Deviation measures in stochastic programming with mixed-integer recourse

Von der Fakultät für Naturwissenschaften
der Universität Duisburg-Essen
(Campus Duisburg)

zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigte Dissertation

von

Hans-Jürgen Andreas Märkert
aus
Wurzen

Referent: Prof. Dr. Rüdiger Schultz (Duisburg)

Korreferent: Prof. Dr. Maarten H. van der Vlerk (Groningen)

Datum der Einreichung: 21.01.2004

Tag der mündlichen Prüfung: 04.06.2004

Preface

During the past four years I have been working at the Institute of Mathematics of the University Duisburg-Essen (Campus Duisburg). This thesis documents the main part of my research activities in this period of time. As regards content, it falls into the mathematical field of stochastic programming – a field which I entered in 1998, when I became a member of Prof. Dr. Rüdiger Schultz' group at the University of Leipzig. I am grateful for his support provided since then, which included many aspects beyond mathematical research and which has been based on a cooperative personal relationship. At this place, I would like to thank the further members of my committee Prof. Dr. Maarten H. van der Vlerk, Prof. Dr. Gerlind Plonka, Prof. Dr. Günther Törner, and Prof. Dr. Martin Rumpf.

I continuously stood to benefit from the pleasant working environment in our research group. Questions were answered before they were raised, and common research was just business as usual. In this spirit, Stephan Tiedemann and I jointly worked out the concept of random cost functions to be presented in Chapter 3 of this thesis and already reported in two common conference lectures, Märkert and Tiedemann (2003). Sharing an office with Markus Westphalen did not only provide a likable atmosphere but also lead to constant improvements of my abilities in mathematical modelling and of my understanding of diverse production processes. Last but not least in this section, I would like to thank Ralf Gollmer for his support in all hard- and software issues.

My work also comprehended research projects in cooperation with practitioners and researchers from other groups. In particular, I would like to mention the fruitful cooperation with Guido Sand form the Process Control Laboratory of the University Dortmund within a program of the German Research Foundation. In this regard, I gratefully acknowledge the funding I received from the German Research Foundation.

In particular during the last year, my friends and family had to do without my presence on many occasions. Numerous appointments were cancelled at short notice. I promise to better myself and thank for the understanding.

Duisburg, 5. Mai 2004.

Abstract

This thesis aims to contribute to the systematic measurement of risk in decision problems under uncertainty. In particular, we intend to support the choice of a risk measure in stochastic linear programming with mixed-integer recourse. We restrict our discussion to *deviation measures* which include in our terminology such frequently used risk measures as the standard deviation, the standard semideviation, the absolute semideviation, and the expected excess of a fixed target.

The choice among risky alternatives is one major issue in decision theory. We review the axiomatic approaches of stochastic dominance orders and coherent risk measures. Then, we discuss deviation measures and the associated mean-risk models in the context of these concepts.

A mean-risk model is a bicriteria optimization problem on a family of random variables. In stochastic programming, the random variables are linked by a cost function. We provide a general view on the underlying decision problem, derive results on the structure and the stability of the mean-risk models from the structure of the cost function, and apply these results to stochastic programming with recourse. More precisely, we conclude properties concerning the continuity and the convexity of the optimal value functions and concerning the qualitative stability of the optimization problems. These properties lead to different implications for the different deviation measures. We investigate stochastic programs with and without integer variables but focus on the former.

Once we have characterized a number of suitable risk measures, we consider the situations in which the underlying probability distributions are discrete and finite. Then, for some risk measures the mean-risk problems turn into large-scale deterministic mixed-integer linear programs. We propose decomposition algorithms based on the Lagrangian relaxation of the nonanticipativity constraints. All algorithms are branch-and-bound algorithms; the employed lower bounds depend on the risk measure.

We report on the numerical experience gained with two real-life applications and thereby show the ability of the algorithms to find risk averse solutions. Both applications are production planning problems. The first one stems from chemical engineering and is the optimization of a plant producing several variants of a particular polymer. The second one is the optimization of gas transport through a pipeline system run by a large gas supplier. We introduce the model background and the mathematical programming models in detail.

Along with this thesis, we deliver a C-implementation of the used algorithms. Informations concerning the availability of the implementation are compiled at the end of the text.

Zusammenfassung (Abstract in German)

Diese Arbeit befasst sich mit der Auswahl von Risikomaßen in Entscheidungsproblemen, die als stochastische lineare Optimierungsprobleme mit gemischt-ganzzahliger Kompensation modelliert werden können. Wir betrachten dabei ausschließlich Risikomaße, die die Abweichung vom Erwartungswert beziehungsweise von einem festen Zielwert messen und bezeichnen sie als *Abweichungsmaße*. Dazu gehören unter anderem die Standardabweichung, die Standardsemiabweichnung und das erwartete Überschreiten eines Zielwertes.

Wir ordnen die Abweichungsmaße in den Kontext der stochastischen Dominanz sowie der kohärenten Risikomaße ein. Diese Konzepte stellen axiomatische Ansätze zur Auswahl risikoaverser Alternativen in Entscheidungsproblemen unter Unsicherheit dar und führen unter anderem zu Aussagen über die Eignung einzelner Mean-Risk-Probleme, risikoaverse Lösungen zu erzeugen.

Ein Mean-Risk-Problem ist ein bikriterielles Optimierungsproblem auf einer Familie von Zufallsvariablen. In der stochastischen linearen Optimierung mit Kompensation besteht zwischen diesen Zufallsvariablen ein funktionaler Zusammenhang, der durch die Kostenfunktion gegeben ist. Wir leiten Ergebnisse zur Struktur und Stabilität der Mean-Risk-Probleme von den Eigenschaften der Kostenfunktionen ab. Im Einzelnen erhalten wir Aussagen zur Stetigkeit und Konvexität der Optimalwertfunktionen und zur qualitativen Stabilität der Probleme bezüglich der zu Grunde liegenden Wahrscheinlichkeitsverteilung. Diese Resultate wenden wir auf unsere speziellen Optimierungsprobleme an, wobei wir uns auf Probleme mit Ganzzahligkeitsforderungen konzentrieren.

Bestimmte Mean-Risk Probleme werden unter der Annahme diskreter endlicher Wahrscheinlichkeitsverteilungen zu hochdimensionalen gemischt-ganzzahligen deterministischen Optimierungsproblemen. Zur Lösung derselben schlagen wir Dekompositionsalgorithmen vor, die auf der Lagrange-Relaxation der Nichtantizipativitätsbedingungen basieren. Alle Algorithmen sind Branch-und-Bound-Algorithmen; die eingesetzten unteren Schranken hängen vom Risikomaß ab.

Wir dokumentieren numerische Ergebnisse an zwei Produktionsplanungsproblemen. Es handelt sich dabei zum einen um ein Problem aus der chemischen Verfahrenstechnik; die Produktionsplanung einer Anlage zur Herstellung von Varianten eines bestimmten Polymers. Zum anderen beschäftigen wir uns mit dem optimalen Gastransport im Netz eines großen Versorgers. Wir zeigen, dass wir mit den entwickelten Dekompositionsalgorithmen in der Lage sind, effiziente Lösungen der Mean-Risk-Probleme zu finden.

Mit der Arbeit stellen wir auch eine C-Implementation der benutzten Algorithmen bereit. Informationen zur Verfügbarkeit des Programms finden sich im Anhang.

Contents

List of symbols

Number ranges

$\mathbb{N}$	the set $\{1, 2, \dots\}$
$\mathbb{Z}_+$	the set $\{0, 1, 2, \dots\}$
$\mathbb{R}$	the set $(-\infty, \infty)$
$\mathbb{R}_+$	the set $[0, \infty)$

Probability and measure

Ω	event space
ω	element of Ω, event
$\mathcal{A}$	sigma algebra in Ω
$\mathbb{P}$	probability measure on Ω
$(\Omega, \mathcal{A}, \mathbb{P})$	probability space
ξ	random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ with realizations in $\mathbb{R}^l$
μ	image measure induced by ξ
$\mathcal{B}^l$	Borel algebra in $\mathbb{R}^l$
E	expected value
$\mathcal{R}$	generic risk measure
$\mathcal{D}_2$	standard deviation
$\mathcal{D}_\mathrm{p}$	central deviation of order p
$\mathcal{D}_\mathrm{p}^+$	semideviation of order p
$\mathrm{E}_\mathrm{p}^\eta$	expected excess of a target η of order p

Random cost function

x	decision variable in $\mathbb{R}^n$
ξ	random parameter in $\mathbb{R}^l$
Z	random cost function mapping from $\mathbb{R}^n \times \mathbb{R}^l$ to $\mathbb{R}$
Q_E	expected value function mapping form $\mathbb{R}^n$ to $\mathbb{R}$
$Q_\mathcal{R}$	generic risk function mapping form $\mathbb{R}^n$ to $\mathbb{R}$

Stochastic programming

x	first-stage variable
y	second-stage variable
T	technology matrix
W	recourse matrix
ϕ	recourse function

List of abbreviations

General abbreviations

e.g.	for example
i.e.	that is
pp.	pages
w.r.t.	with respect to

Mathematical abbreviations

a.s.	almost surely
l.s.c.	lower semicontinuous
MILP	mixed-integer linear program
s.t.	subject to
u.s.c.	upper semicontinuous
w.l.o.g.	without loss of generality

1 Introduction and outline

1.1 Introduction

The future is uncertain. Many decision problems involve future data and thus are also subject to uncertainty. We arrive at the field of optimization under uncertainty if we intend to take decisions in an *optimal* way. In this thesis we investigate some questions arising in stochastic linear programming with mixed-integer recourse, a problem class which falls into the broad field of optimization under uncertainty.

Since the development of the simplex algorithm in Dantzig (1951), deterministic linear programming plays an important role in the field of mathematical programming as there are – in contrast to deterministic nonlinear programming – practically and theoretically efficient solution algorithms. For this reason, linear programs with random parameters have also attained a particular attention.

A stochastic linear program is the composition of a random linear program and a stochastic model. The random linear program

$$\inf_{x \in \mathbb{R}^n_+} \{cx \ : \ A(\omega)x \geq b(\omega)\} \tag{1.1}$$

is characterized by the random parameters A and b mapping from the event space Ω into $\mathbb{R}^{s \times n}$ and $\mathbb{R}^s$, respectively. We note that the case of random cost coefficients $c(\omega)$ is covered by the above formulation since for each $\omega \in \Omega$ the equivalent reformulation

$$\inf_{z \in \mathbb{R}} \{z \ : \ z \geq c(\omega)x, \ A(\omega)x \geq b(\omega)\} \tag{1.2}$$

yields a problem of type (1.1).

Throughout we assume that we are given a probability distribution for the random parameters A and b. This differs from the settings in online and robust optimization, see Albers (2003) and Ben-Tal and Nemirovski (2002), respectively. In contrast to stochastic scheduling models, we additionally assume that the probability distribution of the random parameters does not depend on the decision vector x, cf. Niño-Mora (2001).

It is the stochastic model's task to define what we want to consider as a solution of problem (1.1). The distribution problem, for instance, asks for the probability distribution of the optimal value and the optimal solutions of problem (1.1) given a probability measure $\mathbb{P}$ on the measure space $(\Omega, \mathcal{A})$. Models that call for some moment of these probability distributions are closely related to the distribution problem. Among them, there is the *wait-and-see* problem

$$\int_\Omega \inf_{x \in \mathbb{R}^n_+} \{cx \ : \ A(\omega)x \geq b(\omega)\} \mathbb{P}(d\omega), \tag{1.3}$$

which yields the mathematical expectation w.r.t. the probability distribution of the optimal value.

In practice, we often replace the random parameters $A(\omega)$ and $b(\omega)$ by there expected values $\bar{A} := \int_\Omega A(\omega)\mathbb{P}(d\omega)$ and $\bar{b} := \int_\Omega b(\omega)\mathbb{P}(d\omega)$, respectively. Clearly, the resulting deterministic problem

$$\inf_{x \in X \subset \mathbb{R}^n} \{cx \ : \ \bar{A}x \geq \bar{b}\} \tag{1.4}$$

neglects a great part of the information provided by the probability measure $\mathbb{P}$ and should therefore be reserved for large-scale or complex situations where other models are not applicable.

Stochastic models with *chance constrains* were originated by Charnes and Cooper (1959). These models emphasize the aspect of reliability. Each individual group of constraints is required to hold with prescribed probabilities

$$\inf_{x \in X \subset \mathbb{R}^n} \{cx \ : \ \mathbb{P}(\{\sum_{i=1}^n a_{ij}(\omega)x_i \geq b_j(\omega) \quad \forall j \in I_k\}) \geq \alpha_k \quad k = 1, \ldots, K\}, \tag{1.5}$$

where the sets I_k are disjoint and their union equals $\{1, \ldots, s\}$. We refer to Prekopa (2003) for a recent overview of theory and algorithms for this model class.

In this thesis we discuss *recourse models.* Such problems were first investigated by Dantzig (1955) and Beale (1955). The conceptual idea behind recourse models is the following; assume the decisions are two-stage in the sense that some of them, say x, have to be taken immediately whilst others, say y, may be delayed to a time when uncertainty has revealed. We can write a random linear program of this type as

$$\inf_{x \in X,\, y(\omega) \in \mathbb{R}^m_+} \{cx + q(\omega)y(\omega) \ : \ T(\omega)x + W(\omega)y(\omega) = h(\omega)\}. \tag{1.6}$$

Again, the random parameters $\xi := (q, T, W, h) : \Omega \to \mathbb{R}^m \times \mathbb{R}^{s \times n} \times \mathbb{R}^{s \times m} \times \mathbb{R}^s$ are defined on the probability space $(\Omega, \mathbb{P}, \mathcal{A})$. The set $X \subset \{x \in \mathbb{R}^n_+ : Ax = b\}$ contains all deterministic constraints on x. Inequality constraints can be handled in (1.6) by the introduction of appropriate slack variables. In fact, we will exclusively deal with problems in which the *recourse matrix* W and the *recourse costs* q are deterministic. The case of a deterministic recourse matrix W is called *fixed recourse.* It allows for stronger results concerning the problem structure. For nonfixed (or random) recourse models we refer to Walkup and Wets (1967).

The model (1.6) is also referred to as *two-stage model*. When the decision process has a multi-stage nature we arrive at multi-stage models. A survey of theory and algorithms for these models is given in Römisch and Schultz (2001).

We note that the problem (1.6) is not yet well-defined. As the constraints include random parameters, the meaning of feasibility and thus of optimality is not clear. We complete the recourse model by adding an objective function criterion. Before we do so, we rewrite problem (1.6) – now assuming a fixed recourse matrix W and fixed recourse costs q – as

$$\inf_{x \in X} \{cx + \tilde{\phi}(x, \xi(\omega))\} \tag{1.7}$$

where

$$\tilde{\phi}(x, \xi(\omega)) = \inf_{y \in \mathbb{R}^m_+} \{qy \; : \; T(\omega)x + Wy = h(\omega)\}. \tag{1.8}$$

We note that, provided $\tilde{\phi} : \mathbb{R}^n \times \mathbb{R}^{s \times n} \times \mathbb{R}^s \to \mathbb{R}$ is measurable, we can regard $\mathcal{Z} := \{cx + \tilde{\phi}(x, \xi(\omega)) \; : \; x \in X\}$ as a family of random variables. Now, each function $\mathcal{R} : \mathcal{Z} \to \mathbb{R}$, e.g. the expected value, some measure of risk, or a weighted sum of both, can serve as objective criterion

$$\inf_{x \in X} \mathcal{R}[cx + \tilde{\phi}(x, \xi(\omega))]. \tag{1.9}$$

Our focus is on integer models, i.e. in addition to the constraints employed in problem (1.6) we may have integrality requirements on the variables x and y. It is well-known that in the deterministic case the presence of integer variables significantly influences the complexity of linear programs. In particular, the class of mixed-integer linear programs belongs to the class of $\mathcal{NP}$-hard problems, i.e. in some sense they do not allow an efficient algorithmic treatment, cf. Garey and Johnson (1979). When the random variable ξ has a discrete and finite probability distribution, stochastic linear mixed-integer programs correspond to special-structured large-scale deterministic programs and are then also $\mathcal{NP}$-hard. More results on the complexity of stochastic integer programs and related topics can be found in Stougie and van der Vlerk (2003).

The measurement of risk associated with uncertain investments has been a major issue in mathematical finance for several decades. The focus of discussion has been on the conceptual appropriateness of the risk measures, i.e. the central question to answer is: does a risk measure reflect the risk aversion of a decision maker?

In recent years, risk aversion has entered the field of stochastic programming, see e.g. Ogryczak and Ruszczynski (1999) and Rockafellar and Uryasev (2002). Here, besides the conceptual appropriateness, the consistency of risk measures with mathematical programming structures is also of interest. We aim to contribute to the latter issue through the investigation of a number of risk measures in the framework of stochastic linear programming with mixed-integer recourse. From this perspective, we follow the articles of Schultz and Tiedemann (2003) and Tiedemann (2004) where the risk measures *excess probabilities* and *conditional value-at-risk* are considered.

1.2 Outline

We will proceed as follows. In Chapter 2, we introduce the list of risk measures under consideration. We review some concepts that are relevant for the choice of the function $\mathcal{R}$. The risk measures are discussed w.r.t. their accordance with a list of axiomatic properties including *stochastic dominance orders*. In this chapter, we also revisit the mean-risk model as a bicriteria mathematical program.

In Chapter 3, we investigate mean-risk models as optimization problems on a general class of random variables. These random variables correspond to decision variables in stochastic optimization problems. Motivated by the structure of mixed-integer value functions (to be discussed in Chapter 5) we focus on families of random variables that are linked via a lower semicontinuous function. We derive properties concerning structure and stability of mean-risk models on such families. Results obtained here shall be helpful in the following two chapters.

In Chapter 4, we introduce the basic notations of stochastic programming and briefly summarize the structural properties of the classical (purely expectation based) two-stage stochastic linear programs. Mean-risk models are then discussed in the context by applying the results of Chapter 3. The main purpose of this chapter is to provide a comparison to the mixed-integer counterpart of Chapter 5.

In Chapter 5, we turn to decision problems involving integer variables. Again, we review the purely expectation based model. Then, we derive some results concerning the structure and stability of mean-risk models in the context of stochastic programming with mixed-integer recourse.

In Chapter 6, we investigate the mean-risk models assuming a discrete and finite probability distribution of the random variable ξ. We confine ourselves to *linear* mean-risk models which have attractive theoretical properties. For these models we present equivalent mixed-integer linear programs and algorithms.

In the final chapter, we introduce two stochastic programming applications which both include integer variables. We set out the technical and economical background and discuss the mathematical programming models. We indicate the potential of the two-stage stochastic programming approach by numerical results.

2 Some aspects of risk minimization

2.1 Preliminary notes

The question of how risk should be measured in decision problems under uncertainty has been extensively discussed in the past decades, see Pflug (1999), the introductory section of Steinbach (2001) and the references therein. Most of the work in this area is based on the theory of rational behavior under uncertainty developed in Morgenstern and von Neumann (1947), and so is the research on mean-risk models originated in Markowitz (1959).

Another fundamental branch inspired by Morgenstern and von Neumann (1947) is the one of stochastic dominance orders, see Fishburn (1964), Hadar and Russell (1969) and Quirk and Saposnik (1962) as well as the survey Levy (1992) and the bibliography Bawa (1982). Stochastic dominance leads to mean-risk models involving asymmetric risk measures, see Fishburn (1977) and the recent articles of Ogryczak and Ruszczynski (2001, 2002). A comprehensive overview of stochastic orders including stochastic dominance is provided in Müller and Stoyan (2002). In Chapter 6, we shall use the concept of stochastic dominance to design an algorithm for special stochastic optimization problems.

In more recent years, the axiomatic concepts of coherent and convex risk measures have been developed and extended, see the articles Artzner et al. (1999, 2002) and Föllmer and Schied (2003). In Chapter 4, we shall point out the interplay of convexity in stochastic optimization and the axiomatic setting of Föllmer and Schied (2003).

Here, we do not intend to discuss the validity of the different approaches. However, for the convenience of the reader, we review the risk measures under consideration w.r.t. the above theoretical concepts.

We remark that the class of risk measures covered in this thesis differs from the class of risk measures considered in Rockafellar et al. (2002). However, there is a nonempty intersection and some results in this chapter can also be found there.

2.2 Deviation measures

This section serves to introduce and motivate the list of risk measures that we want to discuss in this thesis. We use the term *deviation measures* for those risk measures that are based on the expected deviation of the random variable from its mean or from a fixed target. We investigate a subset of this class of risk measures.

Throughout this thesis we impose a minimization framework, i.e. we prefer small values (low costs). This is also reflected in the definitions of the risk measures given in Table 2.1, where X is a random variable defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and η_o a target value in $\mathbb{R}$.

Name	Symbol	Definition $(p = 1, 2, \ldots)$
Central deviation of order p	$\mathcal{D}_{\mathrm{p}}(X)$	$\left(\int_\Omega \lvert X - \mathrm{E}X\rvert^p \ \mathbb{P}(d\omega)\right)^{\frac{1}{p}}$
Semideviation of order p	$\mathcal{D}^+_{\mathrm{p}}(X)$	$\left(\int_\Omega \max\{X - \mathrm{E}X, 0\}^p \, \mathbb{P}(d\omega)\right)^{\frac{1}{p}}$
Expected excess of order p	$\mathrm{E}^\eta_{\mathrm{p}}(X)$	$\left(\int_\Omega \max\{X - \eta_o, 0\}^p \ \mathbb{P}(d\omega)\right)^{\frac{1}{p}}$

Table 2.1: List of risk measures

Particularly important are the cases where p takes the values 1 and 2. For p equal to 2, the central deviation turns into the *standard deviation*. Due to the seminal work of Markowitz (1959), the latter and its square the *variance* have become frequently applied measures of risk.

The semideviation of order 1 is called *absolute semideviation* and the semideviation of order 2 is referred to as *standard semideviation*. Semideviations are often the first choice when the focus is on preventing the excess of the mean rather than both the excess of the mean and the shortfall below it.

There is no uniform nomenclature for the expected excess of a target in the literature. In the maximization context, it is referred to as the *expected shortfall of a target* and the *below-target returns*, see Ogryczak and Ruszczynski (1999) and Fishburn (1977), respectively. The expected excess of a target can be seen as a substitute for the semideviation in situations, when problems involving the latter are hard to solve. However, it requires substantial knowledge about the random variable to specify a reasonable target.

2.3 Stochastic dominance

Introduction The question of selecting an appropriate (partial) order on a family of random variables arises when decision problems require the comparison of random variables. In this section we review the concept of stochastic dominance orders which are partial orders that have attracted some attention in the literature.

We impose a minimization framework, i.e. small values are preferred to big ones. To avoid confusing notations when switching from minimization to maximization and backwards, we adapt the concepts of stochastic ordering and decision theory to our minimization set-up. In particular, we say that a random variable dominates another one when it is smaller in some sense.

The comparison of random variables is related to the comparison of the associated utility. In Morgenstern and von Neumann (1947), the question whether the utility of events is measurable has been addressed. This rather conceptual question is answered with *yes* and is followed by the formulation of the commonly accepted *axioms of rational decision making.* Among others, these axioms imply that each rational decision maker possesses a utility function f which serves to compare random variables. In particular, the *expected utility principle* says that the decision maker prefers a random variable X to a random variable Y if it holds $\mathrm{E}f(X) \leq \mathrm{E}f(Y)$, where $\mathrm{E}f(X) = \int_\Omega f(X)\ \mathbb{P}(d\omega)$.

In practice, it turns out to be impossible to determine exactly the utility function of an individual decision maker. However, the function might be known to belong to some family of functions. For instance, we expect that the utility function of a rational decision makers preferring small values belongs to the class of nondecreasing functions. The fact, that many decision problems call for solutions valid for a number of decision makers with different utility functions, provides another argument for the consideration of a family of functions rather than a single function.

The previous elaborations lead us to the first stochastic order relation, cf. Fishburn (1964) and Quirk and Saposnik (1962).

Definition 2.1 (first stochastic order, $X \preceq_{\mathrm{FSD}} Y$) *The random variable X is less than or equal to the random variable Y w.r.t. first stochastic order iff* $\mathrm{E}f(X) \leq \mathrm{E}f(Y)$ *for all nondecreasing functions f for which both expectations exist. We write $X \preceq_{FSD} Y$.*

Since f is nondecreasing iff $-f$ is nonincreasing, we have the equivalence

$$X \preceq_{\mathrm{FSD}} Y \iff \mathrm{E}f(X) \geq \mathrm{E}f(Y) \quad \text{for all nonincreasing functions } f, \tag{2.1}$$

provided both expectations exist. The cumulative distribution function F_X of a random variable X is defined by $F_X(t) := \mathbb{P}(X \leq t)$ for all $t \in \mathbb{R}$. We can express the first stochastic order relation in terms of the cumulative distribution functions F_X and F_Y of the random variables X and Y, respectively.

Lemma 2.2 (Fishburn (1964)) *It holds*

$$X \preceq_{FSD} Y \qquad \iff \qquad F_X(t) \geq F_Y(t) \qquad \forall t \in \mathbb{R}.$$

We note that $F_X(t) \geq F_Y(t)$ also implies $\mathbb{P}(X > t) \leq \mathbb{P}(Y > t)$. Thus, the preference of small values is reflected by the fact that the probability of

X exceeding an arbitrary real target t is smaller than the probability of Y exceeding this target. We say X dominates Y by first stochastic order iff $X \preceq_{\text{FSD}} Y$ and $Y \not\preceq_{\text{FSD}} X$, i.e. there is a nonincreasing function g such that $\mathrm{E}g(X) > \mathrm{E}g(Y)$.

If, in addition, the decision maker is risk averse, it is commonly accepted to assume that she prefers the constant random variable $\mathrm{E}X$ to the actual random variable X no matter what distribution X has. In other words a sure outcome is evaluated higher than any risky alternative. In terms of the utility function it holds $f(\mathrm{E}X) \leq \mathrm{E}f(X)$. By Jensen's inequality this is the case iff f is convex. This gives rise to the second stochastic order relation.

Definition 2.3 (second stochastic order, $X \preceq_{\text{SSD}} Y$) *The random variable X is less than or equal to the random variable Y w.r.t. second stochastic order iff $\mathrm{E}f(X) \leq \mathrm{E}f(Y)$ for all nondecreasing convex functions f for which both expectations exist. We write $X \preceq_{SSD} Y$.*

Since f is nondecreasing convex iff $-f$ is nonincreasing concave, an equivalent definition is given by

$$X \preceq_{\text{SSD}} Y \iff \mathrm{E}f(X) \leq \mathrm{E}f(Y) \quad \text{for all nonincreasing concave functions } f, \tag{2.2}$$

provided both expectations exist. Analogously to the first stochastic order, there is an equivalent condition for $X \preceq_{\text{SSD}} Y$ involving functions related to the cumulative distribution functions of X and Y.

Lemma 2.4 (Hadar and Russell (1969)) *It holds*

$$X \preceq_{SSD} Y \iff \mathrm{E}\max\{X-t,0\} \leq \mathrm{E}\max\{Y-t,0\} \qquad \forall t \in \mathbb{R}.$$

We say X dominates Y by second stochastic order iff $X \preceq_{\text{SSD}} Y$ and $Y \not\preceq_{\text{SSD}} X$. The relation to the cumulative distribution function mentioned above is given by

$$\mathrm{E}\max\{X-t,0\} = \int_{\Omega} \mathbb{P}(X > t)\ \mathbb{P}(d\omega),$$

see Müller and Stoyan (2002). This motivates the definition of p-th order stochastic dominance for $p \in \mathbb{N}$. Recursively, we assign $F_X^1(t) := \mathbb{P}(X > t)$ and $F_X^{p+1}(t) := \int_\Omega F_X^p(t)\ \mathbb{P}(d\omega)$ for all $t \in \mathbb{R}$.

Definition 2.5 (p-th stochastic order, $X \preceq_{\text{p}} Y$) *The random variable X is less than or equal to the random variable Y w.r.t. p-th stochastic order iff $F_X^p(t) \geq F_Y^p(t)$ for all $t \in \mathbb{R}$. We write $X \preceq_p Y$.*

The recursive definition of F_X^p and a monotonicity argument yield the following result.

Lemma 2.6 *Let $p \in \mathbb{N}$. It holds* $X \preceq_p Y \implies X \preceq_{p+1} Y$.

While the Definitions 2.1 and 2.3 of the stochastic orders are more intuitive w.r.t. the preferences they express, the equivalent conditions in terms of the (aggregated) cumulative distribution functions provide accessible optimization criteria. However, if the random variables are continuously distributed both stochastic orders lead to multicriteria optimization problems with a continuum of criteria which correspond to the comparison of nonincreasing and nonincreasing concave functions, respectively.

For discrete distributions, the functions $f(t) := \mathbb{P}(X > t)$ and $g(t) := \mathrm{E}\max\{X - t, 0\}$ are piecewise constant and piecewise linear respectively. The optimization w.r.t. first and second stochastic order reduces to multicriteria mathematical programming with a finite number of criteria depending on the number of probability atoms of the random variables. Let us clarify this point for the first order stochastic dominance.

Assume we are given a class $\mathcal{X}$ of discretely distributed random variables. Suppose that all random variables X in $\mathcal{X}$ have S probability atoms attained with probability $\frac{1}{S}$. For X in $\mathcal{X}$ we denote the probability atoms as $X_1, \ldots, X_S$ and an order list (starting with the smallest) of these probability atoms by $X_{(1)}, \ldots, X_{(S)}$. Then, a random variable which is optimal w.r.t. first order stochastic dominance can be found by the multicriteria optimization problem

$$\min_{X \in \mathcal{X}} [X_{(1)}, \ldots, X_{(S)}]. \tag{2.3}$$

We prove a Lemma which verifies this equivalence.

Lemma 2.7 *Let X and Y be two discretely distributed random variables with S probability atoms attained with probability $\frac{1}{S}$. Let their order statistics be given by $X_{(1)}, \ldots, X_{(S)}$ and $Y_{(1)}, \ldots, Y_{(S)}$, i.e. it holds $X_{(1)} \leq \cdots \leq X_{(S)}$ and $Y_{(1)} \leq \cdots \leq Y_{(S)}$. Then we have*

$$X_{(j)} \leq Y_{(j)} \quad j = 1, \ldots, S \quad \iff \quad X \preceq_{FSD} Y.$$

Proof Let $F_X(t)$ and $F_Y(t)$ be the cumulative distribution functions of X and Y, respectively.

Assume $X_{(j)} \leq Y_{(j)}$ for $j = 1, \ldots, S$. Let $t \in \mathbb{R}$. W.l.o.g. we assume that $X_{(j)} \leq t$ for all $j \leq j^*$ and $X_{(j)} > t$ for all $j > j^*$, $j^* \in \{1, \ldots, S\}$. We obtain

$$\mathbb{P}(X \leq t) = F_X(t) = \frac{j^*}{S} \geq F_Y(t) = \mathbb{P}(Y \leq t).$$

Thus X dominates Y in first stochastic order.

Conversely, assume $X \preceq_{\text{FSD}} Y$. Assume there exists $j^* \in \{1, \ldots, S\}$ such that $X_{(j^*)} > Y_{(j^*)}$. Let $t := Y_{(j^*)}$. We obtain

$$\mathbb{P}(X \leq t) = F_X(t) \leq \frac{j^* - 1}{S} < \frac{j^*}{S} = F_Y(t) = \mathbb{P}(Y \leq t).$$

This contradicts $X \preceq_{\text{FSD}} Y$. Therefore, it holds $X_{(j)} \leq Y_{(j)}$ for $j = 1, \dots, S$. □

Efficient algorithms for multicriteria integer programs with a large number of criteria are currently not available. Mean-risk models are computationally and algorithmically more amenable. Random variables are compared based on two scalar criteria - the expected value and a measure of risk. In Section 2.5, we introduce this approach in some detail. We want to close this section by investigating our list of risk measures w.r.t. their consistency with stochastic dominance of first and second order.

Definition 2.8 (consistency of risk measures with stochastic dominance) *Let $\alpha > 0$. A risk measure $\mathcal{R}$ is α-consistent with stochastic dominance of order p iff*

$$X \preceq_p Y \implies \mathrm{E}X + \alpha\mathcal{R}(X) \leq \mathrm{E}Y + \alpha\mathcal{R}(Y).$$

The definition is intuitive. We do not select dominated random variables by the weighted sum approach to the mean-risk problem. The α-consistency of a risk measure with some stochastic dominance order implies its α'-consistency with the same order for all α' with $0 < \alpha' \leq \alpha$, cf. Ogryczak and Ruszczynski (1999). A risk measure $\mathcal{R}$, which is α-consistent with stochastic dominance of order p for all $\alpha \in \mathbb{R}_+$, is called consistent with stochastic dominance of order p.

Application In the remainder of this section, we discuss the risk measures introduced in Section 2.2 w.r.t. to stochastic dominance. We note that the expected value of a random variable is consistent with first as well as second order stochastic dominance.

Lemma 2.9 *It holds* $X \preceq_{FSD(SSD)} Y \implies \mathrm{E}X \leq \mathrm{E}Y$.

Proof The function $f(x) = -x$ is nonincreasing and concave. Plugging this function into the Definitions 2.1 and 2.3 yields $\mathrm{E}X \leq \mathrm{E}Y$ in both cases. □

The central deviation of order p is neither consistent with first nor with second order stochastic dominance. We give an example.

Example 2.10 *Let X and Y be two random variables defined on the probability space $(\Omega, \mathcal{A}, \mathrm{I\!P})$ and let $k > 0$. Assume $q \in (0,1]$, $\mathrm{I\!P}(X = -k) = 1 - \mathrm{I\!P}(X = 0) = q$, and $Y \equiv 0$. Then, it holds*

$$\mathrm{I\!P}(X > t) = \mathrm{I\!P}(Y > t) \quad t \in \mathbb{R} \setminus [-k, 0)$$

and

$$\mathrm{I\!P}(X > t) = q \leq 1 = \mathrm{I\!P}(Y > t) \quad t \in [-k, 0).$$

Consequently, X dominates Y in first stochastic order and therefore also in second stochastic order, cf. Lemma 2.6. Now, we consider the mean-risk model for $p \in \mathbb{N}$ and $\alpha > 0$. We compute $\mathrm{E}X = -qk$ and

$$\mathcal{D}_p(X) = \big(\mathrm{E}|X - \mathrm{E}X|^p\big)^{\frac{1}{p}} = \big(q|-k+qk|^p + (1-q)|qk|^p\big)^{\frac{1}{p}} = \big(q(1-q)^p + q^p(1-q)\big)^{\frac{1}{p}} k$$

as well as $\mathrm{E}Y = \mathcal{D}_p(Y) = 0$. We show that there exists $\alpha > 0$ such that $\mathrm{E}X + \alpha\mathcal{D}_p(X) > \mathrm{E}Y + \alpha\mathcal{D}_p(Y)$ or equivalently, $f(q) := \big(q^{-p+1}(1-q)^p + (1-q)\big)^{\frac{1}{p}} > \frac{1}{\alpha}$. If $p > 1$ then $f(q)$ tends to ∞ as q tends to 0 and $f(q)$ tends to 0 as q tends to 1. Thus, for every α there exists $q \in (0,1]$ such that the central deviation of order p, $p > 1$, is not consistent with first and second stochastic order.

For $p = 1$ we obtain the inconsistency when the inequality $(1-q) > \frac{1}{2\alpha}$ holds true. This yields an counterexample for the consistency of the risk measure with first and second stochastic order for the central deviation of order 1 if $\alpha > \frac{1}{2}$.

The central deviation of order p is 1-consistent with the stochastic dominance of order $p+1$ if the distributions of the considered random variables are symmetric w.r.t. the mean, see Ogryczak and Ruszczynski (2001). In the following lemma we close the gap obtained in the above example for the central deviation of order 1.

Lemma 2.11 *$\mathcal{D}_1$ is $\frac{1}{2}$-consistent with first and second order stochastic dominance.*

Proof We reformulate the central deviation of order 1

$$\begin{aligned}\mathcal{D}_1(X) = \mathrm{E}|X - \mathrm{E}X| &= \mathrm{E}\big(\max\{X - \mathrm{E}X, 0\} + \max\{\mathrm{E}X - X, 0\}\big)\\ &= \mathrm{E}\max\{X - \mathrm{E}X, 0\} + \mathrm{E}\max\{\mathrm{E}X - X, 0\}\\ &= \mathrm{E}\max\{X, \mathrm{E}X\} - \mathrm{E}X + \mathrm{E}\big(\max\{\mathrm{E}X, X\} - X\big)\\ &= 2\mathrm{E}\max\{X, \mathrm{E}X\} - 2\mathrm{E}X.\end{aligned}$$

Thus, it holds

$$\mathcal{R}(X) := \mathrm{E}X + \alpha\mathcal{D}_1(X) = (1-2\alpha)\mathrm{E}X + 2\alpha\mathrm{E}\max\{X, \mathrm{E}X\}.$$

The expected value is consistent with first and second order stochastic dominance, cf. Lemma 2.9. We show the consistency of the second term. Assume X dominates Y in first (second) stochastic order. This implies $\mathrm{E}X \leq \mathrm{E}Y$ and consequently $\mathrm{E}\max\{X, \mathrm{E}X\} \leq \mathrm{E}\max\{X, \mathrm{E}Y\}$. Since the function $f(x) := \max\{x, \mathrm{E}Y\}$ is nondecreasing and convex, we obtain by $X \preceq_{\text{FSD(SSD)}} Y$ the relation

$$\mathrm{E}\max\{X, \mathrm{E}Y\} \leq \mathrm{E}\max\{Y, \mathrm{E}Y\}.$$

Thus, the risk measure $\mathcal{R}'(X) := \mathrm{E}\max\{X, \mathrm{E}X\}$ is consistent with first as well as second order stochastic dominance. Therefore, it also holds $\mathcal{R}(X) \leq \mathcal{R}(Y)$ when X dominates Y in first or second stochastic order and when $1 - \alpha \geq 0$, i.e. , when $\alpha \leq \frac{1}{2}$. □

For the remaining risk measures we cite two results.

Lemma 2.12 (Ogryczak and Ruszczynski (2001), Theorem 2) *The semideviation of order p is 1-consistent with stochastic dominance of order* $1, \ldots, p+1$. □

Lemma 2.13 (Fishburn (1977), Theorem 3) *For all* $p \in \mathbb{N}$*, the expected excess of order p is consistent with first and second order stochastic dominance.* □

The statements are as general as required in our context. Consult Lemma 2.4 to see the close relation of the expected excess of order 1 and second order stochastic dominance. The semideviation of order p is 1-consistent but not α-consistent with stochastic dominance of order $1, \ldots, p+1$ for $\alpha > 1$. This can be seen using Example 2.10.

Example 2.14 *Let* $p \in \mathbb{N}$*. We leave the specifications of Example 2.10 unchanged and calculate* $\mathcal{D}_p^+(X) = qk\sqrt[p]{1-q}$*. Then the inequality* $\mathrm{E}X + \alpha\mathcal{D}_p(X) > \mathrm{E}Y + \alpha\mathcal{D}_p(Y)$ *is fulfilled if* $\sqrt[p]{1-q}$ *is greater than* $\frac{1}{\alpha}$*. For* $\alpha > 1$*, this holds when* q *is greater than* $\frac{\alpha-1}{\alpha}$*. Since* X *dominates* Y *in first stochastic order,* $\mathcal{D}_p^+$ *can not be* α*-consistent with stochastic dominance of order* p *for* $\alpha > 1$ *and* $p \in \mathbb{N}$*.*

We summarize some of the results of this section in Table 2.2. The risk measures are consistent with first and second order stochastic dominance for the displayed values of α and p. We have given counterexamples for the remaining weights and orders.

Risk Measure	Weight	Order
$\mathcal{D}_\mathrm{p}$	$0 < \alpha \leq \frac{1}{2}$	$p = 1$
$\mathcal{D}_\mathrm{p}^+$	$0 < \alpha \leq 1$	$p \in \mathbb{N}$
$\mathrm{E}_\mathrm{p}^\eta$	$\alpha \in \mathbb{R}_+$	$p \in \mathbb{N}$

Table 2.2: Consistency of the risk measures with stochastic dominance

2.4 Convex and coherent measures of risk

Introduction In this section we discuss the conformity of the listed risk measures with some axiomatic properties that have recently attracted some attention in the literature. The class of coherent risk measures as defined by Artzner et al. (1999) is motivated by the needs of financial markets.

Definition 2.15 (coherent risk measure) *Let X and Y belong to some family of random variables on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We call a risk measure $\mathcal{R}$ coherent if it fulfills the four axioms*

- *Translation invariance:* $\mathcal{R}(X + C) = \mathcal{R}(X) + C, \quad \forall C \in \mathbb{R}$,
- *Positive homogeneity:* $\mathcal{R}(\lambda X) = \lambda \mathcal{R}(X), \quad \forall \lambda \in \mathbb{R}_+$,
- *Subadditivity:* $\mathcal{R}(X + Y) \leq \mathcal{R}(X) + \mathcal{R}(Y)$,
- *Monotonicity:* $X \leq Y$ *a.s.* $\implies \mathcal{R}(X) \leq \mathcal{R}(Y)$.

In the next section the following relation will turn out useful.

Lemma 2.16 $X \leq Y$ *a.s.* $\implies X \preceq_{FSD} Y$.

Proof Let $X \leq Y$ *a.s.* This means $\mathbb{P}(X > Y) = 0$. Let $t \in \mathbb{R}$. We have $\mathbb{P}(X > t) = \mathbb{P}(Y \geq X > t) + \mathbb{P}(X > Y > t) + \mathbb{P}(X > t \geq Y)$ and $\mathbb{P}(Y > t) = \mathbb{P}(Y \geq X > t) + \mathbb{P}(X > Y > t) + \mathbb{P}(Y > t \geq X)$. Our assumption implies that the probabilities $\mathbb{P}(X > Y > t)$ and $\mathbb{P}(X > t \geq Y)$ are zero. Therefore, it holds

$$\mathbb{P}(X > t) = \mathbb{P}(Y \geq X > t) \leq \mathbb{P}(Y \geq X > t) + \mathbb{P}(Y > t \geq X) = \mathbb{P}(Y > t).$$

Since $t \in \mathbb{R}$ was chosen arbitrarily, the latter inequality verifies $X \preceq_{\text{FSD}} Y$. □

The class of convex risk measures as defined in Föllmer and Schied (2003) drops the properties of positive homogeneity and subadditivity and adds the weaker property of convexity instead.

Definition 2.17 (convex risk measure) *Let X and Y belong to some family of random variables on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We refer to a risk measure $\mathcal{R}$ as convex if it fulfills the axioms*

- *Translation invariance:* $\mathcal{R}(X + C) = \mathcal{R}(X) + C, \quad \forall C \in \mathbb{R}$,
- *Convexity:* $\mathcal{R}(\lambda X + (1 - \lambda) Y) \leq \lambda \mathcal{R}(X) + (1 - \lambda) \mathcal{R}(Y) \qquad \forall \lambda \in [0, 1]$,
- *Monotonicity:* $X \leq Y$ *a.s.* $\implies \mathcal{R}(X) \leq \mathcal{R}(Y)$.

Since convexity follows from positive homogeneity and subadditivity, coherent risk measures are convex risk measures, too.

Application Now we shall show whether the risk measures listed in Section 2.2 are convex and/or coherent. We note that the expected value is coherent because the integration is a linear and monotonous operation.

For the central deviation of order p we reconsider Example 2.10.

Example 2.18 *Given the specifications of Example 2.10 it also holds* $P(X > Y) = 0$. *However, we obtain* $\mathcal{D}_p(X) > 0 = \mathcal{D}_p(Y)$ *and* $\mathcal{R}(X) := \mathrm{E}X + \alpha\mathcal{D}_p(X) > \mathcal{R}(Y) := \mathrm{E}Y + \alpha\mathcal{D}_p(Y)$ *for all* $p \in \mathbb{N}$ *by chosing the probability* q *appropriately (with the restriction* $\alpha > \frac{1}{2}$ *if* $p = 1$*), cf. Example 2.10. Thus, the central deviation of order* p *and the composite risk measure* $\mathcal{R}$ *do not fulfill the monotonicity axiom and are therefore neither convex nor coherent risk measures.*

Again, the central deviation of order 1 plays a special role. The compound object

$$\mathcal{R}(X) := \mathrm{E}X + \alpha\mathcal{D}_1(X) = (1 - 2\alpha)\mathrm{E}X + 2\alpha\mathrm{E}\max\{X, \mathrm{E}X\}$$

fulfills the monotonicity axiom for $\alpha \in (0, \frac{1}{2}]$, cf. the Lemmas 2.11 and 2.16. For the term $\mathrm{E}\max\{X, \mathrm{E}X\}$, the remaining axioms follow from the positive homogeneity and the subadditivity of the maximum. Thus, $\mathcal{R}$ is a convex and coherent risk measure for $\alpha \in (0, \frac{1}{2}]$.

The expected excess of a target η_o of order p is monotonous, because it holds

$$X \le Y \ a.s. \implies X \preceq_{\mathrm{FSD}} Y \implies \mathrm{E}\max\{X - \eta_o, 0\}^p \le \mathrm{E}\max\{Y - \eta_o, 0\}^p,$$

cf. Lemma 2.16 and note that $f(x) := \max\{x - \eta_o, 0\}^p$ is nondecreasing for all $\eta_o \in \mathbb{R}$ and $p \in \mathbb{N}$. However, the expected excess of a target is not coherent. The fixed target η needs to be adjusted when turning from one random variable to another. This only amounts to a concept similar to coherent risk measures if we consider the risk measure as a function of η. Let $C, \eta \in \mathbb{R}$, $\lambda \in \mathbb{R}_+$, and $\rho \in [0, 1]$. We define

- Translation invariance w.r.t. fixed targets: $\mathcal{R}_\eta(X + C) = \mathcal{R}_{\eta - C}(X)$,
- Positive homogeneity w.r.t. fixed targets: $\mathcal{R}_\eta(\lambda X) = \lambda\mathcal{R}_{\frac{\eta}{\lambda}}X$,
- Subadditivity w.r.t. fixed targets: $\mathcal{R}_\eta(X + Y) \le \mathcal{R}_{\rho\eta}X + \mathcal{R}_{\eta - \rho\eta}Y$.

Then, a risk measure is *coherent w.r.t. fixed targets* when it fulfills the above three axioms together with the monotonicity axiom. In this sense, the expected excess of a target η is coherent w.r.t. fixed targets. In addition, the expected excess of a target fulfills the convexity axiom. We directly verify this by the convexity inequality. Let $\lambda_1, \lambda_2 \ge 0$, $\lambda_1 + \lambda_2 = 1$. We use $\max\{a + b, 0\} \le$

$\max\{a,0\}+\max\{b,0\}$ for $a,b\in\mathbb{R}$. It holds

$$\begin{aligned}\mathrm{E}_{\mathrm{p}}^{\eta}(\lambda_1 X+\lambda_2 Y) &= \big(\mathrm{E}\max\{\lambda_1(X-\eta)+\lambda_2(Y-\eta),0\}^p\big)^{\frac{1}{p}}\\ &\le \big(\mathrm{E}(\max\{\lambda_1(X-\eta),0\}+\max\{\lambda_2(Y-\eta),0\})^p\big)^{\frac{1}{p}}\\ &\le \big(\mathrm{E}\max\{\lambda_1(X-\eta),0\}^p\big)^{\frac{1}{p}}+\big(\mathrm{E}\max\{\lambda_2(Y-\eta),0\}^p\big)^{\frac{1}{p}}\\ &= \lambda_1\mathrm{E}_{\mathrm{p}}^{\eta}(X)+\lambda_2\mathrm{E}_{\mathrm{p}}^{\eta}(Y),\end{aligned}$$

where the second estimate is due to the Minkowski inequality, see Hardy et al. (1934), Theorem 198.

The semideviation of order p is not translation invariant, since it holds $\mathcal{D}_{\mathrm{p}}^{+}(X+C)=\mathcal{D}_{\mathrm{p}}^{+}(X)$. It is also not monotonous, cf. the Examples 2.14 and 2.18. However, the composite risk measure $\mathcal{R}(X):=\mathrm{E}X+\alpha\mathcal{D}_{\mathrm{p}}^{+}(X)$ is coherent for $\alpha\in(0,1]$, cf. Remark 1 in Ogryczak and Ruszczynski (2002). The absolute semideviation fulfills the convexity axiom too. We prove the coherence of the composite risk measure

$$\mathcal{R}(X):=\mathrm{E}X+\alpha\mathcal{D}_{\mathrm{p}}^{+}X=\mathrm{E}X+\alpha\big(\mathrm{E}\max\{X-\mathrm{E}X,0\}^p\big)^{\frac{1}{p}}\qquad p\in\mathbb{N},\tag{2.4}$$

in the next lemma.

Lemma 2.19 *Let $\alpha\in(0,1]$ and $p\in\mathbb{N}$. The risk measure $\mathcal{R}$ as defined in equation* (2.4) *is a coherent risk measure. Moreover, the risk measure $\mathcal{D}_p^{+}$ fulfills the convexity axiom.*

Proof We start the proof by showing the translation invariance of the risk measure $\mathcal{R}$. It holds $\mathrm{E}(X+C)=C+\mathrm{E}X$ and

$$\mathcal{D}_{\mathrm{p}}^{+}(X+C)=\big(\mathrm{E}\max\{X+C-(C+\mathrm{E}X),0\}^p\big)^{\frac{1}{p}}=\mathcal{D}_{\mathrm{p}}^{+}X,$$

which gives $\mathcal{R}(X+C)=\mathrm{E}(X+C)+\alpha\mathcal{D}_{\mathrm{p}}^{+}(X+C)=C+\mathrm{E}X+\alpha\mathcal{D}_{\mathrm{p}}^{+}X=C+\mathcal{R}(X)$ for $C\in\mathbb{R}$. The positive homogeneity of $\mathcal{R}$ follows from $\mathrm{E}(\lambda X)=\lambda\mathrm{E}X$ and

$$\begin{aligned}\mathcal{D}_{\mathrm{p}}^{+}(\lambda X) &= \big(\mathrm{E}\max\{\lambda X-\mathrm{E}(\lambda X),0\}^p\big)^{\frac{1}{p}}\\ &= \big(\mathrm{E}\lambda^p(\max\{X-\mathrm{E}X,0\})^p\big)^{\frac{1}{p}}\\ &= \lambda\big(\mathrm{E}\max\{X-\mathrm{E}X,0\}^p\big)^{\frac{1}{p}} = \lambda\mathcal{D}_{\mathrm{p}}^{+}X\end{aligned}\tag{2.5}$$

for $\lambda\in\mathbb{R}_{+}$. In order to show the subadditivity we employ the relation $\max\{a+b,0\}\le\max\{a,0\}+\max\{b,0\}$, again. It holds

$$\max\{X+Y-\mathrm{E}X-\mathrm{E}Y,0\}\le\max\{X-\mathrm{E}X,0\}+\max\{Y-\mathrm{E}Y,0\}$$

and consequently,

$$\begin{aligned}\mathcal{D}_{\mathrm{p}}^{+}(X+Y) &\le \big(\mathrm{E}(\max\{X-\mathrm{E}X,0\}+\max\{Y-\mathrm{E}Y,0\})^p\big)^{\frac{1}{p}}\\ &\le \big(\mathrm{E}\max\{X-\mathrm{E}X,0\}^p\big)^{\frac{1}{p}}+\big(\mathrm{E}\max\{Y-\mathrm{E}Y,0\}^p\big)^{\frac{1}{p}}\\ &= \mathcal{D}_{\mathrm{p}}^{+}X+\mathcal{D}_{\mathrm{p}}^{+}Y,\end{aligned}\tag{2.6}$$

where the second estimation is due to the Minkowski inequality. By the equations (2.5) and (2.6), the risk measure $\mathcal{D}_\mathrm{p}^+$ is positive homogenous and subadditive. Thus, the risk measure fulfills the convexity axiom.

The monotonicity is a consequence of the 1-consistency of the semideviation of order p with the stochastic dominance of order $p+1$, see Theorem 2 in Ogryczak and Ruszczynski (2001). By Lemma 2.16, it holds

$$X \leq Y a.s. \implies X \preceq_{\mathrm{FSD}} Y. \tag{2.7}$$

The first order stochastic dominance is sufficient for the stochastic dominance of order $p \geq 1$,

$$X \preceq_{\mathrm{FSD}} Y \implies X \preceq_{p+1} Y, \tag{2.8}$$

cf. Lemma 2.6. The semideviation of order p is 1-consistent with the stochastic dominance of order $p+1$, i.e. we have

$$X \preceq_{p+1} Y \implies \mathcal{R}(X) \leq \mathcal{R}(Y). \tag{2.9}$$

Together, the implications (2.7), (2.8), and (2.9) yield the monotonicity axiom. Thus, the risk measure $\mathcal{R}$ is coherent. □

In Table 2.3, we summarize those results of this section that will be used later on. The results are valid for the compound objects $\mathrm{E}X + \alpha\mathcal{R}(X)$ associated with the risk measure $\mathcal{R}$ in the first column.

The second column of the table displays whether these compound objects fulfill the monotonicity axiom for the values of α and p in the last two columns. The column 'Coherent RM' has a '+'-mark if the compound object $\mathrm{E}X+\alpha\mathcal{R}(X)$ belongs to the class of coherent risk measures. Again, this is restricted to the values of α and p in the last two columns.

The second row, for instance, documents that the risk measure $\mathcal{R}(X) := \mathrm{E}X + \alpha\mathcal{D}_\mathrm{p}^+(X)$ fulfills the monotonicity axiom and that it is coherent for all $\alpha \in (0,1]$ and all $p \in \mathbb{N}$.

Risk Measure	Monotone	Coherent RM	Weight	Order
$\mathcal{D}_\mathrm{p}(X)$	+	+	$0 < \alpha \leq \frac{1}{2}$	$p = 1$
$\mathcal{D}_\mathrm{p}^+(X)$	+	+	$0 < \alpha \leq 1$	$p \in \mathbb{N}$
$\mathrm{E}_\mathrm{p}^\eta(X)$	+	(−)	$\alpha \in \mathbb{R}_+$	$p \in \mathbb{N}$

Table 2.3: Coherent risk measures

We remark that the risk measures in column one fulfill the convexity axiom for all $p \in \mathbb{N}$.

2.5 Mean-risk models

The mean-risk model dates back to Markowitz (1959). It is the following bicriteria optimization problem

$$\inf_{x \in X} [\mathrm{E}Z(x), \mathcal{R}Z(x)], \tag{2.10}$$

where $\{Z(x) : x \in X\}$ is a family of random variables. As above, E and $\mathcal{R}$ denote the expected value and some measure of risk, respectively. We consider any Pareto optimal (efficient) point as a solution of problem (2.10).

Definition 2.20 (Pareto optimal or efficient solution) *A solution $x \in X$ is called Pareto optimal or efficient if there exists no other point $y \in X$ for which one pair or both pairs of the following inequalities hold:*

(i) $\mathrm{E}Z(y) < \mathrm{E}Z(x)$ *and* $\mathcal{R}Z(y) \leq \mathcal{R}Z(x)$,

(ii) $\mathrm{E}Z(y) \leq \mathrm{E}Z(x)$ *and* $\mathcal{R}Z(y) < \mathcal{R}Z(x)$.

The points $(\mathrm{E}Z(x), \mathcal{R}Z(x))$ in the image space associated with efficient solutions $x \in X$ are called *nondominated*. The set of efficient solutions is nonempty when, for instance, the set $X \neq \emptyset$ is compact and the functions $x \mapsto \mathrm{E}Z(x)$ and $x \mapsto \mathcal{R}Z(x)$ are lower semicontinuous (l.s.c.) on X, cf. Tanino and Kuk (2002).

The field of multicriteria optimization offers a number of scalarization methods to find efficient solutions, see Hwang and Masud (1979), Tanino and Kuk (2002), and the references therein. Among them, there is the weighted sum approach

$$\inf_{x \in X} \mathrm{E}Z(x) + \alpha \mathcal{R}Z(x) \qquad \alpha \in \mathbb{R}_+. \tag{2.11}$$

For any $\alpha > 0$, the problem (2.11) yields an efficient solution, see Korhonen (2001). If the set X is convex and both of the functions $x \mapsto \mathrm{E}Z(x)$ and $x \mapsto \mathcal{R}Z(x)$ are convex, then the weighted sum approach bears the potential to compute all efficient solutions as the parameter α is varied, cf. Proposition 3.8. in Ehrgott (2000).

However, in more general cases, for instance, when the functions $x \mapsto \mathrm{E}Z(x)$ and $x \mapsto \mathcal{R}Z(x)$ are nonconvex or even discontinuous, the set of nondominated solutions may be nonconvex and disconnected. Then, only a subset of the efficient set is computable by means of the weighted sum approach. Following Alves (2001), we refer to these solutions as *supported*.

In this thesis, we will not consider other approaches to compute efficient solutions. In particular, we will not cover the currently very popular approach of a Chebyshev-type scalarization Korhonen (2001), since this method conflicts with the decomposition algorithms we aim to apply to our special optimization problem, see Chapter 6.

It is straightforward how one should vary the parameter α in a sequence of problems of type (2.11). Given two nondominated solutions $(\mathrm{E}Z(x_1), \mathcal{R}Z(x_1))$ and $(\mathrm{E}Z(x_2), \mathcal{R}Z(x_2))$ with $\mathrm{E}Z(x_1) < \mathrm{E}Z(x_2)$ and $\mathcal{R}Z(x_1) > \mathcal{R}Z(x_2)$, the problem (2.11) with

$$\alpha = \frac{\mathrm{E}Z(x_2) - \mathrm{E}Z(x_1)}{\mathcal{R}Z(x_1) - \mathcal{R}Z(x_2)}$$

yields a nondominated solution $(\mathrm{E}Z(\bar{x}), \mathcal{R}Z(\bar{x}))$ with $\mathrm{E}Z(x_1) \leq \mathrm{E}Z(\bar{x}) \leq \mathrm{E}Z(x_2)$ and $\mathcal{R}Z(x_2) \leq \mathcal{R}Z(\bar{x}) \leq \mathcal{R}Z(x_1)$. This sketches the idea of an iterative algorithm. However, the algorithm may not produce all supported solutions if the solution of problem (2.11) with fixed α is not unique.

Initial efficient solutions can be obtained by minimizing the risk function $\mathcal{R}Z$ on the optimal set $\arg\min \mathrm{E}Z(x)$, and by minimizing the expected value function $\mathrm{E}Z$ on the optimal set $\arg\min \mathcal{R}Z(x)$, provided the optimal sets are nonempty.

3 Mean-risk models on families of random variables

3.1 Scope

In this chapter we leave the field of stochastic programming with recourse and provide a more general view on the underlying decision problem. In Chapter 1 we have observed that we can regard stochastic programming with recourse as an ordering problem on a family of random variables $\mathcal{Z} := \{Z(x, \xi(\omega)) := cx + \tilde{\phi}(x, \xi(\omega)) : x \in X\}$. The random cost function $Z(x, \xi(\omega))$ depends on a decision vector x as well as on a random variable $\xi : \Omega \to \mathbb{R}^l$. The random variables $Z(x, \xi(\omega))$ in $\mathcal{Z}$ are linked via the set of feasible decisions X, the linear function c, and the value function $\tilde{\phi}$. As we will see in Chapter 5, $\tilde{\phi}$ is merely lower semicontinuous (l.s.c.) in x when integer variables are present. Here we aim at studying some scalar characteristics of the family $\mathcal{Z}$ when the function linking the random variables is l.s.c.

In the following chapters we will apply the obtained results to stochastic programs with recourse. The results are also valid for other stochastic decision problems that fit in the above configuration. This more general approach is mainly motivated by the fact that leaving the field of stochastic programming reveals the properties essential to optimize on the family $\mathcal{Z}$ w.r.t. some (partial) order.

In this thesis, we are concerned with partial orders on families of random variables defined by mean-risk models. Mean-risk models are bicriteria optimization problems

$$\inf_{x \in X} \big(\mathrm{E}[Z(x, \xi(\omega))], \mathcal{R}[Z(x, \xi(\omega))]\big), \tag{3.1}$$

where $\mathrm{E}[Z(x, \xi(\omega))]$ is the expected value $\int_\Omega Z(x, \xi(\omega)) \, \mathbb{P}(d\omega)$ and $\mathcal{R}$ is some measure of risk. The weighted sum approach

$$\inf_{x \in X} \mathrm{E}[Z(x, \xi(\omega))] + \alpha \mathcal{R}[Z(x, \xi(\omega))] \qquad \alpha > 0 \tag{3.2}$$

yields efficient points of the problem (3.1) and will be studied throughout, cf. Section 2.5. When we consider continuity properties which are important for optimization it will be more convenient to use the notion of functions $Q_{\mathrm{E}}(x) := \mathrm{E}[Z(x, \xi(\omega))]$ and $Q_{\mathcal{R}}(x) := \mathcal{R}[Z(x, \xi(\omega))]$ mapping from $\mathbb{R}^n$ to $\mathbb{R}$.

Let us fix some more notations first. We will sometimes write $Z_x(\xi(\omega))$ instead of $Z(x,\xi(\omega))$ as well as $Z_k(\xi(\omega))$ instead of $Z(x_k,\xi(\omega))$. We state the results in terms of the image measure $\mu := \mathbb{P}\circ\xi^{-1}$, a probability measure on the Borel measure space $(\mathbb{R}^l,\mathcal{B}^l)$.

In Table 3.1 we have listed the risk functions under consideration, cf. Section 2.2 for the corresponding risk measures.

Name	Symbol	Definition $(\eta_o\in\mathbb{R},\ p=1,2,\ldots)$
Expected Value	$Q_{\mathrm{E}}(x)$	$\int_{\mathbb{R}^l} Z(x,\xi)\ \mu(d\xi)$
Central deviation	$Q_{\mathcal{D}_\mathrm{p}}(x)$	$\left(\int_{\mathbb{R}^l} \lvert Z(x,\xi)-Q_{\mathrm{E}}(x)\rvert^p\ \mu(d\xi)\right)^{\frac{1}{p}}$
Semideviation	$Q_{\mathcal{D}^+_\mathrm{p}}(x)$	$\left(\int_{\mathbb{R}^l} \max\{Z(x,\xi)-Q_{\mathrm{E}}(x),0\}^p\,\mu(d\xi)\right)^{\frac{1}{p}}$
Expected excess	$Q_{\mathrm{E}^\eta_\mathrm{p}}(x)$	$\left(\int_{\mathbb{R}^l} \max\{Z(x,\xi)-\eta_o,0\}^p\ \mu(d\xi)\right)^{\frac{1}{p}}$

Table 3.1: Expected value function and risk functions

3.2 Structure

First, we address the conditions under which the risk functions are well-defined. Therefore we define the set of p-integrable functions

$$\mathcal{L}^p := \{g:\mathbb{R}^l\to\mathbb{R}\ :\ \int_{\mathbb{R}^l}\lvert g(\xi)\rvert^p\ \mu(d\xi)<\infty\}$$

for a fixed probability measure μ.

Proposition 3.1 *Let $x_o\in\mathbb{R}^n$ and $\eta_o\in\mathbb{R}$. Assume $Z(x_o,\cdot)$ is a μ-measurable function. Then*

(i) *$Q_{\mathrm{E}}(x_o)$ is real-valued if $Z(x_o,\cdot)$ is in $\mathcal{L}^1$, and*

(ii) *$Q_{\mathcal{D}_p}(x_o)$, $Q_{\mathcal{D}_p^+}(x_o)$, and $Q_{\mathrm{E}_p^\eta}(x_o)$ are real-valued if $Z(x_o,\cdot)$ is in $\mathcal{L}^p$.*

Proof The composition of functions is measurable if all the components are. Thus, the μ-measurability of the integrands $\lvert Z(x_o,\cdot)-Q_{\mathrm{E}}(x_o)\rvert^p$, $\max\{Z(x_o,\cdot)-Q_{\mathrm{E}}(x_o),0\}^p$, and $\max\{Z(x_o,\cdot)-\eta_o,0\}^p$ follows from the μ-measurability of $Z(x_o,\cdot)$ and $x\mapsto\max\{f(\xi),g(\xi)\}$ with measurable functions f and g.
ad (i) The finiteness of the expected value function follows from the assumption $Z(x_o,\cdot)\in\mathcal{L}^1$. It holds

$$\lvert Q_{\mathrm{E}}(x_o)\rvert \le \int_{\mathbb{R}^l}\lvert Z(x_o,\xi)\rvert\ \mu(d\xi)<\infty.$$

ad (ii) To see the finiteness of the central deviation we use the estimation $|Z(x_o,\xi) - Q_{\mathrm{E}}(x_o)|^p \le 2^p|Z(x_o,\xi)|^p + 2^p|Q_{\mathrm{E}}(x_o)|^p$ for all $\xi \in \mathbb{R}^l$ and the Minkowski inequality. We obtain

$$\begin{aligned}
|Q_{\mathcal{D}_{\mathrm{p}}}(x_o)| &\le \Big(\int_{\mathbb{R}^l} |Z(x_o,\xi) - Q_{\mathrm{E}}(x_o)|^p \ \mu(d\xi)\Big)^{\frac{1}{p}} \\
&\le 2\Big(\int_{\mathbb{R}^l} |Z(x_o,\xi)|^p \ \mu(d\xi)\Big)^{\frac{1}{p}} + 2\Big(\int_{\mathbb{R}^l} |Q_{\mathrm{E}}(x_o)|^p \ \mu(d\xi)\Big)^{\frac{1}{p}} \\
&= 2\Big(\int_{\mathbb{R}^l} |Z(x_o,\xi)|^p \ \mu(d\xi)\Big)^{\frac{1}{p}} + 2|Q_{\mathrm{E}}(x_o)|.
\end{aligned}$$

By assumption we have $\int_{\mathbb{R}^l} |Z(x_o,\xi)|^p \ \mu(d\xi) < \infty$ and the assertion (i) yields $|Q_{\mathrm{E}}(x_o)| < \infty$. This proves $|Q_{\mathcal{D}_{\mathrm{p}}}(x_o)| < \infty$.

We can estimate the semideviation of order p as follows

$$\begin{aligned}
|Q_{\mathcal{D}_{\mathrm{p}}^+}(x_o)| &= \Big(\int_{\mathbb{R}^l} \max\{Z(x_o,\xi) - Q_{\mathrm{E}}(x_o), 0\}^p \ \mu(d\xi)\Big)^{\frac{1}{p}} \\
&= \Big(\int_{M} |Z(x_o,\xi) - Q_{\mathrm{E}}(x_o)|^p \ \mu(d\xi)\Big)^{\frac{1}{p}} \\
&\le \Big(\int_{\mathbb{R}^l} |Z(x_o,\xi) - Q_{\mathrm{E}}(x_o)|^p \ \mu(d\xi)\Big)^{\frac{1}{p}},
\end{aligned}$$

where $M := \{\xi \in \mathbb{R}^l : Z(x,\xi) \ge Q_{\mathrm{E}}(x_o)\}$ and where we use the monotonicity of the function $x \mapsto \sqrt[p]{x}$. In the estimation of $Q_{\mathcal{D}_{\mathrm{p}}}(x_o)$ we have shown that $\big(\int_{\mathbb{R}^l} |Z(x_o,\xi) - Q_{\mathrm{E}}(x_o)|^p \ \mu(d\xi)\big)^{\frac{1}{p}}$ is finite. This yields $Q_{\mathcal{D}_{\mathrm{p}}^+}(x_o) \in \mathbb{R}$.

Finally, we show the finiteness of the expected excess of a target η_o of order p by means of the estimation $|Z(x,\xi) - \eta_o|^p \le 2^p|Z(x,\xi)|^p + 2^p|\eta_o|^p$ and the Minkowski inequality

$$\begin{aligned}
|Q_{\mathrm{E}_{\mathrm{p}}^{\eta}}(x_o)| &\le \Big(\int_{M} |Z(x,\xi) - \eta_o|^p \ \mu(d\xi)\Big)^{\frac{1}{p}} \\
&\le \Big(\int_{\mathbb{R}^l} |Z(x,\xi) - \eta_o|^p \ \mu(d\xi)\Big)^{\frac{1}{p}} \\
&\le 2\Big(\int_{\mathbb{R}^l} |Z(x,\xi)|^p \ \mu(d\xi)\Big)^{\frac{1}{p}} + 2|\eta_o|,
\end{aligned}$$

where $M := \{\xi \in \mathbb{R}^l : Z(x,\xi) \ge \eta_o\}$. Thus, the expected excess of order p is also finite under the posed assumption and the proof is complete. □

According to Weierstrass' theorem, a lower semicontinuous function attains its minimum on each nonempty compact set. So, lower semicontinuity is an essential property for minimization problems. The question arises whether the lower semicontinuity of the random variables $Z(\cdot,\xi)$ is preserved by a risk function. This carries over to the question as to whether the basic operations involved in the definitions of the risk function preserve this property.

Consider a function $f : \mathbb{R}^l \to \mathbb{R}$ which is l.s.c. but not continuous at t_o, i.e. $\liminf_{t\to t_o} f(t) > f(t_o)$. Since $\liminf_{t\to t_o} -f(t) < -f(t_o)$, the function $-f$ is upper semicontinuous (u.s.c.) but not l.s.c. at t. Likewise, consider a function $f : \mathbb{R}^l \to \mathbb{R}$ which is l.s.c. but not continuous on the set $T^- := \{t \in \mathbb{R}^l : f(t) < 0\} \neq \emptyset$, i.e. there exists $t \in T^-$ such that $\liminf_{t\to t_o} f(t) > f(t_o)$. Then the square $g := f^2$ of f is not l.s.c. at t if $|\liminf_{t\to t_o} f(t)| < |f(t_o)|$.

The first observation shows that, in general, the risk measures which involve the subtraction of a l.s.c. function are not l.s.c. This is the case for the central deviation and the semideviation of order $p \in \mathbb{N}$. Both risk functions involve the subtraction from the mean $Q_{\mathrm{E}}(x_o)$ which will be shown to be l.s.c. in Proposition 3.5. We will provide counterexamples for the lower semicontinuity of these measures in Chapter 5. However, for the semideviation of order 1 we can fix this defect in a mean-risk model when α is in $(0, 1]$.

Before we prove the lower semicontinuity of the risk functions we state two lemmas.

Lemma 3.2 *Let $f, g : \mathbb{R}^n \to \mathbb{R}$ and $x_o \in \mathbb{R}^n$. Assume f and g are l.s.c. at x_o. Then the function $x \mapsto \max\{f(x), g(x)\}$ is l.s.c. at x_o.*

Proof Let $\{x_k\}$ be a sequence in $\mathbb{R}^n$ converging to x_o. For all $k \in \mathbb{N}$ it holds $\max\{f(x_k), g(x_k)\} \geq f(x_k)$ and $\max\{f(x_k), g(x_k)\} \geq g(x_k)$. Thus, we have

$$\liminf_{k\to\infty} \max\{f(x_k), g(x_k)\} \geq \liminf_{k\to\infty} f(x_k),$$

$$\liminf_{k\to\infty} \max\{f(x_k), g(x_k)\} \geq \liminf_{k\to\infty} g(x_k),$$

and consequently

$$\liminf_{k\to\infty} \max\{f(x_k), g(x_k)\} \geq \max\{\liminf_{k\to\infty} f(x_k), \liminf_{k\to\infty} g(x_k)\}. \tag{3.3}$$

The functions f and g are l.s.c. at x_o which means $\liminf_{k\to\infty} f(x_k) \geq f(x_o)$ and $\liminf_{k\to\infty} g(x_k) \geq g(x_o)$. Therefore, the inequality (3.3) yields

$$\liminf_{k\to\infty} \max\{f(x_k), g(x_k)\} \geq \max\{f(x_o), g(x_o)\}. \tag{3.4}$$

Since the sequence $\{x_k\}$ was chosen arbitrarily, inequality (3.4) corresponds to the lower semicontinuity of the maximum. □

Lemma 3.3 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$. Assume f is l.s.c. at x_o and g is l.s.c. at $f(x_o)$ and nondecreasing. Then the function h defined as $h(x) := g(f(x))$ for all $x \in \mathbb{R}^n$ is l.s.c. at x_o.*

Proof Let $\{x_k\}$ be a sequence in $\mathbb{R}^n$ converging to x_o. Since g is l.s.c. at $f(x_o)$ we obtain $\liminf_{k\to\infty} g(f(x_k)) \geq g(\liminf_{k\to\infty} f(x_k))$. The lower semicontinuity of f yields $\liminf_{k\to\infty} f(x_k) \geq f(x_o)$. Hence, as g is nondecreasing we have

$$\liminf_{k\to\infty} g(f(x_k)) \geq g(f(x_o)).$$

This means h is l.s.c. at x_o. □

Remark 3.4 *The functions $x \mapsto x^p$ and $x \mapsto \sqrt[p]{x}$ are nondecreasing on $\mathbb{R}_+$. We can apply Lemma 3.3 to these functions if the inner function f is nonnegative and l.s.c. This holds true for $f(x) := \max\{h(x), 0\}$ with a l.s.c. function $h(x)$, cf. Lemma 3.2.*

Proposition 3.5 *Let $x_o \in \mathbb{R}^n$, $\eta_o \in \mathbb{R}$, $\alpha \in \mathbb{R}_+$, and $r, p \in \mathbb{N}$. Assume that there exists a neighborhood $U_\delta(x_o)$ of x_o with $\delta > 0$ such that $Z(x, \cdot)$ is a μ-measurable function for all $x \in U_\delta(x_o)$ and that $Z(., \xi)$ is l.s.c. at x_o for all $\xi \in \mathbb{R}^l$. Assume further that $Z(x, \cdot) \in \mathcal{L}^r$ for all $x \in U_\delta(x_o)$ and that there is a μ-measurable function $g \in \mathcal{L}^1$ such that $g(\xi) \leq Z(x, \xi)$ for all $\xi \in \mathbb{R}^l$ and all $x \in U_\delta(x_o)$. Then*

(i) $Q_{\mathrm{E}} : \mathbb{R}^n \to \mathbb{R}$ *is l.s.c. at x_o,*

(ii) $Q_{\mathrm{E}} + \alpha Q_{\mathcal{D}_1} : \mathbb{R}^n \to \mathbb{R}$ *is l.s.c. at x_o for all $\alpha \in (0, \frac{1}{2}]$,*

(iii) $Q_{\mathrm{E}} + \alpha Q_{\mathcal{D}_1^+} : \mathbb{R}^n \to \mathbb{R}$ *is l.s.c. at x_o for all $\alpha \in (0, 1]$, and*

(iv) $Q_{\mathrm{E}_p^\eta} : \mathbb{R}^n \to \mathbb{R}$ *is l.s.c. at x_o if it holds $r \geq p$.*

Proof In the proof we use Fatou's Lemma, see Billingsley (1995). To apply the lemma we need an integrable minorant of the integrand. Let $\{x_k\}$ be a sequence converging to x_o. Without loss of generality we can assume $\{x_k\} \subset U_\delta(x_o)$.
ad (i) Together Fatou's Lemma and the lower semicontinuity of $Z(., \xi)$ at x_o yield

$$\liminf_{k\to\infty} Q_{\mathrm{E}}(x_k) \geq \int_{\mathbb{R}^l} \liminf_{k\to\infty} Z(x_k, \xi)\ \mu(d\xi) \geq \int_{\mathbb{R}^l} Z(x_o, \xi)\ \mu(d\xi) = Q_{\mathrm{E}}(x_o),$$

whereby a minorant g of $Z(., \xi)$ exists by assumption. Thus the expected value function is l.s.c. at x_o.
ad (ii) To verify the lower semicontinuity of the function $Q_{\mathrm{E}} + \alpha Q_{\mathcal{D}_1}$ at x_o for $\alpha \in (0, \frac{1}{2}]$ we once again employ the reformulation $\mathcal{D}_1 X = 2\mathrm{E}\max\{X, \mathrm{E}X\} - 2\mathrm{E}X$, cf. Lemma 2.11. We obtain

$$Q_{\mathrm{E}}(x_o) + \alpha Q_{\mathcal{D}_1}(x_o) = (1 - 2\alpha) Q_{\mathrm{E}}(x_o) + 2\alpha \int_{\mathbb{R}^l} \max\{Z(x_o, \xi), Q_{\mathrm{E}}(x_o)\}\ \mu(d\xi)$$

The expected value function Q_{E} is l.s.c. at x_o by (i), the function $Z(\cdot,\xi)$ is l.s.c. at x_o by assumption, and so is also their maximum $x \mapsto \max\{Z(\cdot,\xi), Q_{\mathrm{E}}(\cdot)\}$, cf. Lemma 3.2. For $\alpha \in (0, \frac{1}{2}]$ the function $(1-2\alpha)Q_{\mathrm{E}}$ is also l.s.c. at x_o and this yields the assertion.

ad (iii) Let us turn to the semideviation of order 1 (absolute semideviation). We use $\max\{a,0\} = \max\{a+b,b\} - b$ for $a,b \in \mathbb{R}$ to reformulate

$$\begin{aligned} Q_{\mathrm{E}}(x_o) + \alpha Q_{\mathcal{D}_1^+}(x_o) &= Q_{\mathrm{E}}(x_o) + \alpha \int_{\mathbb{R}^l} \max\{Z(x_o,\xi) - Q_{\mathrm{E}}(x_o), 0\}\ \mu(d\xi) \\ &= (1-\alpha)Q_{\mathrm{E}}(x_o) + \alpha \int_{\mathbb{R}^l} \max\{Z(x_o,\xi), Q_{\mathrm{E}}(x_o)\}\ \mu(d\xi) \end{aligned}$$

Following the argumentation under (ii) we obtain that the sum of the functions $(1-\alpha)Q_{\mathrm{E}}(\cdot)$ and $\alpha\int_{\mathbb{R}^l} \max\{Z(\cdot,\xi), Q_{\mathrm{E}}(\cdot)\}$ is also l.s.c. at x_o for α in $(0,1]$. This verifies the lower semicontinuity of the absolute semideviation.

ad (iii) Finally, we come to the expected excess of the target η_o. The function $x \mapsto \max\{Z(x,\xi)-\eta_o, 0\}$ is l.s.c. at x_o since $Z(x,\xi)$ is l.s.c. at x_o, cf. Lemma 3.2. The p-th power of a function f is l.s.c. on the domain where the function f is nonnegative and l.s.c, cf. Remark 3.4. Therefore $x \mapsto \max\{Z(x,\xi) - \eta_o, 0\}^p$ is l.s.c. at x_o.

Since $g'(\xi) \equiv 0$ is a minorant for the integrand (we don't need the existence of the minorant g here), Fatou's Lemma yields

$$\begin{aligned} & \liminf_{k\to\infty} \int_{\mathbb{R}^l} \max\{Z(x_k,\xi) - \eta_o, 0\}^p\ \mu(d\xi) \\ \geq & \int_{\mathbb{R}^l} \liminf_{k\to\infty} \max\{Z(x_k,\xi) - \eta_o, 0\}^p\ \mu(d\xi) \\ \geq & \int_{\mathbb{R}^l} \max\{Z(x_o,\xi) - \eta_o, 0\}^p\ \mu(d\xi). \end{aligned}$$

The function $x \mapsto \int_{\mathbb{R}^l} \max\{Z(x,\xi) - \eta_o, 0\}^p$ is nonnegative. By Lemma 3.2 the lower semicontinuity of $x \mapsto \sqrt[p]{x}$ yields the assertion. □

As a matter of course, we are also interested in conditions under which the risk functions are continuous. We give a sufficient condition on the measure μ. Therefore we define the set of discontinuity points of $Z(\cdot,\xi)$ at x

$$\begin{aligned} D'_Z(x) &:= \{\xi \in \mathbb{R}^l : Z(\cdot,\xi) \text{ is discontinuous at } x\} \\ &= \{\xi \in \mathbb{R}^l : \exists\ \{x_k\} \subset \mathbb{R}^n,\ x_k \to x \ \text{ s.t. }\ Z(x_k,\xi) \not\to Z(x,\xi)\}. \end{aligned}$$

Note that the set D'_Z is measurable, cf. p. 225 in Billingsley (1968).

Proposition 3.6 *Let $x_o \in \mathbb{R}^n$ and $\eta_o \in \mathbb{R}$. Assume that there exists a neighborhood $U_\delta(x_o)$ of x_o with $\delta > 0$ such that $Z(x,\cdot)$ is a μ-measurable function for all $x \in U_\delta(x_o)$. Assume further that there is a μ-measurable function $g \in \mathcal{L}^1$ such that $|Z(x,\xi)| \leq g(\xi)$ for all $\xi \in \mathbb{R}^l$ and all $x \in U_\delta(x_o)$. Let $D'_Z(x_o)$ be a μ-null set. Then*

(i) $Q_{\mathrm{E}} : \mathbb{R}^n \to \mathbb{R}$ *is continuous at* x_o *and*

(ii) *if* g *is in* $\mathcal{L}^p$ *then* $Q_{\mathcal{D}_p}$, $Q_{\mathcal{D}_p^+}$, *and* $Q_{\mathrm{E}_p^\eta}$ *as functions from* $\mathbb{R}^n$ *to* $\mathbb{R}$ *are continuous at* x_o.

Proof The proof relies basically on Lebesgue's dominated convergence theorem, see Billingsley (1995). For its application it is necessary to show that the corresponding integrands converge μ-almost surely and that there exists an integrable majorant on the integrands of each risk function. Let $\{x_k\}$ be a sequence converging to x_o. W.l.o.g. we assume $\{x_k\} \subset U_\delta(x_o)$.
ad (i) By the assumption $\mu(D'_{\mathrm{Z}}(x_o)) = 0$, we have

$$\lim_{n\to\infty} Z(x_n, \xi) = Z(x_o, \xi) \qquad \forall \xi \in \mathbb{R}^l \setminus D'_{\mathrm{Z}}(x_o), \tag{3.5}$$

i.e. , μ almost sure convergence of $Z(x_n, \cdot)$ to $Z(x_o, \cdot)$. This, together with the assumed existence of a majorant $g \in \mathcal{L}^1$ yields

$$\lim_{k\to\infty} Q_{\mathrm{E}}(x_k) = \int_{\mathbb{R}^l} \lim_{k\to\infty} Z(x_k, \xi)\ \mu(d\xi) = \int_{\mathbb{R}^l} Z(x_o, \xi)\ \mu(d\xi) = Q_{\mathrm{E}}(x_o)$$

and thus, the continuity of Q_{E} at x_o.
ad (ii) By equation (3.5) and the continuity of Q_{E} at x_o we obtain

$$\lim_{k\to\infty} |Z(x_k, \xi) - Q_{\mathrm{E}}(x_k)| = |Z(x_o, \xi) - Q_{\mathrm{E}}(x_o)| \qquad \forall \xi \in \mathbb{R}^l \setminus D'_{\mathrm{Z}}(x_o).$$

Since $x \mapsto x^p$ is also continuous, we have the μ-almost sure convergence of the integrand of the central deviation of order p

$$\lim_{k\to\infty} |Z(x_k, \xi) - Q_{\mathrm{E}}(x_k)|^p = |Z(x_o, \xi) - Q_{\mathrm{E}}(x_o)|^p \qquad \forall \xi \in \mathbb{R}^l \setminus D'_{\mathrm{Z}}(x_o).$$

For the integrand it holds

$$|Z(x_k, \xi) - Q_{\mathrm{E}}(x_k)|^p \le 2^p\big(|Z(x_k, \xi)|^p + |Q_{\mathrm{E}}(x_k)|^p\big) \le 2^p\big(g(\xi)^p + |Q_{\mathrm{E}}(x_k)|^p\big)$$

for all $k \in \mathbb{N}$ and all $\xi \in \mathbb{R}^l$. By assumption we have $|Z(x_k, \xi)| \le g(\xi)$ for all $\xi \in \mathbb{R}^l$ and all $k \in \mathbb{N}$ thus we obtain

$$|Q_{\mathrm{E}}(x_k)| \le \int_{\mathbb{R}^l} |Z(x_k, \xi)|\ \mu(d\xi) \le \int_{\mathbb{R}^l} g(\xi)\ \mu(d\xi) \le C < \infty.$$

Since g is in $\mathcal{L}^p$, the function $2^p(g^p + C^p)$ is an integrable majorant for $|Z(x_k, \xi) - Q_{\mathrm{E}}(x_k)|^p$ for all $k \in \mathbb{N}$. Lebesgue's dominated convergence theorem gives

$$\begin{aligned} \lim_{k\to\infty} \int_{\mathbb{R}^l} |Z(x_k, \xi) - Q_{\mathrm{E}}(x_k)|^p\ \mu(d\xi) &= \int_{\mathbb{R}^l} \lim_{k\to\infty} |Z(x_k, \xi) - Q_{\mathrm{E}}(x_k)|^p\ \mu(d\xi) \\ &= \int_{\mathbb{R}^l} |Z(x_o, \xi) - Q_{\mathrm{E}}(x_o)|^p\ \mu(d\xi). \end{aligned}$$

Finally, the continuity of the function $x \mapsto \sqrt[p]{x}$ gives the continuity of $Q_{\mathcal{D}_\mathrm{p}}$ at x_o.

We turn to the semideviation of order p. It follows from the continuity of Q_E at x_o and the μ-almost sure convergence of $Z(x_n, \xi)$ that $\max\{Z(x_n, \xi) - Q_\mathrm{E}(x_n), 0\}^p$ converges μ-almost surely to $\max\{Z(x_o, \xi) - Q_\mathrm{E}(x_o), 0\}^p$ for all $p > 0$. We obtain an integrable majorant by the estimation

$$\begin{aligned} |\max\{Z(x_k,\xi) - Q_\mathrm{E}(x_k), 0\}| &\leq \max\{|Z(x_k,\xi) - Q_\mathrm{E}(x_k)|, 0\} \\ &\leq |Z(x_k,\xi)| + |Q_\mathrm{E}(x_k)| \\ &\leq g'(\xi) := g(\xi) + \int_{\mathbb{R}^l} g(\xi)\ \mu(d\xi) \end{aligned}$$

for all $\xi \in \mathbb{R}^l$ and all $k \in \mathbb{N}$. Since g' is a majorant of the nonnegative function $\max\{Z(x_k,\xi) - Q_\mathrm{E}(x_k), 0\}$, its power $(g')^p$ is also a majorant of $(\max\{Z(x_k,\xi) - Q_\mathrm{E}(x_k), 0\})^p$ for all $\xi \in \mathbb{R}^l$ and all $k \in \mathbb{N}$. The latter majorant is integrable by assumption. We apply Lebesgue's dominated convergence theorem

$$\begin{aligned} \lim_{k\to\infty} \int_{\mathbb{R}^l} &\max\{Z(x_k,\xi) - Q_\mathrm{E}(x_k), 0\}^p\ \mu(d\xi) = \\ &\int_{\mathbb{R}^l} \max\{Z(x_o,\xi) - Q_\mathrm{E}(x_o), 0\}^p\ \mu(d\xi), \end{aligned}$$

and by the continuity of $x \mapsto \sqrt[p]{x}$ we obtain the continuity of $Q_{\mathcal{D}_\mathrm{p}^+}$ at x_o.

Finally, we consider the expected excess. The μ-almost sure convergence of $\max\{Z(x_n,\xi) - \eta_o, 0\}$ follows from the continuity of $x \mapsto \max\{x, 0\}$ and the μ-almost sure convergence of $Z(x_n,\xi)$. We define $g'(\xi) := g(\xi) + |\eta_o|$ for all $\xi \in \mathbb{R}^l$. By assumption it holds $g'(\xi) \geq |Z(x_k,\xi)| + |\eta_o| \geq \max\{Z(x_k,\cdot) - \eta_o, 0\}$ for all $\xi \in \mathbb{R}^l$ and all $k \in \mathbb{N}$. Since $\max\{Z(x_k,\cdot) - \eta_o, 0\}$ is nonnegative, $(g')^p$ is a majorant of $\max\{Z(x_k,\cdot) - \eta_o, 0\}^p$ for all $\xi \in \mathbb{R}^l$ and all $k \in \mathbb{N}$. Lebesgue's dominated convergence theorem yields

$$\lim_{k\to\infty} \int_{\mathbb{R}^l} \max\{Z(x_k,\xi) - \eta_o, 0\}^p\ \mu(d\xi) = \int_{\mathbb{R}^l} \max\{Z(x_o,\xi) - \eta_o, 0\}^p\ \mu(d\xi)$$

and together with the continuity of $x \mapsto \sqrt[p]{x}$ we have the continuity of $Q_{\mathrm{E}_\mathrm{p}^\eta}$ at x_o. □

Remark 3.7 *If $Z(.,\xi)$ is continuous at x_o for all $\xi \in \mathbb{R}^l$, we have $D_Z^l(x_o) = \emptyset$ and Proposition 3.6 applies.*

When the decision problem is convex we are interested in using risk measures that preserve convexity, cf. also Rockafellar and Uryasev (2002). There is a strong relation between this question and the class of convex risk measures defined in Section 2.4. In fact, we consider risk functions which are composite mappings of type $Q = \mathcal{R} \circ \tilde{Z}$ where $\tilde{Z} : \mathbb{R}^n \to \mathcal{Z}$ assigns x to $Z(x, \xi(\omega))$ and where $\mathcal{R} : \mathcal{Z} \to \mathbb{R}$ is a risk measure. Provided $\tilde{Z}$ is convex, a sufficient condition for the convexity of the composite mapping Q is that the risk measure $\mathcal{R}$ is monotonous nondecreasing and convex. Then for $\lambda_1, \lambda_2 \in [0,1]$, $\lambda_1 + \lambda_2 = 1$ it holds

$$\mathcal{R}(\tilde{Z}(\lambda_1 x + \lambda_2 y)) \leq \mathcal{R}(\lambda_1 \tilde{Z}(x) + \lambda_2 \tilde{Z}(y)) \leq \lambda_1 \mathcal{R}(\tilde{Z}(x)) + \lambda_2 \mathcal{R}(\tilde{Z}(y)) \quad (3.6)$$

where the first inequality is due to the convexity of Z and the monotonicity of $\mathcal{R}$ and the second one is due to the convexity of $\mathcal{R}$.

Proposition 3.8 *Let $\eta_o \in \mathbb{R}$, $\alpha \in \mathbb{R}_+$, and $r \in \mathbb{N}$. Assume $Z(x,\cdot)$ is μ-measurable and element of $\mathcal{L}^r$ for all $x \in X$. Assume further that $Z(\cdot,\xi)$ is convex on X for all $\xi \in \mathbb{R}^l$. Then*

(i) $Q_{\mathrm{E}} : \mathbb{R}^n \to \mathbb{R}$ *is convex on X for $r \geq 1$,*

(ii) $Q_{\mathrm{E}} + \alpha Q_{\mathcal{D}_1} : \mathbb{R}^n \to \mathbb{R}$ *is convex on X for $\alpha \in (0, \frac{1}{2}]$ and $r \geq 1$,*

(iii) $Q_{\mathrm{E}} + \alpha Q_{\mathcal{D}_p^+} : \mathbb{R}^n \to \mathbb{R}$ *is convex on X for $\alpha \in (0,1]$ and $r \geq p$, and*

(iv) $Q_{\mathrm{E}_p^\eta} : \mathbb{R}^n \to \mathbb{R}$ *is convex on X for $r \geq p$.*

Proof In Section 2.4, we have seen that the expected value, the expected excess of a target, and the risk measures $\mathrm{E}X + \frac{\alpha}{2}\mathcal{D}_1(X)$ and $\mathrm{E}X + \alpha(\mathrm{E}\max\{X - \mathrm{E}X, 0\}^p)^{\frac{1}{p}}$ for $\alpha \in (0,1]$ fulfill the monotonicity as well as the convexity axiom, cf. in particular Lemma 2.19. Thus, the equation (3.6) is valid for the risk functions in (i)-(iv). □

In general, the risk functions $Q_{\mathcal{D}_\mathrm{p}}$ and $Q_{\mathcal{D}_\mathrm{p}^+}$ are not convex. We provide counterexamples in Chapter 4.

3.3 Stability

In this section, we will ask for the consequences of perturbations of the probability measure μ for the problem

$$\min_{x \in X} Q_{\mathrm{E}}(x,\mu) + Q_{\mathcal{R}}(x,\mu). \quad (3.7)$$

This requires to study the joint continuity of the functions Q_{E} and $Q_{\mathcal{R}}$ in x and μ, where $Q_{\mathcal{R}}$ is one of the risk functions listed in Table 3.1.

To this end we consider the set $\mathcal{P}(\mathbb{R}^l)$ of Borel probability measures on $\mathbb{R}^l$ endowed with the notion of weak convergence of probability measures.

Definition 3.9 (weak convergence of probability measures) *We say that a sequence* $\{\mu_k\} \subset \mathcal{P}(\mathbb{R}^l)$ *of probability measures converges weakly to* $\mu_o \in \mathcal{P}(\mathbb{R}^l)$ *as* k *tends to infinity* ($\mu_k \xrightarrow{w} \mu_o$) *if the equation*

$$\lim_{k\to\infty} \int_{\mathbb{R}^l} g(\xi)\ \mu_k(d\xi) = \int_{\mathbb{R}^l} g(\xi)\ \mu_o(d\xi)$$

is fulfilled for any bounded continuous function $g : \mathbb{R}^l \to \mathbb{R}$.

Note that the convergence of probability measures is equivalent to the convergence in distribution of the corresponding random variables ($Z_k \xrightarrow{D} Z_o$), i.e. to the pointwise convergence of the distribution functions F_k and F_o of Z_k and Z_o, respectively, in all continuity points of F_o.

We define the set of discontinuity points of $Z(\cdot,\cdot)$ by

$$\begin{aligned} D_Z(x) &:= \{\xi \in \mathbb{R}^l : Z(\cdot,\cdot) \text{ is discontinuous at } (x,\xi)\} \\ &= \{\xi \in \mathbb{R}^l : \text{there exist sequences } \{x_k\} \subset \mathbb{R}^n \text{ and } \{\xi_k\} \subset \mathbb{R}^l \text{ with} \\ &\qquad \lim_{k\to\infty} x_k = x,\ \lim_{k\to\infty} \xi_k = \xi,\ \text{but } \lim_{k\to\infty} Z(x_k,\xi_k) \neq Z(x,\xi)\}. \end{aligned}$$

The set D_Z is measurable, cf. p. 225 in Billingsley (1968). The relation of D_Z and the set D'_Z defined above will be important in the Chapters 4 and 5 when we turn to stochastic programming.

Lemma 3.10 *Let* $x \in \mathbb{R}^n$. *It holds* $D'_Z(x) \subset D_Z(x)$.

Proof Let $\xi \in D'_Z(x)$. Then, there exists $x_k \to x$ such that $Z(x_k,\xi) \not\to Z(x,\xi)$. Let $\{\xi_k\}$ be a sequence in $\mathbb{R}^l$ defined as $\xi_k := \xi$ for all $k \in \mathbb{N}$. We obtain $(x_k,\xi_k) \to (x,\xi)$ such that $Z(x_k,\xi_k) \not\to Z(x,\xi)$. Thus, ξ is in $D_Z(x)$. □

The next lemma uses different notations to the original theorem, namely $P_k = \mu_k$, $P = \mu_o$, $h_k = Z(x_k,\cdot)$, $h = Z(x_o,\cdot)$, and $E = D_Z(x)$, but otherwise the statements coincide. We write Z_k instead of $Z(x_k,\cdot)$ and Z_o instead of $Z(x_o,\cdot)$.

Lemma 3.11 (cf. Theorem 5.5, Billingsley (1968)) *Let* $x_o \in \mathbb{R}^n$ *and* $\mu_o \in \mathcal{P}(\mathbb{R}^l)$. *Assume that there is a neighborhood* $U_\delta(x_o)$ *of* x_o *with* $\delta > 0$ *such that* $Z(x,\cdot)$ *is a* μ*-measurable function for all* $x \in U_\delta(x_o)$ *and all* $\mu \in \mathcal{P}(\mathbb{R}^l)$. *Furthermore, assume* $D_Z(x_o)$ *has* μ_o*-measure zero.*

Then the sequence of induced measures $\mu_k \circ Z_k^{-1}$ *converges weakly to the induced measure* $\mu_o \circ Z_o^{-1}$ *for any sequences* $\{x_k\} \subset \mathbb{R}^n$ *converging to* x_o *and* $\{\mu_k\} \subset \mathcal{P}(\mathbb{R}^l)$ *converging weakly to* μ_o. □

The random variables $\xi_k, \xi_o : \Omega \to \mathbb{R}^l$ induce the probability measures $\mu_k := \mathbb{P} \circ \xi_k^{-1}$ and $\mu_o := \mathbb{P} \circ \xi_o^{-1}$, respectively. Lemma 3.11 states that under the posed assumptions it holds

$$\xi_k \xrightarrow{D} \xi_o \implies Z(x_k,\xi_k(\omega)) \xrightarrow{D} Z(x_o,\xi_o(\omega)).$$

The convergence in distributions of random variables is not sufficient for the convergence of the related expected values, see the counterexample on p. 158 in Stojanov (1987). The additional condition of uniform integrability of the sequence of random variables establishes the convergence of the expected values.

Definition 3.12 (uniformly integrable) *A family of functions $\mathcal{F}$ is called uniformly integrable w.r.t. the set $\Delta \subset \mathcal{P}(\mathbb{R}^l)$ of probability measures if it holds*

$$\lim_{a\to\infty} \sup_{Z\in\mathcal{F}} \int_{|Z(\xi)|\geq a} |Z(\xi)| \; \mu(d\xi) = 0 \qquad \forall \mu \in \Delta.$$

Equipped with this definition and the above lemma, we are able to state the joint continuity of the expected value functions.

Proposition 3.13 *Let $x_o \in \mathbb{R}^n$, $\Delta \subset \mathcal{P}(\mathbb{R}^l)$, and $\mu_o \in \Delta$. Assume there is a $\delta > 0$ such that $Z(x,\cdot)$ is a μ-measurable function for all $x \in U_\delta(x_o)$ and all $\mu \in \Delta$. Assume the family of functions $\{Z(x,\cdot) : x \in U_\delta(x_o)\}$ is uniformly integrable w.r.t. the set Δ. Assume furthermore $\mu_o(D_Z(x_o)) = 0$. Then $Q_{\mathrm{E}} : \mathbb{R}^n \times \Delta \to \mathbb{R}$ is continuous at (x_o, μ_o).*

Proof Let $\{x_k\} \subset U_\delta(x_o)$ be a sequence converging to x_o and $\{\mu_k\} \subset \Delta$ be a sequence converging weakly to μ_o. Then the assumption $\mu_o(D_Z(x_o)) = 0$ implies $\mu_k \circ Z_k^{-1} \xrightarrow{w} \mu_o \circ Z_o^{-1}$, see Lemma 3.11. The sequence $\{Z_k\}$ is uniformly integrable by assumption and Theorem 5.4 in Billingsley (1968) yields the desired equation

$$\lim_{k\to\infty} \int_{\mathbb{R}^l} Z_k(\xi) \; \mu_k(d\xi) = \int_{\mathbb{R}^l} Z_o(\xi) \; \mu_o(d\xi).$$

□

Let us turn to the risk functions. In order to apply Theorem 5.5 in Billingsley (1968) we need to make sure that the sets of discontinuity points of the integrands of $Q_{\mathcal{D}_\mathrm{p}}$, $Q_{\mathcal{D}_\mathrm{p}^+}$, and $Q_{\mathrm{E}_\mathrm{p}^\eta}$ at x_o have μ_o-measure zero, where μ_o denotes the limit measure. For the expected excess of a target η_o the continuity of the maximum directly yields

$$D_Z(x) = \{\xi \in \mathbb{R}^l : \max\{Z(\cdot,\cdot) - \eta_o, 0\}^p \text{ is discontinuous at } (x,\xi)\} \qquad (3.8)$$

for all $x \in \mathbb{R}^n$. For the central deviation and the semideviation, in addition, Q_{E} has to be jointly continuous. Let $h_k^{\mathrm{C}}(\xi) := |Z(x_k,\xi) - Q_{\mathrm{E}}(x_k,\mu_k)|^p$ and $h_k^{\mathrm{S}}(\xi) := \max\{Z(x_k,\xi) - Q_{\mathrm{E}}(x_k,\mu_k), 0\}^p$. We define the discontinuity sets

$$D_{\mathrm{C}}(x) \;:=\; \{\xi \in \mathbb{R}^l : \exists \xi_k \to \xi, x_k \to x_o, \mu_k \xrightarrow{w} \mu_o \text{ such that } h_k^{\mathrm{C}}(\xi_k) \not\to h_o^{\mathrm{C}}(\xi)\},$$

and

$$D_{\mathrm{S}}(x) \;:=\; \{\xi \in \mathbb{R}^l : \exists \xi_k \to \xi, x_k \to x, \mu_k \xrightarrow{w} \mu \text{ such that } h_k^{\mathrm{S}}(\xi_k) \not\to h_o^{\mathrm{S}}(\xi)\}.$$

Lemma 3.14 *Let $x_o \in \mathbb{R}^n$, $\Delta \subset \mathcal{P}(\mathbb{R}^l)$, and $\mu_o \in \Delta$. Assume there is a $\delta > 0$ such that $Z(x, \cdot)$ is a μ-measurable function for all $x \in U_\delta(x_o)$ and all $\mu \in \Delta$. Assume the family of functions $\{Z(x, \cdot) : x \in U_\delta(x_o)\}$ is uniformly integrable w.r.t. the set Δ. Assume furthermore $\mu_o(D_Z(x_o)) = 0$. Then the sets $D_C(x_o)$ and $D_S(x_o)$ have μ_o-measure zero.*

Proof We show that the sets $D_C(x_o)$ and $D_S(x_o)$ are subsets of $D_Z(x_o)$. In fact, we verify that their complements contain the complement of $D_Z(x_o)$. Let $\xi \in \mathbb{R}^l \setminus D_Z(x_o)$. Then, for any sequences $\xi_k \to \xi$ and $x_k \to x$ we have $Z(x_k, \xi_k) \to Z(x_o, \xi)$. Moreover, by Proposition 3.13 it holds $Q_E(x_k, \mu_k) \to Q_E(x_o, \mu_o)$ for any sequence $\{\mu_k\}$ in Δ converging weakly to μ_o. The continuity of the absolute value and the maximum yield $\xi \in \mathbb{R}^l \setminus D_C(x_o)$ and $\xi \in \mathbb{R}^l \setminus D_S(x_o)$. Thus, we have $D_C(x_o) \subset D_Z(x_o)$ and $D_S(x_o) \subset D_Z(x_o)$. By the assumption $\mu_o\big(D_Z(x_o)\big) = 0$ we obtain the desired. □

Now we are able to apply Theorem 5.5. in Billingsley (1968) to the sequences $\{|Z_k(\xi) - Q_E(x_k, \mu_k)|^p\}$, $\{\max\{Z_k(\xi) - Q_E(x_k, \mu_k), 0)\}^p\}$, and $\{\max\{Z_k(\xi) - \eta_o, 0)\}^p\}$. For ease of exposition, we follow the standard proceeding in measure theory and use a stronger condition than the uniform integrability of a family of functions in the next statement.

Proposition 3.15 *Let $\eta_o \in \mathbb{R}$ and $p \in \mathbb{N}$. Let $x_o \in \mathbb{R}^n$, $\Delta \subset \mathcal{P}(\mathbb{R}^l)$ and $\mu_o \in \Delta$. Assume there is a $\delta > 0$ such that $Z(x, \cdot)$ is a μ-measurable function for all $x \in U_\delta(x_o)$ and all $\mu \in \Delta$. Assume there are real numbers r and C, where $r > p$ and $C > 0$, such that*

$$\int_{\mathbb{R}^l} |Z(x, \xi)|^r \; \mu(d\xi) < C \qquad \forall x \in U_\delta(x_o), \forall \mu \in \Delta. \tag{3.9}$$

Furthermore assume $\mu_o\big(D_Z(x_o)\big) = 0$. Then $Q_{\mathcal{D}_p}$, $Q_{\mathcal{D}_p^+}$, and $Q_{E_p^\eta}$ as functions from $\mathbb{R}^n \times \Delta$ to $\mathbb{R}$ are continuous at (x_o, μ_o).

Proof Again, we consider sequences $\{x_k\} \subset U_\delta(x_o)$ and $\{\mu_k\} \subset \Delta$ converging to x_o and converging weakly to μ_o, respectively. Under the assumption $\mu_o(D_Z(x_o)) = 0$, Lemma 3.14 and Theorem 5.5 in Billingsley (1968) together yield $\mu_k \circ Y_k^{-1} \xrightarrow{w} \mu_o \circ Y_o^{-1}$ where the sequence of random variables $\{Y_k(\xi)\}$ (and the limit $Y_o(\xi)$) can be replaced by each of the sequences $\{|Z_k(\xi) - Q_E(x_k, \mu_k)|^p\}$, $\{\max\{Z_k(\xi) - Q_E(x_k, \mu_k), 0)\}^p\}$, and $\{\max\{Z_k(\xi) - \eta_o, 0)\}^p\}$ (and their limits, respectively).

We head for the statement

$$\lim_{k \to \infty} \int_{\mathbb{R}} t \; \mu_k \circ Y_k^{-1}(dt) = \int_{\mathbb{R}} t \; \mu_o \circ Y_o^{-1}(dt)$$

or equivalently

$$\lim_{k\to\infty}\int_{\mathbb{R}^l} Y_k(\xi)\ \mu_k(d\xi) = \int_{\mathbb{R}^l} Y_o(\xi)\ \mu_o(d\xi).$$

The latter is true provided that the sequence $\{Y_k\}$ is uniformly integrable w.r.t. $\{\mu_k\}$. The relation of the integrability assumption and the uniform integrability is as follows; assume $r > 1$ and $a \geq 0$. Then, it holds

$$\begin{aligned} a^{1-r}\int_{\mathbb{R}^l}|Y_k(\xi)|^r\ \mu_k(d\xi) &\geq \int_{|Y_k(\xi)|\geq a} a^{1-r}|Y_k(\xi)|^{r-1}|Y_k(\xi)|\ \mu_k(d\xi) \\ &\geq \int_{|Y_k(\xi)|\geq a} |Y_k(\xi)|\ \mu_k(d\xi). \end{aligned}$$

Thus, the condition $\sup_{k\in\mathbb{N}}\int_{\mathbb{R}^l}|Y_k(\xi)|^r\ \mu_k(d\xi) < \infty$ is sufficient for the uniform integrability of the sequence $\{Y_k\}$ w.r.t. the sequence $\{\mu_k\}$. We verify this condition for the three risk functions.

First we consider the central deviation and the semideviation. We estimate

$$\begin{aligned} \max\{Z_k(\xi) - Q_{\mathrm{E}}(x_k,\mu_k), 0\}^r &\leq |Z_k(\xi) - Q_{\mathrm{E}}(x_k,\mu_k)|^r \\ &\leq 2^r(|Z_k(\xi)|^r + |Q_{\mathrm{E}}(x_k,\mu_k)|^r) \end{aligned}$$

for all $k \in \mathbb{N}$ and all $\xi \in \mathbb{R}^l$. Then, the integrability assumption (3.9) yields

$$\int_{\mathbb{R}^l} 2^r(|Z_k(\xi)|^r + |Q_{\mathrm{E}}(x_k,\mu_k)|^r)\ \mu_k(d\xi) \leq 2^r C + 2^r|Q_{\mathrm{E}}(x_k,\mu_k)|^r \qquad \forall k \in \mathbb{N}. \tag{3.10}$$

Under the posed assumptions, Q_{E} is jointly continuous at (x_o,μ_o), cf. Proposition 3.13. Thus, we have $\max_{k\in N}|Q_{\mathrm{E}}(x_k,\mu_k)|^r < C' < \infty$ and consequently the term on the right of equation (3.10) is less than the constant term $2^rC + 2^rC'$. Since r is greater than p this gives the uniform integrability of the sequences $\{|Z_k(\xi) - Q_{\mathrm{E}}(x_k,\mu_k)|^p\}$ and $\{\max\{Z_k(\xi) - Q_{\mathrm{E}}(x_k,\mu_k), 0\}^p\}$ w.r.t. $\{\mu_k\}$.

Finally, we consider the expected excess $\max\{Z_k(\xi) - \eta_o, 0\}^r$. For all $k \in \mathbb{N}$ and all $\xi \in \mathbb{R}^l$ it holds $\max\{Z_k(\xi)-\eta_o, 0\}^r \leq |Z_k(\xi)-\eta_o|^r \leq 2^r|Z_k(\xi)|^r + 2^r|\eta_o|^r$, and by assumption (3.9) we can estimate

$$2^r\int_{\mathbb{R}^l}|Z_k(\xi)|^r + |\eta_o|^r\ \mu_k(d\xi) \leq 2^rC + 2^r|\eta_o|^r \qquad \forall k \in \mathbb{N}. \tag{3.11}$$

Again, by the assumption $r > p$ we obtain the uniform integrability of the sequence $\{\max\{Z_k(\xi) - \eta_o, 0\}^p\}$ w.r.t. $\{\mu_k\}$.

Thus, the $p-$th powers of the functions $Q_{\mathcal{D}_{\mathrm{p}}}$, $Q_{\mathcal{D}_{\mathrm{p}}^+}$, and $Q_{\mathrm{E}_{\mathrm{p}}^\eta}$ are jointly continuous at (x_o,μ_o). Since $x \mapsto \sqrt[p]{x}$ is continuous, we can conclude the assertions. □

The joint continuity of the risk functions is the key to derive stability results. In particular, we can immediately conclude continuity results on the optimal value functions and the optimal set mappings.

In this thesis we predominantly deal with nonconvex problems. Therefore, we study the stability of the localized optimal value functions

$$\varphi_{\mathrm{V}}(\mu) := \inf\{Q_{\mathcal{R}}(x,\mu) : x \in X \cap clV\}$$

and the localized optimal set mappings

$$\psi_{\mathrm{V}}(\mu) := \{x \in X \cap clV : Q_{\mathcal{R}}(x,\mu) = \varphi_{\mathrm{V}}(\mu)\},$$

where V is a subset of $\mathbb{R}^n$, clV denotes the closure of V, and $Q_{\mathcal{R}}$ stands for one of the risk functions discussed in this section.

When we want to transfer stability results for global optimizers to local ones, we have to restrict our investigations to local minimizers of a special type, see Klatte (1985) and Robinson (1987) for details and pathologies where local minimizers of parametric programs behave unstably.

Definition 3.16 (complete local minimizing (CLM) set) *Let $f : X \to \mathbb{R}$. A nonempty subset M of X is a complete local minimizing set for f w.r.t. an open set $V \supset M$, if the set of minimizers of f on clV equals M.*

In our terminology, this definition reads: a nonempty set $M \subset X$ is a CLM set of $Q_{\mathcal{R}}$ w.r.t. an open set $V \subset \mathbb{R}^n$ if it holds $\psi_{\mathrm{V}}(\mu) = M$.

Before we can state the stability results, we need to introduce the notion of Berge semicontinuity, see Berge (1966). In perturbation analysis, it is frequently used to describe continuity aspects of multifunctions (set-valued mappings).

Definition 3.17 (Berge lower and upper semicontinuity) *A multifunction Θ from a topological space Y into a topological space Z is called Berge lower semicontinuous in μ_o if for any open set G with $G \cap \Theta(\mu_o) \neq \emptyset$, there exists a neighborhood $U(\mu_o)$ such that $\mu \in U(\mu_o)$ implies $\Theta(\mu) \cap G \neq \emptyset$.*

A multifunction Θ from a topological space Y into a topological space Z is called Berge upper semicontinuous in μ_o if for each open set G containing $\Theta(\mu_o)$, there exists a neighborhood $U(\mu_o)$ such that $\mu \in U(\mu_o)$ implies $\Theta(\mu) \subset G$.

The next proposition is originally due to Berge (1966) (Theorem du maximum). We follow the localized versions of Klatte (1985) and Robinson (1987), see also Bank et al. (1982). To state the result in a single proposition, we make the conventions; $r(\mathrm{E}) > 1$ and $r(\mathcal{R}) > p$.

Proposition 3.18 *Let $X \subset \mathbb{R}^n$, $\Delta \subset \mathcal{P}(\mathbb{R}^s)$ and $\mu_o \in \Delta$. Assume that $Z(x,\cdot)$ is a μ-measurable function for all $x \in X$ and all $\mu \in \Delta$. Assume there is a $C > 0$ such that $\int_{\mathbb{R}^l} |Z(x,\xi)|^{r(\mathcal{R})}\ \mu(d\xi) < C$ for all $x \in X$ and all $\mu \in \Delta$. Assume further $\mu_o(D_Z(x)) = 0$ for all $x \in X$.*

Let V be some bounded open subset of X. Let $\varphi(\mu) := \inf_{x\in clV} Q_{\mathcal{R}}(x,\mu)$ and $\psi(\mu) := \{x \in clV : Q_{\mathcal{R}}(x,\mu) = \varphi(\mu)\}$. Assume $\psi(\mu_o)$ is a CLM set w.r.t. V for $\mu_o \in \Delta$. Then

(i) *$\varphi : \Delta \to \mathbb{R}$ is continuous at μ_o,*

(ii) *$\psi : \Delta \to 2^{\mathbb{R}^n}$ is Berge u.s.c. at μ_o, and*

(iii) *if $Q_{\mathcal{R}}(.,\mu)$ is l.s.c. on X for all $\mu \in \Delta$ then there is a neighborhood U of μ_o in Δ such that $\psi(\mu)$ is a CLM set w.r.t. V for all $\mu \in U$.*

Proof In the Propositions 3.13 and 3.15 we have shown that the posed integrability assumption and $\mu_o(D_Z(x)) = 0$ for all $x \in X$ are sufficient for the joint continuity of the individual risk functions on $X \times \{\mu_o\}$. This is the key to show (i) and (ii). The actual proof corresponds to the one of Theorem 4.2.2. in Bank et al. (1982) and is not repeated here.

In (iii) the lower semicontinuity of $Q_{\mathcal{R}}(.,\mu)$ on X for all $\mu \in \Delta$ guarantees that $\psi(\mu)$ is nonempty. Thus and by (ii), $\psi(\mu)$ is a CLM set w.r.t. V for all $\mu \in U$. □

In Proposition 3.5 we have worked out assumptions leading to the lower semicontinuity of the individual risk functions, in the Propositions 3.13 and 3.15 the same has been established for the joint continuity at (x_o, μ_o). Thus, it is straightforward how the previous proposition reads for a particular risk function.

Consult Bank et al. (1982) for a counterexample when the compactness condition on clV is dropped. We note that if the required lower semicontinuity fails to hold, the assertion (ii) may turn meaningless. Because then we may have $\psi(\mu) = \emptyset$ for all μ in some neighborhood of μ_o and for all $G \supset \psi(\mu_o)$ it holds trivially $\psi(\mu) \subset G$.

Remark 3.19 *Under the assumptions of Proposition 3.18 the ε-optimal set mapping*

$$\psi_\varepsilon(\mu) := \{x \in clV : Q_{\mathcal{R}}(x,\mu) \le \varphi(\mu) + \varepsilon\}$$

as a multifunction from Δ to $2^{\mathbb{R}^n}$ is Berge l.s.c., see Corollary 4.2.4.1. in Bank et al. (1982).

4 Stochastic programs with linear recourse

4.1 Scope

It is the aim of this section to introduce the basic notions of two-stage stochastic programming and to identify the subclass which we want to investigate. We remark that the text books Birge and Louveaux (1997), Kall and Wallace (1994), and Prekopa (1995) can serve the first purpose in a much more comprehensive way. We repeat the two-stage model (1.9) given in Chapter 1

$$\inf_{x \in X} \mathcal{R}[cx + \tilde{\phi}(x, \xi(\omega))], \tag{4.1}$$

where $c \in \mathbb{R}^n$, $X = \{x \in \mathbb{R}^n_+ : Ax = b\}$,

$$\tilde{\phi}(x, \xi(\omega)) = \min_{y \in \mathbb{R}^m_+} \{qy \ : \ Wy = h(\omega) - T(\omega)x\}, \tag{4.2}$$

$q \in \mathbb{R}^m$, and $\xi := (h, T) : \Omega \to \mathbb{R}^s \times \mathbb{R}^{s \times n}$ is a random variable defined on the probability space $(\Omega, \mathbb{P}, \mathcal{A})$. We shall assume that X is nonempty and closed. The function $\mathcal{R}$ maps from the family of random variables $\mathcal{Z} := \{cx + \tilde{\phi}(x, \xi(\omega)) : x \in X\}$ to $\mathbb{R}$. $\mathcal{R}$ is either the expected value, a measure of risk, or a weighted sum of both.

Along with the risk measure $\mathcal{R} : \mathcal{Z} \to \mathbb{R}$ we consider the associated functions $Q_{\mathcal{R}} : \mathbb{R}^n \to \mathbb{R}$ defined on the space of decision variables. For the expected value these functions read $\mathrm{E} : \mathcal{Z} \to \mathbb{R}$ and $Q_{\mathrm{E}} : \mathbb{R}^n \to \mathbb{R}$, respectively, where $Q_{\mathrm{E}}(x) := \mathrm{E}[cx + \phi(h - Tx)]$ for all $x \in X$. In two-stage stochastic programming, the function Q_{E} is called the *expected recourse function.*

Structural investigations of the mathematical program (4.1) rely on results from parametric optimization. Only weak structural statements concerning the dependency of the optimal value of a linear program on a parametric constraint matrix are available. This also holds true for the special case of two-stage models where the matrix W is random, cf. Walkup and Wets (1967). When W is deterministic, all parameters (x and ξ) enter the value function $\tilde{\phi}$ at the right-hand side of the mathematical program. This gives rise to (re-)write the function $\tilde{\phi}$ as

$$\phi(t) = \min_{y \in \mathbb{R}^m_+} \{qy \ : \ Wy = t\} \qquad t \in \mathbb{R}^s. \tag{4.3}$$

ϕ is referred to as the *recourse function* and the mathematical program defining ϕ as the *recourse program.* Now, the minimization problem (4.1) reads

$$\inf_{x \in X} \mathcal{R}[cx + \phi(h(\omega) - T(\omega)x)]. \tag{4.4}$$

The decisions x are referred to as *first stage*, the decisions y as *second stage.* In addition to the modelling concept mentioned in Chapter 1 – namely that first-stage decisions have to be taken prior to the realization of the random variable ξ and second-stage decisions can be taken afterwards – there is another (similar) motivation for the two-stage model. Assume we are given a random linear program $\min\{cx : T(\omega)x \geq h(\omega)\}$ along with a recourse program that defines the costs for infeasibilities associated with a decision x. Then, in contrast to the chance constraint model (1.5), infeasible decisions for the random program are explicitly accepted but have to be paid for.

We note that in addition to the constraints in X there is a constraint on the variables x induced by the recourse program, namely the constraint $Q_{\mathcal{R}}(x) < \infty$. For the expected value, the special situation when every feasible first-stage decision can be compensated by recourse decisions, i.e. $X \subset X_2 := \{x \in \mathbb{R}^n : Q_{\mathrm{E}}(x) < \infty\}$, is called *relative complete recourse.* In general, it is hard to describe X_2 and to check whether a particular program has relative complete recourse, see Wets (1974). A stronger condition, that is much easier to verify, is the following; the stochastic program (4.1) is said to have *complete recourse* if $W(\mathbb{R}^m_+) = \mathbb{R}^s$, where $W(\mathbb{R}^m_+)$ is the cone $\{t \in \mathbb{R}^s \ : \ \exists y \in \mathbb{R}^m_+ \quad Wy = t\}$. As this thesis is devoted to complete recourse models, we adhere to

Assumption A 4.1 *Complete recourse:* $W(\mathbb{R}^m_+) = \mathbb{R}^s$.

In the following sections we use the results of Chapter 3 to derive properties concerning the structure and the stability of the considered mean-risk models.

We state the results in terms of the image measure $\mu := \mathbb{P} \circ \xi^{-1}$, a measure in the set $\mathcal{P}(\mathbb{R}^l)$, of Borel probability measures on $\mathbb{R}^l$, where $l := (n+1) \cdot s$. For the ease of exposition, we will write $\mu(d\xi)$ instead of $\mu(d(h,T))$ even when we separate ξ into h and T in the integrands. We identify the first s entries of ξ with h and the remaining ones with the columns of T.

4.2 Prerequisites

In order to obtain results on structure and stability of the mean-risk model (4.4), we need to study the recourse function ϕ as the value function of a linear program in some more detail. Let us assume complete recourse (Assumption A 4.1) and

Assumption A 4.2 *Primal boundedness of* ϕ*:* $\{u \in \mathbb{R}^s : uW \leq q\} \neq \emptyset$, i.e. feasibility of the dual of the linear program defining ϕ. Then, the set $\{u \in \mathbb{R}^s : uW \leq q\}$ is bounded, and thus, has vertices d_j, $j = 1, \ldots, J$, such that via linear programming duality the value function ϕ is given by

$$\phi(t) = \max_{u \in \mathbb{R}^s} \{tu \ : \ uW \leq q\} = \max_{j=1,\ldots,J} d_j t, \tag{4.5}$$

cf. Theorem 4.5. of Chapter I.4. in Nemhauser and Wolsey (1988). Consequently, ϕ is the pointwise maximum of finitely many linear functions, which provides the following structural properties of the value function.

Lemma 4.1 *Assume A 4.1 and A 4.2. Then, $\phi : \mathbb{R}^s \to \mathbb{R}$ is real-valued, piecewise linear, and convex on $\mathbb{R}^s$.* □

Before we apply the results of Chapter 3 we state two lemmas that support the proofs of the following sections. The first one provides an estimate for ϕ.

Lemma 4.2 *Let $p \in \mathbb{R}_+$ and $x \in \mathbb{R}^n$. Assume A 4.1 and A 4.2. Then for all $h \in \mathbb{R}^s$ and all $T \in \mathbb{R}^{s \times n}$ it holds*

$$|cx + \phi(h - Tx)|^p < C(x)\big(\|c\|^p + \|h\|^p + \|T\|\big)^p$$

with a function $C : \mathbb{R}^n \to \mathbb{R}$ that is real-valued and continuous on $\mathbb{R}^n$.

Proof The assumptions A 4.1 and A 4.2 allow us to restate the recourse function $\phi(h - Tx) = \max_{j=1,\ldots,J} d_j(h - Tx)$ with the vertices d_j, $j = 1, \ldots, J$, of the set $\{u \in \mathbb{R}^s : uW \leq q\}$. Let $j^* \in \arg\max\{d_j(h - Tx) : j = 1, \ldots, J\}$. This yields the estimate

$$\begin{aligned} |cx + \phi(h - Tx)|^p &= |cx + d_{j^*}(h - Tx)|^p \\ &\leq \big(\|c\|\|x\| + \|d_{j^*}\|\|h\| + \|d_{j^*}\|\|T\|\|x\|\big)^p \\ &\leq C(x)(\|c\| + \|h\| + \|T\|)^p \end{aligned}$$

where $C(x) := \max\{\|x\|, \|d_{j^*}\|, \|d_{j^*}\|\|x\|\}^p$ is real-valued and continuous. Since we have

$$(\|c\| + \|h\| + \|T\|)^p \leq 3^p(\|c\|^p + \|h\|^p + \|T\|^p)$$

the assertion holds true. □

The second lemma shows how the properties of ϕ read in the notation of the previous Chapter.

Lemma 4.3 *Assume A 4.1 and A 4.2. Let $Z : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}$ be defined as $Z(x, \xi) := cx + \phi(h - Tx)$ for all $x \in \mathbb{R}^n$ and all $(h, T) \in \mathbb{R}^s \times \mathbb{R}^{s \times n}$. Then*

(i) *$Z(x, \cdot) : \mathbb{R}^l \to \mathbb{R}$ is real-valued and convex on $\mathbb{R}^l$ for all $x \in \mathbb{R}^n$.*

(ii) $Z(\cdot, \xi) : \mathbb{R}^n \to \mathbb{R}$ *is real-valued and convex on* $\mathbb{R}^n$ *for all* $\xi \in \mathbb{R}^l$.

(iii) $Z : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}$ *is jointly continuous on* $\mathbb{R}^n \times \mathbb{R}^l$.

Proof Under the Assumptions A 4.1 and A 4.2, ϕ is real-valued, continuous, and convex on $\mathbb{R}^s$, cf. Lemma 4.1. This carries over to $Z(x, \cdot)$ for all $x \in \mathbb{R}^n$ and $Z(\cdot, \xi)$ for for all $\xi \in \mathbb{R}^l$. Given $h \in \mathbb{R}^s$, $T \in \mathbb{R}^{s \times n}$, and $x \in \mathbb{R}^n$, ϕ is continuous at $h - Tx$. This also implies that Z is continuous at (x, ξ), where $\xi = (h, T)$. □

In particular, the latter Lemma implies that $Z(x, \cdot)$ is also measurable for $x \in \mathbb{R}^n$ if the assumptions A 4.1 and A 4.2 are fulfilled.

4.3 Expected value model

In this section we review the expected value model

$$\inf_{x \in X} Q_{\mathrm{E}}(x), \tag{4.6}$$

where $Q_{\mathrm{E}}(x) = \int_{\mathbb{R}^l} cx + \phi(h - Tx)\ \mu(d\xi)$ and where we stick to the specifications given for problem (4.1).

Assuming complete recourse A 4.1 and dual feasibility of ϕ allows us to restate problem (4.6) as

$$\inf_{x \in X} \int_{\mathbb{R}^l} \max_{j=1,\dots,J} d_j(h - Tx)\ \mu(d\xi), \tag{4.7}$$

where d_j, $j = 1, \dots, J$, are the vertices of the set $\{u \in \mathbb{R}^s : uW \le q\}$, cf. Section 4.2. The integrand is piecewise linear and therefore measurable, cf. Lemma 4.1. Thus, the finiteness of the first absolute moments of h and T is sufficient for Q_{E} to be real-valued.

Assumption A 4.3 *Existence of expectation:* $\int_{\mathbb{R}^l} \|h\| + \|T\|\ \mu(d\xi) < \infty$.

For the following result we refer to Theorem 5 of Chapter 3 in Birge and Louveaux (1997).

Proposition 4.4 *Assume A 4.1, A 4.2, and A 4.3. Then*

(i) $Q_{\mathrm{E}}(x)$ *is real-valued for all* $x \in \mathbb{R}^n$,

(ii) $Q_{\mathrm{E}} : \mathbb{R}^n \to \mathbb{R}$ *is convex, and*

(iii) *if* μ *is a discrete and finite probability measure then* $Q_{\mathrm{E}} : \mathbb{R}^n \to \mathbb{R}$ *is a polyhedral function.* □

Together with Weierstrass' theorem, Proposition 4.4 yields that – given X is compact – the minimum in (4.6) exists. Moreover, due to the convexity of Q_{E} and X the set $\arg\min\{Q_{\mathrm{E}}(x) : x \in X\}$ is convex and every local minimum is a global one.

As soon as the dimension of the problem grows it is hopeless to evaluate the multidimensional integral that defines Q_{E}. A discrete approximation, i.e. an approximation of the underlying continuous probability distributions by discrete ones, is necessary. This motivates the perturbation analysis w.r.t. the probability measure μ. Beyond the corresponding chapters in Birge and Louveaux (1997) and Prekopa (1995), stability results for two-stage stochastic programs can be found in Robinson and Wets (1987) and Römisch and Schultz (1993). They include results as those obtained in Proposition 3.18 for a more general situation.

4.4 Mean-risk models

In this section we study the mean-risk models

$$\inf_{x\in X} Q_{\mathrm{E}}(x) + \alpha\ Q_{\mathcal{R}}(x) \qquad \alpha > 0 \tag{4.8}$$

that incorporate the risk measures central deviation, semideviation, and expected excess of a target $\eta_o \in \mathbb{R}$, each of order $p \in \mathbb{N}$. The corresponding risk functions are

$$Q_{\mathcal{D}_{\mathrm{p}}}(x) := \left(\int_{\mathbb{R}^l} |cx + \phi(h - Tx) - Q_{\mathrm{E}}(x)|^p\ \mu(d\xi)\right)^{\frac{1}{p}},$$

$$Q_{\mathcal{D}_{\mathrm{p}}^+}(x) := \left(\int_{\mathbb{R}^l} \max\{cx + \phi(h - Tx) - Q_{\mathrm{E}}(x), 0\}^p\ \mu(d\xi)\right)^{\frac{1}{p}},$$

and

$$Q_{\mathrm{E}_{\mathrm{p}}^{\eta}}(x) := \left(\int_{\mathbb{R}^l} \max\{cx + \phi(h - Tx) - \eta_o, 0\}^p\ \mu(d\xi)\right)^{\frac{1}{p}}.$$

To obtain the finiteness of the risk functions it is necessary to require the finiteness of the p-th moments of ξ

Assumption A 4.4 *Existence of p-th moment:* $\int_{\mathbb{R}^l} \|h\|^p + \|T\|^p\ \mu(d\xi) < \infty$.

Some structural properties of the risk functions are immediate consequences of the results of Chapter 3.

Proposition 4.5 *Let* $p \in \mathbb{N}$ *and* $\eta_o \in \mathbb{R}$. *Assume A 4.1, A 4.2, and A 4.4. Then*

(i) $Q_{\mathcal{D}_p} : \mathbb{R}^n \to \mathbb{R}$ *is real-valued and continuous on* $\mathbb{R}^n$, $Q_{\mathrm{E}} + \alpha Q_{\mathcal{D}_1} : \mathbb{R}^n \to \mathbb{R}$ *is convex on* $\mathbb{R}^n$ *for* $\alpha \in (0, \frac{1}{2}]$,

(ii) $Q_{\mathcal{D}_p^+} : \mathbb{R}^n \to \mathbb{R}$ *is real-valued and continuous on* $\mathbb{R}^n$, $Q_{\mathrm{E}} + \alpha Q_{\mathcal{D}_p^+} : \mathbb{R}^n \to \mathbb{R}$ *is convex on* $\mathbb{R}^n$ *for* $\alpha \in (0,1]$, *and*

(iii) $Q_{\mathrm{E}_p^\eta} : \mathbb{R}^n \to \mathbb{R}$ *is real-valued and convex on* $\mathbb{R}^n$.

Proof We verify the assumptions of the Propositions 3.1, 3.6, and 3.8 for the function $Z(x,\xi) = cx + \phi(h - Tx)$. Under the Assumptions A 4.1 and A 4.2, $Z(x,\cdot)$ is measurable for all $x \in \mathbb{R}^n$ and $Z(\cdot,\xi)$ is continuous on $\mathbb{R}^n$ for all $\xi \in \mathbb{R}^s$, cf. (i) and (ii) of Lemma 4.3. Thus, the set of discontinuity points of $Z(\cdot,\xi)$ has μ-measure zero.

Lemma 4.2 yields $|Z(x,\xi)|^p \le C(x)(\|c\|^p + \|h\|^p + \|T\|^p)$ for all $x \in \mathbb{R}^n$. This gives the integrability assumption on Z. Altogether, we have the finiteness and the continuity of the three functions.

The function $Z(\cdot,\xi)$ is convex under the Assumptions A 4.1 and A 4.2, cf. (ii) of Lemma 4.3. Proposition 3.8 yields the convexity of $Q_{\mathrm{E}_\mathrm{p}^\eta}$ and $Q_{\mathrm{E}} + \alpha Q_{\mathcal{D}_\mathrm{p}^+}$ for $\alpha \in (0,1]$. □

The central deviation $\mathcal{D}_\mathrm{p}$ and semidevation $\mathcal{D}_\mathrm{p}^+$ of order p are not monotonous, and therefore not convex in the sense of Föllmer and Schied (2003), cf. the Examples 2.10, 2.14, and 2.18. Thus, in general, the risk functions $Q_{\mathcal{D}_\mathrm{p}}$ and $Q_{\mathcal{D}_\mathrm{p}^+}$ are not convex. We give an example where the corresponding mean-risk models turn out nonconvex, too.

Example 4.6 *Let* $c = 0$ *and* $\phi(t) := \min\{y^+ + y^- : y^+ - y^- \ge t,\ y^+, y^- \in \mathbb{R}_+\} = |t|$ *for all* $t \in \mathbb{R}$. *Let* $\Omega = \{\omega_1, \omega_2\}$ *and let the random variable* $\xi : \Omega \to \mathbb{R}^2$ *have the distribution* $\mathbb{P}(\xi = (h,T) = (0,-2)) = \mathbb{P}(\xi = (h,T) = (-1,-1)) = \frac{1}{2}$. *We obtain the risk functions* $Q_{\mathcal{D}_p}(x) = \frac{1}{2}||2x| - |x-1||$ *and* $Q_{\mathcal{D}_p^+}(x) = \frac{1}{2}\sqrt[p]{\frac{1}{2}}||2x| - |x-1||$ *for all* $x \in \mathbb{R}$. *These functions are nonconvex on any open interval containing* 0. *In fact, we are more interested in the compound functions* $f(x) := Q_{\mathrm{E}}(x) + \alpha Q_{\mathcal{D}_p}(x)$ *and* $g(x) := Q_{\mathrm{E}}(x) + \alpha Q_{\mathcal{D}_p^+}(x)$, *where* $Q_{\mathrm{E}}(x) = \frac{1}{2}|2x| + \frac{1}{2}|x-1|$. *Let* $x_1 = -1$, $x_2 = \frac{1}{3}$, *and* $\lambda = \frac{1}{4}$. *The convexity inequality is not fulfilled for the points* x_1 *and* x_2 *because it holds*

$$\lambda f(x_1) + (1-\lambda) f(x_2) = 1 < \frac{1}{2} + \frac{1}{2}\alpha = f(\lambda x_1 + (1-\lambda)x_2)$$

for $\alpha > 1$ *and*

$$\lambda g(x_1) + (1-\lambda) g(x_2) = 1 < \frac{1}{2} + \frac{1}{2}\sqrt[p]{\frac{1}{2}}\alpha = g(\lambda x_1 + (1-\lambda)x_2)$$

for $\alpha > \sqrt[p]{2}$.

For the qualitative stability analysis we consider $Q_{\mathcal{D}_\mathrm{p}}$, $Q_{\mathcal{D}_\mathrm{p}^+}$, and $Q_{\mathrm{E}_\mathrm{p}^\eta}$ as functions of x as well as μ. We apply Proposition 3.18.

Proposition 4.7 *Let* $p \in \mathbb{N}$ *and* $\eta_o, r \in \mathbb{R}$. *Let* $Q_{\mathcal{R}}$ *be one of the risk functions* $Q_{\mathcal{D}_p}$, $Q_{\mathcal{D}_p^+}$, *or* $Q_{\mathrm{E}_p^\eta}$. *Let* $r > p$, $C > 0$, $X \subset \mathbb{R}^n$, $\Delta \subset \mathcal{P}(\mathbb{R}^l)$, *and* $\mu_o \in \Delta$. *Assume A 4.1, A 4.2, and* $\int_{\mathbb{R}^l} \|h\|^r + \|T\|^r \; \mu(d\xi) < C$ *for all* $\mu \in \Delta$.

Let V *be some bounded open subset of* X. *Let* $\varphi(\mu) := \inf_{x \in clV} Q_{\mathcal{R}}(x, \mu)$ *and* $\psi(\mu) := \{x \in clV : Q_{\mathcal{R}}(x, \mu) = \varphi(\mu)\}$. *Assume* $\psi(\mu_o)$ *is a CLM set w.r.t.* V *for* $\mu_o \in \Delta$. *Then*

(i) $\varphi : \Delta \to \mathbb{R}$ *is continuous at* μ_o,

(ii) $\psi : \Delta \to 2^{\mathbb{R}^n}$ *is Berge u.s.c. at* μ_o, *and*

(iii) *there is a neighborhood* U *of* μ_o *in* Δ *such that* $\psi(\mu)$ *is a CLM set w.r.t.* V *for all* $\mu \in U$.

Proof Again, let $Z(x, \xi) = cx + \phi(h - Tx)$. First of all $Z(x, \cdot) : \mathbb{R}^l \to \mathbb{R}$ is μ-measurable for all $x \in \mathbb{R}^n$ and all $\mu \in \Delta$, cf. (i) of Lemma 4.3. Since Z is continuous on $\mathbb{R}^n \times \mathbb{R}^l$, see (iii) of Lemma 4.3, the set of discontinuity points of $Z(\cdot, \cdot)$ has μ_o-measure zero.

By the assumptions A 4.1 and A 4.2 and the compactness of clV, Lemma 4.2 gives us an estimate of $|Z(x, \xi)|^p$ independent of x. The integrability assumption and this estimate yield $\int_{\mathbb{R}^l} |Z(x, \xi)|^p \; \mu(d\xi) < C$ for all $\mu \in \Delta$ and all $x \in V$. Thus, the requirements of (i) and (ii) in proposition 3.18 are fulfilled.

Since $Q_{\mathcal{D}_\mathrm{p}}$, $Q_{\mathcal{D}_\mathrm{p}^+}$, and $Q_{\mathrm{E}_\mathrm{p}^\eta}$ are continuous on $\mathbb{R}^n$, cf. Proposition 4.5, the set $\psi(\mu)$ is nonempty. This yields the CLM property. □

4.5 Summary

When we turn from the purely expected value based model to a mean-risk model, one hope is that the model does not become significantly more complicated. In particular, we wish to apply existing (or similar) algorithms.

For the central deviation this is certainly not the case. The loss of convexity excludes the tools of convex analysis. In particular, the class of existing algorithms significantly shrinks. Thus, the variance and the standard deviation also appear inappropriate in the modelling context.

The semideviation appears attractive when we consider the results concerning its structure and stability as well as its consistency with the concepts discussed in Section 2. However, from an algorithmic point of view, risk measures that measure the (semi-)deviation from the mean, introduce a coupling of the recourse programs for different realizations of the random variable. Thus, decomposition algorithms based on the separate calculation of scenario subproblems (as for instance the L-shape algorithm) do not work.

The expected excess fulfills all desired properties. The drawback of the risk measure is a conceptual one. The specification of a reasonable target η requires at least some a priori knowledge of the distribution of the optimal value of the recourse program.

We summarize some of our results in Table 4.1, cf. also the Tables 2.2 and 2.3. The columns should be read as follows; if $Q_{\mathcal{R}}$ is convex for the corresponding

Mean-risk problem	Convex	No Coupling	SSD	Coherence	α	p
$Q_{\mathrm{E}} + \alpha\, Q_{\mathcal{D}_{\mathrm{p}}}$	$+$	$-$	$+$	$+$	$(0, \frac{1}{2}]$	1
$Q_{\mathrm{E}} + \alpha\, Q_{\mathcal{D}_{\mathrm{p}}^{+}}$	$+$	$-$	$+$	$+$	$(0, 1]$	$\mathbb{N}$
$Q_{\mathrm{E}} + \alpha\, Q_{\mathrm{E}_{\mathrm{p}}^{\eta}}$	$+$	$+$	$+$	$(-)$	$\mathbb{R}$	$\mathbb{N}$

Table 4.1: Properties of the risk measures

risk measure $\mathcal{R}$, the second column displays a '+'. If the risk measure introduces a coupling of second-stage variables, then the third column displays a '$-$'. If the risk measure is consistent with second order stochastic dominance or coherent, the corresponding columns display a '+'. The last two columns indicate the weights and orders for which the positive results hold true. We have given counterexamples for the remaining values of α and p.

The focus of this thesis is on integer models. Chapter 6 deals with algorithms for stochastic programs with mixed-integer recourse. We refer to the survey Birge (1997) for algorithms for stochastic programs with linear recourse.

5 Stochastic programs with mixed-integer recourse

5.1 Scope

In this section, integer variables come into play. In the next section, we will see that the structure of the recourse function ϕ significantly changes when integer variables are present. A piecewise linear and convex function turns into a function that is merely lower semicontinuous. This naturally carries over to the structure of the two-stage problem.

We shall be concerned with the problem

$$\min_{x \in X} \mathcal{R}[cx + \phi(h(\omega) - T(\omega)x)], \tag{5.1}$$

where $c \in \mathbb{R}^n$ and $\xi = (h, T) : \Omega \to \mathbb{R}^s \times \mathbb{R}^{s \times n}$ is a random variable defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We set $l := s \cdot (n+1)$ and identify the first s entries of ξ with h and the remaining ones with the columns of T. As before, $\mathcal{R}$ will be replaced by the expected value, by some measure of risk, or a weighted sum of both.

Now, in contrast to Chapter 4, the second stage contains integer variables

$$\phi(t) = \inf_{y,y'} \{qy + q'y' \ : \ Wy + W'y' = t,\ y \in \mathbb{Z}_+^m,\ y' \in \mathbb{R}_+^{m'}\}. \tag{5.2}$$

The set $X \subset \mathbb{R}^n$ is nonempty, closed, and may also contain integer requirements.

For the investigations of the expected recourse function Q_{E} and the risk function $Q_{\mathcal{R}}$ we will assume the existence of the corresponding moments of the random variable ξ, i.e. we stick to the Assumptions A 4.3 and A 4.4. Moreover, we continue to assume that the recourse function ϕ is bounded

Assumption A 5.1 $\quad \{u \in \mathbb{R}^s \ : \ uW \leq q,\ uW' \leq q'\} \neq \emptyset$

and that the recourse program has complete recourse

Assumption A 5.2 $\quad W(Z_+^m) + W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s,$

cf. also the integer-free counterparts A 4.1 and A 4.2.

As in the previous two chapters, we state the results in terms of the image measure $\mu = \mathbb{P} \circ \xi^{-1}$.

5.2 Prerequisites

Value functions play a central role in the analysis of stochastic programs with recourse, cf. Schultz (2002) and see Chapter 4. In this section we provide some prerequisites on the structure of value functions of mixed-integer programs.

Beforehand, we provide conditions on the existence of solutions of the general mixed-integer linear program (MILP)

$$\inf_{y,y'}\{qy + q'y' \; : \; Wy + W'y' = t,\, y \in \mathbb{Z}_+^m,\, y' \in \mathbb{R}_+^{m'}\}, \tag{5.3}$$

where $q \in \mathbb{R}^m$, $q' \in \mathbb{R}^{m'}$, $t \in \mathbb{R}^s$, $W \in \mathbb{R}^{s\times m}$, and $W' \in \mathbb{R}^{s\times m'}$.

Proposition 5.1 (Meyer (1974), Theorem 2.1) *Let at least one of the following three conditions be fulfilled*

(i) $m' = 0$*, i.e. the mathematical program* (5.3) *is a pure integer program,*

(ii) *there exists* K *such that* $\|y\| < K$ *for all feasible vectors* y*,*

(iii) *the matrices* W *and* W' *are rational.*

Then the mathematical program (5.3) *is either infeasible or unbounded or possesses an optimal solution.* □

An example given on p.224 in Meyer (1974) clarifies the difficulties encountered with irrational coefficients in W and W'. In this example, a finite infimum is not attained. We follow the standard proceeding in mixed-integer linear programming and exclude this case.

Assumption A 5.3 *W and W' are rational matrices.*

Note, that this assumption does not effect the applicability of the model (5.3) to real world problems; computers take rational input, anyway. Without further notice in the individual statements we stick to Assumption A 5.3 throughout this chapter.

Let us turn to the value function of the mixed-integer linear program (5.3). For $t \in \mathbb{R}^s$ it is given by

$$\phi(t) = \min_{y,y'}\{qy + q'y' \; : \; Wy + W'y' = t,\, y \in \mathbb{R}_+^m,\, y' \in \mathbb{Z}_+^{m'}\}. \tag{5.4}$$

As usual we define $\phi(t) := \infty$ if the mathematical program is infeasible for some $t \in \mathbb{R}^s$ and $\phi(t) := -\infty$ if it is unbounded. We discuss the properties of ϕ under the Assumptions A 5.1 and A 5.2 ensuring dual and primal feasibility of the linear programming relaxation of the mathematical program (5.3). Let us for the moment also assume $W'(\mathbb{R}_+^{m'}) = \mathbb{R}^s$, i.e. complete recourse with respect to the continuous components. Then, we have the representation of the value function of a linear program as pointwise maximum of linear functions, cf. equation (4.5) in Section 4.2, and we are able to rewrite ϕ as

$$\phi(t) = \min_y \{qy + \max_{j=1,\dots,J} d_j(t - Wy) \ : \ y \in \mathbb{Z}^m_+\} \qquad \forall t \in \mathbb{R}^s, \qquad (5.5)$$

where $\{d_j\}_{j=1,\dots,J}$ are the vertices of the polyhedral set $\{u \in \mathbb{R}^s : uW' \leq q'\}$. For a given $y \in \mathbb{Z}^m_+$, the function $\phi_y(t) = qy + \max_j d_j(t - Wy)$ is the maximum of finitely many linear functions. It is finite on the cone $M_y := \{t \in \mathbb{R}^s \ : \ \exists y' \in \mathbb{R}^{m'}_+ : Wy + W'y' = t\}$. The recourse function ϕ is the pointwise minimum of the functions ϕ_y, $y \in \mathbb{Z}^m_+$. These observations lead to a number of structural properties of ϕ. We confine ourselves to the properties needed in Chapter 5. For more details we refer to Theorem 3.3 in Blair and Jeroslow (1977) and Theorem 8.1 in Bank and Mandel (1988).

Proposition 5.2 *Assume A 5.1 and A 5.2. Then*

(i) *ϕ is real-valued and lower semicontinuous on $\mathbb{R}^s$ and*

(ii) *there exist positive constants γ and β such that it holds*

$$|\phi(t') - \phi(t'')| \leq \gamma \|t' - t''\| + \beta \qquad \forall t', t'' \in \mathbb{R}^s \qquad \square$$

Remark 5.3 *The assumption $\{u \in \mathbb{R}^s : uW \leq q,\ uW' \leq q'\} \neq \emptyset$ is equivalent to $\phi(0) = 0$, see Proposition 6.7 on p. 168 in Nemhauser and Wolsey (1988). Thus, the quasi Lipschitz property (ii) of the previous proposition also provides $|\phi(t)| \leq \gamma\|t\| + \beta$ for all $t \in \mathbb{R}^s$.*

With the above results and the results of Chapter 3 at hand, we can identify situations in which the risk functions are finite, l.s.c., or continuous. We will verify the assumptions made in Chapter 3 for the current setting $Z(x, \xi) := cx + \phi(h - Tx)$ for all $x \in \mathbb{R}^n$ and all $\xi \in \mathbb{R}^l$. We note that both x and ξ enter the recourse function ϕ at the right-hand side of the defining mathematical program. Moreover, the function $f(x, h, T) := h - Tx$ is continuous in all its components. This differs from the general setting in Chapter 3 where nothing is said about the relation of x and ξ in Z.

It also has consequences for the relation of the discontinuity sets D'_Z and D_Z of $Z : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}$ defined in Chapter 3 and the set of discontinuity points of ϕ. Recall that $D_Z(x)$ contains those $\xi \in \mathbb{R}^l$ for which there exist sequences $\xi_k \to \xi$ and $x_k \to x$ such that $Z(x_k, \xi_k) \not\to Z(x, \xi)$. $D'_Z(x)$ was defined as the set of those $\xi \in \mathbb{R}^l$ for which there exists a sequence $x_k \to x$ such that $Z(x_k, \xi) \not\to Z(x, \xi)$. It holds $D'_Z(x) \subset D_Z(x)$ for all $x \in \mathbb{R}^n$, cf. Lemma 3.10.

Now, we also consider the set of discontinuity points of ϕ

$$D_\phi(x) := \{(h, T) \in \mathbb{R}^s \times \mathbb{R}^{s \times n} : \phi \text{ is discontinuous at } h - Tx\}.$$

Lemma 5.4 *With $Z : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}$ defined as $Z(x,\xi) := cx + \phi(h - Tx)$ for all $x \in \mathbb{R}^n$ and all $(h,T) \in \mathbb{R}^s \times \mathbb{R}^{s\times n}$, the sets $D_Z(x)$ and $D_\phi(x)$ coincide for all $x \in \mathbb{R}^n$.*

Proof Let $(h,T) \in D_Z(x)$ for some $x \in \mathbb{R}^n$. Then, there exist sequences $(h_k,T_k) \to (h,T)$ and $x_k \to x$ such that $cx_k + \phi(h_k - T_kx_k) \not\to cx + \phi(h - Tx)$. Since c is continuous, it has to be ϕ which is discontinuous at $h-Tx$. Therefore, (h,T) is in $D_\phi(x)$.

Conversely, let $(h,T) \in D_\phi(x)$ for some $x \in \mathbb{R}^n$. Then, there is a sequence $t_k \to h - Tx$ such that $\phi(t_k) \not\to \phi(h - Tx)$. Let $\{T_k\}$ be a sequence converging to T and $\{x_k\}$ be a sequence converging to x. We define the sequence $\{h_k\}$ by $h_k := t_k + T_kx_k$ for all $k \in \mathbb{N}$ and obtain $h_k \to h$, therefore $h_k - T_kx_k \to h - Tx$ but $cx_k + \phi(h_k - T_kx_k) = cx_k + \phi(t_k) \not\to cx + \phi(h - Tx)$. Consequently, (h,T) is in $D_Z(x)$. □

For later use we make available some results concerning the structure of Z.

Lemma 5.5 *Assume A 5.1 and A 5.2. Let $Z : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}$ be defined as $Z(x,\xi) := cx + \phi(h - Tx)$ for all $x \in \mathbb{R}^n$ and all $(h,T) \in \mathbb{R}^s \times \mathbb{R}^{s\times n}$. Then $Z(\cdot,\cdot) : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}$ is real-valued and l.s.c. on $\mathbb{R}^n \times \mathbb{R}^l$.*

Proof The recourse function ϕ is real-valued and l.s.c. on $\mathbb{R}^s$, see Proposition 5.2. Let $x \in \mathbb{R}^n$ and $\xi \in \mathbb{R}^l$. Since ϕ is real-valued at $h-Tx$ for all x, where $(h,T) = \xi$, $Z(x,\xi)$ is also real-valued at x.

Let $\{\xi_k\} = \{(h_k,T_k)\}$ and $\{x_k\}$ be sequences converging to x and $\xi = (h,T)$, respectively. Due to the continuity of c and the lower semicontinuity of ϕ we obtain

$$\liminf_{k\to\infty} Z(x_k,\xi_k) = \liminf_{k\to\infty} cx_k + \phi(h_k - T_kx_k) \geq cx + \phi(h - Tx) = Z(x,\xi).$$

Thus, Z is l.s.c. at (x,ξ). □

Note that, the previous lemma implies that $Z(x,\cdot)$ as a function from $\mathbb{R}^l$ to $\mathbb{R}$ is real-valued and l.s.c. on $\mathbb{R}^l$, and thus measurable, for all $x \in \mathbb{R}^n$, and that $Z(\cdot,\xi)$ as a function from $\mathbb{R}^n$ to $\mathbb{R}$ is real-valued and l.s.c. on $\mathbb{R}^n$ for all $\xi \in \mathbb{R}^l$.

In the following statements, we will need an estimation of $Z(x,\xi)$ which we provide here.

Lemma 5.6 *Let $p \in \mathbb{R}_+$ and $\xi \in \mathbb{R}^l$. Assume B is a bounded subset of $\mathbb{R}^n$. Then, there exist positive constants C and C_1 such that $|Z(x,\xi)|^p \leq C(\|h\|^p + \|T\|^p) + C_1$ for all $x \in B$.*

Proof Let $r := \sup_{x \in B} \|x\|$. Since B is bounded this value is finite. We use $|a+b+c|^p \leq 3^p(|a|^p + |b|^p + |c|^p)$ for $a, b, c \in \mathbb{R}$, the quasi Lipschitz property of ϕ, cf. (ii) in Proposition 5.2 and Remark 5.3, and obtain

$$|\phi(h - Tx)|^p \leq \big(\beta\|h\| + \beta r\|T\| + \gamma\big)^p \leq 3^p\beta^p\big(\|h\|^p + r^p\|T\|^p\big) + 3^p\gamma^p,$$

where β and γ are positive constants. Let $C' := 3^p\beta^p \max\{1, r\}$. For Z this implies

$$\begin{aligned} |Z(x,\xi)|^p &\leq 2^p\big(|cx|^p + |\phi(h - Tx)|^p\big) \\ &\leq 2^p\big(r^p\|c\|^p + C'(\|h\|^p + \|T\|^p) + 3^p\gamma^p\big). \end{aligned}$$

We set $C := 2^pC'$ and $C_1 := 2^pr^p\|c\|^p + 6^p\gamma^p$ and obtain the assertion. □

5.3 Expected value model

In this section we will review some properties of the expected recourse function $Q_{\mathrm{E}}(x) := \int_{\mathbb{R}^l} cx + \phi(h - Tx)\ \mu(d\xi)$ and the mathematical program

$$\inf_{x \in X} Q_{\mathrm{E}}(x), \tag{5.6}$$

where we employ the specifications given in the previous section. The results for the expected recourse function and the expected value problem are due to Schultz (1995) (Propositions 3.1, 3.2, and Corollary 3.3).

Proposition 5.7 *Let $x_o \in \mathbb{R}^n$. Assume A 4.3, A 5.1, and A 5.2. Then*

(i) *$Q_{\mathrm{E}}(x_o)$ is real-valued and $Q_{\mathrm{E}} : \mathbb{R}^n \to \mathbb{R}$ is l.s.c. at x_o,*

(ii) *if $\mu(D_\phi(x_o)) = 0$ then $Q_{\mathrm{E}} : \mathbb{R}^n \to \mathbb{R}$ is continuous at x_o.* □

In general, Q_{E} is not continuous. When Ω is single-valued, the expected recourse function is given by $Q_{\mathrm{E}}(x) = cx + \phi(h - Tx)$. In view of the lower semicontinuity of ϕ, Q_{E} is discontinuous at x iff ϕ is discontinuous at $h - Tx$, precisely, if there is a sequence $\{x_k\} \subset \mathbb{R}^n$ with $\lim_{k\to\infty} x_k = x$ such that

$$\lim_{k\to\infty} \phi(h - Tx_k) > \phi(h - Tx).$$

Now, assume ξ has a finite discrete probability distribution with S probability atoms (h_s, T_s) and corresponding probabilities π_s, $s = 1, \ldots, S$. The expected recourse function takes the form

$$Q_{\mathrm{E}}(x) = cx + \sum_{s=1}^{S} \pi_s \cdot \phi(h_s - T_s x)$$

for $x \in \mathbb{R}^n$. Suppose there exists $s' \in \{1,\dots,S\}$ and $x_o \in \mathbb{R}^n$ such that ϕ is discontinuous at $h_{s'} - T_{s'}x_o$. Suppose furthermore that there is a sequence $\{x_k\} \subset \mathbb{R}^n$ converging to x_o with $\lim_{k\to\infty} \phi(h_{s'} - T_{s'}x_k) > \phi(h_{s'} - T_{s'}x)$. Then we have

$$\begin{aligned}\lim_{k\to\infty} Q_{\mathrm{E}}(x_k) &= cx + \pi_{s'} \cdot \underbrace{\lim_{k\to\infty} \phi(h_{s'} - T_{s'}x_k)}_{>\phi(h_{s'}-T_{s'}x_o)} + \sum_{\substack{s=1\\ s\neq s'}}^{S} \pi_s \cdot \underbrace{\lim_{k\to\infty} \phi(h_s - T_s x_k)}_{\geq\phi(h_s - T_s x_o)}\big) \\ &> Q_{\mathrm{E}}(x_o),\end{aligned}$$

and thus Q_{E} is discontinuous at x_o. The jump of Q_{E} at x_o depends on the value $\pi_{s'}$ and on the jump of ϕ at $h_{s'} - T_{s'}x_o$. When ξ has a discrete probability distribution $\pi_{s'}$ is strictly positive. However, when the image measure $P \circ \xi^{-1}$ is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^s \times \mathbb{R}^{s\times n}$, the measure of a single point is zero, and thus Q_{E} is continuous on $\mathbb{R}^n$. By Proposition 5.7, it suffices to require the set of potential discontinuity points to be of μ measure zero. Using stronger conditions on the measure μ, we are able to establish the Lipschitz continuity of Q_{E}. We refer to Schultz (1995) for a proof and more details.

Let us turn to the stability properties of the expected value problem (5.6). We now consider Q_{E} as a function of $x \in \mathbb{R}^n$ as well as of $\mu \in \mathcal{P}(B^l)$ - the set of Borel probability measures on $\mathbb{R}^l$ endowed with the topology of weak convergence (see definition 3.9)

$$Q_{\mathrm{E}}(x,\mu) = \int_{\mathbb{R}^l} cx + \phi(h - Tx)\ \mu(d\xi). \tag{5.7}$$

In general, the expected value problem (5.6) is nonconvex. Thus, we use the concept of complete local minimizing sets, again, cf. Definition 3.16. Let V be some subset of $\mathbb{R}^n$. We define the localized optimal value function

$$\varphi_{\mathrm{V}}(\mu) := \inf\{Q_{\mathrm{E}}(x,\mu) : x \in X \cup clV\}$$

and the localized optimal set mapping

$$\psi_{\mathrm{V}}(\mu) := \{x \in X \cup clV : Q_{\mathrm{E}}(x,\mu) = \varphi(\mu)\}.$$

We cite results of Schultz (1995), see Proposition 3.8, Remark 3.9, and Proposition 4.1 therein.

Proposition 5.8 *Let $p > 1$, $C > 0$, $\Delta \subset \mathcal{P}(\mathbb{R}^l)$, and $\mu_o \in \Delta$. Let X be a bounded subset of $\mathbb{R}^n$. Assume A 5.1, A 5.2, $\int_{\mathbb{R}^l} \|h\|^p + \|T\|^p\ \mu(d\xi) < C$ for all $\mu \in \Delta$, and $\mu_o(D_\phi(x)) = 0$ for all $x \in X$. Let $\psi_V(\mu_o)$ be a CLM set w.r.t. some open set $V \subset X$. Then*

(i) $\varphi_V : \Delta \to \mathbb{R}$ *is continuous at* μ_o,

(ii) $\psi_V : \Delta \to 2^{\mathbb{R}^m}$ *is Berge upper semicontinuous at* μ_o,

(iii) *there is a neighborhood* $U(\mu_o) \subset \Delta$ *of* μ_o *such that* $\psi_V(\mu)$ *is a CLM set w.r.t.* V *for all* $\mu \in U(\mu_o)$. □

For the quantitative stability analysis of problem (5.6) we refer to Schultz (1995), again.

5.4 Mean-risk models

In this section we investigate the mean-risk models

$$\inf_{x \in X} Q_{\mathrm{E}}(x) + \alpha\ Q_{\mathcal{R}}(x) \qquad \alpha > 0 \tag{5.8}$$

for each of the risk measures central deviation, semideviation, and expected excess of a target $\eta_o \in \mathbb{R}$, all of them of order $p \in \mathbb{N}$. The corresponding risk functions are

$$Q_{\mathcal{D}_{\mathrm{p}}}(x) := \left(\int_{\mathbb{R}^l} |cx + \phi(h - Tx) - Q_{\mathrm{E}}(x)|^p \ \mu(d\xi) \right)^{\frac{1}{p}},$$

$$Q_{\mathcal{D}_{\mathrm{p}}^+}(x) := \left(\int_{\mathbb{R}^l} \max\{cx + \phi(h - Tx) - Q_{\mathrm{E}}(x), 0\}^p \ \mu(d\xi) \right)^{\frac{1}{p}},$$

and

$$Q_{\mathrm{E}_{\mathrm{p}}^{\eta}}(x) := \left(\int_{\mathbb{R}^l} \max\{cx + \phi(h - Tx) - \eta_o, 0\}^p \ \mu(d\xi) \right)^{\frac{1}{p}}.$$

Recall that Assumption A 4.4 ensures the finiteness of the p-th moment of the random variable $\xi = (h, T)$.

Proposition 5.9 *Let* $\alpha \in (0, 1]$ *and* $x_o \in \mathbb{R}^n$. *Assume A 4.4, A 5.1, and A 5.2. Then*

(i) $Q_{\mathcal{D}_p}(x_o)$, $Q_{\mathcal{D}_p^+}(x_o)$, *and* $Q_{\mathrm{E}_p^{\eta}}(x_o)$ *are real-valued,*

(ii) $Q_{\mathrm{E}} + \frac{\alpha}{2}\ Q_{\mathcal{D}_1}$, $Q_{\mathrm{E}} + \alpha\ Q_{\mathcal{D}_1^+}$ *and* $Q_{\mathrm{E}_p^{\eta}}$ *are l.s.c. at* x_o, *and*

(iii) *if* $D_\phi(x_o)$ *has* μ *measure zero then* $Q_{\mathcal{D}_p}$, $Q_{\mathcal{D}_p^+}$, *and* $Q_{\mathrm{E}_p^{\eta}}$ *are continuous at* x_o.

Proof We verify the assumptions of the Propositions 3.1, 3.5, and 3.6 for the function $Z(x, \xi) = cx + \phi(h - Tx)$. Lemma 5.5 tells us that $Z(x, \cdot)$ is measurable for all $x \in \mathbb{R}^n$ and that $Z(\cdot, \xi)$ is l.s.c. on $\mathbb{R}^n$ for all $\xi \in \mathbb{R}^l$. Let $\delta > 0$ and $x \in U_\delta(x_o)$. Lemma 5.6 provides the majorant $C(\|h\| + \|T\|) + C_1$

of $Z(x,\xi)$ for all $\xi \in \mathbb{R}^l$ and all $x \in U_\delta(x_o)$ with positive constants C and C_1 independent of x. By Assumption A 4.4 this majorant is in $\mathcal{L}^p$. Using the assumption $\mu(D_\phi(x_o)) = 0$ we conclude that the set $D'_Z(x_o)$ of discontinuity points of $Z(\cdot,\xi)$ at x_o has μ-measure zero, cf. Lemma 5.4. Consequently, all the assumptions of the Propositions 3.1, 3.5, and 3.6 hold true. This yields the assertions. □

In general, neither $Q_{\mathcal{D}_p}$ nor $Q_{\mathcal{D}_p^+}$ are l.s.c. Basically, this is due to the fact that both risk functions involve the subtraction from a function which is merely l.s.c., namely Q_{E}. Let us display the lack of lower semicontinuity and its consequences by a number of examples.

Example 5.10 *Let $p \in \mathbb{N}$ and $\alpha \in \mathbb{R}_+$. Consider the problem*

$$\inf_x \{Q_{\mathrm{E}}(x) + \alpha Q_{\mathcal{D}_p}(x) \;:\; 0 \le x \le \frac{1}{2}\} \tag{5.9}$$

with the specifications $n = 1$, $m = 1$, $c = 1$, $T \equiv -1$, and

$$\phi(t) = \min_y \{y \;:\; y \ge t,\, y \in \mathbb{Z}\} = \lceil t \rceil \qquad \forall t \in \mathbb{R}.$$

For the mean and the central deviation of order p we obtain $Q_{\mathrm{E}}(x) = x + \mathrm{E}\lceil h+x \rceil$ and $Q_{\mathcal{D}_p}(x) = \big(\mathrm{E}|x+\lceil h+x\rceil - Q_{\mathrm{E}}(x)|^p\big)^{\frac{1}{p}}$, respectively. For notational convenience we define $f(x) = Q_{\mathrm{E}}(x) + \alpha Q_{\mathcal{D}_p}(x)$.

Let $\Omega = \{\omega_1, \omega_2\}$ and let the distribution of the random variable $\xi = h$ be given by $\mathbb{P}(h = 0) = 1 - \mathbb{P}(h = \frac{1}{2}) = q$ for some $q \in (0,1)$. Then, we have $Q_{\mathrm{E}}(x) = x + q\lceil x\rceil + (1-q)\lceil x + \frac{1}{2}\rceil$ and $Q_{\mathcal{D}_p}(x) = \big(q(1-q)^p + (1-q)q^p\big)^{\frac{1}{p}} |\lceil x\rceil - \lceil x+\frac{1}{2}\rceil|$.

Let $\{x_k\}$ be a nonnegative sequence converging to zero. For k' sufficiently large the members of the sequence $\{x_k\}$ become smaller than $\frac{1}{2}$ for all $k > k'$. Thus, $Q_{\mathcal{D}_p}(x_k)$ equals zero for all $k > k'$ and it holds $\lim_{k\to\infty} Q_{\mathcal{D}_p}(x_k) = 0$. However, $Q_{\mathcal{D}_p}(0)$ is strictly positive for $q \in (0,1)$. This means $Q_{\mathcal{D}_p}$ is not l.s.c. at 0.

Now, we consider f. We have $\lim_{k\to\infty} f(x_k) = \lim_{k\to\infty} Q_{\mathrm{E}}(x_k) = 1 + \lim_{k\to\infty} x_k = 1$. For the limit $x_o = 0$ we compute $f(x_o) = 1 - q + \alpha\big(q(1-q)^p + (1-q)q^p\big)^{\frac{1}{p}}$. As soon as α gets greater than the expression

$$g(q) := \big(\sqrt[p]{q^{1-p}(1-q)^p + (1-q)}\big)^{-1},$$

we obtain $f(0) > 1$, f is not l.s.c., and the global infimum 1 of problem (5.9) is not attained.

Note that it holds $\lim_{q\to 0} g(q) = 0$ for $p > 1$. Thus, for any $\alpha > 0$ we can choose $q \in (0,1)$ such that the pathologies occur. For $p = 1$, the function f is not l.s.c. when α is greater than $(2-2q)^{-1}$ which is possible for $\alpha > \frac{1}{2}$, only.

Likewise, the semideviation of order p is not l.s.c.

Example 5.11 *For $\alpha > 0$ and $p \in \mathbb{N}$ we consider the problem*

$$\inf_x \{Q_{\mathrm{E}}(x) + \alpha Q_{\mathcal{D}_p^+}(x) \ : \ 0 \leq x \leq \frac{1}{2}\} \tag{5.10}$$

with the specifications of Example 5.10. We obtain $Q_{\mathcal{D}_p^+}(x) = (1-q)^{\frac{1}{p}} q(\lceil x + \frac{1}{2}\rceil - \lceil x \rceil)^{\frac{1}{p}}$. *Let $\{x_k\}$ be a nonnegative sequence converging to 0. $Q_{\mathcal{D}_p^+}$ is not l.s.c. at $x_o = 0$ because it holds $\lim_{k\to\infty} Q_{\mathcal{D}_p^+}(x_k) = 0$ and $Q_{\mathcal{D}_p^+}(0) = (1-q)^{\frac{1}{p}} q > 0$.*

Let us define $f(x) := Q_{\mathrm{E}}(x) + \alpha Q_{\mathcal{D}_p^+}(x)$, where $Q_{\mathrm{E}}(x) = x + q\lceil x \rceil + (1-q)\lceil x + \frac{1}{2}\rceil$. We have $\lim_{k\to\infty} f(x_k) = \lim_{k\to\infty} Q_{\mathrm{E}}(x_k) = 1 + \lim_{k\to\infty} x_k = 1$. For the limit $x_o = 0$ we compute $f(0) = 1 - q + \alpha(1-q)^{\frac{1}{p}} q$.

If α is greater than 1 then there exists $q \in (0,1)$ such that $f(0) > \lim_{k\to\infty} f(x_k)$. Consequently, f is not l.s.c. at 0 and the infimum of problem (5.10) is not attained.

Based on the Example 5.10 we give another one that provides a sequence of probability measures for which the corresponding optimization problems have no minimum. In fact, we consider the random variables and probability distributions associated with the probability measures. Recall that the convergence in distributions of random variables is equivalent to the weak convergence of the associated probability measures, cf. Section 3.3.

Example 5.12 *We take on the dimensions, matrices, and the recourse problem of Example 5.10. But now we consider the sequence of random variables $\{h_k\}$ with distributions defined by $\mathbb{P}(h_k = \frac{j}{2k}) = \frac{1}{k+1}$ for $j = 0, \ldots, k$ and for all $k \in \mathbb{N}$. Consequently, it holds $\mathbb{P}(h_k = 0) = \frac{1}{k+1} = 1 - \mathbb{P}(h_k \in (0, \frac{1}{2}])$ and therefore*

$$\mathbb{P}(\phi(h_k + x) = \lceil x \rceil) = 1 - \mathbb{P}(\phi(h_k + x) = \lceil x + \frac{1}{2}\rceil) = \frac{1}{k+1}$$

for $x \in [0, \frac{1}{2}]$ and $k \in \mathbb{N}$. For each k, this is just a special case of Example 5.10 with $p = \frac{1}{k+1}$. Thus, the expected value and the risk function are $Q_{\mathrm{E}}^k(x) = x + \frac{1}{k+1}\lceil x \rceil + \frac{k}{k+1}\lceil x + \frac{1}{2}\rceil$ and $Q_{\mathcal{D}_p}^k(x) = \left(\frac{1}{k}(\frac{k}{k+1})^p + (\frac{k}{k+1})\frac{1}{k}^p\right)^{\frac{1}{p}} |\lceil x \rceil - \lceil x + \frac{1}{2}\rceil|$ for all $k \in \mathbb{N}$. The minimum

$$\min_{x\in X}(f_k(x) := Q_{\mathrm{E}}^k(x) + \alpha Q_{\mathcal{D}_p}^k(x)) \qquad k \in \mathbb{N} \tag{5.11}$$

does not exist for weights α greater than $g(\frac{1}{k+1})$ with g defined as in Example 5.10. Since $g(\frac{1}{k+1})$ is bounded by 1 for all $k \in \mathbb{N}$, none of the problems (5.11) attains its minimum if α is greater than 1.

Consider now a random variable h which is uniformly distributed on the interval $[0, \frac{1}{2}]$. For $x \in [0, \frac{1}{2}]$ the distribution function of the random variable

$\lceil h + x \rceil$ is given by

$$F^x(z) = \begin{cases} 0 & z < 1 \\ 1 & z \geq 1 \end{cases}$$

The sequence of distribution functions of the random variables $(\lceil h_k + x \rceil)_{k \in \mathbb{N}}$

$$F_k^x(z) = \begin{cases} 0 & z < 0 \\ \frac{1}{k+1} & 0 \leq z < 1 \\ 1 & z \geq 1 \end{cases} \quad \text{if } x = 0 \text{ and } F_k^x(z) = \begin{cases} 0 & z < 1 \\ 1 & z \geq 1 \end{cases} \quad \text{otherwise}$$

converges to the distribution function F^x for each $x \in [0, \frac{1}{2}]$. This implies the weak convergence of the corresponding probability measures.

Plugging the random variable h in the above optimization problem and calculating $\mathrm{E}\lceil h + x \rceil = 1$ as well as $\mathrm{E}|\lceil h + x \rceil - 1|^p = 0$ for all $x \in [0, \frac{1}{2}]$ yields $f(x) = x + 1$. Obviously $\min\{f(x) : x \in [0, \frac{1}{2}]\}$ exists.

The latter example displays the consequences of $Q_{\mathcal{D}_p}$ not being lower semicontinuous for the stability. While the problem $\min\{Q_{\mathrm{E}}(x) + \alpha Q_{\mathcal{D}_p}(x),\, x \in X\}$ has a solution for the limit probability measure μ it lacks a solution for all members of a sequence of probability measures weakly converging to μ. Therefore an approximation of continuous probability distributions by discrete ones is not justified. In general, the problem may occur whenever the risk measure involves operations not preserving lower semicontinuity.

The above pathologies have consequences for the stability of the corresponding mean-risk models. We consider Q_{E} and $Q_{\mathcal{R}}$ as functions of $x \in \mathbb{R}^n$ as well as of $\mu \in \mathcal{P}(B^l)$ - the set of Borel probability measures on $\mathbb{R}^l$ endowed with the topology of weak convergence, see Definition 3.9.

In general, the mean-risk problem (5.8) is nonconvex. Again, we consider the stability of CLM sets.

Proposition 5.13 *Let $p \in \mathbb{N}$ and $\eta_o, r \in \mathbb{R}$. Let $\mathcal{R}$ be one of the risk measures $\mathcal{D}_p$, $\mathcal{D}_p^+$, or E_p^η. Let $r > p$, $C > 0$, $X \subset \mathbb{R}^n$, $\Delta \subset \mathcal{P}(\mathbb{R}^l)$, and $\mu_o \in \Delta$. Assume A 5.1, A 5.2, and $\int_{\mathbb{R}^l} \|h\|^r + \|T\|^r \; \mu(d\xi) < C$ for all $\mu \in \Delta$.*

Let V be some bounded open subset of X. Assume p and α are such that $Q_{\mathrm{E}}(\cdot, \xi) + \alpha Q_{\mathcal{R}}(\cdot, \xi)$ is l.s.c. on X for all $\mu \in \Delta$. Let $\varphi(\mu) := \inf\{Q_{\mathrm{E}}(x, \mu) + \alpha Q_{\mathcal{R}}(x, \mu) : x \in clV\}$ and $\psi(\mu) := \{x \in clV : Q_{\mathrm{E}}(x, \mu) + \alpha Q_{\mathcal{R}}(x, \mu) = \varphi(\mu)\}$. Assume $\psi(\mu_o)$ is a CLM set w.r.t. V for $\mu_o \in \Delta$. Then

(i) *$\varphi : \Delta \to \mathbb{R}$ is continuous at μ_o,*

(ii) *$\psi : \Delta \to 2^{\mathbb{R}^n}$ is Berge upper semicontinuous at μ_o, and*

(iii) *there is a neighborhood U of μ_o in Δ such that $\psi(\mu)$ is a CLM set w.r.t. V for all $\mu \in U$.*

Proof We use Proposition 3.18. First, $Z(x, \cdot)$ is measurable for all $x \in \mathbb{R}^n$, cf. Lemma 5.5. For $x \in X$, Lemma 5.6 yields the estimate $|Z(x,\xi)|^r \leq C(\|h\|^r + \|T\|^r) + C_1$ with positive constants C and C_1 independent of x. This gives the uniform integrability of $|Z(x,\xi)|^r$ w.r.t Δ.

By Lemma 5.4 and the assumption $\mu_o(D_\phi(x)) = 0$ for all $x \in X$ we obtain that the set of discontinuity points $D_{\mathrm{Z}}(x)$ of $Z(\cdot,\cdot)$ is of μ-measure zero for all $x \in X$. Thus, the assumptions of Proposition 3.18 are fulfilled. □

In Example 5.12 we have shown that the mean-central-deviation problem does not behave stable. In the example, the assertion (ii) of Proposition 5.13 turns meaningless because the sets $\psi(\mu_k)$ are empty for all $k \in \mathbb{N}$, and thus, for any supset G of $\psi(\mu_o)$ we trivially obtain $G \supset \psi(\mu_k)$, too. Moreover $Q_{\mathcal{D}_\mathrm{p}}$ is not l.s.c., and therefore, the CLM assertion (iii) does not hold.

Remark 5.14 *Under the assumptions of Proposition 5.13 the ε-optimal set mapping*

$$\psi_\varepsilon(\mu) := \{x \in clV : Q_\mathrm{E} + \alpha Q_\mathcal{R}(x,\mu) \leq \varphi(\mu) + \varepsilon\}$$

as a multifunction from Δ to $2^{\mathbb{R}^n}$ is Berge l.s.c., see Remark 3.19. This also holds true for weights α and orders p for which there exists $\mu \in \Delta$ such that the function $Q_\mathrm{E}(\cdot,\xi) + \alpha Q_\mathcal{R}(\cdot,\xi)$ is not l.s.c. on X.

5.5 Summary

We have shown that beyond conceptual considerations, the choice of risk measures in special optimization problems requires additional care. We summarize some of our results in Table 5.1, cf. also the Tables 2.2, 2.3, and 4.1.

The last two columns document the weights and the orders for which the mean-risk models in the first column are l.s.c. We have seen that the lower semicontinuity leads to positive results concerning the qualitative stability of the models. In the next chapter, we will propose algorithms for all *linear* l.s.c.

Mean-risk problem	Weight	Order
$Q_\mathrm{E} + \alpha\ Q_{\mathcal{D}_\mathrm{p}}$	$\alpha \in (0, \frac{1}{2}]$	$p = 1$
$Q_\mathrm{E} + \alpha\ Q_{\mathcal{D}_\mathrm{p}^+}$	$\alpha \in (0, 1]$	$p = 1$
$Q_\mathrm{E} + \alpha\ Q_{\mathrm{E}_\mathrm{p}^\eta}$	$\alpha \in \mathbb{R}_+$	$p \in \mathbb{N}$

Table 5.1: Lower semicontinuity of the risk measures

mean-risk models.

6 Algorithms

6.1 Scope

Once we have clarified some structural properties of the mean-risk problems in the framework of stochastic programming with mixed-integer recourse, we turn to algorithms for these problems.

Throughout this section we assume that ξ has a discrete probability distribution with a finite number S of probability atoms (h_j, T_j) and corresponding probabilities π_j, $\sum_{j=1}^{S} \pi_j = 1$. Moreover, we assume that all the problems under consideration possess a finite solution. As shown in Chapter 5, this can be established by the requirements of complete recourse and dual feasibility of the linear programming relaxation of the recourse program, and by taking X a nonempty, compact subset of $\mathbb{R}^n$, cf. Proposition 5.7 and 5.9. Then, the optimal values $z_j(x)$ of the S single-scenario problems associated with the probability atoms (h_j, T_j)

$$\phi_j(x) := \min_{y_j \in \mathbb{Z}_+^m \times \mathbb{R}_+^{m'}} \{cx + qy_j \ : \ T_j x + W y_j = h_j\}, \qquad j = 1, \ldots, S, \qquad (6.1)$$

and therefore all the statistics under consideration like the moments and the central moments are finite for all $x \in X$.

Recall that both, the first- and the second-stage variables may take mixed-integer values. In this chapter, we suspend the splitting of the second-stage variables into real and integer part, by the assignments $W := (W, W')$ and $q := (q, q')$.

As we want to benefit from the algorithms available for linear and mixed-integer linear programs, we focus in this section on the expected value problem and on mean-risk problems involving risk measures of order 1.To keep notations simple, we refer to the mean-risk problem with the risk measure expected excess of a target of order 1 as the mean-expected-excess problem.

We exclude the mean-central-deviation model since there is no mixed-integer linear or mixed-integer quadratic program with rational data for it. If there was one, then the infimum of such a program had to be attained provided the infimum is finite, cf. Proposition 5.1 and Theorem 2.2. in Bank and Hansel (1984). However, Example 5.10 shows that this is not the case with the exception of the central deviation of order 1 and weights $\alpha \in (0, \frac{1}{2}]$, cf. Proposition 5.9.

As shown in Lemma 2.11, the exceptional case is already covered by the mean-absolute-semideviation-model.

6.2 Expected value model

When the probability distribution of ξ is discrete, we can write the expected value problem as

$$\min_{x \in X,\, y_j \in \mathbb{Z}_+^m \times \mathbb{R}_+^{m'}} \{cx + \sum_{j=1}^{S} \pi_j q y_j \ : \ T_j x + W y_j = h_j, \quad \forall j\}, \tag{6.2}$$

cf. Birge and Louveaux (1997), Kall and Wallace (1994), and Prekopa (1995). Now, the expected recourse function Q_{E} reads

$$Q_{\mathrm{E}}(x) = cx + \min_{y_j \in \mathbb{Z}_+^m \times \mathbb{R}_+^{m'}} \{\sum_{j=1}^{S} \pi_j q y_j \ : \ T_j x + W y_j = h_j, \quad \forall j\},$$

for all $x \in \mathbb{R}^n$. The program (6.2) is a large-scale deterministic mixed-integer linear program (MILP) with a block-angular structure. A recent and comprehensive overview of existing algorithms for problem (6.2) is provided in Louveaux and Schultz (2003). To mention some of the algorithmic approaches we refer to van der Vlerk (1995) for simple recourse models ($W = (I, -I)$), to Laporte and Louveaux (1993) for two-stage models with a binary first stage, and to Ahmed et al. (2000) for models with an integer second stage and a fixed technology matrix T. We shall see that our applications have a rather general structure in the sense that both the first and second stage are (mixed)-integer, cf. Chapter 7. An algorithm for problem (6.2) in its general form has been proposed in Carøe and Schultz (1999). It works on the expense of a branching on continuous first-stage variables. Here, we describe the latter algorithm and in the following sections we clarify its applicability to the mean-risk models under consideration.

By introducing copies of the first-stage variables, an equivalent formulation of (6.2) is given by

$$\min_{x_j, y_j} \{\sum_{j=1}^{S} \pi_j (c x_j + q y_j) \ : \ x_1 = \ldots = x_S,\, (x_j, y_j) \in M_j,\, \forall j\} \tag{6.3}$$

where $M_j = \{(x_j, y_j) : T x_j + W y_j = h_j, x_j \in X, y_j \in Y\},\ j = 1, \ldots, S$.

Considering the constraint matrix of (6.3) (cf. Figure 6.1), we can identify S single-scenario subproblems solely coupled by the equality (*nonanticipativity*) constraints on the copies of the first-stage variables and written as $\sum_{j=1}^{S} H_j x_j = 0$, where $H = (H_1, \ldots, H_j)$. The problem decomposes when we relax the nonanticipativity constraints.

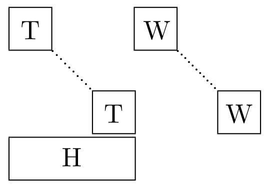

Figure 6.1: Constraints of (6.3)

Upper bounds on the optimal value can be obtained by heuristics based on the solutions for the subproblems. We get a lower bound by solving the Lagrangian dual, which is a nonlinear concave maximization

$$z_{\mathrm{LD}} := \max_{\lambda \in \mathbb{R}^l} \min_{x_j, y_j} \{ \sum_{j=1}^{S} \pi_j (c x_j + d y_j) + \lambda \sum_{j=1}^{S} H_j x_j \ : \ (x_j, y_j) \in M_j, \, \forall j \}. \tag{6.4}$$

In general, the involved integrality restrictions lead to an optimality gap. If we are not satisfied with the bounds given by the above method, we can elaborate a branch-and-bound algorithm that successively reestablishes the equality of the components of the first-stage vector. Let $\mathcal{P}$ denote a list of problems.

`Algorithm SD`: Scenario decomposition (Carøe and Schultz (1999))

STEP 1 Initialization: Set $z^* = \infty$ and let $\mathcal{P}$ consist of problem (6.2).

STEP 2 Termination: If $\mathcal{P} = \emptyset$ then x^* with $z^* = Q_{\mathrm{E}}(x^*)$ is optimal.

STEP 3 Node selection: Select and delete a problem P from $\mathcal{P}$ and solve its Lagrangian dual. If the associated optimal value $z_{\mathrm{LD}}(P)$ equals infinity (infeasibility of a subproblem) go to *STEP 2*.

STEP 4 Bounding: If $z_{\mathrm{LD}}(P)$ is greater than z^* go to *STEP 2*. Otherwise proceed as follows; if the first-stage solutions x_j, $j = 1, \ldots, S$, of the subproblems are

- identical, then set $z^* := \min\{z^*, Q_{\mathrm{E}}(x_j)\}$, delete all $P' \in \mathcal{P}$ with $z_{\mathrm{LD}}(P') \geq z^*$ and go to *STEP 2*.
- not identical, then compute a suggestion $\hat{x} := Heu(x_1, \ldots, x_S)$ using some heuristic. Set $z^* := \min\{z^*, Q_{\mathrm{E}}(\hat{x})\}$ and delete all $P' \in \mathcal{P}$ with $z_{\mathrm{LD}}(P') \geq z^*$.

STEP 5 Branching: Select a component $x_{(k)}$ of x and add two new problems to $\mathcal{P}$ that differ from P by the additional constraint $x_{(k)} \leq \lfloor x_{(k)} \rfloor$ and $x_{(k)} \geq \lfloor x_{(k)} \rfloor + 1$, respectively, if $x_{(k)}$ is integer, or $x_{(k)} \leq x_{(k)} - \varepsilon$ and $x_{(k)} \geq x_{(k)} + \varepsilon$, respectively, if $x_{(k)}$ is continuous. $\varepsilon > 0$ has to be chosen such that the two new problems have disjoint subdomains. Go to *STEP 3*.

The algorithm is finite if X is bounded and if some stopping criterion is employed that prevents the algorithm from endless branching on the continuous components of x, see Carøe and Schultz (1999).

The heuristic in *STEP 4* can be of general nature, for instance, rounding the average of the first-stage solutions x_j, $j = 1, \ldots, S$, of the subproblems or using the subsolution that occurred most frequently, or it can be problem specific. The function Q_{E} is evaluated at x by fixing the first stage to x, solving the scenario many subproblems, and calculating the expected value of the corresponding optimal values. Thus, infeasible suggestion are identified immediately.

6.3 Mean-expected-excess model

We consider the mean-expected-excess model

$$\min_{x \in X} Q_{\mathrm{E}}(x) + \alpha\; Q_{\mathrm{E}_{\mathrm{p}}^{\eta}}(x) \qquad \alpha > 0, \tag{6.5}$$

where the risk function is

$$Q_{\mathrm{E}_{\mathrm{p}}^{\eta}}(x) = \sum_{j=1}^{S} \pi_j \max\{cx + \phi(h_j - T_j x) - \eta_o, 0\}.$$

for some fixed target $\eta_o \in \mathbb{R}$. Recall the recourse function $\phi(t) = \min\{qy \;:\; Wy = t,\, y \in \mathbb{Z}_+^m \times \mathbb{R}_+^{m'}\}$. We look for a mixed-integer linear program equivalent to problem (6.5).

Lemma 6.1 *Let ξ have a finite discrete probability distribution with the probability atoms (h_j, T_j), associated probabilities π_j for $j = 1, \ldots, S$, and $\sum_{j=1}^{S} \pi_j = 1$. Let $x \in \mathbb{R}^n$. Then, it holds*

$$Q_{\mathrm{E}_{p}^{\eta}}(x) = \min_{\substack{v_j \in \mathbb{R}_+,\\ y_j \in \mathbb{Z}_+^m \times \mathbb{R}_+^{m'}}} \{\sum_{j=1}^{S} \pi_j v_j \;:\; v_j \geq cx + qy_j - \eta_o,\, Wy_j = h_j - T_j x \quad \forall j\}.$$

Proof Let $x \in \mathbb{R}^n$. By definition we have

$$Q_{\mathrm{E}_{\mathrm{p}}^{\eta}}(x) = \sum_{j=1}^{S} \pi_j \max\{cx + \min_{y}\{qy \;:\; Wy = h_j - T_j x,\, y \in \mathbb{Z}_+^m \times \mathbb{R}_+^{m'}\} - \eta_o, 0\}.$$

First, we insert the constant term $cx - \eta_o$ (x is fixed) into the recourse program ϕ and obtain

$$Q_{\mathrm{E}_{\mathrm{p}}^{\eta}}(x) = \sum_{j=1}^{S} \pi_j \max\{\min_{y}\{cx + qy - \eta_o \;:\; Wy = h_j - T_j x,\, y \in \mathbb{Z}_+^m \times \mathbb{R}_+^{m'}\}, 0\}.$$

Then, we introduce variables v_j and y_j for each scenario and use $\max\{a, b\} = \min_v\{v : v \geq a,\, v \geq b\}$. This yields

$$Q_{\mathrm{E}_{\mathrm{p}}^{\eta}}(x) = \sum_{j=1}^{S} \pi_j \min_{v_j}\big\{v_j \;:\; v_j \geq \min_{y_j}\{cx + qy_j - \eta_o \;:\; Wy_j = h_j - T_j x\},\, v_j \geq 0\big\},$$

where the inner minimization is carried out over the set $\mathbb{Z}_+^m \times \mathbb{R}_+^{m'}$. Merging the two minimizations, we can restate the function $Q_{\mathrm{E}_{\mathrm{p}}^{\eta}}$ as

$$Q_{\mathrm{E}_{\mathrm{p}}^{\eta}}(x) = \sum_{j=1}^{S} \pi_j \min_{v_j, y_j}\{v_j \;:\; v_j \geq cx + qy_j - \eta_o,\, Wy_j = h_j - T_j x,\, v_j \geq 0\}.$$

Finally, we use the independence among the different single-scenario problems

$$Q_{\mathrm{E}_{\mathrm{p}}^{\eta}}(x) = \min_{v_j, y_j}\{\sum_{j=1}^{S} \pi_j v_j \;:\; v_j \geq cx + qy_j - \eta_o,\, Wy_j = h_j - T_j x,\, v_j \geq 0 \quad \forall j\},$$

where y_j is in the set $\mathbb{Z}_+^m \times \mathbb{R}_+^{m'}$, again. This verifies the assertion. □

Using Lemma 6.1, the mean-risk problem (6.5) takes the form

$$\begin{aligned} \min_{\substack{x \in X,\, v_j \in \mathbb{R}_+, \\ y_j \in \mathbb{Z}_+^m \times \mathbb{R}_+^{m'}}} \quad & cx + \sum_{j=1}^{S} \pi_j q y_j + \alpha \sum_{j=1}^{S} \pi_j v_j \\ \text{s.t.} \quad & v_j \geq cx + qy_j - \eta_o,\, T_j x + W y_j \geq h_j,\, \forall j. \end{aligned} \tag{6.6}$$

The mathematical program (6.6) has the same block-angular structure as the expected value program (6.2). In particular, there are no constraints linking the individual scenarios. Thus, the scenario decomposition algorithm introduced in the previous section can also be applied to the mean-expected-excess problem.

Using the variable transformation $v_j := v_j + \eta_o$ we obtain a reformulation of problem (6.6)

$$\begin{aligned} \min_{\substack{x \in X,\, v_j \in \mathbb{R}, \\ y_j \in \mathbb{Z}_+^m \times \mathbb{R}_+^{m'}}} \quad & cx + \sum_{j=1}^{S} \pi_j q y_j - \alpha\eta_o + \alpha \sum_{j=1}^{S} \pi_j v_j \\ \text{s.t.} \quad & v_j \geq cx + qy_j,\, v_j \geq \eta_o,\, T_j x + W y_j \geq h_j,\, \forall j. \end{aligned} \tag{6.7}$$

In a similar form, this second formulation will reappear as an auxiliary problem in the next section.

6.4 Mean-absolute-semideviation model

Lower bounds Before we turn to the discrete model, we provide a result that will be used to construct lower bounds in a branch-and-bound algorithm.

Let $\alpha \in (0,1]$ throughout this section. Recall the mean-absolute-semideviation model

$$\min_{x\in X} Q_{\mathrm{E}}(x) + \alpha\ Q_{\mathcal{D}_1^+}(x) \tag{6.8}$$

and the risk function

$$Q_{\mathcal{D}_1^+}(x) = \int_{\mathbb{R}^l} \max\{cx + \phi(h - Tx) - Q_{\mathrm{E}}(x), 0\}\ \mu(d\xi)$$

defined for general probability distributions of ξ. Once again we employ the relation $\max\{a-b,0\} = \max\{a,b\} - b$ for $a, b \in \mathbb{R}$, and obtain

$$Q_{\mathcal{D}_1^+}(x) = \int_{\mathbb{R}^l} \max\{cx + \phi(h - Tx), Q_{\mathrm{E}}(x)\}\ \mu(d\xi) - Q_{\mathrm{E}}(x).$$

Thus, problem (6.8) is equivalent to

$$z_{\mathrm{AS}} := \min_{x\in X}(1-\alpha)Q_{\mathrm{E}}(x) + \alpha \int_{\mathbb{R}^l} \max\{cx + \phi(h - Tx), Q_{\mathrm{E}}(x)\}\ \mu(d\xi). \tag{6.9}$$

Along with this reformulation we consider the auxiliary problem

$$z_\eta := \min_{x\in X}(1-\alpha)Q_{\mathrm{E}}(x) + \alpha \int_{\mathbb{R}^l} \max\{cx + \phi(h - Tx), \eta\}\ \mu(d\xi). \tag{6.10}$$

where we replace the expected value in the second term by a target $\eta \in \mathbb{R}$.

Lemma 6.2 *Assume A 4.3, A 5.1, and A 5.2. Let* $z_E := \min_{x\in X} Q_{\mathrm{E}}(x)$*. Then, it holds*

(i) $z_E \leq z_\eta$ *for all* $\eta \in \mathbb{R}$ *and*

(ii) $z_\eta \leq z_{AS}$ *for all* $\eta \leq z_E$.

Proof Let $\eta \in \mathbb{R}$. We define the optimal value functions

$$F_{\mathrm{AS}}(x) := (1-\alpha)Q_{\mathrm{E}}(x) + \alpha Q_{\mathcal{D}_1^+}(x)$$

and

$$F_\eta(x) := (1-\alpha)Q_{\mathrm{E}}(x) + \alpha \int_{\mathbb{R}^l} \max\{cx + \phi(h - Tx), \eta\}\ \mu(d\xi).$$

of the problems (6.9) and (6.10), respectively.
ad (i) We have $\max\{cx + \phi(h - Tx), \eta\} \geq cx + \phi(h - Tx)$ for all $x \in \mathbb{R}^n$, all $h \in \mathbb{R}^s$, and all $T \in \mathbb{R}^{s\times n}$. Thus, it holds

$$F_\eta(x) - Q_{\mathrm{E}}(x) = \alpha\Big(\int_{\mathbb{R}^l} \max\{cx + \phi(h - Tx), \eta\} - cx - \phi(h - Tx)\ \mu(d\xi)\Big) \geq 0$$

for all $x \in \mathbb{R}^n$. This proves (i).

ad (ii) If it holds $\eta \leq Q_{\mathrm{E}}(x)$ then we have $\max\{cx+\phi(h-Tx),\eta\} \leq \max\{cx+\phi(h-Tx), Q_{\mathrm{E}}(x)\}$ for all $x \in \mathbb{R}^n$, all $h \in \mathbb{R}^s$, and all $T \in \mathbb{R}^{s\times n}$. This yields

$$\begin{aligned} F_{\mathrm{AS}}(x) - F_\eta(x) \;&=\; \alpha\Big(\int_{\mathbb{R}^l} \max\{cx+\phi(h-Tx), Q_{\mathrm{E}}(x)\}\;\mu(d\xi) \\ &\qquad - \int_{\mathbb{R}^l} \max\{cx+\phi(h-Tx),\eta\}\;\mu(d\xi)\Big) \;\geq\; 0 \end{aligned}$$

for all $x \in \mathbb{R}^n$. This proves assertion (ii). □

Remark 6.3 *The second statement of the previous lemma is a consequence of the monotonicity of $F_\eta(x)$ w.r.t. the parameter η. Thus, the quality of the lower bound z_η that we obtain by solving problem (6.10) increases with η. We may ask when this lower bound is strictly better than the trivial bound z_E. In general, we have $z_\eta \geq z_E$.*

Let $x \in X$. It holds $F_\eta(x) > Q_{\mathrm{E}}(x)$ if there exists a set $M \subset \mathbb{R}^l$ with strictly positive μ-measure such that $cx+\phi(h-Tx) < \eta$ for all $\xi \in M$. Because then we obtain

$$\begin{aligned} &\int_{\mathbb{R}^l} \max\{cx+\phi(h-Tx),\eta\} - cx+\phi(h-Tx)\;\mu(d\xi) \\ \geq\;& \int_M \max\{cx+\phi(h-Tx),\eta\} - cx+\phi(h-Tx)\;\mu(d\xi) \\ \geq\;& \int_M \eta - cx+\phi(h-Tx)\;\mu(d\xi) \quad > \quad 0. \end{aligned}$$

If this condition is fulfilled for all $x \in X$, we have $z_\eta > z_E$.

Let us consider the case $\eta = z_E$. Assume there is no $x \in X$ for which the above condition holds true, i.e. $cx+\phi(h-Tx) \geq \eta = z_E$ for all $\xi \in \mathbb{R}^l \setminus N(x)$ and all $x \in X$ where $\mu(N(x)) = 0$ for all $x \in X$. Let $x_E \in \arg\min\{Q_{\mathrm{E}}(x) : x \in X\}$. Compiling the conditions $Q_{\mathrm{E}}(x_E) = z_E$ and $cx_E + \phi(h-Tx_E) \geq z_E$ for all $\xi \in \mathbb{R}^l \setminus N(x_E)$ yields $\phi(h-Tx_E) = C$, $C \in \mathbb{R}$, for all $(h,T) = \xi \in \mathbb{R}^l \setminus N(x_E)$. This is the exceptional case when there is no deviation among the scenarios and when it holds $Q_{\mathcal{D}_1^+}(x_E) = 0$.

Algorithm Now, we return to the case when ξ is a discrete random variable with S probability atoms and corresponding probabilities π_j. Let $\phi_j(x) := \phi(h_j - T_j x)$ for $j = 1,\ldots,S$. Then, problem (6.9) turns into

$$\min_{x\in X}\Big\{(1-\alpha)\big(cx+\sum_{j=1}^S \pi_j\phi_j(x)\big)+\alpha\sum_{j=1}^S \pi_j \max\{cx+\phi_j(x), cx+\sum_{i=1}^S \pi_i\phi_i(x)\}\Big\},$$

or by writing the scenario independent term cx separately into

$$\min_{x\in X}\Big\{cx + (1-\alpha)\sum_{j=1}^S \pi_j\phi_j(x) + \alpha\sum_{j=1}^S \pi_j \max\{\phi_j(x), \sum_{i=1}^S \pi_i\phi_i(x)\}\Big\}. \qquad (6.11)$$

An argument following similar lines as the proof of Lemma 6.1 confirms that this is equivalent to the mixed-integer linear program

$$\min_{\substack{x \in X,\ v_j \in \mathbb{R},\\ y_j \in \mathbb{Z}_+^m \times \mathbb{R}_+^{m'}}} \quad cx + (1-\alpha)\sum_{j=1}^{S} \pi_j q y_j + \alpha \sum_{j=1}^{S} \pi_j v_j \tag{6.12}$$

$$\text{s.t.} \quad v_j \geq q y_j,\ v_j \geq \sum_{i=1}^{S} \pi_i q y_i,\ T_j x + W y_j = h_j,\ \forall j.$$

The scenario decomposition does not work for problems in which the second-stage variables are linked by common constraints as in the mean-semideviation model (6.12). Here the constraints $v_j \geq \sum_{i=1}^{S} \pi_i q y_i$ for $j = 1, \ldots, S$, are the linking ones. Dropping these constraints transfers problem (6.12) into the expected value problem (6.2). Therefore, the solution to the expected value problem provides a lower bound for the mean-absolute-semideviation problem, cf. also Lemma 6.2. It is possible to use this bound in a branch-and-bound algorithm similar to the scenario decomposition algorithm, cf. Section 6.5. However, we may obtain a strictly better bound by solving the parametric mathematical program

$$(P_\eta) \qquad \min_{\substack{x \in X,\ v_j \in \mathbb{R},\\ y_j \in \mathbb{Z}_+^m \times \mathbb{R}_+^{m'}}} \quad (1-\alpha)cx + (1-\alpha)\sum_{j=1}^{S} \pi_j q y_j + \alpha \sum_{j=1}^{S} \pi_j v_j$$

$$\text{s.t.} \quad v_j \geq cx + q y_j,\ v_j \geq \eta,\ T_j x + W y_j = h_j,\ \forall j$$

for suitable values of η, cf. (ii) in Lemma 6.2 and Remark 6.3. Following the lines of Lemma 6.1, we can verify that problem (6.10) turns into problem (P_η) when the probability distribution of ξ is discrete.

We will use this lower bound for the mean-absolute-semideviation model in a branch-and-bound algorithm similar to the scenario decomposition. The selection of the parameter $\eta \in \mathbb{R}$ can be specified in different ways. Recall the requirements $\eta \leq \min\{Q_{\mathrm{E}}(x) : x \in X\}$. So one option would be $\eta := z_{\mathrm{LD}}$, where z_{LD} is the optimal value of the Lagrangian dual to the expected value model. Another computationally less costly option is to put η equal to z_{WS}, where z_{WS} is the optimal value of the wait-and-see model

$$z_{\mathrm{WS}} := \sum_{j=1}^{S} \pi_j \min\{cx_j + q y_j \ :\ T_j x_j + W y_j = h_j,\ x_j \in X,\ y_j \in \mathbb{Z}_+^m \times \mathbb{R}_+^{m'}\}. \tag{6.13}$$

Before we give a formal description of the algorithm we state the Lagrangian dual of problem (P_η)

$$\max_{\lambda\in\mathbb{R}^l} \min_{\substack{x_j \in X,\, v_j \in \mathbb{R},\\ y_j \in \mathbb{Z}_+^m \times \mathbb{R}_+^{m'}}} (1-\alpha)\sum_{j=1}^{S}\pi_j(cx_j+qy_j)+\alpha\sum_{j=1}^{S}\pi_j v_j+\lambda\sum_{j=1}^{S}H_j x_j$$
$$\text{s.t.} \quad v_j \geq cx_j+qy_j,\ v_j\geq \eta,\ Tx_j+Wy_j=h_j \quad \forall j.$$

Let $\mathcal{P}$ denote a list of problems in a branch-and-bound tree.

`Algorithm ASD`: Mean-absolute-semideviation model

STEP 1 Initialization: Select η according to one of the above options. Set $z^* := \infty$ and $\mathcal{P} := \{(P_\eta)\}$.

STEP 2 Termination: If $\mathcal{P}$ is empty then x^* with $z^* = Q_{\mathrm{E}}(x^*) + \alpha Q_{\mathcal{D}_1^+}(x^*)$ is optimal.

STEP 3 Node selection: Select and delete a problem $P \in \mathcal{P}$. Solve its Lagrangian dual. If the optimal value thereof $z_{\mathrm{LD}}(P)$ is equal to infinity (infeasibility of a subproblem) go to *STEP 2*.

STEP 4 Bounding: If $z_{\mathrm{LD}}(P)$ is greater than z^* go to *STEP 2*. Otherwise proceed as follows; if the first-stage solutions x_j, $j = 1, \ldots, S$, of the subproblems are

- identical, then x_j is a feasible solution. Update the objective function value $z^* := \min\{z^*, Q_{\mathrm{E}}(x_j) + \alpha Q_{\mathcal{D}_1^+}(x_j)\}$.
- not identical, then compute a suggestion $\hat{x} := Heu(x_1, \ldots, x_S)$ using some heuristic. Set $z^* := \min\{z^*, Q_{\mathrm{E}}(\hat{x}) + \alpha Q_{\mathcal{D}_1^+}(\hat{x})\}$.

Delete all problems $P' \in \mathcal{P}$ with $z_{\mathrm{LD}}(P') \geq z^*$.

STEP 5 Branching and constraint adaptation: Select a component $x_{(k)}$ of x and add two new problems P_1 and P_2 to $\mathcal{P}$ that differ from P by the additional constraint $x_{(k)} \leq \lfloor x_{(k)} \rfloor$ and $x_{(k)} \geq \lfloor x_{(k)} \rfloor + 1$ if $x_{(k)}$ is integer, or $x_{(k)} \leq x_{(k)} - \varepsilon$ and $x_{(k)} \geq x_{(k)} + \varepsilon$, respectively, if $x_{(k)}$ is continuous. $\varepsilon > 0$ has to be chosen such that the two new problems have disjoint subdomains.

For each of the subproblems P_1 and P_2 update η by following one of the above options (either based on the Lagrangian dual or the wait-and-see solution of the expected value problem). Go to *STEP 3*.

Just like `Algorithm SD`, the above algorithm is finite if X is bounded and if the branching on the continuous components is finite. The evaluation of a solution suggestion $\hat{x}$ in *STEP 4* is realized by fixing the first-stage vector to $\hat{x}$, solving the subproblems

$$\phi_j(\hat{x}) := \min_{y_j \in \mathbb{Z}_+^m \times \mathbb{R}_+^{m'}} \{c\hat{x} + qy_j \ : \ T_j\hat{x} + Wy_j = h_j\}$$

for $j = 1, \dots, S$, and calculating the term $\mathrm{E}X + \alpha\mathcal{D}_1^+ X$ for a random variable X that takes on the S values $\phi_j(\hat{x})$ with probabilities π_j. Thus, we take care of infeasibilities.

We add the constraint adaptation in *STEP 5* to emulate the constraints $v_j \geq cx + \sum_{j=1}^{S} \pi_j q y_j$, $j = 1, \dots, S$. In order to show that the algorithm is correct, it is sufficient to verify that the employed lower bounds are valid, i.e. that no optimal node is cut off. This however follows from the fact that in the initialization step as well as in *STEP 5*, η is chosen such that $P(\eta)$ provides a lower bound for the mean-absolute-semideviation problem (6.12), cf. Lemma 6.2.

6.5 Mean-risk models with FSD-consistent risk measures

For a random variable X and a nonnegative risk measures it holds

$$\mathrm{E}X \leq \mathrm{E}X + \alpha\mathcal{R}(X) \qquad \alpha \geq 0. \tag{6.14}$$

For risk measures that are consistent with first stochastic order, there is another lower bound available. To get this bound we need to assume $\pi_j = \frac{1}{S}$ for $j = 1, \dots, S$. Let the risk measure $\mathcal{R}$ be α-consistent with first order stochastic dominance. We consider the corresponding mean-risk model

$$\min_{x \in X} Q_{\mathrm{E}}(x) + \alpha Q_{\mathcal{R}}(x) \qquad \alpha \in \mathbb{R}_+ \tag{6.15}$$

Assume $\hat{x}$ is an optimal solution of problem (6.15). The distribution of ξ with the S probability atoms (h_j, T_j) defines via $z_j(\hat{x}) := c\hat{x} + \phi(h_j - T_j\hat{x})$ a random variable Y that takes on the S values $z_j(\hat{x})$ with probability $\frac{1}{S}$.

Assume we have lower bounds z_j^{LB} on each $z_j(\hat{x})$ for $j = 1, \dots, S$. Lemma 2.7 tells us that a random variable X taking on the S values z_j^{LB} with probability $\frac{1}{S}$ dominates Y in first stochastic order. If $\mathcal{R}$ is α-consistent with first stochastic order, we obtain

$$\mathrm{E}X + \alpha\mathcal{R}(X) \leq \mathrm{E}Y + \alpha\mathcal{R}(Y).$$

The subproblem solutions

$$z_j^{\mathrm{LB}} := \min_{x_j, y_j} \{cx_j + qy_j \ : \ T_j x_j + Wy_j = h_j,\ x_j \in X,\ y_j \in \mathbb{Z}_+^m \times \mathbb{R}_+^{m'}\}$$

of the wait-and-see problem provide such lower bounds, i.e. it holds

$$z_j^{\mathrm{LB}} \leq c\hat{x} + \phi(h_j - T_j\hat{x}) \qquad j = 1, \dots, S.$$

This constitutes the basis for the following algorithm.

Let $\mathcal{P}$ denote a list of problems. The symbol $Q'_{\mathbb{R}}(x,y)$ refers to the value $\mathcal{R}(X)$ of a random variable X that takes on the S values $cx_j + qy_j$ with probabilities $\pi_j = \frac{1}{S}$, $j = 1, \ldots, S$.

`Algorithm FSD`: FSD-consistent risk measures

STEP 1 Initialization: Set $z^* := \infty$ and let $\mathcal{P}$ contain the expected value problem (6.2), only.

STEP 2 Termination: If $\mathcal{P} = \emptyset$ then x^* with $z^* = Q_{\mathrm{E}}(x^*) + \alpha Q_{\mathcal{R}}(x^*)$ is optimal.

STEP 3 Node selection: Select and delete a problem $P \in \mathcal{P}$ and solve its wait-and-see problem. If a subproblem is infeasible go to *STEP 2*. Otherwise we set $z_{\mathrm{LB}}(P) := \sum_{j=1}^{S} \pi_j(cx_j + qy_j) + Q'_{\mathbb{R}}(x,y)$ using the subproblem solutions $x = (x_1, \ldots, x_S)$ and $y = (y_1, \ldots, y_S)$.

STEP 4 Bounding: If $z_{\mathrm{LB}}(P)$ is greater than z^* go to *STEP 2*. Otherwise proceed as follows; if the first-stage solutions x_j, $j = 1, \ldots, S$, of the subproblems are

- identical, then x_j is a feasible solution. Update the objective function value $z^* := \min\{z^*, Q_{\mathrm{E}}(x_j) + \alpha Q_{\mathcal{R}}(x_j)\}$.
- not identical, then compute a suggestion $\hat{x} := Heu(x_1, \ldots, x_S)$ using some heuristic. Set $z^* := \min\{z^*, Q_{\mathrm{E}}(\hat{x}) + \alpha Q_{\mathcal{R}}(\hat{x})\}$.

Delete all problems $P' \in \mathcal{P}$ with $z_{\mathrm{LB}}(P') \geq z^*$.

STEP 5 Branching: Select a component $x_{(k)}$ of x and add two new problems to $\mathcal{P}$ that differ from P by the additional constraint $x_{(k)} \leq \lfloor x_{(k)} \rfloor$ and $x_{(k)} \geq \lfloor x_{(k)} \rfloor + 1$, respectively if $x_{(k)}$ is integer, or $x_{(k)} \leq x_{(k)} - \varepsilon$ and $x_{(k)} \geq x_{(k)} + \varepsilon$ respectively if $x_{(k)}$ is continuous. $\varepsilon > 0$ has to be chosen such that the two new problems have disjoint subdomains. Go to *STEP 3*.

Although we solve subproblems of the expected value problem (6.2) along the way, the optimal solution and the optimal values produced by the above algorithm belong to problem 6.15. This is guaranteed by the solution evaluation and the bounding in *STEP 4*.

The algorithm is finite if X is bounded and if the branching on the continuous components is finite, cf. `Algorithm SD`.

Given a problem (P) in a node of the branch-and-bound tree, the corresponding wait-and-see problem arises from (P) by dropping the nonantizipativity constraint, cf. *STEP 3*. The subsolutions of these wait-and-see problems do not reflect the risk measure employed. Only the heuristic can produce solutions different from those obtained in the scenarios decomposition (`Algorithm SD`).

The evaluation of a solution suggestion $\hat{x}$ in *STEP 4* is realized by fixing the first-stage vector to $\hat{x}$, solving the subproblems

$$z_j := \min_{y_j \in \mathbb{Z}_+^m \times \mathbb{R}_+^{m'}} \{c\hat{x} + qy_j \ : \ T_j\hat{x} + Wy_j = h_j\}$$

for $j = 1, \ldots, S$, and calculating the term $\mathrm{E}X + \alpha\mathcal{R}X$ for a random variable X that takes on the S values z_j with probabilities π_j. Thus, we take care of infeasibilities.

Replacing the lower bounds z_{LB} by the expected value $\sum_{j=1}^{S} \pi_j(cx_j + qy_j)$ of the subproblems provides an algorithm valid for nonnegative risk measures, cf. inequality (6.14). We will refer to the latter as `Algorithm NFSD` (not consistent with first order stochastic dominance) because it is inferior to `Algorithm FSD` and should be exclusively applied to mean-risk models including non-FSD-consistent risk measures.

7 Applications

In this chapter we present two applications modelled as stochastic programs with mixed-integer recourse. Both problems are motivated by real-life production plants. The first one is a model of a chemical batch plant run by Dow Chemical in Buna, see Dow Chemical (2004), the second one represents an attempt towards the optimization of the gas network operation of the German gas supplier Ruhrgas, see Ruhrgas (2004).

Integer requirements and uncertainties arise quite naturally in both problems. We use integer variables to model the operation mode of single production units and – in the first problem – to model the batch mode production. This leads to integer requirements on the second stage of the recourse problem in both applications and to integer requirements on the first-stage vector in the first application.

Both applications possess several sources of uncertainties, see Engell et al. (2001) and Westphalen (2004). Among them, customer demand is certainly the most relevant. As we focus on the consequences of using different mean-risk models, we restrict our investigation to problems with stochastic demand.

7.1 Scheduling of a multiproduct batch plant under uncertainty

7.1.1 Introduction

In the chemical processing industry, batch processing is a concept referring to the fact that production units are run in a discontinuous mode. Generally, batch processes occur where small amounts of similar, typically high-valued products are manufactured. While in a continuously driven single-product plant the processing units are designed to serve a specific market capacity, the units of a multiproduct batch plant may perform different tasks in different situations. The set of operating equipment items, the tasks they perform and the product recipes in use depend on the information about market requirements available to the operator of the plant, cf. Rauch (1998) and Reklaitis (1996).

A natural way to model batch processes is the assignment of integer variables to the number of produced batches. There is vast literature on deterministic mixed-integer linear programs (MILP) for the scheduling of batch plants, see e.g. the surveys Applequist et al. (1997), Pinto and Grossmann (1998), Shah (1998), and the references therein.

We formulate a deterministic model motivated by a batch plant that produces different types of polymer (EPS). Following Blömer and Günther (1999) this model can be classified by the properties: fixed batch size, fixed processing time, continuous and batch mode production units, nonpreemptive operation mode, discrete time representation, networked material flow, multipurpose and special-purpose production units. Moreover we are confronted with coupled production.

The operator of the EPS plant faces uncertainties about customer orders. According to her/his order acceptance policy, decision variables belong to time intervals with deterministic or stochastic demand. We propose a two-stage stochastic programming model to schedule the plant.

The composition of the first-stage vector of our model depends on the order acceptance policy of the operator. The design of the plant is not subject to optimization. In our approach the first stage comprises early decisions, the second stage late decisions in terms of the scheduling horizon.

Examples for stochastic models of batch plants can be found in the articles of Pistikopoulos et al. (1996) and Subrahmanyam et al. (1994). In both papers 'here-and-now' decisions refer to the design and recourse decisions refer to the operation of the plant.

As a discrete time model, our two-stage stochastic program is based on a fixed grid with uniform time discretization. In this context, we refer to related work (for deterministic batch scheduling) involving nonuniform grids and event-driven time representations in Mockus and Reklaitis (1997), Schilling and Pantelides (1996), and Schulz (2001).

7.1.2 Problem description

Process We consider a multiproduct batch plant that produces expandable polystyrene (EPS) in different chemical qualities and grain sizes. The latter two attributes are referred to as product groups and fractions. The plant comprises three stages - preparation, polymerization, and finishing, see Figure 7.1[1].

In the preparation stage raw material is converted into three different types of intermediates. Due to the small processing times and the moderate storage capacities for all intermediates, the preparation stage does not restrict the further production process and is therefore neglected.

A polymerization is performed in one out of a given number of congeneric, batch-wise driven reactors. It is characterized by a recipe comprising the setting of physical parameters and the composition of the input batch. The recipe

[1]Figure 7.1 was created by G. Sand and originally part of Engell et al. (2001). We thank for the friendly permission to use it here.

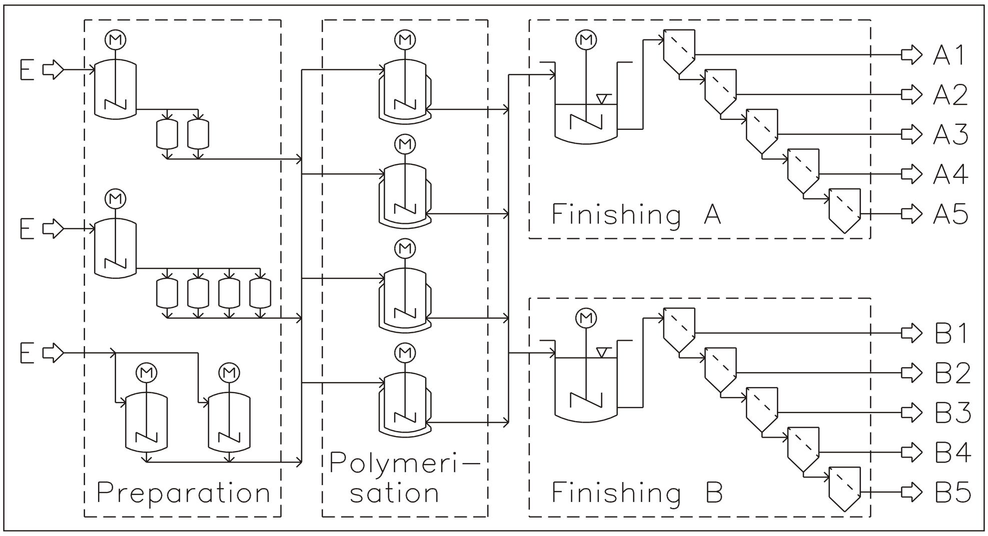

Figure 7.1: The EPS production process

determines the product group and the ratio of each fraction in the output batch. Due to the difficulties generally encountered with polymerization processes, not every grain size distribution can be produced. However, the recipes can be adjusted such that they yield output batches with a high proportion of a single fraction and small proportions of the remaining ones.

The capacity of the polymerization stage in a given time interval depends on the number of available reactors, on the fixed processing time of a polymerization, and on a safety parameter determining the minimal delay between successive polymerizations. It can be calculated offline. The same parameters further restrict the capacity of the polymerization stage in successive time intervals.

For each product group there is a finishing line consisting of a mixer tank and a separation unit. Mixing tanks realize buffering of the material flow on the edge of polymerization reactors (batch mode) and separation units (continuous feed). A detailed model of the mixing characteristics involves nonlinear equations. However, in the model we will use an approximate approach viewing the time the material remains in the mixing tanks as idle processing time.

The content of the mixing tanks may not exceed an upper bound. As soon as the mixer content falls below a lower bound, an expensive shut-down operation of the corresponding finishing line has to be performed. The separation units serve to split the output batch of the polymerizations into the single fractions.

The main production goal is to avoid production deficit, i.e. to comply with customer orders. Further we aim at minimizing the number of polymerizations and the number of changes of the state of the finishing lines.

Uncertainties The operator is interested in reacting flexibly to changing market requirements. Therefore it is necessary to consider uncertain demand in the planning phase. We assume that the following is known about the nature of the demand uncertainty.

At the date of scheduling, the operator of the EPS plant faces uncertain customer orders. The operator accepts new orders or order modifications arriving at least δ days before the due day. Consequently, the demand for the next δ days is deterministic. The remaining demand obeys a probability distribution known to the operator. The probability distribution takes into account changes of the amounts of single orders and shifts of due dates of orders with respect to a 'base' demand. The base demand is specified by experts on the base of past data. In addition, we assume that uncertainty increases with time.

Aggregated and detailed scheduling The MILP in Section 7.1.3 is part of a two-level approach to the scheduling of the EPS production plant. We describe the aggregated scheduling step that provides information about the states of the finishing lines, the number and the specifications of polymerizations to run in a time interval, and the resulting production profile to a detailed scheduling model. The latter model involves a nonlinear mixer model but no uncertainties. For details on the short-term scheduling model and on the interface of the two models we refer to Engell et al. (2001), Löhl et al. (1998), and Schulz (2001).

7.1.3 Single-scenario model

Model description

In this subsection we introduce a MILP for the mid-term scheduling of the EPS production process, cf. Engell et al. (2000), Engell et al. (2001), and Schulz (2001). We refer to the MILP as (*EPS*) and to the list of parameters and variables in Appendix A.1.

Parameters We are given a number of product groups $P \in \mathbb{Z}_+$, a number of fractions $F^{\mathrm{p}} \in \mathbb{Z}_+$ and recipes $R^{\mathrm{p}} \in \mathbb{Z}_+$ for each product group, and a finite number $I \in \mathbb{Z}_+$ of time intervals resulting from an equidistant subdivision of the scheduling horizon. Unless otherwise specified, sums and indices run as follows:

$$i = 1, \ldots, I, \quad p = 1, \ldots, P, \quad f_{\mathrm{p}} = 1, \ldots, F^{\mathrm{p}}, \quad r_{\mathrm{p}} = 1, \ldots, R^{\mathrm{p}}.$$

Constraints Let $N_{i,p,r_{\mathrm{p}}} \in \mathbb{Z}_+$ be a variable indicating the number of polymerizations of product group p and recipe r_{p} in time interval i. The number of polymerization starts in any series of successive time intervals is bounded above by the capacity of the polymerization stage:

$$\sum_{j=i}^{k} \sum_{p,r_\mathrm{p}} N_{j,p,r_\mathrm{p}} \leq N_{i,k}^{max} \qquad k = i, \dots, I,\ i = 1, \dots, I. \tag{7.1}$$

By $x_{i,p} \in \{0,1\}$, $i = 1, \dots, I+1$, we denote a variable that is equal to 1 if the finishing plant of product group p is on-duty in time interval i and equal to 0 otherwise. Frequent start-ups and shut-downs of the finishing lines are avoided by two groups of bounding constraints. For the individual finishing units, the parameter δ_p (ε_p) determines the minimal number of time intervals in idle (operation) state after a shut-down (start-up)

$$x_{i-j,p} - x_{i-j+1,p} + x_{i,p} \leq 1 \qquad 2 \leq j \leq \delta_p,\ i = 2, \dots, I+1, \tag{7.2}$$

$$x_{i-j,p} - x_{i-j+1,p} + x_{i,p} \geq 0 \qquad 2 \leq j \leq \varepsilon_p,\ i = 2, \dots, I+1. \tag{7.3}$$

In case $i - j \leq 0$ the entities $x_{i-j,p}$ correspond to the states of the finishing lines before and at the beginning of the planning horizon. These values are assumed to be known parameters of the problem. Note that we decrease the degrees of freedom of a subsequent optimization interval by fixing the variable $x_{I+1,p}$.

The mass balances of the mixer tanks result in lower and upper bounds for the mixer contents according to the number of polymerization starts and the feed to the separation units. Minimal and maximal feeds are denoted by F_p^{min} and F_p^{max}, the mixer contents by $C_{i,p}$. The initial mixer contents $C_{0,p}$ are model parameters. Therewith, we have

$$C_{i,p} \leq C_{i-1,p} + \sum_{r_\mathrm{p}} N_{i,p,r_\mathrm{p}} - F_p^{min} x_{i,p}, \tag{7.4}$$

$$C_{i,p} \geq C_{i-1,p} + \sum_{r_\mathrm{p}} N_{i,p,r_\mathrm{p}} - F_p^{max} x_{i,p}. \tag{7.5}$$

At the boundaries of idle-state intervals the mixer tanks must be empty. At the remaining boundaries the mixer tanks must contain at least C_p^{min} units of material

$$C_{i,p} \geq C_p^{min} y_{i,p}. \tag{7.6}$$

The variables $y_{i,p} \in \{0,1\}$ correspond to the boundaries. They are introduced via

$$y_{i,p} - x_{i,p} \leq 0, \;\; y_{i,p} - x_{i+1,p} \leq 0, \;\; x_{i,p} + x_{i+1,p} - y_{i,p} \leq 1,$$

cf. Nemhauser and Wolsey (1988). The mixer contents are subject to the capacity limits C_p^{max}

$$C_{i,p} \leq C_p^{max} y_{i,p}. \tag{7.7}$$

The mass balances of polymerizations containing a ratio of $\rho^{\mathrm{p}}_{\mathrm{f_p},\mathrm{r_p}} \in [0,1]$ of fraction f_{p} and orders $B_{j,p,f_{\mathrm{p}}} \in \mathbb{R}_+$ for this fraction are transferred to storage tanks containing $M_{i,p,f_{\mathrm{p}}}$ units of material:

$$M_{i,p,f_{\mathrm{p}}} = M_{i-1,p,f_{\mathrm{p}}} + \sum_{r_{\mathrm{p}}} \rho^{\mathrm{p}}_{\mathrm{f_p},r_{\mathrm{p}}} N_{i,p,r_{\mathrm{p}}} - B_{i,p,f_{\mathrm{p}}}. \tag{7.8}$$

Here $M_{0,p,f_{\mathrm{p}}}$ is the initial content of the storage tank for fraction f_{p} of product group p. Storage tanks have moderate capacities which do not restrict the production process. Thus $M_{i,p,f_{\mathrm{p}}}$ is not bounded from above.

The variables $M_{i,p,f_{\mathrm{p}}}$ are split into a positive $M^+_{i,p,f_{\mathrm{p}}} \geq 0$ (representing surplus production) and a negative $M^-_{i,p,f_{\mathrm{p}}} \geq 0$ (representing production deficit) part

$$M_{i,p,f_{\mathrm{p}}} = M^+_{i,p,f_{\mathrm{p}}} - M^-_{i,p,f_{\mathrm{p}}}. \tag{7.9}$$

Objective function We count the number of start-ups $w^-_{i,p} \in \{0,1\}$ and shutdowns $w^+_{i,p} \in \{0,1\}$ of the finishing lines in order to be able to minimize switching costs

$$x_{i-1,p} - x_{i,p} = w^+_{i,p} - w^-_{i,p} \qquad i = 1,\dots,I+1. \tag{7.10}$$

The prior production goal is the minimization of production deficit. In addition we aim at minimizing polymerization costs and costs caused by the changes of the states of the finishing lines

$$\min_{M^-,N,w^+,w^-} \sum_{i,p,f_{\mathrm{p}}} a_{i,p,f_{\mathrm{p}}} M^-_{i,p,f_{\mathrm{p}}} + \sum_{i,p,r_{\mathrm{p}}} b_{i,p,r_{\mathrm{p}}} N_{i,p,r_{\mathrm{p}}} + \sum_{i,p} (d^+_{i,p} w^+_{i,p} + d^-_{i,p} w^-_{i,p}). \tag{7.11}$$

Since changing the states of the finishing units is already restricted by the constraints (7.2) and (7.3), the latter goal is of minor importance. Therefore, the relative magnitudes of the cost coefficients are chosen as follows

$$a_{i,p,f_{\mathrm{p}}} >> b_{i,p,r_{\mathrm{p}}} >> d^+_{i,p},\ d^-_{i,p}.$$

7.1.4 Multi-scenario model

First-stage variables The choice of first-stage variables is based upon the subdivision of the planning horizon in time intervals with deterministic demand ($i \leq i_{\mathrm{fs}}$) and time intervals with stochastic demand. Therefore, the decision variables $N_{i,p,r_{\mathrm{p}}}$ (polymerization starts) and $x_{i,p}$ (states of the finishing lines) for $i = 1,\dots,i_{\mathrm{fs}}$ are the first-stage variables in our model. The variables $C_{i,p}$ (mixer contents), $M^+_{i,p,f_{\mathrm{p}}}$, $M^-_{i,p,f_{\mathrm{p}}}$ (production surplus and deficit) and $y_{i,p}$ (auxiliary variables) for $i = 1,\dots,i_{\mathrm{fs}}$ formally also belong to the first stage. However, they are uniquely determined by the variables $N_{i,p,r_{\mathrm{p}}}$ and $x_{i,p}$, $i = 1,\dots,i_{\mathrm{fs}}$. Hence it is possible to shift $C_{i,p}$, $M^+_{i,p,f_{\mathrm{p}}}$, $M^-_{i,p,f_{\mathrm{p}}}$ and $y_{i,p}$ for $i = 1,\dots,i_{\mathrm{fs}}$ into the second stage without changing the problem. Nonanticipativity of these variables then follows from the nonanticipativity of $N_{i,p,r_{\mathrm{p}}}$ and $x_{i,p}$, $i = 1,\dots,i_{\mathrm{fs}}$.

Two-stage stochastic model Assume we are given S scenarios with probabilities π_s, $s = 1, \ldots, S$. Then the above problem specifications result in the following two-stage stochastic integer programming model of type (6.3):

$$\min \sum_s \pi_s \Big(\sum_{i,p,f_\mathrm{p}} a_{i,p,f_\mathrm{p}} M^-_{s,i,p,f_\mathrm{p}} + \sum_{i,p,r_\mathrm{p}} b_{i,p,r_\mathrm{p}} N_{s,i,p,r_\mathrm{p}} + \sum_{i,p} (d^+_{i,p} w^+_{s,i,p} + d^-_{i,p} w^-_{s,i,p}) \Big) \tag{7.12}$$

subject to

$$\left.\begin{array}{ll} \sum_{j=i}^{k} \sum_{p,r_\mathrm{p}} N_{s,j,p,r_\mathrm{p}} \le N^{max}_{i,k} & k = i, \ldots, I \\ x_{s,i-j,p} - x_{s,i-j+1,p} + x_{s,i,p} \le 1 & \\ x_{s,i-j,p} - x_{s,i-j+1,p} + x_{s,i,p} \ge 0 & \\ C_{s,i,p} \le C_{s,i-1,p} + \sum_{r_\mathrm{p}} N_{s,i,p,r_\mathrm{p}} - F^{min}_p x_{s,i,p} & \\ C_{s,i,p} \ge C_{s,i-1,p} + \sum_{r_\mathrm{p}} N_{s,i,p,r_\mathrm{p}} - F^{max}_p x_{s,i,p} & \\ C_{s,i,p} \ge C^{min}_p y_{s,i,p} & \\ C_{s,i,p} \le C^{max}_p y_{s,i,p} & \\ M_{s,i,p,f_\mathrm{p}} = M_{s,i-1,p,f_\mathrm{p}} + \sum_{r_\mathrm{p}} \rho_{f,r_\mathrm{p}} N_{s,i,p,r_\mathrm{p}} - B_{s,i,p,f_\mathrm{p}} & \\ M_{s,i,p,f_\mathrm{p}} = M^+_{s,i,p,f_\mathrm{p}} - M^-_{s,i,p,f_\mathrm{p}} & \end{array}\right\} s = 1, \ldots, S$$

$$\left.\begin{array}{ll} N_{1,i,p,r_\mathrm{p}} = N_{s,i,p,r_\mathrm{p}} & i = 1, \ldots, i_\mathrm{fs} \\ x_{1,i,p} = x_{s,i,p} & i = 1, \ldots, i_\mathrm{fs} \end{array}\right\} s = 2, \ldots, S$$

The two latter equations are the nonanticipativity constraints which require the equality of the first-stage variables. The demand for the first i_fs time intervals is deterministic, i.e.

$$B_{1,i,p,f_\mathrm{p}} = \ldots = B_{S,i,p,f_\mathrm{p}} \qquad i = 1, \ldots, i_\mathrm{fs}.$$

As in Section 7.1.3 the nonquantified indices run as follows

$$i = 1, \ldots, I, \quad p = 1, \ldots, P, \quad f_\mathrm{p} = 1, \ldots, F^\mathrm{p}, \quad r_\mathrm{p} = 1, \ldots, R^\mathrm{p}.$$

7.1.5 Numerical results

Computational details All computations that we report in this section were carried out on a SUN V880 with a 880 MHz processor and 4GB of main memory. Our C-implementation `DDSIP` of the algorithms `SD`, `ASD`, `FSD`, and `NFSD` uses `CPLEX` in version 8.0 to solve the MILP subproblems during the branch-and-bound procedure and `ConicBundle` to solve the nonlinear master problem. `CPLEX 8.0` is currently one of the most powerful MILP solvers. It implements the state-of-the-art branch-and-cut procedures, cf. CPLEX (2004).

ConicBundle is C. Helmberg's implementation of a bundle method for the optimization of nonsmooth, nonlinear functions. There is no detailed description of ConicBundle available, yet. We refer to the habilitation of Helmberg (2000), where the used method is explained in the context of semidefinite programming. For more details on our implementation DDSIP we refer to Märkert (2004).

ConicBundle is a method which iteratively approximates the objective function on base of informations on the subgradients and on the objective function values in the iteration points. The calculation of the subgradient Hx involves the solving of scenario-many subproblems, see Chapter 6. Thus, performing several iterations is relatively time consuming. In addition, preliminary computational experiments indicated that the gain of using the solution to the Lagrangian dual is small in terms of the lower bounds. Therefore, we decided to fix the Lagrangian multipliers to 0 in all our computations. This leads to the evaluation of a higher number of nodes in the branch-and-bound tree.

Most of the scenario subproblems were not solved to optimality. To guarantee that the objective value of the relaxation still is a lower bound to the full problem, lower bounds instead of best solutions to the scenario subproblems were employed.

As we are in the comfortable situation that demand scenarios do not affect feasibility, every solution to a subproblem provides a feasible solution. Therefore, simple rounding procedures (e.g. rounding the average to nearest integers) should not be used. In our numerical experiments we use a heuristic that provides the subsolution closest to the average of all subsolutions. The distance of a solution to the average was measured as the sum of the square distance over all components of the first-stage vector (l^2-distance). Of course, there are other possible heuristics such as using the subsolution occurring most frequently, using the subsolution with the best objective value or with the maximal/minimal number of polymerizations. This provides some flexibility for selecting a proper heuristic depending on the problem specifications encountered.

Data sets and scenarios In our numerical experiments we use a set of 10 scenarios for each problem instance. For numerical experiments on the EPS problem with larger numbers of scenarios we refer to Engell et al. (2003). Here, we focus on the stochastic nature of the problem and on the effect of using different measures of risk.

It should be noted that the generation of appropriate scenarios for stochastic programs is a field of extensive research itself. We refer to Dupačová et al. (2000), Gröwe-Kuska et al. (2001) and the references therein.

We have employed a two-step procedure to generate the scenario sets. In the first step, we randomly generate a base scenario taking into account the capacity of the plant. In the second step we derive the remaining scenarios from the base scenario. This step involves a random procedure which determines the quality of the change of each scenario entry (no change, shift of demand, change of amount of demand) and another random procedure which specifies the quantity of the change for each entry of the base scenario (shift to which time interval, amount to add/subtract). It is assumed that uncertainty increases with time, i.e. the probability for changes of the base scenario for early time intervals is smaller than the one for late time intervals. Unless otherwise specified, all the scenarios in each scenario set have the same probability. The scenario sets are available on our website, see Appendix A.3.

First Stage				Second stage			
Int.	Bin.	Cont.	Constraints	Int.	Bin.	Cont.	Constraints
20	4	0	3	50	58	224	311

Table 7.1: Dimension of the two-stage model

We report computational results for a scheduling horizon of 2 weeks. To comply with the detailed scheduling model in Engell et al. (2001) the single time intervals have a length of 2 days, i.e. $I = 7$. In all instances the plant manufactures 10 final products consisting of 2 product groups ($P = 2$) and 5 fractions for each product group ($F^p = 5$ for $p = 1, 2$). There are 5 recipes available for both product groups ($R^p = 5$ for $p = 1, 2$). Table 7.1 displays the dimension of the single-scenario problem together with the separation into first-stage and second-stage variables and constraints.

Our aim was to obtain an acceptable solution within a computing time of 4 hours which is the minimum time between the starts of two successive polymerizations and, thus, the shortest decision period in the considered degree of aggregation, cf. Engell et al. (2000). All the numerical results reported in this section are obtained within 4 hours of computing time. We neglect this information in the individual tables.

Uncertainty and sensitivity The basic idea behind two-stage stochastic programs is heading for an optimal compromise first-stage decision given the uncertainty of data and the resulting variability of the first-stage components of single-scenario optimal solutions. Clearly, variability of data may but does not necessarily have to imply variability of first-stage solutions. In this sense, a random problem with identical first-stage components of optimal solutions throughout the single-scenario problems is not truly random and should be dealt with by a deterministic model.

To show that the latter is not the case with our batch scheduling model, in other words that variability of data indeed leads to variability of optimal first-

Var.	Single-scenario solutions of instance 4										First stage
	1	2	3	4	5	6	7	8	9	10	
$N_{1,1,1}$	3	3	3	3	3	0	0	3	3	2	3
$N_{1,1,2}$	2	2	2	2	2	3	3	2	2	2	2
$N_{1,1,3}$	0	0	0	0	0	0	0	0	0	1	0
$N_{1,1,4}$	0	0	0	0	0	0	0	0	0	0	0
$N_{1,1,5}$	2	2	2	2	2	2	2	2	2	2	2
$N_{1,2,1}$	0	0	0	0	0	2	0	0	0	0	0
$N_{1,2,2}$	0	0	0	0	1	0	2	0	0	0	0
$N_{1,2,3}$	4	4	3	3	3	4	2	4	3	3	3
$N_{1,2,4}$	1	1	2	0	1	0	0	1	0	0	0
$N_{1,2,5}$	0	0	0	2	0	1	3	0	2	2	2
$N_{2,1,1}$	5	6	6	6	4	8	8	4	4	6	5
$N_{2,1,2}$	0	0	0	0	0	0	0	0	0	0	0
$N_{2,1,3}$	0	0	0	0	0	0	0	0	0	0	0
$N_{2,1,4}$	0	0	0	0	0	0	0	0	0	0	0
$N_{2,1,5}$	0	0	0	0	0	0	0	0	0	0	0
$N_{2,2,1}$	1	0	1	0	0	2	0	0	0	0	0
$N_{2,2,2}$	2	5	0	5	2	0	1	0	6	1	6
$N_{2,2,3}$	0	0	0	0	0	1	0	0	0	2	0
$N_{2,2,4}$	0	0	0	0	1	0	1	2	0	0	0
$N_{2,2,5}$	3	0	4	0	4	0	1	5	1	2	0
`Dev`	12	6	14	2	6	14	18	18	2	12	0
`Obj`	322	152	272	52	51	382	202	2034	51	553	51

Table 7.2: Optimal polymerizations for different demand vectors

stage components, we have set up two problem instances with a planning horizon of two weeks and compare optimal first-stage solutions to 10 single-scenario models with best known first-stage solutions to the corresponding stochastic program. The results are displayed in the Tables 7.2 and 7.3.

We neglect the first-stage components associated with the states of the finishing lines because there are no substantial differences between the single-scenario solutions. In general, the optimal states of the finishing lines are relatively easy to guess from the single-scenario solutions or the demand scenarios, see also the preprocessing proposed in Engell et al. (2003).

However, the Tables 7.2 and 7.3 indicate a substantial variability regarding polymerization starts. The rows `Dev` display the l^1-distance of the vectors of the single-scenario solutions and the solution of the recourse program. In the rows `Obj`, we have listed the values of the stochastic programming objective function for the individual subsolutions. Clearly, all entries corresponding to single-scenario solutions are above the entries for the best known stochastic programming solutions. This reflects inferiority of single-scenario solutions to the stochastic programming solution when it comes to make the best compromise. The costs in the rows `Obj` illustrate the drastic consequences of a blind following of single-scenario strategies in a stochastic environment and therewith provide a justification to apply a stochastic rather than a deterministic model.

Var.	Single-scenario solutions of instance 5										First
	1	2	3	4	5	6	7	8	9	10	stage
$N_{1,1,1}$	0	0	0	0	0	0	1	1	1	0	2
$N_{1,1,2}$	0	0	0	0	0	2	0	0	0	0	0
$N_{1,1,3}$	0	0	0	0	0	0	0	0	0	0	0
$N_{1,1,4}$	4	5	4	5	6	4	3	4	4	4	4
$N_{1,1,5}$	1	0	1	0	0	0	1	0	0	1	0
$N_{1,2,1}$	1	1	0	0	0	0	1	1	0	0	0
$N_{1,2,2}$	0	0	0	0	0	0	0	0	0	1	0
$N_{1,2,3}$	0	0	0	0	0	0	0	0	0	0	0
$N_{1,2,4}$	0	0	0	1	0	0	0	0	0	0	0
$N_{1,2,5}$	6	6	7	5	6	6	6	6	7	6	6
$N_{2,1,1}$	3	1	0	4	1	4	1	0	0	3	0
$N_{2,1,2}$	3	2	0	0	1	2	0	4	0	3	4
$N_{2,1,3}$	0	1	0	0	0	0	0	0	0	0	0
$N_{2,1,4}$	0	4	0	2	2	0	6	1	5	0	1
$N_{2,1,5}$	0	0	7	0	2	0	0	0	0	0	0
$N_{2,2,1}$	1	1	2	2	2	2	1	0	1	1	2
$N_{2,2,2}$	0	0	0	0	0	0	0	5	4	2	0
$N_{2,2,3}$	1	0	2	0	0	2	2	0	0	0	2
$N_{2,2,4}$	0	0	0	0	0	0	0	0	0	0	0
$N_{2,2,5}$	3	2	0	2	3	1	1	1	0	2	1
`Dev`	13	15	17	16	15	11	15	11	18	15	0
`Obj`	978	1989	2332	2090	1938	79	2086	9384	589	1158	78

Table 7.3: Optimal polymerizations for different demand vectors

Another indicator for the sensitivity of a mathematical program w.r.t. the stochastic parameters is the `EEV` value (*expected result of using the expected value problem solution*). Here, the expected value problem is the problem (1.4) mentioned in the introduction, where all stochastic parameters are replaced by their expected values. Then, its optimal 'first-stage' solution is evaluated by the optimal value of the recourse program (6.2) with fixed first stage. A comprehensive introduction to the `EEV` concept is given in Birge and Louveaux (1997).

In Table 7.4, we have compiled the `EEV` value, the best known upper bound `RP` of the recourse problem (6.2), and the solution `WS` to the wait-and-see problem (1.3) for 10 problem instances with 10 scenarios each. The difference of the first two values is called the *value of the stochastic solution* (`VSS`), the difference of the last two ones the *expected value of perfect information* (`EVPI`). Clearly, both values are nonnegative. The value in the column `LP` is the objective of the recourse problem when we relax integrality.

For all problem instances the magnitude of the `VSS` is large compared to the magnitude of the best known upper bound. So, simply replacing the stochastic parameters by their expected values is not promising. The `EVPI` is relatively small for most of the instances. In fact, the optimal values of the single-scenario problems are relatively close to the optimal values of single-scenario problems

Instance	EEV	RP	WS	LP	VSS	EVPI
1	1508.8	347.7	325.3	43.9	1161.1	22.4
2	652.9	473.2	64.5	47.7	179.7	408.7
3	605.4	54.7	52.2	51.9	550.7	2.5
4	152.3	51.3	48.3	46.8	101.0	3.0
5	1159.3	78.4	55.2	54.7	1080.9	23.2
6	280.2	48.4	46.4	46.2	231.8	2.0
7	176.4	46.1	43.4	43.0	130.3	2.7
8	1100.3	58.7	56.3	56.1	1041.6	2.4
9	3611.4	47.0	45.1	44.2	3564.4	2.8
10	582.2	60.9	58.2	57.5	521.3	2.7

Table 7.4: VSS and EVPI

with a first stage fixed to an optimal solution of the recourse problem. However, a low EVPI does not provide information on the quality of the single-scenario solutions of the wait-and-see problem. In the Tables 7.2 and 7.3, we have seen that the objective value of the recourse problem is sensitive w.r.t. these single-scenario solutions. Thus, a solution method exclusively based on the single-scenario solutions does not appear attractive.

Altogether, the current section indicates that the EPS problem has a stochastic nature and should be dealt with using a stochastic program.

Target measures Before we come to the numerical results for the mean-risk models, we turn to the target measure $Q_{\mathrm{E}_1^\eta}$. In order to obtain a meaningful target for a particular problem instance, some additional considerations are necessary. Situations in which the efficient frontier contains exactly one point are easy to identify and thus, computationally not particularly interesting. We are more interested in problems with several efficient points. Consequently, we may try to find a target which bears the potential of a high number of efficient points.

Let z_j^{LB} be the optimal values of the single-scenario problems

$$z_j^{\mathrm{LB}} := \min_{x,y_j}\{cx + qy_j \ : \ T_jx + Wy_j = h_j,\, x \in X,\ y_j \in \mathbb{R}_+^m \times \mathbb{Z}_+^{m'}\} \qquad (7.13)$$

for $j \in S_0 := \{1, \dots, S\}$ and let $x_{\mathrm{E}} \in \arg\min\{Q_{\mathrm{E}}(x) : x \in X\}$, where we assume that the latter set is nonempty. Upper bounds on the single-scenario problems are given by

$$\phi_j(x_{\mathrm{E}}) := \min_{y_j}\{cx_{\mathrm{E}} + qy_j \ : \ T_jx_{\mathrm{E}} + Wy_j = h_j,\, y_j \in \mathbb{R}_+^m \times \mathbb{Z}_+^{m'}\}. \qquad (7.14)$$

for $j \in S_0$. Let us assume that the values $\phi_j(x_{\mathrm{E}})$ are finite for all $j \in S_0$. The efficient frontier contains exactly one point if the target η_o is chosen greater than or equal to $\max\{\phi_j(x_{\mathrm{E}}) : j \in S_0\}$ or smaller than $\min\{z_j^{\mathrm{LB}} : j \in S_0\}$. The potential improvements realized by the optimal solution of a mean-risk model is high, if η_o lies in many of the intervals $\big[z_j^{\mathrm{LB}}, \phi_j(x_{\mathrm{E}})\big)_{j\in S_0}$. We maximize the number of such intervals by solving the MILP

$$\max_{\eta_o\in\mathbb{R},\, u\in\{0,1\}^S} \quad \{\tfrac{\eta_o}{M} + \textstyle\sum_{j=1}^{S} u_j \ : \quad \eta_o + M(1-u_j) \geq z_j^{\mathrm{LB}},\ j \in S_0 \qquad (7.15)$$
$$\eta_o - M(1-u_j) \leq \phi_j(x_{\mathrm{E}}) - \varepsilon,\ j \in S_0\}$$

where M is a constant greater then $\max\{\phi_j(x_{\mathrm{E}}) : j \in S_0\} - \min\{z_j^{\mathrm{LB}} : j \in S_0\}$ and ε is a nonnegative constant smaller than $\min\{\phi_j(x_{\mathrm{E}}) - z_{j'}^{\mathrm{LB}} : j, j' \in S_0\}$. Thus, the two constraints force u_j to take the value 0 if η_o is not in the interval $[z_j^{\mathrm{LB}}, \phi_j(x_{\mathrm{E}}) - \varepsilon]$. We may exclude a solution x_{E} and reoptimize, if the parameter ε turns out to be zero and if the optimal η_o to problem (7.15) is equal to $\phi_j(x_{\mathrm{E}})$ for some j in S_0.

In a cost minimization context, the risk measure 'expected excess of a target' expresses the aversion against the excess of a certain high cost level (possibly leading to ruin). We try to take care of this fact by maximizing η_o within the values of high potential. The choice of M guarantees that $\frac{\eta_o}{M}$ is smaller than 1 and therewith that the maximization of η_o lying in some of the interesting intervals is prior.

Clearly, this is a heuristic approach. For experimental purposes, however, we believe that this systematic way is superior to the arbitrary fixing of targets. Therefore, we apply it throughout.

Algorithm evaluation As pointed out in Chapter 5, the investigated mean-risk problems turn into large-scale mixed-integer programs for random variables having a discrete and finite probability distribution. The first interesting question is whether standard mixed-integer solvers are able to handle these problems adequately. In other words, do we need new algorithms for the problems at hand. Therefore, we compare the scenario decomposition algorithm with the state-of-the-art mixed-integer solver `CPLEX 8.0`. This has been done for the purely expected value based problem (6.2), only. The results for 10 problem instances with 10 scenarios each are displayed in Table 7.5.
It is no surprise that a decomposition algorithm which exploits the problem structure is superior to a general purpose method. The smallest gap reached by `CPLEX 8.0` is larger than the largest gap obtained by the scenarios decomposition algorithm. Moreover, there is not a single instance for which `CPLEX 8.0` was able to produce a better lower or upper bound than the decomposition algorithm. This is reason enough to use decomposition algorithms throughout.

Next, we compare the decomposition algorithms for mean-risk models introduced in Chapter 6 from a computational point of view, see Table 7.6. That is,

	Upper bound		Lower bound		Gap in %	
Instance	SD	CPLEX	SD	CPLEX	SD	CPLEX
1	347.7	351.2	332.5	204.0	4.4	41.9
2	473.2	1027.0	471.2	49.9	0.4	95.1
3	54.7	81.3	52.7	52.4	3.7	35.6
4	51.3	54.2	50.3	48.4	1.8	10.7
5	78.4	79.8	77.2	56.0	1.5	29.8
6	48.4	48.9	46.9	46.4	3.0	5.0
7	46.1	46.9	44.5	43.7	3.4	6.7
8	58.7	62.4	56.8	56.4	3.2	9.6
9	47.0	47.5	45.6	44.8	3.0	5.7
10	60.9	914.0	58.9	58.0	3.2	93.7

Table 7.5: Scenarios decomposition (`SD`) versus `CPLEX 8.0` (CPU time: 4 h)

we study the numerical behavior of the algorithms for different instances of the EPS problem and different mean-risk models. We are particularly interested in the additional effort for solving mean-risk problems compared to the expected value problem. Table 7.6 shows the gaps of the individual mean-risk models for 10 problem instances with 10 scenarios each.

The first row of the table contains the abbreviations for the different algorithms introduced in the previous chapter. The second row displays the risk model, where the risk measure $\mathcal{R}$ is used briefly for the mean-risk model $\min\{Q_{\mathrm{E}}(x) + Q_{\mathcal{R}}(x) : x \in X\}$. Note that the weight α is equal to 1 in all cases.

We observe that the mean-risk models perform (at least slightly) worse than the expected value model. However, only the mean-standard-deviation model in the last column seems to have significantly greater gaps than the expected value model. This is not surprising as the `Algorithm NFSD` employs the weakest lower bounds, cf. Section 6.5. Another observation is that the `Algorithm FSD` yields slightly better results (for instance 5 even significantly better) than the `Algorithm ASD` for the mean-absolute-deviation model, see the columns 4 and 5. Remember that the lower bounds employed in `Algorithm FSD` use the fact that the absolute semideviation is consistent with first order stochastic dominance. The lower bounds employed in `Algorithm ASD` stem from an auxiliary parametric problem and are often better. However, we obtain the latter bounds on the expense of solving an additional problem. In our numerical experiments, solving these additional problems did not pay off.

Tracing the efficient frontier In this section, we document the results of the tracing of the efficient frontier for two problem instances and two different mean-

	SD		ASD	FSD	NFSD
Instance	E	E_1^η	$\mathcal{D}_1^+$	$\mathcal{D}_1^+$	$\mathcal{D}_2$
1	4.4	3.9	1.8	2.3	3.9
2	0.4	27.2	39.6	27.5	79.1
3	3.7	5.2	4.2	3.9	6.7
4	1.8	4.1	4.4	2.5	0.7
5	1.5	3.0	21.4	2.6	42.6
6	3.0	4.4	3.5	2.5	8.0
7	3.4	6.2	5.3	4.2	3.6
8	3.2	5.5	4.1	3.5	5.4
9	3.0	3.4	3.8	3.4	7.2
10	3.2	6.4	6.6	3.6	66.3

Table 7.6: Decomposition algorithms, relative gaps in % (CPU time: 4 h)

risk models. It turned out that problem instances possess only a single efficient point when we use the above scenario sets. We decided to modify the scenario sets to obtain more interesting examples. The modification was simple; we leave out the last scenario in each set and add a scenario of zeros instead (no demand at all). The new scenario gets the probability $\frac{8}{17}$ and the remaining ones the probability $\frac{1}{17}$. Moreover, we decrease the cost coefficients for the variables M^- representing production deficit. Then, the solution 0 is optimal for the resulting expected value model. This solution results in a high dispersion among the single-scenario objectives. The first scenario has an objective value of 0, the remaining ones an objective value corresponding to the individual production deficit. Then, the implementation of a mean-risk models corresponds to the trade-off between no production and a compromise production which fulfills the scenarios 2 to 9 optimally.

We remark that we are not able to proof optimality within the given time limit of four hours. Thus, the points which we refer to as efficient are in fact merely ε-efficient. However, during the iterative procedure for the finding of ε-efficient points, we can pass start solutions obtained in a previous iteration to the algorithms. This leads to gaps below 2 % for most instances and below 4.5 % for all instances.

First, we document the numerical results for the mean-absolute-semideviation model. We have used the `Algorithm ASD` to solve it. To determine efficient points, we can not proceed exactly as described in Section 2.5. In chapter 5 we have seen that the mathematical program

$$\min_{x \in X} Q_{\mathrm{E}}(x) + \alpha Q_{\mathcal{D}_1^+}(x) \qquad (7.16)$$

Instance 5				Instance 7			
α	Q_{E}	$Q_{\mathcal{D}_1^+}$	It.	α	Q_{E}	$Q_{\mathcal{D}_1^+}$	It.
0	29.75	14.0	1	0	26.03	12.25	1
0.26	29.75	14.0	4	0.30	26.03	12.25	4
0.58	30.86	9.81	3	0.59	27.27	8.12	3
0.99	34.04	6.61	5	0.96	30.39	4.89	5
1	34.04	6.61	2	1	30.39	4.89	2

Table 7.7: Efficient frontiers for two instances of problem (7.16)

possesses unfavorable properties when α is greater than 1. Therefore we exclude these programs from our numerical experiments. This however, makes it impossible to obtain supported efficient points x with $Q_{\mathcal{D}_1^+}(x) < Q_{\mathcal{D}_1^+}(\hat{x})$ and $\hat{x} \in \arg\min\{Q_{\mathrm{E}}(x) + Q_{\mathcal{D}_1^+}(x) : x \in X\}$ by solving instances of problem (7.16).

We start by solving the problems $\min Q_{\mathrm{E}}(x)$ and $\min\{Q_{\mathrm{E}} + Q_{\mathcal{D}_1^+}(x) : x \in X\}$ for two instances of the EPS model. Then, we use the iterative method sketched in Section 2.5 to get the supported efficient points x with $Q_{\mathrm{E}}(x_1) < Q_{\mathrm{E}}(x) < Q_{\mathrm{E}}(x_2)$ where $x_1 \in \arg\min\{Q_{\mathrm{E}}(x) : x \in X\}$ and $x_2 \in \arg\min\{Q_{\mathrm{E}}(x) + Q_{\mathcal{D}_1^+}(x) : x \in X\}$. The results are displayed in Table 7.7.

For both instances we could trace three efficient points by the iterative method within a CPU time of 4 hours. We explain our proceeding for instance 1 in more detail. After having solved the two initial problems, we obtain the new weight 0.58 by calculating $\alpha(x_1, x_2)$ via the formula

$$\alpha(x, y) := \frac{Q_{\mathrm{E}}(y) - Q_{\mathrm{E}}(x)}{Q_{\mathcal{D}_1^+}(x) - Q_{\mathcal{D}_1^+}(y)} \tag{7.17}$$

with x_1 and x_2 as defined above. The new problem $\min\{Q_{\mathrm{E}} + \alpha_1 Q_{\mathcal{D}_1^+}(x) : x \in X\}$ yields the nondominated solution $(30.86, 9.81)$ and a corresponding efficient point x_3. The results of the remaining two problems verify that there are no other nondominated solutions and that x_1 is efficient. The columns 'It.' indicate the order in which we have solve the problems.

In Table 7.8 we document the supported efficient frontier of two instances of the mean-expected-excess model

$$\min_{x \in X} Q_{\mathrm{E}}(x) + \alpha Q_{\mathrm{E}_1^\eta}(x) \tag{7.18}$$

obtained within a CPU time of 4 hours. The table entries correspond to those of Table 7.7. However, for the mean-expected-excess model we are able to compute $\min\{Q_{\mathrm{E}_1^\eta}(x) : x \in X\}$ and thus, to obtain efficient points in a wider range. The problems $\min\{Q_{\mathrm{E}_1^\eta}(x) : x \in X\}$ appear in the last rows.

Instance 5

α	Q_{E}	$Q_{\mathrm{E}_1^\eta}$	It.
0.00	29.75	4.34	1
0.38	29.75	4.34	4
1.60	30.85	1.46	3
4.42	34.09	0.73	6
5.95	34.09	0.73	5
20.34	35.66	0.65	7
∞	35.66	0.65	2

Instance 7

α	Q_{E}	$Q_{\mathrm{E}_1^\eta}$	It.
0.00	26.03	3.79	1
0.44	26.03	3.79	4
1.59	27.27	0.94	3
3.78	30.39	0.12	8
4.90	30.39	0.12	6
75.67	30.39	0.12	9
111.96	31.37	0.1064	5
1926.54	31.87	0.1061	7
∞	31.87	0.1061	2

Table 7.8: Efficient frontiers for two instances of problem (7.18)

7.2 Gas transport optimization under uncertainty

7.2.1 Introduction

In the course of the liberalization of the European gas market, competition and consequently, the cost pressure on gas suppliers have increased. Thus, today more than ever, gas companies are interested in reducing their production costs. We consider a supplier that distributes gas to small local energy companies such as municipal suppliers. In fact, our research is motivated by the network of the German gas company Ruhrgas which does not produce gas itself but obtains gas from several sources (North Sea, Russia, etc.). At both inflow and outflow nodes contracts specify the bandwidth of acceptable gas parameters. The goal is to transport gas from inflow to outflow nodes at minimal costs. Gas transportation is realized by fuel consuming compressors which are able to increase the gas pressure and thereby to increase the gas flow through a pipeline. Therefore, the goal of minimizing transport costs can be translated into the goal of fuel cost minimization in the compressors. The model incorporates a number of technical and physical constraints, some of them linear by nature and some of them linearized, see Section 7.2.2.

In order to guarantee a certain flexibility to customers, the gas company offers short time changes to the specifications of the gas parameters at outflow nodes. At the time of planning these parameters have to be considered as uncertain. Nevertheless, the gas company is obliged to meet customer demand at all times, with penalty costs in case of failure. We assume that it is possible to approximate the probability distribution of the random demand parameters. The stochastic model aims at minimizing a weighted sum of expected costs and the probability that a certain fixed cost level is exceeded.

7.2.2 The test network

We consider a small pipeline system consisting of a source, delivery nodes, compressors, pipelines, connections, valves, and controllers (control valves), see Figure 7.2. The principal composition of our test network is motivated by the

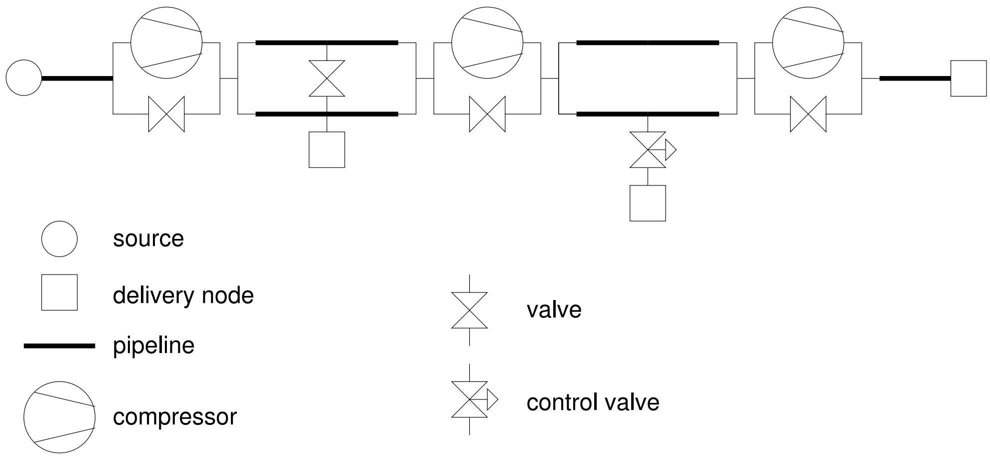

Figure 7.2: Pipeline system

real-world gas network of Ruhrgas, see Ruhrgas (2004). We incorporate the key components of their network but neglect storages and mixing stations. The goal is to transport gas at minimal total cost from the source to the delivery nodes over a time horizon of T hours. The time horizon is divided into time intervals of 1 hour. We use t, $t \in \{1, \dots, T\} =: T_0$, to indicate the time interval.

In addition to the time discretization we employ a spatial discretization. The pipeline system is modeled by a directed graph in which each pipeline component corresponds to an edge, cf. Figure 7.3.

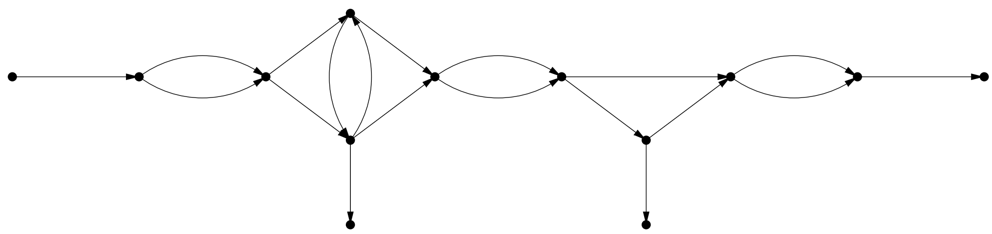

Figure 7.3: Pipeline system as a directed graph

7.2.3 Model equations

Pipelines The basic variables of our model are the variables for pressure and standard gas flow. To state the physical constraints, in particular the continuity and the momentum equations, it is necessary to introduce a number of physical and technical quantities and constants, see Table 7.9. The continuity equation

Constant	Description	Unit/Magnitude
L	pipeline length	$[m]$
D	pipeline diameter	$[m]$
A	pipeline slice plane	$[m^2]$
θ^o	standard temperature	$[273.15K]$
p^o	standard pressure	$[1.01325bar]$
z^o	standard z-factor	$[1.005]$
ρ^o	standard density	$[0.785\frac{kg}{nm^3}]$
h	geodetical altitude	$[m]$
g	gravity	$[9.80665\frac{m}{s^2}]$

Variable	Description	Unit
x	spatial coordinate	$[m]$
t	time coordinate	$[s]$
$\theta(x,t)$	gas temperature	$[K]$
$Q(x,t)$	gas flow density	$[\frac{kg}{m^2 s}]$
$q(x,t) := \frac{3.6\ A\ Q(x,t)}{p^o}$	standard gas flow	$[10^3\frac{nm^3}{h}]$
$p(x,t)$	pressure	$[bar]$
$\tilde{p}(x,t)$	pressure	$[Pa]$
$v(x,t)$	flow velocity	$[m/s]$
$z(p,\theta)$	z-factor	-
$\rho(x,t)$	density	$[kg/m^3]$
$\lambda(q)$	friction factor	-

Table 7.9: Physical quantities and constants

expresses the conservation of mass and describes the compressibility of a gaseous medium in a pipeline in terms of flow density and density

$$\frac{\partial Q}{\partial x} + \frac{\partial \rho}{\partial t} = 0. \tag{7.19}$$

By the thermodynamic state equation

$$\rho(x,t) = \frac{\rho^o z^o \theta^o}{p^o} \frac{p(x,t)}{z(p,\theta)\theta(x,t)}, \tag{7.20}$$

the latter quantity is related to the gas pressure. By elementary algebraic manipulation we obtain the continuity equation in terms of standard gas flow and pressure.

$$\frac{\partial q}{\partial x} + A\frac{z^o\theta^o}{p^o}\frac{\partial\left(\frac{p}{z\theta}\right)}{\partial t} = 0 \tag{7.21}$$

When we assume a constant temperature θ_m and a time independent z-factor which is approximated based on the technical pressure restrictions for every individual pipeline $\bar{z}_e := z(\frac{p_{o_e}^{min}+p_{o_e}^{max}+p_{d_e}^{min}+p_{d_e}^{max}}{4}, \theta_m)$, the above spatial and time discretization yields

$$q_{et}^{out} - q_{et}^{in} + A \cdot L \frac{z^o \theta^o}{\bar{z}_e \; p^o \theta_m} (\bar{p}_{et} - \bar{p}_{et-1}) = 0, \tag{7.22}$$

where we assign a pressure variable $p_{nt} \in [p_n^{min}, p_n^{max}]$ to each node n of the graph in Figure 7.3 and an inflow and an outflow variable $q_{et}^{in} \in [q_e^{min}, q_e^{max}]$ and $q_{et}^{out} \in [q_e^{min}, q_e^{max}]$, respectively, to each edge e for all time intervals. We introduce the symbols o_e and d_e to refer to the entry and exit nodes of edge e, respectively. The average pressure $\bar{p}_{et}$ in pipeline e in time interval t is calculated as the arithmetic mean of the pressures at the two end nodes of edge e.

Let us turn to the momentum equation

$$\frac{\partial \tilde{p}}{\partial x} + \frac{\lambda |Q| Q}{2 D \rho} + \underbrace{\frac{\partial Q}{\partial t} + \frac{\partial (vQ)}{\partial x} + \frac{\partial (g \rho h)}{\partial x}}_{*} = 0$$

which provides a description of the forces effecting the gas molecules in a pipeline. We neglect short term forces and forces caused by the impact pressure as well as gravity forces caused by the slope of the pipeline (the three marked terms) and assume that friction is the only relevant force, cf. Reith and Sekirnjak (2001). In a model without back flow ($Q(x,t) \geq 0 \; \forall t, x$) the above discretization scheme yields the quadratic equation

$$p_{o_e t}^2 - p_{d_e t}^2 = \phi_e (q_{et}^{in} + q_{et}^{out})^2, \tag{7.23}$$

where the constants are given by

$$\phi_e = 3.1272 \cdot 10^{-7} \frac{L_e p^o \rho^o \lambda_e \bar{z}_e \theta_e}{D_e^5 z^o \theta^o}.$$

Note that the standard gas flow q in a pipeline is approximated by the average of standard gas inflow q_{et}^{in} and standard gas outflow q_{et}^{out}. The friction factor is computed as $\lambda_e := \lambda(\frac{q_e^{min}+q_e^{max}}{2})$ where q_e^{min} and q_e^{max} are the technical flow restrictions of pipeline e.

In the three indeterminates $p_{o_e t}$, $p_{d_e t}$, and $q_{et}^{in} + q_{et}^{out}$ equation (7.23) describes the surface of a twin cone with vertex at zero and centered around the $p_{o_e t}$-axis. All variables being nonnegative, we confine ourselves to the intersection of the surface with the nonnegative orthant. The idea is to approximate this surface by the surface of a pyramid. Precision of the approximation is controlled by the number of facets in the pyramidal representation. In this way, the equations (7.23) are approximated by the inequalities

$$\beta_{1ej} p_{o_e t} - \beta_{2ej} p_{d_e t} + \beta_{3ej} (q_{et}^{in} + q_{et}^{out}) \leq 0 \qquad j = 1, \ldots, J, \tag{7.24}$$

where J denotes the number of pyramidal facets and β_{iej} are appropriate coefficients. Apart from exceptional situations, optimization drives the iterations to the boundary of at least one of the half spaces in (7.24). For problem instances where computations indicate invalidity of this pragmatic argument, a refined model can be used where additional Boolean variables ensure that at least one of the inequalities in (7.24) holds as an equation.

Compressors The compressors increase the pressure of the gas and therewith serve to transport the gas from source to delivery nodes. Since compressors consume gas they cause costs. The consumption of gas b_{et} is the difference of inflow and outflow at the end nodes of the compressor

$$b_{et} = q_{et}^{in} - q_{et}^{out}. \tag{7.25}$$

It also depends on the realized increase of pressure and is given by

$$b_{et} = \frac{\kappa}{\kappa - 1} \frac{p^o B_e \theta_m z(p_{o_e t}, \theta_m) q_{et}^{out}}{36 \; \theta^o z^o \eta_e H_u} \left(\left(\frac{p_{d_e t}}{p_{o_e t}} \right)^{\frac{\kappa}{\kappa-1}} - 1 \right), \tag{7.26}$$

where H_u specifies the gas quality and B_e as well as η_e specify the efficiency of compressor e. The fuel consumption is a linear function of standard gas outflow and a nonlinear function of pressure. We use a simple linearization of equation (7.26) in our model

$$b_{et} \geq \gamma_{1e} p_{d_e t} + \gamma_{2e} p_{o_e t} + \gamma_{3e} q_{et}^{out}. \tag{7.27}$$

In operation mode, the fuel consumption of a compressors has to lie in the interval $[b_e^{min}, b_e^{max}]$ where b_e^{min} is a strictly positive amount of fuel. Therefore, compressors are modeled as switchable edges. We introduce a variable $u_{et} \in \{0, 1\}$ which represents the mode of compressor e in time interval t with a value of 1 indicating operation mode. Then, the standard gas flow through a compressors is bounded in the following way

$$u_{et} q_e^{min} \leq q_{et}^{in}, q_{et}^{out} \leq u_{et} q_e^{max}. \tag{7.28}$$

The previous elaborations also lead to bounds for the fuel consumption

$$u_{et} b_e^{min} \leq b_{et} \leq u_{et} b_e^{max}. \tag{7.29}$$

When a compressor is switched off, it blocks the gas. In this situation the gas has to use a bypass valve or the network is disconnected. For a compressor e and its bypass valve e_b it has to hold

$$u_{et} + u_{e_b t} \leq 1. \tag{7.30}$$

In order to avoid material fatigue and expensive candle times, a smooth operation of compressors is desired. Two additional sets of constraints take care about this fact

$$u_{et} - u_{et-1} + u_{et-r} \leq 1 \qquad r = 2, \ldots, \delta, \ \forall t \in T_0, \tag{7.31}$$
$$u_{et} - u_{et-1} + u_{et-r} \geq 0 \qquad r = 2, \ldots, \varepsilon, \ \forall t \in T_0, \tag{7.32}$$

where δ and ε are parameters determining the minimal number of time intervals in idle mode after a shut-down and in operation mode after a start-up, respectively.

Valves Valves are short devices with equal in- and outflow

$$q_{et}^{in} = q_{et}^{out}. \tag{7.33}$$

All valves are switchable edges. Thus the flow is bounded in the same way as the flow through compressors, cf. equation (7.28).

There is no pressure drop in valves. However, when a valve e is closed in time interval t ($u_{et} = 0$) the pressure at the entry and exit nodes of the valve o_{e} and d_{e}, respectively, can take arbitrary values (within the technical restrictions)

$$p_{o_{\text{e}}t} - p_{d_{\text{e}}t} \leq K(1 - u_{et}), \tag{7.34}$$
$$p_{d_{\text{e}}t} - p_{o_{\text{e}}t} \leq K(1 - u_{et}), \tag{7.35}$$

where $K = \max\{p_{o_{\text{e}}}^{max}, p_{d_{\text{e}}}^{max}\} - \min\{p_{o_{\text{e}}}^{min}, p_{d_{\text{e}}}^{min}\}$.

A control valve e is able to decrease the pressure in a certain bandwidth $[p_e^{cmin}, p_e^{cmax}]$. Thus for control valves the pressure constraints (7.34) and (7.35) take the form

$$p_{o_{\text{e}}t} - p_{d_{\text{e}}t} - p_e^{cmin} \leq (K - p_e^{cmin})(1 - u_{et}), \tag{7.36}$$
$$p_{d_{\text{e}}t} - p_{o_{\text{e}}t} + p_e^{cmax} \leq (K - p_e^{cmax})(1 - u_{et}). \tag{7.37}$$

Connections A connection e is a short pipeline without pressure drop between entry and exit node and with flow conservation

$$p_{o_{\text{e}}t} = p_{d_{\text{e}}t}, \tag{7.38}$$
$$q_{et}^{in} = q_{et}^{out}. \tag{7.39}$$

Nodes We require that inflow and outflow of each inner node n (no source or delivery node) are identical

$$\sum_{n=o_{\text{e}}} q_{et}^{in} = \sum_{n=d_{\text{e}}} q_{et}^{out}. \tag{7.40}$$

Final state Instead of fixing the values of pressure in each node and the values of standard gas flow on each edge at time interval T, we require a constant gas volume to obtain an appropriate final state of the gas network

$$\sum_{e \in E} v_{0e} = \sum_{e \in E} v_{Te}, \tag{7.41}$$

where E is the set of edges of the graph in Figure 7.3. The gas volume in a segment e mainly depends on the geometric volume and the pressure in the considered time interval t, see Reith and Sekirnjak (2001),

$$v_{te} = \zeta_e(p_{o_e t} + p_{d_e t}). \tag{7.42}$$

Uncertainty As mentioned above, customers specify an interval of acceptable flow at each delivery node n for each time interval $[f^{min}_{nt\omega}, f^{max}_{nt\omega}]$. The additional index ω indicates that we treat these parameters as random variables. We introduce variables $q^{out-}_{et} \geq 0$ and $q^{out+}_{et} \geq 0$ which are positive when the actual flow q^{out}_{et} lies outside the interval $[f^{min}_{nt\omega}, f^{max}_{nt\omega}]$

$$f^{min}_{nt\omega} \leq q^{out}_{et} - q^{out-}_{et} + q^{out+}_{et} \qquad \forall e \in D, \tag{7.43}$$

$$f^{max}_{nt\omega} \geq q^{out}_{et} - q^{out-}_{et} + q^{out+}_{et} \qquad \forall e \in D. \tag{7.44}$$

where D is the set of edges e with delivery exit nodes ($d_e = n$, n delivery node).

Objective function Production goal is to meet customer requirements on minimal costs. Costs are caused by fuel consumption in compressors and by changes of the operation mode of the compressors. For the latter we introduce variables $s^+_{et} \in \mathbb{R}_+$ and $s^-_{et} \in \mathbb{R}_+$ indicating switching operations of the compressors

$$s^+_{et} - s^-_{et} = u_{et} - u_{et-1} \tag{7.45}$$

In addition, contracts specify costs for failing to meet customer requirements such that our objective function reads

$$\min \sum_{e \in D,\, t \in T_0} (c^-_{et}\; q^{out-}_{et} + c^+_{et}\; q^{out+}_{et}) + \sum_{e \in C,\, t \in T_0} c^b_{et}\; b_{et}. \tag{7.46}$$

C is the set of edges corresponding to compressors. The cost coefficients are chosen such that meeting customer requirements is the primal goal, i.e. $c^-_{et},\ c^+_{et} >> c^b_{lt}$ for all $e \in D,\ l \in C,\ t \in T_o$.

7.2.4 Numerical results

Computational details All computations were carried out on a SUN V880 with a 880 MHz processor and 4GB of main memory. We use the same implementation as for the EPS problem, cf. Section 7.1.5 and Märkert (2004).

Preliminary numerical experiments indicated that the computations can be advanced using the lower bound provided by the Lagrangian dual. However, the gain of using the Lagrangian dual from one node of the branch-and-bound tree to the next was relatively small. This corresponds to the observations made in Carøe and Schultz (1998). We decided to use the Lagrangian multipliers obtained in the root node of the branch-and-bound tree throughout the branch-and-bound algorithm.

The proceeding in the numerical examples is similar to the one in Section 7.1.5. We will not repeat every detail but often refer to the corresponding details above. Again, the model is set up such that demand scenarios do not affect feasibility and every solution to a subproblem provides a feasible solution. Unless otherwise specified we used the heuristic that provides the subsolution closest to the average of all subsolutions, cf. Section 7.1.5.

For the determination of the targets for the mean-expected-excess model we once again refer to Section 7.1.5.

Data sets and scenarios Both the test gas network and therefore the number of variables to represent it are relatively small. However, it takes a large number of constraints to represent all physical and technical details and to reach a satisfactory accuracy of the linearizations. In table 7.10 we have compiled the dimension of the problem for a time horizon of 12 hours.

First Stage				Second stage			
Int.	Bin.	Cont.	Constraints	Int.	Bin.	Cont.	Constraints
0	0	96	0	0	96	902	2149

Table 7.10: Dimension of the two-stage model

As explained above, the first-stage contains the variables describing the state of valves and compressors in the first $\bar{t}$ intervals. Precisely, these are the pressure variables $p_{o_e t}$ and $p_{d_e t}$ at entry and exit nodes of valves and compressors for $t = 1, \ldots, \bar{t}$. It is not necessary to include the binary variables u_{et} indicating the operation mode of the compressors into the first stage. The latter variables are uniquely determined by the pressure variables $p_{o_e t}$ and $p_{d_e t}$ at entry and exit nodes of the compressors. Thus, nonanticipativity of the binary variables u_{et} follows from nonanticipativity of the pressure variables.

For our numerical experiment we have chosen $\bar{t} = 4$. In practice, it is desired to have a moderate time interval between readjustments of the pipeline system. An hourly readjustment is an acceptable operation mode. Thus, all computational results documented were obtained within a time limit of 1 hour. We hope to stabilize the solution for the first time interval by incorporating additional 3

Instance	EEV	RP	WS	VSS	EVPI
1	100.25	99.46	95.02	0.79	4.44
2	99.26	98.29	94.46	0.97	3.83
3	100.44	99.39	95.29	1.05	4.10
4	98.70	97.77	93.79	0.93	3.98
5	99.65	98.52	94.51	1.13	4.01
6	99.96	98.92	95.00	1.04	3.92
7	89.94	88.90	84.95	1.04	3.95
8	104.44	104.02	103.32	0.42	0.70
9	821.87	551.23	548.81	270.64	2.42
10	∞	174.74	172.74	∞	2.00

Table 7.11: VSS and EVPI

time intervals into the first stage. For all problem instances this approach leads to a first-stage vector that contains 96 elements.

Numerical experiments were carried out with 10 sets of 10 scenarios each. We have used two different random procedures to generate the scenario sets. The first one assumes a uniform demand distribution within the technical constraints and samples out of this distribution (scenario sets 6 to 10). For the second one we assume that we have relatively high load in half of the scenarios and relatively low load in the remaining ones. This is realized by assuming the demand is uniformly distributed in the upper (lower) third of the interval specified by the technical constraints. Then, we sample from these distributions (scenario sets 1 to 5).

Uncertainty and sensitivity The `EEV` problem and the wait-and-see problem give an indication of the influence of the stochastic right-hand sides on the optimal solution and the optimal values. In Table 7.11 we see that the difference between the best value of the recourse problem and the optimal value of the `EEV` problem is relatively small. Exceptions are the instances 9 and 10 where the optimal solution to the recourse problem is not able to meet all demand. The high penalties for not meeting the demand result in the large differences. For instance 10, the solution produced by replacing the right-hand-sides with the expected value over all scenarios is infeasible for the recourse problem. The optimal value of the wait-and-see problem is relatively close to the best value found by the recourse problem. The columns `VSS` and `EVPI` indicate that the influence of the demand on the optimal value is relatively small for the test network. Indeed, we also observe that the deviation among the optimal solutions is moderate.

	SD		ASD	FSD	NFSD	CPLEX
Instance	E	E_1^η	$\mathcal{D}_1^+$	$\mathcal{D}_1^+$	$\mathcal{D}_2$	E
1	2.4	3.0	2.0	1.2	22.85	0.92
2	2.3	2.9	2.0	1.3	21.70	1.36
3	2.2	2.6	1.9	1.2	22.65	2.17
4	2.3	3.2	2.2	1.4	22.14	0.65
5	2.3	2.3	2.0	1.3	22.33	0.49
6	2.1	2.7	1.8	1.2	23.12	0.65
7	2.7	3.4	3.1	2.8	30.70	3.84
8	0.7	0.6	0.6	0.6	53.18	5.75
9	0.4	0.4	2.8	0.9	67.04	0.38
10	0.9	1.0	77.9	0.8	44.26	0.96

Table 7.12: Decomposition algorithms and `CPLEX 8.0`, Relative gaps reached within a CPU time of 4 hours (in %)

Algorithm evaluation In this section we compare the different algorithms introduced in Chapter 6 to each other and to the standard solver `CPLEX 8.0` from a computational point of view. This is to study the use of the decomposition algorithms and the additional effort for solving mean-risk problems. Table 7.12 displays the gaps of the individual mean-risk models for 10 problem instances with 10 scenarios each.

The first row of the table contains the abbreviations for the different algorithms introduced in the previous chapter. The second row displays the risk model, where the risk measure $\mathcal{R}$ is used as an abbreviation for the mean-risk model $\min\{Q_{\mathrm{E}}(x) + Q_{\mathcal{R}}(x) : x \in X\}$. Note that the weight α is equal to 1 in all cases. The last column reports the gaps obtained by `CPLEX 8.0` for the expected value model.

There is no significant difference among the decomposition algorithms and `CPLEX 8.0`. An exception is the `Algorithm NFSD` used for the mean-standard-deviation model. Here the weak lower bound employed in the algorithm shows its drawbacks, cf. Section 6.5. `CPLEX 8.0` possibly benefits from the relative weak sensitivity of the problem w.r.t. the stochastic parameters.

Tracing the efficient frontier In this section, we document the results of the tracing of the efficient frontier for two problem instances and two different mean-risk models. Again, we remark that we are not able to proof optimality within the given time limit of one hour. Thus, the points which we refer to as efficient are in fact merely ε-efficient. The gaps are below 3.5 % for all instances.

First, we document the numerical results for the mean-absolute-semideviation

Instance 2

α	Q_{E}	$Q_{\mathcal{D}_1^+}$
0	98.29	11.40
0.10	98.29	11.40
0.99	98.39	11.16
1	98.68	11.00
1.89	-	-

Instance 5

α	Q_{E}	$Q_{\mathcal{D}_1^+}$
0	98.52	11.63
0.26	98.58	11.43
0.65	98.58	11.43
1	98.78	11.24
1.05	-	-

Table 7.13: Efficient frontiers for two instances of problem (7.47)

model

$$\min_{x \in X} Q_{\mathrm{E}}(x) + \alpha Q_{\mathcal{D}_1^+}(x) \qquad \alpha \in [0,1]. \tag{7.47}$$

We use the `Algorithm ASD` to solve it. To determine efficient points, we proceed as described in Section 7.1.5, i.e. we exclude problems with weights outside the interval $[0,1]$.

We start by solving the problems $\min Q_{\mathrm{E}}(x)$ and $\min\{Q_{\mathrm{E}} + Q_{\mathcal{D}_1^+}(x) : x \in X\}$ for two instances of the GAS model. Then, we use the iterative method sketched in Section 2.5 to get the supported efficient points x with $Q_{\mathrm{E}}(x_1) < Q_{\mathrm{E}}(x) < Q_{\mathrm{E}}(x_2)$ where $x_1 \in \arg\min\{Q_{\mathrm{E}}(x) : x \in X\}$ and $x_2 \in \arg\min\{Q_{\mathrm{E}}(x) + Q_{\mathcal{D}_1^+}(x) : x \in X\}$. The results are displayed in Table 7.13.

For both instances we could trace 3 efficient points by the iterative method. We explain our proceeding for instance 2 in more detail. After having solved the two initial problems, we obtain the next weight $\alpha(x_1, x_2)_1 = 0.99$ by the formula

$$\alpha(x,y) := \frac{Q_{\mathrm{E}}(y) - Q_{\mathrm{E}}(x)}{Q_{\mathcal{D}_1^+}(x) - Q_{\mathcal{D}_1^+}(y)} \qquad x, y \in X$$

with x_1 and x_2 as above. The new problem $\min\{Q_{\mathrm{E}} + \alpha_1 Q_{\mathcal{D}_1^+}(x) : x \in X\}$ yields the nondominated solution $(98.39, 11.16)$ and a corresponding efficient point x_3. The two new resulting problems have the weights $\alpha(x_3, x_1) = 0.10$ and $\alpha(x_2, x_3) = 1.89$. We skip the latter problem for the reasons given above. The mean-risk problem $\min\{Q_{\mathrm{E}} + 0.1\, Q_{\mathcal{D}_1^+}(x) : x \in X\}$ gives the nondominated solution $(98.29, 11.4)$ which has been obtained before and which verifies that the optimal solution of the expected value problem ($\alpha = 0$) is efficient.

Table 7.14 documents the supported efficient frontier of two instances of the mean-expected-excess model

$$\min_{x \in X} Q_{\mathrm{E}}(x) + \alpha Q_{\mathrm{E}_1^\eta}(x) \qquad \alpha > 0. \tag{7.48}$$

The table entries correspond to those of the Table 7.13. For the mean-expected-excess model we compute

$$\min\{Q_{\mathrm{E}_1^\eta}(x) : x \in X\} \tag{7.49}$$

Instance 1				Instance 11			
α	Q_{E}	$Q_{\mathrm{E}_1^\eta}$	It.	α	Q_{E}	$Q_{\mathrm{E}_1^\eta}$	It.
0	99.24	11.72	1	0	88.90	20.86	1
.16	99.24	11.72	4	0.04	88.90	20.86	4
2.25	99.29	11.45	3	0.21	88.92	20.32	3
5.93	99.29	11.45	6	0.31	88.96	20.20	6
8.51	99.59	11.40	5	0.65	88.96	20.20	5
11.78	99.59	11.40	7	1.03	88.96	20.20	7
1000	100.01	11.36	8	1000	89.06	20.10	8
∞	100.01	11.36	2	∞	89.06	20.10	2

Table 7.14: Efficient frontiers for two instances of problem (7.48)

and thus, obtain efficient points in a wider range of weights. The problems with weight $\alpha = 1000$ are solved to verify that the solutions of the problems (7.49) are nondominated. The results for the pure risk problems (7.49) are displayed in the rows $\alpha = \infty$.

A Appendix

A.1 Parameters and variables of the EPS model

Parameters

$I \in \mathbb{Z}_+$	Number of time intervals
$P \in \mathbb{Z}_+$	Number of product groups
$R_p \in \mathbb{Z}_+$	Number of recipes for product group p
$F_p \in \mathbb{Z}_+$	Number of fractions (grain sizes) of product group p
$\rho^{\mathrm{p}}_{r_{\mathrm{p}},f_{\mathrm{p}}} \in \mathbb{R}_+$	Ratio of fraction f_{p} of product group p in an output batch produced with recipe r_{p}
$N^{max}_{i,k} \in \mathbb{Z}_+$	Maximal number of polymerization starts from time interval i to k ($i \leq k$)
$x_{j,p} \in \{0,1\}$	State of finishing lines in the time intervals before the scheduling horizon ($j = 0, \ldots, -\max(\varepsilon_p, \delta_p)$)
$C_{0,p} \in \mathbb{R}_+$	Initial mixer content
$C^{max}_p \in \mathbb{R}_+$	Capacity of mixing tanks
$C^{min}_p \in \mathbb{R}_+$	Minimal mixer content in on-duty time intervals
F^{min}_p, $F^{max}_p \in \mathbb{R}_+$	Minimal and maximal feed into separation units
$\varepsilon_p, \delta_p \in \mathbb{Z}_+$	Minimal number of off- and on-duty intervals after a state change of the finishing stage
$B_{i,p,f_{\mathrm{p}}} \in \mathbb{R}_+$	Costumer orders
$M_{0,p,f_{\mathrm{p}}} \in \mathbb{R}$	Initial storage content

Decision variables

$N_{i,p,r_{\mathrm{p}}} \in \{0, N^{max}_{1,1}\}$	Number of polymerization starts
$x_{i,p} \in \{0,1\}$	State of finishing lines (1: on-duty, operating / 0: off-duty, idle), $i = 1, \ldots, I+1$

Other variables

$y_{i,p} \in \{0,1\}$	Logical variable: $x_{i,p} \wedge x_{i+1,p}$
$C_{i,p} \in \mathbb{R}_+$	Mixer content
$M_{i,p,f_{\mathrm{p}}} \in \mathbb{R}$	Storage content
$M^+_{i,p,f_{\mathrm{p}}}, M^-_{i,p,f_{\mathrm{p}}} \in \mathbb{R}_+$	Production surplus/deficit
$w^+_{i,p}, w^-_{i,p} \in \{0,1\}$	Number of start-ups and shut-downs of the finishing

A.2 Parameters and variables of the GAS model

Physical parameters

θ^o	Standard temperature
p^o	Standard pressure
z^o	Standard z-factor
ρ^o	Standard density
h	Geodetical altitude
g	Gravity

Technical parameters

L	Pipeline length
D	Pipeline diameter
A	Pipeline slice plane
H_u	Gas quality constant
η_e	Efficiency of compressor
p_e^{cmin}	Minimal pressure decrease of compressor
p_e^{cmax}	Maximal pressure decrease of compressor

Model parameters

$t \in [1, T]$	Time interval
$e \in E$	Edge
$n \in N$	Node
$o_e \in N$	Entry node of edge e
$d_e \in N$	Exit node of edge e
$f_{nt\omega}^{min} \in \mathbb{R}_+$	Minimal gas flow at delivery node
$f_{nt\omega}^{max} \in \mathbb{R}_+$	Maximal gas flow at delivery node
D	Set of edges e with delivery exit nodes ($d_e = n$, n delivery node)
C	Set of edges corresponding to compressors
c_{et}^-, c_{et}^+, $c_{lt}^b \in \mathbb{R}_+$	Cost coefficients

Variables

$p_{nt} \in [p_n^{min}, p_n^{max}]$	Pressure
$q_{et}^{in} \in [q_e^{min}, q_e^{max}]$	Inflow
$q_{et}^{out} \in [q_e^{min}, q_e^{max}]$	Outflow
$b_{et} \in \mathbb{R}$	Gas consumption
$u_{et} \in \{0, 1\}$	Operation mode of compressor
$q_{et}^{out-} \in \mathbb{R}_+$	Production deficit
$q_{et}^{out+} \in \mathbb{R}_+$	Production surplus

A.3 Data sets

We have made available all data sets used in the numerical experiments in Chapter 7 on the website

http://www.uni-duisburg.de/FB11/disma/maerkert/diss.html .

There, the files *eps.tar.gz* and *gas.tar.gz* can be downloaded. We briefly describe the proceeding on a UNIX or LINUX operating system. The files need to be uncompressed by the commands

gunzip eps.tar.gz;   gunzip gas.tar.gz

This yields the two archive files *eps.tar* and *gas.tar*. They can be extracted using the commands

tar xvf eps.tar;   tar xvf gas.tar

The two resulting directories *eps* and *gas* contain the scenario sets and the model files in the subdirectories *scenarios* and the parameter sets in the subdirectories *parameters*. The tracing of the efficient frontiers of the individual EPS instances has been carried out with modified scenario sets and a modified model file. They can be found in the directory *eps/scenarios/efficient-frontier*. Each scenario file contains 10 scenarios. The single scenarios start with the identifier *scen* and with the probability of the scenario.

The parameter files include the specifications of the two-stage model, the `CPLEX 8.0` parameters, the parameters for the decomposition algorithm, and the parameters for `ConicBundle`. The format of the file depends upon the requirements of the used implementation, cf. Märkert (2004). The targets for the mean-risk models with the risk measure expected excess of a target are summarized in the files *targets.txt* and *efficient-frontier/targets.txt*.

The C-implementation of the algorithms is available on the website, too. Simply download the file *ddsip-src.tar.gz* and uncompressed and extracted it as described above. The text file *readme.txt* contains the installation instructions.

Bibliography

Ahmed, S.; M. Tawarmalani; N. V. Sahinidis (2000), A Finite Branch and Bound Algorithm for Two-Stage Stochastic Integer Programs, Stochastic Programming E-Print Series, http://www.speps.info.

Albers, S. (2003), Online Algorithms: A Survey, In: J. Tind (Ed.), Mathematical Programming 97/1-2, Series B, pp. 3-26.

Alves, M.J. (2001), Multi-Objective Mixed-Integer Programming, In: C.A. Floudas, P.M. Pardalos (Eds.), Encyclopedia of Optimization, Vol. III, Kluwer, Dordrecht, pp. 466-472.

Applequist, G.; O. Samikoglu; J.F. Pekny; G. Reklaitis (1997), Issues in the Use, Design and Evolution of Process Scheduling and Planning Systems, ISA Transactions 36/2, pp. 81-121.

Artzner, P.; F. Delbaen; J.-M. Eber; D. Heath (1999), Coherent Measures of Risk, Mathematical Finance 9/3, pp. 203-228.

Artzner, P.; F. Delbaen; J.-M. Eber; D. Heath; H. Ku (2003), Coherent Multiperiod Risk Measurement, Preprint, http://www.math.ethz.ch/~delbaen/.

Bank, B.; R. Hansel (1984), Stability of Mixed-Integer Quadratic Programming Problems, Mathematical Programming Study 21, pp. 1-17.

Bank, B.; R. Mandel (1988), Parametric Integer Optimization, Akademie-Verlag, Berlin.

Bank, B.; J. Guddat; D. Klatte; B. Kummer; K. Tammer (1982), Non-Linear Parametric Optimization, Akademie-Verlag, Berlin.

Bawa, V.S. (1982), Stochastic Dominance: A Research Bibliography, Management Science 28, pp. 698-712.

Beale, E.M.L. (1955), On Minimizing a Convex Function Subject to Linear Inequalities, Journal of Royal Statistical Society 17, pp. 173-184.

Ben-Tal, A.; A. Nemirovski (2002), A Robust Optimization – Methodology and Applications, Mathematical Programming 92, Series B, pp. 453-480.

Berge, C. (1966), Espaces Topologiques, Dunod, Paris.

Billingsley, P. (1968), Convergence of Probability Measures, Wiley, New York.

Billingsley, P. (1995), Probability and Measure. Wiley, New York.

Blömer, F.; H.-O. Günther (1999), Numerical Evaluation of MILP Models for Scheduling Chemical Batch Processes, Berlin, Technical Report 1999/05, Department of Industrial Management, TU Berlin.

Birge, J.R.; F.V. Louveaux (1997), Introduction to Stochastic Programming, Springer, New York.

Birge, J.R. (1997), Stochastic Programming Computations and Applications, INFORMS Journal on Computing 9/2, pp. 111-133.

Blair, C.E.; R.G. Jeroslow (1977), The Value Function of a Mixed Integer Program I, Discrete Mathematics 19, pp. 121-138.

Blair, C.E.; R.G. Jeroslow (1979), The Value Function of a Mixed Integer Program II, Discrete Mathematics 25, pp. 7-19.

Carøe, C.C.; R. Schultz (1998), A two-stage stochastic program for unit commitment under uncertainty in a hydro-thermal power system, Konrad-Zuse-Zentrum für Informationstechnik Berlin, Preprint SC 98-11.

Carøe, C.C.; J. Tind (1998), L-Shaped Decomposition of Two-Stage Stochastic Programs with Integer Recourse, Mathematical Programming 83/3, pp. 451-464.

Carøe, C.C.; R. Schultz (1999), Dual Decomposition in Stochastic Integer Programming, Operations Research Letters 24, pp. 37-45.

Charnes, A.; W.W. Cooper (1959), Chance Constrained Programming, Management Science 6/1, pp. 73-79.

ILOG (2004), ILOG CPLEX 8.0 Reference Manual, ILOG, Gentilly, France.

Dantzig, G.B. (1951), Maximization of a Linear Function of Variables Subject to Linear Inequalities, In: T.C. Koopmans (Ed.), Activity Analysis of Production and Allocation, Wiley, New York, pp. 339-347.

Dantzig, G.B. (1955), Linear Programming under Uncertainty, Management Science 1, pp. 197-206.

Dow Chemical (2004), SCONAPOR^{TM} – Expandable Polystyrene Website, http://www.dow.com/sconapor/english/index.htm.

Dupačová, J.; G. Consigli; S.W. Wallace (2000), Scenarios for Multistage Stochastic Programs, Annals of Operations Research 100, pp. 25-53.

Ehrgott, M. (2000), Multicriteria Optimization, Springer, Berlin.

Engell, S.; A. Märkert; G. Sand; R. Schultz; C. Schulz (2000), A Hierarchical Approach to Real-Time Scheduling of a Multiproduct Batch Plant with Uncertainties, In: Pierucci, S. (Ed.), Computer-Aided Chemical Engineering 8, European Symposium on Computer-Aided Process Engineering 10, pp. 1075-1080.

Engell, S.; A. Märkert; G. Sand; R. Schultz; C. Schulz (2001), Online Scheduling of Multiproduct Batch Plants under Uncertainty, In: M. Grötschel, S.O. Krumke, J. Rambau, Online Optimization of Large Scale Systems, Springer, Berlin, pp. 649-676.

Engell, S.; A. Märkert; G. Sand; R. Schultz (2003), Aggregated Scheduling of a Multiproduct Batch Plant by Two-Stage Stochastic Integer Programming, To appear in Optimization and Engineering.

Fishburn, P.C. (1964), Decision and Value Theory, Wiley, New York.

Fishburn, P.C. (1977), Mean-Risk Analysis with Risk Associated with Below-Target Returns, American Economic Review 67/2, pp. 116-26.

Föllmer, H.; A. Schied (2003), Convex Measures of Risk and Trading Constraints, Preprint, http://wws.mathematik.hu-berlin.de/˜foellmer.

Gröwe-Kuska, N.; M.P. Nowak; I. Wegner (2001), Modeling of Uncertainty for the Real-Time Management of Power Systems, In: M. Grötschel, S.O. Krumke, J. Rambau, Online Optimization of Large Scale Sytems, Springer, Berlin, pp. 623-647.

Hardy, G.H.; J.E. Littlewood; G. Pólya (1934), Inequalities, Cambridge University Press, Cambridge (UK).

Hadar, J.; W. Russell (1969), Rules for Ordering Uncertain Prospect, The American Economic Review 59, pp. 25-34.

Helmberg, C. (2000), Semidefinite Programming for Combinatorial Optimization, ZIB-Report ZR-00-34, Konrad-Zuse-Zentrum Berlin.

Hwang, C.-L.; A.S.M. Masud (1979), Multiple Objective Decision Making - Methods and Applications, Springer, Berlin.

Garey, M.R.; D.S. Johnson (1979), Computers and Intractability - A Guide to the Theory of NP-Completeness, New Jersey, Bell Laboratories Murray Hill.

Kall, P.; S.W. Wallace (1994), Stochastic Programming, Wiley, Chichester.

Kelling, C.; K. Reith; E. Sekirnjak (2000), A Practical Approach to Transient Optimization for Gas Networks, Pipeline Simulation Interest Group, http://www.psig.org.

Klatte, D. (1985), On the Stability of Local and Global Optimal Solutions in Parametric Problems of Nonlinear Programming Part I and II, Seminarbericht No. 75 of the Sektion Mathematik of the Humbold-University Berlin.

Klein Haneveld, W.K.; M.H. van der Vlerk (1999), Stochastic Integer Programming: General Models and Algorithms, Annals of Operations Research 85, pp. 39-57.

Korhonen, P. (2001), Multiple Objective Programming Support, In: C.A. Floudas, P.M. Pardalos (Eds.), Encyclopedia of Optimization III, Kluwer, Dordrecht, pp. 566-574.

Laporte, G.; F.V. Louveaux (1993), The Integer L-Shape Method for Stochastic Integer Programs with Complete Recourse, Operations Research Letters 13, pp. 133-142.

Levy, H. (1992), Stochastic Dominance and Expected Utility: Survey and Analysis, Management Science 38, pp. 555-593.

Löhl, T.; C. Schulz; S. Engell (1998), Sequencing of Batch Operations for a Highly Coupled Production Process: Genetic Algorithms vs. Mathematical Programming, Computers & Chemical Engineering 22, pp. 579–585.

Louveaux, F.V.; R. Schultz (2003), Stochastic Integer Programming, In: A. Ruszczynski, A. Shapiro (Eds.), Stochastic Programming, Elsevier, North-Holland.

Markowitz, H.M. (1959), Portfolio selection: Efficient Diversification of Investments, Wiley, New York.

Märkert, A.; S. Tiedemann (2003), Risk Aversion in Stochastic Programming with Mixed-Integer Recourse I and II, Lectures given at the 18th International Symposium on Mathematical Programming (ISMP), Copenhagen.

Märkert, A. (2004), User's guide to `ddsip` – A C Package for the Dual Decomposition of Stochastic Programs with Mixed-Integer Recourse, Preprint, http://www.uni-duisburg.de/FB11/disma/ddsip.shtml.

Meyer, R.R. (1974), On the Existence of an Optimal Solution to Integer and Mixed Integer Programs, Mathematical Programming 7, pp. 223-235.

Mockus, L.; G.V. Reklaitis (1997), Mathematical Programming Formulation for Scheduling of Batch Operations based on Nonuniform Time Discretization, Computers & Chemical Engineering 21, pp. 1147-1156.

Morgenstern, O.; J. von Neumann (1947), Theory of Games and Economic Behavior. Princeton University Press.

Müller, A.; D. Stoyan (2002), Comparison Methods for Stochastic Models and Risk, Wiley, Chichester.

Nemhauser, G.; L. Wolsey (1988), Integer and Combinatorial Optimization, Wiley, New York.

Niño-Mora, J. (2001), Stochastic Scheduling, In: C.A. Floudas, P.M. Pardalos (Eds.), Encyclopedia of Optimization, Vol. III, Kluwer, Dordrecht, pp. 367-372.

Nowak, M.P.; W. Römisch (2000), Stochastic Lagrangian Relaxation Applied to Power Scheduling in a Hydro-Thermal System under Uncertainty, Annals of Operations Research 100, pp. 251-272.

Ogryszak, W.; A. Ruszczinsky (1999), From Stochastic Dominance to Mean-Risk Models: Semideviations as Risk Measures, European Journal of Operations Research 116, 33-50.

Ogryszak, W.; A. Ruszczinsky (2001), On Consistency of Stochastic Dominance and Mean-Semideviation Models, Mathematical Programming 89, Series B, pp. 217-232.

Ogryszak, W.; A. Ruszczinsky (2002), Dual Stochastic Dominance and Related Mean-Risk Models, SIAM Journal on Optimization 13, pp. 60-78.

Pflug, G. (1999), How to Measure Risk, Festschrift for F. Ferschl, Physica Verlag, Würzburg.

Pinto, J.M.; I.E. Grossmann (1996), A Continuous Time MILP Model for Short Term Scheduling of Batch Plants with Pre-ordering Constraints, Proc. Sixth European Symposium on Computer Aided Process Engineering (ESCAPE–6), pp. 1197–1202.

Pinto, J.M.; I.E. Grossmann (1998), Assignment and Sequencing Models for the Scheduling of Process Systems, Annals of Operations Research 81, pp. 433-466.

Pistikopoulos, E.N.; T.V. Thomaidis, A. Melix, M.G. Ierapetritous (1996), Flexibility, Reliability and Maintenance Considerations in Batch Plant Design under Uncertainty, Computers & Chemical Engineering 20, pp. 1209-1214.

Prekopa, A. (1995), Stochastic Programming, Kluwer, Dordrecht.

Prekopa, A. (2003), Probabilistic Programming, In: A. Ruszczynski, A. Shapiro (Eds.), Stochastic Programming, Elsevier, North-Holland.

Qurik, J.P.; R. Saposnik (1962), Admissibility and Measurability of Utility Functions, Review of Economical Studies 29, pp. 140-146.

Rauch, J. (1998), Mehrproduktanlagen, Wiley, Weinheim.

Reith, K.; E. Sekirnjak (2001), Transiente Technische Optimierung (TTO-Prototyp), Confidential Documentation, Ruhrgas AG Essen, PSI Berlin.

Reklaitis, G.V. (1996), Overview of Scheduling and Planning of Batch Process Operations, In: G.V. Reklaitis, A.K. Sunol, D.W.T. Rippin, D. Hortasçu (Eds.) (1996), Batch Processing System Engineering: Fundamentals and Applications for Chemical Engineering, Springer, Berlin, pp. 660-705.

Robinson, S.M. (1987), Local Epi-Continuity and Local Optimization, Mathematical Programming 37, pp. 208-222.

Robinson, S.M.; R.J.-B. Wets (1987), Stability in Two-Stage Stochastic Programming, SIAM Journal on Control and Optimization 25/6, pp. 1409-1416.

Rockafellar, R.T.; R.J.-B. Wets (1998), Variational Analysis, Springer, Berlin.

Rockafellar, R.T.; S. Uryasev (2002), Conditional Value at Risk for General Loss Distributions, Journal of Banking and Finance 26/7, pp. 1443-1471.

Rockafellar, R.T.; S. Uryasev, M. Zabarankin (2002), Deviation Measures in Risk Analysis and Optimization, Research Report #2002-07, Risk Management and Financial Engineering Lab, University of Florida, Gainesville.

Römisch, W.; R. Schultz (1993), Stability of Solutions for Stochastic Programs with Complete Recourse, Mathematics of Operations Research 18/3, pp. 590-609.

Römisch, W.; R. Schultz (2001), Multi-stage Stochastic Integer Programs, In: M. Grötschel, S.O. Krumke, J. Rambau, Online Optimization of Large Scale Sytems, Springer, Berlin, pp. 579-598.

Ruhrgas AG (2004), Technology Website, http://www.ruhrgas.de/englisch.

Schilling, G.; C.C. Pantelides (1996), A Simple Continuous-Time Process Scheduling Formulation and A Novel Solution Algorithm, Computers & Chemical Engineering 20, pp. S1221-S1226.

Schulz, C. (2001), Modellierung und Optimierung einer verfahrenstechnischen Mehrproduktanlage, Doctoral thesis, Process Control Laboratory, University of Dortmund, Shaker-Verlag, Aachen.

Schultz, R. (1995), On Structure and Stability in Stochastic Programs with Random Technology Matrix and Complete Integer Recourse, Mathematical Programming 70, pp. 73-86.

Schultz, R. (2002), Mixed-Integer Value Functions in Stochastic Programming, Preprint 514-2002, Mathematical Institute, University Duisburg-Essen.

Schultz, R.; S. Tiedemann (2003), Risk Aversion via Excess Probabilities in Stochastic Programs with Mixed-Integer Recourse, SIAM Journal on Optimization 14, pp. 115-138.

Shah, N.; C.C. Pantelides, R.W.H. Sargent (1993), A General Algorithm for Short-Term Scheduling of Batch Operations – II. Computational Issues, Computers & Chemical Engineering 17, pp. 229-244.

Shah, N. (1998), Single and Multisite Planning and Scheduling: Current Status and Future Challenges, In J.F. Pekny, G.F. Blau (Eds.), AIChE Symposium Series 320, Foundations of Computer-Aided Process Operations, pp. 75-90.

Steinbach, M.C. (2001), Markowitz Revisited: Mean-Variance Models in Financial Portfolio Analysis, SIAM Review 43/1, pp. 31-85.

Stojanov, J.M. (1987), Counterexamples in Probability, Wiley, Chichester.

Stougie, L.; M.H. van der Vlerk (2003), Approximation in Stochastic Integer Programming, SOM Research Report 03A14, http://som.rug.nl.

Subrahmanyam, S.; J.F. Pekny, G.V. Reklaitis (1994), Design of Batch Chemical Plants under Market Uncertainty, Industrial & Engineering Chemistry Research 33, pp. 2688-2701.

Tanino, T.; H. Kuk (2002), Nonlinear Multiobjective Programming, In: M. Ehrgott, X. Gandibleux (Eds.), Multiple Criteria Optimization: State of the Art Annotated Bibliographic Surveys, Kluwer, Boston, pp. 71-128.

Tiedemann, S. (2004), Conditional Value-at-Risk in Stochastic Programs with Mixed-Integer Recourse, Manuscript in preparation.

van der Vlerk, M.H. (1995), Stochastic Programming with Integer Recourse, PhD thesis, University of Groningen, The Netherlands.

van Slyke, R.; R.J.-B. Wets (1969), L-shaped Linear Programs with Applications to Control and Stochastic Programming, SIAM Journal on Applied Mathematics 17, pp. 638-663.

Walkup, D.; R.J.-B. Wets (1967), Stochastic Programs with Recourse, SIAM Journal on Applied Mathematics 15, pp. 1299-1314.

Westphalen, M. (2004), Anwendungen der stochastischen Optimierung im Stromhandel und Gastransport, Doctoral thesis, Institute of Mathematics, University Duisburg-Essen, Campus Duisburg.

Wets, R.J.-B. (1974), Stochastic Programs with Fixed Recourse: The Equivalent Deterministic Program, SIAM Review 16, pp. 309-339.