LECTURES ON STOCHASTIC PROGRAMMING

Modeling and Theory

Alexander Shapiro
Darinka Dentcheva
Andrzej Ruszczyński

MPS-SIAM Series on Optimization

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

siam.
Society for Industrial
and Applied Mathematics
Philadelphia

mps
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

SIAM is a registered trademark. mps Mathematical Programming Society is a registered trademark.

Lectures on Stochastic Programming

MPS-SIAM Series on Optimization

This series is published jointly by the Mathematical Programming Society and the Society for Industrial and Applied Mathematics. It includes research monographs, books on applications, textbooks at all levels, and tutorials. Besides being of high scientific quality, books in the series must advance the understanding and practice of optimization. They must also be written clearly and at an appropriate level.

Editor-in-Chief

Philippe Toint, *University of Namur (FUNDP)*

Editorial Board

Oktay Gunluk, *IBM T.J. Watson Research Center*
Matthias Heinkenschloss, *Rice University*
C.T. Kelley, *North Carolina State University*
Adrian S. Lewis, *Cornell University*
Pablo Parrilo, *Massachusetts Institute of Technology*
Daniel Ralph, *University of Cambridge*
Mike Todd, *Cornell University*
Laurence Wolsey, *Université Catholique de Louvain*
Yinyu Ye, *Stanford University*

Series Volumes

Shapiro, Alexander, Dentcheva, Darinka, and Ruszczyński, Andrzej, *Lectures on Stochastic Programming: Modeling and Theory*

Conn, Andrew R., Scheinberg, Katya, and Vicente, Luis N., *Introduction to Derivative-Free Optimization*

Ferris, Michael C., Mangasarian, Olvi L., and Wright, Stephen J., *Linear Programming with MATLAB*

Attouch, Hedy, Buttazzo, Giuseppe, and Michaille, Gérard, *Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization*

Wallace, Stein W. and Ziemba, William T., editors, *Applications of Stochastic Programming*

Grötschel, Martin, editor, *The Sharpest Cut: The Impact of Manfred Padberg and His Work*

Renegar, James, *A Mathematical View of Interior-Point Methods in Convex Optimization*

Ben-Tal, Aharon and Nemirovski, Arkadi, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*

Conn, Andrew R., Gould, Nicholas I. M., and Toint, Phillippe L., *Trust-Region Methods*

试读结束：需要全本请在线购买： www.ertongbook.com

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

to Tsonka, Konstatin, and Marek;

and to the memory of Feliks, Maria, and Dentcho

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
$IF_{\mathfrak{F}}$, influence function, 304
$L^{\perp}$, orthogonal of (linear) space L, 41
$O(1)$, generic constant, 188
$O_p(\cdot)$, term, 382
S^{ε}, the set of ε-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
Π_X, metric projection onto set X, 231
$\xrightarrow{\mathcal{D}}$, convergence in distribution, 163
$\mathcal{T}^2_X(x, h)$, second order tangent set, 348
AV@R, Average Value-at-Risk, 258
$\bar{\mathfrak{P}}$, 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 ρ, 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
$\mathrm{cl}(A)$, topological closure of set A, 334
$\mathrm{conv}(C)$, convex hull of set C, 337
$\mathbb{C}\mathrm{orr}(X, Y)$, correlation of X and Y, 200
$\mathbb{C}\mathrm{ov}(X, Y)$, covariance of X and Y, 180
$\mathfrak{q}_\alpha$, weighted mean deviation, 256
$\mathfrak{s}_C(\cdot)$, support function of set C, 337
$\mathrm{dist}(x, A)$, distance from point x to set A, 334
$\mathrm{dom} f$, domain of function f, 333
$\mathrm{dom}\, \mathcal{G}$, domain of multifunction $\mathcal{G}$, 365
$\overline{\mathbb{R}}$, set of extended real numbers, 333
$\mathrm{epi} f$, epigraph of function f, 333
$\xrightarrow{e}$, epiconvergence, 377
$\hat{S}_N$, the set of optimal solutions of the SAA problem, 156
$\hat{S}^{\varepsilon}_N$, the set of ε-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
$\mathrm{int}(C)$, interior of set C, 336
$\lfloor a \rfloor$, integer part of $a \in \mathbb{R}$, 219
$\mathrm{lsc} f$, 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
∂, subdifferential, 338
∂°, Clarke generalized gradient, 336
∂_ε, epsilon subdifferential, 380
pos W, positive hull of matrix W, 29
$\Pr(A)$, probability of event A, 360
ri, relative interior, 337
σ_p^+, upper semideviation, 255
σ_p^-, lower semideviation, 255
$\mathrm{V@R}_\alpha$, Value-at-Risk, 256
$\mathbb{V}\mathrm{ar}[X]$, variance of X, 14
ϑ^*, 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 derivative, 336
$g'(x, h)$, directional derivative, 334
$o_p(\cdot)$, term, 382
p-efficient point, 116
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

试读结束：需要全本请在线购买： www.ertongbook.com

Contents

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[1]

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

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

$$\operatorname*{Min}_{x\geq 0} F(x,d). \tag{1.2}$$

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

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

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

$$\operatorname*{Min}_{x\geq 0} \big\{f(x) := \mathbb{E}[F(x,D)]\big\}. \tag{1.4}$$

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:

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

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

试读结束：需要全本请在线购买： www.ertongbook.com