  $x \in X$  it holds that*

$$\Pr\{\xi \in \Xi : C(x, \xi) = 0\} = 0, \quad (5.202)$$

*i.e.,  $C(x, \xi) \neq 0$  w.p. 1. Then the function  $p(x)$  is continuous on  $X$  and  $\hat{p}_N(x)$  converges to  $p(x)$  w.p. 1 uniformly on any compact subset of  $X$ .*

**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 7.42). If the sample is iid, the epiconvergence  $\hat{p}_N \xrightarrow{e} p$  w.p. 1 follows by Theorem 7.51. 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  $x \in X$ ,  $\psi(\cdot, \xi)$  is continuous at  $x$  w.p. 1. It follows by the Lebesgue dominated convergence theorem that  $p(\cdot)$  is continuous at  $x$  (see Theorem 7.43). Finally, the uniform convergence w.p. 1 follows by Theorem 7.48.  $\square$

Since the function  $p(x)$  is lower semicontinuous and the set  $X$  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  $S$  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  $X$  is compact, the function  $f(x)$  is continuous,  $C(x, \xi)$  is a Carathéodory function, the sample is iid, and the following condition holds: (a) there is an optimal solution  $\bar{x}$  of the true problem such that for any  $\epsilon > 0$  there is  $x \in X$  with  $\|x - \bar{x}\| \leq \epsilon$  and  $p(x) < \alpha$ . Then  $\hat{\vartheta}_N \rightarrow \vartheta^*$  and  $\mathbb{D}(\hat{S}_N, S) \rightarrow 0$  w.p. 1 as  $N \rightarrow \infty$ .*

**Proof.** By condition (a), the set  $S$  is nonempty and there is  $x' \in X$  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  $x' \in X$  be a point sufficiently close to  $\bar{x} \in S$  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,

$$\limsup_{N \rightarrow \infty} \hat{\vartheta}_N \leq f(x') \leq f(\bar{x}) + \varepsilon = \vartheta^* + \varepsilon \text{ w.p. 1.} \quad (5.203)$$

Since  $\varepsilon > 0$  is arbitrary, it follows that

$$\limsup_{N \rightarrow \infty} \hat{\vartheta}_N \leq \vartheta^* \text{ w.p. 1.} \quad (5.204)$$

Now let  $\hat{x}_N \in \hat{S}_N$ , i.e.,  $\hat{x}_N \in X$ ,  $\hat{p}_N(\hat{x}_N) \leq \alpha$  and  $\hat{\vartheta}_N = f(\hat{x}_N)$ . Since the set  $X$  is compact, we can assume by passing to a subsequence if necessary that  $\hat{x}_N$  converges to a point  $\bar{x} \in X$  w.p. 1. Also by Proposition 5.29 we have that  $\hat{p}_N \xrightarrow{c} p$  w.p. 1, and hence

$$\liminf_{N \rightarrow \infty} \hat{p}_N(\hat{x}_N) \geq p(\bar{x}) \text{ w.p. 1.}$$

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  $f(\hat{x}_N) \rightarrow f(\bar{x})$  w.p. 1, and hence

$$\liminf_{N \rightarrow \infty} \hat{\vartheta}_N \geq \vartheta^* \text{ w.p. 1.} \quad (5.205)$$

It follows from (5.204) and (5.205) that  $\hat{\vartheta}_N \rightarrow \vartheta^*$  w.p. 1. It also follows that the point  $\bar{x}$  is an optimal solution of the true problem and consequently we obtain that  $\mathbb{D}(\hat{S}_N, S) \rightarrow 0$  w.p. 1.  $\square$

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  $N \rightarrow \infty$ , 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  $\alpha^{-1}$ . For small  $\alpha$  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  $\xi^1, \dots, \xi^N$  consider the problem

$$\text{Min}_{x \in X} f(x) \quad \text{s.t. } C(x, \xi^j) \leq 0, \quad j = 1, \dots, N. \quad (5.206)$$



Note that for  $\alpha = 0$  the SAA problem (5.201) coincides with problem (5.206). If the set  $X$  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  $N \rightarrow \infty$  the feasible set of 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  $X$  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  $J \subset \{1, \dots, N\}$  such that  $|J| \leq n$  and  $\bar{x}_N$  is an optimal solution of the problem*

$$\min_{x \in X} f(x) \quad \text{s.t.} \quad C(x, \xi^j) \leq 0, \quad j \in J. \quad (5.207)$$

**Proof.** Consider sets  $A_0 := \{x \in X : f(x) < f(\bar{x}_N)\}$  and  $A_j := \{x \in X : C(x, \xi^j) \leq 0\}$ ,  $j = 1, \dots, N$ . Since  $X$ ,  $f(\cdot)$  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  $A_j$ ,  $j \in \{1, \dots, N\}$ , 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  $A_j$ ,  $j \in \{0, 1, \dots, N\}$ , is nonempty. By Helly's theorem (Theorem 7.3) this implies that the intersection of all sets  $A_j$ ,  $j \in \{0, 1, \dots, N\}$ , 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  $\bar{x}_N$ .  $\square$

We will use the following assumptions.

**(F1)** For any  $N \in \mathbb{N}$  and any  $(\xi_1, \dots, \xi_N) \in \Xi^N$ , problem (5.206) attains the unique optimal solution  $\bar{x}_N = \bar{x}(\xi_1, \dots, \xi_N)$ .

Recall that sometimes we use the same notation for a random vector and its particular value (realization). In the above assumption we view  $\xi_1, \dots, \xi_N$  as an element of the set  $\Xi^N$  and  $\bar{x}_N$  as a function of  $\xi_1, \dots, \xi_N$ . Of course, if  $\xi_1, \dots, \xi_N$  is a random sample, then  $\bar{x}_N$  becomes a random vector.

Let  $\mathcal{J} = \mathcal{J}(\xi^1, \dots, \xi^N) \subset \{1, \dots, N\}$  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  $\Xi^N$ , i.e.,  $P^N$  is the probability distribution of the iid sample  $\xi^1, \dots, \xi^N$ .

**(F2)** There is an integer  $n \in \mathbb{N}$  such that, for any  $N \geq n$ , w.p. 1 the minimal set  $\mathcal{J} = \mathcal{J}(\xi^1, \dots, \xi^N)$  is uniquely defined and has constant cardinality  $n$ , i.e.,  $P^N\{|\mathcal{J}| = n\} = 1$ .

By Lemma 5.31 we have that  $n \leq n$ .

Assumption (F1) holds, for example, if the set  $X$  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 to show an equality in the estimate (5.209) of the following theorem.

The following result is due to Campi and Garatti [30], building on work of Calafiore and Campi [29]. Denote

$$b(k; \alpha, N) := \sum_{i=0}^k \binom{N}{i} \alpha^i (1 - \alpha)^{N-i}, \quad k = 0, \dots, N. \quad (5.208)$$

That is,  $b(k; \alpha, N) = \Pr(W \leq k)$ , where  $W \sim B(\alpha, N)$  is a random variable having binomial distribution.

**Theorem 5.32.** *Suppose that the set  $X$  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  $\xi^1, \dots, \xi^N$  of size  $N \geq n$  we have that*

$$\Pr\{p(\bar{x}_N) > \alpha\} = b(n - 1; \alpha, N). \quad (5.209)$$

**Proof.** Let  $\mathfrak{J}_n$  be the family of all sets  $J \subset \{1, \dots, N\}$  of cardinality  $n$ . We have that  $|\mathfrak{J}_n| = \binom{N}{n}$ . For  $J \in \mathfrak{J}_n$  define the set

$$\Sigma_J := \{(\xi^1, \dots, \xi^N) \in \Xi^N : \mathcal{J}(\xi^1, \dots, \xi^N) = J\} \quad (5.210)$$

and denote by  $\hat{x}_J = \hat{x}_J(\xi^1, \dots, \xi^N)$  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

$$\Sigma_J = \{(\xi^1, \dots, \xi^N) \in \Xi^N : \hat{x}_J = \bar{x}_N\}. \quad (5.211)$$

Note that for any permutation of vectors  $\xi^1, \dots, \xi^N$ , 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  $\Sigma_J$  are disjoint and, because of condition (F2), union of all these sets is equal to  $\Xi^N$  up to a set of  $P^N$ -measure zero. Since there are  $\binom{N}{n}$  such sets, we obtain that

$$P^N(\Sigma_J) = \frac{1}{\binom{N}{n}}. \quad (5.212)$$

Consider the optimal solution  $\bar{x}_n = \bar{x}(\xi^1, \dots, \xi^n)$  for  $N = n$ , and let  $H(z)$  be the cdf of the random variable  $p(\bar{x}_n)$ , i.e.,

$$H(z) := P^n\{p(\bar{x}_n) \leq z\}. \quad (5.213)$$

Let us show that for  $N \geq n$ ,

$$P^N(\Sigma_J) = \int_0^1 (1 - z)^{N-n} dH(z). \quad (5.214)$$

Indeed, for  $z \in [0, 1]$  and  $J \in \mathfrak{J}_n$  consider the sets

$$\begin{aligned}\Delta_z &:= \{(\xi_1, \dots, \xi_N) : p(\bar{x}_N) \in [z, z + dz]\}, \\ \Delta_{J,z} &:= \{(\xi_1, \dots, \xi_N) : p(\hat{x}_J) \in [z, z + dz]\}.\end{aligned}\quad (5.215)$$

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  $J = \{1, \dots, n\}$ . Note that  $\hat{x}_J$  depends on  $(\xi^1, \dots, \xi^n)$  only. Therefore  $\Delta_{J,z} = \Delta_z^* \times \Xi^{N-n}$ , where  $\Delta_z^*$  is a subset of  $\Xi^n$  corresponding to the event  $p(\hat{x}_J) \in [z, z + dz]$ . Conditional on  $(\xi^1, \dots, \xi^n)$ , the event  $\Delta_{J,z} \cap \Sigma_J$  happens iff the point  $\hat{x}_J = \hat{x}_J(\xi^1, \dots, \xi^n)$  remains feasible for the remaining  $N - n$  constraints, i.e., iff  $C(\hat{x}_J, \xi^j) \leq 0$  for all  $j = n+1, \dots, N$ . 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  $(\xi^1, \dots, \xi^n) \in \Delta_z^*$ , probability of the event  $\Delta_{J,z} \cap \Sigma_J$  is equal to  $(1 - z)^{N-n}$ . Consequently, the unconditional probability

$$P^N(\Delta_z \cap \Sigma_J) = P^N(\Delta_{J,z} \cap \Sigma_J) = (1 - z)^{N-n} dH(z), \quad (5.216)$$

and hence (5.214) follows.

It follows from (5.212) and (5.214) that

$$\binom{N}{n} \int_0^1 (1 - z)^{N-n} dH(z) = 1, \quad N \geq n. \quad (5.217)$$

Let us observe that  $H(z) := z^n$  satisfies (5.217) for all  $N \geq n$ . Indeed, using integration by parts, we have

$$\begin{aligned}\binom{N}{n} \int_0^1 (1 - z)^{N-n} dz^n &= -\binom{N}{n} \frac{n}{N-n+1} \int_0^1 z^{n-1} d(1 - z)^{N-n+1} \\ &= \binom{N}{n-1} \int_0^1 (1 - z)^{N-n+1} dz^{n-1} = \dots = 1.\end{aligned}\quad (5.218)$$

We also have that (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  $H(z) = z^n$ ,  $0 \leq z \leq 1$ , is the cdf of  $p(\bar{x}_n)$ .

We also have by (5.216) that

$$P^N\{p(\bar{x}_N) \in [z, z + dz]\} = \sum_{J \in \mathfrak{J}_n} P^N(\Delta_z \cap \Sigma_J) = \binom{N}{n} (1 - z)^{N-n} dH(z). \quad (5.219)$$

Therefore, since  $H(z) = z^n$  and using integration by parts similar to (5.218), we can write

$$\begin{aligned}P^N\{p(\bar{x}_N) > \alpha\} &= \binom{N}{n} \int_\alpha^1 (1 - z)^{N-n} dH(z) = \binom{N}{n} n \int_\alpha^1 (1 - z)^{N-n} z^{n-1} dz \\ &= \binom{N}{n} \frac{n}{N-n+1} \left[ -(1 - z)^{N-n+1} z^{n-1} \Big|_\alpha^1 + \int_\alpha^1 (1 - z)^{N-n+1} dz^{n-1} \right] \\ &= \binom{N}{n-1} (1 - \alpha)^{N-n+1} \alpha^{n-1} + \binom{N}{n-1} \int_\alpha^1 (1 - z)^{N-n+1} dz^{n-1} \\ &= \dots = \sum_{i=0}^{n-1} \binom{N}{i} \alpha^i (1 - \alpha)^{N-i}.\end{aligned}\quad (5.220)$$

Since  $\Pr\{p(\bar{x}_N) > \alpha\} = P^N\{p(\bar{x}_N) > \alpha\}$ , this completes the proof.  $\square$

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  $n \leq n$ . Therefore, given  $\beta \in (0, 1)$ , the inequality (5.209) implies that for sample size  $N \geq n$  such that

$$b(n-1; \alpha, N) \leq \beta, \quad (5.221)$$

we have with probability at least  $1 - \beta$  that  $\bar{x}_N$  is a feasible solution of the true problem (5.196).

Recall that

$$b(n-1; \alpha, N) = \Pr(W \leq n-1),$$

where  $W \sim B(\alpha, N)$  is a random variable having binomial distribution. For “not too small”  $\alpha$  and large  $N$ , good approximation of that probability is suggested by the CLT. That is,  $W$  has approximately normal distribution with mean  $N\alpha$  and variance  $N\alpha(1-\alpha)$ , and hence<sup>32</sup>

$$b(n-1; \alpha, N) \approx \Phi\left(\frac{n-1-N\alpha}{\sqrt{N\alpha(1-\alpha)}}\right). \quad (5.222)$$

For  $N\alpha \geq n-1$ , the Hoeffding inequality (7.188) gives the estimate

$$b(n-1; \alpha, N) \leq \exp\left\{-\frac{2(N\alpha - n + 1)^2}{N}\right\}, \quad (5.223)$$

and the Chernoff inequality (7.190) gives

$$b(n-1; \alpha, N) \leq \exp\left\{-\frac{(N\alpha - n + 1)^2}{2\alpha N}\right\}. \quad (5.224)$$

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 a significantly bigger estimate of  $N = 40372$ . The estimate derived from the Chernoff inequality (5.224) gives a much better estimate of  $N = 13410$ .

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 guarantees only its feasibility.

### 5.7.2 Validation of an Optimal Solution

We discuss now an approach to a practical validation of a candidate point  $\bar{x} \in X$  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.

<sup>32</sup>Recall that  $\Phi(\cdot)$  is the cdf of standard normal distribution.

### Upper Bounds

Let us start with verification of the 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  $\xi^1, \dots, \xi^N$ , let  $m$  be the number of times that the constraints  $C(\bar{x}, \xi^j) \leq 0$ ,  $j = 1, \dots, N$ , are violated, i.e.,

$$m := \sum_{j=1}^N \mathbf{1}_{(0, \infty)}(C(\bar{x}, \xi^j)).$$

Then  $\hat{p}_N(\bar{x}) = m/N$  is an unbiased estimator of  $p(\bar{x})$ , and  $m$  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 probability<sup>33</sup>  $p(\bar{x})$ :

$$\hat{p}_N(\bar{x}) + z_\beta \sqrt{\frac{\hat{p}_N(\bar{x})(1 - \hat{p}_N(\bar{x}))}{N}}. \quad (5.225)$$

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

$$\mathcal{U}_{\beta, N}(\bar{x}) := \sup_{\rho \in [0, 1]} \{\rho : b(m; \rho, N) \geq \beta\}. \quad (5.226)$$

We have that  $\mathcal{U}_{\beta, N}(\bar{x})$  is a function of  $m$  and hence is a random variable. Note that  $b(m; \rho, N)$  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  $\mathcal{U}_{\beta, N}(\bar{x})$  is equal to such  $\bar{\rho}$  that  $b(m; \bar{\rho}, N) = \beta$ . Denoting  $V := b(m; p(\bar{x}), N)$ , we have that

$$\begin{aligned} \Pr\{p(\bar{x}) < \mathcal{U}_{\beta, N}(\bar{x})\} &= \Pr\left\{V > \overbrace{b(m; \bar{\rho}, N)}^\beta\right\} \\ &= 1 - \Pr\{V \leq \beta\} = 1 - \sum_{k=0}^N \Pr\{V \leq \beta | m = k\} \Pr(m = k). \end{aligned}$$

Since

$$\Pr\{V \leq \beta | m = k\} = \begin{cases} 1 & \text{if } b(k; p(\bar{x}), N) \leq \beta, \\ 0 & \text{otherwise,} \end{cases}$$

and  $\Pr(m = k) = \binom{N}{k} p(\bar{x})^k (1 - p(\bar{x}))^{N-k}$ , it follows that

$$\sum_{k=0}^N \Pr\{V \leq \beta | m = k\} \Pr(m = k) \leq \beta,$$

and hence

$$\Pr\{p(\bar{x}) < \mathcal{U}_{\beta, N}(\bar{x})\} \geq 1 - \beta. \quad (5.227)$$

<sup>33</sup>Recall that  $z_\beta := \Phi^{-1}(1 - \beta) = -\Phi^{-1}(\beta)$ , where  $\Phi(\cdot)$  is the cdf of the standard normal distribution.

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  $m = 0$ , 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  $\vartheta^*$  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  $\alpha$  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  $b(k; \bar{\rho}, N)$  can be approximated by  $\Phi(\frac{k - \bar{\rho}N}{\sqrt{N\bar{\rho}(1-\bar{\rho})}})$ . Therefore, approximately  $\bar{\rho}$  is the solution of the equation  $\Phi(\frac{m - \rho N}{\sqrt{N\rho(1-\rho)}}) = \beta$ , which can be written as

$$\rho = \frac{m}{N} + z_\beta \sqrt{\frac{\rho(1-\rho)}{N}}.$$

By approximating  $\rho$  in the right-hand side of the above equation by  $m/N$  we obtain the bound (5.225).

### Lower Bounds

It is more tricky to construct a valid lower statistical bound for  $\vartheta^*$ . One possible approach is to apply a general methodology of the SAA method. (See the discussion at the end of section 5.6.1.) We have that for any  $\lambda \geq 0$  the following inequality holds (compare with (5.183)):

$$\vartheta^* \geq \inf_{x \in X} [f(x) + \lambda(p(x) - \alpha)]. \quad (5.228)$$

We also have that expectation of

$$\hat{v}_N(\lambda) := \inf_{x \in X} [f(x) + \lambda(\hat{p}_N(x) - \alpha)] \quad (5.229)$$

gives a valid lower bound for the right-hand side of (5.228), and hence for  $\vartheta^*$ . 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  $\lambda \geq 0$  the optimal value of the right-hand side of (5.228) is smaller than  $\vartheta^*$ , 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  $\xi^{1,m}, \dots, \xi^{N,m}$ ,  $m = 1, \dots, M$ , each of size  $N$ , of random vector  $\xi$ . For each sample, solve the associated optimization problem

$$\text{Min}_{x \in X} f(x) \quad \text{s.t.} \quad \sum_{j=1}^N \mathbf{1}_{(0,\infty)}(C(x, \xi^{j,m})) \leq \gamma N \quad (5.230)$$

and hence calculate its optimal value  $\hat{\vartheta}_{\gamma,N}^m$ ,  $m = 1, \dots, M$ . That is, we solve  $M$  times the corresponding SAA problem at the significance level  $\gamma$ . In particular, for  $\gamma = 0$ , problem (5.230) takes the form

$$\text{Min}_{x \in X} f(x) \quad \text{s.t.} \quad C(x, \xi^{j,m}) \leq 0, \quad j = 1, \dots, N. \quad (5.231)$$

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  $\hat{\vartheta}_{\gamma,N}^m$ ,  $m = 1, \dots, M$ , 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  $\gamma$ . Next we rearrange the calculated optimal values in the nondecreasing order,  $\hat{\vartheta}_{\gamma,N}^{(1)} \leq \dots \leq \hat{\vartheta}_{\gamma,N}^{(M)}$ ; i.e.,  $\hat{\vartheta}_{\gamma,N}^{(1)}$  is the smallest,  $\hat{\vartheta}_{\gamma,N}^{(2)}$  is the second smallest, etc., among the values  $\hat{\vartheta}_{\gamma,N}^m$ ,  $m = 1, \dots, M$ . By definition, we choose an integer  $L \in \{1, \dots, M\}$  and use the random quantity  $\hat{\vartheta}_{\gamma,N}^{(L)}$  as a lower bound of the true optimal value  $\vartheta^*$ .

Let us analyze the resulting bounding procedure. Let  $\tilde{x} \in X$  be a feasible point of the true problem, i.e.,

$$\Pr\{C(\tilde{x}, \xi) > 0\} \leq \alpha.$$

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 least<sup>34</sup>

$$\sum_{i=0}^{\lfloor \gamma N \rfloor} \binom{N}{i} \alpha^i (1-\alpha)^{N-i} = \mathbf{b}(\lfloor \gamma N \rfloor; \alpha, N) =: \theta_N.$$

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  $m \in \{1, \dots, M\}$  we have

$$\theta := \Pr\left\{\hat{\vartheta}_{\gamma,N}^m \leq \vartheta^* + \varepsilon\right\} \geq \Pr\left\{\hat{\vartheta}_{\gamma,N}^m \leq f(\tilde{x})\right\} \geq \theta_N.$$

Now, in the case of  $\hat{\vartheta}_{\gamma,N}^{(L)} > \vartheta^* + \varepsilon$ , the corresponding realization of the random sequence  $\hat{\vartheta}_{\gamma,N}^1, \dots, \hat{\vartheta}_{\gamma,N}^M$  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  $\mathbf{b}(L-1; \theta, M)$ . Since  $\theta \geq \theta_N$ , we have that  $\mathbf{b}(L-1; \theta, M) \leq \mathbf{b}(L-1; \theta_N, M)$ . Thus,  $\Pr\{\hat{\vartheta}_{\gamma,N}^{(L)} > \vartheta^* + \varepsilon\} \leq \mathbf{b}(L-1; \theta_N, M)$ . Since the resulting inequality is valid for any  $\varepsilon > 0$ , we arrive at the bound

$$\Pr\left\{\hat{\vartheta}_{\gamma,N}^{(L)} > \vartheta^*\right\} \leq \mathbf{b}(L-1; \theta_N, M). \quad (5.232)$$

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*

$$\mathbf{b}(L-1; \theta_N, M) \leq \beta, \quad (5.233)$$

<sup>34</sup>Recall that the notation  $\lfloor a \rfloor$  stands for the largest integer less than or equal to  $a \in \mathbb{R}$ .

where  $\theta_N := \mathfrak{b}(\lfloor \gamma N \rfloor; \alpha, N)$ . Then

$$\Pr \left\{ \hat{\vartheta}_{\gamma, N}^{(L)} > \vartheta^* \right\} \leq \beta. \quad (5.234)$$

For given sample sizes  $N$  and  $M$ , it is better to take the largest integer  $L \in \{1, \dots, M\}$  satisfying condition (5.233). That is, for

$$L^* := \max_{1 \leq L \leq M} \{L : \mathfrak{b}(L - 1; \theta_N, M) \leq \beta\},$$

we have that the random quantity  $\hat{\vartheta}_{\gamma, N}^{(L^*)}$  gives a lower bound for the true optimal value  $\vartheta^*$  with probability at least  $1 - \beta$ . If no  $L \in \{1, \dots, M\}$  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  $\gamma$ . In the convex case it is advantageous to take  $\gamma = 0$ , since then we 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

$$\sum_{k=0}^{L-1} \binom{M}{k} (1 - \alpha)^{Nk} [1 - (1 - \alpha)^N]^{M-k} \leq \beta. \quad (5.235)$$

Suppose that  $N$  and  $\gamma \geq 0$  are given (fixed). Then the larger  $M$  is, the better. We can view  $\hat{\vartheta}_{\gamma, N}^m$ ,  $m = 1, \dots, M$ , 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)  $\beta$ -quantile of the binomial distribution  $B(\theta_N, M)$ . By the CLT we have that

$$\lim_{M \rightarrow \infty} \frac{L^* - \theta_N M}{\sqrt{M \theta_N (1 - \theta_N)}} = \Phi^{-1}(\beta),$$

and  $L^*/M$  tends to  $\theta_N$  as  $M \rightarrow \infty$ . It follows that the lower bound  $\hat{\vartheta}_{\gamma, N}^{(L^*)}$  converges to the  $\theta_N$ -quantile of the distribution of  $\hat{\vartheta}_N$  as  $M \rightarrow \infty$ .

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 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  $\alpha N$  is large, 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  $\xi_1, \xi_2, \dots, \xi_T$ . 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  $\xi_2^1, \dots, \xi_2^{N_1}$  of  $N_1$  realizations of random vector  $\xi_2$ . Conditional on each  $\xi_2^i, i = 1, \dots, N_1$ , generate a random sample  $\xi_3^{ij}, j = 1, \dots, N_2$ , of  $N_2$  realizations of  $\xi_3$  according to conditional distribution of  $\xi_3$  given  $\xi_2 = \xi_2^i$ . Conditional on each  $\xi_3^{ij}$ , generate a random sample of size  $N_3$  of  $\xi_4$  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 otherwise,<sup>35</sup> we assume that, at the first stage, the sample  $\xi_2^1, \dots, \xi_2^{N_1}$  is iid and the following samples, at each stage  $t = 2, \dots, T - 1$ , 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  $\xi_1, \dots, \xi_T$ , 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  $\xi_t^1, \dots, \xi_t^{N_{t-1}}$  of respective  $\xi_t, t = 2, \dots, T$ , 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  $\xi_t^1, \dots, \xi_t^{N_{t-1}}$ . 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  $N_1, \dots, N_{T-1}$  tend to infinity. We denote  $\mathcal{N} := \{N_1, \dots, N_{T-1}\}$  and by  $\vartheta^*$  and  $\hat{\vartheta}_{\mathcal{N}}$  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  $\xi_2^i, i = 1, \dots, N_1$ , of  $\xi_2$  and for each  $i = 1, \dots, N_1$  of conditional samples  $\xi_3^{ij}, j = 1, \dots, N_2$ , of  $\xi_3$  given  $\xi_2 = \xi_2^i$ . Let us write dynamic programming equations for the true problem. We have

$$Q_3(x_2, \xi_3) = \inf_{x_3 \in \mathcal{X}_3(x_2, \xi_3)} f_3(x_3, \xi_3), \quad (5.236)$$

$$Q_2(x_1, \xi_2) = \inf_{x_2 \in \mathcal{X}_2(x_1, \xi_2)} \{f_2(x_2, \xi_2) + \mathbb{E}[Q_3(x_2, \xi_3) | \xi_2]\}, \quad (5.237)$$

<sup>35</sup>It 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 at the first stage we solve the problem

$$\text{Min}_{x_1 \in \mathcal{X}_1} \{f_1(x_1) + \mathbb{E}[Q_2(x_1, \xi_2)]\}. \quad (5.238)$$

If we could calculate values  $Q_2(x_1, \xi_2)$ , we could approximate problem (5.238) by the sample average problem

$$\text{Min}_{x_1 \in \mathcal{X}_1} \left\{ \hat{f}_{N_1}(x_1) := f_1(x_1) + \frac{1}{N_1} \sum_{i=1}^{N_1} Q_2(x_1, \xi_2^i) \right\}. \quad (5.239)$$

However, values  $Q_2(x_1, \xi_2^i)$  are not given explicitly and are approximated by

$$\hat{Q}_{2,N_2}(x_1, \xi_2^i) := \inf_{x_2 \in \mathcal{X}_2(x_1, \xi_2^i)} \left\{ f_2(x_2, \xi_2^i) + \frac{1}{N_2} \sum_{j=1}^{N_2} Q_3(x_2, \xi_3^{ij}) \right\}, \quad (5.240)$$

$i = 1, \dots, N_1$ . That is, the SAA method approximates the first stage problem (5.238) by the problem

$$\text{Min}_{x_1 \in \mathcal{X}_1} \left\{ \tilde{f}_{N_1, N_2}(x_1) := f_1(x_1) + \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{Q}_{2,N_2}(x_1, \xi_2^i) \right\}. \quad (5.241)$$

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 the analysis of section 5.1.1.) That is, we need to show that

$$\lim_{N_1, N_2 \rightarrow \infty} \sup_{x_1 \in X} \left| \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{Q}_{2,N_2}(x_1, \xi_2^i) - \mathbb{E}[Q_2(x_1, \xi_2)] \right| = 0 \text{ w.p. } 1. \quad (5.242)$$

For that it suffices to show that

$$\lim_{N_1 \rightarrow \infty} \sup_{x_1 \in X} \left| \frac{1}{N_1} \sum_{i=1}^{N_1} Q_2(x_1, \xi_2^i) - \mathbb{E}[Q_2(x_1, \xi_2)] \right| = 0 \text{ w.p. } 1 \quad (5.243)$$

and

$$\lim_{N_1, N_2 \rightarrow \infty} \sup_{x_1 \in X} \left| \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{Q}_{2,N_2}(x_1, \xi_2^i) - \frac{1}{N_1} \sum_{i=1}^{N_1} Q_2(x_1, \xi_2^i) \right| = 0 \text{ w.p. } 1. \quad (5.244)$$

Condition (5.243) can be verified by applying a version of the 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  $N_2 \rightarrow \infty$  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  $\xi_2$  and  $\xi_3$  are independent. In that case distribution of random sample  $\xi_3^{ij}$ ,  $j = 1, \dots, N_2$ , does not depend on  $i$  (in both sampling schemes whether samples  $\xi_3^{ij}$  are the same for all  $i = 1, \dots, N_1$ , or independent of each other), and we can apply Theorem 7.48 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  $N_2 \rightarrow \infty$  uniformly in  $x_2$  on any compact subset of  $\mathbb{R}^{m_2}$ . 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  $N_1, \dots, N_{T-1}$  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)

$$\vartheta^* \geq \mathbb{E}[\hat{\vartheta}_{\mathcal{N}}]. \quad (5.245)$$

That is, the SAA optimal value  $\hat{\vartheta}_{\mathcal{N}}$  is a downward biased estimator of the true optimal value  $\vartheta^*$ .

Suppose now that the data process  $\xi_1, \dots, \xi_T$  is stagewise independent. As 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  $\xi_2^i$ ,  $i = 1, \dots, N_1$ , of  $\xi_2$  and sample  $\xi_3^j$ ,  $j = 1, \dots, N_2$ , of  $\xi_3$  independent of  $\xi_2^i$ . In that case formula (5.240) takes the form

$$\hat{Q}_{2,N_2}(x_1, \xi_2^i) = \inf_{x_2 \in \mathcal{X}_2(x_1, \xi_2^i)} \left\{ f_2(x_2, \xi_2^i) + \frac{1}{N_2} \sum_{j=1}^{N_2} Q_3(x_2, \xi_3^j) \right\}. \quad (5.246)$$

Because of independence of  $\xi_2$  and  $\xi_3$  we have that conditional distribution of  $\xi_3$  given  $\xi_2$  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  $\hat{\vartheta}_{\mathcal{N}}$  has a similar bias. Variance of  $\hat{\vartheta}_{\mathcal{N}}$ , however, could be quite different. In the case of independent conditional sampling we have that  $\hat{Q}_{2,N_2}(x_1, \xi_2^i)$ ,  $i = 1, \dots, N_1$ , are independent of each other, and hence

$$\mathbb{V}\text{ar} \left[ \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{Q}_{2,N_2}(x_1, \xi_2^i) \right] = \frac{1}{N_1} \mathbb{V}\text{ar} \left[ \hat{Q}_{2,N_2}(x_1, \xi_2^i) \right]. \quad (5.247)$$

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  $i = 1, \dots, N_1$ . 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

$\hat{\vartheta}_{\mathcal{N}}$  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  $\xi_1^i, i = 1, \dots, N_0$ , of returns at stage  $t = 1$ , and conditional samples  $\xi_2^{ij}, j = 1, \dots, N_1$ , of returns at stage  $t = 2$ . The dynamic programming equations for the SAA problem can be written as follows (see (1.50)–(1.52)). At stage  $t = 1$  for  $i = 1, \dots, N_0$ , we have

$$\hat{Q}_{1,N_1}(W_1, \xi_1^i) = \sup_{x_1 \geq 0} \left\{ \frac{1}{N_1} \sum_{j=1}^{N_1} U((\xi_2^{ij})^\top x_1) : e^\top x_1 = W_1 \right\}, \quad (5.248)$$

where  $e \in \mathbb{R}^n$  is vector of ones, and at stage  $t = 0$  we solve the problem

$$\text{Max}_{x_0 \geq 0} \frac{1}{N_0} \sum_{i=1}^{N_0} \hat{Q}_{1,N_1}((\xi_1^i)^\top x_0, \xi_1^i) \quad \text{s.t.} \quad e^\top x_0 = W_0. \quad (5.249)$$

Now let  $U(W) := \ln W$  be the logarithmic utility function. Suppose that the data process is stagewise independent. Then the optimal value  $\vartheta^*$  of the true problem is (see (1.58))

$$\vartheta^* = \ln W_0 + \sum_{t=0}^{T-1} v_t, \quad (5.250)$$

where  $v_t$  is the optimal value of the problem

$$\text{Max}_{x_t \geq 0} \mathbb{E} [\ln ((\xi_{t+1}^\top) x_t)] \quad \text{s.t.} \quad e^\top x_t = 1. \quad (5.251)$$

Let the SAA method be applied with the identical conditional sampling, with respective sample  $\xi_t^j, j = 1, \dots, N_{t-1}$ , of  $\xi_t, t = 1, \dots, T$ . In that case, the corresponding SAA problem is also stagewise independent and the optimal value of the SAA problem

$$\hat{\vartheta}_{\mathcal{N}} = \ln W_0 + \sum_{t=0}^{T-1} \hat{v}_{t,N_t}, \quad (5.252)$$

where  $\hat{v}_{t,N_t}$  is the optimal value of the problem

$$\text{Max}_{x_t \geq 0} \frac{1}{N_t} \sum_{j=1}^{N_t} \ln ((\xi_{t+1}^j)^\top x_t) \quad \text{s.t.} \quad e^\top x_t = 1. \quad (5.253)$$

We can view  $\hat{v}_{t,N_t}$  as an SAA estimator of  $v_t$ . Since here we solve a maximization rather than a minimization problem,  $\hat{v}_{t,N_t}$  is an upward biased estimator of  $v_t$ , i.e.,  $\mathbb{E}[\hat{v}_{t,N_t}] \geq v_t$ . We also have that  $\mathbb{E}[\hat{\vartheta}_{\mathcal{N}}] = \ln W_0 + \sum_{t=0}^{T-1} \mathbb{E}[\hat{v}_{t,N_t}]$ , and hence

$$\mathbb{E}[\hat{\vartheta}_{\mathcal{N}}] - \vartheta^* = \sum_{t=0}^{T-1} (\mathbb{E}[\hat{v}_{t,N_t}] - v_t). \quad (5.254)$$

## 5.8. SAA Method Applied to Multistage Stochastic Programming

225

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

$$\mathbb{V}\text{ar}[\hat{\vartheta}_{\mathcal{N}}] = \sum_{t=0}^{T-1} \mathbb{V}\text{ar}[\hat{v}_{t,N_t}]. \quad (5.255)$$

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))

$$\vartheta^* = W_0^\gamma \prod_{t=0}^{T-1} \eta_t, \quad (5.256)$$

where  $\eta_t$  is the optimal value of problem

$$\text{Max}_{x_t \geq 0} \mathbb{E}[(\xi_{t+1}^\top x_t)^\gamma] \quad \text{s.t.} \quad e^\top x_t = 1. \quad (5.257)$$

For the corresponding SAA method with the identical conditional sampling, we have that

$$\hat{\vartheta}_{\mathcal{N}} = W_0^\gamma \prod_{t=0}^{T-1} \hat{\eta}_{t,N_t}, \quad (5.258)$$

where  $\hat{\eta}_{t,N_t}$  is the optimal value of problem

$$\text{Max}_{x_t \geq 0} \frac{1}{N_t} \sum_{j=1}^{N_t} ((\xi_{t+1}^j)^\top x_t)^\gamma \quad \text{s.t.} \quad e^\top x_t = 1. \quad (5.259)$$

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

$$\mathbb{E}[\hat{\vartheta}_{\mathcal{N}}] = \vartheta^* \prod_{t=0}^{T-1} (1 + \beta_{t,N_t}), \quad (5.260)$$

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  $\hat{\vartheta}_{\mathcal{N}}$  grows with increase of the number of stages in a *multiplicative* way. In particular, if the relative biases  $\beta_{t,N_t}$  are constant, then bias of  $\hat{\vartheta}_{\mathcal{N}}$  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  $\vartheta^*$ . Therefore, in order to construct a lower bound for  $\vartheta^*$  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  $\mathbf{x}_t(\xi_{[t]})$  be a feasible policy. Recall that a policy is feasible if it satisfies the feasibility constraints (3.2). Since the multistage problem can be formulated as the minimization problem (3.3) we have that

$$\mathbb{E}[f_1(x_1) + f_2(\mathbf{x}_2(\xi_{[2]}), \xi_2) + \cdots + f_T(\mathbf{x}_T(\xi_{[T]}), \xi_T)] \geq \vartheta^*, \quad (5.261)$$

and equality in (5.261) holds iff the policy  $\mathbf{x}_t(\xi_{[t]})$  is optimal. The expectation in the left-hand side of (5.261) can be estimated in a straightforward way. That is, generate random sample  $\xi_1^j, \dots, \xi_T^j$ ,  $j = 1, \dots, N$ , of  $N$  realizations (scenarios) of the random data process  $\xi_1, \dots, \xi_T$  and estimate this expectation by the average

$$\frac{1}{N} \sum_{j=1}^N [f_1(x_1) + f_2(\mathbf{x}_2(\xi_{[2]}^j), \xi_2^j) + \cdots + f_T(\mathbf{x}_T(\xi_{[T]}^j), \xi_T^j)]. \quad (5.262)$$

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  $\vartheta^*$ . Of course, the quality of this upper bound depends on a successful choice of the feasible policy, i.e., on how small the optimality gap is 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

$$\begin{aligned} \text{Min}_{\mathbf{x}_2, \dots, \mathbf{x}_T} \quad & f_1(\bar{x}_1) + \mathbb{E}[f_2(\mathbf{x}_2(\xi_{[2]}), \xi_2) + \cdots + f_T(\mathbf{x}_T(\xi_{[T]}), \xi_T)] \\ \text{s.t.} \quad & \mathbf{x}_t(\xi_{[t]}) \in \mathcal{X}_t(\mathbf{x}_{t-1}(\xi_{[t-1]}), \xi_t), \quad t = 2, \dots, T. \end{aligned} \quad (5.263)$$

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 to compute value of two-stage stochastic program  $\min_{x \in X} \mathbb{E}[F(x, \xi)]$ , where  $F(x, \xi)$  is the optimal value of the corresponding second-stage problem, at a feasible point  $\bar{x} \in X$  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  $\xi$ . Monte Carlo techniques allow us 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 detail 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  $\xi$  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  $\xi_2$ . For that we need to solve the two-stage stochastic programming problem given in the right-hand side of (5.237). As 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  $N = N_1 N_2$ . 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  $\varepsilon$  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 now discuss in detail the 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  $\xi_2$  and  $\xi_3$  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  $\xi_2$  and  $\xi_3$  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

$$|Q_2(x'_1, \xi_2) - Q_2(x_1, \xi_2)| \leq L_1 \|x'_1 - x_1\| \quad (5.264)$$

for all  $x'_1, x_1 \in \mathcal{X}_1$  and a.e.  $\xi_2$ .

**(M'5)** There exists a constant  $\sigma_1 > 0$  such that for any  $x_1 \in \mathcal{X}_1$  it holds that

$$M_{1, x_1}(t) \leq \exp \left\{ \sigma_1^2 t^2 / 2 \right\}, \quad \forall t \in \mathbb{R}, \quad (5.265)$$

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  $\mathcal{C}$  of finite diameter  $D_2$  such that for every  $x_1 \in \mathcal{X}_1$  and a.e.  $\xi_2$ , the set  $\mathcal{X}_2(x_1, \xi_2)$  is contained in  $\mathcal{C}$ .

(M'7) There is a constant  $L_2 > 0$  such that

$$|Q_3(x'_2, \xi_3) - Q_3(x_2, \xi_3)| \leq L_2 \|x'_2 - x_2\| \quad (5.266)$$

for all  $x'_2, x_2 \in \mathcal{C}$  and a.e.  $\xi_3$ .

(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.  $\xi_2$  it holds that

$$M_{2,x_2}(t) \leq \exp \left\{ \sigma_2^2 t^2 / 2 \right\}, \quad \forall t \in \mathbb{R}, \quad (5.267)$$

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*

$$\left[ \frac{O(1)D_1L_1}{\varepsilon} \right]^{n_1} \exp \left\{ -\frac{O(1)N_1\varepsilon^2}{\sigma_1^2} \right\} + \left[ \frac{O(1)D_2L_2}{\varepsilon} \right]^{n_2} \exp \left\{ -\frac{O(1)N_2\varepsilon^2}{\sigma_2^2} \right\} \leq \alpha, \quad (5.268)$$

*we have that any  $\varepsilon/2$ -optimal solution of the SAA problem (5.241) is an  $\varepsilon$ -optimal solution of the first stage (5.238) of the true problem with probability at least  $1 - \alpha$ .*

**Proof.** The proof of this theorem is based on the uniform exponential bound of Theorem 7.67. Let us sketch the arguments. Assume that the conditional sampling is identical. We have that for every  $x_1 \in \mathcal{X}_1$  and  $i = 1, \dots, N_1$ ,

$$\left| \hat{Q}_{2,N_2}(x_1, \xi_2^i) - Q_2(x_1, \xi_2^i) \right| \leq \sup_{x_2 \in \mathcal{C}} \left| \frac{1}{N_2} \sum_{j=1}^{N_2} Q_3(x_2, \xi_3^j) - \mathbb{E}[Q_3(x_2, \xi_3)] \right|,$$

where  $\mathcal{C}$  is the set postulated in assumption (M'6). Consequently,

$$\begin{aligned} & \sup_{x_1 \in \mathcal{X}_1} \left| \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{Q}_{2,N_2}(x_1, \xi_2^i) - \frac{1}{N_1} \sum_{i=1}^{N_1} Q_2(x_1, \xi_2^i) \right| \\ & \leq \frac{1}{N_1} \sum_{i=1}^{N_1} \sup_{x_1 \in \mathcal{X}_1} \left| \hat{Q}_{2,N_2}(x_1, \xi_2^i) - Q_2(x_1, \xi_2^i) \right| \\ & \leq \sup_{x_2 \in \mathcal{C}} \left| \frac{1}{N_2} \sum_{j=1}^{N_2} Q_3(x_2, \xi_3^j) - \mathbb{E}[Q_3(x_2, \xi_3)] \right|. \end{aligned} \quad (5.269)$$

By the uniform exponential bound (7.217) we have that

$$\begin{aligned} & \Pr \left\{ \sup_{x_2 \in \mathcal{C}} \left| \frac{1}{N_2} \sum_{j=1}^{N_2} Q_3(x_2, \xi_3^j) - \mathbb{E}[Q_3(x_2, \xi_3)] \right| > \varepsilon/2 \right\} \\ & \leq \left[ \frac{O(1)D_2L_2}{\varepsilon} \right]^{n_2} \exp \left\{ -\frac{O(1)N_2\varepsilon^2}{\sigma_2^2} \right\}, \end{aligned} \quad (5.270)$$

and hence

$$\begin{aligned} & \Pr \left\{ \sup_{x_1 \in \mathcal{X}_1} \left| \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{Q}_{2,N_2}(x_1, \xi_2^i) - \frac{1}{N_1} \sum_{i=1}^{N_1} Q_2(x_1, \xi_2^i) \right| > \varepsilon/2 \right\} \\ & \leq \left[ \frac{O(1)D_2L_2}{\varepsilon} \right]^{n_2} \exp \left\{ -\frac{O(1)N_2\varepsilon^2}{\sigma_2^2} \right\}. \end{aligned} \quad (5.271)$$



## 5.8. SAA Method Applied to Multistage Stochastic Programming

229

By the uniform exponential bound (7.217) we also have that

$$\Pr \left\{ \sup_{x_1 \in \mathcal{X}_1} \left| \frac{1}{N_1} \sum_{i=1}^{N_1} Q_2(x_1, \xi_2^i) - \mathbb{E}[Q_2(x_1, \xi_2)] \right| > \varepsilon/2 \right\} \leq \left[ \frac{O(1)D_1L_1}{\varepsilon} \right]^{n_1} \exp \left\{ -\frac{O(1)N_1\varepsilon^2}{\sigma_1^2} \right\}. \quad (5.272)$$

Let us observe that if  $Z_1, Z_2$  are random variables, then

$$\Pr(Z_1 + Z_2 > \varepsilon) \leq \Pr(Z_1 > \varepsilon/2) + \Pr(Z_2 > \varepsilon/2).$$

Therefore it follows from (5.271) and (5.271) that

$$\Pr \left\{ \sup_{x_1 \in \mathcal{X}_1} |\tilde{f}_{N_1, N_2}(x_1) - f_1(x_1) - \mathbb{E}[Q_2(x_1, \xi_2)]| > \varepsilon \right\} \leq \left[ \frac{O(1)D_1L_1}{\varepsilon} \right]^{n_1} \exp \left\{ -\frac{O(1)N_1\varepsilon^2}{\sigma_1^2} \right\} + \left[ \frac{O(1)D_2L_2}{\varepsilon} \right]^{n_2} \exp \left\{ -\frac{O(1)N_2\varepsilon^2}{\sigma_2^2} \right\}, \quad (5.273)$$

which implies the assertion of the theorem.

In the case of the independent conditional sampling the proof can be completed in a similar way.  $\square$

**Remark 17.** We have, of course, that

$$|\hat{\vartheta}_{\mathcal{N}} - \vartheta^*| \leq \sup_{x_1 \in \mathcal{X}_1} |\tilde{f}_{N_1, N_2}(x_1) - f_1(x_1) - \mathbb{E}[Q_2(x_1, \xi_2)]|. \quad (5.274)$$

Therefore bound (5.273) also implies that

$$\Pr\{|\hat{\vartheta}_{\mathcal{N}} - \vartheta^*| > \varepsilon\} \leq \left[ \frac{O(1)D_1L_1}{\varepsilon} \right]^{n_1} \exp \left\{ -\frac{O(1)N_1\varepsilon^2}{\sigma_1^2} \right\} + \left[ \frac{O(1)D_2L_2}{\varepsilon} \right]^{n_2} \exp \left\{ -\frac{O(1)N_2\varepsilon^2}{\sigma_2^2} \right\}. \quad (5.275)$$

In particular, suppose that  $N_1 = N_2$ . Then for

$$n := \max\{n_1, n_2\}, \quad L := \max\{L_1, L_2\}, \quad D := \max\{D_1, D_2\}, \quad \sigma := \max\{\sigma_1, \sigma_2\},$$

the estimate (5.268) implies the following estimate of the required sample size  $N_1 = N_2$ :

$$\left( \frac{O(1)DL}{\varepsilon} \right)^n \exp \left\{ -\frac{O(1)N_1\varepsilon^2}{\sigma^2} \right\} \leq \alpha, \quad (5.276)$$

which is equivalent to

$$N_1 \geq \frac{O(1)\sigma^2}{\varepsilon^2} \left[ n \ln \left( \frac{O(1)DL}{\varepsilon} \right) + \ln \left( \frac{1}{\alpha} \right) \right]. \quad (5.277)$$

The estimate (5.277), for three-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 three-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 indicates only that 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 four) number of stages.

## 5.9 Stochastic Approximation Method

To an extent, this section is based on Nemirovski et al. [133]. Consider the 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  $x \in X$  and that the set  $X \subset \mathbb{R}^n$  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  $X$  is compact and  $f(x)$  is continuous. Clearly,  $\vartheta^* = f(\bar{x})$ . (Recall that  $\vartheta^*$  denotes the optimal value of problem (5.1).) We also assume throughout this section that the set  $X$  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  $x \in X$  and  $\xi \in \Xi$  returns value  $F(x, \xi)$  and a stochastic subgradient, a vector  $G(x, \xi)$  such that  $g(x) := \mathbb{E}[G(x, \xi)]$  is well defined and is a subgradient of  $f(\cdot)$  at  $x$ , i.e.,  $g(x) \in \partial f(x)$ .

**Remark 18.** Recall that if  $F(\cdot, \xi)$  is convex for every  $\xi \in \Xi$ , and  $x$  is an interior point of  $X$ , i.e.,  $f(\cdot)$  is finite valued in a neighborhood of  $x$ , then

$$\partial f(x) = \mathbb{E}[\partial_x F(x, \xi)] \quad (5.278)$$

(see Theorem 7.47). 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  $\xi^j$ ,  $j = 1, \dots$ , of realizations of the random vector  $\xi$ , 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  $\|x\|_2 = (x^\top x)^{1/2}$  the Euclidean norm of vector  $x \in \mathbb{R}^n$  and by

$$\Pi_X(x) := \arg \min_{z \in X} \|x - z\|_2 \quad (5.279)$$

## 5.9. Stochastic Approximation Method

231

the metric projection of  $x$  onto the set  $X$ . Since  $X$  is convex and closed, the minimizer in the right-hand side of (5.279) exists and is unique. Note that  $\Pi_X$  is a nonexpanding operator, i.e.,

$$\|\Pi_X(x') - \Pi_X(x)\|_2 \leq \|x' - x\|_2, \quad \forall x', x \in \mathbb{R}^n. \quad (5.280)$$

The classical stochastic approximation (SA) algorithm solves problem (5.1) by mimicking a simple subgradient descent method. That is, for chosen initial point  $x_1 \in X$  and a sequence  $\gamma_j > 0$ ,  $j = 1, \dots$ , of stepsizes, it generates the iterates by the formula

$$x_{j+1} = \Pi_X(x_j - \gamma_j G(x_j, \xi^j)). \quad (5.281)$$

The crucial question of that approach is how to choose the stepsizes  $\gamma_j$ . Also, the set  $X$  should be simple enough so that the corresponding projection can be easily calculated. We now analyze convergence of the iterates, generated by this procedure, to an optimal solution  $\bar{x}$  of problem (5.1). Note that the iterate  $x_{j+1} = x_{j+1}(\xi_{[j]})$ ,  $j = 1, \dots$ , is a function of the history  $\xi_{[j]} = (\xi^1, \dots, \xi^j)$  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

$$\mathbb{E}[\|G(x, \xi)\|_2^2] \leq M^2, \quad \forall x \in X. \quad (5.282)$$

Note that since for a random variable  $Z$  it holds that  $\mathbb{E}[Z^2] \geq (\mathbb{E}|Z|)^2$ , it follows from (5.282) that  $\mathbb{E}\|G(x, \xi)\| \leq M$ .

Denote

$$A_j := \frac{1}{2}\|x_j - \bar{x}\|_2^2 \quad \text{and} \quad a_j := \mathbb{E}[A_j] = \frac{1}{2}\mathbb{E}[\|x_j - \bar{x}\|_2^2]. \quad (5.283)$$

By (5.280) and since  $\bar{x} \in X$  and hence  $\Pi_X(\bar{x}) = \bar{x}$ , we have

$$\begin{aligned} A_{j+1} &= \frac{1}{2}\|\Pi_X(x_j - \gamma_j G(x_j, \xi^j)) - \bar{x}\|_2^2 \\ &= \frac{1}{2}\|\Pi_X(x_j - \gamma_j G(x_j, \xi^j)) - \Pi_X(\bar{x})\|_2^2 \\ &\leq \frac{1}{2}\|x_j - \gamma_j G(x_j, \xi^j) - \bar{x}\|_2^2 \\ &= A_j + \frac{1}{2}\gamma_j^2\|G(x_j, \xi^j)\|_2^2 - \gamma_j(x_j - \bar{x})^\top G(x_j, \xi^j). \end{aligned} \quad (5.284)$$

Since  $x_j = x_j(\xi_{[j-1]})$  is independent of  $\xi_j$ , we have

$$\begin{aligned} \mathbb{E}[(x_j - \bar{x})^\top G(x_j, \xi^j)] &= \mathbb{E}\left\{\mathbb{E}[(x_j - \bar{x})^\top G(x_j, \xi^j) \mid \xi_{[j-1]}]\right\} \\ &= \mathbb{E}\left\{(x_j - \bar{x})^\top \mathbb{E}[G(x_j, \xi^j) \mid \xi_{[j-1]}]\right\} \\ &= \mathbb{E}[(x_j - \bar{x})^\top g(x_j)]. \end{aligned}$$

Therefore, by taking expectation of both sides of (5.284) and since (5.282) we obtain

$$a_{j+1} \leq a_j - \gamma_j \mathbb{E}[(x_j - \bar{x})^\top g(x_j)] + \frac{1}{2}\gamma_j^2 M^2. \quad (5.285)$$

Suppose, further, that the expectation function  $f(x)$  is differentiable and strongly convex on  $X$  with parameter  $c > 0$ , i.e.,

$$(x' - x)^\top (\nabla f(x') - \nabla f(x)) \geq c\|x' - x\|_2^2, \quad \forall x, x' \in X. \quad (5.286)$$

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

$$(x - \bar{x})^\top \nabla f(\bar{x}) \geq 0, \quad \forall x \in X, \quad (5.287)$$

which together with (5.286) implies that

$$\begin{aligned} \mathbb{E}[(x_j - \bar{x})^\top \nabla f(x_j)] &\geq \mathbb{E}[(x_j - \bar{x})^\top (\nabla f(x_j) - \nabla f(\bar{x}))] \\ &\geq c \mathbb{E}[\|x_j - \bar{x}\|_2^2] = 2ca_j. \end{aligned} \quad (5.288)$$

Therefore it follows from (5.285) that

$$a_{j+1} \leq (1 - 2c\gamma_j)a_j + \frac{1}{2}\gamma_j^2 M^2. \quad (5.289)$$

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

$$a_{j+1} \leq (1 - 2c\theta/j)a_j + \frac{1}{2}\theta^2 M^2/j^2. \quad (5.290)$$

Suppose now that  $\theta > 1/(2c)$ . Then it follows from (5.290) by induction that for  $j = 1, \dots$ ,

$$2a_j \leq \frac{\max\{\theta^2 M^2(2c\theta - 1)^{-1}, 2a_1\}}{j}. \quad (5.291)$$

Recall that  $2a_j = \mathbb{E}[\|x_j - \bar{x}\|_2^2]$  and, since  $x_1$  is deterministic,  $2a_1 = \|x_1 - \bar{x}\|_2^2$ . Therefore, by (5.291) we have that

$$\mathbb{E}[\|x_j - \bar{x}\|_2^2] \leq \frac{Q(\theta)}{j}, \quad (5.292)$$

where

$$Q(\theta) := \max\{\theta^2 M^2(2c\theta - 1)^{-1}, \|x_1 - \bar{x}\|_2^2\}. \quad (5.293)$$

The constant  $Q(\theta)$  attains its optimal (minimal) value at  $\theta = 1/c$ .

Suppose, further, that  $\bar{x}$  is an *interior* point of  $X$  and  $\nabla f(x)$  is Lipschitz continuous, i.e., there is constant  $L > 0$  such that

$$\|\nabla f(x') - \nabla f(x)\|_2 \leq L\|x' - x\|_2, \quad \forall x', x \in X. \quad (5.294)$$

Then

$$f(x) \leq f(\bar{x}) + \frac{1}{2}L\|x - \bar{x}\|_2^2, \quad \forall x \in X, \quad (5.295)$$

and hence by (5.292)

$$\mathbb{E}[f(x_j) - f(\bar{x})] \leq \frac{1}{2}L \mathbb{E}[\|x_j - \bar{x}\|_2^2] \leq \frac{Q(\theta)L}{2j}. \quad (5.296)$$

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  $X := [-1, 1] \subset \mathbb{R}$  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

$$x_{j+1} = x_j - f'(x_j)/j = \left(1 - \frac{\kappa}{j}\right)x_j. \quad (5.297)$$

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,

$$\sum_{s=1}^j \frac{\kappa}{s - \kappa} \leq \frac{\kappa}{1 - \kappa} + \int_1^j \frac{\kappa}{t - \kappa} dt < \frac{\kappa}{1 - \kappa} + \kappa \ln j - \kappa \ln(1 - \kappa).$$

It follows that

$$x_{j+1} > O(1) j^{-\kappa} \quad \text{and} \quad f(x_{j+1}) > O(1) j^{-2\kappa}, \quad j = 1, \dots \quad (5.298)$$

(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  $\kappa$  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  $\theta$ .

## 5.9.2 Robust SA Approach

It was argued in 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 to the SA method was developed by Polyak [152] and Polyak and Juditsky [153], 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” step-size policy. The main ingredients of Polyak’s scheme (long steps and averaging) were, in

a different form, proposed in Nemirovski and Yudin [135] for problems with general-type Lipschitz continuous convex objectives and for convex-concave saddle point problems. Results of this section go back to Nemirovski and Yudin [135], [136].

Recall that  $g(x) \in \partial f(x)$  and  $a_j = \frac{1}{2} \mathbb{E} [\|x_j - \bar{x}\|_2^2]$ , and we assume the boundedness condition (5.282). By convexity of  $f(x)$  we have that  $f(x) \geq f(x_j) + (x - x_j)^\top g(x_j)$  for any  $x \in X$ , and hence

$$\mathbb{E}[(x_j - \bar{x})^\top g(x_j)] \geq \mathbb{E}[f(x_j) - f(\bar{x})]. \quad (5.299)$$

Together with (5.285) this implies

$$\gamma_j \mathbb{E}[f(x_j) - f(\bar{x})] \leq a_j - a_{j+1} + \frac{1}{2} \gamma_j^2 M^2.$$

It follows that whenever  $1 \leq i \leq j$ , we have

$$\sum_{t=i}^j \gamma_t \mathbb{E}[f(x_t) - f(\bar{x})] \leq \sum_{t=i}^j [a_t - a_{t+1}] + \frac{1}{2} M^2 \sum_{t=i}^j \gamma_t^2 \leq a_i + \frac{1}{2} M^2 \sum_{t=i}^j \gamma_t^2. \quad (5.300)$$

Denote

$$v_t := \frac{\gamma_t}{\sum_{\tau=i}^j \gamma_\tau} \quad \text{and} \quad D_X := \max_{x \in X} \|x - x_1\|_2. \quad (5.301)$$

Clearly  $v_t \geq 0$  and  $\sum_{t=i}^j v_t = 1$ . By (5.300) we have

$$\mathbb{E} \left[ \sum_{t=i}^j v_t f(x_t) - f(\bar{x}) \right] \leq \frac{a_i + \frac{1}{2} M^2 \sum_{t=i}^j \gamma_t^2}{\sum_{t=i}^j \gamma_t}. \quad (5.302)$$

Consider points

$$\tilde{x}_{i,j} := \sum_{t=i}^j v_t x_t. \quad (5.303)$$

Since  $X$  is convex, it follows that  $\tilde{x}_{i,j} \in X$  and by convexity of  $f(\cdot)$  we have

$$f(\tilde{x}_{i,j}) \leq \sum_{t=i}^j v_t f(x_t).$$

Thus, by (5.302) and in view of  $a_1 \leq D_X^2$  and  $a_i \leq 4D_X^2$ ,  $i > 1$ , we get

$$\mathbb{E}[f(\tilde{x}_{1,j}) - f(\bar{x})] \leq \frac{D_X^2 + M^2 \sum_{t=1}^j \gamma_t^2}{2 \sum_{t=1}^j \gamma_t} \quad \text{for } 1 \leq j, \quad (5.304)$$

$$\mathbb{E}[f(\tilde{x}_{i,j}) - f(\bar{x})] \leq \frac{4D_X^2 + M^2 \sum_{t=i}^j \gamma_t^2}{2 \sum_{t=i}^j \gamma_t} \quad \text{for } 1 < i \leq j. \quad (5.305)$$

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.,  $\gamma_t = \gamma$ ,  $t = 1, \dots, N$ . It follows then from (5.304) that

$$\mathbb{E}[f(\tilde{x}_{1,N}) - f(\bar{x})] \leq \frac{D_X^2 + M^2 N \gamma^2}{2N\gamma}. \quad (5.306)$$

Minimizing the right-hand side of (5.306) over  $\gamma > 0$ , we arrive at the *constant* stepsize policy

$$\gamma_t = \frac{D_X}{M\sqrt{N}}, \quad t = 1, \dots, N, \quad (5.307)$$

along with the associated efficiency estimate

$$\mathbb{E}[f(\tilde{x}_{1,N}) - f(\bar{x})] \leq \frac{D_X M}{\sqrt{N}}. \quad (5.308)$$

By (5.305), with the constant stepsize policy (5.307), we also have for  $1 \leq K \leq N$

$$\mathbb{E}[f(\tilde{x}_{K,N}) - f(\bar{x})] \leq \frac{C_{N,K} D_X M}{\sqrt{N}}, \quad (5.309)$$

where

$$C_{N,K} := \frac{2N}{N - K + 1} + \frac{1}{2}.$$

When  $K/N \leq 1/2$ , 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

$$\gamma_t = \frac{\theta D_X}{M\sqrt{N}}, \quad t = 1, \dots, N, \quad (5.310)$$

then the efficiency estimate (5.309) becomes

$$\mathbb{E}[f(\tilde{x}_{K,N}) - f(\bar{x})] \leq \max\{\theta, \theta^{-1}\} \frac{C_{N,K} D_X M}{\sqrt{N}}. \quad (5.311)$$

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  $X$ . 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  $\theta$  results just in rescaling of the corresponding error estimate (5.311). This is in a sharp contrast to 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  $\varepsilon > 0$ ,

$$\Pr\{f(\tilde{x}_{1,N}) - f(\bar{x}) \geq \varepsilon\} \leq \varepsilon^{-1} \mathbb{E}[f(\tilde{x}_{1,N}) - f(\bar{x})]. \quad (5.312)$$

Together with (5.308) this implies that, for the constant stepsize policy (5.307),

$$\Pr \{f(\tilde{x}_{1,N}) - f(\bar{x}) \geq \varepsilon\} \leq \frac{D_X M}{\varepsilon \sqrt{N}}. \quad (5.313)$$

It follows that for  $\alpha \in (0, 1)$  and sample size

$$N \geq \frac{D_X^2 M^2}{\varepsilon^2 \alpha^2} \quad (5.314)$$

we are guaranteed that  $\tilde{x}_{1,N}$  is an  $\varepsilon$ -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 estimate (5.314) is of the same order with respect to parameters  $D_X, M$ , and  $\varepsilon$ . On the other hand, the dependence on the significance level  $\alpha$  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 the large deviations theory; we 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 [136], 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

$$\|x\|_* := \sup_{\|y\| \leq 1} y^\top x.$$

By  $\|x\|_p := (|x_1|^p + \cdots + |x_n|^p)^{1/p}$  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  $\ell_1$  norm  $\|x\|_1 = |x_1| + \cdots + |x_n|$  is the  $\ell_\infty$  norm  $\|x\|_\infty = \max\{|x_1|, \dots, |x_n|\}$ .

**Definition 5.37.** We say that a function  $\mathfrak{d} : X \rightarrow \mathbb{R}$  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  $X$ , the set

$$X^* := \{x \in X : \partial \mathfrak{d}(x) \neq \emptyset\} \quad (5.315)$$

is convex,  $\mathfrak{d}(\cdot)$  is continuously differentiable on  $X^*$ , and

$$(x' - x)^\top (\nabla \mathfrak{d}(x') - \nabla \mathfrak{d}(x)) \geq \kappa \|x' - x\|^2, \quad \forall x, x' \in X^*. \quad (5.316)$$

Note that the set  $X^*$  includes the relative interior of the set  $X$ , and hence condition (5.316) implies that  $\mathfrak{d}(\cdot)$  is strongly convex on  $X$  with the parameter  $\kappa$  taken with respect to the considered norm  $\|\cdot\|$ .



## 5.9. Stochastic Approximation Method

237

A simple example of a distance generating function (with modulus 1 with respect to the Euclidean norm) is  $\vartheta(x) := \frac{1}{2}x^\top x$ . Of course, this function is continuously differentiable at every  $x \in \mathbb{R}^n$ . Another interesting example is the *entropy function*

$$\vartheta(x) := \sum_{i=1}^n x_i \ln x_i, \quad (5.317)$$

defined on the standard simplex  $X := \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1, x_i \geq 0\}$ . (Note that by continuity,  $x \ln x = 0$  for  $x = 0$ .) Here the set  $X^*$  is formed by points  $x \in X$  having all coordinates different from zero. The set  $X^*$  is the subset of  $X$  of those points at which the entropy function is differentiable with  $\nabla \vartheta(x) = (1 + \ln x_1, \dots, 1 + \ln x_n)$ . The entropy function is strongly convex with modulus 1 on standard simplex with respect to  $\|\cdot\|_1$  norm.

Indeed, it suffices to verify that  $h^\top \nabla^2 \vartheta(x) h \geq \|h\|_1^2$  for every  $h \in \mathbb{R}^n$  and  $x \in X^*$ . This, in turn, is verified by

$$\begin{aligned} \left[ \sum_i |h_i| \right]^2 &= \left[ \sum_i (x_i^{-1/2} |h_i|) x_i^{1/2} \right]^2 \leq \left[ \sum_i h_i^2 x_i^{-1} \right] \left[ \sum_i x_i \right] \\ &= \sum_i h_i^2 x_i^{-1} = h^\top \nabla^2 \vartheta(x) h, \end{aligned} \quad (5.318)$$

where the inequality follows by Cauchy inequality.

Let us define function  $V : X^* \times X \rightarrow \mathbb{R}_+$  as follows:

$$V(x, z) := \vartheta(z) - [\vartheta(x) + \nabla \vartheta(x)^\top (z - x)]. \quad (5.319)$$

In what follows we refer to  $V(\cdot, \cdot)$  as the *prox-function*<sup>36</sup> associated with the distance-generating function  $\vartheta(x)$ . Note that  $V(x, \cdot)$  is nonnegative and is strongly convex with modulus  $\kappa$  with respect to the norm  $\|\cdot\|$ . Let us define *prox-mapping*  $P_x : \mathbb{R}^n \rightarrow X^*$ , associated with the distance-generating function and a point  $x \in X^*$ , viewed as a parameter, as follows:

$$P_x(y) := \arg \min_{z \in X} \{y^\top (z - x) + V(x, z)\}. \quad (5.320)$$

Observe that the minimum in the right-hand side of (5.320) is attained since  $\vartheta(\cdot)$  is continuous on  $X$  and  $X$  is compact, and a corresponding minimizer is unique since  $V(x, \cdot)$  is strongly convex on  $X$ . Moreover, by the definition of the set  $X^*$ , all these minimizers belong to  $X^*$ . Thus, the prox-mapping is well defined.

For the (Euclidean) distance-generating function  $\vartheta(x) := \frac{1}{2}x^\top x$ , we have that  $P_x(y) = \Pi_X(x - y)$ . In that case the iteration formula (5.281) of the SA algorithm can be written as

$$x_{j+1} = P_{x_j}(\gamma_j G(x_j, \xi^j)), \quad x_1 \in X^*. \quad (5.321)$$

Our goal is to demonstrate that the main properties of the recurrence (5.281) are inherited by (5.321) for any distance-generating function  $\vartheta(x)$ .

**Lemma 5.38.** *For every  $u \in X$ ,  $x \in X^*$  and  $y \in \mathbb{R}^n$  one has*

$$V(P_x(y), u) \leq V(x, u) + y^\top (u - x) + (2\kappa)^{-1} \|y\|_*^2. \quad (5.322)$$

<sup>36</sup>The function  $V(\cdot, \cdot)$  is also called Bregman divergence.

**Proof.** Let  $x \in X^*$  and  $v := P_x(y)$ . Note that  $v$  is of the form  $\operatorname{argmin}_{z \in X} [h^\top z + \mathfrak{d}(z)]$  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

$$(\nabla \mathfrak{d}(v) - \nabla \mathfrak{d}(x) + y)^\top (v - u) \leq 0, \quad \forall u \in X. \quad (5.323)$$

Therefore, for  $u \in X$  we have

$$\begin{aligned} V(v, u) - V(x, u) &= [\mathfrak{d}(u) - \nabla \mathfrak{d}(v)^\top (u - v) - \mathfrak{d}(v)] - [\mathfrak{d}(u) - \nabla \mathfrak{d}(x)^\top (u - x) - \mathfrak{d}(x)] \\ &= (\nabla \mathfrak{d}(v) - \nabla \mathfrak{d}(x) + y)^\top (v - u) + y^\top (u - v) - [\mathfrak{d}(v) - \nabla \mathfrak{d}(x)^\top (v - x) - \mathfrak{d}(x)] \\ &\leq y^\top (u - v) - V(x, v), \end{aligned}$$

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  $\|a\|_* \|b\| \geq a^\top b$  and hence

$$(\|a\|_*^2/\kappa + \kappa \|b\|^2)/2 \geq \|a\|_* \|b\| \geq a^\top b. \quad (5.324)$$

Applying this inequality with  $a = y$  and  $b = x - v$  we obtain

$$y^\top (x - v) \leq \frac{\|y\|_*^2}{2\kappa} + \frac{\kappa}{2} \|x - v\|^2.$$

Also due to the strong convexity of  $V(x, \cdot)$  and since  $V(x, x) = 0$  we have

$$\begin{aligned} V(x, v) &\geq V(x, x) + (x - v)^\top \nabla_v V(x, v) + \frac{1}{2} \kappa \|x - v\|^2 \\ &= (x - v)^\top (\nabla \mathfrak{d}(v) - \nabla \mathfrak{d}(x)) + \frac{1}{2} \kappa \|x - v\|^2 \\ &\geq \frac{1}{2} \kappa \|x - v\|^2, \end{aligned} \quad (5.325)$$

where the last inequality holds by convexity of  $\mathfrak{d}(\cdot)$ . We get

$$\begin{aligned} V(v, u) - V(x, u) &\leq y^\top (u - v) - V(x, v) = y^\top (u - x) + y^\top (x - v) - V(x, v) \\ &\leq y^\top (u - x) + (2\kappa)^{-1} \|y\|_*^2, \end{aligned}$$

as required in (5.322).  $\square$

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

$$\gamma_j (x_j - \bar{x})^\top G(x_j, \xi^j) \leq V(x_j, \bar{x}) - V(x_{j+1}, \bar{x}) + \frac{\gamma_j^2}{2\kappa} \|G(x_j, \xi^j)\|_*^2. \quad (5.326)$$

Let us observe that for the Euclidean distance-generating function  $\mathfrak{d}(x) = \frac{1}{2} x^\top x$ , one has  $V(x, z) = \frac{1}{2} \|x - z\|_2^2$  and  $\kappa = 1$ . That is, in the Euclidean case (5.326) becomes

$$\frac{1}{2} \|x_{j+1} - \bar{x}\|_2^2 \leq \frac{1}{2} \|x_j - \bar{x}\|_2^2 + \frac{1}{2} \gamma_j^2 \|G(x_j, \xi^j)\|_2^2 - \gamma_j (x_j - \bar{x})^\top G(x_j, \xi^j). \quad (5.327)$$

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.

## 5.9. Stochastic Approximation Method

239

Specifically, setting

$$\Delta_j := G(x_j, \xi^j) - g(x_j), \quad (5.328)$$

we can rewrite (5.326), with  $j$  replaced by  $t$ , as

$$\gamma_t(x_t - \bar{x})^\top g(x_t) \leq V(x_t, \bar{x}) - V(x_{t+1}, \bar{x}) - \gamma_t \Delta_t^\top (x_t - \bar{x}) + \frac{\gamma_t^2}{2\kappa} \|G(x_t, \xi^t)\|_*^2. \quad (5.329)$$

Summing up over  $t = 1, \dots, j$ , and taking into account that  $V(x_{j+1}, u) \geq 0$ ,  $u \in X$ , we get

$$\sum_{t=1}^j \gamma_t(x_t - \bar{x})^\top g(x_t) \leq V(x_1, \bar{x}) + \sum_{t=1}^j \frac{\gamma_t^2}{2\kappa} \|G(x_t, \xi^t)\|_*^2 - \sum_{t=1}^j \gamma_t \Delta_t^\top (x_t - \bar{x}). \quad (5.330)$$

Set  $v_t := \frac{\gamma_t}{\sum_{\tau=1}^j \gamma_\tau}$ ,  $t = 1, \dots, j$ , and

$$\tilde{x}_{1,j} := \sum_{t=1}^j v_t x_t. \quad (5.331)$$

By convexity of  $f(\cdot)$  we have  $f(x_t) - f(\bar{x}) \leq (x_t - \bar{x})^\top g(x_t)$ , and hence

$$\begin{aligned} \sum_{t=1}^j \gamma_t(x_t - \bar{x})^\top g(x_t) &\geq \sum_{t=1}^j \gamma_t [f(x_t) - f(\bar{x})] \\ &= \left( \sum_{t=1}^j \gamma_t \right) \left[ \sum_{t=1}^j v_t f(x_t) - f(\bar{x}) \right] \\ &\geq \left( \sum_{t=1}^j \gamma_t \right) [f(\tilde{x}_{1,j}) - f(\bar{x})]. \end{aligned}$$

Combining this with (5.330) we obtain

$$f(\tilde{x}_{1,j}) - f(\bar{x}) \leq \frac{V(x_1, \bar{x}) + \sum_{t=1}^j (2\kappa)^{-1} \gamma_t^2 \|G(x_t, \xi^t)\|_*^2 - \sum_{t=1}^j \gamma_t \Delta_t^\top (x_t - \bar{x})}{\sum_{t=1}^j \gamma_t}. \quad (5.332)$$

- Assume from now on that the procedure starts with the minimizer of  $\mathfrak{d}(\cdot)$ , that is,

$$x_1 := \operatorname{argmin}_{x \in X} \mathfrak{d}(x). \quad (5.333)$$

Since by the optimality of  $x_1$  we have that  $(u - x_1)^\top \nabla \mathfrak{d}(x_1) \geq 0$  for any  $u \in X$ , it follows from the definition (5.319) of the function  $V(\cdot, \cdot)$  that

$$\max_{u \in X} V(x_1, u) \leq D_{\mathfrak{d}, X}^2, \quad (5.334)$$

where

$$D_{\mathfrak{d}, X} := \left[ \max_{u \in X} \mathfrak{d}(u) - \min_{x \in X} \mathfrak{d}(x) \right]^{1/2}. \quad (5.335)$$

Together with (5.332) this implies

$$f(\tilde{x}_{1,j}) - f(\bar{x}) \leq \frac{D_{\mathfrak{d}, X}^2 + \sum_{t=1}^j (2\kappa)^{-1} \gamma_t^2 \|G(x_t, \xi^t)\|_*^2 - \sum_{t=1}^j \gamma_t \Delta_t^\top (x_t - \bar{x})}{\sum_{t=1}^j \gamma_t}. \quad (5.336)$$

We also have (see (5.325)) that  $V(x_1, u) \geq \frac{1}{2}\kappa \|x_1 - u\|^2$ , and hence it follows from (5.334) that for all  $u \in X$ ,

$$\|x_1 - u\| \leq \sqrt{\frac{2}{\kappa}} D_{\mathfrak{d}, X}. \quad (5.337)$$

Let us assume, as in the previous section (see (5.282)), that there is a positive number  $M_*$  such that

$$\mathbb{E} [\|G(x, \xi)\|_*^2] \leq M_*^2, \quad \forall x \in X. \quad (5.338)$$

**Proposition 5.39.** *Let  $x_1 := \operatorname{argmin}_{x \in X} \mathfrak{d}(x)$  and suppose that condition (5.338) holds. Then*

$$\mathbb{E} [f(\tilde{x}_{1,j}) - f(\bar{x})] \leq \frac{D_{\mathfrak{d}, X}^2 + (2\kappa)^{-1} M_*^2 \sum_{t=1}^j \gamma_t^2}{\sum_{t=1}^j \gamma_t}. \quad (5.339)$$

**Proof.** Taking expectations of both sides of (5.336) and noting that (i)  $x_t$  is a deterministic function of  $\xi_{[t-1]} = (\xi^1, \dots, \xi^{t-1})$ , (ii) conditional on  $\xi_{[t-1]}$ , the expectation of  $\Delta_t$  is 0, and (iii) the expectation of  $\|G(x_t, \xi^t)\|_*^2$  does not exceed  $M_*^2$ , we obtain (5.339).  $\square$

### Constant Stepsize Policy

Assume that the total number of steps  $N$  is given in advance and the constant stepsize policy  $\gamma_t = \gamma$ ,  $t = 1, \dots, N$ , is employed. Then (5.339) becomes

$$\mathbb{E} [f(\tilde{x}_{1,j}) - f(\bar{x})] \leq \frac{D_{\mathfrak{d}, X}^2 + (2\kappa)^{-1} M_*^2 N \gamma^2}{N \gamma}. \quad (5.340)$$

Minimizing the right-hand side of (5.340) over  $\gamma > 0$  we arrive at the constant stepsize policy

$$\gamma_t = \frac{\sqrt{2\kappa} D_{\mathfrak{d}, X}}{M_* \sqrt{N}}, \quad t = 1, \dots, N, \quad (5.341)$$

and the associated efficiency estimate

$$\mathbb{E} [f(\tilde{x}_{1,N}) - f(\bar{x})] \leq D_{\mathfrak{d}, X} M_* \sqrt{\frac{2}{\kappa N}}. \quad (5.342)$$

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

$$\gamma_t = \frac{\theta \sqrt{2\kappa} D_{\mathfrak{d}, X}}{M_* \sqrt{N}}, \quad t = 1, \dots, N, \quad (5.343)$$

with rescaling parameter  $\theta > 0$ , the efficiency estimate becomes

$$\mathbb{E} [f(\tilde{x}_{1,N}) - f(\bar{x})] \leq \max \{\theta, \theta^{-1}\} D_{\mathfrak{d}, X} M_* \sqrt{\frac{2}{\kappa N}}, \quad (5.344)$$

similar to the Euclidean case. We refer to the SA method based on (5.321), (5.331), and (5.343) as the 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  $X := \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1, x \geq 0\}$  be the standard simplex. Consider two setups for the mirror descent SA, namely, the *Euclidean setup*, where the considered norm  $\|\cdot\| := \|\cdot\|_2$  and  $\mathfrak{d}(x) := \frac{1}{2}x^\top x$ , and  $\ell_1$ -*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  $X$  is  $\sqrt{2}$  and hence is independent of  $n$ . The corresponding efficiency estimate is

$$\mathbb{E}[f(\tilde{x}_{1,N}) - f(\bar{x})] \leq O(1) \max\{\theta, \theta^{-1}\} M N^{-1/2} \quad (5.345)$$

with  $M^2 = \sup_{x \in X} \mathbb{E}[\|G(x, \xi)\|_2^2]$ .

The  $\ell_1$ -setup corresponds to  $X^\star = \{x \in X : x > 0\}$ ,  $D_{\mathfrak{d}, X} = \sqrt{\ln n}$ ,

$$x_1 := \operatorname{argmin}_{x \in X} \mathfrak{d}(x) = n^{-1}(1, \dots, 1)^\top,$$

$\|x\|_* = \|x\|_\infty$ , 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

$$[P_x(y)]_i = \frac{x_i e^{-y_i}}{\sum_{k=1}^n x_k e^{-y_k}}, \quad i = 1, \dots, n.$$

The respective efficiency estimate of the  $\ell_1$ -setup is

$$\mathbb{E}[f(\tilde{x}_{1,N}) - f(\bar{x})] \leq O(1) \max\{\theta, \theta^{-1}\} (\ln n)^{1/2} M_* N^{-1/2} \quad (5.346)$$

with  $M_*^2 = \sup_{x \in X} \mathbb{E}[\|G(x, \xi)\|_\infty^2]$ , 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  $\ell_1$ -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,

$$\sqrt{\frac{1}{\ln n}} \leq \frac{M}{\sqrt{\ln n} M_*} \leq \sqrt{\frac{n}{\ln n}},$$

with both the upper and lower bounds being achievable. Thus, when  $X$  is a standard simplex of large dimension, we have strong reasons to prefer the  $\ell_1$ -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  $\varepsilon$ -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

$$\mathbb{E} \left[ \exp \left\{ \|G(x, \xi)\|_*^2 / M_*^2 \right\} \right] \leq \exp\{1\}, \quad \forall x \in X. \quad (5.347)$$

Note that condition (5.347) is stronger than (5.338). Indeed, if a random variable  $Y$  satisfies  $\mathbb{E}[\exp\{Y/a\}] \leq \exp\{1\}$  for some  $a > 0$ , then by Jensen inequality

$$\exp\{\mathbb{E}[Y/a]\} \leq \mathbb{E}[\exp\{Y/a\}] \leq \exp\{1\},$$

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  $(x, \xi) \in X \times \Xi$ .

**Theorem 5.41.** *Suppose that condition (5.347) is fulfilled. Then for the constant stepsizes (5.343), the following holds for any  $\Theta \geq 0$ :*

$$\Pr \left\{ f(\tilde{x}_{1,N}) - f(\bar{x}) \geq \frac{C(1 + \Theta)}{\sqrt{\kappa N}} \right\} \leq 4 \exp\{-\Theta\}, \quad (5.348)$$

where  $C := (\max\{\theta, \theta^{-1}\} + 8\sqrt{3})M_*D_{\mathfrak{d},X}/\sqrt{2}$ .

**Proof.** By (5.336) we have

$$f(\tilde{x}_{1,N}) - f(\bar{x}) \leq A_1 + A_2, \quad (5.349)$$

where

$$A_1 := \frac{D_{\mathfrak{d},X}^2 + (2\kappa)^{-1} \sum_{t=1}^N \gamma_t^2 \|G(x_t, \xi^t)\|_*^2}{\sum_{t=1}^N \gamma_t} \quad \text{and} \quad A_2 := \sum_{t=1}^N \nu_t \Delta_t^\top (\bar{x} - x_t).$$

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),

$$\mathbb{E}[\exp\{Y_i/c_i\}] \leq \exp\{1\}, \quad i = 1, \dots, N. \quad (5.350)$$

Since  $\exp\{\cdot\}$  is a convex function we have

$$\exp \left\{ \frac{\sum_{i=1}^N Y_i}{\sum_{i=1}^N c_i} \right\} = \exp \left\{ \sum_{i=1}^N \frac{c_i(Y_i/c_i)}{\sum_{i=1}^N c_i} \right\} \leq \sum_{i=1}^N \frac{c_i}{\sum_{i=1}^N c_i} \exp\{Y_i/c_i\}.$$

By taking expectation of both sides of the above inequality and using (5.350) we obtain

$$\mathbb{E} \left[ \exp \left\{ \frac{\sum_{i=1}^N Y_i}{\sum_{i=1}^N c_i} \right\} \right] \leq \exp\{1\}.$$

Consequently by Chebyshev's inequality we have for any number  $\Theta$

$$\Pr \left[ \exp \left\{ \frac{\sum_{i=1}^N Y_i}{\sum_{i=1}^N c_i} \right\} \geq \exp\{\Theta\} \right] \leq \frac{\exp\{1\}}{\exp\{\Theta\}} = \exp\{1 - \Theta\},$$

and hence

$$\Pr \left\{ \sum_{i=1}^N Y_i \geq \Theta \sum_{i=1}^N c_i \right\} \leq \exp\{1 - \Theta\} \leq 3 \exp\{-\Theta\}. \quad (5.351)$$

That is, for any  $\Theta$ ,

$$\Pr \left\{ \sum_{t=1}^N \gamma_t^2 \|G(x_t, \xi^t)\|_*^2 \geq \Theta M_*^2 \sum_{t=1}^N \gamma_t^2 \right\} \leq 3 \exp\{-\Theta\}. \quad (5.352)$$

For the constant stepsize policy (5.343), we obtain by (5.352) that

$$\Pr \left\{ A_1 \geq \max\{\theta, \theta^{-1}\} \frac{M_* D_{\mathfrak{d},X}(1+\Theta)}{\sqrt{2\kappa N}} \right\} \leq 3 \exp\{-\Theta\}. \quad (5.353)$$

Consider now the random variable  $A_2$ . By (5.337) we have that

$$\|\bar{x} - x_t\| \leq \|x_1 - \bar{x}\| + \|x_1 - x_t\| \leq 2\sqrt{2\kappa}^{-1/2} D_{\mathfrak{d},X},$$

and hence

$$|\Delta_t^\top (\bar{x} - x_t)|^2 \leq \|\Delta_t\|_*^2 \|\bar{x} - x_t\|^2 \leq 8\kappa^{-1} D_{\mathfrak{d},X}^2 \|\Delta_t\|_*^2.$$

We also have that

$$\mathbb{E}[(\bar{x} - x_t)^\top \Delta_t | \xi_{[t-1]}] = (\bar{x} - x_t)^\top \mathbb{E}[\Delta_t | \xi_{[t-1]}] = 0 \quad \text{w.p. 1,}$$

and by condition (5.347) that

$$\mathbb{E}[\exp\{\|\Delta_t\|_*^2 / (4M_*^2)\} | \xi_{[t-1]}] \leq \exp\{1\} \quad \text{w.p. 1.}$$

Consequently, by applying inequality (7.194) of Proposition 7.64 with  $\phi_t := v_t \Delta_t^\top (\bar{x} - x_t)$  and  $\sigma_t^2 := 32\kappa^{-1} M_*^2 D_{\mathfrak{d},X}^2 v_t^2$ , we obtain for any  $\Theta \geq 0$

$$\Pr \left\{ A_2 \geq 4\sqrt{2\kappa}^{-1/2} M_* D_{\mathfrak{d},X} \Theta \sqrt{\sum_{t=1}^N v_t^2} \right\} \leq \exp\{-\Theta^2/3\}. \quad (5.354)$$

Since for the constant stepsize policy we have that  $v_t = 1/N$ ,  $t = 1, \dots, N$ , by changing variables  $\Theta^2/3$  to  $\Theta$  and noting that  $\Theta^{1/2} \leq 1 + \Theta$  for any  $\Theta \geq 0$ , we obtain from (5.354) that for any  $\Theta \geq 0$

$$\Pr \left\{ A_2 \geq \frac{8\sqrt{3} M_* D_{\mathfrak{d},X}(1+\Theta)}{\sqrt{2\kappa N}} \right\} \leq \exp\{-\Theta\}. \quad (5.355)$$

Finally, (5.348) follows from (5.349), (5.353), and (5.355).  $\square$

By setting  $\varepsilon = \frac{C(1+\Theta)}{\sqrt{\kappa N}}$ , we can rewrite the estimate (5.348) in the form<sup>37</sup>

$$\Pr \{f(\tilde{x}_{1,N}) - f(\bar{x}) > \varepsilon\} \leq 12 \exp\{-\varepsilon C^{-1} \sqrt{\kappa N}\}. \quad (5.356)$$

<sup>37</sup>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  $\varepsilon$ -optimal solution of the true problem with probability at least  $1 - \alpha$ :

$$N \geq O(1)\varepsilon^{-2}\kappa^{-1}M_*^2D_{0,X}^2\ln^2(12/\alpha). \quad (5.357)$$

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, 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  $x_1, \dots, x_N$  computed according to formula (5.321). As before, set

$$v_t := \frac{\gamma_t}{\sum_{\tau=1}^N \gamma_\tau}, \quad t = 1, \dots, N, \quad \text{and} \quad \tilde{x}_{1,N} := \sum_{t=1}^N v_t x_t.$$

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

$$f_*^N := \min_{x \in X} f^N(x) \quad \text{and} \quad f^{*N} := \sum_{t=1}^N v_t f(x_t), \quad (5.358)$$

where

$$f^N(x) := \sum_{t=1}^N v_t [f(x_t) + g(x_t)^\top (x - x_t)]. \quad (5.359)$$

Since  $v_t > 0$  and  $\sum_{t=1}^N v_t = 1$ , by convexity of  $f(x)$  we have that the function  $f^N(x)$  underestimates  $f(x)$  everywhere on  $X$ , and hence<sup>38</sup>  $f_*^N \leq \vartheta^*$ . Since  $\tilde{x}_{1,N} \in X$  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  $\xi^1, \dots$ , we have that

$$f_*^N \leq \vartheta^* \leq f^{*N}. \quad (5.360)$$

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

$$\begin{aligned} \underline{f}^N &:= \min_{x \in X} \left\{ \sum_{t=1}^N v_t [F(x_t, \xi^t) + G(x_t, \xi^t)^\top (x - x_t)] \right\}, \\ \overline{f}^N &:= \sum_{t=1}^N v_t F(x_t, \xi^t), \end{aligned} \quad (5.361)$$

<sup>38</sup>Recall that  $\vartheta^*$  denotes the optimal value of the true problem (5.1).



which will be referred to as *online* bounds. The bound  $\bar{f}^N$  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  $X$ . If the set  $X$  is defined by linear constraints, this is a linear programming problem.

Since  $x_t$  is a function of  $\xi_{[t-1]}$  and  $\xi^t$  is independent of  $\xi_{[t-1]}$ , we have that

$$\mathbb{E}[\bar{f}^N] = \sum_{t=1}^N v_t \mathbb{E}\{\mathbb{E}[F(x_t, \xi^t) | \xi_{[t-1]}]\} = \sum_{t=1}^N v_t \mathbb{E}[f(x_t)] = \mathbb{E}[f^{*N}]$$

and

$$\begin{aligned} \mathbb{E}[\underline{f}^N] &= \mathbb{E}\left[\mathbb{E}\left\{\min_{x \in X} \left\{ \sum_{t=1}^N v_t [F(x_t, \xi^t) + G(x_t, \xi^t)^\top (x - x_t)] \right\} \middle| \xi_{[t-1]} \right\}\right] \\ &\leq \mathbb{E}\left[\min_{x \in X} \left\{ \mathbb{E}\left[ \sum_{t=1}^N v_t [F(x_t, \xi^t) + G(x_t, \xi^t)^\top (x - x_t)] \right] \middle| \xi_{[t-1]} \right\}\right] \\ &= \mathbb{E}\left[\min_{x \in X} f^N(x)\right] = \mathbb{E}[f_*^N]. \end{aligned}$$

It follows that

$$\mathbb{E}[\underline{f}^N] \leq \vartheta^* \leq \mathbb{E}[\bar{f}^N]. \quad (5.362)$$

That is, on average  $\underline{f}^N$  and  $\bar{f}^N$  give, respectively, a lower and an upper bound for the optimal value  $\vartheta^*$  of the optimization problem (5.1).

In order to see how good the bounds  $\underline{f}^N$  and  $\bar{f}^N$  are, let us estimate expectations of the corresponding errors. We will need the following result.

**Lemma 5.42.** *Let  $\zeta_t \in \mathbb{R}^n$ ,  $v_1 \in X^*$ , and  $v_{t+1} = P_{v_t}(\zeta_t)$ ,  $t = 1, \dots, N$ . Then*

$$\sum_{t=1}^N \zeta_t^\top (v_t - u) \leq V(v_1, u) + (2\kappa)^{-1} \sum_{t=1}^N \|\zeta_t\|_*^2, \quad \forall u \in X. \quad (5.363)$$

**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  $u \in X$ :

$$V(v_{t+1}, u) \leq V(v_t, u) + \zeta_t^\top (u - v_t) + (2\kappa)^{-1} \|\zeta_t\|_*^2. \quad (5.364)$$

Summing this over  $t = 1, \dots, N$ , we obtain

$$V(v_{N+1}, u) \leq V(v_1, u) + \sum_{t=1}^N \zeta_t^\top (u - v_t) + (2\kappa)^{-1} \sum_{t=1}^N \|\zeta_t\|_*^2. \quad (5.365)$$

Since  $V(v_{N+1}, u) \geq 0$ , (5.363) follows.  $\square$

Consider again condition (5.338), that is,

$$\mathbb{E}[\|G(x, \xi)\|_*^2] \leq M_*^2, \quad \forall x \in X, \quad (5.366)$$

and the following condition: there is a constant  $Q > 0$  such that

$$\mathbb{V}\text{ar}[F(x, \xi)] \leq Q^2, \quad \forall x \in X. \quad (5.367)$$

Note that, of course,  $\mathbb{V}\text{ar}[F(x, \xi)] = \mathbb{E}[(F(x, \xi) - f(x))^2]$ .

**Theorem 5.43.** *Suppose that conditions (5.366) and (5.367) hold. Then*

$$\mathbb{E}[f^{*N} - f_*^N] \leq \frac{2D_{\mathfrak{d},X}^2 + \frac{5}{2}\kappa^{-1}M_*^2 \sum_{t=1}^N \gamma_t^2}{\sum_{t=1}^N \gamma_t}, \quad (5.368)$$

$$\mathbb{E}[|\bar{f}^N - f^{*N}|] \leq Q \sqrt{\sum_{t=1}^N v_t^2}, \quad (5.369)$$

$$\begin{aligned} \mathbb{E}[|\underline{f}^N - f_*^N|] &\leq \left(Q + 4\sqrt{2}\kappa^{-1/2}M_*D_{\mathfrak{d},X}\right) \sqrt{\sum_{t=1}^N v_t^2} \\ &\quad + \frac{D_{\mathfrak{d},X}^2 + 2\kappa^{-1}M_*^2 \sum_{t=1}^N \gamma_t^2}{\sum_{t=1}^N \gamma_t}. \end{aligned} \quad (5.370)$$

**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  $V(x_1, u) \leq D_{\mathfrak{d},X}^2$  that

$$\sum_{t=1}^N \gamma_t (x_t - u)^\top G(x_t, \xi^t) \leq D_{\mathfrak{d},X}^2 + (2\kappa)^{-1} \sum_{t=1}^N \gamma_t^2 \|G(x_t, \xi^t)\|_*^2, \quad \forall u \in X. \quad (5.371)$$

It follows that for any  $u \in X$  (compare with (5.330)),

$$\begin{aligned} &\sum_{t=1}^N v_t [-f(x_t) + (x_t - u)^\top g(x_t)] + \sum_{t=1}^N v_t f(x_t) \\ &\leq \frac{D_{\mathfrak{d},X}^2 + (2\kappa)^{-1} \sum_{t=1}^N \gamma_t^2 \|G(x_t, \xi^t)\|_*^2}{\sum_{t=1}^N \gamma_t} + \sum_{t=1}^N v_t \Delta_t^\top (x_t - u), \end{aligned}$$

where  $\Delta_t := G(x_t, \xi^t) - g(x_t)$ . Since

$$f^{*N} - f_*^N = \sum_{t=1}^N v_t f(x_t) + \max_{u \in X} \sum_{t=1}^N v_t [-f(x_t) + (x_t - u)^\top g(x_t)],$$

it follows that

$$f^{*N} - f_*^N \leq \frac{D_{\mathfrak{d},X}^2 + (2\kappa)^{-1} \sum_{t=1}^N \gamma_t^2 \|G(x_t, \xi^t)\|_*^2}{\sum_{t=1}^N \gamma_t} + \max_{u \in X} \sum_{t=1}^N v_t \Delta_t^\top (x_t - u). \quad (5.372)$$

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

$$\sum_{t=1}^N \gamma_t \Delta_t^\top (v_t - u) \leq D_{\mathfrak{d},X}^2 + (2\kappa)^{-1} \sum_{t=1}^N \gamma_t^2 \|\Delta_t\|_*^2, \quad \forall u \in X. \quad (5.373)$$

Moreover,

$$\Delta_t^\top(v_t - u) = \Delta_t^\top(x_t - u) + \Delta_t^\top(v_t - x_t),$$

and hence it follows by (5.373) that

$$\max_{u \in X} \sum_{t=1}^N v_t \Delta_t^\top(x_t - u) \leq \sum_{t=1}^N v_t \Delta_t^\top(x_t - v_t) + \frac{D_{\mathfrak{d},X}^2 + (2\kappa)^{-1} \sum_{t=1}^N \gamma_t^2 \|\Delta_t\|_*^2}{\sum_{t=1}^N \gamma_t}. \quad (5.374)$$

Moreover,  $\mathbb{E}[\Delta_t | \xi_{[t-1]}] = 0$  and  $v_t$  and  $x_t$  are functions of  $\xi_{[t-1]}$ , and hence

$$\mathbb{E}[(x_t - v_t)^\top \Delta_t] = \mathbb{E}\{(x_t - v_t)^\top \mathbb{E}[\Delta_t | \xi_{[t-1]}]\} = 0. \quad (5.375)$$

In view of condition (5.366), we have that  $\mathbb{E}[\|\Delta_t\|_*^2] \leq 4M_*^2$ , and hence it follows from (5.374) and (5.375) that

$$\mathbb{E} \left[ \max_{u \in X} \sum_{t=1}^N v_t \Delta_t^\top(x_t - u) \right] \leq \frac{D_{\mathfrak{d},X}^2 + 2\kappa^{-1} M_*^2 \sum_{t=1}^N \gamma_t^2}{\sum_{t=1}^N \gamma_t}. \quad (5.376)$$

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 prove (5.369), let us observe that

$$\bar{f}^N - f^{*N} = \sum_{t=1}^N v_t (F(x_t, \xi^t) - f(x_t)),$$

and that for  $1 \leq s < t \leq N$ ,

$$\begin{aligned} & \mathbb{E}[(F(x_s, \xi^s) - f(x_s))(F(x_t, \xi^t) - f(x_t))] \\ &= \mathbb{E}\{\mathbb{E}[(F(x_s, \xi^s) - f(x_s))(F(x_t, \xi^t) - f(x_t)) | \xi_{[t-1]}]\} \\ &= \mathbb{E}\{(F(x_s, \xi^s) - f(x_s))\mathbb{E}[(F(x_t, \xi^t) - f(x_t)) | \xi_{[t-1]}]\} = 0. \end{aligned}$$

Therefore

$$\begin{aligned} \mathbb{E}[(\bar{f}^N - f^{*N})^2] &= \sum_{t=1}^N v_t^2 \mathbb{E}[(F(x_t, \xi^t) - f(x_t))^2] \\ &= \sum_{t=1}^N v_t^2 \mathbb{E}\left\{\mathbb{E}[(F(x_t, \xi^t) - f(x_t))^2 | \xi_{[t-1]}]\right\} \\ &\leq Q^2 \sum_{t=1}^N v_t^2, \end{aligned} \quad (5.377)$$

where the last inequality is implied by condition (5.367). Since for any random variable  $Y$  we have that  $\sqrt{\mathbb{E}[Y^2]} \geq \mathbb{E}[|Y|]$ , the inequality (5.369) follows from (5.377).

Let us now look at (5.370). Denote

$$\tilde{f}^N(x) := \sum_{t=1}^N v_t [F(x_t, \xi^t) + G(x_t, \xi^t)^\top (x - x_t)].$$

Then

$$|\underline{f}^N - f_*^N| = \left| \min_{x \in X} \tilde{f}^N(x) - \min_{x \in X} f^N(x) \right| \leq \max_{x \in X} |\tilde{f}^N(x) - f^N(x)|$$

and

$$\tilde{f}^N(x) - f^N(x) = \bar{f}^N - f^{*N} + \sum_{t=1}^N v_t \Delta_t^\top(x_t - x),$$

and hence

$$\left| \underline{f}^N - f_*^N \right| \leq \left| \bar{f}^N - f^{*N} \right| + \left| \max_{x \in X} \sum_{t=1}^N v_t \Delta_t^\top(x_t - x) \right|. \quad (5.378)$$

For  $\mathbb{E}[\bar{f}^N - f^{*N}]$  we already have the estimate (5.369).

By (5.373) we have

$$\left| \max_{x \in X} \sum_{t=1}^N v_t \Delta_t^\top(x_t - x) \right| \leq \left| \sum_{t=1}^N v_t \Delta_t^\top(x_t - v_t) \right| + \frac{D_{\mathfrak{d},X}^2 + (2\kappa)^{-1} \sum_{t=1}^N \gamma_t^2 \|\Delta_t\|_*^2}{\sum_{t=1}^N \gamma_t}. \quad (5.379)$$

Let us observe that for  $1 \leq s < t \leq N$

$$\begin{aligned} \mathbb{E}[(\Delta_s^\top(x_s - v_s))(\Delta_t^\top(x_t - v_t))] &= \mathbb{E}\{\mathbb{E}[(\Delta_s^\top(x_s - v_s))(\Delta_t^\top(x_t - v_t)) | \xi_{[t-1]}]\} \\ &= \mathbb{E}\{(\Delta_s^\top(x_s - v_s))\mathbb{E}[(\Delta_t^\top(x_t - v_t)) | \xi_{[t-1]}]\} = 0. \end{aligned}$$

Therefore, by condition (5.366) we have

$$\begin{aligned} \mathbb{E}\left[\left(\sum_{t=1}^N v_t \Delta_t^\top(x_t - v_t)\right)^2\right] &= \sum_{t=1}^N v_t^2 \mathbb{E}\left[|\Delta_t^\top(x_t - v_t)|^2\right] \\ &\leq \sum_{t=1}^N v_t^2 \mathbb{E}\left[\|\Delta_t\|_*^2 \|x_t - v_t\|^2\right] = \sum_{t=1}^N v_t^2 \mathbb{E}\left[\|x_t - v_t\|^2 \mathbb{E}[\|\Delta_t\|_*^2 | \xi_{[t-1]}]\right] \\ &\leq 4M_*^2 \sum_{t=1}^N v_t^2 \mathbb{E}\left[\|x_t - v_t\|^2\right] \leq 32\kappa^{-1} M_*^2 D_{\mathfrak{d},X}^2 \sum_{t=1}^N v_t^2, \end{aligned}$$

where the last inequality follows by (5.337). It follows that

$$\mathbb{E}\left[\left|\sum_{t=1}^N v_t \Delta_t^\top(x_t - v_t)\right|\right] \leq 4\sqrt{2}\kappa^{-1/2} M_* D_{\mathfrak{d},X} \sqrt{\sum_{t=1}^N v_t^2}. \quad (5.380)$$

Putting together (5.378), (5.379), (5.380), and (5.369), we obtain (5.370).  $\square$

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, the difference between the upper  $\bar{f}^N$  and lower  $\underline{f}^N$  bounds converges 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 [114]).

**Remark 19.** 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  $\xi^1, \dots, \xi^N$  is employed for both the SA and the SAA method, that  $F(\cdot, \xi)$  is convex for all  $\xi \in \Xi$ , and  $G(x, \xi) \in \partial_x F(x, \xi)$  for all  $(x, \xi) \in X \times \Xi$ . By convexity of  $F(\cdot, \xi)$  and definition of  $\underline{f}^N$ , we have

$$\begin{aligned} \hat{\vartheta}_N &= \min_{x \in X} \left\{ N^{-1} \sum_{t=1}^N F(x, \xi^t) \right\} \\ &\geq \min_{x \in X} \left\{ \sum_{t=1}^N v_t [F(x_t, \xi^t) + G(x_t, \xi^t)^\top(x - x_t)] \right\} = \underline{f}^N. \end{aligned} \quad (5.381)$$

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.

## Exercises

- 5.1. Suppose that set  $X$  is defined by constraints in the form (5.11) with constraint functions given as expectations as in (5.12) and the set  $X_N$  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,  $i = 1, \dots, p$ , 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  $X \subset \mathbb{R}^n$  be a closed convex set. Show that the multifunction  $x \mapsto \mathcal{N}_X(x)$  is closed.
- 5.4. Prove the following extension of Theorem 5.7. Let  $g : \mathbb{R}^m \rightarrow \mathbb{R}$  be a continuously differentiable function,  $F_i(x, \xi)$ ,  $i = 1, \dots, m$ , be a random lower semicontinuous functions,  $f_i(x) := \mathbb{E}[F_i(x, \xi)]$ ,  $i = 1, \dots, m$ ,  $f(x) = (f_1(x), \dots, f_m(x))$ ,  $X$  be a nonempty compact subset of  $\mathbb{R}^n$ , and consider the optimization problem

$$\text{Min}_{x \in X} g(f(x)). \quad (5.382)$$

Moreover, let  $\xi^1, \dots, \xi^N$  be an iid random sample,  $\hat{f}_{iN}(x) := N^{-1} \sum_{j=1}^N F_i(x, \xi^j)$ ,  $i = 1, \dots, m$ ,  $\hat{f}_N(x) = (\hat{f}_{1N}(x), \dots, \hat{f}_{mN}(x))$  be the corresponding sample average functions, and

$$\text{Min}_{x \in X} g(\hat{f}_N(x)) \quad (5.383)$$

be the associated SAA problem. Suppose that conditions (A1) and (A2) (used in Theorem 5.7) hold for every function  $F_i(x, \xi)$ ,  $i = 1, \dots, m$ . Let  $\vartheta^*$  and  $\hat{\vartheta}_N$  be the optimal values of problems (5.382) and (5.383), respectively, and  $S$  be the set of optimal solutions of problem (5.382). Show that

$$\hat{\vartheta}_N - \vartheta^* = \inf_{x \in S} \left( \sum_{i=1}^m w_i(x) [\hat{f}_{iN}(x) - f_i(x)] \right) + o_p(N^{-1/2}), \quad (5.384)$$

where

$$w_i(x) := \frac{\partial g(y_1, \dots, y_m)}{\partial y_i} \Big|_{y=f(x)}, \quad i = 1, \dots, m.$$

Moreover, if  $S = \{\bar{x}\}$  is a singleton, then

$$N^{1/2} (\hat{\vartheta}_N - \vartheta^*) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2), \quad (5.385)$$

where  $\bar{w}_i := w_i(\bar{x})$  and

$$\sigma^2 = \text{Var} \left[ \sum_{i=1}^m \bar{w}_i F_i(\bar{x}, \xi) \right]. \quad (5.386)$$

*Hint:* Consider function  $V : C(X) \times \cdots \times C(X) \rightarrow \mathbb{R}$  defined as  $V(\psi_1, \dots, \psi_m) := \inf_{x \in X} g(\psi_1(x), \dots, \psi_m(x))$ , and apply the functional CLT together with the Delta and Danskin theorems.

- 5.5. Consider matrix  $\begin{bmatrix} H & A \\ A^\top & 0 \end{bmatrix}$  defined in (5.44). Assuming that matrix  $H$  is positive definite and matrix  $A$  has full column rank, verify that

$$\begin{bmatrix} H & A \\ A^\top & 0 \end{bmatrix}^{-1} = \begin{bmatrix} H^{-1} - H^{-1}A(A^\top H^{-1}A)^{-1}A^\top H^{-1} & H^{-1}A(A^\top H^{-1}A)^{-1} \\ (A^\top H^{-1}A)^{-1}A^\top H^{-1} & -(A^\top H^{-1}A)^{-1} \end{bmatrix}.$$

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

$$\Delta_N := \sup_{x \in X, y \in Y} |\hat{f}_N(x, y) - f(x, y)|. \quad (5.387)$$

(i) Show that  $|\hat{\vartheta}_N - \vartheta^*| \leq \Delta_N$ , and that if  $\hat{x}_N$  is a  $\delta$ -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 7.65 conclude that, under appropriate regularity conditions, for any  $\varepsilon > 0$  there exist positive constants  $C = C(\varepsilon)$  and  $\beta = \beta(\varepsilon)$  such that

$$\Pr \left\{ |\hat{\vartheta}_N - \vartheta^*| \geq \varepsilon \right\} \leq C e^{-N\beta}. \quad (5.388)$$

(iii) By using bounds (7.216) and (7.217) derive an estimate, similar to (5.116), of the sample size  $N$  which guarantees with probability at least  $1 - \alpha$  that a  $\delta$ -optimal solution  $\hat{x}_N$  of the SAA problem (5.47) is an  $\varepsilon$ -optimal solution of the minimax problem (5.46). Specify required regularity conditions.

- 5.7. Consider the multistage SAA method based on iid conditional sampling. For corresponding sample sizes  $\mathcal{N} = (N_1, \dots, N_{T-1})$  and  $\mathcal{N}' = (N'_1, \dots, N'_{T-1})$ , we say that  $\mathcal{N}' \succeq \mathcal{N}$  if  $N'_t \geq N_t$ ,  $t = 1, \dots, T-1$ . 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  $\mathbb{E}[\hat{\vartheta}_{\mathcal{N}'}] \geq \mathbb{E}[\hat{\vartheta}_{\mathcal{N}}]$ .
- 5.8. Consider the chance constrained problem

$$\min_{x \in X} f(x) \quad \text{s.t.} \quad \Pr\{T(\xi)x + h(\xi) \in C\} \geq 1 - \alpha, \quad (5.389)$$

where  $X \subset \mathbb{R}^n$  is a closed convex set,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is a convex function,  $C \subset \mathbb{R}^m$  is a convex closed set,  $\alpha \in (0, 1)$ , and matrix  $T(\xi)$  and vector  $h(\xi)$  are functions of random vector  $\xi$ . For example, if

$$C := \{z : z = -Wy - w, \quad y \in \mathbb{R}^\ell, \quad w \in \mathbb{R}_+^m\}, \quad (5.390)$$

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 (5.389).

5.9. Consider the following extension of the chance constrained problem (5.196):

$$\min_{x \in X} f(x) \quad \text{s.t.} \quad p_i(x) \leq \alpha_i, \quad i = 1, \dots, p, \quad (5.391)$$

with several (individual) chance constraints. Here  $X \subset \mathbb{R}^n$ ,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $\alpha_i \in (0, 1)$ ,  $i = 1, \dots, p$ , are given significance levels, and

$$p_i(x) = \Pr\{C_i(x, \xi) > 0\}, \quad i = 1, \dots, p,$$

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 (5.391). Use SAA problems based on *independent* samples. (See Remark 6 on page 162 and (5.18) in particular.) That is, estimate  $p_i(x)$  by

$$\hat{p}_{iN_i}(x) := \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{1}_{(0, \infty)}(C_i(x, \xi^{ij})), \quad i = 1, \dots, p.$$

In order to verify feasibility of a point  $\bar{x} \in X$ , show that

$$\Pr\{p_i(\bar{x}) < U_i(\bar{x}), \quad i = 1, \dots, p\} \geq \prod_{i=1}^p (1 - \beta_i),$$

where  $\beta_i \in (0, 1)$  are chosen constants and

$$U_i(\bar{x}) := \sup_{\rho \in [0, 1]} \{\rho : \mathbf{b}(\mathbf{m}_i; \rho, N_i) \geq \beta_i\}, \quad i = 1, \dots, p,$$

with  $\mathbf{m}_i := \hat{p}_{iN_i}(\bar{x})$ .

In order to construct a lower bound, generate  $M$  independent realizations of the corresponding SAA problems, each of the same sample size  $\mathcal{N} = (N_1, \dots, N_p)$  and significance levels  $\gamma_i \in [0, 1]$ ,  $i = 1, \dots, p$ , and compute their optimal values  $\hat{\vartheta}_{\gamma, \mathcal{N}}^1, \dots, \hat{\vartheta}_{\gamma, \mathcal{N}}^M$ . Arrange these values in the increasing order  $\hat{\vartheta}_{\gamma, \mathcal{N}}^{(1)} \leq \dots \leq \hat{\vartheta}_{\gamma, \mathcal{N}}^{(M)}$ . Given significance level  $\beta \in (0, 1)$ , consider the following rule for choice of the corresponding integer  $L$ :

- Choose the largest integer  $L \in \{1, \dots, M\}$  such that

$$\mathbf{b}(L - 1; \theta_{\mathcal{N}}, M) \leq \beta, \quad (5.392)$$

where  $\theta_{\mathcal{N}} := \prod_{i=1}^p \mathbf{b}(r_i; \alpha_i, N_i)$  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  $\vartheta^*$ .

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  $\mathcal{S}_{N_2}$  be the set of optimal solutions of the problem

$$\text{Min}_{x_1 \in \mathcal{X}_1} \left\{ f_1(x_1) + \mathbb{E}[\hat{Q}_{2, N_2}(x_1, \xi_2^i)] \right\}, \quad (5.393)$$

where the expectation is taken with respect to the distribution of the random vector  $(\xi_2^i, \xi_3^{i1}, \dots, \xi_3^{iN_2})$ .

- (i) By using results of section 5.1.1 show that  $\tilde{\vartheta}_{N_1, N_2} \rightarrow \vartheta_{N_2}^*$  w.p. 1 and distance from  $\tilde{x}_{N_1, N_2}$  to  $\mathcal{S}_{N_2}$  tends to 0 w.p. 1 as  $N_1 \rightarrow \infty$ . Specify required regularity conditions.  
(ii) Show that, under appropriate regularity conditions,

$$\tilde{\vartheta}_{N_1, N_2} = \inf_{x_1 \in \mathcal{S}_{N_2}} \tilde{f}_{N_1, N_2}(x_1) + o_p(N_1^{-1/2}). \quad (5.394)$$

Conclude that if, moreover,  $\mathcal{S}_{N_2} = \{\bar{x}_1\}$  is a singleton, then

$$N_1^{1/2}(\tilde{\vartheta}_{N_1, N_2} - \vartheta_{N_2}^*) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(\bar{x}_1)), \quad (5.395)$$

where  $\sigma^2(\bar{x}_1) := \mathbb{V}\text{ar}[\hat{Q}_{2, N_2}(x_1, \xi_2^i)]$ . *Hint:* Use Theorem 5.7.



## Chapter 6

# Risk Averse Optimization

*Andrzej Ruszczyński and Alexander Shapiro*

## 6.1 Introduction

So far, we have discussed stochastic optimization problems, in which the objective function was defined as the expected value  $f(x) := \mathbb{E}[F(x, \omega)]$ . The function  $F : \mathbb{R}^n \times \Omega \rightarrow \mathbb{R}$  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  $X \subset \mathbb{R}^n$ , the stochastic optimization model

$$\min_{x \in X} f(x) \quad (6.1)$$

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 concentrating on 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 invested money. In this section we discuss stochastic optimization from a point of view of risk averse optimization.

A classical approach to risk averse preferences is based on the expected utility theory, which has its roots in mathematical economics (we touched on this subject in section 1.4). In this theory, in order to compare two random outcomes we consider expected values of some scalar transformations  $u : \mathbb{R} \rightarrow \mathbb{R}$  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

$$\text{Min}_{x \in X} \mathbb{E}[u(F(x, \omega))]. \quad (6.2)$$

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

$$u(\mathbb{E}[F(x, \omega)]) \leq \mathbb{E}[u(F(x, \omega))].$$

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:

$$u(\mathbb{E}[F(x, \omega)]) \geq \mathbb{E}[u(F(x, \omega))].$$

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. A 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.

## 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*  $\mathbb{E}[Z]$ , describing the expected outcome, and the *risk (dispersion measure)*  $\mathbb{D}[Z]$ , 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 they maximize 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  $\mathbb{D}[Z] := \text{Var}[Z]$ , which is well defined for  $Z \in \mathcal{L}_2(\Omega, \mathcal{F}, P)$ . 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  $x \in X$ . This can be accomplished by techniques of multiobjective optimization. Most convenient, from

our perspective, is the idea of *scalarization*. For a coefficient  $c \geq 0$ , we form a composite objective functional

$$\rho[Z] := \mathbb{E}[Z] + c\mathbb{D}[Z]. \quad (6.3)$$

The coefficient  $c$  plays the role of the price of risk. We formulate the problem

$$\text{Min}_{x \in X} \mathbb{E}[Z_x] + c\mathbb{D}[Z_x]. \quad (6.4)$$

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  $\mathbb{D}[Z] := \mathbb{V}\text{ar}[Z]$ , 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 the minimization case, we are not concerned if a particular realization of  $Z$  is significantly below its mean; 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 semideviations*. The *upper* semideviation of order  $p$  is defined as

$$\sigma_p^+[Z] := \left( \mathbb{E} \left[ (Z - \mathbb{E}[Z])_+^p \right] \right)^{1/p}, \quad (6.5)$$

where  $p \in [1, \infty)$  is a fixed parameter. It is natural to assume here that considered random variables (uncertain outcomes)  $Z : \Omega \rightarrow \mathbb{R}$  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

$$\text{Min}_{x \in X} \mathbb{E}[Z_x] + c\sigma_p^+[Z_x]. \quad (6.6)$$

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

$$\sigma_p^-[Z] := \left( \mathbb{E} \left[ (\mathbb{E}[Z] - Z)_+^p \right] \right)^{1/p}, \quad (6.7)$$

where  $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ . The resulting mean–risk model has the form

$$\text{Max}_{x \in X} \mathbb{E}[Z_x] - c\sigma_p^-[Z_x]. \quad (6.8)$$

In the special case of  $p = 1$ , both left and right first order semideviations are related to the mean absolute deviation

$$\sigma_1(Z) := \mathbb{E}|Z - \mathbb{E}[Z]|. \quad (6.9)$$

**Proposition 6.1.** *The following identity holds:*

$$\sigma_1^+[Z] = \sigma_1^-[Z] = \frac{1}{2}\sigma_1[Z], \quad \forall Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P). \quad (6.10)$$

**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  $\sigma_1[Z] = \sigma_1^-[Z] + \sigma_1^+[Z]$ .  $\square$

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  $H_Z(z) = \Pr(Z \leq z)$  be the cdf of the random variable  $Z$  and  $\alpha \in (0, 1)$ . Recall that the *left-side  $\alpha$ -quantile* of  $H_Z$  is defined as

$$H_Z^{-1}(\alpha) := \inf\{t : H_Z(t) \geq \alpha\} \quad (6.11)$$

and the *right-side  $\alpha$ -quantile* as

$$\sup\{t : H_Z(t) \leq \alpha\}. \quad (6.12)$$

If  $Z$  represents losses, the (left-side) quantile  $H_Z^{-1}(1 - \alpha)$  is also called *Value-at-Risk* and denoted  $V@R_{\alpha}(Z)$ , i.e.,

$$V@R_{\alpha}(Z) = H_Z^{-1}(1 - \alpha) = \inf\{t : \Pr(Z \leq t) \geq 1 - \alpha\} = \inf\{t : \Pr(Z > t) \leq \alpha\}.$$

Its meaning is the following: *losses larger than  $V@R_{\alpha}(Z)$  occur with probability not exceeding  $\alpha$* . Note that

$$V@R_{\alpha}(Z + \tau) = V@R_{\alpha}(Z) + \tau, \quad \forall \tau \in \mathbb{R}. \quad (6.13)$$

The weighted mean deviation from a quantile is defined as follows:

$$q_{\alpha}[Z] := \mathbb{E}[\max\{(1 - \alpha)(H_Z^{-1}(\alpha) - Z), \alpha(Z - H_Z^{-1}(\alpha))\}]. \quad (6.14)$$

The functional  $q_{\alpha}[Z]$  is well defined and finite valued for all  $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$ . It can be easily shown that

$$q_{\alpha}[Z] = \min_{t \in \mathbb{R}} \{\varphi(t) := \mathbb{E}[\max\{(1 - \alpha)(t - Z), \alpha(Z - t)\}]\}. \quad (6.15)$$

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

$$\Pr[Z < t] \leq \alpha \leq \Pr[Z \leq t].$$

This means that every  $\alpha$ -quantile is a minimizer in (6.15).

The risk functional  $q_\alpha[\cdot]$  can be used in mean–risk models, both in the case of minimization

$$\text{Min}_{x \in X} \mathbb{E}[Z_x] + c q_{1-\alpha}[Z_x] \quad (6.16)$$

and in the case of maximization

$$\text{Max}_{x \in X} \mathbb{E}[Z_x] - c q_\alpha[Z_x]. \quad (6.17)$$

We use  $1 - \alpha$  in the minimization problem and  $\alpha$  in the maximization problem, because in practical applications we are interested in these quantities for small  $\alpha$ .

### 6.2.4 Average Value-at-Risk

The mean-deviation from quantile model is closely related to the concept of Average Value-at-Risk.<sup>39</sup> Suppose that  $Z$  represents losses and we want to satisfy the chance constraint:

$$V@R_\alpha[Z_x] \leq 0. \quad (6.18)$$

Recall that

$$V@R_\alpha[Z] = \inf \{t : \Pr(Z \leq t) \geq 1 - \alpha\},$$

and hence constraint (6.18) is equivalent to the constraint  $\Pr(Z_x \leq 0) \geq 1 - \alpha$ . We have that<sup>40</sup>  $\Pr(Z_x > 0) = \mathbb{E}[\mathbf{1}_{(0,\infty)}(Z_x)]$ , and hence constraint (6.18) can also be written as the expected value constraint:

$$\mathbb{E}[\mathbf{1}_{(0,\infty)}(Z_x)] \leq \alpha. \quad (6.19)$$

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 (6.19).

Let  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  be a nonnegative valued, nondecreasing, convex function such that  $\psi(z) \geq \mathbf{1}_{(0,\infty)}(z)$  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  $\psi(tz) \geq \mathbf{1}_{(0,\infty)}(z)$  and hence the following inequality holds:

$$\inf_{t>0} \mathbb{E}[\psi(tZ)] \geq \mathbb{E}[\mathbf{1}_{(0,\infty)}(Z)].$$

Consequently, the constraint

$$\inf_{t>0} \mathbb{E}[\psi(tZ_x)] \leq \alpha \quad (6.20)$$

is a *conservative* approximation of the chance constraint (6.18) in the sense that the feasible set defined by (6.20) is contained in the feasible set defined by (6.18).

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  $\psi(z) :=$

<sup>39</sup>Average 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.

<sup>40</sup>Recall that  $\mathbf{1}_{(0,\infty)}(z) = 0$  if  $z \leq 0$  and  $\mathbf{1}_{(0,\infty)}(z) = 1$  if  $z > 0$ .

$[1 + \gamma z]_+$  for some  $\gamma > 0$ . Since constraint (6.20) 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 (6.20) is equivalent to

$$\inf_{t>0} \{t\mathbb{E}[t^{-1} + Z]_+ - \alpha\} \leq 0,$$

or equivalently

$$\inf_{t>0} \{\alpha^{-1}\mathbb{E}[Z + t^{-1}]_+ - t^{-1}\} \leq 0.$$

Now replacing  $t$  with  $-t^{-1}$  we get the form

$$\inf_{t<0} \{t + \alpha^{-1}\mathbb{E}[Z - t]_+\} \leq 0. \quad (6.21)$$

The quantity

$$\text{AV@R}_\alpha(Z) := \inf_{t \in \mathbb{R}} \{t + \alpha^{-1}\mathbb{E}[Z - t]_+\} \quad (6.22)$$

is called the *Average Value-at-Risk*<sup>41</sup> of  $Z$  (at level  $\alpha$ ). Note that  $\text{AV@R}_\alpha(Z)$  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

$$t^* := \inf\{z : H_Z(z) \geq 1 - \alpha\} \text{ and } t^{**} := \sup\{z : H_Z(z) \leq 1 - \alpha\} \quad (6.23)$$

are the respective left- and right-side quantiles. Recall that the left-side quantile  $t^* = \text{V@R}_\alpha(Z)$ .

Since the minimum of  $\varphi(t)$  is attained at  $t^* = \text{V@R}_\alpha(Z)$ , we have that  $\text{AV@R}_\alpha(Z)$  is bigger than  $\text{V@R}_\alpha(Z)$  by the nonnegative amount of  $\alpha^{-1}\mathbb{E}[Z - t^*]_+$ . Therefore

$$\inf_{t \in \mathbb{R}} \{t + \alpha^{-1}\mathbb{E}[Z - t]_+\} \leq 0 \text{ implies that } t^* \leq 0,$$

and hence constraint (6.21) is equivalent to  $\text{AV@R}_\alpha(Z) \leq 0$ . Therefore, the constraint

$$\text{AV@R}_\alpha[Z_x] \leq 0 \quad (6.24)$$

is equivalent to the constraint (6.21) and gives a conservative approximation<sup>42</sup> of the chance constraint (6.18).

The function  $\rho(Z) := \text{AV@R}_\alpha(Z)$ , 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

$$\rho(tZ + (1 - t)Z') \leq t\rho(Z) + (1 - t)\rho(Z').$$

<sup>41</sup>In some publications the concept of Average Value-at-Risk is called Conditional Value-at-Risk and is denoted  $\text{CV@R}_\alpha$ .

<sup>42</sup>It is easy to see that for any  $\tau \in \mathbb{R}$ ,

$$\text{AV@R}_\alpha(Z + \tau) = \text{AV@R}_\alpha(Z) + \tau. \quad (6.25)$$

Consequently, the constraint  $\text{AV@R}_\alpha[Z_x] \leq \tau$  gives a conservative approximation of the chance constraint  $\text{V@R}_\alpha[Z_x] \leq \tau$ .

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 monotone, i.e., if  $Z$  and  $Z'$  are two random variables such that with probability one  $Z \geq Z'$ , then  $\rho(Z) \geq \rho(Z')$ . 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

$$\rho[G(tZ + (1-t)Z', \xi)] \leq \rho[tG(Z, \xi) + (1-t)G(Z', \xi)]$$

and hence by convexity of  $\rho(\cdot)$  that

$$\rho[G(tZ + (1-t)Z', \xi)] \leq t\rho[G(Z, \xi)] + (1-t)\rho[G(Z', \xi)].$$

Consequently, (6.24) is a *convex* conservative approximation of the chance constraint (6.18). Moreover, from the considered point of view, (6.24) is the best convex conservative approximation of the chance constraint (6.18).

We can now relate the concept of Average Value-at-Risk to mean deviations from quantiles. Recall that (see (6.14))

$$q_\alpha[Z] := \mathbb{E}\left[\max\left\{(1-\alpha)(H_Z^{-1}(\alpha) - Z), \alpha(Z - H_Z^{-1}(\alpha))\right\}\right].$$

**Theorem 6.2.** *Let  $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$  and  $H(z)$  be its cdf. Then the following identities hold true:*

$$AV@R_\alpha(Z) = \frac{1}{\alpha} \int_{1-\alpha}^1 V@R_{1-\tau}(Z) d\tau = \mathbb{E}[Z] + \frac{1}{\alpha} q_{1-\alpha}[Z]. \quad (6.26)$$

Moreover, if  $H(z)$  is continuous at  $z = V@R_\alpha(Z)$ , then

$$AV@R_\alpha(Z) = \frac{1}{\alpha} \int_{V@R_\alpha(Z)}^{+\infty} z dH(z) = \mathbb{E}[Z | Z \geq V@R_\alpha(Z)]. \quad (6.27)$$

**Proof.** As discussed earlier, the minimum in (6.22) is attained at  $t^* = H^{-1}(1-\alpha) = V@R_\alpha(Z)$ . Therefore

$$AV@R_\alpha(Z) = t^* + \alpha^{-1}\mathbb{E}[Z - t^*]_+ = t^* + \alpha^{-1} \int_{t^*}^{+\infty} (z - t^*) dH(z).$$

Moreover,

$$\int_{t^*}^{+\infty} dH(z) = \Pr(Z \geq t^*) = 1 - \Pr(Z \leq t^*) = \alpha,$$

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 (6.27), and then the second equality in (6.27) follows provided that  $\Pr(Z = t^*) = 0$ .

The first equality in (6.26) follows from the first equality in (6.27) by the substitution  $\tau = H(z)$ . Finally, we have

$$\begin{aligned} AV@R_\alpha(Z) &= t^* + \alpha^{-1}\mathbb{E}[Z - t^*]_+ = \mathbb{E}[Z] + \mathbb{E}\{-Z + t^* + \alpha^{-1}[Z - t^*]_+\} \\ &= \mathbb{E}[Z] + \mathbb{E}\left[\max\left\{\alpha^{-1}(1-\alpha)(Z - t^*), t^* - Z\right\}\right]. \end{aligned}$$

This proves the last equality in (6.26).  $\square$

The first equation in (6.26) motivates the term *Average Value-at-Risk*. The last equation in (6.27) explains the origin of the alternative term *Conditional Value-at-Risk*.

Theorem 6.2 allows us to show an important relation between the absolute semideviation  $\sigma_1^+[Z]$  and the mean deviation from quantile  $q_\alpha[Z]$ .

**Corollary 6.3.** *For every  $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$  we have*

$$\sigma_1^+[Z] = \max_{\alpha \in [0,1]} q_\alpha[Z] = \min_{t \in \mathbb{R}} \max_{\alpha \in [0,1]} \mathbb{E} \{ (1 - \alpha)[t - Z]_+ + \alpha[Z - t]_+ \}. \quad (6.28)$$

**Proof.** From (6.26) we get

$$q_{1-\alpha}[Z] = \int_{1-\alpha}^1 H_Z^{-1}(\tau) d\tau - \alpha \mathbb{E}[Z].$$

The right derivative of the right-hand side with respect to  $\alpha$  equals  $H_Z^{-1}(1 - \alpha) - \mathbb{E}[Z]$ . As it is nonincreasing, the function  $\alpha \mapsto q_{1-\alpha}[Z]$  is concave. Moreover, its maximum is achieved at  $\alpha^*$  for which  $\mathbb{E}[Z]$  is the  $(1 - \alpha^*)$ -quantile of  $Z$ . Substituting the minimizer  $t^* = \mathbb{E}[Z]$  into (6.22) we conclude that

$$\text{AV@R}_{\alpha^*}(Z) = \mathbb{E}[Z] + \frac{1}{\alpha^*} \sigma_1^+[Z].$$

Comparison with (6.26) yields the first equality in (6.28). To prove the second equality we recall relation (6.15) and note that

$$\max \{ (1 - \alpha)(t - Z), \alpha(Z - t) \} = (1 - \alpha)[t - Z]_+ + \alpha[Z - t]_+.$$

Thus

$$\sigma_1^+[Z] = \max_{\alpha \in [0,1]} \min_{t \in \mathbb{R}} \mathbb{E} \{ (1 - \alpha)[t - Z]_+ + \alpha[Z - t]_+ \}.$$

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 (6.28).  $\square$

It also follows from (6.26) that the minimization problem (6.16) can be equivalently written as follows:

$$\begin{aligned} \min_{x \in X} \mathbb{E}[Z_x] + c q_{1-\alpha}[Z_x] &= \min_{x \in X} (1 - c\alpha) \mathbb{E}[Z_x] + c\alpha \text{AV@R}_\alpha[Z_x] \\ &= \min_{x \in X, t \in \mathbb{R}} \mathbb{E} \left[ (1 - c\alpha) Z_x + c(\alpha t + [Z_x - t]_+) \right]. \end{aligned} \quad (6.29)$$

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 (6.29) is convex.

The maximization problem (6.17) can be equivalently written as follows:

$$\max_{x \in X} \mathbb{E}[Z_x] - c q_\alpha[Z_x] = - \min_{x \in X} \mathbb{E}[-Z_x] + c q_{1-\alpha}[-Z_x] \quad (6.30)$$

$$\begin{aligned} &= - \min_{x \in X, t \in \mathbb{R}} \mathbb{E} \left[ -(1 - c\alpha) Z_x + c(\alpha t + [-Z_x - t]_+) \right] \\ &= \max_{x \in X, t \in \mathbb{R}} \mathbb{E} \left[ (1 - c\alpha) Z_x + c(\alpha t - [t - Z_x]_+) \right]. \end{aligned} \quad (6.31)$$



In the last problem we replaced  $t$  by  $-t$  to stress the similarity with (6.29). Again, if  $c \in [0, \alpha^{-1}]$  and the function  $x \mapsto Z_x$  is convex, problem (6.30) is convex.

## 6.3 Coherent Risk Measures

Let  $(\Omega, \mathcal{F})$  be a sample space, equipped with the sigma algebra  $\mathcal{F}$ , on which considered uncertain outcomes (random functions  $Z = Z(\omega)$ ) are defined. By a *risk measure* we understand a function  $\rho(Z)$  which maps  $Z$  into the extended real line  $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ . 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  $Z : \Omega \rightarrow \mathbb{R}$ . However, typically, this space is too large for development of a meaningful theory. 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 of these spaces.) By assuming that  $Z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$ , we assume that random variable  $Z(\omega)$  has a finite  $p$ th order moment with respect to the reference probability measure  $P$ . Also, by considering function  $\rho$  to be defined on the space  $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ , it is implicitly assumed that actually  $\rho$  is defined on classes of functions which can differ on sets of  $P$ -measure zero, i.e.,  $\rho(Z) = \rho(Z')$  if  $P\{\omega : Z(\omega) \neq Z'(\omega)\} = 0$ .

We assume throughout this chapter that risk measures  $\rho : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  are *proper*. That is,  $\rho(Z) > -\infty$  for all  $Z \in \mathcal{Z}$  and the domain

$$\text{dom}(\rho) := \{Z \in \mathcal{Z} : \rho(Z) < +\infty\}$$

is nonempty. We consider the following axioms associated with a risk measure  $\rho$ . For  $Z, Z' \in \mathcal{Z}$  we denote by  $Z \succeq Z'$  the pointwise partial order,<sup>43</sup> meaning  $Z(\omega) \geq Z'(\omega)$  for a.e.  $\omega \in \Omega$ . We also assume that the smaller the realizations of  $Z$ , the better; for example,  $Z$  may represent a random cost.

**(R1)** Convexity:

$$\rho(tZ + (1-t)Z') \leq t\rho(Z) + (1-t)\rho(Z')$$

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  $\rho(Z) \geq \rho(Z')$ .

**(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  $\rho(tZ) = t\rho(Z)$ .

It is said that a risk measure  $\rho$  is *coherent* if it satisfies the above conditions (R1)–(R4). An example of a coherent risk measure is the Average Value-at-Risk  $\rho(Z) := \text{AV@R}_\alpha(Z)$ . (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  $Z \in \mathcal{Z}$ .

<sup>43</sup>This partial order corresponds to the cone  $\mathcal{C} := \mathcal{L}_p^+(\Omega, \mathcal{F}, P)$ . See the discussion of section 7.3, page 404, following (7.245).

If the random outcome represents a reward, i.e., larger realizations of  $Z$  are preferred, we can define a risk measure  $\varrho(Z) = \rho(-Z)$ , where  $\rho$  satisfies axioms (R1)–(R4). In this case, the function  $\varrho$  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), and (R4).

With each space  $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$  is associated its dual space  $\mathcal{Z}^* := \mathcal{L}_q(\Omega, \mathcal{F}, P)$ , where  $q \in (1, +\infty]$  is such that  $1/p + 1/q = 1$ . For  $Z \in \mathcal{Z}$  and  $\zeta \in \mathcal{Z}^*$  their scalar product is defined as

$$\langle \zeta, Z \rangle := \int_{\Omega} \zeta(\omega) Z(\omega) dP(\omega). \quad (6.32)$$

Recall that the conjugate function  $\rho^* : \mathcal{Z}^* \rightarrow \overline{\mathbb{R}}$  of a risk measure  $\rho$  is defined as

$$\rho^*(\zeta) := \sup_{Z \in \mathcal{Z}} \{ \langle \zeta, Z \rangle - \rho(Z) \} \quad (6.33)$$

and the conjugate of  $\rho^*$  (the biconjugate function) as

$$\rho^{**}(Z) := \sup_{\zeta \in \mathcal{Z}^*} \{ \langle \zeta, Z \rangle - \rho^*(\zeta) \}. \quad (6.34)$$

By the Fenchel–Moreau theorem (Theorem 7.71) we have that if  $\rho : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  is convex, proper and lower semicontinuous, then  $\rho^{**} = \rho$ , i.e.,  $\rho(\cdot)$  has the representation

$$\rho(Z) = \sup_{\zeta \in \mathcal{Z}^*} \{ \langle \zeta, Z \rangle - \rho^*(\zeta) \}, \quad \forall Z \in \mathcal{Z}. \quad (6.35)$$

Conversely, if the representation (6.35) holds for some proper function  $\rho^* : \mathcal{Z}^* \rightarrow \overline{\mathbb{R}}$ , then  $\rho$  is convex, proper, and lower semicontinuous. Note that if  $\rho$  is convex, proper, and lower semicontinuous, then its conjugate function  $\rho^*$  is also proper. Clearly, we can write the representation (6.35) in the following equivalent form:

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \{ \langle \zeta, Z \rangle - \rho^*(\zeta) \}, \quad \forall Z \in \mathcal{Z}, \quad (6.36)$$

where  $\mathfrak{A} := \text{dom}(\rho^*)$  is the domain of the conjugate function  $\rho^*$ .

The following basic duality result for convex risk measures is a direct consequence of the Fenchel–Moreau theorem.

**Theorem 6.4.** *Suppose that  $\rho : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  is convex, proper, and lower semicontinuous. Then the representation (6.36) holds with  $\mathfrak{A} := \text{dom}(\rho^*)$ . Moreover, we have that: (i) condition (R2) holds iff every  $\zeta \in \mathfrak{A}$  is nonnegative, i.e.,  $\zeta(\omega) \geq 0$  for a.e.  $\omega \in \Omega$ ; (ii) condition (R3) holds iff  $\int_{\Omega} \zeta dP = 1$  for every  $\zeta \in \mathfrak{A}$ ; and (iii) condition (R4) holds iff  $\rho(\cdot)$  is the support function of the set  $\mathfrak{A}$ , i.e., can be represented in the form*

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \langle \zeta, Z \rangle, \quad \forall Z \in \mathcal{Z}. \quad (6.37)$$

**Proof.** If  $\rho : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  is convex, proper, and lower semicontinuous, then representation (6.36) is valid by virtue of the Fenchel–Moreau theorem (Theorem 7.71).

Now suppose that assumption (R2) holds. It follows that  $\rho^*(\zeta) = +\infty$  for every  $\zeta \in \mathcal{Z}^*$  which is not nonnegative. Indeed, if  $\zeta \in \mathcal{Z}^*$  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  $\bar{Z} := \mathbf{1}_\Delta$  we have that  $\langle \zeta, \bar{Z} \rangle < 0$ . Take any  $Z$  in the domain of  $\rho$ , i.e., such that  $\rho(Z)$  is finite, and consider  $Z_t := Z - t\bar{Z}$ . Then for  $t \geq 0$ , we have that  $Z \succeq Z_t$ , and assumption (R2) implies that  $\rho(Z) \geq \rho(Z_t)$ . Consequently,

$$\rho^*(\zeta) \geq \sup_{t \in \mathbb{R}_+} \{ \langle \zeta, Z_t \rangle - \rho(Z_t) \} \geq \sup_{t \in \mathbb{R}_+} \{ \langle \zeta, Z \rangle - t \langle \zeta, \bar{Z} \rangle - \rho(Z) \} = +\infty.$$

Conversely, suppose that every  $\zeta \in \mathcal{A}$  is nonnegative. Then for every  $\zeta \in \mathcal{A}$  and  $Z \succeq Z'$ , we have that  $\langle \zeta, Z' \rangle \geq \langle \zeta, Z \rangle$ . By (6.36), this implies that if  $Z \succeq Z'$ , then  $\rho(Z) \geq \rho(Z')$ . This completes the proof of assertion (i).

Suppose that assumption (R3) holds. Then for every  $Z \in \text{dom}(\rho)$  we have

$$\rho^*(\zeta) \geq \sup_{a \in \mathbb{R}} \{ \langle \zeta, Z + a \rangle - \rho(Z + a) \} = \sup_{a \in \mathbb{R}} \left\{ a \int_{\Omega} \zeta dP - a + \langle \zeta, Z \rangle - \rho(Z) \right\}.$$

It follows that  $\rho^*(\zeta) = +\infty$  for any  $\zeta \in \mathcal{Z}^*$  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 (6.36). This completes the proof of (ii).

Clearly, if (6.37) holds, then  $\rho$  is positively homogeneous. Conversely, if  $\rho$  is positively homogeneous, then its conjugate function is the indicator function of a convex subset of  $\mathcal{Z}^*$ . Consequently, the representation (6.37) follows by (6.36).  $\square$

It follows from the above theorem that if  $\rho$  is a risk measure satisfying conditions (R1)–(R3) and is proper and lower semicontinuous, then the representation (6.36) holds with  $\mathcal{A}$  being a subset of the set of probability density functions,

$$\mathfrak{P} := \left\{ \zeta \in \mathcal{Z}^* : \int_{\Omega} \zeta(\omega) dP(\omega) = 1, \zeta \succeq 0 \right\}. \quad (6.38)$$

If, moreover,  $\rho$  is positively homogeneous (i.e., condition (R4) holds), then its conjugate  $\rho^*$  is the indicator function of a convex set  $\mathcal{A} \subset \mathcal{Z}^*$ , and  $\mathcal{A}$  is equal to the subdifferential  $\partial\rho(0)$  of  $\rho$  at  $0 \in \mathcal{Z}$ . Furthermore,  $\rho(0) = 0$  and hence by the definition of  $\partial\rho(0)$  we have that

$$\mathcal{A} = \{ \zeta \in \mathfrak{P} : \langle \zeta, Z \rangle \leq \rho(Z), \quad \forall Z \in \mathcal{Z} \}. \quad (6.39)$$

The set  $\mathcal{A}$  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  $\rho$  is positively homogeneous and continuous, then  $\mathcal{A} = \partial\rho(0)$  is a *bounded* (and weakly\* compact) subset of  $\mathcal{Z}^*$  (see Proposition 7.74).

We have that if  $\rho$  is a *coherent* risk measure, then the corresponding set  $\mathcal{A}$  is a set of probability density functions. Consequently, for any  $\zeta \in \mathcal{A}$  we can view  $\langle \zeta, Z \rangle$  as the expectation  $\mathbb{E}_{\zeta}[Z]$  taken with respect to the probability measure  $\zeta dP$ , defined by the density  $\zeta$ . Consequently representation (6.37) can be written in the form

$$\rho(Z) = \sup_{\zeta \in \mathcal{A}} \mathbb{E}_{\zeta}[Z], \quad \forall Z \in \mathcal{Z}. \quad (6.40)$$

Definition of a risk measure  $\rho$  depends on a particular choice of the corresponding space  $\mathcal{Z}$ . In many cases there is a natural choice of  $\mathcal{Z}$  which ensures that  $\rho(Z)$  is finite valued for all  $Z \in \mathcal{Z}$ . We shall see such examples in section 6.3.2. By Theorem 7.79 we have the following result, which shows that for real valued convex and monotone risk measures, the assumption of lower semicontinuity in Theorem 6.4 holds automatically.

**Proposition 6.5.** *Let  $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$  with  $p \in [1, +\infty]$  and  $\rho : \mathcal{Z} \rightarrow \mathbb{R}$  be a real valued risk measure satisfying conditions (R1) and (R2). Then  $\rho$  is continuous and subdifferentiable on  $\mathcal{Z}$ .*

Theorem 6.4 together with Proposition 6.5 imply the following basic duality result.

**Theorem 6.6.** *Let  $\rho : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$ , where  $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$  with  $p \in [1, +\infty)$ . Then  $\rho$  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 (6.37) holds.*

**Proof.** If  $\rho : \mathcal{Z} \rightarrow \mathbb{R}$  is a real valued coherent risk measure, then by Proposition 6.5 it is continuous, and hence by Theorem 6.4 the representation (6.37) 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 (6.37) holds with the set  $\mathfrak{A}$  being a convex subset of  $\mathfrak{P}$  and weakly\* compact, then  $\rho$  is real valued and satisfies conditions (R1)–(R4).  $\square$

The following result shows that if a risk measure satisfies conditions (R1)–(R3), then either it is finite valued and continuous on  $\mathcal{Z}$  or it takes value  $+\infty$  on a dense subset of  $\mathcal{Z}$ .

**Proposition 6.7.** *Let  $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ , with  $p \in [1, +\infty)$ , and  $\rho : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  be a proper risk measure satisfying conditions (R1), (R2) and (R3). Suppose that the domain of  $\rho$  has a nonempty interior. Then  $\rho$  is finite valued and continuous on  $\mathcal{Z}$ .*

**Proof.** Consider the level sets of  $\rho$ :

$$\mathcal{A}_c := \{Z \in \mathcal{Z} : \rho(Z) \leq c\}.$$

We have that  $\bigcup_{c \in \mathbb{R}} \mathcal{A}_c = \text{dom}(\rho)$ . Since  $\text{dom}(\rho)$  has a nonempty interior, it follows by Baire's lemma that for some  $c \in \mathbb{R}$  the set  $\mathcal{A}_c$  has a nonempty interior. Because of condition (R3) we have that  $Z \in \mathcal{A}_0$  iff  $Z + c \in \mathcal{A}_c$ , i.e.,  $\mathcal{A}_c = \mathcal{A}_0 + c$  (here  $c$  denotes the constant function  $Z(\cdot) = c$ ). Therefore  $\mathcal{A}_0$  has a nonempty interior. That is, there exist  $Z_0 \in \mathcal{Z}$  and  $r > 0$  such that  $B(Z_0, r) \subset \mathcal{A}_0$ , where

$$B(Z_0, r) := \{Z \in \mathcal{Z} : \|Z - Z_0\| \leq r\}.$$

By changing variables  $Z \mapsto Z - Z_0$ , we can assume without loss of generality that  $Z_0 = 0$ , i.e.,  $B(0, r) \subset \mathcal{A}_0$ .

Consider a point  $Z \in \mathcal{Z}$ . For  $c \in \mathbb{R}$  we have that  $Z = Z_c^- + Z_c^+$ , where  $Z_c^-(\cdot) := \min\{Z(\cdot), c\}$  and  $Z_c^+(\cdot) := [Z(\cdot) - c]_+$ . Note that for  $c$  large enough, the norm of  $Z_c^+$  can be made arbitrarily small. Therefore we can choose  $c$  such that  $\|Z_c^+\| < r$ . Since  $\mathcal{A}_c = \mathcal{A}_0 + c$ , we have that  $B(c, r) \subset \mathcal{A}_c$ . Consequently,  $c + Z_c^+ \in \mathcal{A}_c$ . It follows by the monotonicity

condition (R2) that  $\rho(Z) \leq \rho(c + Z_c^+) \leq c$ , and hence  $\rho(Z)$  is finite. That is, we showed that  $\rho(\cdot)$  is finite valued on  $\mathcal{Z}$ . Continuity of  $\rho(\cdot)$  follows by Proposition 6.5.  $\square$

It is not difficult to show (we leave this as an exercise) that for  $\mathcal{Z} := \mathcal{L}_\infty(\Omega, \mathcal{F}, P)$  any risk measure  $\rho : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  satisfying conditions (R1)–(R3), and having a finite value in at least one point of  $\mathcal{Z}$ , is finite valued and hence is continuous by Proposition 6.5.

Of course, the analysis simplifies considerably if the space  $\Omega$  is finite, say,  $\Omega := \{\omega_1, \dots, \omega_K\}$  equipped with sigma algebra of all subsets of  $\Omega$  and respective (positive) probabilities  $p_1, \dots, p_K$ . Then every function  $Z : \Omega \rightarrow \mathbb{R}$  is measurable and the space  $\mathcal{Z}$  of all such functions can be identified with  $\mathbb{R}^K$  by identifying  $Z \in \mathcal{Z}$  with the vector  $(Z(\omega_1), \dots, Z(\omega_K)) \in \mathbb{R}^K$ . The dual of the space  $\mathbb{R}^K$  can be identified with itself by using the standard scalar product in  $\mathbb{R}^K$ , and the set  $\mathfrak{P}$  becomes

$$\mathfrak{P} = \left\{ \zeta \in \mathbb{R}^K : \sum_{k=1}^K p_k \zeta_k = 1, \zeta \geq 0 \right\}. \quad (6.41)$$

The above set  $\mathfrak{P}$  forms a convex bounded subset of  $\mathbb{R}^K$ , and hence the set  $\mathfrak{A}$  is also bounded.

### 6.3.1 Differentiability Properties of Risk Measures

Let  $\rho : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  be a convex proper lower semicontinuous risk measure. By convexity and lower semicontinuity of  $\rho$  we have that  $\rho^{**} = \rho$  and hence by Proposition 7.73 that

$$\partial\rho(Z) = \arg \max_{\zeta \in \mathfrak{A}} \{ \langle \zeta, Z \rangle - \rho^*(\zeta) \}, \quad (6.42)$$

provided that  $\rho(Z)$  is finite. If, moreover,  $\rho$  is positively homogeneous, then  $\mathfrak{A} = \partial\rho(0)$  and

$$\partial\rho(Z) = \arg \max_{\zeta \in \mathfrak{A}} \langle \zeta, Z \rangle. \quad (6.43)$$

As we know, 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 also have that if  $\rho$  is finite valued and continuous at  $Z$ , then  $\partial\rho(Z)$  is a nonempty bounded and weakly\* compact subset of  $\mathcal{Z}^*$ ,  $\rho$  is Hadamard directionally differentiable and subdifferentiable at  $Z$ , and

$$\rho'(Z, H) = \sup_{\zeta \in \partial\rho(Z)} \langle \zeta, H \rangle, \quad \forall H \in \mathcal{Z}. \quad (6.44)$$

In particular, if  $\rho$  is continuous at  $Z$  and  $\partial\rho(Z)$  is a singleton, i.e.,  $\partial\rho(Z)$  consists of unique point denoted  $\nabla\rho(Z)$ , then  $\rho$  is Hadamard differentiable at  $Z$  and

$$\rho'(Z, \cdot) = \langle \nabla\rho(Z), \cdot \rangle. \quad (6.45)$$

We often have to deal with composite functions  $\rho \circ F : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ , where  $F : \mathbb{R}^n \rightarrow \mathcal{Z}$  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  $(\Omega, \mathcal{F})$ . We say that the mapping  $F$  is *convex* if the function  $f(\cdot, \omega)$  is convex for every  $\omega \in \Omega$ .

**Proposition 6.8.** *If the mapping  $F : \mathbb{R}^n \rightarrow \mathcal{Z}$  is convex and  $\rho : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  satisfies conditions (R1)–(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

$$\rho(F(tx + (1-t)x')) \leq \rho(tF(x) + (1-t)F(x')).$$

Hence convexity of  $\rho(\cdot)$  implies that

$$\rho(F(tx + (1-t)x')) \leq t\rho(F(x)) + (1-t)\rho(F(x')).$$

This proves convexity of  $\rho(F(\cdot))$ .  $\square$

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  $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ . The mapping  $F : \mathbb{R}^n \rightarrow \mathcal{Z}$  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  $\Omega$ , 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 6.9.** *Let  $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$  and  $F : \mathbb{R}^n \rightarrow \mathcal{Z}$  be a convex mapping. Then  $F$  is continuous and directionally differentiable, and*

$$[F'(\bar{x}, h)](\omega) = f'_\omega(\bar{x}, h), \quad \omega \in \Omega, \quad h \in \mathbb{R}^n. \quad (6.46)$$

**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  $x \rightarrow \bar{x}$ . By the Lebesgue dominated convergence theorem and continuity of  $f(\cdot, \omega)$  we can write that

$$\lim_{x \rightarrow \bar{x}} \int_{\Omega} |f(x, \omega) - f(\bar{x}, \omega)|^p dP(\omega) = \int_{\Omega} \lim_{x \rightarrow \bar{x}} |f(x, \omega) - f(\bar{x}, \omega)|^p dP(\omega) = 0, \quad (6.47)$$

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  $x_1, \dots, x_{n+1} \in \mathbb{R}^n$  be such points that the set  $U := \text{conv}\{x_1, \dots, x_{n+1}\}$  forms a neighborhood of the point  $\bar{x}$ , and let  $g(\omega) := \max\{f(x_1, \omega), \dots, f(x_{n+1}, \omega)\}$ . By convexity of  $f(\cdot, \omega)$  we have that  $f(x, \cdot) \leq g(\cdot)$  for all  $x \in U$ . Also since every  $f(x_i, \cdot)$ ,  $i = 1, \dots, n+1$ , 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

$$f(x, \omega) \geq 2f(\bar{x}, \omega) - f(2\bar{x} - x, \omega).$$

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  $\tilde{g}(\omega) := 2f(\bar{x}, \omega) - g(\omega)$  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 (6.47).

Now for  $h \in \mathbb{R}^n$  and  $t > 0$  denote

$$R_t(\omega) := t^{-1} [f(\bar{x} + th, \omega) - f(\bar{x}, \omega)] \quad \text{and} \quad Z(\omega) := f'_\omega(\bar{x}, h), \quad \omega \in \Omega.$$

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  $R_t(\cdot) \geq Z(\cdot)$  for any  $t > 0$ . Again by convexity of  $f(\cdot, \omega)$ , we have that for  $t > 0$ ,

$$Z(\cdot) \geq t^{-1} [f(\bar{x}, \cdot) - f(\bar{x} - th, \cdot)].$$

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 (6.46) follows.  $\square$

The following theorem can be viewed as an extension of Theorem 7.46, where a similar result is derived for  $\rho(\cdot) := \mathbb{E}[\cdot]$ .

**Theorem 6.10.** *Let  $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$  and  $F : \mathbb{R}^n \rightarrow \mathcal{Z}$  be a convex mapping. Suppose that  $\rho$  is convex, finite valued, and continuous at  $\bar{Z} := F(\bar{x})$ . Then the composite function  $\phi = \rho \circ F$  is directionally differentiable at  $\bar{x}$ ,  $\phi'(\bar{x}, h)$  is finite valued for every  $h \in \mathbb{R}^n$ , and*

$$\phi'(\bar{x}, h) = \sup_{\zeta \in \partial \rho(\bar{Z})} \int_{\Omega} f'_\omega(\bar{x}, h) \zeta(\omega) dP(\omega). \quad (6.48)$$

**Proof.** Since  $\rho$  is continuous at  $\bar{Z}$ , it follows that  $\rho$  is subdifferentiable and Hadamard directionally differentiable at  $\bar{Z}$  and formula (6.44) holds. Also by Lemma 6.9, mapping  $F$  is directionally differentiable. Consequently, we can apply the chain rule (see Proposition 7.58) to conclude that  $\phi(\cdot)$  is directionally differentiable at  $\bar{x}$ ,  $\phi'(\bar{x}, h)$  is finite valued and

$$\phi'(\bar{x}, h) = \rho'(\bar{Z}, F'(\bar{x}, h)). \quad (6.49)$$

Together with (6.44) and (6.46), the above formula (6.49) implies (6.48).  $\square$

It is also possible to write formula (6.48) in terms of the corresponding subdifferentials.

**Theorem 6.11.** Let  $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$  and  $F : \mathbb{R}^n \rightarrow \mathcal{Z}$  be a convex mapping. Suppose that  $\rho$  satisfies conditions (R1) and (R2) and is finite valued and continuous at  $\bar{Z} := F(\bar{x})$ . Then the composite function  $\phi = \rho \circ F$  is subdifferentiable at  $\bar{x}$  and

$$\partial\phi(\bar{x}) = \text{cl} \left( \bigcup_{\zeta \in \partial\rho(\bar{Z})} \int_{\Omega} \partial f_{\omega}(\bar{x}) \zeta(\omega) dP(\omega) \right). \quad (6.50)$$

**Proof.** Since, by Lemma 6.9,  $F$  is continuous at  $\bar{x}$  and  $\rho$  is continuous at  $F(\bar{x})$ , we have that  $\phi$  is continuous at  $\bar{x}$ , and hence  $\phi(x)$  is finite valued for all  $x$  in a neighborhood of  $\bar{x}$ . Moreover, by Proposition 6.8,  $\phi(\cdot)$  is convex and hence is continuous in a neighborhood of  $\bar{x}$  and is subdifferentiable at  $\bar{x}$ . Also, formula (6.48) holds. It follows that  $\phi'(\bar{x}, \cdot)$  is convex, continuous, and positively homogeneous, and

$$\phi'(\bar{x}, \cdot) = \sup_{\zeta \in \partial\rho(\bar{Z})} \eta_{\zeta}(\cdot), \quad (6.51)$$

where

$$\eta_{\zeta}(\cdot) := \int_{\Omega} f'_{\omega}(\bar{x}, \cdot) \zeta(\omega) dP(\omega). \quad (6.52)$$

Because of condition (R2), we have that every  $\zeta \in \partial\rho(\bar{Z})$  is nonnegative. Consequently, the corresponding function  $\eta_{\zeta}$  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 (6.51), is the support function of the set  $\bigcup_{\zeta \in \partial\rho(\bar{Z})} \partial\eta_{\zeta}(0)$ . Applying Theorem 7.47 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

$$\partial\eta_{\zeta}(0) = \int_{\Omega} \partial f_{\omega}(\bar{x}) \zeta(\omega) dP(\omega). \quad (6.53)$$

Since  $\partial\rho(\bar{Z})$  is convex, it is straightforward to verify that the set  $\bigcup_{\zeta \in \partial\rho(\bar{Z})} \partial\eta_{\zeta}(0)$  is also convex. Consequently it follows by (6.51) and (6.53) that the subdifferential of  $\phi'(\bar{x}, \cdot)$  at  $0 \in \mathbb{R}^n$  is equal to the topological closure of the set  $\bigcup_{\zeta \in \partial\rho(\bar{Z})} \partial\eta_{\zeta}(0)$ , i.e., is given by the right-hand side of (6.50). It remains to note that the subdifferential of  $\phi'(\bar{x}, \cdot)$  at  $0 \in \mathbb{R}^n$  coincides with  $\partial\phi(\bar{x})$ .  $\square$

Under the assumptions of the above theorem, we have that the composite function  $\phi$  is convex and is continuous (in fact, even Lipschitz continuous) in a neighborhood of  $\bar{x}$ . It follows that  $\phi$  is differentiable<sup>44</sup> at  $\bar{x}$  iff  $\partial\phi(\bar{x})$  is a singleton. This leads to the following result, where for  $\zeta \geq 0$  we say that a property holds for  $\zeta$ -a.e.  $\omega \in \Omega$  if the set of points  $A \in \mathcal{F}$  where it does not hold has  $\zeta dP$  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  $\zeta$ -a.e.  $\omega \in \Omega$ .

<sup>44</sup>Note that since  $\phi(\cdot)$  is Lipschitz continuous near  $\bar{x}$ , the notions of Gâteaux and Fréchet differentiability at  $\bar{x}$  are equivalent here.



**Corollary 6.12.** *Let  $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$  and  $F : \mathbb{R}^n \rightarrow \mathcal{Z}$  be a convex mapping. Suppose that  $\rho$  satisfies conditions (R1) and (R2) and is finite valued and continuous at  $\bar{Z} := F(\bar{x})$ . Then the composite function  $\phi = \rho \circ F$  is differentiable at  $\bar{x}$  iff the following two properties hold: (i) for every  $\zeta \in \partial\rho(\bar{Z})$  the function  $f(\cdot, \omega)$  is differentiable at  $\bar{x}$  for  $\zeta$ -a.e.  $\omega \in \Omega$ , and (ii)  $\int_{\Omega} \nabla f_{\omega}(\bar{x}) \zeta(\omega) dP(\omega)$  is the same for every  $\zeta \in \partial\rho(\bar{Z})$ .*

*In particular, if  $\partial\rho(\bar{Z}) = \{\bar{\zeta}\}$  is a singleton, then  $\phi$  is differentiable at  $\bar{x}$  iff  $f(\cdot, \omega)$  is differentiable at  $\bar{x}$  for  $\bar{\zeta}$ -a.e.  $\omega \in \Omega$ , in which case*

$$\nabla\phi(\bar{x}) = \int_{\Omega} \nabla f_{\omega}(\bar{x}) \bar{\zeta}(\omega) dP(\omega). \quad (6.54)$$

**Proof.** By Theorem 6.11 we have that  $\phi$  is differentiable at  $\bar{x}$  iff the set on the right-hand side of (6.50) 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  $\phi$  is differentiable at  $\bar{x}$  iff conditions (i) and (ii) hold. The second assertion then follows.  $\square$

Of course, if the set inside the parentheses on the right-hand side of (6.50) is closed, then there is no need to take its topological closure. This holds true in the following case.

**Corollary 6.13.** *Suppose that the assumptions of Theorem 6.11 are satisfied and for every  $\zeta \in \partial\rho(\bar{Z})$  the function  $f_{\omega}(\cdot)$  is differentiable at  $\bar{x}$  for  $\zeta$ -a.e.  $\omega \in \Omega$ . Then*

$$\partial\phi(\bar{x}) = \bigcup_{\zeta \in \partial\rho(\bar{Z})} \int_{\Omega} \nabla f_{\omega}(\bar{x}) \zeta(\omega) dP(\omega). \quad (6.55)$$

**Proof.** In view of Theorem 6.11 we only need to show that the set on the right-hand side of (6.55) is closed. As  $\rho$  is continuous at  $\bar{Z}$ , the set  $\partial\rho(\bar{Z})$  is weakly\* compact. Also, the mapping  $\zeta \mapsto \int_{\Omega} \nabla f_{\omega}(\bar{x}) \zeta(\omega) dP(\omega)$ , 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  $\partial\rho(\bar{Z})$  by this mapping is compact and hence is closed, i.e., the set at the right-hand side of (6.55) is closed.  $\square$

### 6.3.2 Examples of Risk Measures

In this section we discuss several examples of risk measures which are commonly used in applications. In each of 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  $1 \leq p' \leq p$  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  $p' < p$ . Note, however, that this embedding is not continuous. 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  $\rho : \mathcal{Z} \rightarrow \mathbb{R}$  and define

$$\tilde{\rho}(Z) := \mathbb{E}[Z] + \inf_{t \in \mathbb{R}} \rho(Z - t). \quad (6.56)$$

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) irrespective of whether  $\rho$  does. It is not difficult to see that if  $\rho$  satisfies conditions (R1) and (R2), then  $\tilde{\rho}$  satisfies these conditions as well. Also, if  $\rho$  is positively homogeneous, then so is  $\tilde{\rho}$ . Let us calculate the conjugate of  $\tilde{\rho}$ . We have

$$\begin{aligned} \tilde{\rho}^*(\zeta) &= \sup_{Z \in \mathcal{Z}} \{ \langle \zeta, Z \rangle - \tilde{\rho}(Z) \} = \sup_{Z \in \mathcal{Z}} \left\{ \langle \zeta, Z \rangle - \mathbb{E}[Z] - \inf_{t \in \mathbb{R}} \rho(Z - t) \right\} \\ &= \sup_{Z \in \mathcal{Z}, t \in \mathbb{R}} \{ \langle \zeta, Z \rangle - \mathbb{E}[Z] - \rho(Z - t) \} \\ &= \sup_{Z \in \mathcal{Z}, t \in \mathbb{R}} \{ \langle \zeta - 1, Z \rangle + t(\mathbb{E}[\zeta] - 1) - \rho(Z) \}. \end{aligned}$$

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  $\rho : \mathcal{Z} \rightarrow \mathbb{R}$ . Define

$$\check{\rho}(Z) := \inf_{\tau > 0} \tau \rho(\tau^{-1} Z). \quad (6.57)$$

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  $\rho$  is or isn't. Clearly, if  $\rho$  is positively homogeneous to start with, then  $\rho = \check{\rho}$ .

If  $\rho$  is convex, then so is  $\check{\rho}$ . Indeed, observe that if  $\rho$  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  $Z_1, Z_2 \in \mathcal{Z}$ , and setting  $\tau := t\tau_1 + (1-t)\tau_2$  and  $Z := tZ_1 + (1-t)Z_2$ , we have

$$\begin{aligned} t[\tau_1 \rho(\tau_1^{-1} Z_1)] + (1-t)[\tau_2 \rho(\tau_2^{-1} Z_2)] &= \tau \left[ \frac{t\tau_1}{\tau} \rho(\tau_1^{-1} Z_1) + \frac{(1-t)\tau_2}{\tau} \rho(\tau_2^{-1} Z_2) \right] \\ &\geq \tau \rho \left( \frac{t}{\tau} Z_1 + \frac{(1-t)}{\tau} Z_2 \right) = \tau \rho(\tau^{-1} Z). \end{aligned}$$

Minimizing convex function  $\varphi(\tau, Z)$  over  $\tau > 0$ , we obtain a convex function. It is also not difficult to see that if  $\rho$  satisfies conditions (R2) and (R3), then so does  $\check{\rho}$ .

Let us calculate the conjugate of  $\check{\rho}$ . We have

$$\begin{aligned} \check{\rho}^*(\zeta) &= \sup_{Z \in \mathcal{Z}} \{ \langle \zeta, Z \rangle - \check{\rho}(Z) \} = \sup_{Z \in \mathcal{Z}, \tau > 0} \{ \langle \zeta, Z \rangle - \tau \rho(\tau^{-1} Z) \} \\ &= \sup_{Z \in \mathcal{Z}, \tau > 0} \{ \tau [\langle \zeta, Z \rangle - \rho(Z)] \}. \end{aligned}$$

It follows that  $\check{\rho}^*$  is the indicator function of the set

$$\mathfrak{A} := \{ \zeta \in \mathcal{Z}^* : \langle \zeta, Z \rangle \leq \rho(Z), \quad \forall Z \in \mathcal{Z} \}. \quad (6.58)$$

If, moreover,  $\check{\rho}$  is lower semicontinuous and 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 6.14 (Utility Model).** It is possible to relate the theory of convex risk measures with the utility model. Let  $g : \mathbb{R} \rightarrow \overline{\mathbb{R}}$  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* function.<sup>45</sup>

**Proposition 6.15.** *Let  $g : \mathbb{R} \rightarrow \overline{\mathbb{R}}$  be a proper convex nondecreasing lower semicontinuous function. Suppose that the risk measure*

$$\rho(Z) := \mathbb{E}[g(Z)] \quad (6.59)$$

*is well defined and proper. Then  $\rho$  is convex and lower semicontinuous and satisfies the monotonicity condition (R2), and the representation (6.35) holds with*

$$\rho^*(\zeta) = \mathbb{E}[g^*(\zeta)]. \quad (6.60)$$

*Moreover, if  $\rho(Z)$  is finite, then*

$$\partial\rho(Z) = \{\zeta \in \mathcal{Z}^* : \zeta(\omega) \in \partial g(Z(\omega)) \text{ a.e. } \omega \in \Omega\}. \quad (6.61)$$

**Proof.** Since  $g$  is lower semicontinuous and convex, we have by the Fenchel–Moreau theorem that

$$g(z) = \sup_{\alpha \in \mathbb{R}} \{\alpha z - g^*(\alpha)\},$$

where  $g^*$  is the conjugate of  $g$ . As  $g$  is proper, the conjugate function  $g^*$  is also proper. It follows that

$$\rho(Z) = \mathbb{E} \left[ \sup_{\alpha \in \mathbb{R}} \{\alpha Z - g^*(\alpha)\} \right]. \quad (6.62)$$

By the interchangeability principle (Theorem 7.80) for the space  $\mathfrak{M} := \mathcal{Z}^* = \mathcal{L}_q(\Omega, \mathcal{F}, P)$ , which is decomposable, we obtain

$$\rho(Z) = \sup_{\zeta \in \mathcal{Z}^*} \{\langle \zeta, Z \rangle - \mathbb{E}[g^*(\zeta)]\}. \quad (6.63)$$

It follows that  $\rho$  is convex and lower semicontinuous, and representation (6.35) holds with the conjugate function given in (6.60). Moreover, since the function  $g$  is nondecreasing, it follows that  $\rho$  satisfies the monotonicity condition (R2).

Since  $\rho$  is convex proper and lower semicontinuous, and hence  $\rho^{**} = \rho$ , we have by Proposition 7.73 that

$$\partial\rho(Z) = \arg \max_{\zeta \in \mathfrak{A}} \{\mathbb{E}[\zeta Z - g^*(\zeta)]\}, \quad (6.64)$$

assuming that  $\rho(Z)$  is finite. Together with formula (7.247) of the interchangeability principle (Theorem 7.80), this implies (6.61).  $\square$

<sup>45</sup>We consider here minimization problems, and that is why we speak about disutility. Any disutility function  $g$  corresponds to a utility function  $u : \mathbb{R} \rightarrow \mathbb{R}$  defined by  $u(z) = -g(-z)$ . Note that the function  $u$  is concave and nondecreasing since the function  $g$  is convex and nondecreasing.

The above risk measure  $\rho$ , defined in (6.59), does not satisfy condition (R3) unless  $g(z) \equiv z$ . We can consider the corresponding risk measure  $\tilde{\rho}$ , defined in (6.56), which in the present case can be written as

$$\tilde{\rho}(Z) = \inf_{t \in \mathbb{R}} \mathbb{E}[Z + g(Z - t)]. \quad (6.65)$$

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

$$\tilde{\rho}(Z) = \sup_{\substack{\zeta \in \mathcal{Z}^* \\ \mathbb{E}[\zeta] = 1}} \{ \langle \zeta - 1, Z \rangle - \mathbb{E}[g^*(\zeta)] \} \quad (6.66)$$

holds. ■

**Example 6.16 (Average Value-at-Risk).** The risk measure  $\rho$  associated with disutility function  $g$ , defined in (6.59), is positively homogeneous only if  $g$  is positively homogeneous. Suppose now that  $g(z) := \max\{az, bz\}$ , where  $b \geq a$ . 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 6.15 that the representation (6.37) 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 (6.37) since, of course, the condition that “ $\zeta(\omega) \in [a, b]$  for a.e.  $\omega \in \Omega$ ” implies that  $\zeta$  is essentially bounded.

Consider now the risk measure

$$\tilde{\rho}(Z) := \mathbb{E}[Z] + \inf_{t \in \mathbb{R}} \mathbb{E} \{ \beta_1[t - Z]_+ + \beta_2[Z - t]_+ \}, \quad Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P), \quad (6.67)$$

where  $\beta_1 \in [0, 1]$  and  $\beta_2 \geq 0$ . This risk measure can be recognized as risk measure defined in (6.65), associated with function  $g(z) := \beta_1[-z]_+ + \beta_2[z]_+$ . For specified  $\beta_1$  and  $\beta_2$ , 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

$$\tilde{\rho}(Z) = (1 - \beta_1)\mathbb{E}[Z] + \beta_1 \text{AV@R}_\alpha(Z), \quad (6.68)$$

where  $\alpha := \beta_1/(\beta_1 + \beta_2)$ . Note that the right-hand side of (6.67) attains its minimum at  $t^* = \text{V@R}_\alpha(Z)$ . Therefore, the second term on the right-hand side of (6.67) is the weighted measure of deviation from the quantile  $\text{V@R}_\alpha(Z)$ , discussed in section 6.2.3.

The respective conjugate function is the indicator function of the set  $\mathfrak{A} := \text{dom}(\tilde{\rho}^*)$ , and  $\tilde{\rho}$  can be represented in the dual form (6.37) with

$$\mathfrak{A} = \{ \zeta \in \mathcal{L}_\infty(\Omega, \mathcal{F}, P) : \zeta(\omega) \in [1 - \beta_1, 1 + \beta_2] \text{ a.e. } \omega \in \Omega, \mathbb{E}[\zeta] = 1 \}. \quad (6.69)$$

In particular, for  $\beta_1 = 1$  we have that  $\tilde{\rho}(\cdot) = \text{AV@R}_\alpha(\cdot)$ , and hence the dual representation (6.37) of  $\text{AV@R}_\alpha$  holds with the set

$$\mathfrak{A} = \{ \zeta \in \mathcal{L}_\infty(\Omega, \mathcal{F}, P) : \zeta(\omega) \in [0, \alpha^{-1}] \text{ a.e. } \omega \in \Omega, \mathbb{E}[\zeta] = 1 \}. \quad (6.70)$$

### 6.3. Coherent Risk Measures

273

Since  $AV@R_\alpha(\cdot)$  is convex and continuous, it is subdifferentiable and its subdifferentials can be calculated using formula (6.43). That is,

$$\partial(AV@R_\alpha)(Z) = \arg \max_{\zeta \in Z^*} \{ \langle \zeta, Z \rangle : \zeta(\omega) \in [0, \alpha^{-1}] \text{ a.e. } \omega \in \Omega, \mathbb{E}[\zeta] = 1 \}. \quad (6.71)$$

Consider the maximization problem on the right-hand side of (6.71). 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

$$\text{Min}_{\lambda \in \mathbb{R}} \sup_{\zeta(\cdot) \in [0, \alpha^{-1}]} \{ \langle \zeta, Z - \lambda \rangle + \lambda \}. \quad (6.72)$$

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 (6.72) can be written as

$$\text{Min}_{\lambda \in \mathbb{R}} \alpha^{-1} \mathbb{E}([Z - \lambda]_+) + \lambda. \quad (6.73)$$

The set of optimal solutions of problem (6.73) is the interval with the end points given by the left and right side  $(1 - \alpha)$ -quantiles of the cdf  $H_Z(z) = \Pr(Z \leq z)$  of  $Z(\omega)$ . Since the set of optimal solutions of the dual problem (6.72) is a compact subset of  $\mathbb{R}$ , there is no duality gap between the maximization problem on the right hand side of (6.71) and its dual (6.72) (see Theorem 7.10). It follows that the set of optimal solutions of the right-hand side of (6.71), and hence the subdifferential  $\partial(AV@R_\alpha)(Z)$ , is given by such feasible  $\tilde{\zeta}$  that  $(\tilde{\zeta}, \tilde{\lambda})$  is a saddle point of the Lagrangian  $L(\zeta, \lambda)$  for any  $(1 - \alpha)$ -quantile  $\tilde{\lambda}$ . 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

$$\partial(AV@R_\alpha)(Z) = \left\{ \zeta : \mathbb{E}[\zeta] = 1, \begin{array}{ll} \zeta(\omega) = \alpha^{-1} & \text{if } Z(\omega) > V@R_\alpha(Z), \\ \zeta(\omega) = 0 & \text{if } Z(\omega) < V@R_\alpha(Z), \\ \zeta(\omega) \in [0, \alpha^{-1}] & \text{if } Z(\omega) = V@R_\alpha(Z). \end{array} \right. \quad (6.74)$$

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 (6.74) 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:

$$\Pr(Z < V@R_\alpha(Z)) = 1 - \alpha \text{ or } \Pr(Z > V@R_\alpha(Z)) = \alpha. \quad (6.75)$$

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 (6.75) can be replaced by any  $(1 - \alpha)$ -quantile of  $H_Z$ . Note that condition (6.75) 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 6.17 (Exponential Utility Function Risk Measure).** Consider utility risk measure  $\rho$ , defined in (6.59), 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  $\mathcal{L}_p(\Omega, \mathcal{F}, P)$  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  $\rho$  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  $\rho$  is proper and, by Proposition 6.15, is convex and lower semicontinuous. Note that if  $Z : \Omega \rightarrow \mathbb{R}$  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  $p \geq 1$ . Therefore, by formula (6.61) of Proposition 6.15, 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  $e^Z$ , it is discontinuous and nondifferentiable at  $Z$  unless  $\mathcal{L}_p(\Omega, \mathcal{F}, P)$  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 risk measure

$$\rho_e(Z) := \ln \mathbb{E}[e^Z], \quad (6.76)$$

defined on  $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$  for some  $p \in [1, +\infty)$ . Since  $\ln(\cdot)$  is continuous on the positive half of the real line and  $\mathbb{E}[e^Z] > 0$ , it follows from the above that  $\rho_e$  has the same domain as  $\rho(Z) = \mathbb{E}[e^Z]$  and is lower semicontinuous and proper. It is also can be verified that  $\rho_e$  is convex. (See derivations of section 7.2.8 following (7.175).) Moreover, for any  $a \in \mathbb{R}$ ,

$$\ln \mathbb{E}[e^{Z+a}] = \ln(e^a \mathbb{E}[e^Z]) = \ln \mathbb{E}[e^Z] + a,$$

i.e.,  $\rho_e$  satisfies condition (R3).

Let us calculate the conjugate of  $\rho_e$ . We have that

$$\rho_e^*(\zeta) = \sup_{Z \in \mathcal{Z}} \{ \mathbb{E}[\zeta Z] - \ln \mathbb{E}[e^Z] \}. \quad (6.77)$$

Since  $\rho_e$  satisfies conditions (R2) and (R3), it follows that  $\text{dom}(\rho_e) \subset \mathfrak{P}$ , where  $\mathfrak{P}$  is the set of density functions (see (6.38)). By writing (first order) optimality conditions for the optimization problem on the right-hand side of (6.77), it is straightforward to verify that for  $\zeta \in \mathfrak{P}$  such that  $\zeta(\omega) > 0$  for a.e.  $\omega \in \Omega$ , a point  $\bar{Z}$  is an optimal solution of that problem if  $\bar{Z} = \ln \zeta + a$  for some  $a \in \mathbb{R}$ . Substituting this into the right-hand side of (6.77), and noting that the obtained expression does not depend on  $a$ , we obtain

$$\rho_e^*(\zeta) = \begin{cases} \mathbb{E}[\zeta \ln \zeta] & \text{if } \zeta \in \mathfrak{P}, \\ +\infty & \text{if } \zeta \notin \mathfrak{P}. \end{cases} \quad (6.78)$$

Note that  $x \ln x$  tends to zero as  $x \downarrow 0$ . Therefore, we set  $0 \ln 0 = 0$  in the above formula (6.78). Note also that  $x \ln x$  is bounded for  $x \in [0, 1]$ . Therefore,  $\text{dom}(\rho_e^*) = \mathfrak{P}$  for any  $p \in [1, +\infty)$ .

Furthermore, we can apply the homogenization procedure to  $\rho_e$  (see (6.57)). That is, consider the following risk measure:

$$\check{\rho}_e(Z) := \inf_{\tau > 0} \tau \ln \mathbb{E}[e^{\tau^{-1}Z}]. \quad (6.79)$$

Risk measure  $\check{\rho}_e$  satisfies conditions (R1)–(R4), i.e., it is a coherent risk measure. Its conjugate  $\check{\rho}_e^*$  is the indicator function of the set (see (6.58)):

$$\mathfrak{A} := \{\zeta \in \mathcal{Z}^* : \mathbb{E}[\zeta Z] \leq \ln \mathbb{E}[e^Z], \quad \forall Z \in \mathcal{Z}\}. \quad (6.80)$$

Note that since  $e^z$  is a convex function it follows by Jensen inequality that  $\mathbb{E}[Z] \leq \ln \mathbb{E}[e^Z]$ . Consequently,  $\zeta(\cdot) = 1$  is an element of the above set  $\mathfrak{A}$ . ■

**Example 6.18 (Mean-Variance Risk Measure).** Consider

$$\rho(Z) := \mathbb{E}[Z] + c \mathbb{V}\text{ar}[Z], \quad (6.81)$$

where  $c \geq 0$  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  $\mathbb{V}\text{ar}[Z]$  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

$$\|Z\|_2^2 = \sup_{\zeta \in \mathcal{Z}} \left\{ \langle \zeta, Z \rangle - \frac{1}{4} \|\zeta\|_2^2 \right\}. \quad (6.82)$$

Indeed, it is not difficult to verify that the maximum on the right-hand side of (6.82) is attained at  $\zeta = 2Z$ .

We have that  $\mathbb{V}\text{ar}[Z] = \|Z - \mathbb{E}[Z]\|_2^2$ , and since  $\|\cdot\|_2^2$  is a convex and continuous function on the Hilbert space  $\mathcal{Z}$ , it follows that  $\rho(\cdot)$  is convex and continuous. Also because of (6.82), we can write

$$\mathbb{V}\text{ar}[Z] = \sup_{\zeta \in \mathcal{Z}} \left\{ \langle \zeta, Z - \mathbb{E}[Z] \rangle - \frac{1}{4} \|\zeta\|_2^2 \right\}.$$

Since

$$\langle \zeta, Z - \mathbb{E}[Z] \rangle = \langle \zeta, Z \rangle - \mathbb{E}[\zeta] \mathbb{E}[Z] = \langle \zeta - \mathbb{E}[\zeta], Z \rangle, \quad (6.83)$$

we can rewrite the last expression as follows:

$$\begin{aligned} \mathbb{V}\text{ar}[Z] &= \sup_{\zeta \in \mathcal{Z}} \left\{ \langle \zeta - \mathbb{E}[\zeta], Z \rangle - \frac{1}{4} \|\zeta\|_2^2 \right\} \\ &= \sup_{\zeta \in \mathcal{Z}} \left\{ \langle \zeta - \mathbb{E}[\zeta], Z \rangle - \frac{1}{4} \mathbb{V}\text{ar}[\zeta] - \frac{1}{4} (\mathbb{E}[\zeta])^2 \right\}. \end{aligned}$$

Since  $\zeta - \mathbb{E}[\zeta]$  and  $\mathbb{V}\text{ar}[\zeta]$  are invariant under transformations of  $\zeta$  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

$$\mathbb{V}\text{ar}[Z] = \sup_{\substack{\zeta \in \mathcal{Z} \\ \mathbb{E}[\zeta] = 0}} \left\{ \langle \zeta, Z \rangle - \frac{1}{4} \mathbb{V}\text{ar}[\zeta] \right\}.$$

Therefore the risk measure  $\rho$ , defined in (6.81), can be expressed as

$$\rho(Z) = \mathbb{E}[Z] + c \sup_{\substack{\zeta \in \mathcal{Z} \\ \mathbb{E}[\zeta] = 0}} \left\{ \langle \zeta, Z \rangle - \frac{1}{4} \mathbb{V}\text{ar}[\zeta] \right\}$$

and hence for  $c > 0$  (by making change of variables  $\zeta' = c\zeta + 1$ ) as

$$\rho(Z) = \sup_{\substack{\zeta \in \mathcal{Z} \\ \mathbb{E}[\zeta] = 1}} \left\{ \langle \zeta, Z \rangle - \frac{1}{4c} \text{Var}[\zeta] \right\}. \quad (6.84)$$

It follows that for any  $c > 0$  the function  $\rho$  is convex, continuous, and

$$\rho^*(\zeta) = \begin{cases} \frac{1}{4c} \text{Var}[\zeta] & \text{if } \mathbb{E}[\zeta] = 1, \\ +\infty & \text{otherwise.} \end{cases} \quad (6.85)$$

The function  $\rho$  satisfies the translation equivariance condition (R3), e.g., because the domain of its conjugate contains only  $\zeta$  such that  $\mathbb{E}[\zeta] = 1$ . However, for any  $c > 0$  the function  $\rho$  is not positively homogeneous and it does not satisfy the monotonicity condition (R2), because the domain of  $\rho^*$  contains density functions which are not nonnegative.

Since  $\text{Var}[Z] = \langle Z, Z \rangle - (\mathbb{E}[Z])^2$ , it is straightforward to verify that  $\rho(\cdot)$  is (Fréchet) differentiable and

$$\nabla \rho(Z) = 2cZ - 2c\mathbb{E}[Z] + 1. \quad \blacksquare \quad (6.86)$$

**Example 6.19 (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  $c \geq 0$ , consider

$$\rho(Z) := \mathbb{E}[Z] + c \left( \mathbb{E}[|Z - \mathbb{E}[Z]|^p] \right)^{1/p}. \quad (6.87)$$

We have that  $\left( \mathbb{E}[|Z|^p] \right)^{1/p} = \|Z\|_p$ , where  $\|\cdot\|_p$  denotes the norm of the space  $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ . The function  $\rho$  is convex continuous and positively homogeneous. Also

$$\|Z\|_p = \sup_{\|\zeta\|_q \leq 1} \langle \zeta, Z \rangle, \quad (6.88)$$

and hence

$$\left( \mathbb{E}[|Z - \mathbb{E}[Z]|^p] \right)^{1/p} = \sup_{\|\zeta\|_q \leq 1} \langle \zeta, Z - \mathbb{E}[Z] \rangle = \sup_{\|\zeta\|_q \leq 1} \langle \zeta - \mathbb{E}[\zeta], Z \rangle. \quad (6.89)$$

It follows that representation (6.37) holds with the set  $\mathfrak{A}$  given by

$$\mathfrak{A} = \left\{ \zeta' \in \mathcal{Z}^* : \zeta' = 1 + \zeta - \mathbb{E}[\zeta], \|\zeta\|_q \leq c \right\}. \quad (6.90)$$

We obtain here that  $\rho$  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

$$\zeta'(\omega) = 1 + \zeta(\omega) - \mathbb{E}[\zeta] \geq 1 - |\zeta(\omega)| - \mathbb{E}[\zeta] \geq 1 - 2c.$$

It follows that if  $c \in [0, 1/2]$ , then  $\zeta'(\omega) \geq 0$  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:

$$\text{For any } \varepsilon > 0 \text{ there exists } A \in \mathcal{F} \text{ such that } \varepsilon > P(A) > 0. \quad (6.91)$$

That is, for  $p = 1$  the mean-deviation measure  $\rho$  satisfies (R2) if, and provided that condition (6.91) holds, only if  $c \in [0, 1/2]$ . (The above condition (6.91) 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  $\rho$  does not satisfy (R2) provided that condition (6.91) holds.

Since  $\rho$  is convex continuous, it is subdifferentiable. By (6.43) and because of (6.90) and (6.83) 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,

$$\partial\rho(Z) = \{\zeta' = 1 + c\zeta - c\mathbb{E}[\zeta] : \zeta \in \mathfrak{S}_Y\}, \quad (6.92)$$

where  $Y(\omega) \equiv Z(\omega) - \mathbb{E}[Z]$  and  $\mathfrak{S}_Y$  is the set of contact points of  $Y$ . If  $p \in (1, +\infty)$ , then the set  $\mathfrak{S}_Y$  is a singleton, i.e., there is unique contact point  $\zeta_Y^*$ , provided that  $Y(\omega)$  is not zero for a.e.  $\omega \in \Omega$ . In that case  $\rho(\cdot)$  is Hadamard differentiable at  $Z$  and

$$\nabla\rho(Z) = 1 + c\zeta_Y^* - c\mathbb{E}[\zeta_Y^*]. \quad (6.93)$$

(An explicit form of the contact point  $\zeta_Y^*$  is given in (7.232).) If  $Y(\omega)$  is zero for a.e.  $\omega \in \Omega$ , i.e.,  $Z(\omega)$  is constant w.p. 1, then  $\mathfrak{S}_Y = \{\zeta \in \mathcal{Z}^* : \|\zeta\|_q \leq 1\}$ .

For  $p = 1$  the set  $\mathfrak{S}_Y$  is described in (7.233). 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

$$\nabla\rho(Z) = \left\{ \zeta : \begin{array}{ll} \zeta(\omega) = 1 + 2c(1 - \Pr(Z > \mathbb{E}[Z])) & \text{if } Z(\omega) > \mathbb{E}[Z], \\ \zeta(\omega) = 1 - 2c\Pr(Z > \mathbb{E}[Z]) & \text{if } Z(\omega) < \mathbb{E}[Z]. \end{array} \right. \quad \blacksquare \quad (6.94)$$

**Example 6.20 (Mean-Upper-Semideviation of Order  $p$ ).** Let  $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$  and for  $c \geq 0$  consider<sup>46</sup>

$$\rho(Z) := \mathbb{E}[Z] + c \left( \mathbb{E} \left[ [Z - \mathbb{E}[Z]]_+^p \right] \right)^{1/p}. \quad (6.95)$$

For any  $c \geq 0$  this function satisfies conditions (R1), (R3), and (R4), and similarly to the derivations of Example 6.19 it can be shown that representation (6.37) holds with the set  $\mathfrak{A}$  given by

$$\mathfrak{A} = \{\zeta' \in \mathcal{Z}^* : \zeta' = 1 + \zeta - \mathbb{E}[\zeta], \|\zeta\|_q \leq c, \zeta \geq 0\}. \quad (6.96)$$

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 (6.91) holds, only if  $c \in [0, 1]$ . That is,  $\rho$  is a coherent risk measure if  $c \in [0, 1]$ .

<sup>46</sup>We denote  $[a]_+^p := (\max\{0, a\})^p$ .

Since  $\rho$  is convex continuous, it is subdifferentiable. Its subdifferential can be calculated in a way similar to the derivations of Example 6.19. That is,  $\partial\rho(Z)$  is formed by vectors  $\zeta' = 1 + \zeta - \mathbb{E}[\zeta]$  such that

$$\zeta \in \arg \max \{ \langle \zeta, Y \rangle : \|\zeta\|_q \leq c, \zeta \geq 0 \}, \quad (6.97)$$

where  $Y := Z - \mathbb{E}[Z]$ . Suppose that  $p \in (1, +\infty)$ . Then the set of maximizers on the right-hand side of (6.97) 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

$$\nabla\rho(Z) = 1 + c\zeta_{Y_+}^* - c\mathbb{E}[\zeta_{Y_+}^*], \quad (6.98)$$

where  $\zeta_{Y_+}^*$  is the contact point of  $Y_+$ . (Note that the contact point of  $Y_+$  is nonnegative since  $Y_+ \geq 0$ .)

Suppose now that  $p = 1$  and hence  $q = +\infty$ . Then the set on the right-hand side of (6.97) 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

$$\nabla\rho(Z) = \begin{cases} \zeta(\omega) = 1 + c(1 - \Pr(Z > \mathbb{E}[Z])) & \text{if } Z(\omega) > \mathbb{E}[Z], \\ \zeta(\omega) = 1 - c\Pr(Z > \mathbb{E}[Z]) & \text{if } Z(\omega) < \mathbb{E}[Z]. \end{cases} \quad (6.99)$$

It can be noted that by Lemma 6.1

$$\mathbb{E}(|Z - \mathbb{E}[Z]|) = 2\mathbb{E}([Z - \mathbb{E}[Z]]_+). \quad (6.100)$$

Consequently, formula (6.99) can be derived directly from (6.94). ■

**Example 6.21 (Mean-Upper-Semivariance from a Target).** Let  $Z := \mathcal{L}_2(\Omega, \mathcal{F}, P)$  and for a weight  $c \geq 0$  and a target  $\tau \in \mathbb{R}$  consider

$$\rho(Z) := \mathbb{E}[Z] + c\mathbb{E}\left[[Z - \tau]_+^2\right]. \quad (6.101)$$

This is a convex and continuous risk measure. We can now use (6.63) with  $g(z) := z + c[z - \tau]_+^2$ . Since

$$g^*(\alpha) = \begin{cases} (\alpha - 1)^2/4c + \tau(\alpha - 1) & \text{if } \alpha \geq 1, \\ +\infty & \text{otherwise,} \end{cases}$$

we obtain that

$$\rho(Z) = \sup_{\zeta \in \mathcal{Z}, \zeta(\cdot) \geq 1} \left\{ \mathbb{E}[\zeta Z] - \tau\mathbb{E}[\zeta - 1] - \frac{1}{4c}\mathbb{E}[(\zeta - 1)^2] \right\}. \quad (6.102)$$

Consequently, representation (6.36) holds with  $\mathfrak{A} = \{\zeta \in \mathcal{Z} : \zeta - 1 \geq 0\}$  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  $\rho$  is convex continuous, it is subdifferentiable. Moreover, by using (6.61) we obtain that its subdifferentials are singletons and hence  $\rho(\cdot)$  is differentiable at every  $Z \in \mathcal{Z}$ , and

$$\nabla \rho(Z) = \left\{ \zeta : \begin{array}{ll} \zeta(\omega) = 1 + 2c(Z(\omega) - \tau) & \text{if } Z(\omega) \geq \tau, \\ \zeta(\omega) = 1 & \text{if } Z(\omega) < \tau. \end{array} \right. \quad (6.103)$$

The above formula can also be derived directly, and it can be shown that  $\rho$  is differentiable in the sense of Fréchet. ■

**Example 6.22 (Mean-Upper-Semideviation of Order  $p$  from a Target).** Let  $\mathcal{Z}$  be the space  $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ , and for  $c \geq 0$  and  $\tau \in \mathbb{R}$  consider

$$\rho(Z) := \mathbb{E}[Z] + c \left( \mathbb{E} \left[ [Z - \tau]_+^p \right] \right)^{1/p}. \quad (6.104)$$

For any  $c \geq 0$  and  $\tau$  this risk measure satisfies conditions (R1) and (R2), but not (R3) and (R4) if  $c > 0$ . We have

$$\begin{aligned} \left( \mathbb{E} \left[ [Z - \tau]_+^p \right] \right)^{1/p} &= \sup_{\|\zeta\|_q \leq 1} \mathbb{E}(\zeta[Z - \tau]_+) = \sup_{\|\zeta\|_q \leq 1, \zeta(\cdot) \geq 0} \mathbb{E}(\zeta[Z - \tau]_+) \\ &= \sup_{\|\zeta\|_q \leq 1, \zeta(\cdot) \geq 0} \mathbb{E}(\zeta[Z - \tau]) = \sup_{\|\zeta\|_q \leq 1, \zeta(\cdot) \geq 0} \mathbb{E}[\zeta Z - \tau \zeta]. \end{aligned}$$

We obtain that representation (6.36) holds with

$$\mathfrak{A} = \{\zeta \in \mathcal{Z}^* : \|\zeta\|_q \leq c, \zeta \geq 0\}$$

and  $\rho^*(\zeta) = \tau \mathbb{E}[\zeta]$  for  $\zeta \in \mathfrak{A}$ . ■

### 6.3.3 Law Invariant Risk Measures and Stochastic Orders

As in the previous sections, unless stated otherwise we assume here that  $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ ,  $p \in [1, +\infty)$ . We say that random outcomes  $Z_1 \in \mathcal{Z}$  and  $Z_2 \in \mathcal{Z}$  have the same distribution, with respect to the reference probability measure  $P$ , if  $P(Z_1 \leq z) = P(Z_2 \leq z)$  for all  $z \in \mathbb{R}$ . We write this relation as  $Z_1 \stackrel{\mathcal{D}}{\sim} Z_2$ . In all examples considered in section 6.3.2, the risk measures  $\rho(Z)$  discussed there 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)  $H_Z(t) := P(Z \leq t)$  associated with  $Z \in \mathcal{Z}$ . We call such risk measures law invariant (or law based, or version independent).

**Definition 6.23.** A risk measure  $\rho : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  is law invariant, with respect to the reference probability measure  $P$ , if for all  $Z_1, Z_2 \in \mathcal{Z}$  we have the implication

$$\{Z_1 \stackrel{\mathcal{D}}{\sim} Z_2\} \Rightarrow \{\rho(Z_1) = \rho(Z_2)\}.$$

Suppose for the moment that the set  $\Omega = \{\omega_1, \dots, \omega_K\}$  is finite with respective probabilities  $p_1, \dots, p_K$  such that any two partial sums of  $p_k$  are different, i.e.,  $\sum_{k \in A} p_k =$

$\sum_{k \in B} p_k$  for  $A, B \subset \{1, \dots, K\}$  iff  $A = B$ . Then  $Z_1, Z_2 : \Omega \rightarrow \mathbb{R}$  have the same distribution only if  $Z_1 = Z_2$ . In that case, any risk measure, defined on the space of random variables  $Z : \Omega \rightarrow \mathbb{R}$ , is law invariant. Therefore, for a meaningful discussion of law invariant risk measures it is natural to consider nonatomic probability spaces.

A particular example of law invariant coherent risk measure is the Average Value-at-Risk measure  $\text{AV@R}_\alpha$ . Clearly, a convex combination  $\sum_{i=1}^m \mu_i \text{AV@R}_{\alpha_i}$ , with  $\alpha_i \in (0, 1]$ ,  $\mu_i \geq 0$ ,  $\sum_{i=1}^m \mu_i = 1$ , of Average Value-at-Risk measures is also a law invariant coherent risk measure. Moreover, maximum of several law invariant coherent risk measures is again a law invariant coherent risk measure. It turns out that any law invariant coherent risk measure can be constructed by the operations of taking convex combinations and maximum from the class of Average Value-at-Risk measures.

**Theorem 6.24 (Kusuoka).** *Suppose that the probability space  $(\Omega, \mathcal{F}, P)$  is nonatomic and let  $\rho : \mathcal{Z} \rightarrow \mathbb{R}$  be a law invariant lower semicontinuous coherent risk measure. Then there exists a set  $\mathfrak{M}$  of probability measures on the interval  $(0, 1]$  (equipped with its Borel sigma algebra) such that*

$$\rho(Z) = \sup_{\mu \in \mathfrak{M}} \int_0^1 \text{AV@R}_\alpha(Z) d\mu(\alpha), \quad \forall Z \in \mathcal{Z}. \quad (6.105)$$

In order to prove this we will need the following result.

**Lemma 6.25.** *Let  $(\Omega, \mathcal{F}, P)$  be a nonatomic probability space and  $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$ . Then for  $Z \in \mathcal{Z}$  and  $\zeta \in \mathcal{Z}^*$  we have*

$$\sup_{Y: Y \stackrel{\mathcal{D}}{\sim} Z} \int_{\Omega} \zeta(\omega) Y(\omega) dP(\omega) = \int_0^1 H_\zeta^{-1}(t) H_Z^{-1}(t) dt, \quad (6.106)$$

where  $H_\zeta$  and  $H_Z$  are the cdf's of  $\zeta$  and  $Z$ , respectively.

**Proof.** First we prove formula (6.106) for finite set  $\Omega = \{\omega_1, \dots, \omega_n\}$  with equal probabilities  $P(\{\omega_i\}) = 1/n$ ,  $i = 1, \dots, n$ . For a function  $Y : \Omega \rightarrow \mathbb{R}$  denote  $Y_i := Y(\omega_i)$ ,  $i = 1, \dots, n$ . We have here that  $Y \stackrel{\mathcal{D}}{\sim} Z$  iff  $Y_i = Z_{\pi(i)}$  for some permutation  $\pi$  of the set  $\{1, \dots, n\}$ , and  $\int_{\Omega} \zeta Y dP = n^{-1} \sum_{i=1}^n \zeta_i Y_i$ . Moreover,<sup>47</sup>

$$\sum_{i=1}^n \zeta_i Y_i \leq \sum_{i=1}^n \zeta_{[i]} Y_{[i]}, \quad (6.107)$$

where  $\zeta_{[1]} \leq \dots \leq \zeta_{[n]}$  are numbers  $\zeta_1, \dots, \zeta_n$  arranged in the increasing order, and  $Y_{[1]} \leq \dots \leq Y_{[n]}$  are numbers  $Y_1, \dots, Y_n$  arranged in the increasing order. It follows that

$$\sup_{Y: Y \stackrel{\mathcal{D}}{\sim} Z} \int_{\Omega} \zeta(\omega) Y(\omega) dP(\omega) = n^{-1} \sum_{i=1}^n \zeta_{[i]} Z_{[i]}. \quad (6.108)$$

<sup>47</sup>Inequality (6.107) is called the Hardy–Littlewood–Polya inequality (compare with the proof of Theorem 4.50).

It remains to note that in the considered case the right-hand side of (6.108) coincides with the right-hand side of (6.106).

Now if the space  $(\Omega, \mathcal{F}, P)$  is nonatomic, we can partition  $\Omega$  into  $n$  disjoint subsets, each of the same  $P$ -measure  $1/n$ , and it suffices to verify formula (6.106) for functions which are piecewise constant on such partitions. This reduces the problem to the case considered above.  $\square$

**Proof of Theorem 6.24.** By the dual representation (6.37) of Theorem 6.4, we have that for  $Z \in \mathcal{Z}$ ,

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \int_{\Omega} \zeta(\omega) Z(\omega) dP(\omega), \quad (6.109)$$

where  $\mathfrak{A}$  is a set of probability density functions in  $\mathcal{Z}^*$ . Since  $\rho$  is law invariant, we have that

$$\rho(Z) = \sup_{Y \in \mathcal{D}(Z)} \rho(Y),$$

where  $\mathcal{D}(Z) := \{Y \in \mathcal{Z} : Y \stackrel{\mathcal{D}}{\sim} Z\}$ . Consequently,

$$\rho(Z) = \sup_{Y \in \mathcal{D}(Z)} \left[ \sup_{\zeta \in \mathfrak{A}} \int_{\Omega} \zeta(\omega) Y(\omega) dP(\omega) \right] = \sup_{\zeta \in \mathfrak{A}} \left[ \sup_{Y \in \mathcal{D}(Z)} \int_{\Omega} \zeta(\omega) Y(\omega) dP(\omega) \right]. \quad (6.110)$$

Moreover, by Lemma 6.25 we have

$$\sup_{Y \in \mathcal{D}(Z)} \int_{\Omega} \zeta(\omega) Y(\omega) dP(\omega) = \int_0^1 H_{\zeta}^{-1}(t) H_Z^{-1}(t) dt, \quad (6.111)$$

where  $H_{\zeta}$  and  $H_Z$  are the cdf's of  $\zeta(\omega)$  and  $Z(\omega)$ , respectively.

Recalling that  $H_Z^{-1}(t) = V @ R_{1-t}(Z)$ , we can write (6.111) in the form

$$\sup_{Y \in \mathcal{D}(Z)} \int_{\Omega} \zeta(\omega) Y(\omega) dP(\omega) = \int_0^1 H_{\zeta}^{-1}(t) V @ R_{1-t}(Z) dt, \quad (6.112)$$

which together with (6.110) imply that

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \int_0^1 H_{\zeta}^{-1}(t) V @ R_{1-t}(Z) dt. \quad (6.113)$$

For  $\zeta \in \mathfrak{A}$ , the function  $H_{\zeta}^{-1}(t)$  is monotonically nondecreasing on  $[0,1]$  and can be represented in the form

$$H_{\zeta}^{-1}(t) = \int_{1-t}^1 \alpha^{-1} d\mu(\alpha) \quad (6.114)$$

for some measure  $\mu$  on  $[0,1]$ . Moreover, for  $\zeta \in \mathfrak{A}$  we have that  $\int \zeta dP = 1$ , and hence  $\int_0^1 H_{\zeta}^{-1}(t) dt = \int \zeta dP = 1$ , and therefore

$$1 = \int_0^1 \int_{1-t}^1 \alpha^{-1} d\mu(\alpha) dt = \int_0^1 \int_{1-\alpha}^1 \alpha^{-1} dt d\mu(\alpha) = \int_0^1 d\mu(\alpha).$$

Consequently,  $\mu$  is a probability measure on  $[0, 1]$ . Also (see Theorem 6.2) we have

$$AV@R_\alpha(Z) = \frac{1}{\alpha} \int_{1-\alpha}^1 V@R_{1-t}(Z) dt,$$

and hence

$$\begin{aligned} \int_0^1 AV@R_\alpha(Z) d\mu(\alpha) &= \int_0^1 \int_{1-\alpha}^1 \alpha^{-1} V@R_{1-t}(Z) dt d\mu(\alpha) \\ &= \int_0^1 V@R_{1-t}(Z) \left( \int_{1-t}^1 \alpha^{-1} d\mu(\alpha) \right) dt \\ &= \int_0^1 V@R_{1-t}(Z) H_\zeta^{-1}(t) dt. \end{aligned}$$

By (6.113) this completes the proof, with the correspondence between  $\zeta \in \mathfrak{A}$  and  $\mu \in \mathfrak{M}$  given by (6.114).  $\square$

**Example 6.26.** Consider  $\rho := AV@R_\gamma$  risk measure for some  $\gamma \in (0, 1)$ . Assume that the corresponding probability space is  $\Omega = [0, 1]$  equipped with its Borel sigma algebra and uniform probability measure  $P$ . We have here (see (6.70))

$$\mathfrak{A} = \left\{ \zeta : 0 \leq \zeta(\omega) \leq \gamma^{-1}, \omega \in [0, 1], \int_0^1 \zeta(\omega) d\omega = 1 \right\}.$$

Consequently, the family of cumulative distribution functions  $H_\zeta^{-1}$ ,  $\zeta \in \mathfrak{A}$ , is formed by left-side continuous monotonically nondecreasing on  $[0, 1]$  functions with  $\int_0^1 H_\zeta^{-1}(t) dt = 1$  and range values  $0 \leq H_\zeta^{-1}(t) \leq \gamma^{-1}$ ,  $t \in [0, 1]$ . Since  $V@R_{1-t}(Z)$  is monotonically nondecreasing in  $t$  function, it follows that the maximum in the right-hand side of (6.113) is attained at  $\zeta \in \mathfrak{A}$  such that  $H_\zeta^{-1}(t) = 0$  for  $t \in [0, 1 - \gamma]$ , and  $H_\zeta^{-1}(t) = \gamma^{-1}$  for  $t \in (1 - \gamma, 1]$ . The corresponding measure  $\mu$ , defined by (6.114), is given by function  $\mu(\alpha) = 1$  for  $\alpha \in [0, \gamma]$  and  $\mu(\alpha) = 0$  for  $\alpha \in (\gamma, 1]$ , i.e.,  $\mu$  is the measure of mass 1 at the point  $\gamma$ . By the above proof of Theorem 6.24, this  $\mu$  is the maximizer of the right-hand side of (6.105). It follows that the representation (6.105) recovers the measure  $AV@R_\gamma$ , as it should be.  $\blacksquare$

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  $u : \mathbb{R} \rightarrow \mathbb{R}$  as follows. For (real valued) random variables  $Z_1$  and  $Z_2$  it is said that  $Z_2$  dominates  $Z_1$ , denoted  $Z_2 \succeq_u Z_1$ , if  $\mathbb{E}[u(Z_2)] \geq \mathbb{E}[u(Z_1)]$  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  $u : \mathbb{R} \rightarrow \mathbb{R}$ . Equivalently,  $Z_2 \succeq_{(1)} Z_1$  iff  $H_{Z_2}(t) \leq H_{Z_1}(t)$  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 *increasing* if all functions in the set  $\mathcal{U}$  are nondecreasing. The usual stochastic order is an example of increasing integral stochastic order.

**Definition 6.27.** A law invariant risk measure  $\rho : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  is consistent (monotone) with the integral stochastic order  $\succeq_u$  if for all  $Z_1, Z_2 \in \mathcal{Z}$  we have the implication

$$\{Z_2 \succeq_u Z_1\} \Rightarrow \{\rho(Z_2) \geq \rho(Z_1)\}.$$

For an increasing integral stochastic order we have that if  $Z_2(\omega) \geq Z_1(\omega)$  for a.e.  $\omega \in \Omega$ , then  $u(Z_2(\omega)) \geq u(Z_1(\omega))$  for any  $u \in \mathcal{U}$  and a.e.  $\omega \in \Omega$ , and hence  $\mathbb{E}[u(Z_2)] \geq \mathbb{E}[u(Z_1)]$ . That is, if  $Z_2 \succeq Z_1$  in the almost sure sense, then  $Z_2 \succeq_u Z_1$ . It follows that if  $\rho$  is law invariant and consistent with respect to an increasing integral stochastic order, then it satisfies the monotonicity condition (R2). In other words, if  $\rho$  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 6.20, is not consistent with any increasing integral stochastic order, provided that condition (6.91) holds.

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  $\mathcal{Z}$ , 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  $(\Omega, \mathcal{F}, P)$  is nonatomic. Then there exists a measurable function  $U : \Omega \rightarrow \mathbb{R}$  (uniform random variable) such that  $P(U \leq t) = t$  for all  $t \in [0, 1]$ .

**Theorem 6.28.** *Suppose that the probability space  $(\Omega, \mathcal{F}, P)$  is nonatomic. Then the following holds: if a risk measure  $\rho : \mathcal{Z} \rightarrow \mathbb{R}$  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  $\hat{Z}_2(\omega) \geq \hat{Z}_1(\omega)$  for all  $\omega \in \Omega$ , and hence by virtue of (R2),  $\rho(\hat{Z}_2) \geq \rho(\hat{Z}_1)$ . 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  $\rho(Z_2) \geq \rho(Z_1)$ . Consequently, the risk measure  $\rho$  is consistent with the usual stochastic order.  $\square$

It is said that  $Z_1$  is smaller than  $Z_2$  in the *increasing convex order*, written  $Z_1 \preceq_{\text{icx}} Z_2$ , if  $\mathbb{E}[u(Z_1)] \leq \mathbb{E}[u(Z_2)]$  for all increasing convex functions  $u : \mathbb{R} \rightarrow \mathbb{R}$  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

$$\mathbb{E}\{[Z_1 - \eta]_+\} \leq \mathbb{E}\{[Z_2 - \eta]_+\}, \quad \forall \eta \in \mathbb{R}. \quad (6.115)$$

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 (6.115) implies that  $\mathbb{E}[u(Z_1)] \leq \mathbb{E}[u(Z_2)]$ , as claimed (compare with the statement (4.8)).

**Theorem 6.29.** *Suppose that the probability space  $(\Omega, \mathcal{F}, P)$  is nonatomic. Then any law invariant lower semicontinuous coherent risk measure  $\rho : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  is consistent with the increasing convex order.*

**Proof.** By using definition (6.22) of  $AV@R_\alpha$  and the property that  $Z_1 \preceq_{\text{icx}} Z_2$  iff condition (6.115) holds, it is straightforward to verify that  $AV@R_\alpha$  is consistent with the increasing convex order. Now by using the representation (6.105) of Theorem 6.24 and noting that the operations of taking convex combinations and maximum preserve consistency with the increasing convex order, we can complete the proof.  $\square$

**Remark 20.** For *convex* risk measures (without the positive homogeneity property), Theorem 6.29 in the space  $\mathcal{L}_1(\Omega, \mathcal{F}, P)$  can be derived from Theorem 4.52, which for the increasing convex order can be written as follows:

$$\{Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P) : Z \preceq_{\text{icx}} Y\} = \text{cl conv}\{Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P) : Z \preceq_{(1)} Y\}. \quad (6.116)$$

If  $Z$  is an element of the set in the left-hand side of (6.116), then there exists a sequence of random variables  $Z^k \rightarrow Z$ , which are convex combinations of some elements of the set in the right-hand side of (6.116), that is,

$$Z^k = \sum_{j=1}^{N_k} \alpha_j^k Z_j^k, \quad \sum_{j=1}^{N_k} \alpha_j^k = 1, \quad \alpha_j^k \geq 0, \quad Z_j^k \preceq_{(1)} Y.$$

By convexity of  $\rho$  and by Theorem 6.28, we obtain

$$\rho(Z^k) \leq \sum_{j=1}^{N_k} \alpha_j^k \rho(Z_j^k) \leq \sum_{j=1}^{N_k} \alpha_j^k \rho(Y) = \rho(Y).$$

Passing to the limit with  $k \rightarrow \infty$  and using lower semicontinuity of  $\rho$ , we obtain  $\rho(Z) \leq \rho(Y)$ , as required.

If  $p > 1$  the domain of  $\rho$  can be extended to  $\mathcal{L}_1(\Omega, \mathcal{F}, P)$ , while preserving its lower semicontinuity (cf. Filipović and Svindland [66]).

**Remark 21.** For some measures of risk, in particular, for the mean-semideviation measures, defined in Example 6.20, and for the Average Value-at-Risk, defined in Example 6.16, consistency with the increasing convex order can be proved *without* the assumption that the probability space  $(\Omega, \mathcal{F}, P)$  is nonatomic by using the following construction. Let  $(\Omega, \mathcal{F}, P)$  be a nonatomic probability space; for example, we can take  $\Omega$  as the interval  $[0, 1]$  equipped with its Borel sigma algebra and uniform probability measure  $P$ . Then for any finite set of probabilities  $p_k > 0$ ,  $k = 1, \dots, K$ ,  $\sum_{k=1}^K p_k = 1$ , we can construct a partition of the set  $\Omega = \cup_{k=1}^K A_k$  such that  $P(A_k) = p_k$ ,  $k = 1, \dots, K$ . Consider the linear subspace of the respective space  $\mathcal{L}_p(\Omega, \mathcal{F}, P)$  formed by piecewise constant on the sets  $A_k$  functions  $Z : \Omega \rightarrow \mathbb{R}$ . We can identify this subspace with the space of random variables defined on a finite probability space of cardinality  $K$  with the respective probabilities  $p_k$ ,  $k = 1, \dots, K$ . By the above theorem, the mean-upper-semideviation risk measure (of order  $p$ ) defined on  $(\Omega, \mathcal{F}, P)$  is consistent with the increasing convex order. This property is preserved by restricting it to the constructed subspace. This shows that the mean-upper-semideviation risk measures are consistent with the increasing convex order on any finite probability space. This can be extended to the general probability spaces by continuity arguments.



**Corollary 6.30.** *Suppose that the probability space  $(\Omega, \mathcal{F}, P)$  is nonatomic. Let  $\rho : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  be a law invariant lower semicontinuous coherent risk measure and  $\mathcal{G}$  be a sigma subalgebra of the sigma algebra  $\mathcal{F}$ . Then*

$$\rho(\mathbb{E}[Z|\mathcal{G}]) \leq \rho(Z), \quad \forall Z \in \mathcal{Z}, \quad (6.117)$$

and

$$\mathbb{E}[Z] \leq \rho(Z), \quad \forall Z \in \mathcal{Z}. \quad (6.118)$$

**Proof.** Consider  $Z \in \mathcal{Z}$  and  $Z' := \mathbb{E}[Z|\mathcal{G}]$ . For every convex function  $u : \mathbb{R} \rightarrow \mathbb{R}$  we have

$$\mathbb{E}[u(Z')] = \mathbb{E}[u(\mathbb{E}[Z|\mathcal{G}])] \leq \mathbb{E}[\mathbb{E}(u(Z)|\mathcal{G})] = \mathbb{E}[u(Z)],$$

where the inequality is implied by Jensen's inequality. This shows that  $Z' \preceq_{\text{icx}} Z$ , and hence (6.117) follows by Theorem 6.29.

In particular, for  $\mathcal{G} := \{\Omega, \emptyset\}$ , it follows by (6.117) that  $\rho(Z) \geq \rho(\mathbb{E}[Z])$ , and since  $\rho(\mathbb{E}[Z]) = \mathbb{E}[Z]$  this completes the proof.  $\square$

An intuitive interpretation of property (6.117) 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  $\mathbb{E}[Z'] = \mathbb{E}[Z]$ .

### 6.3.4 Relation to Ambiguous Chance Constraints

Owing to the dual representation (6.36), measures of risk are related to robust and ambiguous models. Consider a chance constraint of the form

$$P\{C(x, \omega) \leq 0\} \geq 1 - \alpha. \quad (6.119)$$

Here  $P$  is a probability measure on a measurable space  $(\Omega, \mathcal{F})$  and  $C : \mathbb{R}^n \times \Omega \rightarrow \mathbb{R}$  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 constraints are called *ambiguous chance constrained* problems. For a specified uncertainty set  $\mathcal{A}$  of probability measures on  $(\Omega, \mathcal{F})$ , the corresponding ambiguous chance constraint defines a feasible set  $X \subset \mathbb{R}^n$ , which can be written as

$$X := \{x : \mu\{C(x, \omega) \leq 0\} \geq 1 - \alpha, \quad \forall \mu \in \mathcal{A}\}. \quad (6.120)$$

The set  $X$  can be written in the following equivalent form:

$$X = \left\{ x \in \mathbb{R}^n : \sup_{\mu \in \mathcal{A}} \mathbb{E}_{\mu} [\mathbf{1}_{A_x}] \leq \alpha \right\}, \quad (6.121)$$

where  $A_x := \{\omega \in \Omega : C(x, \omega) > 0\}$ . Recall that by the duality representation (6.37), with the set  $\mathcal{A}$  is associated a coherent risk measure  $\rho$ , and hence (6.121) can be written as

$$X = \{x \in \mathbb{R}^n : \rho(\mathbf{1}_{A_x}) \leq \alpha\}. \quad (6.122)$$

We discuss now constraints of the form (6.122) where the respective risk measure is defined in a direct way. As before, we use spaces  $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ , where  $P$  is viewed as a reference probability measure.

It is not difficult to see that if  $\rho$  is a law invariant risk measure, then for  $A \in \mathcal{F}$  the quantity  $\rho(\mathbf{1}_A)$  depends only on  $P(A)$ . Indeed, if  $Z := \mathbf{1}_A$  for some  $A \in \mathcal{F}$ , then its cdf  $H_Z(z) := P(Z \leq z)$  is

$$H_Z(z) = \begin{cases} 0 & \text{if } z < 0, \\ 1 - P(A) & \text{if } 0 \leq z < 1, \\ 1 & \text{if } 1 \leq z, \end{cases}$$

which clearly depends only on  $P(A)$ .

- With every law invariant real valued risk measure  $\rho : \mathcal{Z} \rightarrow \mathbb{R}$  we associate function  $\varphi_\rho$  defined as  $\varphi_\rho(t) := \rho(\mathbf{1}_A)$ , where  $A \in \mathcal{F}$  is any event such that  $P(A) = t$ , and  $t \in T := \{P(A) : A \in \mathcal{F}\}$ .

The function  $\varphi_\rho$  is well defined because for law invariant risk measure  $\rho$  the quantity  $\rho(\mathbf{1}_A)$  depends only on the probability  $P(A)$  and hence  $\rho(\mathbf{1}_A)$  is the same for any  $A \in \mathcal{F}$  such that  $P(A) = t$  for a given  $t \in T$ . Clearly  $T$  is a subset of the interval  $[0, 1]$ , and  $0 \in T$  (since  $\emptyset \in \mathcal{F}$ ) and  $1 \in T$  (since  $\Omega \in \mathcal{F}$ ). If  $P$  is a nonatomic measure, then for any  $A \in \mathcal{F}$  the set  $\{P(B) : B \subset A, B \in \mathcal{F}\}$  coincides with the interval  $[0, P(A)]$ . In particular, if  $P$  is nonatomic, then the set  $T = \{P(A) : A \in \mathcal{F}\}$ , on which  $\varphi_\rho$  is defined, coincides with the interval  $[0, 1]$ .

**Proposition 6.31.** *Let  $\rho : \mathcal{Z} \rightarrow \mathbb{R}$  be a (real valued) law invariant coherent risk measure. Suppose that the reference probability measure  $P$  is nonatomic. Then  $\varphi_\rho(\cdot)$  is a continuous nondecreasing function defined on the interval  $[0, 1]$  such that  $\varphi_\rho(0) = 0$  and  $\varphi_\rho(1) = 1$ , and  $\varphi_\rho(t) \geq t$  for all  $t \in [0, 1]$ .*

**Proof.** Since the coherent risk measure  $\rho$  is real valued, it is continuous. Because  $\rho$  is continuous and positively homogeneous,  $\rho(0) = 0$  and hence  $\varphi_\rho(0) = 0$ . Also by (R3), we have that  $\rho(\mathbf{1}_\Omega) = 1$  and hence  $\varphi_\rho(1) = 1$ . By Corollary 6.30 we have that  $\rho(\mathbf{1}_A) \geq P(A)$  for any  $A \in \mathcal{F}$  and hence  $\varphi_\rho(t) \geq t$  for all  $t \in [0, 1]$ .

Let  $t_k \in [0, 1]$  be a monotonically increasing sequence tending to  $t^*$ . Since  $P$  is a nonatomic, there exists a sequence  $A_1 \subset A_2 \subset \dots$  of  $\mathcal{F}$ -measurable sets such that  $P(A_k) = t_k$  for all  $k \in \mathbb{N}$ . It follows that the set  $A := \cup_{k=1}^\infty A_k$  is  $\mathcal{F}$ -measurable and  $P(A) = t^*$ . Since  $\mathbf{1}_{A_k}$  converges (in the norm topology of  $\mathcal{Z}$ ) to  $\mathbf{1}_A$ , it follows by continuity of  $\rho$  that  $\rho(\mathbf{1}_{A_k})$  tends to  $\rho(\mathbf{1}_A)$ , and hence  $\varphi_\rho(t_k)$  tends to  $\varphi_\rho(t^*)$ . In a similar way we have that  $\varphi_\rho(t_k) \rightarrow \varphi_\rho(t^*)$  for a monotonically decreasing sequence  $t_k$  tending to  $t^*$ . This shows that  $\varphi_\rho$  is continuous.

For any  $0 \leq t_1 < t_2 \leq 1$  there exist sets  $A, B \in \mathcal{F}$  such that  $B \subset A$  and  $P(B) = t_1$ ,  $P(A) = t_2$ . Since  $\mathbf{1}_A \geq \mathbf{1}_B$ , it follows by monotonicity of  $\rho$  that  $\rho(\mathbf{1}_A) \geq \rho(\mathbf{1}_B)$ . This implies that  $\varphi_\rho(t_2) \geq \varphi_\rho(t_1)$ , i.e.,  $\varphi_\rho$  is nondecreasing.  $\square$

Now consider again the set  $X$  of the form (6.120). Assuming conditions of Proposition 6.31, we obtain that this set  $X$  can be written in the following equivalent form:

$$X = \{x : P\{C(x, \omega) \leq 0\} \geq 1 - \alpha^*\}, \quad (6.123)$$

where  $\alpha^* := \varphi_\rho^{-1}(\alpha)$ . That is,  $X$  can be defined by a chance constraint with respect to the reference distribution  $P$  and with the respective significance level  $\alpha^*$ . Since  $\varphi_\rho(t) \geq t$ , for any  $t \in [0, 1]$ , it follows that  $\alpha^* \leq \alpha$ . Let us consider some examples.

Consider Average Value-at-Risk measure  $\rho := \text{AV@R}_\gamma$ ,  $\gamma \in (0, 1]$ . By direct calculations it is straightforward to verify that for any  $A \in \mathcal{F}$

$$\text{AV@R}_\gamma(\mathbf{1}_A) = \begin{cases} \gamma^{-1}P(A) & \text{if } P(A) \leq \gamma, \\ 1 & \text{if } P(A) > \gamma. \end{cases}$$

Consequently the corresponding function  $\varphi_\rho(t) = \gamma^{-1}t$  for  $t \in [0, \gamma]$ , and  $\varphi_\rho(t) = 1$  for  $t \in [\gamma, 1]$ . Now let  $\rho$  be a convex combination of Average Value-at-Risk measures, i.e.,  $\rho := \sum_{i=1}^m \lambda_i \rho_i$ , with  $\rho_i := \text{AV@R}_{\gamma_i}$  and positive weights  $\lambda_i$  summing up to one. By the definition of the function  $\varphi_\rho$  we have then that  $\varphi_\rho = \sum_{i=1}^m \lambda_i \varphi_{\rho_i}$ . It follows that  $\varphi_\rho : [0, 1] \rightarrow [0, 1]$  is a piecewise linear nondecreasing concave function with  $\varphi_\rho(0) = 0$  and  $\varphi_\rho(1) = 1$ . More generally, let  $\lambda$  be a probability measure on  $(0, 1]$  and  $\rho := \int_0^1 \text{AV@R}_\gamma d\lambda(\gamma)$ . In that case, the corresponding function  $\varphi_\rho$  becomes a nondecreasing concave function with  $\varphi_\rho(0) = 0$  and  $\varphi_\rho(1) = 1$ . We also can consider measures  $\rho$  given by the maximum of such integral functions over some set  $\mathcal{M}$  of probability measures on  $(0, 1]$ . In that case the respective function  $\varphi_\rho$  becomes the maximum of the corresponding nondecreasing concave functions. By Theorem 6.24 this actually gives the most general form of the function  $\varphi_\rho$ .

For instance, let  $\mathcal{Z} := \mathcal{L}_1(\Omega, \mathcal{F}, P)$  and  $\rho(Z) := (1 - \beta)\mathbb{E}[Z] + \beta \text{AV@R}_\gamma(Z)$ , where  $\beta, \gamma \in (0, 1)$  and the expectations are taken with respect to the reference distribution  $P$ . This risk measure was discussed in example 6.16. Then

$$\varphi_\rho(t) = \begin{cases} (1 - \beta + \gamma^{-1}\beta)t & \text{if } t \in [0, \gamma], \\ \beta + (1 - \beta)t & \text{if } t \in (\gamma, 1]. \end{cases} \quad (6.124)$$

It follows that for this risk measure and for  $\alpha \leq \beta + (1 - \beta)\gamma$ ,

$$\alpha^* = \frac{\alpha}{1 + \beta(\gamma^{-1} - 1)}. \quad (6.125)$$

In particular, for  $\beta = 1$ , i.e., for  $\rho = \text{AV@R}_\gamma$ , we have that  $\alpha^* = \gamma\alpha$ .

As another example consider the mean-upper-semideviation risk measure of order  $p$ . That is,  $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$  and

$$\rho(Z) := \mathbb{E}[Z] + c \left( \mathbb{E} \left[ [Z - \mathbb{E}[Z]]_+^p \right] \right)^{1/p}$$

(see Example 6.20). We have here that  $\rho(\mathbf{1}_A) = P(A) + c[P(A)(1 - P(A))^p]^{1/p}$ , and hence

$$\varphi_\rho(t) = t + c t^{1/p} (1 - t), \quad t \in [0, 1]. \quad (6.126)$$

In particular, for  $p = 1$  we have that  $\varphi_\rho(t) = (1 + c)t - ct^2$ , and hence

$$\alpha^* = \frac{1 + c - \sqrt{(1 + c)^2 - 4\alpha c}}{2c}. \quad (6.127)$$

Note that for  $c > 1$  the above function  $\varphi_\rho(\cdot)$  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.

## 6.4 Optimization of Risk Measures

As before, we use 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  $F : \mathbb{R}^n \rightarrow \mathcal{Z}$  and a risk measure  $\rho : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$ . 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  $(\Omega, \mathcal{F})$ . 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

$$\text{Min}_{x \in X} \{ \phi(x) := \rho(F(x)) \}. \quad (6.128)$$

Unless stated otherwise, we assume that the feasible set  $X$  is a nonempty convex closed subset of  $\mathbb{R}^n$ . Of course, if we use  $\rho(\cdot) := \mathbb{E}[\cdot]$ , then problem (6.128) 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 6.18). We shall discuss this later.

Unless stated otherwise we assume that the risk measure  $\rho$  is proper and lower semicontinuous and satisfies conditions (R1)–(R2). By Theorem 6.4 we can use representation (6.36) to write problem (6.128) in the form

$$\text{Min}_{x \in X} \sup_{\zeta \in \mathfrak{A}} \Phi(x, \zeta), \quad (6.129)$$

where  $\mathfrak{A} := \text{dom}(\rho^*)$  and the function  $\Phi : \mathbb{R}^n \times \mathcal{Z}^* \rightarrow \overline{\mathbb{R}}$  is defined by

$$\Phi(x, \zeta) := \int_{\Omega} f(x, \omega) \zeta(\omega) dP(\omega) - \rho^*(\zeta). \quad (6.130)$$

If, moreover,  $\rho$  is positively homogeneous, then  $\rho^*$  is the indicator function of the set  $\mathfrak{A}$  and hence  $\rho^*(\cdot)$  is identically zero on  $\mathfrak{A}$ . That is, if  $\rho$  is a proper lower semicontinuous coherent risk measure, then problem (6.128) can be written as the minimax problem

$$\text{Min}_{x \in X} \sup_{\zeta \in \mathfrak{A}} \mathbb{E}_{\zeta}[f(x, \omega)], \quad (6.131)$$

where

$$\mathbb{E}_{\zeta}[f(x, \omega)] := \int_{\Omega} f(x, \omega) \zeta(\omega) dP(\omega)$$

denotes the expectation with respect to  $\zeta dP$ . Note that, by the definition,  $F(x) \in \mathcal{Z}$  and  $\zeta \in \mathcal{Z}^*$ , and hence

$$\mathbb{E}_{\zeta}[f(x, \omega)] = \langle F(x), \zeta \rangle$$

is finite valued.

Suppose that the mapping  $F : \mathbb{R}^n \rightarrow \mathcal{Z}$  is convex, i.e., for a.e.  $\omega \in \Omega$  the function  $f(\cdot, \omega)$  is convex. This implies that for every  $\zeta \geq 0$  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  $\zeta \in \mathcal{Z}^*$ , and hence for every  $x \in X$  the function  $\Phi(x, \cdot)$  is concave. Therefore, under various regularity conditions, there is no duality gap between problem (6.128) and its dual

$$\text{Max}_{\zeta \in \mathfrak{A}} \inf_{x \in X} \left\{ \int_{\Omega} f(x, \omega) \zeta(\omega) dP(\omega) - \rho^*(\zeta) \right\}, \quad (6.132)$$

which is obtained by interchanging the min and max operators in (6.129). (Recall that the set  $X$  is assumed to be nonempty closed and convex.) In particular, if there exists a saddle point  $(\bar{x}, \bar{\zeta}) \in X \times \mathfrak{A}$  of the minimax problem (6.129), then there is no duality gap between problems (6.129) and (6.132), and  $\bar{x}$  and  $\bar{\zeta}$  are optimal solutions of (6.129) and (6.132), respectively.

**Proposition 6.32.** *Suppose that mapping  $F : \mathbb{R}^n \rightarrow \mathcal{Z}$  is convex and risk measure  $\rho : \mathcal{Z} \rightarrow \bar{\mathbb{R}}$  is proper and lower semicontinuous and satisfies conditions (R1)–(R2). Then  $(\bar{x}, \bar{\zeta}) \in X \times \mathfrak{A}$  is a saddle point of  $\Phi(x, \zeta)$  iff  $\bar{\zeta} \in \partial \rho(\bar{Z})$  and*

$$0 \in \mathcal{N}_X(\bar{x}) + \mathbb{E}_{\bar{\zeta}}[\partial f_{\omega}(\bar{x})], \quad (6.133)$$

where  $\bar{Z} := F(\bar{x})$ .

**Proof.** By the definition,  $(\bar{x}, \bar{\zeta})$  is a saddle point of  $\Phi(x, \zeta)$  iff

$$\bar{x} \in \arg \min_{x \in X} \Phi(x, \bar{\zeta}) \quad \text{and} \quad \bar{\zeta} \in \arg \max_{\zeta \in \mathfrak{A}} \Phi(\bar{x}, \zeta). \quad (6.134)$$

The first of the above conditions means that  $\bar{x} \in \arg \min_{x \in X} \psi(x)$ , where

$$\psi(x) := \int_{\Omega} f(x, \omega) \bar{\zeta}(\omega) dP(\omega).$$

Since  $X$  is convex and  $\psi(\cdot)$  is convex real valued, by the standard optimality conditions this holds iff  $0 \in \mathcal{N}_X(\bar{x}) + \partial \psi(\bar{x})$ . Moreover, by Theorem 7.47 we have  $\partial \psi(\bar{x}) = \mathbb{E}_{\bar{\zeta}}[\partial f_{\omega}(\bar{x})]$ . Therefore, condition (6.133) and the first condition in (6.134) are equivalent. The second condition (6.134) and the condition  $\bar{\zeta} \in \partial \rho(\bar{Z})$  are equivalent by (6.42).  $\square$

Under the assumptions of Proposition 6.32, existence of  $\bar{\zeta} \in \partial \rho(\bar{Z})$  in (6.133) can be viewed as an optimality condition for problem (6.128). 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 (6.129). In order for that condition to be necessary we need to verify existence of a saddle point for problem (6.129).

**Proposition 6.33.** *Let  $\bar{x}$  be an optimal solution of the problem (6.128). Suppose that the mapping  $F : \mathbb{R}^n \rightarrow \mathcal{Z}$  is convex and risk measure  $\rho : \mathcal{Z} \rightarrow \bar{\mathbb{R}}$  is proper and lower semicontinuous and satisfies conditions (R1)–(R2) and is continuous at  $\bar{Z} := F(\bar{x})$ . Then there exists  $\bar{\zeta} \in \partial \rho(\bar{Z})$  such that  $(\bar{x}, \bar{\zeta})$  is a saddle point of  $\Phi(x, \zeta)$ .*

**Proof.** By monotonicity of  $\rho$  (condition (R2)) it follows from the optimality of  $\bar{x}$  that  $(\bar{x}, \bar{Z})$  is an optimal solution of the problem

$$\min_{(x, Z) \in S} \rho(Z), \quad (6.135)$$

where  $S := \{(x, Z) \in X \times \mathcal{Z} : F(x) \geq Z\}$ . Since  $F$  is convex, the set  $S$  is convex, and since  $F$  is continuous (see Lemma 6.9), the set  $S$  is closed. Also because  $\rho$  is convex and continuous at  $\bar{Z}$ , the following (first order) optimality condition holds at  $(\bar{x}, \bar{Z})$  (see Remark 34, page 403):

$$0 \in \partial\rho(\bar{Z}) \times \{0\} + \mathcal{N}_S(\bar{x}, \bar{Z}). \quad (6.136)$$

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

$$\langle \bar{\zeta}, Z - \bar{Z} \rangle \geq 0, \quad \forall (x, Z) \in S. \quad (6.137)$$

Setting  $Z := F(x)$  we obtain that

$$\langle \bar{\zeta}, F(x) - F(\bar{x}) \rangle \geq 0, \quad \forall x \in X. \quad (6.138)$$

It follows that  $\bar{x}$  is a minimizer of  $\langle \bar{\zeta}, F(x) \rangle$  over  $x \in X$ , and hence  $\bar{x}$  is a minimizer of  $\Phi(x, \bar{\zeta})$  over  $x \in X$ . That is,  $\bar{x}$  satisfies first of the two conditions in (6.134). Moreover, as it was shown in the proof of Proposition 6.32, this implies condition (6.133), and hence  $(\bar{x}, \bar{\zeta})$  is a saddle point by Proposition 6.32.  $\square$

**Corollary 6.34.** Suppose that problem (6.128) has optimal solution  $\bar{x}$ , the mapping  $F : \mathbb{R}^n \rightarrow \mathcal{Z}$  is convex and risk measure  $\rho : \mathcal{Z} \rightarrow \mathbb{R}$  is proper and lower semicontinuous and satisfies conditions (R1)–(R2), and is continuous at  $\bar{Z} := F(\bar{x})$ . Then there is no duality gap between problems (6.129) and (6.132), and problem (6.132) has an optimal solution.

Propositions 6.32 and 6.33 imply the following optimality conditions.

**Theorem 6.35.** Suppose that mapping  $F : \mathbb{R}^n \rightarrow \mathcal{Z}$  is convex and risk measure  $\rho : \mathcal{Z} \rightarrow \mathbb{R}$  is proper and lower semicontinuous and satisfies conditions (R1)–(R2). Consider a point  $\bar{x} \in X$  and let  $\bar{Z} := F(\bar{x})$ . Then a sufficient condition for  $\bar{x}$  to be an optimal solution of the problem (6.128) is existence of  $\bar{\zeta} \in \partial\rho(\bar{Z})$  such that (6.133) holds. This condition is also necessary if  $\rho$  is continuous at  $\bar{Z}$ .

It could be noted that if  $\rho(\cdot) := \mathbb{E}[\cdot]$ , then its subdifferential consists of unique subgradient  $\bar{\zeta}(\cdot) \equiv 1$ . In that case condition (6.133) takes the form

$$0 \in \mathcal{N}_X(\bar{x}) + \mathbb{E}[\partial f_\omega(\bar{x})]. \quad (6.139)$$

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 7.47).

### 6.4.1 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  $F : \mathbb{R}^n \rightarrow \mathcal{Z}$  is convex and  $\rho : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  is proper lower semicontinuous and satisfies conditions (R1) and (R2). A way to represent problem (6.128) is to consider the decision vector  $x$  as a function of the elementary event  $\omega \in \Omega$  and then to impose an appropriate nonanticipativity constraint. That is, let  $\mathfrak{M}$  be a linear space of  $\mathcal{F}$ -measurable mappings  $\chi : \Omega \rightarrow \mathbb{R}^n$ . Define  $F_\chi(\omega) := f(\chi(\omega), \omega)$  and

$$\mathfrak{M}_X := \{\chi \in \mathfrak{M} : \chi(\omega) \in X, \text{ a.e. } \omega \in \Omega\}. \quad (6.140)$$

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 (6.128) in the following equivalent form:

$$\text{Min}_{(\chi, x) \in \mathfrak{M}_X \times \mathbb{R}^n} \rho(F_\chi) \quad \text{s.t. } \chi(\omega) = x, \text{ a.e. } \omega \in \Omega. \quad (6.141)$$

Formulation (6.141) 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 (6.141) can be written in the form

$$\text{Max}_{\lambda \in \mathfrak{M}^*} \left\{ \inf_{(\chi, x) \in \mathfrak{M}_X \times \mathbb{R}^n} L(\chi, x, \lambda) \right\}, \quad (6.142)$$

where

$$L(\chi, x, \lambda) := \rho(F_\chi) + \mathbb{E}[\lambda^\top (\chi - x)], \quad (\chi, x, \lambda) \in \mathfrak{M} \times \mathbb{R}^n \times \mathfrak{M}^*. \quad (6.143)$$

Note that

$$\inf_{x \in \mathbb{R}^n} L(\chi, x, \lambda) = \begin{cases} L(\chi, 0, \lambda) & \text{if } \mathbb{E}[\lambda] = 0, \\ -\infty & \text{if } \mathbb{E}[\lambda] \neq 0. \end{cases}$$

Therefore the dual problem (6.143) can be rewritten in the form

$$\text{Max}_{\lambda \in \mathfrak{M}^*} \left\{ \inf_{\chi \in \mathfrak{M}_X} L_0(\chi, \lambda) \right\} \quad \text{s.t. } \mathbb{E}[\lambda] = 0, \quad (6.144)$$

where  $L_0(\chi, \lambda) := L(\chi, 0, \lambda) = \rho(F_\chi) + \mathbb{E}[\lambda^\top \chi]$ .

We have that the optimal value of problem (6.141) (which is the same as the optimal value of problem (6.128)) is greater than or equal to the optimal value of its dual (6.144). 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 (6.141) and (6.144), and  $(\bar{\chi}, \bar{x})$  and  $\bar{\lambda}$  are optimal solutions of problems

(6.141) and (6.144), respectively. Noting that  $L(\chi, 0, \lambda)$  is linear in  $x$  and in  $\lambda$ , we have that  $((\bar{\chi}, \bar{x}), \bar{\lambda})$  is a saddle point of  $L(\chi, x, \lambda)$  iff the following conditions hold:

$$\begin{aligned} \bar{\chi}(\omega) &= \bar{x}, \text{ a.e. } \omega \in \Omega, \text{ and } \mathbb{E}[\bar{\lambda}] = 0, \\ \bar{\chi} &\in \arg \min_{\chi \in \mathfrak{M}_X} L_0(\chi, \bar{\lambda}). \end{aligned} \quad (6.145)$$

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

$$\text{Min}_{(\chi, x) \in \mathfrak{M}_X \times \mathbb{R}^n} \rho(F_\chi) \text{ s.t. } \chi(\omega) = x + y(\omega), \quad (6.146)$$

and let  $\vartheta(y)$  be its optimal value. Note that a perturbation in the vector  $x$ , in the constraints of problem (6.141), can be absorbed into  $y(\omega)$ . Clearly for  $y = 0$ , problem (6.146) coincides with the unperturbed problem (6.141), and  $\vartheta(0)$  is the optimal value of the unperturbed problem (6.141). Assume that  $\vartheta(0)$  is finite. Then there is no duality gap between problem (6.141) and its dual (6.142) iff  $\vartheta(y)$  is lower semicontinuous at  $y = 0$ . Again it may be not easy to verify lower semicontinuity of the optimal value function  $\vartheta : \mathfrak{M} \rightarrow \mathbb{R}$ . By the general theory of conjugate duality we have the following result.

**Proposition 6.36.** *Suppose that  $F : \mathbb{R}^n \rightarrow \mathcal{Z}$  is convex,  $\rho : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  satisfies conditions (R1)–(R2) and the function  $\rho(F_\chi)$ , from  $\mathfrak{M}$  to  $\overline{\mathbb{R}}$ , 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 (6.141) and (6.142), and the dual problem (6.142) has an optimal solution.*

**Proof.** Since  $\rho$  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 7.77).  $\square$

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  $\Omega$  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 (6.141) is finite, and the assumption that  $\vartheta(y) < +\infty$  means that the corresponding problem (6.146) has a feasible solution.

### Interchangeability Principle for Risk Measures

By removing the nonanticipativity constraint  $\chi(\cdot) = x$ , we obtain the following relaxation of the problem (6.141):

$$\text{Min}_{\chi \in \mathfrak{M}_X} \rho(F_\chi), \quad (6.147)$$



where  $\mathfrak{M}_X$  is defined in (6.140). Similarly to the interchangeability principle for the expectation operator (Theorem 7.80), we have the following result for monotone risk measures. By  $\inf_{x \in X} F(x)$  we denote the pointwise minimum, i.e.,

$$\left[ \inf_{x \in X} F(x) \right] (\omega) := \inf_{x \in X} f(x, \omega), \quad \omega \in \Omega. \quad (6.148)$$

**Proposition 6.37.** *Let  $\mathcal{Z} := \mathcal{L}_p(\Omega, \mathcal{F}, P)$  and  $\mathfrak{M} := \mathcal{L}_{p'}(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ , where  $p, p' \in [1, +\infty]$ ,  $\mathfrak{M}_X$  be defined in (6.140),  $\rho : \mathcal{Z} \rightarrow \bar{\mathbb{R}}$  be a proper risk measure satisfying monotonicity condition (R2), and  $F : \mathbb{R}^n \rightarrow \mathcal{Z}$  be such that  $\inf_{x \in X} F(x) \in \mathcal{Z}$ . Suppose that  $\rho$  is continuous at  $\Psi := \inf_{x \in X} F(x)$ . Then*

$$\inf_{\chi \in \mathfrak{M}_X} \rho(F_\chi) = \rho \left( \inf_{x \in X} F(x) \right). \quad (6.149)$$

**Proof.** For any  $\chi \in \mathfrak{M}_X$  we have that  $\chi(\cdot) \in X$ , and hence the following inequality holds:

$$\left[ \inf_{x \in X} F(x) \right] (\omega) \leq F_\chi(\omega) \quad \text{a.e. } \omega \in \Omega.$$

By monotonicity of  $\rho$  this implies that  $\rho(\Psi) \leq \rho(F_\chi)$ , and hence

$$\rho(\Psi) \leq \inf_{\chi \in \mathfrak{M}_X} \rho(F_\chi). \quad (6.150)$$

Since  $\rho$  is proper we have that  $\rho(\Psi) > -\infty$ . If  $\rho(\Psi) = +\infty$ , then by (6.150) the left-hand side of (6.149) is also  $+\infty$  and hence (6.149) holds. Therefore we can assume that  $\rho(\Psi)$  is finite.

Let us derive now the converse of (6.150) 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  $\chi_k(\omega) \in X$  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  $\chi_k$  belongs to  $\mathfrak{M}_X$ , 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  $F_{\chi_k}$  converges to  $\Psi$  in the norm topology of  $\mathcal{Z}$ . Since  $\rho$  is continuous at  $\Psi$ , it follows that  $\rho(F_{\chi_k})$  tends to  $\rho(\Psi)$ . Also  $\inf_{\chi \in \mathfrak{M}_X} \rho(F_\chi) \leq \rho(F_{\chi_k})$ , and hence the required converse inequality

$$\inf_{\chi \in \mathfrak{M}_X} \rho(F_\chi) \leq \rho(\Psi) \quad (6.151)$$

follows.  $\square$

**Remark 22.** It follows from (6.149) that if

$$\bar{\chi} \in \arg \min_{\chi \in \mathfrak{M}_X} \rho(F_\chi), \quad (6.152)$$

then

$$\bar{\chi}(\omega) \in \arg \min_{x \in X} f(x, \omega) \quad \text{a.e. } \omega \in \Omega. \quad (6.153)$$

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 (6.153) can be chosen to be measurable. If, moreover,  $\bar{\chi} \in \mathfrak{M}$  (this holds, in particular, if the set  $X$  is bounded and hence  $\bar{\chi}(\cdot)$  is bounded), then the inclusion (6.152) 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

$$\text{Min}_{x \in X} \rho(F(x)) \quad (6.154)$$

is given by the optimal value of the second-stage problem

$$\text{Min}_{y \in \mathcal{G}(x, \omega)} g(x, y, \omega), \quad (6.155)$$

where  $g : \mathbb{R}^n \times \mathbb{R}^m \times \Omega \rightarrow \mathbb{R}$  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 (6.155) to obtain

$$\rho(F(x)) = \inf_{y(\cdot) \in \mathcal{G}(x, \cdot)} \rho(g(x, y(\omega), \omega)), \quad (6.156)$$

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 (6.156) holds, then the two-stage problem (6.154)–(6.155) can be written as one large optimization problem:

$$\text{Min}_{x \in X, y(\cdot) \in \mathcal{G}(x, \cdot)} \rho(g(x, y(\omega), \omega)). \quad (6.157)$$

In particular, suppose that the set  $\Omega$  is finite, say  $\Omega = \{\omega_1, \dots, \omega_K\}$ , i.e., there is a finite number  $K$  of scenarios. In that case we can view function  $Z : \Omega \rightarrow \mathbb{R}$  as vector  $(Z(\omega_1), \dots, Z(\omega_K)) \in \mathbb{R}^K$  and hence identify the space  $\mathcal{Z}$  with  $\mathbb{R}^K$ . Then problem (6.157) takes the form

$$\text{Min}_{x \in X, y_k \in \mathcal{G}(x, \omega_k), k=1, \dots, K} \rho[(g(x, y_1, \omega_1), \dots, g(x, y_K, \omega_K))]. \quad (6.158)$$

Moreover, consider the linear case where  $X := \{x : Ax = b, x \geq 0\}$ ,  $g(x, y, \omega) := c^\top x + q(\omega)^\top y$

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

Assume that  $\rho$  satisfies conditions (R1)–(R3) and the set  $\Omega = \{\omega_1, \dots, \omega_K\}$  is finite. Then problem (6.158) takes the form

$$\begin{aligned} \text{Min}_{x, y_1, \dots, y_K} \quad & c^\top x + \rho[(q_1^\top y_1, \dots, q_K^\top y_K)] \\ \text{s.t.} \quad & Ax = b, x \geq 0, T_k x + W_k y_k = h_k, y_k \geq 0, k = 1, \dots, K, \end{aligned} \quad (6.159)$$

where  $(q_k, T_k, W_k, h_k) := (q(\omega_k), T(\omega_k), W(\omega_k), h(\omega_k)), k = 1, \dots, K$ .

### 6.4.2 Examples

Let  $\mathcal{Z} := \mathcal{L}_1(\Omega, \mathcal{F}, P)$  and consider

$$\rho(Z) := \mathbb{E}[Z] + \inf_{t \in \mathbb{R}} \mathbb{E}\{\beta_1[t - Z]_+ + \beta_2[Z - t]_+\}, \quad Z \in \mathcal{Z}, \quad (6.160)$$

where  $\beta_1 \in [0, 1]$  and  $\beta_2 \geq 0$  are some constants. Properties of this risk measure were studied in Example 6.16 (see (6.67) and (6.68) in particular). We can write the corresponding optimization problem (6.128) in the following equivalent form:

$$\text{Min}_{(x, t) \in X \times \mathbb{R}} \mathbb{E}\{f_\omega(x) + \beta_1[t - f_\omega(x)]_+ + \beta_2[f_\omega(x) - t]_+\}. \quad (6.161)$$

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, backordering, and holding, respectively. Again we assume that  $b > c > 0$ , i.e., the backorder 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

$$\text{Min}_{x \geq 0} \{f(x) := \rho[F(x, D)]\}, \quad (6.162)$$

where  $\rho$  is a specified risk measure.

Assume that the risk measure  $\rho$  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  $\rho : \mathcal{Z} \rightarrow \mathbb{R}$  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

$$\rho(Z) = \sup_{\zeta \in \mathfrak{A}} \int_{\Omega} Z(\omega) \zeta(\omega) dP(\omega), \quad Z \in \mathcal{Z}.$$

Consequently we have that

$$\rho[F(x, D)] = \sup_{\zeta \in \mathfrak{A}} \int_{\Omega} F(x, D(\omega)) \zeta(\omega) dP(\omega). \quad (6.163)$$

To each  $\zeta \in \mathfrak{P}$  corresponds the cumulative distribution function  $H$  of  $D$  with respect to the measure  $Q := \zeta dP$ , that is,

$$H(z) = Q(D \leq z) = \mathbb{E}_{\zeta}[\mathbf{1}_{D \leq z}] = \int_{\{\omega: D(\omega) \leq z\}} \zeta(\omega) dP(\omega). \quad (6.164)$$

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  $\zeta \in \mathfrak{A}$ . The correspondence between  $\zeta \in \mathfrak{A}$  and  $H \in \mathfrak{M}$  is given by formula (6.164) and depends on  $D(\cdot)$  and the reference probability measure  $P$ . Then we can rewrite (6.163) in the form

$$\rho[F(x, D)] = \sup_{H \in \mathfrak{M}} \int F(x, z) dH(z) = \sup_{H \in \mathfrak{M}} \mathbb{E}_H[F(x, D)]. \quad (6.165)$$

This leads to the following minimax formulation of the risk averse optimization problem (6.162):

$$\text{Min}_{x \geq 0} \sup_{H \in \mathfrak{M}} \mathbb{E}_H[F(x, D)]. \quad (6.166)$$

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 (6.166), rather than the risk averse formulation (6.162), viewing  $\mathfrak{M}$  as a given set of cumulative distribution functions. We show next that the minimax problem (6.166), and hence the risk averse problem (6.162), 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 6.38.** *Let  $\mathfrak{M}$  be a set of cumulative distribution functions such that  $H(z) = 0$  for 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  $\bar{H}$ , depending on the set  $\mathfrak{M}$  and  $\eta := b/(b+h)$ , such that  $\bar{H}(z) = 0$  for any  $z < 0$ , and the function  $f(x)$  can be written in the form*

$$f(x) = b \sup_{H \in \mathfrak{M}} \mathbb{E}_H[D] + (c-b)x + (b+h) \int_{-\infty}^x \bar{H}(z) dz. \quad (6.167)$$

**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  $f(x) = (c-b)x + (b+h)g(x)$ , where

$$g(x) := \sup_{H \in \mathfrak{M}} \left\{ \eta \mathbb{E}_H[D] + \int_{-\infty}^x H(z) dz \right\}. \quad (6.168)$$

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  $g(x)$  is given by the maximum of convex functions and hence is convex. Moreover,  $g(x) \geq 0$  and

$$g(x) \leq \eta \sup_{H \in \mathfrak{M}} \mathbb{E}_H[D] + [x]_+, \quad (6.169)$$

and hence  $g(x)$  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  $g(x) = \eta \sup_{H \in \mathfrak{M}} \mathbb{E}_H[D]$  for any  $x < 0$ .

Consider the right-hand-side derivative of  $g(x)$ :

$$g^+(x) := \lim_{t \downarrow 0} \frac{g(x+t) - g(x)}{t},$$

and define  $\bar{H}(\cdot) := g^+(\cdot)$ . Since  $g(x)$  is real valued convex, its right-hand-side derivative  $g^+(x)$  exists and is finite, and for any  $x \geq 0$  and  $a < 0$ ,

$$g(x) = g(a) + \int_a^x g^+(z) dz = \eta \sup_{H \in \mathfrak{M}} \mathbb{E}_H[D] + \int_{-\infty}^x \bar{H}(z) dz. \quad (6.170)$$

Note that definition of the function  $g(\cdot)$ , and hence  $\bar{H}(\cdot)$ , involves the constant  $\eta$  and set  $\mathfrak{M}$  only. Let us also observe that the right-hand-side derivative  $g^+(x)$ , of a real valued convex function, is monotonically nondecreasing and right-side continuous. Moreover,  $g^+(x) = 0$  for  $x < 0$  since  $g(x)$  is constant for  $x < 0$ . We also have that  $g^+(x)$  tends to one as  $x \rightarrow +\infty$ . Indeed, since  $g^+(x)$  is monotonically nondecreasing it tends to a limit, denoted  $r$ , as  $x \rightarrow +\infty$ . We have then that  $g(x)/x \rightarrow r$  as  $x \rightarrow +\infty$ . It follows from (6.169) that  $r \leq 1$ , and by (6.168) that for any  $H \in \mathfrak{M}$ ,

$$\liminf_{x \rightarrow +\infty} \frac{g(x)}{x} \geq \liminf_{x \rightarrow +\infty} \frac{1}{x} \int_{-\infty}^x H(z) dz \geq 1,$$

and hence  $r \geq 1$ . It follows that  $r = 1$ .

We obtain that  $\bar{H}(\cdot) = g^+(\cdot)$  is a cumulative distribution function of some probability distribution and the representation (6.167) holds.  $\square$

It follows from the representation (6.167) that the set of optimal solutions of the risk averse problem (6.162) is an interval given by the set of  $\kappa$ -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  $\bar{H}$  in a closed form. Consider the risk measure  $\rho$  defined in (6.160),

$$\rho(Z) := \mathbb{E}[Z] + \inf_{t \in \mathbb{R}} \mathbb{E}\{\beta_1[t - Z]_+ + \beta_2[Z - t]_+\},$$

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

$$(1 - \beta_1) \int_S dH^* \leq \int_S dH \leq (1 + \beta_2) \int_S dH^* \quad (6.171)$$

for any Borel set  $S \subset \mathbb{R}$ . (Compare with formula (6.69).) Recall that for  $\beta_1 = 1$  this risk measure is  $\rho(Z) = \text{AV@R}_\alpha(Z)$  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

$$\mathbb{E}_H[D] = \int_0^1 z dH(z) = zH(z)|_0^1 - \int_0^1 H(z) dz = 1 - \int_0^1 H(z) dz.$$

Consequently we can write function  $g(x)$ , defined in (6.168), for  $x \in [0, 1]$  in the form

$$g(x) = \eta + \sup_{H \in \mathfrak{M}} \left\{ (1 - \eta) \int_0^x H(z) dz - \eta \int_x^1 H(z) dz \right\}. \quad (6.172)$$

Suppose, further, that  $h = 0$  (i.e., there are no holding costs) and hence  $\eta = 1$ . In that case

$$g(x) = 1 - \inf_{H \in \mathfrak{M}} \int_x^1 H(z) dz \quad \text{for } x \in [0, 1]. \quad (6.173)$$

By using the first inequality of (6.171) with  $S := [0, z]$  we obtain that  $H(z) \geq (1 - \beta_1)z$  for any  $H \in \mathfrak{M}$  and  $z \in [0, 1]$ . Similarly, by the second inequality of (6.171) with  $S := [z, 1]$  we have that  $H(z) \geq 1 + (1 + \beta_2)(z - 1)$  for any  $H \in \mathfrak{M}$  and  $z \in [0, 1]$ . Consequently, the cdf

$$\bar{H}(z) := \max\{(1 - \beta_1)z, (1 + \beta_2)z - \beta_2\}, \quad z \in [0, 1], \quad (6.174)$$

is dominated by any other cdf  $H \in \mathfrak{M}$ , and it can be verified that  $\bar{H} \in \mathfrak{M}$ . Therefore, the minimum on the right-hand side of (6.173) is attained at  $\bar{H}$  for any  $x \in [0, 1]$ , and hence this cdf  $\bar{H}$  fulfills (6.167).

Note that for any  $\beta_1 \in (0, 1)$  and  $\beta_2 > 0$ , the cdf  $\bar{H}(\cdot)$  defined in (6.174) is strictly less than the reference cdf  $H^*(\cdot)$  on the interval  $(0, 1)$ . Consequently, the corresponding risk averse optimal solution  $\bar{H}^{-1}(\kappa)$  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

$$\text{Min}_{x \in X} \rho\left(-\sum_{i=1}^n \xi_i x_i\right), \quad (6.175)$$

where  $\rho$  is a chosen risk measure and  $X := \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = W_0, x \geq 0\}$ . 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 a possible problem with using risk measures with dispersions measured by variance or standard deviation.

**Example 6.39.** Let  $n = 2$ ,  $W_0 = 1$  and the risk measure  $\rho$  be of the form

$$\rho(Z) := \mathbb{E}[Z] + c \mathbb{D}[Z], \quad (6.176)$$

where  $c > 0$  and  $\mathbb{D}[\cdot]$  is a dispersion measure. Let the dispersion measure be either  $\mathbb{D}[Z] := \sqrt{\text{Var}[Z]}$  or  $\mathbb{D}[Z] := \text{Var}[Z]$ . 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  $\xi_1, \xi_2 : \Omega \rightarrow \mathbb{R}$  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, it is better to

invest in asset 1 than asset 2. Now, for  $\mathbb{D}[Z] := \sqrt{\text{Var}[Z]}$ , 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  $\mathbb{D}[Z] := \text{Var}[Z]$  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  $\xi_2$  dominates  $\xi_1$  in the sense that  $\xi_1(\omega) \geq \xi_2(\omega)$  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 6.18 and 6.19.) ■

Suppose now that  $\rho$  is a real valued coherent risk measure. We can then write problem (6.175) in the corresponding min-max form (6.131), that is,

$$\text{Min sup}_{x \in X} \sum_{\zeta \in \mathfrak{A}}^n (-\mathbb{E}_{\zeta}[\xi_i]) x_i.$$

Equivalently,

$$\text{Max inf}_{x \in X} \sum_{\zeta \in \mathfrak{A}}^n (\mathbb{E}_{\zeta}[\xi_i]) x_i. \quad (6.177)$$

Since the feasible set  $X$  is compact, problem (6.175) always has an optimal solution  $\bar{x}$ . Also (see Proposition 6.33), the min-max problem (6.177) has a saddle point, and  $(\bar{x}, \bar{\zeta})$  is a saddle point iff

$$\bar{\zeta} \in \partial\rho(\bar{Z}) \text{ and } \bar{x} \in \arg \max_{x \in X} \sum_{i=1}^n \bar{\mu}_i x_i, \quad (6.178)$$

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  $\zeta$  represents the mixed strategy of the opponent (the market). It is chosen not from the set of all possible mixed strategies but rather from the set  $\mathfrak{A}$ . The risk averse solution (6.178) corresponds to the equilibrium of the game.

It is not difficult to see that the set  $\arg \max_{x \in X} \sum_{i=1}^n \bar{\mu}_i x_i$  is formed by all convex combinations of vectors  $W_0 e_i$ ,  $i \in \mathcal{I}$ , where  $e_i \in \mathbb{R}^n$  denotes the  $i$ th coordinate vector (with zero entries except the  $i$ th entry equal to 1), and

$$\mathcal{I} := \{i' : \bar{\mu}_{i'} = \max_{1 \leq i \leq n} \bar{\mu}_i, i' = 1, \dots, n\}.$$

Also  $\partial\rho(Z) \subset \mathfrak{A}$ ; see formula (6.43) for the subdifferential  $\partial\rho(Z)$ .

## 6.5 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.

### 6.5.1 Average Value-at-Risk

Recall that the Average Value-at-Risk,  $\text{AV@R}_\alpha(Z)$ , at a level  $\alpha \in (0, 1)$  of a random variable  $Z$ , is given by the optimal value of the minimization problem

$$\text{Min}_{t \in \mathbb{R}} \mathbb{E} \{ t + \alpha^{-1} [Z - t]_+ \}, \quad (6.179)$$

where the expectation is taken with respect to the probability distribution  $P$  of  $Z$ . We assume that  $\mathbb{E}|Z| < +\infty$ , which implies that  $\text{AV@R}_\alpha(Z)$  is finite. Suppose now that we have an iid random sample  $Z^1, \dots, Z^N$  of  $N$  realizations of  $Z$ . Then we can estimate  $\theta^* := \text{AV@R}_\alpha(Z)$  by replacing distribution  $P$  with its empirical estimate<sup>48</sup>  $P_N := \frac{1}{N} \sum_{j=1}^N \Delta(Z^j)$ . This leads to the sample estimate  $\hat{\theta}_N$ , of  $\theta^* = \text{AV@R}_\alpha(Z)$ , given by the optimal value of the following problem:

$$\text{Min}_{t \in \mathbb{R}} \left\{ t + \frac{1}{\alpha N} \sum_{j=1}^N [Z^j - t]_+ \right\}. \quad (6.180)$$

Let us observe that problem (6.179) can be viewed as a stochastic programming problem and problem (6.180) as its sample average approximation. That is,

$$\theta^* = \inf_{t \in \mathbb{R}} f(t) \quad \text{and} \quad \hat{\theta}_N = \inf_{t \in \mathbb{R}} \hat{f}_N(t),$$

where

$$f(t) = t + \alpha^{-1} \mathbb{E}[Z - t]_+ \quad \text{and} \quad \hat{f}_N(t) = t + \frac{1}{\alpha N} \sum_{j=1}^N [Z^j - t]_+.$$

Therefore, results of section 5.1 can be applied here in a straightforward way. Recall that the set of optimal solutions of problem (6.179) is the interval  $[t^*, t^{**}]$ , where

$$t^* = \inf \{ z : H_Z(z) \geq 1 - \alpha \} = \text{V@R}_\alpha(Z) \quad \text{and} \quad t^{**} = \sup \{ z : H_Z(z) \leq 1 - \alpha \}$$

are the respective left- and right-side  $(1 - \alpha)$ -quantiles of the distribution of  $Z$  (see page 258). Since for any  $\alpha \in (0, 1)$  the interval  $[t^*, t^{**}]$  is finite and problem (6.179) is convex, we have by Theorem 5.4 that

$$\hat{\theta}_N \rightarrow \theta^* \quad \text{w.p. 1 as } N \rightarrow \infty. \quad (6.181)$$

That is,  $\hat{\theta}_N$  is a consistent estimator of  $\theta^* = \text{AV@R}_\alpha(Z)$ .

<sup>48</sup>Recall that  $\Delta(z)$  denotes measure of mass one at point  $z$ .



Assume now that  $\mathbb{E}[Z^2] < +\infty$ . Then the assumptions (A1) and (A2) of Theorem 5.7 hold, and hence

$$\hat{\theta}_N = \inf_{t \in [t^*, t^{**}]} \hat{f}_N(t) + o_p(N^{-1/2}). \quad (6.182)$$

Moreover, if  $t^* = t^{**}$ , i.e., the left- and right-side  $(1 - \alpha)$ -quantiles of the distribution of  $Z$  are the same, then

$$N^{1/2} (\hat{\theta}_N - \theta^*) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2), \quad (6.183)$$

where  $\sigma^2 = \alpha^{-2} \mathbb{V}\text{ar}([Z - t^*]_+)$ .

The estimator  $\hat{\theta}_N$  has a negative bias, i.e.,  $\mathbb{E}[\hat{\theta}_N] - \theta^* \leq 0$ , and (see Proposition 5.6)

$$\mathbb{E}[\hat{\theta}_N] \leq \mathbb{E}[\hat{\theta}_{N+1}], \quad N = 1, \dots, \quad (6.184)$$

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  $f(t)$  here are  $f'(t) = 1 + \alpha^{-1}(H_Z(t) - 1)$ , provided that the cumulative distribution function  $H_Z(\cdot)$  is continuous at  $t$ , and  $f''(t) = \alpha^{-1}h_Z(t)$ , 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 168), under appropriate regularity conditions, in particular if  $t^* = t^{**} = \mathbb{V}@\mathbb{R}_\alpha(Z)$  and the density  $h_Z(t^*) = \partial H_Z(t^*)/\partial t$  exists and  $h_Z(t^*) \neq 0$ , that

$$\begin{aligned} \hat{\theta}_N - \hat{f}_N(t^*) &= N^{-1} \inf_{\tau \in \mathbb{R}} \left\{ \tau Z + \frac{1}{2} \tau^2 f''(t^*) \right\} + o_p(N^{-1}) \\ &= -\frac{\alpha Z^2}{2N h_Z(t^*)} + o_p(N^{-1}), \end{aligned} \quad (6.185)$$

where  $Z \sim \mathcal{N}(0, \gamma^2)$  with

$$\gamma^2 = \mathbb{V}\text{ar} \left( \alpha^{-1} \frac{\partial [Z - t^*]_+}{\partial t} \right) = \frac{H_Z(t^*)(1 - H_Z(t^*))}{\alpha^2} = \frac{1 - \alpha}{\alpha}.$$

Consequently, under appropriate regularity conditions,

$$N \left[ \hat{\theta}_N - \hat{f}_N(t^*) \right] \xrightarrow{\mathcal{D}} - \left[ \frac{1 - \alpha}{2h_Z(t^*)} \right] \chi_1^2 \quad (6.186)$$

and (see Remark 32 on page 382)

$$\mathbb{E}[\hat{\theta}_N] - \theta^* = -\frac{1 - \alpha}{2N h_Z(t^*)} + o(N^{-1}). \quad (6.187)$$

### 6.5.2 Absolute Semideviation Risk Measure

Consider the mean absolute semideviation risk measure

$$\rho_c(Z) := \mathbb{E} \{ Z + c[Z - \mathbb{E}(Z)]_+ \}, \quad (6.188)$$

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

$Z^1, \dots, Z^N$  of  $Z$ , the corresponding estimator of  $\theta^* := \rho_c(Z)$  is

$$\hat{\theta}_N = N^{-1} \sum_{j=1}^N (Z^j + c[Z^j - \bar{Z}]_+), \quad (6.189)$$

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

$$\text{Min}_{t \in \mathbb{R}} \max_{\gamma \in [0,1]} \mathbb{E}[F(t, \gamma, Z)], \quad (6.190)$$

where

$$\begin{aligned} F(t, \gamma, z) &:= z + c\gamma[z - t]_+ + c(1 - \gamma)[t - z]_+ \\ &= z + c[z - t]_+ + c(1 - \gamma)(z - t). \end{aligned} \quad (6.191)$$

This follows by virtue of Corollary 6.3. More directly we can argue as follows. Denote  $\mu := \mathbb{E}[Z]$ . We have that

$$\begin{aligned} \sup_{\gamma \in [0,1]} \mathbb{E}\{Z + c\gamma[Z - t]_+ + c(1 - \gamma)[t - Z]_+\} \\ = \mathbb{E}[Z] + c \max\{\mathbb{E}([Z - t]_+), \mathbb{E}([t - Z]_+)\}. \end{aligned}$$

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 (6.190) is attained at unique point  $t^* = \mu$ . It also follows that the set of saddle points of the minimax problem (6.190) is given by  $\{\mu\} \times [\gamma^*, \gamma^{**}]$ , where

$$\gamma^* = \Pr(Z < \mu) \text{ and } \gamma^{**} = \Pr(Z \leq \mu) = H_Z(\mu). \quad (6.192)$$

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

$$\text{Min}_{t \in \mathbb{R}} \max_{\gamma \in [0,1]} N^{-1} \sum_{j=1}^N F(t, \gamma, Z^j). \quad (6.193)$$

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  $\theta^*$  as  $N \rightarrow \infty$ . Moreover, assuming that  $\mathbb{E}[Z^2] < +\infty$  we have by Theorem 5.10 that

$$\begin{aligned} \hat{\theta}_N &= \max_{\gamma \in [\gamma^*, \gamma^{**}]} N^{-1} \sum_{j=1}^N F(\mu, \gamma, Z^j) + o_p(N^{-1/2}) \\ &= \bar{Z} + cN^{-1} \sum_{j=1}^N [Z^j - \mu]_+ + c\Psi(\bar{Z} - \mu) + o_p(N^{-1/2}), \end{aligned} \quad (6.194)$$

where  $\bar{Z} = N^{-1} \sum_{j=1}^N Z^j$  and function  $\Psi(\cdot)$  is defined as

$$\Psi(z) := \begin{cases} (1 - \gamma^*)z & \text{if } z > 0, \\ (1 - \gamma^{**})z & \text{if } z \leq 0. \end{cases}$$

If, moreover, the cdf  $H_Z(\cdot)$  is continuous at  $\mu$ , and hence  $\gamma^* = \gamma^{**} = H_Z(\mu)$ , then

$$N^{1/2}(\hat{\theta}_N - \theta^*) \xrightarrow{\mathcal{D}} N(0, \sigma^2), \quad (6.195)$$

where  $\sigma^2 = \text{Var}[F(\mu, H_Z(\mu), Z)]$ .

This analysis can be extended to risk averse optimization problems of the form (6.128). That is, consider problem

$$\text{Min}_{x \in X} \left\{ \rho_c[G(x, \xi)] = \mathbb{E}\{G(x, \xi) + c[G(x, \xi) - \mathbb{E}(G(x, \xi))]_+\} \right\}, \quad (6.196)$$

where  $X \subset \mathbb{R}^n$  and  $G : X \times \Xi \rightarrow \mathbb{R}$ . Its SAA is obtained by replacing the true distribution of the random vector  $\xi$  with the empirical distribution associated with a random sample  $\xi^1, \dots, \xi^N$ , that is,

$$\text{Min}_{x \in X} \frac{1}{N} \sum_{j=1}^N \left\{ G(x, \xi^j) + c \left[ G(x, \xi^j) - \frac{1}{N} \sum_{j=1}^N G(x, \xi^j) \right]_+ \right\}. \quad (6.197)$$

Assume that the set  $X$  is convex compact and function  $G(\cdot, \xi)$  is convex for a.e.  $\xi$ . Then, for  $c \in [0, 1]$ , problems (6.196) and (6.197) are convex. By using the min-max representation (6.190), problem (6.196) can be written as the minimax problem

$$\text{Min}_{(x, t) \in X \times \mathbb{R}} \max_{\gamma \in [0, 1]} \mathbb{E}[F(t, \gamma, G(x, \xi))], \quad (6.198)$$

where function  $F(t, \gamma, z)$  is defined in (6.191). 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 (6.198) is a convex-concave minimax problem. Consequently, results of section 5.1.4 can be applied.

Let  $\vartheta^*$  and  $\hat{\vartheta}_N$  be the optimal values of the true problem (6.196) and the SAA problem (6.197), respectively, and  $S$  be the set of optimal solutions of the true problem (6.196). By Theorem 5.10 and the above analysis we obtain, assuming that conditions specified in Theorem 5.10 are satisfied, that

$$\hat{\vartheta}_N = N^{-1} \inf_{\substack{x \in S \\ t = \mathbb{E}[G(x, \xi)]}} \max_{\gamma \in [\gamma^*, \gamma^{**}]} \left\{ \sum_{j=1}^N F(t, \gamma, G(x, \xi^j)) \right\} + o_p(N^{-1/2}), \quad (6.199)$$

where

$$\gamma^* := \Pr\{G(x, \xi) < \mathbb{E}[G(x, \xi)]\} \quad \text{and} \quad \gamma^{**} := \Pr\{G(x, \xi) \leq \mathbb{E}[G(x, \xi)]\}, \quad x \in S.$$

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 (6.198), 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 (6.196) 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

$$\text{Var}\{G(\bar{x}, \xi) + c[G(\bar{x}, \xi) - \mu]_+ + c(1 - \gamma^*)(G(\bar{x}, \xi) - \mu)\}.$$

### 6.5.3 Von Mises Statistical Functionals

In the two examples, of  $AV@R_\alpha$  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 6.24 (such approach was developed in [147]). 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  $\mathfrak{F}(P)$  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

$$\mathfrak{F}(P) := \left( \mathbb{E}_P \left[ (Z - \mathbb{E}_P[Z])_+^p \right] \right)^{1/p}. \quad (6.200)$$

The sample estimate of  $\mathfrak{F}(P)$  is obtained by replacing probability measure  $P$  with the empirical measure  $P_N$ . That is, we estimate  $\theta^* = \mathfrak{F}(P)$  by  $\hat{\theta}_N = \mathfrak{F}(P_N)$ .

Let  $Q$  be an arbitrary probability measure, defined on the same probability space as  $P$ , and consider the convex combination  $(1-t)P + tQ = P + t(Q - P)$ , with  $t \in [0, 1]$ , of  $P$  and  $Q$ . Suppose that the following limit exists:

$$\mathfrak{F}'(P, Q - P) := \lim_{t \downarrow 0} \frac{\mathfrak{F}(P + t(Q - P)) - \mathfrak{F}(P)}{t}. \quad (6.201)$$

The above limit is just the directional derivative of  $\mathfrak{F}(\cdot)$  at  $P$  in the direction  $Q - P$ . If, moreover, the directional derivative  $\mathfrak{F}'(P, \cdot)$  is linear, then  $\mathfrak{F}(\cdot)$  is Gâteaux differentiable at  $P$ . Consider now the approximation

$$\mathfrak{F}(P_N) - \mathfrak{F}(P) \approx \mathfrak{F}'(P, P_N - P). \quad (6.202)$$

By this approximation,

$$N^{1/2}(\hat{\theta}_N - \theta^*) \approx \mathfrak{F}'(P, N^{1/2}(P_N - P)), \quad (6.203)$$

and we can use  $\mathfrak{F}'(P, N^{1/2}(P_N - P))$  to derive asymptotics of  $N^{1/2}(\hat{\theta}_N - \theta^*)$ .

Suppose, further, that  $\mathfrak{F}'(P, \cdot)$  is linear, i.e.,  $\mathfrak{F}(\cdot)$  is Gâteaux differentiable at  $P$ . Then, since  $P_N = N^{-1} \sum_{j=1}^N \Delta(Z^j)$ , we have that

$$\mathfrak{F}'(P, P_N - P) = \frac{1}{N} \sum_{j=1}^N IF_{\mathfrak{F}}(Z^j), \quad (6.204)$$

where

$$IF_{\mathfrak{F}}(z) := \sum_{j=1}^N \mathfrak{F}'(P, \Delta(z) - P) \quad (6.205)$$

is the so-called *influence function* (also called influence curve) of  $\mathfrak{F}$ .

It follows from the linearity of  $\mathfrak{F}'(P, \cdot)$  that  $\mathbb{E}_P[IF_{\mathfrak{F}}(Z)] = 0$ . Indeed, linearity of  $\mathfrak{F}'(P, \cdot)$  means that it is a linear functional and hence can be represented as

$$\mathfrak{F}'(P, Q - P) = \int g d(Q - P) = \int g dQ - \mathbb{E}_P[g(Z)]$$

for some function  $g$  in an appropriate functional space. Consequently,  $IF_{\mathfrak{F}}(z) = g(z) - \mathbb{E}_P[g(Z)]$ , and hence

$$\mathbb{E}_P[IF_{\mathfrak{F}}(Z)] = \mathbb{E}_P\{g(Z) - \mathbb{E}_P[g(Z)]\} = 0.$$

Then by the CLT we have that  $N^{-1/2} \sum_{j=1}^N IF_{\mathfrak{F}}(Z^j)$  converges in distribution to normal with zero mean and variance  $\mathbb{E}_P[IF_{\mathfrak{F}}(Z)^2]$ . This suggests the following asymptotics:

$$N^{1/2}(\hat{\theta}_N - \theta^*) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbb{E}_P[IF_{\mathfrak{F}}(Z)^2]). \quad (6.206)$$

It should be mentioned at this point that the above derivations *do not* prove in a rigorous way validity of the asymptotics (6.206). The main technical difficulty is to give a rigorous justification for the approximation (6.203) leading to the corresponding convergence in distribution. This can be compared with the Delta method, discussed in section 7.2.7 and applied in section 5.1, where first (and second) order approximations were derived in functional spaces rather than spaces of measures. Anyway, formula (6.206) gives correct asymptotics and is routinely used in statistical applications.

Let us consider, for example, the statical functional

$$\mathfrak{F}(P) := \mathbb{E}_P[Z - \mathbb{E}_P[Z]]_+, \quad (6.207)$$

associated with  $\sigma_1^+[Z]$ . Denote  $\mu := \mathbb{E}_P[Z]$ . Then

$$\begin{aligned} \mathfrak{F}(P + t(Q - P)) - \mathfrak{F}(P) &= t \left( \mathbb{E}_Q[Z - \mu]_+ - \mathbb{E}_P[Z - \mu]_+ \right) \\ &\quad + \mathbb{E}_P[Z - \mu - t(\mathbb{E}_Q[Z] - \mu)]_+ + o(t). \end{aligned}$$

Moreover, the right-side derivative at  $t = 0$  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

$$\mathfrak{F}'(P, Q - P) = \mathbb{E}_Q[Z - \mu]_+ - \mathbb{E}_P[Z - \mu]_+ + (1 - H_Z(\mu))(\mathbb{E}_Q[Z] - \mu),$$

and hence

$$IF_{\mathfrak{F}}(z) = [z - \mu]_+ - \mathbb{E}_P[Z - \mu]_+ + (1 - H_Z(\mu))(z - \mu). \quad (6.208)$$

It can be seen now that  $\mathbb{E}_P[IF_{\mathfrak{F}}(Z)] = 0$  and

$$\mathbb{E}_P[IF_{\mathfrak{F}}(Z)^2] = \mathbb{V}\text{ar}\{[Z - \mu]_+ + (1 - H_Z(\mu))(Z - \mu)\}.$$

That is, the asymptotics (6.206) here are exactly the same as the ones derived in the previous section 6.5.2 (compare with (6.195)).

In a similar way, it is possible to compute the influence function of the statistical functional defined in (6.200), 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

$$IF_{\mathfrak{F}}(z) = \frac{1}{2\theta^*} \left( [z - \mu]_+^2 - \theta^{*2} + 2\kappa(1 - H_Z(\mu))(z - \mu) \right), \quad (6.209)$$

where  $\theta^* := \mathfrak{F}(P) = (\mathbb{E}_P[Z - \mu]_+^2)^{1/2}$  and  $\kappa := \mathbb{E}_P[Z - \mu]_+ = \frac{1}{2}\mathbb{E}_P|Z - \mu|$ .

## 6.6 The Problem of Moments

Due to the duality representation (6.37) of a coherent risk measure, the corresponding risk averse optimization problem (6.128) can be written as the minimax problem (6.131). 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

$$\text{Min}_{x \in X} \sup_{P \in \mathfrak{M}} \mathbb{E}_P[f(x, \omega)] \quad (6.210)$$

in a direct way for a specified set  $\mathfrak{M}$  of probability measures on a measurable space  $(\Omega, \mathcal{F})$ . 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  $(\Omega, \mathcal{F})$ . We denote by  $\tilde{\mathfrak{P}}$  the set of probability measures<sup>49</sup> on  $(\Omega, \mathcal{F})$  and  $\mathbb{E}_P[f(x, \omega)]$  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. Of course, there are various ways to define the uncertainty set  $\mathfrak{M}$ . In some situations, it is reasonable to assume that we have knowledge about certain moments of the corresponding probability distribution. That is, the set  $\mathfrak{M}$  is defined by moment constraints as follows:

$$\mathfrak{M} := \left\{ P \in \tilde{\mathfrak{P}} : \begin{array}{l} \mathbb{E}_P[\psi_i(\omega)] = b_i, \quad i = 1, \dots, p, \\ \mathbb{E}_P[\psi_i(\omega)] \leq b_i, \quad i = p+1, \dots, q \end{array} \right\}, \quad (6.211)$$

where  $\psi_i : \Omega \rightarrow \mathbb{R}$ ,  $i = 1, \dots, q$ , are measurable functions. Note that the condition  $P \in \tilde{\mathfrak{P}}$ , i.e., that  $P$  is a probability measure, can be formulated explicitly as the constraint<sup>50</sup>  $\int_{\Omega} dP = 1$ ,  $P \geq 0$ .

We assume that every finite subset of  $\Omega$  is  $\mathcal{F}$ -measurable. This is a mild assumption. For example, if  $\Omega$  is a metric space equipped with its Borel sigma algebra, then this certainly holds true. We denote by  $\tilde{\mathfrak{P}}_m^*$  the set of probability measures on  $(\Omega, \mathcal{F})$  having a finite support of at most  $m$  points. That is, every measure  $P \in \tilde{\mathfrak{P}}_m^*$  can be represented in the form  $P = \sum_{i=1}^m \alpha_i \Delta(\omega_i)$ , where  $\alpha_i$  are nonnegative numbers summing up to one and  $\Delta(\omega)$  denotes measure of mass one at the point  $\omega \in \Omega$ . Similarly, we denote  $\mathfrak{M}_m^* := \mathfrak{M} \cap \tilde{\mathfrak{P}}_m^*$ . Note that the set  $\mathfrak{M}$  is convex while, for a fixed  $m$ , the set  $\mathfrak{M}_m^*$  is not necessarily convex. By Theorem 7.32, to any  $P \in \mathfrak{M}$  corresponds a probability measure  $Q \in \tilde{\mathfrak{P}}$  with a finite support of at most  $q+1$  points such that  $\mathbb{E}_P[\psi_i(\omega)] = \mathbb{E}_Q[\psi_i(\omega)]$ ,  $i = 1, \dots, q$ . That is, if the set  $\mathfrak{M}$  is nonempty, then its subset  $\mathfrak{M}_{q+1}^*$  is also nonempty. Consider the function

$$g(x) := \sup_{P \in \mathfrak{M}} \mathbb{E}_P[f(x, \omega)]. \quad (6.212)$$

**Proposition 6.40.** *For any  $x \in X$  we have that*

$$g(x) = \sup_{P \in \mathfrak{M}_{q+1}^*} \mathbb{E}_P[f(x, \omega)]. \quad (6.213)$$

<sup>49</sup>The set  $\tilde{\mathfrak{P}}$  of probability measures should be distinguished from the set  $\mathfrak{P}$  of probability density functions used before.

<sup>50</sup>Recall that the notation  $P \geq 0$  means that  $P$  is a nonnegative (not necessarily probability) measure on  $(\Omega, \mathcal{F})$ .

**Proof.** If the set  $\mathfrak{M}$  is empty, then its subset  $\mathfrak{M}_{q+1}^*$  is also empty, and hence  $g(x)$  as well as the optimal value of the right-hand side of (6.213) are equal to  $+\infty$ . So suppose that  $\mathfrak{M}$  is nonempty. Consider a point  $x \in X$  and  $P \in \mathfrak{M}$ . By Theorem 7.32 there exists  $Q \in \mathfrak{M}_{q+2}^*$  such that  $\mathbb{E}_P[f(x, \omega)] = \mathbb{E}_Q[f(x, \omega)]$ . It follows that  $g(x)$  is equal to the maximum of  $\mathbb{E}_P[f(x, \omega)]$  over  $P \in \mathfrak{M}_{q+2}^*$ , which in turn is equal to the optimal value of the problem

$$\begin{aligned} & \text{Max}_{\substack{\omega_1, \dots, \omega_m \in \Omega \\ \alpha \in \mathbb{R}_+^m}} \sum_{j=1}^m \alpha_j f(x, \omega_j) \\ & \text{s.t. } \sum_{j=1}^m \alpha_j \psi_i(\omega_j) = b_i, \quad i = 1, \dots, p, \\ & \sum_{j=1}^m \alpha_j \psi_i(\omega_j) \leq b_i, \quad i = p+1, \dots, q, \\ & \sum_{j=1}^m \alpha_j = 1, \end{aligned} \tag{6.214}$$

where  $m := q+2$ . For fixed  $\omega_1, \dots, \omega_m \in \Omega$ , the above is a linear programming problem. Its feasible set is bounded and its optimum is attained at an extreme point of its feasible set which has at most  $q+1$  nonzero components of  $\alpha$ . Therefore it suffices to take the maximum over  $P \in \mathfrak{M}_{q+1}^*$ .  $\square$

For a given  $x \in X$ , the (Lagrangian) dual of the problem

$$\text{Max}_{P \in \mathfrak{M}} \mathbb{E}_P[f(x, \omega)] \tag{6.215}$$

is the problem

$$\text{Min}_{\lambda \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}_+^{q-p}} \sup_{P \geq 0} L_x(P, \lambda), \tag{6.216}$$

where

$$L_x(P, \lambda) := \int_{\Omega} f(x, \omega) dP(\omega) + \lambda_0 \left(1 - \int_{\Omega} dP(\omega)\right) + \sum_{i=1}^q \lambda_i \left(b_i - \int_{\Omega} \psi_i(\omega) dP(\omega)\right).$$

It is straightforward to verify that

$$\sup_{P \geq 0} L_x(P, \lambda) = \begin{cases} \lambda_0 + \sum_{i=1}^q b_i \lambda_i & \text{if } f(x, \omega) - \lambda_0 - \sum_{i=1}^q \lambda_i \psi_i(\omega) \leq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

The last assertion follows since for any  $\bar{\omega} \in \Omega$  and  $\alpha > 0$  we can take  $P := \alpha \Delta(\bar{\omega})$ , in which case

$$\mathbb{E}_P \left[ f(x, \omega) - \lambda_0 - \sum_{i=1}^q \lambda_i \psi_i(\omega) \right] = \alpha \left[ f(x, \bar{\omega}) - \lambda_0 - \sum_{i=1}^q \lambda_i \psi_i(\bar{\omega}) \right].$$

Consequently, we can write the dual problem (6.216) in the form

$$\begin{aligned} \text{Min}_{\lambda \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}_+^{q-p}} \quad & \lambda_0 + \sum_{i=1}^q b_i \lambda_i \\ \text{s.t.} \quad & \lambda_0 + \sum_{i=1}^q \lambda_i \psi_i(\omega) \geq f(x, \omega), \quad \omega \in \Omega. \end{aligned} \quad (6.217)$$

If the set  $\Omega$  is finite, then problem (6.215) and its dual (6.217) are linear programming problems. In that case, there is no duality gap between these problems unless both are infeasible. If the set  $\Omega$  is infinite, then the dual problem (6.217) becomes a linear semi-infinite programming problem. In that case, one needs to verify some regularity conditions in order to ensure the no-duality-gap property. One such regularity condition will be, “the dual problem (6.217) has a nonempty and bounded set of optimal solutions” (see Theorem 7.8). Another regularity condition ensuring the no-duality-gap property is, “the set  $\Omega$  is a compact metric space equipped with its Borel sigma algebra and functions  $\psi_i(\cdot)$ ,  $i = 1, \dots, q$ , and  $f(x, \cdot)$  are continuous on  $\Omega$ .”

If for every  $x \in X$  there is no duality gap between problems (6.215) and (6.217), then the corresponding min-max problem (6.210) is equivalent to the following semi-infinite programming problem:

$$\begin{aligned} \text{Min}_{x \in X, \lambda \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}_+^{q-p}} \quad & \lambda_0 + \sum_{i=1}^q b_i \lambda_i \\ \text{s.t.} \quad & \lambda_0 + \sum_{i=1}^q \lambda_i \psi_i(\omega) \geq f(x, \omega), \quad \omega \in \Omega. \end{aligned} \quad (6.218)$$

**Remark 23.** Let  $\Omega$  be a nonempty measurable subset of  $\mathbb{R}^d$ , equipped with its Borel sigma algebra, and let  $\mathcal{M}$  be the set of *all* probability measures supported on  $\Omega$ . Then by the above analysis we have that it suffices in problem (6.210) to take the maximum over measures of mass one, and hence problem (6.210) is equivalent to the following (deterministic) minimax problem:

$$\text{Min}_{x \in X} \sup_{\omega \in \Omega} f(x, \omega). \quad (6.219)$$

## 6.7 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 discrete process in which evolution of the state of the system is represented by a scenario tree.

### 6.7.1 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 next possible realizations of the data process, which are represented by its children nodes at stage  $t + 1$ . In the risk neutral approach we optimize the corresponding conditional expectation of the objective function. This allows us 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 now discuss such construction in detail.

As in section 3.1.3, we denote by  $\Omega_t$  the set of all nodes at stage  $t = 1, \dots, T$ , by  $K_t := |\Omega_t|$  the cardinality of  $\Omega_t$  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  $\Omega_{t+1}$ , i.e.,  $C_a \cap C_{a'} = \emptyset$  if  $a \neq a'$  and  $\Omega_{t+1} = \bigcup_{a \in \Omega_t} C_a$ ,  $t = 1, \dots, T - 1$ . With the set  $\Omega_T$  we associate sigma 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  $\Omega_T$ .) By this construction, there is a one-to-one correspondence between elementary events of  $\mathcal{F}_{T-1}$  and the set  $\Omega_{T-1}$  of nodes at stage  $T - 1$ . By continuing this process we construct a sequence of sigma algebras  $\mathcal{F}_1 \subset \dots \subset \mathcal{F}_T$ . (Such a sequence of nested sigma algebras is 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 nodes of  $\Omega_t$  and elementary events of the sigma algebra  $\mathcal{F}_t$ , 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  $\Omega_T$ .

Suppose, further, that there is a probability distribution defined on the scenario tree. As 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 vector<sup>51</sup>  $p^a \in \mathbb{R}^{|C_a|}$  of conditional probabilities of moving from  $a$  to nodes of  $C_a$ . 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  $Z : C_a \rightarrow \mathbb{R}$  can be viewed as a random variable. Since the space of functions  $Z : C_a \rightarrow \mathbb{R}$  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  $t = 1, \dots, T - 1$  we associate a risk measure  $\rho^a(Z)$  defined on the space of functions  $Z : C_a \rightarrow \mathbb{R}$ , that is, we choose a family of risk measures

$$\rho^a : \mathbb{R}^{|C_a|} \rightarrow \mathbb{R}, \quad a \in \Omega_t, \quad t = 1, \dots, T - 1. \quad (6.220)$$

Of course, there are many ways to define such risk measures. For instance, for a given probability distribution on the scenario tree, we can use conditional expectations

$$\rho^a(Z) := \mathbb{E}_{p^a}[Z], \quad a \in \Omega_t, \quad t = 1, \dots, T - 1. \quad (6.221)$$

Such choice of risk measures  $\rho^a$  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,

$$\rho^a[Z] := \inf_{t \in \mathbb{R}} \{t + \lambda_a^{-1} \mathbb{E}_{p^a}[Z - t]_+\}, \quad \lambda_a \in (0, 1), \quad (6.222)$$

<sup>51</sup>A vector  $p = (p_1, \dots, p_n) \in \mathbb{R}^n$  is said to be a *probability vector* if all its components  $p_i$  are nonnegative and  $\sum_{i=1}^n p_i = 1$ . If  $Z = (Z_1, \dots, Z_n) \in \mathbb{R}^n$  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 and

$$\rho^a[Z] := \mathbb{E}_{p^a}[Z] + c_a \mathbb{E}_{p^a}[Z - \mathbb{E}_{p^a}[Z]]_+, \quad c_a \in [0, 1], \quad (6.223)$$

corresponding to the absolute semideviation risk measure.

Since  $\Omega_{t+1}$  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,  $\mathbb{R}^{K_{t+1}} = \mathbb{R}^{|C_{a_1}|} \times \dots \times \mathbb{R}^{|C_{a_{K_t}}|}$ , where  $\{a_1, \dots, a_{K_t}\} = \Omega_t$ . Define the mappings

$$\rho_{t+1} := (\rho^{a_1}, \dots, \rho^{a_{K_t}}) : \mathbb{R}^{K_{t+1}} \rightarrow \mathbb{R}^{K_t}, \quad t = 1, \dots, T-1, \quad (6.224)$$

associated with risk measures  $\rho^a$ . Recall that the set  $\Omega_{t+1}$  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  $\mathcal{Z}_T$  the space of all functions  $Z : \Omega_T \rightarrow \mathbb{R}$ . As mentioned, we can identify every such function with a vector of the space  $\mathbb{R}^{K_T}$ , i.e., the space  $\mathcal{Z}_T$  can be identified with the space  $\mathbb{R}^{K_T}$ . We have that a function  $Z : \Omega_T \rightarrow \mathbb{R}$  is  $\mathcal{F}_{T-1}$ -measurable iff it is constant on every set  $C_a$ ,  $a \in \Omega_{T-1}$ . We denote by  $\mathcal{Z}_{T-1}$  the subspace of  $\mathcal{Z}_T$  formed by  $\mathcal{F}_{T-1}$ -measurable functions. The space  $\mathcal{Z}_{T-1}$  can be identified with  $\mathbb{R}^{K_{T-1}}$ . And so on, we can construct a sequence  $\mathcal{Z}_t$ ,  $t = 1, \dots, T$ , of spaces of  $\mathcal{F}_t$ -measurable functions  $Z : \Omega_T \rightarrow \mathbb{R}$  such that  $\mathcal{Z}_1 \subset \dots \subset \mathcal{Z}_T$  and each  $\mathcal{Z}_t$  can be identified with the space  $\mathbb{R}^{K_t}$ . Recall that  $K_1 = 1$ , and hence  $\mathcal{Z}_1$  can be identified with  $\mathbb{R}$ . We view the mapping  $\rho_{t+1}$ , defined in (6.224), as a mapping from the space  $\mathcal{Z}_{t+1}$  into the space  $\mathcal{Z}_t$ . Conversely, with any mapping  $\rho_{t+1} : \mathcal{Z}_{t+1} \rightarrow \mathcal{Z}_t$  we can associate a family of risk measures of the form (6.220).

We say that a mapping  $\rho_{t+1} : \mathcal{Z}_{t+1} \rightarrow \mathcal{Z}_t$  is a *conditional risk mapping* if it satisfies the following conditions:<sup>52</sup>

**(R'1)** Convexity:

$$\rho_{t+1}(\alpha Z + (1 - \alpha)Z') \leq \alpha \rho_{t+1}(Z) + (1 - \alpha) \rho_{t+1}(Z')$$

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  $\alpha \geq 0$  and  $Z \in \mathcal{Z}_{t+1}$ , then  $\rho_{t+1}(\alpha Z) = \alpha \rho_{t+1}(Z)$ .

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  $\rho^a$  associated with  $\rho_{t+1}$ . Also by construction of  $\rho_{t+1}$ , we have that condition (R'3) holds iff condition (R3) holds for all  $\rho^a$ . That is,  $\rho_{t+1}$  is a *conditional risk mapping* iff every corresponding risk measure  $\rho^a$  is a *coherent risk measure*.

By Theorem 6.4 with each coherent risk measure  $\rho^a$ ,  $a \in \Omega_t$ , is associated a set  $\mathfrak{A}(a)$  of probability measures (vectors) such that

$$\rho^a(Z) = \max_{p \in \mathfrak{A}(a)} \mathbb{E}_p[Z]. \quad (6.225)$$

<sup>52</sup>For  $Z_1, Z_2 \in \mathcal{Z}_t$  the inequality  $Z_2 \succeq Z_1$  is understood componentwise.

## 6.7. Multistage Risk Averse Optimization

311

Here  $Z \in \mathbb{R}^{K_{t+1}}$  is a vector corresponding to function  $Z : \Omega_{t+1} \rightarrow \mathbb{R}$ , 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  $\rho_{t+1}$  as a maximum of conditional expectations as follows. Let  $\nu = (\nu_a)_{a \in \Omega_t}$  be a probability distribution on  $\Omega_t$ , assigning *positive* probability  $\nu_a$  to every  $a \in \Omega_t$ , and define

$$\mathfrak{C}_{t+1} := \left\{ \mu = \sum_{a \in \Omega_t} \nu_a p^a : p^a \in \mathfrak{A}_{t+1}(a) \right\}. \quad (6.226)$$

It is not difficult to see that  $\mathfrak{C}_{t+1} \subset \mathbb{R}^{K_{t+1}}$  is a convex set of probability vectors. Moreover, since each  $\mathfrak{A}_{t+1}(a)$  is compact, the set  $\mathfrak{C}_{t+1}$  is also compact and hence is closed. Consider a probability distribution (measure)  $\mu = \sum_{a \in \Omega_t} \nu_a p^a \in \mathfrak{C}_{t+1}$ . We have that for  $a \in \Omega_t$ , the corresponding conditional distribution given the event  $C_a$  is  $p^a$ , and<sup>53</sup>

$$\mathbb{E}_\mu [Z | \mathcal{F}_t](a) = \mathbb{E}_{p^a} [Z], \quad Z \in \mathcal{Z}_{t+1}. \quad (6.227)$$

It follows then by (6.225) that

$$\rho_{t+1}(Z) = \max_{\mu \in \mathfrak{C}_{t+1}} \mathbb{E}_\mu [Z | \mathcal{F}_t], \quad (6.228)$$

where the maximum on the right-hand side of (6.228) is taken pointwise in  $a \in \Omega_t$ . That is, formula (6.228) means that

$$[\rho_{t+1}(Z)](a) = \max_{p \in \mathfrak{A}_{t+1}(a)} \mathbb{E}_p [Z], \quad Z \in \mathcal{Z}_{t+1}, \quad a \in \Omega_t. \quad (6.229)$$

Note that in this construction, choice of the distribution  $\nu$  is arbitrary and any distribution of  $\mathfrak{C}_{t+1}$  agrees with the distribution  $\nu$  on  $\Omega_t$ .

We are ready now to give a formulation of risk averse multistage programs. For a sequence  $\rho_{t+1} : \mathcal{Z}_{t+1} \rightarrow \mathcal{Z}_t$ ,  $t = 1, \dots, T-1$ , of conditional risk mappings, consider the following risk averse formulation analogous to the nested risk neutral formulation (3.1):

$$\begin{aligned} \text{Min}_{x_1 \in \mathcal{X}_1} f_1(x_1) + \rho_2 \left[ \inf_{x_2 \in \mathcal{X}_2(x_1, \omega)} f_2(x_2, \omega) + \dots \right. \\ \left. + \rho_{T-1} \left[ \inf_{x_{T-1} \in \mathcal{X}_{T-1}(x_{T-2}, \omega)} f_{T-1}(x_{T-1}, \omega) \right. \right. \\ \left. \left. + \rho_T \left[ \inf_{x_T \in \mathcal{X}_T(x_{T-1}, \omega)} f_T(x_T, \omega) \right] \right] \right]. \end{aligned} \quad (6.230)$$

Here  $\omega$  is an element of  $\Omega := \Omega_T$ , the objective functions  $f_t : \mathbb{R}^{n_{t-1}} \times \Omega \rightarrow \mathbb{R}$  are real valued functions, and  $\mathcal{X}_t : \mathbb{R}^{n_{t-1}} \times \Omega \rightrightarrows \mathbb{R}^{n_t}$ ,  $t = 2, \dots, T$ , 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  $x_{t-1}$ . Note that if the corresponding risk measures  $\rho^a$  are defined as conditional expectations (6.221), then the multistage problem (6.230) coincides with the risk neutral multistage problem (3.1).

<sup>53</sup>Recall that the conditional expectation  $\mathbb{E}_\mu[\cdot | \mathcal{F}_t]$  is a mapping from  $\mathcal{Z}_{t+1}$  into  $\mathcal{Z}_t$ .

There are several ways in which the nested formulation (6.230) can be formalized. Similarly to (3.3), we can write problem (6.230) in the form

$$\begin{aligned} \text{Min}_{x_1, x_2, \dots, x_T} \quad & f_1(x_1) + \rho_2 \left[ f_2(x_2(\omega), \omega) + \dots \right. \\ & \left. + \rho_{T-1} [f_{T-1}(x_{T-1}(\omega), \omega) + \rho_T [f_T(x_T(\omega), \omega)]] \right] \\ \text{s.t.} \quad & x_1 \in \mathcal{X}_1, \quad x_t(\omega) \in \mathcal{X}_t(x_{t-1}(\omega), \omega), \quad t = 2, \dots, T. \end{aligned} \quad (6.231)$$

Optimization in (6.231) is performed over functions  $x_t : \Omega \rightarrow \mathbb{R}, t = 1, \dots, T$ , satisfying the corresponding constraints, which imply that each  $x_t(\omega)$  is  $\mathcal{F}_t$ -measurable and hence each  $f_t(x_t(\omega), \omega)$  is  $\mathcal{F}_t$ -measurable. The requirement for  $x_t(\omega)$  to be  $\mathcal{F}_t$ -measurable is another way of formulating the *nonanticipativity constraints*. Therefore, it can be viewed that the optimization in (6.231) is performed over feasible policies.

Consider the function  $\varrho : \mathcal{Z}_1 \times \dots \times \mathcal{Z}_T \rightarrow \mathbb{R}$  defined as

$$\varrho(Z_1, \dots, Z_T) := Z_1 + \rho_2 \left[ Z_2 + \dots + \rho_{T-1} [Z_{T-1} + \rho_T [Z_T]] \right]. \quad (6.232)$$

By condition (R'3) we have that

$$\rho_{T-1} [Z_{T-1} + \rho_T [Z_T]] = \rho_{T-1} \circ \rho_T [Z_{T-1} + Z_T].$$

By continuing this process we obtain that

$$\varrho(Z_1, \dots, Z_T) = \bar{\rho}(Z_1 + \dots + Z_T), \quad (6.233)$$

where  $\bar{\rho} := \rho_2 \circ \dots \circ \rho_T$ . We refer to  $\bar{\rho}$  as the *composite risk measure*. That is,

$$\bar{\rho}(Z_1 + \dots + Z_T) = Z_1 + \rho_2 \left[ Z_2 + \dots + \rho_{T-1} [Z_{T-1} + \rho_T [Z_T]] \right], \quad (6.234)$$

defined for  $Z_t \in \mathcal{Z}_t, t = 1, \dots, T$ . Recall that  $\mathcal{Z}_1$  is identified with  $\mathbb{R}$ , and hence  $Z_1$  is a real number and  $\bar{\rho} : \mathcal{Z}_T \rightarrow \mathbb{R}$  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  $f_{T-1}(x_{T-1}(\omega), \omega)$  is  $\mathcal{F}_{T-1}$ -measurable, it follows by condition (R'3) that

$$f_{T-1}(x_{T-1}(\omega), \omega) + \rho_T [f_T(x_T(\omega), \omega)] = \rho_T [f_{T-1}(x_{T-1}(\omega), \omega) + f_T(x_T(\omega), \omega)].$$

Continuing this process backward, we obtain that the objective function of (6.231) can be formulated using the composite risk measure. That is, problem (6.231) can be written in the form

$$\begin{aligned} \text{Min}_{x_1, x_2, \dots, x_T} \quad & \bar{\rho} [f_1(x_1) + f_2(x_2(\omega), \omega) + \dots + f_T(x_T(\omega), \omega)] \\ \text{s.t.} \quad & x_1 \in \mathcal{X}_1, \quad x_t(\omega) \in \mathcal{X}_t(x_{t-1}(\omega), \omega), \quad t = 2, \dots, T. \end{aligned} \quad (6.235)$$

If the conditional risk mappings are defined as the respective conditional expectations, then the composite risk measure  $\bar{\rho}$  becomes the corresponding expectation operator, and (6.235) coincides with the multistage program written in the form (3.3). Unfortunately, it is not easy

## 6.7. Multistage Risk Averse Optimization

313

to write the composite risk measure  $\bar{\rho}$  in a closed form even for relatively simple conditional risk mappings other than conditional expectations.

An alternative approach to formalize the nested formulation (6.230) is to write *dynamic programming equations*. That is, for the last period  $T$  we have

$$Q_T(x_{T-1}, \omega) := \inf_{x_T \in \mathcal{X}_T(x_{T-1}, \omega)} f_T(x_T, \omega), \quad (6.236)$$

$$\mathcal{Q}_T(x_{T-1}, \omega) := \rho_T[Q_T(x_{T-1}, \omega)], \quad (6.237)$$

and for  $t = T - 1, \dots, 2$ , we recursively apply the conditional risk measures

$$\mathcal{Q}_t(x_{t-1}, \omega) := \rho_t[\mathcal{Q}_t(x_{t-1}, \omega)], \quad (6.238)$$

where

$$\mathcal{Q}_t(x_{t-1}, \omega) := \inf_{x_t \in \mathcal{X}_t(x_{t-1}, \omega)} \left\{ f_t(x_t, \omega) + \mathcal{Q}_{t+1}(x_t, \omega) \right\}. \quad (6.239)$$

Of course, equations (6.238) and (6.239) can be combined into one equation:<sup>54</sup>

$$\mathcal{Q}_t(x_{t-1}, \omega) = \inf_{x_t \in \mathcal{X}_t(x_{t-1}, \omega)} \left\{ f_t(x_t, \omega) + \rho_{t+1}[\mathcal{Q}_{t+1}(x_t, \omega)] \right\}. \quad (6.240)$$

Finally, at the first stage we solve the problem

$$\text{Min}_{x_1 \in \mathcal{X}_1} f_1(x_1) + \rho_2[\mathcal{Q}_2(x_1, \omega)]. \quad (6.241)$$

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  $\mathcal{Q}_t(x_{t-1}, \omega)$  are real valued. In particular, this implies that the considered problem should have relatively complete recourse. Also, in the above development of the dynamic programming equations, the monotonicity condition (R'2) plays a crucial role, because only then we can move the optimization under the risk operation.

**Remark 24.** By using representation (6.228), we can write the dynamic programming equations (6.240) in the form

$$\mathcal{Q}_t(x_{t-1}, \omega) = \inf_{x_t \in \mathcal{X}_t(x_{t-1}, \omega)} \left\{ f_t(x_t, \omega) + \sup_{\mu \in \mathcal{C}_{t+1}} \mathbb{E}_\mu[\mathcal{Q}_{t+1}(x_t) | \mathcal{F}_t](\omega) \right\}. \quad (6.242)$$

Note that the left- and right-hand-side functions in (6.242) 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  $\mu \in \mathcal{C}_{t+1}$  is representable in the form  $\mu = \sum_{a \in \Omega_t} v_a p^a$  (see (6.226)) and that

$$\mathbb{E}_\mu[\mathcal{Q}_{t+1}(x_t) | \mathcal{F}_t](a) = \mathbb{E}_{p^a}[\mathcal{Q}_{t+1}(x_t)], \quad a \in \Omega_t. \quad (6.243)$$

We say that the problem is *convex* if the functions  $f_t(\cdot, \omega)$ ,  $\mathcal{Q}_t(\cdot, \omega)$  and the sets  $\mathcal{X}_t(x_{t-1}, \omega)$  are convex for every  $\omega \in \Omega$  and  $t = 1, \dots, T$ . If the problem is convex, then (since the

<sup>54</sup>With some abuse of the notation we write  $\mathcal{Q}_{t+1}(x_t, \omega)$  for the value of  $\mathcal{Q}_{t+1}(x_t)$  at  $\omega \in \Omega$ , and  $\rho_{t+1}[\mathcal{Q}_{t+1}(x_t, \omega)]$  for  $\rho_{t+1}[\mathcal{Q}_{t+1}(x_t)](\omega)$ .

set  $\mathfrak{C}_{t+1}$  is convex compact) the inf and sup operators on the right-hand side of (6.242) 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  $\bar{\mu}_{t+1} := \sum_{a \in \Omega_t} v_a \bar{p}^a$  an optimal solution of the original problem and the corresponding cost-to-go functions satisfy the following dynamic programming equations:

$$Q_t(x_{t-1}, \omega) = \inf_{x_t \in \mathfrak{X}_t(x_{t-1}, \omega)} \left\{ f_t(x_t, \omega) + \mathbb{E}_{\bar{\mu}_{t+1}}[Q_{t+1}(x_t) | \mathcal{F}_t](\omega) \right\}. \quad (6.244)$$

Moreover, it is possible to choose the “worst case” distributions  $\bar{\mu}_{t+1}$  in a consistent way, i.e., such that each  $\bar{\mu}_{t+1}$  coincides with  $\bar{\mu}_t$  on  $\mathcal{F}_t$ . That is, consider the first-stage problem (6.241). We have that (recall that at the first stage there is only one node,  $\mathcal{F}_1 = \{\emptyset, \Omega\}$  and  $\mathfrak{C}_2 = \mathfrak{A}_2$ )

$$\rho_2[Q_2(x_1)] = \sup_{\mu \in \mathfrak{C}_2} \mathbb{E}_\mu[Q_2(x_1) | \mathcal{F}_1] = \sup_{\mu \in \mathfrak{C}_2} \mathbb{E}_\mu[Q_2(x_1)]. \quad (6.245)$$

By convexity and since  $\mathfrak{C}_2$  is compact, we have that there is  $\bar{\mu}_2 \in \mathfrak{C}_2$  (an optimal solution of the dual problem) such that the optimal value of the first-stage problem is equal to the optimal value and the set of optimal solutions of the first-stage problem is contained in the set of optimal solutions of the problem

$$\text{Min}_{x_1 \in \mathfrak{X}_1} \mathbb{E}_{\bar{\mu}_2}[Q_2(x_1)]. \quad (6.246)$$

Let  $\bar{x}_1$  be an optimal solution of the first-stage problem. Then we can choose  $\bar{\mu}_3 \in \mathfrak{C}_3$ , of the form  $\bar{\mu}_3 := \sum_{a \in \Omega_2} v_a \bar{p}^a$  such that (6.244) holds with  $t = 2$  and  $x_1 = \bar{x}_1$ . Moreover, we can take the probability measure  $v = (v_a)_{a \in \Omega_2}$  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  $\bar{\mu}_4 \in \mathfrak{C}_4$ , and so on for later stages.

In that way, assuming existence of optimal solutions, we can construct a probability distribution  $\bar{\mu}_2, \dots, \bar{\mu}_T$  on the considered 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 (6.230) and has an optimal solution which also is an optimal solution of the problem (6.230). (In that sense, the obtained multistage problem, driven by dynamic programming equations (6.244), is almost equivalent to the original problem.)

**Remark 25.** Let us define, for every node  $a \in \Omega_t$ ,  $t = 1, \dots, T-1$ , 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 (6.225) is attained at a measure of mass one at a point of the set  $C_a$ . Consequently, by (6.243), for such choice of the sets  $\mathfrak{A}_{t+1}(a)$  the dynamic programming equations (6.242) can be written as

$$Q_t(x_{t-1}, a) = \inf_{x_t \in \mathfrak{X}_t(x_{t-1}, a)} \left\{ f_t(x_t, a) + \max_{\omega \in C_a} Q_{t+1}(x_t, \omega) \right\}, \quad a \in \Omega_t. \quad (6.247)$$

It is interesting to note (see Remark 24, page 313) that if the problem is convex, then it is possible to construct a probability distribution (on the considered scenario tree), defined by a sequence  $\bar{\mu}_t$ ,  $t = 2, \dots, T$ , of consistent probability distributions, such that the obtained (risk neutral) multistage program is almost equivalent to the min-max formulation (6.247).

### 6.7.2 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  $\Omega$  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  $(\Omega, \mathcal{F}_2)$ . 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  $\mathcal{Z}_1 \subset \mathcal{Z}_2$ .

We say that a mapping  $\rho : \mathcal{Z}_2 \rightarrow \mathcal{Z}_1$  is a *conditional risk mapping* if it satisfies the following conditions:

**(R'1)** Convexity:

$$\rho(\alpha Z + (1 - \alpha)Z') \leq \alpha \rho(Z) + (1 - \alpha)\rho(Z')$$

for any  $Z, Z' \in \mathcal{Z}_2$  and  $\alpha \in [0, 1]$ .

**(R'2)** Monotonicity: If  $Z, Z' \in \mathcal{Z}_2$  and  $Z \geq Z'$ , then  $\rho(Z) \geq \rho(Z')$ .

**(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  $\alpha \geq 0$  and  $Z \in \mathcal{Z}_2$ , 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 (R'1)–(R'4) define a coherent risk measure. Examples of coherent risk measures, discussed in section 6.3.2, have conditional risk mapping analogues which are obtained by replacing the expectation operator with the corresponding conditional expectation  $\mathbb{E}[\cdot | \mathcal{F}_1]$  operator. Let us look at some examples.

**Conditional Expectation.** In itself, the conditional expectation mapping  $\mathbb{E}[\cdot | \mathcal{F}_1] : \mathcal{Z}_2 \rightarrow \mathcal{Z}_1$  is a conditional risk mapping. Indeed, for any  $p \geq 1$  and  $Z \in \mathcal{L}_p(\Omega, \mathcal{F}_2, P)$  we have by Jensen inequality that  $\mathbb{E}[|Z|^p | \mathcal{F}_1] \geq |\mathbb{E}[Z | \mathcal{F}_1]|^p$ , and hence

$$\int_{\Omega} |\mathbb{E}[Z | \mathcal{F}_1]|^p dP \leq \int_{\Omega} \mathbb{E}[|Z|^p | \mathcal{F}_1] dP = \mathbb{E}[|Z|^p] < +\infty. \quad (6.248)$$

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, and condition (R'3) is a property of conditional expectation.

**Conditional AV@R.** An analogue of the AV@R risk measure 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  $\text{AV@R}_{\alpha}(\cdot | \mathcal{F}_1) : \mathcal{Z}_2 \rightarrow \mathcal{Z}_1$  as

$$[\text{AV@R}_{\alpha}(Z | \mathcal{F}_1)](\omega) := \inf_{Y \in \mathcal{Z}_1} \{Y(\omega) + \alpha^{-1} \mathbb{E}[(Z - Y)_+ | \mathcal{F}_1](\omega)\}, \quad \omega \in \Omega. \quad (6.249)$$

It is not difficult to verify that, indeed, this mapping satisfies conditions (R'1)–(R'4). Similarly to (6.68), for  $\beta \in [0, 1]$  and  $\alpha \in (0, 1)$ , we can also consider the following conditional risk mapping:

$$\rho_{\alpha, \beta | \mathcal{F}_1}(Z) := (1 - \beta)\mathbb{E}[Z | \mathcal{F}_1] + \beta \text{AV@R}_\alpha(Z | \mathcal{F}_1). \quad (6.250)$$

Of course, the above conditional risk mapping  $\rho_{\alpha, \beta | \mathcal{F}_1}$  corresponds to the coherent risk measure  $\rho_{\alpha, \beta}(Z) := (1 - \beta)\mathbb{E}[Z] + \beta \text{AV@R}_\alpha(Z)$ .

**Conditional Mean-Upper-Semideviation.** An analogue of the mean-upper-semideviation 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

$$\rho_{c | \mathcal{F}_1}(Z) := \mathbb{E}[Z | \mathcal{F}_1] + c \left( \mathbb{E} \left[ [Z - \mathbb{E}[Z | \mathcal{F}_1]]_+^p | \mathcal{F}_1 \right] \right)^{1/p}. \quad (6.251)$$

In particular, for  $p = 1$  this gives an analogue of the absolute semideviation risk measure.

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  $t = 1, \dots, T$ .

Consider a conditional risk mapping  $\rho : \mathcal{Z}_2 \rightarrow \mathcal{Z}_1$ . With a set  $A \in \mathcal{F}_1$ , such that  $P(A) \neq 0$ , we associate the function

$$\rho_A(Z) := \mathbb{E}[\rho(Z) | A], \quad Z \in \mathcal{Z}_2, \quad (6.252)$$

where  $\mathbb{E}[Y | A] := \frac{1}{P(A)} \int_A Y dP$  denotes the conditional expectation of random variable  $Y \in \mathcal{Z}_1$  given event  $A \in \mathcal{F}_1$ . Clearly conditions (R'1)–(R'4) imply that the corresponding conditions (R1)–(R4) hold for  $\rho_A$ , and hence  $\rho_A$  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

$$\rho_A(Z + \alpha \mathbf{1}_B) := \mathbb{E}[\rho(Z) + \alpha \mathbf{1}_B | A] = \rho_A(Z) + \alpha P(B | A) \quad \forall \alpha \in \mathbb{R}, \quad (6.253)$$

where  $P(B | A) = P(B \cap A) / P(A)$ .

Since  $\rho_A$  is a coherent risk measure, by Theorem 6.4 it can be represented in the form

$$\rho_A(Z) = \sup_{\zeta \in \mathfrak{A}(A)} \int_{\Omega} \zeta(\omega) Z(\omega) dP(\omega) \quad (6.254)$$

for some set  $\mathfrak{A}(A) \subset \mathcal{L}_q(\Omega, \mathcal{F}_2, P)$  of probability density functions. Let us make the following observation:

- Each density  $\zeta \in \mathfrak{A}(A)$  is supported on the set  $A$ .

Indeed, for any  $B \in \mathcal{F}_1$ , such that  $P(B \cap A) = 0$ , and any  $\alpha \in \mathbb{R}$ , we have by (6.253) that  $\rho_A(Z + \alpha \mathbf{1}_B) = \rho_A(Z)$ . On the other hand, if there exists  $\zeta \in \mathfrak{A}(A)$  such that  $\int_B \zeta dP > 0$ , then it follows from (6.254) that  $\rho_A(Z + \alpha \mathbf{1}_B)$  tends to  $+\infty$  as  $\alpha \rightarrow +\infty$ .

Similarly to (6.228), we show now that a conditional risk mapping can be represented as a maximum of a family of conditional expectations. We consider a situation where 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  $\Omega$  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  $Z : \Omega \rightarrow \mathbb{R}$  is  $\mathcal{F}_1$ -measurable iff it is constant on every set  $A_i$ ,  $i \in \mathbb{N}$ .

Consider a conditional risk mapping  $\rho : \mathcal{Z}_2 \rightarrow \mathcal{Z}_1$ . Let

$$\mathfrak{N} := \{i \in \mathbb{N} : P(A_i) \neq 0\}$$

and  $\rho_{A_i}$ ,  $i \in \mathfrak{N}$ , be the corresponding coherent risk measures defined in (6.252). By (6.254) with every  $\rho_{A_i}$ ,  $i \in \mathfrak{N}$ , is associated set  $\mathfrak{A}(A_i)$  of probability density functions, supported on the set  $A_i$ , such that

$$\rho_{A_i}(Z) = \sup_{\zeta \in \mathfrak{A}(A_i)} \int_{\Omega} \zeta(\omega) Z(\omega) dP(\omega). \quad (6.255)$$

Now let  $\nu = (\nu_i)_{i \in \mathbb{N}}$  be a probability distribution (measure) on  $(\Omega, \mathcal{F}_1)$ , assigning probability  $\nu_i$  to the event  $A_i$ ,  $i \in \mathbb{N}$ . Assume that  $\nu$  is such that  $\nu(A_i) = 0$  iff  $P(A_i) = 0$  (i.e.,  $\mu$  is absolutely continuous with respect to  $P$  and  $P$  is absolutely continuous with respect to  $\nu$  on  $(\Omega, \mathcal{F}_1)$ ); otherwise the probability measure  $\nu$  is arbitrary. Define the following family of probability measures on  $(\Omega, \mathcal{F}_2)$ :

$$\mathfrak{C} := \left\{ \mu = \sum_{i \in \mathfrak{N}} \nu_i \mu_i : d\mu_i = \zeta_i dP, \zeta_i \in \mathfrak{A}(A_i), i \in \mathfrak{N} \right\}. \quad (6.256)$$

Note that since  $\sum_{i \in \mathfrak{N}} \nu_i = 1$ , every  $\mu \in \mathfrak{C}$  is a probability measure. For  $\mu \in \mathfrak{C}$ , with respective densities  $\zeta_i \in \mathfrak{A}(A_i)$  and  $d\mu_i = \zeta_i dP$ , and  $Z \in \mathcal{Z}_2$  we have that

$$\mathbb{E}_{\mu}[Z|\mathcal{F}_1] = \sum_{i \in \mathfrak{N}} \mathbb{E}_{\mu_i}[Z|\mathcal{F}_1]. \quad (6.257)$$

Moreover, since  $\zeta_i$  is supported on  $A_i$ ,

$$\mathbb{E}_{\mu_i}[Z|\mathcal{F}_1](\omega) = \begin{cases} \int_{A_i} Z \zeta_i dP & \text{if } \omega \in A_i, \\ 0 & \text{otherwise.} \end{cases} \quad (6.258)$$

By the max-representations (6.255) it follows that for  $Z \in \mathcal{Z}_2$  and  $\omega \in A_i$ ,

$$\sup_{\mu \in \mathfrak{C}} \mathbb{E}_{\mu}[Z|\mathcal{F}_1](\omega) = \sup_{\zeta_i \in \mathfrak{A}(A_i)} \int_{A_i} Z \zeta_i dP = \rho_{A_i}(Z). \quad (6.259)$$

Also since  $[\rho(Z)](\cdot)$  is  $\mathcal{F}_1$ -measurable, and hence is constant on every set  $A_i$ , we have that  $[\rho(Z)](\omega) = \rho_{A_i}(Z)$  for every  $\omega \in A_i$ ,  $i \in \mathfrak{N}$ . We obtain the following result.

**Proposition 6.41.** *Let  $\mathcal{Z}_i := \mathcal{L}_p(\Omega, \mathcal{F}_i, P)$ ,  $i = 1, 2$ , with  $\mathcal{F}_1 \subset \mathcal{F}_2$ , and let  $\rho : \mathcal{Z}_2 \rightarrow \mathcal{Z}_1$  be a conditional risk mapping. Suppose that  $\mathcal{F}_1$  has a countable number of elementary events. Then*

$$\rho(Z) = \sup_{\mu \in \mathfrak{C}} \mathbb{E}_{\mu}[Z|\mathcal{F}_1], \quad \forall Z \in \mathcal{Z}_2, \quad (6.260)$$

where  $\mathfrak{C}$  is a family of probability measures on  $(\Omega, \mathcal{F}_2)$ , specified in (6.256), corresponding to a probability distribution  $\nu$  on  $(\Omega, \mathcal{F}_1)$ .

### 6.7.3 Risk Averse Multistage Stochastic Programming

There are several ways in which risk averse stochastic programming can be formulated in a multistage setting. We now discuss a nested formulation similar to the derivations of section 6.7.1. Let  $(\Omega, \mathcal{F}, P)$  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  $\mathcal{Z}_t := \mathcal{L}_p(\Omega, \mathcal{F}_t, P)$ ,  $t = 1, \dots, T$ , be the corresponding sequence of spaces of  $\mathcal{F}_t$ -measurable and  $p$ -integrable functions, and let  $\rho_{t+1|\mathcal{F}_t} : \mathcal{Z}_{t+1} \rightarrow \mathcal{Z}_t$ ,  $t = 1, \dots, T-1$ , be a selected family of conditional risk mappings. It is straightforward to verify that the composition

$$\rho_{t|\mathcal{F}_{t-1}} \circ \cdots \circ \rho_{T|\mathcal{F}_{T-1}} : \mathcal{Z}_T \xrightarrow{\rho_{T|\mathcal{F}_{T-1}}} \mathcal{Z}_{T-1} \xrightarrow{\rho_{T-1|\mathcal{F}_{T-2}}} \cdots \xrightarrow{\rho_{t|\mathcal{F}_{t-1}}} \mathcal{Z}_{t-1}, \quad (6.261)$$

$t = 2, \dots, T$ , of such conditional risk mappings is also a conditional risk mapping. In particular, the space  $\mathcal{Z}_1$  can be identified with  $\mathbb{R}$  and hence the composition  $\rho_{2|\mathcal{F}_1} \circ \cdots \circ \rho_{T|\mathcal{F}_{T-1}} : \mathcal{Z}_T \rightarrow \mathbb{R}$  is a real valued coherent risk measure.

Similarly to (6.230), we consider the following nested risk averse formulation of multistage programs:

$$\begin{aligned} \text{Min}_{x_1 \in \mathcal{X}_1} \quad & f_1(x_1) + \rho_{2|\mathcal{F}_1} \left[ \inf_{x_2 \in \mathcal{X}_2(x_1, \omega)} f_2(x_2, \omega) + \cdots \right. \\ & + \rho_{T-1|\mathcal{F}_{T-2}} \left[ \inf_{x_{T-1} \in \mathcal{X}_{T-1}(x_{T-2}, \omega)} f_{T-1}(x_{T-1}, \omega) \right. \\ & \left. \left. + \rho_{T|\mathcal{F}_{T-1}} \left[ \inf_{x_T \in \mathcal{X}_T(x_{T-1}, \omega)} f_T(x_T, \omega) \right] \right] \right]. \end{aligned} \quad (6.262)$$

Here  $f_t : \mathbb{R}^{n_t-1} \times \Omega \rightarrow \mathbb{R}$  and  $\mathcal{X}_t : \mathbb{R}^{n_t-1} \times \Omega \rightrightarrows \mathbb{R}^{n_t}$ ,  $t = 2, \dots, T$ , 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 was discussed in section 6.7.1, the above nested formulation (6.262) has two equivalent interpretations. Namely, it can be formulated as

$$\begin{aligned} \text{Min}_{x_1, x_2, \dots, x_T} \quad & f_1(x_1) + \rho_{2|\mathcal{F}_1} \left[ f_2(\mathbf{x}_2(\omega), \omega) + \cdots \right. \\ & + \rho_{T-1|\mathcal{F}_{T-2}} [f_{T-1}(\mathbf{x}_{T-1}(\omega), \omega) \\ & \left. + \rho_{T|\mathcal{F}_{T-1}} [f_T(\mathbf{x}_T(\omega), \omega)]] \right] \\ \text{s.t. } & x_1 \in \mathcal{X}_1, \quad \mathbf{x}_t(\omega) \in \mathcal{X}_t(\mathbf{x}_{t-1}(\omega), \omega), \quad t = 2, \dots, T, \end{aligned} \quad (6.263)$$

where the optimization is performed over  $\mathcal{F}_t$ -measurable  $\mathbf{x}_t : \Omega \rightarrow \mathbb{R}$ ,  $t = 1, \dots, T$ , satisfying the corresponding constraints, and such that  $f_t(\mathbf{x}_t(\cdot), \cdot) \in \mathcal{Z}_t$ . Recall that the nonanticipativity is enforced here by the  $\mathcal{F}_t$ -measurability of  $\mathbf{x}_t(\cdot)$ . 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 (6.263) in the form

$$\begin{aligned} \text{Min}_{x_1, x_2, \dots, x_T} \quad & \bar{\rho} [f_1(x_1) + f_2(\mathbf{x}_2(\omega), \omega) + \cdots + f_T(\mathbf{x}_T(\omega), \omega)] \\ \text{s.t. } & x_1 \in \mathcal{X}_1, \quad \mathbf{x}_t(\omega) \in \mathcal{X}_t(\mathbf{x}_{t-1}(\omega), \omega), \quad t = 2, \dots, T. \end{aligned} \quad (6.264)$$

## 6.7. Multistage Risk Averse Optimization

319

Recall that for  $Z_t \in \mathcal{Z}_t$ ,  $t = 1, \dots, T$ ,

$$\bar{\rho}(Z_1 + \dots + Z_T) = Z_1 + \rho_{2|\mathcal{F}_1} \left[ Z_2 + \dots + \rho_{T-1|\mathcal{F}_{T-2}} \left[ Z_{T-1} + \rho_{T|\mathcal{F}_{T-1}} [Z_T] \right] \right], \quad (6.265)$$

and that conditions (R'1)–(R'4) imply that  $\bar{\rho} : \mathcal{Z}_T \rightarrow \mathbb{R}$  is a coherent risk measure.

Alternatively we can write the corresponding dynamic programming equations (compare with (6.236)–(6.241)):

$$Q_T(x_{T-1}, \omega) = \inf_{x_T \in \mathcal{X}_T(x_{T-1}, \omega)} f_T(x_T, \omega), \quad (6.266)$$

$$Q_t(x_{t-1}, \omega) = \inf_{x_t \in \mathcal{X}_t(x_{t-1}, \omega)} \left\{ f_t(x_t, \omega) + Q_{t+1}(x_t, \omega) \right\}, \quad t = T-1, \dots, 2, \quad (6.267)$$

where

$$Q_t(x_{t-1}, \omega) = \rho_{t|\mathcal{F}_{t-1}} [Q_{t+1}(x_t, \omega)], \quad t = T, \dots, 2. \quad (6.268)$$

Finally, at the first stage we solve the problem

$$\text{Min}_{x_1 \in \mathcal{X}_1} f_1(x_1) + \rho_{2|\mathcal{F}_1} [Q_2(x_1, \omega)]. \quad (6.269)$$

We need to ensure here that the cost-to-go functions are  $p$ -integrable, i.e.,  $Q_t(x_{t-1}, \cdot) \in \mathcal{Z}_t$  for  $t = 1, \dots, T-1$  and all feasible  $x_{t-1}$ .

In applications we often deal with a data process represented by a sequence of random vectors  $\xi_1, \dots, \xi_T$ , say, defined on a probability space  $(\Omega, \mathcal{F}, P)$ . We can associate with this data process filtration  $\mathcal{F}_t := \sigma(\xi_1, \dots, \xi_t)$ ,  $t = 1, \dots, T$ , where  $\sigma(\xi_1, \dots, \xi_t)$  denotes the smallest sigma algebra with respect to which  $\xi_{[t]} = (\xi_1, \dots, \xi_t)$  is measurable. However, it is more convenient to deal with conditional risk mappings defined directly in terms of the data process rather than the respective sequence of sigma algebras. For example, consider

$$\rho_{t|\xi_{[t-1]}}(Z) := (1 - \beta_t) \mathbb{E} [Z | \xi_{[t-1]}] + \beta_t \text{AV@R}_{\alpha_t}(Z | \xi_{[t-1]}), \quad t = 2, \dots, T, \quad (6.270)$$

where

$$\text{AV@R}_{\alpha_t}(Z | \xi_{[t-1]}) := \inf_{Y \in \mathcal{Z}_{t-1}} \left\{ Y + \alpha_t^{-1} \mathbb{E} [ [Z - Y]_+ | \xi_{[t-1]} ] \right\}. \quad (6.271)$$

Here  $\beta_t \in [0, 1]$  and  $\alpha_t \in (0, 1)$  are chosen constants,  $\mathcal{Z}_t := \mathcal{L}_1(\Omega, \mathcal{F}_t, P)$ , where  $\mathcal{F}_t$  is the smallest filtration associated with the process  $\xi_t$ , and the minimum on the right-hand side of (6.271) is taken pointwise in  $\omega \in \Omega$ . Compared with (6.249), the conditional AV@R is defined in (6.271) 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

$$\rho_{t|\xi_{[t-1]}}(Z) := \mathbb{E} [Z | \xi_{[t-1]}] + c_t \left( \mathbb{E} [ [Z - \mathbb{E} [Z | \xi_{[t-1]}]]_+^p | \xi_{[t-1]} ] \right)^{1/p}, \quad (6.272)$$

defined in terms of the data process. Note that with  $\rho_{t|\xi_{[t-1]}}$ , defined in (6.270) or (6.272), is associated coherent risk measure  $\rho_t$  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 (6.270)–(6.272) coincide with the respective unconditional expectations, and hence  $\rho_{t|\xi_{[t-1]}}(Z)$  does not depend on  $\xi_{[t-1]}$  and coincides with  $\rho_t(Z)$ .

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 (6.263) takes the form

$$\begin{aligned} \text{Min}_{x_1, x_2, \dots, x_T} \quad & f_1(x_1) + \rho_{2|\xi_{[1]}} \left[ f_2(x_2(\xi_{[2]}), \xi_2) + \dots \right. \\ & + \rho_{T-1|\xi_{[T-2]}} \left[ f_{T-1}(x_{T-1}(\xi_{[T-1]}), \xi_{T-1}) \right. \\ & \left. \left. + \rho_{T|\xi_{[T-1]}} [f_T(x_T(\xi_{[T]}), \xi_T)] \right] \right] \\ \text{s.t. } \quad & x_1 \in \mathcal{X}_1, \quad x_t(\xi_{[t]}) \in \mathcal{X}_t(x_{t-1}(\xi_{[t-1]}), \xi_t), \quad t = 2, \dots, T, \end{aligned} \quad (6.273)$$

where the optimization is performed over feasible policies.

The corresponding dynamic programming equations (6.267)–(6.268) take the form

$$Q_t(x_{t-1}, \xi_{[t]}) = \inf_{x_t \in \mathcal{X}_t(x_{t-1}, \xi_t)} \left\{ f_t(x_t, \xi_t) + Q_{t+1}(x_t, \xi_{[t+1]}) \right\}, \quad (6.274)$$

where

$$Q_{t+1}(x_t, \xi_{[t+1]}) = \rho_{t+1|\xi_{[t]}} [Q_{t+1}(x_t, \xi_{[t+1]})]. \quad (6.275)$$

Note that if the process  $\xi_t$  is stagewise independent, then the conditional expectations coincide with the respective unconditional expectations, and hence (similar to the risk neutral case) 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  $\xi_t$  rather than  $\xi_{[t]}$ .

Of course, if we set  $\rho_{t|\xi_{[t-1]}}(\cdot) := \mathbb{E}[\cdot | \xi_{[t-1]}]$ , then the above equations (6.274) 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 (6.264) coincides with the respective risk neutral formulation (3.3). Unfortunately, in the general case it is quite difficult to write the composite measure  $\bar{\rho}$  in an explicit form.

### Multiperiod Coherent Risk Measures

It is possible to approach risk averse multistage stochastic programming in the following framework. As before, let  $\mathcal{F}_t$  be a filtration and  $Z_t := \mathcal{L}_p(\Omega, \mathcal{F}_t, P)$ ,  $t = 1, \dots, T$ . Consider the space  $\mathcal{Z} := \mathcal{Z}_1 \times \dots \times \mathcal{Z}_T$ . Recall that since  $\mathcal{F}_1 = \{\emptyset, \Omega\}$ , the space  $\mathcal{Z}_1$  can be identified with  $\mathbb{R}$ . With space  $\mathcal{Z}$  we can associate its dual space  $\mathcal{Z}^* := \mathcal{Z}_1^* \times \dots \times \mathcal{Z}_T^*$ , where  $\mathcal{Z}_t^* = \mathcal{L}_q(\Omega, \mathcal{F}_t, P)$  is the dual of  $\mathcal{Z}_t$ . For  $Z = (Z_1, \dots, Z_T) \in \mathcal{Z}$  and  $\zeta = (\zeta_1, \dots, \zeta_T) \in \mathcal{Z}^*$  their scalar product is defined in the natural way:

$$\langle \zeta, Z \rangle := \sum_{t=1}^T \int_{\Omega} \zeta_t(\omega) Z_t(\omega) dP(\omega). \quad (6.276)$$

Note that  $\mathcal{Z}$  can be equipped with a norm, consistent with  $\|\cdot\|_p$  norms of its components, which makes it a Banach space. For example, we can use  $\|Z\| := \sum_{t=1}^T \|Z_t\|_p$ . This norm induces the dual norm  $\|\zeta\|^* = \max\{\|\zeta_1\|_q, \dots, \|\zeta_T\|_q\}$  on the space  $\mathcal{Z}^*$ .

## 6.7. Multistage Risk Averse Optimization

321

Consider a function  $\varrho : \mathcal{Z} \rightarrow \mathbb{R}$ . For such a function it makes sense to talk about conditions (R1), (R2), and (R4) defined in section 6.3, with  $Z \succeq Z'$  understood componentwise. We say that  $\varrho(\cdot)$  is a *multi-period risk measure* if it satisfies the respective conditions (R1), (R2), and (R4). Similarly to the analysis of section 6.3, we have the following results. By Theorem 7.79 it follows from convexity (condition (R1)) and monotonicity (condition (R2)), and since  $\varrho(\cdot)$  is real valued, that  $\varrho(\cdot)$  is continuous. By the Fenchel–Moreau theorem, we have that convexity, continuity, and positive homogeneity (condition (R4)) imply the dual representation

$$\varrho(Z) = \sup_{\zeta \in \mathfrak{A}} \langle \zeta, Z \rangle, \quad \forall Z \in \mathcal{Z}, \quad (6.277)$$

where  $\mathfrak{A}$  is a convex, bounded, and weakly\* closed subset of  $\mathcal{Z}^*$  (and hence, by the Banach–Alaoglu theorem,  $\mathfrak{A}$  is weakly\* compact). Moreover, it is possible to show, exactly in the same way as in the proof of Theorem 6.4, that condition (R2) holds iff  $\zeta \succeq 0$  for every  $\zeta \in \mathfrak{A}$ . Conversely, if  $\varrho$  is given in the form (6.277) with  $\mathfrak{A}$  being a convex weakly\* compact subset of  $\mathcal{Z}^*$  such that  $\zeta \succeq 0$  for every  $\zeta \in \mathfrak{A}$ , then  $\varrho$  is a (real valued) multi-period risk measure. An analogue of the condition (R3) (translation equivariance) is more involved; we will discuss this later.

For any multi-period risk measure  $\varrho$ , we can formulate the risk averse multistage program

$$\begin{aligned} \text{Min}_{x_1, x_2, \dots, x_T} \quad & \varrho(f_1(x_1), f_2(x_2(\omega), \omega), \dots, f_T(x_T(\omega), \omega)) \\ \text{s.t.} \quad & x_1 \in \mathcal{X}_1, \quad x_t(\omega) \in \mathcal{X}_t(x_{t-1}(\omega), \omega), \quad t = 2, \dots, T, \end{aligned} \quad (6.278)$$

where optimization is performed over  $\mathcal{F}_t$ -measurable  $x_t : \Omega \rightarrow \mathbb{R}, t = 1, \dots, T$ , satisfying the corresponding constraints, and such that  $f_t(x_t(\cdot), \cdot) \in \mathcal{Z}_t$ . The nonanticipativity is enforced here by the  $\mathcal{F}_t$ -measurability of  $x_t(\omega)$ .

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 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 the current state of the system, we already know that only scenarios passing through this node can happen in the future. Therefore, apart from the nonanticipativity constraint, it is also reasonable to think about the following concept, which we refer to as the *time consistency* principle:

- At every state of the system, optimality of our decisions should not depend on scenarios which we already know cannot happen in the future.

In order to formalize this concept of time consistency we need to say, of course, what we optimize (say, minimize) at every state of the process, i.e., to formulate a respective optimality criterion associated with every state of the system. The risk neutral formulation (3.3) of multistage stochastic programming, discussed in Chapter 3, automatically satisfies the time consistency requirement (see below). The risk averse case is more involved and needs discussion. We say that multi-period risk measure  $\varrho$  is *time consistent* if the corresponding multistage problem (6.278) satisfies the above principle of time consistency.

Consider the class of functionals  $\varrho : \mathcal{Z} \rightarrow \mathbb{R}$  of the form (6.232), i.e., functionals representable as

$$\varrho(Z_1, \dots, Z_T) = Z_1 + \rho_{2|\mathcal{F}_1} \left[ Z_2 + \dots + \rho_{T-1|\mathcal{F}_{T-2}} \left[ Z_{T-1} + \rho_{T|\mathcal{F}_{T-1}} [Z_T] \right] \right], \quad (6.279)$$

where  $\rho_{t+1|\mathcal{F}_t} : \mathcal{Z}_{t+1} \rightarrow \mathcal{Z}_t$ ,  $t = 1, \dots, T-1$ , is a sequence of conditional risk mappings. It is not difficult to see that conditions (R'1), (R'2), and (R'4) (defined in section 6.7.2), applied to every conditional risk mapping  $\rho_{t+1|\mathcal{F}_t}$ , imply respective conditions (R1), (R2), and (R4) for the functional  $\varrho$  of the form (6.279). That is, (6.279) defines a particular class of multiperiod risk measures.

Of course, for  $\varrho$  of the form (6.279), optimization problem (6.278) coincides with the nested formulation (6.263). Recall that if the set  $\Omega$  is finite, then we can formulate multistage risk averse optimization in the framework of scenario trees. As it was discussed in section 6.7.1, nested formulation (6.263) is implied by the approach where with every node of the scenario tree is associated a coherent risk measure applied to the next stage of the scenario tree. In particular, this allows us to write the corresponding dynamic programming equations and implies that an associated optimal policy has the decomposition property. That is, if the process reached a certain node at stage  $t$ , then the remaining decisions of the optimal policy are also optimal with respect to this node considered as the starting point of the process. It follows that the multiperiod risk measure of the form (6.279) is time consistent and the corresponding approach to risk averse optimization satisfies the time consistency principle.

It is interesting and important to give an intrinsic characterization of the nested approach to multiperiod risk measures. Unfortunately, this seems to be too difficult and we will give only a partial answer to this question. Let observe first that for any  $Z = (Z_1, \dots, Z_T) \in \mathcal{Z}$ ,

$$\mathbb{E}[Z_1 + \dots + Z_T] = Z_1 + \mathbb{E}_{|\mathcal{F}_1} \left[ Z_2 + \dots + \mathbb{E}_{|\mathcal{F}_{T-1}} [Z_{T-1} + \mathbb{E}_{|\mathcal{F}_T} [Z_T]] \right], \quad (6.280)$$

where  $\mathbb{E}_{|\mathcal{F}_t}[\cdot] = \mathbb{E}[\cdot | \mathcal{F}_t]$  are the corresponding conditional expectation operators. That is, the expectation risk measure  $\varrho(Z_1, \dots, Z_T) := \mathbb{E}[Z_1 + \dots + Z_T]$  is time consistent and the risk neutral formulation (3.3) of multistage stochastic programming satisfies the time consistency principle.

Consider the following condition:

**(R3-d)** For any  $Z = (Z_1, \dots, Z_T) \in \mathcal{Z}$ ,  $Y_t \in \mathcal{Z}_t$ ,  $t = 1, \dots, T-1$ , and  $a \in \mathbb{R}$  it holds that

$$\varrho(Z_1, \dots, Z_t, Z_{t+1} + Y_t, \dots, Z_T) = \varrho(Z_1, \dots, Z_t + Y_t, Z_{t+1}, \dots, Z_T), \quad (6.281)$$

$$\varrho(Z_1 + a, \dots, Z_T) = a + \varrho(Z_1, \dots, Z_T). \quad (6.282)$$

**Proposition 6.42.** Let  $\varrho : \mathcal{Z} \rightarrow \mathbb{R}$  be a multiperiod risk measure. Then the following conditions (i)–(iii) are equivalent:

(i) There exists a coherent risk measure  $\bar{\rho} : \mathcal{Z}_T \rightarrow \mathbb{R}$  such that

$$\varrho(Z_1, \dots, Z_T) = \bar{\rho}(Z_1 + \dots + Z_T) \quad \forall (Z_1, \dots, Z_T) \in \mathcal{Z}. \quad (6.283)$$

(ii) Condition (R3-d) is fulfilled.

## 6.7. Multistage Risk Averse Optimization

323

(iii) *There exists a nonempty, convex, bounded, and weakly\* closed subset  $\mathfrak{A}_T$  of probability density functions  $\mathfrak{P}_T \subset \mathfrak{Z}_T^*$  such that the dual representation (6.277) holds with the corresponding set  $\mathfrak{A}$  of the form*

$$\mathfrak{A} = \{(\zeta_1, \dots, \zeta_T) : \zeta_T \in \mathfrak{A}_T, \zeta_t = \mathbb{E}[\zeta_T | \mathcal{F}_t], t = 1, \dots, T-1\}. \quad (6.284)$$

**Proof.** If condition (i) is satisfied, then for any  $Z = (Z_1, \dots, Z_T) \in \mathfrak{Z}$  and  $Y_t \in \mathfrak{Z}_t$ ,

$$\begin{aligned} \varrho(Z_1, \dots, Z_t, Z_{t+1} + Y_t, \dots, Z_T) &= \bar{\rho}(Z_1 + \dots + Z_T + Y_t) \\ &= \varrho(Z_1, \dots, Z_t + Y_t, Z_{t+1}, \dots, Z_T). \end{aligned}$$

Property (6.282) also follows by condition (R3) of  $\bar{\rho}$ . That is, condition (i) implies condition (R3-d).

Conversely, suppose that condition (R3-d) holds. Then for  $Z = (Z_1, Z_2, \dots, Z_T)$  we have that  $\varrho(Z_1, Z_2, \dots, Z_T) = \varrho(0, Z_1 + Z_2, \dots, Z_T)$ . Continuing in this way, we obtain that

$$\varrho(Z_1, \dots, Z_T) = \varrho(0, \dots, 0, Z_1 + \dots + Z_T).$$

Define

$$\bar{\rho}(W_T) := \varrho(0, \dots, 0, W_T), \quad W_T \in \mathfrak{Z}_T.$$

Conditions (R1), (R2), and (R4) for  $\varrho$  imply the respective conditions for  $\bar{\rho}$ . Moreover, for  $a \in \mathbb{R}$  we have

$$\begin{aligned} \bar{\rho}(W_T + a) &= \varrho(0, \dots, 0, W_T + a) = \varrho(0, \dots, a, W_T) = \dots = \varrho(a, \dots, 0, W_T) \\ &= a + \varrho(0, \dots, 0, W_T) = \bar{\rho}(W_T) + a. \end{aligned}$$

That is,  $\bar{\rho}$  is a coherent risk measure, and hence (ii) implies (i).

Now suppose that condition (i) holds. By the dual representation (see Theorem 6.4 and Proposition 6.5), there exists a convex, bounded, and weakly\* closed set  $\mathfrak{A}_T \subset \mathfrak{P}_T$  such that

$$\bar{\rho}(W_T) = \sup_{\zeta_T \in \mathfrak{A}_T} \langle \zeta_T, W_T \rangle, \quad W_T \in \mathfrak{Z}_T. \quad (6.285)$$

Moreover, for  $W_T = Z_1 + \dots + Z_T$  we have  $\langle \zeta_T, W_T \rangle = \sum_{t=1}^T \mathbb{E}[\zeta_T Z_t]$ , and since  $Z_t$  is  $\mathcal{F}_t$ -measurable,

$$\mathbb{E}[\zeta_T Z_t] = \mathbb{E}[\mathbb{E}[\zeta_T Z_t | \mathcal{F}_t]] = \mathbb{E}[Z_t \mathbb{E}[\zeta_T | \mathcal{F}_t]]. \quad (6.286)$$

That is, (i) implies (iii). Conversely, suppose that (iii) holds. Then (6.285) defines a coherent risk measure  $\bar{\rho}$ . The dual representation (6.277) together with (6.284) imply (6.283). This shows that conditions (i) and (iii) are equivalent.  $\square$

As we know, condition (i) of the above proposition is necessary for the multiperiod risk measure  $\varrho$  to be representable in the nested form (6.279). (See section 6.7.3 and equation (6.265) in particular.) This condition, however, is not sufficient. It seems to be quite difficult to give a complete characterization of coherent risk measures  $\bar{\rho}$  representable in the form

$$\bar{\rho}(Z_1 + \dots + Z_T) = Z_1 + \rho_{2|\mathcal{F}_1} \left[ Z_2 + \dots + \rho_{T-1|\mathcal{F}_{T-2}} \left[ Z_{T-1} + \rho_{T|\mathcal{F}_{T-1}} [Z_T] \right] \right] \quad (6.287)$$

for all  $Z = (Z_1, \dots, Z_T) \in \mathcal{Z}$ , and some sequence  $\rho_{t+1|\mathcal{F}_t} : \mathcal{Z}_{t+1} \rightarrow \mathcal{Z}_t$ ,  $t = 1, \dots, T-1$ , of conditional risk mappings.

**Remark 26.** Of course, condition  $\zeta_t = \mathbb{E}[\zeta_T|\mathcal{F}_t]$ ,  $t = 1, \dots, T-1$ , of (6.284) can be written as

$$\zeta_t = \mathbb{E}[\zeta_{t+1}|\mathcal{F}_t], \quad t = 1, \dots, T-1. \quad (6.288)$$

That is, if representation (6.283) holds for some coherent risk measure  $\bar{\rho}(\cdot)$ , then any element  $(\zeta_1, \dots, \zeta_T)$  of the dual set  $\mathcal{A}$ , in the representation (6.277) of  $\varrho(\cdot)$ , forms a *martingale* sequence.

**Example 6.43.** Let  $\rho_{\tau|\mathcal{F}_{\tau-1}} : \mathcal{Z}_{\tau} \rightarrow \mathcal{Z}_{\tau-1}$  be a conditional risk mapping for some  $2 \leq \tau \leq T$ , and let  $\rho_1(Z_1) := Z_1$ ,  $Z_1 \in \mathbb{R}$ , and  $\rho_t|\mathcal{F}_{t-1} := \mathbb{E}_{|\mathcal{F}_{t-1}}$ ,  $t = 2, \dots, T$ ,  $t \neq \tau$ . That is, we take here all conditional risk mappings to be the respective conditional expectations except (an arbitrary) conditional risk mapping  $\rho_{\tau|\mathcal{F}_{\tau-1}}$  at the period  $t = \tau$ . It follows that

$$\begin{aligned} \varrho(Z_1, \dots, Z_T) &= \mathbb{E}[Z_1 + \dots + Z_{\tau-1} + \rho_{\tau|\mathcal{F}_{\tau-1}}[\mathbb{E}_{|\mathcal{F}_{\tau}}[Z_{\tau} + \dots + Z_T]]] \\ &= \mathbb{E}[\rho_{\tau|\mathcal{F}_{\tau-1}}[\mathbb{E}_{|\mathcal{F}_{\tau}}[Z_1 + \dots + Z_T]]]. \end{aligned} \quad (6.289)$$

That is,

$$\bar{\rho}(W_T) = \mathbb{E}[\rho_{\tau|\mathcal{F}_{\tau-1}}[\mathbb{E}_{|\mathcal{F}_{\tau}}[W_T]]], \quad W_T \in \mathcal{Z}_T, \quad (6.290)$$

is the corresponding (composite) coherent risk measure.

Coherent risk measures of the form (6.290) have the following property:

$$\bar{\rho}(W_T + Y_{\tau-1}) = \bar{\rho}(W_T) + \mathbb{E}[Y_{\tau-1}], \quad \forall W_T \in \mathcal{Z}_T, \quad \forall Y_{\tau-1} \in \mathcal{Z}_{\tau-1}. \quad (6.291)$$

By (6.284) the above condition (6.291) means that the corresponding set  $\mathcal{A}$ , defined in (6.284), has the additional property that  $\zeta_t = \mathbb{E}[\zeta_T] = 1$ ,  $t = 1, \dots, \tau-1$ , i.e., these components of  $\zeta \in \mathcal{A}$  are constants (equal to one).

In particular, for  $\tau = T$  the composite risk measure (6.290) becomes

$$\bar{\rho}(W_T) = \mathbb{E}[\rho_{T|\mathcal{F}_{T-1}}[W_T]], \quad W_T \in \mathcal{Z}_T. \quad (6.292)$$

Further, let  $\rho_{T|\mathcal{F}_{T-1}} : \mathcal{Z}_T \rightarrow \mathcal{Z}_{T-1}$  be the conditional mean absolute deviation, i.e.,

$$\rho_{T|\mathcal{F}_{T-1}}[Z_T] := \mathbb{E}_{|\mathcal{F}_{T-1}}[Z_T + c | Z_T - \mathbb{E}_{|\mathcal{F}_{T-1}}[Z_T]|], \quad (6.293)$$

$c \in [0, 1/2]$ . The corresponding composite coherent risk measure here is

$$\bar{\rho}(W_T) = \mathbb{E}[W_T] + c \mathbb{E}[|W_T - \mathbb{E}_{|\mathcal{F}_{T-1}}[W_T]|], \quad W_T \in \mathcal{Z}_T. \quad (6.294)$$

For  $T > 2$  the risk measure (6.294) is different from the mean absolute deviation measure

$$\tilde{\rho}(W_T) := \mathbb{E}[W_T] + c \mathbb{E}[|W_T - \mathbb{E}[W_T]|], \quad W_T \in \mathcal{Z}_T, \quad (6.295)$$

and that the multiperiod risk measure

$$\varrho(Z_1, \dots, Z_T) := \tilde{\rho}(Z_1 + \dots + Z_T) = \mathbb{E}[Z_1 + \dots + Z_T] + c \mathbb{E}[|Z_1 + \dots + Z_T - \mathbb{E}[Z_1 + \dots + Z_T]|]$$

corresponding to (6.295) is not time consistent. ■



### Risk Averse Multistage Portfolio Selection

We discuss now the example of portfolio selection. A nested formulation of multistage portfolio selection can be written as

$$\begin{aligned} \text{Min } & \left\{ \bar{\rho}(-W_T) := \rho_1 \left[ \cdots \rho_{T-1|W_{T-2}} \left[ \rho_{T|W_{T-1}} [-W_T] \right] \right] \right\} \\ \text{s.t. } & W_{t+1} = \sum_{i=1}^n \xi_{i,t+1} x_{it}, \quad \sum_{i=1}^n x_{it} = W_t, \quad x_t \geq 0, \quad t = 0, \dots, T-1. \end{aligned} \quad (6.296)$$

We use here conditional risk mappings formulated in terms of the respective conditional expectations, like the conditional AV@R (see (6.270)) and conditional mean semideviations (see (6.272)), 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 $_{\alpha}(\cdot | \xi_{[t-1]})$  corresponds the respective (unconditional) AV@R $_{\alpha}(\cdot)$ . If we set  $\rho_{t|W_{t-1}} := \mathbb{E}_{|W_{t-1}}, t = 1, \dots, T$ , then since

$$\mathbb{E}[\cdots \mathbb{E}[\mathbb{E}[-W_T | W_{T-1}] | W_{T-2}]] = \mathbb{E}[-W_T],$$

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  $\xi_{it}$ .

Suppose that the random process  $\xi_t$  is *stagewise independent*. Let us write dynamic programming equations. At the last stage we have to solve problem

$$\begin{aligned} \text{Min}_{x_{T-1} \geq 0, W_T} & \rho_{T|W_{T-1}} [-W_T] \\ \text{s.t. } & W_T = \sum_{i=1}^n \xi_{iT} x_{i,T-1}, \quad \sum_{i=1}^n x_{i,T-1} = W_{T-1}. \end{aligned} \quad (6.297)$$

Since  $W_{T-1}$  is a function of  $\xi_{[T-1]}$ , by the stagewise independence we have that  $\xi_T$ , and hence  $W_T$ , are independent of  $W_{T-1}$ . It follows by positive homogeneity of  $\rho_T$  that the optimal value of (6.297) is  $Q_{T-1}(W_{T-1}) = W_{T-1} v_{T-1}$ , where  $v_{T-1}$  is the optimal value of

$$\begin{aligned} \text{Min}_{x_{T-1} \geq 0, W_T} & \rho_T [-W_T] \\ \text{s.t. } & W_T = \sum_{i=1}^n \xi_{iT} x_{i,T-1}, \quad \sum_{i=1}^n x_{i,T-1} = 1, \end{aligned} \quad (6.298)$$

and an optimal solution of (6.297) is  $\bar{x}_{T-1}(W_{T-1}) = W_{T-1} x_{T-1}^*$ , where  $x_{T-1}^*$  is an optimal solution of (6.298). Continuing in this way, 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

$$\begin{aligned} \text{Min}_{x_t \geq 0, W_{t+1}} & \rho_{t+1} [-W_{t+1}] \\ \text{s.t. } & W_{t+1} = \sum_{i=1}^n \xi_{i,t+1} x_{it}, \quad \sum_{i=1}^n x_{it} = 1 \end{aligned} \quad (6.299)$$

(compare with section 1.4.3). Note that the composite risk measure  $\bar{\rho}$  can be quite complicated here.

An alternative, multiperiod risk averse approach can be formulated as

$$\begin{aligned} & \text{Min } \rho[-W_T] \\ & \text{s.t. } W_{t+1} = \sum_{i=1}^n \xi_{i,t+1} x_{it}, \quad \sum_{i=1}^n x_{it} = W_t, \quad x_t \geq 0, \quad t = 0, \dots, T-1, \end{aligned} \quad (6.300)$$

for an explicitly defined risk measure  $\rho$ . Let, for example,

$$\rho(\cdot) := (1 - \beta)\mathbb{E}[\cdot] + \beta \text{AV@R}_\alpha(\cdot), \quad \beta \in [0, 1], \quad \alpha \in (0, 1). \quad (6.301)$$

Then problem (6.300) becomes

$$\begin{aligned} & \text{Min } (1 - \beta)\mathbb{E}[-W_T] + \beta(-r + \alpha^{-1}\mathbb{E}[r - W_T]_+) \\ & \text{s.t. } W_{t+1} = \sum_{i=1}^n \xi_{i,t+1} x_{it}, \quad \sum_{i=1}^n x_{it} = W_t, \quad x_t \geq 0, \quad t = 0, \dots, T-1, \end{aligned} \quad (6.302)$$

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

$$\begin{aligned} & \text{Min}_{x_{T-1} \geq 0, W_T} \mathbb{E}[-(1 - \beta)W_T + \beta\alpha^{-1}[r - W_T]_+] \\ & \text{s.t. } W_T = \sum_{i=1}^n \xi_{iT} x_{i,T-1}, \quad \sum_{i=1}^n x_{i,T-1} = W_{T-1}. \end{aligned} \quad (6.303)$$

Proceeding in this way, at stages  $t = T-2, \dots, 1$  we consider the problems

$$\begin{aligned} & \text{Min}_{x_t \geq 0, W_{t+1}} \mathbb{E}\{Q_{t+1}(W_{t+1}, r)\} \\ & \text{s.t. } W_{t+1} = \sum_{i=1}^n \xi_{i,t+1} x_{it}, \quad \sum_{i=1}^n x_{it} = W_t, \end{aligned} \quad (6.304)$$

whose optimal value is denoted  $Q_t(W_t, r)$ . Finally, at stage  $t = 0$  we solve the problem

$$\begin{aligned} & \text{Min}_{x_0 \geq 0, r, W_1} -\beta r + \mathbb{E}[Q_1(W_1, r)] \\ & \text{s.t. } W_1 = \sum_{i=1}^n \xi_{i1} x_{i0}, \quad \sum_{i=1}^n x_{i0} = W_0. \end{aligned} \quad (6.305)$$

In the above multiperiod risk averse approach, the optimal policy is not myopic and the property of time consistency is not satisfied.

### Risk Averse Multistage Inventory Model

Consider the multistage inventory problem (1.17). The nested risk averse formulation of that problem can be written as

$$\begin{aligned} \text{Min}_{x_t \geq y_t} \quad & c_1(x_1 - y_1) + \rho_1 \left[ \psi_1(x_1, D_1) + c_2(x_2 - y_2) + \rho_{2|D_{[1]}} [\psi_2(x_2, D_2) + \cdots \right. \\ & \left. + c_{T-1}(x_{T-1} - y_{T-1}) + \rho_{T-1|D_{[T-2]}} [\psi_{T-1}(x_{T-1}, D_{T-1}) \right. \\ & \left. + c_T(x_T - y_T) + \rho_{T|D_{[T-1]}} [\psi_T(x_T, D_T)] \right] \\ \text{s.t.} \quad & y_{t+1} = x_t - D_t, \quad t = 1, \dots, T-1, \end{aligned} \quad (6.306)$$

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  $\rho_{t|D_{[t-1]}}(\cdot)$ ,  $t = 2, \dots, T$ , 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 discussed earlier, there are two equivalent interpretations of problem (6.306). We can write it as an optimization problem with respect to feasible policies  $\mathbf{x}_t(d_{[t-1]})$  (compare with (6.273)):

$$\begin{aligned} \text{Min}_{x_1, \mathbf{x}_2, \dots, \mathbf{x}_T} \quad & c_1(x_1 - y_1) + \rho_1 \left[ \psi_1(x_1, D_1) + c_2(\mathbf{x}_2(D_1) - x_1 + D_1) \right. \\ & \left. + \rho_{2|D_1} [\psi_2(\mathbf{x}_2(D_1), D_2) + \cdots \right. \\ & \left. + c_{T-1}(\mathbf{x}_{T-1}(D_{[T-2]}) - \mathbf{x}_{T-2}(D_{[T-3]}) + D_{T-2}) \right. \\ & \left. + \rho_{T-1|D_{[T-2]}} [\psi_{T-1}(\mathbf{x}_{T-1}(D_{[T-2]}), D_{T-1}) \right. \\ & \left. + c_T(\mathbf{x}_T(D_{[T-1]}) - \mathbf{x}_{T-1}(D_{[T-2]}) + D_{T-1}) \right. \\ & \left. + \rho_{T|D_{[T-1]}} [\psi_T(\mathbf{x}_T(D_{[T-1]}), D_T)] \right] \\ \text{s.t.} \quad & x_1 \geq y_1, \quad \mathbf{x}_2(D_1) \geq x_1 - D_1, \\ & \mathbf{x}_t(D_{[t-1]}) \geq \mathbf{x}_{t-1}(D_{[t-2]}) - D_{t-1}, \quad t = 3, \dots, T. \end{aligned} \quad (6.307)$$

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

$$\text{Min}_{x_T \geq y_T} c_T(x_T - y_T) + \rho_{T|D_{[T-1]}} [\psi_T(x_T, D_T)]. \quad (6.308)$$

The optimal value of problem (6.308) is denoted  $Q_T(y_T, D_{[T-1]})$ . Continuing in this way, we write for  $t = T-1, \dots, 2$  the following dynamic programming equations:

$$Q_t(y_t, D_{[t-1]}) = \min_{x_t \geq y_t} c_t(x_t - y_t) + \rho_{t|D_{[t-1]}} [\psi_t(x_t, D_t) + Q_{t+1}(x_t - D_t, D_{[t]})]. \quad (6.309)$$

Finally, at the first stage we need to solve the problem

$$\text{Min}_{x_1 \geq y_1} c_1(x_1 - y_1) + \rho_1 [\psi_1(x_1, D_1) + Q_2(x_1 - D_1, D_1)]. \quad (6.310)$$

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  $Q_t(y_t, d_{[t-1]}) = Q_t(y_t)$ ,  $t = 2, \dots, T$ , 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

$$\min_{x_t} c_t x_t + \rho_t [\psi(x_t, D_t) + Q_{t+1}(x_t - D_t)]. \quad (6.311)$$

Recall that  $\rho_t$  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  $\beta_1, \beta_2$  are positive constants. Show that  $\phi_1, \phi_2$ , and  $\phi$  are real valued convex functions with subdifferentials

$$\partial \phi_1(t) = [H^-(t), H(t)] \quad \text{and} \quad \partial \phi_2(t) = [-1 + H^-(t), -1 + H(t)],$$

$$\partial \phi(t) = [(\beta_1 + \beta_2)H^-(t) - \beta_2, (\beta_1 + \beta_2)H(t) - \beta_2].$$

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. (i) Let  $Y \sim \mathcal{N}(\mu, \sigma^2)$ . Show that

$$\text{V@R}_\alpha(Y) = \mu + z_\alpha \sigma, \quad (6.312)$$

where  $z_\alpha := \Phi^{-1}(1 - \alpha)$ , and

$$\text{AV@R}_\alpha(Y) = \mu + \frac{\sigma}{\alpha \sqrt{2\pi}} e^{-z_\alpha^2/2}. \quad (6.313)$$

(ii) Let  $Y^1, \dots, Y^N$  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  $\theta^* = \text{AV@R}_\alpha(Y)$ , defined on page 300.

- 6.3. Consider the chance constraint

$$\Pr \left\{ \sum_{i=1}^n \xi_i x_i \geq b \right\} \geq 1 - \alpha, \quad (6.314)$$

where  $\xi \sim \mathcal{N}(\mu, \Sigma)$  (see problem (1.43)). Note that this constraint can be written as

$$\text{V@R}_\alpha \left( b - \sum_{i=1}^n \xi_i x_i \right) \leq 0. \quad (6.315)$$

Consider the following constraint:

$$\text{AV@R}_\gamma \left( b - \sum_{i=1}^n \xi_i x_i \right) \leq 0. \quad (6.316)$$

Show that constraints (6.314) and (6.316) are equivalent if  $z_\alpha = \frac{1}{\gamma\sqrt{2\pi}} e^{-z_\gamma^2/2}$ .

- 6.4. Consider the function  $\phi(x) := \text{AV@R}_\alpha(F_x)$ , where  $F_x = F_x(\omega) = F(x, \omega)$  is a real valued random variable, on a probability space  $(\Omega, \mathcal{F}, P)$ , 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, and (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

$$\phi'(x_0, d) = \alpha^{-1} \inf_{t \in [a, b]} \mathbb{E} \{ d^\top \nabla_x ([F(x_0, \omega) - t]_+) \}, \quad (6.317)$$

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,  $a = b = \text{V@R}_\alpha(F_{x_0})$ , then  $\phi(\cdot)$  is differentiable at  $x_0$  and

$$\nabla \phi(x_0) = \alpha^{-1} \mathbb{E} [\mathbf{1}_{\{F_{x_0} > a\}}(\omega) \nabla_x F(x_0, \omega)]. \quad (6.318)$$

*Hint:* Use Theorem 7.44 together with the Danskin theorem, Theorem 7.21.

- 6.5. Show that the set of saddle points of the minimax problem (6.190) is given by  $\{\mu\} \times [\gamma^*, \gamma^{**}]$ , where  $\gamma^*$  and  $\gamma^{**}$  are defined in (6.192).
- 6.6. Consider the absolute semideviation risk measure

$$\rho_c(Z) := \mathbb{E} \{ Z + c[Z - \mathbb{E}(Z)]_+ \}, \quad Z \in \mathcal{L}_1(\Omega, \mathcal{F}, P),$$

where  $c \in [0, 1]$ , and the following risk averse optimization problem:

$$\text{Min}_{x \in X} \underbrace{\mathbb{E} \{ G(x, \xi) + c[G(x, \xi) - \mathbb{E}(G(x, \xi))]_+ \}}_{\rho_c[G(x, \xi)]}. \quad (6.319)$$

Viewing the optimal value of problem (6.319) as the Von Mises statistical functional of the probability measure  $P$ , compute its influence function.

*Hint:* Use derivations of section 6.5.3 together with the Danskin theorem.

- 6.7. Consider the risk averse optimization problem (6.162) related to the inventory model. Let the corresponding risk measure be of the form  $\rho_\lambda(Z) = \mathbb{E}[Z] + \lambda \mathbb{D}(Z)$ , where  $\mathbb{D}(Z)$  is a measure of variability of  $Z = Z(\omega)$  and  $\lambda$  is a nonnegative trade-off coefficient between expectation and variability. Higher values of  $\lambda$  reflect a higher degree of risk aversion. Suppose that  $\rho_\lambda$  is a coherent risk measure for all  $\lambda \in [0, 1]$  and let  $S_\lambda$  be the set of optimal solutions of the corresponding risk averse problem. Suppose that the sets  $S_0$  and  $S_1$  are nonempty.

Show that if  $S_0 \cap S_1 = \emptyset$ , then  $S_\lambda$  is monotonically nonincreasing or monotonically nondecreasing in  $\lambda \in [0, 1]$  depending on 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  $\lambda \in (0, 1)$ .

6.8. Consider the news vendor problem with cost function

$$F(x, d) = cx + b[d - x]_+ + h[x - d]_+, \quad \text{where } b > c \geq 0, \quad h > 0,$$

and the minimax problem

$$\text{Min}_{x \geq 0} \sup_{H \in \mathfrak{M}} \mathbb{E}_H[F(x, D)], \quad (6.320)$$

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  $\bar{d} \in [l, u]$ . 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  $\bar{H} = p\Delta(l) + (1 - p)\Delta(u)$ , where  $p = (u - \bar{d})/(u - l)$ , i.e., the cdf  $\bar{H}(\cdot)$  is the step function

$$\bar{H}(z) = \begin{cases} 0 & \text{if } z < l, \\ p & \text{if } l \leq z < u, \\ 1 & \text{if } u \leq z. \end{cases}$$

Conclude that  $\bar{H}$  is the cdf specified in Proposition 6.38 and that  $\bar{x} = \bar{H}^{-1}(\kappa)$ , where  $\kappa = (b - c)/(b + h)$ , is the optimal solution of problem (6.320). That is,  $\bar{x} = l$  if  $\kappa < p$  and  $\bar{x} = u$  if  $\kappa > p$ , where  $\kappa = \frac{b-c}{b+h}$ .

6.9. Consider the following version of the news vendor problem. A news vendor 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  $0 \leq r < c < s$ . If the demand  $D$  turns out to be greater than or equal to the order quantity  $x$ , then he makes 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

$$F(x, D) = \begin{cases} (s - c)x & \text{if } x \leq D, \\ (r - c)x + (s - r)D & \text{if } x > D. \end{cases} \quad (6.321)$$

(a) Assuming that demand  $D \geq 0$  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

$$f(x) = (s - c)x - (s - r) \int_0^x H(z) dz. \quad (6.322)$$

Conclude that the set of optimal solutions of the problem

$$\text{Max}_{x \geq 0} \{f(x) := \mathbb{E}_H[F(x, D)]\} \quad (6.323)$$

is an interval given by the set of  $\kappa$ -quantiles of the cdf  $H(\cdot)$  with  $\kappa := (s - c)/(s - r)$ .

(b) Consider the following risk averse version of the news vendor problem:

$$\text{Min}_{x \geq 0} \{\phi(x) := \rho[-F(x, D)]\}. \quad (6.324)$$

Here  $\rho$  is a real valued coherent risk measure representable in the form (6.165) and  $H^*$  is the corresponding reference cdf.

(i) Show that the function  $\phi(x) = \rho[-F(x, D)]$  can be represented in the form

$$\phi(x) = (c - s)x + (s - r) \int_0^x \bar{H}(z) dz \quad (6.325)$$

for some cdf  $\bar{H}$ .

(ii) Show that if  $\rho(\cdot) := \text{AV@R}_\alpha(\cdot)$ , then  $\bar{H}(z) = \max\{\alpha^{-1}H^*(z), 1\}$ . Conclude that in that case, optimal solutions of the risk averse problem (6.324) are smaller than the risk neutral problem (6.323).

6.10. Let  $\mathcal{Z}_i := \mathcal{L}_p(\Omega, \mathcal{F}_i, P)$ ,  $i = 1, 2$ , with  $\mathcal{F}_1 \subset \mathcal{F}_2$ , and let  $\rho : \mathcal{Z}_2 \rightarrow \mathcal{Z}_1$ .

(a) Show that if  $\rho$  is a conditional risk mapping,  $Y \in \mathcal{Z}_1$  and  $Y \geq 0$ , then  $\rho(YZ) = Y\rho(Z)$  for any  $Z \in \mathcal{Z}_2$ .

(b) Suppose that the mapping  $\rho$  satisfies conditions (R'1)–(R'3), but not necessarily the positive homogeneity condition (R'4). Show that it can be represented in the form

$$[\rho(Z)](\omega) = \sup_{\mu \in \mathfrak{C}} \{\mathbb{E}_\mu[Z|\mathcal{F}_1](\omega) - [\rho^*(\mu)](\omega)\}, \quad (6.326)$$

where  $\mathfrak{C}$  is a set of probability measures on  $(\Omega, \mathcal{F}_2)$  and

$$[\rho^*(\mu)](\omega) = \sup_{Z \in \mathcal{Z}_2} \{\mathbb{E}_\mu[Z|\mathcal{F}_1](\omega) - [\rho(Z)](\omega)\}. \quad (6.327)$$

You may assume that  $\mathcal{F}_1$  has a countable number of elementary events.

6.11. Consider the following risk averse approach to multistage portfolio selection. Let  $\xi_1, \dots, \xi_T$  be the respective data process (of random returns) and consider the following chance constrained nested formulation:

$$\begin{aligned} & \text{Max } \mathbb{E}[W_T] \\ & \text{s.t. } W_{t+1} = \sum_{i=1}^n \xi_{i,t+1} x_{it}, \quad \sum_{i=1}^n x_{it} = W_t, \quad x_t \geq 0, \\ & \quad \Pr\{W_{t+1} \geq \kappa W_t \mid \xi_{[t]}\} \geq 1 - \alpha, \quad t = 0, \dots, T-1, \end{aligned} \quad (6.328)$$

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

$$\begin{aligned} & \text{Max}_{x_{T-1} \geq 0, W_T} \mathbb{E}[W_T \mid \xi_{[T-1]}] \\ & \text{s.t. } W_T = \sum_{i=1}^n \xi_{iT} x_{i,T-1}, \quad \sum_{i=1}^n x_{i,T-1} = W_{T-1}, \\ & \quad \Pr\{W_T \geq \kappa W_{T-1} \mid \xi_{[T-1]}\}, \end{aligned} \quad (6.329)$$

and at stage  $t = T - 2, \dots, 1$ , the cost-to-go function  $Q_t(W_t, \xi_{[t]})$  is given by the optimal value of the problem

$$\begin{aligned} \text{Max}_{x_t \geq 0, W_{t+1}} \quad & \mathbb{E}[Q_{t+1}(W_{t+1}, \xi_{[t+1]}) \mid \xi_{[t]}] \\ \text{s.t.} \quad & W_{t+1} = \sum_{i=1}^n \xi_{i,t+1} x_{i,t}, \quad \sum_{i=1}^n x_{i,t} = W_t, \\ & \Pr \{W_{t+1} \geq \kappa W_t \mid \xi_{[t]}\}. \end{aligned} \quad (6.330)$$

Assuming that the process  $\xi_t$  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

$$\begin{aligned} \text{Max}_{x_t \geq 0} \quad & \sum_{i=1}^n \mathbb{E}[\xi_{i,t+1}] x_{i,t} \\ \text{s.t.} \quad & \sum_{i=1}^n x_{i,t} = 1, \quad \Pr \left\{ \sum_{i=1}^n \xi_{i,t+1} x_{i,t} \geq \kappa \right\} \geq 1 - \alpha. \end{aligned} \quad (6.331)$$



## Chapter 7

# Background Material

*Alexander Shapiro*

In this chapter we discuss some concepts and results from convex analysis, probability, functional analysis, and optimization theories needed for a development of the material in this book. Of course, a careful derivation of the required material goes far beyond the scope of this book. We give or outline proofs of some results while others are referred to the literature. Of course, this choice is somewhat subjective.

We denote by  $\mathbb{R}^n$  the standard  $n$ -dimensional vector space, of (column) vectors  $x = (x_1, \dots, x_n)^\top$ , equipped with the scalar product  $x^\top y = \sum_{i=1}^n x_i y_i$ . Unless stated otherwise, we denote by  $\|\cdot\|$  the Euclidean norm  $\|x\| = \sqrt{x^\top x}$ . The notation  $A^\top$  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  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is defined as

$$\text{dom } f := \{x \in \mathbb{R}^n : f(x) < +\infty\}.$$

It is said that  $f$  is *proper* if  $f(x) > -\infty$  for all  $x \in \mathbb{R}^n$  and its domain,  $\text{dom } f$ , is nonempty. The function  $f$  is said to be *lower semicontinuous* at a point  $x_0 \in \mathbb{R}^n$  if  $f(x_0) \leq \liminf_{x \rightarrow x_0} f(x)$ . It is said that  $f$  is lower semicontinuous if it is lower semicontinuous at every point of  $\mathbb{R}^n$ . The largest lower semicontinuous function which is less than or equal to  $f$  is denoted  $\text{lsc } f$ . It is not difficult to show that  $f$  is lower semicontinuous iff its epigraph

$$\text{epi } f := \{(x, \alpha) \in \mathbb{R}^{n+1} : f(x) \leq \alpha\}$$

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  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  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*<sup>55</sup> function

$$\mathbf{1}_A(x) := \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A \end{cases} \quad (7.1)$$

and by  $\mathbb{I}_A(\cdot)$  the *indicator* function

$$\mathbb{I}_A(x) := \begin{cases} 0 & \text{if } x \in A, \\ +\infty & \text{if } x \notin A \end{cases} \quad (7.2)$$

of set  $A$ .

By  $\text{cl}(A)$  we denote the topological closure of set  $A \subset \mathbb{R}^n$ . For sets  $A, B \subset \mathbb{R}^n$  we denote by

$$\text{dist}(x, A) := \inf_{x' \in A} \|x - x'\| \quad (7.3)$$

the distance from  $x \in \mathbb{R}^n$  to  $A$ , and by

$$\mathbb{D}(A, B) := \sup_{x \in A} \text{dist}(x, B) \quad \text{and} \quad \mathbb{H}(A, B) := \max \{\mathbb{D}(A, B), \mathbb{D}(B, A)\} \quad (7.4)$$

the *deviation* of the set  $A$  from the set  $B$  and the *Hausdorff distance* between the sets  $A$  and  $B$ , respectively. By the definition,  $\text{dist}(x, A) = +\infty$  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  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . 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

$$g'(x_0, h) := \lim_{t \downarrow 0} \frac{g(x_0 + th) - g(x_0)}{t} \quad (7.5)$$

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  $t \geq 0$ . 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

$$g(x_0 + h) = g(x_0) + g'(x_0, h) + r(h), \quad (7.6)$$

where the remainder term  $r(h)$  is such that  $r(th)/t \rightarrow 0$ , 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  $r(h)/\|h\| \rightarrow 0$  as  $h \rightarrow 0$ , 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  $x_0$ .

Clearly, Fréchet differentiability implies Gâteaux differentiability. The converse of that is not necessarily true. However, the following theorem shows that for locally Lipschitz

<sup>55</sup>Function  $\mathbf{1}_A(\cdot)$  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  $\mathbb{I}_A(\cdot)$ .

continuous mappings both concepts do coincide. Recall that a mapping (function)  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is said to be *Lipschitz continuous* on a set  $X \subset \mathbb{R}^n$  if there is a constant  $c \geq 0$  such that

$$\|g(x_1) - g(x_2)\| \leq c\|x_1 - x_2\|, \quad \forall x_1, x_2 \in X.$$

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  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  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*

$$\lim_{h \rightarrow 0} \frac{g(x_0 + h) - g(x_0) - g'(x_0, h)}{\|h\|} = 0. \quad (7.7)$$

**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

$$\|g(x_0 + th_1) - g(x_0 + th_2)\| \leq ct\|h_1 - h_2\|.$$

It follows that  $\|g'(x_0, h_1) - g'(x_0, h_2)\| \leq c\|h_1 - h_2\|$  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

$$g(x_0 + t_k h_k) - g(x_0) = (g(x_0 + t_k h) - g(x_0)) + (g(x_0 + t_k h_k) - g(x_0 + t_k h))$$

and

$$\|g(x_0 + t_k h_k) - g(x_0 + t_k h)\| \leq ct_k \|h_k - h\|$$

for all  $k$  large enough. It follows that

$$g'(x_0, h) = \lim_{k \rightarrow \infty} \frac{g(x_0 + t_k h_k) - g(x_0)}{t_k}. \quad (7.8)$$

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.  $\square$

We have that  $g$  is differentiable at a point  $x \in \mathbb{R}^n$  iff

$$g(x + h) - g(x) = [\nabla g(x)]h + o(h), \quad (7.9)$$

where  $\nabla g(x)$  is the so-called  $m \times n$  Jacobian matrix of partial derivatives  $[\partial g_i(x)/\partial x_j]$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, n$ . If  $m = 1$ , i.e.,  $g(x)$  is real valued, we call  $\nabla g(x)$  the *gradient* of  $g$  at  $x$ . In that case, (7.9) takes the form

$$g(x + h) - g(x) = h^\top \nabla g(x) + o(h). \quad (7.10)$$

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  $f : U \rightarrow \mathbb{R}$  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  $\lim_{k \rightarrow \infty} \nabla f(x_k)$  such that  $x_k \rightarrow \bar{x}$  and  $f$  is differentiable at  $x_k$ . This set is nonempty and compact, and 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

$$f^\circ(\bar{x}, d) := \limsup_{\substack{x \rightarrow \bar{x} \\ t \downarrow 0}} \frac{f(x + td) - f(x)}{t}. \quad (7.11)$$

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,

$$f^\circ(\bar{x}, d) = \sup_{z \in \partial^\circ f(\bar{x})} z^\top d, \quad \forall d \in \mathbb{R}^n. \quad (7.12)$$

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  $h : \mathbb{R}^m \rightarrow \mathbb{R}^n$  and  $g : \mathbb{R}^n \rightarrow \mathbb{R}$ , and assume that  $g$  and  $h$  are locally Lipschitz continuous. Then

$$\partial^\circ f(x) \subset \text{cl} \left\{ \text{conv} \left( \sum_{i=1}^n \alpha_i v_i : \alpha \in \partial^\circ g(y), v_i \in \partial^\circ h_i(x), i = 1, \dots, n \right) \right\}, \quad (7.13)$$

where  $\alpha = (\alpha_1, \dots, \alpha_n)$ ,  $y = h(x)$  and  $h_1, \dots, h_n$  are components of  $h$ . The equality in (7.13) holds true if any one of the following conditions is satisfied: (i)  $g$  and  $h_i, i = 1, \dots, n$ , are Clarke-regular and every element in  $\partial^\circ g(y)$  has nonnegative components, (ii)  $g$  is differentiable and  $n = 1$ , and (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  $\text{int}(C)$ .

The *convex hull* of  $C$ , denoted  $\text{conv}(C)$ , is the smallest convex set including  $C$ . It is said that  $C$  is a *cone* if for any  $x \in C$  and  $t \geq 0$  it follows that  $tx \in C$ . The *polar cone* of a cone  $C \subset \mathbb{R}^n$  is defined as

$$C^* := \{z \in \mathbb{R}^n : z^\top x \leq 0, \quad \forall x \in C\}. \quad (7.14)$$

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  $\text{ri}(C)$ . Note that if the interior of  $C$  is nonempty, then the affine space generated by  $C$  coincides with  $\mathbb{R}^n$ , and hence in that case  $\text{ri}(C) = \text{int}(C)$ . 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

$$s(h) := \sup_{z \in C} z^\top h. \quad (7.15)$$

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  $\text{cl}(\text{conv} C)$ . 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

$$\mathcal{N}_C(x_0) := \{z : z^\top (x - x_0) \leq 0, \quad \forall x \in C\}. \quad (7.16)$$

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  $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ . It is not difficult to show that  $f$  is convex iff its epigraph  $\text{epi} f$  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  $x_0$ , and its directional derivative  $f'(x_0, \cdot)$  is an extended real valued convex positively homogeneous function and can be written in the form

$$f'(x_0, h) = \inf_{t>0} \frac{f(x_0 + th) - f(x_0)}{t}. \quad (7.17)$$

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

$$f(x) - f(x_0) \geq z^T(x - x_0), \quad \forall x \in \mathbb{R}^n. \quad (7.18)$$

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 the duality theory of convex analysis we have that if the directional derivative  $f'(x_0, \cdot)$  is lower semicontinuous, then

$$f'(x_0, h) = \sup_{z \in \partial f(x_0)} z^T h, \quad \forall h \in \mathbb{R}^n, \quad (7.19)$$

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  $f_i : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ ,  $i = 1, \dots, m$ , be proper convex functions,  $f(\cdot) := f_1(\cdot) + \dots + f_m(\cdot)$  and  $x_0$  be a point such that  $f_i(x_0)$  are finite, i.e.,  $x_0 \in \cap_{i=1}^m \text{dom } f_i$ . Then*

$$\partial f_1(x_0) + \dots + \partial f_m(x_0) \subset \partial f(x_0). \quad (7.20)$$

Moreover,

$$\partial f_1(x_0) + \dots + \partial f_m(x_0) = \partial f(x_0) \quad (7.21)$$

*if any one of the following conditions holds: (i) the set  $\cap_{i=1}^m \text{ri}(\text{dom } f_i)$  is nonempty, (ii) the functions  $f_1, \dots, f_k$ ,  $k \leq m$ , are polyhedral and the intersection of the sets  $\cap_{i=1}^k \text{dom } f_i$  and  $\cap_{i=k+1}^m \text{ri}(\text{dom } f_i)$  is nonempty, or (iii) there exists a point  $\bar{x} \in \text{dom } f_m$  such that  $\bar{x} \in \text{int}(\text{dom } f_i)$ ,  $i = 1, \dots, m-1$ .*

In particular, if all functions  $f_1, \dots, f_m$  in the above theorem are polyhedral, then (7.21) holds without an additional regularity condition.

Let  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  be an extended real valued function. The *conjugate function* of  $f$  is

$$f^*(z) := \sup_{x \in \mathbb{R}^n} \{z^T x - f(x)\}. \quad (7.22)$$

The conjugate function  $f^* : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  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  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  be a proper extended real valued convex function. Then*

$$f^{**} = \text{lsc } f. \quad (7.23)$$

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

$$z \in \partial f(x) \text{ iff } f^*(z) + f(x) = z^\top x.$$

By applying that to the function  $f^{**}$ , instead of  $f$ , we obtain that  $z \in \partial f^{**}(x)$  iff  $f^{***}(z) + f^{**}(x) = z^\top x$ . Now by the Fenchel–Moreau theorem we have that  $f^{***} = f^*$ , and hence  $z \in \partial f^{**}(x)$  iff  $f^*(z) + f^{**}(x) = z^\top x$ . Consequently, we obtain

$$\partial f^{**}(x) = \arg \max_{z \in \mathbb{R}^n} \{z^\top x - f^*(z)\}, \quad (7.24)$$

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  $f : X \rightarrow \mathbb{R}$  is *strongly convex*, with parameter  $c > 0$ , if<sup>56</sup>

$$tf(x') + (1-t)f(x) \geq f(tx' + (1-t)x) + \frac{1}{2}ct(1-t)\|x' - x\|^2 \quad (7.25)$$

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  $\psi$  means that the inequality

$$\begin{aligned} tf(x') - \frac{1}{2}ct\|x'\|^2 + (1-t)f(x) - \frac{1}{2}c(1-t)\|x\|^2 \\ \geq f(tx' + (1-t)x) - \frac{1}{2}c\|tx' + (1-t)x\|^2 \end{aligned}$$

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  $f : X \rightarrow \mathbb{R}$  is continuous and differentiable at every point  $x \in \text{int}(X)$ , then  $f$  is strongly convex iff

$$f(x') \geq f(x) + (x' - x)^\top \nabla f(x) + \frac{1}{2}c\|x' - x\|^2, \quad \forall x, x' \in \text{int}(X) \quad (7.26)$$

or, equivalently, iff

$$(x' - x)^\top (\nabla f(x') - \nabla f(x)) \geq c\|x' - x\|^2, \quad \forall x, x' \in \text{int}(X). \quad (7.27)$$

### 7.1.3 Optimization and Duality

Consider a real valued function  $L : X \times Y \rightarrow \mathbb{R}$ , where  $X$  and  $Y$  are arbitrary sets. We can associate with the function  $L(x, y)$  the following two optimization problems:

$$\text{Min}_{x \in X} \{f(x) := \sup_{y \in Y} L(x, y)\}, \quad (7.28)$$

$$\text{Max}_{y \in Y} \{g(y) := \inf_{x \in X} L(x, y)\}, \quad (7.29)$$

<sup>56</sup>Unless stated otherwise, we denote by  $\|\cdot\|$  the Euclidean norm on  $\mathbb{R}^n$ .

viewed as dual to each other. We have that for any  $x \in X$  and  $y \in Y$ ,

$$g(y) = \inf_{x' \in X} L(x', y) \leq L(x, y) \leq \sup_{y' \in Y} L(x, y') = f(x),$$

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

$$L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \leq L(x, \bar{y}), \quad \forall (x, y) \in X \times Y. \quad (7.30)$$

**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 iff 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

$$\text{Min}_{x \in X} L(x, \bar{y}), \quad (7.31)$$

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

$$\text{Min}_{x \in \mathbb{R}} x \quad \text{s.t. } x \geq 0. \quad (7.32)$$

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  $\psi : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \bar{\mathbb{R}}$ . Let  $\vartheta(y)$  be the optimal value of the parameterized problem

$$\text{Min}_{x \in \mathbb{R}^n} \psi(x, y), \quad (7.33)$$



i.e.,  $\vartheta(y) := \inf_{x \in \mathbb{R}^n} \psi(x, y)$ . Note that implicitly the optimization in the above problem is performed over the domain of the function  $\psi(\cdot, y)$ , i.e.,  $\text{dom } \psi(\cdot, y)$  can be viewed as the feasible set of problem (7.33).

The conjugate of the function  $\vartheta(y)$  can be expressed in terms of the conjugate of  $\psi(x, y)$ . That is, the conjugate of  $\vartheta$  is

$$\psi^*(x^*, y^*) := \sup_{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m} \{ (x^*)^\top x + (y^*)^\top y - \psi(x, y) \},$$

and hence the conjugate of  $\vartheta$  can be written as

$$\begin{aligned} \vartheta^*(y^*) &:= \sup_{y \in \mathbb{R}^m} \{ (y^*)^\top y - \vartheta(y) \} = \sup_{y \in \mathbb{R}^m} \{ (y^*)^\top y - \inf_{x \in \mathbb{R}^n} \psi(x, y) \} \\ &= \sup_{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m} \{ (y^*)^\top y - \psi(x, y) \} = \psi^*(0, y^*). \end{aligned}$$

Consequently, the conjugate of  $\vartheta^*$  is

$$\vartheta^{**}(y) = \sup_{y^* \in \mathbb{R}^m} \{ (y^*)^\top y - \psi^*(0, y^*) \}. \quad (7.34)$$

This leads to the following dual of (7.33):

$$\text{Max}_{y^* \in \mathbb{R}^m} \{ (y^*)^\top y - \psi^*(0, y^*) \}. \quad (7.35)$$

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. Suppose that  $\vartheta(\cdot)$  is convex. Then we have by the Fenchel–Moreau theorem that either  $\vartheta^{**}(\cdot)$  is identically  $-\infty$ , or

$$\vartheta^{**}(y) = (\text{lsc } \vartheta)(y), \quad \forall y \in \mathbb{R}^m. \quad (7.36)$$

It follows that  $\vartheta^{**}(y) \leq \vartheta(y)$  for any  $y \in \mathbb{R}^m$ . It is said that there is no duality gap between (7.33) and its dual (7.35) if  $\vartheta^{**}(y) = \vartheta(y)$ .

Suppose now that the function  $\psi(x, y)$  is *convex* (as a function of  $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ ). Then it is straightforward to verify that the optimal value function  $\vartheta(y)$  is also convex. It is said that the problem (7.33) is *subconsistent* for a given value of  $y$  if  $\text{lsc } \vartheta(y) < +\infty$ . If problem (7.33) is feasible, i.e.,  $\text{dom } \psi(\cdot, y)$  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  $\vartheta^{**}(y) = \vartheta(y)$  iff 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 iff  $\vartheta(y)$  is finite and  $\vartheta(\cdot)$  is continuous at  $y$ .*

A few words about the above statements are now in order. 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  $\partial\vartheta^{**}(y) = \partial\vartheta(y)$ .

The above analysis can be also used 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  $y \in \mathbb{R}^m$  be a given point. Then the following holds: (i) The optimal value function  $\vartheta(\cdot)$  is subdifferentiable at  $y$  iff  $\vartheta(\cdot)$  is lower semicontinuous at  $y$  and the dual problem (7.35) possesses an optimal solution. (ii) The subdifferential  $\partial\vartheta(y)$  is nonempty and bounded iff  $\vartheta(y)$  is finite and the set of optimal solutions of the dual problem (7.35) is nonempty and bounded. (iii) In both above cases  $\partial\vartheta(y)$  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  $y \in \text{int}(\text{dom } \vartheta)$ . The condition  $y \in \text{int}(\text{dom } \vartheta)$  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.

As an example, let us consider the problem

$$\begin{aligned} \text{Min}_{x \in X} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) + y_i \leq 0, \quad i = 1, \dots, m, \end{aligned} \quad (7.37)$$

where  $X$  is a subset of  $\mathbb{R}^n$ ,  $f(x)$  and  $g_i(x)$  are real valued functions, and  $y = (y_1, \dots, y_m)$  is a vector of parameters. We can formulate this problem in the form (7.33) by defining

$$\psi(x, y) := \bar{f}(x) + F(G(x) + y),$$

where  $\bar{f}(x) := f(x) + \mathbb{I}_X(x)$  (recall that  $\mathbb{I}_X$  denotes the indicator function of the set  $X$ ) and  $F(\cdot)$  is the indicator function of the negative orthant, i.e.,  $F(z) := 0$  if  $z_i \leq 0, i = 1, \dots, m$ , and  $F(z) := +\infty$  otherwise, and  $G(x) := (g_1(x), \dots, g_m(x))$ .

Suppose that the problem (7.37) is convex, that is, the set  $X$  and the functions  $f(x)$  and  $g_i(x), i = 1, \dots, m$ , 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  $\psi(x, y)$ ,

$$\begin{aligned} \psi^*(x^*, y^*) &= \sup_{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m} \{((x^*)^\top x + (y^*)^\top y - \bar{f}(x) - F(G(x) + y))\} \\ &= \sup_{x \in \mathbb{R}^n} \left\{ (x^*)^\top x - \bar{f}(x) - (y^*)^\top G(x) + \sup_{y \in \mathbb{R}^m} [(y^*)^\top (G(x) + y) - F(G(x) + y)] \right\}. \end{aligned}$$

By change of variables  $z = G(x) + y$  we obtain that

$$\sup_{y \in \mathbb{R}^m} [(y^*)^\top (G(x) + y) - F(G(x) + y)] = \sup_{z \in \mathbb{R}^m} [(y^*)^\top z - F(z)] = \mathbb{I}_{\mathbb{R}_+^m}(y^*).$$

Therefore we obtain

$$\psi^*(x^*, y^*) = \sup_{x \in X} \{(x^*)^\top x - L(x, y^*)\} + \mathbb{I}_{\mathbb{R}_+^m}(y^*),$$

where  $L(x, y^*) := f(x) + \sum_{i=1}^m y_i^* g_i(x)$  is the Lagrangian of the problem. Consequently, the dual of the problem (7.37) can be written in the form

$$\text{Max}_{\lambda \geq 0} \left\{ \lambda^T y + \inf_{x \in X} L(x, \lambda) \right\}. \quad (7.38)$$

Note that we changed the notation from  $y^*$  to  $\lambda$  in order to emphasize that the above problem (7.38) is the standard Lagrangian dual of (7.37) with  $\lambda$  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.

As another example, consider a function  $L : \mathbb{R}^n \times Y \rightarrow \overline{\mathbb{R}}$ , where  $Y$  is a vector space (not necessarily finite dimensional), and the corresponding pair of dual problems (7.28) and (7.29). Define

$$\varphi(y, z) := \sup_{x \in \mathbb{R}^n} \{ z^T x - L(x, y) \}, \quad (y, z) \in Y \times \mathbb{R}^n. \quad (7.39)$$

Note that the problem

$$\text{Max}_{y \in Y} \{-\varphi(y, 0)\} \quad (7.40)$$

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 lower semicontinuous. 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 (7.40), of the form (7.35), coincides with the problem (7.28). This leads to the following result.

**Theorem 7.10.** *Let  $Y$  be an abstract vector space and  $L : \mathbb{R}^n \times Y \rightarrow \overline{\mathbb{R}}$ . 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, and (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 (7.39), and the corresponding optimal value function

$$\vartheta(z) := \inf_{y \in Y} \varphi(y, z). \quad (7.41)$$

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  $\vartheta^{**} : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$

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.  $\square$

**Remark 27.** Note that it follows from the lower semicontinuity of  $L(\cdot, y)$  that the max-function  $f(x) = \sup_{y \in Y} L(x, y)$  is also lower semicontinuous. 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  $a = (a_1, \dots, a_m)^\top \in \mathbb{R}^m$ , we use notation  $(a)_+$  componentwise, i.e.,  $(a)_+ := ([a_1]_+, \dots, [a_m]_+)^\top$ , where  $[a_i]_+ := \max\{0, a_i\}$ .

**Theorem 7.11 (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  $\kappa$ , depending on  $A$ , such that for any  $x \in \mathbb{R}^n$  and any  $b \in \text{dom } \mathcal{M}$ ,*

$$\text{dist}(x, \mathcal{M}(b)) \leq \kappa \|(Ax - b)_+\|. \quad (7.42)$$

**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  $\|a\| = \sup_{\|z\|^* \leq 1} z^\top a$ , where  $\|\cdot\|^*$  is the dual of the norm  $\|\cdot\|$ . Then we have

$$\text{dist}(x, \mathcal{M}(b)) = \inf_{x' \in \mathcal{M}(b)} \|x - x'\| = \inf_{Ax' \leq b} \sup_{\|z\|^* \leq 1} z^\top (x - x') = \sup_{\|z\|^* \leq 1} \inf_{Ax' \leq b} z^\top (x - x'),$$

where the interchange of the min and max operators can be justified, for example, by applying Theorem 7.10 (see Remark 27 on page 344). By making change of variables  $y = x - x'$  and using linear programming duality we obtain

$$\inf_{Ax' \leq b} z^\top (x - x') = \inf_{Ay \geq Ax - b} z^\top y = \sup_{\lambda \geq 0, A^\top \lambda = z} \lambda^\top (Ax - b).$$

It follows that

$$\text{dist}(x, \mathcal{M}(b)) = \sup_{\lambda \geq 0, \|A^\top \lambda\|^* \leq 1} \lambda^\top (Ax - b). \quad (7.43)$$

Since any two norms on  $\mathbb{R}^n$  are equivalent, we can assume without loss of generality that  $\|\cdot\|$  is the  $\ell_1$  norm, and hence its dual is the  $\ell_\infty$  norm. For such choice of a polyhedral norm, we have that the set  $S := \{\lambda : \lambda \geq 0, \|A^\top \lambda\|^* \leq 1\}$  is polyhedral. We obtain that the right-hand side of (7.43) 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 (7.43) is finite), and hence has an optimal solution  $\bar{\lambda}$ . It follows that

$$\text{dist}(x, \mathcal{M}(b)) = \bar{\lambda}^\top (Ax - b) \leq \bar{\lambda}^\top (Ax - b)_+ \leq \|\bar{\lambda}\|^* \|(Ax - b)_+\|.$$

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  $\bar{\lambda}$  can be taken to be an extreme point of the polyhedral set  $S_0$ . Consequently, (7.42) follows with  $\kappa := \max_{\lambda \in S_0} \|\lambda\|^*$ .  $\square$

The term  $\|(Ax - b)_+\|$ , in the right-hand side of (7.42), measures the infeasibility of the point  $x$ .

Consider now the following linear programming problem:

$$\text{Min}_{x \in \mathbb{R}^n} c^\top x \quad \text{s.t.} \quad Ax \leq b. \quad (7.44)$$

A slight variation of the proof of Hoffman's lemma leads to the following result.

**Theorem 7.12.** *Let  $S(b)$  be the set of optimal solutions of problem (7.44). Then there exists a positive constant  $\gamma$ , depending only on  $A$ , such that for any  $b, b' \in \text{dom } S$  and any  $x \in S(b)$ ,*

$$\text{dist}(x, S(b')) \leq \gamma \|b - b'\|. \quad (7.45)$$

**Proof.** Problem (7.44) can be written in the following equivalent form:

$$\text{Min}_{t \in \mathbb{R}} t \quad \text{s.t.} \quad Ax \leq b, \quad c^\top x - t \leq 0. \quad (7.46)$$

Denote by  $\mathcal{M}(b)$  the set of feasible points of problem (7.46), i.e.,

$$\mathcal{M}(b) := \{(x, t) : Ax \leq b, \quad c^\top x - t \leq 0\}.$$

Let  $b, b' \in \text{dom } S$  and consider a point  $(x, t) \in \mathcal{M}(b)$ . Proceeding as in the proof of Theorem 7.11 we can write

$$\text{dist}((x, t), \mathcal{M}(b')) = \sup_{\|(z, a)\|^* \leq 1} \inf_{Ax' \leq b', \quad c^\top x' \leq t'} z^\top (x - x') + a(t - t').$$

By changing variables  $y = x - x'$  and  $s = t - t'$  and using linear programming duality, we have

$$\inf_{Ax' \leq b', \quad c^\top x' \leq t'} z^\top (x - x') + a(t - t') = \sup_{\lambda \geq 0, \quad A^\top \lambda + ac = z} \lambda^\top (Ax - b') + a(c^\top x - t)$$

for  $a \geq 0$ , and for  $a < 0$  the above minimum is  $-\infty$ . By using  $\ell_1$  norm  $\|\cdot\|$ , and hence  $\ell_\infty$  norm  $\|\cdot\|^*$ , we obtain that

$$\text{dist}((x, t), \mathcal{M}(b')) = \bar{\lambda}^\top (Ax - b') + \bar{a}(c^\top x - t),$$

where  $(\bar{\lambda}, \bar{a})$  is an optimal solution of the problem

$$\text{Max}_{\lambda \geq 0, a \geq 0} \lambda^\top (Ax - b') + a(c^\top x - t) \quad \text{s.t.} \quad \|A^\top \lambda + ac\|^* \leq 1, \quad a \leq 1. \quad (7.47)$$

By normalizing  $c$  we can assume without loss of generality that  $\|c\|^* \leq 1$ . Then by replacing the constraint  $\|A^\top \lambda + ac\|^* \leq 1$  with the constraint  $\|A^\top \lambda\|^* \leq 2$  we increase the feasible set

of problem (7.47) 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  $S := \{\lambda : \|A^\top \lambda\|^* \leq 2\}$ . 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  $\gamma$  which depends only on  $A$ . Since  $(x, t) \in \mathcal{M}(b)$ , and hence  $Ax - b \leq 0$  and  $c^\top x - t \leq 0$ , we have

$$\hat{\lambda}^\top (Ax - b') = \hat{\lambda}^\top (Ax - b) + \hat{\lambda}^\top (b - b') \leq \hat{\lambda}^\top (b - b') \leq \|\hat{\lambda}\|^* \|b - b'\|$$

and  $\hat{a}(c^\top x - t) \leq 0$ , and hence

$$\text{dist}((x, t), \mathcal{M}(b')) \leq \|\hat{\lambda}\|^* \|b - b'\| \leq \gamma \|b - b'\|. \quad (7.48)$$

The above inequality implies (7.45).  $\square$

### 7.1.4 Optimality Conditions

Consider the optimization problem

$$\text{Min}_{x \in X} f(x), \quad (7.49)$$

where  $X \subset \mathbb{R}^n$  and  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is an extended real valued function.

#### First Order Optimality Conditions

**Convex Case.** Suppose that the function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  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  $f(x) \geq f(\bar{x})$  for all  $x \in \mathbb{R}^n$  iff

$$0 \in \partial f(\bar{x}). \quad (7.50)$$

That is, condition (7.50) 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  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is proper and convex, and consider a point  $\bar{x} \in X \cap \text{dom } f$ . It follows that the function  $\bar{f}(x) := f(x) + \mathbb{I}_X(x)$  is convex, and of course the point  $\bar{x}$  is an optimal solution of the problem (7.49) iff  $\bar{x}$  is a (global) minimizer of  $\bar{f}(x)$ . Suppose that

$$\text{ri}(X) \cap \text{ri}(\text{dom } f) \neq \emptyset. \quad (7.51)$$

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 (7.49) iff

$$0 \in \partial f(\bar{x}) + \mathcal{N}_X(\bar{x}), \quad (7.52)$$

provided that the regularity condition (7.51) holds. Note that (7.51) holds, in particular, if  $\bar{x} \in \text{int}(\text{dom } f)$ .

**Nonconvex Case.** Assume that the function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is real valued continuously differentiable and the set  $X$  is closed (not necessarily convex).

**Definition 7.13.** 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  $h_k \rightarrow h$  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 (7.49).

**Proposition 7.14.** Let  $\bar{x} \in X$  be a locally optimal solution of problem (7.49). Then

$$h^\top \nabla f(\bar{x}) \geq 0, \quad \forall h \in \mathcal{T}_X(\bar{x}). \quad (7.53)$$

**Proof.** Consider  $h \in \mathcal{T}_X(\bar{x})$  and let  $h_k \rightarrow h$  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  $f(x_k) - f(\bar{x}) \geq 0$ . We also have that

$$f(x_k) - f(\bar{x}) = t_k h^\top \nabla f(\bar{x}) + o(t_k),$$

and hence (7.53) follows.  $\square$

Condition (7.53) 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 (7.52) and (7.53) are equivalent.

Suppose now that the set  $X$  is given in the form

$$X := \{x \in \mathbb{R}^n : G(x) \in K\}, \quad (7.54)$$

where  $G(\cdot) = (g_1(\cdot), \dots, g_m(\cdot)) : \mathbb{R}^n \rightarrow \mathbb{R}^m$  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 (7.54) becomes

$$X = \{x \in \mathbb{R}^n : g_i(x) = 0, \ i = 1, \dots, q, \ g_i(x) \leq 0, \ i = q+1, \dots, m\}. \quad (7.55)$$

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  $\bar{x} \in X$ :

$$\mathcal{T}_X(\bar{x}) = \{h \in \mathbb{R}^n : [\nabla G(\bar{x})]h \in \mathcal{T}_K(G(\bar{x}))\}, \quad (7.56)$$

where  $\nabla G(\bar{x}) = [\nabla g_1(\bar{x}), \dots, \nabla g_m(\bar{x})]^\top$  is the corresponding  $m \times n$  Jacobian matrix. The following condition is called the *Robinson constraint qualification*:

$$[\nabla G(\bar{x})]\mathbb{R}^n + \mathcal{T}_K(G(\bar{x})) = \mathbb{R}^m. \quad (7.57)$$

If the cone  $K$  has a nonempty interior, Robinson constraint qualification is equivalent to the following condition:

$$\exists h : G(\bar{x}) + [\nabla G(\bar{x})]h \in \text{int}(K). \quad (7.58)$$

In case  $X$  is given in the form (7.55), Robinson constraint qualification is equivalent to the *Mangasarian–Fromovitz constraint qualification*:

$$\begin{aligned} \nabla g_i(\bar{x}), \quad i = 1, \dots, q, \text{ are linearly independent,} \\ \exists h : \quad h^\top \nabla g_i(\bar{x}) = 0, \quad i = 1, \dots, q, \\ \quad \quad h^\top \nabla g_i(\bar{x}) < 0, \quad i \in \mathcal{I}(\bar{x}), \end{aligned} \quad (7.59)$$

where  $\mathcal{I}(\bar{x}) := \{i \in \{q+1, \dots, m\} : g_i(\bar{x}) = 0\}$  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 (7.49) and the constraint mapping  $G(x)$ . Under a constraint qualification ensuring validity of formula (7.56), the first order necessary optimality condition (7.53) can be written in the following dual form: there exists a vector  $\lambda \in \mathbb{R}^m$  of Lagrange multipliers such that

$$\nabla_x L(\bar{x}, \lambda) = 0, \quad G(\bar{x}) \in K, \quad \lambda \in K^*, \quad \lambda^\top G(\bar{x}) = 0. \quad (7.60)$$

Denote by  $\Lambda(\bar{x})$  the set of Lagrange multipliers vectors  $\lambda$  satisfying (7.60).

**Theorem 7.15.** *Let  $\bar{x}$  be a locally optimal solution of problem (7.49). 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 (7.55), then optimality conditions (7.60) 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 we denote by  $\nabla^2 f(x)$  the Hessian matrix of second order partial derivatives of  $f$  at  $x$ . Let  $\bar{x}$  be a locally optimal solution of problem (7.49). Consider the set (cone)

$$C(\bar{x}) := \{h \in \mathcal{T}_X(\bar{x}) : h^\top \nabla f(\bar{x}) = 0\}. \quad (7.61)$$

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

$$\mathcal{T}_X^2(x, h) := \{z \in \mathbb{R}^n : \text{dist}(x + th + \tfrac{1}{2}t^2 z, X) = o(t^2), \quad t \geq 0\} \quad (7.62)$$

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  $x + th + \frac{1}{2}t^2 z + r(t) \in X$  for some  $r(t) = o(t^2)$ ,  $t \geq 0$ . 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 7.16.** *Let  $\bar{x}$  be a locally optimal solution of problem (7.49). Then<sup>57</sup>*

$$h^\top \nabla^2 f(\bar{x}) h - s(-\nabla f(\bar{x}), \mathcal{T}_X^2(\bar{x}, h)) \geq 0, \quad \forall h \in C(\bar{x}). \quad (7.63)$$

<sup>57</sup>Recall that  $s(v, A) = \sup_{z \in A} z^\top v$  denotes the support function of set  $A$ .



**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  $x(t) + r(t) \in X$ ,  $t \geq 0$ . It follows by local optimality of  $\bar{x}$  that  $f(x(t) + r(t)) - f(\bar{x}) \geq 0$  for all  $t \geq 0$  small enough. Since  $r(t) = o(t^2)$ , by the second order Taylor expansion we have

$$f(x(t) + r(t)) - f(\bar{x}) = th^\top \nabla f(\bar{x}) + \frac{1}{2}t^2 [z^\top \nabla f(\bar{x}) + h^\top \nabla^2 f(\bar{x})h] + o(t^2).$$

Since  $h \in C(\bar{x})$ , the first term in the right-hand side of the above equation vanishes. It follows that

$$z^\top \nabla f(\bar{x}) + h^\top \nabla^2 f(\bar{x})h \geq 0, \quad \forall h \in C(\bar{x}), \forall z \in \mathcal{T}_X^2(\bar{x}, h). \quad (7.64)$$

Condition (7.64) can be written in the form

$$\inf_{z \in \mathcal{T}_X^2(\bar{x}, h)} \{z^\top \nabla f(\bar{x}) + h^\top \nabla^2 f(\bar{x})h\} \geq 0, \quad \forall h \in C(\bar{x}). \quad (7.65)$$

Since

$$\inf_{z \in \mathcal{T}_X^2(\bar{x}, h)} z^\top \nabla f(\bar{x}) = - \sup_{z \in \mathcal{T}_X^2(\bar{x}, h)} z^\top (-\nabla f(\bar{x})) = -s(-\nabla f(\bar{x}), \mathcal{T}_X^2(\bar{x}, h)),$$

the second order necessary conditions (7.64) can be written in the form (7.63).  $\square$

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  $h^\top \nabla f(\bar{x}) = 0$  and because of the first order optimality conditions (7.53), it follows that if the set  $X$  is polyhedral, then the term  $s(-\nabla f(\bar{x}), \mathcal{T}_X^2(\bar{x}, h))$  in (7.63) 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 (7.54) with the mapping  $G(x)$  being twice continuously differentiable, then the second order optimality conditions (7.63) can be written in the following dual form.

**Theorem 7.17.** *Let  $\bar{x}$  be a locally optimal solution of problem (7.49). Suppose that the Robinson constraint qualification (7.57) is fulfilled. Then the following second order necessary conditions hold:*

$$\sup_{\lambda \in \Lambda(\bar{x})} \{h^\top \nabla_{xx}^2 L(\bar{x}, \lambda)h - s(\lambda, \mathfrak{T}(h))\} \geq 0, \quad \forall h \in C(\bar{x}), \quad (7.66)$$

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 (7.66) vanishes. In general,  $s(\lambda, \mathfrak{T}(h)) \leq 0$  and the second order necessary conditions (7.66) are stronger than the “standard” second order conditions:

$$\sup_{\lambda \in \Lambda(\bar{x})} h^\top \nabla_{xx}^2 L(\bar{x}, \lambda)h \geq 0, \quad \forall h \in C(\bar{x}). \quad (7.67)$$

**Second Order Sufficient Conditions.** Consider condition

$$h^T \nabla^2 f(\bar{x}) h - s(-\nabla f(\bar{x}), \mathcal{T}_X^2(\bar{x}, h)) > 0, \quad \forall h \in C(\bar{x}), h \neq 0. \quad (7.68)$$

This condition is obtained from the second order necessary condition (7.63) by replacing the “ $\geq$ ” sign with the strict inequality sign “ $>$ ”. Necessity of second order conditions (7.63) 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 (7.68) we need an additional condition.

**Definition 7.18.** *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  $t_k r_k \rightarrow 0$ , it follows that*

$$\lim_{k \rightarrow \infty} \text{dist}(r_k, \mathcal{T}_X^2(\bar{x}, h)) = 0. \quad (7.69)$$

Note that in the above definition the term  $\frac{1}{2} t_k^2 r_k = o(t_k)$ , and hence such a sequence  $x_k \in X$  can exist only if  $h \in T_X(\bar{x})$ . 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 [22, 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

$$f(x) \geq f(\bar{x}) + c \|x - \bar{x}\|^2, \quad \forall x \in X \cap N. \quad (7.70)$$

Of course, the quadratic growth condition implies that  $\bar{x}$  is a locally optimal solution of problem (7.49).

**Proposition 7.19.** *Let  $\bar{x} \in X$  be a feasible point of problem (7.49) satisfying first order necessary conditions (7.53). Suppose that  $X$  is second order regular at  $\bar{x}$ . Then the second order conditions (7.68) are necessary and sufficient for the quadratic growth at  $\bar{x}$  to hold.*

**Proof.** Suppose that conditions (7.68) 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

$$f(x_k) - f(\bar{x}) \leq c_k \|x_k - \bar{x}\|^2. \quad (7.71)$$

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 (7.71) we have

$$c_k t_k^2 \geq f(x_k) - f(\bar{x}) = t_k h^T \nabla f(\bar{x}) + o(t_k),$$

and hence  $h^T \nabla f(\bar{x}) \leq 0$ . Because of the first order necessary conditions it follows that  $h^T \nabla f(\bar{x}) = 0$ , 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  $t_k r_k \rightarrow 0$ . 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  $r_k - z_k \rightarrow 0$ . Since  $h^\top \nabla f(\bar{x}) = 0$ , by the second order Taylor expansion we have

$$f(x_k) = f(\bar{x} + t_k h + \frac{1}{2}t_k^2 r_k) = f(\bar{x}) + \frac{1}{2}t_k^2 [z_k^\top \nabla f(\bar{x}) + h^\top \nabla^2 f(\bar{x})h] + o(t_k^2).$$

Moreover, since  $z_k \in \mathcal{T}_X^2(\bar{x}, h)$  we have that

$$z_k^\top \nabla f(\bar{x}) + h^\top \nabla^2 f(\bar{x})h \geq c,$$

where  $c$  is equal to the left-hand side of (7.68), which by the assumption is positive. It follows that

$$f(x_k) \geq f(\bar{x}) + \frac{1}{2}c \|x_k - \bar{x}\|^2 + o(\|x_k - \bar{x}\|^2),$$

a contradiction with (7.71).

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  $h^\top \nabla^2 \phi(\bar{x})h = h^\top \nabla^2 f(\bar{x})h - c\|h\|^2$ . Therefore, by the second order necessary conditions (7.63), applied to the function  $\phi$ , it follows that the left-hand side of (7.68) is greater than or equal to  $c\|h\|^2$ . This completes the proof.  $\square$

If the set  $X$  is given in the form (7.54), then similar to Theorem 7.17 it is possible to formulate second order sufficient conditions (7.68) in the following dual form.

**Theorem 7.20.** *Let  $\bar{x} \in X$  be a feasible point of problem (7.49) satisfying first order necessary conditions (7.60). Suppose that the Robinson constraint qualification (7.57) 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:*

$$\sup_{\lambda \in \Lambda(\bar{x})} \{h^\top \nabla_{xx}^2 L(\bar{x}, \lambda)h - s(\lambda, \mathfrak{T}(h))\} > 0, \quad \forall h \in C(\bar{x}), \quad h \neq 0, \quad (7.72)$$

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 (7.72) vanishes.

### 7.1.5 Perturbation Analysis

#### Differentiability Properties of Max-Functions

We often have to deal with optimal value functions, say, max-functions of the form

$$\phi(x) := \sup_{\theta \in \Theta} g(x, \theta), \quad (7.73)$$

where  $g : \mathbb{R}^n \times \Theta \rightarrow \mathbb{R}$ . In applications the set  $\Theta$  usually is a subset of a finite dimensional vector space. At this point, however, this is not important and we can assume that  $\Theta$  is an abstract topological space. Denote

$$\bar{\Theta}(x) := \arg \max_{\theta \in \Theta} g(x, \theta).$$

The following result about directional differentiability of the max-function is often called the Danskin theorem.

**Theorem 7.21 (Danskin).** *Let  $\Theta$  be a nonempty, compact topological space and  $g : \mathbb{R}^n \times \Theta \rightarrow \mathbb{R}$  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*

$$\phi'(x, h) = \sup_{\theta \in \bar{\Theta}(x)} h^T \nabla_x g(x, \theta). \quad (7.74)$$

*In particular, if for some  $x \in \mathbb{R}^n$  the set  $\bar{\Theta}(x) = \{\bar{\theta}\}$  is a singleton, then the max-function is differentiable at  $x$  and*

$$\nabla \phi(x) = \nabla_x g(x, \bar{\theta}). \quad (7.75)$$

In the convex case we have the following result giving a description of subdifferentials of max-functions.

**Theorem 7.22 (Levin–Valadier).** *Let  $\Theta$  be a nonempty compact topological space and  $g : \mathbb{R}^n \times \Theta \rightarrow \mathbb{R}$  be a real valued function. Suppose that (i) for every  $\theta \in \Theta$  the function  $g_\theta(\cdot) = g(\cdot, \theta)$  is convex on  $\mathbb{R}^n$  and (ii) for every  $x \in \mathbb{R}^n$  the function  $g(x, \cdot)$  is upper semicontinuous on  $\Theta$ . Then the max-function  $\phi(x)$  is convex real valued and*

$$\partial \phi(x) = \text{cl} \left\{ \text{conv} \left( \bigcup_{\theta \in \bar{\Theta}(x)} \partial g_\theta(x) \right) \right\}. \quad (7.76)$$

Let us make the following observations regarding the above theorem. Since  $\Theta$  is compact and by the assumption (ii), we have that the set  $\bar{\Theta}(x)$  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 (7.76) that the set  $A := \bigcup_{\theta \in \bar{\Theta}(x)} \partial g_\theta(x)$  is bounded. Suppose further that

(iii) For every  $x \in \mathbb{R}^n$  the function  $g(x, \cdot)$  is continuous on  $\Theta$ .

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  $\theta_k \in \bar{\Theta}(x)$ . Since  $\bar{\Theta}(x)$  is compact and  $A$  is bounded, by passing to a subsequence if necessary, we can assume that  $\theta_k$  converges to a point  $\bar{\theta} \in \bar{\Theta}(x)$  and  $z_k$  converges to a point  $\bar{z} \in \mathbb{R}^n$ . By the definition of subgradients  $z_k$  we have that for any  $x' \in \mathbb{R}^n$  the following inequality holds

$$g_{\theta_k}(x') - g_{\theta_k}(x) \geq z_k^T (x' - x).$$

By passing to the limit in the above inequality as  $k \rightarrow \infty$ , 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, the convex hull of a compact subset of  $\mathbb{R}^n$  is also compact, and hence is closed, we obtain that if assumption (ii) in the above theorem is strengthened to assumption (iii), then the set inside the parentheses in (7.76) is closed, and hence formula (7.76) takes the form

$$\partial \phi(x) = \text{conv} \left( \bigcup_{\theta \in \bar{\Theta}(x)} \partial g_\theta(x) \right). \quad (7.77)$$

### Second Order Perturbation Analysis

Consider the following parameterization of problem (7.49):

$$\text{Min}_{x \in X} f(x) + t\eta_t(x), \quad (7.78)$$

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  $X \subset \text{int}(U)$ . It follows, of course, that the set  $U$  has a nonempty interior. Consider the space  $W^{1,\infty}(U)$  of Lipschitz continuous functions  $\psi : U \rightarrow \mathbb{R}$  equipped with the norm

$$\|\psi\|_{1,U} := \sup_{x \in U} |\psi(x)| + \sup_{x \in U'} \|\nabla \psi(x)\| \quad (7.79)$$

with  $U' \subset \text{int}(U)$  being the set of points where  $\psi(\cdot)$  is differentiable. Recall that by the 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  $\eta_t$  converges (in the norm topology) to a function  $\delta \in W^{1,\infty}(U)$ , that is,  $\|\eta_t - \delta\|_{1,U} \rightarrow 0$  as  $t \downarrow 0$ .

Denote by  $v(t)$  the optimal value and by  $\tilde{x}(t)$  an optimal solution of (7.78), i.e.,

$$v(t) := \inf_{x \in X} \{f(x) + t\eta_t(x)\} \quad \text{and} \quad \tilde{x}(t) \in \arg \min_{x \in X} \{f(x) + t\eta_t(x)\}.$$

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 (7.49) 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 (7.78) is continuous, it has an optimal solution for any  $t$ .

The following result is taken from [22, section 4.10.3].

**Theorem 7.23.** *Let  $\bar{x}$  be unique optimal solution of problem (7.49). Suppose that: (i) the set  $X$  is compact and second order regular at  $\bar{x}$ , (ii)  $\eta_t$  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  $\bar{x}$ , and (iv) the quadratic growth condition (7.70) holds. Then*

$$v(t) = v(0) + t\eta_t(\bar{x}) + \frac{1}{2}t^2\mathfrak{V}_f(\delta) + o(t^2), \quad t \geq 0, \quad (7.80)$$

where  $\mathfrak{V}_f(\delta)$  is the optimal value of the auxiliary problem

$$\text{Min}_{h \in C(\bar{x})} \{2h^\top \nabla \delta(\bar{x}) + h^\top \nabla^2 f(\bar{x})h - s(-\nabla f(\bar{x}), \mathcal{T}_{\bar{X}}^2(\bar{x}, h))\}. \quad (7.81)$$

Moreover, if (7.81) has unique optimal solution  $\bar{h}$ , then

$$\tilde{x}(t) = \bar{x} + t\bar{h} + o(t), \quad t \geq 0. \quad (7.82)$$

**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

$$v(t) = f(\tilde{x}(t)) + t\eta_t(\tilde{x}(t)) \geq f(\bar{x}) + c\|\tilde{x}(t) - \bar{x}\|^2 + t\eta_t(\tilde{x}(t)).$$

Since  $\bar{x} \in X$  we also have that  $v(t) \leq f(\bar{x}) + t\eta_t(\bar{x})$ . Consequently,

$$t|\eta_t(\tilde{x}(t)) - \eta_t(\bar{x})| \geq c\|\tilde{x}(t) - \bar{x}\|^2.$$

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

$$f(x(t)) = f(\bar{x}) + th^\top \nabla f(\bar{x}) + \frac{1}{2}t^2w^\top \nabla f(\bar{x}) + \frac{1}{2}h^\top \nabla^2 f(\bar{x})h + o(t^2),$$

and since  $h \in C(\bar{x})$  we have that  $h^\top \nabla f(\bar{x}) = 0$ . Also since  $\|\eta_t - \delta\|_{1,\infty} \rightarrow 0$ , 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

$$\delta(x(t)) = \delta(\bar{x}) + th^\top \nabla \delta(\bar{x}) + o(t).$$

Putting this all together and noting that  $f(\bar{x}) = v(0)$ , we obtain that

$$f(x(t)) + t\eta_t(x(t)) = v(0) + t\eta_t(\bar{x}) + t^2h^\top \nabla \delta(\bar{x}) + \frac{1}{2}t^2h^\top \nabla^2 f(\bar{x})h + \frac{1}{2}t^2w^\top \nabla f(\bar{x}) + o(t^2).$$

Consequently,

$$\limsup_{t \downarrow 0} \frac{v(t) - v(0) - t\eta_t(\bar{x})}{\frac{1}{2}t^2} \leq 2h^\top \nabla \delta(\bar{x}) + h^\top \nabla^2 f(\bar{x})h + w^\top \nabla f(\bar{x}). \quad (7.83)$$

Since the above inequality (7.83) holds for any  $w \in \mathcal{T}_X^2(\bar{x}, h)$ , by taking minimum (with respect to  $w$ ) in the right-hand side of (7.83) we obtain for any  $h \in C(\bar{x})$ ,

$$\limsup_{t \downarrow 0} \frac{v(t) - v(0) - t\eta_t(\bar{x})}{\frac{1}{2}t^2} \leq 2h^\top \nabla \delta(\bar{x}) + h^\top \nabla^2 f(\bar{x})h - s(-\nabla f(\bar{x}), \mathcal{T}_X^2(\bar{x}, h)).$$

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,

$$v(t_k) = f(x_k) + t_k\eta_{t_k}(x_k) = f(\bar{x}) + t_kh^\top \nabla f(\bar{x}) + t_k\delta(\bar{x}) + o(t_k),$$

and by the Danskin theorem  $v'(0) = \delta(\bar{x})$ . It follows that  $h^\top \nabla f(\bar{x}) = 0$ , 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  $t_k r_k \rightarrow 0$  and  $x_k \in X$  and hence, by the second order regularity of  $X$ , there exists  $w_k \in \mathcal{T}_X^2(\bar{x}, h)$  such that  $\|r_k - w_k\| \rightarrow 0$ . Finally,

$$\begin{aligned} v(t_k) &= f(x_k) + t_k \eta_{t_k}(x_k) \\ &= f(\bar{x}) + t_k \eta_{t_k}(\bar{x}) + t_k^2 h^\top \nabla \delta(\bar{x}) + \frac{1}{2} t_k^2 h^\top \nabla^2 f(\bar{x}) h + \frac{1}{2} t_k^2 w_k^\top \nabla f(\bar{x}) + o(t_k^2) \\ &\geq v(0) + t_k \eta_{t_k}(\bar{x}) + t_k^2 h^\top \nabla \delta(\bar{x}) + \frac{1}{2} t_k^2 h^\top \nabla^2 f(\bar{x}) h \\ &\quad + \frac{1}{2} t_k^2 \inf_{w \in \mathcal{T}_X^2(\bar{x}, h)} w^\top \nabla f(\bar{x}) + o(t_k^2). \end{aligned}$$

It follows that

$$\liminf_{t \downarrow 0} \frac{v(t) - v(0) - t \eta_t(\bar{x})}{\frac{1}{2} t^2} \geq 2h^\top \nabla \delta(\bar{x}) + h^\top \nabla^2 f(\bar{x}) h - s(-\nabla f(\bar{x}), \mathcal{T}_X^2(\bar{x}, h)).$$

This completes the proof of (7.80).

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 (7.81). Since  $(\tilde{x}(t) - \bar{x})/t$  is bounded, the assertion (7.82) follows by compactness arguments.  $\square$

As in the case of second order optimality conditions, we have here that if the set  $X$  is polyhedral, then the curvature term  $s(-\nabla f(\bar{x}), \mathcal{T}_X^2(\bar{x}, h))$  in (7.81) vanishes.

Suppose now that the set  $X$  is given in the form (7.54) with the mapping  $G(x)$  being twice continuously differentiable. Suppose further that the Robinson constraint qualification (7.57), for the unperturbed problem, holds. Then the optimal value of problem (7.81) can be written in the following dual form:

$$\mathfrak{V}_f(\delta) = \inf_{h \in C(\bar{x})} \sup_{\lambda \in \Lambda(\bar{x})} \left\{ 2h^\top \nabla \delta(\bar{x}) + h^\top \nabla_{xx}^2 L(\bar{x}, \lambda) h - s(\lambda, \mathfrak{T}(h)) \right\}, \quad (7.84)$$

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 (7.84) vanishes.

### Minimax Problems

In this section we consider the minimax problem

$$\text{Min}_{x \in X} \left\{ \phi(x) := \sup_{y \in Y} f(x, y) \right\} \quad (7.85)$$

and its dual

$$\text{Max}_{y \in Y} \left\{ \iota(y) := \inf_{x \in X} f(x, y) \right\}. \quad (7.86)$$

We assume that the sets  $X \subset \mathbb{R}^n$  and  $Y \subset \mathbb{R}^m$  are convex and compact, and the function  $f : X \times Y \rightarrow \mathbb{R}$  is continuous,<sup>58</sup> i.e.,  $f \in C(X, Y)$ . Moreover, assume that  $f(x, y)$  is

<sup>58</sup>Recall that  $C(X, Y)$  denotes the space of continuous functions  $\psi : X \times Y \rightarrow \mathbb{R}$  equipped with the sup-norm  $\|\psi\| = \sup_{(x, y) \in X \times Y} |\psi(x, y)|$ .

convex in  $x \in X$  and concave in  $y \in Y$ . Under these conditions there is no duality gap between problems (7.85) and (7.86), 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 (7.85) has a nonempty set of optimal solutions, denoted  $X^*$ , the min-function  $\iota(y)$  is continuous on  $Y$ , and problem (7.86) has a nonempty set of optimal solutions, denoted  $Y^*$ , and  $X^* \times Y^*$  forms the set of saddle points of the minimax problems (7.85) and (7.86).

Consider the following perturbation of the minimax problem (7.85):

$$\text{Min sup}_{x \in X, y \in Y} \{f(x, y) + t\eta_t(x, y)\}, \quad (7.87)$$

where  $\eta_t \in C(X, Y)$ ,  $t \geq 0$ . Denote by  $v(t)$  the optimal value of the parameterized problem (7.87). Clearly  $v(0)$  is the optimal value of the unperturbed problem (7.85). We assume that  $\eta_t$  converges uniformly (i.e., in the sup-norm) as  $t \downarrow 0$  to a function  $\gamma \in C(X, Y)$ , that is

$$\lim_{t \downarrow 0} \sup_{x \in X, y \in Y} |\eta_t(x, y) - \gamma(x, y)| = 0.$$

**Theorem 7.24.** Suppose that (i) the sets  $X \subset \mathbb{R}^n$  and  $Y \subset \mathbb{R}^m$  are convex and compact, (ii) for all  $t \geq 0$  the function  $\zeta_t := f + t\eta_t$  is continuous on  $X \times Y$ , convex in  $x \in X$  and concave in  $y \in Y$ , and (iii)  $\eta_t$  converges uniformly as  $t \downarrow 0$  to a function  $\gamma \in C(X, Y)$ . Then

$$\lim_{t \downarrow 0} \frac{v(t) - v(0)}{t} = \inf_{x \in X^*} \sup_{y \in Y^*} \gamma(x, y). \quad (7.88)$$

**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

$$v(0) = \sup_{y \in Y} f(x^*, y) \quad \text{and} \quad v(t_k) \leq \sup_{y \in Y} \zeta_k(x^*, y).$$

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  $\zeta_k$  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  $k \rightarrow \infty$ ). By passing to a subsequence if necessary we can assume that  $y_k$  converges to a point  $y^* \in Y$  as  $k \rightarrow \infty$ . It follows that  $y^* \in Y^*$ , and of course we have that

$$\sup_{y \in Y} f(x^*, y) \geq f(x^*, y_k).$$



Also since  $\eta_k$  tends uniformly to  $\gamma$ , it follows that  $\eta_k(x^*, y_k) \rightarrow \gamma(x^*, y^*)$ . Consequently

$$v(t_k) - v(0) \leq \zeta_k(x^*, y_k) - f(x^*, y_k) = t_k \eta_k(x^*, y_k) = t_k \gamma(x^*, y^*) + o(t_k).$$

We obtain that for any  $x^* \in X^*$  there exists  $y^* \in Y^*$  such that

$$\limsup_{k \rightarrow \infty} \frac{v(t_k) - v(0)}{t_k} \leq \gamma(x^*, y^*).$$

It follows that

$$\limsup_{k \rightarrow \infty} \frac{v(t_k) - v(0)}{t_k} \leq \inf_{x \in X^*} \sup_{y \in Y^*} \gamma(x, y). \quad (7.89)$$

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  $\theta_k : X \rightarrow \mathbb{R}$  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 convex-concave, 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  $y \in Y^*$ ,

$$v(t_k) - v(0) \geq \zeta_k(x_k, y) - f(x_k, y) = t_k \gamma_k(x^*, y) + o(t_k)$$

holds, and hence

$$\liminf_{k \rightarrow \infty} \frac{v(t_k) - v(0)}{t_k} \geq \gamma(x^*, y).$$

Since  $y$  was an arbitrary element of  $Y^*$ , we obtain that

$$\liminf_{k \rightarrow \infty} \frac{v(t_k) - v(0)}{t_k} \geq \sup_{y \in Y^*} \gamma(x^*, y),$$

and hence

$$\liminf_{k \rightarrow \infty} \frac{v(t_k) - v(0)}{t_k} \geq \inf_{x \in X^*} \sup_{y \in Y^*} \gamma(x, y). \quad (7.90)$$

The assertion of the theorem follows from (7.89) and (7.90).  $\square$

### 7.1.6 Epiconvergence

Consider a sequence  $f_k : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ ,  $k = 1, \dots$ , of extended real valued functions. It is said that the functions  $f_k$  *epiconverge* to a function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ , written  $f_k \xrightarrow{e} f$ , 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 7.25.** 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

$$\liminf_{k \rightarrow \infty} f_k(x_k) \geq f(x); \quad (7.91)$$

(ii) there exists a sequence  $x_k$  converging to  $x$  such that<sup>59</sup>

$$\limsup_{k \rightarrow \infty} f_k(x_k) \leq f(x). \quad (7.92)$$

Epiconvergence  $f_k \xrightarrow{e} f$  implies that the function  $f$  is lower semicontinuous.

For  $\varepsilon \geq 0$  we say that a point  $\bar{x} \in \mathbb{R}^n$  is an  $\varepsilon$ -minimizer<sup>60</sup> of  $f$  if  $f(\bar{x}) \leq \inf f(x) + \varepsilon$ . (We write here  $\inf f(x)$  for  $\inf_{x \in \mathbb{R}^n} f(x)$ .) Clearly, for  $\varepsilon = 0$  the set of  $\varepsilon$ -minimizers of  $f$  coincides with the set  $\arg \min f$  (of minimizers of  $f$ ).

**Proposition 7.26.** Suppose that  $f_k \xrightarrow{e} f$ . Then

$$\limsup_{k \rightarrow \infty} [\inf f_k(x)] \leq \inf f(x). \quad (7.93)$$

Suppose, further, that (i) for some  $\varepsilon_k \downarrow 0$  there exists an  $\varepsilon_k$ -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

$$\lim_{k \rightarrow \infty} [\inf f_k(x)] = \inf f(x). \quad (7.94)$$

**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 (7.92) holds. Clearly  $f_k(x_k) \geq \inf f_k(x)$  for all  $k$ . Together with (7.92) this implies that

$$f(\bar{x}) \geq \limsup_{k \rightarrow \infty} f_k(x_k) \geq \limsup_{k \rightarrow \infty} [\inf f_k(x)].$$

Since the above holds for any  $\bar{x}$ , the inequality (7.93) follows.

Now let  $x_k$  be a sequence of  $\varepsilon_k$ -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 (7.93) we obtain

$$\liminf_{k \rightarrow \infty} [\inf f_k(x)] = \liminf_{k \rightarrow \infty} [\inf f_k(x) + \varepsilon_k] \geq \liminf_{k \rightarrow \infty} f_k(x_k) \geq f(\bar{x}) \geq \inf f(x).$$

Together with (7.93) this implies (7.94) and  $f(\bar{x}) = \inf f(x)$ . This completes the proof.  $\square$

Assumption (i) in the above proposition can be ensured by various boundedness conditions. Proof of the following theorem can be found in [181, Theorem 7.17].

**Theorem 7.27.** Let  $f_k : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  be a sequence of convex functions and  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  be a convex lower semicontinuous function such that  $\text{dom } f$  has a nonempty interior. Then the following are equivalent: (i)  $f_k \xrightarrow{e} f$ , (ii) there exists a dense subset  $D$  of  $\mathbb{R}^n$  such that

<sup>59</sup>Note that here some (all) points  $x_k$  can be equal to  $x$ .

<sup>60</sup>For the sake of convenience, we allow in this section for a minimizer, or  $\varepsilon$ -minimizer,  $\bar{x}$  to be such that  $f(\bar{x})$  is not finite, i.e., can be equal to  $+\infty$  or  $-\infty$ .

$f_k(x) \rightarrow f(x)$  for all  $x \in D$ , and (iii)  $f_k(\cdot)$  converges uniformly to  $f(\cdot)$  on every compact set  $C$  that does not contain a boundary point of  $\text{dom } f$ .

## 7.2 Probability

### 7.2.1 Probability Spaces and Random Variables

Let  $\Omega$  be an abstract set. It is said that a set  $\mathcal{F}$  of subsets of  $\Omega$  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  $\Omega$  belongs to  $\mathcal{F}$ , and (iii) if<sup>61</sup>  $A_i \in \mathcal{F}$ ,  $i \in \mathbb{N}$ , then  $\cup_{i \in \mathbb{N}} A_i \in \mathcal{F}$ . The set  $\Omega$  equipped with a sigma algebra  $\mathcal{F}$  is called a *sample* or *measurable* space and denoted  $(\Omega, \mathcal{F})$ . 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  $\Omega$ , 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  $\Omega$  consists of two elements  $\Omega$  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  $A_i \subset \Omega$ ,  $i = 1, \dots, n$ , 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  $\Xi$ .

A function  $P : \mathcal{F} \rightarrow \mathbb{R}_+$  is called a (sigma-additive) *measure* on  $(\Omega, \mathcal{F})$  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

$$P\left(\cup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} P(A_i). \quad (7.95)$$

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  $(\Omega, \mathcal{F})$  equipped with a probability measure  $P$  is called a *probability space* and denoted  $(\Omega, \mathcal{F}, P)$ . 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 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  $P$  and  $Q$  be two measures on a measurable space  $(\Omega, \mathcal{F})$ . 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  $Q(A) < \varepsilon$ .

<sup>61</sup>By  $\mathbb{N}$  we denote the set of positive integers.

**Theorem 7.28 (Radon–Nikodym).** *If  $P$  and  $Q$  are measures on  $(\Omega, \mathcal{F})$ , then  $Q$  is absolutely continuous with respect to  $P$  iff there exists a function  $f : \Omega \rightarrow \mathbb{R}_+$  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  $V : \Omega \rightarrow \mathbb{R}^m$  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.<sup>62</sup> A measurable mapping  $V(\omega)$  from probability space  $(\Omega, \mathcal{F}, P)$  into  $\mathbb{R}^m$  is called a *random vector*. Note that the mapping  $V$  generates the probability measure<sup>63</sup> (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  $(\Xi, \mathcal{B})$  equipped with probability measure  $P$  as a probability space  $(\Xi, \mathcal{B}, P)$ . 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  $\xi$  data vector of a considered problem. Sometimes we view  $\xi$  as a random vector  $\xi : \Omega \rightarrow \mathbb{R}^m$  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  $\xi$  as a random vector, we sometimes write  $\xi = \xi(\omega)$ .

A measurable mapping (function)  $Z : \Omega \rightarrow \mathbb{R}$  is called a *random variable*. Its probability distribution is completely defined by the cumulative distribution function (cdf)  $H_Z(z) := \Pr\{Z \leq z\}$ . 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  $\{\omega \in \Omega : Z(\omega) \leq z\}$  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 dependence on  $\omega \in \Omega$ . The coordinate functions  $V_1(\omega), \dots, V_m(\omega)$  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)  $V_1, \dots, V_m$ .

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  $Z : \Omega \rightarrow \overline{\mathbb{R}}$ , where  $\overline{\mathbb{R}}$  denotes the set of extended real numbers. Such functions  $Z : \Omega \rightarrow \overline{\mathbb{R}}$  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  $\lim_{z \rightarrow +\infty} F_Z(z)$  is equal to the probability of the event  $\{\omega \in \Omega : Z(\omega) < +\infty\}$  and can be less than 1 if the event  $\{\omega \in \Omega : Z(\omega) = +\infty\}$  has a positive probability.

<sup>62</sup>In fact it suffices to verify  $\mathcal{F}$ -measurability of  $V^{-1}(A)$  for any family of sets generating the Borel sigma algebra of  $\mathbb{R}^m$ .

<sup>63</sup>With some abuse of notation we also denote here by  $P$  the probability distribution induced by the probability measure  $P$  on  $(\Omega, \mathcal{F})$ .

The *expected value* or *expectation* of an (extended) random variable  $Z : \Omega \rightarrow \overline{\mathbb{R}}$  is defined by the integral

$$\mathbb{E}_P[Z] := \int_{\Omega} Z(\omega) dP(\omega). \quad (7.96)$$

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  $\Upsilon$  is positive, then, by definition,  $\mathbb{E}[Z] = +\infty$ . For a general (not necessarily nonnegative valued) random variable we would like to define<sup>64</sup>  $\mathbb{E}[Z] := \mathbb{E}[Z_+] - \mathbb{E}[(-Z)_+]$ . 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  $z_1, z_2, \dots$ , 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  $(\Omega, \mathcal{F}, P)$ . 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.

**Theorem 7.29 (Monotone Convergence Theorem).** *Suppose that  $f_n \uparrow f$  a.e. and there exists a  $P$ -integrable function  $g(\omega)$  such that  $f_n(\cdot) \geq g(\cdot)$ . Then  $\int_{\Omega} f dP$  is well defined and  $\int_{\Omega} f_n dP \uparrow \int_{\Omega} f dP$ .*

**Theorem 7.30 (Fatou's Lemma).** *Suppose that there exists a  $P$ -integrable function  $g(\omega)$  such that  $f_n(\cdot) \geq g(\cdot)$ . Then*

$$\liminf_{n \rightarrow \infty} \int_{\Omega} f_n dP \geq \int_{\Omega} \liminf_{n \rightarrow \infty} f_n dP. \quad (7.97)$$

**Theorem 7.31 (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  $\int_{\Omega} f dP$  is well defined and  $\int_{\Omega} f_n dP \rightarrow \int_{\Omega} f dP$ .*

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  $A \in \mathcal{F}$ .

<sup>64</sup>Recall that  $Z_+ := \max\{0, Z\}$ .

**Theorem 7.32 (Richter–Rogosinski).** *Let  $(\Omega, \mathcal{F})$  be a measurable space,  $f_1, \dots, f_m$  be measurable on  $(\Omega, \mathcal{F})$  real valued functions, and  $\mu$  be a measure on  $(\Omega, \mathcal{F})$  such that  $f_1, \dots, f_m$  are  $\mu$ -integrable. Suppose that every finite subset of  $\Omega$  is  $\mathcal{F}$ -measurable. Then there exists a measure  $\eta$  on  $(\Omega, \mathcal{F})$  with a finite support of at most  $m$  points such that  $\int_{\Omega} f_i d\mu = \int_{\Omega} f_i d\eta$  for all  $i = 1, \dots, m$ .*

**Proof.** 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  $(\int f_1 d\mu', \dots, \int f_m d\mu')$  with  $\mu'$  being a measure on  $\Omega$  with a finite support. It is not difficult to show that it suffices to take measures  $\mu'$  with support of at most  $m$  points in the definition of the set  $S$  (we leave this as an exercise). We have to show that vector  $a := (\int f_1 d\mu, \dots, \int f_m d\mu)$  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  $c^T a \leq c^T x$ , for all  $x \in S$ . Since  $S$  is a cone, it follows that  $c^T a \leq 0$ . This implies that for  $f := \sum_{i=1}^m c_i f_i$  we have that  $\int f d\mu \leq 0$  and  $\int f d\mu \leq \int f d\mu'$  for any measure  $\mu'$  with a finite support. In particular, by taking measures of the form<sup>65</sup>  $\mu' = \alpha \Delta(\omega)$ , with  $\alpha > 0$  and  $\omega \in \Omega$ , we obtain from the second inequality that  $\int f d\mu \leq \alpha f(\omega)$ . This implies that  $f(\omega) \geq 0$  for all  $\omega \in \Omega$ , since otherwise if  $f(\omega) < 0$  we can make  $\alpha f(\omega)$  arbitrary small by taking  $\alpha$  large enough. 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  $\Omega'$  is a support of  $\mu$ , 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  $\Omega'$  is nonempty. Moreover, the functions  $f_i, i = 1, \dots, m$ , are linearly dependent on  $\Omega'$ . Consequently, by the induction assumption there exists a measure  $\mu'$  with a finite support on  $\Omega'$  such that  $\int f_i d\mu^* = \int f_i d\mu'$  for all  $i = 1, \dots, m$ , where  $\mu^*$  is the restriction of the measure  $\mu$  to the set  $\Omega'$ . Moreover, since  $\mu$  is supported on  $\Omega'$  we have that  $\int f_i d\mu^* = \int f_i d\mu$ , and hence the proof is complete.  $\square$

Let us remark that if the measure  $\mu$  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  $\eta$  on  $(\Omega, \mathcal{F})$  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  $i = 1, \dots, m$ .

Also let us recall two famous inequalities. The *Chebyshev inequality*<sup>66</sup> says that if  $Z : \Omega \rightarrow \mathbb{R}_+$  is a nonnegative valued random variable, then

$$\Pr(Z \geq \alpha) \leq \alpha^{-1} \mathbb{E}[Z], \quad \forall \alpha > 0. \quad (7.98)$$

Proof of (7.98) is rather simple. We have

$$\Pr(Z \geq \alpha) = \mathbb{E}[\mathbf{1}_{[\alpha, +\infty)}(Z)] \leq \mathbb{E}[\alpha^{-1} Z] = \alpha^{-1} \mathbb{E}[Z].$$

The *Jensen inequality* says that if  $V : \Omega \rightarrow \mathbb{R}^m$  is a random vector,  $v := \mathbb{E}[V]$  and  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  is a convex function, then

$$\mathbb{E}[f(V)] \geq f(v), \quad (7.99)$$

<sup>65</sup>We denote by  $\Delta(\omega)$  measure of mass one at the point  $\omega$  and refer to such measures as *Dirac* measures.

<sup>66</sup>Sometimes (7.98) is called the Markov inequality, while the Chebyshev inequality is referred to as the inequality (7.98) applied to the function  $(Z - \mathbb{E}[Z])^2$ .

provided the above expectations are finite. Indeed, for a subgradient  $g \in \partial f(v)$  we have that

$$f(V) \geq f(v) + g^T(V - v). \quad (7.100)$$

By taking expectation of the both sides of (7.100) we obtain (7.99).

Finally, let us mention the following simple inequality. Let  $Y_1, Y_2 : \Omega \rightarrow \mathbb{R}$  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  $\{\omega : Y_1(\omega) + Y_2(\omega) \geq a_1 + a_2\}$  is included in the union of the events  $\{\omega : Y_1(\omega) \geq a_1\}$  and  $\{\omega : Y_2(\omega) \geq a_2\}$ . It follows that

$$\Pr\{Y_1 + Y_2 \geq a_1 + a_2\} \leq \Pr\{Y_1 \geq a_1\} + \Pr\{Y_2 \geq a_2\}. \quad (7.101)$$

## 7.2.2 Conditional Probability and Conditional Expectation

For two events  $A$  and  $B$  the *conditional probability* of  $A$  given  $B$  is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad (7.102)$$

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

$$p_{x|y}(x|y) := P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p(x, y)}{p_y(y)} \quad (7.103)$$

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 (7.103) for all values of  $y$  such that  $f_y(y) > 0$  as

$$f_{x|y}(x|y) := \frac{f(x, y)}{f_y(y)}. \quad (7.104)$$

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

$$\mathbb{E}[X|Y = y] := \int_{-\infty}^{+\infty} x f_{x|y}(x|y) dx. \quad (7.105)$$

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:

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]. \quad (7.106)$$

In the continuous case, for example, we have

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xf(x, y)dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xf_{x|y}(x|y)dx f_y(y)dy,$$

and hence

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} \mathbb{E}[X|Y = y]f_y(y)dy. \quad (7.107)$$

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  $(\Omega, \mathcal{F}, P)$ , 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, and (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, and
- (ii)  $\mathbb{E}[X|\mathcal{G}]$  satisfies the functional equation

$$\int_G \mathbb{E}[X|\mathcal{G}] dP = \int_G X dP, \quad \forall G \in \mathcal{G}. \quad (7.108)$$

The above construction shows 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  $X$ .

Many random variables will satisfy 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

$$\mathbb{E}[X] = \int_{\Omega} \mathbb{E}[X|\mathcal{G}] dP = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]]. \quad (7.109)$$

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, and
- (ii')  $P(A|\mathcal{G})$  satisfies the functional equation

$$\int_G P(A|\mathcal{G}) dP = P(A \cap G), \quad \forall G \in \mathcal{G}. \quad (7.110)$$



### 7.2.3 Measurable Multifunctions and Random Functions

Let  $\mathcal{G}$  be a mapping from  $\Omega$  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

$$\text{dom } \mathcal{G} := \{\omega \in \Omega : \mathcal{G}(\omega) \neq \emptyset\} = \mathcal{G}^{-1}(\mathbb{R}^n)$$

of  $\mathcal{G}$  is an  $\mathcal{F}$ -measurable subset of  $\Omega$ .

**Proposition 7.33.** *A closed valued multifunction  $\mathcal{G} : \Omega \rightrightarrows \mathbb{R}^n$  is measurable iff the (extended real valued) function  $d(\omega) := \text{dist}(x, \mathcal{G}(\omega))$  is measurable for any  $x \in \mathbb{R}^n$ .*

**Proof.** Recall that by the definition  $\text{dist}(x, \mathcal{G}(\omega)) = +\infty$  if  $\mathcal{G}(\omega) = \emptyset$ . Note also that  $\text{dist}(x, \mathcal{G}(\omega)) = \|x - y\|$  for some  $y \in \mathcal{G}(\omega)$ , because of closedness of set  $\mathcal{G}(\omega)$ . Therefore, for any  $t \geq 0$  and  $x \in \mathbb{R}^n$  we have that

$$\{\omega \in \Omega : \text{dist}(x, \mathcal{G}(\omega)) \leq t\} = \mathcal{G}^{-1}(x + tB),$$

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  $A = x + tB$ ,  $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^n$ .  $\square$

**Remark 28.** Suppose now that  $\Omega$  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  $\omega_k \rightarrow \omega$ ,  $x_k \in \mathcal{G}(\omega_k)$  and  $x_k \rightarrow x$ , 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  $\{\omega \in \Omega : \text{dist}(x, \mathcal{G}(\omega)) \leq t\}$  is closed, and hence the function  $d(\omega) := \text{dist}(x, \mathcal{G}(\omega))$  is measurable. Consequently we obtain that any closed multifunction  $\mathcal{G} : \Omega \rightrightarrows \mathbb{R}^n$  is measurable.

It is said that a mapping  $G : \text{dom } \mathcal{G} \rightarrow \mathbb{R}^n$  is a *selection* of  $\mathcal{G}$  if  $G(\omega) \in \mathcal{G}(\omega)$  for all  $\omega \in \text{dom } \mathcal{G}$ . If, in addition, the mapping  $G$  is measurable, it is said that  $G$  is a *measurable selection* of  $\mathcal{G}$ .

**Theorem 7.34 (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  $\Omega$  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 [181, Theorem 14.5] the result of the above theorem is called *Castaing representation*.

Consider a function  $F : \mathbb{R}^n \times \Omega \rightarrow \overline{\mathbb{R}}$ . 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 7.35.** *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 lower semicontinuous, then the multifunction  $\omega \mapsto \text{dom } F(\cdot, \omega)$  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  $\text{epi } F(\cdot, \omega)$  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  $\omega$  does not imply measurability of the corresponding epigraphical multifunction and random lower semicontinuity of  $F(x, \omega)$ . A large class of random lower semicontinuous is given by the so-called Carathéodory functions, i.e., real valued functions  $F : \mathbb{R}^n \times \Omega \rightarrow \mathbb{R}$  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 7.36.** *Suppose that the sigma algebra  $\mathcal{F}$  is  $P$ -complete. Then an extended real valued function  $F : \mathbb{R}^n \times \Omega \rightarrow \mathbb{R}$  is random lower semicontinuous iff the following two properties hold: (i) for every  $\omega \in \Omega$ , the function  $F(\cdot, \omega)$  is lower semicontinuous, and (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*  $X^*(\omega) := \arg \min_{x \in \mathbb{R}^n} F(x, \omega)$ .

**Theorem 7.37.** *Let  $F : \mathbb{R}^n \times \Omega \rightarrow \overline{\mathbb{R}}$  be a random lower semicontinuous function. Then the optimal value function  $\vartheta(\omega)$  and the optimal solution multifunction  $X^*(\omega)$  are both measurable.*

Since we assume that the considered sigma algebras are complete, it follows from condition (ii) of Theorem 7.36 that the optimal value function is measurable. We assume in the remainder of this chapter, sometimes without explicitly saying this, that the function  $F(x, \omega)$  is measurable in the sense of the above condition (ii), and hence considered max- and min-functions are measurable. In case the set  $\Omega$  is a subset of a finite dimensional vector space equipped with its Borel sigma algebra, the optimal value functions are Lebesgue, rather than Borel, measurable (see, e.g., [181, p. 649] for a discussion of a delicate difference between Borel and Lebesgue measurability).

Note that it follows from lower semicontinuity of  $F(\cdot, \omega)$  that the optimal solution multifunction  $X^*(\omega)$  is closed valued. Note also that if  $F(x, \omega)$  is random lower semicontinuous and  $\mathcal{G} : \Omega \rightrightarrows \mathbb{R}^n$  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 lower semicontinuous. 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 7.38.** *Let  $F : \mathbb{R}^{n+m} \times \Omega \rightarrow \bar{\mathbb{R}}$  be a random lower semicontinuous function and*

$$\vartheta(x, \omega) := \inf_{y \in \mathbb{R}^m} F(x, y, \omega) \quad (7.111)$$

*be the associated optimal value function. Suppose that there exists a bounded set  $S \subset \mathbb{R}^m$  such that  $\text{dom} F(x, \cdot, \omega) \subset S$  for all  $(x, \omega) \in \mathbb{R}^n \times \Omega$ . Then the optimal value function  $\vartheta(x, \omega)$  is random lower semicontinuous.*

Let us observe that the above framework of random lower semicontinuous 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  $\text{conv } \mathcal{G}(\omega)$  the convex hull of set  $\mathcal{G}(\omega)$ . If the set  $\Omega = \{\omega_1, \dots, \omega_K\}$  is finite and equipped with respective probabilities  $p_k, k = 1, \dots, K$ , then it is natural to define the integral

$$\int_{\Omega} \mathcal{G}(\omega) dP(\omega) := \sum_{k=1}^K p_k \mathcal{G}(\omega_k), \quad (7.112)$$

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  $(\Omega, \mathcal{F})$ , the corresponding integral is defined as follows.

**Definition 7.39.** *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 nonatomic.<sup>67</sup> The following theorem often is due to Aumann (1965).

**Theorem 7.40 (Aumann).** *Suppose that the measure  $P$  is nonatomic and let  $\mathcal{G} : \Omega \rightrightarrows \mathbb{R}^n$  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  $\|\mathcal{G}(\omega)\| \leq g(\omega)$  for a.e.  $\omega \in \Omega$ . Then*

$$\int_{\Omega} \mathcal{G}(\omega) dP(\omega) = \int_{\Omega} (\text{conv } \mathcal{G}(\omega)) dP(\omega). \quad (7.113)$$

The above theorem is a consequence of a theorem due to Lyapunov (1940).

<sup>67</sup>It is said that measure  $P$ , and the space  $(\Omega, \mathcal{F}, P)$ , is *nonatomic* if any set  $A \in \mathcal{F}$ , such that  $P(A) > 0$ , contains a subset  $B \in \mathcal{F}$  such that  $P(A) > P(B) > 0$ .

**Theorem 7.41 (Lyapunov).** *Let  $\mu_1, \dots, \mu_n$  be a finite collection of nonatomic measures on a measurable space  $(\Omega, \mathcal{F})$ . Then the set  $\{(\mu_1(S), \dots, \mu_n(S)) : S \in \mathcal{F}\}$  is a closed and convex subset of  $\mathbb{R}^n$ .*

## 7.2.4 Expectation Functions

Consider a random function  $F : \mathbb{R}^n \times \Omega \rightarrow \overline{\mathbb{R}}$  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 shown in the next theorem, the lower semicontinuity of the expected value function follows from the lower semicontinuity of  $F(\cdot, \omega)$ .

**Theorem 7.42.** *Suppose that for  $P$ -almost every  $\omega \in \Omega$  the function  $F(\cdot, \omega)$  is lower semicontinuous at a point  $x_0$  and there exists  $P$ -integrable function  $Z(\omega)$  such that  $F(x, \omega) \geq Z(\omega)$  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 lower semicontinuous 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

$$\liminf_{x \rightarrow x_0} \int_{\Omega} F(x, \omega) dP(\omega) \geq \int_{\Omega} \liminf_{x \rightarrow x_0} F(x, \omega) dP(\omega). \quad (7.114)$$

Together with lower semicontinuity of  $F(\cdot, \omega)$  this implies lower semicontinuity of  $f$  at  $x_0$ .  $\square$

With stronger assumptions, we can show that the expectation function is continuous.

**Theorem 7.43.** *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

$$\lim_{x \rightarrow x_0} \int_{\Omega} F(x, \omega) dP(\omega) = \int_{\Omega} \lim_{x \rightarrow x_0} F(x, \omega) dP(\omega) = \int_{\Omega} F(x_0, \omega) dP(\omega). \quad (7.115)$$

This shows the continuity of  $f(x)$  at  $x_0$ .  $\square$

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  $\xi$ . 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:

$$|F(x_1, \omega) - F(x_2, \omega)| \leq C(\omega)\|x_1 - x_2\|. \quad (7.116)$$

- (A3) For almost every  $\omega$  the function  $F_\omega(\cdot)$  is directionally differentiable at  $x_0$ .
- (A4) For almost every  $\omega$  the function  $F_\omega(\cdot)$  is differentiable at  $x_0$ .

**Theorem 7.44.** *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*

$$f'(x_0, h) = \mathbb{E}[F'_\omega(x_0, h)], \quad \forall h. \quad (7.117)$$

- (c) *If conditions (A1), (A2), and (A4) hold, then  $f(x)$  is differentiable at  $x_0$  and*

$$\nabla f(x_0) = \mathbb{E}[\nabla_x F(x_0, \omega)]. \quad (7.118)$$

**Proof.** It follows from (7.116) that for any  $x_1, x_2$  in a neighborhood of  $x_0$ ,

$$|f(x_1) - f(x_2)| \leq \int_{\Omega} |F(x_1, \omega) - F(x_2, \omega)| dP(\omega) \leq c\|x_1 - x_2\|,$$

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

$$R_t(\omega) := t^{-1}[F(x_0 + th, \omega) - F(x_0, \omega)].$$

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 (7.117). This proves assertion (b).

Finally, if  $F'_\omega(x_0, h)$  is linear in  $h$  for almost every  $\omega$ , i.e., the function  $F_\omega(\cdot)$  is differentiable at  $x_0$  w.p. 1, then (7.117) implies that  $f'(x_0, h)$  is linear in  $h$ , and hence (7.118) 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).  $\square$

The above analysis shows that two basic conditions for interchangeability of the expectation and differentiation operators, i.e., for the validity of formula (7.118), are the above conditions (A2) and (A4). The following lemma shows that if, in addition to assumptions (A1)–(A3), the directional derivative  $F'_\omega(x_0, h)$  is *convex* in  $h$  w.p. 1, then  $f(x)$  is differentiable at  $x_0$  iff  $F(\cdot, \omega)$  is differentiable at  $x_0$  w.p. 1.

**Lemma 7.45.** *Let  $\psi : \mathbb{R}^n \times \Omega \rightarrow \mathbb{R}$  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 iff 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  $e_1, \dots, e_n$  be a basis of the space  $\mathbb{R}^n$ . Since  $\phi(\cdot)$  is convex and positively homogeneous, it follows that  $\phi(e_i) + \phi(-e_i) \geq \phi(0) = 0$ ,  $i = 1, \dots, n$ . Furthermore, since  $\phi(\cdot)$  is finite valued, it is the support function of a convex compact set. This convex set is a singleton iff

$$\phi(e_i) + \phi(-e_i) = 0, \quad i = 1, \dots, n. \quad (7.119)$$

Therefore,  $\phi(\cdot)$  is linear iff condition (7.119) holds. Consider the sets

$$A_i := \{\omega \in \Omega : \psi(e_i, \omega) + \psi(-e_i, \omega) > 0\}.$$

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.  $\square$

Regularity conditions which are required for formula (7.117) 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 7.46.** *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 and directionally differentiable at  $x_0$  and formula (7.117) holds. Moreover,  $f(x)$  is differentiable at  $x_0$  iff  $F_\omega(x)$  is differentiable at  $x_0$  w.p. 1, in which case formula (7.118) 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

$$R_t(\omega) := t^{-1}[F(x_0 + th, \omega) - F(x_0, \omega)]$$

is monotonically decreasing to  $F'_\omega(x_0, h)$  as  $t \downarrow 0$ . Also we have that

$$\mathbb{E}|R_{t_0}(\omega)| \leq t_0^{-1}(\mathbb{E}|F(x_0 + t_0 h, \omega)| + \mathbb{E}|F(x_0, \omega)|) < +\infty.$$

Then it follows by the monotone convergence theorem that

$$\lim_{t \downarrow 0} \mathbb{E}[R_t(\omega)] = \mathbb{E}\left[\lim_{t \downarrow 0} R_t(\omega)\right] = \mathbb{E}[F'_\omega(x_0, h)]. \quad (7.120)$$

Since  $\mathbb{E}[R_t(\omega)] = t^{-1}[f(x_0 + th) - f(x_0)]$ , we have that the left-hand side of (7.120) is equal to  $f'(x_0, h)$ , and hence formula (7.117) follows.

The last assertion follows then from Lemma 7.45.  $\square$

**Remark 29.** 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)  $\mathbb{E}[F(x_0 + t_0 h, \omega)] < +\infty$  for some  $t_0 > 0$ . Then  $f'(x_0, h) < +\infty$  and formula (7.117) holds. Note also that if 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

$$df(x_0)/dx = \mathbb{E}[\partial F(x_0, \omega)/\partial x] = \Pr(\xi < x_0) - \Pr(\xi > x_0). \quad (7.121)$$

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,

$$f'(x_0, -1) + f'(x_0, 1) = 2\Pr(\xi = x_0) > 0. \quad (7.122)$$

We can also investigate differentiability properties of the expectation function by studying the subdifferentiability of the integrand. Suppose for the moment that the set  $\Omega$  is finite, say,  $\Omega := \{\omega_1, \dots, \omega_K\}$  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  $\text{dom } f = \bigcap_{k=1}^K \text{dom } F_k$ , 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  $\Omega = \{\omega_1, \dots, \omega_K\}$  is finite, (ii) for every  $\omega_k \in \Omega$  the function  $F_k(\cdot) := F(\cdot, \omega_k)$  is proper and convex, and (iii) the sets  $\text{ri}(\text{dom } F_k)$ ,  $k = 1, \dots, K$ , have a common point. Then for any  $x_0 \in \text{dom } f$ ,

$$\partial f(x_0) = \sum_{k=1}^K p_k \partial F(x_0, \omega_k). \quad (7.123)$$

Note that the above regularity assumption (iii) holds, in particular, if the interior of  $\text{dom } f$  is nonempty.

The subdifferentials at the right-hand side of (7.123) are taken with respect to  $x$ . Note that  $\partial F(x_0, \omega_k)$ , and hence  $\partial f(x_0)$ , in (7.123) can be unbounded or empty. Suppose that all probabilities  $p_k$  are positive. It follows then from (7.123) that  $\partial f(x_0)$  is a singleton iff all subdifferentials  $\partial F(x_0, \omega_k)$ ,  $k = 1, \dots, K$ , 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 30.** In the case of a finite set  $\Omega$  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

$$F^*(z, \omega) := \sup_{x \in \mathbb{R}^n} \{z^\top x - F(x, \omega)\}$$

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))

$$\partial F(x, \omega) = \arg \max_{z \in \mathbb{R}^n} \{z^\top x - F^*(z, \omega)\}.$$

Then it follows by Theorem 7.37 that the multifunction  $\omega \mapsto \partial F(x, \omega)$  is measurable.

In general we have the following extension of formula (7.123).

**Theorem 7.47.** 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, and (iv) the domain of  $f$  has a nonempty interior. Then for any  $x_0 \in \text{dom } f$ ,

$$\partial f(x_0) = \int_{\Omega} \partial F(x_0, \omega) dP(\omega) + \mathcal{N}_{\text{dom } f}(x_0). \quad (7.124)$$

**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:

$$F(x, \omega) - F(x_0, \omega) \geq G(\omega)^T (x - x_0) \quad \forall x \in \mathbb{R}^n.$$

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

$$\int_{\Omega} \partial F(x_0, \omega) dP(\omega) \subset \partial f(x_0). \quad (7.125)$$

In particular, it follows from (7.125) that if  $\partial f(x_0)$  is empty, then the set at the right-hand side of (7.124) 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 (7.125) that

$$\int_{\Omega} \partial F(x_0, \omega) dP(\omega) + \mathcal{N}_{\text{dom } f}(x_0) \subset \partial f(x_0). \quad (7.126)$$

Note that inclusion (7.126) holds irrespective of assumption (iv).

Proving the converse of inclusion (7.126) 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 (7.125) 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 (7.126), respectively. By virtue of inclusion (7.125),  $\mathcal{N}_{\text{dom } f}(x_0)$  forms the recession cone of the set at the left-hand side of (7.126) 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 (7.124) does not hold, i.e., inclusion (7.126) 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  $f'(x_0, \bar{h})$  is finite for all  $h$  in a neighborhood of  $\bar{h}$ ,  $f'(x_0, \bar{h}) = s_2(\bar{h})$  and (see Remark 29 on page 371)

$$f'(x_0, \bar{h}) = \int_{\Omega} F'_{\omega}(x_0, \bar{h}) dP(\omega).$$

Also,  $F'_{\omega}(x_0, h)$  is finite for a.e.  $\omega$  and for all  $h$  in a neighborhood of  $\bar{h}$ , and hence  $F'_{\omega}(x_0, \bar{h}) = \bar{h}^T G(\omega)$  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,

$$\int_{\Omega} F'_{\omega}(x_0, \bar{h}) dP(\omega) = \bar{h}^T \int_{\Omega} G(\omega) dP(\omega).$$

Since  $\int_{\Omega} G(\omega) dP(\omega)$  is a point of the set at the left-hand side of (7.125), we obtain that  $s_1(\bar{h}) \geq f'(x_0, \bar{h}) = s_2(\bar{h})$ , a contradiction.  $\square$

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

$$\partial f(x_0) = \int_{\Omega} \partial F(x_0, \omega) dP(\omega). \quad (7.127)$$

Also, it follows from formula (7.124) 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  $(\Omega, \mathcal{F}, P)$ . As it was discussed in section 7.2.1, we can view  $\xi^i$  as random vectors supported on a (closed) set  $\Xi \subset \mathbb{R}^d$  equipped with its Borel sigma algebra  $\mathcal{B}$ . We say that  $\xi^i$ ,  $i \in \mathbb{N}$ , are *identically distributed* if each  $\xi^i$  has the same probability distribution on  $(\Xi, \mathcal{B})$ . If, moreover,  $\xi^i$ ,  $i \in \mathbb{N}$ , are independent, we say that they are *independent identically distributed* (iid). Consider a measurable function  $F : \Xi \rightarrow \mathbb{R}$  and the sequence  $F(\xi^i)$ ,  $i \in \mathbb{N}$ , of random variables. If  $\xi^i$  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  $\xi^i$  are identically distributed and the expectation  $\mathbb{E}[F(\xi^1)]$  is well defined, then, under some regularity conditions,<sup>68</sup>

$$N^{-1} \sum_{i=1}^N F(\xi^i) \rightarrow \mathbb{E}[F(\xi^1)] \text{ w.p. 1 as } N \rightarrow \infty. \quad (7.128)$$

In particular, the classical LLN states that the convergence (7.128) holds if the sequence  $\xi^i$  is iid.

Consider now a random function  $F : X \times \Xi \rightarrow \mathbb{R}$ , where  $X$  is a nonempty subset of  $\mathbb{R}^n$  and  $\xi = \xi(\omega)$  is a random vector supported on the set  $\Xi$ . 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 distribution as the random vector  $\xi$ , and let

$$\hat{f}_N(x) := N^{-1} \sum_{i=1}^N F(x, \xi^i) \quad (7.129)$$

be the so-called sample average functions. Note that the sample average function  $\hat{f}_N(x)$  depends on the random sequence  $\xi^1, \dots, \xi^N$  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.,

$$\hat{f}_N(x) \rightarrow f(x) \text{ w.p. 1 as } N \rightarrow \infty. \quad (7.130)$$

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  $N \geq N^*$ . It should be emphasized that  $N^*$  depends on  $\varepsilon$  and  $\omega$ , and also on  $x \in X$ .

<sup>68</sup>Sometimes (7.128) is referred to as the *strong* LLN to distinguish it from the *weak* LLN where the convergence is ensured in probability instead of w.p. 1. Unless stated otherwise, we deal with the strong LLN.

We may refer to (7.130) 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

$$\sup_{x \in X} |\hat{f}_N(x) - f(x)| \rightarrow 0 \text{ w.p. 1 as } N \rightarrow \infty \quad (7.131)$$

and refer to this as the *uniform* LLN. Note that maximum of a countable number of measurable functions is measurable. Since the maximum (supremum) in (7.131) can be taken over a countable and dense subset of  $X$ , this supremum is a measurable function on  $(\Omega, \mathcal{F})$ .

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 7.48.** *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, and (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 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

$$\lim_{k \rightarrow \infty} \mathbb{E}[F(x_k, \xi)] = \mathbb{E}\left[\lim_{k \rightarrow \infty} F(x_k, \xi)\right].$$

Since, by (i),  $F(x_k, \xi) \rightarrow F(x, \xi)$  w.p. 1, it follows that  $f(x_k) \rightarrow f(x)$ , and hence  $f(x)$  is continuous.

Choose now a point  $\bar{x} \in X$  and a sequence  $\gamma_k$  of positive numbers converging to zero, and define  $V_k := \{x \in X : \|x - \bar{x}\| \leq \gamma_k\}$  and

$$\delta_k(\xi) := \sup_{x \in V_k} |F(x, \xi) - F(\bar{x}, \xi)|. \quad (7.132)$$

Because of the standing assumption of measurability of  $F(x, \xi)$ , we have that  $\delta_k(\xi)$  is Lebesgue measurable (see the discussion after Theorem 7.37). By assumption (i) we have that for a.e.  $\xi \in \Xi$ ,  $\delta_k(\xi)$  tends to zero as  $k \rightarrow \infty$ . Moreover, by 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

$$\lim_{k \rightarrow \infty} \mathbb{E}[\delta_k(\xi)] = \mathbb{E}\left[\lim_{k \rightarrow \infty} \delta_k(\xi)\right] = 0. \quad (7.133)$$

We also have that

$$|\hat{f}_N(x) - \hat{f}_N(\bar{x})| \leq \frac{1}{N} \sum_{i=1}^N |F(x, \xi^i) - F(\bar{x}, \xi^i)|,$$

and hence

$$\sup_{x \in V_k} |\hat{f}_N(x) - \hat{f}_N(\bar{x})| \leq \frac{1}{N} \sum_{i=1}^N \delta_k(\xi^i). \quad (7.134)$$

Since the sequence  $\xi^i$  is iid, it follows by the LLN that the right-hand side of (7.134) converges w.p. 1 to  $\mathbb{E}[\delta_k(\xi)]$  as  $N \rightarrow \infty$ . Together with (7.133) this implies that for any given  $\varepsilon > 0$  there exists a neighborhood  $W$  of  $\bar{x}$  such that w.p. 1 for sufficiently large  $N$ ,

$$\sup_{x \in W \cap X} |\hat{f}_N(x) - \hat{f}_N(\bar{x})| < \varepsilon.$$

Since  $X$  is compact, there exists a finite number of points  $x_1, \dots, x_m \in X$  and corresponding neighborhoods  $W_1, \dots, W_m$  covering  $X$  such that w.p. 1 for  $N$  large enough, the following holds:

$$\sup_{x \in W_j \cap X} |\hat{f}_N(x) - \hat{f}_N(x_j)| < \varepsilon, \quad j = 1, \dots, m. \quad (7.135)$$

Furthermore, since  $f(x)$  is continuous on  $X$ , these neighborhoods can be chosen in such a way that

$$\sup_{x \in W_j \cap X} |f(x) - f(x_j)| < \varepsilon, \quad j = 1, \dots, m. \quad (7.136)$$

Again by the LLN we have that  $\hat{f}_N(x)$  converges pointwise to  $f(x)$  w.p. 1. Therefore,

$$|\hat{f}_N(x_j) - f(x_j)| < \varepsilon, \quad j = 1, \dots, m, \quad (7.137)$$

w.p. 1 for  $N$  large enough. It follows from (7.135)–(7.137) that w.p. 1 for  $N$  large enough

$$\sup_{x \in X} |\hat{f}_N(x) - f(x)| < 3\varepsilon. \quad (7.138)$$

Since  $\varepsilon > 0$  was arbitrary, we obtain that (7.131) follows and the proof is complete.  $\square$

**Remark 31.** It could be noted that 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  $x \geq \xi$  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  $\Pr(\xi \leq x)$ , i.e.,  $f(x) = \mathbb{E}[F(x, \xi)]$  is the cumulative distribution function (cdf) of  $\xi$ . Assumption (i) means here that for any given  $x$ , probability of the event “ $x = \xi$ ” is zero, i.e., that the cdf of  $\xi$  is continuous at  $x$ . In this example, the sample average function  $\hat{f}_N(\cdot)$  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  $(\Omega, \mathcal{F}, P)$ . Recall definition 7.25

of epiconvergence of extended real valued functions. We say that functions  $\hat{f}_N$  epiconverge to  $f$  w.p. 1, written  $\hat{f}_N \xrightarrow{e} f$  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  $F(x, \xi) : \mathbb{R}^n \times \Xi \rightarrow \bar{\mathbb{R}}$  is an extended real valued function, i.e., can take values  $\pm\infty$ .

**Theorem 7.49.** *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,  $\text{dom } f$ , has a nonempty interior; and the pointwise LLN holds. Then  $\hat{f}_N \xrightarrow{e} f$  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  $N \rightarrow \infty$ . 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  $N \rightarrow \infty$ . Consider the set  $\Upsilon := \cup_{x \in D} \Upsilon_x$ . 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  $N \rightarrow \infty$ , pointwise on  $D$ . It follows then by Theorem 7.27 that  $\hat{f}_N(\cdot, \omega) \xrightarrow{e} f(\cdot)$  for any  $\omega \in \Omega \setminus \Upsilon$ . That is,  $\hat{f}_N(\cdot) \xrightarrow{e} f(\cdot)$  w.p. 1.  $\square$

We also have the following result. It can be proved in a way similar to the proof of the above theorem by using Theorem 7.27.

**Theorem 7.50.** *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. The following theorem is due to Artstein and Wets [7, Theorem 2.3]. Recall that we always assume measurability of  $F(x, \xi)$  (see the discussion after Theorem 7.37).

**Theorem 7.51.** *Suppose that: (a) the function  $F(x, \xi)$  is random lower semicontinuous, (b) for every  $\bar{x} \in \mathbb{R}^n$  there exists a neighborhood  $V$  of  $\bar{x}$  and  $P$ -integrable function  $h : \Xi \rightarrow \bar{\mathbb{R}}$  such that  $F(x, \xi) \geq h(\xi)$  for all  $x \in V$  and a.e.  $\xi \in \Xi$ , and (c) the sample is iid. Then  $\hat{f}_N \xrightarrow{e} f$  w.p. 1.*

### Uniform LLN for Derivatives

Let us discuss now uniform LLN for derivatives of the sample average function. By Theorem 7.44 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 (7.118) 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 7.48 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 29 on page 371.

**Theorem 7.52.** *Let  $U \subset \mathbb{R}^n$  be an open set,  $X$  a nonempty compact subset of  $U$ , and  $F : U \times \Xi \rightarrow \mathbb{R}$  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*

$$|F(x', \xi) - F(x, \xi)| \leq C(\xi) \|x' - x\| \quad \text{a.e. } \xi \in \Xi, \quad \forall x, x' \in U, \quad (7.139)$$

and (iii) for every  $x \in X$  the function  $F(\cdot, \xi)$  is continuously differentiable at  $x$  w.p. 1. Then the following hold: (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.,

$$\nabla f(x) = \mathbb{E}[\nabla_x F(x, \xi)], \quad (7.140)$$

and (c) Clarke generalized gradient  $\partial^\circ \hat{f}_N(x)$  converges to  $\nabla f(x)$  w.p. 1 uniformly on  $X$ , i.e.,

$$\lim_{N \rightarrow \infty} \sup_{x \in X} \mathbb{D}(\partial^\circ \hat{f}_N(x), \{\nabla f(x)\}) = 0 \quad \text{w.p. 1.} \quad (7.141)$$

**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 7.44. Consequently, it follows by Theorem 7.44 that  $f(\cdot)$  is differentiable at every point  $x \in X$  and the interchangeability formula (7.140) holds. Moreover, it follows from (7.139) that  $\|\nabla_x F(x, \xi)\| \leq C(\xi)$  for a.e.  $\xi$  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

$$\lim_{k \rightarrow \infty} \nabla f(x_k) = \mathbb{E} \left[ \lim_{k \rightarrow \infty} \nabla_x F(x_k, \xi) \right] = \mathbb{E}[\nabla_x F(x, \xi)] = \nabla f(x).$$

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 7.48. 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  $\delta_k(\xi) \rightarrow 0$  for a.e.  $\xi$ . Also,

$$\delta_k(\xi) \leq \|\nabla_x F(\bar{x}, \xi)\| + \sup_{x \in V_k^*(\xi)} \|\nabla_x F(x, \xi)\| \leq 2C(\xi),$$

and hence  $\delta_k(\xi)$ ,  $k \in \mathbb{N}$ , are dominated by the integrable function  $2C(\xi)$ . Consequently,

$$\lim_{k \rightarrow \infty} \mathbb{E}[\delta_k(\xi)] = \mathbb{E} \left[ \lim_{k \rightarrow \infty} \delta_k(\xi) \right] = 0,$$

and the remainder of the proof can be completed in the same way as the proof of Theorem 7.48 using compactness arguments.  $\square$

### 7.2.6 Law of Large Numbers for Random Sets and Subdifferentials

Consider a measurable multifunction  $\mathcal{A} : \Omega \rightrightarrows \mathbb{R}^n$ . 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  $\Omega$  into  $\mathfrak{C}_n$ . We have that the multifunction  $\mathcal{A} : \Omega \rightrightarrows \mathbb{R}^n$  is measurable iff the corresponding mapping  $\mathcal{A} : \Omega \rightarrow \mathfrak{C}_n$  is measurable.

We say  $A_i : \Omega \rightarrow \mathfrak{C}_n$ ,  $i \in \mathbb{N}$ , is an iid sequence of realizations of  $\mathcal{A}$  if each  $A_i = \mathcal{A}(\omega)$  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 7.53 (Artstein–Vitale).** *Let  $A_i$ ,  $i \in \mathbb{N}$ , be an iid sequence of realizations of a measurable mapping  $\mathcal{A} : \Omega \rightarrow \mathfrak{C}_n$  such that  $\mathbb{E}[\|\mathcal{A}(\omega)\|] < \infty$ . Then*

$$N^{-1}(A_1 + \cdots + A_N) \rightarrow \mathbb{E}[\text{conv}(\mathcal{A})] \text{ w.p. 1 as } N \rightarrow \infty, \quad (7.142)$$

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  $A \mapsto s_A$ , where  $s_A(h) := \sup_{z \in A} z^\top h$  is the support function of  $A$ . Note that for any two convex sets  $A, B \in \mathfrak{C}_n$  we have that  $s_{A+B}(\cdot) = s_A(\cdot) + s_B(\cdot)$ , and  $A \subset B$  iff  $s_A(\cdot) \leq s_B(\cdot)$ . Consequently, for convex sets  $A_1, A_2 \in \mathfrak{C}_n$  and  $B_r := \{x : \|x\| \leq r\}$ ,  $r \geq 0$ , we have

$$\mathbb{D}(A_1, A_2) = \inf \{r \geq 0 : A_1 \subset A_2 + B_r\} \quad (7.143)$$

and

$$\inf \{r \geq 0 : A_1 \subset A_2 + B_r\} = \inf \{r \geq 0 : s_{A_1}(\cdot) \leq s_{A_2}(\cdot) + s_{B_r}(\cdot)\}. \quad (7.144)$$

Moreover,  $s_{B_r}(h) = \sup_{\|z\| \leq r} z^\top h = r \|h\|^*$ , where  $\|\cdot\|^*$  is the dual of the norm  $\|\cdot\|$ . We obtain that

$$\mathbb{H}(A_1, A_2) = \sup_{\|h\|^* \leq 1} |s_{A_1}(h) - s_{A_2}(h)|. \quad (7.145)$$

It follows that if the multifunction  $\mathcal{A}(\omega)$  is compact and *convex* valued, then the convergence assertion (7.142) is equivalent to

$$\sup_{\|h\|^* \leq 1} \left| N^{-1} \sum_{i=1}^N s_{A_i}(h) - \mathbb{E}[s_{\mathcal{A}}(h)] \right| \rightarrow 0 \text{ w.p. 1 as } N \rightarrow \infty. \quad (7.146)$$

Therefore, for compact and convex valued multifunction  $\mathcal{A}(\omega)$ , Theorem 7.53 is a direct consequence of Theorem 7.50. For general compact valued multifunctions, the averaging operation (in the left-hand side of (7.142)) makes a “convexification” of the limiting set.

Consider now a random lower semicontinuous convex function  $F : \mathbb{R}^n \times \Xi \rightarrow \overline{\mathbb{R}}$  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 (7.129)). Recall that for any  $x \in \mathbb{R}^n$ , the multifunction  $\xi \mapsto \partial F(x, \xi)$  is measurable (see Remark 30 on page 372). In a sense the following result can be viewed as a particular case of Theorem 7.53 for compact convex valued multifunctions.

**Theorem 7.54.** *Let  $F : \mathbb{R}^n \times \Xi \rightarrow \overline{\mathbb{R}}$  be a random lower semicontinuous convex function and  $\hat{f}_N(x)$  be the corresponding sample average functions based on an iid sequence  $\xi^i$ . Suppose that the expectation function  $f(x)$  is well defined and finite valued in a neighborhood of a point  $\bar{x} \in \mathbb{R}^n$ . Then*

$$\mathbb{H}(\partial \hat{f}_N(\bar{x}), \partial f(\bar{x})) \rightarrow 0 \text{ w.p. 1 as } N \rightarrow \infty. \quad (7.147)$$

**Proof.** By Theorem 7.46 we have that  $f(x)$  is directionally differentiable at  $\bar{x}$  and

$$f'(\bar{x}, h) = \mathbb{E}[F'_\xi(\bar{x}, h)]. \quad (7.148)$$

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

$$\hat{f}'_N(\bar{x}, h) = N^{-1} \sum_{i=1}^N F'_{\xi^i}(\bar{x}, h). \quad (7.149)$$

Therefore, by the LLN it follows that  $\hat{f}'_N(\bar{x}, \cdot)$  converges to  $f'(\bar{x}, \cdot)$  pointwise w.p. 1 as  $N \rightarrow \infty$ . Consequently, by Theorem 7.50 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 (7.145) 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 7.47 we have  $\mathbb{E}[\partial F(\bar{x}, \xi)] = \partial f(\bar{x})$ .  $\square$

The problem in trying to extend the pointwise convergence (7.147) 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.<sup>69</sup>

Let us consider now the  $\varepsilon$ -subdifferential,  $\varepsilon \geq 0$ , of a convex real valued function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , defined as

$$\partial_\varepsilon f(\bar{x}) := \{z \in \mathbb{R}^n : f(x) - f(\bar{x}) \geq z^\top(x - \bar{x}) - \varepsilon, \quad \forall x \in \mathbb{R}^n\}. \quad (7.150)$$

Clearly for  $\varepsilon = 0$ , the  $\varepsilon$ -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  $\mathbb{R}^n$ .

<sup>69</sup>This multifunction is upper semicontinuous in the sense that if the function  $f(\cdot)$  is convex and continuous at  $\bar{x}$ , then  $\lim_{x \rightarrow \bar{x}} \mathbb{D}(\partial f(x), \partial f(\bar{x})) = 0$ .



**Theorem 7.55.** Let  $g_k : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $k \in \mathbb{N}$ , 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  $\varepsilon$ -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.,

$$\lim_{k \rightarrow \infty} \sup_{x \in X} \mathbb{H}(\partial_\varepsilon g_k(x), \partial_\varepsilon g(x)) = 0. \quad (7.151)$$

**Proof.** Convexity of  $g(\cdot)$  means that

$$g(tx_1 + (1-t)x_2) \leq tg(x_1) + (1-t)g(x_2), \quad \forall x_1, x_2 \in \mathbb{R}^n, \quad \forall t \in [0, 1].$$

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 (7.151) 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  $k \rightarrow \infty$ . Consider the  $\varepsilon$ -directional derivative of  $g$  at  $x$ :

$$g'_\varepsilon(x, h) := \inf_{t > 0} \frac{g(x + th) - g(x) + \varepsilon}{t}. \quad (7.152)$$

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  $h \in \mathbb{R}^n$ ,

$$\lim_{k \rightarrow \infty} g'_{k\varepsilon}(x_k, h) = g'_\varepsilon(\bar{x}, h).$$

Let us fix  $t > 0$ . Then

$$\limsup_{k \rightarrow \infty} g'_{k\varepsilon}(x_k, h) \leq \limsup_{k \rightarrow \infty} \frac{g_k(x_k + th) - g_k(x_k) + \varepsilon}{t} = \frac{g(\bar{x} + th) - g(\bar{x}) + \varepsilon}{t}.$$

Since  $t > 0$  was arbitrary, this implies that

$$\limsup_{k \rightarrow \infty} g'_{k\varepsilon}(x_k, h) \leq g'_\varepsilon(\bar{x}, h).$$

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  $\text{dist}(t_k, T_\varepsilon) \rightarrow 0$ . Note that  $\inf T_\varepsilon > 0$ . Consequently,

$$\liminf_{k \rightarrow \infty} g'_{k\varepsilon}(x_k, h) = \liminf_{k \rightarrow \infty} \frac{g_k(x_k + t_k h) - g_k(x_k) + \varepsilon}{t_k} \geq g'_\varepsilon(\bar{x}, h).$$

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  $\alpha \downarrow 0$ .  $\square$

The above result is deterministic. It can be easily translated into the stochastic framework as follows.

**Theorem 7.56.** *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  $\varepsilon$ -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.,*

$$\sup_{x \in X} \mathbb{H}(\partial_\varepsilon \hat{f}_N(x), \partial_\varepsilon f(x)) \rightarrow 0 \text{ w.p. 1 as } N \rightarrow \infty. \quad (7.153)$$

**Proof.** In a way similar to the proof of Theorem 7.50 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 7.27), and hence the proof can be completed by applying Theorem 7.55.  $\square$

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.  $\xi$ , 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  $N > 1$ .

### 7.2.7 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 32.** 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  $k \rightarrow \infty$ , even if all these expected values are finite. This implication holds under the additional condition that  $Z_k$  are *uniformly integrable*, that is,

$$\lim_{c \rightarrow \infty} \sup_{k \in \mathbb{N}} \mathbb{E}[Z_k(c)] = 0, \quad (7.154)$$

where  $Z_k(c) := |Z_k|$  if  $|Z_k| \geq c$ , 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  $\mathbb{E}[Z_k] \rightarrow \mathbb{E}[Z]$ , is the following: there exists  $\varepsilon > 0$  such that  $\sup_{k \in \mathbb{N}} \mathbb{E}[|Z_k|^{1+\varepsilon}] < \infty$ . Indeed, for  $c > 0$  we have

$$\mathbb{E}[Z_k(c)] \leq c^{-\varepsilon} \mathbb{E}[Z_k(c)^{1+\varepsilon}] \leq c^{-\varepsilon} \mathbb{E}[|Z_k|^{1+\varepsilon}],$$

from which the assertion follows.

**Remark 33 (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  $\Pr(|Z_k/X_k| > c) \leq \varepsilon$  for all  $k \in \mathbb{N}$ . It is written that  $Z_k = o_p(X_k)$  if for any  $\varepsilon > 0$  it holds that  $\lim_{k \rightarrow \infty} \Pr(|Z_k/X_k| > \varepsilon) = 0$ . 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 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  $\tau_N$  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  $G : \mathbb{R}^d \rightarrow \mathbb{R}^m$  be a vector valued function, differentiable at  $\mu$ . That is,

$$G(y) - G(\mu) = J(y - \mu) + r(y), \quad (7.155)$$

where  $J := \nabla G(\mu)$  is the  $m \times d$  Jacobian matrix of  $G$  at  $\mu$ , and the remainder  $r(y)$  is of order  $o(\|y - \mu\|)$ , i.e.,  $r(y)/\|y - \mu\| \rightarrow 0$  as  $y \rightarrow \mu$ . It follows from (7.155) that

$$\tau_N [G(Y_N) - G(\mu)] = J [\tau_N(Y_N - \mu)] + \tau_N r(Y_N). \quad (7.156)$$

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 (7.156) that

$$\tau_N [G(Y_N) - G(\mu)] \xrightarrow{\mathcal{D}} JY. \quad (7.157)$$

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  $\Sigma$ , 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 (7.157) that

$$N^{1/2} [G(Y_N) - G(\mu)] \xrightarrow{\mathcal{D}} \mathcal{N}(0, J \Sigma J^T). \quad (7.158)$$

We need to extend this method in several directions. The random functions  $\hat{f}_N(\cdot)$  can be viewed as random elements in an appropriate functional space. This motivates us to extend formula (7.157) to a Banach space setting. Let  $B_1$  and  $B_2$  be two Banach spaces, and let  $G : B_1 \rightarrow B_2$  be a mapping. Suppose that  $G$  is directionally differentiable at a considered point  $\mu \in B_1$ , i.e., the limit

$$G'_\mu(d) := \lim_{t \downarrow 0} \frac{G(\mu + td) - G(\mu)}{t} \quad (7.159)$$

exists for all  $d \in B_1$ . If, in addition, the directional derivative  $G'_\mu : B_1 \rightarrow B_2$  is linear and continuous, then it is said that  $G$  is Gâteaux differentiable at  $\mu$ . 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  $\alpha \geq 0$  and  $d \in B_1$ .

It follows from (7.159) 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.,  $r(td)/t \rightarrow 0$  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  $\mu$  in the sense of Hadamard if the directional derivative  $G'_\mu(d)$  exists for all  $d \in B_1$  and, moreover,

$$G'_\mu(d) = \lim_{\substack{t \downarrow 0 \\ d' \rightarrow d}} \frac{G(\mu + td') - G(\mu)}{t}. \quad (7.160)$$

**Proposition 7.57.** *Let  $B_1$  and  $B_2$  be Banach spaces,  $G : B_1 \rightarrow B_2$ , and  $\mu \in B_1$ . Then the following hold: (i) If  $G(\cdot)$  is Hadamard directionally differentiable at  $\mu$ , then the directional derivative  $G'_\mu(\cdot)$  is continuous. (ii) If  $G(\cdot)$  is Lipschitz continuous in a neighborhood of  $\mu$  and directionally differentiable at  $\mu$ , then  $G(\cdot)$  is Hadamard directionally differentiable at  $\mu$ .*

The above properties can be proved in a way similar to the proof of Theorem 7.2. We also have the following *chain rule*.

**Proposition 7.58 (Chain Rule).** *Let  $B_1$ ,  $B_2$ , and  $B_3$  be Banach spaces and  $G : B_1 \rightarrow B_2$  and  $F : B_2 \rightarrow B_3$  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  $F \circ G : B_1 \rightarrow B_3$  is directionally differentiable at  $\mu$  and*

$$(F \circ G)'(\mu, d) = F'(\eta, G'(\mu, d)), \quad \forall d \in B_1. \quad (7.161)$$

**Proof.** Since  $G$  is directionally differentiable at  $\mu$ , we have for  $t \geq 0$  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  $\mu$  and formula (7.161) holds.  $\square$

Now let  $B_1$  and  $B_2$  be equipped with their Borel  $\sigma$ -algebras  $\mathcal{B}_1$  and  $\mathcal{B}_2$ , respectively. An  $\mathcal{F}$ -measurable mapping from a probability space  $(\Omega, \mathcal{F}, P)$  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  $N \rightarrow \infty$ , for any bounded and continuous function  $f : B_1 \rightarrow \mathbb{R}$ . 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 7.59 (Delta Theorem).** *Let  $B_1$  and  $B_2$  be Banach spaces, equipped with their Borel  $\sigma$ -algebras,  $Y_N$  be a sequence of random elements of  $B_1$ ,  $G : B_1 \rightarrow B_2$  be a mapping, and  $\tau_N$  be a sequence of positive numbers tending to infinity as  $N \rightarrow \infty$ . 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

$$\tau_N [G(Y_N) - G(\mu)] \xrightarrow{\mathcal{D}} G'_\mu(Y) \quad (7.162)$$

and

$$\tau_N [G(Y_N) - G(\mu)] = G'_\mu(X_N) + o_p(1). \quad (7.163)$$

Note that because of the Hadamard directional differentiability of  $G$ , the mapping  $G'_\mu : B_1 \rightarrow B_2$  is continuous, and hence is measurable with respect to the Borel  $\sigma$ -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 7.60 (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  $X'_N \xrightarrow{\mathcal{D}} X_N$  for all  $N$ ,  $Y' \xrightarrow{\mathcal{D}} Y$ , and  $X'_N \rightarrow Y'$  w.p. 1.*

Here  $Y' \xrightarrow{\mathcal{D}} Y$  means that the probability measures induced by  $Y'$  and  $Y$  coincide.

**Proof of Theorem 7.59.** 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  $X'_N \xrightarrow{\mathcal{D}} X_N$ ,  $Y' \xrightarrow{\mathcal{D}} Y$ , and  $X'_N \rightarrow Y'$  w.p. 1. Consequently for  $Y'_N := \mu + \tau_N^{-1} X'_N$ , we have  $Y'_N \xrightarrow{\mathcal{D}} Y_N$ . It follows then from Hadamard directional differentiability of  $G$  that

$$\tau_N [G(Y'_N) - G(\mu)] \rightarrow G'_\mu(Y') \quad \text{w.p. 1.} \quad (7.164)$$

Since convergence w.p. 1 implies convergence in distribution and the terms in (7.164) have the same distributions as the corresponding terms in (7.162), the asymptotic result (7.162) follows.

Now since  $G'_\mu(\cdot)$  is continuous and  $X'_N \rightarrow Y'$  w.p. 1, we have that

$$G'_\mu(X'_N) \rightarrow G'_\mu(Y') \quad \text{w.p. 1.} \quad (7.165)$$

Together with (7.164) this implies that the difference between  $G'_\mu(X'_N)$  and the left-hand side of (7.164) tends w.p. 1, and hence in probability, to zero. We obtain that

$$\tau_N [G(Y'_N) - G(\mu)] = G'_\mu [\tau_N(Y'_N - \mu)] + o_p(1),$$

which implies (7.163).  $\square$

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  $\mu$  *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  $d_N \rightarrow d$ , the following limit exists:

$$G'_\mu(d) = \lim_{N \rightarrow \infty} \frac{G(\mu + t_N d_N) - G(\mu)}{t_N}. \quad (7.166)$$

Equivalently, condition (7.166) can be written in the form

$$G'_\mu(d) = \lim_{\substack{t \downarrow 0 \\ d' \rightarrow_K d}} \frac{G(\mu + td') - G(\mu)}{t}, \quad (7.167)$$

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  $\mu$ , denoted  $\mathcal{T}_K(\mu)$ , is formed by vectors  $d \in B$  such that there exist sequences  $d_N \rightarrow d$  and  $t_N \downarrow 0$  such that  $\mu + t_N d_N \in K$ . Note that  $\mathcal{T}_K(\mu)$  is nonempty only if  $\mu$  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 7.59.

**Theorem 7.61 (Delta Theorem).** *Let  $B_1$  and  $B_2$  be Banach spaces,  $K$  be a subset of  $B_1$ ,  $G : K \rightarrow B_2$  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  $\mu$  tangentially to the set  $K$ , (iii) for some sequence  $\tau_N$  of positive numbers tending to infinity, the sequence  $X_N := \tau_N(Y_N - \mu)$  converges in distribution to a random element  $Y$ , and (iv)  $Y_N \in K$  w.p. 1 for all  $N$  large enough. Then*

$$\tau_N [G(Y_N) - G(\mu)] \xrightarrow{\mathcal{D}} G'_\mu(Y). \quad (7.168)$$

Moreover, if the set  $K$  is convex, then (7.163) holds.

Note that it follows from 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  $\mu$  and define

$$G''_\mu(d) := \lim_{\substack{t \downarrow 0 \\ d' \rightarrow d}} \frac{G(\mu + td') - G(\mu) - tG'_\mu(d')}{\frac{1}{2}t^2}. \quad (7.169)$$

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  $\mu$  if the second order directional derivative  $G''_\mu(d)$ , defined in (7.169), exists for all  $d \in B_1$ . We say that  $G$  is *second order* Hadamard directionally differentiable at  $\mu$  tangentially to a set  $K \subset B_1$  if for all  $d \in \mathcal{T}_K(\mu)$  the limit

$$G''_\mu(d) = \lim_{\substack{t \downarrow 0 \\ d' \rightarrow_K d}} \frac{G(\mu + td') - G(\mu) - tG'_\mu(d')}{\frac{1}{2}t^2} \quad (7.170)$$

exists.

Note that if  $G$  is first and second order Hadamard directionally differentiable at  $\mu$  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 7.62 (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  $B_1$ ,  $G : K \rightarrow B_2$  be a mapping, and  $\tau_N$  be a sequence of positive numbers tending to infinity as  $N \rightarrow \infty$ . Suppose that (i) the space  $B_1$  is separable, (ii)  $G$  is first and second order Hadamard directionally differentiable at  $\mu$  tangentially to the set  $K$ , (iii) the sequence  $X_N := \tau_N(Y_N - \mu)$  converges in distribution to a random element  $Y$  of  $B_1$ , and (iv)  $Y_N \in K$  w.p. 1 for  $N$  large enough. Then*

$$\tau_N^2 [G(Y_N) - G(\mu) - G'_\mu(Y_N - \mu)] \xrightarrow{\mathcal{D}} \frac{1}{2} G''_\mu(Y) \quad (7.171)$$

and

$$G(Y_N) = G(\mu) + G'_\mu(Y_N - \mu) + \frac{1}{2} G''_\mu(Y_N - \mu) + o_p(\tau_N^{-2}). \quad (7.172)$$

**Proof.** Let  $X'_N$ ,  $Y'$ , and  $Y'_N$  be elements as in the proof of Theorem 7.59. Recall that their existence is guaranteed by the representation theorem. Then by the definition of  $G''_\mu$  we have

$$\tau_N^2 [G(Y'_N) - G(\mu) - \tau_N^{-1} G'_\mu(X'_N)] \rightarrow \frac{1}{2} G''_\mu(Y') \quad \text{w.p. 1.}$$

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 (7.171) 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 (7.172) then follows.  $\square$

## 7.2.8 Exponential Bounds of the Large Deviations Theory

Consider an iid sequence  $Y_1, \dots, Y_N$  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  $\Pr(Z_N \geq a) = \Pr(e^{tZ_N} \geq e^{ta})$ , and hence, by Chebyshev's inequality,

$$\Pr(Z_N \geq a) \leq e^{-ta} \mathbb{E}[e^{tZ_N}] = e^{-ta} [M(t/N)]^N,$$

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  $a \geq \mu$ . By taking the logarithm of both sides of the above inequality, changing variables  $t' = t/N$  and minimizing over  $t' > 0$ , we obtain

$$\frac{1}{N} \ln [\Pr(Z_N \geq a)] \leq -I(a), \quad (7.173)$$

where

$$I(z) := \sup_{t \in \mathbb{R}} \{tz - \Lambda(t)\} \quad (7.174)$$

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 (7.173) corresponds to the upper bound of Cramér's LD theorem.

Note that the moment-generating function  $M(\cdot)$  is convex and positive valued,  $M(0) = 1$ , and its domain  $\text{dom} M$  is a subinterval of  $\mathbb{R}$  containing zero. It follows by Theorem 7.44 that  $M(\cdot)$  is infinitely differentiable at every interior point of its domain. Moreover, if  $a := \inf(\text{dom} M)$  is finite, then  $M(\cdot)$  is right-side continuous at  $a$ , and similarly for the  $b := \sup(\text{dom} M)$ . 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,  $\text{dom} \Lambda = \text{dom} M$  and at an interior point  $t$  of  $\text{dom} \Lambda$ ,

$$\Lambda''(t) = \frac{\mathbb{E}[Y^2 e^{tY}] \mathbb{E}[e^{tY}] - \mathbb{E}[Y e^{tY}]^2}{M(t)^2}. \quad (7.175)$$

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.,

$$\mathbb{E}[Y^2 e^{tY}] \mathbb{E}[e^{tY}] - \mathbb{E}[Y e^{tY}]^2 \geq 0.$$

We obtain that  $\Lambda''(\cdot)$  is nonnegative at every point of the interior of  $\text{dom} \Lambda$ , 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  $\bar{t} > 0$ . Then  $M(t)$  is finite for all  $t \in [0, \bar{t}]$  and right-side differentiable at  $t = 0$ . Moreover, the right-side derivative of  $M(t)$  at  $t = 0$  is  $\mu$ , 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 (7.174) is not changed if the constraint  $t > 0$  is removed. If  $a = \mu$ , then the supremum in (7.174) is attained at  $t = 0$  and hence  $I(a) = 0$ . In that case the inequality (7.173) trivially holds. Now if  $M(t) = +\infty$  for all  $t > 0$ , then  $I(a) = 0$  for any  $a \geq \mu$  and the inequality (7.173) trivially holds.

For  $a \leq \mu$  the upper bound (7.173) takes the form

$$\frac{1}{N} \ln [\Pr(Z_N \leq a)] \leq -I(a), \quad (7.176)$$

which of course can be written as

$$\Pr(Z_N \leq a) \leq e^{-I(a)N}. \quad (7.177)$$

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 (7.174) 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  $\sigma^2 := \mathbb{V}\text{ar}[Y]$ . 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,

$$I(a) = \frac{(a - \mu)^2}{2\sigma^2} + o(|a - \mu|^2). \quad (7.178)$$

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

$$I(a) = \frac{(a - \mu)^2}{2\sigma^2}. \quad (7.179)$$

The constant  $I(a)$  in (7.173) gives, in a sense, the best possible exponential rate at which the probability  $\Pr(Z_N \geq a)$  converges to zero. This follows from the lower bound

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \ln [\Pr(Z_N \geq a)] \geq -I(a) \quad (7.180)$$

of Cramér's LD theorem, which holds for  $a \geq \mu$ .

Other closely related, exponential-type inequalities can be derived for bounded random variables.

**Proposition 7.63.** *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*

$$\mathbb{E}[e^{tY}] \leq e^{t^2(b-a)^2/8}, \quad \forall t \geq 0. \quad (7.181)$$

**Proof.** If  $Y$  is identically zero, then (7.181) 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

$$e^Y \leq \frac{b - Y}{b - a} e^a + \frac{Y - a}{b - a} e^b. \quad (7.182)$$

Taking expectation from both sides of (7.182) and using  $\mathbb{E}[Y] = 0$ , we obtain

$$\mathbb{E}[e^Y] \leq \frac{b}{b - a} e^a - \frac{a}{b - a} e^b. \quad (7.183)$$

The right-hand side of (7.182) 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

$$g''(x) = \frac{\alpha(1 - \alpha)}{(1 - \alpha)^2 e^{-x} + \alpha^2 e^x + 2\alpha(1 - \alpha)}. \quad (7.184)$$

Moreover,  $(1 - \alpha)^2 e^{-x} + \alpha^2 e^x \geq 2\alpha(1 - \alpha)$ , and hence  $g''(x) \leq 1/4$  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  $g(u) \leq u^2/8 = (b - a)^2/8$ , and hence

$$\mathbb{E}[e^Y] \leq e^{(b-a)^2/8}. \quad (7.185)$$

Finally, (7.181) follows from (7.185) by rescaling  $Y$  to  $tY$  for  $t \geq 0$ .  $\square$

In particular, if  $|Y| \leq b$  and  $\mathbb{E}[Y] = 0$ , then  $|-Y| \leq b$  and  $\mathbb{E}[-Y] = 0$  as well, and hence by (7.181) we have

$$\mathbb{E}[e^{tY}] \leq e^{t^2 b^2/2}, \quad \forall t \in \mathbb{R}. \quad (7.186)$$

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 (7.181) that the rate function of  $Y - \mu$  satisfies

$$I(z) \geq \sup_{t \in \mathbb{R}} \{tz - t^2(b-a)^2/8\} = 2z^2/(b-a)^2.$$

Together with (7.177) this implies the following. Let  $Y_1, \dots, Y_N$  be an iid sequence of realizations of  $Y$  and  $Z_N$  be the corresponding average. Then for  $\tau > 0$  it holds that

$$\Pr(Z_N \geq \mu + \tau) \leq e^{-2\tau^2 N/(b-a)^2}. \quad (7.187)$$

The bound (7.187) is often referred to as the *Hoeffding inequality*.

In particular, let  $W \sim B(p, n)$  be a random variable having Binomial distribution, i.e.,  $\Pr(W = k) = \binom{n}{k} p^k (1-p)^{n-k}$ ,  $k = 0, \dots, n$ . Recall that  $W$  can be represented as  $W = Y_1 + \dots + Y_n$ , where  $Y_1, \dots, Y_n$  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  $k \leq np$ ,

$$\Pr(W \leq k) \leq \exp \left\{ -\frac{2(np - k)^2}{n} \right\}. \quad (7.188)$$

For small  $p$  it is possible to improve the above estimate as follows. For  $Y \sim \text{Bernoulli}(p)$  we have

$$\mathbb{E}[e^{tY}] = pe^t + 1 - p = 1 - p(1 - e^t).$$

By using the inequality  $e^{-x} \geq 1 - x$  with  $x := p(1 - e^t)$ , we obtain

$$\mathbb{E}[e^{tY}] \leq \exp[p(e^t - 1)],$$

and hence for  $z > 0$ ,

$$I(z) := \sup_{t \in \mathbb{R}} \{tz - \ln \mathbb{E}[e^{tY}]\} \geq \sup_{t \in \mathbb{R}} \{tz - p(e^t - 1)\} = z \ln \frac{z}{p} - z + p.$$

Moreover, since  $\ln(1 + x) \geq x - x^2/2$  for  $x \geq 0$ , we obtain

$$I(z) \geq \frac{(z - p)^2}{2p} \quad \text{for } z \geq p.$$

By (7.173) it follows that

$$\Pr(n^{-1}W \geq z) \leq \exp\{-n(z-p)^2/(2p)\} \quad \text{for } z \geq p. \quad (7.189)$$

Alternatively, this can be written as

$$\Pr(W \leq k) \leq \exp\left\{-\frac{(np-k)^2}{2pn}\right\} \quad (7.190)$$

for a nonnegative integer  $k \leq np$ . The above inequality (7.190) is often called the *Chernoff inequality*. For small  $p$  it can be significantly better than the Hoeffding inequality (7.188).

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 (7.174) with the supremum taken over  $t \in \mathbb{R}^d$  and  $tz$  denoting 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:

$$\begin{aligned} -\inf_{z \in \text{int}(A)} I(z) &\leq \liminf_{N \rightarrow \infty} N^{-1} \ln [\Pr(Z_N \in A)] \\ &\leq \limsup_{N \rightarrow \infty} N^{-1} \ln [\Pr(Z_N \in A)] \\ &\leq -\inf_{z \in \text{cl}(A)} I(z), \end{aligned} \quad (7.191)$$

where  $\text{int}(A)$  and  $\text{cl}(A)$  denote the interior and topological closure, respectively, of the set  $A$ . In the above one-dimensional setting, the LD principle (7.191) was derived for sets  $A := [a, +\infty)$ .

We have that if  $\mu \in \text{int}(A)$  and the moment-generating function  $M(t)$  is finite valued for all  $t$  in a neighborhood of  $0 \in \mathbb{R}^d$ , then  $\inf_{z \in \mathbb{R}^d \setminus \text{int}(A)} I(z)$  is positive. Moreover, if the sequence is iid, then

$$\limsup_{N \rightarrow \infty} N^{-1} \ln [\Pr(Z_N \notin A)] < 0, \quad (7.192)$$

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 7.64.** *Let  $\xi^1, \xi^2, \dots$  be a sequence of iid random variables (vectors),  $\sigma_t > 0$ ,  $t = 1, \dots$ , be a sequence of deterministic numbers, and  $\phi_t = \phi_t(\xi_{[t]})$  be (measurable) functions of  $\xi_{[t]} = (\xi^1, \dots, \xi^t)$  such that*

$$\mathbb{E}[\phi_t | \xi_{[t-1]}] = 0 \quad \text{and} \quad \mathbb{E}[\exp\{\phi_t^2/\sigma_t^2\} | \xi_{[t-1]}] \leq \exp\{1\} \quad \text{w.p. 1.} \quad (7.193)$$

Then for any  $\Theta \geq 0$ ,

$$\Pr\left\{\sum_{t=1}^N \phi_t \geq \Theta \sqrt{\sum_{t=1}^N \sigma_t^2}\right\} \leq \exp\{-\Theta^2/3\}. \quad (7.194)$$

**Proof.** Let us set  $\tilde{\phi}_t := \phi_t/\sigma_t$ . By condition (7.193) we have that  $\mathbb{E}[\tilde{\phi}_t|\xi_{[t-1]}] = 0$  and  $\mathbb{E}[\exp\{\tilde{\phi}_t^2|\xi_{[t-1]}\}] \leq \exp\{1\}$  w.p. 1. By the Jensen inequality it follows that for any  $a \in [0, 1]$ ,

$$\mathbb{E}[\exp\{a\tilde{\phi}_t^2|\xi_{[t-1]}\}] = \mathbb{E}[(\exp\{\tilde{\phi}_t^2\})^a|\xi_{[t-1]}] \leq \left(\mathbb{E}[\exp\{\tilde{\phi}_t^2|\xi_{[t-1]}\}]\right)^a \leq \exp\{a\}.$$

We also have that  $\exp\{x\} \leq x + \exp\{9x^2/16\}$  for all  $x$  (this inequality can be verified by direct calculations), and hence for any  $\lambda \in [0, 4/3]$ ,

$$\mathbb{E}[\exp\{\lambda\tilde{\phi}_t|\xi_{[t-1]}\}] \leq \mathbb{E}[\exp\{(9\lambda^2/16)\tilde{\phi}_t^2|\xi_{[t-1]}\}] \leq \exp\{9\lambda^2/16\}. \quad (7.195)$$

Moreover, we have that  $\lambda x \leq \frac{3}{8}\lambda^2 + \frac{2}{3}x^2$  for any  $\lambda$  and  $x$ , and hence

$$\mathbb{E}[\exp\{\lambda\tilde{\phi}_t|\xi_{[t-1]}\}] \leq \exp\{3\lambda^2/8\}\mathbb{E}[\exp\{2\tilde{\phi}_t^2/3|\xi_{[t-1]}\}] \leq \exp\{2/3 + 3\lambda^2/8\}.$$

Combining the latter inequality with (7.195), we get

$$\mathbb{E}[\exp\{\lambda\tilde{\phi}_t|\xi_{[t-1]}\}] \leq \exp\{3\lambda^2/4\}, \quad \forall \lambda \geq 0.$$

Going back to  $\phi_t$ , the above inequality reads

$$\mathbb{E}[\exp\{\gamma\phi_t|\xi_{[t-1]}\}] \leq \exp\{3\gamma^2\sigma_t^2/4\}, \quad \forall \gamma \geq 0. \quad (7.196)$$

Now, since  $\phi_\tau$  is a deterministic function of  $\xi_{[\tau]}$  and using (7.196), we obtain for any  $\gamma \geq 0$ ,

$$\begin{aligned} \mathbb{E}[\exp\{\gamma \sum_{t=1}^N \phi_t\}] &= \mathbb{E}[\exp\{\gamma \sum_{t=1}^{N-1} \phi_t\} \mathbb{E}(\exp\{\gamma\phi_N|\xi_{[N-1]}\})] \\ &\leq \exp\{3\gamma^2\sigma_N^2/4\} \mathbb{E}[\exp\{\gamma \sum_{t=1}^{N-1} \phi_t\}] \end{aligned}$$

and hence

$$\mathbb{E}[\exp\{\gamma \sum_{t=1}^N \phi_t\}] \leq \exp\{3\gamma^2 \sum_{t=1}^N \sigma_t^2/4\}. \quad (7.197)$$

By Chebyshev's inequality, we have for  $\gamma > 0$  and  $\Theta$ ,

$$\begin{aligned} \Pr\left\{\sum_{t=1}^N \phi_t \geq \Theta \sqrt{\sum_{t=1}^N \sigma_t^2}\right\} &= \Pr\left\{\exp\left[\gamma \sum_{t=1}^N \phi_t\right] \geq \exp\left[\gamma \Theta \sqrt{\sum_{t=1}^N \sigma_t^2}\right]\right\} \\ &\leq \exp\left[-\gamma \Theta \sqrt{\sum_{t=1}^N \sigma_t^2}\right] \mathbb{E}\left\{\exp\left[\gamma \sum_{t=1}^N \phi_t\right]\right\}. \end{aligned}$$

Together with (7.197) this implies for  $\Theta \geq 0$ ,

$$\begin{aligned} \Pr\left\{\sum_{t=1}^N \phi_t \geq \Theta \sqrt{\sum_{t=1}^N \sigma_t^2}\right\} &\leq \inf_{\gamma>0} \exp\left\{\frac{3}{4}\gamma^2 \sum_{t=1}^N \sigma_t^2 - \gamma \Theta \sqrt{\sum_{t=1}^N \sigma_t^2}\right\} \\ &= \exp\{-\Theta^2/3\}. \end{aligned}$$

This completes the proof.  $\square$

### 7.2.9 Uniform Exponential Bounds

Consider the setting of section 7.2.5 with a sequence  $\xi^i, i \in \mathbb{N}$ , of random realizations of an  $d$ -dimensional random vector  $\xi = \xi(\omega)$ , a function  $F : X \times \Xi \rightarrow \mathbb{R}$ , and the corresponding sample average function  $\hat{f}_N(x)$ . We assume here that the sequence  $\xi^i, i \in \mathbb{N}$ , 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 now discuss uniform exponential rates of convergence of  $\hat{f}_N(x)$  to  $f(x)$ . Denote by

$$M_x(t) := \mathbb{E}[e^{t(F(x, \xi) - f(x))}]$$

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  $\kappa : \Xi \rightarrow \mathbb{R}_+$  such that

$$|F(x', \xi) - F(x, \xi)| \leq \kappa(\xi) \|x' - x\| \quad (7.198)$$

for all  $\xi \in \Xi$  and all  $x', x \in X$ .

(C3) The moment-generating function  $M_\kappa(t) := \mathbb{E}[e^{t\kappa(\xi)}]$  of  $\kappa(\xi)$  is finite valued for all  $t$  in a neighborhood of zero.

**Theorem 7.65.** *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*

$$\Pr \left\{ \sup_{x \in X} |\hat{f}_N(x) - f(x)| \geq \varepsilon \right\} \leq C e^{-N\beta}. \quad (7.199)$$

**Proof.** By the upper bound (7.173) of Cramér's LD theorem, we have that for any  $x \in X$  and  $\varepsilon > 0$  it holds that

$$\Pr \{ \hat{f}_N(x) - f(x) \geq \varepsilon \} \leq \exp \{ -N I_x(\varepsilon) \}, \quad (7.200)$$

where

$$I_x(z) := \sup_{t \in \mathbb{R}} \{ zt - \ln M_x(t) \} \quad (7.201)$$

is the LD rate function of random variable  $F(x, \xi) - f(x)$ . Similarly,

$$\Pr \{ \hat{f}_N(x) - f(x) \leq -\varepsilon \} \leq \exp \{ -N I_x(-\varepsilon) \},$$

and hence

$$\Pr \left\{ |\hat{f}_N(x) - f(x)| \geq \varepsilon \right\} \leq \exp \{ -N I_x(\varepsilon) \} + \exp \{ -N I_x(-\varepsilon) \}. \quad (7.202)$$

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  $\bar{x}_1, \dots, \bar{x}_K \in X$  be such that for every  $x \in X$  there exists  $\bar{x}_i$ ,  $i \in \{1, \dots, K\}$ , such that  $\|x - \bar{x}_i\| \leq \nu$ , i.e.,  $\{\bar{x}_1, \dots, \bar{x}_K\}$  is a  $\nu$ -net in  $X$ . We can choose this net in such a way that

$$K \leq [\varrho D/\nu]^n, \quad (7.203)$$

where

$$D := \sup_{x', x \in X} \|x' - x\|$$

is the diameter of  $X$  and  $\varrho$  is a constant depending on the chosen norm  $\|\cdot\|$ . By (7.198) we have that

$$|f(x') - f(x)| \leq L\|x' - x\|, \quad (7.204)$$

where  $L := \mathbb{E}[\kappa(\xi)]$  is finite by assumption (C3). Moreover,

$$|\hat{f}_N(x') - \hat{f}_N(x)| \leq \hat{\kappa}_N \|x' - x\|, \quad (7.205)$$

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

$$\Pr \{\hat{\kappa}_N \geq L'\} \leq \exp\{-N\ell\}. \quad (7.206)$$

Consider

$$Z_i := \hat{f}_N(\bar{x}_i) - f(\bar{x}_i), \quad i = 1, \dots, K.$$

We have that the event  $\{\max_{1 \leq i \leq K} |Z_i| \geq \varepsilon\}$  is equal to the union of the events  $\{|Z_i| \geq \varepsilon\}$ ,  $i = 1, \dots, K$ , and hence

$$\Pr \{\max_{1 \leq i \leq K} |Z_i| \geq \varepsilon\} \leq \sum_{i=1}^K \Pr \{|Z_i| \geq \varepsilon\}.$$

Together with (7.202) this implies that

$$\Pr \left\{ \max_{1 \leq i \leq K} |\hat{f}_N(\bar{x}_i) - f(\bar{x}_i)| \geq \varepsilon \right\} \leq 2 \sum_{i=1}^K \exp \{ -N[I_{\bar{x}_i}(\varepsilon) \wedge I_{\bar{x}_i}(-\varepsilon)] \}. \quad (7.207)$$

For an  $x \in X$  let  $i(x) \in \arg \min_{1 \leq i \leq K} \|x - \bar{x}_i\|$ . By construction of the  $\nu$ -net we have that  $\|x - \bar{x}_{i(x)}\| \leq \nu$  for every  $x \in X$ . Then

$$\begin{aligned} |\hat{f}_N(x) - f(x)| &\leq |\hat{f}_N(x) - \hat{f}_N(\bar{x}_{i(x)})| + |\hat{f}_N(\bar{x}_{i(x)}) - f(\bar{x}_{i(x)})| + |f(\bar{x}_{i(x)}) - f(x)| \\ &\leq \hat{\kappa}_N \nu + |\hat{f}_N(\bar{x}_{i(x)}) - f(\bar{x}_{i(x)})| + L\nu. \end{aligned}$$

Let us take now a  $\nu$ -net with such  $\nu$  that  $L\nu = \varepsilon/4$ , i.e.,  $\nu := \varepsilon/(4L)$ . Then

$$\Pr \left\{ \sup_{x \in X} |\hat{f}_N(x) - f(x)| \geq \varepsilon \right\} \leq \Pr \left\{ \hat{\kappa}_N \nu + \max_{1 \leq i \leq K} |\hat{f}_N(\bar{x}_i) - f(\bar{x}_i)| \geq 3\varepsilon/4 \right\}.$$

Moreover, we have that

$$\Pr \{\hat{\kappa}_N \nu \geq \varepsilon/2\} \leq \exp\{-N\ell\},$$

where  $\ell$  is a positive constant specified in (7.206) for  $L' := 2L$ . Consequently

$$\begin{aligned} & \Pr \left\{ \sup_{x \in X} |\hat{f}_N(x) - f(x)| \geq \varepsilon \right\} \\ & \leq \exp\{-N\ell\} + \Pr \left\{ \max_{1 \leq i \leq K} |\hat{f}_N(\bar{x}_i) - f(\bar{x}_i)| \geq \varepsilon/4 \right\} \\ & \leq \exp\{-N\ell\} + 2 \sum_{i=1}^K \exp \left\{ -N [I_{\bar{x}_i}(\varepsilon/4) \wedge I_{\bar{x}_i}(-\varepsilon/4)] \right\}. \end{aligned} \quad (7.208)$$

Since the above choice of the  $\nu$ -net does not depend on the sample (although it depends on  $\varepsilon$ ), and both  $I_{\bar{x}_i}(\varepsilon/4)$  and  $I_{\bar{x}_i}(-\varepsilon/4)$  are positive,  $i = 1, \dots, K$ , we obtain that (7.208) implies (7.199), and hence completes the proof.  $\square$

In the convex case the (Lipschitz continuity) condition (C2) holds, in a sense, automatically. That is, we have the following result.

**Theorem 7.66.** *Let  $U \subset \mathbb{R}^n$  be a convex open set. Suppose that (i) for a.e.  $\xi \in \Xi$  the function  $F(\cdot, \xi) : U \rightarrow \mathbb{R}$  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*

$$\Pr \left\{ \sup_{x \in X} |\hat{f}_N(x) - f(x)| \geq \varepsilon \right\} \leq C e^{-N\beta}. \quad (7.209)$$

**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  $\gamma \geq 0$  consider the set

$$X_\gamma := \{x \in \mathbb{R}^n : \text{dist}(x, X) \leq \gamma\}.$$

Since the set  $U$  is open, we can choose  $\gamma > 0$  such that  $X_\gamma \subset U$ . The set  $X_\gamma$  is compact and by convexity of  $f(\cdot)$  we have that  $f(\cdot)$  is continuous and hence is bounded on  $X_\gamma$ . That is, there is constant  $c > 0$  such that  $|f(x)| \leq c$  for all  $x \in X_\gamma$ . Also by convexity of  $f(\cdot)$  we have for any  $\tau \in [0, 1]$  and  $x, y \in \mathbb{R}^n$  such that  $x + y, x - y/\tau \in U$ :

$$f(x) = f\left(\frac{1}{1+\tau}(x+y) + \frac{\tau}{1+\tau}(x-y/\tau)\right) \leq \frac{1}{1+\tau}f(x+y) + \frac{\tau}{1+\tau}f(x-y/\tau).$$

It follows that if  $x, x+y, x-y/\tau \in X_\gamma$ , then

$$f(x+y) \geq (1+\tau)f(x) - \tau f(x-y/\tau) \geq f(x) - 2\tau c. \quad (7.210)$$

Now we proceed similar to the proof of Theorem 7.65. Let  $\varepsilon > 0$  and  $\nu > 0$ , and let  $\bar{x}_1, \dots, \bar{x}_K \in X_{\gamma/2}$  be a  $\nu$ -net for  $X_{\gamma/2}$ . As in the proof of Theorem 7.65, this  $\nu$ -net will be dependent on  $\varepsilon$  but not on the random sample  $\xi^1, \dots, \xi^N$ . Consider the event

$$A_N := \left\{ \max_{1 \leq i \leq K} |\hat{f}_N(\bar{x}_i) - f(\bar{x}_i)| \leq \varepsilon \right\}.$$

By (7.200) and (7.202) we have similar to (7.207) that  $\Pr(A_N) \geq 1 - \alpha_N$ , 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  $\mathcal{I} \subset \{1, \dots, K\}$  be such an 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  $\left\{ \hat{f}_N(x) \leq \sum_{i \in \mathcal{I}} t_i f(\bar{x}_i) + \varepsilon \right\}$ . By (7.210) we also have that

$$f(x) \geq f(\bar{x}_i) - 2\tau c, \quad \forall i \in \mathcal{I},$$

provided that  $a\nu \leq \tau\gamma/2$ . 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 that<sup>70</sup>  $\nu \leq O(1)\varepsilon$ . It follows that the event  $A_N$  is included in the event  $\cap_{x \in X} B_x$ , and hence

$$\Pr \left\{ \sup_{x \in X} \left( \hat{f}_N(x) - f(x) \right) \leq 2\varepsilon \right\} = \Pr \{ \cap_{x \in X} B_x \} \geq \Pr \{ A_N \} \geq 1 - \alpha_N, \quad (7.211)$$

provided that  $\nu \leq O(1)\varepsilon$ .

In order to derive the converse to (7.211) estimate let us observe that by convexity of  $\hat{f}_N(\cdot)$  we have with probability at least  $1 - \alpha_N$  that  $\sup_{x \in X_\gamma} \hat{f}_N(x) \leq c + \varepsilon$ . Also, by using (7.210) we have with probability at least  $1 - \alpha_N$  that  $\inf_{x \in X_\gamma} \hat{f}_N(x) \geq -(c + \varepsilon)$ , provided that  $\nu \leq O(1)\varepsilon$ . That is, with probability at least  $1 - 2\alpha_N$  we have that

$$\sup_{x \in X_\gamma} |\hat{f}_N(x)| \leq c + \varepsilon,$$

provided that  $\nu \leq O(1)\varepsilon$ . We can now proceed in the same way as above to show that

$$\Pr \left\{ \sup_{x \in X} \left( f(x) - \hat{f}_N(x) \right) \leq 2\varepsilon \right\} \geq 1 - 3\alpha_N. \quad (7.212)$$

Since by condition (ii)  $I_{\bar{x}_i}(\varepsilon)$  and  $I_{\bar{x}_i}(-\varepsilon)$  are positive, this completes the proof.  $\square$

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:

$$M_x(t) \leq \exp \{ \sigma^2 t^2 / 2 \}, \quad \forall t \in \mathbb{R}. \quad (7.213)$$

It follows from condition (7.213) that  $\ln M_x(t) \leq \sigma^2 t^2 / 2$ , and hence<sup>71</sup>

$$I_x(z) \geq \frac{z^2}{2\sigma^2}, \quad \forall z \in \mathbb{R}. \quad (7.214)$$

Consequently, inequality (7.208) implies

$$\Pr \left\{ \sup_{x \in X} |\hat{f}_N(x) - f(x)| \geq \varepsilon \right\} \leq \exp \{ -N\ell \} + 2K \exp \left\{ -\frac{N\varepsilon^2}{32\sigma^2} \right\}, \quad (7.215)$$

<sup>70</sup>Recall that  $O(1)$  denotes a generic constant, here  $O(1) = \gamma/(2ca)$ .

<sup>71</sup>Recall that if random variable  $F(x, \xi) - f(x)$  has normal distribution with variance  $\sigma^2$ , then its moment generating function is equal to the right-hand side of (7.213), and hence the inequalities (7.213) and (7.214) hold as equalities.



where  $\ell$  is a constant specified in (7.206) with  $L' := 2L$ ,  $K = [\varrho D/\nu]^n$ ,  $\nu = \varepsilon/(4L)$ , and hence

$$K = [4\varrho DL/\varepsilon]^n. \quad (7.216)$$

If we assume further that the Lipschitz constant in (7.198) does not depend on  $\xi$ , i.e.,  $\kappa(\xi) \equiv L$ , then the first term in the right-hand side of (7.215) can be omitted. Therefore we obtain the following result.

**Theorem 7.67.** *Suppose that conditions (C2)–(C4) hold and that the set  $X$  has finite diameter  $D$ . Then*

$$\Pr \left\{ \sup_{x \in X} |\hat{f}_N(x) - f(x)| \geq \varepsilon \right\} \leq \exp\{-N\ell\} + 2 \left[ \frac{4\varrho DL}{\varepsilon} \right]^n \exp \left\{ -\frac{N\varepsilon^2}{32\sigma^2} \right\}. \quad (7.217)$$

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 (7.217) can be omitted.

As shown in the proof of Theorem 7.66, in the convex case estimates of the form (7.217), with different constants, can be obtained without assuming the (Lipschitz continuity) condition (C2).

### 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., [38, Theorem 2.7.2]).

**Theorem 7.68.** *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*

$$\partial^\circ f(\bar{x}) = \int_{\Xi} \partial^\circ F(\bar{x}, \xi) dP(\xi), \quad (7.218)$$

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 (7.218) 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,  $\gamma$  can be considered as an integral of a measurable selection from  $\partial^\circ F(\bar{x}, \cdot)$ .

**Theorem 7.69.** *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 that<sup>72</sup>*

$$\Pr \left\{ \sup_{d \in S^{n-1}} |\hat{f}'_N(\bar{x}, d) - f'(\bar{x}, d)| > \varepsilon \right\} \leq Ce^{-N\beta}. \quad (7.219)$$

Moreover, suppose that for a.e.  $\xi \in \Xi$  the function  $F(\cdot, \xi)$  is Clarke-regular at  $\bar{x}$ . Then

$$\Pr \left\{ \mathbb{H} \left( \partial^\circ \hat{f}_N(\bar{x}), \partial^\circ f(\bar{x}) \right) > \varepsilon \right\} \leq Ce^{-N\beta}. \quad (7.220)$$

Furthermore, if in condition (C2)  $\kappa(\xi) \equiv L$  is constant, then

$$\Pr \left\{ \mathbb{H} \left( \partial^\circ \hat{f}_N(\bar{x}), \partial^\circ f(\bar{x}) \right) > \varepsilon \right\} \leq 2 \left[ \frac{4\varrho L}{\varepsilon} \right]^n \exp \left\{ -\frac{N\varepsilon^2}{128L^2} \right\}. \quad (7.221)$$

**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 7.44). We also have here that  $\hat{f}'_N(\bar{x}, \cdot) = \hat{\eta}_N(\cdot)$ , where

$$\hat{\eta}_N(d) := \frac{1}{N} \sum_{i=1}^N \eta(d, \xi^i), \quad d \in \mathbb{R}^n, \quad (7.222)$$

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 (7.219) follows directly from Theorem 7.65. 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 (7.219) together with (7.145) for sets  $A_1 := \partial^\circ \hat{f}_N(\bar{x})$  and  $A_2 := \partial^\circ f(\bar{x})$ , we obtain (7.220).

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  $d \in$

<sup>72</sup>By  $S^{n-1} := \{d \in \mathbb{R}^n : \|d\| = 1\}$  we denote the unit sphere taken with respect to a norm  $\|\cdot\|$  on  $\mathbb{R}^n$ .

$S^{n-1}$  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  $M_d(t) \leq \exp\{2t^2 L^2\}$ , for all  $t \in \mathbb{R}$  (see (7.186)). It follows by Theorem 7.67 that

$$\Pr \left\{ \sup_{d \in S^{n-1}} |\hat{f}'_N(\bar{x}, d) - f'(\bar{x}, d)| > \varepsilon \right\} \leq 2 \left[ \frac{4\varrho L}{\varepsilon} \right]^n \exp \left\{ -\frac{N\varepsilon^2}{128L^2} \right\}, \quad (7.223)$$

and hence (7.221) follows.  $\square$

### 7.3 Elements of Functional Analysis

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 (convergence, continuity, lower continuity, etc.) will be made with respect to the norm topology of  $\mathcal{Z}$ .

The space of all linear continuous functionals  $\zeta : \mathcal{Z} \rightarrow \mathbb{R}$  forms the dual of space  $\mathcal{Z}$  and is denoted  $\mathcal{Z}^*$ . For  $\zeta \in \mathcal{Z}^*$  and  $z \in \mathcal{Z}$  we denote  $\langle \zeta, z \rangle := \zeta(z)$  and view it as a scalar product on  $\mathcal{Z}^* \times \mathcal{Z}$ . The space  $\mathcal{Z}^*$ , equipped with the dual norm

$$\|\zeta\|_* := \sup_{\|z\| \leq 1} \langle \zeta, z \rangle, \quad (7.224)$$

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

$$|\langle \zeta, z \rangle| \leq \|\zeta\|_* \|z\|, \quad z \in \mathcal{Z}, \zeta \in \mathcal{Z}^*. \quad (7.225)$$

Also to every  $z \in \mathcal{Z}$  corresponds set

$$\mathfrak{S}_z := \arg \max \{ \langle \zeta, z \rangle : \zeta \in \mathcal{Z}^*, \|\zeta\| \leq 1 \}. \quad (7.226)$$

The set  $\mathfrak{S}_z$  is always nonempty and will be referred to as the *set of contact points* of  $z \in \mathcal{Z}$ . Every point of  $\mathfrak{S}_z$  will be called a *contact point* of  $z$ .

An important class of Banach spaces are  $\mathcal{L}_p(\Omega, \mathcal{F}, P)$  spaces, where  $(\Omega, \mathcal{F})$  is a sample space, equipped with sigma algebra  $\mathcal{F}$  and probability measure  $P$ , and  $p \in [1, +\infty)$ . The space  $\mathcal{L}_p(\Omega, \mathcal{F}, P)$  consists of all  $\mathcal{F}$ -measurable functions  $\phi : \Omega \rightarrow \mathbb{R}$  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

$$\|\phi\|_p := \left( \int_{\Omega} |\phi(\omega)|^p dP(\omega) \right)^{1/p}, \quad (7.227)$$

$\mathcal{L}_p(\Omega, \mathcal{F}, P)$  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)  $\phi : \Omega \rightarrow \mathbb{R}$  which are  $\mathcal{F}$ -measurable and essentially bounded. A function  $\phi$  is said to be essentially bounded if its sup-norm

$$\|\phi\|_{\infty} := \operatorname{ess\,sup}_{\omega \in \Omega} |\phi(\omega)| \quad (7.228)$$

is finite, where

$$\operatorname{ess\,sup}_{\omega \in \Omega} |\phi(\omega)| := \inf \left\{ \sup_{\omega \in \Omega} |\psi(\omega)| : \phi(\omega) = \psi(\omega) \text{ a.e. } \omega \in \Omega \right\}.$$

In particular, suppose that the set  $\Omega := \{\omega_1, \dots, \omega_K\}$  is finite, and let  $\mathcal{F}$  be the sigma algebra of all subsets of  $\Omega$  and  $p_1, \dots, p_K$  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  $(z(\omega_1), \dots, z(\omega_K))$ , and  $\mathcal{L}_p(\Omega, \mathcal{F}, P)$  can be identified with the space  $\mathbb{R}^K$  equipped with the corresponding norm

$$\|z\|_p := \left( \sum_{k=1}^K p_k |z(\omega_k)|^p \right)^{1/p}. \quad (7.229)$$

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)  $\psi : \Omega \rightarrow \mathbb{R}^m$  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  $\psi : \Omega \rightarrow \mathbb{R}^m$ .

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

$$\int_{\Omega} |\zeta(\omega)z(\omega)| dP(\omega) \leq \left( \int_{\Omega} |\zeta(\omega)|^q dP(\omega) \right)^{1/q} \left( \int_{\Omega} |z(\omega)|^p dP(\omega) \right)^{1/p}. \quad (7.230)$$

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

$$\langle \zeta, z \rangle := \int_{\Omega} \zeta(\omega)z(\omega) dP(\omega). \quad (7.231)$$

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 (7.230) holds iff  $\zeta(\omega)$  is proportional<sup>73</sup> to  $\operatorname{sign}(z(\omega))|z(\omega)|^{1/(q-1)}$ . It follows that for  $p \in (1, +\infty)$ , with every nonzero  $z \in \mathcal{L}_p(\Omega, \mathcal{F}, P)$  is associated unique contact point, denoted  $\tilde{\zeta}_z$ , which can be written in the form

$$\tilde{\zeta}_z(\omega) = \frac{\operatorname{sign}(z(\omega))|z(\omega)|^{1/(q-1)}}{\|z\|_p^{q/p}}. \quad (7.232)$$

In particular, for  $p = 2$  and  $q = 2$  the contact point is  $\tilde{\zeta}_z = \|z\|_2^{-1}z$ . Of course, if  $z = 0$ , then  $\mathfrak{S}_0 = \{\zeta \in \mathcal{Z}^* : \|\zeta\|_* \leq 1\}$ .

For  $p = 1$  and  $z \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$  the corresponding set of contact points can be described as follows:

$$\mathfrak{S}_z = \left\{ \zeta \in \mathcal{L}_{\infty}(\Omega, \mathcal{F}, P) : \begin{array}{ll} \zeta(\omega) = 1 & \text{if } z(\omega) > 0, \\ \zeta(\omega) = -1 & \text{if } z(\omega) < 0, \\ \zeta(\omega) \in [-1, 1] & \text{if } z(\omega) = 0. \end{array} \right. \quad (7.233)$$

It follows that  $\mathfrak{S}_z$  is a singleton iff  $z(\omega) \neq 0$  for a.e.  $\omega \in \Omega$ .

<sup>73</sup>For  $a \in \mathbb{R}$ ,  $\operatorname{sign}(a)$  is equal to 1 if  $a > 0$ , to  $-1$  if  $a < 0$ , and to 0 if  $a = 0$ .

Together with the strong (norm) topology of  $\mathcal{Z}$  we sometimes need to consider its weak topology, which is the weakest topology in which all linear functionals  $\langle \zeta, \cdot \rangle$ ,  $\zeta \in \mathcal{Z}^*$ , are continuous. The dual space  $\mathcal{Z}^*$  can be also equipped with its weak\* topology, which is the weakest topology in which all linear functionals  $\langle \cdot, z \rangle$ ,  $z \in \mathcal{Z}$ , are continuous. If the space  $\mathcal{Z}$  is reflexive, then  $\mathcal{Z}^*$  is also reflexive and its weak\* and weak topologies do coincide. Note also that a convex subset of  $\mathcal{Z}$  is closed in the strong topology iff it is closed in the weak topology of  $\mathcal{Z}$ .

**Theorem 7.70 (Banach–Alaoglu).** *Let  $\mathcal{Z}$  be Banach space. The closed unit ball  $\{\zeta \in \mathcal{Z}^* : \|\zeta\|_* \leq 1\}$  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.

### 7.3.1 Conjugate Duality and Differentiability

Let  $\mathcal{Z}$  be a Banach space,  $\mathcal{Z}^*$  be its dual space and  $f : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  be an extended real valued function. Similar to the finite dimensional case we define the *conjugate function* of  $f$  as

$$f^*(\zeta) := \sup_{z \in \mathcal{Z}} \{\langle \zeta, z \rangle - f(z)\}. \quad (7.234)$$

The conjugate function  $f^* : \mathcal{Z}^* \rightarrow \overline{\mathbb{R}}$  is always convex and lower semicontinuous. The biconjugate function  $f^{**} : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$ , i.e., the conjugate of  $f^*$ , is

$$f^{**}(z) := \sup_{\zeta \in \mathcal{Z}^*} \{\langle \zeta, z \rangle - f^*(\zeta)\}. \quad (7.235)$$

The basic duality theorem still holds in the considered infinite dimensional framework.

**Theorem 7.71 (Fenchel–Moreau).** *Let  $\mathcal{Z}$  be a Banach space and  $f : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  be a proper extended real valued convex function. Then*

$$f^{**} = \text{lsc } f. \quad (7.236)$$

It follows from (7.236) 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 7.72.** *Let  $\mathcal{Z}$  be a Banach space and  $f : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  be a convex lower semicontinuous function having a finite value in at least one point. Then  $f$  is proper and is continuous on  $\text{int}(\text{dom } f)$ .*

In particular, it follows from the above proposition that if  $f : \mathcal{Z} \rightarrow \mathbb{R}$  is real valued convex and lower semicontinuous, then  $f$  is continuous on  $\mathcal{Z}$ .

The subdifferential of a function  $f : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$ , 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,

$$\partial f(z_0) := \{\zeta \in \mathcal{Z}^* : f(z) - f(z_0) \geq \langle \zeta, z - z_0 \rangle, \quad \forall z \in \mathcal{Z}\}. \quad (7.237)$$

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  $z_0 \in \mathcal{Z}$ , then  $f$  is proper and lower semicontinuous at  $z_0$ . Similar to the finite dimensional case, we have the following.

**Proposition 7.73.** *Let  $\mathcal{Z}$  be a Banach space and  $f : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  be a convex function and  $z \in \mathcal{Z}$  be such that  $f^{**}(z)$  is finite. Then*

$$\partial f^{**}(z) = \arg \max_{\zeta \in \mathcal{Z}^*} \{ \langle \zeta, z \rangle - f^*(\zeta) \}. \quad (7.238)$$

Moreover, if  $f^{**}(z) = f(z)$ , then  $\partial f^{**}(z) = \partial f(z)$ .

**Proposition 7.74.** *Let  $\mathcal{Z}$  be a Banach space,  $f : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  be a convex function. Suppose that  $f$  is finite valued and continuous at a point  $z_0 \in \mathcal{Z}$ . Then  $f$  is subdifferentiable at  $z_0$ ,  $\partial f(z_0)$  is nonempty, convex, bounded, and weakly\* compact subset of  $\mathcal{Z}^*$ ,  $f$  is Hadamard directionally differentiable at  $z_0$  and*

$$f'(z_0, h) = \sup_{\zeta \in \partial f(z_0)} \langle \zeta, h \rangle. \quad (7.239)$$

Note that by the definition, every element of the subdifferential  $\partial f(z_0)$  (called *subgradient*) is a continuous linear functional on  $\mathcal{Z}$ . A linear (not necessarily continuous) functional  $\ell : \mathcal{Z} \rightarrow \mathbb{R}$  is called an *algebraic subgradient* of  $f$  at  $z_0$  if

$$f(z_0 + h) - f(z_0) \geq \ell(h), \quad \forall h \in \mathcal{Z}. \quad (7.240)$$

Of course, if the algebraic subgradient  $\ell$  is also continuous, then  $\ell \in \partial f(z_0)$ .

**Proposition 7.75.** *Let  $\mathcal{Z}$  be a Banach space and  $f : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  be a proper convex function. Then the set of algebraic subgradients at any point  $z_0 \in \text{int}(\text{dom } f)$  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

$$f'(z_0, h) = \inf_{t > 0} \frac{f(z_0 + th) - f(z_0)}{t}, \quad (7.241)$$

and  $\delta(\cdot)$  is convex, positively homogeneous. Moreover, since  $z_0 \in \text{int}(\text{dom } f)$  and hence  $f(z)$  is finite valued for all  $z$  in a neighborhood of  $z_0$ , it follows by (7.241) that  $\delta(h)$  is finite valued for all  $h \in \mathcal{Z}$ . That is,  $\delta(\cdot)$  is a real valued subadditive and positively homogeneous function. Consequently, by the Hahn–Banach theorem we have that there exists a linear functional  $\ell : \mathcal{Z} \rightarrow \mathbb{R}$  such that  $\delta(h) \geq \ell(h)$  for all  $h \in \mathcal{Z}$ . Since  $f(z_0 + h) \geq f(z_0) + \delta(h)$  for any  $h \in \mathcal{Z}$ , it follows that  $\ell$  is an algebraic subgradient of  $f$  at  $z_0$ .  $\square$

There is also the following version of the Moreau–Rockafellar theorem in the infinite dimensional setting.

**Theorem 7.76 (Moreau–Rockafellar).** *Let  $f_1, f_2 : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  be convex proper lower semicontinuous functions,  $f := f_1 + f_2$  and  $\bar{z} \in \text{dom}(f_1) \cap \text{dom}(f_2)$ . Then*

$$\partial f(\bar{z}) = \partial f_1(\bar{z}) + \partial f_2(\bar{z}), \quad (7.242)$$

provided that the following regularity condition holds:

$$0 \in \text{int} \{ \text{dom}(f_1) - \text{dom}(f_2) \}. \quad (7.243)$$

In particular, (7.242) holds if  $f_1$  is continuous at  $\bar{z}$ .

**Remark 34.** It is possible to derive the following (first order) necessary optimality condition from the above theorem. Let  $S$  be a convex closed subset of  $\mathcal{Z}$  and  $f : \mathcal{Z} \rightarrow \bar{\mathbb{R}}$  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  $z \in \mathcal{Z}$ . 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

$$0 \in \partial f(z_0) + \mathcal{N}_S(z_0), \quad (7.244)$$

provided that  $0 \in \text{int} \{ \text{dom}(f) - S \}$ . In particular, (7.244) 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,  $\psi : \mathcal{X} \times \mathcal{Y} \rightarrow \bar{\mathbb{R}}$  and  $\vartheta(y) := \inf_{x \in \mathcal{X}} \psi(x, y)$ .

**Theorem 7.77.** 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.,  $\bar{y} \in \text{int}(\text{dom } \vartheta)$ .

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 Lattice Structure

Let  $\mathcal{C} \subset \mathcal{Z}$  be a closed convex pointed<sup>74</sup> cone. It defines an order relation on the space  $\mathcal{Z}$ . That is,  $z_1 \geq z_2$  if  $z_1 - z_2 \in \mathcal{C}$ . It is not difficult to verify that this order relation defines a *partial order* on  $\mathcal{Z}$ , i.e., the following conditions hold for any  $z, z', z'' \in \mathcal{Z}$ : (i)  $z \geq z$ , (ii) if  $z \geq z'$  and  $z' \geq z''$ , then  $z \geq z''$  (transitivity), and (iii) if  $z \geq z'$  and  $z' \geq z$ , then  $z = z'$ . This partial order relation is also compatible with the algebraic operations, i.e., the following conditions hold: (iv) if  $z \geq z'$  and  $t \geq 0$ , then  $tz \geq tz'$ , and (v) if  $z' \geq z''$  and  $z \in \mathcal{Z}$ , then  $z' + z \geq z'' + z$ .

It is said that  $u \in \mathcal{Z}$  is the *least upper bound* (or *supremum*) of  $z, z' \in \mathcal{Z}$ , written  $u = z \vee z'$ , if  $u \geq z$  and  $u \geq z'$  and, moreover, if  $u' \geq z$  and  $u' \geq z'$  for some  $u' \in \mathcal{Z}$ , then  $u' \geq u$ . By the above property (iii) we have that if the least upper bound  $z \vee z'$  exists, then it is unique. It is said that the considered partial order induces a *lattice structure* on  $\mathcal{Z}$  if the least upper bound  $z \vee z'$  exists for any  $z, z' \in \mathcal{Z}$ . Denote  $z_+ := z \vee 0$ ,  $z_- := (-z) \vee 0$ , and  $|z| := z_+ \vee z_- = z \vee (-z)$ . It is said that Banach space  $\mathcal{Z}$  with lattice structure is a *Banach lattice* if  $z, z' \in \mathcal{Z}$  and  $|z| \geq |z'|$  implies  $\|z\| \geq \|z'\|$ .

<sup>74</sup>Recall that cone  $\mathcal{C}$  is said to be *pointed* if  $z \in \mathcal{C}$  and  $-z \in \mathcal{C}$  implies that  $z = 0$ .

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

$$\mathcal{L}_p^+(\Omega, \mathcal{F}, P) := \{z \in \mathcal{L}_p(\Omega, \mathcal{F}, P) : z(\omega) \geq 0 \text{ for a.e. } \omega \in \Omega\}. \quad (7.245)$$

This cone  $\mathcal{C}$  is closed, convex, and pointed. The corresponding partial order means that  $z \geq z'$  iff  $z(\omega) \geq z'(\omega)$  for a.e.  $\omega \in \Omega$ . It has a lattice structure with

$$(z \vee z')(\omega) = \max\{z(\omega), z'(\omega)\}$$

and  $|z|(\omega) = |z(\omega)|$ . Also, the property, “if  $z \geq z' \geq 0$ , then  $\|z\| \geq \|z'\|$ ,” 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 7.78 (Klee–Nachbin–Namioka).** *Let  $\mathcal{Z}$  be a Banach lattice and  $\ell : \mathcal{Z} \rightarrow \mathbb{R}$  be a linear functional. Suppose that  $\ell$  is positive, i.e.,  $\ell(z) \geq 0$  for any  $z \geq 0$ . Then  $\ell$  is continuous.*

**Proof.** We have that linear functional  $\ell$  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  $z \geq 0$ . Recall that  $z \geq 0$  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  $\bar{z} := \sum_{k=1}^{\infty} 2^{-k} z_k$ . 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  $\bar{z}$  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  $\ell$  that  $\ell(\sum_{k=m}^{\infty} 2^{-k} z_k) \geq 0$  for any  $m \in \mathbb{N}$ . Therefore, we have

$$\begin{aligned} \ell(\bar{z}) &= \ell\left(\sum_{k=1}^n 2^{-k} z_k\right) + \ell\left(\sum_{k=n+1}^{\infty} 2^{-k} z_k\right) \geq \ell\left(\sum_{k=1}^n 2^{-k} z_k\right) \\ &= \sum_{k=1}^n 2^{-k} \ell(z_k) \geq n, \end{aligned}$$

for any  $n \in \mathbb{N}$ . This gives a contradiction.

Now for any  $z \in \mathcal{Z}$  we have

$$|z| = z_+ \vee z_- \geq z_+ = |z_+|.$$

It follows that for  $v = |z|$  we have that  $\|v\| \geq \|z_+\|$ , and similarly  $\|v\| \geq \|z_-\|$ . Since  $z = z_+ - z_-$  and  $\ell(z_-) \geq 0$  by positivity of  $\ell$ , it follows that

$$\ell(z) = \ell(z_+) - \ell(z_-) \leq \ell(z_+) \leq c\|z_+\| \leq c\|z\|,$$

and similarly

$$-\ell(z) = -\ell(z_+) + \ell(z_-) \leq \ell(z_-) \leq c\|z_-\| \leq c\|z\|.$$

It follows that  $|\ell(z)| \leq c\|z\|$ , which completes the proof.  $\square$

Suppose that Banach space  $\mathcal{Z}$  has a lattice structure. It is said that a function  $f : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  is *monotone* if  $z \geq z'$  implies that  $f(z) \geq f(z')$ .

**Theorem 7.79.** *Let  $\mathcal{Z}$  be a Banach lattice and  $f : \mathcal{Z} \rightarrow \overline{\mathbb{R}}$  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 7.75, function  $f$  possesses an algebraic subgradient  $\ell$  at  $z_0$ . It follows from monotonicity of  $f$  that  $\ell$  is positive. Indeed, if  $\ell(h) < 0$  for some  $h \in \mathcal{C}$ , then it follows by (7.240) that

$$f(z_0 - h) \geq f(z_0) - \ell(h) > f(z_0),$$

which contradicts monotonicity of  $f$ . It follows by Theorem 7.78 that  $\ell$  is continuous, and hence  $\ell \in \partial f(z_0)$ . This shows that  $f$  is subdifferentiable at every point of  $\text{int}(\text{dom } f)$ . This, in turn, implies that  $f$  is lower semicontinuous on  $\text{int}(\text{dom } f)$  and hence by Proposition 7.72 is continuous on  $\text{int}(\text{dom } f)$ .  $\square$

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  $\mathcal{C} := \mathcal{L}_p^+(\Omega, \mathcal{F}, P)$ .

### Interchangeability Principle

Let  $(\Omega, \mathcal{F}, P)$  be a probability space. It is said that a linear space  $\mathfrak{M}$  of  $\mathcal{F}$ -measurable functions (mappings)  $\psi : \Omega \rightarrow \mathbb{R}^m$  is *decomposable* if for every  $\psi \in \mathfrak{M}$  and  $A \in \mathcal{F}$ , and every bounded and  $\mathcal{F}$ -measurable function  $\gamma : \Omega \rightarrow \mathbb{R}^m$ , 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. Proof of the following theorem can be found in [181, Theorem 14.60].

**Theorem 7.80.** *Let  $\mathfrak{M}$  be a decomposable space and  $f : \mathbb{R}^m \times \Omega \rightarrow \mathbb{R}$  be a random lower semicontinuous function. Then*

$$\mathbb{E} \left[ \inf_{x \in \mathbb{R}^m} f(x, \omega) \right] = \inf_{\chi \in \mathfrak{M}} \mathbb{E}[F_\chi], \quad (7.246)$$

where  $F_\chi(\omega) := f(\chi(\omega), \omega)$ , provided that the right-hand side of (7.246) is less than  $+\infty$ . Moreover, if the common value of both sides in (7.246) is not  $-\infty$ , then

$$\bar{\chi} \in \underset{\chi \in \mathfrak{M}}{\text{argmin}} \mathbb{E}[F_\chi] \text{ iff } \bar{\chi}(\omega) \in \underset{x \in \mathbb{R}^m}{\text{argmin}} f(x, \omega) \text{ for a.e. } \omega \in \Omega \text{ and } \bar{\chi} \in \mathfrak{M}. \quad (7.247)$$

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 6.37.

### Exercises

- 7.1. Show that function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is lower semicontinuous iff its epigraph  $\text{epi } f$  is a closed subset of  $\mathbb{R}^{n+1}$ .
- 7.2. Show that a function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  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  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  which is Gâteaux but not Fréchet differentiable.
- 7.4. Show that if  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  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  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  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  $s(\cdot)$  be the support function of a nonempty set  $C \subset \mathbb{R}^n$ . Show that the conjugate of  $s(\cdot)$  is the indicator function of the set  $\text{cl}(\text{conv}(C))$ .
- 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. 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  $\text{dom } \mathcal{G}$  is compact, then  $\mathcal{G}$  is upper semicontinuous.
- 7.9. Consider function  $F(x, \omega)$  used in Theorem 7.44. Show that if  $F(\cdot, \omega)$  is differentiable for a.e.  $\omega$ , then condition (A2) of that theorem is equivalent to the following condition: there exists a neighborhood  $V$  of  $x_0$  such that

$$\mathbb{E} \left[ \sup_{x \in V} \|\nabla_x F(x, \omega)\| \right] < \infty. \quad (7.248)$$

- 7.10. Show that if  $f(x) := \mathbb{E}|x - \xi|$ , then formula (7.121) holds. Conclude that  $f(\cdot)$  is differentiable at  $x_0 \in \mathbb{R}$  iff  $\Pr(\xi = x_0) = 0$ .
- 7.11. Verify equalities (7.143) and (7.144) and hence conclude (7.145).
- 7.12. Show that the estimate (7.199) of Theorem 7.65 still holds if the bound (7.198) in condition (C2) is replaced by

$$|F(x', \xi) - F(x, \xi)| \leq \kappa(\xi) \|x' - x\|^\gamma \quad (7.249)$$

for some constant  $\gamma > 0$ . Show how the estimate (7.217) of Theorem 7.67 should be corrected in that case.

## Chapter 8

# Bibliographical Remarks

## Chapter 1

The news vendor problem (sometimes called the newsboy problem), portfolio selection, and supply chain models are classical, and numerous papers have been written on each subject. It would be far beyond the scope of 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, and dynamic programming equations. We give below just a few basic references.

The news vendor problem is a classical model used in inventory management. Its origin is in the paper by Edgeworth [62]. In the stochastic setting, study of the news vendor problem started with the classical paper by Arrow, Harris, and Marchak [5]. The optimality of the basestock policy for the multistage inventory model was first proved in Clark and Scarf [37]. The worst-case distribution approach to the news vendor problem was initiated by Scarf [193], 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 multistage inventory models, see Zipkin [230].

Modern portfolio theory was introduced by Markowitz [125, 126]. The concept of utility function has a long history. Its origins go back as far as the work of Daniel Bernoulli (1738). The axiomatic approach to the expected utility theory was introduced by von Neumann and Morgenstern [221].

For an introduction to supply chain network design, see, e.g., Nagurney [132]. The material of section 1.5 is based on Santoso et al. [192].

For a thorough discussion of robust optimization we refer to the forthcoming book by Ben-Tal, El Ghaoui, and Nemirovski [15].

## Chapters 2 and 3

Stochastic programming with recourse originated in the works of Beale [14], Dantzig [41], and Tintner [215].

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 [99, 100], Walkup and Wets [223, 224], and Wets [226, 227]. Example 2.5 is discussed in Birge and Louveaux [19]. 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 [19], Kall and Wallace [102], and Wallace and Ziemba [225]. For a thorough analysis of simple recourse models, see Kall and Mayer [101].

Duality analysis of stochastic problems, and in particular dualization of the nonanticipativity constraints, was developed by Eisner and Olsen [64], Wets [228], and Rockafellar and Wets [179, 180]. (See also Rockafellar [176] and the references therein.)

Expected value of perfect information is a classical concept in decision theory (see Raiffa and Schlaifer [165] and Raiffa [164]). In stochastic programming this and related concepts were analyzed by Madansky [122], Spivey [210], Avriel and Williams [11], Dempster [46], and Huang, Vertinsky, and Ziemba [95].

Numerical methods for solving two- and multistage stochastic programming problems are extensively discussed in Birge and Louveaux [19], Ruszczyński [186], and Kall and Mayer [101], 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, pseudorandom sequences, lower and upper bounding trees, and moment matching. The reader is referred to Kall and Mayer [101], Heitsch and Römisches [82], Hochreiter and Pflug [91], Casey and Sen [31], Pennanen [145], Dupačová, Growe-Kuska, and Römisches [59], and the references therein.

An extensive stochastic programming bibliography can be found at the website <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 [32], Miller and Wagner [129], and Prékopa [154]. Problems with integrated chance constraints are considered in [81]. Models with stochastic dominance constraints were introduced and analyzed by Dentcheva and Ruszczyński in [52, 54, 55]. The notion of stochastic ordering or stochastic dominance of first order has been introduced in statistics in Mann and Whitney [124] and Lehmann [116] and further applied and developed in economics (see Quirk and Saposnik [163], Fishburn [67], and Hadar and Russell [80].)

An essential contribution to the theory and solutions of problems with chance constraints was the theory of  $\alpha$ -concave measures and functions. In Prékopa [155, 156] the concept of logarithmic concave measures was introduced and studied. This notion was generalized to  $\alpha$ -concave measures and functions in Borell [23, 24], Brascamp and Lieb [26], and Rinott [168] and further analyzed in Tamm [214] and Norkin [141]. Approximations of the probability function by Steklov–Sobolev transformation was suggested by Norkin in [139]. Differentiability properties of probability functions were studied in Uryasev [216, 217], Kibzun and Tretyakov [104], Kibzun and Uryasev [105], and Raik [166]. The first definition of  $\alpha$ -concave discrete multivariate distributions was introduced in Barndorff-Nielsen [13]. The generalized definition of  $\alpha$ -concave functions on a set, which we have adopted here, was introduced in Dentcheva, Prékopa, and Ruszczyński [49]. It facilitates the development of optimality and duality theory of probabilistic optimization. Its consequences

for probabilistic optimization were explored in Dentcheva, Prékopa, and Ruszczyński [50]. The notion of  $p$ -efficient points was first introduced in Prékopa [157]. A similar concept was used in Sen [195]. The concept was studied and applied in the context of discrete distributions and linear problems in the papers of Dentcheva, Prékopa, and Ruszczyński [49, 50] and Prékopa, Vízvári, Badics [160] and in the context of general distributions in Dentcheva, Lai, and Ruszczyński [48]. Optimization problems with probabilistic set-covering constraint were investigated in Beraldi and Ruszczyński [16, 17], where efficient enumeration procedures of  $p$ -efficient points of 0–1 variable are employed. There is a wealth of research on estimating probabilities of events. We refer to Boros and Prékopa [25], Bukszár [27], Bukszár and Prékopa [28], Dentcheva, Prékopa, and Ruszczyński [50], Prékopa [158], and Szántai [212], where probability bounds are used in the context of chance constraints.

Statistical approximations of probabilistically constrained problems were analyzed in Sainetti [191], Kankova [103], Deák [43], and Gröwe [77]. Stability of models with probabilistic constraints was addressed in Dentcheva [47], Henrion [84, 83], and Henrion and Römisich [183, 85]. Nonlinear probabilistic problems were investigated in Dentcheva, Lai, and Ruszczyński [48], where optimality conditions are established.

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), 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 [57], Pflug [148], and Pflug and Römisich [149]). The concept of stochastic dominance plays a fundamental role in economics and statistics. We refer to Mosler and Scarsini [131], Shaked and Shanthikumar [196], and Szekli [213] 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 [222], consistency properties of the ML method and M-estimators were studied in numerous publications. The epi-convergence approach to studying consistency of statistical estimators was discussed in Dupačová and Wets [60]. In the context of stochastic programming, consistency of SAA estimators was also investigated by tools of epi-convergence analysis in King and Wets [108] and Robinson [173].

Proposition 5.6 appeared in Norkin, Pflug, and Ruszczyński [140] and Mak, Morton, and Wood [123]. Theorems 5.7, 5.11, and 5.10 are taken from Shapiro [198] and [204], respectively. The approach to second order asymptotics, discussed in section 5.1.3, is based on Dentcheva and Römisich [51] and Shapiro [199]. 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 [96]. An extension of the SAA method to

stochastic generalized equations is a natural one. Stochastic variational inequalities were discussed by Gürkan, Özge, and Robinson [79]. Proposition 5.14 and Theorem 5.15 are similar to the results obtained in [79, Theorems 1 and 2]. Asymptotics of SAA estimators of optimal solutions of stochastic programs were discussed in King and Rockafellar [107] and Shapiro [197].

The idea of using Monte Carlo sampling for solving stochastic optimization problems of the form (5.1) certainly is not new. A variety of sampling-based optimization techniques have been suggested in the literature. It is beyond the scope of this chapter to give a comprehensive survey of these methods, but 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 [90], Glasserman [75], Kushner and Clark [112], and Nevelson and Hasminskii [137], respectively. For an application of this approach to optimization of queueing systems see Chong and Ramadge [36] and L'Ecuyer and Glynn [115], for example. Closely related to this approach is the stochastic quasi-gradient method (see Ermoliev [65]).

Another class of methods uses sample average estimates of the values of the objective function, and maybe 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 [97] and the stochastic decomposition method of Higle and Sen [88].

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 an 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  $\xi^1, \dots, \xi^N$ , the obtained optimization problem can be considered as a stochastic program with the associated scenarios  $\xi^1, \dots, \xi^N$ , 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, a quite well-developed statistical inference of the SAA method is available. 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 [184], [185]) and sample-path optimization (Plambeck et al. [151] and Robinson [173]), 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 [72] 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, 120, 123, 220].

The complexity analysis of the SAA method, discussed in section 5.3, is motivated by the following observations. Suppose for the moment that components  $\xi_i$ ,  $i = 1, \dots, d$ , 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  $\xi_i$ . Then the resulting number of scenarios is  $K = r^d$ , i.e., it grows exponentially with an increase of the number of random parameters. Already with, say,  $r = 4$  and  $d = 20$  we 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  $x \in X$ , much less to solve the corresponding optimization problem.<sup>75</sup> And, indeed, it was shown in Dyer and Stougie [61] that under the assumption that the stochastic parameters are independently distributed, two-stage linear stochastic programming problems are  $\sharp P$ -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 inexact modeling, distribution approximations, etc., could be far bigger. In some situations the randomization approach based on Monte Carlo sampling techniques allows one to solve stochastic programs with reasonable accuracy and a reasonable computational effort.

The material of section 5.3.1 is based on Kleywegt, Shapiro, and Homem-De-Mello [109]. The extension of that analysis to general feasible sets, given in section 5.3.2, was discussed in Shapiro [200, 202, 205] and Shapiro and Nemirovski [206]. The material of section 5.3.3 is based on Shapiro and Homem-de-Mello [208], 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 see Niederreiter [138]. For applications of quasi-Monte Carlo techniques to stochastic programming see, e.g., Koivu [110], Homem-de-Mello [94], and Pennanen and Koivu [146].

For a discussion of variance reduction techniques in Monte Carlo sampling we refer to Fishman [68] 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 [185], Dantzig and Infanger [42], Higle [86] and Bailey, Jensen, and Morton [12], for example.

The statistical bounds of section 5.6.1 were suggested in Norkin, Pflug, and Ruszczyński [140] and developed in Mak, Morton, and Wood [123]. The common random

<sup>75</sup>Of 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.

numbers estimator  $\widehat{\text{gap}}_{N,M}(\bar{x})$  of the optimality gap was introduced in [123]. The KKT statistical test, discussed in section 5.6.2, was developed in Shapiro and Homem-de-Mello [207], so that the material of that section is based on [207]. See also Higle and Sen [87].

The estimate of the sample size derived in Theorem 5.32 is due to Campi and Garatti [30]. This result builds on a previous work of Calafiore and Campi [29], 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 [134]. For some numerical experiments with these bounds see Luedtke and Ahmed [121].

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 [201]. Discussion of the portfolio selection (Example 5.34) is based on Blomvall and Shapiro [21]. Complexity of the SAA approach to multistage programming was discussed in Shapiro and Nemirovski [206] and Shapiro [203].

Section 5.9 is based on Nemirovski et al. [133]. The origins of the stochastic approximation algorithms go back to the pioneering paper by Robbins and Monro [169]. For a thorough discussion of the asymptotic theory of the SA method, we refer to Kushner and Clark [112] and Nevelson and Hasminskii [137]. The robust SA approach was developed in Polyak [152] and Polyak and Juditsky [153]. The main ingredients of Polyak's scheme (long steps and averaging) were, in a different form, proposed in Nemirovski and Yudin [135].

## Chapter 6

Foundations of the expected utility theory were developed in von Neumann and Morgenstern [221]. The dual utility theory was developed in Quiggin [161, 162] and Yaari [229].

The mean-variance model was introduced and analyzed in Markowitz [125, 126, 127]. Deviations and semideviations in mean-risk analysis were analyzed in Kijima and Ohnishi [106], Konno [111], Ogryczak and Ruszczyński [142, 143], and Ruszczyński and Vanderbei [190]. Weighted deviations from quantiles, relations to stochastic dominance, and Lorenz curves are discussed in Ogryczak and Ruszczyński [144]. For Conditional (Average) Value-at-Risk see Acerbi and Tasche [1], Rockafellar and Uryasev [177], and Pflug [148]. A general class of convex approximations of chance constraints was developed in Nemirovski and Shapiro [134].

The theory of coherent measures of risk was initiated in Artzner et al. [8] and further developed, inter alia, by Delbaen [44], Föllmer and Schied [69], Leitner [117], and Rockafellar, Uryasev, and Zabarankin [178]. Our presentation is based on Ruszczyński and Shapiro [187, 189]. The Kusuoka representation of law invariant coherent risk measures (Theorem 6.24) was derived in [113] 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 [149]. Theory of consistency with stochastic orders was initiated in [142] and developed in [143, 144]. An alternative approach to asymptotic analysis of law invariant coherent risk measures (see section 6.5.3), was developed in Pflug and Wozabal [147] based on Kusuoka representation. Application to portfolio optimization was discussed in Miller and Ruszczyński [130].

The theory of conditional risk mappings was developed in Riedel [167] and Ruszczyński and Shapiro [187, 188]. For the general theory of dynamic measures of



risk, see Artzner et al. [9], Cheridito, Delbaen, and Kupper [33, 34], Frittelli and Rosazza Gianin [71, 70], Eichhorn and Römisch [63], and Pflug and Römisch [149]. Our inventory example is based on Ahmed, Cakmak, and Shapiro [2].

## Chapter 7

There are many monographs where concepts of directional differentiability are discussed in detail; see, e.g., [22]. A thorough discussion of the Clarke generalized gradient and regularity in the sense of Clarke can be found in Clarke [38]. Classical references on (finite dimensional) convex analysis are books by Rockafellar [174] and Hiriart-Urruty and Lemaréchal [89]. For a proof of the Fenchel–Moreau theorem (in an infinite dimensional setting) see, e.g., [175]. For a development of conjugate duality (in an infinite dimensional setting) we refer to Rockafellar [175]. Theorem 7.11 (Hoffman’s lemma) appeared in [93].

Theorem 7.21 appeared in Danskin [40]. Theorem 7.22 goes back to Levin [118] and Valadier [218] (see also Ioffe and Tihomirov [98, page 213]). For a general discussion of second order optimality conditions and perturbation analysis of optimization problems we refer to Bonnans and Shapiro [22] and references therein. Theorem 7.24 is an adaptation of a result going back to Gol’shtein [76]. For a thorough discussion of epiconvergence we refer to Rockafellar and Wets [181]. Theorem 7.27 is taken from [181, Theorem 7.17].

There are many books on probability theory. Of course, it is 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 [18]. Theorem 7.32 appeared in Rogosinski [182]. A thorough discussion of measurable multifunctions and random lower semicontinuous functions can be found in Rockafellar and Wets [181, Chapter 14], to which the interested reader is referred for further reading. For a proof of the Aumann and Lyapunov theorems (Theorems 7.40 and 7.41) see, e.g., [98, section 8.2].

Theorem 7.47 originated in Strassen [211], where the interchangeability of the sub-differential and integral operators was shown in the case when the expectation function is continuous. The present formulation of Theorem 7.47 is taken from [98, Theorem 4, page 351].

Uniform Laws of Large Numbers (LLN) take their origin in the Glivenko–Cantelli theorem. For a further discussion of the uniform LLN we refer to van der Vaart and Wellner [219]. Epi-convergence LLN, formulated in Theorem 7.51, is due to Artstein and Wets [7]. The uniform convergence w.p. 1 of Clarke generalized gradients, specified in part (c) of Theorem 7.52, was obtained in [197]. The LLN for random sets (Theorem 7.53) appeared in Artstein and Vitale [6]. The uniform convergence of  $\varepsilon$ -subdifferentials (Theorem 7.55) was derived in [209].

The finite dimensional Delta method is well known and routinely used in theoretical statistics. The infinite dimensional version (Theorem 7.59) goes back to Grübel [78], Gill [74], and King [107]. The tangential version (Theorem 7.61) appeared in [198].

There is a large literature on large deviations theory (see, e.g., a book by Dembo and Zeitouni [45]). The Hoeffding inequality appeared in [92] and the Chernoff inequality in [35]. Theorem 7.68, about interchangeability of Clarke generalized gradient and integral operators, can be derived by using the interchangeability formula (7.117) for directional derivatives, Strassen’s Theorem 7.47, and the fact that in the Clarke-regular case the directional derivative is the support function of the corresponding Clarke generalized gradient (see [38] for details).

A classical reference for functional analysis is Dunford and Schwartz [58]. The concept of algebraic subgradient and Theorem 7.78 are taken from Levin [119]. (Unfortunately, this excellent book was not translated from Russian.) Theorem 7.79 is from Ruszczyński and Shapiro [189]. The interchangeability principle (Theorem 7.80) is taken from [181, Theorem 14.60]. Similar results can be found in [98, Proposition 2, page 340] and [119, Theorem 0.9].

## Bibliography

- [1] C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance*, 26:1491–1507, 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 H. Ku. Coherent multiperiod risk adjusted values and Bellman’s 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. *Communications in 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] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, Princeton, NJ, 2009.
- [16] P. Beraldi and A. Ruszczyński. The probabilistic set covering problem. *Operations Research*, 50:956–967, 1999.
- [17] 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.
- [18] P. Billingsley. *Probability and Measure*. John Wiley & Sons, New York, 1995.
- [19] J.R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer-Verlag, New York, 1997.
- [20] G. Birkhoff. Tres observaciones sobre el algebra lineal. *Univ. Nac. Tucumán Rev. Ser. A*, 5:147–151, 1946.
- [21] J. Blomvall and A. Shapiro. Solving multistage asset investment problems by Monte Carlo based optimization. *Mathematical Programming*, 108:571–595, 2007.
- [22] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer-Verlag, New York, 2000.
- [23] C. Borell. Convex measures on locally convex spaces. *Ark. Mat.*, 12:239–252, 1974.
- [24] C. Borell. Convex set functions in d-space. *Periodica Mathematica Hungarica*, 6:111–136, 1975.
- [25] 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. Oper. Res.*, 14:317–342, 1989.
- [26] 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.
- [27] J. Bukszár. Probability bounds with multitrees. *Adv. Appl. Probab.*, 33:437–452, 2001.
- [28] J. Bukszár and A. Prékopa. Probability bounds with cherry trees. *Mathematics of Operations Research*, 26:174–192, 2001.
- [29] G. Calafiore and M.C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51:742–753, 2006.
- [30] M.C. Campi and S. Garatti. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM J. Optimization*, 19:1211–1230, 2008.

- [31] M.S. Casey and S. Sen. The scenario generation algorithm for multistage stochastic linear programming. *Mathematics of Operations Research*, 30:615–631, 2005.
- [32] 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.
- [33] 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.
- [34] P. Cheridito, F. Delbaen, and M. Kupper. Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability*, 11:57–106, 2006.
- [35] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum observations. *Annals Math. Statistics*, 23:493–507, 1952.
- [36] 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.
- [37] A. Clark and H. Scarf. Optimal policies for a multi-echelon inventory problem. *Management Science*, 6:475–490, 1960.
- [38] F.H. Clarke. *Optimization and nonsmooth analysis*. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons, New York, 1983.
- [39] R. Cranley and T.N.L. Patterson. Randomization of number theoretic methods for multiple integration. *SIAM J. Numer. Anal.*, 13:904–914, 1976.
- [40] J.M. Danskin. *The Theory of Max-Min and Its Applications to Weapons Allocation Problems*. Springer-Verlag, New York, 1967.
- [41] G.B. Dantzig. Linear programming under uncertainty. *Management Science*, 1:197–206, 1955.
- [42] G.B. Dantzig and G. Infanger. Large-scale stochastic linear programs—importance sampling and Benders decomposition. In *Computational and Applied Mathematics I (Dublin, 1991)*, North-Holland, Amsterdam, 1992, 111–120.
- [43] I. Deák. Linear regression estimators for multinormal distributions in optimization of stochastic programming problems. *European Journal of Operational Research*, 111:555–568, 1998.
- [44] P. Delbaen. Coherent risk measures on general probability spaces. In *Essays in Honour of Dieter Sondermann*. Springer-Verlag, Berlin, 2002, 1–37.
- [45] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer-Verlag, New York, 1998.

- [46] 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, Lecture Notes in Information and Control* 36, Berkeley, CA, 1982, 25–40.
- [47] D. Dentcheva. Regular Castaing representations of multifunctions with applications to stochastic programming. *SIAM J. Optimization*, 10:732–749, 2000.
- [48] D. Dentcheva, B. Lai, and A. Ruszczyński. Dual methods for probabilistic optimization. *Mathematical Methods of Operations Research*, 60:331–346, 2004.
- [49] D. Dentcheva, A. Prékopa, and A. Ruszczyński. Concavity and efficient points of discrete distributions in probabilistic programming. *Mathematical Programming*, 89:55–77, 2000.
- [50] D. Dentcheva, A. Prékopa, and A. Ruszczyński. Bounds for probabilistic integer programming problems. *Discrete Applied Mathematics*, 124:55–65, 2002.
- [51] D. Dentcheva and W. Römisch. Differential stability of two-stage stochastic programs. *SIAM J. Optimization*, 11:87–112, 2001.
- [52] D. Dentcheva and A. Ruszczyński. Optimization with stochastic dominance constraints. *SIAM J. Optimization*, 14:548–566, 2003.
- [53] D. Dentcheva and A. Ruszczyński. Convexification of stochastic ordering. *Comptes Rendus de l'Academie Bulgare des Sciences*, 57:11–16, 2004.
- [54] D. Dentcheva and A. Ruszczyński. Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints. *Mathematical Programming*, 99:329–350, 2004.
- [55] D. Dentcheva and A. Ruszczyński. Semi-infinite probabilistic optimization: First order stochastic dominance constraints. *Optimization*, 53:583–601, 2004.
- [56] D. Dentcheva and A. Ruszczyński. Portfolio optimization under stochastic dominance constraints. *Journal of Banking and Finance*, 30:433–451, 2006.
- [57] K. Dowd. *Beyond Value at Risk. The Science of Risk Management*. Wiley & Sons, New York, 1997.
- [58] N. Dunford and J. Schwartz. *Linear Operators, Vol I*. Interscience, New York, 1958.
- [59] J. Dupacova, N. Growe-Kuska, and W. Römisch. Scenario reduction in stochastic programming: An approach using probability metrics. *Mathematical Programming*, 95:493–511, 2003.
- [60] J. Dupačová and R.J.B. Wets. Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Annals of Statistics*, 16:1517–1549, 1988.
- [61] M. Dyer and L. Stougie. Computational complexity of stochastic programming problems. *Mathematical Programming*, 106:423–432, 2006.

- [62] F. Edgeworth. The mathematical theory of banking. *Royal Statistical Society*, 51:113–127, 1888.
- [63] A. Eichhorn and W. Römisch. Polyhedral risk measures in stochastic programming. *SIAM J. Optimization*, 16:69–95, 2005.
- [64] 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.
- [65] Y. Ermoliev. Stochastic quasi-gradient methods and their application to systems optimization. *Stochastics*, 4:1–37, 1983.
- [66] D. Filipović and G. Svindland. The canonical model space for law-invariant convex risk measures is  $L^1$ . *Mathematical Finance*, to appear.
- [67] P.C. Fishburn. *Utility Theory for Decision Making*. John Wiley & Sons, New York, 1970.
- [68] G.S. Fishman. *Monte Carlo, Concepts, Algorithms and Applications*. Springer-Verlag, New York, 1999.
- [69] H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6:429–447, 2002.
- [70] M. Frittelli and G. Scandolo. Risk measures and capital requirements for processes. *Mathematical Finance*, 16:589–612, 2006.
- [71] M. Frittelli and E. Rosazza Gianin. Dynamic convex risk measures. In G. Szegö, editor, *Risk Measures for the 21st Century*, John Wiley & Sons, Chichester, UK, 2005, pages 227–248.
- [72] 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.
- [73] J.E. Littlewood, G.H. Hardy, and G. Pólya. *Inequalities*. Cambridge University Press, Cambridge, UK, 1934.
- [74] R.D. Gill. Non-and-semiparametric maximum likelihood estimators and the von Mises method (Part I). *Scandinavian Journal of Statistics*, 16:97–124, 1989.
- [75] P. Glasserman. *Gradient Estimation via Perturbation Analysis*. Kluwer Academic Publishers, Norwell, MA, 1991.
- [76] E.G. Gol’shtein. *Theory of Convex Programming*. Translations of Mathematical Monographs, AMS, Providence, RI, 1972.
- [77] N. Gröwe. Estimated stochastic programs with chance constraint. *European Journal of Operations Research*, 101:285–305, 1997.
- [78] R. Grübel. The length of the short. *Annals of Statistics*, 16:619–628, 1988.
- [79] G. Gurkan, A.Y. Ozge, and S.M. Robinson. Sample-path solution of stochastic variational inequalities. *Mathematical Programming*, 21:313–333, 1999.

- [80] J. Hadar and W. Russell. Rules for ordering uncertain prospects. *American Economic Review*, 59:25–34, 1969.
- [81] W. K. Klein Haneveld. *Duality in Stochastic Linear and Dynamic Programming, Lecture Notes in Economics and Mathematical Systems* 274. Springer-Verlag, New York, 1986.
- [82] H. Heitsch and W. Römisch. Scenario tree modeling for multistage stochastic programs. *Mathematical Programming*, 118:371–406, 2009.
- [83] R. Henrion. Perturbation analysis of chance-constrained programs under variation of all constraint data. In K. Marti et al., editors, *Dynamic Stochastic Optimization, Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag, Heidelberg, pages 257–274.
- [84] R. Henrion. On the connectedness of probabilistic constraint sets. *Journal of Optimization Theory and Applications*, 112:657–663, 2002.
- [85] R. Henrion and W. Römisch. Metric regularity and quantitative stability in stochastic programs with probabilistic constraints. *Mathematical Programming*, 84:55–88, 1998.
- [86] J.L. Higle. Variance reduction and objective function evaluation in stochastic linear programs. *INFORMS Journal on Computing*, 10(2):236–247, 1998.
- [87] 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.
- [88] J.L. Higle and S. Sen. *Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.
- [89] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I and II*. Springer-Verlag, New York, 1993.
- [90] Y.C. Ho and X.R. Cao. *Perturbation Analysis of Discrete Event Dynamic Systems*. Kluwer Academic Publishers, Norwell, MA, 1991.
- [91] R. Hochreiter and G. Ch. Pflug. Financial scenario generation for stochastic multi-stage decision processes as facility location problems. *Annals of Operations Research*, 152:257–272, 2007.
- [92] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.
- [93] 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.
- [94] T. Homem-de-Mello. On rates of convergence for stochastic optimization problems under non-independent and identically distributed sampling. *SIAM J. Optimization*, 19:524–551, 2008.



- [95] C. C. Huang, I. Vertinsky, and W. T. Ziemba. Sharp bounds on the value of perfect information. *Operations Research*, 25:128–139, 1977.
- [96] P.J. Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability, Vol. I*, University of California Press, Berkeley, CA, 1967, 221–233.
- [97] G. Infanger. *Planning Under Uncertainty: Solving Large Scale Stochastic Linear Programs*. Boyd and Fraser, Danvers, MA, 1994.
- [98] A.D. Ioffe and V.M. Tihomirov. *Theory of Extremal Problems*. North-Holland, Amsterdam, 1979.
- [99] P. Kall. Qualitative aussagen zu einigen problemen der stochastischen programmierung. *Z. Warscheinlichkeitstheorie u. Vervandte Gebiete*, 6:246–272, 1966.
- [100] P. Kall. *Stochastic Linear Programming*. Springer-Verlag, Berlin, 1976.
- [101] P. Kall and J. Mayer. *Stochastic Linear Programming*. Springer, New York, 2005.
- [102] P. Kall and S.W. Wallace. *Stochastic Programming*. John Wiley & Sons, Chichester, UK, 1994.
- [103] V. Kankova. On the convergence rate of empirical estimates in chance constrained stochastic programming. *Kybernetika (Prague)*, 26:449–461, 1990.
- [104] A.I. Kibzun and G.L. Tretyakov. Differentiability of the probability function. *Doklady Akademii Nauk*, 354:159–161, 1997. Russian.
- [105] A.I. Kibzun and S. Uryasev. Differentiability of probability function. *Stochastic Analysis and Applications*, 16:1101–1128, 1998.
- [106] 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.
- [107] 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.
- [108] A.J. King and R.J.-B. Wets. Epi-consistency of convex stochastic programs. *Stochastics Stochastics Rep.*, 34(1–2):83–92, 1991.
- [109] A.J. Kleywegt, A. Shapiro, and T. Homem-De-Mello. The sample average approximation method for stochastic discrete optimization. *SIAM J. Optimization*, 12:479–502, 2001.
- [110] M. Koivu. Variance reduction in sample approximations of stochastic programs. *Mathematical Programming*, 103:463–485, 2005.
- [111] H. Konno and H. Yamazaki. Mean–absolute deviation portfolio optimization model and its application to Tokyo stock market. *Management Science*, 37:519–531, 1991.

- [112] H.J. Kushner and D.S. Clark. *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer-Verlag, Berlin, 1978.
- [113] S. Kusuoka. On law-invariant coherent risk measures. In S. Kusuoka and T. Maruyama, editors, *Advances in Mathematical Economics, Vol. 3*, Springer, Tokyo, 2001, pages 83–95.
- [114] G. Lan, A. Nemirovski, and A. Shapiro. Validation analysis of robust stochastic approximation method. E-print available at <http://www.optimization-online.org>, 2008.
- [115] 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.
- [116] E. Lehmann. Ordered families of distributions. *Annals of Mathematical Statistics*, 26:399–419, 1955.
- [117] J. Leitner. A short note on second-order stochastic dominance preserving coherent risk measures. *Mathematical Finance*, 15:649–651, 2005.
- [118] 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. Russian.
- [119] V.L. Levin. *Convex Analysis in Spaces of Measurable Functions and Its Applications in Economics*. Nauka, Moscow, 1985. Russian.
- [120] J. Linderoth, A. Shapiro, and S. Wright. The empirical behavior of sampling methods for stochastic programming. *Annals of Operations Research*, 142:215–241, 2006.
- [121] J. Luedtke and S. Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM J. Optimization*, 19:674–699, 2008.
- [122] A. Madansky. Inequalities for stochastic linear programming problems. *Management Science*, 6:197–204, 1960.
- [123] 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.
- [124] 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.
- [125] H. M. Markowitz. Portfolio selection. *Journal of Finance*, 7:77–91, 1952.
- [126] H. M. Markowitz. *Portfolio Selection*. Wiley, New York, 1959.
- [127] H. M. Markowitz. *Mean–Variance Analysis in Portfolio Choice and Capital Markets*. Blackwell, Oxford, UK, 1987.
- [128] M. Meyer and S. Reisner. Characterizations of affinely-rotation-invariant log-concave measures by section-centroid location. In *Geometric Aspects of Functional Analysis, Lecture Notes in Mathematics* 1469, Springer-Verlag, Berlin, 1989–90, pages 145–152.

- [129] L.B. Miller and H. Wagner. Chance-constrained programming with joint constraints. *Operations Research*, 13:930–945, 1965.
- [130] 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.
- [131] K. Mosler and M. Scarsini. *Stochastic Orders and Decision Under Risk*. Institute of Mathematical Statistics, Hayward, CA, 1991.
- [132] A. Nagurney. *Supply Chain Network Economics: Dynamics of Prices, Flows, and Profits*. Edward Elgar Publishing, 2006.
- [133] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM J. Optimization*, 19:1574–1609, 2009.
- [134] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM J. Optimization*, 17:969–996, 2006.
- [135] A. Nemirovski and D. Yudin. On Cezari’s convergence of the steepest descent method for approximating saddle point of convex-concave functions. *Soviet Math. Dokl.*, 19: 1978.
- [136] A. Nemirovski and D. Yudin. *Problem Complexity and Method Efficiency in Optimization*. John Wiley, New York, 1983.
- [137] M.B. Nevelson and R.Z. Hasminskii. *Stochastic Approximation and Recursive Estimation*. American Mathematical Society Translations of Mathematical Monographs 47, AMS, Providence, RI, 1976.
- [138] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia, 1992.
- [139] V.I. Norkin. *The Analysis and Optimization of Probability Functions*. IIASA Working Paper, WP-93-6, Laxenburg (Austria), 1993.
- [140] V.I. Norkin, G.Ch. Pflug, and A. Ruszczyński. A branch and bound method for stochastic global optimization. *Mathematical Programming*, 83:425–450, 1998.
- [141] V.I. Norkin and N.V. Roenko.  $\alpha$ -Concave functions and measures and their applications. *Kibernet. Sistem. Anal.*, 189:77–88, 1991. Russian. Translation in *Cybernet. Systems Anal.*, 27:860–869, 1991.
- [142] 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.
- [143] W. Ogryczak and A. Ruszczyński. On consistency of stochastic dominance and mean–semideviation models. *Mathematical Programming*, 89:217–232, 2001.
- [144] W. Ogryczak and A. Ruszczyński. Dual stochastic dominance and related mean-risk models. *SIAM J. Optimization*, 13:60–78, 2002.

- [145] T. Pennanen. Epi-convergent discretizations of multistage stochastic programs. *Mathematics of Operations Research*, 30:245–256, 2005.
- [146] T. Pennanen and M. Koivu. Epi-convergent discretizations of stochastic programs via integration quadratures. *Numerische Mathematik*, 100:141–163, 2005.
- [147] G.Ch. Pflug and N. Wozabal. Asymptotic distribution of law-invariant risk functionals. *Finance and Stochastics*, to appear.
- [148] G.Ch. Pflug. Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic Constrained Optimization—Methodology and Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000, 272–281.
- [149] G.Ch. Pflug and W. Römisch. *Modeling, Measuring and Managing Risk*. World Scientific, Singapore, 2007.
- [150] R.R. Phelps. *Convex functions, monotone operators, and differentiability, Lecture Notes in Mathematics* 1364. Springer-Verlag, Berlin, 1989.
- [151] 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.
- [152] B.T. Polyak. New stochastic approximation type procedures. *Automat. i Telemekh.*, 7:98–107, 1990.
- [153] B.T. Polyak and A.B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM J. Control and Optimization*, 30:838–855, 1992.
- [154] A. Prékopa. On probabilistic constrained programming. In *Proceedings of the Princeton Symposium on Mathematical Programming*. Princeton University Press, Princeton, NJ, 1970, 113–138.
- [155] A. Prékopa. Logarithmic concave measures with applications to stochastic programming. *Acta Scientiarum Mathematicarum (Szeged)*, 32:301–316, 1971.
- [156] A. Prékopa. On logarithmic concave measures and functions. *Acta Scientiarum Mathematicarum (Szeged)*, 34:335–343, 1973.
- [157] A. Prékopa. Dual method for the solution of a one-stage stochastic programming problem with random rhs obeying a discrete probability distribution. *ZOR-Methods and Models of Operations Research*, 34:441–461, 1990.
- [158] A. Prékopa. Sharp bound on probabilities using linear programming. *Operations Research*, 38:227–239, 1990.
- [159] A. Prékopa. *Stochastic Programming*. Kluwer Academic Publishers, Boston, 1995.
- [160] A. Prékopa, B. Vízvári, and T. Badics. Programming under probabilistic constraint with discrete random variable. In L. Grandinetti et al., editors, *New Trends in Mathematical Programming*. Kluwer, Boston, 2003, 235–255.

- [161] J. Quiggin. A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3:225–243, 1982.
- [162] J. Quiggin. *Generalized Expected Utility Theory—The Rank-Dependent Expected Utility Model*. Kluwer, Dordrecht, The Netherlands, 1993.
- [163] J.P. Quirk and R. Saposnik. Admissibility and measurable utility functions. *Review of Economic Studies*, 29:140–146, 1962.
- [164] H. Raiffa. *Decision Analysis*. Addison–Wesley, Reading, MA, 1968.
- [165] H. Raiffa and R. Schlaifer. *Applied Statistical Decision Theory. Studies in Managerial Economics*. Harvard University, Cambridge, MA, 1961.
- [166] 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 Akadeemia Toimetised. Füüsika-Matemaatika*, 24:860–869, 1975. Russian.
- [167] F. Riedel. Dynamic coherent risk measures. *Stochastic Processes and Their Applications*, 112:185–200, 2004.
- [168] Y. Rinott. On convexity of measures. *Annals of Probability*, 4:1020–1026, 1976.
- [169] H. Robbins and S. Monro. A stochastic approximation method. *Annals of Math. Stat.*, 22:400–407, 1951.
- [170] S.M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5:43–62, 1980.
- [171] S.M. Robinson. Generalized equations and their solutions, Part II: Applications to nonlinear programming. *Mathematical Programming Study*, 19:200–221, 1982.
- [172] S.M. Robinson. Normal maps induced by linear transformations. *Mathematics of Operations Research*, 17:691–714, 1992.
- [173] S.M. Robinson. Analysis of sample-path optimization. *Mathematics of Operations Research*, 21:513–528, 1996.
- [174] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.
- [175] R.T. Rockafellar. *Conjugate Duality and Optimization. CBMS-NSF Regional Conference Series in Applied Mathematics* 16, SIAM, Philadelphia, 1974.
- [176] R.T. Rockafellar. Duality and optimality in multistage stochastic programming. *Ann. Oper. Res.*, 85:1–19, 1999.
- [177] R.T. Rockafellar and S. Uryasev. Optimization of conditional value at risk. *Journal of Risk*, 2:21–41, 2000.
- [178] R.T. Rockafellar, S. Uryasev, and M. Zabarankin. Generalized deviations in risk analysis. *Finance and Stochastics*, 10:51–74, 2006.

- [179] R.T. Rockafellar and R.J.-B. Wets. Stochastic convex programming: Basic duality. *Pacific J. Math.*, 62(1):173–195, 1976.
- [180] 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.
- [181] R.T. Rockafellar and R.J.-B. Wets. *Variational Analysis*. Springer, Berlin, 1998.
- [182] W.W. Rogosinski. Moments of non-negative mass. *Proc. Roy. Soc. London Ser. A*, 245:1–27, 1958.
- [183] W. Römisch. Stability of stochastic programming problems. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming, Handbooks in Operations Research and Management Science* 10. Elsevier, Amsterdam, 2003, 483–554.
- [184] 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.
- [185] R.Y. Rubinstein and A. Shapiro. *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*. John Wiley & Sons, Chichester, UK, 1993.
- [186] A. Ruszczyński. Decomposition methods. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming, Handbooks in Operations Research and Management Science* 10. Elsevier, Amsterdam, 2003, 141–211.
- [187] A. Ruszczyński and A. Shapiro. Optimization of risk measures. In G. Calafiore and F. Dabbene, editors, *Probabilistic and Randomized Methods for Design under Uncertainty*. Springer-Verlag, London, 2005, 117–158.
- [188] A. Ruszczyński and A. Shapiro. Conditional risk mappings. *Mathematics of Operations Research*, 31:544–561, 2006.
- [189] A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Mathematics of Operations Research*, 31:433–452, 2006.
- [190] A. Ruszczyński and R. Vanderbei. Frontiers of stochastically nondominated portfolios. *Econometrica*, 71:1287–1297, 2003.
- [191] G. Salinetti. Approximations for chance constrained programming problems. *Stochastics*, 10:157–169, 1983.
- [192] 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.
- [193] H. Scarf. A min-max solution of an inventory problem. In *Studies in the Mathematical Theory of Inventory and Production*. Stanford University Press, Stanford, CA, 1958, 201–209.

- [194] L. Schwartz. *Analyse Mathématique*, Volume I, Mir, Moscow, 1967; Volume II, Hermann, Paris, 1972.
- [195] S. Sen. Relaxations for the probabilistically constrained programs with discrete random variables. *Operations Research Letters*, 11:81–86, 1992.
- [196] M. Shaked and J. G. Shanthikumar. *Stochastic Orders and Their Applications*. Academic Press, Boston, 1994.
- [197] A. Shapiro. Asymptotic properties of statistical estimators in stochastic programming. *Annals of Statistics*, 17:841–858, 1989.
- [198] A. Shapiro. Asymptotic analysis of stochastic programs. *Annals of Operations Research*, 30:169–186, 1991.
- [199] A. Shapiro. Statistical inference of stochastic optimization problems. In S. Uryasev, editor, *Probabilistic Constrained Optimization: Methodology and Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000, 282–304.
- [200] 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*. 2001, 428–431.
- [201] A. Shapiro. Inference of statistical bounds for multistage stochastic programming problems. *Mathematical Methods of Operations Research*, 58:57–68, 2003.
- [202] A. Shapiro. Monte Carlo sampling methods. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming. Handbooks in Operations Research and Management Science* 10. North-Holland, Dordrecht, The Netherlands, 2003, 353–425.
- [203] A. Shapiro. On complexity of multistage stochastic programs. *Operations Research Letters*, 34:1–8, 2006.
- [204] A. Shapiro. Asymptotics of minimax stochastic programs. *Statistics and Probability Letters*, 78:150–157, 2008.
- [205] A. Shapiro. Stochastic programming approach to optimization under uncertainty. *Mathematical Programming, Series B*, 112:183–220, 2008.
- [206] 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*. Springer-Verlag, New York, 2005, 111–144.
- [207] A. Shapiro and T. Homem-de-Mello. A simulation-based approach to two-stage stochastic programming with recourse. *Mathematical Programming*, 81:301–325, 1998.
- [208] A. Shapiro and T. Homem-de-Mello. On the rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs. *SIAM J. Optimization*, 11:70–86, 2000.
- [209] A. Shapiro and Y. Wardi. Convergence analysis of stochastic algorithms. *Mathematics of Operations Research*, 21:615–628, 1996.

- [210] W. A. Spivey. Decision making and probabilistic programming. *Industrial Management Review*, 9:57–67, 1968.
- [211] V. Strassen. The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, 38:423–439, 1965.
- [212] 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.
- [213] R. Szekli. *Stochastic Ordering and Dependence in Applied Probability*. Springer-Verlag, New York, 1995.
- [214] 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.
- [215] G. Tintner. Stochastic linear programming with applications to agricultural economics. In H. A. Antosiewicz, editor, *Proc. 2nd Symp. Linear Programming*. National Bureau of Standards, Washington, D.C., 1955, 197–228.
- [216] S. Uryasev. A differentiation formula for integrals over sets given by inclusion. *Numerical Functional Analysis and Optimization*, 10:827–841, 1989.
- [217] S. Uryasev. Derivatives of probability and integral functions. In P. M. Pardalos and C. M. Floudas, editors, *Encyclopedia of Optimization*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001, 267–352.
- [218] 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.
- [219] A.W. van der Vaart and A. Wellner. *Weak Convergence and Empirical Processes*. Springer-Verlag, New York, 1996.
- [220] 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.
- [221] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 1944.
- [222] A. Wald. Note on the consistency of the maximum likelihood estimates. *Annals of Mathematical Statistics*, 20:595–601, 1949.
- [223] D.W. Walkup and R.J.-B. Wets. Some practical regularity conditions for nonlinear programs. *SIAM J. Control*, 7:430–436, 1969.
- [224] D.W. Walkup and R.J.B. Wets. Stochastic programs with recourse: Special forms. In *Proceedings of the Princeton Symposium on Mathematical Programming*. Princeton University Press, Princeton, NJ, 1970, pages 139–161.



- [225] S.W. Wallace and W.T. Ziemba, editors. *Applications of Stochastic Programming*. SIAM, Philadelphia, 2005.
- [226] R.J.B. Wets. Programming under uncertainty: The equivalent convex program. *SIAM J. Applied Mathematics*, 14:89–105, 1966.
- [227] R.J.B. Wets. Stochastic programs with fixed recourse: The equivalent deterministic program. *SIAM Review*, 16:309–339, 1974.
- [228] R.J.B. Wets. Duality relations in stochastic programming. In *Symposia Mathematica, Vol. XIX (Convegno sulla Programmazione Matematica e sue Applicazioni)*, INDAM, Rome. Academic Press, London, 1976, pages 341–355.
- [229] M. E. Yaari. The dual theory of choice under risk. *Econometrica*, 55:95–115, 1987.
- [230] P.H. Zipkin. *Foundations of Inventory Management*. McGraw–Hill, New York, 2000.

# Index

- approximation
  - conservative, 257
- Average Value-at-Risk, 257, 258, 260, 272
  - dual representation, 272
- Banach lattice, 403
- Borel set, 359
- bounded in probability, 382
- Bregman divergence, 237
- capacity expansion, 31, 42, 59
- chain rule, 384
- chance constrained problem
  - ambiguous, 285
  - disjunctive semi-infinite formulation, 117
- chance constraints, 5, 11, 15, 210
- Clarke generalized gradient, 336
- CLT (central limit theorem), 143
- common random number generation method, 180
- complexity
  - of multistage programs, 227
  - of two-stage programs, 181, 187
- conditional expectation, 363
- conditional probability, 363
- conditional risk mapping, 310, 315
- conditional sampling, 221
  - identical, 221
  - independent, 221
- Conditional Value-at-Risk, 257, 258, 260
- cone
  - contingent, 347, 386
  - critical, 178, 348
  - normal, 337
  - pointed, 403
  - polar, 29
  - recession, 29
  - tangent, 337
- confidence interval, 163
- conjugate duality, 340, 403
- constraint
  - nonanticipativity, 53, 291
- constraint qualification
  - linear independence, 169, 179
  - Mangasarian–Fromovitz, 347
  - Robinson, 347
  - Slater, 162
- contingent cone, 347, 386
- convergence
  - in distribution, 163, 382
  - in probability, 382
  - weak, 384
  - with probability one, 374
- convex hull, 337
- cumulative distribution function, 2
  - of random vector, 11
- decision rule, 21
- Delta theorem, 384, 386
  - finite dimensional, 383
  - second order, 387
- deviation of a set, 334
- diameter
  - of a set, 186
- differential uniform dominance condition, 145
- directional derivative, 334
  - $\varepsilon$ -directional derivative, 381
  - generalized, 336
  - Hadamard, 384
  - second order, 386
  - tangentially to a set, 387
- distribution
  - asymptotically normal, 163

- Binomial, 390
- conditional, 363
- Dirichlet, 98
- discrete, 361
- discrete with a finite support, 361
- empirical, 156
- gamma, 102
- log-concave, 97
- log-normal, 107
- multivariate normal, 16, 96
- multivariate Student, 150
- normal, 163
- Pareto, 151
- uniform, 96
- Wishart, 103
- domain
  - of a function, 333
  - of multifunction, 365
- dual feasibility condition, 128
- duality gap, 340, 341
- dynamic programming equations, 7, 64, 313
- empirical cdf, 3
- empirical distribution, 156
- entropy function, 237
- epiconvergence, 357
  - with probability one, 377
- epigraph of a function, 333
- $\varepsilon$ -subdifferential, 380
- estimator
  - common random number, 205
  - consistent, 157
  - linear control, 200
  - unbiased, 156
- expected value, 361
  - well defined, 361
- expected value of perfect information, 60
- Fatou's lemma, 361
- filtration, 71, 74, 309, 318
- floating body of a probability measure, 105
- Fréchet differentiability, 334
- function
  - $\alpha$ -concave, 94
  - $\alpha$ -concave on a set, 105
  - biconjugate, 262, 401
  - Carathéodory, 156, 170, 366
  - characteristic, 334
  - Clarke-regular, 103, 336
  - composite, 265
  - conjugate, 262, 338, 401
  - continuously differentiable, 336
  - cost-to-go, 65, 67, 313
  - cumulative distribution (cdf), 2, 360
  - distance generating, 236
  - disutility, 254, 271
  - essentially bounded, 399
  - extended real valued, 360
  - indicator, 29, 334
  - influence, 304
  - integrable, 361
  - likelihood ratio, 200
  - log-concave, 95
  - logarithmically concave, 95
  - lower semicontinuous, 333
  - moment-generating, 387
  - monotone, 404
  - optimal value, 366
  - polyhedral, 28, 42, 333, 405
  - proper, 333
  - quasi-concave, 96
  - radical-inverse, 197
  - random, 365
  - random lower semicontinuous, 366
  - random polyhedral, 42
  - sample average, 374
  - strongly convex, 339
  - subdifferentiable, 338, 402
  - utility, 254, 271
  - well defined, 368
- Gâteaux differentiability, 334, 383
- generalized equation
  - sample average approximation, 175
- generic constant  $O(1)$ , 188
- gradient, 335
- Hadamard differentiability, 384
- Hausdorff distance, 334
- here-and-now solution, 10
- Hessian matrix, 348
- higher order distribution functions, 90
- Hoffman's lemma, 344

- identically distributed, 374
- importance sampling, 201
- independent identically distributed, 374
- inequality
  - Chebyshev, 362
  - Chernoff, 391
  - Hölder, 400
  - Hardy–Littlewood–Polya, 280
  - Hoeffding, 390
  - Jensen, 362
  - Markov, 362
  - Minkowski for matrices, 101
- inf-compactness condition, 158
- interchangeability principle, 405
  - for risk measures, 293
  - for two-stage programming, 49
- interior of a set, 336
- inventory model, 1, 295
- Jacobian matrix, 335
- Lagrange multiplier, 348
- large deviations rate function, 388
- lattice, 403
- Law of Large Numbers, 2, 374
  - for random sets, 379
  - pointwise, 375
  - strong, 374
  - uniform, 375
  - weak, 374
- least upper bound, 403
- Lindeberg condition, 143
- Lipschitz continuous, 335
- lower bound
  - statistical, 203
- Lyapunov condition, 143
- mapping
  - convex, 50
  - measurable, 360
- Markov chain, 70
- Markovian process, 63
- martingale, 324
- mean absolute deviation, 255
- measurable selection, 365
- measure
  - $\alpha$ -concave, 97
  - absolutely continuous, 359
  - complete, 359
  - Dirac, 362
  - finite, 359
  - Lebesgue, 359
  - nonatomic, 367
  - sigma-additive, 359
- metric projection, 231
- mirror descent SA, 241
- model state equations, 68
- model state variables, 68
- moment-generating function, 387
- multifunction, 365
  - closed, 175, 365
  - closed valued, 175, 365
  - convex, 50
  - convex valued, 50, 367
  - measurable, 365
  - optimal solution, 366
  - upper semicontinuous, 380
- news vendor problem, 1, 330
- node
  - ancestor, 69
  - children, 69
  - root, 69
- nonanticipativity, 7, 52, 63
- nonanticipativity constraints, 72, 312
- nonatomic probability space, 367
- norm
  - dual, 236, 399
- normal cone, 337
- normal integrands, 366
- optimality conditions
  - first order, 207, 346
  - Karush–Kuhn–Tucker (KKT), 174, 207, 348
  - second order, 179, 348
- partial order, 403
- point
  - contact, 399
  - saddle, 340
- polar cone, 337
- policy
  - basestock, 8, 328
  - feasible, 8, 17, 64
  - fixed mix, 21

- implementable, 8, 17, 64
- myopic, 19, 325
- optimal, 8, 65, 67
- portfolio selection, 13, 298
- positive hull, 29
- positively homogeneous, 178
- probabilistic constraints, 5, 11, 87, 162
  - individual, 90
  - joint, 90
- probabilistic liquidity constraint, 94
- probability density function, 360
- probability distribution, 360
- probability measure, 359
- probability vector, 309
- problem
  - chance constrained, 87, 210
  - first stage, 10
  - of moments, 306
  - piecewise linear, 192
  - second stage, 10
  - semi-infinite programming, 308
  - subconsistent, 341
  - two stage, 10
- prox-function, 237
- prox-mapping, 237
- quadratic growth condition, 190, 350
- quantile, 16
  - left-side, 3, 256
  - right-side, 3, 256
- radial cone, 337
- random function
  - convex, 369
- random variable, 360
- random vector, 360
- recession cone, 337
- recourse
  - complete, 33
  - fixed, 33, 45
  - relatively complete, 10, 33
  - simple, 33
- recourse action, 2
- relative interior, 337
- risk measure, 261
  - absolute semideviation, 301, 329
  - coherent, 261
  - composite, 312, 318
  - consistency with stochastic orders, 282
  - law based, 279
  - law invariant, 279
  - mean-deviation, 276
  - mean-upper-semideviation, 277
  - mean-upper-semideviation from a target, 278
  - mean-variance, 275
  - multiperiod, 321
  - proper, 261
  - version independent, 279
- robust optimization, 11
- saddle point, 340
- sample
  - independently identically distributed (iid), 156
  - random, 155
- sample average approximation (SAA), 155
  - multistage, 221
- sample covariance matrix, 208
- sampling
  - Latin Hypercube, 198
  - Monte Carlo, 180
- scenario tree, 69
- scenarios, 3, 30
- second order regularity, 350
- second order tangent set, 348
- semi-infinite probabilistic problem, 144
- semideviation
  - lower, 255
  - upper, 255
- separable space, 384
- sequence
  - Halton, 197
  - log-concave, 106
  - low-discrepancy, 197
  - van der Corput, 197
- set
  - elementary, 359
  - of contact points, 399
- sigma algebra, 359
  - Borel, 359
  - trivial, 359
- significance level, 5
- simplex, 237

- Slater condition, 162
- solution
  - $\varepsilon$ -optimal, 181
  - sharp, 190, 191
- space
  - Banach, 399
  - decomposable, 405
  - dual, 399
  - Hilbert, 275
  - measurable, 359
  - probability, 359
  - reflexive, 399
  - sample, 359
- stagewise independence, 7, 63
- star discrepancy, 195
- stationary point of  $\alpha$ -concave function, 104
- stochastic approximation, 231
- stochastic dominance
  - $k$ th order, 91
  - first order, 90, 282
  - higher order, 91
  - second order, 283
- stochastic dominance constraint, 91
- stochastic generalized equations, 174
- stochastic order, 90, 282
  - increasing convex, 283
  - usual, 282
- stochastic ordering constraint, 91
- stochastic programming
  - nested risk averse multistage, 311, 318
- stochastic programming problem
  - minimax, 170
  - multiperiod, 66
  - multistage, 64
  - multistage linear, 67
  - two-stage convex, 49
  - two-stage linear, 27
  - two-stage polyhedral, 42
- strict complementarity condition, 179, 209
- strongly regular solution of a generalized equation, 176
- subdifferential, 338, 401
- subgradient, 338, 402
  - algebraic, 402
  - stochastic, 230
- supply chain model, 22
- support
  - of a set, 337
  - of measure, 360
- support function, 28, 337, 338
- support of a measure, 36
- tangent cone, 337
- theorem
  - Artstein–Vitale, 379
  - Aumann, 367
  - Banach–Alaoglu, 401
  - Birkhoff, 111
  - central limit, 143
  - Cramér’s large deviations, 388
  - Danskin, 352
  - Fenchel–Moreau, 262, 338, 401
  - functional CLT, 164
  - Glivenko–Cantelli, 376
  - Helly, 337
  - Hlawka, 196
  - Klee–Nachbin–Namioka, 404
  - Koksma, 195
  - Kusuoka, 280
  - Lebesgue dominated convergence, 361
  - Levin–Valadier, 352
  - Lyapunov, 368
  - measurable selection, 365
  - monotone convergence, 361
  - Moreau–Rockafellar, 338, 402
  - Rademacher, 336, 353
  - Radon–Nikodym, 360
  - Richter–Rogosinski, 362
  - Skorohod–Dudley almost sure representation, 385
- time consistency, 321
- topology
  - strong (norm), 401
  - weak, 401
  - weak\*, 401
- uncertainty set, 11, 306
- uniformly integrable, 382
- upper bound
  - consecutive, 204
  - statistical, 204
- utility model, 271

---

value function, 7  
Value-at-Risk, 16, 256, 273  
    constraint, 16  
variation  
    of a function, 195  
variational inequality, 174  
    stochastic, 174  
Von Mises statistical functional, 304  
  
wait-and-see solution, 10, 60  
weighted mean deviation, 256