Models and Algorithms for Dominance-Constrained Stochastic Programs with Recourse

Von der Fakultät für Mathematik der
Universität Duisburg-Essen
zur Erlangung des akademischen Grades eines
Dr. rer. nat.

angenommene Dissertation

von

Herrn Dipl.-Math. Dimitri Drapkin

Referent: Prof. Dr. Rüdiger Schultz

Koreferent: Prof. Dr. Maarten H. van der Vlerk

Datum der Disputation: 30. April 2014

Acknowledgements

It gives me great pleasure to thank first and foremost my supervisor Professor Dr. Rüdiger Schultz for his constant support throughout my time as his student - without his support this thesis would not have been possible. I have been extremely lucky to have a supervisor who cared so much about the advancements of my work above and beyond pure mathematical questions.

I owe my deepest gratitude to Dr. Ralf Gollmer for introducing me to the world of C, Linux, CPLEX, etc., and providing invaluable advice in many desperate situations.

Completing this work would have been all the more difficult were it not for the support and friendship provided by the other members of the working group "Optimization and Algorithmic Discrete Mathematics".

A very special word of thanks goes to the former members of this group Dr. Uwe Gotzes and Oliver Klaar for their contributions to the foundations of this thesis.

Finally, I must express my gratitude to my wife, Svetlana, for her continued support and encouragement during all the ups and downs of my research.

Abstract

We consider optimization problems with stochastic order constraints of first and second order posed on random variables coming from two-stage stochastic programs with recourse. We clarify the theoretical relevance of these specific problems, and contribute to improving their computational tractability. For the latter, we review and enhance mixed-integer linear programming (MILP) equivalents. These exist for either mixed-integer or continuous variables in the second stage. Algorithmically, our focus is on developing tailored cutting-plane decomposition methods for these models.

Stochastic mixed-integer programming, stochastic dominance, decomposition methods, cutting-plane methods, risk aversion.

Zusammenfassung

In der vorliegenden Dissertationsschrift befassen wir uns mit stochastischen Optimierungsproblemen unter Nebenbedingungen, die mithilfe stochastischer Ordnungen formuliert sind. Hierbei konzentrieren wir uns auf stochastische Dominanz erster Ordnung und die steigende konvexe Ordnung, wobei beide Ordnungen in unserem Fall auf Zufallsgrößen operieren, welche Optimalwerten zweistufiger stochastischer Optimierungsprobleme mit Kompensation entsprechen.

Wir stellen die theoretische Relevanz der vorliegenden Problemklasse heraus und tragen zur Entwicklung von effizienten Lösungsverfahren bei. Um Letzteres zu erreichen untersuchen und erweitern wir bestehende gemischt-ganzzahlige lineare Repräsentationen dieser Probleme und entwickeln maßgeschneiderte Dekompositionsverfahren. Der Schwerpunkt dieser Arbeit liegt dabei auf der Entwicklung und Implementierung besonders effizienter Lösungsansätze für den Fall mit linearer Kompensation.

Stochastische gemischt-ganzzahlige Optimierung, Stochastische Dominanz, Dekompositionsverfahren, Schnittebenenverfahren, Risikoaversion.

Contents

List of Figures

List of Tables

List of Abbreviations

cdf cumulative distribution function. 10

CPT cumulative prospect theory. 6

DARA declining absolute risk aversion. 5

EUT expected utility theory. 5

FSD first-order stochastic dominance. 8

ICC integrated chance constraints. 27

ICV increasing concave order. 12

ICX increasing convex order. 18

MILP mixed-integer linear programming. 2

PSD prospect stochastic dominance. 9

RDEUT rank-dependent expected utility theory. 6

rv random variable. 10

SD stochastic dominance. 7

SP stochastic programming. 1

SSD second-order stochastic dominance. 9

Symbol Index

Chapter 1

Introduction

Good decisions have always been connected with mastering some kind of uncertainty. In former times experience and common sense used to be the only aids to find a good path. Starting from the middle of the twentieth century, stochastic programming (SP) emerged, at the interface of probability and optimization theory, to become a discipline of science aiming at exploration, development and improvement of models for decision making under uncertainty.

Today, stochastic programming has a great variety of applications from sports, e.g., yacht racing [Phi05], over management of risks related to natural disasters, e.g., of flood and seismic risks [EE05], to manifold applications in finance [DHv02]. Even the problem of finding an optimal shape for an elastic body (e.g., a cantilever) under uncertain loading configurations was recently formulated and solved by means of (infinite-dimensional) stochastic programming, see [CHP+09].

Though the achievements in this field have been remarkable, a sensible handling of uncertainty and risks seems to be more important than ever to master the challenges of the day. Various aspects of risk management are subject to constant debate in science and society. Also the mathematical models are constantly getting larger and gain complexity, providing a motivation for ongoing research.

In the present thesis, we will deal with a specific decision making framework of dominance-constrained stochastic programs with recourse. Our aim will be, on the one hand, to clarify the theoretical relevance of this problem class in view of recent developments in the adjacent fields of decision theory, risk modeling and stochastic programming. On the other hand, to strengthen the relevance of dominance-constrained problems in practical decision making under uncertainty, we will concentrate on the algorithmic aspects of these problems proposing new and enhancing existing solution techniques.

The thesis is structured as follows. In Chapter 2, we define relations of stochastic dominance and review their decision theoretical background. Also some computationally tractable representations for SD and an outline of its connections with the related concept of risk measures are presented. In Chapter 3, we expose how to incorporate SD into the established optimization framework of SP, and introduce our problem class of dominance-constrained stochastic problems with recourse. In Chapter 4, mixed integer linear programming (MILP) equivalent formulations for such problems are developed and enhanced. Starting from Chapter 5, we concentrate on the case of linear recourse, i.e., on problems without integer variables in the second stage. Model equivalents based on duality are derived for these problems in Chapter 5, whereas cutting plane decomposition methods are proposed in Chapter 6. Lastly, computational results presented in Chapter 7 indicate the effectiveness of our approach and conclude the thesis.

Chapter 2

Comparing Risks for Decision Making under Uncertainty

2.1 Stochastic Dominance and Decision Theory

Decisions we are making today tend to have prospects observable in the future, only. The problem of making good decisions *prior* to having the exact information from the future, is the fundamental matter of stochastic programming. Thus, stochastic programming can be understood as optimization under information or nonanticipativity constraints. Though we cannot anticipate the future, in the present thesis we assume, that all relevant uncertain quantities can be identified and modeled as random variables with distributions known to the decision maker.[1]

Once the basic model is set, the question of a sensible comparison of random outcomes arises, since it has to be clarified what a "good" decision should be. The first sound idea of such a comparison yielded the concept of the *expected*

[1] The problem of selecting an appropriate basic probability model is referred to as the *ambiguity problem*, cf. [RP07] and the discussion therein, in contrast to the so-called *uncertainty problem* treated here.

value, introduced in the 17th century by Blaise Pascal as a "fair" solution for the "Problem of Points"[2]. Symptomatically, the question of what should be considered "good" or "fair" was even then not only a mathematical question.

In today's terms, and at least in case of many repetitions of the same setting, optimizing the expected value of the prospects is justified by the Law of Large Numbers. In fact, the average of a prospect will converge with the number of repetitions to its expected value, meaning that the obtained solution would be optimal on average.

The drawback of the expectation based approach is its complete neglect of the risk incurred by concrete realizations of the random outcome. Ignoring the risk, however, may easily lead to inferior or even completely implausible decisions. One famous example of such a situation is that of the St. Petersburg Lottery: a game with infinite expected payoff, for which there is "no person of good sense, who would wish to give 20 coins"[3].

To resolve this difficulty, Daniel Bernoulli proposed to measure the utility of an outcome numerically (as the logarithm of one's monetary possessions) and to optimize the "mean utility" instead of the expectation itself.[4] Bernoulli's work inspired the concept of *marginal utility* and gave rise to a *cardinal utility* theory, which made the notion of utility indispensable in economics.

However, it proved problematic (if not impossible) to determine utility functions explicitly.[5] This difficulty led to the development of normative models, starting from the beginning of the 20th century. In these models, systems of a

[2]According to [Kat98, Chapter 11.3], the discussion of this problem belongs to the earliest beginnings of probability theory.

[3]As Gabriel Cramer put it in his correspondence with Nicolas Bernoulli in 1728, cf. [Ber75] for the edifying discussion of St. Petersburg paradox.

[4]Cf. [Ber54] for the English translation of D. Bernoulli's seminal "Exposition of a New Theory on the Measurement of Risk", originally published in 1738 in Latin.

[5]The effort in utility measurement was considerable and produced some interesting concepts (e.g., Edgeworth's hedonimeter), cf. [Col07]. In the modern economic discourse this *experienced utility* was largely replaced by *decision utility* which refers to the prospect's weight in decisions and can be inferred from observed choice, cf. [KWS97].

few axioms were proposed to describe preferences which should be consistent across different choice problems.

A (preliminary) culmination of these efforts was the development of the *expected utility theory* (EUT) by von Neumann and Morgenstern. In their fundamental work [VNM44], a notion of a *rational decision maker* - defined as an agent obedient towards a given set of four intuitive axioms - was introduced. This rational decision maker was proven to possess a utility function $u(\cdot)$[6] such that he would prefer a prospect Y to X iff

$$\mathbb{E}[u(\mathsf{Y})] \geq \mathbb{E}[u(\mathsf{X})].^{7} \tag{2.1}$$

In the framework of EUT the study of attitudes towards risk is of special importance. A decision maker is called *risk-averse* if he prefers the expected value of a prospect to the random prospect itself, i.e., his utility function is concave with

$$\mathbb{E}(u(\mathsf{X})) \leq u(\mathbb{E}(\mathsf{X})). \tag{2.2}$$

Otherwise, he is called *risk-seeking* with a convex utility.[8] To compare risk aversion between individuals[9] some measures of risk aversion were introduced, most notably the *Arrow-Pratt measure of absolute risk-aversion* $\rho(x) := -\frac{u''(x)}{u'(x)}$, cf. [Pra64, Arr65]. With the help of this measure it is possible to formulate the plausible assumption of *declining absolute risk aversion* (DARA) which is characterized by a non-increasing ρ ($\rho' \leq 0$), cf. [Vic75].[10]

Being a powerful tool for decision making under uncertainty, EUT recently came under pressure, both from the descriptive and the normative side. On

[6] Such utility functions can be determined up to an affine transformation through the analysis of a decision maker's preferences between *simple lotteries*. For a more practically successful approach the author refers to [ADH77].

[7] From now on we assume the utility functions to be differentiable sufficiently often, all expected values are assumed to exist.

[8] Both observations are a direct consequence of Jensen's inequality.

[9] This task is not straightforward because utility functions are lacking uniqueness.

[10] E.g., Bernoulli's logarithmic utility function exposed DARA.

the one hand, there is empirical evidence of behavioral patterns which systematically violate EUT, cf. [All53, KT79].[11] For example, individuals systematically overweight low-probability events and show different attitudes towards gains and losses with respect to the status quo (e.g., buy lottery tickets and insurance contracts simultaneously). On the other hand, fundamentally distinct notions of attitude towards risk and attitude towards wealth coincide in EUT thus leading to the question whether these concepts could be decoupled. Several generalizations of EUT have recently been proposed to resolve these drawbacks.

The *rank-dependent expected utility theory* (RDEUT) elaborated on the observed subjective probability distortion. It was originally presented by Quiggin ([Qui82]) and developed for a special case by Yaari ([Yaa87])[12] under the name *dual utility theory*. From a (weaker) set of axioms a utility function $u(\cdot)$ and a nondecreasing probability transformation function $q(\cdot) : [0,1] \to [0,1]$ were proven to exist, s.t. a prospect Y is preferred to X iff

$$\int u(t)\, d(q \circ F_{\mathsf{Y}})(t) \geq \int u(t)\, d(q \circ F_{\mathsf{X}})(t). \tag{2.3}$$

A further generalization of EUT is the highly praised *cumulative prospect theory* (CPT) of Kahneman and Tversky, cf. [KT92].[13] In this theory, gains and losses are considered separately: the utility function is assumed to be convex for the losses and concave for the gains (i.e., $u(\cdot)$ is S-shaped); separate reverse S-shaped probability distortion functions are proposed for either case. While most experimental studies support CPT (thus explaining its popularity), we refer to [LL02a, LL02b] and the references therein for some critical findings.[14]

So far, we have compared the riskiness of a prospect from the individual

[11] The work [KT79] is regarded as the fundamental paper in *behavioral economics*.

[12] In Yaari's version, $u(\cdot)$ is assumed to be the identity function and the probability transformation function is referred to as a *dual utility function*.

[13] In particular for the development of CPT, Kahneman received the Nobel Prize in Economics in 2002.

[14] In the quoted articles particularly the S-shape of a utility function is rejected by studies envolving mixed (partly positive, partly negative) prospects. However, [Wak03] shows, that

perspective of a given decision maker, pointing out the properties of his utility functions connected with his attitude towards risk. In practice, the knowledge of the concrete shape of a utility function is at best partial and it makes sense to consider classes of utility functions characterizing typical risk attitudes.

This idea leads us to a more general approach of *stochastic dominance* (SD), which will be the main subject of the present thesis. In the context of decision theory, SD enables a direct comparison of prospects by means of EUT. More precisely, one prospect will be said to *dominate* the other if it is preferred by all individuals with utility functions in a given class.

Under the assumed preference of more money to less, the most general class $\mathcal{U}_1$ of utility functions will include all nondecreasing functions u (with $u' \geq 0$). We have already seen that in EUT risk-aversion is equivalent to concavity of the utility function, hence the class $\mathcal{U}_2$ will contain all concave utility functions from $\mathcal{U}_1$ ($u' \geq 0$, $u'' \leq 0$).

More generally, one could consider utility functions whose derivatives alternate in sign, i.e, belong to the class $\mathcal{U}_n := \{u \in \mathcal{U}_{n-1} : (-1)^n u^{(n)} \leq 0\}$, $n > 1$, whereby the economic interpretation for $n > 3$ is not evident.[15] The importance of $\mathcal{U}_3$ is related to the fact that $\mathcal{U}_3 \supset \mathcal{U}_{DARA} := \{u \in \mathcal{U}_2 : u' \neq 0,\, \rho' \leq 0\}$ and the conditions in $\mathcal{U}_3$ are easier to check.[16]

Corresponding stochastic dominance relations are then defined in a straightforward way:

Definition 2.1.1. *For random variables* X *and* Y*, we define* Y *to dominate* X *w.r.t. nth order stochastic dominance, written* $\mathsf{X} \preceq_{(n)} \mathsf{Y}$ *iff*

$$\mathbb{E}u(\mathsf{X}) \leq \mathbb{E}u(\mathsf{Y})\, \forall u \in \mathcal{U}_n. \tag{2.4}$$

due to probability distortion, the studies actually support the CPT - an argument countered in [LL03] for some special cases.

[15]Such classes of utility functions were generalized for all real numbers $n > 0$, cf. [Fis76, Fis80].

[16]For more details the author refers to [Whi70, FV78, Baw75].

Stochastic dominance w.r.t. DARA utility functions is defined analogously.[17]

Remark 2.1.2. *In the present thesis, we define all dominance relations in the so-called* weak *form, i.e., we do not exclude the possibility of simultaneously* $\mathsf{X} \preceq \mathsf{Y}$ *and* $\mathsf{Y} \preceq \mathsf{X}$*. For any (weak) dominance relation "*$\preceq$*", the corresponding* strict *form "*$\prec$*" is given through the standard rule*

$$\mathsf{X} \prec \mathsf{Y} \Leftrightarrow \mathsf{X} \preceq \mathsf{Y} \text{ and } \mathsf{Y} \not\preceq \mathsf{X}. \tag{2.5}$$

In the generalizations of EUT the defined dominance rules only make sense if they are consistent with the underlying decision model. More precisely, if Y is preferred to X w.r.t. a dominance relation, it should also be preferred in the corresponding model by all decision makers with utility functions in the given class.[18]

In the rank-dependent expected utility theory (RDEUT), consistency follows for first-order stochastic dominance (FSD) from monotonicity of the probability weighting function q in view of (2.3), cf. [Qui82, Proposition 3]. Due to its interpretation (as a general preference of more to less), consistency with FSD is such a fundamental property, that it holds for most generalizations of EUT.[19] FSD is also sometimes referred to as the axiom of "absolute preference", because it can be used to axiomatize RDEUT ([Qui92, Yaa87]).

Lacking consistency with FSD of the original version of the prospect theory ([KT79]), was considered such a large drawback that it partly inspired the development of CPT ([KT92]) to elaborate on this issue.

The situation with second-order stochastic dominance (SSD) in the generalized models is more complex. In RDEUT a decision maker is risk-averse (in the sense of preferring certainty over risk, see above) iff he is characterized

[17] For more details on this dominance relation the author refers to [Vic75, Vic77].

[18] For EUT this consistency requirement is immediate from the definition of dominance relations.

[19] Especially those generalizations of EUT concerning relaxation of the independence axiom and partly even the transitivity axiom, cf. [Lev92, p. 559 - 560].

by a concave utility function and a pessimistic transformation of probabilities.[20] SSD consistency is preserved if both $u(\cdot)$ and $q(\cdot)$ are concave, which is a weaker condition than risk-aversion, cf. [Qui92, p. 80].

Due to the proposed S-shape of utility functions, which explicitly assumes a partly risk-seeking behavior of the decision maker, CPT is of course not consistent with SSD. The notion of *prospect stochastic dominance* (PSD) which considers all S-shaped utility functions and is consistent with prospect theory was proposed in [LW98, Lev06]. However, consistency problems with CPT arise in view of reverse S-shaped probability distortion functions proposed there, cf. [LL02a, p. 1065]. Recently, several even more general classes of dominance relations have been proposed which also take account of the relevant probability distortion functions, cf. [BH06]. Nevertheless, the author takes the view that a broadly accepted and computationally tractable stochastic dominance theory for the CPT still has to be developed.

In view of the above discussion, in the present thesis we will consider the FSD relation, because it is the most fundamental SD rule consistent with all most prominent decision models, and the SSD relation because of its account for risk-aversion in EUT and its good mathematical and computational properties, cf. [DR03].

In the context of decision theory, SD was introduced by a number of authors starting from 1960's, most notably Quirk and Saposnik ([QS62]), Fishburn ([Fis64]), Hadar and Russell ([HR69]), Hanoch and Levy ([HL69]) and Rothschild and Stiglitz ([RS69]).[21] A detailed survey of SD rules mainly from the economical perspective can be found, e.g, in [WF78] and [Lev92].

[20]Pessimism means that bad outcomes receive larger and good outcomes smaller probabilities. For the exact definition and the proof see [Qui92, pp. 77].

[21]For a more complete list of early contributions we refer to [Baw82].

2.2 Stochastic Orders and Measures of Risk

The concept of SD is the decision-theoretical counterpart of the more general concept of *stochastic orders*, which was developed in statistics starting from the late 1940's, cf. [MW47], [Bla51, Bla53] and [Leh55] for early references[22] and [MS02], [SS07] for a contemporary discussion.

Stochastic ordering aims at imposing sensible orders[23] on the set of cumulative distribution functions (cdf) of real-valued random variables (rv) defined on a probability space $(\Omega, \mathcal{F}, I\!\!P)$. In the present thesis, ordering of rvs will not be distinguished from ordering of their corresponding cdfs. In these terms, we have seen in the previous section that EUT imposes a total ordering (through (2.1)) in case the utility function is given, and a partial ordering through SD rules. Another possibility to obtain ordering relations will employ *acceptability* and *risk* functionals defined on rvs[24], which we will discuss in the present section focusing on their relations to SD.

As we have pointed out above, SD relations were developed for decision makers preferring big outcomes to small, so that a large body of literature exists for this case. On the other hand, some important risk functionals have more natural interpretations for rvs representing losses instead of gains. Starting from the classical setting, we will illustrate here the transition from the one case to the other. In this way, we will obtain the intuition and the representations we will need in the main part of the thesis for the discussion of our minimization framework.

[22]These works were in turn inspired by earlier findings in majorization theory, cf. [HLP34] and [MOA11] for an overview.

[23]Formally, a (partial) order is a reflexive, transitive and antisymmetric binary relation over an arbitrary set. The order is called total if any two elements in the set are comparable under the relation.

[24]More precisely, we will use only *law-invariant* or *version-independent* functionals, which depend on the cdf only, cf. [RP07, Definition 2.1].

2.2.1 Preference of Large Outcomes

In the context of statistics, SD rules are equivalently expressed through a pointwise comparison of some performance functions constructed from a rv's cdf. For FSD, this performance function is the cdf itself and the following (primal) characterizations hold.

Proposition 2.2.1. *For rvs* $\mathsf{X}, \mathsf{Y} \in (\Omega, \mathcal{F}, I\!P)$ *with cdfs* F_{X} *and* F_{Y} *the following statements are equivalent:*

(i) $\mathsf{X} \preceq_{(1)} \mathsf{Y}$;

(ii) $F_{\mathsf{X}}(t) \geq F_{\mathsf{Y}}(t) \ \forall \ t \in \mathbb{R}$;

(iii) there exists a probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{I\!P})$ *and random variables* $\hat{\mathsf{X}}$ *and* $\hat{\mathsf{Y}}$ *with marginals* F_{X} *and* F_{Y} *such that* $\hat{\mathsf{X}}(\hat{\omega}) \leq \hat{\mathsf{Y}}(\hat{\omega})$ *for all* $\hat{\omega} \in \hat{\Omega}$;

(iv) $I\!P(\mathsf{X} > t) \leq I\!P(\mathsf{Y} > t) \ \forall \ t \in \mathbb{R}$.

Proof. (i)⇔(ii) Theorem 1.2.8 in [MS02], (ii)⇔(iii) Theorem 1.2.4 in [MS02], (iv)⇔(ii) clear. □

In other words, the preference of more to less in EUT can be equivalently described by a pointwise comparison of cdfs (which implies that the smaller rv should take smaller values with a higher probability) and is closely related to the simple pointwise comparison of rvs. Thus, being the most fundamental version-independent ordering concept for rvs, FSD is often just referred to as the *(usual) stochastic order*, cf. [SS07, p. 3].

The function $\bar{F}_{\mathsf{X}}(t) := I\!P(\mathsf{X} > t)$ from the representation (iv), which we will use to derive computationally tractable representations for FSD starting from Chapter 4, denotes the so-called *survival function*, well-known, e.g., in actuarial sciences, cf. [Pro11, p. 194].

For SSD, computationally more tractable representations can be derived by means of the second performance function

$$F_{\mathsf{X}}^{(2)}(t) := \int_{-\infty}^{t} F_{\mathsf{X}}(\alpha)\, d\alpha \quad \forall\, t \in \mathbb{R} \tag{2.6}$$

as follows.

Proposition 2.2.2. *For rvs* $\mathsf{X}, \mathsf{Y} \in \mathcal{L}_1(\Omega, \mathcal{F}, I\!\!P)$ *with cdfs* F_{X} *and* F_{Y} *the following statements are equivalent:*

(i) $\mathsf{X} \preceq_{(2)} \mathsf{Y}$*;*

(ii) $I\!\!E[t - \mathsf{X}]_+ \;\geq\; I\!\!E[t - \mathsf{Y}]_+$ *for all* $t \in \mathbb{R}$*, where* $[\alpha]_+ := \max(\alpha, 0)$ *is the positive part of* α*;*

(iii) $F_{\mathsf{X}}^{(2)}(t) \geq F_{\mathsf{Y}}^{(2)}(t)$ *for all* $t \in \mathbb{R}$*.*

Proof. See Theorem 4.A.2 in [SS07]. □

Due to the definition of SSD by nondecreasing concave utility functions in Definition 2.1.1 this order is also often called the *increasing concave order* (ICV). The observation that only a small subset of these functions is sufficient to fully characterize ICV leads to (ii). From the integral condition (iii) it is here again immediate that FSD implies SSD, because (iii) can be interpreted as a requirement for the area enclosed between the two cdfs to be non-negative up to every point t, which is a weaker condition than a pointwise comparison of the cdfs.

Though the above representations are more tractable, checking both dominance relations implies comparison of performance functions on infinitely many points, which is a complex task. In fact, a much easier approach to compare distributions was developed in probability theory from its very beginnings: namely the study and comparison of cdfs by their relevant parameters. These parameters are distinguished between a value dimension and a risk dimension.

Since Pascal, the value dimension is typically represented by the expected value of the prospect. At this, it follows directly from the definitions that

$$\mathsf{X} \preceq_{(i)} \mathsf{Y} \Longrightarrow \mathbb{E}(\mathsf{X}) \leq \mathbb{E}(\mathsf{Y}), \text{ for } i = 1, 2, \tag{2.7}$$

because the identity function is both increasing and concave.

To address the risk dimension, a wide variety of (version-independent) risk functionals $\mathcal{R}(\cdot)$ on the space of rvs has been defined which characterize the riskiness of the whole cdf by a scalar. Keeping both dimensions separate leads to a bi-criteria decision problem, while the relative importance of value to risk represents the *risk aversion* in this context.[25] These considerations lead to the definition of the following relations.

Definition 2.2.3. *For rvs* $\mathsf{X}, \mathsf{Y} \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$*, we define*

$$\mathsf{X} \preceq_{\mathbb{E},\mathcal{R}} \mathsf{Y} \text{ iff } \mathbb{E}(\mathsf{X}) \leq \mathbb{E}(\mathsf{Y}) \text{ and } \mathcal{R}(\mathsf{X}) \geq \mathcal{R}(\mathsf{Y}); \tag{2.8}$$

and

$$\mathsf{X} \preceq_{\mathbb{E}-\lambda\mathcal{R}} \mathsf{Y} \text{ iff } \mathbb{E}(\mathsf{X}) - \lambda\mathcal{R}(\mathsf{X}) \leq \mathbb{E}(\mathsf{Y}) - \lambda\mathcal{R}(\mathsf{Y}), \tag{2.9}$$

where $\lambda > 0$ *is an assumed degree of risk aversion.*

The relation (2.8) is called *mean-risk dominance*, whereas the approach of maximizing $\mathbb{E}(\mathsf{X}) - \lambda\mathcal{R}(\mathsf{X})$ constitutes the so-called *mean-risk approach.* This approach was pioneered by Markowitz in his seminal work [Mar52] and still is very popular among practitioners and researchers due to its excellent computational tractability.

However, to justify the mean-risk approach theoretically, compatibility with the findings of decision theory has to be verified. In particular, the risk functional should be such that the model becomes *consistent* with FSD and preferably also with SSD (to account for risk-aversion in EUT) in the sense of the following definitions.

[25] See (3.12) for the formulation of this bi-criteria problem for the minimization case.

Definition 2.2.4. *The mean-risk model* $(\mathbb{E}, \mathcal{R})$ *is said to be consistent with i-th order SD if*

$$\mathsf{X} \preceq_{(i)} \mathsf{Y} \Longrightarrow \mathsf{X} \preceq_{\mathbb{E},\mathcal{R}} \mathsf{Y}, \tag{2.10}$$

and $\lambda-$*consistent with i-th order SD if*

$$\mathsf{X} \preceq_{(i)} \mathsf{Y} \Longrightarrow \mathsf{X} \preceq_{\mathbb{E}-\lambda\mathcal{R}} \mathsf{Y} \tag{2.11}$$

for some $\lambda > 0$ *and* $i = 1, 2$.

Thus, a strict maximum of a $\lambda-$consistent model will not be dominated w.r.t. the corresponding dominance relation, cf. Section 3.1.3. Clearly, (2.10) implies (2.11) for all $\lambda > 0$ and consistency with SSD implies consistency with FSD (but not vice versa).

The seminal mean-risk model presented in [Mar52] considered the variance as a risk functional. This model was heavily criticized for being inconsistent with FSD. Also variance as a risk measure is in many ways not adequate.[26] On the other hand, e.g., the expected shortfall[27] below some fixed target t defined as $\mathbb{E}[t-\mathsf{X}]_+$ is a sensible risk measure, which yields an SSD-consistent mean-risk model in view of Proposition 2.2.2 (ii). Moreover, in this way SSD can be described as a continuum of constraints on this risk measure.[28]

Examples of sensible risk functionals taking into account all gains *below* the mean, which are not consistent but only 1-consistent with SSD (cf. [OR99, OR02]) are *lower absolute semideviation*

$$ASD^-(\mathsf{X}) := \mathbb{E}([\mathbb{E}(\mathsf{X}) - \mathsf{X}]_+) = \frac{1}{2}\int_{-\infty}^{\infty} |t - \mathbb{E}(\mathsf{X})| \, d\mathbb{P}_{\mathsf{X}}(t) \tag{2.12}$$

[26] Despite its many drawbacks, the mean-variance model attracted much attention and led, e.g., to the development of the highly praised Capital Asset Pricing Model of portfolio optimization, cf. [Sha64]. For special classes of distributions (most commonly normal distributed rvs are assumed) the model is also even consistent with SSD, cf. [Big93].

[27] The term *expected shortfall* is also frequently used in a different meaning, cf. Remark 2.2.8.

[28] These constraints are closely related to *integrated chance constraints* (3.5), also cf. [KH86].

and *lower standard semideviation*

$$STD^{-}(\mathsf{X}) := \sqrt{\mathbb{E}([\mathbb{E}(\mathsf{X}) - \mathsf{X}]_{+}^{2})} = \sqrt{\int_{-\infty}^{\mathbb{E}(\mathsf{X})} (\mathbb{E}(\mathsf{X}) - t)^2 \, d\mathbb{P}_{\mathsf{X}}(t)}. \tag{2.13}$$

To elaborate on the desirable properties for risk functionals, axiomatic definitions were proposed in [ADEH99], where the notion of *coherence* was introduced. A coherent risk functional then complies with the following axioms:

(R1) Antimonotonicity: $\mathsf{X} \leq \mathsf{Y}$ a.s. implies that $\mathcal{R}(\mathsf{X}) \geq \mathcal{R}(\mathsf{Y})$;

(R2) Convexity: $\mathcal{R}(t\mathsf{X} + (1-t)\mathsf{Y}) \leq t\mathcal{R}(\mathsf{X}) + (1-t)\mathcal{R}(\mathsf{Y}) \, \forall t \in [0,1]$;

(R3) Translation antivariance: $\mathcal{R}(\mathsf{X} + a) = \mathcal{R}(\mathsf{X}) - a \, \forall a \in \mathbb{R}$;

(R4) Positive homogenity: $\mathcal{R}(t\mathsf{X}) = t\mathcal{R}(\mathsf{X}) \, \forall t \geq 0$.

A mirror image to risk functionals are *acceptability functionals* or *safety measures*, which assess the acceptability of the cdf instead of its riskiness. These functionals should comply with monotonicity ($\mathcal{A}1$), concavity ($\mathcal{A}2$) and translation equivariance ($\mathcal{A}3$) axioms, which are the counterparts to (R1)-(R3), cf. [RP07].[29] Moreover, if $\mathcal{A}$ is a positively homogeneous acceptability functional then $\mathcal{R}(\mathsf{X}) := -\mathcal{A}(\mathsf{X})$ will be a coherent risk functional. Of course, higher acceptability will imply higher preference of a decision maker.

In view of Proposition 2.2.1 (iii), the monotonicity axiom for acceptability functionals is equivalent to the following requirement of *isotonicity with FSD*

$$\mathsf{X} \preceq_{(1)} \mathsf{Y} \Longrightarrow \mathcal{A}(\mathsf{X}) \leq \mathcal{A}(\mathsf{Y}), \tag{2.14}$$

which once again underlines the importance of the order. Analogously, consistency of the corresponding mean-risk models with FSD has an axiomatic meaning for coherent risk functionals.

Two of the most important acceptability functionals have close relations to the so-called *dual* representations of the stochastic orders proposed in [OR02], which are characterized with the help of quantiles.[30]

[29]The axioms for acceptability functionals can be found in A.3.1.

[30]For the definition of a quantile we refer to A.1.

Let $F_{\mathsf{X}}^{-1} : [0,1] \to \bar{\mathbb{R}}$ denote the (left-continuous) *inverse distribution* or *first quantile function* of a distribution function F_{X}, defined as

$$F_{\mathsf{X}}^{-1}(p) := \inf\{t : F_{\mathsf{X}}(t) \geq p\} \text{ for } 0 < p \leq 1. \tag{2.15}$$

The infimum is attained for $0 < p < 1$ since cdfs are continuous from the right, and we can define the first acceptability functional as follows.

Definition 2.2.5. *The left α-quantile*

$$V@R_\alpha(\mathsf{X}) := F_{\mathsf{X}}^{-1}(\alpha) \tag{2.16}$$

is called value-at-risk at level α.

Though $\mathrm{V@R}_\alpha$ is not concave, it is widely used and very relevant in many decision models, cf. [RP07, pp. 57] and the references therein.[31] Directly from Proposition 2.2.1 (ii) we now get another characterization for FSD as a continuum of constraints on the $V@R_\alpha$ acceptability functional.

Proposition 2.2.6. *For random variables* X *and* Y *the following statements are equivalent:*

(i) $\mathsf{X} \preceq_{(1)} \mathsf{Y}$*;*

(ii) $V@R_\alpha(\mathsf{X}) \leq V@R_\alpha(\mathsf{Y})$ *for all* $\alpha \in]0,1]$.

The average of the left quantiles below α now gives another important acceptability functional with nice mathematical properties.[32]

Definition 2.2.7. *The tail value-at-risk at level α, written $TV@R_\alpha$, with* $0 < \alpha \leq 1$ *is defined as*

$$TV@R_\alpha(\mathsf{X}) := \frac{1}{\alpha}\int_0^\alpha F_{\mathsf{X}}^{-1}(t)\,dt. \tag{2.17}$$

[31] In the case of rvs representing losses, $\mathrm{V@R}_\alpha$ has a natural interpretation of a quantile risk measure, that we will discuss in Section 2.2.2.

[32] In fact, $-TV@R_\alpha$ defines a coherent risk functional, see Section 2.2.2.

Remark 2.2.8. *Definition 2.2.7 is due to [Ace02], where* $-TV@R_\alpha$ *was called* α*-Expected Shortfall.* $TV@R_\alpha$ *is also referred to as the average value-at-risk in [FS11], which is a major reference in financial mathematics. In the literature, different names are frequently used synonymously, which is, unfortunately, a constant source of confusion. To distinguish between the functionals, we will use the term* $TV@R_\alpha$ *as above (cf. [OR02]) and* $CV@R_\alpha$ *as in Definition 2.2.14 (cf. [Pfl00]).*

Interestingly, the *second quantile function*

$$F_{\mathsf{X}}^{(-2)}(p) := \int_0^p F_{\mathsf{X}}^{-1}(t)\,dt \text{ for } 0 < p \leq 1, \tag{2.18}$$

which we used for the definition of the $TV@R_\alpha$, is a Fenchel conjugate to the second performance function $F_{\mathsf{X}}^{(2)}$, cf. [OR02, Theorem 3.1].

Proposition 2.2.9. *For every rv* X *with* $\mathbb{E}(|\mathsf{X}|) < \infty$ *the following duality relations hold*

(i) $F_{\mathsf{X}}^{(-2)} = [F_{\mathsf{X}}^{(2)}]^\star$;

(ii) $F_{\mathsf{X}}^{(2)} = [F_{\mathsf{X}}^{(-2)}]^\star$.

Proof. For the proof and background on dual dominance relations the author refers to [OR02]. An excellent exposition of convex analysis can be found in [Roc97]. Some basic notions and results of convex analysis are provided in A.2. □

Since the conjugacy operation is order-reversing, cf. [BL06, p.49], Proposition 2.2.9 yields a dual representation of SSD as a continuum of constraints on the $TV@R_\alpha$ acceptability functional.

Proposition 2.2.10. *For random variables* X *and* Y *the following statements are equivalent:*

(i) $\mathsf{X} \preceq_{(2)} \mathsf{Y}$;

(ii) $TV@R_\alpha(\mathsf{X}) \leq TV@R_\alpha(\mathsf{Y})$ *for all* $\alpha \in]0,1]$.

From this proposition it is immediate that the mean-risk model with $-TV@R_\alpha$ as a risk functional is consistent with SSD.

2.2.2 Preference of Small Outcomes

For the decision maker minimizing losses, in complete analogy to the Definition 2.1.1, we could define the FSD relation for small outcomes ($\mathsf{X} \preceq_{\substack{FSD \\ small}} \mathsf{Y}$) through

$$\mathbb{E}u(\mathsf{X}) \leq \mathbb{E}u(\mathsf{Y}) \tag{2.19}$$

for all *nonincreasing* utility functions $u(\cdot)$. The same inequality would then hold for all nondecreasing functions $-u(\cdot)$, which is equivalent to changing sides in (2.19), thus yielding

$$\mathsf{X} \preceq_{\substack{FSD \\ small}} \mathsf{Y} \Longleftrightarrow \mathsf{Y} \preceq_{(1)} \mathsf{X}. \tag{2.20}$$

In view of this relation, we will stick to the original definition of FSD and just regard the dominated variable as the better one.

For SSD, the argumentation is analogous, however we will also have to change sides in Jensen's inequality (2.2), which expresses the risk aversion. In other words, the SSD relation for the minimization case, would mean the preference of a *dominated* variable in the *increasing convex order* (ICX), defined as follows.

Definition 2.2.11. *For rvs* $\mathsf{X}, \mathsf{Y} \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$*, we define* Y *to dominate* X *w.r.t. the increasing convex order, written* $\mathsf{X} \preceq_{icx} \mathsf{Y}$ *iff*

$$\mathbb{E}u(\mathsf{X}) \leq \mathbb{E}u(\mathsf{Y})\, \forall u \text{ nondecreasing and convex.} \tag{2.21}$$

Remark 2.2.12. *Usually, the notation* $\mathsf{X} \preceq \mathsf{Y}$ *automatically implies that* Y *should be the preferred variable in the* $\preceq$*-order. As we have seen above, for the minimization case that would imply introducing new ordering relations specially* *for the minimization case. Since both FSD and ICX have a long tradition in literature, after some discussion with the community, it was decided to prefer the dominated rvs in the established orders as a compromise. Since then, these notions became state of the art in the literature dealing with the SD relations in the minimization case.*

In analogy to Proposition 2.2.2, the following proposition presents a computationally more tractable representation for ICX.

Proposition 2.2.13. *For rvs* $\mathsf{X}, \mathsf{Y} \in \mathcal{L}_1(\Omega, \mathcal{F}, I\!P)$, *the following statements are equivalent:*

(i) $\mathsf{X} \preceq_{icx} \mathsf{Y}$*;*

(ii) $I\!E[\mathsf{X} - t]_+ \leq I\!E[\mathsf{Y} - t]_+$ *for all* $t \in \mathbb{R}$, *where* $[\alpha]_+ := \max(\alpha, 0)$ *is the positive part of* α.

Proof. See Theorem 4.A.2 in [SS07]. □

The function $H_{\mathsf{X}}(t) := I\!E[\mathsf{X} - t]_+ = \int_t^\infty \bar{F}_{\mathsf{X}}(\alpha)\, d\alpha$ is called the *integrated survival function* or *excess function* and is referred to as the *stop-loss transform* in actuarial sciences. Here, we will chose the interpretation of $I\!E[\mathsf{X} - t]_+$ as the *expected excess above some fixed target* t, which is the counterpart risk functional to $I\!E[t - \mathsf{X}]_+$ we used to characterize the SSD.

In fact, negative losses can be interpreted as gains and vice versa. Replacing X with $-\mathsf{X}$ in characterization (ii) of the above proposition yields

$$\mathsf{X} \preceq_{icx} \mathsf{Y} \Longleftrightarrow -\mathsf{Y} \preceq_{(2)} -\mathsf{X}, \tag{2.22}$$

which allows us to transfer all results from one order to the other. Clearly, for FSD we have $-\mathsf{X} \preceq_{(1)} -\mathsf{Y} \Longleftrightarrow \mathsf{Y} \preceq_{(1)} \mathsf{X}$.

With the risk functionals in the minimization case the situation is more diverse. For some functionals the switch of preference is reflected just in replacing X with $-\mathsf{X}$. E.g., the risk functional corresponding to $STD^-(\mathsf{X})$ is the *upper* standard semideviation, which measures the risk of losses *above* the mean. For this measure the following relation holds

$$STD^{+}(\mathsf{X}) := \sqrt{\mathbb{E}([\mathsf{X} - \mathbb{E}(\mathsf{X})]_{+}^{2})} = STD^{-}(-\mathsf{X}). \tag{2.23}$$

Other important risk functionals arise as counterparts to the acceptability functionals $V@R_\alpha$ and $TV@R_\alpha$. In fact, $\mathrm{V@R}_\alpha$ can be used here directly as a risk functional, because the value $V@R_\alpha(\mathsf{X}) = \inf\{t : \mathbb{P}(\mathsf{X} > t) \leq 1 - \alpha\}$ has a natural interpretation as the smallest loss, such that the probability of exceeding this loss lies below $1 - \alpha$. In other words, $\mathrm{V@R}_\alpha$ describes the minimum potential loss in the "$(1 - \alpha) \cdot 100\,\%$ worst cases".

This functional is extremely popular in finance, being, e.g., the standard in the Basel II accord, cf. [BAS06]. However, $\mathrm{V@R}_\alpha$ has undesirable mathematical and computational properties. Particularly, its failing convexity[33] may strongly discourage diversification of risks, cf. [FS11], [ADEH99] and the references therein.

The counterpart to the $TV@R_\alpha$ usually is called *conditional value-at-risk* ($CV@R_\alpha$), since the intention behind this functional was to consider the conditional expectation of losses in the "$(1 - \alpha) \cdot 100\,\%$ worst cases".

Definition 2.2.14. *The conditional value-at-risk at level α, written $CV@R_\alpha$, with $0 < \alpha \leq 1$ is defined as*

$$CV@R_\alpha(\mathsf{X}) := \frac{1}{1-\alpha} \int_{\alpha}^{1} F_{\mathsf{X}}^{-1}(t)\, dt. \tag{2.24}$$

For continuous distributions or, generally, in case α is in the range of F_{X}, it holds that

$$CV@R_\alpha(\mathsf{X}) = \mathbb{E}(\mathsf{X} | \mathsf{X} > F_{\mathsf{X}}^{-1}(\alpha)). \tag{2.25}$$

Since an η with $F_{\mathsf{X}}(\eta) = \alpha$ need not exist,[34] the above equality will not hold in

[33] Convexity stands here for the axiom (C2), which can be found, together with the other axioms of coherence for risk functionals in the minimization case, in Definition A.3.2.

[34] Which, in particular, may be the case for a discretely distributed rv with a probability atom at $\mathrm{V@R}_\alpha$.

general, and the correct definition of the $CV@R_\alpha$ is (2.24). The importance of the $CV@R_\alpha$ now is on the one hand due to the fact, that it is the most fundamental[35] coherent risk functional, cf. [Pfl00] and [AT02]. On the other hand, a computationally tractable representation of this measure is given through the minimization rule

$$CV@R_\alpha(\mathsf{X}) = \min\{a + \frac{1}{1-\alpha}\mathbb{E}(\mathsf{X} - a)_+ : a \in \mathbb{R}\}, \tag{2.26}$$

which was established in [RU00] and [RU02]. Thus, to obtain $CV@R_\alpha$ a continuous convex function has to be minimized, which opens a variety of possibilities to construct tractable stochastic programming models, cf. [ST04].

Also, the following relations to the $TV@R_\alpha$ can be derived

$$\mathbb{E}(\mathsf{X}) = \alpha \cdot TV@R_\alpha(\mathsf{X}) + (1-\alpha) \cdot CV@R_\alpha(\mathsf{X}) \tag{2.27}$$

and both functionals can be transformed into each other:

$$CV@R_\alpha(\mathsf{X}) = -TV@R_{1-\alpha}(-\mathsf{X}). \tag{2.28}$$

In fact, the last equality combined with Proposition 2.2.10 and the relation (2.22) implies the following dual characterization of ICX.

Proposition 2.2.15. *For rvs* $\mathsf{X}, \mathsf{Y} \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$*, the following statements are equivalent:*

(i) $\mathsf{X} \preceq_{icx} \mathsf{Y}$*;*

(ii) $CV@R_\alpha(\mathsf{X}) \leq CV@R_\alpha(\mathsf{Y})$ *for all* $\alpha \in]0,1]$*.*

In view of the relation (2.20), the dual characterization of FSD presented in Proposition 2.2.6 remains intact in the minimization case. Thus, FSD has a representation as a continuum of constraints on the $\mathrm{V@R}_\alpha$ and ICX as a continuum of constraints on the $CV@R_\alpha$.

[35] Many other coherent risk functionals can be represented through functions of the $CV@R_\alpha$ due to Choquet and Kusuoka representations, cf. [RP07, pp. 58].

The definitions of mean-risk dominance and its consistency with the stochastic orders are analogous to the maximization case and are presented in A.3. In view of the results above, it is immediately clear that the model with $\mathrm{V@R}_\alpha$ as a risk functional is consistent with FSD, and the model with $CV@R_\alpha$ even with ICX.

Chapter 3

Stochastic Orders and Contemporary Stochastic Programming

3.1 Stochastic Programming: Models

In the Introduction, we discussed various ways to compare random prospects. In context of SP, these prospects are now generally assumed to depend on a *policy* or *decision variable* $x \in \mathcal{X}$, so that the decision problem could consist in selecting the "best" rv out of the family

$$\{f(x, \omega) : x \in \mathcal{X}\}. \tag{3.1}$$

For the following it is further assumed that all underlying probability distributions are independent of the decisions x, and that these decisions have to be made *before* uncertainty is revealed. Thus, SP can be seen as optimization under *information constraints*, with the latter assumption also referred to as *nonanticipativity*.[1]

[1] While most SP models comply with nonanticipativity constraints, exceptions are discussed, e.g., in [GG06].

Since stochastic problems typically arise as generalizations of deterministic problems with random data, the concrete shape of the rvs $f(x, \omega)$ is usually not given explicitly. In the present thesis, we will restrict ourselves to *stochastic linear programs* which are the random counterparts to (mixed-integer) linear programs. Stochastic linear programs were pioneered in the famous papers of Dantzig [Dan55] and Beale [Bea55] and can be considered fundamental in contemporary SP.

To become more specific, let us consider the following "random" linear program as the starting point of our discussion:

$$\text{"}\min\text{"}\{c^\top(\omega)x : Ax = b, T(\omega)x = h(\omega),\ x \geq 0\}. \tag{3.2}$$

Here, x denotes a decision variable defined on a given polyhedron $\mathcal{X} := \{x : Ax = b,\ x \geq 0\}$.[2] Let the cost vector $c(\omega)$ associated with x, the constraint matrix $T(\omega)$ and the right-hand side vector $h(\omega)$ denote rvs defined on a probability space $(\Omega, \mathcal{F}, I\!P)$.

Clearly, under nonanticipativity constraints, i.e., if we have to make decisions *here-and-now*, the problem (3.2) gets ill-posed: it is neither specified for which ω the "random" constraints have to hold, nor what the meaning of the "min" should be.

Before we elaborate on these issues, let us recall that already in deterministic programming the notions of optimality and feasibility are closely interlinked, especially in the face of conflicting objectives. Since the attitude towards risk has to be additionally considered in a stochastic model, the situation in SP is bound to become even more intricate.

Basically, any absolute order could be used to address optimality, so for the moment we assume an appropriate order to be fixed and will come back to the treatment of optimality in Section 3.1.3. To define feasibility, maybe the most obvious way is to impose the random constraints to hold for ($I\!P$-almost) all ω, or to specify a set $\mathcal{D}$ of possible values for the random components over which

[2] From now on, we assume all matrices and vectors to be of consistent dimensions.

the constraints are to be fulfilled. Models of such a kind are called *robust*.[3]

Depending on the properties of the set $\mathcal{D}$, robust optimization programs can be reformulated and treated as linear or convex problems, cf. [BTN02], [BTBN06] and the references therein. Leading to "safe" decisions, the disadvantage of robust solutions is that they might be very expensive or not even exist at all.[4]

Therefore, some kind of infeasibility is usually allowed in stochastic programs. This infeasibility could be treated in the objective, e.g., the probability of constraint satisfaction may be maximized or penalty costs for infeasibility could be specified. However, the first technique has close relations to problems with *probabilistic constraints*, cf. [Pré95, Chapter 10], and the second to problems with *recourse*, cf. [KH86, Remark 3.6]), which are the two classical modeling techniques of SP. A schematic overview of SP models can be found in Figure 3.1.

Without aiming at any completeness, we will now introduce both techniques to pave the way for problems with SD-constraints, which constitute the main part of the present thesis.

3.1.1 Probabilistic Constraints

For stochastic systems subject to high uncertainty and where reliability is a central issue, so-called *probabilistic* or *chance constraints* pioneered by Charnes, Cooper and Symonds, cf. [CCS58], might be the right choice to handle infeasibility.

In this approach, for a prespecified (usually large) probability level p we obtain the well-defined *joint probabilistic constraints* through

$$\mathbb{P}(\{\omega : T(\omega)x \geq h(\omega)\}) \geq p \tag{3.3}$$

[3]In fact, for robust programs we make an assumption on the range of the rvs instead of the exact shapes of their distributions.

[4]That is why such solutions are also called "worst-case" or "fat", cf. [KW94], [BL97].

Stochastic Optimization Models	
Objective	**Constraints**
Absolute Orders	Robust
Risk-neutral:	Probabilistic
-Expectation	ICC
Risk-averse:	Risk measures
-Risk measures	Recourse
-Mean-risk	-two-stage
-Utility functions	-multistage
Partial Orders	SD-Constraints
SD (multi-criteria)	
Mean-risk dominance (bi-criteria)	

Underlying Optimization Problem
Linear
Nonlinear (Convex/Non-Convex)
Mixed-Integer Linear/Nonlinear
Finite Dimensional/Infinite Dimensional

Figure 3.1: Overview of SP Models

and, analogously, the simpler *individual probabilistic constraints* through

$$\mathbb{P}(\{\omega : T_{i\bullet}(\omega)x \geq h_i(\omega)\}) \geq p_i, \; i = 1, \ldots, m, \tag{3.4}$$

where $T_{i\bullet}$ denotes the i-th row of the matrix T.[5] Due to their intuitive interpretation as a safety requirement, probabilistic constraints are appealing to practitioners and widely used in applications, see, e.g., [HLM+01], [TL99] and [Pré03, pp. 338] for a brief overview.

Unfortunately, especially the practically more interesting problems with joint probabilistic constraints possess rather bad mathematical properties, since their feasible regions may become non-convex and even disconnected,

[5]For chance constraints, the inequality form of the corresponding constraints is assumed.

in general. Only under special requirements on the underlying probability distributions convexity can be guaranteed.[6] Once convexity is obtained, the problems can be efficiently solved directly or incorporated into larger optimization models as a part of constraints – a view we will adopt in the present thesis.

As it was pointed out in [KH86], probabilistic constraints are based upon the *qualitative risk concept*, which accounts only for the probability of an infeasibility but not for its extent. Clearly, in some situations placing an upper bound on the *amount* of risk could be a plausible alternative. This idea led to the development of so-called *integrated chance constraints* (ICC) in [KH86].

By analogy with (3.4), *individual integrated chance constraints* are defined through

$$\mathbb{E}([h_i(\omega) - T_{i\bullet}(\omega)x]_+) = \int_{-\infty}^{0} \mathbb{P}\{w : T_{i\bullet}(\omega)x - h_i(\omega) < t\}\; dt \leq \beta_i, \qquad (3.5)$$

where β_i is the prespecified risk aversion parameter.[7] Apparently, ICC are of the same form as constraints on the expected shortfall. It is also important to mention that problems with ICC are tractable computationally due to the intrinsic convexity of their feasible regions.[8]

3.1.2 Recourse Problems

The second classical modeling approach of SP is completely based on the *quantitative risk concept*. This technique is appropriate if we may assume that infeasibility can be corrected *after* the realization of the random outcomes. To this end, the model is extended with a so-called *second stage*, where recourse actions can be taken to correct the infeasibility.

[6] A well-known convexity result was achieved for a nonrandom matrix T and a quasi-concave distribution of $h(\omega)$. For more general results, we refer to the seminal paper of Prékopa [Pré71] and to the textbooks [SDR09], [Pré03].

[7] The representation with the integral explains the "*integrated*" as part of the name. A joint version of ICC was introduced in [KH86] as well.

[8] For the joint ICC case, convexity also persists, cf. [KH86].

Thus, the information constraints are modified in such a way that the here-and-now or *first-stage* decisions x are followed by the *wait-and-see* or *second-stage* decisions $y = y(x, \omega)$, carried out after the randomness is observed. These decisions are governed by the *fixed recourse*[9] matrix W and penalized with *recourse costs* $q(\omega)$ to enter the objective function. Then, we get the two-stage information scheme of alternating decision and observation depicted in Figure 3.2. This scheme can be naturally extended to a multistage framework involving sequential decisions for situations where randomness is subsequently revealed over time, cf. [SDR09], [Pré95].

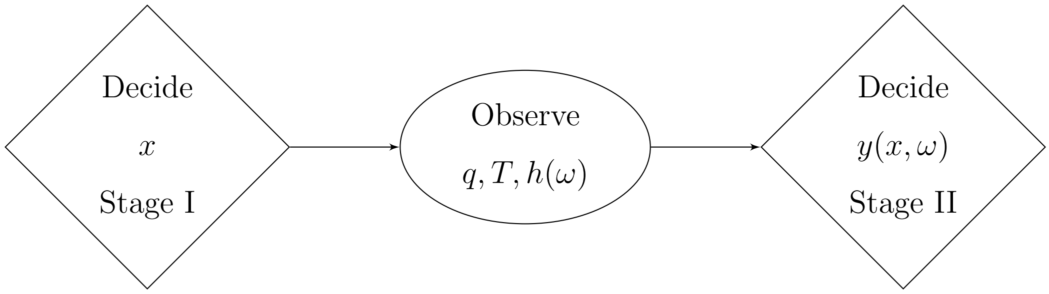

Figure 3.2: Information Constraints

In the present work, we will restrict ourselves to the two-stage approach, which yields the formulation

$$\text{"}\min\text{"} \left\{ \begin{array}{ll} c(\omega)^\top x + q(\omega)^\top y : & T(\omega)x + Wy = h(\omega) \\ & x \in \mathcal{X},\, y \in \mathcal{Y} \end{array} \right\}, \tag{3.6}$$

where $\mathcal{X}, \mathcal{Y}$ are polyhedral sets possibly with integer requirements. The constraints of this model are now posed for ($\mathbb{P}$-almost) all ω and hence are well-defined. The second-stage decisions y are solutions of a (parametric) linear program (LP) with the following value function

$$\Phi(x, \omega) := \min_{y \in \mathcal{Y}} \{q(\omega)^\top y : Wy = h(\omega) - T(\omega)x\}. \tag{3.7}$$

The study of this function and its properties plays the key role in the theory of problems with recourse. The most widely studied problem in this class is

[9]The randomness of the recourse matrix may lead to extreme numerical instability for discrete distributions, cf. [RS03, p. 80] and [BL97, pp. 109] for a more general view.

obtained if we minimize the expectation of the overall, i.e., of first- and second-stage, costs. For this purpose, we define the *expected recourse function*

$$Q_{\mathbb{E}}(x) := \mathbb{E}_\omega \Phi(x, \omega), \tag{3.8}$$

and get a so-called *deterministic equivalent* formulation[10] of the expectation based stochastic program as

$$\min_{x \in \mathcal{X}} \{c^\top x + Q_{\mathbb{E}}(x)\}.^{11} \tag{3.9}$$

Clearly, the program is well-posed, and if the function $Q_{\mathbb{E}}$ was given, it would translate into a deterministic nonlinear program. In fact, if the second-stage problem (3.7) is solvable for at least one x, the function $\Phi(\cdot, \omega)$ is convex and even piecewise linear in the pure linear case, cf. [SDR09, Proposition 2.1]. Under certain assumptions, convexity results can also be transferred to the expected recourse function $Q_{\mathbb{E}}$, cf. [SDR09, Propositions 2.3, 2.7], which enables an efficient algorithmic treatment of the model (3.9) and hence contributes to the popularity of the mean-based approach.

3.1.3 Risk Aversion and Dominance Constraints

In what follows, we assume the two-stage framework with recourse presented above to be the modeling technique of choice, i.e., probabilistic constraints are assumed to be already included as linear constraints in the polyhedron $\mathcal{X}$ containing all first-stage restrictions. Thus, it is assumed that the decisions we make will ensure feasibility of our model with a sufficient reliability level and in exchange for certain costs.

Staying in this framework, we now return to the treatment of optimality. To this end, we represent the overall cost for each first-stage decision x as a rv

[10] The term *deterministic equivalent* implies that all symbols of rvs are eliminated in the formulation. We will stick to this term though it may be misleading, because stochastic problems are deterministic regardless of their formulation, as was argued in [Pré95, p. 234].

[11] For ease of exposition, we assume from this point onwards $c(\omega) = c$ to be deterministic. Due to linearity of the expectation, here we could simply take the average of the values.

$$f(x,\omega) := c^\top x + \Phi(x,\omega), \tag{3.10}$$

with $\Phi(x,\omega)$ as in (3.7), and thus obtain the family (3.1) of rvs with the specific structure. Selecting the "best" rv out of this family through taking the expectation of the costs, as it was done in the formulation (3.9), rests upon a risk neutral decision model and hence may lead to first-stage decisions incurring ruinous costs for unfavorable random outcomes.

A popular way to consider risk aversion consists in the application of a mean-risk optimization model that here takes the shape

$$\min_{x\in\mathcal{X}} \mathbb{E}(f(x,\omega)) + \lambda\rho(f(x,\omega)) \tag{3.11}$$

and was studied in [Ahm06], [Tie05] and [ST04] in more detail. If the risk functional ρ is λ-consistent with SD, a strict minimum of this model will be non-dominated in SD in view of (A.4) and (A.6). Being computationally attractive, this approach is thus justified on grounds of the decision theory.

However, a drawback of mean-risk models is the need to specify the risk aversion parameter λ directly or to employ a sensitivity analysis on this parameter.[12] To avoid this difficulty, SD-consistent mean-risk dominance, see (A.3) and (A.5), may be used instead. As a partial order, it yields the multi-criteria, here bi-criteria, optimization problem

$$\min_{x\in\mathcal{X}} \{\mathbb{E}(f(x,\omega)), \rho(f(x,\omega))\} \tag{3.12}$$

with conflicting objectives. Such a problem is not likely to possess a solution that simultaneously optimizes each of the objectives. Hence, the set of *Pareto optimal* solutions, called the *efficient frontier*, can be considered.[13] For (3.12), a solution $\bar{x} \in \mathcal{X}$ is Pareto optimal if there is no other x with $f(x,\omega) \prec_{\mathbb{E},\rho} f(\bar{x},\omega)$. Thus, these solutions will also be non-dominated w.r.t. SD.

[12]Because the risk measure itself is typically not given naturally, families of other risk measures may have to be considered as well.

[13]For an introduction to multiobjective optimization we refer to [Ste89] and [Ehr06].

It is also possible to look for solutions preferable w.r.t. SD directly. Resulting optimization problems then possess a continuum of objectives. Though for finite, discrete distributions the number of objectives can be reduced to a finite number, cf. [Ogr02] and [RDDM06], such models remain relatively hard to solve. By contrast, if the utility function of the decision maker is given explicitly, its optimization yields a computationally attractive risk-averse technique referred to as the *Bernoulli principle*, cf. [Pré95, pp. 221]. Unfortunately, this method is not universally applicable, as we have argued in Section 2.1.

In the present thesis, we will now concentrate on an alternative way to incorporate risk aversion through shifting its treatment to the constraints. This trick is well-known from deterministic programming with conflicting objectives, all but one of which can be transformed into *goal restrictions*.[14]

Here, this idea implies imposing constraints on the risk, thus defining decisions with "acceptable" risk as feasible solutions. To ensure such "economic" feasibility, probabilistic constraints and also ICC or constraints on risk measures could be used, once corresponding data on probability and risk thresholds is available. Instead, in some practical situations a *reference* random outcome Y - a so-called *benchmark* - is available. We will concentrate on such situations and look for decisions producing outcomes preferable to the benchmark.

The seminal model for this type of problems was proposed by Dentcheva and Ruszczyński in [DR03], [DR04a]. Inspired by applications in portfolio optimization where benchmarks naturally arise from stock indexes like [SP], the authors employed stochastic orders to characterize the preferable outcomes.

Thus, the following *dominance-constrained* model was obtained:

$$\max\{g(\mathsf{X}) : \mathsf{Y} \preceq_{(i)} \mathsf{X}, \mathsf{X} \in \mathcal{C}\}, \qquad (3.13)$$

where Y was the benchmark rv, $\mathcal{C}$ a convex and closed set and g a concave

[14] An intermediate approach is followed in the so-called *goal programming* where violations of the goal restrictions are additionally penalized in the objective, cf. [CCF55]. However, this approach does not seem applicable here.

continuous functional. For dominance constraints of first and second order the feasible regions of this problem were shown to be closed, and for SSD even convex, cf. [DR03]. Moreover, under rather weak assumptions[15] it was shown in [DR04b] that problems with SSD constraints are convexifications of problems with first order constraints (which are not convex in general). For stability and sensitivity analysis of these problems the author refers to [DHR07] and [DR13].

In the framework of EUT, the dominance-constrained model (3.13) guarantees the preference of the solution X by all decision makers with utility functions in the corresponding class $\mathcal{U}_i$, see Definition 2.1.1. In particular, no risk-averse decision maker will strictly prefer the benchmark outcome over a feasible solution to the second order model.

Conversely, it was shown in [DR04b] and [DR03] that utility functions of EUT can be identified with Lagrange multipliers associated with the dominance constraints of (3.13). Moreover, dual representations of SD from Propositions 2.2.6 and 2.2.10 possess Lagrange multipliers that can be identified with dual utility functions in the sense of RDEUT, cf. [DR05]. In this way, problems with SD-constraints can be regarded as dual to EUT and RDEUT, thus providing a link between both theories.

Links to SD-consistent mean-risk and mean-risk dominance models exist, as we have already discussed above, because their optimal values yield feasible solutions for dominance-constrained problems. Due to the dual representations of SD, dominance constraints can be also viewed as continua of constraints on important risk measures. By Proposition 2.2.1 (ii), first order dominance constraints are nothing else than a continuum of probabilistic constraints. In view of Proposition 2.2.2 (ii) and (3.5), SSD constraints can in turn be viewed as a continuum of ICC.

Recent applications of dominance-constrained models include, e.g., finan-

[15] E.g., for a discrete distribution of Y with equiprobable outcomes or if its distribution is continuous, cf. [DR04b].

cial optimization [DR06] and power generation capacity expansion problems [VBZE13]. For recent interesting theoretical developments we refer to contributions on multivariate SD [DR09, HdMM10, AL10], and on dynamic optimization models with dominance constraints [DR08, HJ13].

Finally, we introduce the problem class we will work with starting from Chapter 4. It results from the specialization of problem (3.13) for the cost minimization framework, where the decision dependent random outcome is given with $f(x, \omega)$ from (3.10). A typical problem of the relevant class then takes the shape

$$\min\{g(x) \;:\; f(x, \omega) \preceq_{(i)} d(\omega),\; x \in \mathcal{X}\} \tag{3.14}$$

where $\preceq_{(i)}$ refers to the orders FSD and ICX, cf. (4.1). Thus, model feasibility is assured through recourse actions, whereas the "economic" feasibility of a decision is defined w.r.t. a cost benchmark $d(\omega)$ by means of the SD-rules. In this way, the model combines one of the most popular feasibility modeling techniques of SP with the decision theoretical benefits of SD.

Models of the form (3.14) were proposed in [GNS08, GGS07]. The special case with a linear second stage was considered in [DS10]. Applications were carried out, e.g., in energy trading [CGS09] and in operation planning of virtual power plants [DGG$^+$11].

Building up on these works, in the following we will develop appropriate solution techniques for this interesting but demanding problem class. Since our tailored methods are based on the standard methods of SP, we first outline their main ideas in the following section.

3.2 Stochastic Programming: Methods

3.2.1 Deterministic equivalents

A simple way to approach a two-stage stochastic program is to tackle its deterministic equivalent formulation directly. For the mean-based problem, the deterministic equivalent (3.9) can be written down as the linear program

$$\min \left\{ c^\top x + \sum_{\ell=1}^{L} \pi_\ell q^\top y_\ell : \begin{array}{l} Ax = b,\ x \geq 0, \\ \\ Tx + W y_\ell = h_\ell,\ y_\ell \geq 0,\ \ell = 1, \ldots, L \end{array} \right\} \tag{3.15}$$

for a finite discrete distribution of the right-hand side with probabilities $\pi_1, \ldots, \pi_L$ for the scenarios $h_1, \ldots, h_L$. While we assume the recourse matrix W to be fixed for numerical reasons, cf. [RS03, p. 80], the matrix T and the cost vector q will be presented as fixed only for the ease of exposition. Also the equality form of the second-stage constraints and the restrictions on first and second-stage variables are presented exemplary.[16]

To formulate risk-averse models as (mixed-integer) linear programs, auxiliary (sometimes binary) variables and a so-called *Big M*, which is a large number associated with these variables, can be used. For example, under certain assumptions, cf. [Tie05], there exists a constant $M > 0$, such that the pure risk problem $\min\{\mathrm{V@R}_\alpha\,(\mathrm{x})\colon x \in \mathcal{X}\}$ can be equivalently restated as

$$\min \left\{ \eta : \begin{array}{lcll} c^\top x + q^\top y_\ell - M\theta_\ell & \leq & \eta, & \forall \ell \\ Tx + W y_\ell & = & h_\ell, & \forall \ell \\ \sum_{\ell=1}^{L} \pi_\ell \theta_\ell & \leq & 1 - \alpha, & \\ x \in \mathcal{X},\ \eta \in \mathbb{R},\ y_\ell \geq 0,\ \theta_\ell \in \{0,1\} & & & \forall \ell \end{array} \right\}. \tag{3.16}$$

With the help of the minimization rule (2.26), an analogous representation can be gained for the problem of $\mathrm{CV@R}_\alpha$ minimization:

$$\min \left\{ \eta + \frac{1}{1-\alpha} \sum_{\ell=1}^{L} \pi_\ell \theta_\ell : \begin{array}{lcll} c^\top x + q^\top y_\ell - \theta_\ell & \leq & \eta, & \forall \ell \\ Tx + W y_\ell & = & h_\ell, & \forall \ell \\ x \in \mathcal{X},\ \eta \in \mathbb{R},\ y_\ell \geq 0,\ \theta_\ell \in \mathbb{R}_+ & & & \forall \ell \end{array} \right\}, \tag{3.17}$$

cf. [Tie05] for more details on such equivalents and other risk measures. Clearly, MILP equivalents of mean-risk models directly arise as combinations of the mean-based problem (3.15) with a corresponding pure risk problem.

[16] More generally, the variables are assumed to be contained in nonempty closed convex polyhedra, which arise as solution sets to systems of linear inequalities (later also involving certain integer requirements).

In view of the close relations between $\mathrm{V@R}_\alpha$ and FSD and $\mathrm{CV@R}_\alpha$ and ICX, it is not surprising that formulations (3.16) and (3.17) have similarities with the MILP equivalents

$$\min \left\{ g^\top x : \begin{array}{lcll} c^\top x + q^\top y_{\ell k} - M\theta_{\ell k} & \leq & d_k & \forall \ell \; \forall k \\ Tx + Wy_{\ell k} & = & h_\ell & \forall \ell \; \forall k \\ \sum\limits_{\ell=1}^{L} \pi_\ell \theta_{\ell k} & \leq & \sum\limits_{j=k+1}^{K} p_j & \forall k \\ x \in \mathcal{X},\; y_{\ell k} \geq 0,\; \theta_{\ell k} \in \{0,1\} & & & \forall \ell \; \forall k \end{array} \right\} \tag{3.18}$$

and

$$\min \left\{ g^\top x : \begin{array}{lcll} c^\top x + q^\top y_{\ell k} - \theta_{\ell k} & \leq & d_k & \forall \ell \; \forall k \\ Tx + Wy_{\ell k} & = & h_\ell & \forall \ell \; \forall k \\ \sum\limits_{\ell=1}^{L} \pi_\ell \theta_{\ell k} & \leq & \sum\limits_{j=k+1}^{K} p_j (d_j - d_k) & \forall k \\ x \in \mathcal{X},\; y_{\ell k} \geq 0,\; \theta_{\ell k} \geq 0 & & & \forall \ell \; \forall k \end{array} \right\}, \tag{3.19}$$

which were derived in [GNS08] and [GGS11] for the dominance-constrained problems

$$\min\{g^\top x : f(x,\omega) \preceq_{(1)} d(\omega), x \in \mathcal{X}\} \tag{3.20}$$

and

$$\min\{g^\top x : f(x,\omega) \preceq_{icx} d(\omega), x \in \mathcal{X}\}. \tag{3.21}$$

Here, the objective function $g(x)$ is assumed to be linear, and the distribution of the benchmark rv $d(\omega)$ is finite, discrete with scenarios $d_1, \ldots, d_K$ and probabilities $p_1, \ldots, p_K$. We will discuss these so-called *lifting-representations*[17] in Chapter 4 in more detail also presenting some possible improvements and modifications. In Section 5.2, then the so-called *polyhedral representation* for the problem with ICX will be discussed. This representation avoids any auxiliary variables and goes back to a representation obtained in [KH86] for ICC.

Thus, a solution technique for risk-averse stochastic programs may consist in applying readily available efficient solvers, like [CPL13] and [GUR13], to

[17]This term was proposed by Fábián in [Fáb12] to reflect the usage of auxiliary variables which lift the problem dimension.

corresponding MILP equivalents. The drawback of this approach is that such equivalents grow in size quickly with the number of data scenarios. Here, decomposition methods exploiting specific structures of stochastic programs may go into action.

3.2.2 Primal Decomposition Methods

To explain the first group of so-called *primal decomposition methods*, let us go back to the mean-based problem. For this problem, the idea behind the primal methods consists in tackling the formulation (3.9) using the structural properties of the value functions

$$Q_\ell(x) := \min_{y \geq 0} \{q^\top y : Wy = h_\ell - Tx\}, \tag{3.22}$$

and of the expected recourse function $Q_{\mathbb{E}}(x) = \sum_{\ell=1}^{L} \pi_\ell Q_\ell(x)$, which are, as we have already mentioned, convex and piecewise linear under certain assumptions. These properties enable us to construct suitable approximations for the functions Q_ℓ and $Q_{\mathbb{E}}$ by solving the *subproblems* (3.22) for the current first stage solution x^k. These approximations are then used in a so-called *master problem* to calculate the next iterate x^{k+1}.

While primal methods mainly differ in the way how master problems are constructed and solved, in the present thesis we will employ the idea of *cutting plane methods*, which we now sketch below. To focus on the points essential here, we assume both $Q_\ell(x) > -\infty$ and $Q_\ell(x) < +\infty$ for all $x \in \mathcal{X}$.[18] Alternatively, first stage solutions yielding infeasible subproblems could be cut off algorithmically with the so-called *feasibility cuts*. For a discussion of these cuts and for an overview of other primal methods, we refer the interested reader to [RS03, Chapter 3].

To design a cutting plane method, let us consider the LP-dual problem to

[18] These assumptions are well-known in SP and will be discussed later, cf. assumptions (A3) and (A4) of Chapter 5.

the subproblem (3.22)

$$\max\{u^\top(h_\ell - Tx) : W^\top u \leq q\}, \tag{3.23}$$

which is solvable due to the assumed finiteness of $Q_\ell(x)$. With $\Delta := \{u : W^\top u \leq q\}$ denoting the feasible domain of (3.23), the subdifferential of the function $Q_\ell(x)$ at any fixed point $\bar{x}$ can be expressed as

$$\partial Q_\ell(\bar{x}) = -T^\top \operatorname*{arg\,max}_{u\in\Delta} u^\top(h_\ell - T\bar{x}), \tag{3.24}$$

cf. [RS03, pp. 72]. Let $\bar{u}_{\ell,k}$ denote an optimal solution to the subproblem (3.23) for the current first stage iterate x^k. From the definition of the subdifferential we then get a lower approximation of $Q_\ell(x)$ over $\mathcal{X}$ with the inequality

$$Q_\ell(x) \geq Q_\ell(x^k) - \bar{u}_{\ell,k}^\top T(x - x^k) = \bar{u}_{\ell,k}^\top(h_\ell - Tx) \quad \forall x \in \mathcal{X}. \tag{3.25}$$

This approximation leads to *objective cuts*, which are gradually included into the master problems

$$\min\left\{\begin{array}{ll} c^\top x + \sum\limits_{\ell=1}^{L}\pi_\ell\theta_\ell : & x \in \mathcal{X},\, \theta_\ell \in \mathbb{R}\,, \ell = 1,\ldots,L, \\ & \qquad\qquad \bar{u}_\ell^\top(h_\ell - Tx) \leq \theta_\ell \quad \forall \bar{u}_\ell \in \Delta_\ell^k \end{array}\right\} \tag{3.26}$$

of the cutting plane method. Here, Δ_ℓ^k denotes the set containing the optimal solutions to the subproblems for each scenario ℓ which we have computed in the previous k iterations of the algorithm. This so-called *disaggregate form* of the cutting plane method was developed by Dantzig and Mandansky, cf. [DM61], as a dual method to Dantzig-Wolfe decomposition, cf. [DW60].

The cutting plane method in the *aggregate form* operates with the master problems of the form

$$\min\left\{\begin{array}{ll} c^\top x + \theta : & x \in \mathcal{X},\, \theta \in \mathbb{R}, \\ & \sum\limits_{\ell=1}^{L}\pi_\ell\bar{u}_\ell^\top(h_\ell - Tx) \leq \theta \quad \forall(\bar{u}_1,\ldots,\bar{u}_L) \in \bar{\Delta}^k \end{array}\right\}, \tag{3.27}$$

where $\bar{\Delta}^k \subset \Delta_1^k \times \cdots \times \Delta_L^k$. In this form, the method is usually referred to as Benders decomposition, cf. [Ben63]. It was also proposed by Van Slyke and

Wets in [VSW69] under the name of the *L-shaped method*, due to the shape of the block structure in the constraints of the formulation (3.15). For a more detailed exposition of the cutting plane methods we refer to [BL97], [KW94], [RS03].

Analogously, cutting plane methods can be applied to convex risk-averse problems, like (3.17), cf. [Ahm06], [KBM06]. In this context, it is interesting to note that not only for the mean-based problems lifting-representations have structural similarities with master problems of disaggregate cutting plane methods, whereas polyhedral representations and aggregate cutting plane methods get along well, cf. [KBM06] and [Fáb12].

In this way, representations and methods of both kinds that we will develop and discuss for the dominance-constrained models in Chapters 5 and 6 fit into the general framework of the cutting plane methodology.

3.2.3 Dual Decomposition Methods

In Chapter 7, we will come across two *dual decomposition methods* proposed in [GNS08] and [GGS07] especially for the dominance-constrained problems. To explain the functioning of these methods, let us once again go back to the mean-based problem, but now in its formulation (3.15) as an LP.

As we just mentioned, this formulation has a special L-shaped block structure, which is especially amenable for decomposition, in its constraints. In fact, the blocks $Tx + Wy_\ell = h_\ell$ are not interconnected, i.e., for two scenarios $\ell \neq \ell'$ the corresponding second-stage variables do not appear together in one constraint. Hence, only the implicit nonanticipativity conditions - which require the first-stage decisions x to be chosen independently of the scenarios - prevent the mean-based problem from decomposing directly into scenario-specific subproblems.

To make this nonanticipativity requirement explicit, we can split the first stage variable x into copies $x_1, \ldots, x_L$ and add the constraints $x_1 = \ldots = x_L$

to the formulation (3.15), which then reads

$$\min \left\{ \begin{array}{lll} \sum_{\ell=1}^{L} \pi_\ell (c^\top x_\ell + q^\top y_\ell): & A x_\ell = b,\ x_\ell \geq 0, & \forall \ell \\ & T x_\ell + W y_\ell = h_\ell,\ y_\ell \geq 0, & \forall \ell \\ & x_1 = \ldots = x_L & \end{array} \right\}. \tag{3.28}$$

The idea behind dual decomposition now consists in the application of Lagrangean relaxation to the last group of constraints. To facilitate notation, we rewrite these constraints as $\sum_{\ell=1}^{L} H_\ell x_\ell = 0$, with suitable matrices H_ℓ. The Lagrangean function of (3.28) is then given by[19]

$$L(x, y, \lambda) := \sum_{\ell=1}^{L} \pi_\ell (c^\top x_\ell + q^\top y_\ell + \lambda^\top H_\ell x_\ell) \tag{3.29}$$

with the associated dual function

$$D(\lambda) := \min \left\{ \begin{array}{lll} L(x, y, \lambda): & A x_\ell = b,\ x_\ell \geq 0 & \forall \ell \\ & T x_\ell + W y_\ell = h_\ell,\ y_\ell \geq 0 & \forall \ell \end{array} \right\}. \tag{3.30}$$

For an arbitrary λ the value $D(\lambda)$ provides a lower bound for (3.15). To compute this bound, the optimization problem (3.30) - which now has the desired structure with independent blocks - can be solved by decomposition into the scenario subproblems

$$D_\ell(\lambda) := \min \left\{ \begin{array}{lll} c^\top x_\ell + q^\top y_\ell + \lambda^\top H_\ell x_\ell: & A x_\ell = b,\ x_\ell \geq 0 & \forall \ell \\ & T x_\ell + W y_\ell = h_\ell,\ y_\ell \geq 0 & \forall \ell \end{array} \right\}, \tag{3.31}$$

since it holds that

$$D(\lambda) = \sum_{\ell=1}^{L} \pi_\ell D_\ell(\lambda). \tag{3.32}$$

By the duality theory, cf. [RS03, pp. 119], the problem

$$\max D(\lambda) \tag{3.33}$$

is the dual problem to (3.28), which implies that their optimal values are equal unless both problems are infeasible. For a dual optimal solution $\bar{\lambda}$ a primal

[19]The factors π_ℓ in front of $\lambda^\top H_\ell x_\ell$ preserve the expectation like nature of the objective and may prevent the ill-conditioning of the Lagrangean dual, cf. [LS03, p. 234].

optimal solution can be obtained as a solution to the scenario subproblems (3.31).

To solve the dual problem (3.33), which is a non-smooth concave maximization problem with a piecewise linear objective, appropriate nonlinear methods like, e.g, bundle methods (cf. [HK02]), can be applied. Subgradients for these methods applied to the corresponding convex minimization problem can be calculated from the subproblems as well. In fact, for the subdifferential of $-D(\cdot)$ at a fixed $\bar{\lambda}$ it holds

$$\partial(-D(\bar{\lambda})) = -\left(\sum_{\ell=1}^{L} H_\ell \arg\min_{x_\ell} D_\ell(\bar{\lambda})\right), \tag{3.34}$$

cf. [RS03, pp. 190].[20]

Analogously, dual decomposition methods can be applied to other risk-averse problems with a similar block structure in the constraints, cf. [ST04]. However, some risk-averse problems may have additional scenario-coupling in the constraints. For example, the pure risk problem (3.16) - unlike problem (3.17) - has a scenario dependent constraint $\sum_{\ell=1}^{L} \pi_\ell \theta_\ell \leq 1 - \alpha$, which hampers decomposition. Unfortunately, both SD-constrained problems possess similar groups of coupling constraints, cf. third group of constraints in (3.18) and (3.19). One of the remedies proposed in [GNS08, GGS11] then consists in expanding the Lagrangean relaxation to these restrictions as well.

Further, dual decomposition becomes especially interesting once integer requirements are present in both the first and the second stage (as it is the case in [GNS08]). In fact, an integer first stage would not cause much trouble for the primal methods: the subproblems (3.22) remain linear, duality is preserved, and hence the convexity properties of the recourse functions needed for the construction of the cuts remain intact. With an integer second stage, however, both convexity and duality, which are essential for the primal methods, cannot

[20] In this reference, the properties of the function $-D(\cdot)$ are derived from the fact that it is the Fenchel conjugate of the expected recourse function. The formula for the subdifferential then results from the Fenchel-Moreau Theorem, cf. [RS03, Theorem 52].

be maintained for reasonable problem classes, cf. [LS03, pp. 215].

At this, the advantage of the dual methods is based on the fact that the Lagrangean dual problem (3.33) retains its structural properties in the integer case, so that it still could be tackled by the bundle methods. Of course, an optimal solution to the dual problem would not necessarily yield a primal optimal solution in this case. However, the procedure described in this subsection represents an important option to generate adequate lower bounds. These bounds can be used together with feasibility heuristics in a branch-and-bound scheme. This approach, which is also referred to as *scenario decomposition*, is due to [CS99].

Chapter 4

Lifting-Representations of Dominance Constraints with Mixed-Integer Recourse

In the present chapter, we consider different lifting-representations for the dominance-constrained problems (3.20) and (3.21) from a unified point-of-view.[1] To this end, we discuss and enhance MILP equivalents (3.18), (3.19) proposed in [GGS11, GNS08], and adapt model formulations from [Lue07, Lue08] to our framework.

Putting together all model ingredients introduced in Section 3.1, we may write down the problems of interest as

$$\min\{g^\top x : f(x, h(\omega)) \preceq_{(1)} d(\omega), x \in \mathcal{X}\} \tag{4.1}$$

and

$$\min\{g^\top x : f(x, h(\omega)) \preceq_{icx} d(\omega), x \in \mathcal{X}\} \tag{4.2}$$

where

$$f(x, h(\omega)) := c^\top x + \Phi(h(\omega) - Tx) \tag{4.3}$$

[1]The results of the present chapter were originally published by the author in [DKS11].

with the second-stage value function rewritten as

$$\Phi(t) := \min\{q^\top y \ :\ Wy \geq t,\ y \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}\} \qquad \text{for all } t \in \mathbb{R}^s.$$

Here, second-stage variables are allowed to possess integer components, which is why the models are said to have *mixed-integer recourse*. From now on, we again assume W to be fixed, while T and q are only depicted as fixed for the ease of exposition.

So far, we have always assumed the solvability of the optimization problem behind Φ, and did not bother about its possible infeasibility or unboundedness. To ensure that the entities

$$f(x, h_\ell) \ = \ c^\top x + \Phi(h_\ell - Tx)$$

are well-defined real numbers for all $\ell = 1, \ldots, L$, $x \in \mathcal{X}$, we now make the solvability requirements explicit. To this end, we pose the assumptions

(A1) (complete mixed-integer-recourse)

for any $t \in \mathbb{R}^s$ there exists a $y \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}$ such that $Wy \geq t$,

(A2) (dual feasibility of the linear programming relaxation)

$\{u \in \mathbb{R}^s \ :\ W^\top u \leq q,\ u \geq 0\} \neq \emptyset$,

which are well-known in SP, see [Sch93], and ensure that $\Phi(t)$ is a well-defined real number for any $t \in \mathbb{R}^s$.

Further, we assume that the objective $g^\top x$ is linear, the polyhedron $\mathcal{X}$ nonempty, possibly with integer requirements on components of x. Both the random right-hand side $h(\omega)$ and the benchmark $d(\omega)$ follow discrete distributions with realizations $h_\ell,\ \ell = 1, \ldots, L$, and $d_k,\ k = 1, \ldots, K$, as well as probabilities $\pi_\ell,\ \ell = 1, \ldots, L$, and $p_k,\ k = 1, \ldots, K$, respectively.

Out of all different representations we presented in Section 2.2 to characterize the SD relations, in the following we will use Proposition 2.2.1 (iv) for FSD and Proposition 2.2.13 for ICX.

For a discrete and finite benchmark Y with realizations Y_k, $k = 1, \ldots, K$ it is possible to reduce the continua of constraints in both characterizations to a finite number as follows:

$$\mathsf{X} \preceq_{(1)} \mathsf{Y} \quad \text{iff} \quad I\!\!P\,[\mathsf{X} > \mathsf{Y}_k] \;\leq\; I\!\!P\,[\mathsf{Y} > \mathsf{Y}_k]\,, \quad k = 1, \ldots, K, \tag{4.4}$$

$$\mathsf{X} \preceq_{icx} \mathsf{Y} \quad \text{iff} \quad I\!\!E[\mathsf{X} - \mathsf{Y}_k]_+ \;\leq\; I\!\!E[\mathsf{Y} - \mathsf{Y}_k]_+, \quad k = 1, \ldots, K, \tag{4.5}$$

see [GNS08, Got09] for the proofs.

MILP equivalents to (4.1) and (4.2) are now obtained by representing the feasible sets

$$C_1 := \{x \in \mathcal{X} : f(x, h) \preceq_{(1)} d\} \tag{4.6}$$

and

$$C_2 := \{x \in \mathcal{X} : f(x, h) \preceq_{icx} d\} \tag{4.7}$$

as polyhedra in lifted dimensions, possibly involving integer requirements to some of the variables.

4.1 Lifting-Representations of Type I

For convenience we denote the probability measures induced by the random variables $h(\omega)$ and $d(\omega)$ on $\mathbb{R}^s$ and $\mathbb{R}$ by μ and ν, respectively. Without loss of generality, we assume that the realizations of the benchmark, d_k, are arranged in ascending order, i.e.,

$$d_1 < d_2 < \ldots < d_K. \tag{4.8}$$

Now, let us have a closer look at the feasibility region C_1 starting from the following observation.

Observation 4.1.1. *The function values $f(x, h_\ell), \ell = 1, \ldots, L$, $x \in C_1$, are bounded above by the biggest realization d_K of the benchmark. The reason is that, for $k = K$, inequality (4.4) becomes $\mu[f(x, h) > d_K] = 0$.*

Adjusting the steps of the proof of Proposition 3.1 in [GNS08] in the light of Observation 4.1.1, we obtain the following equivalent representation for C_1.

Proposition 4.1.2. *For finite discrete distributions of $h(\omega)$ and $d(\omega)$ and under the assumptions (A1) and (A2) it holds that*

$$C_1 = \left\{ x \in \mathcal{X} : \begin{array}{lcll} \exists y_\ell \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'} & & & \\ \exists \theta_{\ell k} \in \{0,1\} & & & \\ c^\top x + q^\top y_\ell & \leq & d_k + \theta_{\ell k}(d_K - d_k) & \forall \ell \; \forall k \\ Tx + W y_\ell & \geq & h_\ell & \forall \ell \\ \sum_{\ell=1}^{L} \pi_\ell \theta_{\ell k} & \leq & \sum_{j=k+1}^{K} p_j & \forall k \end{array} \right\}. \tag{4.9}$$

Proof: Employing (4.4) and calculating explicitly the probabilities $\nu[d > d_k]$, provides

$$\begin{aligned} C_1 &= \{x \in \mathcal{X} \; : \; \mu[f(x,h) > d_k] \; \leq \; \nu[d > d_k], \quad k = 1, \ldots, K\} \\ &= \left\{ x \in \mathcal{X} \; : \; \mu[f(x,h) > d_k] \leq \sum_{j=k+1}^{K} p_j, \quad k = 1, \ldots, K \right\}. \end{aligned}$$

(Throughout, we make the convention that $\sum_{j=K+1}^{K}$ corresponds to summing over the empty set and yielding zero.)

Now, for

$$\mu[f(x,h) > d_k] \leq \sum_{j=k+1}^{K} p_j, \qquad k = 1, \ldots, K, \tag{4.10}$$

to hold, it is sufficient to find $y_{\ell k} \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}$ and $\theta_{\ell k} \in \{0,1\}$, $k = 1, \ldots, K$, $\ell = 1, \ldots, L$, such that

$$Tx + W y_{\ell k} \;\; \geq \;\; h_\ell, \;\; \ell = 1, \ldots, L, \tag{4.11}$$

$$c^\top x + q^\top y_{\ell k} \;\; \leq \;\; d_k + \theta_{\ell k}(d_K - d_k), \;\; \ell = 1, \ldots, L, \tag{4.12}$$

and

$$\sum_{\ell=1}^{L} \pi_\ell \theta_{\ell k} \; \leq \; \sum_{j=k+1}^{K} p_j. \tag{4.13}$$

Indeed, if (4.11), (4.12) can be fulfilled with $\theta_{\ell k} = 0$, then $f(x, h_\ell) \leq d_k$. Hence, $f(x, h_\ell) > d_k$ is possible for $\theta_{\ell k} = 1$, only. This implies

$$\mu[f(x,h) > d_k] \; \leq \; \sum_{\ell=1}^{L} \pi_\ell \theta_{\ell k}.$$

On the other hand, if (4.10) is valid, then setting $\theta_{\ell k} = 1$ if and only if $f(x, h_\ell) > d_k$ and choosing $y_{\ell k} \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}, \ell = 1, \ldots, L$, as optimal solutions to

$$\min\{q^\top y \;:\; Wy \geq h_\ell - Tx,\; y \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}\},\; \ell = 1, \ldots, L,$$

together with Observation 4.1.1, demonstrate that (4.11) - (4.13) can be fulfilled. The following Observation 4.1.3 from [Nei08] completes the proof. □

Observation 4.1.3. *The vectors $y_{\ell k} \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}, \ell = 1, \ldots, L$, from the above proof can be selected independently on k. In the necessity part the selection has already been made this way. In the sufficiency part, passing for each $\ell = 1, \ldots, L$ to $y_\ell := \arg\min\{q^\top y_{\ell k} \;:\; k = 1, \ldots, K\}$ yields the desired.*

Now, let us turn our attention to the feasibility region C_2 starting from a result similar to Observation 4.1.1.

Observation 4.1.4. *The function values $f(x, h_\ell), \ell = 1, \ldots, L$, $x \in C_2$, are bounded above by d_K, because, for $k = K$, the expectation on the right in (4.5) becomes zero, so that $\mathbb{E}\left[f(x,h) - d_K\right]_+ = \sum_{\ell=1}^{L} \pi_\ell \left[f(x,h_\ell) - d_K\right]_+ = 0$ yielding $\left[f(x,h_\ell) - d_K\right]_+ = 0$ for $\ell = 1, \ldots, L$.*

Adjusting the steps of the proof of Proposition 3.1 in [GGS11] according to Observation 4.1.4, we obtain the following equivalent representation for C_2.

Proposition 4.1.5. *For finite discrete distributions of $h(\omega)$ and $d(\omega)$ and under the assumptions (A1) and (A2) it holds that*

$$C_2 = \left\{ x \in \mathcal{X} : \begin{array}{lcll} \exists y_\ell \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'} & & & \\ \exists \theta_{\ell k} \in [0,1] & & & \\ c^\top x + q^\top y_\ell & \leq & d_k + \theta_{\ell k}(d_K - d_k) & \forall \ell\ \forall k \\ Tx + Wy_\ell & \geq & h_i & \forall \ell \\ \sum_{\ell=1}^{L} \pi_\ell \theta_{\ell k}(d_K - d_k) & \leq & \sum_{j=k+1}^{K} p_j(d_j - d_k) & \forall k \end{array} \right\}. \tag{4.14}$$

Proof: By (4.5) we have

$$C_2 \;=\; \left\{ x \in \mathcal{X} \;:\; \sum_{\ell=1}^{L} \pi_\ell \left[f(x,h_\ell) - d_k \right]_+ \le \sum_{j=k+1}^{K} p_j(d_j - d_k), \quad k = 1,\dots,K \right\}.$$

Then, for

$$\sum_{\ell=1}^{L} \pi_\ell \left[f(x,h_\ell) - d_k \right]_+ \le \sum_{j=k+1}^{K} p_j(d_j - d_k) \tag{4.15}$$

to hold, it is sufficient to find $y_{\ell k} \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}$ and $\theta_{\ell k} \ge 0, \ell = 1,\dots,L$, such that

$$Tx + Wy_{\ell k} \;\ge\; h_\ell, \;\; \ell = 1,\dots,L, \tag{4.16}$$

$$c^\top x + q^\top y_{\ell k} \;\le\; d_k + \theta_{\ell k}, \;\; \ell = 1,\dots,L, \tag{4.17}$$

and

$$\sum_{\ell=1}^{L} \pi_\ell \theta_{\ell k} \;\le\; \sum_{j=k+1}^{K} p_j(d_j - d_k). \tag{4.18}$$

Indeed, (4.16) and (4.17) imply

$$f(x,h_\ell) - d_k \;\le\; \theta_{\ell k},$$

and, together with (4.18),

$$\sum_{\ell=1}^{L} \pi_\ell [f(x,h_\ell) - d_k]_+ \le \sum_{\ell=1}^{L} \pi_\ell \theta_{\ell k} \;\le\; \sum_{j=k+1}^{K} p_j(d_j - d_k).$$

If, vice versa, (4.15) is valid, then picking $y_{\ell k} \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}$ as optimal solutions to

$$\min\{q^\top y \;:\; Wy \ge h_\ell - Tx,\; y \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}\},\; \ell = 1,\dots,L,$$

and setting

$$\theta_{\ell k} = [f(x,h_\ell) - d_k]_+,$$

for $\ell = 1,\dots,L$, fulfills (4.16) - (4.18).

In view of Observation 4.1.4, there cannot occur arbitrarily large $\theta_{\ell k} \ge 0$ in relations (4.17), (4.18), although this were formally feasible. With new variables $\theta_{\ell k}$ for which $0 \leq \theta_{\ell k} \leq 1$, the relations can be expressed as follows:

$$\begin{aligned} c^\top x + q^\top y_{\ell k} &\leq d_k + \theta_{\ell k}(d_K - d_k), \quad \ell = 1, \dots, L, \\ \sum_{\ell=1}^{L} \pi_\ell \theta_{\ell k}(d_K - d_k) &\leq \sum_{j=k+1}^{K} p_j(d_j - d_k), \quad k = 1, \dots, K. \end{aligned}$$

For the same reasons as in Observation 4.1.3 the vectors $y_{\ell k} \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}, \ell = 1, \dots, L$, can be selected independently on k. This argument completes the proof. $\square$

Remark 4.1.6. *The representations of C_1 and C_2 given in Propositions 4.1.2 and 4.1.5 slightly improve earlier results in [GNS08] and [GGS11]. In [GNS08] a big-M technique is employed which is often problematic for practical and numerical reasons, whereas Propositions 4.1.2 now works with the improved upper bounds $d_K - d_k$. In both [GNS08] and [GGS11], the y-variables still depend on the benchmark scenarios k. This makes the model dimensions bigger, while matrix fill-in is bigger in the present models.*

Remark 4.1.7. *The closedness of the sets C_1 and C_2 is apparent from their MILP equivalent formulations. In a more general context, this was shown in [GNS08, Proposition 2.1] and [GGS11, Proposition 2.2]. Once the set $\mathcal{X}$ is bounded and $C_1, C_2 \neq \emptyset$, the optimization problems (4.1) and (4.2) become well-defined, i.e, the corresponding infima are finite and attained. While the boundedness of $\mathcal{X}$ was required in [GNS08, Proposition 2.1] for the calculation of big - M, here it may assure the solvability of our optimization problems.*

4.2 Lifting-Representations of Type II

Both the representations (4.9) and (4.14) rely on variables $\theta_{\ell k}$ indicating whether $f(x, h_\ell)$ exceeds the benchmark d_k. This is the crucial factor for launching the model dimension as well as the number of constraints into the order of $K \cdot L$.

However, since the benchmark scenarios are ascending, for each data scenario ℓ it holds that $f(x,h_\ell) \le d_k$ implies $f(x,h_\ell) \le d_{k'}$ for all $k' > k$.

Therefore, one could introduce Boolean variables $\theta_{\ell k}$ to indicate the change from $f(x,h_\ell) \le d_k$ to $f(x,h_\ell) > d_k$ which, inspired by [Lue08], now leads us to an alternative mixed-integer linear formulation of C_1.

Proposition 4.2.1. *(adapted from [Lue08]) For finite discrete distributions of $h(\omega)$ and $d(\omega)$ and under the assumptions (A1) and (A2) it holds that*

$$C_1 = \left\{ x \in \mathcal{X} : \begin{array}{lcll} \exists y_\ell \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'} & & & \\ \exists \theta_{\ell k} \in \{0,1\} & & & \\ c^\top x + q^\top y_\ell & \le & \sum_{k=1}^{K} \theta_{\ell k} d_k & \forall \ell \\ Tx + W y_\ell & \ge & h_\ell & \forall \ell \\ \sum_{\ell=1}^{L} \pi_\ell \sum_{j=k+1}^{K} \theta_{\ell j} & \le & \sum_{j=k+1}^{K} p_j & \forall k \\ \sum_{k=1}^{K} \theta_{\ell k} & = & 1 & \forall \ell \end{array} \right\}. \tag{4.19}$$

Proof: Let $x \in \mathcal{X}$. Then for (4.10) to hold it is sufficient to find $y_\ell \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}$, $\theta_{\ell k} \in \{0,1\}$ $(\forall\, \ell, k)$ such that

$$c^\top x + q^\top y_\ell \;\le\; \sum_{k=1}^{K} \theta_{\ell k} d_k \quad \forall\, \ell, \tag{4.20}$$

$$Tx + W y_\ell \;\ge\; h_\ell \quad \forall\, \ell, \tag{4.21}$$

$$\sum_{k=1}^{K} \theta_{\ell k} \;=\; 1 \quad \forall\, \ell \tag{4.22}$$

and

$$\sum_{\ell=1}^{L} \pi_\ell \sum_{j=k+1}^{K} \theta_{\ell j} \;\le\; \sum_{j=k+1}^{K} p_j \quad \forall\, k. \tag{4.23}$$

This can be seen as follows: In view of (4.22), for each $\ell \in \{1,\dots,L\}$ there exists $k^* = k^*(\ell)$ such that $\theta_{\ell k^*} = 1$. By (4.20) and (4.21) it follows that $f(x,h_\ell) \le d_{k^*}$. Since the d_k are in ascending order, the relation $f(x,h_\ell) > d_k$ can be valid for $k < k^*$, only. For $k < k^*$ it holds that

$$\sum_{j=k+1}^{K} \theta_{\ell j} = 1.$$

Therefore,

$$\{l \,:\, f(x,h_\ell) > d_k\} \;\subseteq\; \left\{l \,:\, \sum_{j=k+1}^{K} \theta_{\ell j} = 1\right\}$$

implying

$$\mu[f(x,h) > d_k] \;\leq\; \sum_{\ell=1}^{L} \pi_\ell \sum_{j=k+1}^{K} \theta_{\ell j} \;\leq\; \sum_{j=k+1}^{K} p_j \quad \forall\, k.$$

For necessity, let $x \in \mathcal{X}$ fulfill (4.10) for all k. The variables $y_\ell \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}$ again are taken as optimal solutions to

$$\min\{q^\top y \,:\, Wy \geq h_\ell - Tx,\ y \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}\}, \quad \ell = 1, \ldots, L.$$

For $\ell \in \{1, \ldots, L\}$ let $k^* = k^*(\ell)$ be the minimal index k with $f(x,h_\ell) \leq d_k$. Put $\theta_{\ell k} = 1$ if $k = k^*(\ell)$ and $\theta_{\ell k} = 0$, otherwise. It is easily seen that (4.20)-(4.22) are valid with these choices. To check (4.23), observe that

$$\sum_{j=k+1}^{K} \theta_{\ell j} = 1 \qquad \text{iff} \qquad k \,<\, k^*(\ell) \qquad \text{iff} \qquad f(x,h_\ell) \,>\, d_k.$$

Hence, for all k,

$$\sum_{\ell=1}^{L} \pi_\ell \sum_{j=k+1}^{K} \theta_{\ell j} \;=\; \mu[f(x,h) > d_k] \;\leq\; \sum_{j=k+1}^{K} p_j.$$

□

Following [Lue07], with the help of Observation 4.1.4 and (4.8) the information about the quantities $[f(x,h_\ell) - d_k]_+$ can be stored in the auxiliary variables $\theta_{\ell k}$ more efficiently than in the representation (4.14). This leads to another equivalent representation of C_2.

Proposition 4.2.2. *(adapted from [Lue07][2]) For finite discrete distributions*

[2] In [Lue08], Luedtke also proposed another proof - based on Strassen's theorem ([MS02, pp. 25]) - for the original result. To preserve the analogy to our previous proofs, we followed the more constructive version of the proof given in [Lue07].

of $h(\omega)$ and $d(\omega)$ and under the assumptions (A1) and (A2) it holds that

$$C_2 = \left\{ x \in \mathcal{X} : \begin{array}{llll} \exists y_\ell \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'} & & & \\ \exists \theta_{\ell k} \in [0,1] & & & \\ c^\top x + q^\top y_\ell & \le & \sum_{k=1}^{K} \theta_{\ell k} d_k & \forall \ell \\ Tx + Wy_\ell & \ge & h_\ell & \forall \ell \\ \sum_{\ell=1}^{L} \pi_\ell \sum_{j=k+1}^{K} \theta_{\ell j}(d_j - d_k) & \le & \sum_{j=k+1}^{K} p_j(d_j - d_k) & \forall k \\ \sum_{k=1}^{K} \theta_{\ell k} & = & 1 & \forall \ell \end{array} \right\}. \tag{4.24}$$

Proof: For $x \in \mathcal{X}$ to satisfy

$$\sum_{\ell=1}^{L} \pi_\ell \left[f(x, h_\ell) - d_k \right]_+ \le \sum_{j=k+1}^{K} p_j(d_j - d_k) \ \ \forall k \tag{4.25}$$

it is sufficient to find $y_\ell \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}$, $\theta_{\ell k} \in [0,1]$ $(\forall\, \ell, k)$ such that

$$c^\top x + q^\top y_\ell \ \le \ \sum_{j=1}^{K} \theta_{\ell j} d_j \quad \forall\, \ell, \tag{4.26}$$

$$Tx + Wy_\ell \ \ge \ h_\ell \quad \forall\, \ell, \tag{4.27}$$

$$\sum_{k=1}^{K} \theta_{\ell k} \ = \ 1 \quad \forall\, \ell \tag{4.28}$$

and

$$\sum_{\ell=1}^{L} \pi_\ell \sum_{j=k+1}^{K} \theta_{\ell j}(d_j - d_k) \ \le \ \sum_{j=k+1}^{K} p_j(d_j - d_k) \quad \forall\, k. \tag{4.29}$$

Indeed, (4.26) and (4.27) imply

$$f(x, h_\ell) \ \le \ \sum_{j=1}^{K} \theta_{\ell j} d_j,$$

and, together with (4.28) and (4.8),

$$f(x, h_\ell) - d_k \ \le \ \sum_{j=1}^{K} \theta_{\ell j}(d_j - d_k) \ \le \ \sum_{j=k+1}^{K} \theta_{\ell j}(d_j - d_k).$$

Then, (4.29) delivers the desired

$$\sum_{\ell=1}^{L} \pi_\ell [f(x,h_\ell) - d_k]_+ \le \sum_{\ell=1}^{L} \pi_\ell \sum_{j=k+1}^{K} \theta_{\ell j}(d_j - d_k) \ \le \ \sum_{j=k+1}^{K} p_j(d_j - d_k).$$

If, vice versa, (4.25) is valid, the variables $y_\ell \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}$ are once again taken as optimal solutions to

$$\min\{q^\top y \ : \ Wy \ge h_\ell - Tx, \ y \in \mathbb{Z}_+^{\bar{m}} \times \mathbb{R}_+^{m'}\}, \ \ell = 1, \dots, L.$$

For the selection of $\theta_{\ell k}$ we first consider the case of $f(x,h_\ell) \le d_1$. Here, we set $\theta_{\ell 1} = 1, \ \theta_{\ell k} = 0 \ \forall \ k > 1$, which fulfills (4.26)-(4.29). Otherwise, in view of Observation 4.1.4, we may select an index $k^* = k^*(\ell)$ with $1 < k^* \le K$ such that

$$d_{k^*-1} < f(x,h_\ell) \le d_{k^*}. \tag{4.30}$$

Now, we define $\theta_{\ell k} = 0 \ \forall \ k < k^* - 1, \ k > k^*$, and

$$\theta_{\ell k^*-1} := \frac{d_{k^*} - f(x,h_\ell)}{d_{k^*} - d_{k^*-1}}, \quad \theta_{\ell k^*} := \frac{f(x,h_\ell) - d_{k^*-1}}{d_{k^*} - d_{k^*-1}}. \tag{4.31}$$

Clearly, $\theta_{\ell k} \ge 0$ and fulfill (4.28). The constraint (4.26) is fulfilled since

$$\sum_{j=1}^{K} \theta_{\ell j} d_j = \frac{d_{k^*-1}(d_{k^*} - f(x,h_\ell)) + d_{k^*}(f(x,h_\ell) - d_{k^*-1})}{d_{k^*} - d_{k^*-1}} = f(x,h_\ell).$$

For the remaining constraint (4.29), we first compute that

$$\begin{aligned} k \ge k^* : \qquad & \sum_{j=k+1}^{K} \theta_{\ell j}(d_j - d_k) = 0 = [f(x,h_\ell) - d_k]_+, \\ k = k^* - 1 : \quad & \sum_{j=k+1}^{K} \theta_{\ell j}(d_j - d_k) = \theta_{\ell k^*}(d_{k^*} - d_{k^*-1}) = [f(x,h_\ell) - d_k]_+, \end{aligned}$$

hold. After a lengthier computation we see that also

$$\begin{aligned} k < k^* - 1 : \quad \sum_{j=k+1}^{K} \theta_{\ell j}(d_j - d_k) \ &= \theta_{\ell k^*-1}(d_{k^*-1} - d_k) + \theta_{\ell k^*}(d_{k^*} - d_k) \\ &= [f(x,h_\ell) - d_k]_+ \end{aligned}$$

holds. Together with (4.25) this leads to

$$\sum_{\ell=1}^{L} \pi_\ell \sum_{j=k+1}^{K} \theta_{\ell j}(d_j - d_k) = \sum_{\ell=1}^{L} \pi_\ell \left[f(x, h_\ell) - d_k\right]_+ \leq \sum_{j=k+1}^{K} p_j(d_j - d_k) \quad \forall\, k$$

completing the proof. □

Observation 4.2.3. *While the model dimension in (4.19), (4.24) still is in the order of $K \cdot L$, compared with (4.9), (4.14), the number of constraints now is in the order of $K + L$.*

Remark 4.2.4. *For the Propositions 4.1.2, 4.1.5, 4.2.1 and 4.2.2 to hold we have posed the assumptions (A1) and (A2), which are not used in the proofs explicitly. Indeed, if (A1) were missing for some $t = h_\ell - Tx$, then one would associate $\Phi(t) = +\infty$ yielding x neither in C_1, C_2 nor in their corresponding MILP equivalents. If (A2) were violated, then $\Phi(t) = -\infty$, and the dominance constraints would hold trivially.*

The MILP formulations discussed so far, in principle, can be tackled by readily available mixed-integer linear programming solvers such as the open source solver GLPK [GLP12] or one of the commercial solvers such as CPLEX [CPL13], GUROBI [GUR13] or Xpress [XPR13], just to mention a few. However, with the growth of the model size these MILP formulations quickly become large-scale. Especially, models with numbers of constraints in the order of $K \cdot L$ are problematic in this respect. The formulations inspired by [Lue07, Lue08] with constraint numbers in the order of $K + L$ provide some relief, but, as will be seen later on, come to their limits where decomposition methods still perform quite decently.

Chapter 5

Linear Recourse: Model Equivalents Tailored for Decomposition

In the MILP equivalents (4.9), (4.14), (4.19), and (4.24), the presence of the second-stage variables $y_\ell, \ell = 1, \ldots, L$, and their constraints is a critical factor for the model sizes being huge and quickly leading to intractability by general-purpose MILP solvers. As we have pointed out in Section 3.2.3, one possible remedy, still working for mixed-integer y_ℓ, is dual decomposition by means of Lagrangian relaxation combined with branch-and-bound. For dominance-constrained problems with mixed-integer recourse, this approach has been proven successful in [GNS08, GGS11].

If the second-stage variables y_ℓ all are continuous, i.e., if

$$\Phi(t) := \min\{q^\top y \ : \ Wy \geq t, \ y \geq 0\}, \tag{5.1}$$

also referred to as *linear recourse*, then linear programming duality allows for considerable shortcuts in both the MILP equivalents and algorithms for their solution. A prominent class of such algorithms, namely cutting plane or Benders' decomposition methods, was discussed in Section 3.2.2 for the case of expectation based problems (3.9) with linear recourse.

In the following, we will adapt the idea of these cutting plane methods to the dominance-constrained problems (3.20), (3.21). To this end, in the present chapter we derive the MILP equivalent formulations whose relaxations will reappear as master problems in the cutting plane methods proposed consecutively in Chapter 6.

To avoid technical difficulties, we assume

(A3) (complete recourse) For any $t \in \mathbb{R}^s$ there exists a $y \geq 0$ such that $Wy \geq t$;

(A4) (dual feasibility) $\{u \in \mathbb{R}^s \ : \ W^\top u \leq q, \, u \geq 0\} \neq \emptyset$;

which are the counterpart assumptions to (A1), (A2) and imply that $\Phi(t)$ is a well-defined real number for any $t \in \mathbb{R}^s$, see [BL97, KW94, Pré95, RS03].

5.1 Equivalent Formulations Based on the Lifting-Representations

Assuming linear recourse and (A3), (A4), we now derive MILP equivalents to the first- and second-order dominance constraints specified in C_1 and C_2. In particular, we study linear-recourse versions of (4.9), (4.14), (4.19), and (4.24). This refines and extends our earlier work in [DS10].[1] Here, refinement concerns (4.9) where the "big-M" is replaced by tighter bounds. Extension arises from including second-order dominance as in (4.14), (4.24) and from incorporating the modeling techniques of [Lue08], as in (4.19), (4.24). The following lemma from [DS10], providing a characterization of $f(x,h) \leq \eta$ ($\eta \in \mathbb{R}$), will have a pivotal role. To be self-contained concerning this detail and for convenient use later on, we include the short proof of the lemma.

Lemma 5.1.1. *Let* $(\delta_i, \delta_{io}) \in \mathbb{R}^{s+1}$, $i = 1, \ldots, I$, *denote the vertices of*

$$\Delta \ := \ \left\{(u, u_o) \in \mathbb{R}^{s+1} \ : \ 0 \leq u \leq \mathbf{1}, \ 0 \leq u_o \leq 1, \ W^\top u - u_o q \leq 0\right\}$$

[1] The results of the present section were originally published by the author in [DKS11].

with $\mathbf{1} \in \mathbb{R}^s$ *denoting the vector of all ones. Then*

$$f(x,h) \;=\; c^\top x + \Phi(h - Tx) \;\;\le\;\; \eta$$

if and only if

$$(h - Tx)^\top \delta_i \;+\; (c^\top x - \eta)\delta_{io} \;\;\le\;\; 0 \qquad \text{for all } i = 1, \ldots, I. \tag{5.2}$$

Proof: Since Δ is a nonempty and bounded polyhedron, it has vertices. By the definition of Φ in (5.1), the relation $c^\top x + \Phi(h - Tx) \le \eta$ is equivalent to claiming that the feasibility problem

$$\min_{y,\tau,\tau_o} \left\{ \mathbf{1}^\top \tau + \tau_o \;:\; Wy + \tau \;\ge\; h - Tx,\; c^\top x + q^\top y - \tau_o \;\le\; \eta,\; y \ge 0,\; \tau \ge 0,\; \tau_o \ge 0 \right\}$$

has optimal value zero. This problem is always solvable, and, by linear programming duality, its optimal value coincides with the optimal value of

$$\max_{u,u_o} \left\{ (h - Tx)^\top u + (c^\top x - \eta)u_o \;:\; 0 \le u \le \mathbf{1},\; 0 \le u_o \le 1,\; W^\top u - u_o q \le 0 \right\}. \tag{5.3}$$

Since Δ is the constraint set of the above linear program, the proof is complete. $\square$

With this result, the formulations (4.9) and (4.19) for the problems with FSD arise in the following shape.

Proposition 5.1.2. *Assume (A3), (A4), and let $h(\omega), d(\omega)$ follow the finite discrete probability distributions introduced at the beginning of Chapter 4. Then the following are equivalent:*

The linear-recourse first-order dominance model

$$\min\{g^\top x \;:\; f(x, h(\omega)) \preceq_{(1)} d(\omega),\; x \in \mathcal{X}\} \tag{5.4}$$

- *and the mixed-integer linear program*

$$\min \left\{ g^\top x : \begin{array}{llll} (h_\ell - Tx)^\top \delta_i + & & & \\ (c^\top x - d_k - \theta_{\ell k}(d_K - d_k))\delta_{i0} & \le & 0 & \forall \ell, k, i \\ \sum\limits_{\ell=1}^{L} \pi_\ell \theta_{\ell k} & \le & \sum\limits_{j=k+1}^{K} p_j & \forall k \\ x \in \mathcal{X},\; \theta_{\ell k} \in \{0,1\} & & & \forall \ell, k \end{array} \right\} \tag{5.5}$$

- *and the mixed-integer linear program*

$$\min \left\{ g^\top x : \begin{array}{lcll} (h_\ell - Tx)^\top \delta_i + (c^\top x - \sum_{k=1}^{K} \theta_{\ell k} d_k)\delta_{i0} & \leq & 0 & \forall \ell, i \\ \sum_{\ell=1}^{L} \pi_\ell \sum_{j=k+1}^{K} \theta_{\ell j} & \leq & \sum_{j=k+1}^{K} p_j & \forall k \\ \sum_{k=1}^{K} \theta_{\ell k} & = & 1 & \forall \ell \\ x \in \mathcal{X},\ \theta_{\ell k} \in \{0,1\} & & & \forall \ell, k \end{array} \right\}. \tag{5.6}$$

Proof: The proof of the first claim is given by confirming that the constraint set of (5.5) coincides with the set of all $x \in \mathcal{X}$ fulfilling (4.10):

$$\mu[f(x,h) > d_k] \leq \sum_{j=k+1}^{K} p_j, \quad k = 1, \ldots, K.$$

For (4.10) to hold it is sufficient to find $\theta_{\ell k} \in \{0,1\}$, $k = 1, \ldots, K$, $\ell = 1, \ldots, L$, such that

$$(h_\ell - Tx)^\top \delta_i + (c^\top x - d_k - \theta_{\ell k}(d_K - d_k))\delta_{i0} \ \leq\ 0 \quad \forall \ell, k, i \tag{5.7}$$

and

$$\sum_{\ell=1}^{L} \pi_\ell \theta_{\ell k} \ \leq\ \sum_{j=k+1}^{K} p_j \qquad \forall k.$$

Indeed, if the above relations can be fulfilled with $\theta_{\ell k} = 0$, then, by Lemma 5.1.1, it holds that $f(x, h_\ell) \leq d_k$. Hence,

$$\mu[f(x,h) > d_k] \ \leq\ \sum_{\ell=1}^{L} \pi_\ell \theta_{\ell k}.$$

For necessity, set $\theta_{\ell k} = 1$ if and only if $f(x, h_\ell) > d_k$. Then, for any k,

$$\sum_{\ell=1}^{L} \pi_\ell \theta_{\ell k} \ =\ \mu[f(x,h) > d_k] \ \leq\ \sum_{j=k+1}^{K} p_j.$$

Moreover, for $\theta_{\ell k} = 0$ we have $f(x, h_\ell) \leq d_k$, and Lemma 5.1.1 provides the validity of (5.7). For $\theta_{\ell k} = 1$, the relevant relations in (5.7) yield $f(x, h_\ell) \leq d_K$

which is valid by Observation 4.1.1.

To prove the second claim we first confirm that, for (4.10) to hold, it is sufficient to find $\theta_{\ell k} \in \{0,1\}$, $\ell = 1,\ldots,L$, such that

$$\begin{aligned}
(h_\ell - Tx)^\top \delta_i + \left(c^\top x - \sum_{k=1}^{K} \theta_{\ell k} d_k \right) \delta_{i0} \quad &\leq \quad 0 \qquad \forall \ell, i; \qquad (5.8)\\
\sum_{\ell=1}^{L} \pi_\ell \sum_{j=k+1}^{K} \theta_{\ell j} \quad &\leq \quad \sum_{j=k+1}^{K} p_j \quad \forall k; \qquad (5.9)\\
\sum_{k=1}^{K} \theta_{\ell k} \quad &= \quad 1 \qquad \forall \ell. \qquad (5.10)
\end{aligned}$$

Indeed, we conclude from (5.8) and (5.10) that for each $\ell = 1,\ldots,L$, there exists an index $k^*(\ell)$ with $f(x,h_\ell) \leq d_{k^*(\ell)}$. Since the d_k are ascending, $f(x,h_\ell) > d_k$ is possible for $k < k^*(\ell)$, only. For $k < k^*(\ell)$, however, $\sum_{j=k+1}^{K} \theta_{\ell j} = 1$. Therefore,

$$\mu[f(x,h) > d_k] \;=\; \sum_{\ell=1}^{L} \pi_\ell \sum_{j=k+1}^{K} \theta_{\ell j} \;\leq\; \sum_{j=k+1}^{K} p_j.$$

For the necessity part put $\theta_{\ell k} = 1$ if and only if, given ℓ, k equals the minimal index $k^*(\ell)$ with $f(x,h_\ell) \leq d_{k^*}$. □

For the second-order problems, Lemma 5.1.1 brings about the following reformulations of (4.14), (4.24).

Proposition 5.1.3. *Assume (A3), (A4), and let $h(\omega), d(\omega)$ follow the finite discrete probability distributions introduced at the beginning of Chapter 4. Then the following are equivalent:*
The linear-recourse, second-order dominance model

$$\min\{g^\top x \;:\; f(x,h(\omega)) \preceq_{icx} d(\omega),\; x \in \mathcal{X}\} \qquad (5.11)$$

- *and the mixed-integer linear program*

$$\min\left\{ g^\top x: \begin{array}{lll} (h_\ell - Tx)^\top \delta_i + & & \\ (c^\top x - d_k - \theta_{\ell k}(d_K - d_k))\delta_{i0} & \leq \; 0 & \forall \ell, k, i \\ \sum_{\ell=1}^{L} \pi_\ell \theta_{\ell k}(d_K - d_k) & \leq \; \sum_{j=k+1}^{K} p_j(d_j - d_k) & \forall k \\ x \in \mathcal{X},\; \theta_{\ell k} \in [0,1] & & \forall \ell, k \end{array} \right\} \tag{5.12}$$

- *and the mixed-integer linear program*

$$\min\left\{ g^\top x: \begin{array}{lll} (h_\ell - Tx)^\top \delta_i + & & \\ (c^\top x - \sum_{k=1}^{K} \theta_{\ell k} d_k)\delta_{i0} & \leq \; 0 & \forall \ell, i \\ \sum_{\ell=1}^{L} \pi_\ell \sum_{j=k+1}^{K} \theta_{\ell j}(d_j - d_k) & \leq \; \sum_{j=k+1}^{K} p_j(d_j - d_k) & \forall k \\ \sum_{k=1}^{K} \theta_{\ell k} & = \; 1 & \forall \ell \\ x \in \mathcal{X},\; \theta_{\ell k} \in [0,1] & & \forall \ell, k \end{array} \right\}. \tag{5.13}$$

Proof: Very similarly to the proof of Propositions 5.1.2, the principal lines of arguments for the mixed-integer recourse models in Section 4 are repeated with some adaptations. Let us demonstrate this at (5.12). To obtain the desired relation characterizing (this time) C_2, namely,

$$\sum_{\ell=1}^{L} \pi_\ell \left[f(x, h_\ell) - d_k\right]_+ \leq \sum_{j=k+1}^{K} p_j(d_j - d_k) \quad \forall\, k, \tag{5.14}$$

cf. (4.15), it is sufficient to fulfill the constraints of the prospective equivalent, i.e.,

$$(h_\ell - Tx)^\top \delta_i + (c^\top x - d_k - \theta_{\ell k}(d_K - d_k))\delta_{i0} \;\leq\; 0 \qquad \forall \ell, k, i; \tag{5.15}$$

$$\sum_{\ell=1}^{L} \pi_\ell \theta_{\ell k}(d_K - d_k) \;\leq\; \sum_{j=k+1}^{K} p_j(d_j - d_k) \;\; \forall k; \tag{5.16}$$

$$x \in \mathcal{X},\; \theta_{\ell k} \in [0,1] \quad \forall \ell, k$$

Now (5.15) implies in light of Lemma 5.1.1

$$f(x, h_\ell) - d_k \leq \theta_{\ell k}(d_K - d_k) \quad \forall\, \ell, k.$$

Together with (5.16) this yields the desired relation (5.14):

$$\sum_{\ell=1}^{L} \pi_\ell \left[f(x,h_\ell) - d_k\right]_+ \;\leq\; \sum_{\ell=1}^{L} \pi_\ell \theta_{\ell k}(d_K - d_k) \;\leq\; \sum_{j=k+1}^{K} p_j(d_j - d_k) \quad \forall\; k.$$

Vice versa, the desired relation (5.14) implies that the constraints of the prospected equivalent can be fulfilled. Given $x \in \mathcal{X}$, choose $\theta_{\ell k}$ such that

$$\theta_{\ell k}(d_K - d_k) \;=\; \left[f(x,h_\ell) - d_k\right]_+ \quad \forall\; \ell, k.$$

Then (5.14) implies (5.16). Due to Observation 4.1.4, we have $\theta_{\ell k} \in [0,1] \;\; \forall\, \ell, k$. To verify (5.15), notice that

$$f(x,h_\ell) - d_k \;\leq\; \left[f(x,h_\ell) - d_k\right]_+ \;=\; \theta_{\ell k}(d_K - d_k).$$

Thus, $f(x,h_\ell) \;\leq\; d_k + \theta_{\ell k}(d_K - d_k)$, from where Lemma 5.1.1 yields (5.15).

To prove the equivalence of C_2 and (5.13), it is sufficient to fulfill the constraints

$$(h_\ell - Tx)^\top \delta_i \;+ (c^\top x - \sum_{k=1}^{K} \theta_{\ell k} d_k)\delta_{i0} \leq 0 \quad \forall \ell, i; \tag{5.17}$$

$$\sum_{\ell=1}^{L} \pi_\ell \sum_{j=k+1}^{K} \theta_{\ell j}(d_j - d_k) \leq \sum_{j=k+1}^{K} p_j(d_j - d_k) \quad \forall k; \tag{5.18}$$

$$\sum_{k=1}^{K} \theta_{\ell k} = 1 \quad \forall \ell; \tag{5.19}$$

$$x \in \mathcal{X},\; \theta_{\ell k} \in [0,1] \quad \forall \ell, k. \tag{5.20}$$

Indeed, (5.17) and Lemma 5.1.1 imply

$$f(x,h_\ell) \;\leq\; \sum_{j=1}^{K} \theta_{\ell j} d_j, \tag{5.21}$$

and, together with (5.19) and (4.8),

$$f(x,h_\ell) - d_k \;\leq\; \sum_{j=1}^{K} \theta_{\ell j}(d_j - d_k) \;\leq\; \sum_{j=k+1}^{K} \theta_{\ell j}(d_j - d_k).$$

Then, (5.18) delivers the desired

$$\sum_{\ell=1}^{L} \pi_\ell [f(x,h_\ell) - d_k]_+ \leq \sum_{\ell=1}^{L} \pi_\ell \sum_{j=k+1}^{K} \theta_{\ell j}(d_j - d_k) \;\leq\; \sum_{j=k+1}^{K} p_j(d_j - d_k).$$

For the necessity part, we can choose $\theta_{\ell k}$ as in the proof of Proposition 4.2.2. There it was shown, that this choice of $\theta_{\ell k}$ fulfills (5.18), (5.19), (5.20) and that (5.21) holds. Since (5.21) implies (5.17) by Lemma 5.1.1, we are done.

□

Remark 5.1.4. *While in the MILP equivalents of Chapter 4 second-stage optimality is captured by explicit conditions involving the second-stage variables, the MILP equivalents of Propositions 5.1.2 and 5.1.3 rely on dual characterization of second-stage optimality by implicit constraints of the type (5.2).*

The formulation presented in the next section is based upon implicit constraints as well. However, in contrast to all formulations presented above, it is devoid of any auxiliary variables.

5.2 An Equivalent Formulation Based on the Polyhedral Representation

As we have discussed in Section 3.1.3, dominance constraints have close relations to constraints on some risk measures. Particularly, SSD constraints can be viewed as ICC with the risk thresholds defined through the benchmark. This implies, that the MILP equivalent formulations we developed for C_2 so far, can be applied to solve the problems with ICC as well.[2]

By contrast, in the present section, we consider an alternative formulation for C_2 which was originally developed for ICC in [KHvdV06] and then adapted for SSD in [RR08]. As opposed to the usual approach to model this relation, which allows to construct linear representations for the nonlinear shortfall terms $[t - \mathsf{X}]_+$ by means of additional indicator variables $(\theta_{\ell k})$, the alternative formulation will forgo any auxiliary variables.

[2]More precisely, our representations can be applied to solve problems with constraints on expected excess, cf. Proposition 2.2.13 and below.

We start from the representation of SSD which builds upon the characterization from Proposition 2.2.2 (ii) and is due to [RR08]. Analogously, we construct the corresponding representation for ICX basing upon Proposition 2.2.13.

Proposition 5.2.1. *Assume that* $\mathsf{X}, \mathsf{Y} \in \mathcal{L}_1(\Omega, \mathcal{F}, I\!P)$. *Then*

(i) $\mathsf{Y} \preceq_{(2)} \mathsf{X}$ *iff* $I\!E((t-\mathsf{X}) \cdot 1_A) \leq I\!E[t-\mathsf{Y}]_+ \quad \forall t \in \mathbb{R}, A \in \mathcal{F}$;

(ii) $\mathsf{X} \preceq_{icx} \mathsf{Y}$ *iff* $I\!E((\mathsf{X}-t) \cdot 1_A) \leq I\!E[\mathsf{Y}-t]_+ \quad \forall t \in \mathbb{R}, A \in \mathcal{F}$.

Proof: The proof of (i) can be found in [RR08, Theorem 2.2]. For (ii) we, analogously, prove necessity and sufficiency:

"$\Longrightarrow$" For all $A \in \mathcal{F},\ t \in \mathbb{R}$ it holds $(\mathsf{X}(\omega)-t) \cdot 1_A(\omega) \leq [\mathsf{X}(\omega)-t]_+$ for all $\omega \in \Omega$. By Proposition 2.2.13 then $I\!E((\mathsf{X}-t) \cdot 1_A) \leq I\!E[\mathsf{X}-t]_+ \leq I\!E[\mathsf{Y}-t]_+$;

"$\Longleftarrow$" For all $t \in \mathbb{R}$ and $A := [\mathsf{X} > t]$ we have $(\mathsf{X}(\omega)-t) \cdot 1_A(\omega) = [\mathsf{X}(\omega)-t]_+$ for all $\omega \in \Omega$, which yields the desired in view of Proposition 2.2.13.

□

These representations were termed *polyhedral* in [Fáb12], as opposed to the lifting-representations we considered so far. For a finite and discrete benchmark variable Y with realizations $\mathsf{Y}_k,\ k = 1, \dots, K$ the continua of constraints in Proposition 5.2.1 again reduce to a finite number, as the following proposition shows.

Proposition 5.2.2.

(i) $\mathsf{Y} \preceq_{(2)} \mathsf{X}$ *iff* $I\!E((\mathsf{Y}_k-\mathsf{X}) \cdot 1_A) \leq I\!E[\mathsf{Y}_k-\mathsf{Y}]_+ \quad \forall A \in 2^\Omega,\ k = 1, \dots, K$;

(ii) $\mathsf{X} \preceq_{icx} \mathsf{Y}$ *iff* $I\!E((\mathsf{X}-\mathsf{Y}_k) \cdot 1_A) \leq I\!E[\mathsf{Y}-\mathsf{Y}_k]_+ \quad \forall A \in 2^\Omega, \quad k = 1, \dots, K$.

Proof:

"$\Longrightarrow$" Proposition 5.2.1 for $t = \mathsf{Y}_k$ and $\mathcal{F} = 2^\Omega$;

"$\Longleftarrow$" For each k define $A_k := [\mathsf{X} > \mathsf{Y}_k]$ and it holds $(\mathsf{X}(\omega) - \mathsf{Y}_k) \cdot 1_{A_k}(\omega) = [\mathsf{X}(\omega) - \mathsf{Y}_k]_+$ for all $\omega \in \Omega$. By (4.5) this implies $\mathsf{X} \preceq_{icx} \mathsf{Y}$. To see (i) choose $A_k := [\mathsf{X} < \mathsf{Y}_k]$.

□

If we assume (A3), (A4), which ensure the finiteness of our rvs $f(x, h_\ell)$, the set C_2 has the representation with nonlinear entities as

$$C_2 = \left\{ x \in \mathcal{X} : \sum_{\ell=1}^{L} \pi_\ell (f(x, h_\ell) - d_k)_+ \leq \sum_{j=k+1}^{K} p_j (d_j - d_k),\ \forall\ k = 1, \ldots, K \right\}. \tag{5.22}$$

In view of Proposition 5.2.2, the corresponding polyhedral representation of this set immediately arises in the following proposition.

Proposition 5.2.3. *Assume (A3), (A4), and let $h(\omega), d(\omega)$ follow the finite discrete probability distributions introduced at the beginning of Chapter 4. Then it holds that*

$$C_2 = \left\{ x \in \mathcal{X} : \sum_{\ell \in A} \pi_\ell (f(x, h_\ell) - d_k) \leq v_k,\ \forall\ k = 1, \ldots, K,\ \ \forall\ A \subset \{1, \ldots, L\} \right\}, \tag{5.23}$$

where $v_k := \sum\limits_{j=k+1}^{K} p_j (d_j - d_k)$ from now on.

Unfortunately, (5.23) cannot be used for decomposition directly because the involved rvs are not given explicitly. As in the previous section, in case of linear recourse we can use LP-duality to derive a polyhedral formulation amenable for decomposition. In fact, under the assumptions (A3) and (A4) it holds that

$$\begin{aligned} f(x, h_\ell) &= c^\top x + \min\{q^\top y \ :\ Wy \geq h_\ell - Tx,\ y \geq 0\} \\ &= c^\top x + \max\{(h_\ell - Tx)^\top u : W^\top u \leq q,\ u \geq 0\} \end{aligned}$$

with both optimization problems being solvable. For the dual problem it means that we can restrict optimization over the vertices $\delta_i \in \mathbb{R}^s, i = 1, \dots, I$, of

$$\bar{\Delta} := \left\{ u \in \mathbb{R}^s \ : \ W^\top u \leq q, \ u \geq 0 \right\}.$$

For each $x \in \mathcal{X}, \ \ell = 1, \dots, L$ it holds

$$f(x, h_\ell) \ \geq \ c^\top x + (h_\ell - Tx)^\top \delta_i \quad \forall\, i = 1, \dots, I \tag{5.24}$$

and there exists a vertex $\delta(x, h_\ell)$ of $\bar{\Delta}$ such that

$$f(x, h_\ell) = c^\top x + \max_{\delta_i \in \bar{\Delta}} \{(h_\ell - Tx)^\top \delta_i : i = 1, \dots, I\} = c^\top x + (h_\ell - Tx)^\top \delta(x, h_\ell). \tag{5.25}$$

With this representation of the rvs, we obtain the desired formulation for C_2 in the following proposition.

Proposition 5.2.4. *Assume (A3), (A4), and let $h(\omega), d(\omega)$ follow the finite discrete probability distributions introduced at the beginning of Chapter 4. Then it holds that*

$$C_2 = \left\{ x \in \mathcal{X} : \sum_{\ell \in A} \pi_\ell((h_\ell - Tx)^\top \boldsymbol{\delta} + c^\top x - d_k) \leq v_k, \quad \forall\ A \subset \{1, \dots, L\}, \ \forall\ k \right\}, \tag{5.26}$$

where $\boldsymbol{\delta}$ in each addend is an arbitrary vertex of $\bar{\Delta}$, so that we have $I^{|A|}$ constraints for any fixed k and A.

Proof: We call the set C_2 in representation (5.23) S_1 and in representation (5.26) S_2 and show $S_1 = S_2$.
"$\subset$": For an $x \in S_1$, (5.24) yields

$$\sum_{\ell \in A} \pi_\ell((h_\ell - Tx)^\top \boldsymbol{\delta} + c^\top x - d_k) \leq \sum_{\ell \in A} \pi_\ell(f(x, h_\ell) - d_k) \leq v_k$$

for any fixed k and A.
"$\supset$": For $x \in S_2$ the constraint holds for arbitrary vertices δ in each addend, so also for the ones with $f(x, h_\ell) = c^\top x + (h_\ell - Tx)^\top \delta(x, h_\ell)$. $\square$

Chapter 6

Cutting Plane Methods

The representations (5.5), (5.6), (5.12), and (5.13) all contain constraints (or cuts) of the type (5.2)

$$(h - Tx)^\top \delta_i + (c^\top x - \eta)\delta_{io} \leq 0$$

induced by vertices (δ_i, δ_{io}) of Δ, see Lemma 5.1.1, whereas the representation (5.26) employs, roughly speaking, the aggregated version of those cuts, see Proposition 5.2.4. Because the number of vertices in a polyhedron may grow exponentially in the problem dimension, all the MILP equivalents derived in the previous chapter are about to have a vast number of such constraints for realistic problem dimensions. Moreover, these constraints are not available a priorily.

Rather, we will generate them as they are needed in the course of the subsequent algorithms. In these algorithms, the idea of traditional Benders' decomposition, or L-shaped method in stochastic programming terms, see Section 3.2.2 and [Ben63, BL97, KW94, Pré95, RS03], has a central role.

The algorithms work with master problems, i.e., relaxations of the full problems (5.5), (5.6), (5.12), (5.13) or (5.26) involving subsets of the above cuts. For an optimal solution to the master problem it is checked algorithmically whether it violates cuts not yet included in the master. If so, then violated cuts are added. If not, then the optimal solution to the master problem is optimal for the full problem as well. The algorithms differ in the MILP equivalents they employ and in the ways cuts are managed.[1]

6.1 Decomposition Methods for the Lifting-Representations

For the lifting-representations of Section 5.1, the algorithmic check of cut violation is based on solving the dual feasibility problem (5.3) from the proof of Lemma 5.1.1:

$$D(x,h,\eta) \quad \max_{u,u_o} \left\{ \begin{array}{ll} (h-Tx)^\top u + (c^\top x - \eta)u_o \; : & 0 \le u \le \mathbf{1},\ 0 \le u_o \le 1, \\ & W^\top u - q u_o \le 0 \end{array} \right\}.$$

By the notation $D(x,h,\eta)$ we indicate that the problem will be solved for suitable specifications of the parameters (x,h,η). Note that changes in the parameters do not influence the feasible set which is always Δ.[2]

We begin with an algorithm tackling the first-order model (5.5):

$$\min \left\{ g^\top x : \begin{array}{lll} (h_\ell - Tx)^\top \delta_i + & & \\ (c^\top x - d_k - \theta_{\ell k}(d_K - d_k))\delta_{i0} \le 0 & & \forall \ell, k, i \\ \sum_{\ell=1}^{L} \pi_\ell \theta_{\ell k} \le \sum_{j=k+1}^{K} p_j & & \forall k \\ x \in \mathcal{X},\ \theta_{\ell k} \in \{0,1\} & & \forall \ell, k \end{array} \right\}.$$

Algorithm 6.1.1. *Cutting Plane Algorithm for (5.5)*

Step 1 (Initialization):
Set $\nu := 0$ *and* $\mathcal{I}^0 = \emptyset$.

Step 2 (Master Problem):

[1] The results of Section 6.1 were originally published by the author in [DKS11].

[2] In practice, this set depends on the scenario index once q is allowed to be stochastic.

Solve the current master problem $(MP)^\nu$

$$\min \left\{ g^\top x : \begin{array}{lll} (h_\ell - Tx)^\top \delta_i + & & \\ (c^\top x - d_k - \theta_{\ell k}(d_K - d_k))\delta_{i0} & \leq \; 0 & \forall (\ell, k, i) \in \mathcal{I}^\nu \\ \sum\limits_{\ell=1}^{L} \pi_\ell \theta_{\ell k} & \leq \; \sum\limits_{j=k+1}^{K} p_j & \forall k \\ x \in \mathcal{X}, \; \theta_{\ell k} \in \{0,1\} & & \forall \ell, k \end{array} \right\}$$

and let (x^ν, θ^ν) *be an optimal solution.*

Step 3 (Subproblems):

Set $\mathcal{I}_+ = \emptyset$.

Solve the subproblems $D(x^\nu, h_\ell, d_k)$ *for all* (ℓ, k) *such that* $\theta^\nu_{\ell k} = 0$. *Distinguish between the following situations:*

3.1) If all subproblems have optimal value zero, then STOP. The current solution (x^ν, θ^ν) *is optimal for (5.5).*

3.2) For each subproblem $D(x^\nu, h_\ell, d_k)$*, whose optimal value is greater than zero, pick an optimal vertex* (δ_i, δ_{i0}) *yielding a cut*

$$(h_\ell - Tx)^\top \delta_i + (c^T x - d_k - \theta_{\ell k}(d_K - d_k))\delta_{i0} \;\; \leq \;\; 0$$

which is added to the master problem. Collect these triplets (ℓ, k, i) *in a set* $\mathcal{I}_+$.

Set $\nu := \nu + 1$ *and* $\mathcal{I}^{\nu+1} := \mathcal{I}^\nu \cup \mathcal{I}_+$; *GOTO Step 2.*

Proposition 6.1.2. *Assume (A3), (A4) and that* $\mathcal{X}$ *is bounded. If the feasible set of (5.5) is nonempty, then Algorithm 6.1.1 terminates with an optimal solution to (5.5) after a finite number of steps.*

Proof: Recall that $\mathcal{X}$ is a non-empty polyhedron, possibly with integer requirements to components of x. By the boundedness assumption, this set is compact. Hence, if the feasible set of (5.5) is nonempty, it is compact as well, implying that (5.5) and all master problems arising in the course of the algorithm possess optimal solutions. Clearly, these master problems are relaxations of (5.5).

Consider (ℓ, k) such that $\theta^{\nu}_{\ell k} = 0$ and the optimal value of $D(x^{\nu}, h_{\ell}, d_k)$ equals zero. Then $(x, \theta) = (x^{\nu}, \theta^{\nu})$ satisfies the constraints

$$(h_{\ell} - Tx)^{\top}\delta_i + (c^{\top}x - d_k - \theta_{\ell k}(d_K - d_k))\delta_{i0} \;\; \leq \;\; 0 \qquad (6.1)$$

for all vertices (δ_i, δ_{i0}) of Δ.

For $k = K$, the constraint

$$\sum_{\ell=1}^{L} \pi_{\ell}\theta_{\ell k} \;\; \leq \;\; \sum_{j=k+1}^{K} p_j$$

implies $\theta_{\ell K} = 0$ for all $\ell = 1, \ldots .L$. Inserting this into (6.1) yields

$$(h_{\ell} - Tx)^{\top}\delta_i + (c^{\top}x - d_K)\delta_{i0} \;\; \leq \;\; 0 \quad \forall\ (\delta_i, \delta_{i0}),$$

coinciding with the constraint (6.1) for (ℓ, k) with $\theta^{\nu}_{\ell k} = 1$.

Hence, (x^{ν}, θ^{ν}) is both feasible to (5.5) and optimal to a relaxation of (5.5). Thus, (x^{ν}, θ^{ν}) is optimal to (5.5).

The algorithm must terminate after finitely many steps, since altogether there are only finitely many cuts of the type (5.2), and at least one new cut is added per loop. □

Algorithm 6.1.1 does not prescribe a specific order to deal with the subproblems $D(x^{\nu}, h_{\ell}, d_k)$, corresponding to (ℓ, k) with $\theta^{\nu}_{\ell k} = 0$. By the following observation, it is possible to establish such an order which leads to a substantial reduction of subproblems to be considered.

Observation 6.1.3. *If for some (x, ℓ, k) the optimal value of $D(x, \ell, d_k)$ is zero then the same is true for all (x, ℓ, k') with $k' \geq k$. Indeed, by Lemma 5.1.1, $f(x, h_{\ell}) \leq d_k$ if and only if the optimal value of $D(x, h_{\ell}, d_k)$ is zero. Since the d_k are in ascending order, it follows that $f(x, h_{\ell}) \leq d_{k'}$ for all $k' \geq k$. Therefore, the optimal value of $D(x, h_{\ell}, d_{k'})$ is zero for all $k' \geq k$.*

The next algorithm, again for the MILP equivalent (5.5), makes use of this observation. After having solved the master problem and picked an optimal solution (x^ν, θ^ν), the subproblems $D(x^\nu, h_\ell, d_k)$ are inspected with indizes $\ell \in \{1, \dots, L\}$ iterating in an outer and $k \in \{1, \dots, K\}$ in an inner loop.

For given ℓ, the index k is increased starting from $k = 1$ up to the first k with $\theta^\nu_{\ell k} = 0$. If $D(x^\nu, h_\ell, d_k)$ has optimal value zero then the same is true for all $k' \geq k$, and the iteration continues with $\ell := \ell + 1,\ k := 1$. If the optimal value of $D(x^\nu, h_\ell, d_k)$ is positive, then, before continuing with $\ell := \ell + 1,\ k := 1$, a cut is constructed and added to the master problem. Thus, for each ℓ precisely one subproblem is solved and at most one cut generated per iteration.

Algorithm 6.1.4. *Enhanced Cutting Plane Algorithm for (5.5)*

STEP 1 (INITIALIZATION):

Set $\nu := 0$ *and* $\mathcal{I}^0 = \emptyset$.

STEP 2 (MASTER PROBLEM):

Solve the current master problem $(MP)^\nu$

$$\min \left\{ g^\top x : \begin{array}{lcll} (h_\ell - Tx)^\top \delta_i + & & & \\ (c^\top x - d_k - \theta_{\ell k}(d_K - d_k))\delta_{i0} & \leq & 0 & \forall (\ell, k, i) \in \mathcal{I}^\nu \\ \sum\limits_{\ell=1}^{L} \pi_\ell \theta_{\ell k} & \leq & \sum\limits_{j=k+1}^{K} p_j & \forall k \\ x \in \mathcal{X},\ \theta_{\ell k} \in \{0,1\} & & & \forall \ell, k \end{array} \right\}$$

and let (x^ν, θ^ν) *be an optimal solution.*

STEP 3 (SUBPROBLEMS):

Set $\mathcal{I}_+ = \emptyset$.

For $\ell = 1, \dots, L$

For $k = 1, \dots, K$

IF $\theta^\nu_{\ell k} = 0$ *solve* $D(x^\nu, h_\ell, d_k)$

IF optimal value of $D(x^\nu, h_\ell, d_k)$ *equals zero*

BREAK;

ELSE Pick a vertex (δ_i, δ_{i0}) *of* Δ *optimal to* $D(x^\nu, h_\ell, d_k)$

Add $(h_\ell - Tx)^\top \delta_i + (c^T x - d_k - \theta_{\ell k}(d_K - d_k))\delta_{i0} \leq 0$

to the master and (ℓ, k, i) to $\mathcal{I}_+$

BREAK;

3.1) If all subproblems considered have optimal value zero, then STOP. The current solution x^ν is optimal for (5.5);

3.2) Otherwise, set $\nu := \nu + 1$ and $\mathcal{I}^{\nu+1} := \mathcal{I}^\nu \cup \mathcal{I}_+$; GOTO Step 2.

Sketched Proof of Correctnes: The algorithm terminates after a finite number of steps for the same reason as the previous one did so.

In each loop ν of the algorithm and for each data scenario ℓ exactly one subproblem $D(x^\nu, h_\ell, d_k)$ is solved. Upon termination, according to Substep 3.1, all these subproblems have optimal value zero.

For fixed ℓ, by Observation 6.1.3 all subproblems $D(x^\nu, h_\ell, d_{k'})$ with $k' \geq k$ have optimal value zero, too. As in the proof of Proposition 6.1.2,

$$(h_\ell - Tx^\nu)^\top \delta_i \, + \, (c^\top x^\nu - d_k - \theta_{\ell k}(d_K - d_k))\delta_{i0} \; \leq \; 0$$

then holds for all $k' \geq k$ and all (δ_i, δ_{i0}). For $k' < k$, by construction, $\theta^\nu_{\ell k'} = 1$, whose feasibility is equivalent to the feasibility of $\theta^\nu_{\ell K} = 0$ we have just checked.

□

For second-order dominance, Algorithm 6.1.4 may suggest a counterpart method addressing representation (5.12). The key distinction between the models (5.5) and (5.12) is that the former has constraints $\theta_{\ell k} \in \{0,1\} \; \forall \ell, k$ where the latter has $\theta_{\ell k} \in [0,1] \; \forall \ell, k$.

The correctness proof of a hypothetical counterpart algorithm would follow verbatim the above arguments, except for the last conclusion where feasibility for $k' < k$ is verified. The constraints that remain to be checked have $\theta^\nu_{\ell k'} \neq 0$ and read

$$(h_\ell - Tx^\nu)^\top \delta_i \, + \, (c^\top x^\nu - d_k - \theta_{\ell k}(d_K - d_k))\delta_{i0} \; \leq \; 0.$$

For $\theta_{\ell k} \in \{0,1\}$ the requirement that $\theta_{\ell k} \neq 0$ implies $\theta_{\ell k} = 1$ leading to

$$(h_\ell - Tx^\nu)^\top \delta_i \, + \, (c^\top x^\nu - d_K))\delta_{i0} \; \leq \; 0,$$

which coincides with the constraint for $k' := K \geq k$ that was checked already. So no additional effort is needed.

For $\theta_{\ell k} \in [0,1]$, however, the requirement that $\theta_{\ell k} \neq 0$ merely gives $0 < \theta_{\ell k} \leq 1$, leading to

$$(h_\ell - Tx^\nu)^\top \delta_i \;+\; (c^\top x^\nu - d_k - \theta_{\ell k}(d_K - d_k)\delta_{i0} \;\leq\; 0,$$

which has to be checked explicitly by solving the subproblems $D(x^\nu, h_\ell, d_{k'} + \theta_{\ell k'}(d_K - d_{k'}))$ for all $k' < k$, amounting to complete enumeration of the k'.

The next algorithm again is devoted to first-order dominance, but now for the model (5.6) where the Boolean variables $\theta_{\ell k}$ have a slightly different role. So far, these variables indicated whether, for given x, ℓ, k, the function values $f(x, h_\ell)$ exceed d_k or not. Now these variables, for given x, ℓ, indicate at which index k the change from $f(x, h_\ell) \leq d_k$ to $f(x, h_\ell) > d_k$ takes place:

$$\min \left\{ g^\top x : \begin{array}{lcll} (h_\ell - Tx)^\top \delta_i + (c^\top x - \sum\limits_{k=1}^{K} \theta_{\ell k} d_k)\delta_{i0} & \leq & 0 & \forall \ell, i \\ \sum\limits_{\ell=1}^{L} \pi_\ell \sum\limits_{j=k+1}^{K} \theta_{\ell j} & \leq & \sum\limits_{j=k+1}^{K} p_j & \forall k \\ \sum\limits_{k=1}^{K} \theta_{\ell k} & = & 1 & \forall \ell \\ x \in \mathcal{X},\ \theta_{\ell k} \in \{0,1\} & & & \forall \ell, k \end{array} \right\}.$$

Algorithm 6.1.5. *Cutting Plane Algorithm for (5.6)*

STEP 1 (INITIALIZATION):
Set $\nu := 0$ *and* $\mathcal{I}^0 = \emptyset$.

STEP 2 (MASTER PROBLEM):

Solve the current master problem $(MP)^\nu$

$$\min \left\{ g^\top x : \begin{array}{lcll} (h_\ell - Tx)^\top \delta_i + (c^\top x - \sum_{k=1}^{K} \theta_{\ell k} d_k)\delta_{i0} & \leq & 0 & \forall (\ell, i) \in \mathcal{I}^\nu \\ \sum_{\ell=1}^{L} \pi_\ell \sum_{j=k+1}^{K} \theta_{\ell j} & \leq & \sum_{j=k+1}^{K} p_j & \forall k \\ \sum_{k=1}^{K} \theta_{\ell k} & = & 1 & \forall \ \ell \\ x \in \mathcal{X}, \ \theta_{\ell k} \in \{0,1\} & & & \forall \ell, k \end{array} \right\}$$

and let (x^ν, θ^ν) *be an optimal solution.*

STEP 3 (SUBPROBLEMS):

Set $\mathcal{I}_+ = \emptyset$.

Solve the subproblems $D(x^\nu, h_\ell, \sum_{k=1}^{K} \theta^\nu_{\ell k} d_k)$ *for all* $\ell \in \{1, \ldots, L\}$. *Consider the situations:*

3.1) If all subproblems have optimal value zero, then STOP. The current solution (x^ν, θ^ν) *is optimal for (5.6).*

3.2) If some of these subproblems have optimal value greater than zero, then their optimal solutions yield vertices (δ_i, δ_{i0}) *of* Δ. *Add the resulting cuts*

$$(h_\ell - Tx)^\top \delta_i + (c^T x - \sum_{k=1}^{K} \theta_{\ell k} d_k)\delta_{i0} \leq 0$$

to the master problem, and add (ℓ, i) *to* $\mathcal{I}_+$.

Set $\nu := \nu + 1$ *and* $\mathcal{I}^{\nu+1} := \mathcal{I}^\nu \cup \mathcal{I}_+$; *GOTO Step 2.*

Proposition 6.1.6. *Assume (A3), (A4) and that* $\mathcal{X}$ *is bounded. If the feasible set of (5.6) is nonempty, then Algorithm 6.1.5 terminates with an optimal solution to (5.6) after a finite number of steps.*

Proof: As with Algorithm 6.1.1, the assumptions ensure that (5.6) as well as all master problems in the course of the iteration possess optimal solutions.

If all subproblems $D(x^\nu, h_\ell, \sum_{k=1}^{K} \theta^\nu_{\ell k} d_k)$

$$\max_{u, u_o} \left\{ \begin{array}{ll} (h_\ell - Tx^\nu)^\top u + \left(c^\top x^\nu - \sum_{k=1}^{K} \theta^\nu_{\ell k} d_k \right) u_o : & 0 \leq u \leq \mathbf{1}, \ 0 \leq u_o \leq 1, \\ & W^\top u - q u_o \leq 0, \end{array} \right\}$$

for $\ell = 1, \ldots, L$, have optimal value zero, then

$$(h_\ell - Tx^\nu)^\top \delta_i + \left(c^\top x^\nu - \sum_{k=1}^{K} \theta^\nu_{\ell k} d_k \right) \delta_{i0} \;\leq\; 0$$

for all vertices (δ_i, δ_{i0}) of Δ, implying that (x^ν, θ^ν) is feasible for (5.6). Simultaneously, (x^ν, θ^ν) is optimal for a relaxation of (5.6), and thus optimal for (5.6).

Again, termination of the method is granted since there are only finitely many cuts in total, and per loop at least one new cut is added to the master problem. □

We conclude the present section with an algorithm for the second-order dominance model (5.13)

$$\min \left\{ g^\top x : \begin{array}{lcll} (h_\ell - Tx)^\top \delta_i + (c^\top x - \sum\limits_{k=1}^{K} \theta_{\ell k} d_k)\delta_{i0} & \leq & 0 & \forall \ell, i \\ \sum\limits_{\ell=1}^{L} \pi_\ell \sum\limits_{j=k+1}^{K} \theta_{\ell j}(d_j - d_k) & \leq & \sum\limits_{j=k+1}^{K} p_j (d_j - d_k) & \forall k \\ \sum\limits_{k=1}^{K} \theta_{\ell k} & = & 1 & \forall \ell \\ x \in \mathcal{X},\; \theta_{\ell k} \in [0,1] & & & \forall \ell, k \end{array} \right\}.$$

Algorithm 6.1.7. *Cutting Plane Algorithm for (5.13)*

STEP 1 (INITIALIZATION):
Set $\nu := 0$ *and* $\mathcal{I}^0 = \emptyset$.

STEP 2 (MASTER PROBLEM):
Solve the current master problem $(MP)^\nu$

$$\min \left\{ g^\top x : \begin{array}{lcll} (h_\ell - Tx)^\top \delta_i + & & & \\ (c^\top x - \sum\limits_{k=1}^{K} \theta_{\ell k} d_k)\delta_{i0} & \leq & 0 & \forall (\ell, i) \in \mathcal{I}^\nu \\ \sum\limits_{\ell=1}^{L} \pi_\ell \sum\limits_{j=k+1}^{K} \theta_{\ell j}(d_j - d_k) & \leq & \sum\limits_{j=k+1}^{K} p_j (d_j - d_k) & \forall k \\ \sum\limits_{k=1}^{K} \theta_{\ell k} & = & 1 & \forall \ell \\ x \in \mathcal{X},\; \theta_{\ell k} \in [0,1] & & & \forall \ell, k \end{array} \right\}$$

and let (x^ν, θ^ν) *be an optimal solution.*

STEP 3 (SUBPROBLEMS):
Set $\mathcal{I}_+ = \emptyset$.
Solve the subproblems $D(x^\nu, h_\ell, \sum_{k=1}^K \theta^\nu_{\ell k} d_k)$ *for all* $\ell \in \{1, \ldots, L\}$. *Consider the following situations:*

3.1) If all subproblems have optimal value zero, then STOP. The current solution (x^ν, θ^ν) *is optimal for (5.13).*

3.2) If some of these subproblems have optimal value greater than zero, then their optimal solutions yield vertices (δ_i, δ_{i0}) *of* Δ. *Add the resulting cuts*

$$(h_\ell - Tx)^\top \delta_i + (c^T x - \sum_{k=1}^{K} \theta_{\ell k} d_k)\delta_{i0} \ \leq \ 0$$

to the master problem and (ℓ, i) *to* $\mathcal{I}_+$.
Set $\nu := \nu + 1$ *and* $\mathcal{I}^{\nu+1} := \mathcal{I}^\nu \cup \mathcal{I}_+$; *GOTO Step 2.*

To see that Algorithm 6.1.7 works correctly, we resort to the proof of Proposition 6.1.6, which transfers verbatim to the present situation.

6.2 Decomposition for the Polyhedral Representation

In the present section, we propose a decomposition method for the polyhedral representation from Section 5.2. To check the feasibility of the current iterate, i.e., $x = x^\nu \ \in \ C_2$, the method uses the formulation (5.22) by computing for each $\ell = 1, \ldots, L$ the quantities

$$f(x, h_\ell) = c^\top x + (h_\ell - Tx)^\top \delta(x, h_\ell)$$

with $\delta(x, h_\ell)$ being the vertices of $\bar{\Delta} \ = \ \left\{u \in \mathbb{R}^s \ : \ W^\top u \leq q, \ u \geq 0\right\}$ which arise as optimal solutions to the subproblems

$$\max_u \{(h_\ell - Tx)^\top u : W^\top u \leq q, \ u \geq 0\}. \tag{6.2}$$

To reach optimality, the algorithm works with master problems which are relaxations of the full problem in shape (5.26) from Proposition 5.2.4. If a solution to the master problem is checked to be infeasible to the full problem, the violated cuts can be constructed with the help of the vertices $\delta(x, h_\ell)$ already computed during the feasibility check. The following algorithm arises.

Algorithm 6.2.1. *Cutting Plane Algorithm for (5.26)*

Step 1 (Initialization):
Set $\nu := 0$ *and* $\mathcal{K}^0 := \emptyset$.

Step 2 (Master Problem):
Solve the current master problem $(MP)^\nu$*:*

$$\min \left\{ g^\top x : \begin{array}{ll} \sum\limits_{\ell \in A_k^n} \pi_\ell((h_\ell - Tx)^\top \delta_\ell^n + c^\top x - d_k) & \leq v_k, \qquad \forall\, n < \nu,\ \forall\, k \in \mathcal{K}^n \\ x \in \mathcal{X} & \end{array} \right\},$$

where $v_k := \sum\limits_{j=k+1}^{K} p_j(d_j - d_k)$*. Let* x^ν *be an optimal solution.*

Step 3 (Subproblems):
Solve the subproblems (6.2) for $\ell = 1, \ldots, L$ *and obtain optimal vertices* δ_ℓ^ν *of* $\bar{\Delta}$*. Check the feasibility for* x^ν *by verifying constraints from (5.22):*

$$\sum_{\ell=1}^{L} \pi_\ell (f(x^\nu, h_\ell) - d_k)_+ = \sum_{\ell=1}^{L} \pi_\ell (c^\top x + (h_\ell - Tx^\nu)^\top \delta_\ell^\nu - d_k)_+ \leq v_k \quad \forall\, k.$$

Distinguish between the following situations:

3.1) If x^ν *satisfies the constraints for all* k*, then STOP. The current solution* x^ν *is optimal for (5.26).*

3.2) Let $\mathcal{K}^\nu$ *be the set of all* k*, where the constraints are violated. For all* $k \in \mathcal{K}^\nu$ *define* $A_k^\nu := \{\ell : f(x^\nu, h_\ell) > d_k\}$ *and introduce cuts*

$$\sum_{\ell \in A_k^\nu} \pi_\ell((h_\ell - Tx)^\top \delta_\ell^\nu + c^\top x - d_k) \leq v_k,\ \forall\, k \in \mathcal{K}^\nu, \tag{6.3}$$

which are added to the master problem.

Set $\nu := \nu + 1$*; GOTO Step 2.*

Proposition 6.2.2. *Assume (A3), (A4) and that $\mathcal{X}$ is bounded. If the feasible set of (5.26) is nonempty, then Algorithm 6.2.1 terminates with an optimal solution to (5.26) after a finite number of steps.*

Proof: Recall that $\mathcal{X}$ is a non-empty polyhedron, possibly with integer requirements to components of x. By the boundedness assumption, this set is compact. Hence, if the feasible set of (5.26) is nonempty, it is compact as well, implying that (5.26) and all master problems arising in the course of the algorithm possess optimal solutions. Clearly, these master problems are relaxations of (5.26). All subproblems are solvable due to assumptions (A3) and (A4).

In view of Proposition 5.2.4 the formulations (5.22) and (5.26) are equivalent, so if $x = x^\nu$ satisfies the constraints in (5.22) it is both feasible to (5.26) and optimal to a relaxation of (5.26). Thus, x^ν is optimal to (5.26).

The algorithm must terminate after finitely many steps, since altogether there are only finitely many cuts of the type (6.3), and at least one new cut is added per loop. □

For a more detailed understanding of the presented method and in order to get a better comparison to the decomposition methods developed in [DKS11] and Section 6.1 we will now provide some details on our implementation of the cut generation routine from Step 3 of the above algorithm.

Algorithm 6.2.3. *(An improved cut generation routine for Algorithm 6.2.1) Let $x = x^\nu$ be the solution of the current master problem $(MP)^\nu$ from Algorithm 6.2.1:*

For $\ell = 1, \ldots, L$

Solve subproblem (6.2) $\Longrightarrow \delta_\ell^\nu$ *and* $f(x^\nu, h_\ell)$

For $k = 1, \ldots, K$

IF $(f(x^\nu, h_\ell) > d_k)$

Add scenario ℓ *to* A_k^ν;

θ_k+=$\pi_\ell(f(x^\nu, h_\ell) - d_k)$;

ELSE

IF $(\ell < L)$

BREAK; (see Remark 6.2.4 (1))

IF $(\ell = L)$

IF $(\theta_k = 0)$

BREAK; (see Remark 6.2.4 (2))

ELSE

IF $(\theta_k > \upsilon_k)$

Add a cut with A_k^ν *to the MASTER; (see Remark 6.2.4 (3))*

Remark 6.2.4. *In each step of the algorithm L subproblems have to be solved. Only one path though* (ℓ, k) *combinations is carried out whereby only the needed combinations are treated, see (1) and (2). At most* K *cuts are created, usually much less due to (3).*

(1) θ_k *is increased by* $\pi_\ell(f(x^\nu, h_\ell) - d_k)_+$ *for each scenario* ℓ*, so that it equals the left-hand side in the constraints (5.22), i.e.,* $\theta_k = \sum_{\ell=1}^{L} \pi_\ell(f(x^\nu, h_\ell) - d_k)_+$*, for* $\ell = L$ *of the loop. If* $f(x^\nu, h_\ell) < d_k$ *the increment is zero, but not only for the current* $\bar{k}$ *but also for all* $k > \bar{k}$ *due to the ascending ordering of* d_k.

(2) *For the same reason,* θ_k *is decreasing in* k*. Since* υ_k *is nonnegative, all constraints from (5.22) would be fulfilled for* $k \geq \bar{k}$ *and no cuts are to be applied.*

(3) *In contrast to the primal cutting plane method presented in [RR08] for the polyhedral representation, the sets* A_k^ν *generated here may repeat. More precisely, these sets may repeat for different* k *because we generate multiple cuts per iteration. However, the following observations hold:*

a) *the coefficients in the cuts (6.3) can be calculated as* $\sum_{\ell \in A_k^\nu} \pi_\ell (c^\top - \delta_\ell^{\nu\top} T)$*;*

b) *the right-hand side in the cuts (6.3) is* $v_k - \sum_{\ell \in A_k^\nu} \pi_\ell (\delta_\ell^{\nu\top} h_\ell - d_k)$*; both* v_k *and the sum are decreasing in* k *but the right-hand side, in general, is not;*

c) $A_1^\nu \supset A_2^\nu \supset \cdots \supset A_K^\nu$ *and because of a), cuts with* $A_k^\nu = A_{\bar{k}}^\nu$ *have identical coefficients;*

d) *due to a) - c), out of all cut candidates with the same set* A_k^ν*, we may generate a single cut by choosing the cut with the smallest right-hand side. With this modification, the number of cuts added in one iteration* ν *is simply the number of different sets* A_k^ν *generated here.*

Remark 6.2.5. *In the recent publication [DM12], a method similar to Algorithm 6.2.1 was developed independently in context of the so-called* quantile decomposition*, which is based on yet another representation of ICX (related to the dual representation of Proposition 2.2.15):*

$$\mathsf{X} \preceq_{icx} \mathsf{Y} \Longleftrightarrow \mathbb{E}(\mathsf{X}|\mathsf{X} > \alpha) \leq \frac{1}{\mathbb{P}(\mathsf{X} > \alpha)} \int_{F_{\mathsf{X}}(\alpha)}^{1} F_{\mathsf{Y}}^{-1}(t)\, dt \;\forall \alpha \in \mathbb{R}\,, \mathbb{P}(\mathsf{X} > \alpha) > 0.$$

Similar to Algorithm 6.2.1, cuts approximating the dominance constraints are generated with the help of some sets $\{\mathsf{X} > \alpha\}$.[3] *The algorithmic framework developed for the resulting quantile decomposition method is then applied for*

[3] The integral on the right-hand side defines a function which is closely related to the second quantile function (2.18). If it can be evaluated on all the required points, e.g., with a minimization rule similar to (2.26), the benchmark rv does not need to be discrete for this method.

the representation from Proposition 5.2.2 (ii). The differences to our approach of Algorithm 6.2.1 lie in the treatment of feasibility and in the way the cuts are created and managed.

Chapter 7

Computational Experiments

7.1 Test Problem Formulations

There are (at least) three options for solving the dominance-constrained stochastic programs studied in the present thesis. The first is to apply a general-purpose mixed-integer linear programming solver such as CPLEX, [CPL13], to the large-scale MILP equivalents displayed in Chapter 4. The next option is to resort to decomposition methods, combining relaxation of nonanticipativity with branch-and-bound. These methods, see [GNS08, GGS11], work for dominance constraints induced by mixed-integer recourse. Finally, there are the decomposition methods developed in Chapter 6 which require, however, that the underlying model has no integer variables in the second stage. In the present chapter, we report on computational experiments in light of these options. Accent is placed on models with dominance constraints induced by linear recourse.

For first-order models, we compare the performance of the Algorithms 6.1.1, 6.1.4, and 6.1.5 proposed in the present thesis with the computational behavior of CPLEX, [CPL13], and with the decomposition method from [GNS08] which even is able to cope with models involving mixed-integer linear recourse. Our test instances stem from a two-stage investment planning problem for electricity generation under uncertainty, inspired by a multi-stage model introduced in [LS88]. A second, more demanding but also more academic, group of test problems was formed on the basis of Sudoku puzzling.

For second-order models, we illustrate the performance of the Algorithms 6.1.7 and 6.2.1 compared to the direct application of a general-purpose MILP solver to the representations (4.14) and (4.24). The test instances concern an electricity retailer problem analyzed in [CGS09].

7.2 First-Order Models

Let us start with a few words on the practical background of our first group of test instances. They are derived from a two-stage investment planning problem for electricity generation under uncertainty, which was inspired by a multi-stage model introduced in [LS88].

In the first stage, (integer) decisions on capacity expansions for different generation technologies with budget constraints and supply guarantee are made under uncertainty of power demand. The second stage concerns the minimization of production costs for electricity under the constraints that electricity demand is met and the available capacity is not exceeded. Here, the decision variables are continuous. The random variable $f(x, h(\omega))$ arises as the minimum of the costs incurred by the investment decisions in the first stage and the production plans in the second. With the random benchmark $d(\omega)$, the constraint $f(x, h(\omega)) \preceq_{(1)} d(\omega),\ x \in \mathcal{X}$, singles out those investment policies x that are economically feasible and lead to costs $f(x, h(\omega))$ which are preferable to the benchmark in terms of stochastic order.

The dominance constrained stochastic program (4.1) is completed by an objective function referring to preferences among different pre-designated technologies for capacity expansion.

Table 7.1 displays the dimensions of (4.9) for test instances with different numbers K of benchmark and L of demand scenarios.

K	L	Variables			Constraints
		General Integer	Boolean	Continuous	
10	150	4	1500	3001	2863
	300	4	3000	6001	5713
	1000	4	10000	20001	19013
20	80	4	1600	1601	2343
	100	4	2000	2001	2923
	150	4	3000	3001	4373
	300	4	6000	6001	8723

Table 7.1: Dimensions of mixed-integer linear programming equivalents (investment planning)

Table 7.2 reports computing times for the individual solvers:

- A.(6.1.1): Algorithm 6.1.1 applied to (4.9) without utilizing relations between subproblems, cf. [DS10];
- A.(6.1.4): Algorithm 6.1.4 applied to (4.9) and taking advantage of relations between subproblems (Observation 6.1.3), resulting in substantial reductions of the numbers of subproblems to be inspected;
- C.(4.9): general-purpose MILP solver (CPLEX 12.4.01) applied to representation (4.9);
- C.(4.19): CPLEX applied to the alternative representation (4.19);
- A.(6.1.5): Algorithm 6.1.5 developed for the alternative representation (4.19);
- A.[GNS08]: decomposition method for mixed-integer recourse models from [GNS08].

Computations were carried out on a Linux PC with a 2.67GHz Core i5 processor and 4GB RAM. The time limit for all computations was set to three hours, with "no feas. Sol." meaning that no feasible solution could be found during this time.

K	L	A.(6.1.1)	A.(6.1.4)	C.(4.9)	C.(4.19)	A.(6.1.5)	A.[GNS08]
10	150	1.09	0.24	1.04	1.52	2.58	2.24
	300	4.21	0.62	5.28	60.21	11.73	5.97
	1000	42.54	4.43	175.91	1004.10	49.47	5.83
20	80	19.82	3.72	8.1	19.70	51.50	no feas. Sol.
	100	35.09	4.79	10.71	135.60	153.24	no feas. Sol.
	150	4.11	1.42	19.65	196.15	112	no feas. Sol.
	300	6.89	1.55	9.17	167.36	25.42	0.52

Table 7.2: CPU times in seconds for investment planning instances

From these computational results several conclusions can be drawn:

- Algorithms 6.1.1 and 6.1.4 demonstrate superior performance over direct application of the MILP solver CPLEX to the full model (4.9). Algorithm 6.1.5 is superior over solving (4.19) for the larger instances.
- Solving the formulation (4.19) directly takes significantly more time than the solution of (4.9), which is surprising since the former model has less constraints, see Observation 4.2.3. One explanation could be the more complex structure of the constraints (4.23) in (4.19) which may lead to some combinatorial problems ($\theta_{\ell k}$ are binary).
- Algorithm 6.1.4 demonstrates a significant improvement over Algorithm 6.1.1. The cuts generated in both methods have the same shape, but the former method solves less subproblems and introduces less of these cuts.
- By construction, both Algorithms 6.1.4 and 6.1.5 introduce at most L cuts per iteration. Though the number of generated cuts and solved master problems was less for Algorithms 6.1.5, its performance was weaker. Again, this could be explained with the more complex structure of the

constructed cuts, which seems to cause combinatorial problems for this group of test instances.[1]

- The decomposition method from [GNS08], which is valid for a much more comprehensive class of problems (mixed-integer linear recourse compared to linear recourse), leaves mixed impressions. While the fastest of all for the last test instance, it performs worst of all for the other instances with $K = 20$. The reason is in the different successes of the primal heuristics which are part of the algorithm.[2] Anything from the strong performance to a complete failure seems possible.

To create more demanding test instances we consider a somewhat academic group of test problems derived from Sudoku puzzling, see also [GGS11]. Sudoku is a logic game, which is played over a 9×9 grid, canonically divided into nine 3×3 sub grids. It begins with some of the grid cells already filled with numbers. The task of a Sudoku player is to fill the remaining empty cells with numbers between 1 and 9 (one number only in each cell), such that each number occurs exactly once in each row, each column and each of the nine sub blocks. The Sudoku rules can easily be represented with 729 Boolean variables and a system of linear inequalities (cf. [KK06]).

A two-stage random optimization problem now comes up as follows: the entries on the main diagonal are chosen as first stage decisions. A scenario is formed by a single Sudoku puzzle with a small number of prescribed entries and the property that a solution with joint elements on the main diagonal exists. The random variable $f(x, h(\omega))$ arises as the minimum of the sum of the elements on the secondary diagonal (north-east to south-west) over all feasible Sudokus.

[1] In a cut of the Algorithm 6.1.5 each $\theta_{\ell k}$ may have a coefficient different from zero, whereas the coefficients for all but one $\theta_{\ell k}$ will vanish in Algorithm 6.1.4.

[2] Since we manually tuned the heuristics for this method, the comparison with the other techniques is not completely fair. The numbers should only demonstrate that - given a good heuristic - this more general method may solve even large problems in a competitive time.

To obtain a dominance constrained model (4.1), the objective $g^\top x$ is taken as the sum of the elements on the main diagonal. With a benchmark $d(\omega)$, the constraint $f(x, h(\omega)) \preceq_{(1)} d(\omega),\ x \in \mathcal{X}$ models those first-stage decisions x leading to minimal sums $f(x, h(\omega))$ along the secondary diagonal that are preferable to $d(\omega)$ with respect to the stochastic order $\preceq_{(1)}$. With finite discrete data and benchmark distributions, the dominance constrained stochastic program can be represented as a MILP. In our computations, we relax integrality in the representation of the Sudoku rules, thus arriving at models with linear recourse.

Table 7.3 shows dimensions of the mixed-integer linear programming equivalents (4.9) for the Sudoku-inspired test instances. Table 7.4 reports our computational results.

K	L	Boolean variables	Continuous Variables	Constraints
10	20	200	14581	8331
	50	500	36451	20811
	100	1000	72901	41611
	200	2000	145801	83211
	300	3000	218701	124811
	500	5000	364501	208011

Table 7.3: Dimensions of mixed-integer linear programming equivalents (Sudoku puzzling)

The results displayed in Table 7.4 help clarify the effects observed in the Table 7.2. With the increasing size of test instances the superiority of all proposed decomposition methods becomes apparent. Since the Sudoku-inspired models are less restrictive, combinatorial issues do not play a major role any more, so that the model formulation (4.19) gains attraction together with the corresponding decomposition method 6.1.5. Here, similar numbers of cuts and master problems between the Algorithms 6.1.4 and 6.1.5 imply a very similar performance of both methods. The advantage of Algorithm 6.1.4 over 6.1.1 is

K	L	A. (6.1.1)	A. (6.1.4)	C. (4.9)	C. (4.19)	A. (6.1.5)
10	20	29.29	3.08	1.48	5.40	3.20
	50	70.52	7.14	6.95	34.79	7.25
	100	140.55	14.22	26.99	133.83	14.38
	200	279.93	28.82	312.39	339.10	28.96
	300	561.44	71.70	892.91	1409.21	57.63
	500	924.42	94.94	6385.31	4036.49	94.82

Table 7.4: CPU times in seconds for Sudoku instances

again considerable.

7.3 Second-Order Models

The decision problem, from where our test instances for second-order problems were derived, is that of an electricity retailer who has to fix forward contracting portfolios and selling prices to clients at the beginning of the year. Over the year, the retailer has to satisfy the electricity demand of the customers and may trade at a day-ahead electricity pool market. Customer demand and pool prices are uncertain at the moment of decision about forward portfolios and selling prices. The objective is to maximize the retailer's revenue.

A two-stage random optimization problem arises when taking the decisions on forward contracting portfolios (base and peak contracts) together with selling prices for clients (industrial, commercial, residential) into the first stage and decisions on day-ahead trading as well as delivery to clients into the second. Customer demand and pool prices form the stochastic data. The stochastic benchmark is given by a just acceptable pre-specified profit profile. The dominance constrained stochastic program is completed by the objective which is given as the sum of the selling prices to the different categories of clients. For further details on the modeling background of the retailer problem see [CGS09].

The problem dimensions of the mixed-integer linear programming equivalents (4.14) are listed in Table 7.5.

K	L	Boolean variables	Continuous Variables	Constraints
200	600	20	163265	163592
	800	20	217665	217992
	1000	20	272065	272392
	2000	20	544065	544392

Table 7.5: Dimensions of mixed-integer linear programming equivalents (Energy Retailer Problems)

L	K	A. (6.1.7)	A. (6.2.1)	C. (4.14)	C. (4.24)
600	200	218.12	170.60	no feas. Sol.	236.05
	+30%	119.58	150.39	8645.1	205.80
800	200	383.22	223.74	no feas. Sol.	488.60
	+30%	192.39	247.49	no feas. Sol.	301.24
1000	200	554.78	278.15	OOM	801.25
	+30%	269.31	260.95	OOM	368.39
2000	200	1950.43	551.28	OOM	2897.95
	+30%	670.37	549.22	OOM	1437.06

Table 7.6: CPU times in seconds for Energy Retailer Problems

In Table 7.6 we present computational results for the decomposition methods 6.1.7 and 6.2.1 as opposed to CPLEX applied to the MILPs (4.14) and (4.24) directly.[3]

In our test instances the number of data scenarios varied between 600 and 2000, while the number of benchmark scenarios was 200 throughout. Additional instances were created by introducing a lighter benchmark with the same number of realizations.

[3] "OOM" means that the computations ran out of RAM without having found a feasible solution before.

Table 7.6 shows that the model formulation (4.24) outperforms (4.14) when applying CPLEX directly, an observation made in [Lue08] already. Both decomposition methods, however, are again increasingly attractive with increasing problem sizes compared to all direct applications of CPLEX. The Algorithm 6.2.1 based on the polyhedral representation seems to show a better performance for large instances than the Algorithm 6.1.7 based on the lifting-representation. A similar observation was also made in [KHvdV06], [RR08], [KBM06]. A relation to the more general discussion of aggregate vs. disaggregate methods was drawn in [Fáb12].

Appendix A

A.1 Selected Facts of Probability Theory

A probability space is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the scenario set, $\mathcal{F}$ is a σ-algebra and $\mathbb{P}$ is a probability measure. A (real-valued) random variable X then denotes a real-valued measurable function on such a probability space, i.e., $\mathsf{X} : \Omega \mapsto \mathbb{R}$. Random variables are characterized by their cumulative distribution functions (cdf), which are defined as $F_{\mathsf{X}}(t) := \mathbb{P}(\mathsf{X} \leq t)$. Such functions are only continuous from the right, in general. The closed interval

$$\{t \in \mathbb{R} : P(\mathsf{X} \leq t) \geq p \text{ and } P(\mathsf{X} \geq t) \geq 1 - p\}$$

is called the *set of all p-quantiles of* X, its left end

$$F_{\mathsf{X}}^{-1}(p) := \inf\{t \in \mathbb{R} : P(\mathsf{X} \leq t) \geq p\}$$

is called a *p-quantile*. The function $F_{\mathsf{X}}^{-1} : [0, 1] \to \bar{\mathbb{R}}$ then denotes the (left-continuous) *inverse distribution* or *first quantile function* of a distribution function F_{X}. It is defined as

$$F_{\mathsf{X}}^{-1}(p) := \inf\{t : F_{\mathsf{X}}(t) \geq p\} \text{ for } 0 < p \leq 1.$$

Since F_{X} is continuous from the right, the infimum is attained for $0 < p < 1$. The relations between a cdf and its quantile function can be characterized with the following properties:

- for all $0 < p < 1$, $F_{\mathsf{X}}(F_{\mathsf{X}}^{-1}(p)) \geq p$ and equality holds if p is in the range of F_{X} or, equivalently, if $F_{\mathsf{X}}^{-1}(p)$ is a continuity point of F_{X};
- for all $t \in \mathbb{R}$, $F_{\mathsf{X}}^{-1}(F_{\mathsf{X}}(t)) \leq t$ and equality holds if t is in the range of F_{X}^{-1} or, equivalently, if $F_{\mathsf{X}}(t)$ is a continuity point of F_{X}^{-1};
- $F_{\mathsf{X}}^{-1}(p) \leq t \quad \text{iif} \quad p \leq F_{\mathsf{X}}(t)$.

A.2 Selected Facts of Convex Analysis

Dual stochastic dominance relations were explored in [OR02] with help of some fundamental results of convex analysis, cf. [Roc97] and [BL06].

Definition A.2.1. *For $f : \mathbb{R}^n \supset S \to \mathbb{R} \cup \{\pm\infty\}$ we define the following sets:*

epi $f := \{(x, \alpha) \in S \times \mathbb{R} | f(x) \leq \alpha\}$ as the epigraph of f;

gph $f := \{(x, \alpha) \in S \times \mathbb{R} | f(x) = \alpha\}$ as the graph of f.

Definition A.2.2.

a) $A \subset \mathbb{R}^n$ is said to be convex, if

$$\forall\, x, y \in A,\, \lambda \in (0,1): \quad (1-\lambda)x + \lambda y \,\in\, A \tag{A.1}$$

b) f is said to be convex, if epi f is a convex set;

c) a convex function f is said to be proper, if epi $f \neq \emptyset$, $f(x) < \infty$ for at least one x and $f(x) > -\infty$ for all x;

d) a proper convex function f is said to be closed, if it is lower semi-continuous.

Definition A.2.3. *The Fenchel conjugate of a function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is the function*

$$f^*(x^*) := \sup_{\xi} \{\xi x^* - f(\xi)\}. \tag{A.2}$$

The conjugate function $f^* : \mathbb{R}^n \to \bar{\mathbb{R}}$ is always convex and lower semicontinuous.

A.3 Axiomatic Definitions for Risk and Acceptability Functionals and their Consistency with Stochastic Orders

Acceptability functionals describe preferences of decision makers maximizing gains in the sense that higher values of acceptability yield higher preference. That is why the axioms for these functionals are counterparts to the axioms (R1) - (R3), cf. Section 2.2.

Definition A.3.1. *An acceptability functional $\mathcal{A}$ should comply with the following axioms*

(A1) Monotonicity: $\mathsf{X} \leq \mathsf{Y}$ *a.s. implies that* $\mathcal{A}(\mathsf{X}) \leq \mathcal{A}(\mathsf{Y})$*;*

(A2) Concavity: $\mathcal{A}(t\mathsf{X} + (1-t)\mathsf{Y}) \geq t\mathcal{A}(\mathsf{X}) + (1-t)\mathcal{A}(\mathsf{Y})\, \forall t \in [0,1]$*;*

(A3) Translation Equivariance: $\mathcal{A}(\mathsf{X} + a) = \mathcal{A}(\mathsf{X}) + a\, \forall a \in \mathbb{R}$.

The axiom of positive homogenity (R4) can be defined here in the same way, since it is responsible for stability of the functionals under scaling of the units. Since some important acceptability functionals lack positive homogenity, this property is not required in [RP07].

The axioms of concavity and convexity are connected with the additivity properties responsible for the diversification of risk. A positive homogenous functional is known to be convex (concave) iff it is subadditive (superadditive), cf. [RP07].

For the minimization case, we obtain the axioms of *coherence* in analogy to those developed in [ADEH99] for variables representing gains as follows.

Definition A.3.2. *A coherent risk functional ρ defined on rvs representing losses should comply with*

(C1) Monotonicity: $\mathsf{X} \leq \mathsf{Y}$ *a.s. implies that* $\rho(\mathsf{X}) \leq \rho(\mathsf{Y})$*;*

(C2) Convexity: $\rho(t\mathsf{X} + (1-t)\mathsf{Y}) \leq t\rho(\mathsf{X}) + (1-t)\rho(\mathsf{Y})\, \forall t \in [0,1]$*;*

(C3) Translation Equivariance: $\rho(\mathsf{X} + a) = \rho(\mathsf{X}) + a\, \forall a \in \mathbb{R}$*;*

(C4) Positive Homogenity: $\rho(t\mathsf{X}) = t\rho(\mathsf{X})\, \forall t \geq 0$.

Further, we define *mean-risk dominance* for the minimization case in the same way similarly with the SD rules, i.e., the dominated variable in this relation should be the smaller and hence the better one.

Definition A.3.3. *For random variables* X *and* Y*, we define*

$$\mathsf{X} \preceq_{\mathbb{E},\rho} \mathsf{Y} \text{ iff } \mathbb{E}(\mathsf{X}) \leq \mathbb{E}(\mathsf{Y}) \text{ and } \rho(\mathsf{X}) \leq \rho(\mathsf{Y}) \tag{A.3}$$

and

$$\mathsf{X} \preceq_{\mathbb{E}+\lambda\rho} \mathsf{Y} \text{ iff } \mathbb{E}(\mathsf{X}) + \lambda\rho(\mathsf{X}) \leq \mathbb{E}(\mathsf{Y}) + \lambda\rho(\mathsf{Y}), \tag{A.4}$$

where $\lambda > 0$ *is an assumed degree of risk aversion.*

Definition A.3.4. *The mean-risk model* $(\mathbb{E}, \rho)$ *is said to be consistent with* $\preceq_{(i)}$ *if*

$$\mathsf{X} \preceq_{(i)} \mathsf{Y} \Longrightarrow \mathsf{X} \preceq_{\mathbb{E},\rho} \mathsf{Y} \tag{A.5}$$

and $\lambda-$*consistent with* $\preceq_{(i)}$ *if*

$$\mathsf{X} \preceq_{(i)} \mathsf{Y} \Longrightarrow \mathsf{X} \preceq_{\mathbb{E}+\lambda\rho} \mathsf{Y}, \tag{A.6}$$

for some $\lambda > 0$ *and* $i = 1$ *and ICX.*

As in (2.7) we have

$$\mathsf{X} \preceq_{(i)} \mathsf{Y} \Longrightarrow \mathbb{E}(\mathsf{X}) \leq \mathbb{E}(\mathsf{Y}), \text{ for } i = 1 \text{ and ICX}, \tag{A.7}$$

and the following implications hold true

$$\mathsf{X} \preceq_{\mathbb{E},\rho} \mathsf{Y} \Rightarrow \mathsf{X} \preceq_{\mathbb{E}+\lambda\rho} \mathsf{Y} \Rightarrow \mathsf{X} \preceq_{\mathbb{E}+\alpha\rho} \mathsf{Y}\, \forall \alpha \in [0, \lambda]. \tag{A.8}$$

Since FSD implies ICX, it is again immediate that consistency with ICX implies consistency with FSD. The property

$$\mathsf{X} \preceq_{(1)} \mathsf{Y} \Longrightarrow \rho(\mathsf{X}) \leq \rho(\mathsf{Y}), \text{ for } i = 1 \text{ and ICX} \tag{A.9}$$

is referred to as *isotonicity* of a risk measure with the corresponding dominance relation. In view of Proposition 2.2.1 (iii), the monotonicity axiom (C1) is equivalent to isotonicity with FSD, which in turn directly implies FSD-consistency of all coherent mean-risk models.

Bibliography

[Ace02] C. Acerbi. Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 26:1505–1518, 2002.

[ADEH99] P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Coherent Measures of Risk. *Mathematical Finance*, 9(3), 1999.

[ADH77] J. R. Anderson, J. L. Dillon, and J. B. Hardaker. *Agricultural decision analysis*. Iowa State University Press, 1977.

[Ahm06] S. Ahmed. Convexity and decomposition of mean-risk stochastic programs. *Mathematical Programming*, 106(3):433–446, 2006.

[AL10] B. Armbruster and J. Luedtke. Models and formulations for multivariate dominance constrained stochastic programs. Submitted for publication, available here: http://www.optimization-online.org/DB_FILE/2010/05/2613.pdf, 2010.

[All53] M. Allais. Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. *Econometrica*, 21(4):503–546, 1953.

[Arr65] K. J. Arrow. *Aspects of the theory of risk-bearing*. Yrjö Jahnsson lectures. Yrjö Jahnssonin Säätiö, 1965.

[AT02] C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance*, 26:1487–1503, 2002.

[BAS06] International Convergence of Capital Measurement and Capital Standards. Basel Committee on Banking Supervision, available here: http://www.bis.org/publ/bcbs128.pdf, 2006.

[Baw75] V. S. Bawa. Optimal rules for ordering uncertain prospects. *Journal of Financial Economics*, 2(1):95–121, 1975.

[Baw82] V. S. Bawa. Research bibliography—stochastic dominance: A research bibliography. *Management Science*, 28(6):698–712, June 1982.

[Bea55] E. M. L. Beale. On minimizing a convex function subject to linear inequalities. *Journal of the Royal Statistical Society. Series B (Methodological)*, 17(2):173–184, 1955.

[Ben63] J. F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4:238–252, 1962/63.

[Ber54] D. Bernoulli. Exposition of a new theory on the measurement of risk. *Econometrica*, 22(1):23–36, 1954.

[Ber75] J. Bernoulli. *Die Werke von Jakob Bernoulli, Band 3*. Birkhäuser Verlag Basel, 1975.

[BH06] M. Baucells and F.H. Heukamp. Stochastic dominance and cumulative prospect theory. *Management Science*, 52(9):1409–1423, 2006.

[Big93] J. P. Bigelow. Consistency of mean-variance analysis and expected utility analysis: A complete characterization. *Economics Letters*, 43(2):187–192, 1993.

[BL97] J. R. Birge and F. V. Louveaux. *Introduction to stochastic programming*. Springer-Verlag, New York, 1997.

[BL06] J. M. Borwein and A. S. Lewis. *Convex Analysis and Nonlinear Optimization: Theory and Examples*. CMS books in mathematics. Springer-Verlag New York, 2006.

[Bla51] D. Blackwell. Comparison of experiments. In Jerzy Neyman, editor, *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, pages 93–102. University of California Press, 1951. Available here: http://projecteuclid.org/euclid.bsmsp/1200500222.

[Bla53] D. Blackwell. Equivalent comparisons of experiments. *Ann. Math. Statistics*, 24(2):265–272, 1953. Available here: http://dx.doi.org/10.1214/aoms/1177729032.

[BTBN06] A. Ben-Tal, S. Boyd, and A. Nemirovski. Extending scope of robust optimization: Comprehensive robust counterparts of uncertain problems. *Mathematical Programming*, 107(1-2):63–89, 2006.

[BTN02] A. Ben-Tal and A. Nemirovski. Robust optimization - methodology and applications. *Mathematical Programming*, 92(3):453–480, 2002.

[CCF55] A. Charnes, W. W. Cooper, and R. O. Ferguson. Optimal estimation of executive compensation by linear programming. *Management Science*, 1(2):138–151, 1955.

[CCS58] A. Charnes, W. W. Cooper, and G. H. Symonds. Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil. *Management Science*, 4(3):235–263, 1958.

[CGS09] M. Carrión, U. Gotzes, and R. Schultz. Risk aversion for an electricity retailer with second-order stochastic dominance constraints. *Computational Management Science*, 6(2):233–250, May 2009.

[CHP+09] S. Conti, H. Held, M. Pach, M. Rumpf, and R. Schultz. Shape optimization under uncertaintya stochastic programming perspective. *SIAM Journal on Optimization*, 19(4):1610–1632, January 2009.

[Col07] D. Colander. Retrospectives: Edgeworth's hedonimeter and the quest to measure utility. *Journal of Economic Perspectives*, 21(2):215–226, September 2007.

[CPL13] IBM ILOG CPLEX Optimization Studio V12.4 Documentation. http://pic.dhe.ibm.com/infocenter/cosinfoc/v12r4, 2013.

[CS99] C. C. Carøe and R. Schultz. Dual decomposition in stochastic integer programming. *Operations Research Letters*, 24(1):37–45, 1999.

[Dan55] G. B. Dantzig. Linear programming under uncertainty. *Management Science*, 1(3/4):197–206, 1955.

[DGG+11] D. Drapkin, R. Gollmer, U. Gotzes, F. Neise, and R. Schultz. Risk management with stochastic dominance models in energy systems with dispersed generation. In M. Bertocchi, G. Consigli, and M. A. H. Dempster, editors, *Stochastic Optimization Methods in Finance and Energy*, volume 163 of *International Series in Operations Research and Management Science*, pages 253–271. Springer New York, 2011.

[DHR07] D. Dentcheva, R. Henrion, and A. Ruszczyński. Stability and sensitivity of optimization problems with first order stochastic dominance constraints. *SIAM Journal on Optimization*, 18(1):322–337, 2007.

[DHv02] J. Dupačová, J. Hurt, and J. Štěpán. *Stochastic Modeling in Economics and Finance.* Applied Optimization. Springer, 2002.

[DKS11] D. Drapkin, O. Klaar, and R. Schultz. Model equivalents and cutting-plane decomposition methods for dominance-constrained two-stage stochastic programs. In A. Ridha Mahjoub, editor, *Progress in Combinatorial Optimization*, pages 283 –310. ISTE-Wiley, 2011.

[DM61] G. B. Dantzig and A. Madansky. In J. Neyman, editor, *On the solution of two-stage linear programs under uncertainty*, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, pages 165–176. University of California Press, Berkeley, 1961.

[DM12] Darinka Dentcheva and Gabriela Martinez. Two-stage stochastic optimization problems with stochastic ordering constraints on the recourse. *European Journal of Operational Research*, 219(1):1 – 8, 2012.

[DR03] D. Dentcheva and A. Ruszczyński. Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, 14:548 – 566, 2003.

[DR04a] D. Dentcheva and A. Ruszczyński. Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints. *Math. Program.*, 99(2, Ser. A):329–350, 2004.

[DR04b] D. Dentcheva and A. Ruszczyński. Semi-infinite probabilistic optimization: first-order stochastic dominance constraints. *Optimization*, 53(5-6):583–601, 2004.

[DR05] D. Dentcheva and A. Ruszczyński. Inverse stochastic dominance constraints and rank dependent expected utility theory. Ge, growth, math methods, EconWPA, 2005.

[DR06] D. Dentcheva and A. Ruszczyński. Portfolio optimization with stochastic dominance constraints. *Journal on Banking and Finance*, 30:433–451, 2006.

[DR08] D. Dentcheva and A. Ruszczyński. Stochastic dynamic optimization with discounted stochastic dominance constraints. *SIAM Journal on Control and Optimization*, 47(5):2540–2556, 2008.

[DR09] D. Dentcheva and A. Ruszczyński. Optimization with multivariate stochastic dominance constraints. *Mathematical Programming*, 117(1-2):111–127, 2009.

[DR13] D. Dentcheva and W. Römisch. Stability and sensitivity of stochastic dominance constrained optimization models. *SIAM Journal on Optimization*, 23(3):1672–1688, 2013.

[DS10] D. Drapkin and R. Schultz. An algorithm for stochastic programs with first-order dominance constraints induced by linear recourse. *Discrete Applied Mathematics*, 158(4):291 – 297, 2010. 6th Cologne/Twente Workshop on Graphs and Combinatorial Optimization (CTW 2007).

[DW60] G. B. Dantzig and P. Wolfe. Decomposition principle for linear programs. *Operations Research*, 8:101–111, 1960.

[EE05] T. Ermolieva and Y. Ermoliev. Catastrophic risk management: Flood and seismic risks case studies. In S. W. Wallace and W. T. Ziemba, editors, *Applications of Stochastic Programming*, MPS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, 2005.

[Ehr06] M. Ehrgott. *Multicriteria Optimization*. Springer, 2006.

[Fáb12] C. I. Fábián. Computational aspects of risk-averse optimisation in two-stage stochastic models. *Optimization Online*, August 2012.

[Fis64] P. C. Fishburn. *Decision and value theory.* Publications in operations research. Wiley, 1964.

[Fis76] P. C. Fishburn. Continua of stochastic dominance relations for bounded probability distributions. *Journal of Mathematical Economics*, 3(3):295 – 311, 1976.

[Fis80] P. C. Fishburn. Continua of stochastic dominance relations for unbounded probability distributions. *Journal of Mathematical Economics*, 7(3):271 – 285, 1980.

[FS11] H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time.* De Gruyter Textbook. De Gruyter, 2011.

[FV78] P. C. Fishburn and R. Vickson. Theoretical foundations of stochastic dominance. In G. Whitmore and M. Findlay, editors, *Stochastic dominance: An approach to decision-making under risk.* Lexington Books, D. C. Heath and Company., 1978.

[GG06] Vikas Goel and Ignacio E. Grossmann. A class of stochastic programs with decision dependent uncertainty. *Mathematical Programming*, 108(2-3):355–394, 2006.

[GGS07] R. Gollmer, U. Gotzes, and R. Schultz. Second-order stochastic dominance constraints induced by mixed-integer linear recourse. Number 6 in Stochastic Programming E-Print Series. http://edoc.hu-berlin.de/docviews/abstract.php?id=27962, 2007.

[GGS11] R. Gollmer, U. Gotzes, and R. Schultz. A note on second-order stochastic dominance constraints induced by mixed-integer linear recourse. *Mathematical Programming*, 127:179–190, 2011.

[GLP12] GLPK (GNU Linear Programming Kit). http://www.gnu.org/software/glpk, 2012.

[GNS08] R. Gollmer, F. Neise, and R. Schultz. Stochastic programs with first-order dominance constraints induced by mixed-integer linear recourse. *SIAM Journal on Optimization*, 19:552–571, 2008.

[Got09] Uwe Gotzes. *Decision Making with Dominance Constraints in Two-Stage Stochastic Integer Programming*. Vieweg and Teubner, Germany, 2009.

[GUR13] GUROBI OPTIMIZER. http://www.gurobi.com, 2013.

[HdMM10] T. Homem-de Mello and S. Mehrotra. A cutting-surface method for uncertain linear programs with polyhedral stochastic dominance constraints. *SIAM Journal on Optimization*, 20(3):1250–1273, 2010.

[HJ13] W. Haskell and R. Jain. Stochastic dominance-constrained markov decision processes. *SIAM Journal on Control and Optimization*, 51(1):273–303, 2013.

[HK02] C. Helmberg and K. C. Kiwiel. A spectral bundle method with bounds. *Mathematical Programming*, 93:173–194, 2002.

[HL69] G. Hanoch and H. Levy. The efficiency analysis of choices involving risk. *Review of Economic Studies*, 36(107):335–46, 1969.

[HLM$^+$01] R. Henrion, P. Li, A. Möller, M. C. Steinbach, M. Wendt, and G. Wozny. Stochastic optimization for operating chemical processes under uncertainty. In M. Grötschel, S. O. Krumke, and J. Rambau, editors, *Online Optimization of Large Scale Systems*, pages 457–478. Springer Berlin Heidelberg, 2001.

[HLP34] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. The University Press, 1934.

[HR69] J. Hadar and W. R. Russell. Rules for ordering uncertain prospects. *American Economic Review*, 59(1):25–34, March 1969.

[Kat98] V. J. Katz. *A History of Mathematics: An Introduction*. Addison Wesley Educational Publishers, Inc., 1998.

[KBM06] A. Künzi-Bay and J. Mayer. Computational aspects of minimizing conditional value-at-risk. *Computational Management Science*, 3(1):3–27, 2006.

[KH86] W. K. Klein Haneveld. *Duality in Stochastic Linear and Dynamic Programming*. Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, 1986.

[KHvdV06] W. K. Klein Haneveld and M. H. van der Vlerk. Integrated chance constraints: Reduced forms and an algorithm. *Computational Management Science*, 3(4):245–269, 2006.

[KK06] V. Kaibel and T. Koch. Mathematik für den Volkssport. *DMV-Mitteilungen*, 14:93–96, 2006.

[KT79] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47:263–291, 1979.

[KT92] D. Kahneman and A. Tversky. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.

[KW94] P. Kall and S. W. Wallace. *Stochastic Programming*. Wiley, Chichester etc., 1994.

[KWS97] D. Kahneman, P. P. Wakker, and R. Sarin. Back to Bentham? Explorations of Experienced Utility. *The Quarterly Journal of Economics*, 112(2):375–405, May 1997.

[Leh55] E. L. Lehmann. Ordered families of distributions. *The Annals of Mathematical Statistics*, 26(3):399–419, 1955.

[Lev92] H. Levy. Stochastic dominance and expected utility: Survey and analysis. *Management Science*, 38(4):555–593, 1992.

[Lev06] H. Levy. *Stochastic dominance: investment decision making under uncertainty.* Studies in risk and uncertainty. Springer Science+Business Media Inc., 2006.

[LL02a] M. Levy and H. Levy. Experimental test of the prospect theory value function: A stochastic dominance approach. *Organizational Behavior and Human Decision Processes*, 89(2):1058 – 1081, 2002.

[LL02b] M. Levy and H. Levy. Prospect theory: Much ado about nothing? *Management Science*, 48(10):1334–1349, October 2002.

[LL03] M. Levy and H. Levy. Risk-Aversion, Risk-Seeking, and Subjective Probability Weighting: A Response to Wakker's Critique. Available here: http://people.few.eur.nl/wakker/miscella/levy.levy.2003.reply/reply/reply.of.ll.htm, 2003.

[LS88] F. V. Louveaux and Y. Smeers. Optimal investments for electricity generation: A stochastic model and a test-problem. In Y. Ermoliev and R. J.-B. Wets, editors, *Numerical techniques for stochastic optimization, Springer Series in Comput. Math. 10, 445-453*, 1988.

[LS03] F. V. Louveaux and R. Schultz. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*. Elsevier, 2003.

[Lue07] J. Luedtke. *Integer programming approaches for some non-convex and stochastic optimization problems.* PhD thesis, Georgia Institute of Technology, 2007.

[Lue08] J. Luedtke. New formulations for optimization under stochastic dominance constraints. *SIAM Journal on Optimization*, 19(3):1433–1450, dec 2008.

[LW98] H. Levy and Z. Wiener. Stochastic dominance and prospect dominance with subjective weighting functions. *Journal of Risk and Uncertainty*, 16(2):147–163, 1998.

[Mar52] H. Markowitz. Portfolio Selection. *The Journal of Finance*, 7(1):77–91, March 1952.

[MOA11] A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities: Theory of Majorization and Its Applications.* Springer Series in Statistics. Springer Science+Business Media, LLC, 2011.

[MS02] A. Müller and D. Stoyan. *Comparison Methods for Stochastic Models and Risks.* Wiley, Chichester, 2002.

[MW47] H. B. Mann and D. R. Whitney. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 18(1):50–60, 1947.

[Nei08] F. Neise. *Risk Management in Stochastic Integer Programming: With Application to Dispersed Power Generation.* Vieweg+Teubner research. Vieweg + Teubner, 2008.

[Ogr02] W. Ogryczak. Multiple criteria optimization and decisions under risk. *Control and Cybernetics*, (31):975–1003, 2002.

[OR99] W. Ogryczak and A. Ruszczyński. From stochastic dominance to mean-risk models: Semideviations as risk measures. *European Journal of Operational Research*, 116(1):33–50, 1999.

[OR02] W. Ogryczak and A. Ruszczyński. Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization*, 13(1):60–78 (electronic), 2002.

[Pfl00] G. Ch. Pflug. Some remarks on the value-at-risk and the conditional value-at-risk. In S. P. Uryasev, editor, *Probabilistic Constrained Optimization*, volume 49 of *Nonconvex Optimization and Its Applications*, pages 272–281. Springer US, 2000.

[Phi05] A. B. Philpott. Stochastic optimization and yacht racing. In S. W. Wallace and W. T. Ziemba, editors, *Applications of Stochastic Programming*, MPS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, 2005.

[Pra64] J. W. Pratt. Risk Aversion in the Small and in the Large. *Econometrica*, 32(1/2):122–136, 1964.

[Pré71] A. Prékopa. Logarithmic concave measures with application to stochastic programming. *Acta Scientiarum Mathematicarum*, 32:301–316, 1971.

[Pré95] A. Prékopa. *Stochastic Programming*. Kluwer, Dodrecht, 1995.

[Pré03] A. Prékopa. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*. Elsevier, 2003.

[Pro11] S.D. Promislow. *Fundamentals of Actuarial Mathematics*. Wiley, 2011.

[QS62] J. P. Quirk and R. Saposnik. Admissibility and measurable utility functions. *The Review of Economic Studies*, 29(2):140–146, 1962.

[Qui82] J. Quiggin. A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343, 1982.

[Qui92] J. Quiggin. *Generalized Expected Utility Theory: The Rank Dependent Model*. Kluwer Academic Publishers, 1992.

[RDDM06] D. Roman, K. Darby-Dowman, and G Mitra. Portfolio construction based on stochastic dominance and target return distributions. *Math. Program.*, pages 541–569, 2006.

[Roc97] R. T. Rockafellar. *Convex Analysis*. Princeton landmarks in mathematics and physics series. Princeton University Press, 1997.

[RP07] W. Römisch and G. Ch. Pflug. *Modeling, measuring and managing risk*. World Scientific, New Jersey [u.a.], 2007.

[RR08] G. Rudolf and A. Ruszczyński. Optimization problems with second order stochastic dominance constraints: Duality, compact formulations, and cut generation methods. *SIAM Journal on Optimization*, 19(3):1326–1343, 2008.

[RS69] M. Rothschild and J. E. Stiglitz. Increasing risk: A definition and its economic consequences. Technical report, 1969.

[RS03] A. Ruszczyński and A. Shapiro, editors. volume 10 of *Handbooks in Operations Research and Management Science*. Elsevier, 2003.

[RU00] R. T. Rockafellar and S. P. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

[RU02] R. T. Rockafellar and S. P. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26:1443–1471, 2002.

[Sch93] R. Schultz. Continuity properties of expectation functions in stochastic integer programming. *Math. Oper. Res.*, 18(3):578–589, 1993.

[SDR09] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory.* MPS-SIAM Series on Optimization. SIAM, Philadelphia, 2009.

[Sha64] W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964.

[SP] S&P 500 Stock Market Index. http://www.spindices.com/indices/equity/sp-500.

[SS07] M. Shaked and J. G. Shanthikumar. *Stochastic orders*. Springer Series in Statistics. Springer-Verlag New York Inc., 2007.

[ST04] R. Schultz and S. Tiedemann. Conditional value-at-risk in stochastic programs with mixed-integer recourse. Number 20 in Stochastic Programming E-Print Series. Institut für Mathematik, 2004.

[Ste89] R. E. Steuer. *Multiple Criteria Optimization: Theory, Computation, and Application.* Wiley Series in Probability and Mathematical Statistics. Krieger Publishing Company, 1989.

[Tie05] S. Tiedemann. *Risk Measures with Preselected Tolerance Levels in Two-Stage Stochastic Mixed-Integer programming.* Cuvillier, 2005.

[TL99] A. K. Takyi and B. J. Lence. Surface water quality management using a multiple-realization chance constraint method. *Water Resources Research*, 35(5):1657–1670, 1999.

[VBZE13] M. T. Vespucci, M. Bertocchi, S. Zigrino, and L. F. Escudero. Stochastic optimization models for power generation capacity expansion with risk management. In *10th International Conference on the European Energy Market*, pages 1–8, 2013.

[Vic75] R. G. Vickson. Stochastic Dominance Tests for Decreasing Absolute Risk Aversion. I. Discrete Random Variables. *Management Science*, 21(12):1438–1446, 1975.

[Vic77] R. G. Vickson. Stochastic Dominance Tests for Decreasing Absolute Risk Aversion. II. General Random Variables. *Management Science*, 23(5):478–489, 1977.

[VNM44] J. Von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Science Editions. Princeton University Press, 1944.

[VSW69] R. M. Van Slyke and R. Wets. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17(4):638–663, 1969.

[Wak03] P. P. Wakker. The Data of Levy and Levy (2002), "Prospect Theory Much Ado About Nothing?" Actually Support Prospect Theory. *Management Science*, 2003.

[WF78] G. Whitmore and M. Findlay, editors. *Stochastic dominance: An approach to decision-making under risk.* Lexington Books, D.C. Heath and Company., 1978.

[Whi70] G. A. Whitmore. Third-degree stochastic dominance. *The American Economic Review*, 60(3):457–459, June 1970.

[XPR13] FICO Xpress-Optimizer. http://www.fico.com/en/Products/DMTools/xpress-overview, 2013.

[Yaa87] M. E. Yaari. The dual theory of choice under risk. *Econometrica*, 55(1):95–115, 1987.