Christian Küchler

Stability, Approximation, and Decomposition in Two- and Multistage Stochastic Programming

VIEWEG+
TEUBNER

Christian Küchler

Stability, Approximation, and Decomposition
in Two- and Multistage Stochastic Programming

VIEWEG+TEUBNER RESEARCH

Stochastic Programming

Editor:

Prof. Dr. Rüdiger Schultz

Uncertainty is a prevailing issue in a growing number of optimization problems in science, engineering, and economics. Stochastic programming offers a flexible methodology for mathematical optimization problems involving uncertain parameters for which probabilistic information is available. This covers model formulation, model analysis, numerical solution methods, and practical implementations. The series ”Stochastic Programming“ presents original research from this range of topics.

Christian Küchler

Stability, Approximation, and Decomposition in Two- and Multistage Stochastic Programming

VIEWEG+TEUBNER RESEARCH

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über
<http://dnb.d-nb.de> abrufbar.

Dissertation Humboldt-Universität zu Berlin, 2009

1. Auflage 2009

Lektorat: Dorothee Koch | Anita Wilke

Vieweg+Teubner ist Teil der Fachverlagsgruppe Springer Science+Business Media.
www.viewegteubner.de

Umschlaggestaltung: KünkelLopka Medienentwicklung, Heidelberg
Druck und buchbinderische Verarbeitung: STRAUSS GMBH, Mörlenbach
Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier.
Printed in Germany

ISBN 978-3-8348-0921-6

To my parents

Preface

The present doctoral thesis has been developed during my employment at the Department of Mathematics at Humboldt-Universität zu Berlin. This employment was enabled and supported by the *Bundesministerium für Bildung und Forschung* and the *Wiener Wissenschafts-, Forschungs- und Technologiefonds.*

First of all, it is a great pleasure to express my gratitude to my advisor Prof. Werner Römisch who enabled for me a most pleasant and inspiring time at Humboldt-Universität. Prof. Römisch supported me with numerous discussions and suggestions, incessantly encouraged me to follow my own ideas, and made me attend various workshops and conferences that gave me many opportunities to present and discuss my work.

Very special thanks also go to my colleague Stefan Vigerske from Humboldt-Universität for the intensive collaboration whose results are also part of this thesis. In particular, it was his profound knowledge in optimization and programming which had allowed to implement the recombining tree decomposition approach in the present form. I would also like to thank him for never tiring to discuss the many facets of computer manipulation as well as mathematics.

Special thanks also go to Prof. René Henrion from Weierstraß-Institut in Berlin for the very pleasant and fruitful collaboration on scenario reduction, resulting in parts of this thesis. I also want to thank him for numerous motivating and inspiring discussions.

I would also like to thank my colleagues from the BMBF network for the successful cooperation, especially Alexa Epe from Ruhr-Universität Bochum and Oliver Woll from Universität Duisburg-Essen. Furthermore, I want to thank all of my friends and colleagues from Humboldt-Universität for their support, many helpful discussions, and the enjoyable common coffee breaks. In particular, I thank Dr. Andreas Eichhorn, Thomas Surowiec, and, once again, Stefan Vigerske for proofreading parts of this thesis. I also want to thank Dr. Holger Heitsch for providing me his *scenred* software.

Finally, I want to express my deepest gratitude to my family for all their love and support, and, in particular, to Konny for having been so strong and patient during the last years.

Christian Küchler

Contents

List of Figures

List of Tables

Index of Notation

$\mathbf{1}_B(\cdot)$	the indicator function of the set B
$\langle\cdot,\cdot\rangle$	the standard scalar product in $\mathbb{R}^m$
$\lvert I\rvert$	the cardinality of the finite set I
$[\![\underline{a},\bar{a}]\!]$	a closed polyhedron in $\mathbb{R}^s$ induced by $\underline{a},\bar{a}\in\bar{\mathbb{R}}^k$, see p.116
$\alpha_{\mathcal{B}}$	the discrepancy distance w.r.t. the system $\mathcal{B}$, see p.98
$\mathcal{B}$	a system of Borel subsets of $\mathbb{R}^s$, see p.98
$\mathcal{B}_{cell}$	the system of closed cells in $\mathbb{R}^s$, see p.98
$\mathcal{B}_{cl}$	the system of closed subsets of $\mathbb{R}^s$, see p.98
$\mathcal{B}_{conv}$	the system of closed, convex subsets of $\mathbb{R}^s$, see p.98
$\mathcal{B}_{ph,k}$	the system of polyhedra in $\mathbb{R}^s$ having at most k vertices, see p.98
$\mathcal{B}_{ph,W}$	the system of polyhedra in $\mathbb{R}^s$ each of whose facets parallels a facet of $[0,1]^s$ or $\operatorname{pos} W$, see p.98
$\mathcal{B}_{rect}$	the system of closed, s-dimensional rectangles in $\mathbb{R}^s$
C_{feas}	a set of feasibility cuts, see p.43
C_{opt}	a set of optimality cuts, see p.42
δ_ξ	the Dirac measure in ξ
∂B	the topological boundary of the Borel set B
$\boldsymbol{d}$	the Pompeiu-Hausdorff distance, see p.12
$\mathbb{D}_{\mathbb{F}}$	a probability metric with ζ-structure, see p.100
$\Delta_{\mathcal{B}}$	the optimal value of the scenario reduction problem, see p.107
$\operatorname{dist}(a,B)$	the distance between the point $a\in\mathbb{R}^n$ and the set $B\subset\mathbb{R}^n$, see p.12
$\mathbb{E}[\boldsymbol{\xi}]$	the expectation of the random variable $\boldsymbol{\xi}$
$\mathbb{F}$	a class of Borel measurable mappings, see p.100
$\mathbb{F}_p$	a class of locally Lipschitz continuous mappings, see p.101
$\bar{\mathbb{F}}_p$	a class of uniformly bounded, piecewise locally Lipschitz continuous mappings, see p.101
$\bar{\mathbb{F}}_{p,\mathcal{B}'}$	a class of uniformly bounded, locally Lipschitz continuous mappings, see p.101
$\mathcal{I}_{\mathcal{B}}$	the system of critical index sets, see p.114
$\mathcal{I}^*_{\mathcal{B}}$	the system of reduced critical index sets, see p.115

$\operatorname{int} S$	the topological interior of the set S
Λ_{R_j}	the set of representative nodes at time R_j, see p.40
$\mathcal{M}^m_{[1,T]}$	a set of Borel measurable mappings, see p.11
$\operatorname{pos} W$	the positive cone of a $(d \times m)$-matrix W, i.e., $\operatorname{pos} W = \{Wy : y \in \mathbb{R}^m_+\}$
$\mathcal{Q}_t(\cdot,\cdot)$	the recourse function at time t, see p.13
$\mathbb{P}, \mathbb{Q}$	Borel probability measures
$\mathbb{P}_t$	the probability distribution of the random variable $\boldsymbol{\xi}_t$ under the measure $\mathbb{P}$, see p.11
$\mathbb{P}_{[t]}$	the probability distribution of the random variable $\boldsymbol{\xi}_{[t]}$ under the measure $\mathbb{P}$, see p.11
$\mathfrak{P}$	the set of supporting polyhedra, see p.117
$\mathcal{P}_p(\Xi)$	the set of all Borel probability measures on $\Xi \subset \mathbb{R}^s$ with finite absolute moments of order $p \geq 1$, see p.101
$\mathbb{R}$	the set of real numbers
$\mathbb{R}_+$	the set of non-negative real numbers, i.e., $\mathbb{R}_+ = [0, \infty)$
$\bar{\mathbb{R}}$	the set of extended real numbers, $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$
$\boldsymbol{\xi}$	an $\mathbb{R}^s$-valued random variable or stochastic process
$\boldsymbol{\xi}_{[t]}$	the random vector $(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_t)$
$\boldsymbol{\xi}_{[s,t]}$	the random vector $(\boldsymbol{\xi}_s, \ldots, \boldsymbol{\xi}_t)$ for $s, t \in \mathbb{N}$ with $s \leq t$
$\mathcal{S}(\boldsymbol{\xi})$	the set of decisions that are feasible w.r.t. the process $\boldsymbol{\xi}$, see p.12
S_n	the standard simplex in $\mathbb{R}^n$, see p.105
$v(\boldsymbol{\xi}), v(\mathbb{P})$	the optimal value of a stochastic program, see p.12
Ξ_t	the support of the measure $\mathbb{P}_t$
$\Xi_{[t]}$	the support of the measure $\mathbb{P}_{[t]}$
ζ_p	the p-th order Fortet-Mourier metric, see p.101
$\zeta_{p,\mathcal{B}_{ph,k}}$	an extended polyhedral discrepancy, see p.101
$\zeta_{p,\mathcal{B}_{ph,W}}$	an extended polyhedral discrepancy, see p.101

Chapter 1

Introduction

1.1 Stochastic Programming Models

In modern decision theory, it is often the case that at least some of the considered components of a given model are uncertain. Such problems arise in a variety of applications, such as inventory control, financial planning and portfolio optimization, airline revenue management, scheduling and operation of power systems, and supply chain management. Dealing with such decision problems, it is reasonable (and sometimes inevitable) to consider possible uncertainties within an optimization and decision-making process.

Stochastic programming provides a framework for modeling, analyzing, and solving optimization problems with some parameters being not known up to a probability distribution. Stochastic programming has its origin in the early work of Dantzig (1955). It was initially motivated to allow uncertain demand in an optimization model of airline scheduling to be taken into account. Since its beginnings, the field has grown and extended in various directions. Introductory textbooks that give an impression of the diversity of stochastic programming are due to Kall and Wallace (1994), Prékopa (1995), Birge and Louveaux (1997), and Ruszczyński and Shapiro (2003b). A variety of applications are discussed by Wallace and Ziemba (2005).

In particular, Dantzig (1955) introduced the concept of *two-stage linear stochastic programs*, which is today regarded as the classical stochastic programming framework. Two-stage stochastic programs model the situation of a decision maker who must first make (first-stage) decisions without knowing some uncertain parameters, which, e.g., may affect the costs or constraints on future decisions. In the second stage, the unknown parameters are revealed and the decision maker then makes a *recourse decision* that is allowed to depend (in a measurable way) on the realization of the stochastic parameters. In some applications, the first and second stage decisions stand for investment and operation decisions, respectively.

One of several possible mathematical formulations of a two-stage linear stochastic program reads as follows.

$$\inf \langle b_1, x_1\rangle + \mathbb{E}\left[\langle b_2(\boldsymbol{\xi}), x_2(\boldsymbol{\xi})\rangle\right] \tag{1.1}$$

s.t.

$$x_1 \in X_1, x_2(\boldsymbol{\xi}) \in X_2, \tag{1.2}$$

$$A_{2,1}(\boldsymbol{\xi})x_1 + A_{2,0}(\boldsymbol{\xi})x_2(\boldsymbol{\xi}) = h_2(\boldsymbol{\xi}) \tag{1.3}$$

Here, $\boldsymbol{\xi}$ is a random vector on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and models the stochastic parameters of the optimization problem. The variables x_1 and x_2 denote the first- and second-stage decision, respectively. For $i = 1, 2$, the decision x_i has to lie in some Borel constraint set $X_i \subset \mathbb{R}^m$. The first-stage decision x_1 is a constant, whereas the second-stage decision $x_2 = x_2(\cdot)$ is assumed to be a measurable mapping from $\Xi \triangleq \operatorname{supp} \mathbb{P}[\boldsymbol{\xi} \in \cdot]$ to $\mathbb{R}^m$. The decision x_i at stage i causes linear costs $\langle b_i, x_i\rangle$ with some coefficients $b_i \in \mathbb{R}^m$, where b_2 is allowed to depend affinely on the realization of $\boldsymbol{\xi}$. The decisions x_1 and x_2 are intertwined by the time coupling constraint (1.3). Finally, we note that the *technology matrix* $A_{2,1}$, the *recourse matrix* $A_{2,0}$, and the *right-hand side* h_2 may again depend affinely on $\boldsymbol{\xi}$ and take values in $\mathbb{R}^{n \cdot m}$ and $\mathbb{R}^n$, respectively. Note that the objective of the optimization problem (1.1) is to minimize the *expected value* of the total costs, and the constraints (1.2) and (1.3) are assumed to hold $\mathbb{P}$-almost surely.

Dantzig's framework has been extended during the last few decades in various directions. If some of the components of the decision variables in problem (1.1) are required to be integer, i.e.,

$$X_1, X_2 \subset \mathbb{Z}^{m_1} \times \mathbb{R}^{m_2} \tag{1.4}$$

with $m_1, m_2 \in \mathbb{N}, m_1 + m_2 = m$, one arrives at *mixed-integer two-stage linear stochastic programs.* Such integrality constraints may arise in a variety of practical situations, e.g., by modeling technical or economical systems that allow only for discrete decisions. Furthermore, integer variables can be helpful to describe discontinuities or piecewise linear functions by means of linear expressions.

Under integrality constraints, continuity and convexity properties of problem (1.1) are generally lost and thus the structure of mixed-integer stochastic programs is more intricate. Despite their practical relevance, mixed-integer stochastic programs have received only limited attention compared to the non-integer case, see Stougie (1985) for an early reference, and Römisch and

Schultz (2001), Louveaux and Schultz (2003), Schultz (2003), Sen and Sherali (2006) for more recent results.

The constraints in problem (1.1) are claimed to hold $\mathbb{P}$-almost surely. However, in several technical or economical decision problems almost-sure constraints may be too restrictive and may lead to unacceptably expensive solutions, or even to infeasibility of the decision problem. Such problems may be modeled by a further class of stochastic programs considering constraints that are assumed to hold (at least) with a certain probability, i.e., so-called *chance constraints*. Chance constraints are also a modeling tool for regulatory terms as the *Value-at-Risk constraints* in financial applications. A simple example for an optimization problem including chance constraints is the following.

$$\inf \langle b_1, x_1 \rangle \tag{1.5}$$

$$\text{s.t.}$$

$$x_1 \in X_1,$$

$$\mathbb{P}\left[A_{2,1}(\boldsymbol{\xi})x_1 \geq h_2(\boldsymbol{\xi})\right] \geq p, \tag{1.6}$$

where $p \in [0,1]$ denotes some probability threshold, and b_1, X_1, $A_{2,1}(\cdot)$, and $h_2(\cdot)$ are defined as above. Further formulations and various results on chance-constrained stochastic programming as well as numerous references are provided by Prékopa (1995, 2003).

A natural extension of the two-stage framework (1.1) is the consideration of a multi-stage setting. The latter corresponds to a situation where information about the unknown parameters is revealed sequentially and decisions have to be made at certain time points. A multi-stage extension of (1.1) can be formulated as follows:

$$\inf \langle b_1, x_1 \rangle + \sum_{t=2}^{T} \mathbb{E}\left[\langle b_t(\boldsymbol{\xi}_{[t]}), x_t(\boldsymbol{\xi}_{[t]}) \rangle\right] \tag{1.7}$$

$$\text{s.t.}$$

$$\begin{aligned} & x_1 \in X_1, \\ & x_t(\boldsymbol{\xi}_{[t]}) \in X_t, && t = 2, \ldots, T, \\ & \textstyle\sum_{\tau=0}^{t-1} A_{t,\tau}(\boldsymbol{\xi}_{[t]}) x_{t-\tau}(\boldsymbol{\xi}_{[t-\tau]}) = h_t(\boldsymbol{\xi}_{[t]}), && t = 2, \ldots, T, \end{aligned} \tag{1.8}$$

where $\boldsymbol{\xi} = (\boldsymbol{\xi}_t)_{t=1,\ldots,T}$ is a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$ with *time horizon* $T \in \mathbb{N}$ and $\boldsymbol{\xi}_{[t]}$ denotes the vector $(\boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_t)$. Note that, in particular, the decision x_t at time t is allowed to depend (in a measurable way) on $\boldsymbol{\xi}_{[t]}$, i.e., on the information obtained by observing $\boldsymbol{\xi}$ until time t.

A further extension of the classical framework is to replace (or, to adjust) the expectation operator $\mathbb{E}[\cdot]$ by some *risk functional* $\mathbb{F}[\cdot]$, i.e., the objective of (1.1) becomes

$$\inf \mathbb{F}\left[\langle b_1, x_1\rangle + \langle b_2(\boldsymbol{\xi}), x_2(\boldsymbol{\xi})\rangle\right]. \tag{1.9}$$

A variety of risk functionals have been proposed and studied in the literature. We refer to, e.g., the classical mean-variance approach due to Markowitz (1952), the widely applied (Average-)Value-at-Risk functionals, several (semi-)deviation measures, as well as functionals based on *utility functions*. Risk functionals for the multistage case have arisen and studied intensively during the last years; we refer to the recent book of Pflug and Römisch (2007) as well as the work of Eichhorn (2007) and the numerous references therein.

1.2 Approximations, Stability, and Decomposition

A common feature of the stochastic programming models considered in the previous section is that in most practical applications analytic solutions are rarely available. In such cases, one has to resort to numerical optimization methods to find optimal (or, at least, acceptable) solutions. While there are approaches that embed the construction of solutions into a sampling scheme, most of the numerical methods require the underlying stochastic entities to take only a finite number of values. Furthermore, in order to enable acceptable solution times, the number of possible values of the stochastic variables has to be very limited in many cases. In particular, this is the case for multistage and mixed-integer stochastic programs.

Approximations

Whenever the underlying probability measure does not fulfill the aforementioned finiteness requirements, a common approach is to approximate it by a measure that is supported by a suitable number of atoms (or, *scenarios*). For this purpose, several techniques have been developed. These techniques are based on different principles like random sampling (Shapiro, 2003b), Quasi Monte-Carlo sampling (Pennanen, 2005), and moment matching (Høyland et al., 2003; Høyland and Wallace, 2001). Accordingly, convergence properties of optimal values and/or solution sets for specific techniques as well as bounds for statistical estimates have been established, cf. Pflug (2003), Shapiro (2003b), and the references therein.

Another established approximation approach relies on the usage of specific probability metrics[1], see, e.g., Pflug (2001), Dupačová et al. (2003), Henrion et al. (2009), Heitsch and Römisch (2008). For such methods, the approximation of the initial measure in terms of a specific metric is considered reasonable whenever the optimal value and solution set of the considered stochastic program are known to possess some regularity with respect to the given metric (e.g., in form of Lipschitz or Hölder continuity). In order to identify distances that are suitable for specific problem classes, perturbation and stability issues become relevant.

Stability

In Stochastic Programming, the term stability usually refers to calmmess and continuity properties of optimal values and solution sets of a stochastic program under perturbations (or, approximations) of the underlying probability measure (cf. the recent survey by Römisch (2003)). For such regularity properties, the particular probability metric must be adapted to the structure of the stochastic program under consideration. In particular, *Fortet-Mourier* and *Wasserstein metrics* are relevant for two-stage stochastic programs (cf. Römisch and Schultz (1991); Rachev and Römisch (2002)). These distances have been used for the approximation of discrete probability distributions in two-stage stochastic programs without integrality requirements (Dupačová et al., 2003; Heitsch and Römisch, 2003, 2007). For two-stage mixed-integer models *discrepancy distances* are useful, see Schultz (1996), Römisch (2003), Römisch and Vigerske (2008). Discrepancy distances are also relevant for chance-constrained problems, see Römisch and Wakolbinger (1987), Henrion and Römisch (1999, 2004).

Heitsch et al. (2006) established a general stability result for linear multistage stochastic programs involving a specific *filtration distance*. The latter measures the distance between the information flows of the initial and the perturbed stochastic process. This distance is taken into account by the techniques for scenario tree generation developed by Heitsch and Römisch (2008).

While consistency and stability results have turned out to be useful for approximation purposes, they usually require the optimization problems and underlying random variables to fulfill specific boundedness and regularity properties, which, however, may be hard to verify in cases of practical interest. Furthermore, due to the numerical complexity of solving stochastic optimization problems, it may be necessary to use approximations that are

[1] The term *probability metric* refers to a distance on some space of probability measures.