Integer Programming Approaches to Risk-Averse Optimization

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Xiao Liu, B.E.

Graduate Program in Industrial and Systems Engineering

The Ohio State University

2016

Dissertation Committee:

Simge Küçükyavuz, Advisor

Güzin Bayraksan, Academic Advisor

Ramteen Sioshansi

Sam Davanloo Tajbakhsh

Abstract

Risk-averse stochastic optimization problems widely exist in practice, but are generally challenging computationally. In this dissertation, we conduct both theoretical and computational research on these problems. First, we study chance-constrained two-stage stochastic optimization problems where second-stage feasible recourse decisions incur additional cost. We also propose a new model, where recovery decisions are made for the infeasible scenarios, and develop strong decomposition algorithms. Our computational results show the effectiveness of the proposed method. Second, we study the static probabilistic lot-sizing problem (SPLS), as an application of a two-stage chance-constrained problem in supply chains. We propose a new formulation that exploits the simple recourse structure, and give two classes of strong valid inequalities, which are shown to be computationally effective. Third, we study two-sided chance-constrained programs with a finite probability space. We reformulate this class of problems as a mixed-integer program. We study the polyhedral structure of the reformulation and propose a class of facet-defining inequalities. We propose a polynomial dynamic programming algorithm for the separation problem. Preliminary computational results are encouraging. Finally, we study risk-averse models for multicriteria stochastic optimization problems. We propose a new model that optimizes the worst-case multivariate conditional value-at-risk (CVaR), and develop a finitely convergent delayed cut generation algorithm.

To my family

Acknowledgments

At this moment I have almost completed my Ph.D. study. This four and a half years have been wonderful and exciting. I would like to express my sincerest gratitude to those who helped and supported me.

First of all I could not ever thank my Ph.D adviser Dr. Simge Kücükyavuz enough. When I came here in 2012, I was a Bachelor student who had weak background in Optimization and almost no experience in conducting original research. And right now I have almost finished my fourth peraper, I have been to many wonderful cities to present our research, and I become very confident in the areas of mixed-integer programing, risk-averse optimization and large-scale optimization. Without Dr. Kücükyavuz, none of these is possible. Dr. Kücükyavuz has taught me so many things, not only in research, but also in other aspects of life. She is dedicated, confident, helping, patient, forgiving and inspiring. To me, she is not only an ideal adviser to work with, but also the exact kind of person that I really want to become. In particular, I greatly appreciate the Graduate Research Assistantship and travel support from her (National Science Foundation grants # 1055668 and #1537317). With her support, I can fully focus on my research and attend many national conferences to present our work and build my career network.

I am also grateful to Dr. Güzin Bayraksan for serving as my co-adviser during the last semester and also as the committee member of my candidacy exam and Ph.D.

dissertation. Her advices about the scope of my research and my presentation skill are invaluable. I also really appreciate her patience for all of my emergency emails, endless paperworks, and other last minute changes happened in this semester.

Moreover, I want to thank Dr. Ramteen Sioshansi and Dr. Sam Davanloo Tajbakhsh for serving as the committee members of my Ph.D dissertation. Their helpful and constructive suggestions really help me to improve my dissertation and my fourth paper. In addition, I would like to thank Dr. Barbara Keyfitz and Dr. Luis Rademacher for their services in my candidacy exam. I would also like to thank all the professors from whom I took courses during my Ph.D. study.

Many thanks to my dear friends and officemates Zhong chen, Mingzhi Leng, Haixi Shi, Yixian Liu, Fei Wu, Hao-Hsiang Wu, Merve Merakli, Kimiebi James, Hamed Rahimian, Cheng-Lung Chen, Huseyin Bahtiyar, Jangho Park. I have enjoyed every moment with them and I really appreciate their help and accompany. Special thanks to Minjiao Zhang and Pelin Damci-Kurt. They helped me a lot when I started out as a Ph.D student. From them, I have learned what are needed to become a successful Ph.D. student.

Particularly, I want to thank my parents, Zhengmin Liu and Xiaoning Shi and my grandparents, Zhaolu Liu and Shugui Liu, and Ruifa Shi and Shuying Wang. In the last four and a half years, I have only spent two months with them. I deeply regret about it, and I know that there is no way I could ever repay them for their love. I have always been so proud of my family, and I will keep fighting for them.

At last, I would like to express my special thanks to my girl friend Junshi Wang. I cherish every moment with her, and I hope this feeling will last forever.

Vita

2012	B.E. Reliability and Systems Engineering Beijing University of Aeronautics and Astronautics, Beijing, China
2016	Ph.D. Operations Research The Ohio State University, Columbus, OH

Publications

Research Publications

X. Liu , S. Küçükyavuz and J. Luedtke "Decomposition Algorithms for Two-Stage Chance-Constrained Programs". *Mathematical Programming*, 157(1), 219-243, 2016.

- (INFORMS Computing Society) Student Paper Prize Runner-Up, 2015.

Fields of Study

Major Field: Industrial and Systems Engineering
Specialization: Operations Research
Minor Fields: Computer Sciences, Mathematics

Table of Contents

Page

Appendices 151

List of Figures

List of Tables

Chapter 1: Introduction

Risk-averse optimization problems involving uncertainty widely exist in practice, but are generally challenging computationally because of their large scale and non-convexity. For example, consider a call center staffing problem in which the customer arrival rate is uncertain. We want to decide the staffing levels of different types of servers in the first stage, before knowing the actual arrival rates of customers, to meet the stochastic demands while minimizing the operating cost. One of the challenges of this problem comes from the uncertainty of the parameters (random arrival rate). One way to handle this difficulty is to use classical two-stage stochastic programming to formulate this problem (see Birge and Louveaux [16] for an introduction to stochastic programming). However, this modeling choice enforces that every possible scenario has to be satisfied by our staffing plan, even including the one that is very unlikely but causes conservative first-stage decision and unnecessarily expensive operation costs.

An alternative modeling choice is introducing a probabilistic (chance) constraint. In the resulting joint chance-constrained model, it is not required to find a staffing plan that satisfies all possible outcomes. Instead, the model enforces that the probability of the staffing plan is successful is at least $1 - \epsilon$, where $0 < \epsilon < 1$ is a user-given risk rate. This modeling choice gives us the flexibility to ignore some of those extreme scenarios, as long as the reliability of our plan is maintained at the desired level $(1-\epsilon)$. In this way, we can greatly reduce the operating cost caused by those unlikely but extreme scenarios.

However, a potential drawback of the joint chance-constrained model is that if the first-stage plan turns out to be unsuccessful, then we do nothing in the second stage. In addition, the level of the infeasibility caused by the first-stage decisions is ignored. To deal with these shortcomings of the joint chance-constrained model, we also consider another measure of risk: Conditional Value-at-Risk. This measure captures the expected shortfall of our decision, and thus provide more information about the performance of our decision in the worst-case scenarios.

In practice, many other problems that involve risk or reliability considerations appear in production planning, power systems, disaster management and homeland security problems. Motivated by these problems, in this dissertation, we conduct both theoretical and computational research on risk-averse optimization problems.

In this dissertation, we use two methods to measure the risk: joint chance constraint and conditional value-at-risk (CVaR). The joint chance constraint, as discussed earlier, is a qualitative risk measure, which ensures that the quality of service of the solution is maintained at a high level. The CVaR, on the other hand, is a quantitative risk measure that captures the magnitude of the risk in the worst cases. For example, let X be the random profit, and let $\mathrm{CVaR}_{0.05}(X) = 100$. This indicates that in the worst 5% scenarios, the expected profit is 100.

In the remainder of this chapter, we provide a brief introduction and literature review of the classical two-stage stochastic programming, and these two risk measures:

joint chance-constraint and CVaR. Then, we introduce the research scope and the outline of this dissertation.

1.1 Two-Stage Stochastic Programs

First, we briefly introduce the classical two-stage stochastic programs (a detailed introduction can be found in [16]). If we consider a stochastic facility location problem: we need to decide the optimal facility location, production level at each facility, and the delivery plan for customers at different locations, in order to satisfy the stochastic demand at each location. In this example, due to the time limit of setting up the production facilities, we cannot wait until the random demand is observed to open the facilities and start producing. Hence, we need to consider two stages in the decision process. In the first stage, we make decisions on the facility locations and the production levels. In the second stage, after the random demand at each location is observed, we make the transportation plan according to the demand realization and the strategic decision from the first stage.

Formally, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. In addition, let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^{n'}$ be the vector of the first-stage and second-stage decision variables, respectively. A generic 2-stage stochastic program is given as follows:

$$\min \mathbf{c}^\top \mathbf{x} + \mathbb{E}_\omega[f(x, \omega)]$$
$$\text{s.t. } \mathbf{x} \in \mathcal{X},$$

where $\mathbf{c}$ is the n-dimensional cost vector for the first-stage decision variables, and $\mathcal{X} \subseteq \mathbb{R}^n$ is the set of constraints for the first-stage decision variables. In addition, the second-stage cost $f(x, \omega)$, for all $\omega \in \Omega$ is given by:

$$
\begin{aligned}
f(x, \omega) = \min\ & \mathbf{d}_\omega^\top \mathbf{y} \\
\text{s.t. } & T_\omega \mathbf{x} + W_\omega \mathbf{y} \geq \mathbf{h}_\omega, \\
& \mathbf{y} \in \mathbb{R}^{n'},
\end{aligned}
$$

where $\mathbf{d}_\omega$ is the cost vector for the second-stage decision variables parameterized by the random variable $\omega \in \Omega$. In addition, T_ω is the technology matrix, W_ω is the recourse matrix, and $\mathbf{h}_\omega$ is the right-hand side vector with appropriate dimensions, that are parameterized by the random variable $\omega \in \Omega$, respectively. If $W_\omega = \binom{\mathbf{I_1}}{-\mathbf{I_2}}$, where $\mathbf{I_1}$ and $\mathbf{I_2}$ are identity matrices with appropriate dimensions, then the second-stage problem has *simple recourse* structure. We refer to [16] for more details about the properties of the generic two-stage stochastic programs. In addition, for solution algorithms, we refer to [98], [11] and [17].

As we can see, the objective of the classical two-stage stochastic program is to minimize the first-stage cost and the expected second-stage cost, while all possible outcomes need to be satisfied. As we discussed earlier, this may lead to very conservative or even infeasible first-stage decisions. In the next section, we briefly introduce the classic single-stage chance-constrained programs, in which we gain the freedom to ignore some of the "extreme" scenarios, while maintaining the quality of the solution at an acceptable level.

1.1.1 Stochastic Program with Simple Recourse

1.2 Chance-constrained Programs

Chance-constrained mathematical programs (CCMPs) aim to find optimal solutions to problems where the probability of an undesirable outcome is limited by a given threshold, ϵ. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. In addition, let $\mathbf{x} \in \mathbb{R}^n$ be the vector of the decision variables, and let $\mathbf{c}$ be its cost vector. A classical single-stage chance-constrained program is stated as follows:

$$
\begin{aligned}
& \min \mathbf{c}^\top \mathbf{x} \\
& \text{s.t. } \mathbb{P}(A\mathbf{x} \geq \mathbf{b}(\omega)) \geq 1 - \epsilon \qquad (1.1a) \\
& \quad x \in \mathbb{R}^n,
\end{aligned}
$$

where $\mathbf{b}(\omega)$ is a m-dimensional random right-hand side vector parameterized by ω, for all $\omega \in \Omega$. In addition, A is a $n \times m$ deterministic matrix. Constraint (1.1a) is the joint chance-constraint which enforces that the probability of our solution to be infeasible should be less than the risk rate ϵ.

The first optimization problem with disjoint chance constraints is defined by Charnes et al. [25]. Charnes and Cooper [24] establish the deterministic equivalent for chance-constrained programs. Miller and Wagner [70] study the mathematical properties of joint chance-constrained programs with independent random variables. Prékopa [80] studies joint probabilistic constraints with dependent random variables and proposes an equivalent deterministic convex program under certain assumptions on the distribution of the random right-hand side.

1.2.1 Deterministic Equivalent Program

Luedtke and Ahmed [66] show that for more general distributions, sample-average approximation (SAA) can be applied to find good feasible solutions and statistical bounds to the original CCMPs. Related studies can also be found in [21, 22, 23, 73]. The resulting sampled problem can be formulated as a large-scale deterministic mixed-integer program by introducing a big-M term for each inequality in the chance constraint and a binary variable for each scenario.

Let $\Omega := \{1, 2, \ldots, N\}$ be the probability space of the finitely sampled problem, and let $\pi_j = \mathbb{P}(\omega = j)$, where $j \in \Omega$ and $\sum_{j=1}^{N} \pi_j = 1$. In addition, let z_j, for all $j \in \Omega$, be the binary variable which indicates that scenario j is satisfied if $z_j = 0$, for all $j \in \Omega$. Furthermore, to simplify notation, let $\mathbf{b}_j = \mathbf{b}(\omega = j)$, for all $j \in \Omega$. Then we can reformulate the generic problem (1.1) as follows:

$$\min \mathbf{c}^\top \mathbf{x}$$

$$\text{s.t. } A\mathbf{x} + Mz_j \geq \mathbf{b}_j \tag{1.2a}$$

$$\sum_{j=1}^{N} \pi_j z_j \geq 1 - \epsilon \tag{1.2b}$$

$$x \in \mathbb{R}^n,$$

where M is a sufficiently large constant to make inequality (1.2b) redundant when $z_j = 1$, for all $j \in \Omega$. Inequalities (1.2a) and (1.2b) represent the joint chance constraints for the finitely sampled problem: if $z_j = 0$, then the constraint set $A\mathbf{x} \geq \mathbf{b}_j$ is enforced for scenario $j \in \Omega$. Otherwise, if $z_j = 1$, then the big-M constant deactivates the constraint set for scenario $j \in \Omega$, which indicates that we do not have to satisfy the constraints of scenario j.

However, the weakness of the linear programming relaxation of this big-M formulation and its large size make it hard to solve. Luedtke et al. [67], Küçükyavuz [52] and Abdi and Fukasawa [1] study strong valid inequalities for the deterministic equivalent formulation of chance-constrained problems with random right-hand sides. An alternative reformulation for this class of problems involves using the concept of $(1 - \epsilon)$-efficient points [81]. Sen [90] studies a disjunctive programming reformulation by using $(1-\epsilon)$-efficient points. Dentcheva et al. [26] give reformulations of CCMPs based on the $(1-\epsilon)$-efficient points, and obtain valid bounds for the objective value. Beraldi and Ruszczyński [13] propose a branch-and-bound algorithm based on the enumeration of the exponentially many $(1-\epsilon)$-efficient points. See also Beraldi and Ruszczyński [12], Ruszczyński [86] and Saxena et al. [88] for algorithms based on the $(1-\epsilon)$-efficient points reformulation. For problems with special structures, formulations that do not involve additional binary variables are developed in Song and Luedtke [95] and Song et al. [94].

There are two potential shortcomings of the joint chance-constrained model. As we can see from formulation (1.2), although the joint chance constraint ensures that the quality of the solution must be maintained at a certain level, it does not provide any information about the magnitude of the violation if the first-stage plan is not successful (in the $\epsilon \times 100\%$ worst case). In addition, if the plan turns out to be unsuccessful, we can do nothing to "recover" from the infeasible solution. In Chapter 2, we propose a two-stage chance-constrained program with recovery, which allows us to recover from a infeasible solution. Furthermore, in the next section, we briefly introduce the concept of CVaR, which measures the "excepted shortfall" of our decision in the worst cases.

1.3 The Conditional Value-at-Risk

Risk measures are functionals that represent the risk associated with a random variable by a scalar value, and provide a direct way to define preference relations between the random outcomes. Among the risk measures that have desirable properties such as coherence [3], CVaR, introduced by [83], has been very popular in a wide range of decision making problems under uncertainty. It is also important to note that it serves as a fundamental building block for a large class of risk measures [56]. In this dissertation, we give the CVaR definition based on acceptability functionals, which follows the lines of [78] and [75]. As a result, larger values of the random variable is preferred (as in random profit).

We present some relevant definitions and relations based on CVaR; for more detailed discussions we refer to [78], [82], and [75].

Definition 1 ([83, 84])**.** *For a random variable X, the* conditional value-at-risk *at confidence level $\alpha \in (0, 1]$ is given by*

$$\mathrm{CVaR}_\alpha(X) = \max\left\{\eta - \frac{1}{\alpha}\mathbb{E}\left([\eta - X]_+\right) \ : \ \eta \in \mathbb{R}\right\}, \tag{1.3}$$

where $[a]_+ = \max\{a, 0\}$, for all $a \in \mathbb{R}$.

For risk-averse decision makers typical choices for the confidence level are small values such as $\alpha = 0.05$. Note that $\mathrm{CVaR}_\alpha(X)$ is concave in X. Suppose X is a random variable with (not necessarily distinct) realizations $x_1, \ldots, x_N$ and corresponding probabilities $p_1, \ldots, p_N$. Then, the optimization problem in (5.1) can equivalently be formulated as the following linear program:

$$\max\{\eta - \frac{1}{\alpha}\sum_{i\in[N]} p_i w_i \ : \ w_i \geq \eta - x_i, \ \forall\, i \in [N], \quad \mathbf{w} \in \mathbb{R}^N_+\}, \tag{1.4}$$

where $[a] = \{1, 2, \ldots, a\}$, for all $a \in \mathbb{Z}_+$.

It is well known that the maximum in definition (5.2) is attained at the α-quantile, which is known as the *value-at-risk* (VaR) at confidence level α:

$$\mathrm{VaR}_\alpha(X) = \min\{\eta \in \mathbb{R} \ : \ F_X(\eta) \geq \alpha\}, \tag{1.5}$$

where F_X is the cumulative distribution function of X.

Moreover, observing (for any given confidence level $\alpha \in (0,1]$) $\mathrm{VaR}_\alpha(X) = x_k$ for at least one $k \in [N]$ provides an alternative expression of CVaR:

$$\mathrm{CVaR}_\alpha(X) = \max_{k\in[N]} \left\{ x_k - \frac{1}{\alpha} \sum_{i\in[N]} p_i [x_k - x_i]_+ \right\}. \tag{1.6}$$

Then, we present the notation of CVaR-preferability for scalar-valued random variables. Let X and Y be two random variables with respective cumulative distribution functions F_X and F_Y. We say that X is CVaR-*preferable* to Y at confidence level α, denoted as $X \succcurlyeq_{\mathrm{CVaR}_\alpha} Y$, if

$$\mathrm{CVaR}_\alpha(X) \geq \mathrm{CVaR}_\alpha(Y). \tag{1.7}$$

Noyan and Rudolf [75] extend the univariate CVaR preference relation to vector-valued random variables by considering a polyhedral scalarization set and requiring that all scalarized versions of the random variables conform to the univariate CVaR-preferability relation; we next provide the formal definition of this *multivariate* CVaR *relation*:

Definition 2 ([75]). *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two* d*-dimensional random vectors,* $C \subset \mathbb{R}^d_+$ *a set of scalarization vectors, and* $\alpha \in (0,1]$ *a specified confidence level. We say that* $\mathbf{X}$ *is* CVaR-preferable *to* $\mathbf{Y}$ *at confidence level* α *with respect to* C*, denoted as* $\mathbf{X} \succcurlyeq^C_{\mathrm{CVaR}_\alpha} \mathbf{Y}$*, if*

$$\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) \geq \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y}) \quad \textit{for all } \mathbf{c} \in C. \tag{1.8}$$

1.4 Research Scope and Outline

In the first part of this dissertation, we study a class of chance-constrained two-stage stochastic optimization problems where second-stage feasible recourse decisions incur additional cost. In addition, we propose a new model, where recovery decisions are made for the infeasible scenarios to obtain feasible solutions to a relaxed second-stage problem. We develop decomposition algorithms with specialized optimality and feasibility cuts to solve this class of problems. Computational results on a chance-constrained resource planing problem indicate that our algorithms are highly effective in solving these problems compared to a mixed-integer programming reformulation and a naive decomposition method.

Next, we study the polyhedral structure of the static probabilistic lot-sizing problem and propose valid inequalities that integrate information from the chance constraint and the binary setup variables. We prove that the proposed inequalities subsume existing inequalities for this problem, and they are facet-defining under certain conditions. In addition, we show that they give the convex hull description of a related stochastic lot-sizing problem. We propose a new formulation that exploits the simple recourse structure, which significantly reduces the number of variables and constraints of the deterministic equivalent program. This reformulation can be applied to general chance-constrained programs with simple recourse. The computational results show that the proposed inequalities and the new formulation are effective for the the static probabilistic lot-sizing problems.

Next, we study two-sided chance-constrained programs with a finite probability space. We reformulate this class of problems as a mixed-integer program. We study the polyhedral structure of the reformulation and propose valid inequalities that define the convex hull of solutions. Furthermore, we propose polynomial optimization and separation algorithms for the optimization problem over a substructure of this problem.

Finally, we study risk-averse models for multicriteria optimization problems under uncertainty. We model the risk aversion of the decision makers via the concept of *multivariate conditional value-at-risk (CVaR)*. We use a weighted sum-based scalarization and take a robust approach by considering a set of scalarization vectors to address the ambiguity and inconsistency in the relative weights of each criterion. First, we introduce a model that optimizes the *worst-case multivariate CVaR*, and develop a finitely convergent delayed cut generation algorithm for finite probability spaces. We also show that this model can be reformulated as a compact linear program under certain assumptions. In addition, for the cut generation problem, which is in general a mixed-integer program, we give a stronger formulation for the equiprobable case. Next, we observe that similar polyhedral enhancements are also useful for a related class of *multivariate CVaR-constrained* optimization problems that has attracted attention recently. In our computational study, we use a budget allocation application to compare the decisions from our proposed maximin type risk-averse model with those from its risk-neutral version and the multivariate CVaR-constrained model. Finally, we illustrate the effectiveness of the proposed solution methods for both classes of models.

The remainder of this dissertation is organized as follows. In Chapter 2, we study the two-stage chance-constrained mathematical programs. In Chapter 3, we explore a special case of two-stage chance-constrained program: static probabilistic lot-sizing problem. In Chapter 4, we investigate the two-sided chance-constrained program. In Chapter 5, we propose the worst-case multivariate Conditional Value-at-Risk model. We conclude this dissertation in Chapter 6.

Chapter 2: Decomposition Algorithm for 2-stage Chance-constrained Programs

2.1 Introduction

This chapter is based on [63]. Most of the earlier work in CCMPs, including the aforementioned studies, can be seen as single-stage (i.e., static) decision-making problems where the decisions are made such that after the uncertain data is realized, there is a low probability of an undesirable outcome. Zhang et al. [107] consider multi-stage CCMPs and give valid inequalities for the deterministic equivalent formulation, and observe that decomposition algorithms are needed to solve large-scale instances of these problems. In this chapter, we study such algorithms for *two-stage* CCMPs, where after the uncertain parameters are revealed, we would like to determine recourse actions that incur additional cost. This is similar to traditional two-stage stochastic programs (*without* chance constraints), where some decisions are made in the first stage before the uncertain parameters are revealed. In the second stage, *recourse* decisions are made to satisfy the second stage problems for all possible scenarios at a minimum total expected cost. Van Slyke and Wets [98] propose the L-shaped decomposition method, which is an adaptation of the Benders decomposition algorithm [11] to such stochastic programs. (See also Birge and Louveaux [17] for a multi-cut implementation.) However, these methods cannot be directly applied to the two-stage CCMP, since both the feasibility and optimality cuts of the Benders method work on the assumption that all second stage problems should be feasible, which is not the case for CCMPs. Luedtke [65] overcomes one of these difficulties by developing a valid "feasibility cut" for a special case of two-stage CCMPs with no additional costs for the second stage variables. A few studies [102, 103] have attempted to integrate Benders decomposition to solve two-stage CCMP, but the optimality cuts in these algorithms involve undesirable "big-M" coefficients, which lead to weak lower bounds and computational difficulties. Hence, an interesting research question is whether we can derive strong optimality cuts for two-stage CCMPs. In this chapter, we answer this question in the affirmative. In a concurrent work, Zeng et al. [106] propose a decomposition algorithm for two-stage CCMP, which is based on bilinear feasibility and optimality cuts. However, these cuts need to be linearized by adding additional variables which are constrained by "big-M" inequalities. This yields a significantly larger master problem formulation, and in our experiments we found that this approach also yielded weak lower bounds, although more promising results were reported in Zeng et al. [106] on their test instances.

Similar to a static CCMP, the two-stage CCMP model assumes that we are allowed to ignore the outcomes of a small fraction of the scenarios. This assumption is appropriate when constraint satisfaction is "all or nothing" or if the magnitude of constraint violation is not important, e.g., if this simply results in lost demands. However, in some other problems the magnitude of violation, and not just the probability that a violation occurs, is also important. For example, in an emergency preparedness problem it is important to have a plan that meets all the needs with high probability, but also that does not have excessive shortages in case the plan is not successful. As another example, a power system operator wishes to have a plan in which all energy supply needs are met with high probability, but also wishes to control the amount of shortage in the cases when a shortage occurs. To deal with such problems, we introduce in §2.2 an alternative model for risk management, which models the need to *recover* from an undesirable outcome. We refer to this model as two-stage CCMP with recovery (CCMPR). We show that a standard two-stage CCMP is a special case of the two-stage CCMPR, and thus algorithms for two-stage CCMPR can be used to solve either problem. In §2.3 we propose a branch-and-cut based decomposition algorithm for two-stage CCMPR based on optimality cuts that do not involve big-M terms. In §3.5, we summarize the performance of the proposed decomposition algorithm on a resource planning example. We conclude with §2.5.

2.2 Mathematical Models

2.2.1 Two-Stage Chance-Constrained Programs with Recourse

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We consider a two-stage problem with first-stage decision variables $x \in X \subseteq \mathbb{R}^{n_1}$, where X is assumed to be a polyhedron representing the deterministic constraints of the problem. For each $x \in X$ and random outcome $\omega \in \Omega$, the second stage problem is defined as:

$$f(x,\, \omega) := \min_{y} \{{q_\omega}^\top y \,:\, T_\omega x + W_\omega y \geq h_\omega,\, y \in \mathbb{R}^{n_2}_{+}\}. \tag{2.1}$$

Here, for each $\omega \in \Omega$, T_ω is a $d \times n_1$ matrix, $h_\omega \in \mathbb{R}^d$, W_ω is a $d \times n_2$ matrix, $q_\omega \in \mathbb{R}^{n_2}_{+}$, and y is the vector of second stage decision variables. We adopt the convention that

$f(x,\omega) = +\infty$ if (2.1) is infeasible. For any $\omega \in \Omega$, we define

$$P(\omega) = \{x \in X \,:\, \exists y \in \mathbb{R}^{n_2}_+, T_\omega x + W_\omega y \geq h_\omega\},$$

as the set of first-stage solutions x for which the second-stage problem (2.1) has a feasible solution (i.e., $f(x,\omega) < +\infty$ if and only if $x \in P(\omega)$).

Given a cost vector $c \in \mathbb{R}^{n_1}$, the traditional two-stage stochastic program has the form:

$$\min\{c^\top x + \mathbb{E}_\omega[f(x,\omega)] \;:\; x \in X\}.$$

This problem implicitly enforces that $x \in P(\omega)$ for almost every $\omega \in \Omega$, since otherwise the objective value is infinite. In contrast, a traditional CCMP without costs in the second stage has the form:

$$\min\{c^\top x \;:\; \mathbb{P}\{x \in P(\omega)\} \geq 1 - \varepsilon, x \in X\}.$$

The motivation behind the chance constraint model is that enforcing $x \in P(\omega)$ for almost every $\omega \in \Omega$ may either lead to an infeasible model, or may lead to a model in which the first-stage solutions are too costly. Thus, the constraint that the second-stage must be feasible in all possible realizations is replaced with the relaxed version that enforces this to hold with high probability. This chance-constrained model does not account for the cost of second-stage solutions in the case when they are feasible.

We propose a two-stage CCMP that extends both the traditional two-stage and chance-constrained models. The model uses indicator decision variables $\mathbb{I}_\omega$ for $\omega \in \Omega$, where $\mathbb{I}_\omega = 0$ implies that $x \in P(\omega)$:

$$\min_{x,\mathbb{I}} c^\top x + \mathbb{P}(\mathbb{I}_\omega = 0)\mathbb{E}_\omega[f(x,\omega) \mid \mathbb{I}_\omega = 0] \tag{2.2a}$$

$$\text{s.t. } \mathbb{I}_\omega = 0 \Rightarrow x \in P(\omega), \quad \omega \in \Omega \tag{2.2b}$$

$$\mathbb{P}\{x \in P(\omega)\} \geq 1 - \varepsilon \tag{2.2c}$$

$$x \in X,\, \mathbb{I}_\omega \in \mathbb{B}, \omega \in \Omega. \tag{2.2d}$$

The idea behind this model is that the recourse model (2.1) represents the "normal" system operation, which is the desired state. The constraint (2.2c) together with the logical condition (2.2b) enforces the chance constraint, which states that the probability that the system has a feasible second stage solution is at least $1 - \varepsilon$. The objective in (2.2) minimizes the sum of the first-stage costs and the expected second stage costs, averaged only over the outcomes ω for which $I_\omega = 0$. This ensures that the objective is finite for any feasible solution.

Throughout the chapter we make the following assumptions:

A1: The random vector ω has finite support, i.e., $\Omega := \{\omega_1, \omega_2, \ldots, \omega_m\}$, and each outcome is equally likely ($\mathbb{P}\{\omega = \omega_k\} = 1/m$ for $k = 1, \ldots, m$).

A2: X and $P(\omega)$ for $\omega \in \Omega$ are non-empty polyhedra;

A3: $P(\omega), \omega \in \Omega$ have the same recession cone, i.e., $C := \{r \in \mathbb{R}^n \;:\; x + \lambda r \in P(\omega)\;; \forall x \in P(\omega),\ \lambda \geq 0\}$ for all $\omega \in \Omega$;

A4: There does not exist an extreme ray, $\tilde{r}$, of X with $c^\top \tilde{r} < 0$,

where **(A1)**-**(A3)** directly follow [65]. To simplify notation, let $P_k = P(\omega_k)$, $W_k = W_{\omega_k}$, $h_k = h_{\omega_k}$, $q_k = q_{\omega_k}$ $T_k = T_{\omega_k}$, $z_k = \mathbb{I}_{\omega_k}$, and $f(x,k) = f(x,\omega_k)$ for all $k \in K$, where $K = \{1, 2, \ldots, m\}$. Assumption **(A4)** together with $f(x,\omega) \geq 0$ (because $q, y \geq 0$) ensures that there is a bounded optimal solution to the two-stage CCMP.

2.2.2 Two-Stage Chance-Constrained Programs with Recovery

Model (2.2) ignores the outcome of scenarios that are not enforced to have a feasible second-stage solution under the normal operations ($\mathbb{I}_\omega = 0$), and instead just enforces that the probability of this selected set of "ignored outcomes" be small. In this section, we extend this model to also include a cost for scenarios in which the normal operation is not enforced to be feasible ($\mathbb{I}_\omega = 1$). The idea is to introduce a secondary "recovery" model, that in some way represents the system operation in cases when we do not operate under the normal operation defined by (2.1). For example, the recovery problem may relax some of the constraints of the normal operational model, by including decision variables that measure and penalize the magnitude of the violation of such constraints. Formally, for any $\omega \in \Omega$ and $x \in X$, we define the recovery operation problem as follows:

$$\bar{f}(x, \omega) = \min_{\bar{y}} \{\bar{q}_\omega^\top \bar{y} \, : \, \bar{T}_\omega x + \bar{W}_\omega \bar{y} \geq \bar{h}_\omega, \, \bar{y} \in \mathbb{R}_+^{\bar{n}_2}\} \tag{2.3}$$

where $\bar{T}_\omega$, $\bar{W}_\omega$ are $\bar{d} \times n_1$ and $\bar{d} \times \bar{n}_2$ matrices, respectively, $\bar{h}_\omega \in \mathbb{R}^{\bar{d}}$, $\bar{q} \in \mathbb{R}_+^{\bar{n}_2}$, and $\bar{y}$ is the vector of recovery decisions. Note that the dimension $\bar{n}_2$ of $\bar{y}$ is not necessarily the same as the dimension n_2 of the recourse decision vector y (similarly for d and $\bar{d}$). For example, the recovery problem may be identical to the normal recourse problem except for the addition of new recovery decision variables that guarantee a feasible solution always exists.

We then introduce the two-stage chance-constrained problem with recovery (CCMPR) model as follows:

$$\min_{x,\,\mathbb{I}}\; c^\top x + \mathbb{P}(\mathbb{I}_\omega = 0)\mathbb{E}_\omega[f(x,\,\omega)\,|\,\mathbb{I}_\omega = 0] + \mathbb{P}(\mathbb{I}_\omega = 1)\mathbb{E}_\omega[\bar{f}(x,\,\omega)\,|\,\mathbb{I}_\omega = 1]$$

$$\text{s.t. } \mathbb{I}_\omega = 0 \Rightarrow x \in P(\omega), \quad \omega \in \varOmega \tag{2.4a}$$

$$\mathbb{P}\{x \in P(\omega)\} \geq 1 - \varepsilon \tag{2.4b}$$

$$x \in X,\, \mathbb{I}_\omega \in \mathbb{B},\; \omega \in \varOmega. \tag{2.4c}$$

The only difference between the CCMPR and CCMP models is the inclusion of the term $\mathbb{E}_\omega[\bar{f}(x,\,\omega)\,|\,\mathbb{I}_\omega = 1]$ in the objective, which captures the expected cost of the recovery operations, conditioned over the scenarios that are selected to operate in recovery mode.

We make the following assumptions in addition to assumptions **(A1)** - **(A4)**:

B1: $\bar{f}(x,\omega) < +\infty$ for all $x \in X, \omega \in \varOmega$.

B2: $f(x,\omega) \geq \bar{f}(x,\omega)$ for all $x \in X, \omega \in \varOmega$.

The assumption **(B1)** is analogous to the standard relatively complete recourse assumption used in two-stage stochastic programs, which we apply *only* to the recovery model. The motivation behind assumption **(B2)** is that the recovery operation has a larger feasible region than the normal operation due to the introduction of *additional* recovery actions. The use of these recovery actions is either highly undesirable, or possibly not even physically meaningful (e.g., if they are just used to measure constraint violation), and so the chance constraint enforces that in most scenarios they should not be used. However, when they are allowed to be used, all the operations of the normal recourse model are still feasible, and hence the cost in the recovery operation can only be smaller. On the other hand, if the recovery actions are not allowed to be used, then the recovery model reduces to the normal model, and incurs the same cost.

Based on assumption **(A1)**, we once again simplify notation by letting $\bar{W}_k = \bar{W}_{\omega_k}$, $\bar{h}_k = \bar{h}_{\omega_k}$, $\bar{q}_k = \bar{q}_{\omega_k}$, $\bar{T}_k = \bar{T}_{\omega_k}$, and $\bar{f}(x,k) = \bar{f}(x,\omega_k)$ for all $k \in K$. Using **(A1)** and this new notation, we can re-write (2.4) as:

$$\min_{x,z} c^\top x + \frac{1}{m}\left(\sum_{k=1}^{m}(1-z_k)f(x,k) + \sum_{k=1}^{m} z_k \bar{f}(x,k)\right) \tag{2.5a}$$

$$\text{s.t. } z_k = 0 \Rightarrow x \in P_k, \quad k \in K \tag{2.5b}$$

$$\sum_{k=1}^{m} z_k \le p \tag{2.5c}$$

$$x \in X,\ z \in \mathbb{B}^m, \tag{2.5d}$$

where $p := \lfloor m\epsilon \rfloor$, and (2.5c) represents the chance constraint (2.2c). Throughout the chapter, we adopt the convention that $0 \times \infty = 0$.

The deterministic equivalent formulation for two-stage CCMPR (2.5) is then:

$$\min_{x,y,\bar{y},z} \quad c^\top x + \frac{1}{m}\sum_{k=1}^{m} q_k^\top y_k + \frac{1}{m}\sum_{k=1}^{m} \bar{q}_k^\top \bar{y}_k \tag{2.6a}$$

$$\text{s.t. } T_k x + W_k y_k + M'_k z_k \ge h_k, \quad k \in K \tag{2.6b}$$

$$\bar{T}_k x + \bar{W}_k \bar{y}_k + \bar{M}_k(1-z_k) \ge \bar{h}_k, \quad k \in K \tag{2.6c}$$

$$\sum_{k=1}^{m} z_k \le p \tag{2.6d}$$

$$x \in X,\ y \in \mathbb{R}_+^{n_2 \times m}, \bar{y} \in \mathbb{R}_+^{\bar{n}_2 \times m}, z \in \mathbb{B}^m, \tag{2.6e}$$

where $M'_k, k \in K$ is a vector of sufficiently large numbers to make (2.6b) redundant when z_k equals to 1, assuming it exists (e.g., when X is compact). Similarly, $\bar{M}_k, k \in K$ is a vector of sufficiently large numbers to make (2.6c) redundant when $z_k = 0$ (assuming it exists). Nonnegativity of the coefficient vector q_k, together with (2.6b) and nonnegativity of the recourse variables y imply that when $z_k = 1$ the normal operation cost term ($q_k^\top y_k$) will be zero, and similarly when $z_k = 0$ the recovery operation cost term ($\bar{q}_k^\top \bar{y}_k$) will be zero. In this chapter, we give a decomposition algorithm with the aim of avoiding the constraints (2.6b)-(2.6c), because they lead to weak LP relaxations and also because of the introduction of a large number of variables $y_k, \bar{y}_k$, $k \in K$.

Remark 1. *In another line of work, an alternative risk-averse two-stage optimization problem is defined where the expectation term in* (2.2a) *is replaced with the conditional value-at-risk (CVaR) (see, for example Miller and Ruszczyński [71] and Noyan [74]). Under the assumption that each scenario is equally likely and that* $\epsilon = \frac{p}{m}$*, such a model would minimize the sum of* p *worst outcomes (using the representation of CVaR in Bertsimas and Brown [14]) and hence will not lead to solutions with second-stage infeasibility for any scenario. In contrast, model* (2.2) *minimizes the sum of* $m - p$ *best outcomes and allows the remaining outcomes to be infeasible with the assumption that such outcomes can be ignored.*

Special cases

We first observe that the two-stage CCMP model (2.2) is a special case of the two-stage CCMPR model (2.4), by setting $\bar{f}(x, \omega) \equiv 0$ for all $x \in X$ and $\omega \in \Omega$. In this case, because $f(x, \omega) \geq 0$ for all $x \in X, \omega \in \Omega$, we immediately have that assumptions **(B1)** and **(B2)** are satisfied.

Another interesting special case of two-stage CCMPR is a penalty-based model. We refer to this class of problems as two-stage CCMP with simple recovery. In this case, the recovery model takes the form:

$$\bar{f}(x,\omega) = \min_{y,u}\{q_\omega^\top y + w_\omega^\top u \ : \ T_\omega x + W_\omega y + D_\omega u \geq h_\omega, y \in \mathbb{R}_+^{n_2}, u \in \mathbb{R}_+^{n_2'}\}, \qquad (2.7)$$

where T_ω, W_ω, h_ω, and q_ω are the data associated with the *normal* operation problem under outcome ω. The vector of decision variables $u \in \mathbb{R}_+^{n_2'}$ can be interpreted as slack variables that are introduced to ensure that (2.7) always has a feasible solution. Thus, the full set of recovery variables is $\bar{y} = (y, u)$, with dimension $\bar{n}_2 = n_2 + n_2'$. The use of the slack variables u is penalized in the objective with the nonnegative cost vector w_ω. D_ω is a $d \times n_2'$ matrix. For example, feasibility of (2.7) can be guaranteed by taking $n_2' = d$ and D_ω to be a $d \times d$ identity matrix. To simplify notation, we let $\bar{W}_\omega = (W_\omega, D_\omega)$ and $\bar{q}_\omega = (q_\omega, w_\omega)$. Hence, if a constraint is violated in the normal model, then the corresponding slack variable equals the shortfall. If D_ω is chosen such that (2.7) is feasible for any $x \in X$, $\omega \in \Omega$, then assumption **(B1)** is satisfied. We next show that assumption **(B2)** is also satisfied, and therefore this is a special case of the two-stage CCMPR model (2.4). For any $x \in X \setminus P(\omega)$, $f(x,\omega) = +\infty$, so the assumption trivially holds. Now consider any $\hat{x} \in P(\omega)$ and $\hat{y}$, the optimal second stage solution with objective $f(\hat{x},\omega)$. So we have $T_\omega \hat{x} + W_\omega \hat{y} \geq h_\omega$. Then according to (2.7), the vector $(\hat{y}, 0)$ is a feasible solution (2.7). Hence, $f(\hat{x},\omega) = q_\omega^\top \hat{y}_\omega = \bar{q}_\omega^\top(\hat{y}_\omega, 0) \geq \bar{f}(x,\omega)$, since $(\hat{y}, 0)$ is a feasible solution to the recovery problem. In Section 3.5 we will report computational experiments with two-stage CCMPR on this special case.

Two-stage consistency

Takriti and Ahmed [97] observe that in two-stage stochastic programs with an objective to minimize a weighted sum of the expected cost and some measure of the variability of costs, it is possible to obtain a model in which a second-stage solution obtained when solving a deterministic equivalent of the two-stage problem is *not* an optimal solution of the second-stage problem. They argue that this inconsistency makes such models inappropriate, and in the two-stage stochastic programming setting they provide conditions on the variability measure that assure this inconsistency does not occur.

In our two-stage model, for a given first-stage solution $x \in X$ and scenario $\omega \in \Omega$, the key question is whether we will operate in the normal model of (2.1), or in the recovery model (2.3). The two-stage model introduces decision variables $\mathbb{I}_\omega$ to distinguish these cases. However, the constraints (2.5b) only enforce that if $\mathbb{I}_\omega = 0$ then the normal operation is feasible, and do not enforce the converse that $\mathbb{I}_\omega = 0$ whenever the normal operation is feasible. Thus, the CCMPR model allows the possibility to operate some scenarios in recovery mode even when they could feasibly be operated with the normal recourse model. Thus, when we are actually solving the second-stage problem for a given $x \in X$ and $\omega \in \Omega$, it may be ambiguous which model should be solved. If the normal model is infeasible, then it is clear that we must solve the recovery model. However, if the normal model is feasible, then we need to have a policy to determine whether this is one of the outcomes that is selected to be operated in the recovery mode.

Model (2.5) determines which outcomes will operate according to the recovery model with a threshold policy. The decision maker first attempts to solve the normal operation problem (2.1). If it is infeasible, the recovery model is implemented. Otherwise, the decision-maker compares the cost of the normal operation $f(x, \omega)$ to a threshold v^*. If the cost exceeds v^*, the decision-maker chooses to operate in recovery mode, otherwise the decision-maker implements the optimal normal operation decision. The value of v^* can be obtained from the optimal solution of the CCMPR model (2.5) by setting $v^* = \max\{f(x,k) : z_k = 0\}$. By construction, this policy of operation in the second-stage is consistent with the solution obtained in the CCMPR model, and for the given sample yields a solution that is feasible to the normal model in at least $1 - \epsilon$ fraction of the scenarios. This policy generalizes the traditional use of a chance constraint, in which the "recovery" model amounts to just ignoring the scenario, and this "recovery" option is only used when the normal model is infeasible. If we adopt the convention that an infeasible problem has infinite cost, then the traditional chance constraint model operates in recovery mode only when the cost of the normal model is infinite. Our model instead operates in recovery mode only when the cost of normal operation exceeds a finite value, v^*.

2.3 Decomposition algorithm for solving two-stage CCMP with Recovery

In this section we propose a decomposition algorithm for two-stage CCMPR. We begin this subsection by analyzing the structure of optimal solutions for (2.5).

Proposition 1. *There exists an optimal solution* (x^*, z^*) *of* (2.5) *in which* $\sum_{k=1}^{m} z_k^* = p$.

Proof. This follows immediately from assumption **(B2)**. □

In other words, there exists an optimal solution where we operate under the "normal" mode for exactly $m-p$ scenarios. Thus, we can replace (2.5c) with the constraint $\sum_{k=1}^{m} z_k = p$.

To introduce the branch-and-cut algorithm, we first define the following sets which will be approximated via cuts:

$$F = \Big\{x \in \mathbb{R}^{n_1},\, z \in \mathbb{B}^m : \sum_{k=1}^{m} z_k = p, z_k = 0 \Rightarrow x \in P_k, k \in K\Big\},$$
$$Z = \Big\{(x, z, \theta) \in F \times \mathbb{R}_+ : \theta \geq \frac{1}{m}\sum_{k=1}^{m}\big((1 - z_k)f(x, k) + z_k\bar{f}(x, k)\big)\Big\}. \tag{2.8}$$

With this notation, solving the problem $\min\{c^\top x + \theta : x \in X, (x, z, \theta) \in Z\}$, where θ is the variable which is used to represent the expected second-stage costs (from both normal and recovery operations), gives an optimal solution (x^*, z^*) to (2.5) satisfying Proposition 1.

The branch-and-cut algorithm we propose works with the following master problem:

$$\mathrm{RP}(K_0, K_1) \ = \min_{x,z,\theta}\ c^\top x + \theta \tag{2.9a}$$
$$\text{s.t.} \sum_{k=1}^{m} z_k = p \tag{2.9b}$$
$$(x, z) \in \mathcal{C} \tag{2.9c}$$
$$(x, z, \theta) \in \mathcal{D} \tag{2.9d}$$
$$x \in X, z \in [0, 1]^m,\, \theta \geq 0 \tag{2.9e}$$
$$z_k = 0,\ k \in K_0;\ z_k = 1,\ k \in K_1, \tag{2.9f}$$

where the sets K_0 and K_1 satisfy $K_0 \cap K_1 = \emptyset$ and represent the sets of variables z_k fixed to zero and to one, respectively, during the branch and bound process. The set $\mathcal{C}$ is a polyhedral outer approximation of F, which is defined by feasibility cuts of the following form:

$$\alpha_1^\top x + \delta_1^\top z \geq \beta_1. \tag{2.10}$$

Here α_1 and δ_1 are n_1 and m dimensional coefficient vectors, respectively, and $\beta_1 \in \mathbb{R}$. The set $\mathcal{D}$ is a polyhedral outer approximation of Z, which is defined by optimality cuts of the form:

$$\theta + \delta^\top z \geq \beta - \alpha^\top x, \tag{2.11}$$

where α and δ are n_1 and m dimensional coefficient vectors, respectively, and $\beta \in \mathbb{R}$.

The key of the decomposition approach is to derive strong valid inequalities for the sets F and Z. A class of strong valid feasibility cuts has been proposed by Luedtke [65] based on the so-called mixing set. We use the same class of cuts in the present chapter. However, Luedtke [65] does not consider second-stage costs, and therefore does not require any optimality cuts. In this chapter, we focus on obtaining strong optimality cuts for two-stage CCMPRs.

We first describe a naive way to obtain valid optimality cuts (2.11). Let $(\hat{x}, \hat{z})$ be such that $\hat{x} \in X$, $\hat{z} \in \mathbb{B}^m$, and $\hat{z}$ satisfies (2.9b). If there exists $k \in K$ with $\hat{z}_k = 0$ and $\hat{x} \notin P_k$, then this solution violates the logical condition (2.5b), and hence we seek and add a feasibility cut. Otherwise, for each $k \in K$ with $\hat{z}_k = 0$, we solve the corresponding normal operation subproblem (which is now feasible by assumption):

$$\min_{y \in \mathbb{R}^{n_2}_+} \{q_k^\top y : W_k y \geq h_k - T_k \hat{x}\} = \max_{\pi \in \mathbb{R}^d_+} \{\pi^\top (h_k - T_k \hat{x}) : \pi^\top W_k \leq q_k\} \tag{2.12}$$

and let π_k be an optimal dual solution. Also let Π_k be the set of dual extreme points in (2.12). In addition, for each $k \in K$ such that $\hat{z}_k = 1$ we solve the recovery problem

$$\min_{\bar{y} \in \mathbb{R}^{\bar{n}_2}_+} \{\bar{q}_k^\top \bar{y} : \bar{W}_k \bar{y} \geq \bar{h}_k - \bar{T}_k \hat{x}\} = \max_{\bar{\pi} \in \mathbb{R}^{\bar{d}}_+} \{\bar{\pi}^\top (\bar{h}_k - \bar{T}_k \hat{x}) : \bar{\pi}^\top \bar{W}_k \leq \bar{q}_k\} \tag{2.13}$$

and let $\bar{\pi}_k$ be an optimal dual solution (let $\bar{\pi}_k = 0$ for problem (2.2) without recovery). Also let $\bar{\Pi}_k$ be the set of dual extreme points in (2.13). Then, if we let $S(\hat{z}) = \{k \in$

$K : \hat{z}_k = 0\}$ and $\bar{S}(\hat{z}) = K \setminus S(\hat{z})$, we obtain the following optimality cut, which is valid for Z:

$$\theta + \sum_{k \in S(\hat{z})} M_k z_k + \sum_{k \in \bar{S}(\hat{z})} M_k (1 - z_k) \geq \frac{1}{m} \left(\sum_{k \in S(\hat{z})} \pi_k^\top (h_k - T_k x) + \sum_{k \in \bar{S}(\hat{z})} \bar{\pi}_k^\top (\bar{h}_k - \bar{T}_k x) \right), \quad (2.14)$$

where $M_k, k \in K$ is assumed to be large enough so that inequality (2.14) is redundant whenever $z_k = 1$ for some $k \in S(\hat{z})$ or $z_k = 0$ for some $k \in \bar{S}(\hat{z})$ (let $M_k = 0$ for $k \in \bar{S}(\hat{z})$ for problem (2.2) without recovery). We refer to inequality (2.14) as the big-M optimality cut. To see its validity, observe that for the solution $(\hat{x}, \hat{z})$, this cut gives a correct lower bound on the second-stage costs (normal and recovery). For any other solution, we must have at least one $k \in S(\hat{z})$ with $z_k = 1$ or $k \in \bar{S}(\hat{z})$ with $z_k = 0$, and hence inequality (2.14) is valid for large enough M_k. Note that additional assumptions may be necessary to ensure the existence of such M_k. Once again, our goal in this chapter is to avoid the use of such big-M cuts.

An alternative approach, recently proposed by Zeng et al. [106] for problems without recovery, is to use a *bilinear* cut of the form:

$$\theta \geq \frac{1}{m} \left(\sum_{k \in S(\hat{z})} (1 - z_k) \pi_k^\top (h_k - T_k x) + \sum_{k \in \bar{S}(\hat{z})} z_k \bar{\pi}_k^\top (\bar{h}_k - \bar{T}_k x) \right). \quad (2.15)$$

To use this cut in a branch-and-cut algorithm, the bilinear terms ($z_k x_j$ for $k \in K$, $j = 1, \ldots, n_1$) need to be linearized by adding additional variables s_{kj} and inequalities to enforce $s_{kj} = z_k x_j$. We experimented with this approach but on our test instances we found its performance to be comparable to the performance of the basic method using the cuts (2.14), and so we do not explore this approach further in this work. We think, however, that investigating the integration of these different techniques would be an interesting subject of future study.

2.3.1 Strong optimality cuts for two-stage CCMPR

In this section we derive valid optimality cuts that are stronger than the big-M optimality cuts (2.14). First, we define a secondary subproblem associated with the normal recourse problem for a given $\alpha \in \mathbb{R}^{n_1}$ and $k \in K$:

$$v_k(\alpha) = \min\{f(x,k) + \alpha^\top x : x \in P_k\} \tag{2.16}$$

so that, by definition,

$$f(x,k) \geq v_k(\alpha) - \alpha^\top x, \quad \forall x \in P_k. \tag{2.17}$$

Problem (2.16) always has a feasible solution from assumption **(A2)**. Let $\mathbf{dom}\ v_k = \{\alpha \in \mathbb{R}^{n_1} : v_k(\alpha) > -\infty\}$ be the domain of v_k.

Similarly, we define a secondary subproblem associated with the recovery problem for a given $\alpha \in \mathbb{R}^{n_1}$ and $k \in K$:

$$\bar{v}_k(\alpha) = \min\{\bar{f}(x,k) + \alpha^\top x : x \in X\} \tag{2.18}$$

so that:

$$\bar{f}(x,k) \geq \bar{v}_k(\alpha) - \alpha^\top x, \quad \forall x \in X. \tag{2.19}$$

Problem (2.18) always has a feasible solution from assumption **(A2)**. Let $\mathbf{dom}\ \bar{v}_k$ be the domain of $\bar{v}_k$.

We make the following additional assumption:

B3: There exists $D \subseteq \mathbb{R}^{n_1}$ such that $\mathbf{dom}\ v_k = \mathbf{dom}\ \bar{v}_k = D$ for all $k \in K$.

Assumption **(B3)** is satisfied with $D = \mathbb{R}^{n_1}$ if X is bounded. In our example application in Section 3.5, this assumption is satisfied with $D = \mathbb{R}^{n_1}_+$.

Next, for each $k \in K$, we define the following set:

$$Z_k = \{x \in X, z_k \in \mathbb{B}, \eta_k \in \mathbb{R}_+ : \eta_k \geq (1 - z_k) f(x, k) + z_k \bar{f}(x, k)\},$$

where η_k represents the objective function value of the second-stage problem for scenario k. Using the relationship $\theta = (1/m) \sum_{k \in K} \eta_k$, valid inequalities for the sets Z_k, $k \in K$ can be used to obtain valid inequalities for the set Z (optimality cuts).

Proposition 2. *Let $k \in K$, $\pi \in \Pi_k$, and $\alpha = T_k{}^\top \pi$. Then $\bar{v}_k(\alpha) > -\infty$ and the following inequality is valid for Z_k:*

$$\eta_k + (\pi^\top h_k - \bar{v}_k(\alpha)) z_k \geq \pi^\top h_k - \alpha^\top x. \tag{2.20}$$

Proof. Let $(x, z_k, \eta_k) \in Z_k$. If $z_k = 0$, then

$$\eta_k \geq f(x, k) \geq \pi^\top h_k - \pi^\top T_k x = \pi^\top h_k - \alpha^\top x$$

by weak duality. It also follows then that

$$v_k(\alpha) \geq \min\{\pi^\top h_k - \alpha^\top x + \alpha^\top x : x \in P_k\} = \pi^\top h_k,$$

which shows that $\alpha \in \mathbf{dom}\ v_k = D$, and thus $\bar{v}_k(\alpha) > -\infty$ by assumption **(B3)**. Finally, if $z_k = 1$, then $\eta_k \geq \bar{f}(x, k) \geq \bar{v}_k(\alpha) - \alpha^\top x$ by (2.19). □

We can obtain an analogous cut from dual solutions of the recovery problem.

Proposition 3. *Let $k \in K$, $\bar{\pi} \in \bar{\Pi}_k$, and $\alpha = \bar{T}_k^\top \bar{\pi}$ ($\bar{\pi} = \alpha = 0$ for problem* (2.2) *without recovery). Then $v_k(\alpha) > -\infty$ and the following inequality is valid for Z_k:*

$$\eta_k + (\bar{\pi}^\top \bar{h}_k - v_k(\alpha))(1 - z_k) \geq \bar{\pi}^\top \bar{h}_k - \alpha^\top x. \tag{2.21}$$

Proof. Analogous to the proof of proposition 2. □

We next discuss how we obtain the dual solutions used in inequality (2.20) or (2.21). Suppose that we have a solution $(\hat{x}, \hat{z})$ and that $\hat{z}_k = 0$ for some scenario $k \in K$. Then, we attempt to solve the normal operation problem (2.12). If it is infeasible (i.e., $\hat{x} \notin P_k$), we must add a feasibility cut. If it is feasible, then we choose π in (2.20) to be an optimal dual solution. Then, by construction, when $\hat{z}_k = 0$, (2.20) enforces $\eta_k \geq f(\hat{x}, k)$. If $\hat{z}_k = 1$ for some scenario $k \in K$, then we solve the recovery problem (2.13) and choose $\bar{\pi}$ in (2.21) to be an optimal dual solution. Again, this choice enforces $\eta_k \geq \bar{f}(\hat{x}, k)$ when $\hat{z}_k = 1$.

We can use inequalities (2.20) and (2.21) in a multi-cut implementation of a Benders-type decomposition algorithm. For a single cut implementation, we have the following corollary.

Corollary 4. *Let $S \subseteq K$, $\pi_k \in \Pi_k$ and $\alpha_k = {T_k}^\top \pi_k$ for $k \in S$, and $\bar{\pi}_k \in \bar{\Pi}_k$ and $\alpha_k = \bar{T}_k^\top \bar{\pi}_k$ ($\bar{\pi}_k = \alpha_k = 0$ for problem (2.2) without recovery) for $k \in K \setminus S$. Then, the following inequality is valid for Z:*

$$\theta + \frac{1}{m}\sum_{k \in S}(\pi^\top h_k - \bar{v}_k(\alpha_k))z_k + \frac{1}{m}\sum_{k \in K \setminus S}(\bar{\pi}_k^\top \bar{h}_k - v_k(\alpha_k))(1 - z_k)$$
$$\geq \frac{1}{m}\sum_{k \in S}\pi^\top h_k + \frac{1}{m}\sum_{k \in K \setminus S}\bar{\pi}_k^\top \bar{h}_k - \frac{1}{m}\sum_{k=1}^{m}\alpha_k^\top x. \qquad (2.22)$$

For a given a solution $(\hat{x}, \hat{z})$ with $\hat{x} \in X$ and $\hat{z} \in \mathbb{B}^m$, and such that the logical constraints (2.5b) are satisfied, we choose $S = \{k \in K : \hat{z}_k = 0\}$ and for $k \in S$ choose π_k to be an optimal dual solution of (2.12), and for $k \in K \setminus S$ choose $\bar{\pi}_k$ to be an optimal dual solution of (2.13).

Next we derive another class of optimality cuts for two-stage CCMPR.

Theorem 5. *Let $S \subset K$ have $|S| = m - p$, $\pi_k \in \Pi_k$ and $\alpha_k = {T_k}^\top \pi_k$ for $k \in S$, and $\bar{\pi}_k \in \bar{\Pi}_k$ and $\alpha_k = \bar{T}_k^\top \bar{\pi}_k$ ($\bar{\pi}_k = \alpha_k = 0$ for problem (2.2) without recovery) for $k \in K \setminus S$. Also define $v_*(\alpha_k) = \min\{v_{k'}(\alpha_k) : k' \in K \setminus S\}$ and $\bar{v}_*(\alpha_k) = \min\{\bar{v}_{k'}(\alpha_k) : k' \in S\}$. Then, the following inequality is valid for Z:*

$$\begin{aligned} \theta \geq & \frac{1}{m} \sum_{k \in S} \big(\pi_k^\top h_k - \alpha_k^\top x + (v_*(\alpha_k) - \pi_k^\top h_k) z_k \big) \\ & + \frac{1}{m} \sum_{k \in K \setminus S} \big(\bar{\pi}_k^\top \bar{h}_k - \alpha_k^\top x + (\bar{v}_*(\alpha_k) - \bar{\pi}_k^\top \bar{h}_k)(1 - z_k) \big). \end{aligned} \tag{2.23}$$

Proof. Let $(x, z, \theta) \in Z$. First, note, from **(B3)**, that $v_*(\alpha_k)$ and $\bar{v}_*(\alpha_k)$ are well-defined since $\alpha_k \in D$ for $k \in K$. We prove the two inequalities:

$$\sum_{k=1}^{m} f(x, k)(1 - z_k) \geq \sum_{k \in S} \big(\pi_k^\top h_k - \alpha_k^\top x + (v_*(\alpha_k) - \pi_k^\top h_k) z_k \big), \tag{2.24}$$

$$\sum_{k=1}^{m} \bar{f}(x, k) z_k \geq \sum_{k \in K \setminus S} \big(\bar{\pi}_k^\top \bar{h}_k - \alpha_k^\top x + (\bar{v}_*(\alpha_k) - \bar{\pi}_k^\top \bar{h}_k)(1 - z_k) \big) \tag{2.25}$$

which establishes the claim from the definition of Z in (2.8).

Let $U = \{k \in K : z_k = 0\}$. By (2.9b), we have $|U| = m - p$, and so $|S \setminus U| = |U \setminus S|$. Let $\sigma : U \setminus S \to S \setminus U$ be a one-to-one mapping between $U \setminus S$ and $S \setminus U$, i.e.,

$\{\sigma_k : k \in U \setminus S\} = S \setminus U$. Then, using (2.17) we obtain

$$
\begin{aligned}
\sum_{k=1}^{m} f(x,k)(1-z_k) &= \sum_{k\in U} f(x,k) \\
&= \sum_{k\in U\cap S} f(x,k) + \sum_{k\in U\setminus S} f(x,k) \\
&\geq \sum_{k\in U\cap S} (\pi_k^\top h_k - \alpha_k^\top x) + \sum_{k\in U\setminus S} f(x,k) \\
&\geq \sum_{k\in U\cap S} (\pi_k^\top h_k - \alpha_k^\top x) + \sum_{k\in U\setminus S} (v_k(\alpha_{\sigma_k}) - \alpha_{\sigma_k}^\top x) \\
&\geq \sum_{k\in U\cap S} (\pi_k^\top h_k - \alpha_k^\top x) + \sum_{k\in U\setminus S} (v_*(\alpha_{\sigma_k}) - \alpha_{\sigma_k}^\top x) && (2.26) \\
&= \sum_{k\in U\cap S} (\pi_k^\top h_k - \alpha_k^\top x) + \sum_{k\in S\setminus U} (v_*(\alpha_k) - \alpha_k^\top x) && (2.27) \\
&= \sum_{k\in S} \big(\pi_k^\top h_k - \alpha_k^\top x + (v_*(\alpha_k) - \pi_k^\top h_k) z_k\big),
\end{aligned}
$$

where (2.26) follows because $U \setminus S \subseteq K \setminus S$ and therefore $v_k(\alpha_{\sigma_k}) \geq v_*(\alpha_{\sigma_k})$ and (2.27) follows from the definition of one-to-one mapping σ. This establishes (2.24).

The arguments for (2.25) are similar. Let $\bar{U} = \{k \in K : z_k = 1\}$ so that $|\bar{U}| = p$ and so $|S \setminus \bar{U}| = |\bar{U} \setminus S|$. Let $\bar{\sigma} : \bar{U} \setminus S \to S \setminus \bar{U}$ be a one-to-one mapping between $\bar{U} \setminus S$ and $S \setminus \bar{U}$. Then, using (2.19), we obtain

$$
\begin{aligned}
\sum_{k=1}^{m} \bar{f}(x,k) z_k &= \sum_{k\in \bar{U}} \bar{f}(x,k) \\
&\geq \sum_{k\in \bar{U}\cap S} (\bar{\pi}_k^\top \bar{h}_k - \alpha_k^\top x) + \sum_{k\in \bar{U}\setminus S} \bar{f}(x,k) \\
&\geq \sum_{k\in \bar{U}\cap S} (\bar{\pi}_k^\top \bar{h}_k - \alpha_k^\top x) + \sum_{k\in \bar{U}\setminus S} (\bar{v}_k(\alpha_{\bar{\sigma}_k}) - \alpha_{\bar{\sigma}_k}^\top x) \\
&\geq \sum_{k\in \bar{U}\cap S} (\bar{\pi}_k^\top \bar{h}_k - \alpha_k^\top x) + \sum_{k\in \bar{U}\setminus S} (\bar{v}_*(\alpha_{\bar{\sigma}_k}) - \alpha_{\bar{\sigma}_k}^\top x) \\
&= \sum_{k\in \bar{U}\cap S} (\bar{\pi}_k^\top \bar{h}_k - \alpha_k^\top x) + \sum_{k\in S\setminus \bar{U}} (\bar{v}_*(\alpha_k) - \alpha_k^\top x) \\
&= \sum_{k\in K\setminus S} \big(\bar{\pi}_k^\top \bar{h}_k - \alpha_k^\top x + (\bar{v}_*(\alpha_k) - \bar{\pi}_k^\top \bar{h}_k)(1-z_k)\big)
\end{aligned}
$$

which establishes (2.25). □

Given a solution $(\hat{x}, \hat{z})$ with $\hat{x} \in X$ and $\hat{z} \in \mathbb{B}^m$, and such that the logical constraints (2.5b) are satisfied, in Theorem 5 we use $S = \{k \in K : \hat{z}_k = 0\}$, and for $k \in S$ we choose π_k to be an optimal dual solution of (2.12), and for $k \in K \setminus S$ we choose $\bar{\pi}_k$ to be an optimal dual solution of (2.13). It is easy to see that if $(\hat{x}, \hat{z}) \in F$, with this choice, (2.23) enforces $\theta \geq \frac{1}{m}\sum_{k\in K}\big(f(\hat{x}, k)(1 - \hat{z}_k) + \bar{f}(\hat{x}, k)\hat{z}_k\big)$.

Based on numerical experiments presented in §2.4.2, inequality (2.23) provides a stronger lower bound than optimality cut (2.22) with faster convergence rate, however, at the cost of solving $2p(m - p)$ single scenario subproblems in order to obtain the values $v_*(\alpha_k)$ for each $k \in S$ and $\bar{v}_*(\alpha_k)$ for each $k \in K \setminus S$. Hence, a strategy is to combine the optimality cuts (2.22) and (2.23).

Note that for the special case without recovery given by (2.2), the last term in inequality (2.23) is eliminated. In this case, we need to solve only $p(m-p)$ secondary subproblems to obtain the values $v_*(\alpha_k)$ for $k \in S$.

2.3.2 Strong optimality cut for random right-hand sides problem

Inequalities (2.22) and (2.23) are valid for two-stage CCMPR for which the randomness appears only in the right-hand-side vectors $h_\omega, \bar{h}_\omega, \omega \in \Omega$. However, we can take advantage of the special structure of this class of problems to obtain optimality cuts with less effort.

For (2.12) and (2.13) with $T_k = T, W_k = W, \bar{T}_k = \bar{T}$ and $\bar{W}_k = \bar{W}$ for all $k \in K$, the corresponding dual subproblems share the same polyhedron for all $k \in K$ and hence the same dual extreme point sets Π for the dual of (2.12) and $\bar{\Pi}$ for the dual of (2.13). Furthermore, for each $\pi \in \Pi$ and $k \in K$, we have

$$f(x,k) \geq \pi^\top h_k - \pi^\top T x, \quad \forall x \in P_k$$

and for each $\bar{\pi} \in \bar{\Pi}$ and $k \in K$

$$\bar{f}(x,k) \geq \bar{\pi}^\top \bar{h}_k - \bar{\pi}^\top \bar{T} x, \quad \forall x \in X.$$

Note that the second stage value approximation function generated by the same dual extreme point $\pi \in \Pi$ and $\bar{\pi} \in \bar{\Pi}$ for different scenarios are parallel planes.

Let $S \subseteq K$ have $|S| = m - p$ (e.g., $S = \{k \in K : \hat{z}_k = 0\}$ for some solution $\hat{z}$), and let $G_k(\pi) = \pi^\top h_k$ for $k \in S$. In addition, define $G_*(\pi) = \min\{G_k(\pi) : k \in K \setminus S\}$, for $\pi \in \Pi$. Also let $\bar{G}_k(\bar{\pi}) = \bar{\pi}^\top \bar{h}_k$ for $k \in K \setminus S$. Similarly, define $\bar{G}_*(\bar{\pi}) = \min\{\bar{G}_k(\bar{\pi}) : k \in S\}$, for $\bar{\pi} \in \bar{\Pi}$. For each $k \in S$, let π_k be an optimal dual solution to (2.12) and for each $k \in K \setminus S$, let $\bar{\pi}_k$ be an optimal dual solution to (2.13) ($\bar{\pi}_k = 0$ for problem (2.2) without recovery). Then the proposed optimality cut is :

$$\begin{aligned}\theta \geq & \frac{1}{m} \sum_{k \in S} (G_k(\pi_k) - \pi_k^\top T x + (G_*(\pi_k) - G_k(\pi_k)) z_k) \\ & + \frac{1}{m} \sum_{k \in K \setminus S} (\bar{G}_k(\bar{\pi}_k) - \bar{\pi}_k^\top \bar{T} x + (\bar{G}_*(\bar{\pi}_k) - \bar{G}_k(\bar{\pi}_k))(1 - z_k)). \end{aligned} \tag{2.28}$$

Theorem 6. *Inequality* (2.28) *is valid for Z in the case when the randomness occurs only in the right-hand-side.*

Proof. Based on the definition of $G_k(\pi_k)$, we have:

$$f(x,k) \geq G_k(\pi_k) - \pi_k^\top T x, \quad \forall x \in P_k.$$

In addition, we have:

$$\bar{f}(x,k) \geq \bar{G}_k(\bar{\pi}_k) - \bar{\pi}_k^\top \bar{T} x, \quad \forall x \in \bar{P}_k.$$

Let $\alpha_k = T^\top \pi_k$ for $k \in S$, and $\alpha_k = \bar{T}^\top \bar{\pi}_k$ for $k \in K \setminus S$. Then, the rest of the proof is identical to the validity of (2.23), where the role of $v_k(\alpha)$ in (2.17) is replaced by $G_k(\pi_k)$, and the role of $\bar{v}_k(\bar{\alpha})$ in (2.19) is replaced by $\bar{G}_k(\bar{\pi}_k)$. □

Deriving inequality (2.28) requires only $O(mp)$ vector multiplications in contrast to $O(mp)$ optimization subproblems, so it requires less computational effort.

2.3.3 Decomposition algorithm for two-stage CCMPRs

In this subsection, we present a branch-and-cut based decomposition algorithm. The algorithm is described in Algorithm 1, and the optimality cut generation procedure OptCuts$(\hat{x}, \hat{z}, \hat{\theta}, \mathcal{D})$ is given in Algorithm 2. The feasibility cut separation procedure SepCuts$(\hat{x}, \hat{z}, \mathcal{C})$ used in Algorithm 1 is the same as that in [65], so we do not discuss the details of SepCuts$(\hat{x}, \hat{z}, \mathcal{C})$.

Theorem 7. *Algorithm 1 converges to an optimal solution of* (2.5) *after finitely many iterations.*

Proof. First, as shown in [65], the number of feasibility cuts (2.10) is finite, and (2.10) always cuts off solutions where $z_k = 0$ but $x \notin P_k$. In addition, there are finitely many optimality cuts (2.22), (2.23) and (2.28), because there are finitely many $\pi_k \in \Pi_k, \bar{\pi}_k \in \bar{\Pi}_k$ (and hence α_k) to consider for all $k \in K$. Also, optimality cuts (2.22), (2.23) and (2.28) do not cut off any $(x, z, \theta) \in Z$ with $\sum_{k \in K} z_k = m - p$.

Next, we show that by using (2.22), (2.23) and (2.28), the algorithm converges to an optimal solution (x^*, z^*). If for a current solution $(\hat{x}, \hat{z})$, $\hat{z} = z^*$, then (2.22), (2.23) and (2.28) reduce to Benders optimality cuts, which correctly capture the cost approximation function for $(\hat{x}, \hat{z})$. Otherwise, (2.22), (2.23) and (2.28) provide a lower bound on θ^*. Hence, the convergence of the algorithm directly follows from the

Algorithm 1: Decomposition algorithm for two-stage CCMPRs

$t \leftarrow 0$, $K_0(0) \leftarrow \emptyset$, $K_1(0) \leftarrow \emptyset$, $\mathcal{C} \leftarrow \mathbb{R}^{n_1 \times m}$, $\mathcal{D} \leftarrow \mathbb{R}^{n_1 \times m \times 1}$, OPEN $\leftarrow \{0\}$, $Ub \leftarrow +\infty$, $Lb \leftarrow -\infty$;
while OPEN $\neq \emptyset$ **do**
  Step 1 : Choose $l \in$ OPEN and let OPEN $\leftarrow$ OPEN $\setminus \{l\}$;
  Step 2 : Process node l;
  while CUTFOUND *and* RCUTFOUND $\neq$ TRUE *and* $Lb < Ub$; **do**
    solve (2.9) ;
    if (2.9) *is infeasible* **then**
      CUTFOUND $\leftarrow$ FALSE ;
    else
      Let $(\hat{x}, \hat{z}, \hat{\theta})$ be an optimal solution to (2.9);
      $Lb \leftarrow$ RP($K_0(l)$, $K_1(l)$);
      if $\hat{z} \in \{0,1\}^m$ **then**
        CUTFOUND $\leftarrow$ SepCuts($\hat{x}, \hat{z}, \mathcal{C}$) ;
        if CUTFOUND $\neq$ TRUE **then**
          CUTFOUND $\leftarrow$ OptCuts($\hat{x}, \hat{z}, \hat{\theta}, \mathcal{D}$) ;
        end
        if CUTFOUND $\neq$ TRUE, **then update** $Ub \leftarrow Lb$
      end
    end
  end
  Step 3 : Branch if necessary;
  if $Lb < Ub$ **then**
    Choose $k \in K$ such that $\hat{z_k} \in$ (0,1);
    $K_0(t+1) \leftarrow K_0(l) \cup \{k\}$, $K_1(t+1) \leftarrow K_1(l)$;
    $K_0(t+2) \leftarrow K_0(l)$, $K_1(t+2) \leftarrow K_1(l) \cup \{k\}$;
    $t \leftarrow t+2$;
    OPEN $\leftarrow$ OPEN $\bigcup\{t+1, t+2\}$
  end
end

Algorithm 2: Procedure: OptCuts$(\hat{x}, \hat{z}, \hat{\theta}, \mathcal{D})$

Input: $\hat{\sigma} = (\hat{x}, \hat{z}, \hat{\theta})$.

Output: If a valid optimality cut for (2.5) is found which is violated by $(\hat{x}, \hat{z}, \hat{\theta})$, then add this cut to $\mathcal{D}$, and returns TRUE. Otherwise returns FALSE.

```
for all k ∈ K with ẑ_k = 0 do
    solve (2.12) to obtain f(x̂, k) and a dual optimal solution π_k;
end
for all k ∈ K with ẑ_k = 1 do
    solve (2.13) to obtain f̄(x̂, k) and a dual optimal solution π̄_k if there is recovery.
    Otherwise, π̄_k = 0. ;
end
if θ̂ < (1/m) Σ_{k=1}^m f(x̂, k)(1 − ẑ_k) + (1/m) Σ_{k=1}^m f̄(x̂, k) ẑ_k then
    CUTFOUND ← TRUE;
    Generate an optimality cut: if the randomness appears only in the right hand-side, then
    use (2.28). Otherwise, use (2.22) and (2.23) with α_k = T_k^⊤ π_k for k with ẑ_k = 0, and
    α_k = T̄_k^⊤ π̄_k for k with ẑ_k = 1 (α_k = 0 if there is no recovery);
else
    CUTFOUND ← FALSE
end
```

convergence result of Benders decomposition algorithm. Finally, since the algorithm uses a branch-and cut procedure to solve the master problem, it processes a finite number of nodes. Thus, it terminates finitely. □

2.4 Application and computational experiments

In this section we test our proposed algorithm on a resource planning problem first with no recovery (2.2) and then with recovery (2.5). We implemented our approach with C using CPLEX 12.4. The subroutines SepCuts$(\hat{x}, \hat{z}, \mathcal{C})$ and OptCuts$(\hat{x}, \hat{z}, \hat{\theta}, \mathcal{D})$ were implemented using the CPLEX lazy constraint callback function. These subroutines are called whenever CPLEX finds an integer candidate solution to the master problem (2.9). All the tests were run on a Windows XP operating system with 2.30 GHz Intel QuadCore processor 2356 (2 cores) with 2GB RAM. Both cores were used for testing the deterministic equivalent formulation and only a single core was used for the decomposition algorithm. A time limit of one hour and a tree memory limit of 500 MB were enforced.

2.4.1 Two-Stage CCMP without Recovery

Here we test our approach on a resource planning problem adapted from [65]. It consists of a set of resources (e.g., server types), denoted by $i \in I := \{1, \ldots, n_1\}$, which can be used to meet demands of a set of customer types, denoted by $j \in J := \{1, \ldots, r\}$.

The problem can be stated as:

$$\begin{aligned} \min_{x \in \mathbb{R}_+^{n_1}, z \in \mathbb{B}^m} \ & c^\top x + \frac{1}{m} \sum_{k=1}^{m} \big((1 - z_k) f(x, k) + z_k \bar{f}(x, k) \big) \\ \text{s.t.} \ & z_k = 0 \Rightarrow x \in P_k, k \in K \\ & \sum_{k=1}^{m} z_k \leq p, \end{aligned}$$

where

$$\begin{aligned} P_k = & \{ x \in \mathbb{R}_+^{n_1} \ : \ \exists y \in \mathbb{R}_+^{n_1 \times r}, \\ & \sum_{j=1}^{r} y_{ij} \leq \rho_{ki} x_i, \, \forall i \in I, \, \sum_{i=1}^{n_1} \mu_{kij} y_{ij} \geq \xi_{kj}, \, \forall j \in J \}. \end{aligned}$$

Here the first stage vector x represents the number of servers to employ, and c is its cost. In this problem, ξ, ρ, μ are random vectors following a finite and discrete joint distribution represented by a set of m equally likely scenarios K, where $\rho_{ki} \in (0, 1]$ represents the utilization rate of server type $i \in I$, $\xi_{kj} \geq 0$ represents the demand of customer type $j \in J$, and $\mu_{kij} \geq 0$ represents the rate of serving customer type j with server type i under scenario $k \in K$. Furthermore, the second stage problem for

$k \in K$ is stated as:

$$f(x,k) = \{ \min_{y \in \mathbb{R}_+^{n_1 \times r}} q_k^\top y \ : \ \sum_{j=1}^{r} y_{ij} \le \rho_{ki} x_i, \, \forall i \in I,$$
$$\sum_{i=1}^{n_1} \mu_{kij} y_{ij} \ge \xi_{kj}, \, \forall j \in J\},$$

where y_{ij} is the second-stage decision variable representing the number of server type i allocated to customer j, and q_{kij} is the unit allocation cost under scenario k. In this section, we assume there is no recovery model, and so $\bar{f}(x,k) \equiv 0$. In Section 2.4.2, we consider the problem with a recovery model.

We generate the parameters c, ρ_k, and μ_k following the scheme in [65] for the same type of resource planning problems. To generate the random demands ξ_k, we first generate a base demand $\bar{\xi}_j$ which follow $N(200, 20)$ for all $j \in J$. Then we let ξ_{kj} follow $N(\bar{\xi}_j, 0.1 \times \bar{\xi}_j)$ for all $k \in K$. We let the second stage cost $q_{kij} = \rho_{ki}$, for all $k \in K$ and $j \in J$, which guarantees that the second stage costs associated with the highly reliable servers are more expensive.

In Table 2.1, we summarize our experiments on problems where only the demand (right-hand side) is uncertain. We compare our "Strong" decomposition algorithm which uses the optimality cuts (2.28), against two other approaches: the deterministic equivalent problem (DEP) (2.6) and the "Basic" decomposition approach which uses the strong feasibility cuts (2.10) and the big-M optimality cuts (2.14). To obtain a valid big-M in inequality (2.14), we let $M_k = \sum_{j=1}^{r} \pi_{kj} \xi_{kj}$, where π_k is the dual vector for the subproblem (2.12) for $k \in K$ at the current optimality cut generation iteration.

In all the tables, each row reports the average of three instances under the same settings: the number of server and customer types (n_1, r), the risk level ϵ, and the number of scenarios m. The Time column reports the average solution time in seconds, and the Gap column reports the average percentage end gap of all instances for each setting, given by $(Ub - Lb)/Ub$, where Ub and Lb are the best upper and lower bounds returned by the algorithm, respectively. The # column reports how many of the three instances are solved to optimality within the time limit. We do not include Time and # columns in a table, if an algorithm reached the time limit for all instances tested. The asterisk (*) indicates that there is at least one instance where the tree memory limit is reached. The dash (-) indicates that no instance is solved to optimality, and that no feasible solution is found by the algorithm within the time limit.

Table 2.1: Result for instances with random demand.

Instances		DEP	Basic	Strong		
(m, ε)	(n_1, r)	`Gap (%)`	`Gap (%)`	#	`Time`	`Gap (%)`
(2000, 0.05)	(5,10)	4.60	2.34	3	166	0
	(10,20)	-	2.93	3	483	0
	(15,30)	-	2.69	3	1106	0
(2500, 0.05)	(5,10)	4.64	2.61	3	279	0
	(10,20)	-	3.08	3	711	0
	(15,30)	-	2.88	2	1819	0.09
(2000, 0.1)	(5,10)	7.1	5.46	3	723	0
	(10,20)	-	5.99	3	1069	0
	(15,30)	-	6.27	3	1032	0
(2500, 0.1)	(5,10)	7.63	5.32	3	641	0
	(10,20)	-	5.79	3	1198	0
	(15,30)	-	6.03	2	2112	0.02

From Table 2.1 we see that the deterministic equivalent formulation is not able to solve any instances to optimality within the time limit. Moreover, it even fails to find a feasible solution for instances of larger sizes. The basic decomposition algorithm which uses the big-M optimality cuts, makes a big improvement over the deterministic equivalent formulation. However, because of the weak lower bound resulting from the big-M optimality cuts, it is still not capable of solving any of the instances within the time limit. The end gaps are between 2-6% for the basic decomposition algorithm after an hour. In contrast, the strong decomposition algorithm, based on the proposed strong optimality cuts, is able to solve most of the instances to optimality. For the two unsolved instances, the average end gap is less than 0.1%.

The only difference between the basic and strong decomposition algorithm is the type of optimality cuts used. To illustrate the benefit of the optimality cuts we propose, we report the number of nodes explored (Node) and the optimality cuts (Cut) added to the master problem in the basic and strong decomposition algorithm in Table 2.2 for the instances in Table 2.1. Observe that the strong decomposition algorithm requires significantly fewer nodes than the basic decomposition algorithm. The number of optimality cuts added for the strong decomposition algorithm is also generally smaller than that for the basic decomposition algorithm. Hence, from Tables 2.1 and 2.2, we conclude that the additional computational effort to generate (2.28) pays off.

In Table 2.3, we report results for instances with random demands, second-stage costs, service and utilization rates. Since these instances are much more challenging, we consider smaller instances. We see that the deterministic equivalent formulation and the basic decomposition algorithm were not able to solve most of the instances. In addition, for the basic decomposition method, the memory used by the branch-and-cut tree exploded very fast as we added the big-M optimality cuts into the master

Table 2.2: Number of optimality cuts and nodes for instances with random demand.

Instances		Basic		Strong	
(m, ε)	(n_1, r)	Node	Cut	Node	Cut
(2000, 0.05)	(5,10)	15257	624	67	17
	(10,20)	9266	474	3	58
	(15,30)	10166	203	104	9
(2500, 0.05)	(5,10)	9174	563	6	34
	(10,20)	9833	427	3	70
	(15,30)	8203	190	20	135
(2000, 0.1)	(5,10)	10097	581	853	181
	(10,20)	13139	263	76	160
	(15,30)	8967	196	61	100
(2500, 0.1)	(5,10)	8275	554	598	93
	(10,20)	11703	266	19	134
	(15,30)	5867	132	9	137

problem, and so most of the instances terminated due to the tree memory limit. On the other hand, the strong decomposition algorithm with proposed optimality cuts (2.23) gives the best results. It solves many more instances to optimality within the time limit. In our implementation we choose not to use inequalities (2.22) for the case of no recovery.

Table 2.4 reports the number of nodes and optimality cuts for both of the decomposition algorithms for instances with random demands, second-stage costs, service and utilization rates. As before, the strong decomposition algorithm requires much fewer optimality cuts and generally fewer number of branch-and-cut nodes to find solutions that are of higher quality. Note that the branch-and-cut nodes reported appear to be smaller in some cases for the basic decomposition algorithm than the strong decomposition algorithm, but this is because the former algorithm terminates prematurely for most of the instances due to the memory limit. Tables 2.3 and 2.4

Table 2.3: Results for instances with random ρ, μ, ξ, q.

Instances		DEP			Basic			Strong		
(m, ε)	(n_1, r)	#	Time	Gap	#	Time	Gap	#	Time	Gap
(400, 0.05)	(5,10)	1	2625	0.62	0	1765*	1.55	3	809	0
	(7,14)	0	3600	1.21	1	1020*	0.50	3	1915	0
(500, 0.05)	(5,10)	0	3600	1.08	0	3600	1.63	3	1379	0
	(7,14)	0	3600	1.37	0	3186*	0.36	2	2006	0.006
(600, 0.05)	(5,10)	0	3600	1.17	0	3067*	1.04	3	1346	0
	(7,14)	0	3600	2.81	0	1352*	0.50	2	2731	0.07
(400, 0.1)	(5,10)	0	3600	1.50	0	3600	3.49	1	2633	0.21
	(7,14)	0	3600	4.09	1	860*	0.68	0	1689	0.68
(500, 0.1)	(5,10)	0	3600	4.52	0	3339*	4.15	1	2528*	0.35
	(7,14)	0	3600	5.73	0	2187*	1.11	1	1831	0.74
(600, 0.1)	(5,10)	0	3600	3.89	0	2639*	1.82	1	2041	0.48
	(7,14)	0	3600	6.32	2	1449	1.00	2	2061	0.56

highlight the benefits of obtaining strong optimality cuts (2.23) despite their high computational requirements to solve $O(mp)$ single scenario subproblems. Note also that we can further take advantage of the special structure of the resource planning problem as suggested in [65] to solve these problems more effectively.

2.4.2 Two-Stage CCMP with Recovery

Here we introduce the recovery version of the probabilistic resource planning problem studied in §2.4.1, where the simple recovery operation is given by

$$\begin{aligned} \bar{f}(x,k) = \{ \min_{y \in \mathbb{R}_+^{n_1 \times r}, u \in \mathbb{R}_+^r} q_k^\top y + w_k^\top u : & \sum_{j=1}^{r} y_{ij} \leq \rho_{ki} x_i, \ \forall i \in I, \\ & \sum_{i=1}^{n_1} \mu_{kij} y_{ij} + u_j \geq \xi_{kj}, \ \forall j \in J\}, \end{aligned}$$

where u_j is a variable that represents the level of outsourcing to cover the shortage in servers due to high demand of customer type $j \in J$. We let w_k, the unit cost of outsourcing, follow $N(3, 0.2)$, which is higher than the unit costs of service q_k.

Table 2.4: Number of optimality cuts and nodes for instances with random ρ, μ, ξ, q.

Instances		Basic		Strong	
(m, ε)	(n_1, r)	Node	Cut	Node	Cut
(400, 0.05)	(5,10)	36117	825	2676	207
	(7,14)	20674	367	14619	301
(500, 0.05)	(5,10)	49404	1361	6681	243
	(7,14)	92401	442	41060	177
(600, 0.05)	(5,10)	33291	996	17977	246
	(7,14)	33280	321	116324	162
(400, 0.1)	(5,10)	44402	880	83383	300
	(7,14)	41287	160	27159	169
(500, 0.1)	(5,10)	31864	878	82179	216
	(7,14)	23154	346	12052	124
(600, 0.1)	(5,10)	41218	394	37493	146
	(7,14)	23743	297	152293	82

In Table 2.5, we report the results for instances with random demands. For this class of problems we use two types of optimality cuts (2.22) and (2.28) in the strong decomposition algorithm, and compare its performance against the deterministic equivalent formulation and the basic decomposition method with the big-M optimality cut. In column 'Strengthened', we collected the results for the decomposition algorithm using the optimality cuts (2.22) only. In column 'Strong', we report the results of the decomposition algorithm which uses the optimality cut (2.28) only. The '-' in the gap column indicates that there is at least one instance where CPLEX failed to find any feasible solution within the time limit. In this case, we report the average end gap only for the instances for which it is available. As we can see, neither the deterministic equivalent nor the basic decomposition algorithm terminated with an acceptable gap. In addition, the basic decomposition algorithm performs worse than DEP for this class of problems. In contrast, the strengthened decomposition algorithm which uses (2.22) as optimality cuts results in much smaller end gaps, and solves a few instances to optimality. The strong decomposition algorithm which uses (2.28) does not perform as well on the problems with recovery as it does for the problems without recovery. It reaches the time limit for all instances, but it still provides the smallest average end gaps for every setting.

Table 2.5: Result for 2-stage CCMPR with random demands only.

Instances		DEP	Basic	Strengthened			Strong
(m, ε)	(n_1, r)	`Gap (%)`	`Gap (%)`	#	`Time`	`Gap (%)`	`Gap (%)`
(800, 0.05)	(5,10)	2.16	4.26	1	2484	1.27	0.19
	(10,20)	5.28	10.29	0	3600	2.44	0.19
(1200, 0.05)	(5,10)	5.05	9.07	1	2432	1.32	0.21
	(10,20)	4.05(-)	8.46	0	3600	2.51	0.24
(800, 0.1)	(5,10)	5.33	10.61	1	2423	0.73	0.26
	(10,20)	5.28	10.79	0	3600	1.45	0.39
(1200, 0.1)	(5,10)	7.32	9.08	1	3566	0.85	0.85
	(10,20)	5.02(-)	14.28	0	3600	1.69	0.60

For the two-stage CCMPR with random ρ, μ, ξ, q, we generated the "expensive" optimality cut (2.23) once every $m \times 0.02$ calls to the OptCuts$(\hat{x}, \hat{z}, \hat{\theta}, \mathcal{D})$ function. For the remaining calls, we use the strengthened big-M optimality cut (2.22). For example, for an instance which has 1000 scenarios, inequality (2.23) was generated once every 20 calls to the procedure OptCuts$(\hat{x}, \hat{z}, \hat{\theta}, \mathcal{D})$. We compare the proposed algorithm against the deterministic equivalent formulation and the basic decomposition algorithm.

As we can see from Table 2.6, for the general two-stage CCMPR problems, the deterministic equivalent formulation and the basic decomposition algorithm which

Table 2.6: Result for two-stage CCMPR with random ρ, μ, ξ, q.

Instances		DEP	Basic	Strong		
(m, ε)	(n_1, r)	`Gap (%)`	`Gap (%)`	#	`Time`	`Gap (%)`
(1000, 0.05)	(5,10)	6.94	6.11	0	3600	1.34
	(10,20)	6.91	8.51	2	1381	0.14
(1200, 0.05)	(5,10)	7.61	6.53	0	3600	1.67
	(10,20)	-	8.79	2	1413	0.20
(1000, 0.1)	(5,10)	10.16	11.05	0	3600	2.62
	(10,20)	7.17	14.33	1	1539*	0.60
(1200, 0.1)	(5,10)	12.63	11.76	0	3600	2.95
	(10,20)	-	14.28	0	2316*	0.80

utilizes the big-M optimality cuts both performed poorly on all instances due to large solution times and end gaps. These instances are difficult for the strong decomposition algorithm as well. In most instances our algorithm reaches the time limit, but the end gaps are less than 3% for all instances. In addition, there are some instances where this algorithm reaches the memory limit. Therefore, developing more efficient algorithms for the general two-stage CCMPR continues to be an interesting research question.

2.5 Conclusion

In this chapter, we study a class of chance-constrained two-stage stochastic optimization problems where second-stage feasible recourse decisions incur additional cost. In addition, "recovery" decisions are made for the infeasible scenarios to obtain feasible solutions to a relaxed second-stage problem. We develop Benders-type decomposition algorithms with specialized optimality and feasibility cuts to solve this class of problems. Computational results on a chance-constrained resource planing problem indicate that our algorithms are highly effective in solving these problems compared to a mixed-integer programming reformulation and a basic decomposition method, especially for the cases where the randomness appears only on the right-hand-side. Even though our description assumes that the first-stage feasible region X is a polyhedron, our algorithm can be extended to the case where there are integer restrictions on the first stage variables. Similarly, we currently add optimality cuts when an integral z is found as an optimal solution to the master problem. An interesting extension is to add the optimality cuts at fractional z encountered during the branch-and-bound algorithm.

Chapter 3: A Polyhedral Study of the Static Probabilistic Lot-Sizing Problem

3.1 Introduction

This chapter is based on [61]. In this chapter, we study the static probabilistic lot-sizing (SPLS) problem. Given a joint probability distribution of random demand over a finite planning horizon, and a service level, $1-\epsilon$, SPLS problem aims to find a production plan at the beginning of the planning horizon (before the random demand is realized), so that the expected total cost of production and inventory is minimized, and the probability of stockout does not exceed ϵ. In this study, we focus on finite probability spaces.

[101] introduce the *deterministic* uncapacitated lot-sizing (ULS) problem (*without* backlogging), which is the problem of finding the optimal plan of production and inventory quantities, to satisfy the demand in each period of the planning horizon on time. The authors propose an $O(n^2)$ algorithm for ULS, where n is the number of time periods in the planning horizon. Improved polynomial algorithms can also be found in [34] and [100]. [6] give a complete linear description of the convex hull of ULS in the original space of variables by the so-called (ℓ, S) inequalities. In addition,

[51] propose an extended formulation for the ULS problem, which gives the complete linear description of the convex hull of solutions to ULS in the extended space.

[79] provide the first polyhedral study of a closely related deterministic ULS problem *with* backlogging (ULSB), in which backorders are allowed in intermediate periods and penalized by shortage costs, and demands must be met at the end of the planning horizon. [54] propose a class of inequalities that generalizes the inequalities of [79] and show that this class of inequalities is enough to give a complete linear description of the convex hull of ULSB. [35] give extended formulations for the deterministic ULSB problem when there is a limit on the number of periods in which shortages occur.

The aforementioned studies assume that the demands are known for each time period of the planning horizon. However, in many applications, these parameters are uncertain, and only the joint probability distribution of these data is available. [37] address a multi-stage stochastic integer programming formulation of the uncapacitated lot-sizing problem under uncertainty. They extend the deterministic (ℓ, S) inequalities to the stochastic case. [36] show that these inequalities are sufficient to describe the convex hull of solutions to the two-period problem [see, also 30]. [2] use the tight extended formulation proposed for the deterministic lot-sizing problem to strengthen the deterministic equivalent formulation of the stochastic lot-sizing problem. [70], [46] and [48] propose dynamic programming algorithms for solving stochastic uncapacitated lot-sizing problems that run in polynomial time in the input size (number of scenarios and time periods).

The stochastic lot-sizing model assumes that we have to satisfy the uncertain demand in each time period for every scenario, which may lead to an over-conservative solution with excessive inventory. As an alternative, for a given service level, $1-\epsilon$, a chance-constrained lot-sizing formulation, referred to as the *static* probabilistic lot-sizing problem (SPLS), ensures that the production schedule, which is determined at the beginning of the planning horizon before the realization of random demands, meets the demands on time with probability at least $1 - \epsilon$. [12] consider a variant of SPLS, where the total expected cost is approximated by eliminating the holding cost and inventory variables from the objective function. The authors propose a branch-and-bound method that relies on a partial enumeration of the so-called p-efficient points [see 81, 90, 26, 86]. [See, also 58, for a more general probabilistic production and distribution planning problem]. The SPLS model *with* the inventory costs is solved using a branch-and-cut algorithm in [52]. [107] propose a *dynamic* probabilistic lot-sizing model, in which the production schedule is updated based on the scenario realization of the previous time periods. We refer the reader to [53] for a survey on deterministic, stochastic and probabilistic lot-sizing models.

Chance-constrained programming (CCP) is a class of optimization problems where the probability of an undesirable outcome is limited by a given threshold, ϵ, (see, e.g., Charnes et al. [25], Charnes and Cooper [24], Miller and Wagner [70], Prékopa [80]). Luedtke and Ahmed [66] propose sample-average approximation (SAA) algorithm for CCPs with general probabilistic distribution [see, also 21, 22, 73, 23]. The resulting sampled problem can be formulated as a large-scale deterministic mixed-integer program. However, the weakness of the linear programming relaxation of this formulation makes it inefficient to solve with commercial integer programming solvers.

For unstructured chance-constrained programs (CCP) with random right-hand sides, Luedtke et al. [67], Küçükyavuz [52] and [1] study strong valid inequalities for the deterministic equivalent formulation of the chance constraint. In addition, Luedtke [65] and [63] propose decomposition algorithms for two-stage CCPs with a finite number of scenarios, which show significant improvement in computational performance when solving the deterministic equivalent formulation of the CCPs. CCPs with other special structures are also studied in [95], [94] and [103].

In this chapter, we provide a polyhedral study of the static probabilistic lot-sizing problem. Different from earlier studies (summarized in Section 3.3.1), we derive a class of valid inequalities that synthesize information from the binary production setup variables and the chance constraint (Section 3.3.2). As a result, we obtain inequalities that are stronger than those considering the chance constraint and lot-sizing structures separately. We prove that our inequalities are facet-defining under certain conditions. Furthermore, we show that they are sufficient to provide the complete linear description of a related stochastic lot-sizing problem. In Section 3.4, we propose a new formulation for SPLS, which greatly reduces the number of variables and constraints of the deterministic equivalent formulation. We also show that the proposed new formulation can be extended to general two-stage chance-constrained programs with simple recourse. Our computational experiments summarized in Section 3.5 show that the proposed methods are effective.

3.2 Problem Formulation

Given a planning horizon with length n, let $N := \{1, \ldots, n\}$. Also, let x_i be the production setup variable and f_i be the fixed cost of production at time period i for all $i \in N$. In addition, let y_i be the production quantity and c_i be the unit cost of production at time period i, for all $i \in N$. Let ξ be the uncertain demand. Throughout, we let $[j] = \{1, 2, \ldots, j\}$, for $j \in \mathbb{Z}_+$.

The generic model of the static probabilistic lot-sizing problem, which is introduced in [12], can be formulated as a two-stage optimization problem. The first-stage problem is stated as:

$$\min \mathbf{f}^\top \mathbf{x} + \mathbf{c}^\top \mathbf{y} + \mathbb{E}_\xi\Big(\Theta_\xi(\mathbf{y})\Big) \tag{3.1a}$$

$$\text{s.t. } \mathbb{P}\big(\sum_{i=1}^{t} y_i \geq \sum_{i=1}^{t} \xi_i,\ t \in N\big) \geq 1 - \epsilon \tag{3.1b}$$

$$y_i \leq M_i x_i, \qquad i \in N \tag{3.1c}$$

$$\mathbf{y} \in \mathbb{R}_+^n, \mathbf{x} \in \mathbb{B}^n, \tag{3.1d}$$

where M_i is a large constant to make (3.1c) redundant when x_i equals to one, for all $i \in N$. Constraint (3.1b) enforces that the probability of violating the demands from time 1 to n should be less than the user-given risk rate ϵ. In addition, $\Theta_\xi(\mathbf{y})$ is the value function of the second-stage problem given by:

$$\Theta_\xi(\mathbf{y}) = \ \min \mathbf{h}^\top \mathbf{s}(\xi) \tag{3.2a}$$

$$s_t(\xi) \geq \sum_{i=1}^{t} (y_i - \xi_i) \qquad t \in N \tag{3.2b}$$

$$\mathbf{s}(\xi) \in \mathbb{R}_+^n, \tag{3.2c}$$

where $\mathbf{s}(\xi)$ is the vector of second-stage inventory variables with nonnegative cost vector $\mathbf{h}$. In addition, constraints (3.2b) together with (3.2c) ensure the correct calculation of the inventory level. Note that the second-stage problem has a simple-recourse structure. [107] propose a related model, in which there may be shortages in the intermediate time periods, but all demand must be satisfied by the end of the planning horizon to meet contractual obligations. Our methods are valid for both variations of SPLS.

Given a finite scenario set $\Omega = \{1, \ldots, m\}$, let π_j be the probability of scenario j, for all $j \in \Omega$. In addition, let d_{ji} be the demand for period i under scenario j, for all $i \in N$ and $j \in \Omega$. Let s_{jt} be the inventory at the end of time period $t \in N$ in scenario $j \in \Omega$, which incurs a unit holding cost h_t. As is common in SAA methods, throughout the rest of this chapter, we assume that each scenario is equally likely, i.e., $\pi_j = \frac{1}{m}$, for all $j \in \Omega$. Letting $k = \lfloor m\epsilon \rfloor$, the deterministic equivalent formulation of SPLS is

$$\min \mathbf{f}^\top \mathbf{x} + \mathbf{c}^\top \mathbf{y} + \frac{1}{m}\sum_{j=1}^{m} \mathbf{h}^\top \mathbf{s}_j \tag{3.3a}$$

$$\text{s.t.} \sum_{i=1}^{t} y_i \geq \sum_{i=1}^{t} d_{ji}(1 - z_j), \qquad t \in N, j \in \Omega \tag{3.3b}$$

$$\sum_{j=1}^{m} z_j \leq k \tag{3.3c}$$

$$y_i \leq M_i x_i, \qquad i \in N \tag{3.3d}$$

$$s_{jt} \geq \sum_{i=1}^{t} (y_i - d_{ji}) \qquad t \in N, j \in \Omega \tag{3.3e}$$

$$\mathbf{s}_j \in \mathbb{R}^n_+, j \in \Omega, \mathbf{y} \in \mathbb{R}^n_+, \mathbf{x} \in \mathbb{B}^n, \mathbf{z} \in \mathbb{B}^m, \tag{3.3f}$$

where we introduce additional logical variable z_j, which equals 0 if the demand in each time period under scenario j is satisfied, and 1 otherwise, for all $j \in \Omega$. In addition, $M_i = \max_{j \in \Omega} D_{jin}$, for all $i \in N$, where $D_{jin} = \sum_{p=i}^{n} d_{jp}$, for all $j \in \Omega$. Furthermore, the cardinality constraint (3.3c) along with the big-M constraint (3.3b) represents the chance constraint in the equal probability case. However, this deterministic equivalent formulation is hard to solve due to the large number of scenario-based variables and constraints, and the big-M type of constraints (3.3b) and (3.3d), which yield weak linear programming relaxations. In the next section, we survey the existing valid inequalities for this class of problems, and then propose new valid inequalities.

3.3 Valid Inequalities

In this section, we propose a class of strong valid inequalities for SPLS that subsume known inequalities for this problem. Before we describe the proposed inequalities, we review existing inequalities for SPLS adapted from the (ℓ, S) inequalities for the deterministic lot-sizing problem, and the mixing inequalities for the deterministic equivalent of chance-constrained programs with random right-hand sides.

3.3.1 Existing Studies

Consider the feasible region of (3.3) in the space of $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ variables. Let $P = \{(\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \mathbb{B}^n \times \mathbb{R}^n_+ \times \mathbb{B}^m \mid \text{(3.3b)} - \text{(3.3d)}\}$. First, note that we can adapt the (ℓ, S) inequalities [6] for the deterministic lot-sizing problem, to obtain the following valid inequalities for its chance-constrained counterpart:

$$\sum_{i \in S} y_i + \sum_{i \in \bar{S}} D_{ji\ell} x_i \geq D_{j1\ell}(1 - z_j), \quad j \in \Omega, \tag{3.4}$$

where $\ell \in N$, $S \subseteq [\ell]$, and $\bar{S} = [\ell] \setminus S$. To see the validity of (3.4), note that if $z_j = 0$, then the demand in each time period of the j-th scenario must be met, and (3.4) reduces to (ℓ, S) inequalities for the j-th scenario. Otherwise, if $z_j = 1$, then the inequality is trivially valid. However, this class of inequalities contains the undesirable big-M terms, which lead to weak linear programming relaxations. Furthermore, they only contain information from a single scenario at a time. We will address the question on the strength of inequalities (3.4) for a special case in Proposition 11. Similarly, we can also apply the modified extended formulation of deterministic uncapacitated lot-sizing problem studied in [51] to the SPLS, with the added big-M terms. However, this simple adaption only uses the information from a single scenario, which may not be strong for the deterministic equivalent program where we have to consider the intersection of the whole scenarios set. In addition, the number of variables and constraints explode as we increase m and n.

Second, since the big-M inequalities (3.3b) contain the mixing structure, we can apply the mixing inequalities to strengthen the linear programming relaxation of (3.3). To simplify notation, let $D_{ji} = D_{j1i}$, for all $i \in N$ and $j \in \Omega$. In addition, for all $i \in N$ and $j \in \Omega$, let σ be a permutation of the scenarios such that $D_{\sigma_{i(1)}i} \geq D_{\sigma_{i(2)}i} \geq \cdots \geq D_{\sigma_{i(m)}i}$, where $D_{\sigma_{i(j)}i}$ is the j-th largest cumulative demand for the i-th time period. To further simplify the notation, let $D_{\sigma_{i(j)}} = D_{\sigma_{i(j)}i}$. Let $T_i^* = \{\sigma_{i(1)}, \sigma_{i(2)}, \ldots, \sigma_{i(k)}\}$, for all $i \in N$. Throughout this chapter, when we define a set such as $T := \{t_1, t_2, \ldots, t_a\}$, it should be understood that a is the cardinality of T.

Proposition 8 (adapted from [65])**.** *For $\ell \in N$, let $T_\ell := \{t_{\ell(1)}, t_{\ell(2)}, \ldots, t_{\ell(a_\ell)}\} \subseteq T_\ell^*$, where $D_{t_{\ell(1)}} \geq D_{t_{\ell(2)}} \geq \cdots \geq D_{t_{\ell(a_\ell)}}$. The basic mixing inequalities*

$$\sum_{i=1}^{\ell} y_i + \sum_{j=1}^{a_\ell} (D_{t_{\ell(j)}} - D_{t_{\ell(j+1)}}) z_{t_{\ell(j)}} \geq D_{t_{\ell(1)}}, \tag{3.5}$$

are valid for P, where $t_{\ell(a_\ell+1)} = \sigma_{\ell(k+1)}$.

[65], [52] and [1] provide extensions of the basic mixing inequalities (3.5) for equal and general probability cases. However, the mixing inequalities based on cumulative production quantities do not provide any strengthening for fractional $\mathbf{x}$. Hence, an interesting research question is whether we can combine the mixing inequalities and the (ℓ, S) inequalities to obtain valid inequalities that cut off fractional $(\mathbf{x}, \mathbf{z})$. Next, we provide an affirmative answer to this question.

3.3.2 New Valid Inequalities

In this section, we propose a class of valid inequalities which subsumes inequality (3.5). In addition, we study the strength of the new inequalities and provide a polynomial separation algorithm, which is exact under certain conditions.

Proposition 9. *For $\ell \in N$, let $S \subseteq [\ell]$, $\bar{S} = [\ell] \setminus S$, and let $T_{i-1} := \{t_{i-1(1)}, t_{i-1(2)}, \ldots, t_{i-1(a_{i-1})}\} \subseteq T^*_{i-1}$, where $D_{t_{i-1(1)}} \geq D_{t_{i-1(2)}} \geq \cdots \geq D_{t_{i-1(a_{i-1})}}$, for all $i \in (\bar{S} \setminus \{1\}) \cup \{\ell+1\}$. In addition, we fix $t_{\ell(1)} = \sigma_{\ell(1)}$. Let $\bar{T} = (\cup_{i \in \bar{S}} T_{i-1}) \cup T_\ell$. The inequality*

$$\sum_{i \in S} y_i + \sum_{i \in \bar{S}} (D_{t_{\ell(1)}} - D_{t_{i-1(1)}}) x_i + \sum_{j \in \bar{T}} \bar{\alpha}_j z_j \geq D_{t_{\ell(1)}} \tag{3.6}$$

is valid for P, where $t_{i-1(a_{i-1}+1)} = \sigma_{i-1(k+1)}$, for all $i \in (\bar{S} \setminus \{1\}) \cup \{\ell+1\}$,

$$\alpha_{ji} = \begin{cases} 0, & \text{if } i = \ell, j \notin T_\ell, \text{ or if } i+1 \in \bar{S}, j \notin T_i, \\ D_{t_{\ell(p)}} - D_{t_{\ell(p+1)}}, & \text{if } i = \ell, j = t_{\ell(p)} \in T_\ell \text{ for some } p \in [a_\ell], \\ D_{t_{i(p)}} - D_{t_{i(p+1)}}, & \text{if } i+1 \in \bar{S}, j = t_{i(p)} \in T_i \text{ for some } p \in [a_i], \end{cases}$$

and $\bar{\alpha}_j = \max\left\{ \max_{i \in \bar{S}} \{\alpha_{j(i-1)}\}, \alpha_{j\ell} \right\}$ for $j \in \bar{T}$.

Proof. Suppose that $x_i = 0$, for all $i \in \bar{S}$. Then inequality (3.6) reduces to:

$$\sum_{i \in S} y_i + \sum_{j \in \bar{T}} \bar{\alpha}_j z_j \geq \sum_{i \in S} y_i + \sum_{j=1}^{a_\ell} (D_{t_{\ell(j)}} - D_{t_{\ell(j+1)}}) z_{t_{\ell(j)}} \geq D_{t_{\ell(1)}},$$

where the first inequality follows from the definition of $\bar{\alpha}_j$, and the second inequality follows from the validity of the mixing inequality (3.5) for time period ℓ when $x_i = 0$ for all $i \in \bar{S}$. Otherwise, let $i' \in \bar{S}$ be the smallest index in $\bar{S}$ such that $x_{i'} = 1$. Then we have:

$$\begin{aligned} &\sum_{i \in S} y_i + \sum_{i \in \bar{S} \setminus \{i'\}} (D_{t_{\ell(1)}} - D_{t_{i-1(1)}}) x_i + \sum_{j \in \bar{T}} \bar{\alpha}_j z_j \\ &\quad \geq \sum_{i \in S} y_i + \sum_{j=1}^{a_{i'-1}} (D_{t_{i'-1(j)}} - D_{t_{i'-1(j+1)}}) z_{t_{i'-1(j)}} \geq D_{t_{i'-1(1)}}, \end{aligned}$$

where the first inequality follows from the nonnegativity of $(D_{t_{\ell(1)}} - D_{t_{i-1(1)}})$ and the definition of $\bar{\alpha}_j$, since $t_{\ell(1)} = \sigma_{\ell(1)}$. In addition, the second inequality follows from the validity of mixing inequality (3.5) for period i' given that $y_i = 0$ for $i \in \bar{S}, i < i'$. $\square$

Example 1. *Let $k = 2$, $n = 2$, $m = 5$, and consider the demand data given in Table 3.1.*

Table 3.1: Data for Example 1.

Scenarios	1	2	3	4	5
d_1	6	3	1	2	4
d_2	1	6	10	8	5
$d_1 + d_2$	7	9	11	10	9

For $\ell = 2$, let $T_\ell = \{3, 4\}$, $T_{\ell-1} = T_1 = \{1, 5\}$, $S = \{1\}$, and $\bar{S} = \{2\}$. According to the definition, $\bar{T} = \{1, 3, 4, 5\}$. Since $1 \in \bar{T}$, and scenario 1 is the scenario with the largest demand in the first time period, $\alpha_{11} = D_{t_{1(1)}} - D_{t_{1(2)}} = D_{\sigma_{1(1)}} - D_{\sigma_{1(2)}} = 6 - 4 = 2$. In addition, since $1 \notin T_\ell$, we have $\alpha_{12} = 0$. Hence, we have $\bar{\alpha}_1 = \max\{\alpha_{11}, \alpha_{12}\} = 2$. Since $3 \in \bar{T}$, and $3 \notin T_1$, we have $\alpha_{31} = 0$. In addition, since $3 \in T_\ell$, and it is the scenario with the largest cumulative demand at time period ℓ, we have $\alpha_{32} = D_{t_{2(1)}} - D_{t_{2(2)}} = D_{\sigma_{2(1)}} - D_{\sigma_{2(2)}} = 11 - 10 = 1$. Hence, we have $\bar{\alpha}_3 = \max\{\alpha_{31}, \alpha_{32}\} = 1$. Similarly, $\bar{\alpha}_4 = \max\{\alpha_{41}, \alpha_{42}\} = \max\{0, D_{t_{2(2)}} - D_{t_{2(3)}}\} = \max\{0, D_{\sigma_{2(2)}} - D_{\sigma_{2(3)}}\} = 1$, and $\bar{\alpha}_5 = \max\{\alpha_{51}, \alpha_{52}\} = \max\{D_{t_{1(2)}} - D_{t_{1(3)}}, 0\} = \max\{D_{\sigma_{1(2)}} - D_{\sigma_{1(3)}}, 0\} = 1$. Hence, the proposed inequality for this choice of parameters is:

$$y_1 + 5x_2 + 2z_1 + z_3 + z_4 + z_5 \geq 11. \tag{3.7}$$

Next we show the strength of the proposed inequalities.

Proposition 10. *Inequalities* (3.6) *are facet-defining for conv*(P) *if*

1. $\bar{S} \neq \emptyset$ *and* $1 \in S$;

2. $T_{p-1} \cap (T^*_{q-1} \cup T^*_{\ell}) = \emptyset$, $T_\ell \cap T^*_{q-1} = \emptyset$, *for all* $p \neq q$, *and* $p, q \in \bar{S}$;

3. $D_{t_{\ell(1)}} - D_{t_{i-1(a_{i-1}+1)}} < M_i$, *for all* $i \in \bar{S}$;

4. $d_{ji} > 0$, *for all* $j \in \Omega$ *and* $i \in N$.

Proof. First, note that $dim(P) = 2n + m - 1$, assuming that $d_{ji} > 0$, for all $j \in \Omega$ and $i \in N$, since $x_1 = 1$ when demands are positive at period 1, and backordering is not allowed in $n - k$ scenarios. To show that inequality (3.6) is facet-defining under conditions *(i)*-*(iv)*, we need to find $2n + m - 1$ affinely independent points $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ that satisfy (3.6) at equality. Let $g(t_{i(j)})$, for all $t_{i(j)} \in T_i$ and $i+1 \in \bar{S} \cup \{\ell+1\}$, be a unique mapping such that scenario $t_{i(j)}$ has the $g(t_{i(j)})$-th largest cumulative demand at time period i. Also, for $p \in [\ell]$, $i \in \bar{T}$ and $j \in [a_i + 1]$, let $\bar{\mathbf{y}}^p_j$ be an n-dimensional vector such that $\bar{y}^p_{j1} = D_{t_{p(j)}}$ and $\bar{y}^p_{ji} = 0$, for all $i = 2, \ldots, n$.

First, consider the feasible points: $(\mathbf{e}_1 + \mathbf{e}_{\ell+1}, \bar{\mathbf{y}}^\ell_j + \mathbf{e}_{\ell+1} M_{\ell+1}, \sum_{i=1}^{g(t_{\ell(j)})-1} \mathbf{e}_{\sigma_{\ell(i)}})$, for $j \in [a_\ell + 1]$, where $\mathbf{e}_j$ is the j-th unit vector with appropriate dimension. These $a_\ell + 1$ points are affinely independent and satisfy inequality (3.6) at equality. Next, consider the set of points: $(\mathbf{e}_1 + \mathbf{e}_{\ell+1}, \bar{\mathbf{y}}^\ell_1 + \mathbf{e}_{\ell+1} M_{\ell+1}, \mathbf{e}_j)$, $\forall j = \Omega \setminus \bar{T}$. These $m - a_\ell - \sum_{i \in \bar{S}} a_{i-1}$ points are feasible, affinely independent from all other points, and satisfy inequality (3.6) at equality.

Next, for T_{i-1}, for all $i \in \bar{S}$ we construct the following set of feasible points:

$$(\mathbf{e}_1 + \mathbf{e}_i + \mathbf{e}_{\ell+1}, \bar{\mathbf{y}}^{i-1}_j + \mathbf{e}_i (D_{\sigma_{\ell(1)}} - D_{t_{i-1(j)}}) + \mathbf{e}_{\ell+1} M_{\ell+1}, \sum_{p=1}^{g(t_{i-1(j)})-1} \mathbf{e}_{\sigma_{i-1(p)}}), \quad j \in [a_{i-1} + 1],$$

and

$$(\mathbf{e}_1 + \mathbf{e}_i + \mathbf{e}_{\ell+1}, \bar{\mathbf{y}}^{i-1}_{a_{i-1}+1} + \mathbf{e}_i(D_{\sigma_{\ell(1)}} - D_{t_{i-1(a_{i-1}+1)}} + \triangle) + \mathbf{e}_{\ell+1}M_{\ell+1}, \sum_{p=1}^{k} \mathbf{e}_{\sigma_{i-1(p)}}),$$

where $0 < \triangle \leq M_i - D_{\sigma_{\ell(1)}} + D_{t_{i-1(a_{i-1}+1)}}$. These $\sum_{i \in \bar{S}} a_{i-1} + 2|\bar{S}|$ points are affinely independent from all other points, and satisfy inequality (3.6) at equality.

Next, for each $i \in S \setminus \{1\}$, we construct the following set of feasible points:

$$(\mathbf{e}_1 + \mathbf{e}_i + \mathbf{e}_{\ell+1}, \bar{\mathbf{y}}^{\ell}_1 + \mathbf{e}_{\ell+1}M_{\ell+1}, \mathbf{0}),$$

$$(\mathbf{e}_1 + \mathbf{e}_i + \mathbf{e}_{\ell+1}, \mathbf{e}_1 D_{\sigma_{i-1(1)}} + \mathbf{e}_i(D_{\sigma_{\ell(1)}} - D_{\sigma_{i-1(1)}}) + \mathbf{e}_{\ell+1}M_{\ell+1}, \mathbf{0}).$$

These $2|S| - 2$ points are feasible, affinely independent from all other points, and satisfy inequality (3.6) at equality.

Next, for each $i \in N \setminus [\ell + 1]$, consider the following set of points: $(\mathbf{e}_1 + \mathbf{e}_{\ell+1} + \mathbf{e}_i, \bar{\mathbf{y}}^{\ell}_1 + \mathbf{e}_{\ell+1}M_{\ell+1}, \mathbf{0})$, and $(\mathbf{e}_1 + \mathbf{e}_{\ell+1} + \mathbf{e}_i, \bar{\mathbf{y}}^{\ell}_1 + \mathbf{e}_{\ell+1}M_{\ell+1} + \mathbf{e}_i\triangle, \mathbf{0})$, where $0 < \triangle \leq M_i$. These $2(n - \ell - 1)$ points are feasible, affinely independent from all other points, and satisfy inequality (3.6) at equality.

Finally, for a fixed index $p^* \in \bar{S}$, we construct the remaining two points: $(\mathbf{e}_1 + \mathbf{e}_{p^*}, \bar{\mathbf{y}}^{p^*-1}_1 + \mathbf{e}_{p^*}M_{p^*}, \mathbf{0})$, and $(\mathbf{e}_1 + \mathbf{e}_{p^*} + \mathbf{e}_{\ell+1}, \bar{\mathbf{y}}^{p^*-1}_1 + \mathbf{e}_{p^*}M_{p^*} + \mathbf{e}_{\ell+1}\triangle, \mathbf{0})$, where $0 < \triangle < M_{\ell+1}$. These two points are feasible, affinely independent from all other points, and satisfy inequality (3.6) at equality. Hence, we obtain $2n + m - 1$ affinely independent feasible points that satisfy inequality (3.6) at equality, which completes the proof. $\square$

Example 1. *(Continued.) Inequality* (3.7) *is a facet-defining inequality for* $conv(P)$*, because* $T_{\ell-1} \cap T^*_{\ell} = \emptyset$*,* $T_{\ell} \cap T^*_{\ell-1} = \emptyset$*,* $1 \in S$*, and* $D_{t_{2(1)}} - D_{t_{1(1)}} = D_{\sigma_{2(1)}} - D_{\sigma_{1(1)}} = 5 < M_2 = 10$*.*

Remark 2. *Note that if* $\bar{S} = \emptyset$*, then the proposed inequality* (3.6) *reduces to the mixing inequality* (3.5) *for a given* $\ell \in N$ *and* $T_{\ell} \subseteq T^*_{\ell}$*. In addition, suppose that*

$D_{\sigma_{\ell+1(k+1)}} \geq D_{\sigma_{\ell(1)}}$. *Consider inequality* (3.6) *for the* $(\ell+1)$*-th time period, when* $\bar{S} = \{\ell+1\}$ *and* $T_{\ell+1} = \emptyset$*, for the same choice of* T_ℓ *as inequality* (3.5)*:*

$$\sum_{i=1}^{\ell} y_i + \sum_{j=1}^{a_\ell} (D_{t_{\ell(j)}} - D_{t_{\ell(j+1)}}) z_{t_{\ell(j)}} \geq D_{\sigma_{\ell+1(k+1)}} - (D_{\sigma_{\ell+1(k+1)}} - D_{\sigma_{\ell(1)}}) x_{\ell+1}. \tag{3.8}$$

Because $D_{\sigma_{\ell+1(k+1)}} \geq D_{\sigma_{\ell(1)}}$ *by assumption, the right-hand side of* (3.8) *equals* $D_{\sigma_{\ell+1(k+1)}}(1-x_{\ell+1}) + D_{\sigma_{\ell(1)}} x_{\ell+1} \geq D_{\sigma_{\ell(1)}} = D_{t_{\ell(1)}}$*, the right-hand side of* (3.5)*. Hence, if* $\bar{S} = \emptyset$ *and* $D_{\sigma_{\ell+1(k+1)}} \geq D_{\sigma_{\ell(1)}}$*, then the mixing inequality* (3.5) *is dominated by the proposed inequality* (3.8)*.*

Next, we consider another special case that shows the strength of our inequalities.

Proposition 11. *If* $\epsilon = 0$*, then adding the proposed inequalities* (3.6) *to* P *is sufficient to give the complete linear description of* conv(P)*.*

Proof. If $\epsilon = 0$, then $k = 0$, and we have to satisfy every scenario, i.e., the cumulative production until time period $i \in N$ must be sufficient to satisfy the scenario with largest cumulative demand until time period i. In this case, $\bar{T} = \emptyset$, and the proposed inequalities (3.6) reduce to the following inequalities:

$$\sum_{i \in S} y_i + \sum_{i \in \bar{S}} (D_{\sigma_{\ell(1)}} - D_{\sigma_{i-1(1)}}) x_i \geq D_{\sigma_{\ell(1)}}. \tag{3.9}$$

Furthermore, when $k = 0$, we can fix $\mathbf{z} = \mathbf{0}$ and $s_{jt} = \sum_{i=1}^{t} (y_i - d_{ji})$ for all $t \in N, j \in \Omega$, and rewrite the deterministic equivalent program:

$$\min \mathbf{f}^\top \mathbf{x} + \mathbf{c}^\top \mathbf{y} + \sum_{j=1}^{m} \sum_{t=1}^{n} \pi_j h_t (\sum_{i=1}^{t} (y_i - d_{ji})) \tag{3.10a}$$

$$\text{s.t.} \sum_{i=1}^{t} y_i \geq D_{\sigma_{t(1)}} \qquad t \in N \tag{3.10b}$$

$$y_i \leq M_i x_i, \qquad i \in N \tag{3.10c}$$

$$\mathbf{y} \in \mathbb{R}^n_+, \mathbf{x} \in \mathbb{B}^n. \tag{3.10d}$$

Note that the optimization problem (3.10) is equivalent to a deterministic uncapacitated lot-sizing problem, where the cumulative demand in each time period is given by the largest cumulative demand in each time period over all scenarios. Hence, the (ℓ, S) inequalities for the deterministic equivalent program (3.10) when $k = 0$ are sufficient to describe $conv(P)$ when $\epsilon = 0$ [follows from 6], and they are in the form of inequality (3.9), which is a special case of the proposed inequality (3.6) when $k = 0$. $\square$

In contrast, for the special case of $\epsilon = 0$, when we let $\mathbf{z} = \mathbf{0}$, inequalities (3.4) reduce to (ℓ, S) inequalities for each scenario $j \in \Omega$ individually, which is not sufficient to describe $conv(P)$ in this case. Clearly, inequalities (3.6) combine information across all scenarios and yield stronger inequalities.

Separation of inequalities (3.6): There are exponentially many inequalities (3.6). We have two main questions when dealing with the separation problem for a given $\ell \in N$: first, for any time period $i \neq 1$, we need to decide if $i \in S$ or $i \in \bar{S}$; second, for each $i \in \bar{S}$, we need to find a subset T_{i-1} of T^*_{i-1} so that the term $\sum_{j \in \bar{T}} \bar{\alpha}_j z_j$ is minimized. First, given a fractional solution $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$, for all $i \in N \setminus \{1\}$, we solve the following problems

$$Y_{i-1} = \min_{T_{i-1} \subseteq T^*_{i-1}} \{ - D_{t_{i-1(1)}} \hat{x}_i + \sum_{p=1}^{a_{i-1}} (D_{t_{i-1(p)}} - D_{t_{i-1(p+1)}}) \hat{z}_{t_{i-1(p)}} \}, \tag{3.11}$$

$$\hat{Y}_i = \min_{T_i \subseteq T^*_i, \sigma_{i(1)} \in T_i} \{ \sum_{p=1}^{a_i} (D_{t_{i(p)}} - D_{t_{i(p+1)}}) \hat{z}_{t_{i(p)}} \}. \tag{3.12}$$

Problems (3.11) and (3.12) can be solved similarly to the separation of the mixing inequalities in $O(k \log k)$ time [38] for each $i \in N \setminus \{1\}$. We let $\bar{T}_{i-1}$ and $\hat{T}_i$ be the optimal argument of problems (3.11) and (3.12), respectively. Finally, for each $\ell \in N \setminus \{1\}$ and $i \in [\ell]$ if $\hat{y}_i \leq D_{t_{\ell(1)}}\hat{x}_i + Y_{i-1}$, then we let $i \in S$. Otherwise, we let $i \in \bar{S}$ and $T_{i-1} = \bar{T}_{i-1}$. Then we obtain $\alpha_{j(i-1)}$ for each $i \in \bar{S} \cup \{\ell+1\}$ and $j \in \bar{T}_{i-1}$. In addition, we let $T_\ell = \hat{T}_\ell$ and $\bar{\alpha}_j = \max\Big\{\max_{i\in\bar{S}}\{\alpha_{j(i-1)}\}, \alpha_{j\ell}\Big\}$, for all $j \in \bar{T} = (\cup_{i\in\bar{S}} T_{i-1}) \cup T_\ell$. If $\sum_{i\in S}\hat{y}_i + \sum_{i\in\bar{S}}(D_{t_{\ell(1)}} - D_{t_{i-1(1)}})\hat{x}_i + \sum_{j\in\bar{T}}\bar{\alpha}_j\hat{z}_j < D_{t_{\ell(1)}}$ for this choice of $\ell, S, \bar{T}$, then we have found a violated inequality (3.6).

Proposition 12. *The proposed separation procedure runs in $O(n\max\{n, k\log(k)\})$ time. Suppose that $T^*_{p-1} \cap T^*_{q-1} \cap T_\ell = \emptyset$, for all $p \neq q$, and $p, q \in \bar{S}$, then the proposed separation procedure is exact.*

Proof. For a fixed index $\ell \in N$, if the condition stated in the proposition holds, then we can rewrite inequality (3.6) as:

$$\sum_{i\in S} y_i + \sum_{i\in\bar{S}}\Big((D_{t_{\ell(1)}} - D_{t_{i-1(1)}})x_i + \sum_{p=1}^{a_{i-1}}(D_{t_{i-1(p)}} - D_{t_{i-1(p+1)}})z_{t_{i-1(p)}}\Big) \qquad + \sum_{p=1}^{a_\ell}(D_{t_{\ell(p)}} - D_{t_{\ell(p+1)}})z_{t_{\ell(p)}} \geq D_{t_{\ell(1)}},$$

because $\alpha_{j(i-1)} = 0$, for all but at most one $i \in \bar{S} \cap \{\ell+1\}$ and $j \in T_{i-1}$. As a result, each time period is separable from other time periods, and the separation procedure is exact.

The complexity of the algorithm for solving (3.11) and (3.12) for all $i \in N \setminus \{1\}$ is $O\big(nk\log(k)\big)$. After finding the optimal $\bar{T}_{i-1}$ for $i \in N \setminus \{1\}$, which is independent of the choice of ℓ, identifying the set S for a given $\ell \in N$ takes $O(n)$ time. Therefore, we get an overall run time of $O(n\max\{n, k\log(k)\})$. $\square$

If the conditions in Proposition 12 are not satisfied, then the separation procedure is a heuristic.

In Appendix C we give a second class of valid inequalities that involves the inventory variables, which is valid for the deterministic equivalent formulation, and facet-defining under certain conditions.

3.4 A new formulation that exploits the simple recourse property

The deterministic equivalent formulation contains $O(mn)$ additional variables, which becomes computationally challenging if the number of scenarios, m, or the number of time periods, n increases. One can consider a Benders decomposition algorithm given in Appendix D. However, we may have to add exponentially many optimality cuts, which significantly slow down the solution of the master problem, as we show in our computational study in Section 3.5.

In this section, we propose a new formulation for SPLS that is similar to the master problem used in the Benders decomposition algorithm. However we show that the new formulation only uses polynomially many inequalities to capture the second-stage cost.

For all $i \in N$ and $j \in \Omega$, let $\bar{\sigma}$ be a permutation of the scenarios such that $D_{\bar{\sigma}_{i(1)}i} \leq D_{\bar{\sigma}_{i(2)}i} \leq \cdots \leq D_{\bar{\sigma}_{i(m)}i}$, where $D_{\bar{\sigma}_{i(j)}i}$ is the j-th *smallest* cumulative demand for the i-th time period. To further simplify the notation, let $D_{\bar{\sigma}_{i(j)}} = D_{\bar{\sigma}_{i(j)}i}$.

Proposition 13. *Let Θ'_i be an additional variable that captures the total inventory of i-th time period for all scenarios. In addition, let $[k]^+ = \{0, 1, 2, \ldots, k\}$. The* *formulation*

$$\min \mathbf{f}^\top \mathbf{x} + \mathbf{c}^\top \mathbf{y} + \frac{1}{m}\sum_{i=1}^{n} h_i \Theta'_i \tag{3.13a}$$

$$s.t.\ (3.3\text{b}) - (3.3\text{d}), \tag{3.13b}$$

$$\Theta'_i \geq (m-q)\sum_{p=1}^{i} y_p - \sum_{p=1}^{m-q} D_{\bar{\sigma}_{i(p)}}, \qquad i \in N, q \in [k]^+ \tag{3.13c}$$

$$x \in \mathbb{B}^n, y \in \mathbb{R}^n_+, z \in \mathbb{B}^m, \Theta' \in \mathbb{R}^n_+, \tag{3.13d}$$

is equivalent to the deterministic equivalent of SPLS (3.3) *under equiprobable scenarios.*

Proof. We can rewrite the deterministic equivalent formulation (3.3) as a two-stage problem given by

$$\begin{aligned} &\min \mathbf{f}^\top \mathbf{x} + \mathbf{c}^\top \mathbf{y} + \frac{1}{m}\sum_{i=1}^{n} h_i \Theta'_i(\mathbf{y}) \\ &\text{s.t. } (3.3\text{b}) - (3.3\text{d}), \\ &\qquad x \in \mathbb{B}^n, y \in \mathbb{R}^n_+, z \in \mathbb{B}^m, \Theta' \in \mathbb{R}^n_+, \end{aligned}$$

where $\Theta'_i(\mathbf{y})$, the total inventory level at each period i, is defined by the second-stage simple resource problem with respect to each time period i, stated as

$$\begin{aligned} \Theta'_i(\mathbf{y}) = \ \min \ & \sum_{j=1}^{m} s_{ji} \\ \text{s.t. } & s_{ji} \geq \sum_{p=1}^{i} y_p - D_{ji}, && j \in \Omega \\ & s_{ji} \geq 0 && j \in \Omega. \end{aligned}$$

Let Θ'_i be a variable that captures the correct value of $\Theta'_i(\mathbf{y})$ for any feasible $\mathbf{y}$ through the exponentially many inequalities

$$\Theta'_i \geq (m-q)\sum_{p=1}^{i} y_p - \sum_{j \in R_q} D_{ji}, \quad q \in [k]^+, \tag{3.14}$$

where $R_q \subseteq \Omega$ is a subset of scenarios such that $|R_q| = m - q$. Hence, to show that the proposed formulation (3.13) is equivalent to the deterministic equivalent program (3.3), we show that the polynomial subclass (3.13c) of the exponential class of inequalities (3.14) suffice to give a correct formulation. For a fixed $q \in [k]^+$ and $i \in N$, consider the following chain of inequalities:

$$\Theta'_i \geq (m-q)\sum_{p=1}^{i} y_i - \sum_{p=1}^{m-q} D_{\bar{\sigma}_{i(p)}} \geq (m-q)\sum_{p=1}^{i} y_p - \sum_{j \in R_q} D_{ji},$$

where the first inequality follows from the fact that the set $\{\bar{\sigma}_{i(1)}, \bar{\sigma}_{i(2)}, \ldots, \bar{\sigma}_{i(m-q)}\}$ is a possible choice of R_q, and the second inequality follows from the definition of the permutation $\bar{\sigma}$. Hence, the polynomial class of inequalities (3.13c) implies all inequalities of the form (3.14), which completes the proof. $\square$

Example 1. *(Continued.) Let $i = 2$, then the value of Θ'_2 can be captured by the following $k + 1 = 3$ inequalities*

$$\Theta'_2 \geq 5(y_1 + y_2) - 7 - 9 - 9 - 10 - 11, \tag{3.15a}$$

$$\Theta'_2 \geq 4(y_1 + y_2) - 7 - 9 - 9 - 10, \tag{3.15b}$$

$$\Theta'_2 \geq 3(y_1 + y_2) - 7 - 9 - 9. \tag{3.15c}$$

In the optimal solution, if every scenario is satisfied at time period 2, then inequality (3.15a) *captures the value of Θ'_2, and the other two inequalities provide lower bounds on Θ'_2. Suppose that in the optimal solution, one scenario is violated in time period 2, then the violated scenario must be the scenario with the highest cumulative demand at time period 2. Hence, inequality* (3.15b) *captures the correct value of Θ'_2, and inequalities* (3.15a) *and* (3.15c) *yield lower bounds for Θ'_2.*

Remark 3. *We show that the proposed formulation can also be applied to the general two-stage chance-constrained program with simple recourse, equiprobable scenarios and finite probability space.*

Given a scenario set $\Omega = \{1, 2, \ldots, m\}$*, let* $\mathbf{x}$ *be the vector of the first stage decision variables,* $\mathbf{c}$ *be its cost vector, and* X *be its feasible region. In addition, the following scenario-dependent constraint set:*

$$A_j \mathbf{x} \geq b_j$$

is enforced only when scenario $j \in \Omega$ *is satisfied by the chance constraint, where* A_j *and* b_j *are random coefficient matrix of* $\mathbf{x}$ *and right-hand side vector with appropriate dimensions, respectively. In addition, the d-dimensional simple recourse function, [see, e.g., 16] is defined as:*

$$\sum_{i=1}^{d} h_i [\mathbf{u}_i^\top \mathbf{x} - g_{ji}]_+, \quad j \in \Omega,$$

where g_{ji} *is scenario-dependent parameter, for all* $j \in \Omega$ *and* $i \in [d]$*, and* $\mathbf{u}_i$ *is the coefficient vector of the recourse function for i-th dimension, for all* $i \in [d]$*. Let* $h_i, i \in [d]$ *be a penalty term for the excess* $[\mathbf{u}_i^\top \mathbf{x} - g_{ji}]_+$ *in the second stage.*

Assume that each scenario is equally likely. The deterministic equivalent of a general two-stage chance-constrained program with simple recourse, equiprobable scenarios, and finite probability space is stated as follows:

$$\min \mathbf{c}^\top \mathbf{x} + \frac{1}{m} \sum_{j=1}^{m} \sum_{i=1}^{d} h_i [\mathbf{u}_i^\top \mathbf{x} - g_{ji}]_+ \tag{3.16a}$$

$$s.t.\ A_j x + \bar{M}_j z_j \geq b_j \tag{3.16b}$$

$$\sum_{j=1}^{m} z_j \leq k \tag{3.16c}$$

$$\mathbf{x} \in X, \mathbf{z} \in \mathbb{B}^m, \tag{3.16d}$$

where (3.16b)-(3.16c) enforce the chance constraint, and $\bar{M}_j$ is sufficiently large to make (3.16b) redundant when $z_j = 1$. Since we have to introduce $O(md)$ new variables and constraints to linearize the nonlinear term in the cost function (3.16a), the deterministic equivalent program (3.16) is a large-scale mixed-integer program, which is very hard to solve.

Let $\bar{\Theta}_i$, for all $i \in [d]$, be the additional variable that captures the value of the recourse function for dimension i. In addition, let σ' be the permutation of scenarios such that: $g_{\sigma'_{i(1)}i} \leq g_{\sigma'_{i(2)}i} \leq \cdots \leq g_{\sigma'_{i(m)}i}$. In order to simplify notation, let $g_{\sigma'_{i(j)}i} = g_{\sigma'_{i(j)}}$, for all $j = 1, 2, \ldots, m$. Hence, according to Proposition 13, we can rewrite the deterministic equivalent formulation (3.1) as:

$$
\begin{aligned}
\min\ & \mathbf{c}^\top \mathbf{x} + \frac{1}{m} \sum_{i=1}^{d} h_i \bar{\Theta}_i \\
\text{s.t. } & (3.1\text{b}) - (3.1\text{c}) \\
& \bar{\Theta}_i \geq \sum_{j=1}^{m-q} (m-q)\mathbf{u}_i^\top \mathbf{x} - g_{\sigma'_{i(j)}}, && i \in [d], q \in [k]^+, \\
& \bar{\Theta} \in \mathbb{R}_+^d.
\end{aligned}
$$

Here we only require d new variables and $O(dk)$ many new constraints. In this case, we can greatly reduce the number of variables and constraints in the deterministic equivalent formulation, because $k \ll m$, for small ϵ.

3.5 Computational Experiments

In this section, we summarize our computational experience with various classes of valid inequalities and our new formulation. All runs were executed on a Windows Server 2012 R2 Data Center with 2.40GHZ Intel(R) Xeon(R) CPU and 32.0 GB RAM. The algorithms tested in the computational experiment were implemented using C programming language, with Microsoft Visual Studio 2012 and CPLEX 12.6. A time limit of one hour is set.

In our experiments, we compare the proposed new formulation (3.13) against the deterministic equivalent formulation (3.3) and Benders decomposition algorithm (see Appendix D), with different choices of valid inequalities. The first class of valid inequalities (3.6) and its special case of mixing inequalities (3.5) are valid for the deterministic equivalent formulation, the Benders master problem and the new formulation (3.13). However, the second class of valid inequalities given in Appendix C include the inventory variables, hence they only apply to the deterministic equivalent formulation. In Tables 3.2 and 3.3, each row reports the average of three instances. We let f_i and c_i to be randomly generated from a discrete uniform distribution over [50, 100], and [5, 10], respectively, for all $i \in N$. In addition, we generate the demand in each period randomly, where d_{ji} follows discrete uniform distribution [10, 30], for all $i \in N$ and $j \in \Omega$.

In Table 3.2, the "DEP (3.5), (3.6), (C.1);" "B. D. & Ineq. (3.5)-(3.6);" "N.F. & (3.5);" and "N.F. & (3.5)-(3.6)" columns report the performance of the deterministic equivalent formulation with the additional strengthening from inequalities (3.5), (3.6), and (C.1); Benders decomposition algorithm with valid inequalities (3.5) and (3.6); new formulation with valid mixing inequalities (3.5); and new formulation with mixing inequalities and the proposed inequalities (3.6), respectively. The number of mixing inequalities that can be added to both formulations is limited to 150, and based on the results, this limit is hit by every instance. The "Time" column reports the average solution time in seconds for the instances that are solved to optimality within time limit, and the "Gap" column reports the average optimality gap for the instances that reach the time limit. The " - " sign under the "Time " column indicates that no instance is solved to optimality within time limit. The " * " sign indicates that CPLEX is not able to solve the instance due to memory limit, and no feasible solution is obtained. In addition, we only add the proposed inequalities at the root node level.

Table 3.2: Computational results comparing different formulations.

Instances		DEP (3.5), (3.6), (C.1)		B. D. & Ineq. (3.5)-(3.6)		N.F. & (3.5)		N.F. & (3.5)-(3.6)	
(ϵ, n)	m (10^3)	Time	Gap (%)	Time	Gap (%)	Time	Gap (%)	Time	Gap (%)
(0.01, 5)	10	277	0	199	0	143	0	92	0
	20	*	*	860	0	441	0	387	0
	30	*	*	-	0.47	-	0.12	3534	0.05
(0.01, 10)	10	*	*	-	0.16	-	1.11	-	1.38
	20	*	*	-	2.17	-	0.94	3416	0.90
	30	*	*	*	*	-	3.28	-	2.36
(0.01, 30)	3	*	*	1028	0	185	0	127	0
	4	*	*	1794	6.71	524	0	397	0
	5	*	*	3324	14.66	1472	0	1334	0
(0.01, 40)	3	*	*	1179	0	723	0	606	0
	4	*	*	-	23.02	1864	0.24	1690	0.07
	5	*	*	-	14.51	3321	0.71	2793	0.57

Table 3.3: Additional information for the experiments in Table 3.2.

Instances		B. D. & Ineq. (3.5)-(3.6)		N.F. & (3.5)		N.F. & (3.5)-(3.6)		
(ϵ, n)	m (10^3)	Nodes	Opt.Cut	Nodes	R.Gap (%)	Nodes	R. Gap (%)	Cuts
(0.01, 5)	10	1828	42398	2028	1.08	880	1.00	7
	20	12493	121553	737	3.90	623	3.43	6
	30	48676	373193	54667	3.88	41379	3.41	7
(0.01, 10)	10	55031	98625	36067	3.61	33751	3.46	12
	20	28112	288242	29315	6.76	32529	4.96	10
	30	*	*	6327	7.92	9088	6.78	10
(0.01, 30)	3	5047	29828	837	3.60	359	2.88	23
	4	9696	64784	2233	5.46	1892	3.59	26
	5	12267	88599	5917	7.42	5653	6.08	22
(0.01, 40)	3	5026	31672	2088	3.88	1608	3.76	16
	4	9397	103936	6494	6.36	5729	2.53	19
	5	8150	64154	7375	3.96	7672	3.24	18

As we can see from Table 3.2, the deterministic equivalent formulation cannot solve most of the instances, due to the memory limit. The Benders decomposition provides slightly better results, since it is able to find a feasible solution. However, for the instances with 30 or 40 time periods, the optimality gap of Benders decomposition algorithm is very large. The proposed new formulation provides a big improvement. It can solve most of the instances to optimality. For the instances that reach the time limit, the optimality gap is small. Finally, the effectiveness of the proposed inequalities (3.6) is shown in the last column. It provides the best results, with generally the smallest solution time and optimality gap.

In Table 3.3, we report additional information on the average root gap ("R.Gap %") and number of nodes explored during the branch-and-bound process ("Nodes"). The column "Opt.Cut" reports the number of optimality cuts added to the Benders master problem. The column "Cuts" reports the number of the proposed inequalities (3.6) added to the new formulation in addition to the mixing cuts (3.5), which are special cases of inequalities (3.6). As we can see from Table 3.3, because we only add the proposed inequalities (3.6) at the root node after adding the violated inequalities (3.5), the number of additional inequalities (3.6) is not very large. However, the new cuts are beneficial; the number of branch-and-bound nodes is reduced with the proposed inequalities (3.6), and the root node gap with the new inequalities is also smaller in most cases. As a result, more instances are solved to optimality within the time limit. In addition, compared with the results from Benders decomposition, the proposed new formulation uses much fewer "optimality cuts" to capture the second-stage inventory value. For example, for the instances where $m = 10000$ and $n = 5$, the Benders decomposition algorithm requires 42398 optimality cuts. In contrast, the proposed new formulation only requires $m \times \epsilon \times n = 500$ additional inequalities to fully capture the second-stage inventory value. As a result, the proposed new formulation (3.13) provides a significant improvement in solution time.

We also tested the effectiveness of adapting the extended formulation of [51] for deterministic ULS to strengthen the deterministic equivalent of SPLS. However, we observe that it slows down the deterministic equivalent model further, so we do not report our computations with this formulation.

Chapter 4: Integer Programming Approaches to Two-Sided Chance-Constrained Program

This chapter is based on [60]. To model complex systems operating under uncertainty, chance constraints are used to assure that the quality of service or the reliability of the solution at an acceptable level. Charnes et al. [25] defines the first disjoint chance-constrained program. The deterministic equivalent program of chance-constraint is proposed by Charnes and Cooper [24]. [80] provides the first study of joint chance-constrained programs with dependent variables.

A sample-average approximation (SAA) algorithm for CCPs with general probabilistic distribution is proposed by Luedtke and Ahmed [66] [see, also 21, 22, 73, 23]. The sampled problem (with finitely many scenarios) can be formulated as a deterministic mixed-integer program, where the joint chance constraint is represented by the scenario dependent Big-M type of constraints. This reformulation is generally weak and inefficient to solve for state of the art mixed-integer programming solvers.

Luedtke et al. [67] observe that this reformulation contains a mixing set substructure that is first introduced by Günlük and Pochet [38], and propose valid inequalities that strengthen the basic mixing inequalities studied in [38] and [4]. Küçükyavuz [52] and [1] propose various methods, including strong valid inequalities and extended formulations, for the deterministic equivalent formulation of the chance constraint, where the randomness only appears in the right-hand side of the joint chance-constraint. In addition, Luedtke [65] proposes valid inequalities and a decomposition method for general chance-constrained programs with *linear constraints*. [63] propose decomposition algorithms for general two-stage CCPs.

In this chapter, we study the two-sided chance-constrained programs first introduced in [64]. In particular, we consider an inequality in the chance constraints that contains an absolute value term. This is the most natural extension of a linear chance-constrained program. While individual linear chance constraints are easy to handle and can be linearized using quantile arguments [17], a single inequality containing absolute value terms cannot be linearized, because of its representation as two linear inequalities that contain correlated random variables. Our work is differs from [64] in that we do not assume any distribution on the random variables. In addition, the structure of the two-sided inequality we study generalizes that of these authors under the finite distribution assumption. As a result of the finiteness assumption, we are able to obtain mixed-integer linear reformulations, for which we propose strong valid inequalities.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a finite probability space. We consider the following problem:

$$\min \xi^\top \mathbf{x} \tag{4.1a}$$

$$\text{s.t. } \mathbb{P}\Big(|\mathbf{d}^\top \mathbf{x} - h(\omega)| \leq \mathbf{p}^\top \mathbf{x} - q(\omega)\Big) \geq 1 - \epsilon, \tag{4.1b}$$

$$x \in X, \tag{4.1c}$$

where $\mathbf{x} \in \mathbb{R}^{n_1}$ is the vector of decision variables , ξ is its cost vector, and $\mathbf{d}$ and $\mathbf{p}$ are n_1 dimensional coefficient vectors. In addition, $h(\omega)$ and $q(\omega)$ are random parameters that depend on the random variable $\omega \in \Omega$. The chance constraint (4.1b) enforces that the probability that the solution is feasible should be no less than the risk rate ϵ. In Appendix F, we show that problem (4.1) is a special case of the problems that contains the intersection of multiple mixing sets. In addition, we also show that the inequalities proposed in this chapter can be extended to problems that contain the intersection of multiple mixing sets.

Unlike the problem structures studied in Luedtke et al. [67], Küçükyavuz [52] and Abdi and Fukasawa [1], problem (4.1) involves a absolute value of function inside the joint chance-constraint, which brings more complication in terms of the convex hull structure of this problem.

Before we present our results, we give a summary of the notation and conventions used throughout the chapter.

Notation. Given $a \in \mathbb{R}$, we set $(a)_+ = \max\{a, 0\}$. For a positive integer n, we let $[n] = \{1, \ldots, n\}$. We use bold letters to denote vectors. For a vector $\mathbf{x} \in \mathbb{R}^n$ and an integer $k \in [n]$, x_k denote the k-th coordinate of $\mathbf{x}$. We let $\mathbf{e}_j$ be the j-th unit vector for $j \in [n]$, and $\mathbf{1}_n$ denotes the vector of all 1's in $\mathbb{R}^n$. Given a set $\mathcal{P}$, we denote its dimension, closure, convex hull, and closed convex hull by $\dim(\mathcal{P})$, $\text{cl}(\mathcal{P})$, $\text{conv}(\mathcal{P})$, and $\text{clconv}(\mathcal{P})$, respectively.

4.1 Deterministic Equivalent Formulation

When the probability space Ω is finite, problem (4.1) can be reformulated as a so-called deterministic equivalent program as follows. Let $\Omega := \{1, 2, \ldots, m\}$, and $\mathbb{P}(\omega = j) = \pi_j$, for all $j \in \Omega$ and $\sum_{j=1}^{m} \pi_j = 1$. In addition, to simplify notation, we define $h_j := h(\omega = j)$ and $q_j := q(\omega = j)$ for all $j \in \Omega$. Then (4.1) is equivalent to

$$\min \xi^\top \mathbf{x} \tag{4.2a}$$

$$\text{s.t. } |\mathbf{d}^\top \mathbf{x} - h_j| \le \mathbf{p}^\top \mathbf{x} - q_j + M_j z_j, \qquad \forall j \in \Omega, \tag{4.2b}$$

$$\sum_{j \in \Omega} \pi_j z_j \le \epsilon, \tag{4.2c}$$

$$\mathbf{z} \in \mathbb{B}^m, \tag{4.2d}$$

$$\mathbf{x} \in X,$$

where M_j is a sufficient large constant to make (4.2b) redundant when $z_j = 1$, for all $j \in \Omega$.

Let us define the variables $y_p = \mathbf{p}^\top \mathbf{x}$ and $y_d = \mathbf{d}^\top \mathbf{x}$. Then inequality (4.2b) can be linearized as follows

$$y_p + y_d + M_j^1 z_j \ge q_j + h_j, \qquad \forall j \in \Omega, \tag{4.3a}$$

$$y_p - y_d + M_j^2 z_j \ge q_j - h_j, \qquad \forall j \in \Omega, \tag{4.3b}$$

where M_j^1 and M_j^2 are sufficiently large to make inequalities (4.3a) and (4.3b) redundant when $z_j = 1$, respectively, for all $j \in \Omega$.

Because X is a compact set, we can also derive lower and upper bounds on our new variables $y_d = \mathbf{d}^\top \mathbf{x}$ and $y_p = \mathbf{p}^\top \mathbf{x}$. In particular, we set $u_d := \max_{\mathbf{x} \in X} \; \mathbf{d}^\top \mathbf{x}$, and $l_d := \min_{\mathbf{x} \in X} \; \mathbf{d}^\top \mathbf{x}$. Then $l_d \le y_d \le u_d$. Similarly, we have $y_p \ge l_p$, where $l_p := \min_{\mathbf{x} \in X} \; \mathbf{p}^\top \mathbf{x}$.

Throughout this chapter, in order to simplify our notation, we define

$$w_j := q_j + h_j \quad \text{and} \quad v_j := q_j - h_j \quad \text{for all } j \in \Omega.$$

Based on this, we set the big-M values in (4.3a)–(4.3b) to $M_j^1 := w_j - l_p - l_d$ and $M_j^2 := v_j - l_p + u_d$ for all $j \in \Omega$.

Throughout the rest of this chapter, to simplify notation, without loss of generality, we assume that $l_p = l_d = 0$, see Observation 14. Hence $u_d \geq 0$.

The constraint set (4.3) together with (4.2d) contains an interesting substructure. Based on this notation, the key substructure originating from the inequalities (4.3) that we are interested in is given by the variables $y_p, y_d \in \mathbb{R}$ and $\mathbf{z} \in \mathbb{B}^m$ defined by

$$y_p + y_d + w_j z_j \geq w_j, \qquad \forall j \in \Omega, \tag{4.4a}$$

$$y_p - y_d + (v_j + u_d) z_j \geq v_j, \qquad \forall j \in \Omega, \tag{4.4b}$$

$$y_p \geq 0, \tag{4.4c}$$

$$u_d \geq y_d \geq 0, \tag{4.4d}$$

$$\mathbf{z} \in \mathbb{B}^m. \tag{4.4e}$$

We define

$$\mathcal{P} := \{(y_p, y_d, \mathbf{z}) \mid (4.4\text{a}) - (4.4\text{e})\}.$$

Based on this, two-sided chance constrained problem (4.1) is equivalent to

$$\min_{\mathbf{x}, y_p, y_d, \mathbf{z}} \left\{ \xi^\top \mathbf{x} : \quad x \in X,\ y_p = \mathbf{p}^\top \mathbf{x},\ y_d = \mathbf{d}^\top \mathbf{x},\ \sum_{j \in \Omega} \pi_j z_j \leq \epsilon,\ (y_p, y_d, \mathbf{z}) \in \mathcal{P} \right\}.$$

Remark 4. *$\mathcal{P}$ can be viewed as the intersection of two mixing sets represented by inequalities* (4.4a) *and* (4.4e)*, and inequalities* (4.4b) *and* (4.4e)*, and the bound inequalities* (4.4c)–(4.4d)*.*

Despite Remark 4, interaction between these two mixing sets in $\mathcal{P}$ through the shared continuous variables y_p and y_d, along with their bounds, easily lead to a nontrivial structure. In this chapter, we study the structure of $\text{clconv}(\mathcal{P})$.

4.2 Structure of the Set $\mathcal{P}$

4.2.1 Preliminaries and Main Assumptions

Throughout the chapter, we use $(y_p, y_d, \mathbf{z})$ to express points from $\mathcal{P}$.

Observation 14. *Without loss of generality, we can assume that $l_p = l_d = 0$ in the definition of $\mathcal{P}$.*

Proof. Let $y'_p = y_p + l_p$ and $y'_d = y_d + l_d$. Define $w'_j := w_j + l_p + l_d$ and $v'_j := v_j + l_p - l_d$ for all $j \in \Omega$. Consider the set $\mathcal{P}'$ defined as follows:

$$\begin{aligned}
& y'_p + y'_d + w_j z_j \geq w'_j, && \forall j \in \Omega,\\
& y'_p - y'_d + (v_j + u_d) z_j \geq v'_j, && \forall j \in \Omega,\\
& y'_p \geq l_p, \quad u_d + l_d \geq y'_d \geq l_d, \quad \mathbf{z} \in \mathbb{B}^m. &&
\end{aligned}$$

For any $(y_p, y_d, \mathbf{z}) \in \mathcal{P}$, the corresponding $(y'_p, y'_d, \mathbf{z}) \in \mathcal{P}'$ and vice versa. $\square$

Observation 15. *Without loss of generality, we can assume that $w_j \geq 0$ for all $j \in \Omega$ in the definition of $\mathcal{P}$.*

Proof. Define $\Omega' := \{j \in \Omega : \ w_j < 0\}$. For every $j \in \Omega'$, let $z'_j = 1 - z_j$, $w'_j = 0$, and $v'_j = -u_d$. For every $j \in \Omega \setminus \Omega'$, we set $z'_j = z_j$, $w'_j = w_j$, and $v'_j = v_j$. Consider the set $\mathcal{P}'$ defined as follows:

$$\begin{aligned}
& y_p + y_d + w_j z'_j \geq w_j, && \forall j \in \Omega \setminus \Omega',\\
& y_p - y_d + (v_j + u_d) z'_j \geq v_j, && \forall j \in \Omega \setminus \Omega',\\
& y_p + y_d + (-w_j) z'_j \geq 0, && \forall j \in \Omega',\\
& y_p - y_d + (-v_j - u_d) z'_j \geq -u_d, && \forall j \in \Omega',\\
& y_p \geq 0, \quad u_d \geq y_d \geq 0, \quad \mathbf{z}' \in \mathbb{B}^m. &&
\end{aligned}$$

Then there is a one-to-one correspondence between the vectors in $\mathcal{P}$ and the vectors in $\mathcal{P}'$.

Note that after the transformation, the constraints $y_p + y_d + w'_j z'_j \geq 0 \ \ \forall j \in \Omega'$ are redundant because all of the variables are nonnegative and $w'_j \geq 0$. However, the constraint $y_p - y_d + (-v_j - u_d)z'_j \geq -u_d$ for some $\forall j \in \Omega'$ may be non-redundant. $\square$

While it may be possible to have $v_j < 0$ for some $j \in \Omega$, throughout the rest of this chapter, we work with the following assumption that complements Observation 15:

A*: w_j and v_j are nonnegative for all $j \in \Omega$.

We start by examining some valid inequalities for the set $\mathcal{P}$ and establishing conditions under which these inequalities are facets of clconv($\mathcal{P}$).

4.2.2 Valid Inequalities

Mixing sets have been studied extensively in the literature. From Remark 4, because of the existing mixing set substructure in $\mathcal{P}$, the star inequalities of [4], or the mixing inequalities of [38], can immediately be used to strengthen the formulation of $\mathcal{P}$.

Proposition 16. *[4, 38] Let $S := \{s_1, s_2, \ldots, s_\eta\} \subseteq \Omega$ be a subset of scenarios such that $w_{s_1} \geq w_{s_2} \geq \cdots \geq w_{s_\eta}$, and define $w_{s_{\eta+1}} = 0$. Similarly, let $T := \{t_1, t_2, \ldots, t_\rho\} \subseteq \Omega$ be a subset of scenarios such that $v_{t_1} \geq v_{t_2} \geq \cdots \geq v_{t_\rho}$, and define $v_{t_{\rho+1}} = -u_d$. Then the following* mixing inequalities *are valid for $\mathcal{P}$.*

$$y_p + y_d + \sum_{j=1}^{\eta} (w_{s_j} - w_{s_{j+1}}) z_{s_j} \geq w_{s_1}, \quad \textit{for the given } S \subseteq \Omega, \tag{4.5}$$

and

$$y_p - y_d + \sum_{j=1}^{\rho}(v_{t_j} - v_{t_{j+1}})z_{t_j} \geq v_{t_1}, \quad \textit{for the given } T \subseteq \Omega. \tag{4.6}$$

Proof. The validity of inequality (4.5) directly follows from [4] and [38]. In addition, inequality (4.6) is closely related to the mixing inequalities for the set generated by inequalities (4.4b)–(4.4d). $\square$

In general, the mixing inequalities are not sufficient to describe clconv($\mathcal{P}$) because the intersection of the convex hulls of two mixing sets can create additional extreme points. We next introduce a new class of valid inequalities for $\mathcal{P}$ and we will later on show that these inequalities are facet defining for $\mathcal{P}$ under certain conditions.

Let $\tau \in [m]$, and Π be a *sequence* of τ scenarios given by $\pi_1 \to \pi_2 \to \cdots \to \pi_\tau$, where $\pi_j \in \Omega$, for all $j \in [\tau]$. Given $\Pi := \{\pi_1 \to \pi_2 \to \cdots \to \pi_\tau\}$, consider the following class of inequalities:

$$2y_p + \sum_{j=1}^{\tau}\Big((w_{\pi_j} - \bar{w}_{\pi_j})_+ + (v_{\pi_j} - \bar{v}_{\pi_j})_+\Big)z_{\pi_j} \geq \bar{w}_{\pi_0} + \bar{v}_{\pi_0}, \tag{4.7}$$

where

$$\bar{w}_{\pi_j} = \begin{cases} \max_{j+1 \leq \ell \leq \tau}\{w_{\pi_\ell}\}, & \text{if } j \in [\tau - 1] \cup \{0\}, \\ 0, & \text{if } \quad j = \tau, \end{cases}$$

and

$$\bar{v}_{\pi_j} = \begin{cases} \max_{j+1 \leq \ell \leq \tau}\{v_{\pi_\ell}\}, & \text{if } j \in [\tau - 1] \cup \{0\}, \\ 0, & \text{if } \quad j = \tau. \end{cases}$$

Proposition 17. *For a given $\tau \in [m]$ and a sequence of scenarios $\Pi := \{\pi_1 \to \pi_2 \to \cdots \to \pi_\tau\}$, inequality (4.7) is valid for* clconv($\mathcal{P}$)*.*

Proof. Let $R \subseteq \Pi$ be a subsequence of scenarios given by $r_1 \to r_2 \to \cdots \to r_{\tau_R}$, with $|R| = \tau_R \leq \tau$, such that $r_j \in R$, for all $j \in [\tau_R]$, if $w_{r_j} \geq \bar{w}_{r_j}$. Since $\bar{w}_{\pi_{\tau_R}} = 0$, we know that R cannot be empty. Similarly, let $G \subseteq \Pi$ be a subsequence of scenarios given by $g_1 \to g_2 \to \cdots \to g_{\tau_G}$, with $|G| = \tau_G \le \tau$, such that $g_j \in G$, for all $j \in [\tau_G]$, if $v_{g_j} \ge \bar{v}_{g_j}$. Since $\bar{v}_{\pi_{\tau_G}} = 0$, we know that G cannot be empty. From this construction, we have $w_{r_i} \ge w_{r_{i+1}}$, for all $i \in [\tau_R - 1]$, because $w_{r_i} \ge \bar{w}_{r_i}$, and based on the definition of the subsequence R, scenario r_i precedes scenario r_{i+1} in R if and only if r_i precedes r_{i+1} in Π, which implies $\bar{w}_{r_i} \ge w_{r_{i+1}}$. Similarly, we have $v_{g_i} \ge v_{g_{i+1}}$, for all $i \in [\tau_G - 1]$. Hence, inequality (4.7) is equivalent to

$$2y_p + \sum_{j=1}^{\tau_R} (w_{r_j} - w_{r_{j+1}}) z_{r_j} + \sum_{j=1}^{\tau_G} (v_{g_j} - v_{g_{j+1}}) z_{g_j} \ge w_{r_1} + v_{g_1}, \tag{4.8}$$

where $w_{r_{\tau_R+1}} = 0$, and $v_{g_{\tau_G+1}} = 0$.

For a given solution $(y_p, y_d, \mathbf{z})$, let $j_1 := \arg\min_{i \in [\tau_R]} \{i \mid z_{r_i} = 0\}$ and $j_2 := \arg\min_{i \in [\tau_G]} \{i \mid z_{g_i} = 0\}$.

First, suppose that j_1 and j_2 exist. Then we have $y_p + y_d \ge w_{r_{j_1}}$ and $y_p - y_d \ge v_{g_{j_2}}$ from inequalities (4.4a) and (4.4b). Hence,

$$\begin{aligned}
&2y_p + \sum_{j=1}^{\tau_R} (w_{r_j} - w_{r_{j+1}}) z_{r_j} + \sum_{j=1}^{\tau_G} (v_{g_j} - v_{g_{j+1}}) z_{g_j} \\
&= y_p + y_d + y_p - y_d + \sum_{j=1}^{\tau_R} (w_{r_j} - w_{r_{j+1}}) z_{r_j} + \sum_{j=1}^{\tau_G} (v_{g_j} - v_{g_{j+1}}) z_{g_j} \\
&\ge w_{r_{j_1}} + \sum_{j=1}^{j_1 - 1} (w_{r_j} - w_{r_{j+1}}) z_{r_j} + v_{g_{j_2}} + \sum_{j=1}^{j_2 - 1} (v_{g_j} - v_{g_{j+1}}) z_{g_j} \\
&= w_{r_1} + v_{g_1}.
\end{aligned}$$

This establishes validity of inequality (4.7) when both j_1 and j_2 exist.

Next, suppose that j_1 does not exist, but j_2 exists. Then $y_p + y_d \geq 0$, $y_p - y_d \geq v_{g_{j_2}}$, and

$$\begin{aligned}
&2y_p + \sum_{j=1}^{\tau_R}(w_{r_j} - w_{r_{j+1}})z_{r_j} + \sum_{j=1}^{\tau_G}(v_{g_j} - v_{g_{j+1}})z_{g_j} \\
&= y_p + y_d + y_p - y_d + \sum_{j=1}^{\tau_R}(w_{r_j} - w_{r_{j+1}}) + \sum_{j=1}^{\tau_G}(v_{g_j} - v_{g_{j+1}})z_{g_j} \\
&\geq 0 + w_{r_1} + v_{g_{j_2}} + \sum_{j=1}^{j_2-1}(v_{g_j} - v_{g_{j+1}}) \\
&= w_{r_1} + v_{g_1}.
\end{aligned}$$

Thus, inequality (4.7) is valid if j_2 exists and j_1 does not exist.

Next, if j_1 exists and j_2 does not exist, then $z_{r_{j_1}} = 0$, and we have $y_p + y_d \geq w_{r_{j_1}}$, and $y_p - y_d \geq v_{r_{j_1}}$. As a result, we have:

$$\begin{aligned}
&2y_p + \sum_{j=1}^{\tau_R}(w_{r_j} - w_{r_{j+1}})z_{r_j} + \sum_{j=1}^{\tau_G}(v_{g_j} - v_{g_{j+1}})z_{g_j} \\
&= y_p + y_d + y_p - y_d + \sum_{j=1}^{\tau_R}(w_{r_j} - w_{r_{j+1}})z_{r_j} + \sum_{j=1}^{\tau_G}(v_{g_j} - v_{g_{j+1}})z_{g_j} \\
&\geq w_{r_{j_1}} + v_{r_{j_1}} + \sum_{j=1}^{\tau_R}(w_{r_j} - w_{r_{j+1}})z_{r_j} + v_{g_1} \\
&\geq w_{r_{j_1}} + v_{r_{j_1}} + \sum_{j=1}^{j_1-1}(w_{r_j} - w_{r_{j+1}}) + v_{g_1} \\
&= w_{r_1} + v_{g_1} + v_{r_{j_1}} \geq w_{r_1} + v_{g_1},
\end{aligned}$$

where the last inequality follows Assumption $\mathbf{A}^*$.

Finally, suppose that both j_1 and j_2 do not exist, then inequality (4.7) becomes

$$2y_p \geq 0,$$

which trivially holds.

Therefore, inequality (4.7) is valid for clconv($\mathcal{P}$). $\square$

Remark 5. *From inequality* (4.8)*, it is tempting to think that inequality* (4.7) *is generated by simply adding up two mixing inequalities* (4.5) *and* (4.6)*, for* $S = R$ *and* $T = G$*, respectively. However, because* $v_{t_{\rho+1}} = -u_d$ *in inequality* (4.6) *corresponding to the set* $T = G$ *and* $v_{g_{\tau+1}} = 0$ *in inequality* (4.7) *corresponding to the set* $R \cup G$*, we can see that the new inequality* (4.7) *is stronger than the inequality obtained by adding the two mixing inequalities* (4.5) *for* $S = R$ *and* (4.6) *for* $T = G$*.*

We demonstrate Remark 5 more concretely on an example below.

Example 1. Let $m = 3$, $l_p = l_d = 0$, $u_d = 10$, $\mathbf{w} = (8, 6, 10)$ and $\mathbf{v} = (3, 4, 2)$. Consider $\Pi := 2 \to 1 \to 3$. Then $R = 3$ and $G = 2 \to 1 \to 3$, and the inequality (4.7) is given by

$$\begin{aligned} 14 \leq\ & 2y_p + (4-3)_+ z_2 + (6-10)_+ z_2 + (3-2)_+ z_1 + (8-10)_+ z_1 \\ & + (2-0)_+ z_3 + (10-0)_+ z_3 \\ & = 2y_p + z_2 + z_1 + 12z_3. \end{aligned}$$

In Appendix E, we give the complete convex hull description of $\mathcal{P}$ for Example 1. Indeed, in Example 1, $\text{clconv}(\mathcal{P})$ is simply described by including the classes of inequalities (4.5)-(4.7).

Next, we study the polyhedral structure of $\text{clconv}(\mathcal{P})$ under certain assumptions. In particular, we establish conditions under which $\text{clconv}(\mathcal{P})$ can be obtained by adding only the classes of inequalities characterized in (4.5)-(4.7).

4.2.3 Polyhedral Study Preliminaries

We carry out our polyhedral study of $\text{clconv}(\mathcal{P})$ under two main assumptions:

A1: $w_j \geq v_j$ for all $j \in \Omega$;

A2: $u_d \geq \max_{j \in \Omega}\{w_j\} > 0$.

Assumption **A1** is reasonable because, otherwise, if $v_j > w_j$ and scenario j is satisfied, i.e., $z_j = 0$ fro some $j \in \Omega$, then we must have $y_p + y_d \geq w_j$ and $y_p - y_d \geq v_j$ from (4.4a)–(4.4b), but since $y_d \geq 0$, $y_p \geq v_j + y_d > w_j - y_d$, inequality (4.4a) is redundant in the formulation. Assumption **A2** ensures that the upper bound of y_d is sufficiently large so that it does not cut off any feasible solution with respective to inequalities (4.4a)–(4.4c), and (4.4e).

From now on, we let α and β be the permutations of scenarios such that $w_{\alpha_1} \geq w_{\alpha_2} \geq \cdots \geq w_{\alpha_m}$, and $v_{\beta_1} \geq v_{\beta_2} \geq \cdots \geq v_{\beta_m}$.

First, we present several results that are used to conduct our polyhedral study.

Observation 18. *Consider a point* $(\bar{y}_p, \bar{y}_d, \bar{\mathbf{z}}) \in \mathcal{P}$. *Define the set* $V(\bar{\mathbf{z}}) := \{j \in [m] : \ \bar{z}_j = 0\}$.

1. *For any* $j' \in V(\bar{\mathbf{z}})$, *the point* $(\bar{y}_p, \bar{y}_d, \bar{\mathbf{z}} + \mathbf{e}_{j'})$ *is also in* $\mathcal{P}$.

2. *Whenever* $\bar{y}_d = 0$, *the point* $(\max\{\bar{y}_p, w_{j'}\}, 0, \bar{\mathbf{z}} - \mathbf{e}_{j'})$ *for any* $j' \in [m] \setminus V(\bar{\mathbf{z}})$ *is also in* $\mathcal{P}$.

3. *The point* $(\bar{y}_p + \triangle, \bar{y}_d, \bar{\mathbf{z}})$, *where* $\triangle > 0$, *is also in* $\mathcal{P}$.

Proof. Given $(\bar{y}_p, \bar{y}_d, \bar{\mathbf{z}}) \in \mathcal{P}$, let $V := V(\bar{\mathbf{z}})$, i.e., $j \in V$ if and only if $\bar{z}_j = 0$.

1. Since $(\bar{y}_p, \bar{y}_d, \bar{\mathbf{z}}) \in \mathcal{P}$, we have $\bar{y}_p + \bar{y}_d \geq \max_{j \in V} w_j$, $\bar{y}_p - \bar{y}_d \geq \max_{j \in V} v_j$, $\bar{y}_p \leq 0$, and $u_d \geq \bar{y}_d \geq 0$. Then for any $j' \in V$, the point $(\bar{y}_p, \bar{y}_d, \bar{\mathbf{z}} + \mathbf{e}_{j'})$ satisfies inequalities (4.4a) and (4.4b) because $\max_{j \in V} w_j \geq \max_{j \in V \setminus \{j'\}} w_j$, and $\max_{j \in V} v_j \geq \max_{j \in V \setminus \{j'\}} v_j$. In addition, because $\bar{y}_p$ and $\bar{y}_d$ remain the same,

inequalities (4.4c)–(4.4e) are also trivially satisfied. Hence, the point $(\bar{y}_p, \bar{y}_d, \bar{\mathbf{z}} + \mathbf{e}_{j'})$ is also in $\mathcal{P}$.

2. Since $(\bar{y}_p, 0, \bar{\mathbf{z}}) \in \mathcal{P}$, we have $\bar{y}_p \geq \max_{j \in V} w_j$, $\bar{y}_p \geq \max_{j \in V} v_j$. Hence, $\max\{\bar{y}_p, w_{j'}\} \geq \max_{j \in V \cup \{j'\}} w_j \geq \max_{j \in V \cup \{j'\}} v_j$ where the last inequality follows from Assumption **A1**, and $\max\{\bar{y}_p, w_{j'}\} \geq 0$ holds because $\bar{y}_p \geq 0$. Thus, inequalities (4.4a)–(4.4c) are satisfied. Inequalities (4.4d) and (4.4e) are also trivially satisfied. Hence, the point $(\max\{\bar{y}_p, w_{j'}\}, 0, \bar{\mathbf{z}} - \mathbf{e}_{j'})$ is also in $\mathcal{P}$.

3. This part follows because there are no constraints in $\mathcal{P}$ that can impose an upper bound on the variable y_p.

$\square$

Next, we present classes of points that are critical in our convex hull characterization:

Lemma 19. *The following points are in $\mathcal{P}$:*

$$A(V): \quad \Big(\max_{j \in V} w_j, 0, \sum_{j \in \Omega \setminus V} \mathbf{e}_j\Big), \qquad V \subseteq \Omega, \tag{4.9a}$$

$$B(V): \quad \Big(\max_{j \in V} v_j + u_d, u_d, \sum_{j \in \Omega \setminus V} \mathbf{e}_j\Big), \qquad \emptyset \neq V \subseteq \Omega, \tag{4.9b}$$

$$C(V): \quad \Big(\frac{\max_{j \in V} w_j + \max_{j \in V} v_j}{2}, \frac{\max_{j \in V} w_j - \max_{j \in V} v_j}{2}, \sum_{j \in \Omega \setminus V} \mathbf{e}_j\Big), \qquad V \subseteq \Omega, \tag{4.9c}$$

$$D: \quad (0, u_d, \mathbf{1}), \tag{4.9d}$$

where $A(\emptyset) = C(\emptyset) = (0, 0, \mathbf{1})$.

Proof. The points listed above satisfy inequality (4.4e) trivially.

Recall our convention that for $V = \emptyset$ and $\mathbf{a} \in \mathbb{R}^n$, we define $\max_{j\in V} a_j = 0$. Then because $u_d > 0$ (from Assumption **A2**), all of the points $A(V)$ for $V \subseteq \Omega$, $B(V)$ for $\emptyset \neq V \subseteq \Omega$, and D immediately satisfy inequalities (4.4d). The points $C(V)$ for $V \subseteq \Omega$ also satisfy inequalities (4.4d) because $u_d \geq \max_{j\in V} w_j \geq \frac{\max_{j\in V} w_j - \max_{j\in V} v_j}{2} \geq 0$ holds from Assumptions **A2**, $\mathbf{A}^*$, and **A1**, respectively.

The point D satisfies inequalities (4.4c)–(4.4e) trivially. It is easy to see that the point D also satisfies inequalities (4.4a)–(4.4b) because $u_d > 0$.

Clearly, $A(\emptyset) \in \mathcal{P}$. For a given $\emptyset \neq V \subseteq \Omega$, starting from the fact that $A(\emptyset) \in \mathcal{P}$ and repeatedly applying Observation 18(ii) for the indices $j \in V$, we observe that the point $A(V)$ is feasible.

Next, the point $B(V)$, for any $\emptyset \neq V \subseteq \Omega$, satisfies inequalities (4.4a) and (4.4b), because $y_p + y_d = 2u_d + \max_{j\in V} v_j > \max_{j\in V} w_j$ from Assumptions $\mathbf{A}^*$ (or **A1**) and **A2**, and $y_p - y_d = \max_{j\in V} v_j$, respectively. In addition, $B(V)$ also satisfies inequalities (4.4c) because $u_d \geq \max_{j\in V} w_j \geq \max_{j\in V} v_j$ from Assumptions **A1** and **A2**.

Finally, the point $C(V)$, for any $V \subseteq \Omega$, satisfies inequalities (4.4a) and (4.4b), because $y_p + y_d = \max_{j\in V} w_j$, and $y_p - y_d = \max_{j\in V} v_j$, respectively. In addition, $C(V)$ also satisfies inequalities (4.4c) from Assumption $\mathbf{A}^*$. □

The points in Lemma 19 are also useful in characterization of the extreme points of $\mathrm{clconv}(\mathcal{P})$.

Proposition 20. *The only recessive direction of* $\mathrm{clconv}(\mathcal{P})$ *is* $(1, 0, \mathbf{0})$*. The extreme points of* $\mathrm{clconv}(\mathcal{P})$ *are* $A(V)$ *and* $C(V)$*, for all* $V \subseteq \Omega$*,* $B(V)$*, for all* $\emptyset \neq V \subseteq \Omega$ *and* D*, as defined in equations* (B.3)–(B.6) *in Lemma 19.*

Proof. From Observation 18 (iii), $(1, 0, \mathbf{0})$ is a recessive direction of $\mathcal{P}$. Moreover, there are no other recessive directions of $\mathrm{clconv}(\mathcal{P})$ because y_p is bounded from below, and y_d and $\mathbf{z}$ are bounded from above and below. Moreover, from Lemma 19, the points $A(V)$, $C(V)$ for some $V \subseteq \Omega$, $B(V)$ for some $\emptyset \neq V \subseteq \Omega$, and D are all in $\mathcal{P}$.

Next, for a fixed $V \subseteq \Omega$, let $\widehat{\mathcal{P}}(V)$ be the resulting polyhedron of $(y_p, y_d, \sum_{j \in \Omega \setminus V} \mathbf{e}_j)$, i.e., $\widehat{\mathcal{P}}(V) :=$

$$\left\{ (y_p, y_d, \sum_{j \in \Omega \setminus V} \mathbf{e}_j) \mid \; y_p + y_d \geq \max_{j \in V} w_j, \; y_p - y_d \geq \max_{j \in V} v_j, \; y_p \geq 0, u_d \geq y_d \geq 0 \right\}. \tag{4.10}$$

When $V = \emptyset$, the only extreme points of $\widehat{\mathcal{P}}(\emptyset)$ are $A(\emptyset) = C(\emptyset)$ and $(u_d, u_d, \mathbf{1})$. Moreover, because $u_d > 0$, the point $(u_d, u_d, \mathbf{1})$ can be obtained from the point D and the recessive direction $(1, 0, \mathbf{0})$. For a given $\emptyset \neq V \subseteq \Omega$, under Assumptions $\mathbf{A}^*$, **A1**, and **A2**, Figure 4.1 illustrates the projection of the region $\widehat{\mathcal{P}}(V)$ onto the space of (y_p, y_d). We then immediately observe from Figure 4.1 that $A(V)$, $B(V)$, and $C(V)$ are the only extreme points of $\widehat{\mathcal{P}}(V)$.

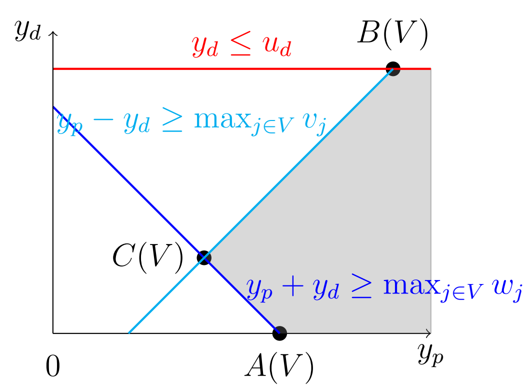

Figure 4.1: Projection of $\widehat{\mathcal{P}}(V)$ onto the space of (y_p, y_d) under Assumptions $\mathbf{A}^*$, **A1**, and **A2**

Note that $\mathcal{P} = \bigcup_{V\subseteq \Omega} \widehat{P}(V)$ and the recessive direction of $\widehat{P}(V)$ for any $V \subseteq \Omega$ is $(1, 0, \mathbf{0})$ for all $V \subseteq \Omega$. As a result, clconv($\mathcal{P}$) is simply convex combinations of the points of the form $A(V)$, $C(V)$, for some $V \subseteq \Omega$, $B(V)$ for some $\emptyset \neq V \subseteq \Omega$, and D, and conical combination of $(1, 0, \mathbf{0})$. □

Next, we address the complexity of optimizing over $\mathcal{P}$. Given a linear objective function $(c_p, c_d, \mathbf{f})$, we denote the cost of a given solution $(y_p, y_d, \mathbf{z})$ by $F((y_p, y_d, \mathbf{z})) := c_p y_p + c_d y_d + \mathbf{f}^\top \mathbf{z}$.

Proposition 21. *Let $(c_p, c_d, \mathbf{f})$ be an arbitrary nonzero cost vector. Then the optimization problem*

$$\min_{(y_p, y_d, \mathbf{z})\in\mathcal{P}} \quad c_p y_p + c_d y_d + \mathbf{f}^\top \mathbf{z} \tag{4.11}$$

can be solved in $O(m^3)$ time.

Proof. Note that if the problem is not unbounded (i.e., $c_p \geq 0$), then there exists an optimal solution that is an extreme point of clconv($\mathcal{P}$). Let $V_A^* \subseteq \Omega$ be such that $A_V^* = \arg\min_{V\subseteq\Omega} F(A(V))$, in other words, $A(V_A^*)$ is the solution among all solutions of the form $A(V)$ that gives the minimum objective. Define V_B^* and V_C^* similarly for the solutions of the form $B(V)$ and $C(V)$, respectively. Then the optimal solution is given by $\min\{F(A(V_A^*)), F(B(V_B^*)), F(C(V_C^*)), F(D)\}$. Finding A_V^* and B_V^* takes $O(m \log m)$ time, because this is equivalent to optimizing over the mixing set [c.f. 4, 38]. Hence, we address the complexity of finding V_C^*. Recall that $A(\emptyset) = C(\emptyset)$. Therefore, we consider a slightly different problem of finding $V_C^* = \arg\min_{\emptyset\neq V\subseteq\Omega} F(C(V))$.

For $i, j \in \Omega$, we define $\Omega_{ij} = \{k \in \Omega : \ w_k \leq w_i, \ v_k \leq v_j\}$ and let $G(i, j)$ be the objective value of the best extreme point of form $C(V)$, for some set V satisfying

$\{i,j\} \subseteq V \subseteq \Omega_{i,j}$. From the definition of $\Omega_{i,j}$, we have $w_i = \max_{\ell \in V} w_\ell$ and $v_j = \max_{\ell \in V} v_\ell$ for any set V satisfying $\{i,j\} \subseteq V \subseteq \Omega_{i,j}$. Therefore, for fixed $i,j \in \Omega$, we have

$$G(i,j) = \min_{\{i,j\} \subseteq V \subset \Omega_{ij}} \left\{ c_p \frac{w_i + v_j}{2} + c_d \frac{w_i - v_j}{2} + \sum_{\ell \in (\Omega \setminus V)} f_\ell \right\}. \tag{4.12}$$

Next, we show that for a given $i,j \in \Omega$, the optimal set $V_{ij} \subseteq \Omega_{ij}$ minimizing (4.12) can be found in polynomial time. By feasibility, $i,j \in V_{ij}$. For all $\ell \in \Omega$ such that $w_\ell > w_i$, or $v_\ell > v_j$, we have $\ell \notin V_{ij}$ from the definition of Ω_{ij}. Next, for all $\ell \in \Omega$ such that $\ell \neq i,j$, and $w_\ell \leq w_i$, and $v_\ell \leq v_j$, if $f_\ell > 0$, we must have $\ell \notin V_{ij}$ to minimize the cost. Otherwise, if $f_\ell \leq 0$, we let $\ell \in V_{ij}$. Hence, from this procedure, for a fixed $i,j \in \Omega$, we can find the optimal $G(i,j)$ in $\mathrm{O}(m)$ time. Finally, the optimal $V_C^* = V_{i^*,j^*}$, where $(i^*,j^*) = \arg\min_{i,j \in \Omega} G(i,j)$. Hence, the overall complexity is $O(m^3)$. □

While Proposition 21 brings good news by demonstrating an efficient algorithm to optimize over $\mathcal{P}$, in the cases where $\mathcal{P}$ arises as a substructure, such as our motivation originating from two-sided chance constrained optimization problems, we cannot immediately use Proposition 21. On the other hand, strong valid inequalities for $\mathcal{P}$ can immediately be employed in the cases where $\mathcal{P}$ arises as a substructure. Consequently, we examine the strength of the inequalities (4.5)-(4.7).

4.2.4 When are Inequalities (4.5)-(4.7) Facets of $\mathrm{clconv}(\mathcal{P})$?

In this section, we establish conditions under which inequalities (4.5)-(4.7) are facet-defining for $\mathrm{clconv}(\mathcal{P})$.

We first establish that $\mathrm{clconv}(\mathcal{P})$ is full dimensional under Assumptions .

Proposition 22. *Consider the points $A(\emptyset)$, $A(\Omega)$, $B(\Omega)$, $(w_{\alpha_1}, 0, \mathbf{e}_j)$ for all $j \in \Omega \setminus \{\alpha_1\}$, and $(w_{\alpha_1} + \triangle, 0, \mathbf{0})$, where $\triangle > 0$ is a small number. All of these points are in $\mathcal{P}$. Moreover,* $\dim\Big(\mathrm{clconv}(\mathcal{P})\Big) = m + 2$.

Proof. From Lemma 19, we know that $A(\emptyset)$, $A(\Omega)$ and $B(\Omega)$ are feasible. Next, using Observation 18(i) starting from $A(\Omega) = (w_{\alpha_1}, 0, \mathbf{0})$, we know that the point $(w_{\alpha_1}, 0, \mathbf{e}_j)$, for all $j \in \Omega \setminus \{\alpha_1\}$, is feasible. Also, the feasibility of the point $A(\Omega)$ along with Observation 18(iii) established that the point $(w_{\alpha_1} + \triangle, 0, \mathbf{0})$ is feasible.

Moreover, these points are affinely independent. Let $P_0 = A(\Omega) = (w_{\alpha_1}, 0, \mathbf{0})$, $P_1 = A(\emptyset) = (0, 0, \mathbf{1})$, $P_2 = (w_{\alpha_1} + \triangle, 0, \mathbf{0})$, $P_{2+i} = (w_{\alpha_1}, 0, \mathbf{e}_j)$, for all $i = [m-1]$ and $j \in \Omega \setminus \{\alpha_1\}$, and $P_{m+2} = B(\Omega) = (v_{\beta_1} + u_d, u_d, \mathbf{0})$. Then, $P_i - P_0$, for all $i \in [m+2]$ are linearly independent. Hence, P_i, for all $i \in [m+2] \cup \{0\}$, are affinely independent. □

Let us next examine the mixing inequalities (4.5) and (4.6).

Proposition 23. *[4, 38] Consider the setup of Proposition 16. Inequalities* (4.5) *and* (4.6) *are facet-defining for* $\mathrm{clconv}(\mathcal{P})$ *if and only if* $w_{s_1} = w_{\alpha_1}$, *and* $v_{t_1} = v_{\beta_1}$, *respectively.*

Proof. To see the necessity condition, $w_{s_1} = w_{\alpha_1}$, for inequality (4.5) to be a facet, if $w_{s_1} < w_{\alpha_1}$, then consider inequality (4.5) for $S = \{\alpha_1, s_1, \ldots, s_\eta\}$ given by

$$y_p + y_d + \sum_{j=1}^{\eta} (w_{s_j} - w_{s_{j+1}}) z_{s_j} \geq w_{s_1} + (w_{\alpha_1} - w_{s_1})(1 - z_{\alpha_1}).$$

The resulting inequality is stronger than the original inequality, because $(w_{\alpha_1} - w_{s_1})(1 - z_{\alpha_1}) \geq 0$. Hence, this establishes the necessity condition, $w_{s_1} = w_{\alpha_1}$, for inequality (4.5). The argument for the necessity condition, $v_{t_1} = v_{\beta_1}$, for inequality (4.6) is identical.

To see that inequality (4.5) is facet defining if $w_{s_1} = w_{\alpha_1}$, first, for all $j \in \Omega \setminus S$, we consider the points $(w_{\alpha_1}, 0, \mathbf{e}_j)$. These points are feasible (see the proof of Proposition 33). In addition, these points satisfy inequality (4.5) at equality and are affinely independent. Next, for all $j \in [\eta]$, we consider the points $(w_{s_j}, 0, \sum_{i \in \Omega \setminus (\cup_{i=j}^{m} s_i)} \mathbf{e}_{s_i}) = A\Big(\cup_{i=j}^{\eta} s_i\Big)$, for all $j \in [\eta]$. From Lemma 19, we know that these points are feasible. In addition, these points satisfy inequality (4.5) at equality and are affinely independent. Finally, we consider the feasible points $A(\emptyset)$ and $C(\Omega)$, which are affinely independent from all other points. In addition, they satisfy inequality (4.5) at equality. Hence, we obtain $m+2$ affinely independent points that are feasible and satisfy inequality (4.5) at equality, which indicates that inequality (4.5) is facet-defining for clconv($\mathcal{P}$).

The sufficiency proof for inequality (4.6) is similar to the sufficiency proof of inequality (4.5), where we consider the points $B(\Omega)$, $C(\Omega)$, $C(\Omega \setminus \{j\})$, for all $j \in \Omega \setminus T$, and $B\Big(\cup_{i=j}^{\rho} t_i\Big)$, for all $j \in [\rho]$. These points are feasible from Lemma 19 and are also affinely independent. $\square$

Next, we study the strength of the proposed inequalities (4.7).

Proposition 24. *Consider the setup of Proposition 17. Given $\tau \in [m]$ and a sequence of scenarios $\Pi := \{\pi_1 \to \pi_2 \to \cdots \to \pi_\tau\}$, let $R \subseteq \Pi$ be a subsequence of scenarios given by $r_1 \to r_2 \to \cdots \to r_{\tau_R}$, with $|R| = \tau_R \leq \tau$, such that $r_j \in R$, for all $j \in [\tau_R]$, if $w_{r_j} \geq \bar{w}_{r_j}$. In addition, let $G \subseteq \Pi$ be a subsequence of scenarios given by $g_1 \to g_2 \to \cdots \to g_{\tau_G}$, with $|G| = \tau_G \leq \tau$, such that $g_j \in G$, for all $j \in [\tau_G]$, if $v_{g_j} \geq \bar{v}_{g_j}$. Then, inequality (4.7) is facet-defining for* clconv($\mathcal{P}$) *if and only if $w_{r_1} = w_{\alpha_1}$ and $v_{g_1} = v_{\beta_1}$.*

Proof. Suppose that $w_{r_1} < w_{\alpha_1}$. We can attach scenario α_1 at the *beginning* of the sequence Π to obtain another valid inequality of form (4.7)

$$2y_p + \sum_{j=1}^{\tau_R}(w_{r_j} - w_{r_{j+1}})z_{r_j} + \sum_{j=1}^{\tau_G}(v_{g_j} - v_{g_{j+1}})z_{g_j} \geq w_{r_1} + v_{g_1} + (w_{\alpha_1} - w_{r_1})(1 - z_{\alpha_1}).$$

The resulting inequality is at least as strong as the original inequality because $w_{\alpha_1} > w_{r_1}$ and $1 - z_{\alpha_1} \geq 0$. We can apply a similar argument for the case where $v_{g_1} < v_{\beta_1}$. This shows the necessity of the facet conditions.

To see the sufficiency, first consider the feasible points $A(\emptyset)$ and D. It can be seen that they satisfy inequality (4.7) at equality. Next, we consider the feasible point $C(\Omega)$, which satisfies inequality (4.7) at equality. Now, consider the points $(\frac{w_{\alpha_1}+v_{\beta_1}}{2}, \frac{w_{\alpha_1}-v_{\beta_1}}{2}, \mathbf{e}_j)$, for all $j \in \Omega \setminus \Pi$. For each $j \in \Omega \setminus \Pi$, using Observation 18(i) and the feasibility of the point $C(\Omega) = (\frac{w_{\alpha_1}+v_{\beta_1}}{2}, \frac{w_{\alpha_1}-v_{\beta_1}}{2}, \mathbf{0})$, we conclude that these points are also feasible. Since $j \notin \Pi$, these points satisfy (4.7) at equality as well. Note that the points considered thus far are affinely independent.

Next, for all $j \in [\tau] \setminus \{1\}$ such that $\pi_j \in \Pi$, if $w_{\pi_j} < \bar{w}_{\pi_j}$ and $v_{\pi_j} < \bar{v}_{\pi_j}$, then we consider the point $(\frac{w_{\alpha_1}+v_{\beta_1}}{2}, \frac{w_{\alpha_1}-v_{\beta_1}}{2}, \mathbf{e}_{\pi_j})$. For each such j, the feasibility of the associated point follows from the feasibility of $C(\Omega)$ and Observation 18(i). In addition, this point also satisfies inequality (4.7) at equality, because $(w_{\pi_j} - \bar{w}_{\pi_j})_+ = (v_{\pi_j} - \bar{v}_{\pi_j})_+ = 0$, so the left-hand side of inequality (4.7), after substituting this point, becomes $w_{\alpha_1} + v_{\beta_1}$. Otherwise, if $w_{\pi_j} \geq \bar{w}_{\pi_j}$ or $v_{\pi_j} \geq \bar{v}_{\pi_j}$, then we consider the following feasible point $(\frac{\bar{w}_{\pi_{j-1}}+\bar{v}_{\pi_{j-1}}}{2}, \frac{\bar{w}_{\pi_{j-1}}-\bar{v}_{\pi_{j-1}}}{2}, \sum_{i=1}^{j-1}\mathbf{e}_{\pi_i} + \sum_{i\in(\Omega\setminus\Pi)}\mathbf{e}_i) = C\Big((\Omega \setminus \Pi) \setminus (\cup_{i=1}^{j-1}\{\pi_i\})\Big)$, for all $j \in [\tau] \setminus \{1\}$. Because

$$\sum_{i=1}^{j-1}\left(\left(w_{\pi_i} - \bar{w}_{\pi_i}\right)_+\right) + \bar{w}_{\pi_{j-1}} = \max_{\ell\in[\tau]} w_{\pi_\ell} = \bar{w}_{\pi_0},$$

and

$$\sum_{i=1}^{j-1}\left(\left(v_{\pi_i}-\bar{v}_{\pi_i}\right)_+\right)+\bar{v}_{\pi_{j-1}}=\max_{\ell\in[\tau]}\ v_{\pi_\ell}=\bar{v}_{\pi_0},$$

this point satisfies inequality (4.7) at equality. In addition, these points are affinely independent from the points listed earlier. Hence, in total, we obtain $m+2$ affinely independent feasible points that satisfy inequality (4.7) at equality. This completes the proof. □

Example 1(continued). Consider the inequality $2y_p + z_2 + z_1 + 12z_3 \geq 14$ derived from $\Pi := 2 \to 1 \to 3$. Because this inequality also satisfies $\alpha_1 \in \Pi$ and $\beta_1 \in \Pi$, it is facet-defining.

4.2.5 Separation of Inequalities of Form (4.7)

In this section, we give a polynomial-time dynamic programming algorithm to separate inequality (4.7). Let $(\hat{y}_p, \hat{y}_d, \hat{\mathbf{z}})$ be a fractional solution. In order to find the most violated inequality (4.7), we need to find a sequence $\Pi = \{\pi_1 \to \pi_2 \to \cdots \to \pi_\tau\}$ that minimizes the value of the term $\sum_{j=1}^{\tau}\Big((w_{\pi_j}-\bar{w}_{\pi_j})_+ + (v_{\pi_j}-\bar{v}_{\pi_j})_+\Big)\hat{z}_{\pi_j}$. Throughout the rest of our discussion, this value is interpreted as *cost*. Without loss of generality, we assume that the sequence Π has length m. We will later show how the resulting sequence can be shortened to a length $\tau \leq m$. Here, we only consider the case where $\alpha_1 \in \Pi$ and $\beta_1 \in \Pi$, because otherwise the resulting inequality can be strengthened by including α_1 and β_1.

In the proposed algorithm, the state function is:

$$\bar{G}_i(j, \bar{w}_{\pi_{i-1}}, \bar{v}_{\pi_{i-1}}), \quad i \in [m],\ j \in \Omega,\ \bar{w}_{\pi_{i-1}} \geq w_j,\ \bar{v}_{\pi_{i-1}} \geq v_j,$$

which is defined as the *minimum* cost of the subsequence $\pi_i \to \pi_{i+1} \to \cdots \to \pi_m$, where scenario j is the first scenario in this subsequence (i.e., $\pi_i = j$), $\max\{w_{\pi_j} : i \leq$

$j \le m\} = \bar{w}_{\pi_{i-1}}$, and $\max\{v_{\pi_j} : i \le j \le m\} = \bar{v}_{\pi_{i-1}}$. Note that there are $\mathrm{O}(m^4)$ many possible states.

Next, the boundary condition is defined as:

$$\bar{G}_m(j, \bar{w}_{\pi_{m-1}}, \bar{v}_{\pi_{m-1}}) = \begin{cases} (w_j + v_j)\hat{z}_j, & \text{if } \bar{w}_{\pi_{m-1}} = w_j, \text{ and } \bar{v}_{\pi_{m-1}} = v_j, \\ +\infty, & \text{if } \bar{w}_{\pi_{m-1}} > w_j, \text{ or } \bar{v}_{\pi_{m-1}} > v_j, \end{cases}$$

where the state $\bar{G}_m(j, \bar{w}_{\pi_{m-1}}, \bar{v}_{\pi_{m-1}})$, in which $\bar{w}_{\pi_{m-1}} > w_j$ or $\bar{v}_{\pi_{m-1}} > v_j$ is infeasible, because if scenario $j = \pi_m$, then we must have $\bar{w}_{\pi_{m-1}} = w_j$ and $\bar{v}_{\pi_{m-1}} = v_j$. The optimal solution is then given by

$$\min\left\{\bar{G}_1(\alpha_1, w_{\alpha_1}, v_{\beta_1}),\ \bar{G}_1(\beta_1, w_{\alpha_1}, v_{\beta_1})\right\},$$

because α_1 and β_1 are in Π, and without loss of generality, we have $w_{\pi_1} = w_{\alpha_1}$ or $v_{\pi_1} = v_{\beta_1}$.

Next, we give the backward transition function

$$\bar{G}_i(j, \bar{w}_{\pi_{i-1}}, \bar{v}_{\pi_{i-1}}) =$$
$$\begin{cases} \min_{j' \in \Omega} \left\{\bar{G}_{i+1}(j', \bar{w}_{\pi_{i-1}}, \bar{v}_{\pi_{i-1}})\right\}, \\ \quad \text{if } \bar{w}_{\pi_{i-1}} > w_j, \text{ and } \bar{v}_{\pi_{i-1}} > v_j, \\ \min_{j' \in \Omega, \bar{w}_{\pi_i} \le w_j, \bar{v}_{\pi_i} \le v_j} \left\{\bar{G}_{i+1}(j', \bar{w}_{\pi_i}, \bar{v}_{\pi_i}) + (w_j + v_j - \bar{w}_{\pi_i} - \bar{v}_{\pi_i})\hat{z}_j\right\}, \\ \quad \text{if } \bar{w}_{\pi_{i-1}} = w_j, \text{ and } \bar{v}_{\pi_{i-1}} = v_j, \\ \min_{j' \in \Omega, \bar{w}_{\pi_i} \le w_j} \left\{\bar{G}_{i+1}(j', \bar{w}_{\pi_i}, \bar{v}_{\pi_{i-1}}) + (w_j - \bar{w}_{\pi_i})\hat{z}_j\right\}, \\ \text{if } \bar{w}_{\pi_{i-1}} = w_j, \text{ and } \bar{v}_{\pi_{i-1}} > v_j, \\ \min_{j' \in \Omega, \bar{v}_{\pi_i} \le v_j} \left\{\bar{G}_{i+1}(j', \bar{w}_{\pi_{i-1}}, \bar{v}_{\pi_i}) + (v_j - \bar{v}_{\pi_i})\hat{z}_j\right\}, \\ \text{if } \bar{w}_{\pi_{i-1}} > w_j, \text{ and } \bar{v}_{\pi_{i-1}} = v_j. \end{cases}$$

Finally, note that this recursion will not lead to sequences with cycles.

The running time of the transition function is $\mathrm{O}(m^3)$, so the total running time of this dynamic programming algorithm is $O(m^7)$.

4.2.6 Special Cases

In this section, we consider variants of polyhedron $\mathcal{P}$. [64] examine a related model for a two-sided chance-constraint given by

$$|y_d - h_j| \leq q_j + M_j z_j, \qquad \forall j \in \Omega \tag{4.13a}$$

$$z \in \mathbb{B}^m, \tag{4.13b}$$

$$l_d \leq y_d \leq u_d, \tag{4.13c}$$

To simplify notation, let $h_j = h_j + q_j$, and $q_j = h_j - q_j$, we obtain the following equivalent constraints

$$y_d \leq h_j + (u_d - h_j) z_j, \qquad \forall j \in \Omega \tag{4.14a}$$

$$y_d \geq q_j - (q_j - l_d) z_j, \qquad \forall j \in \Omega \tag{4.14b}$$

$$z \in \mathbb{B}^m, \tag{4.14c}$$

$$0 \leq y_d \leq u_d, \tag{4.14d}$$

Let α and β be the permutations of scenarios such that $h_{\alpha_1} \leq h_{\alpha_2} \leq \cdots \leq h_{\alpha_m}$, and $q_{\beta_1} \geq q_{\beta_2} \geq \cdots \geq q_{\beta_m}$, respectively.

When $h_{\alpha_1} \geq q_{\beta_1}$(i.e., we cannot find any pair of scenarios that cannot be satisfied at the same time), then conv($\mathcal{P}$) is simply given by the mixing inequalities for inequality (4.14a) and (4.14b).

Proposition 25. *Let $\mathcal{P} = \{(y_p, \mathbf{z}) \mid (y_p, \mathbf{z})$ satisfies (4.14)$\}$. In addition, let $\bar{T} := \{\bar{t}_1, \bar{t}_2, \ldots, \bar{t}_{\bar{\tau}}\}$ be any subset of scenarios such that $h_{\bar{t}_1} \leq h_{\bar{t}_2} \leq \cdots \leq h_{t_{\bar{\tau}}}$. Similarly, let $\underline{T} := \{\underline{t}_1, \underline{t}_2, \ldots, \underline{t}_{\underline{\tau}}\}$ be any subset of scenarios such that $q_{\underline{t}_1} \geq q_{\underline{t}_2} \geq \cdots \geq q_{\underline{t}_{\underline{\tau}}}$. If*

$h_{\alpha_1} \geq q_{\beta_1}$, *then* $\operatorname{conv}(\mathcal{P})$ *is given by:*

$$y_d \leq h_{\alpha_1} + \sum_{i=1}^{\bar{a}} (h_{\bar{t}_{i+1}} - h_{\bar{t}_i}) z_{\bar{t}_i}, \qquad \forall \bar{T} \subset \Omega, \tag{4.15a}$$

$$y_d + \sum_{i=1}^{\underline{\tau}} (q_{\underline{t}_i} - q_{\underline{t}_{i+1}}) z_{\underline{t}_i} \geq q_{\beta_1}, \qquad \forall \underline{T} \subset \Omega, \tag{4.15b}$$

$$0 \leq z_j \leq 1, \qquad \forall j \in \Omega \tag{4.15c}$$

$$0 \leq y_d \leq u_d, \tag{4.15d}$$

where $h_{\bar{t}_{\bar{\tau}+1}} = u_d$, and $q_{\underline{t}_{\underline{\tau}+1}} = -l_d$.

Proof. The inequalities (4.15a) and (4.15b) are mixing inequalities for (4.14a) and (4.14b), respectively. Hence, for any extreme point solution of the polytope defined by (4.15), if the inequalities (4.15a) and (4.15b) are not tight at the same time, then the solution must be integral.

Suppose that inequalities (4.15a) and (4.15b) are tight at the same time, then the only possible condition is when $w_{\alpha_1} = v_{\beta_1}$, and $\mathbf{z} = 0$. To see why, suppose that one of inequalities (4.15a) is tight, then $y_d = w_{\alpha_1} + \sum_{i=1}^{\bar{a}} (w_{\bar{t}_{i+1}} - w_{\bar{t}_i}) z_{\bar{t}_i} \geq w_{\alpha_1}$. Similarly, if one of inequalities (4.15b) is tight, then $y_d = v_{\alpha_1} - \sum_{i=1}^{\underline{a}} (v_{\underline{t}_i} - v_{\underline{t}_{i+1}}) z_{\underline{t}_i} \leq v_{\beta_1}$. If $w_{\alpha_1} \geq v_{\beta_1}$, then the only possible condition which gives us a feasible y_d is when $w_{\alpha_1} = v_{\beta_1}$, and $\mathbf{z} = 0$. Hence, any extreme point solution of (4.15) must be integral, which completes the proof. □

4.3 Preliminary Computations

In this section, we study the computational performance of the proposed inequality (4.7) against adding mixing inequalities (4.5) and (4.6) only. In the tested instances, $y_p = \sum_{i=1}^{5} p_i x_i$, $y_d = \sum_{i=1}^{5} d_i x_i$, and p_i and d_i follow uniform distribution $U[0, 1]$. In addition, c_p follows $U[0, 1]$, c_d follows $U[0, 0.5]$. Furthermore, v_j follows discrete

Table 4.1: Preliminary Computational Results

Instances		DEP & Mix. Ineq. (4.5), (4.6)				DEP & Mix.Ineq. (4.5), (4.6) & New Ineq.			
ϵ	m (10^3)	Time	Gap (%)	R.Gap (%)	Nodes	Time	Gap (%)	R.Gap (%)	Nodes
0.1	1	21	0	20.4	2004	14	0	10.84	1405
	2	63	0	16.4	2042	50	0	16.13	992
	3	182	0	17.8	12452	151	0	16.38	11324
	4	453	0	18.5	33069	366	0	15.5	17081
0.2	1	110	0	22.4	32580	72	0	19.3	17141
	2	1429	0	31.1	258478	329	0	26.7	87335
	3	1536	0	25.3	226028	729	0	23.6	120003
	4	3021(1)	1.12	35.2	400642	2883(2)	1.05	34.2	392047

uniform $[20, 40]$, w_j follows discrete uniform $[60, 80]$, for all $j \in \Omega$, and each scenario is equally likely.

In our computational study, we apply the polynomial separation algorithms studied in [4] for inequalities (4.6) and (4.7), to find the minimum mixing sets S^* and T^*. Although a polynomial separation algorithm for (4.7) is proposed in Section 4.2.5, it is inefficient because its complexity is $O(m^7)$. In our computational experiments, we use heuristic to separate inequality (4.7). First, we get the minimum S^* and T^*. Then, we generate inequality (4.7) using the following sequence: $\Pi = t_1 \to t_2 \to \cdots \to t_\rho \to s_1 \to s_2 \to \cdots \to s_\eta$, where we append every scenario in S^* after set T^*. It can be seen that the sequence Π generated by this procedure minimizes the term $\sum_{j=1}^{\tau}(w_{\pi_j} - \bar{w}_{\pi_j})_+ \hat{z}_{\pi_j}$.

All runs were executed on a Windows Server 2012 R2 Data Center with 2.40GHZ Intel(R) Xeon(R) CPU and 32.0 GB RAM. The algorithms tested in the computational experiment were implemented using C programming language, with Microsoft Visual Studio 2012 and CPLEX 12.6. A time limit of one hour is set. The preliminary computational results are given in Table 4.1.

In Table 4.1, the column "DEP & Mix. Ineq. (4.5), (4.6)" reports the results of instances with mixing inequalities (4.5) and (4.6) only, and the column "DEP & Mix.Ineq. (4.5), (4.6) & New Ineq." reports the results of instances with both mixing inequalities, and the proposed new inequalities (4.7). We only add these inequalities to the root node. In addition, each entry averages the results of three instances. Since inequalities (4.5) - (4.7) may not be facet-defining if we intersect clconv($\mathcal{P}$) and the knapsack constraint (4.2c), in the computation, the number of mixing inequalities (4.5), (4.6) and new inequalities (4.7) that can be added is limited to 150 each. Based on the results, these limits are hit by every instance, for all valid inequalities. The "Time" column reports the average solution time in seconds for the instances that are solved to optimality within the time limit, and the "Gap" column reports the average optimality gap for the instances that reach the time limit. If not all three instances are solved to optimality within time limit, then the number inside the parenthesis under the "Time" column indicates the number of instances that are solved to optimality within time limit. The " - " sign under the "Time" column indicates that no instance is solved to optimality within the time limit. In addition, the "R.Gap" column reports the root node gap for the instances. Furthermore, the "Nodes" column averages the number of branch-and-bound nodes explored during the process.

According to Table 4.1, the proposed inequalities are computationally beneficial: the solution time, ending gap, root node gap and number of branch-and-bound nodes added are generally better for the option with the new inequality (4.7) than the option without the new inequality (4.7).

Chapter 5: Robust Multicriteria Risk-Averse Stochastic Programming Models

5.1 Introduction

This chapter is based on [62]. For many decision making problems under uncertainty, it may be essential to consider multiple possibly conflicting stochastic performance criteria. Stochastic multicriteria decision making problems arise in a wide range of areas, including financial portfolio optimization, humanitarian relief network design, production scheduling, and homeland security budget allocation [see, e.g., 43, 50, 74]. In such problems, we can represent the stochastic outcomes of interest by a random vector, each dimension of which corresponds to a particular decision criterion. Then, comparing the potential decisions requires specifying preference relations among random vectors. It is also crucial to compare the random outcomes based on the decision makers' risk preferences. These concerns call for optimization models that incorporate multivariate risk-averse preference relations into *constraints* and/or *objectives*. The class of models, which incorporates the multivariate risk preferences into the *constraints* using benchmarking relations, has received some attention in the recent literature. Alternatively, in this study, we introduce a new class of models with an *objective* of optimizing a multivariate risk measure.

First, we review the existing literature on risk-averse multicriteria optimization models that feature benchmarking preference relations. In this line of research initiated by [28], two types of benchmarking relations are modeled as constraints: multivariate risk-averse relations based on second-order stochastic dominance (SSD) and conditional value-at-risk (CVaR). These models assume that a benchmark random outcome vector is available and extend univariate (scalar-based) preference rules to the multivariate (vector-based) case by using linear scalarization functions. The linear scalarization corresponds to the *weighted-sum* approach, which is widely used in multicriteria decision making [96, 31]; the scalarization coefficients are interpreted as the weights representing the relative (subjective) importance of each decision criterion.

In many decision making situations, especially those involving multiple decision makers, it can be difficult to determine a single weight vector. There are many alternative methods to elicit relative weights of each criterion, including multiattribute weighting, swing weighting and the analytic hierarchy process [for a review, see 99, 87]. However, the relative weights of even a single expert could be very different depending on which elicitation approach is used as shown in [89] and [19]. The problem of choosing a single weight vector is further exacerbated if multiple experts are involved. To address these ambiguity and inconsistency issues, a so-called *robust approach* considers a collection of weight vectors within a prescribed *scalarization set* instead of a single weight vector. Various scalarization sets are considered in the literature such as the set of all non-negative coefficients, arbitrary polyhedral and arbitrary convex sets [see, e.g., 28, 42, 45, respectively].

While the majority of existing studies focuses on enforcing multivariate SSD relations [see, e.g., 28, 42, 45, 29], this modeling approach can be overly conservative in practice and leads to very demanding constraints that sometimes cannot be satisfied. For example, due to this infeasibility issue, [43] solve such an optimization problem with relaxed SSD constraints. As an alternative, [75] propose to use a multivariate preference relation based on CVaR; their approach is motivated by the fact that the univariate SSD relation is equivalent to a continuum of CVaR inequalities [27]. The authors consider polyhedral scalarization sets and show that their CVaR-based methodology can be extended to optimization problems featuring benchmarking constraints based on a wider class of coherent risk measures. In our study, we follow the line of research of [75], which provides sufficient flexibility to obtain feasible problem formulations and capture a wide range of risk preferences, including risk-neutral and worst-case approaches.

Optimization models under both types of multivariate preference relations (SSD and CVaR) are challenging, since they require introducing infinitely many univariate risk constraints associated with all possible weight vectors in the scalarization set. For polyhedral scalarization sets, [42] and [75] show that enforcing the corresponding univariate risk constraint for a finite (exponential) subset of weight vectors is sufficient to model the multivariate SSD and CVaR relations, respectively. These finite representation results allow them to develop finitely convergent delayed cut generation algorithms, where each cut is obtained by solving a mixed-integer programming (MIP) problem. Since solving these MIP formulations is the main computational bottleneck, [55] develop computationally effective solution methods for the cut generation problems arising in both types of optimization models.

As outlined earlier, the existing literature on risk-averse multicriteria optimization problems mainly focuses on *multivariate risk-constrained models*, where a benchmark random vector is available and the goal is to find a solution with a multivariate outcome vector that is preferable to the benchmark (with respect to the multivariate SSD or CVaR relation). In this chapter, we propose a novel model which *does not require a given benchmark* and aims to optimize the risk associated with the decision-based random vector of outcomes. In this sense, the problem we consider can be seen as a *risk-averse stochastic multiobjective* optimization. There are, in general, two types of approaches to solve stochastic multiobjective problems: 1) to eliminate the stochastic nature of the problem by replacing each random objective function with one of its summary statistics; 2) to eliminate the multiobjective structure of the problem by aggregating the multiple objectives and obtaining a single random objective function. For recent surveys on these two types of approaches we refer to [40] and [7]. The first (non-aggregation based) approach results in a traditional deterministic multiobjective problem and requires the identification of multiple (typically exponential) non-dominated solutions in the efficient frontier. Ultimately, however, the decision makers need to specify the weights for each criterion to choose among the non-dominated solutions. In the second (aggregation-based) approach, one can consider a weighted sum of the multiple objectives and solve the resulting stochastic problem to obtain a solution. However, the weights to be used in either approach can be *ambiguous* and *inconsistent* due to the presence of conflicting criteria and lack of consensus among multiple experts. Alternatively, in the second approach of aggregating multiple objectives into one, one can use an aggregated (but non-scalarized) single objective using stochastic goal programming. This approach considers random and/or deterministic goals (benchmark values) for the different objectives and constructs a single objective based on a function of the deviations from the goals.

However, a benchmark goal may not be immediately available in all practical applications. For problems where the relative importance of the criteria is ambiguous and a benchmark performance vector is not available, we propose to focus on the worst-case CVaR with respect to the prescribed scalarization set and employ a recent notion of CVaR robustness in the context of stochastic multicriteria optimization.

In a related line of work, to address the ambiguity and inconsistency in the weights used to scalarize the multiple criteria in the objective function of a *deterministic* optimization problem, [76] and [44] consider minimax type robustness with respect to a given weight set. Note that such existing *robust weighted-sum models* assume that either the problem parameters are deterministic or the decision makers are risk-neutral. For an overview on minimax robustness for multiobjective optimization problems we refer to [32]. However, some multicriteria decision-making problems of recent interest, such as disaster preparedness [44] and homeland security [43], involve uncertain events with small probabilities but dire consequences. Therefore, it is crucial to incorporate risk aversion into multicriteria optimization models, which is the main focus of our study. Note that the risk-averse model we propose in this chapter features the risk-neutral version as a special case.

In the recent literature, another type of CVaR robustness appears in the univariate case stemming from the distributional robustness. [108] and [105] consider optimizing the worst-case CVaR and a wider class of convex risk measures (of a scalar-based random variable), respectively. However, this line of work assumes that there is ambiguity in the underlying probability distribution and express the worst-case with respect to a specified set of distributions. [see also, 109, for Worst-Case CVaR approximation of joint chance constraints in the distributionally robust optimization framework]. In contrast, we assume that the underlying probability distribution is known but there is ambiguity in the scalarization vector (i.e., relative importance of multiple criteria) within a polyhedral set; this leads to a worst-case multivariate CVaR measure.

5.1.1 Our contributions

We incorporate risk aversion into multicriteria optimization models using the concept of multivariate CVaR. We propose a maximin type model optimizing the worst-case CVaR over a scalarization set. While the worst-case multivariate CVaR measure was recently introduced in the finance literature to assess the risk of portfolio vectors [see, e.g., 85], to the best of our knowledge, there is no model or method to optimize this risk measure. In this chapter, we fill this gap, and give an optimization model that maximizes the worst-case multivariate CVaR. To demonstrate the adequacy of the proposed model, we show that the risk measure of interest is coherent in an appropriate multivariate sense, and an optimal solution of the model is not dominated in an appropriate stochastic sense. These two properties are highly desirable in risk-averse optimization and multicriteria optimization, respectively.

Unlike the risk-neutral version with a polyhedral weight set, in the risk-averse case, the inner minimization problem involves a concave minimization. Hence, the problem in general can no longer be solved as a compact linear program [as in 44]. Therefore, we propose a delayed cut generation-based solution algorithm and show that the cut generation problem can be modeled as a bilinear program that contains the multiplication of the scalarization variables and some binary variables used for representing CVaR. We demonstrate that the assumptions on the scalarization set allow us to employ the reformulation-linearization technique (RLT) [92, 93] to strengthen the resulting MIP formulations of the cut generation problem. This observation in turn speeds up the overall solution time considerably as we show in our extensive computational study.

We observe that the cut generation subproblems in the proposed algorithm have similar structure with those encountered in solving the related multivariate CVaR-constrained optimization model. Therefore, in the second part of the chapter, we show that the RLT technique can be applied to obtain stronger and computationally more efficient formulations for the cut generation problems arising in optimization under multivariate CVaR constraints, especially for the equal probability case.

5.1.2 Outline

The rest of the chapter is organized as follows. In Section 5.2, we introduce the new worst-case CVaR optimization model and provide some analytical results to highlight the appropriateness of the proposed modeling approach. This section also presents a cut generation algorithm and effective mathematical programming formulations of the original optimization problem and the corresponding cut generation problems for some special cases. We describe how to apply some of these algorithmic features to the multivariate CVaR-constrained models in Section 5.3. Section 5.4 gives a hybrid model that includes both the multivariate CVaR-based constraints and objective. We give a unified methodology that solves the hybrid model, integrating the algorithmic developments in Sections 5.2 and 5.3. Section 5.5 is dedicated to the computational study, while Section 5.6 contains concluding remarks.

5.2 Worst-case CVaR Optimization Model

In our study, we consider a multicriteria decision making problem where d random performance measures of interest associated with the decision vector $\mathbf{z}$ are represented by the random outcome vector $\mathbf{G}(\mathbf{z}) = (G_1(\mathbf{z}), \dots, G_d(\mathbf{z}))$. All random variables in this chapter are assumed to be defined on some finite probability spaces; we simplify our exposition accordingly. Let $(\Omega, 2^\Omega, \mathbb{P})$ be a finite probability space with $\Omega = \{\omega_1, \dots, \omega_n\}$ and $\mathbb{P}(\omega_i) = p_i$. In particular, denoting the set of feasible decisions by Z, the random outcomes are determined according to the outcome mapping $\mathbf{G} : Z \times \Omega \to \mathbb{R}^d$, and the random outcome vector $G(\mathbf{z}) : \Omega \to \mathbb{R}^d$ is defined by $G(\mathbf{z})(\omega) = G(\mathbf{z}, \omega)$ for all $\omega \in \Omega$. For a given elementary event ω_i the mapping $\mathbf{g}_i : Z \to \mathbb{R}^d$ is defined by $\mathbf{g}_i(\mathbf{z}) = \mathbf{G}(\mathbf{z}, \omega_i)$. Let $C \subset \mathbb{R}^d_+$ be a polyhedron of scalarization vectors, each component of which corresponds to the relative importance of each criterion. We naturally assume, without loss of generality, that C is a subset of the unit simplex, C_f, i.e., $C \subseteq C_f := \{\mathbf{c} \in \mathbb{R}^d_+ \mid \sum_{i \in [d]} c_i = 1\}$.

Before proceeding to give our definitions and models, we need to make a note of some conventions used throughout this chapter, and recall a basic definition. The set of the first n positive integers is denoted by $[n] = \{1, \dots, n\}$, while the positive part of a number $x \in \mathbb{R}$ is denoted by $[x]_+ = \max(x, 0)$. We assume that *larger values of random variables* are preferred. We quantify the risk associated with a random variable via a risk measure (specifically, CVaR) where *higher values correspond to less risky* random outcomes. In this context, risk measures are often referred to as *acceptability functionals*. Our presentation follows along the lines of [78] and [75]. Recall that for a *univariate* random variable X with (not necessarily distinct) realizations $x_1, \dots, x_n$ and corresponding probabilities $p_1, \dots, p_n$, the *conditional value-at-risk* at confidence level $\alpha \in (0, 1]$ is given by [83]

$$\mathrm{CVaR}_\alpha(X) = \max\left\{\eta - \frac{1}{\alpha}\mathbb{E}\left([\eta - X]_+\right) \ : \ \eta \in \mathbb{R}\right\} \tag{5.1}$$

$$= \max\{\eta - \frac{1}{\alpha}\sum_{i\in[n]} p_i w_i \ : \ w_i \geq \eta - x_i, \ \forall \ i \in [n], \quad \mathbf{w} \in \mathbb{R}^n_+, \ \eta \in \mathbb{R}\} \tag{5.2}$$

$$= \max_{k\in[n]}\left\{x_k - \frac{1}{\alpha}\sum_{i\in[n]} p_i[x_k - x_i]_+\right\}, \tag{5.3}$$

where the last equality follows from the well known result that the maximum in definition (5.2) is attained at the α-quantile, which is known as the *value-at-risk* (VaR) at confidence level α (denoted by $\mathrm{VaR}_\alpha(X)$) and that $\mathrm{VaR}_\alpha(X) = x_k$ for at least one $k \in [n]$. For risk-averse decision makers typical choices for the confidence level are small values such as $\alpha = 0.05$. Note that $\mathrm{CVaR}_\alpha(X)$, as defined in (5.1), is concave in X.

The significance of modeling robustness against the ambiguity and inconsistency in relative weights motivates us to introduce a new robust optimization model for the stochastic multicriteria decision making problem of interest. To model the risk aversion of the decision makers, we use CVaR as the acceptability functional. In particular, we focus on the recently introduced worst-case multivariate CVaR [85] with respect to the specified scalarization set C, which we review next.

Definition 3 (Worst-Case Multivariate Polyhedral CVaR)**.** *Let* $\mathbf{X}$ *be a* d*-dimensional random vector and* $C \subseteq C_f$ *a set of scalarization vectors. The* worst-case multivariate polyhedral CVaR (WCVaR) *at confidence level* $\alpha \in (0, 1]$ *with respect to* C *is defined as*

$$\mathrm{WCVaR}_{C,\alpha}(\mathbf{X}) = \min_{\mathbf{c}\in C} \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}). \tag{5.4}$$

Following a risk-averse approach, we propose to optimize $\mathrm{WCVaR}_{C,\alpha}$ for a given confidence level $\alpha \in (0,1]$ and a scalarization set C, and introduce a new class of robust multicriteria optimization problems of the general form

$$\text{(W-CVaR)}: \quad \max_{\mathbf{z}\in Z} \min_{\mathbf{c}\in C} \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z})). \tag{5.5}$$

We note that the proposed risk-averse W-CVaR problem features the risk-neutral version, proposed in [44], as a special case when $\alpha = 1$. Another special case appears in the literature [31] for a sufficiently small value of α (corresponding to the worst-case); it optimizes the worst value of a particular weighted sum over the set of scenarios. This robust version of the weighted sum scalarization problem is clearly a special case of W-CVaR if we assume that all scenarios are equally likely, $\alpha = 1/n$, and there is a single scalarization vector in the scalarization set C.

It is important to note that the major difficulty of the proposed optimization problem W-CVaR, and the related models in the literature, stems from the presence of the *joint* acceptability functional $\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z}))$. One might wonder why an alternative model that maximizes the scalarization of component-wise acceptability functionals, i.e., $\max_{\mathbf{z}\in Z} \min_{\mathbf{c}\in C} \sum_{i\in[d]} c_i \mathrm{CVaR}_\alpha(G_i(\mathbf{z}))$ is not preferred. After all, this approach would lead to more tractable reformulations; for example, the alternative model can be formulated as a linear program when there is no integrality restriction on the decision vector $\mathbf{z}$, Z is a polyhedral set, and the mapping $\mathbf{g}_i(\mathbf{z})$ is linear in $\mathbf{z}$ for all $i \in [n]$. However, such a model completely ignores the correlation between the random variables $G_i(\mathbf{z}),\ i \in [d]$. The worst α proportion of scenarios with respect to one criterion would most likely not coincide with the worst α proportion of scenarios with respect to the other criteria, except for the very trivial case when $G_i(\mathbf{z}),\ i \in [d]$, are comonotone random variables. Therefore, using the aforementioned alternative modeling approach could only be justified to capture the multivariate risk in the trivial case when the worst-case scenarios of the multiple random outcomes coincide, which does not appear to be the typical situation in optimization with conflicting criteria. In all other cases, it would be a conservative approximation. In this chapter, we are interested in exact models and methods that optimize a multivariate risk measure based on the joint behavior of the random outcomes of interest.

In the remainder of this section, we first provide some analytical results to highlight the appropriateness of the proposed model (Section 5.2.1). Then, in Section 5.2.2, we develop methods to solve this new class of problems.

5.2.1 Coherence and Stochastic Pareto Optimality

We first analyze the properties of $\mathrm{WCVaR}_{C,\alpha}$ as a risk measure and then show that an optimal solution of W-CVaR is Pareto optimal according to a certain stochastic dominance relation.

Desirable properties of risk measures have been axiomatized starting with the work of [3], in which the concept of coherent risk measures for scalar-valued random variables is introduced. There are several approaches to define the concept of coherency for the vector-valued random variables [see, e.g., 49, 20, 85, 41]. For example, [41] introduce set-valued conditional value-at-risk for multivariate random variables; using such set-valued functionals as risk measures is appropriate for financial market models with transaction costs [see, e.g., 49]. Our approach is more aligned with the studies which consider multivariate risk measures that map a random vector to a scalar value; in particular, we consider the following definition of coherence in the multivariate case [33].

We say that a functional $\rho : L^\infty(\Omega, 2^\Omega, \mathbb{P}; \mathbb{R}^d) \to \mathbb{R}$ mapping a d-dimensional random vector to a real number is *coherent in dimension* d (in other words, ρ is a *coherent acceptability functional* in dimension d, equivalently, that $-\rho$ is a *coherent risk measure* in dimension d), if ρ has the following properties (for all d-dimensional random vectors $\mathbf{V}, \mathbf{V}_1, \mathbf{V}_2$):

1. *Normalized*: $\rho(\nvdash) = 0$.

2. *Monotone*: $\mathbf{V}_1 \leq \mathbf{V}_2 \;\Rightarrow\; \rho(\mathbf{V}_1) \leq \rho(\mathbf{V}_2)$.

3. *Positive homogeneous*: $\rho(\lambda \mathbf{V}) = \lambda \rho(\mathbf{V})$ for all $\lambda > 0$.

4. *Superadditive*: $\rho(\mathbf{V}_1 + \mathbf{V}_2) \geq \rho(\mathbf{V}_1) + \rho(\mathbf{V}_2)$.

5. *Translation invariant (equivariant)*: $\rho(\mathbf{V} + \lambda \mathbf{e}) = \rho(\mathbf{V}) + \lambda$ for all $\lambda \in \mathbb{R}$.

The constant vector $\mathbf{e}$ denotes the vector of ones $(1, 1, \ldots, 1)$. It is easy to see that for $d = 1$ the definition coincides with the notion of coherence for scalar-valued random variables [3]; we remind the reader that we provide the definition for acceptability functionals, along the lines of [78]. In monotonicity property we consider the usual component-wise ordering; i.e., $\mathbf{V}_1 \leq \mathbf{V}_2$ if $\mathbf{V}_1(j) \leq \mathbf{V}_2(j)$ for all $j \in [d]$.

The next result indicates that the proposed risk measure is of particular importance since it satisfies the desirable properties axiomatized in the above definition of coherence.

Proposition 26. *Consider a one-dimensional mapping ρ and a scalarization set $C \subseteq C_f$, and let $\rho_C(\mathbf{X}) = \min_{\mathbf{c} \in C} \rho(\mathbf{c}^\top \mathbf{X})$ for a d-dimensional random vector X. If ρ is a coherent acceptability functional (-ρ is a coherent risk measure), then $\rho_C(\mathbf{X})$ denoting the worst-case functional in dimension d (with respect to C) is also coherent.*

Proof. It is easy to verify that ρ_C is normalized, monotone, and positive homogeneous. To show that ρ_C is superadditive, let us consider two d-dimensional random vectors $\mathbf{V}_1$ and $\mathbf{V}_2$. Then, by the supperadditivity of ρ and the minimum operator, we have $\rho_C(\mathbf{V}_1 + \mathbf{V}_2) = \min_{\mathbf{c}\in C} \rho(\mathbf{c}^\top(\mathbf{V}_1 + \mathbf{V}_2)) \geq \min_{\mathbf{c}\in C}(\rho(\mathbf{c}^\top\mathbf{V}_1) + \rho(\mathbf{c}^\top\mathbf{V}_2)) \geq \min_{\mathbf{c}\in C} \rho(\mathbf{c}^\top\mathbf{V}_1) + \min_{\mathbf{c}\in C} \rho(\mathbf{c}^\top\mathbf{V}_2) = \rho_C(\mathbf{V}_1) + \rho_C(\mathbf{V}_2)$. The translation invariance of ρ_C follows from the assumptions that $\sum_{j\in[d]} c_j = 1$ and ρ is translation invariant: for any constant λ, $\rho_C(\mathbf{V} + \lambda\mathbf{e}) = \min_{\mathbf{c}\in C} \rho(\mathbf{c}^\top(\mathbf{V} + \lambda\mathbf{e})) = \min_{\mathbf{c}\in C} \rho(\mathbf{c}^\top\mathbf{V} + \lambda) = \min_{\mathbf{c}\in C} \rho(\mathbf{c}^\top\mathbf{V}) + \lambda = \rho_C(\mathbf{V}) + \lambda$. □

We note that one can also consider a stronger notion of translation invariance in condition 5 of the above definition of coherence; for example, [20] state it as follows: $\rho(\mathbf{V} + \lambda\mathbf{e}_j) = \rho(\mathbf{V}) + \lambda$ for all $j \in [d]$ and $\lambda \in \mathbb{R}$, where $\mathbf{e}_j$ is the standard basis vector (1 in the jth component, 0 elsewhere). [85] claims that $\rho_C(\mathbf{X})$ is coherent when ρ is a coherent acceptability functional, even with the above mentioned stronger translation invariance property. However, this claim is not correct even for the unit simplex ($C = C_f$), as we explain next. Since $\rho(\mathbf{c}^\top\mathbf{V})$ is concave in $\mathbf{c}$, the minimum in the definition of $\rho_C(\mathbf{V})$ is attained at an extreme point of C, i.e., $\rho_C(\mathbf{V}) = \min\{\rho(V_1), \rho(V_2), \ldots, \rho(V_d)\}$ if C is a unit simplex. Suppose that $\rho(V_j),\ j \in [d]$, are not all equal, which implies that there exists an index $j^* \in [d]$ such that $\rho_C(\mathbf{V}) < \rho(V_{j^*})$. Then, for any $\lambda > 0$, by the monotonicity of ρ, we have $\rho(\mathbf{V} + \lambda\mathbf{e}_{j^*}) = \min\{\min_{j\in[d]\setminus\{j^*\}} \rho(V_j), \rho(V_{j^*} + \lambda)\} = \min_{j\in[d]\setminus\{j^*\}} \rho(V_j) = \rho_C(\mathbf{V}) < \rho(V_{j^*}) < \rho(V_{j^*} + \lambda) = \rho(V_{j^*}) + \lambda$. This provides an example where $\rho_C(\mathbf{V} + \lambda\mathbf{e}_j) \neq \rho_C(\mathbf{V}) + \lambda$ for all $j \in [d]$ and $\lambda \in \mathbb{R}$.

We next discuss the *Pareto efficiency/optimality* of the solutions of W-CVaR. For *deterministic* multiobjective optimization problems, the concept of Pareto optimality is well-known and it defines a dominance relation to compare the solutions with respect to the multiple criteria. It is natural to consider the "non-dominated" solutions as potentially good solutions. Here, we recall two widely-used Pareto optimality concepts:

- A point $\mathbf{z}^* \in Z$ is called *Pareto optimal* if there exists no point $\mathbf{z} \in Z$ such that

$$G_j(\mathbf{z}) \geq G_j(\mathbf{z}^*) \text{ for all } j \in [d] \text{ and } G_j(\mathbf{z}) > G_j(\mathbf{z}^*) \text{ for at least one index } j \in [d]. \tag{5.6}$$

- A point $\mathbf{z}^* \in Z$ is called *weakly Pareto optimal* if there exists no point $\mathbf{z} \in Z$ such that

$$G_j(\mathbf{z}) > G_j(\mathbf{z}^*) \text{ for all } j \in [d]. \tag{5.7}$$

In contrast to the deterministic case, in a stochastic context there is no single widely-adopted concept of Pareto optimality. The challenge stems from the stochasticity of the criteria: $G_1(\mathbf{z}), \ldots, G_d(\mathbf{z})$ are in general random variables for any decision vector $\mathbf{z} \in Z$, and one should specify how to compare solutions in terms of these random objective criteria. To this end, in this chapter, we use the stochastic dominance rules and introduce stochastic dominance-based Pareto optimality concepts below for stochastic multiobjective optimization problems. For $k \in \mathbb{N}_0 = \{0, 1, \ldots\}$, let us denote the kth degree stochastic dominance (kSD) relation by $\succeq_{(k)}$; we refer the reader to Appendix A for a brief review of these relations [see, also, 77].

Definition 4 (Stochastic dominance-based Pareto Optimality)**.** *A point* $\mathbf{z}^* \in Z$ *is called* kSD-based Pareto optimal *for some* $k \in \mathbb{N}_0$ *if there exists no point* $\mathbf{z} \in Z$ *such that*

$$G_j(\mathbf{z}) \succeq_{(k)} G_j(\mathbf{z}^*) \textit{ for all } j \in [d] \textit{ and } G_j(\mathbf{z}) \succ_{(k)} G_j(\mathbf{z}^*) \textit{ for at least one index } j \in [d]. \tag{5.8}$$

Definition 5 (Stochastic dominance-based Weak Pareto Optimality)**.** *A point* $\mathbf{z}^* \in Z$ *is called* weakly kSD-based Pareto optimal *for some* $k \in \mathbb{N}_0$ *if there exists no point* $\mathbf{z} \in Z$ *such that*

$$G_j(\mathbf{z}) \succ_{(k)} G_j(\mathbf{z}^*) \text{ for all } j \in [d]. \tag{5.9}$$

These stochastic Pareto optimality concepts are based on comparing the random variables $G_j(\mathbf{z})$ and $G_j(\mathbf{z}^*)$ (in relations (5.6) and (5.7)) using a univariate stochastic dominance rule for each criterion $j \in [d]$. Such a component-wise dominance relation provides a natural and an intuitive approach for extending the concept of traditional Pareto optimality to the stochastic case. A closely related but slightly different notion of efficiency based on the realizations under each scenario is presented in [7]. Alternatively, one can consider a multivariate stochastic dominance relation as in [8]. However, multivariate stochastic dominance relations are very restrictive [see, e.g., 72] and finding a non-dominated solution according to such a multivariate relation may not even be possible. For other generalizations of the Pareto efficiency concept to multiobjective stochastic problems we refer to [7].

We next focus on the zeroth-order stochastic dominance (ZSD) rule (also known as statewise dominance) defined in Appendix A, and present a close analogue of Theorem 2.2 in [44], which provides some managerial insights about our new W-CVaR model.

Proposition 27. *Let* $C \subseteq C_f$ *and* $\mathbf{z}^*$ *be an optimal solution of W-CVaR.*

1. $\mathbf{z}^*$ *is a weakly ZSD-based Pareto optimal solution of W-CVaR.*

2. *If for every* $\mathbf{c} \in C$ *we have* $c_j > 0$ *for all* $j \in [d]$*, then* $\mathbf{z}^*$ *is an ZSD-based Pareto optimal solution of W-CVaR.*

3. *If* $\mathbf{z}^*$ *is a unique optimal solution of W-CVaR, then it is an ZSD-based Pareto optimal solution of W-CVaR.*

Proof. Let us assume for contradiction that $\mathbf{z}^*$ is not a weakly ZSD-based Pareto optimal solution of W-CVaR. Then there exists $\hat{\mathbf{z}} \in Z$ such that $G_j(\hat{\mathbf{z}}, \omega_i) > G_j(\mathbf{z}^*, \omega_i)$ for all $i \in [n]$ and $j \in [d]$. By the non-negativity of $\mathbf{c} \in C$ and the observation that $c_k > 0$ for at least one index $k \in [d]$ for every $\mathbf{c} \in C$, we have $\sum_{j\in[d]} c_j G_j(\hat{\mathbf{z}}, \omega_i) > \sum_{j\in[d]} c_j G_j(\mathbf{z}^*, \omega_i)$ for all $i \in [n]$ and $\mathbf{c} \in C$. Then, by the monotonicity of CVaR it is easy to see that $\mathrm{CVaR}_\alpha\left(\mathbf{c}^\top \mathbf{G}(\hat{\mathbf{z}})\right) > \mathrm{CVaR}_\alpha\left(\mathbf{c}^\top \mathbf{G}(\mathbf{z}^*)\right)$ holds for any $\alpha \in (0, 1]$ and $\mathbf{c} \in C$. Therefore, the following inequalities hold and result in a contradiction:

$$\begin{aligned}\max_{\mathbf{z}\in Z} \min_{\mathbf{c}\in C} \; \mathrm{CVaR}_\alpha\left(\mathbf{c}^\top \mathbf{G}(\mathbf{z})\right) &\geq \min_{\mathbf{c}\in C} \; \mathrm{CVaR}_\alpha\left(\mathbf{c}^\top \mathbf{G}(\hat{\mathbf{z}})\right) \\ &> \min_{\mathbf{c}\in C} \; \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z}^*)) = \max_{\mathbf{z}\in Z} \min_{\mathbf{c}\in C} \; \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z})).\end{aligned}$$

This completes the proof of part 1. The proofs of parts 2 and 3 follow from similar arguments. □

We would like to emphasize that the W-CVaR model keeps the stochastic nature of the weighted-sum, and is novel in terms of incorporating the risk associated with the inherent randomness. Therefore, it calls for the development of stochastic Pareto efficiency concepts discussed above. In contrast, in some of the existing stochastic multiobjective optimization models, summary statistics such as expected value, CVaR or variance are used as the multiple criteria [see, for example, 50, for a stochastic portfolio optimization problem with three criteria: expected return, CVaR and a liquidity measure]. Using these summary statistics, the resulting problem becomes a deterministic multicriteria optimization problem for which the well-defined deterministic Pareto optimality concepts can be applied. One method of obtaining Pareto optimal solutions is to scalarize these multiple criteria using a single weight vector in the scalarization set C. By heuristically searching over C, multiple solutions in the deterministic efficient frontier are generated, and then an interactive method is employed for the decision makers to choose among these solutions. To illustrate this approach, consider a modification of the portfolio optimization problem in [50], where $G_1(\mathbf{z})$ is the uncertain return of the portfolio and $G_2(\mathbf{z})$ is a random liquidity measure. Suppose that two criteria are considered: $\mathrm{CVaR}_\alpha(G_1(\mathbf{z}))$ and $\mathrm{CVaR}_\alpha(G_2(\mathbf{z}))$. Thus, for a fixed $\mathbf{c} \in C$, the problem solved is $\max_{\mathbf{z}\in Z}\{c_1 \, \mathrm{CVaR}_\alpha(G_1(\mathbf{z})) + c_2 \, \mathrm{CVaR}_\alpha(G_2(\mathbf{z}))\}$. In contrast, in our model, we search over $\mathbf{c} \in C$, such that the worst-case multivariate CVaR is maximized: $\max_{\mathbf{z}\in Z} \min_{\mathbf{c}\in C}\{\mathrm{CVaR}_\alpha(c_1 G_1(\mathbf{z}) + c_2 G_2(\mathbf{z}))\}$, eliminating the need for an interactive approach that may be prone to conflict among decision makers. Note also that the term in the minimization is different from the objective of the interactive approach, because the order of CVaR and scalarization operations cannot be changed. Only for the special case that the decision makers are risk-neutral (i.e., $\alpha = 1$), the order of CVaR (expectation) and scalarization operations can be changed. In addition, the interactive approach only considers a single weight vector at a time.

Although it is not directly related, we would like to mention that the concept of Pareto optimality in a different decision-making under uncertainty setting has been studied by [47]. The authors focus on the classical robust optimization (RO) framework and aim to find robustly optimal solutions (so-called Pareto robustly optimal solutions), which perform as well as possible across all uncertainty scenarios. To highlight the differences from our setup, recall that in RO it is assumed that the uncertain parameters belong to particular uncertainty sets, and decisions are made to optimize the worst-case performance among all possible uncertainty realizations without considering the underlying probability distributions. Moreover, [47] do not consider a multicriteria optimization problem, they borrow and adapt the corresponding concept of Pareto optimality by considering RO as a multiobjective optimization problem with an infinite number of objectives, one for each uncertainty scenario. For robust optimization in general the interested reader may refer to [10, 9] and [15].

5.2.2 Solution Methods

In this section, we give reformulations and solution methods for W-CVaR. We also provide improved formulations for the important special case when each scenario has an equal probability. Before proceeding to describe the solution methods we first show that W-CVaR is a convex program under certain conditions.

Proposition 28. *If Z is a convex set and $G_j(\mathbf{z})$ is concave in $\mathbf{z} \in Z$ for all $j \in [d]$, then W-CVaR is a convex program.*

Proof. It is sufficient to prove that the mapping $\mathbf{z} \mapsto \min_{\mathbf{c} \in C} \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z}))$ is concave. Recall that by our assumptions c_j is non-negative and $G_j(\mathbf{z})$ is concave in $\mathbf{z} \in Z$ for all $j \in [d]$ and $\mathbf{c} \in C$. Since any non-negative combination of concave functions is also concave, the mapping $\mathbf{z} \mapsto \mathbf{c}^\top \mathbf{G}(\mathbf{z})$ is concave for any $\mathbf{c} \in C$. Then, by the monotonicity and concavity of CVaR, the mapping $\mathbf{z} \mapsto \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z}))$ is concave, and the claim follows from the superadditivity of the minimum operator.

□

We first observe that the inner optimization problem in (5.5) is a concave minimization over a convex set, which implies that an optimal solution of the inner problem occurs at an extreme point of C. Let $\hat{\mathbf{c}}^1, \ldots, \hat{\mathbf{c}}^N$ be the extreme points of C.

Then, using the definition of CVaR given in (5.2), we can formulate (5.5) as follows:

$$\max \psi \tag{5.10a}$$

$$\text{s.t. } \psi \leq \eta_\ell - \frac{1}{\alpha}\sum_{i\in[n]} p_i w_{\ell i}, \qquad \forall\, \ell \in [N], i \in [n] \tag{5.10b}$$

$$w_{\ell i} \geq \eta_\ell - (\hat{\mathbf{c}}^\ell)^\top \mathbf{g}_i(\mathbf{z}), \qquad \forall\, \ell \in [N], i \in [n] \tag{5.10c}$$

$$\mathbf{z} \in Z, \quad \mathbf{w} \in \mathbb{R}_+^{N\times n}, \quad \eta \in \mathbb{R}^N, \quad \psi \in \mathbb{R}. \tag{5.10d}$$

Note that if the mapping $\mathbf{g}_i(\mathbf{z})$ is linear in $\mathbf{z}$ for all $i \in [n]$, Z is a polyhedral set, and $\mathbf{z}$ is a continuous decision vector, then the formulation (5.10) is a linear program. Under certain assumptions on the scalarization set, the number of extreme points of C may be polynomial (we will discuss these cases in Section 5.2.2), and hence the resulting formulation (5.10) is compact. However, in general, the number of extreme points, N, is exponential. Therefore, we propose a delayed cut generation algorithm to solve (5.10). We start with an initial subset of scalarization vectors $\hat{\mathbf{c}}^1, \cdots, \hat{\mathbf{c}}^L$ and solve an intermediate relaxed master problem (RMP), which is obtained by replacing N with L in (5.10). Solving the RMP provides us with a candidate solution denoted by $(\mathbf{z}^*, \psi^*, \mathbf{w}^*, \eta^*)$. At each iteration, we solve a cut generation problem:

$$(\mathbf{CutGen-Robust}) : \qquad \min_{\mathbf{c}\in C} \text{CVaR}(\mathbf{c}^\top \mathbf{G}(\mathbf{z}^*)).$$

If the optimal objective function value of the cut generation problem is not smaller than ψ^*, then the current solution $(\mathbf{z}^*, \psi^*, \mathbf{w}^*, \eta^*)$ is optimal. Otherwise, the optimal solution $\mathbf{c}^t$ at iteration t gives a violated inequality of the form $\psi \leq \text{CVaR}_\alpha((\mathbf{c}^t)^\top \mathbf{G}(\mathbf{z}))$. We augment the RMP by setting $L \leftarrow L+1$, and $\hat{\mathbf{c}}^{L+1} \leftarrow \mathbf{c}^t$.

Observe that in the multivariate CVaR-constrained problems studied in [75] and [55], given a random benchmark vector $\mathbf{Y}$, the cut generation problems are given by $\min_{\mathbf{c}\in C} \mathrm{CVaR}(\mathbf{c}^\top \mathbf{G}(\mathbf{z}^*)) - \mathrm{CVaR}(\mathbf{c}^\top \mathbf{Y})$ (we will revisit this cut generation problem in Section 5.3). Due to the similar structure, we can use the formulations and enhancements given in [75] and [55] to formulate the cut generation problem $(\mathbf{CutGen}-\mathbf{Robust})$ as a mixed-integer program. Let $\mathbf{X} = \mathbf{G}(\mathbf{z}^*)$ with the realizations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ (i.e., $\mathbf{x}_i = \mathbf{g}_i(\mathbf{z}), i \in [n]$). The representation of CVaR in (5.3) leads to the following formulation of $(\mathbf{CutGen}-\mathbf{Robust})$:

$$
\begin{aligned}
\min \quad & \mu && && (5.11a)\\
\text{s.t.} \quad & \mu \geq \mathbf{c}^\top \mathbf{x}_k - \frac{1}{\alpha} \sum_{i\in[n]} p_i [\mathbf{c}^\top \mathbf{x}_k - \mathbf{c}^\top \mathbf{x}_i]_+, && \forall\, k \in [n], && (5.11b)\\
& \mathbf{c} \in C, \quad \mu \in \mathbb{R}. && && (5.11c)
\end{aligned}
$$

The shortfall terms $[\mathbf{c}^\top \mathbf{x}_k - \mathbf{c}^\top \mathbf{x}_i]_+$ in inequalities (5.11b) present a computational challenge. Introducing additional variables and constraints, we can linearize these terms using big-M type of constraints, and obtain an equivalent MIP formulation similar to the one proposed by [75] for the cut generation problems arising in optimization under multivariate polyhedral CVaR constraints. However, the big-M type constraints may lead to weak LP relaxation bounds and computational difficulties. In order to deal with these difficulties, [55] propose an improved model based on a new representation of VaR_α, which we describe next. Let

$$
M_{ik} = \max\{\max_{\mathbf{c}\in C} \{\mathbf{c}^\top \mathbf{x}_k - \mathbf{c}^\top \mathbf{x}_i\}, 0\} \text{ and } M_{ki} = \max\{\max_{\mathbf{c}\in C} \{\mathbf{c}^\top \mathbf{x}_i - \mathbf{c}^\top \mathbf{x}_k\}, 0\}.
$$

Also let $M_{i*} = \max_{k\in[n]} M_{ik}$ and $M_{*i} = \max_{k\in[n]} M_{ki}$ for $i \in [n]$, and $\tilde{M}_j = \max\{c_j : \mathbf{c} \in C\}$ for $j \in [d]$. Then, the following inequalities hold for any $\mathbf{c} \in C$:

$$z \le \mathbf{c}^\top \mathbf{x}_i + \beta_i M_{i*}, \qquad \forall\ i \in [n] \tag{5.12a}$$

$$z \ge \mathbf{c}^\top \mathbf{x}_i - (1-\beta_i) M_{*i}, \qquad \forall\ i \in [n] \tag{5.12b}$$

$$z = \sum_{i\in[n]} \xi_i^\top \mathbf{x}_i, \tag{5.12c}$$

$$\xi_{ij} \le \tilde{M}_j u_i, \qquad \forall\ i \in [n], j \in [d] \tag{5.12d}$$

$$\sum_{i\in[n]} \xi_{ij} = c_j, \qquad \forall\ j \in [d] \tag{5.12e}$$

$$\sum_{i\in[n]} p_i \beta_i \ge \alpha, \tag{5.12f}$$

$$\sum_{i\in[n]} p_i \beta_i - \sum_{i\in[n]} p_i u_i \le \alpha - \epsilon, \tag{5.12g}$$

$$\sum_{i\in[n]} u_i = 1, \tag{5.12h}$$

$$u_i \le \beta_i, \qquad \forall\ i \in [n] \tag{5.12i}$$

$$\beta,\ \mathbf{u} \in \{0,1\}^n, \quad \xi \in \mathbb{R}_+^{n\times d}, \quad z \in R, \tag{5.12j}$$

if and only if $z = \mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$. Here ϵ is a sufficiently small positive constant to ensure that the constraint (5.12g) is equivalent to the strict inequality $\sum_{i\in[n]} p_i\beta_i - \sum_{i\in[n]} p_i u_i < \alpha$. Denoting the finite set of all non-zero probabilities of events by $\mathcal{K} = \left\{\mathbb{P}(S)\ :\ S \in 2^\Omega,\ \mathbb{P}(S) > 0\right\}$ it is easy to see that ϵ can be taken as any number that satisfies $0 < \epsilon < \min\{\alpha - \kappa\ :\ \kappa \in \mathcal{K} \cup \{0\},\ \kappa < \alpha\}$. For example, for the equiprobable case, we let $0 < \epsilon < \frac{1}{n}$. The logical variable $u_i = 1$ only if the i-th scenario corresponds to $\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$, and the additional variables $\xi_{il} = c_l$ only when $u_i = 1$, for all $i \in [n]$ and $l \in [d]$.

Based on the representation of $\mathrm{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ given in (5.12), we propose an alternative formulation for $(\mathbf{CutGen} - \mathbf{Robust})$:

$$\min \quad z - \frac{1}{\alpha}\sum_{i\in[n]} p_i v_i \tag{5.13a}$$

$$\text{s.t. } (5.12\text{a}) - (5.12\text{i}), \tag{5.13b}$$

$$v_i - \delta_i = z - \mathbf{c}^\top \mathbf{x}_i, \qquad \forall\, i \in [n] \tag{5.13c}$$

$$v_i \le M_{i*}\beta_i, \qquad \forall\, i \in [n] \tag{5.13d}$$

$$\delta_i \le M_{*i}(1-\beta_i), \qquad \forall\, i \in [n] \tag{5.13e}$$

$$\beta,\ \mathbf{u} \in \{0,1\}^n, \quad \xi \in \mathbb{R}_+^{n\times d}, \quad z \in R, \tag{5.13f}$$

$$\mathbf{c} \in C, \quad \mathbf{v},\ \delta \in \mathbb{R}_+^n. \tag{5.13g}$$

In this formulation, it is guaranteed that $v_i = [z - \mathbf{c}^\top \mathbf{x}_i]_+$ and $\delta_i = [\mathbf{c}^\top \mathbf{x}_i - z]_+$ for $i \in [n]$.

Equal Probability Case

To keep our exposition simple, we consider confidence levels of the form $\alpha = k/n$ for some $k \in [n]$, and refer to [75] for an extended MIP formulation with an arbitrary confidence level. In this case, an alternative formulation of $(\mathbf{CutGen} - \mathbf{Robust})$, adapted from [75], is given by the bilinear program

$$\begin{aligned}
\min \quad & \frac{1}{k}\sum_{i\in[n]}\sum_{j\in[d]} x_{ij} c_j \beta_i \\
\text{s.t.} \ & \sum_{i\in[n]} \beta_i = k, \\
& \beta \in [0,1]^n, \quad \mathbf{c} \in C.
\end{aligned}$$

Note that we can relax the integrality of β in this formulation, which follows from the observation that in the special case of equal probabilities and $\alpha = k/n$, $\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X})$

corresponds to the weighted sum of the smallest k out of n realizations ($\mathbf{c}^\top \mathbf{x}_i,\ i \in [n]$). Using McCormick envelopes [69], we can linearize the bilinear terms $c_j \beta_i$ in the objective function. Introducing the additional variables $\gamma_{ij} = c_j \beta_i$, for all $i \in [n]$ and $j \in [d]$, an equivalent MIP formulation is stated as:

$$\min \quad \frac{1}{k} \sum_{i \in [n]} \sum_{j \in [d]} x_{ij} \gamma_{ij} \tag{5.14a}$$

$$\text{s.t. } \gamma_{ij} \leq c_j, \qquad \forall\ i \in [n],\ j \in [d] \tag{5.14b}$$

$$\gamma_{ij} \geq c_j - \tilde{M}_j (1 - \beta_i), \qquad \forall\ i \in [n],\ j \in [d] \tag{5.14c}$$

$$\gamma_{ij} \leq \tilde{M}_j \beta_i, \qquad \forall\ i \in [n],\ j \in [d] \tag{5.14d}$$

$$\sum_{i \in [n]} \beta_i = k, \tag{5.14e}$$

$$\beta \in \{0,1\}^n, \quad \gamma \in \mathbb{R}_+^{n \times d}, \quad \mathbf{c} \in C. \tag{5.14f}$$

For $i \in [n]$, if $\beta_i = 1$, then constraint (5.14b) together with (5.14c) enforces that $\gamma_{ij} = c_j$, for all $j \in [d]$. For $i \in [n]$, if $\beta_i = 0$, then constraint (5.14d) enforces γ_{ij} to be 0.

Let $P := \{(\gamma, \beta, \mathbf{c}) \in \mathbb{R}_+^{n \times d} \times \{0,1\}^n \times C \mid \gamma = \beta \mathbf{c}^\top, \sum_{i \in [n]} \beta_i = k\}$. Then we have $\min_{\mathbf{c} \in C}\ \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) = \min_{(\gamma, \beta, \mathbf{c}) \in P} \sum_{i \in [n]} \sum_{j \in [d]} x_{ij} \gamma_{ij}$. Note that the structure of P also appears in pooling problems [c.f., 39]. The next proposition gives the convex hull of P for a special choice of C using the reformulation-linearization technique (RLT) [92].

Proposition 29. ([93, 39]) *If C is a unit simplex (i.e., $C = C_f$), then the convex hull of P is described by:*

$$\begin{aligned} conv(P) = \{ &(\gamma, \beta, \mathbf{c}) \in \mathbb{R}_+^{n \times d} \times [0,1]^n \times C \mid \gamma_{ij} \leq c_j, i \in [n],\ j \in [d], \gamma_{ij} = \beta_i,\ i \in [n], \\ &\gamma_{ij} = kc_j,\ j \in [d]\}. \end{aligned}$$

Using the fact that $C \subseteq C_f$ and Proposition 29, we can strengthen the formulation (5.14) as follows:

$$\min \quad \frac{1}{k} \sum_{i \in [n]} \sum_{j \in [d]} x_{ij} \gamma_{ij} \tag{5.15a}$$

$$\text{s.t. } \gamma_{ij} \leq c_j, \qquad \forall\, i \in [n],\ j \in [d] \tag{5.15b}$$

$$\sum_{j \in [d]} \gamma_{ij} = \beta_i, \qquad \forall\, i \in [n] \tag{5.15c}$$

$$\sum_{i \in [n]} \gamma_{ij} = kc_j, \qquad \forall\, j \in [d] \tag{5.15d}$$

$$(5.14c) - (5.14d), \tag{5.15e}$$

$$\mathbf{c} \in C, \quad \beta \in \{0,1\}^n, \quad \gamma \in \mathbb{R}_+^{n \times d}. \tag{5.15f}$$

Note also that if C *is* the unit simplex ($C = C_f$), then the integrality restrictions on β can be relaxed in (5.15) and the cut generation problem is an LP. However, recall that if C is the unit simplex, then the extreme points of C are polynomial, given by $\hat{\mathbf{c}}^\ell = \mathbf{e}_\ell$ for $\ell \in [d]$. Hence, in this case, the overall problem formulation (5.10) itself is a compact LP when the mapping $\mathbf{g}_i(\mathbf{z})$ is linear in $\mathbf{z}$ for all $i \in [n]$, and Z is a polyhedral set without integrality restrictions, even under general probabilities.

Furthermore, using the additional information on the structure of the scalarization polytope C and the RLT technique, we can obtain stronger formulations. Suppose that $C = \{\mathbf{c} \in \mathbb{R}_+^d \,|B\mathbf{c} \geq \mathbf{b}\}$, for a given $r \times d$ matrix B and $\mathbf{b} = (b_1, \ldots, b_r)$. Let B_ℓ

be the ℓth row of B. Then, we can strengthen the formulation (5.14) as follows:

$$\begin{aligned}
\min \quad & \frac{1}{k}\sum_{i\in[n]}\sum_{j\in[d]} x_{ij}\gamma_{ij} && && (5.16a)\\
\text{s.t.} \quad & \sum_{j\in[d]} B_{\ell j}\gamma_{ij} - b_\ell\beta_i \le B_\ell\mathbf{c} - b_\ell, && \forall\ i\in[n],\ \ell\in[r] && (5.16b)\\
& \sum_{j\in[d]}\sum_{j\in[d]} B_{\ell j}\gamma_{ij} - b_\ell\beta_i \ge 0, && \forall\ i\in[n],\ \ell\in[r] && (5.16c)\\
& \sum_{i\in[n]}(\sum_{j\in[d]} B_{\ell j}\gamma_{ij} - b_\ell\beta_i) = k(B_\ell\mathbf{c} - b_\ell), && \forall\ \ell\in[r] && (5.16d)\\
& \mathbf{c}\in C, \quad \beta\in\{0,1\}^n, \quad \gamma\in\mathbb{R}_+^{n\times d}. && && (5.16e)
\end{aligned}$$

It is known that if $C = \{\mathbf{c}\in\mathbb{R}_+^d\ |B\mathbf{c}\ge\mathbf{b}\}$ is a d-simplex, then $conv(P) = \{(\gamma,\beta,\mathbf{c})\in\mathbb{R}_+^{n\times d}\times[0,1]^n\times C\ |(5.16b)-(5.16d)\}$ [39]. Therefore, the LP relaxation of (5.16) is integral in this case.

Remark 6. *Note that if $\tilde{M}_j = 1$ for all $j\in[d]$ (as is the case when C is the unit simplex), then constraints* (5.14c)-(5.14d) *are implied by* (5.15c)-(5.15d)*, and can be dropped from the formulation. However, for the situations where $\tilde{M}_j < 1$ for some $j\in[d]$, the constraints* (5.14c)-(5.14d)*, obtained by applying the RLT technique to the constraints $c_j\le\tilde{M}_j, j\in[d]$, can be useful to reduce the solution time.*

Remark 7. *It is also possible to obtain stronger formulations of* (5.12) *by applying the RLT technique for the general probability case. In particular, the RLT procedure based on the constraint $\sum_{i\in[d]} c_i = 1$ provides the following valid inequality*

$$\sum_{j\in[d]}\xi_{ij} = u_i, \tag{5.17}$$

which can be added to the formulation (5.12).

Next we consider an important special case of C that applies to multicriteria optimization when certain criteria are deemed more important than others. In particular, we study the case where C contains ordered preference constraints that take the form

$$C = \{\mathbf{c} \in \mathbb{R}^d_+ \mid \sum_{j\in[d]} c_j = 1, c_j \geq c_{j+1}, \quad \forall\, j \in [d-1]\}. \tag{5.18}$$

If the set C has the ordered preference structure (5.18), then we are able to obtain the convex hull of P, which is stated in the following result.

Proposition 30. *If C is given by* (5.18)*, then the convex hull of P is described by:*

$$\begin{aligned} conv(P) = \{(\gamma, \beta, \mathbf{c}) \in \mathbb{R}^{n\times d}_+ \times [0,1]^n \times C \mid (5.15c), (5.15d), \gamma_{ij} \geq \gamma_{ij+1}, \\ \gamma_{ij} - \gamma_{ij+1} \leq c_j - c_{j+1}, i \in [n],\ j \in [d-1]\}. \end{aligned}$$

Proof. First, we show that the extreme points of C are given by

$$\begin{aligned} \hat{\mathbf{c}}^1 &= (1, 0, 0, \ldots, 0) \\ \hat{\mathbf{c}}^2 &= (\frac{1}{2}, \frac{1}{2}, 0, \ldots, 0) \\ \hat{\mathbf{c}}^3 &= (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \ldots, 0) \\ &\vdots \\ \hat{\mathbf{c}}^d &= (\frac{1}{d}, \frac{1}{d}, \frac{1}{d}, \ldots, \frac{1}{d}). \end{aligned}$$

Let $\tilde{\mathbf{c}} = (\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_d)$ be a feasible point of C, by definition, we have $\tilde{c}_1 \geq \tilde{c}_2 \geq \cdots \geq \tilde{c}_d$. First, we show that $\tilde{c}_j \leq \frac{1}{j}$, for all $j \in [d]$. Suppose that there exists $j \in [d]$ such that $\tilde{c}_j > \frac{1}{j}$, then we have $\sum_{i=1}^{j} \tilde{c}_i \geq j\tilde{c}_j > 1$ since $\tilde{c}_i \geq c_j$, for all $i \in [j-1]$; this results in a contradiction. Hence, for any feasible point, we have $\tilde{c}_j \leq \frac{1}{j}$, for all $j \in [d]$. Next, let $\lambda_j = j(\tilde{c}_j - \tilde{c}_{j+1})$, for all $j \in [d]$, where $\tilde{c}_{d+1} = 0$. Note that $0 \leq \lambda_j \leq 1$, for all $j \in [d]$, and $\sum_{j=1}^{d} \lambda_j = 1$. We have $\tilde{\mathbf{c}} = \sum_{j=1}^{d} \lambda_j \hat{\mathbf{c}}^j$, which indicates that any feasible point $\tilde{\mathbf{c}}$ can be represented as a convex combination of the points $\hat{\mathbf{c}}^j$, for all $j \in [d]$. As a result, C is a $(d-1)$-simplex, and the proposition follows similarly from [39]. $\square$

5.2.3 Finite Convergence

In this section, we study the convergence of the proposed cut generation algorithm.

Proposition 31. *The delayed cut generation algorithm described in Section 5.2.2 to solve W-CVaR is finitely convergent.*

Proof. We show that given a solution to RMP we can find an optimal solution to the cut generation subproblem, which is an extreme point of C. As a result, the proposed cut generation algorithm is finitely convergent, because there are finitely many extreme points of C. For the general probability case, we can obtain such a vertex optimal solution by using the following method: suppose that we solve one of the MIP formulations of $(\mathbf{CutGen} - \mathbf{Robust})$ and obtain an optimal solution $\mathbf{c}^*$. Let π be a permutation describing a non-decreasing ordering of the realizations of the random vector ${\mathbf{c}^*}^\top \mathbf{X}$, i.e., ${\mathbf{c}^*}^\top \mathbf{x}_{\pi(1)} \leq \cdots \leq {\mathbf{c}^*}^\top \mathbf{x}_{\pi(n)}$, and define

$$k^* = \min \left\{ k \in [n] \ : \ \sum_{i \in [k]} p_{\pi(i)} \geq \alpha \right\} \quad \text{and} \quad K^* = \{\pi(1), \ldots, \pi(k^* - 1)\}.$$

Then, we can obtain the desired vertex solution $\hat{\mathbf{c}}$ by finding a vertex optimal solution of the following linear program:

$$\min_{\mathbf{c} \in C} \ \frac{1}{\alpha} \left[\sum_{i \in K^*} p_i \mathbf{c}^\top \mathbf{x}_i + \left(\alpha - \sum_{i \in K^*} p_i \right) \mathbf{c}^\top \mathbf{x}_{\pi(k^*)} \right].$$

This LP relies on the subset-based representation of CVaR [Theorem 1, 75]. The feasible set is the polytope C, so there exists a vertex optimal solution $\hat{\mathbf{c}}$. It is easy to show that $\hat{\mathbf{c}}$ is also an optimal solution of $(\mathbf{CutGen} - \mathbf{Robust})$. □

Furthermore, when equal probability is assumed, by solving the alternative cut generation formulation (5.16) using a branch-and-bound (B&B) method, we are guaranteed to obtain a desired vertex optimal solution $\mathbf{c}$ without solving an additional

LP. To see this, note that once the LP relaxation at a B&B node results in an integral β, the only remaining constraints enforce $\mathbf{c} \in C$.

5.3 Multivariate CVaR-constrained Optimization Model

In this section, we consider a related class of multicriteria decision making problems, where the decision vector $\mathbf{z}$ is selected from a feasible set Z and associated random outcomes are determined by the outcome mapping $\mathbf{G} : Z \times \Omega \to \mathbb{R}^d$. We consider an arbitrary objective function $f : Z \mapsto \mathbb{R}$ and assume that a d-dimensional benchmark random vector $\mathbf{Y}$ is available. We aim to find the best decision vector $\mathbf{z}$ for which the random outcome vector $\mathbf{G}(\mathbf{z})$ is preferable to the benchmark $\mathbf{Y}$ with respect to the multivariate polyhedral CVaR preference relation. Such multivariate CVaR-constrained optimization problems are introduced in [75]. Given a polyhedron of scalarization vectors $C \subseteq C_f$ and a confidence level $\alpha \in (0, 1]$, the problem is of the general form:

$$\max f(\mathbf{z}) \tag{5.19a}$$

$$\text{s.t. } \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z})) \geq \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y}), \qquad \forall\, \mathbf{c} \in C \tag{5.19b}$$

$$\mathbf{z} \in Z. \tag{5.19c}$$

The benchmark random vector can be defined on a different probability space, but in practical applications it often takes the form $\mathbf{Y} = \mathbf{G}(\bar{\mathbf{z}})$, where $\bar{\mathbf{z}}$ is a benchmark decision.

Observe that (5.19b) contains infinitely many inequalities. [75] show that these inequalities can be replaced with those for a finite subset of scalarization vectors corresponding to the vertices of a higher dimensional polyhedron. The authors propose a delayed cut generation algorithm, which involves the solution of a relaxed master problem (RMP-B) to obtain a candidate solution $\hat{\mathbf{z}} \in Z$, and the following cut generation subproblem:

$$(\mathbf{CutGen-Benchmark}): \quad \min_{\mathbf{c}\in C} \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) - \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y}), \tag{5.20}$$

where $\mathbf{X} = \mathbf{G}(\hat{\mathbf{z}})$. If the optimal objective function value of $(\mathbf{CutGen-Benchmark})$ is non-negative, then $\hat{\mathbf{z}}$ is optimal, otherwise we obtain a solution $\mathbf{c}^* \in C$ such that the corresponding CVaR inequality in (5.19b) is violated. We augment RMP-B by adding this violated CVaR constraint and resolve it. According to [75], the main bottleneck of this delayed cut generation algorithm is solving the cut-generation problem (5.20), since it is generally nonconvex. Therefore, the main focus of this section is the cut generation problem. Throughout the rest of this chapter, we assume that $\mathbf{Y}$ is a random vector with (not necessarily distinct) realizations $\mathbf{y}_1, \ldots, \mathbf{y}_m$ and corresponding probabilities $q_1, \ldots, q_m$. As before, we let $\mathbf{g}_i(\hat{\mathbf{z}}) = \mathbf{x}_i = (x_{i1}, \ldots, x_{id})$ for all $i \in [n]$.

To solve (5.20), we first need to represent $\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ and $\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y})$ appropriately. Using the LP representation (5.2) for $\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y})$, we can reformulate $(\mathbf{CutGen-Benchmark})$ as

$$\begin{aligned}
\min \quad & \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) - \eta + \frac{1}{\alpha}\sum_{l\in[m]} q_l w_l \\
\text{s.t. } & w_l \geq \eta - \mathbf{c}^\top \mathbf{y}_l, && \forall\, l \in [m] && (5.21a)\\
& \mathbf{w} \in \mathbb{R}^m_+, \quad \eta \in \mathbb{R}, \quad \mathbf{c} \in C. && && (5.21b)
\end{aligned}$$

The minimization of the concave term $\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ causes computational difficulties. For this cut generation problem, [55] introduce a MIP formulation based on the VaR representation of CVaR (see (5.12)), which is given by

$$\begin{aligned}
\min \quad & z - \frac{1}{\alpha}\sum_{i\in[n]} p_i v_i - \eta + \frac{1}{\alpha}\sum_{l\in[m]} q_l w_l && (5.22a)\\
\text{s.t.}\ & (5.13b) - (5.13e),\ (5.21a), && (5.22b)\\
& \beta,\ \mathbf{u} \in \{0,1\}^n, \quad \xi \in \mathbb{R}_+^{n\times d}, \quad \mathbf{z}, \eta \in \mathbb{R}, && (5.22c)\\
& \mathbf{c} \in C, \quad \mathbf{v}, \delta \in \mathbb{R}_+^n, \quad \mathbf{w} \in \mathbb{R}_+^m. && (5.22d)
\end{aligned}$$

The authors demonstrate that this formulation, which we refer to as (**MIP − CVaR**), along with computational enhancements, outperforms existing models for (**CutGen − Benchmark**) under general probabilities. In this section, we consider the special case of equal probabilities, and propose strengthened MIP formulations for the cut generation problems using the RLT technique.

5.3.1 Equal Probability Case

As in Section 5.2.2, to keep our exposition simple, we consider confidence levels of the form $\alpha = k/n$ and assume that all the outcomes of $\mathbf{X}$ are equally likely. For this special case, similar to the development in Section 5.2.2, [75] give the equivalent formulation below:

$$\begin{aligned}
\min \quad & \frac{1}{k}\sum_{i\in[n]} \gamma_i^\top \mathbf{x}_i - \eta + \frac{1}{\alpha}\sum_{l\in[m]} q_l w_l && (5.23a)\\
\text{s.t.} \quad & (5.14b) - (5.14e),\ (5.21a), && (5.23b)\\
& \beta \in \{0,1\}^n, \quad \gamma \in \mathbb{R}_+^{n\times d}, \quad \mathbf{c} \in C, \quad \mathbf{w} \in \mathbb{R}_+^m, \quad \eta \in \mathbb{R}. && (5.23c)
\end{aligned}$$

As before, $\tilde{M}_j = \max\{c_j : \mathbf{c} \in C\}$. Suppose that the vertices of the polytope C is known and given as $\{\hat{\mathbf{c}}_1, \ldots, \hat{\mathbf{c}}_N\}$. Then, we can simply set $\tilde{M}_j = \max_{\ell\in[N]} \hat{c}_{\ell j}$. Furthermore, we can use the RLT-based strengthening for (5.14b)-(5.14e) and obtain the following MIP formulation:

$$
\begin{aligned}
(\textbf{MIP}-\textbf{Special}):\qquad \min\quad & \frac{1}{k}\sum_{i\in[n]}\gamma_i^\top \mathbf{x}_i-\eta+\frac{1}{\alpha}\sum_{l\in[m]}q_l w_l && (5.24a)\\
\text{s.t.}\quad & (5.15b)-(5.15e),\ (5.21a), && (5.24b)\\
& \beta\in\{0,1\}^n,\quad \gamma\in\mathbb{R}_+^{n\times d},\quad \mathbf{c}\in C, && (5.24c)\\
& \mathbf{w}\in\mathbb{R}_+^m,\quad \eta\in\mathbb{R}. && (5.24d)
\end{aligned}
$$

In addition, we can use the RLT technique to further strengthen this formulation using any additional constraints in C as in (5.16); we will provide some numerical results on the performance of such strengthened versions in the computational study (Section 5.5.2).

From Proposition 29, we can obtain the minimum of $\mathrm{CVaR}_\alpha(\mathbf{c}^\top\mathbf{X})$ by solving a linear program when C is a d-simplex. However, even for the special case of unit simplex, constraints (5.15b)–(5.15d) are not sufficient to describe the convex hull of the set of feasible solutions to (5.24), due to the additional constraints (5.21a)–(5.21b) representing $\mathrm{CVaR}_\alpha(\mathbf{c}^\top\mathbf{Y})$. To show this and develop potentially stronger MIP formulations, we derive a class of valid inequalities that describes facets of the convex hull of feasible solutions to (5.24d).

Let

$$
\begin{aligned}
\mathcal{S}:=\{&(\gamma,\mathbf{c},\beta,\eta,\mathbf{w})\in\mathbb{R}_+^{n\times d}\times\mathbb{R}_+^d\times\{0,1\}^n\times\mathbb{R}\times\mathbb{R}_+^m\mid\gamma=\beta\mathbf{c}^\top,\ \sum_{j\in[d]}c_j=1,\\
&\sum_{i\in[n]}\beta_i=k,\mathbf{c}^\top\mathbf{y}_l\geq\eta-w_l,\ \forall\ l\in[m]\}.
\end{aligned}
$$

Proposition 32. *For any $i\in[n]$, $s\in[m]$, and $t\in[m]\setminus\{s\}$, the inequality*

$$
\mathbf{c}^\top\mathbf{y}_s-\sum_{j\in[d]}(y_{sj}-y_{tj})\gamma_{ij}\geq\eta-w_s-w_t, \tag{5.25}
$$

is valid for $\mathcal{S}$. In addition, inequality (5.25) *defines a facet of conv*($\mathcal{S}$) *if and only if* $s \in [m], t \in [m] \setminus \{s\}$ *are such that* $y_{sj} < y_{tj}$ *and* $y_{si} > y_{ti}$ *for some* $i, j \in [d]$.

Proof. Suppose that $\beta_i = 0$, then $\gamma_{ij} = 0$ for all $j \in [d]$. Hence, inequality (5.25) reduces to

$$\mathbf{c}^\top \mathbf{y}_s \geq \eta - w_s - w_t,$$

which is valid since $w_t \geq 0$. Otherwise, suppose that $\beta_i = 1$, then $\gamma_{ij} = c_j$ for all $j \in [d]$, and inequality (5.25) reduces to

$$\mathbf{c}^\top \mathbf{y}_t \geq \eta - w_t - w_s,$$

which is valid, because $w_s \geq 0$, for all $s \in [m]$. We provide the facet proof in Appendix B (see Proposition 34). □

Note that applying the RLT procedure directly to the additional constraints

$$\mathbf{c}^\top \mathbf{y}_l \geq \eta - w_l, \quad \forall\, l \in [m], \tag{5.26}$$

in the set $\mathcal{S}$, would lead to additional bilinear terms $\eta\beta_i$ and $w_l\beta_i$ that will need to be linearized by introducing additional variables and big-M constraints. The proposed inequalities (5.25) can also be obtained by an indirect application of the RLT procedure as follows. Given $i \in [n]$, $s \in [m]$, and $t \in [m] \setminus \{s\}$, multiply constraint (5.26) for $l = s$ with $(1 - \beta_i)$, constraint (5.26) for $l = t$ with β_i, constraint $0 \geq -w_s$ with β_i and constraint $0 \geq -w_t$ with $(1 - \beta_i)$, and sum the resulting inequalities up to obtain inequality (5.25) (the undesirable nonlinear terms cancel out with this selection of multipliers). It is interesting to note that such an application of RLT yields facet-defining inequalities as claimed in Proposition 32.

Alternative VaR-based formulations

In this section, without loss of generality, we assume that all the realizations of $\mathbf{c}^\top \mathbf{X}$ are non-negative (or equivalently, $\mathbf{x}_i$ is non-negative for all $i \in [n]$). Then, it is easy to show that (**CutGen – Benchmark**) can be formulated as follows:

$$
\begin{aligned}
\min \quad & \frac{1}{k}\sum_{i\in[n]} \theta_i - \eta + \frac{1}{\alpha}\sum_{l\in[m]} q_l w_l \\
\text{s.t.} \quad & \theta_i \geq \mathbf{c}^\top \mathbf{x}_i - (1-\beta_i)M_i, && \forall\, i \in [n] && (5.27\text{a}) \\
& \sum_{i\in[n]} \beta_i = k, && && (5.27\text{b}) \\
& (5.21\text{a}), && && (5.27\text{c}) \\
& \mathbf{c} \in C, \quad \beta \in \{0,1\}^n, \quad \theta \in \mathbb{R}_+^n, \quad \mathbf{w} \in \mathbb{R}_+^m, \quad \eta \in \mathbb{R}. && && (5.27\text{d})
\end{aligned}
$$

In this formulation, M_i is the largest possible value of θ_i (e.g., $M_i = \max_{\mathbf{c}\in C} \mathbf{c}^\top \mathbf{x}_i$). This new formulation again follows from the observation that in the special case of equal probabilities and $\alpha = k/n$, $\text{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ corresponds to the weighted sum of the smallest k realizations of $\mathbf{c}^\top \mathbf{X}$. In this special case, $\text{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$ corresponds to the kth smallest realization, and the model guarantees that $\theta_i = \mathbf{c}^\top \mathbf{x}_i$ if $\mathbf{c}^\top \mathbf{x}_i \leq \text{VaR}_\alpha(\mathbf{c}^\top \mathbf{X})$, and $\theta_i = 0$ otherwise. However, this MIP formulation is weak due to the big-M constraints (5.27a). Hence, we can take advantage of the new representation of VaR given in (5.12) to develop a stronger MIP formulation:

$$
\begin{aligned}
(\textbf{MIP\_VaR\_Special}): \quad \min \quad & \frac{1}{k}\sum_{i\in[n]}\theta_i - \eta + \frac{1}{\alpha}\sum_{l\in[m]} q_l w_l && (5.28a)\\
\text{s.t.} \quad & z \le \mathbf{c}^\top \mathbf{x}_i + \beta_i M_{i*}, \forall\ i\in[n] && (5.28b)\\
& \theta_i \ge \mathbf{c}^\top \mathbf{x}_i - (1-\beta_i)M_i, \forall\ i\in[n] && (5.28c)\\
& z \ge \theta_i, \forall\ i\in[n] && (5.28d)\\
& z = \sum_{i\in[n]} \xi_i^\top \mathbf{x}_i, && (5.28e)\\
& \sum_{i\in[n]} \xi_{ij} = c_j, \forall\ j\in[d] && (5.28f)\\
& \sum_{j\in[d]} \xi_{ij} = u_i, \forall\ i\in[n] && (5.28g)\\
& \sum_{i\in[n]} \beta_i = k, && (5.28h)\\
& (5.12d), (5.12h)-(5.12i), (5.21a), && (5.28i)\\
& \mathbf{c}\in C, \quad \beta, \mathbf{u}\in\{0,1\}^n, \quad \mathbf{w}\in\mathbb{R}^m_+, \quad \eta, z\in\mathbb{R}, && (5.28j)\\
& \xi\in\mathbb{R}^{n\times d}_+, \quad \theta\in\mathbb{R}^n_+. && (5.28k)
\end{aligned}
$$

In this formulation, the variable $z = \mathrm{VaR}_\alpha(\mathbf{c}^\top\mathbf{X})$ is represented by $\sum_{i\in[n]} \xi_i^\top \mathbf{x}_i = \sum_{i\in[n]} u_i \mathbf{c}^\top \mathbf{x}_i$, and it is guaranteed that $\xi_{ij} = c_j u_i$ for all $i\in[n]$ and $j\in[d]$. These bilinear terms are linearized by using the McCormick envelopes and their RLT strengthening based on only the information that C is a subset of the unit simplex. Additional constraints on the scalarization set C can be used to further strengthen the above formulation. Notice that different from (5.12), this formulation includes the RLT strengthening equality (5.17) (or (5.28g)).

Finally, we note that (**MIP_VaR_Special**) can also be applied to solve (**CutGen − Robust**) by dropping the variables and constraints associated with

$\mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y})$; leading to enhanced versions of (5.13) for the equal probability case. We test its computational performance in Section 5.5.2.

5.4 Hybrid Model

In this section, we present a hybrid model that includes both the multivariate CVaR constraints and the robust objective based on the worst-case CVaR. We show that the algorithms in Sections 5.2 and 5.3 can be integrated into a unified methodology to solve the hybrid model of the form

$$\begin{aligned} (\mathbf{Hybrid}):\quad & \max_{\mathbf{z}\in Z} \quad \min_{\mathbf{c}\in C} \ \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) \\ & \text{s.t.} \quad \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z})) \geq \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y}), \qquad \forall\, \mathbf{c} \in C. \end{aligned} \tag{5.29}$$

For a given subset of scalarization vectors $\tilde{C} := \{\tilde{\mathbf{c}}^1, \cdots, \tilde{\mathbf{c}}^{\tilde{L}}\} \subset C$ a relaxed master problem (RMP-H) is given by

$$\max_{\mathbf{z}\in Z} \quad \min_{\mathbf{c}\in C} \ \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) \tag{5.30a}$$

$$\text{s.t.} \quad \mathrm{CVaR}_\alpha((\tilde{\mathbf{c}}^\ell)^\top \mathbf{G}(\mathbf{z})) \geq \mathrm{CVaR}_\alpha((\tilde{\mathbf{c}}^\ell)^\top \mathbf{Y}), \qquad \forall\, \ell \in [\tilde{L}]. \tag{5.30b}$$

We can represent the constraints (5.30b) by linear inequalities, leading to the following equivalent reformulation of RMP-H:

$$\begin{aligned} \max \quad & \min_{\mathbf{c}\in C} \ \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{X}) \\ \text{s.t.} \quad & \tilde{\eta}_r - \frac{1}{\alpha}\sum_{i\in[n]} p_i \tilde{w}_{ri} \geq \mathrm{CVaR}_\alpha((\tilde{\mathbf{c}}^r)^\top \mathbf{Y}), && \forall\, r \in [\tilde{L}] \\ & \tilde{w}_{ri} \geq \tilde{\eta}_r - (\tilde{\mathbf{c}}^r)^\top \mathbf{g}_i(\mathbf{z}), && \forall\, r \in [\tilde{L}],\ i \in [n] \\ & \tilde{\mathbf{w}} \in \mathbb{R}_+^{\tilde{L}\times n}, \quad \tilde{\eta} \in \mathbb{R}_+^{\tilde{L}}, \quad \mathbf{z} \in Z. \end{aligned}$$

As discussed in Section 5.2.2, we can handle the maximin type objective function of interest using a finitely convergent delayed cut generation algorithm. In this spirit, suppose now that $\hat{C} = \{\hat{\mathbf{c}}^1, \ldots, \hat{\mathbf{c}}^L\} \subset C$ is a given subset of scalarization vectors used to calculate the worst-case CVaR. In line with the formulation given in (5.10), RMP-H takes the following form:

$$
\begin{aligned}
\max \quad & \psi && && \text{(5.31a)}\\
\text{s.t.} \quad & \tilde{\eta}_r - \frac{1}{\alpha} \sum_{i \in [n]} \tilde{p}_i \tilde{w}_{ri} \geq \mathrm{CVaR}_\alpha((\tilde{\mathbf{c}}^r)^\top \mathbf{Y}), && \forall\, r \in [\tilde{L}] && \text{(5.31b)}\\
& \tilde{w}_{ri} \geq \tilde{\eta}_r - (\tilde{\mathbf{c}}^r)^\top \mathbf{g}_i(\mathbf{z}), && \forall\, r \in [\tilde{L}],\ i \in [n] && \text{(5.31c)}\\
& \psi \leq \eta_\ell - \frac{1}{\alpha} \sum_{i \in [n]} p_i w_{\ell i}, && \forall\, \ell \in [L],\ i \in [n] && \text{(5.31d)}\\
& w_{\ell i} \geq \eta_\ell - (\hat{\mathbf{c}}^\ell)^\top \mathbf{g}_i(\mathbf{z}), && \forall\, \ell \in [L],\ i \in [n] && \text{(5.31e)}\\
& \tilde{\mathbf{w}} \in \mathbb{R}_+^{\tilde{L} \times n}, \quad \mathbf{w} \in \mathbb{R}_+^{L \times n}, \quad \tilde{\eta} \in \mathbb{R}_+^{\tilde{L}}, \quad \psi \in R, \quad \mathbf{z} \in Z. && && \text{(5.31f)}
\end{aligned}
$$

Given a solution to the RMP-H (5.31), two types of cut generation problems are solved to identify if the current solution is optimal or if there is a scalarization vector $\mathbf{c} \in C$ for which at least one of the following constraints is violated: (5.10b) and (5.29). As discussed in Section 5.2.2, for minimizing the worst-case CVaR, it is sufficient to consider the extreme points of C. On the other hand, for the multivariate CVaR relation, it is sufficient to consider the finitely many $\mathbf{c}$ vectors obtained as the projections of the vertices of the higher dimensional polyhedron $P(C, \mathbf{Y})$ given by [75]

$$P(C, \mathbf{Y}) = \left\{(\mathbf{c}, \eta, \mathbf{w}) \in C \times \mathbb{R} \times \mathbb{R}_+^m \ :\ w_l \geq \eta - \mathbf{c}^\top \mathbf{y}_l, \quad l \in [m]\right\}. \tag{5.32}$$

Thus, generating the violated constraints associated with those particular vertex scalarization vectors at each iteration guarantees the finite convergence of the delayed cut generation algorithm of (**Hybrid**). In other words, the provable finite convergence depends on finding a solution to the cut generation problems (**CutGen – Robust**) and (**CutGen – Benchmark**), which is an extreme point of C and the projection of a vertex of $P(C, \mathbf{Y})$, respectively. In Section 5.2.3, we discuss how to obtain a vertex optimal solution of (**CutGen – Robust**) from an optimal solution obtained by solving one of its MIP formulations (such as (5.13)). For obtaining the desired vertex optimal solution of (**CutGen – Benchmark**), we refer to [75].

5.5 Computational study

In the first part of our computational study, we investigate the value of the proposed W-CVaR model with respect a robust risk-neutral model and a multivariate CVaR-constrained model. We also report on the performance of the cut generation algorithm for the W-CVaR model. In the second part, we demonstrate the computational effectiveness of the MIP formulations and the valid inequalities developed (in Section 5.3) for the cut generation problem arising in multivariate CVaR-constrained optimization models.

5.5.1 Worst-case Multivariate CVaR Optimization

We explore the effectiveness of the proposed W-CVaR model by applying it to a homeland security budget allocation (HSBA) problem [43]. This problem studies the allocation of a fixed budget to ten urban areas, which are classified in three groups: *1) higher risk*: New York; *2) medium risk*: Chicago, San Francisco Bay Area, Washington DC-MD-VA-WV, and Los Angeles-Long Beach; *3) lower risk*: Philadelphia PA-NJ, Boston MA-NH, Houston, Newark, and Seattle-Bellevue-Everett. The risk share of each area is measured based on four criteria: *1) property losses, 2) fatalities,* *3) air departures* and *4) average daily bridge traffic.* To represent the inherent randomness a random risk share matrix $A : \Omega \to \mathbb{R}_+^{4\times 10}$ is considered, where A_{ij} denotes the proportion of losses in urban area j relative to the total losses for criterion i. The set $Z = \{\mathbf{z} \in \mathbb{R}_+^{10} : \sum_{j\in[10]} z_j = 1\}$ represents all the feasible allocations and the associated random performance measures of interest are specified based on a particular type of penalty function for allocations under the risk share. The negatives of these budget misallocations associated with each criterion are used to construct the random outcome vector $\mathbf{G}(\mathbf{z}) = (G_1(\mathbf{z}), \ldots, G_4(\mathbf{z}))$, as given below, in order to be consistent with our setup where the larger values of the random variables are preferred:

$$G_i(\mathbf{z}) = -\sum_{j\in[10]} [A_{ij} - z_j]_+, \quad i \in [4].$$

[43] model this HSBA problem using optimization under multivariate polyhedral SSD constraints based on two benchmarks: one based on average government allocations (Department of Homeland Security's Urban Areas Security Initiative) - denoted by $\mathbf{G}(\mathbf{z}^G)$, and one based on the suggestions in the RAND report [104] - denoted by $\mathbf{G}(\mathbf{z}^R)$. On the other hand, [75] replace the SSD constraints with CVaR-based ones, leading to the following optimization model:

$$\begin{aligned}
\max \quad & \min_{\mathbf{c}\in C} \mathbb{E}(\mathbf{c}^\top \mathbf{G}(\mathbf{z})) && (5.33a)\\
\text{s.t.} \quad & \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z})) \geq \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z}^R)), \qquad \forall\, \mathbf{c} \in C && (5.33b)\\
& \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z})) \geq \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{G}(\mathbf{z}^G)), \qquad \forall\, \mathbf{c} \in C && (5.33c)\\
& \mathbf{z} \in Z. && (5.33d)
\end{aligned}$$

We benchmark our W-CVaR model, defined in (5.5), against two relevant existing models: the first one, which we refer to as B-CVaR, is obtained from (5.33) by dropping (5.33c) (the government benchmark is ignored for simplicity), and the second one is the risk-neutral counterpart of our model [44]:

$$\text{W-Exp}: \quad \max_{\mathbf{z}\in Z} \min_{\mathbf{c}\in C} \mathbb{E}(\mathbf{c}^\top \mathbf{G}(\mathbf{z})).$$

We follow the data generation scheme described in [75] and consider their "base case" scalarization set given by $C = C_{\text{Base}} := \{\mathbf{c} \in \mathbb{R}^4_+ : \sum_{i\in[4]} c_j = 1,\ c_j \geq c_j^* - \frac{\theta}{3},\ j \in [4]\}$, where $\mathbf{c}^* = (1/4, 1/4, 1/4, 1/4)$ and $\theta = 0.25$. Additionally, we also consider a second choice of C, which involves the so-called *ordered preferences* as follows: $C = C_{\text{Ord}} := \{\mathbf{c} \in \mathbb{R}^4_+ : \sum_{i\in[4]} c_j = 1,\ c_2 \geq c_1 \geq c_3 \geq c_4\}$. This choice relies on the assumption that the second criterion (based on fatalities) is the most important one, followed by the first criterion (based on property losses), the third criterion (based on air departures) and the fourth criterion (based on average daily bridge traffic). For further details on data generation, we refer to [43] and [75].

In our benchmarking analysis, we consider the equal probability case, set $n = 500$ and obtain the results for three models W-CVaR, W-Exp, and B-CVaR under each value of $\alpha \in \{0.05, 0.1, 0.15\}$. The results on allocation decisions - averaged over three randomly generated instances - are reported in Table 5.1. As seen from these results, for each setting, B-CVaR provides solutions that allocate most of the budget (at least 51%) to the area with the highest risk (New York). This is primarily due to that fact that New York has a large (58.61%) allocation in the RAND benchmark. On the other hand, the budget percentage allocated to the five urban areas with lower risk cities is less than 18 and 12 for the scalarization sets C_{Base} and C_{Ord}, respectively. Since the set C_{Ord} involves the scalarization vectors giving more priority to the second criterion (based on fatalities), B-CVaR suggests to allocate even more budget to New York, the most populated area with a significantly higher risk share associated with fatalities; for the raw data for fatalities and the remaining three criterion see Table 1 in [43]. As expected, the allocation decisions obtained by the B-CVaR model with benchmarking constraints are sensitive to the particular benchmark allocations. On the other hand, the robust risk-neutral model W-Exp provides a more "averaged" solution compared to B-CVAR and W-CVAR. For both choices of the scalarization set, W-Exp always allocates more budget to the areas with medium risk comparing to the other models. For example, for the instances with C_{Base} and $\alpha = 0.05$, it allocates almost three percent more budget to such areas than W-CVAR, and this behavior is also observed under the other settings. The results of W-Exp are consistent with its "risk-neutral" nature.

Finally, we would like to emphasize that W-CVaR allocates more budget to the areas with lower risk compared to the other models. In particular, for the instances with the scalarization set C_{Base}, W-CVaR allocates on average four percent more budget to such areas than W-Exp. These results are consistent with the risk-averse perspective of W-CVaR. Moreover, it is much less conservative than B-CVaR with respect to its allocation to New York.

We next provide some insights about the solution times of our proposed W-CVaR model for the instances under consideration. All computations in this study are performed on a 64-bit Windows Server 2012 R2 Datacenter with 2.40GHz Intel Xeon CPU E5-2630 processor with 32 GB RAM, unless otherwise stated. The vertices of both types of scalarization sets are known and there are only four of them. Thus, we could easily solve W-CVaR using the compact LP formulation (5.10). For the HSBA instances with C_{Base}, $\alpha = 0.1$, and $n = 500$, it takes at most 20 seconds to obtain an optimal solution; even for $n = 5000$ it takes at most 60 seconds. We observe that while the cut generation algorithm we propose is only essential for cases where the number of extreme points of C is exponential, it could also be useful in cases where the number of extreme points is small. For example, for C_{Base}, the compact LP takes 200 seconds on average for the three hardest instances with $n = 5000$ and $\alpha = 0.15$, whereas the cut generation algorithm takes on average 20 seconds, and generates only three extreme points of C_{Base}. This difference in solution times can be due to the large number of scenario dependent constraints and variables introduced in (5.10b)-(5.10c) for each extreme point of C.

		Allocations (%) for Areas with			Allocations (%) for Areas with		
		Higher Risk	Medium Risk	Lower Risk	Higher Risk	Medium Risk	Lower Risk
		Base Polytope (C_{Base})			Ordered Preferences (C_{Ord})		
$\alpha = 0.05$	W-CVaR	31.33	35.70	32.98	51.30	29.30	19.40
	W-Exp	32.90	38.56	28.53	48.83	34.33	16.83
	B-CVaR	52.03	30.41	17.56	57.10	31.43	11.47
$\alpha = 0.10$	W-CVaR	31.30	35.50	32.20	51.13	31.83	17.03
	W-Exp	32.93	38.63	28.43	48.83	34.33	16.83
	B-CVaR	52.00	30.57	17.43	56.93	31.40	11.67
$\alpha = 0.15$	W-CVaR	31.20	35.93	32.87	50.53	31.23	18.23
	W-Exp	32.90	38.47	28.63	48.73	34.37	16.90
	B-CVaR	51.77	30.92	17.32	56.93	31.17	11.90
RAND Benchmark		58.61	34.31	7.07	58.61	34.31	7.07

Table 5.1: Model benchmarking results for the HSBA data with $n = 500$

5.5.2 Multivariate Polyhedral CVaR-Constrained Optimization

In order to perform a detailed analysis on comparing the computational performance of the alternative MIP formulations of (**CutGen – Benchmark**) under equal probabilities, we consider an additional type of problem and a class of randomly generated instances given by [55]

$$\max\{f(\mathbf{z}) \ : \ \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{R}\mathbf{z}) \geq \mathrm{CVaR}_\alpha(\mathbf{c}^\top \mathbf{Y}) \quad \forall\, \mathbf{c} \in C, \quad \mathbf{z} \in \mathbb{R}^{100}_+\}.$$

Here $\mathbf{R} : \Omega \mapsto [0,1]^{d\times 100}$ is a random matrix and the benchmark vector $\mathbf{Y}$ takes the form of $\bar{\mathbf{R}}\bar{\mathbf{z}}$ for another random matrix $\bar{\mathbf{R}} : \Omega \mapsto [0,1]^{d\times 100}$ and a given benchmark decision $\bar{\mathbf{z}} \in \mathbb{R}^{100}_+$. Following the data generation procedure presented in [55], we independently generate the entries of the matrices $\mathbf{R}$ and $\bar{\mathbf{R}}$ from the uniform distribution on the interval $[0,1]$. Moreover, the decision variables $\mathbf{z}$ and $\bar{\mathbf{z}}$ are are independently and uniformly generated from the interval $[100, 500]$. This data generation scheme directly provides us with the realizations of two d-dimensional random vectors $\mathbf{X} = \mathbf{R}\mathbf{z}$ and $\mathbf{Y} = \bar{\mathbf{R}}\bar{\mathbf{z}}$; such an approach is sufficient since we only focus on solving the cut generation problem given the random vectors $\mathbf{X}$ and $\mathbf{Y}$. On the other hand, for the HSBA instances, $\mathbf{Y}$ is already well-defined (since the benchmark allocations are given) while the realizations of the random vector $\mathbf{X}$ are obtained using a different approach. In particular, we solve the corresponding RMP-B once, and use its optimal solution to calculate the realizations of the associated 4-dimensional random vector $\mathbf{X}$. For more details on both types of data sets (HSBA and random data sets), we refer to [55].

All the optimization problems are modeled with the AMPL mathematical programming language. All runs were executed on 4 threads of a Lenovo(R) workstation with two Intel® Xeon® 2.30 GHz CE5-2630 CPUs and 64 GB memory running on Microsoft Windows Server 8.1 Pro x64 Edition. All reported times are elapsed times, and the time limit is set to 3600 seconds unless otherwise stated. `CPLEX` 12.2 is invoked with its default set of options and parameters. If optimality is not proven within the time allotted, we record both the best lower bound on the optimal objective value (retrieved from `CPLEX` and denoted by LB) and the best available objective value (denoted by UB). Since the optimal objective function can take any value including 0, we report the following relative optimality gap: ROG $= |\,\mathrm{LB} - \mathrm{UB}\,|/(|\,\mathrm{LB}\,|)$.

One can obtain slightly different versions of the presented MIP formulations by applying the RLT techniques for different types of available information (such as the valid lower and upper bounds on the scalarization vectors). We next provide the alternative MIP formulations of (**CutGen − Benchmark**) for which we report results in Tables 5.2-5.3.

- (**MIP − CVaR**): The best available benchmark model proposed by [55]; it is based on the VaR representation (5.12) and its formulation is given by (5.22). For further computational enhancements, we added the valid inequality (5.17), and deleted the set of Big-M constraints (5.12d).

- (**MIP_VaR_Special**): This new formulation is also based on the VaR representation (5.12) but it is valid for the case of equal probabilities. Its formulation is given in (5.28); (5.12d) is deleted as in (**MIP − CVaR**).

- (**MIP − Special**): This new model is obtained by using the RLT-based strengthening for (5.23). The formulation (5.24) involves the inequalities obtained by applying the RLT procedure based on the unit simplex condition and the upper bounding constraints. We also apply the RLT procedure based on the lower bounding information ($c_j \geq L^c_j,\ j \in [d]$), which provides the following valid inequalities:

$$\gamma_{ij} \geq L^c_j \beta_i, \quad \forall\ i \in [n],\ j \in [d], \tag{5.34}$$

$$-\gamma_{ij} + c_j \geq L^c_j(1 - \beta_i), \quad \forall\ i \in [n],\ j \in [d]. \tag{5.35}$$

Unless stated otherwise, $(\mathbf{MIP-Special})$ refers to the formulation obtained by adding the constraints (5.34)-(5.35) to (5.24).

From Remark 6 for the unit simplex case, we drop the redundant constraints (those obtained from the upper and lower bounding information). In Table 5.4, "Base Special" refers to the model obtained from $(\mathbf{MIP-Special})$ by deleting the constraints (5.14c)-(5.14d) and (5.34)-(5.35); it only involves the most effective constraints (obtained from the unit simplex condition).

For the HSBA problem, we report the results averaged over two instances with different benchmarks (based on Government and RAND benchmarks) for each combination of α and n. For the random data set, we randomly generate two instances and report the average statistics. In all the tables in this section, the "Time" column reports the average solution time and the "B&B Nodes" column collects the number of nodes used during the branch-and-cut process.

From Table 5.2, we can see that $(\mathbf{MIP-Special})$ solves a majority of the test instances in the shortest amount of time. However, there are some instances (for example, for HSBA data, unit simplex, $\alpha = 0.01, n = 1000, 1500$) for which $(\mathbf{MIP-Special})$ only solves one out of the four instances within the time limit as opposed to (**MIP_VaR_Special**) which solves three of the instances within the limit. Furthermore, both new formulations we propose significantly outperform the existing formulation $(\mathbf{MIP-CVaR})$ for the equal probability case.

Furthermore, we can apply the computational enhancements proposed in [55] to the proposed formulations, namely variable fixing, bounding and a class of valid inequalities referred to as the ordering inequalities (on the β variables). The variable fixing method recognizes scenarios which are guaranteed to be larger than VaR, and fixes the corresponding β variables to zero. In addition, for the existing MIP (5.23) and (**MIP − Special**), we introduce upper and lower bounds on $\mathrm{CVaR}_\alpha(\mathbf{c}^\top\mathbf{X})$, for the others which involve the z decision variable (representing the VaR) we introduce upper and lower bounds on $\mathrm{VaR}_\alpha(\mathbf{c}^\top\mathbf{X})$. Table 5.3 summarizes our computational experience with using these enhancements. The 'Remaining Bin. Var.' column reports remaining percentage of binary variables after the preprocessing, and the '# of Order. Ineq.' column represents the number of ordering inequalities added to the formulations. Observe that there is a significant reduction in the number of binary variables. Furthermore, many ordering inequalities are added to strengthen the formulation. As a result, instances that were not solvable to optimality by any of the methods (reported in Table 5.2) can now be solved to optimality with at least one of the new formulations. We would also like to note that the total time spent on preprocessing (for calculating the Big-M coefficients and handling all the enhancements - fixing, bounding, and ordering inequalities), which is not included in the times reported, is negligible.

Effectiveness of the MIP formulations strengthened by the RLT procedure

In this section, we use additional information on C to obtain stronger RLT formulations. Our experiments are reported in Table 5.4, for the scalarization sets C_{Base} and C_{Ord}. We observe that the RLT-based strengthening using only the unit simplex information (5.15b)-(5.15d), reported in the column titled Base Simplex, is not very effective. Recall (Remark 6) that when there exists an index $j \in [d]$ such that $\tilde{M}_j = \max\{c_j : \mathbf{c} \in C\} < 1$, the constraints (5.14c)-(5.14d) are not redundant for (**MIP − Special**). In fact, for the HSBA instances, including these inequalities in (**MIP − Special**) leads to a significant reduction in the computational

		(**MIP – CVaR**)		(**MIP_VaR_Special**)		(**MIP – Special**)	
		Time; [ROG]	B&B Nodes	Time; [ROG]	B&B Nodes	Time; [ROG]	B&B Nodes
HSBA Data		**Base Polytope** & $\alpha = 0.01$					
n	1500	705.8	1524.2	447.1	1781.9	3.1	0.1
	2000	1225.1	3095.9	†1840 [50]	3510.0	6.4	0.0
	2500	†2313 [50]	5439.5	†1846 [50]	2995.9	9.1	0.0
	3000	†2275 [50.5]	3712.3	†1970 [50]	2658.1	10.1	0.0
	5000	†[1415.8]	4594.0	†[56.2]	6170.4	34.7	0.0
		Base Polytope & $\alpha = 0.05$					
n	500	109.5	1422.6	109.6	1514.2	0.7	0.0
	1000	1667.5	11976.4	†1829 [50]	8823.9	2.9	0.0
	1500	†[100.3]	14627.9	†2316 [50]	7430.1	9.2	0.5
	2000	†[451.9]	9696.0	†3071 [50]	8230.4	8.5	1.3
	2500	†[174]	5998.7	†[57.2]	6928.7	†1837 [50]	2028.8
	3000	⋄	1008.0	⋄[21.4]	4073.8	95.6	18.7
	5000	†[145.5]	2901.7	†[77]	1332.6	†[50.2]	1047.9
		Unit Simplex & $\alpha = 0.01$					
n	500	96.1	546.5	60.5	441.8	4.7	73.9
	1000	1122.2	4388.1	1130.6	4139.8	†2048 [42.6]	26341.8
	1500	†[129.6]	6424.9	†3245 [50]	7059.8	†[81.3]	29985.7
	2000	†[108.5]	5910.1	†[159.7]	5327.7	†[215.9]	22330.7
	2500	†[106.2]	1704.0	†[118.2]	4784.6	†[188]	19352.6
		Unit Simplex & $\alpha = 0.05$					
n	300	211.4	2796.1	186.3	2833.4	†2001 [42.2]	71177.5
	500	†2425 [145.3]	17459.6	1854.4	15871.0	†[125.1]	54251.9
Random Data		**Unit Simplex** & $d = 4$ & $\alpha = 0.01$					
n	1000	581.4	2588.0	437.1	2800.6	60.3	820.0
	1500	†[45.6]	7961.1	†3131 [47.6]	10201.0	†[56.9]	25972.6
	2000	†[97.5]	6196.6	†[98]	10282.1	†[73.2]	22182.7
		Unit Simplex & $d = 6$ & $\alpha = 0.01$					
n	500	542.2	4639.2	258.9	2540.5	5.3	59.2
	1000	†[89.5]	9098.1	†[91.8]	15183.7	†1890 [4]	15124.4

Table 5.2: Computational performance of the alternative MIPs for (**CutGen – CVaR**) under equal probability case

†: Time limit with integer feasible solution; ⋄: Time limit with no integer feasible solution. B&B Nodes are reported in hundreds.

	Existing-Special (5.23) Time; [ROG]	B&B Nodes	$(\mathbf{MIP-CVaR})$ Time; [ROG]	B&B Nodes	$(\mathbf{MIP\_VaR\_Special})$ Time; [ROG]	B&B Nodes	$(\mathbf{MIP-Special})$ Time; [ROG]	B&B Nodes	Remaining Binary Var. (%)	# of Order. Ineq.
n	**HSBA Data** & **Base Polytope** & $\alpha = 0.01$									
1500	141.4	2580.9	2.9	23.4	1.8	9.8	0.3	0.0	3.6	353.0
2000	†[820.36]	19786.1	16.0	112.5	5.2	33.3	0.7	0.0	4.5	745.0
2500	†[290.89]	14933.2	23.0	160.8	15.2	74.2	1.0	0.0	4.3	1168.5
3000	†[422.16]	13630.5	42.3	245.3	24.3	99.4	1.0	0.0	4.4	1908.5
5000	†[221.54]	5731.1	366.3	1707.8	149.5	521.9	6.7	0.1	4.3	4960.0
n	**HSBA Data** & **Base Polytope** & $\alpha = 0.05$									
300	106.5	4128.0	1.1	23.2	0.7	5.5	0.1	0.0	19.2	325.0
500	†2175 [71.3]	43351.7	5.3	58.5	1.4	9.7	0.3	0.0	16.2	818.0
1000	†[739.6]	28977.7	55.6	544.3	16.2	122.4	1.4	0.1	15.0	2959.0
1500	†[441]	17373.6	470.9	3461.1	90.9	415.8	2.6	0.0	15.7	7620.5
2000	†[351]	9858.6	768.7	3538.1	171.9	510.9	5.3	0.0	14.9	12656.5
2500	†[269.9]	3967.2	†3021 [9.2]	9700.8	813.0	1932.5	13.6	0.6	15.8	22194.5
3000	†[253.1]	2399.3	†[66.5]	8618.1	1029.1	1809.3	20.7	0.3	15.9	30857.0
5000	†[250.9]	632.4	†[138.5]	2248.4	†[74.7]	1671.8	42.7	0.3	15.5	84272.0
n	**HSBA Data** & **Unit Simplex** & $\alpha = 0.01$									
500	788.6	32724.9	5.8	83.5	2.4	41.0	0.6	11.3	22.2	170
1000	†[106]	49324.4	105.1	788.0	81.8	640.7	†1847 [14.7]	37653.9	23.9	771.5
1500	†[105.7]	29495.4	374.5	2054.1	290.1	1631.1	†2002 [49.3]	21375.9	21.9	1837.5
2000	†[106.3]	19970.0	1706.1	7224.5	1172.6	4874.4	†[233.3]	27470.6	22.2	3237.5
2500	†[105]	14320.9	†[393.8]	9854.0	†3469 [50]	10137.6	†[1834.8]	15418.6	22.9	5386.5
n	**HSBA Data** & **Unit Simplex** & $\alpha = 0.05$									
300	†[107]	72202.6	55.2	946.3	39.9	720.5	132.7	4143.7	52.8	761.5
500	†[107.1]	35273.8	473.6	5888.6	352.4	4430.1	†2756 [44.2]	33873.7	53.3	2049
n	**Random Data** & **Unit Simplex** & $d = 4$ & $\alpha = 0.01$									
500	58.8	1482.2	2.5	40.3	3.5	54.0	0.7	12.3	17.4	146.0
1000	†[76.1]	67673.1	33.3	321.0	29.9	334.8	6.7	120.9	14.9	547.0
1500	†[89.4]	48570.4	184.2	1573.3	186.2	1819.4	†3514 [2.2]	48577.2	15.0	1261.5
2000	†[89.9]	30300.3	826.3	4878.4	835.4	6774.6	†[58.5]	32174.8	15.2	2287.5
n	**Random Data** & **Unit Simplex** & $d = 6$ & $\alpha = 0.01$									
300	22.0	492.4	4.2	63.5	4.2	67.7	0.4	3.5	30.0	52.0
500	†2754 [41.3]	37181.5	50.7	503.8	50.1	706.9	2.0	39.9	31.1	191.0
1000	†[80.6]	48415.9	1810.5	10147.2	1464.9	12522.5	276.4	4806.8	30.5	953.5

Table 5.3: Computational performance of the alternative enhanced MIPs (fixing, bounding, ordering inequalities) for $(\mathbf{CutGen-CVaR})$ under equal probability case

†: Time limit with integer feasible solution.
B&B Nodes are reported in hundreds.

time as reported in the second column of Table 5.4. It is surprising to observe that (**MIP − Special**) could solve some instances in very short CPU time, while it reaches the time limit when (5.14c)-(5.14d) are dropped.

When we have the extreme points of C, we can easily obtain the upper and lower bounds on the components of $\mathbf{c}$. For C_{Ord} including the ordered preference constraints $c_j \geq c_{j+1}$, we obtain the corresponding inequalities obtained by using the RLT (see Proposition 30):

$$\gamma_{ij} \geq \gamma_{ij+1}, \quad \forall\, i \in [n],\ j \in [d-1], \tag{5.36}$$

$$\gamma_{ij} - \gamma_{ij+1} \leq c_j - c_{j+1}, \quad \forall\, i \in [n],\ j \in [d-1]. \tag{5.37}$$

In addition, for this case, $\tilde{\mathbf{M}} = (1, 1/2, 1/3, 1/4)$ and $\mathbf{L}^c = (1/4, 0, 0, 0)$. In our computational experiments reported in Table 5.4, we use the RLT strengthening of the upper bounding inequalities and the ordered preference constraints defining C_{Base}.

Table 5.4 demonstrates that the most effective solution method for cut generation under equal probabilities is to use the formulation (**MIP − Special**) with all enhancements: fixing, bounding, ordering inequalities on β, and the RLT-based strengthening using the additional inequalities defining C.

Effectiveness of the new valid inequalities

In this section, we study the computational effectiveness of the proposed valid inequalities (5.25). Although there are polynomially many inequalities (5.25), adding all of them ($O(m^2 n)$) into a MIP formulation of (**CutGen − Benchmark**) may not be effective due to the large number of scenarios, m and n. Therefore, we implement a branch-and-cut algorithm that adds inequalities (5.25) to the MIP formulation as they are violated. To further limit the number of inequalities added, instead

	Base Special		Base & UB Ineq.		Base & UB & LB or RLT Ord. Ineq.		Base Special F&B&Order. Ineq.		Base & UB Ineq. F&B&Order. Ineq.		Base & UB & LB or RLT Ord. Ineq. F&B&Order. Ineq.			
	Time; [ROG]	B&B Nodes	Time; [ROG]	B&B Nodes	Time; [ROG]	B&B Nodes	Time; [ROG]	B&B Nodes	Time; [ROG]	B&B Nodes	Time; [ROG]	B&B Nodes	Remaining Binary Var. (%)	# of Order. Ineq.
n	**HSBA Data** & **Base Polytope** & $\alpha = 0.01$ (Time limit=1800 sec.)													
1500	†921.7 [6.6]	2832.6	3.6	0.2	3.4	0.1	1.2	4.3	0.6	0.3	0.2	0.0	3.6	353
2000	†942.5 [18.7]	2225.0	6.8	0.0	7.2	0.0	18.3	102.4	1.3	0.2	0.4	0.0	4.5	745
2500	†1642 [29.6]	2234.7	7.7	0.0	10.1	0.0	5.1	8.2	1.5	0.0	0.7	0.0	4.3	1168.5
3000	†[87.5]	2052.9	15.5	0.0	12.8	0.0	57.9	247.4	2.2	0.0	0.7	0.0	4.4	1908.5
5000	†[115.1]	832.5	146.3	12.2	37.4	0.0	†908 [4.2]	1848.6	10.3	0.3	5.7	0.1	4.3	4960
n	**HSBA Data** & **Base Polytope** & $\alpha = 0.05$ (Time limit=1800 sec.)													
500	†902 [5.3]	3734.6	1.0	0.0	0.8	0.0	1.3	3.2	0.4	0.0	0.2	0.0	16.2	818
1000	†1241 [15.9]	3432.8	7.9	12.3	3.8	0.0	169.9	681.4	2.5	0.2	1.3	0.1	15.0	2959
1500	†[72.5]	2531.5	28.3	12.1	10.3	0.5	†903 [1.9]	1583.6	8.0	4.0	2.7	0.0	15.7	7620.5
2000	†930 [33.1]	758.9	56.8	12.2	12.3	1.3	†906 [2.2]	790.7	11.0	0.1	6.0	0.0	14.9	12656.5
2500	†[83.5]	931.8	104.6	27.7	†932 [50]	674.0	†929 [5.4]	714.3	29.0	2.0	14.7	0.6	15.8	22194.5
3000	†[87.8]	786.5	317.0	120.7	83.6	18.7	†920.8 [7.3]	431.1	43.9	1.2	22.5	0.3	15.9	30857
5000	†[91.3]	441.5	†[50.3]	390.6	†[50.2]	421.6	†1104 [10.4]	74.8	141.6	1.6	44.2	0.3	15.5	84272
n	**HSBA Data** & [**Unit Simplex** & **Ordered Preference**] & $\alpha = 0.01$ (Time limit=3600 sec.)													
1000	†1809 [50]	21268.5	995.8	4017.4	242.4	746.5	7.8	219.0	4.7	70.1	3.8	61.9	7.6	314.0
1500	†[49.6]	26700.0	†1843 [45.9]	4786.3	†1802 [37.9]	3687.5	93.0	1605.9	36.4	444.6	42.5	521.8	7.2	687.5
2000	†1832 [49.5]	10935.4	†1873 [49]	3325.9	†1803 [48.5]	3643.8	1035.1	15729.2	271.2	2602.5	395.4	3575.9	7.2	1192.5
2500	†[54.4]	14035.7	†1831 [50]	2562.0	†1835 [50]	1063.8	†[50.5]	31553.8	†1914 [50]	7503.4	†1802 [50]	14374.4	7.0	1899.5
3000	† [50.2]	9335.9	†1842 [50]	1244.5	†1820 [50]	940.0	†[50.9]	29419.5	†1929 [50]	10192.9	†1803 [50]	10223.6	6.8	2398.0
n	**Random Data** & [**Unit Simplex** & **Ordered Preference**] & $d = 4$ & $\alpha = 0.01$ (Time limit=3600 sec.)													
1500	†[57.2]	29602.1	270.8	525.3	16.1	11.5	3.9	47.1	2.5	16.0	1.2	2.5	4.0	400.5
2000	†[61.9]	21979.6	†3294 [11.8]	4924.4	39.5	35.8	10.1	140.9	6.1	53.2	2.2	10.9	3.6	555.5
3000	†[84.1]	12807.2	†[68.8]	2756.1	1776.7	2086.5	989.6	13214.0	501.7	4924.0	30.5	187.0	3.9	1623.5

Table 5.4: Computational performance of the RLT procedure for (**CutGen − CVaR**) under equal probability case

†: Time limit with integer feasible solution. B&B Nodes are reported in hundreds.
UB inequalities: (5.14c)-(5.14d). LB inequalities: (5.34)-(5.35). RLT Ord. Ineq.: (5.36)-(5.37); used only for the instances with ordered preference structure - see Proposition 30.

of adding all the violated inequalities (5.25) for each s and $t \in [m] \setminus s$, we use a separation algorithm where we add one violated inequality for the smallest $t > s$ for a given $s \in [m]$. In our implementation, we consider (**MIP − Special**) as the base MIP formulation, and add the violated inequalities (5.25) as user cuts; this approach to solve (**CutGen − Benchmark**) is implemented in C language using CPLEX 12.6. Moreover, we only seek and add the violated inequalities (5.25) in the root node and every 10,000 branch-and-bound nodes for the first 50,000 nodes. We test the computational performance of (**MIP − Special**) with and without the proposed inequalities. For each approach, one core is used and the dynamic search is turned off. In these experiments, we solve random instances where each component of the vectors $\mathbf{X}$ and $\mathbf{Y}$ is independently generated from a discrete uniform distribution on the interval [1,10]. We assume that the number of realizations of $\mathbf{X}$ and $\mathbf{Y}$ are equal (i.e., $n = m$), $\alpha = 0.01$, and all the outcomes are equally likely.

		(**MIP − Special**)		(**MIP − Special**) & Ineq. (5.25)			Remaining	# of
d	n	Time	B&B nodes	Time	B&B nodes	# of Ineq. (5.25)	Bin. Var. (%)	Order. Ineq.
	900	51	91260	31	49228	31	53.3	24822
3	1000	27	39615	17	16984	12	41.0	30975
	1100	101.3	349827	36	124928	15	46.9	33412
	600	44.3	53896	32.7	47387	17	53.0	10924
4	700	378	779551	358	766937	8	38.1	8457
	800	329	382508	292	346429	31	50.1	11065

Table 5.5: Effectiveness of the valid inequalities (5.25) for $\alpha = 0.01$ (with fixing and ordering inequalities)

In Table 5.5, we report the results - averaged over three instances - on the performance of two approaches. In both approaches, we also take advantage of two computational enhancements, namely variable fixing and ordering inequalities. In addition to the columns defined in earlier tables, the " # of Ineq. (5.25)" column collects the number of the proposed inequalities (5.25) that are added to (**MIP − Special**). From this column, we see that only a small subset of inequalities (5.25) is added. Nevertheless, according to the results, we are still able to benefit from these inequalities for these instances. However, we note that as we increase d, the effectiveness of inequalities (5.25) within a branch-and-cut algorithm decreases. We note that it becomes more effective to use the base formulation (**MIP − Special**) for this data set as α increases. In our experience, the base formulation (**MIP − Special**) is also more effective for the HSBA data and the random data described in Section 5.5.2.

5.6 Conclusions

In this chapter, we study risk-averse models for multicriteria optimization problems under uncertainty. First, we introduce a model that optimizes the *worst-case multivariate CVaR*, and develop a finitely convergent delayed cut generation algorithm for finite probability spaces. In addition, for the cut generation problem, which is in general a mixed-integer program, we give a stronger formulation for the equiprobable case using the reformulation linearization technique. Next, we observe that similar polyhedral enhancements are also useful for a related class of *multivariate CVaR-constrained* optimization problems that has attracted attention recently. Our computational study demonstrates the effectiveness of the proposed solution methods for both classes of models.

Chapter 6: Contributions and Future Work

In this chapter, we summarize the research presented in this thesis and propose potential extensions of current work. In this dissertation, we propose novel mathematical models for risk-averse optimization problems under uncertainty. In addition, we propose state-of-the-art algorithms using mixed-integer programming techniques, including cutting plane and decomposition methods, dynamic programming and other related algorithms.

In Chapter 2, we study a two-stage chance-constrained program where feasible second stage solutions incur additional costs. Earlier work in chance-constrained optimization either ignores the second-stage (recourse) decisions or the second-stage costs. In addition, we propose a new model, where we are allowed to recover from a unsuccessful strategic decision. We propose a decomposition algorithm, which uses specialized feasibility and optimality cuts. The computational experiments indicate that the proposed methods are highly efficient.

In our further study, we derived a new class of valid optimality cuts for the problems where the randomness only appears in the right-hand side of the chance-constraint. We prove that this class of inequalities is stronger than the optimality cuts presented in Chapter 2, but uses more computational effort. It would be interesting if we can test this improved version of optimality cuts against the optimality cut used in Chapter 2, to see the difference in performance. In addition, it would also be interesting if we can find more theoretical polyhedral results for this class of improved optimality cuts, and extend the result to general two-stage chance-constrained programs.

In Chapter 3, we study the polyhedral structure of the static probabilistic lot-sizing problem and propose a class of new valid inequalities. We show the strength of this class of inequalities. In addition, we show that it is practically effective. Furthermore, we propose a new formulation for general two-stage chance-constrained programs with simple recourse, which significantly reduces the number of variables and constraints of the deterministic equivalent program. The computational results show that the proposed formulation significantly outperforms the Benders decomposition algorithm.

We observe that there exist other facet-defining inequalities that are not in the form of the proposed inequalities. It would be interesting to characterize new classes of facet-defining inequalities for the static probabilistic lot-sizing problem.

In Chapter 4, we study the structure of the two-sided chance-constrained programs. We propose a class of valid inequalities, and give locally ideal formulation for this class of problem. In addition, we propose polynomial optimization and separation algorithms for the optimization problem over a substructure of the two-sided chance-constrained programs. An interesting question is whether we can further strengthen the proposed inequality using the information of the knapsack constraint in the formulation. In addition, we observe that the two-sided chance-constrained programs is a special case of general chance-constrained second-order conic program in $\mathbb{R}^2$. In the mixed-integer programming reformulation of chance-constrained second-order conic program, a continuous mixing structure with knapsack constraint is observed as the key structure. It would be interesting to further investigate this problem, and extend our results for the two-sided chance-constrained programs to general chance-constrained second-order conic programs.

In Chapter 5, we introduce a model that optimizes the *worst-case multivariate CVaR*, and propose develop a finitely convergent delayed cut generation algorithm for finite probability spaces. In addition, we propose a strong formulation for the cut generation algorithm in the equiprobable case using the reformulation linearization technique. Next, we apply this method to a related class of *multivariate CVaR-constrained* optimization problems. The computational results show that the proposed model and algorithms are effective.

In our experiments, we observe that we obtain the complete linear description of the convex hull for the *multivariate CVaR-constrained* optimization problems where $m = 2$. It would be interesting if we can prove this result formally to show that the proposed method leads to a locally ideal formulation. In addition, we observe that the valid inequality (2.23) that is used in Chapter 2 can be applied in this problem, to strengthen the linear programming reformulation of the cut-generation problem. It would be interesting if we can compare the performance of inequality (2.23) against the reformulation linearization technique in the future.

Appendix A: Stochastic dominance

In this section, we review the well-known stochastic dominance relations, which are essential for the stochastic Pareto optimality definitions presented in Section 5.2.1.

The stochastic dominance relations are fundamental concepts in comparing random variables [68, 57] and have been widely used in economics and finance [see, e.g., 59]. Different from the approaches based on risk measures, in a stochastic dominance based approach, the random variables are compared by a point-wise comparison of some performance functions (constructed from their distribution functions when the order is greater than zero). We note that the lower order dominance relations ($k = 0, 1$, and 2) are the most common ones (referred to as ZSD, FSD, and SSD, respectively). We provide the formal definitions below and refer the reader to [72] and [91] for further details.

- We say that a random variable X *dominates* another random variable Y in the *zeroth order* if $X \geq Y$ everywhere, i.e., $X(\omega) \geq Y(\omega)$ for all $\omega \in \Omega$.

- An integrable random variable X *dominates* another integrable Y in the *first order* (or X is stochastically larger than Y) if $F_1(X, \eta) := \mathbb{P}(X \leq \eta) \leq F_1(Y, \eta) := \mathbb{P}(Y \leq \eta)$ for all $\eta \in \mathbb{R}$.

- For $k \geq 2$ we say that a k-integrable random variable X (i.e., $\in \mathcal{L}^k$) *dominates* another k-integrable random variable Y *in the kth order* if $F_k(X, \eta) \leq F_k(Y, \eta)$ for all $\eta \in \mathbb{R}$, where $F_k(X, \eta) = \int_{-\infty}^{\eta} F_{k-1}(X, t)\, \mathrm{d}t$ for all $\eta \in \mathbb{R}$.

- For $k = 0$, if $X(\omega) > Y(\omega)$ for all $\omega \in \Omega$, we will refer to the relation as "*strong ZSD*" and denote it by $X \succ_{(0)} Y$. For $k \geq 1$, if all the inequalities $F_k(X, \eta) \leq F_k(Y, \eta)$ are strict, then we refer to the relation as "*strong kSD*" and denote it by $X \succ_{(k)} Y$. We remark that the notion of "strong kSD" is not analogous to the notion of strict kSD, which requires that at least one of the inequalities defining the dominance relation is strict.

Appendix B: A class of facets of $conv(\mathcal{S})$.

Before we study the facets of $conv(\mathcal{S})$, we first need to establish its dimension.

Proposition 33. *$Conv(\mathcal{S})$ is a polyhedron with dimension $n + d + m - 1$.*

Proof. First, we show that $conv(S)$ is a polyhedron. First, note that the extreme rays of $conv(\mathcal{S})$ can be enumerated as follows:

$$\delta^l = (\mathbf{0}, \mathbf{0}, \mathbf{0}, 0, \mathbf{e}_l), \qquad \forall\ l \in [m] \tag{B.1a}$$

$$\delta^{m+1} =: (\mathbf{0}, \mathbf{0}, \mathbf{0}, -1, \mathbf{0}), \tag{B.1b}$$

$$\delta^{m+2} = (\mathbf{0}, \mathbf{0}, \mathbf{0}, 1, \mathbf{1}), \tag{B.1c}$$

where $\mathbf{e}_l$ is a standard basis vector of an appropriate dimension with the l-th element equal to 1 and all other elements equal to 0. We have

$$\begin{aligned} conv(\mathcal{S}) = \cup_{h=1}^{H} \{(\gamma, \mathbf{c}, \beta, \eta, \mathbf{w}) \mid \gamma = \beta \hat{\mathbf{c}}^{h^\top},\ (\mathbf{c}, \eta, \mathbf{w}) \\ = (\hat{\mathbf{c}}^h,\ \hat{\eta}^h, \hat{\mathbf{w}}^h),\ \beta \in [0,1]^n,\ \sum_{i \in [n]} \beta_i = k\} + \sum_{l=1}^{m+2} \mu^l \delta^l, \end{aligned}$$

where $(\hat{\mathbf{c}}^h, \hat{\eta}^h, \hat{\mathbf{w}}^h)$ for all $h \in [H]$ are the vertices of the polyhedron $P(C_f, \mathbf{Y})$ (see (5.32)) for a finite H. In addition, μ^l is a non-negative parameter for all $l \in [m+2]$. From [5] we know that we can take the union of these polyhedra parameterized by $\hat{\mathbf{c}}^h$, since each polyhedron shares the same recession cone (B.1). Hence, we obtain the extended formulation of $conv(\mathcal{S})$ as:

$$
\begin{aligned}
&1 = \sum_{h=1}^{H} \lambda^h,\ \beta = \sum_{h=1}^{H} \beta^h, \\
&\mathbf{c} = \sum_{h=1}^{H} \mathbf{c}^h,\ \gamma = \sum_{h=1}^{H} \gamma^h, \\
&\eta = \sum_{h=1}^{H} \eta^h + \sum_{l=m+1}^{m+2} \mu^l \delta^l, \\
&\mathbf{w} = \sum_{h=1}^{H} \mathbf{w}^h + \sum_{l=1}^{m+2} \mu^l \delta^l, \\
&\gamma^h = \hat{\mathbf{c}}^h \beta^h,\ \mathbf{c}^h = \lambda^h \hat{\mathbf{c}}^h, && \forall\ h \in [H] \\
&\eta^h = \lambda^h \hat{\eta}^h,\ \mathbf{w}^h = \lambda^h \hat{\mathbf{w}}^h && \forall\ h \in [H] \\
&\sum_{i\in[n]} \beta_i^h = k\lambda^h, && \forall\ h \in [H] \\
&0 \le \beta^h \le \lambda^h, && \forall\ h \in [H] \\
&\lambda \in \mathbb{R}_+^h, \quad \mu \in \mathbb{R}_+^{m+2}.
\end{aligned}
$$

Therefore, $conv(\mathcal{S})$ is a polyhedron. Next, we show that the dimension of $conv(\mathcal{S})$ is $n + d + m - 1$. Clearly, in the original constraints defining $\mathcal{S}$, there are two linearly independent equalities: $\sum_{j\in[d]} c_j = 1$, $\sum_{i\in[n]} \beta_i = k$. In addition, there are nd implied nontrivial equalities: $\gamma_{ij} = c_j\beta_i$, for all $i \in [n]$ and $j \in [d]$. Hence, $dim(conv(\mathcal{S})) \le n + m + d - 1$.

Consider the following set of points:

$$
\begin{aligned}
&(\mathbf{u}_v \mathbf{e}_1^\top, \mathbf{e}_1, \mathbf{u}_v, 0, 0) && \forall\ v \in [n], \\
&(\mathbf{u}_1 \mathbf{e}_j^\top, \mathbf{e}_j, \mathbf{u}_1, 0, 0) && \forall\ j \in [d] \setminus \{1\}, \\
&(\mathbf{u}_1 \mathbf{e}_1^\top, \mathbf{e}_1, \mathbf{u}_1, 0, e_l) && \forall\ l \in [m], \\
&(\mathbf{u}_1 \mathbf{e}_1^\top, \mathbf{e}_1, \mathbf{u}_1, -1, 0),
\end{aligned}
$$

where $\mathbf{u}_v$, for all $v \in [n]$ are any affinely independent vectors with k elements equal to 1 and the remaining elements equal to 0. These vectors exist because the dimension of the following system:

$$\beta \in \{0,1\}^n, \quad \sum_{i \in [n]} \beta_i = k, \tag{B.2}$$

is $n-1$. Clearly, this set of points is feasible and affinely independent. In addition, the cardinality of this set is $n+m+d$. Hence, $dim(conv(\mathcal{S})) \geq n+m+d-1$, which completes the proof. □

Proposition 34. *For any $i \in [n]$, $s \in [m]$, and $t \in [m] \setminus \{s\}$, inequality* (5.25) *is facet-defining for $conv(\mathcal{S})$ if and only if $s \in [m], t \in [m] \setminus \{s\}$ are such that $y_{sj} < y_{tj}$ and $y_{si} > y_{ti}$ for some $i, j \in [d]$.*

Proof. To show the necessity, we first note that if there exists a pair $s \in [m], t \in [m] \setminus \{s\}$ such that $y_{sj} \geq y_{tj}$ or $y_{sj} \leq y_{tj}$ for all $j \in [d]$, in other words, when the realizations under a scenario are *dominated* by the realizations under another scenario, then the corresponding inequality (5.25) is dominated. To see this, suppose that $y_{sj} \leq y_{tj}$ for all $j \in [d]$ for some pair $\forall\ s \in [m], \forall\ t \in [m] \setminus \{s\}$. Then the corresponding inequality (5.25) is dominated by the original inequality $\mathbf{c}^\top \mathbf{y}_s \geq \eta - w_s$, because the coefficients of γ_{ij} are $y_{tj} - y_{sj} \geq 0$, and $\gamma_{ij}, w_t \geq 0$. Now consider the case that $y_{sj} \geq y_{tj}$ for all $j \in [d]$ for some pair $\forall\ s \in [m], \forall\ t \in [m] \setminus \{s\}$. Then the corresponding inequality (5.25) is dominated by the original inequality $\mathbf{c}^\top \mathbf{y}_t \geq \eta - w_t$. To see this, observe that we can rewrite inequality (5.25) for this choice of s and t as, $\mathbf{c}^\top \mathbf{y}_t + \sum_{j \in [d]} (y_{sj} - y_{tj})(c_j - \gamma_{ij}) \geq \eta - w_t - w_s$. It is now easy to see that the inequality is dominated, because $y_{sj} - y_{tj} \geq 0$, $c_j \geq \gamma_{ij}$ and $w_s \geq 0$.

To show sufficiency, we need to show that for any given $i \in [n]$, $s \in [m]$, and $t \in [m] \setminus \{s\}$, there are $n+m+d-1$ affinely independent points that satisfy (5.25) at equality. From the necessity condition, we only need to consider the cases for which there exists an index $j_1 \in [d]$, such that $y_{sj_1} < y_{tj_1}$, and there exists an index $j_2 \in [d]$, such that $y_{sj_2} > y_{tj_2}$. In order to simplify the notation, and without loss of generality, throughout the rest of the proof, we let $j_1 = 1$, and $j_2 = 2$, or equivalently, $y_{s1} < y_{t1}$, and $y_{s2} > y_{t2}$.

First, we construct a set of points:

$$\mathbf{PT}_v^1 = (\mathbf{u}_v \tilde{\mathbf{e}}_v^\top, \tilde{\mathbf{e}}_v, \mathbf{u}_v, \rho_v^1, \xi_v^1), \qquad \forall\, v \in [n], \tag{B.3}$$

where if $u_{vi} = 0$, then $\tilde{\mathbf{e}}_v = \mathbf{e}_1$ and $\rho_v^1 = y_{s1}$, else if $u_{vi} = 1$, then $\tilde{\mathbf{e}}_v = \mathbf{e}_2$ and $\rho_v^1 = y_{t2}$. In addition, $\xi_{vs}^1 = \xi_{vt}^1 = 0$, and $\xi_{vl}^1 = \max\{\tilde{M}_s, \tilde{M}_t\}$ for all $v \in [n]$ and $l \in [m] \setminus \{s, t\}$. Clearly, the set of points defined in (B.3) are affinely independent feasible points, and satisfy (5.25) at equality. Next, we construct a set of points:

$$\mathbf{PT}_j^2 = (\tilde{\mathbf{u}}_j \mathbf{e}_j^\top, \mathbf{e}_j, \tilde{\mathbf{u}}_j, \rho_j^2, \xi_j^2), \qquad \forall\, j \in [d] \setminus \{1, 2\}, \tag{B.4}$$

where $\tilde{\mathbf{u}}_j$ is any feasible point of (B.2) with $\tilde{u}_{ji} = 0$ if $y_{sj} \leq y_{tj}$, and $\tilde{u}_{ji} = 1$ otherwise (i.e., if $y_{sj} \geq y_{tj}$), for all $j \in [d] \setminus \{1, 2\}$. In addition, $\rho_j^2 = \min\{y_{sj}, y_{tj}\}$, for all $j \in [d] \setminus \{1, 2\}$. Furthermore, $\xi_{js}^2 = \xi_{jt}^2 = 0$, and $\xi_{jl}^2 = \max\{\tilde{M}_s, \tilde{M}_t\}$ for all $j \in [d] \setminus \{1, 2\}$ and $l \in [m] \setminus \{s, t\}$. It is easy to see that the set of points defined in (B.4) are feasible, affinely independent from (B.3), and satisfy (5.25) at equality.

Furthermore, we construct the following set of points:

$$\mathbf{PT}_s^3 = (\bar{\mathbf{u}}_1 \mathbf{e}_1^\top, \mathbf{e}_1, \bar{\mathbf{u}}_1, y_{t1}, \xi_s^3) \tag{B.5a}$$

$$\mathbf{PT}_t^3 = (\bar{\mathbf{u}}_2 \mathbf{e}_2^\top, \mathbf{e}_2, \bar{\mathbf{u}}_2, y_{s2}, \xi_t^3) \tag{B.5b}$$

$$\mathbf{PT}_l^3 = \mathbf{PT}_s^3 + (0, 0, 0, 0, \mathbf{e}_l), \qquad \forall\, l \in [m] \setminus \{s, t\}, \tag{B.5c}$$

where $\bar{\mathbf{u}}_1$ is any feasible point of (B.2) with $\bar{u}_{1i} = 0$, and $\bar{\mathbf{u}}_2$ is any feasible point of (B.2) with $\bar{u}_{2i} = 1$. In addition, $\xi^3_{ss} = y_{t1} - y_{s1}$, $\xi^3_{st} = 0$, and $\xi^3_{sl} = \max\{\tilde{M}_s, \tilde{M}_t\}$ for all $l \in [m] \setminus \{s,t\}$. Similarly, $\xi^3_{ts} = 0$, $\xi^3_{tt} = y_{s2} - y_{t2}$, and $\xi^3_{tl} = \max\{\tilde{M}_s, \tilde{M}_t\}$ for all $l \in [m] \setminus \{s,t\}$. Clearly, the set of points defined by (B.5) are affinely independent feasible points which satisfy (5.25) at equality.

Finally, we construct the single point:

$$\mathbf{PT}^4 = (\mathbf{u}_1 \mathbf{c}^{*\top}, \mathbf{c}^*, \mathbf{u}_1, \eta^*, \xi^4), \tag{B.6}$$

where $\mathbf{c}^* = (c^*_1, c^*_2, 0, \ldots, 0)$, and the parameters (c^*_1, c^*_2, η^*) are uniquely defined by the following linear system:

$$\begin{aligned}
&c^*_1 + c^*_2 = 1\\
&y_{s1}c^*_1 + y_{s2}c^*_2 = \eta^*\\
&y_{t1}c^*_1 + y_{t2}c^*_2 = \eta^*,
\end{aligned}$$

or equivalently, $c^*_1 = \frac{y_{s2} - y_{t2}}{y_{s2} - y_{t2} + y_{t1} - y_{s1}}$, $c^*_2 = 1 - c^*_1$, and $0 < c^*_1, c^*_2 < 1$. In addition, $\xi^4_s = \xi^4_t = 0$, and $\xi^4_l = \max\{\tilde{M}_s, \tilde{M}_t\}$, for all $l \in [m] \setminus \{s,t\}$.

Clearly, $\mathbf{PT}^4$ is affinely independent from the points defined by (B.3), since the following matrix:

$$\begin{bmatrix} 1 & 0 & y_{s1} \\ 0 & 1 & y_{t2} \\ c^*_1 & c^*_2 & \eta^* = y_{s1}c^*_1 + y_{s2}c^*_2 \end{bmatrix}, \tag{B.7}$$

has full rank (due to $y_{t2} < y_{s2}$). In addition, it is easy to check that (B.6) is affinely independent from the points defined by (B.4) and (B.5). Furthermore, it is also a feasible point which satisfies (5.25) at equality. From (B.3)-(B.6), we obtain $n + m + d - 1$ affinely independent feasible points which satisfy (5.25) at equality. Hence, inequalities (5.25) are facet defining. $\square$

Appendix C: Valid inequalities that involve stock variables

In this section, we study the polyhedral structure of the deterministic equivalent formulation which includes the stock variables. Let $P_+ = \{(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{s}) \mid (3.3\text{b})-(3.3\text{f})\}$.

Proposition 35. *For $\ell = 2, \ldots, n$, let $T_\ell := \{t_{\ell(1)}, t_{\ell(2)}, \ldots, t_{\ell(a_\ell)}\} \subseteq T_\ell^*$, where $D_{t_{\ell(1)}} \geq D_{t_{\ell(2)}} \geq \cdots \geq D_{t_{\ell(n)}}$. For $j \in \Omega$, the inequalities*

$$s_{j(\ell-1)} + (D_{t_{\ell(1)}} - D_{j\ell-1})x_\ell + \sum_{p=1}^{a_\ell} (D_{t_{\ell(p)}} - D_{t_{\ell(p+1)}}) z_{t_{\ell(p)}} \geq D_{t_{\ell(1)}} - D_{j\ell-1}, \qquad \text{(C.1)}$$

are valid for P_+.

Proof. If $x_\ell = 1$, then inequality (C.1) is trivially satisfied. Otherwise, $y_\ell = 0$. Because $s_{j(\ell-1)} \geq \sum_{p=1}^{\ell-1} y_p - D_{j(\ell-1)} = \sum_{p=1}^{\ell} y_p - D_{j(\ell-1)}$, the validity of inequality (C.1) follows from the validity of the mixing inequality (3.5) for time period ℓ. $\square$

Example 1. *(Continued.) Let $\ell = 2, j = 1$ and $T_\ell = \{3, 4\}$, then we obtain:*

$$s_{11} + (D_{t_{2(1)}} - D_{t_{1(1)}})x_2 + (D_{t_{2(1)}} - D_{t_{2(2)}})z_3 + (D_{t_{2(2)}} - D_{t_{2(3)}})z_4 \geq (D_{t_{2(1)}} - D_{t_{1(1)}}),$$

which is equivalent to:

$$s_{11} + 5x_2 + z_3 + z_4 \geq 5.$$

In fact, this inequality is a facet-defining inequality for this problem, as we show in Proposition 36.

Next, we show the strength of the proposed inequalities (C.1).

Proposition 36. *For $\ell = 2, \ldots, n$ and $T_\ell \subseteq T_\ell^*$, if $\sigma_{\ell-1(1)} \notin T_\ell^* \cup \{\sigma_{\ell(k+1)}\}$, $j = \sigma_{\ell-1(1)}$ and $t_{\ell(1)} = \sigma_{\ell(1)}$, then inequality* (C.1) *is facet-defining for $conv(P_+)$.*

Proof. First, we show that under the conditions stated in Proposition 36, inequality (C.1) is facet-defining for the convex hull of the polyhedron: $P_{s_{j(\ell-1)}} = \{(\mathbf{x}, \mathbf{y}, \mathbf{z}, s_{j(\ell-1)}) \in B^n \times \mathbb{R}_+^n \times B^m \times \mathbb{R}_+ \mid s_{j(\ell-1)} \geq \sum_{p=1}^{\ell-1} y_p - D_{j(\ell-1)}, (3.3\text{b}) - (3.3\text{d})\}$, in which we only consider the stock variable for scenario $j = \sigma_{\ell-1(1)}$ at time period $\ell - 1$. To show that inequality (C.1) is facet-defining for $conv(P_{s_{j(\ell-1)}})$, we need to find $dim(P_{s_{j(\ell-1)}}) = 2n + m$ affinely independent points $(\mathbf{x}, \mathbf{y}, \mathbf{z}, s_{j(\ell-1)})$ that satisfy inequality (C.1) at equality.

Let $g(t_{i(p)})$, for all $t_{i(p)} \in T_i$ and $i + 1 \in \bar{S} \cup \{\ell + 1\}$, be a unique mapping such that the scenario $t_{i(p)}$ has the $g(t_{i(p)})$-th largest cumulative demand at time period i. We first consider the following set of feasible points:

$$(\mathbf{e}_1 + \mathbf{e}_{\ell+1}, \bar{\mathbf{y}}_p^\ell + \mathbf{e}_{\ell+1} M_{\ell+1}, \sum_{i=1}^{g(t_{\ell(p)})-1} \mathbf{e}_{\sigma_{\ell(i)}}, \bar{y}_{p1}^\ell - D_{\sigma_{\ell-1(1)}}), \quad p \in [a_\ell + 1],$$

where $\bar{\mathbf{y}}_p^q$ is defined in the proof of Proposition 10. To see the feasibility of these points, note that if $\sigma_{\ell-1(1)} \notin T_\ell^* \cup \{\sigma_{\ell(k+1)}\}$, we must have $\bar{y}_{a_\ell+1,1}^\ell - D_{\sigma_{\ell-1(1)}} \geq 0$. Hence, we obtain $a_\ell + 1$ affinely independent points that satisfy inequality (C.1) at equality.

Next, consider the following set of points:

$$(\mathbf{e}_1 + \mathbf{e}_{\ell+1}, \bar{\mathbf{y}}_1^\ell + \mathbf{e}_{\ell+1} M_{\ell+1}, \mathbf{e}_p, \bar{y}_{11}^\ell - D_{\sigma_{\ell-1(1)}}), \quad \forall p = \Omega \setminus T_\ell.$$

These $m - a_\ell$ points are feasible, affinely independent from all other points, and satisfy inequality (C.1) at equality. Next, we consider the following set of points:

$$(\mathbf{e}_1 + \mathbf{e}_\ell, \bar{\mathbf{y}}_1^{\ell-1} + \mathbf{e}_\ell M_\ell, \mathbf{0}, 0),$$

$$(\mathbf{e}_1 + \mathbf{e}_\ell + \mathbf{e}_p, \bar{\mathbf{y}}_1^{\ell-1} + \mathbf{e}_\ell M_\ell, \mathbf{0}, 0), p \in N \setminus [\ell],$$

$$(\mathbf{e}_1 + \mathbf{e}_\ell + \mathbf{e}_p, \bar{\mathbf{y}}_1^{\ell-1} + \mathbf{e}_\ell M_\ell + \mathbf{e}_p \triangle, \mathbf{0}, 0), p \in N \setminus [\ell],$$

where $0 < \triangle < M_p$, for all $p \in N \setminus [\ell]$. These $2(n-\ell)+1$ points are feasible, affinely independent from all other points, and satisfy inequality (C.1) at equality.

Next, we consider the following set of points:

$$(\mathbf{e}_1 + \mathbf{e}_\ell + \mathbf{e}_{\ell+1}, \bar{\mathbf{y}}_1^{\ell-1} + \mathbf{e}_\ell(M_\ell - \triangle_1) + \mathbf{e}_{\ell+1} M_{\ell+1}, \mathbf{0}, 0),$$

$$(\mathbf{e}_1 + \mathbf{e}_p + \mathbf{e}_\ell, \bar{\mathbf{y}}_1^{\ell-1} + \mathbf{e}_\ell M_\ell, \mathbf{0}, 0), p \in [\ell-1] \setminus \{1\},$$

$$(\mathbf{e}_1 + \mathbf{e}_p + \mathbf{e}_\ell, \bar{\mathbf{y}}_1^{\ell-1} + \mathbf{e}_\ell M_\ell + \triangle_2(\mathbf{e}_p - \mathbf{e}_1), \mathbf{0}, 0), p \in [\ell-1] \setminus \{1\},$$

where $0 < \triangle_1 \le \bar{y}_{11}^{\ell-1}$, and $0 < \triangle_2 \le \min\{\bar{y}_{11}^{\ell-1} - D_{\sigma_{1(1)}}, M_p\}$, for all $p \in [\ell-1] \setminus \{1\}$. It is easy to see that these $2\ell - 3$ points are feasible, affinely independent from other points, and satisfy inequality (C.1) at equality. Finally, consider the feasible point: $(\mathbf{e}_1 + \mathbf{e}_\ell, \mathbf{y}^* + \mathbf{e}_\ell M_\ell, \mathbf{e}_{\sigma_{\ell-1(1)}}, 0)$, where $y_1^* = D_{\sigma_{\ell-1(2)}}$, and $y_i^* = 0$, for all $i = 2, \ldots, n$. This point is affinely independent from all other points, and satisfies inequality (C.1) at equality. Hence, we have $2n+m$ affinely independent points that satisfy inequality (C.1) at equality, which shows that the proposed inequality is facet-defining for $conv(P_{s_{j(\ell-1)}})$.

To show that the proposed inequality is also facet-defining for $conv(P_+)$, let: $(\tilde{\mathbf{x}}^p, \tilde{\mathbf{y}}^p, \tilde{\mathbf{z}}^p, \tilde{s}^p_{j(\ell-1)}), p \in [2n+m]$, be the affinely independent points constructed for $conv(P_{s_{j(\ell-1)}})$. Then, we construct the set of points: $(\tilde{\mathbf{x}}^p, \tilde{\mathbf{y}}^p, \tilde{\mathbf{z}}^p, \tilde{\mathbf{s}}^p), p \in [2n+m]$, where $\tilde{s}^p_{qi} = \max\{\sum_{u=1}^i \tilde{y}^p_u - D_{qi}\ ,\ 0\}$ for $q \in \Omega, i \in N$. These "extended" points are feasible, affinely independent, and satisfy inequality (C.1) at equality. Finally, for each inventory variable s_{pi} such that $p \neq \sigma_{\ell-1(1)}$ or $i \neq \ell - 1$, we construct the set of points: $(\tilde{\mathbf{x}}^1, \tilde{\mathbf{y}}^1, \tilde{\mathbf{z}}^1, \tilde{\mathbf{s}}^1) + (\mathbf{0}, \mathbf{0}, \mathbf{0}, \triangle \mathbf{e}_{pi}), p \neq \sigma_{\ell-1(1)}, i \neq \ell - 1$, where $\triangle > 0$, and $\mathbf{e}_{pi}$ is an $m \times n$ dimensional matrix such that the (p, i)-th entry equals 1, and all other entries are 0. These $nm - 1$ points are feasible, affinely independent from other points, and satisfy inequality (C.1) at equality. Hence, we obtain $2n + m + mn - 1$ feasible, affinely independent points that satisfy inequality (C.1) at equality, which completes the proof. □ □

Separation of inequalities (C.1): Given a fractional solution of the deterministic equivalent formulation $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}, \hat{\mathbf{s}})$, we solve the problem (3.12) to obtain $\hat{Y}_i, i \in N \backslash \{1\}$. Then, with a linear pass, we add the violated inequality (C.1) for $\ell \in N \setminus \{1\}, j = \sigma_{\ell-1(1)}$, if $\hat{s}_{j(\ell-1)} + (D_{t_{\ell(1)}} - D_{j(\ell-1)})\hat{x}_\ell + \hat{Y}_\ell < D_{t_{\ell(1)}} - D_{j(\ell-1)}$. Otherwise, there is no violated inequalities (C.1). The overall running time is $O(nk \log(k))$. In addition, since we consider a single time period at a time, the separation procedure is exact.

Appendix D: A Benders decomposition algorithm

There are mn stock variables, which could cause computational difficulty as the size of the problem increases. In this section, we study a Benders decomposition algorithm. Let θ_j, for all $j \in \Omega$, represent the additional variable that captures the second-stage cost of scenario j. The relaxed master problem is

$$\begin{aligned}
\textbf{MASTER: } \min\ & \mathbf{f}^\top \mathbf{x} + \mathbf{c}^\top \mathbf{y} + \frac{1}{m} \sum_{j=1}^{m} \pi_j \theta_j \\
\text{s.t. } & \sum_{i=1}^{t} y_i \geq \sum_{i=1}^{t} d_{ji}(1 - z_j), && t \in N, j \in \Omega \\
& \sum_{j=1}^{m} z_j \leq k \\
& y_i \leq M_i x_i, && i \in N \\
& \theta \in \mathbb{R}_+^m, \mathbf{y} \in \mathbb{R}_+^n, \mathbf{x} \in \mathbb{B}^n, \mathbf{z} \in \mathbb{B}^m,
\end{aligned}$$

where we relax the constraint (3.3e) which captures the second-stage cost of each scenario. Note that the first class of proposed inequalities (3.6) is valid for the master problem. However, since the second class of valid inequalities (C.1) involves the stock variables, it cannot be directly applied to the master problem. For each $j \in \Omega$, the subproblem is stated as:

$$\begin{aligned}
\theta_j = \ & \min \mathbf{h}^\top \mathbf{s}_j \\
& s_{ji} \geq \sum_{i=1}^{t} (y_i - d_{ji}), \quad t \in N \qquad\qquad (\gamma_{ji}) \\
& \mathbf{s} \in \mathbb{R}^n_+,
\end{aligned}$$

where γ_{ji} is the dual variable associated with i-th time period of j-th scenario. Next, the corresponding dual variable for j-th scenario is stated as:

$$\theta_j \geq \ \max \sum_{i=1}^{n} \Big(\sum_{t=1}^{i} (y_t - d_{jt}) \Big) \gamma_{ji} \tag{D.1a}$$

$$\gamma_{ji} \leq h_i, \qquad i \in N, \tag{D.1b}$$

$$\gamma_j \in \mathbb{R}^n_+. \tag{D.1c}$$

Note that according to [18], we can apply the second class of valid inequalities (C.1) to the subproblems, to further strengthen the quality of the Benders optimality cuts added to the master problem. However, this implementation did not lead to improvements in solution time for our test instances, hence we do not report experiments with this version of Benders in our computational study in Section 3.5.

Given a first stage solution $(\hat{y}, \hat{\theta})$, instead of solving the dual problem (D.1) as a linear problem, we can take advantage of the special structure of (D.1) and generate Benders optimality cuts in $O(n)$ time: for each $i \in N$, if the term $\sum_{t=1}^{i}(y_t - d_{jt}) < 0$, then $\gamma_{ji} = 0$, because of the nonnegativity of h_i, for all $i \in N$. Otherwise, $\gamma_{ji} = h_i$. Let γ^*_j be the optimal dual solution for scenario j, if $\hat{\theta}_j < \sum_{i=1}^{n} \Big(\sum_{t=1}^{i}(y_t - d_{jt}) \Big) \gamma^*_{ji}$, then we add the following optimality cut to the master problem:

$$\theta_j \geq \sum_{i=1}^{n} \Big(\sum_{t=1}^{i} (y_t - d_{jt}) \Big) \gamma^*_{ji},$$

to cut off the suboptimal solution.

However, although we can solve the subproblem in $O(n)$ time, there are an exponential number of possible Benders optimality cuts for each scenario. As shown in our computational study in Section 3.5, as the number of time periods (n) grows, the Benders decomposition algorithm becomes ineffective.

Appendix E: Convex Hull of Example 5.1

Recall that $m = 3$, $l_p = l_d = 0$, $u_d = 10$, $\mathbf{w} = (8, 6, 10)$ and $\mathbf{v} = (3, 4, 2)$. The convex hull is given by inequalities (4.7)

$$2y_p + z_1 + z_2 + 12z_3 \geq 14 \qquad (\Pi = \{2 \to 1 \to 3\})$$

$$2y_p + 11z_1 + z_2 + 2z_3 \geq 14 \qquad (\Pi = \{3 \to 1 \to 2\})$$

$$2y_p + 2z_2 + 12z_3 \geq 14 \qquad (\Pi = \{2 \to 3\})$$

$$2y_p + 2z_1 + 10z_2 + 2z_3 \geq 14 \qquad (\Pi = \{1 \to 3 \to 2\})$$

$$2y_p + 10z_2 + 4z_3 \geq 14 \qquad (\Pi = \{3 \to 2\}),$$

inequalities (4.5)

$$y_p + y_d + 10z_3 \geq 10 \qquad (S = \{3\})$$

$$y_p + y_d + 8z_1 + 2z_3 \geq 10 \qquad (S = \{1, 3\})$$

$$y_p + y_d + 2z_1 + 6z_2 + 2z_3 \geq 10 \qquad (S = \{1, 2, 3\})$$

$$y_p + y_d + 6z_2 + 4z_3 \geq 10 \qquad (S = \{2, 3\}),$$

and inequalities (4.6)

$$y_p - y_d + 14z_2 \geq 4 \qquad (T = \{2\})$$

$$y_p - y_d + 13z_1 + z_2 \geq 4 \qquad (T = \{1, 2\})$$

$$y_p - y_d + z_1 + z_2 + 12z_3 \geq 4 \qquad (T = \{1, 2, 3\})$$

$$y_p - y_d + 2z_2 + 12z_3 \geq 4 \qquad (T = \{2, 3\}),$$

and the variable bounds

$$10 \geq y_d \geq 0$$

$$1 \geq z_1 \geq 0$$

$$1 \geq z_2 \geq 0$$

$$1 \geq z_3 \geq 0.$$

Appendix F: Extension for the intersection of multiple mixing sets

In this Chapter, we show that the proposed valid inequalities can be applied to strengthen the linear programming relaxation of the intersection of multiple mixing sets. To simplify the exposition, we only study the intersection of two mixing sets. If we consider the following problem

$$\mathbf{a}_1^\top \mathbf{x} + b_1 y + M_j^1 z_j \geq r_j^1, \qquad \forall j \in \Omega \tag{F.1}$$

$$\mathbf{a}_2^\top \mathbf{x} - b_2 y + M_j^2 z_j \geq r_j^2, \qquad \forall j \in \Omega, \tag{F.2}$$

where $\mathbf{x} \in \mathbb{R}^n$ and y are decision variables, and $\mathbf{a}_1$ and $\mathbf{a}_2$ are arbitrary n-dimensional vectors. In addition, $b_1 > 0$ and $b_2 > 0$ are coefficients of y and r_j^1 and r_j^2 are right-hand side parameters for the first and second sets of inequalities for scenario $j \in \Omega$, respectively. We assume that $\mathbf{a}_1 \geq 0$, and $\mathbf{a}_2 \geq 0$. In addition, to simplify the exposition and without loss of generality, we assume that $\mathbf{x} \geq 0$, and $u \geq y \geq 0$. Hence, we can set $M_j^1 = r_j^1$, and $M_j^2 = r_j^2 + b_2 u$. It can be seen that problem (4.4) is a special case of the problem stated above. Next, we will show that the proposed inequality (4.7) is valid for this formulation.

Let $\tau \in [m]$, and Π' be a *sequence* of τ scenarios given by $\pi'_1 \to \pi'_2 \to \cdots \to \pi'_\tau$, where $\pi_j \in \Omega$, for all $j \in [\tau]$. In addition, let

$$\bar{r}^1_{\pi'_j} = \begin{cases} \max_{j+1 \le \ell \le \tau}\{r^1_{\pi'_\ell}\}, & \text{if } j \in [\tau-1] \cup \{0\}, \\ 0, & \text{if} \quad j = \tau, \end{cases}$$

and

$$\bar{r}^2_{\pi'_j} = \begin{cases} \max_{j+1 \le \ell \le \tau}\{r^2_{\pi'_\ell}\}, & \text{if } j \in [\tau-1] \cup \{0\}, \\ 0, & \text{if} \quad j = \tau. \end{cases}$$

Proposition 37. *The following class of inequalities for a given sequence of scenarios* $\Pi := \{\pi'_1 \to \pi'_2 \to \cdots \to \pi'_\tau\}$*:*

$$(b_2\mathbf{a}_1 + b_1\mathbf{a}_2)^\top \mathbf{x} + \sum_{j=1}^{\tau} \Big(b_2(r^1_{\pi'_j} - \bar{r}^1_{\pi'_j})_+ + b_1(r^2_{\pi'_j} - \bar{r}^2_{\pi'_j})_+\Big) z_{\pi_j} \ge b_2\bar{r}^1_{\pi'_0} + b_1\bar{r}^2_{\pi'_0}, \quad \text{(F.3)}$$

where $r^1_{\pi'_0} = \max_{1\le \ell \le \tau}\{r^1_{\pi'_\ell}\}$*, and* $r^2_{\pi'_0} = \max_{1\le \ell \le \tau}\{r^2_{\pi'_\ell}\}$*, is valid.*

Proof. We scale the set of inequalities (F.1) by b_2, and the set of inequalities (F.2) by b_1. Letting $y_d = b_1 b_2 y$, and $w_j = b_2 r^1_j, v_j = b_1 r^2_j, j \in \Omega$, we obtain a structure similar to (4.4a)-(4.4b). The rest of the proof follows from Proposition 17. □

Example 1. *Consider the following example:*

$$2x + 3y + M^1_j z_j \ge r^1_j,$$

$$x - 2y + M^2_j z_j \ge r^2_j,$$

where $\Omega = \{1, 2, 3\}$, $\mathbf{r}^1 = \{4, 5, 8\}$, and $\mathbf{r}^2 = \{3, 2, 1\}$. In addition, let $u = 10$. The following inequality (F.3)

$$7x + 6z_1 + 19z_3 \ge 25,$$

where $\Pi = \{1 \to 3\}$ is valid, and facet-defining for this example.

Bibliography

[1] Abdi, A., Fukasawa, R.: On the mixing set with a knapsack constraint. Mathematical Programming **157**(1), 191–217 (2016)

[2] Ahmed, S., King, A., Parija, G.: A multi-stage stochastic integer programming approach for capacity expansion under uncertainty. Journal of Global Optimization **26**, 3–24 (2003)

[3] Artzner, P., Delbaen, F., Eber, J., Heath, D.: Coherent measures of risk. Mathematical Finance **9**(3), 203–228 (1999)

[4] Atamtürk, A., Nemhauser, G.L., Savelsbergh, M.W.P.: The mixed vertex packing problem. Math. Program. **89**, 35–53 (2000)

[5] Balas, E.: Disjunctive programming: Properties of the convex hull of feasible points. Discrete Applied Mathematics, **89(1-3)**, 3–44 (1998)

[6] Barany, I., van Roy, T., Wolsey, L.A.: Uncapacitated lot-sizing: The convex hull of solutions. Mathematical Programming Study **22**, 43–43 (1984)

[7] Ben Abdelaziz, F.: Solution approaches for the multiobjective stochastic programming. European Journal of Operational Research **216**(1), 1–16 (2012)

[8] Ben Abdelaziz, F., Lang, P., Nadeau, R.: Distributional efficiency in multiobjective stochastic linear programming. European Journal of Operational Research **85**(2), 399 – 415 (1995)

[9] Ben-Tal, A., Ghaoui, L., Nemirovski, A.: Robust Optimization. Princeton University Press, Princeton, NJ (2009)

[10] Ben-Tal, A., Nemirovski, A.: Robust optimization–methodology and applications. Mathematical Programming **92**(3), 453–480 (2002)

[11] Benders, J.: Partitioning procedures for solving mixed-variables programming problems **4**(1), 238–252 (1962)

[12] Beraldi, P., Ruszczyński, A.: A branch and bound method for stochastic integer programs under probabilistic constraints. Optim. Methods Softw. **17**, 359–382 (2002)

[13] Beraldi, P., Ruszczyński, A.: The probabilistic set covering problem. Oper. Res. **50**(6), 956–967 (2002)

[14] Bertsimas, D., Brown, D.: Constructing uncertainty sets for robust linear optimization. Oper. Res. **57**(6), 1483–1495 (2009)

[15] Bertsimas, D., Brown, D.B., Caramanis, C.: Theory and applications of robust optimization. SIAM Review **53**(3), 464–501 (2011)

[16] Birge, J., Louveaux, F.: Introduction to stochastic programming. Springer, New York (1997)

[17] Birge, J.R., Louveaux, F.V.: A multicut algorithm for two-stage stochastic linear programs. European Journal of Operational Research **34**(3), 384–392 (1988)

[18] Bodur, M., Dash, S., Günlük, O., Luedtke, J.: Strengthened Benders cuts for stochastic integer programs with continuous recourse (2014). `http://www.optimization-online.org/DB_HTML/2014/03/4263.html`

[19] Borcherding, K., Eppel, T., von Winterfeldt, D.: Comparison of weighting judgments in multiattribute utility measurement. Management Science **37**(12), 1603–1619 (1991)

[20] Burgert, C., Rüschendorf, L.: Consistent risk measures for portfolio vectors. Insurance: Mathematics and Economics **38**(2), 289–297 (2006)

[21] Calafiore, G., Campi, M.: Uncertain convex programs: randomized solutions and confidence levels. Math. Program. **102**, 25–46 (2005)

[22] Calafiore, G., Campi, M.: The scenario approach to robust control design. IEEE Trans. Automat. Control **51**, 742–753 (2006)

[23] Campi, M., Garatti, S.: A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality. J. Optim. Theory Appl. **148**, 257–280 (2011)

[24] Charnes, A., Cooper, W.: Deterministic equivalents for optimizing and satisficing under chance constraints. Oper. Res. **11**, 18–39 (1963)

[25] Charnes, A., Cooper, W., Symonds, G.: Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. Manage. Sci. **4**, 235–263 (1958)

[26] Dentcheva, D., Prékopa, A., Ruszczyński, A.: Concavity and efficient points of discrete distributions in probabilistic programming. Math. Program. **89**, 55–77 (2000)

[27] Dentcheva, D., Ruszczyński, A.: Portfolio optimization with stochastic dominance constraints. Journal of Banking and Finance **30**(2), 433–451 (2006)

[28] Dentcheva, D., Ruszczyński, A.: Optimization with multivariate stochastic dominance constraints. Mathematical Programming **117**, 111–127 (2009)

[29] Dentcheva, D., Wolfhagen, E.: Optimization with multivariate dominance constraints. In: Deodatis, G., Ellingwood, B., Frangopol, D. (eds.) Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures. CRC Press LLC (2013)

[30] Di Summa, M., Wolsey, L.A.: Lot-sizing on a tree. Oper. Res. Lett. **36**(1), 7 – 13 (2008)

[31] Ehrgott, M.: Multicriteria Optimization. Springer, Berlin (2005)

[32] Ehrgott, M., Ide, J., Schöbel, A.: Minmax robustness for multi-objective optimization problems. European Journal of Operational Research **239**(1), 17–31 (2014)

[33] Ekeland, I., Schachermayer, W.: Law invariant risk measures on $L^\infty(\mathbb{R}^d)$. Statistics and Risk Modeling with Applications in Finance and Insurance **28**(3), 195–225 (2011)

[34] Federgruen, A., Tzur, M.: A simple forward algorithm to solve general dynamic lot sizing models with n periods in $\mathrm{O}(n\log(n))$ or $\mathrm{O}(n)$ time. Manage. Sci. **37**(8), 909–925 (1991)

[35] Gade, D., Küçükyavuz, S.: Formulations for dynamic lot sizing with service levels. Naval Research Logistics **60**(2), 87–101 (2013)

[36] Guan, Y., Ahmed, S., Miller, A.J., Nemhauser, G.L.: On formulations of the stochastic uncapacitated lot-sizing problem. Oper. Res. Lett. **34**(3), 241 – 250 (2006)

[37] Guan, Y., Ahmed, S., Nemhauser, G., Miller, A.: A branch-and-cut algorithm for the stochastic uncapacitated lot-sizing problem. Math. Program. **105**, 55–84 (2006)

[38] Günlük, O., Pochet, Y.: Mixing mixed-integer inequalities. Math. Program. **90**, 429–457 (2001)

[39] Gupte, A., Ahmed, S., Dey, S., Cheon, M.: Relaxations and discretizations for the pooling problem. Journal of Global Optimization pp. 1–39 (2016). Article in Advance. DOI: 10.1007/s10898-016-0434-4

[40] Gutjahr, W.J., Pichler, A.: Stochastic multi-objective optimization: a survey on non-scalarizing methods. Annals of Operations Research **236**(2), 475–499 (2016)

[41] Hamel, A.H., Rudloff, B., Yankova, M.: Set-valued average value at risk and its computation. Mathematics and Financial Economics **7**(2), 229–246 (2013)

[42] Homem-de-Mello, T., Mehrotra, S.: A cutting surface method for uncertain linear programs with linear stochastic dominance constraints. SIAM Journal on Optimization **20**(3), 1250–1273 (2009)

[43] Hu, J., Homem-de-Mello, T., Mehrotra, S.: Risk-adjusted budget allocation models with application in homeland security. IIE Transactions **43**(12), 819–839 (2011)

[44] Hu, J., Mehrotra, S.: Robust and stochastically weighted multiobjective optimization models and reformulations. Operations Research **60**(4), 936–953 (2012)

[45] Hu, J., Homem-de Mello, T., Mehrotra, S.: Sample average approximation of stochastic dominance constrained programs. Mathematical Programming **133**(1-2), 171–201 (2012)

[46] Huang, K., Küçükyavuz, S.: On stochastic lot-sizing problems with random lead times. Oper. Res. Lett. **36**(3), 303–308 (2008)

[47] Iancu, D.A., Trichakis, N.: Pareto efficiency in robust optimization. Management Science **60**(1), 130–147 (2014)

[48] Jiang, R., Guan, Y.: An O(n^2) time algorithm for the stochastic uncapacitated lot-sizing problem with random lead times. Oper. Res. Lett. **39**, 74–77 (2011)

[49] Jouini, E., Meddeb, M., Touzi, N.: Vector-valued coherent risk measures. Finance and Stochastics **8**(4), 531–552 (2004)

[50] Köksalan, M., Şakar, Tuncer, C.: An interactive approach to stochastic programming-based portfolio optimization. Annals of Operations Research pp. 1–20 (2014). Article in Advance. DOI:10.1007/s10479-014-1719-y

[51] Krarup, K., Bilde, O.: Plant location, set covering and economic lot-sizes: An $O(mn)$ algorithm for structured problems. In: Collatz, L. (ed.) Optimierung bei Graphentheoretischen und Ganzzahligen Probleme. Birkhauser Verlag (1977)

[52] Küçükyavuz, S.: On mixing sets arising in chance-constrained programming. Math. Program. **132**, 31–56 (2012)

[53] Küçükyavuz, S.: Mixed-integer optimization approaches for deterministic and stochastic inventory management. In: INFORMS Tutorials in Operations Research, vol. 6 (2014)

[54] Küçükyavuz, S., Pochet, Y.: Uncapacitated lot-sizing with backlogging: The convex hull. Math. Program. **118**, 151–175 (2009)

[55] Küçükyavuz, S., Noyan, N.: Cut generation for optimization problems with multivariate risk constraints. Mathematical Programming **159**(1), 165–199 (2016)

[56] Kusuoka, S.: On law invariant coherent risk measures. Advances in Mathematical Economics **3**, 83–95 (2001)

[57] Lehmann, E.: Ordered families of distributions. Annals of Mathematical Statistics **26**(3), 399–419 (1955)

[58] Lejeune, M.A., Ruszczyński, A.: An efficient trajectory method for probabilistic inventory-production-distribution problems. Operations Research **55**(2), 378–394 (2007)

[59] Levy, H.: Stochastic dominance and expected utility: Survey and analysis. Management Science **38**(4), 555–593 (1992)

[60] Liu, X., Kılınç-Karzan, F., Küçükyavuz, S.: A polyhedral study of two-sided chance-constrained programs with finite probability space. (2016)

[61] Liu, X., Küçükyavuz, S.: A polyhedral study of the static probabilistic lot-sizing problem. Optimization Online (2016). URL `http://www.optimization-online.org/DB_HTML/2016/03/5354.html`

[62] Liu, X., Küçükyavuz, S., Noyan, N.: Robust multicriteria risk-averse stochastic programming models. Optimization Online (2016). URL `http://www.optimization-online.org/DB_HTML/2015/10/5157.html`

[63] Liu, X., Küçükyavuz, S., Luedtke, J.: Decomposition algorithms for two-stage chance-constrained programs. Mathematical Programming **157**(1), 219–243 (2016)

[64] Lubin, M., Bienstock, D., J.P.Vielma: Two-sided linear chance constraints and extensions. Optimization Online (2016). URL `http://www.optimization-online.org/DB_FILE/2016/02/5345.pdf`

[65] Luedtke, J.: A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. Math. Program. **146**, 219–244 (2014)

[66] Luedtke, J., Ahmed, S.: A sample approximation approach for optimization with probabilistic constraints. SIAM J. Optim. **19**, 674–699 (2008)

[67] Luedtke, J., Ahmed, S., Nemhauser, G.: An integer programming approach for linear programs with probabilistic constraints. Math. Program. **12**, 247–272 (2010)

[68] Mann, H., Whitney, D.: On a test of whether one of two random variables is stochastically larger than the other. Annals of Mathematical Statistics **18**, 50–60 (1947)

[69] McCormick, G.: Computability of global solutions to factorable nonconvex programs: Part I – convex underestimating problems. Mathematical Programming **10**(1), 147–175 (1976)

[70] Miller, B.L., Wagner, H.M.: Chance constrained programming with joint constraints. Oper. Res. **13**(6), 930–965 (1965)

[71] Miller, N., Ruszczyński, A.: Risk-averse two-stage stochastic linear programming: Modeling and decomposition. Operations Research **59**(1), 125–132 (2011)

[72] Müller, A., Stoyan, D.: Comparison Methods for Stochastic Models and Risks. John Wiley & Sons, Chichester (2002)

[73] Nemirovski, A., Shapiro, A.: Scenario approximation of chance constraints. In: Calafiore, G., Dabbene, F. (eds.) Probabilistic and Randomized Methods for Design Under Uncertainty, pp. 3–48. Springer, London (2005)

[74] Noyan, N.: Risk-averse two-stage stochastic programming with an application to disaster management. Computers and Operations Research **39**(3), 541–559 (2012)

[75] Noyan, N., Rudolf, G.: Optimization with multivariate conditional value-at-risk-constraints. Operations Research **61**(4), 990–1013 (2013)

[76] Ogryczak, W.: On robust solutions to multi-objective linear programs. In: Trzaskalik, T., Wachowicz, T. (eds.) Multiple Criteria Decision Making '09, pp. 197–212 (2010)

[77] Ogryczak, W., Ruszczyński, A.: On consistency of stochastic dominance and mean–semideviation models. Mathematical Programming **89**(2), 217–232 (2001)

[78] Pflug, G.C., Römisch, W.: Modelling, managing and measuring risk. World Scientific publishing, Singapore (2007)

[79] Pochet, Y., Wolsey, L.A.: Lot-size models with backlogging: Strong reformulations and cutting planes. Math. Program. **40**, 317–335 (1988)

[80] Prékopa, A.: Contributions to the theory of stochastic programming. Math. Program. **4**, 202–221 (1973)

[81] Prékopa, A.: Dual method for the solution of a one-stage stochastic programming problem with random RHS obeying a discrete probability distribution. ZOR - Methods and Models of Operations Research **34**, 441–461 (1990)

[82] Rockafellar, R.: Coherent approaches to risk in optimization under uncertainty. In: Tutorials in Operations Research, pp. 38–61. INFORMS (2007)

[83] Rockafellar, R., Uryasev, S.: Optimization of conditional value-at-risk. The Journal of Risk **2**(3), 21–41 (2000)

[84] Rockafellar, R., Uryasev, S.: Conditional value-at-risk for general loss distributions. Journal of Banking and Finance **26**(7), 1443–1471 (2002)

[85] Rüschendorf, L.: Risk measures for portfolio vectors. In: Mathematical Risk Analysis, Springer Series in Operations Research and Financial Engineering, pp. 167–188. Springer Berlin Heidelberg (2013)

[86] Ruszczyński, A.: Probabilistic programming with discrete distributions and precedence constrained knapsack polyhedra. Mathematical Programming **93**(2), 195–215 (2002)

[87] Saaty, T.: Decision Making for Leaders; the Analytical Hierarchy Process for Decisions in a Complex World. RWS Publications (2000)

[88] Saxena, A., Goyal, V., Lejeune, M.: MIP reformulations of the probabilistic set covering problem. Math. Program. **121**, 1–31 (2009)

[89] Schoemaker, P.J.H., Waid, C.C.: An experimental comparison of different approaches to determining weights in additive utility models. Management Science **28**(2), 182–196 (1982)

[90] Sen, S.: Relaxations for probabilistically constrained programs with discrete random variables. Oper. Res. Lett. **11**, 81–86 (1992)

[91] Shaked, M., Shanthikumar, J.G.: Stochastic Orders and their Applications. Associated Press, Boston (1994)

[92] Sherali, H.D., Adams, W.P.: A hierarchy of relaxations and convex hull representations for mixed-integer zero-one programming problems. Discrete Appl. Math. **52**(1), 83–106 (1994)

[93] Sherali, H.D., Adams, W.P., Driscoll, P.J.: Exploiting special structures in constructing a hierarchy of relaxations for 0-1 mixed integer problems. Operations Research **46**(3), 396–405 (1998)

[94] Song, Y., Luedtke, J., Küçükyavuz, S.: Chance-constrained binary packing problems. To appear in INFORMS Journal on Computing (2013)

[95] Song, Y., Luedtke, J.R.: Branch-and-cut approaches for chance-constrained formulations of reliable network design problems. Math. Program. Comput. **5**(4), 397–432 (2013)

[96] Steuer, R.E.: Multiple Criteria Optimization: Theory, Computation, and Application. John Wiley & Sons, New York (1986)

[97] Takriti, S., Ahmed, S.: On robust optimization of two-stage systems. Math. Program. **99**, 109–126 (2004)

[98] Van Slyke, R., Wets, R.J.: L-shaped linear programs with applications to optimal control and stochastic programming. SIAM J. Appl. Math **17**, 638–663 (1969)

[99] von Winterfeldt, D., Edwards, W.: Decision Analysis and Behavioral Research. Cambridge University Press (1986)

[100] Wagelmans, A., van Hoesel, S., Kolen, A.: Economic lot sizing: An O($n \log n$) algorithm that runs in linear time in the Wagner-Whitin case. Oper. Res. **40**, S145–S156 (1992)

[101] Wagner, H.M., Whitin, T.: Dynamic version of the economic lot size model. Manage. Sci. **5**, 89–96 (1958)

[102] Wang, J., Shen, S.: Risk and energy consumption tradeoffs in cloud computing service via stochastic optimization models. In: Proceedings of the 5th IEEE/ACM International Conference on Utility and Cloud Computing (UCC 2012). Chicago, IL (2012)

[103] Wang, S., Guan, Y., Wang, J.: A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output. IEEE Transactions on Power Systems **27**, 206–215 (2012)

[104] Willis, H.H., Morral, A.R., Kelly, T.K., Medby, J.J.: Estimating terrorism risk. Tech. rep., The RAND Corporation, Santa Monica, CA (2005). URL `http://www.rand.org/pubs/monographs/MG388`

[105] Wozabal, D.: Robustifying convex risk measures for linear portfolios: A nonparametric approach. Operations Research **62**(6), 1302–1315 (2014)

[106] Zeng, B., An, Y., Kuznia, L.: Chance constrained mixed integer program: Bilinear and linear formulations, and Benders decomposition. Optimization Online (2014). `http://www.optimization-online.org/DB_FILE/2014/03/4295.pdf`

[107] Zhang, M., Küçükyavuz, S., Goel, S.: A branch-and-cut method for dynamic decision making under joint chance constraints. Manage. Sci. **60**(5), 1317–1333 (2014)

[108] Zhu, S., Fukushima, M.: Worst-case conditional value-at-risk with application to robust portfolio management. Operations Research **57**(5), 1155–1168 (2009)

[109] Zymler, S., Kuhn, D., Rustem, B.: Distributionally robust joint chance constraints with second-order moment information. Mathematical Programming **137**(1), 167–198 (2013)