Approximate and exact convexification approaches for solving two-stage mixed-integer recourse models

van der Laan, Niels

DOI:
10.33612/diss.198086429

Publication date:
2022

Approximate and exact convexification approaches for solving two-stage mixed-integer recourse models

Niels van der Laan

Published by: University of Groningen, Groningen, The Netherlands.

Printed by: Ipskamp Printing

university of groningen

Approximate and exact convexification approaches for solving two-stage mixed-integer recourse models

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Thursday 27 January 2022 at 16:15 hours

by

Niels van der Laan

born on 1 March 1995
in Assen

Supervisor
Prof. R.H. Teunter

Co-supervisor
Dr. W. Romeijnders

Assessment committee
Prof. K.J. Roodbergen
Prof. R. Schultz
Prof. M.J. Uetz

Contents

Chapter 1

Introduction

In practical problems in, e.g., healthcare, energy, and logistics, the best course of action typically depends on unknown or uncertain parameters. For example, in logistics, optimal delivery routes depend on travel times, which are unknown in advance; and in healthcare, workforce schedules need to account for patient influx, which is unpredictable. In such situations, scenario analysis may be of use to determine which parameters are crucial, and how they affect the optimal solution. In general, however, scenario analysis cannot be used to obtain solutions that perform well *overall*, e.g., on average. A powerful alternative for coping with parameter uncertainty in decision problems is offered by the stochastic programming (SP) modelling paradigm. Indeed, the practical relevance of SP is well-established [71], and confirmed by a wide range of applications in, e.g., telecommunications, environmental control, and finance, see also [32] and [86].

SP models typically involve integer restrictions on the decision variables for sensible modelling. In particular, such restrictions arise naturally when modelling indivisibilities, and they can be used to represent, e.g., routing, scheduling, and location decisions. The downside of the additional modelling flexibility provided by integer decision variables is that the resulting models are significantly harder to solve than their continuous counterparts. Indeed, integer decision variables introduce non-convexities in the model, and thus we cannot directly use the rich toolbox of convex optimization to develop efficient solution methods.

In this thesis, we consider so-called two-stage mixed-integer recourse (MIR) models, which are of the form

$$\min_{x} \quad c^\top x + \mathbb{E}_\omega[v(\omega, x)] \tag{1.1}$$

$$\begin{aligned} \text{s.t.} \quad & Ax \geq b \\ & x \in X, \end{aligned}$$

where $x \in \mathbb{R}^n$ is a vector of decision variables, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}^n$ are parameter vectors, the matrix A has appropriate dimension, ω is a random vector, and $v(\omega, x)$ is the optimal value of a mixed-integer program (MIP) that depends on ω and x, to be defined in Section 1.2. The function v represents the costs that result from adapting to parameter uncertainty. In this model, the decisions x are subject to linear constraints $Ax \geq b$, and possible non-negativity and integer restrictions imposed by $X \subset \mathbb{R}^n$. Furthermore, the corresponding costs $c^\top x + v(\omega, x)$ are exposed to parameter uncertainty: they depend on the random vector ω, which explicitly models a vector of uncertain parameters. To be specific, the realization of ω is not known at the time that we have to choose x, but we do assume that the probability distribution of ω is known at that time. Finally, the objective in (1.1) is to minimize the total expected costs.

1.1 Approach and main contribution

State-of-the-art solution methods for two-stage MIR models are typically restricted to special cases, or limited in their use from a computational point of view. Before elaborating on the challenges of solving two-stage MIR models and the current state-of-the-art in Sections 1.2 and 1.3, respectively, we describe the main contribution of this thesis on a high level. In particular, we develop computationally efficient algorithms for solving the two-stage MIR model (1.1). The key difficulty of solving (1.1) is that the *recourse function* $Q(x) = \mathbb{E}_\omega[v(\omega, x)]$, $x \in \mathbb{R}^n$, is in general non-convex. In order to overcome this difficulty, we approximate the original model by one that is easier to solve, and we guarantee the performance of the resulting approximate solution. To be specific, we derive convex approximations $\hat{Q}$ of Q, and we use them to obtain an approximate solution $\hat{x}$ by solving

$$\min_x \{c^\top x + \hat{Q}(x) : \ Ax \geq b, \ x \in X\}, \tag{1.2}$$

which is obtained from (1.1) by replacing $\mathbb{E}_\omega[v(\omega, x)]$ by $\hat{Q}(x)$. The advantage is that (1.2) features a convex objective function, and thus it is easier to solve compared to the original model (1.1). Of course, without further restrictions on $\hat{Q}$, we do not expect to obtain good solutions for (1.1) by solving the approximating

model (1.2). For example, if $\hat{Q}$ is not a close approximation of Q, then an optimal solution $\hat{x}$ of (1.2) may be far from optimal in (1.1).

A natural approach, therefore, is to use convex approximations $\hat{Q}$ that closely approximate Q. Indeed, if the approximation error $\sup_x |Q(x) - \hat{Q}(x)|$ is small, then $\hat{x}$ is expected to perform well or even near-optimal in (1.1). That is why we derive error bounds on the approximation error, which can be directly used to guarantee the performance of $\hat{x}$ in (1.1). In particular, we propose novel types of convex approximations and derive corresponding error bounds. A key feature of the convex approximations that we propose is that they are suitable for fast computations. As a result, the approximating model (1.2) lends itself well to effective decomposition strategies, enabling us to develop tractable algorithms for computing a corresponding optimal solution $\hat{x}$. Furthermore, we show that our error bounds carry over to the performance of $\hat{x}$ in the original model. In particular, $\hat{x}$ is near-optimal in (1.1) if the variability of the random parameters in the model is large.

Clearly, if Q is highly non-convex, then it is not possible to derive a convex approximation that closely approximates Q, and thus we cannot obtain provably good solutions by using this approach. We, therefore, also consider an alternative approach for guaranteeing the performance of $\hat{x}$. To be specific, in Chapter 5, we only use convex approximations $\hat{Q}$ of Q such that $\hat{Q}$ is a lower bound of Q. As a result, the approximating model (1.2) is a *convex relaxation* of the original model (1.1), which is easier to solve and provides a lower bound on the optimal value of (1.1). In fact, this lower bound is sharp if $\hat{Q}$ is the convex envelope of Q, defined as the greatest convex function that defines a lower bound of Q. Then, we say that the convex relaxation is *exact*, and moreover, every optimal solution of the original problem is also optimal in the convex relaxation. That is why we derive a sequence of convex lower bounding approximations of Q that converges uniformly to its convex envelope.

1.2 Two-stage mixed-integer recourse models

1.2.1 Definition and background

Two-stage MIR models are mathematically sound reformulations of deterministic MIPs, where some of the constraint parameters are random variables. To see this, consider the mixed-integer linear programming problem

$$\min_x \quad c^\top x$$

$$\begin{aligned} \text{s.t.} \quad & Ax \geq b \\ & Tx \geq h \\ & x \in X, \end{aligned}$$

where $x \in \mathbb{R}^n$ is a vector of decision variables subject to the linear constraints $Ax \geq b$ and $Tx \geq h$, and where the vector $c \in \mathbb{R}^n$ contains the corresponding unit costs. Furthermore, the right-hand side vectors b and h are real-valued, and the matrices A and T have appropriate dimensions. In deterministic optimization, all data elements are known at the time that we have to choose x. Here, however, we assume that the parameters T and h that define the constraints $Tx \geq h$ are uncertain at that time. The SP approach is to assume that T and h are governed by an underlying random vector ω, whose support Ω and probability distribution are assumed to be known, resulting in the *random goal constraints* $T(\omega)x \geq h(\omega)$.

Of course, the resulting optimization model is ill-defined. After all, the interpretation of the random goal constraints is unclear if ω is a random vector. In fact, two-stage MIR models arise if we assume that infeasibilities with respect to $T(\omega)x \geq h(\omega)$ can be repaired after observing ω, by choosing *recourse actions* $y \in \mathbb{R}^{n_2}$ such that

$$Wy \geq h(\omega) - T(\omega)x,$$

where the *recourse matrix* W models the repair technology. The resulting model possesses a two-stage structure: in the first stage, we choose x, and after observing ω, we choose the recourse actions y in such a way that we minimize the costs $q^\top y$ of repairing infeasibilities, by solving the *second-stage problem*

$$\begin{aligned} v(\omega, x) = \min_y \quad & q^\top y \\ \text{s.t.} \quad & Wy \geq h(\omega) - T(\omega)x \\ & y \in Y, \end{aligned}$$

where Y may impose non-negativity and integer restrictions on the recourse actions y, i.e., $Y = \mathbb{Z}_+^p \times \mathbb{R}_+^{n_2 - p}$; see also Example 1.1 for an illustrative example.

In the two-stage model, the total costs resulting from the decision x comprise the immediate costs $c^\top x$ and the random second-stage costs $v(\omega, x)$. It follows that

the problem of minimizing the total expected costs can be stated compactly as

$$\min_x \{c^\top x + Q(x) : \ Ax \geq b, \ x \in X\}, \tag{1.3}$$

where the recourse function Q, defined as

$$Q(x) = \mathbb{E}_\omega \left[\min_y \left\{ q^\top y : \ Wy \geq h(\omega) - T(\omega)x, \ y \in \mathbb{Z}_+^p \times \mathbb{R}_+^{n_2 - p} \right\} \right], \quad x \in \mathbb{R}^n, \tag{1.4}$$

represents the expected second-stage costs. In Chapter 5, we derive results that also hold for a more general formulation, where the cost vector q and recourse matrix W are allowed to depend on ω and thus random. For the sake of exposition, however, we assume *fixed recourse* throughout this introduction, i.e., the cost vector q and recourse matrix W are deterministic. In Example 1.1 below, we consider a nurse scheduling problem which satisfies this fixed recourse assumption.

Example 1.1 (adapted from [40])**.** Consider the problem of scheduling nurses in a hospital, where the number of available nurses should at all times be sufficient to handle demand, which depends on patient volume. For practical reasons, however, the schedule has to be determined several weeks in advance, when the patient volume is still uncertain. We assume that the schedule covers a one-week period, and that nurses can be scheduled according to a number of shifts, e.g., day-, late-, or night-shifts, which cover, say, eight consecutive hours.

To model this problem, let $\mathcal{T}$ and $\mathcal{S}$ denote the set of hours of the week and the set of shifts, respectively, and let $u_{st} = 1$ if shift s includes hour t, and $u_{st} = 0$ otherwise, $s \in \mathcal{S}$, $t \in \mathcal{T}$. In addition, h_t denotes the demand for nurses in hour t, $t \in \mathcal{T}$. If x_s represents the number of nurses that is scheduled according to shift s, $s \in \mathcal{S}$, then the constraint that the number of available nurses should satisfy demand reads

$$\sum_{s \in \mathcal{S}} u_{st} x_s \geq h_t, \quad t \in \mathcal{T}.$$

The issue is, of course, that the schedule has to be determined well in advance, but h_t is not known with full certainty at that time.

Prior to the start of the planning period, however, an updated demand forecast becomes available, and the schedule is adjusted accordingly. For simplicity, we assume that the updated demand forecast is accurate, and thus we can model demand as a random vector with elements $h_t(\omega)$, $t \in \mathcal{T}$, whose realizations are

revealed after deciding the initial schedule. After learning the updated demand forecast, we have the option of scheduling additional nurses according to a set of shifts $\mathcal{S}'$. That is, the recourse actions, denoted by y_s, describe the number of nurses that we assign to shift s, $s \in \mathcal{S}'$. Let q_s denote the corresponding unit costs, and let $w_{st} = 1$ if shift s includes hour t, and $w_{st} = 0$ otherwise, $s \in \mathcal{S}'$, $t \in \mathcal{T}$. Then, the second-stage problem of adjusting the schedule can be represented as

$$\begin{aligned} v(\omega, x) = \min_{y} \; & \sum_{s \in \mathcal{S}'} q_s y_s \\ \text{s.t.} \; & \sum_{s \in \mathcal{S}'} w_{st} y_s \geq h_t(\omega) - \sum_{s \in \mathcal{S}} u_{st} x_s \\ & y_s \in \mathbb{Z}_+ \; \forall s \in \mathcal{S}', \end{aligned}$$

where the integer restrictions on y_s model the indivisibility of nurses. ◊

1.2.2 Structural properties

Structural properties of two-stage MIR models are crucial for designing corresponding solution methods that are computationally efficient. We review such properties, and to this end, we consider the value function v of the second-stage problem, defined as

$$v(\omega, x) = \min_{y} \Big\{ q^\top y : \; Wy \geq h(\omega) - T(\omega)x, \\ y \in \mathbb{Z}_+^p \times \mathbb{R}_+^{n_2 - p} \Big\}, \quad \omega \in \Omega, \; x \in \mathbb{R}^n. \quad (1.5)$$

The behaviour of this value function as a function of the right-hand side vector $h(\omega) - T(\omega)x$ has been studied extensively, see, e.g., [12] or [85]. Before we state these results, we first need assumptions on the recourse data to ensure that for every first-stage decision x, the second-stage costs $v(\omega, x)$ are finite with probability 1. To motivate this, note that if $v(\omega, x) = +\infty$ with positive probability, then the decision x may result in irreparable infeasibilities with respect to the random goal constraints, and thus should be considered infeasible. This situation is undesirable from a computational point of view, and thus we exclude it by assuming *complete recourse*, see Definition 1.1.

Definition 1.1. The recourse is complete if and only if for every $s \in \mathbb{R}^m$, there exists a $y \in Y$ such that $Wy \geq s$. Then, $v(\omega, x) < +\infty$ for every $\omega \in \Omega$ and $x \in \mathbb{R}^n$.

The other extreme where $v(\omega, x) = -\infty$ with positive probability results in an unbounded model if x is feasible, and thus we exclude it. To this end, we consider the linear programming (LP) dual of the second-stage problem (1.5), and we assume that its feasible region $\{\lambda \geq 0 : \ \lambda W \leq q\}$ is non-empty. Then, we say that the recourse is *sufficiently expensive*, see Definition 1.2.

Definition 1.2. The recourse is sufficiently expensive if there exists a dual multiplier $\lambda \geq 0$ such that $\lambda W \leq q$. Then, $v(\omega, x) > -\infty$ with probability 1 for every $\omega \in \Omega$ and $x \in \mathbb{R}^n$.

Throughout, we assume that the recourse is complete and sufficiently expensive, so that v is finite-valued. Then, finiteness of v carries over to $Q(x) = \mathbb{E}_\omega[v(\omega, x)]$ under the technical assumption that the random data $(h(\omega), T(\omega))$ satisfy the *weak covariance condition*, see Definition 1.3.

Definition 1.3. The random data $(h(\omega), T(\omega))$ satisfy the weak covariance condition if $\mathbb{E}_\omega||h(\omega)|| < \infty$ and $\mathbb{E}_\omega||T(\omega)|| < \infty$.

Theorem 1.1 (Schultz [65])**.** *Consider the mixed-integer recourse function Q, defined as*

$$Q(x) = \mathbb{E}_\omega[\min_y\{q^\top y : \ Wy \geq h(\omega) - T(\omega)x, \ y \in \mathbb{Z}_+^p \times \mathbb{R}_+^{n_2-p}\}], \quad x \in \mathbb{R}^n.$$

Assume that the recourse is complete and sufficiently expensive, the elements of W are rational, and the random data $(h(\omega), T(\omega))$ satisfy the weak covariance condition. Then,

(i) Q is finite-valued and lower semi-continuous on $\mathbb{R}^n$, and

(ii) if the elements of $h(\omega)$ and $T(\omega)$ follow continuous distributions, then Q is continuous.

The structural properties of Q in Theorem 1.1 guarantee that, under mild assumptions, the two-stage MIR model (1.3) is well-defined. Indeed, the lower semi-continuity of Q ensures that (1.3) is feasible, bounded, and admits an optimal solution, provided that the first-stage feasible region $\{x \in X : \ Ax \geq b\}$ is non-empty and compact, see, e.g., [7].

Typically, efficient solution methods for two-stage MIR models rely crucially on additional (stronger) properties of Q, such as convexity, which only hold in special cases. In particular, if the recourse is continuous, i.e., if no integer restrictions are imposed on the recourse actions, then such properties follow directly from standard linear programming (LP) theory. To be specific, assuming complete and sufficiently

expensive recourse, strong LP-duality implies that

$$v(\omega, x) = \max_{\lambda} \{\lambda^\top [h(\omega) - T(\omega)x] : \ \lambda W \leq q, \ \lambda \geq 0\},$$

and it follows that $v(\omega, x)$ is a *convex polyhedral* function of x. To see this, note that the dual feasible region $\Lambda := \{\lambda \geq 0 : \ \lambda W \leq q\}$ is polyhedral, and by assumption, non-empty and bounded. Hence, $v(\omega, x)$ is the pointwise maximum of finitely many affine functions:

$$v(\omega, x) = \max_{k=1,\ldots,K} \lambda_k^\top \{h(\omega) - T(\omega)x\},$$

where λ_k, $k = 1, \ldots, K$, are the extreme points of Λ. Moreover, if the random data satisfy the weak covariance condition, then $Q(x) = \mathbb{E}_\omega[v(\omega, x)]$ inherits these properties of $v(\omega, x)$, see Theorem 1.2.

Theorem 1.2 (Shapiro et al. [71])**.** *Consider the continuous recourse function Q, defined for $x \in \mathbb{R}^n$ as $Q(x) = \mathbb{E}_\omega[v(\omega, x)]$, where*

$$v(\omega, x) = \min_{y} \{q^\top y : \ Wy \geq h(\omega) - T(\omega)x, \ y \geq 0\}. \tag{1.6}$$

Assume that the recourse is complete and sufficiently expensive, and the random data $(h(\omega), T(\omega))$ satisfy the weak covariance condition.

(i) Q is a finite-valued, convex, continuous, and subdifferentiable function on $\mathbb{R}^n$,

(ii) a subgradient of Q at $\bar{x} \in \mathbb{R}^n$ is given by

$$-\mathbb{E}_\omega \left[\lambda_\omega^\top T(\omega)\right] \in \partial Q(\bar{x}),$$

where λ_ω is a vector of optimal dual multipliers of the second-stage problem (1.6) *with $x = \bar{x}$, and*

(iii) if ω follows a finite discrete distribution, then Q is a convex polyhedral function on $\mathbb{R}^n$.

The convexity of Q in Theorem 1.2 enables the use of techniques from convex optimization to efficiently solve continuous recourse models, see also Section 1.3.1. If integer restrictions are imposed on the recourse actions y, however, then convexity of Q is lost, see, e.g., [54], resulting in significant computational challenges.

From the perspective of computational complexity, however, the difficulties posed by integer recourse actions are dominated by those caused by the *curse of*

dimensionality. Indeed, even in the absence of integer restrictions on y, solving the MIR model (1.3) is an arduous task. For example, if the random parameters in the model follow continuous distributions, then the problem of evaluating $Q(x)$ for a single value of x is already #P-hard in general [38], as it requires computing a complicated multidimensional integral. This difficulty is not mitigated, however, if we assume instead that the random parameters follow independent discrete distributions [29]. The main issue is that the number of realizations of the random parameters, referred to as *scenarios*, is exponential in the input size, and thus a single evaluation of Q requires solving exponentially many subproblems.

This issue carries over to the computation of the optimal solution and value of the corresponding recourse problem. To see this, suppose that ω follows a finite discrete distribution with mass points ω_s and corresponding probabilities p_s, $s = 1, \dots, S$. Then, the two-stage MIR model (1.1) is equivalent to the following *large-scale deterministic equivalent* (LSDE) program,

$$\begin{aligned} \min \quad & c^\top x + \sum_s p_s q^\top y_s \\ \text{s.t.} \quad & Ax \geq b \\ & T(\omega_s)x + Wy_s \geq h(\omega_s) \quad \forall s = 1, \dots, S \\ & x \in X,\ y_s \in Y \quad \forall s = 1, \dots, S, \end{aligned}$$

where y_s represents a copy of the recourse actions y corresponding to the realization ω_s, $s = 1, \dots, S$. In general, off-the-shelf solvers are of limited use for solving the LSDE program, as its input size depends linearly on the number of scenarios, which is typically (exponentially) large. In practice, however, the extent of this difficulty depends largely on the presence of integer restrictions on the recourse actions y. It turns out, indeed, that the LSDE program permits effective decomposition if the recourse is continuous, by exploiting the special properties of Q in Theorem 1.2.

1.3 Solution methods

1.3.1 Decomposition

Decomposition methods use a divide-and-conquer strategy to efficiently solve two-stage MIR models, where the problem is decomposed into many subproblems that are much smaller, and thus easier to solve. For example, the dual decomposition approach in [24] introduces copies x_s of the first-stage variables x to the LSDE pro-

gram for each scenario, and dualizes the *non-anticipativity* restrictions, which enforce that $x_s = x_{s'}$ for $s \neq s'$, by penalizing violations in the objective function. The advantage is that the resulting relaxation decomposes into S subproblems, one for each scenario, which can be solved separately.

We, however, focus on Benders' decomposition [15], which is widely used to solve (two-stage) MIR models, see, e.g., [31, 72, 84] and surveys in [43, 68]. Here, we briefly outline the main idea. That is, we reformulate the MIR model (1.3) as

$$\min_{x,\theta}\{c^\top x + \theta : \; Ax \geq b, \; \theta \geq Q(x)\}, \tag{1.7}$$

and we approximate the constraint $\theta \geq Q(x)$ by *optimality cuts* of the form $\theta \geq \psi_t(x)$, where ψ_t is an affine function such that $\psi_t \leq Q$, $t \in \mathcal{T}$. The advantage of using affine optimality cuts is that the *master problem* (MP), defined as

$$\min_{x,\theta}\{c^\top x + \theta : \; Ax \geq b, \; \theta \geq \psi_t(x) \; \forall t \in \mathcal{T}\}, \tag{MP}$$

is a linear program and thus can be solved efficiently. Moreover, (MP) is a relaxation of the original problem (1.7), and thus, if an optimal solution $(\bar{x}, \bar{\theta})$ of (MP) is feasible in (1.7), i.e., if $\bar{\theta} \geq Q(\bar{x})$, then $(\bar{x}, \bar{\theta})$ is also optimal in (1.7). If, on the other hand, $\bar{\theta} < Q(\bar{x})$, then we add a constraint $\theta \geq \psi(x)$ to (MP), in order to cut away $(\bar{x}, \bar{\theta})$, and we resolve (MP).

A prime example of Benders' decomposition for MIR models is the L-shaped method by Van Slyke and Wets [84], which solves continuous recourse problems by exploiting the convexity of Q in Theorem 1.2. In particular, if $u \in \partial Q(\bar{x})$ is a subgradient of Q at $\bar{x}$, then an optimality cut for Q is given by

$$Q(x) \geq \psi(x) := Q(\bar{x}) + u^\top (x - \bar{x}) \quad \forall x \in \mathbb{R}^n,$$

which is tight for Q at $\bar{x}$, i.e., $\psi(\bar{x}) = Q(\bar{x})$. Moreover, we can efficiently obtain ψ by using the expression for a subgradient of Q in Theorem 1.2. Finally, if ω follows a finite discrete distribution, then finite convergence can be established under mild conditions [43]. Indeed, we only need a finite number of optimality cuts to completely describe Q, because Q is a convex polyhedral function.

The L-shaped method cannot be used to solve general two-stage MIR models, because convexity of Q is lost if integer restrictions are imposed on the recourse actions [54]. Typically, Benders' decomposition algorithms that solve general MIR models combine ideas from the L-shaped method and deterministic mixed-integer

programming. For example, Carøe and Tind [26] generalize the L-shaped method to general MIR models, by underestimating Q in terms of *non-affine* optimality cuts, which are computed by using generalized duality for MIPs. In particular, they recover Q in terms of so-called Gomory functions, a class of non-convex functions that involve rounding of the decision variables, see also [18]. As a result, however, the generalized L-shaped method is of limited practical use, because the resulting master problem is highly non-convex, and thus hard to solve.

That is why alternative Benders' decomposition algorithms for MIR models typically use affine optimality cuts to underestimate Q, similar to the original L-shaped method. The rationale is to preserve the convexity of the master problem, which enables computationally efficient Benders' decomposition. The downside is, however, that such solution methods are limited to special classes of MIR models. The reason is that Q is non-convex and thus cannot be described completely in terms of affine optimality cuts. For example, the Benders' decomposition algorithms in [8, 31, 44, 48, 49, 51, 53, 69, 70, 72, 89] solve two-stage MIR models with *binary* first-stage decisions and mixed-integer second-stage decisions, and the algorithm in [88] requires *pure integer* decisions in both stages for convergence. To date, indeed, there do not exist computationally efficient Benders' decomposition algorithms for two-stage MIR models that can handle general mixed-integer variables in both stages. We, therefore, propose to apply Benders' decomposition in a fundamentally different way by using novel convexification techniques.

1.3.2 Convexification

Our approach is to *convexify* the recourse function, so that we can use Benders' decomposition in a computationally efficient way. To be specific, we use a convex approximation $\hat{Q}$ of Q, and we solve the resulting approximating model:

$$\min_x \{c^\top x + \hat{Q}(x) : \ Ax \geq b, \ x \in X\}. \tag{1.8}$$

Our reasoning is that the convexity of $\hat{Q}$ ensures that Benders' decomposition can be used to efficiently solve (1.8), similar to the original L-shaped method. Moreover, if $\hat{Q}$ is the convex envelope of Q, then the approximating problem (1.8) defines an exact relaxation of the original two-stage MIR model (1.3). That is, their optimal values coincide, and under mild assumptions, the solution obtained by solving (1.8) is also optimal in (1.3). For this reason, we derive a family of affine optimality cuts that yield the convex envelope of Q, and we use these *scaled cuts* in a Benders'

decomposition that solves general MIR models.

Alternatively, we use a convex approximation $\hat{Q}$ that is not necessarily a lower bound of Q, but one that closely approximates Q. The justification is that the resulting approximate solution $\hat{x}$ is a good solution of the original model if the approximation error $\sup_x |Q(x) - \hat{Q}(x)|$ is small. Indeed, we are able to guarantee the quality of $\hat{x}$ by deriving *error bounds* on this approximation error. This idea is not new, and it was in fact proposed by Van der Vlerk [80] for the special case of simple integer recourse (SIR) models. Over the years, the convex approximations methodology received significant attention in the literature, see [41, 59, 60, 61, 62, 81, 82, 83] and the survey in [58], but this progress is mainly limited to special cases of MIR models. For example, see [41] for SIR models, and [59, 61, 81] for pure integer recourse models with a totally unimodular recourse matrix.

A notable exception is the shifted LP-relaxation in [60], which is developed for two-stage MIR models with mixed-integer decision variables in both stages. The practical use of the shifted LP-relaxation $\hat{Q}$ is limited, however, because the computations that are required in a corresponding Benders' decomposition algorithm cannot be carried out efficiently. In particular, evaluating $\hat{Q}(x)$ and a subgradient $u \in \partial\hat{Q}(x)$ for a single value of x is infeasible from a computational point of view. That is why we propose novel convex approximations for general two-stage MIR models, which do lend themselves to efficient Benders' decomposition.

1.4 Outline

In this thesis, we propose novel solution methods for two-stage MIR models that are inspired by Benders' decomposition. In particular, in Chapters 2-4, we derive new types of convex approximations and corresponding error bounds, and we use these results to construct *approximate* Benders' decomposition algorithms, whereas in Chapter 5, we use *exact* Benders' decomposition. In the remainder, we discuss these chapters in more detail.

In Chapter 2, we derive an error bound for the shifted LP-relaxation approximation of SIR models by Romeijnders et al. [61]. This error bound improves the original error bound by Romeijnders et al., which depends on the *total variation* of the pdf f of the underlying random variable ω. In order to achieve this, we also take into account information on the *higher-order derivatives* of f. To be specific, our improved error bound decreases as the total variation of these derivatives decreases. In fact, we derive a hierarchy of error bounds that becomes sharper as we

take into account more higher-order derivatives. For SIR models, these results directly imply sharper performance guarantees for the solution obtained by solving the shifted LP-relaxation.

In the subsequent chapters, we consider a much larger class of two-stage MIR models: we allow general mixed-integer decision variables in both stages. In particular, in Chapters 3 and 4, we operationalize the use of convex approximations for this class of models by deriving convex approximations that are suitable for efficient Benders' decomposition.

Our approach in Chapter 3 is to approximate the mixed-integer second-stage feasible regions by relaxing the integer restrictions on the recourse actions and adding *pseudo-valid* cutting planes. That is, we use cutting planes that are affine in the first-stage decision variables. The advantage is that the approximating model can be interpreted as a continuous recourse model and thus it can be solved directly by using the L-shaped method. In general, however, this approximating model is not exact, since, contrary to traditional cutting planes for MIPs, pseudo-valid cutting planes may cut away feasible solutions and are allowed to be overly conservative. That is why we guarantee the quality of the approximation by deriving an error bound which is similar to the error bound for the shifted LP-relaxation in [60], that is valid if the cutting planes are exact on a grid of first-stage solutions.

In Chapter 4, we develop a complementary approximation scheme that can be used if such a family of cutting planes is not available. In particular, we propose the so-called generalized α-approximations $\hat{Q}_\alpha$ for two-stage MIR models, whose definition depends on *Gomory relaxations* of the second-stage problems. Moreover, we derive a corresponding error bound which is similar to the one in Chapter 3. The main advantage of the generalized α-approximation over the shifted LP-relaxation in [60] is that it permits fast computations. Indeed, we construct a *loose Benders' decomposition algorithm*, which efficiently solves the corresponding approximating model. We refer to our decomposition algorithm as loose, because the optimality cuts that we use to underestimate $\hat{Q}_\alpha$ are in general not tight. Therefore, the resulting solution $\hat{x}$ is not necessarily optimal in the approximating problem. We are able to prove, nonetheless, that our error bound for the generalized α-approximations carries over to the performance of $\hat{x}$ in the original model.

In Chapters 3 and 4, we also conduct extensive numerical experiments to empirically assess the proposed convex approximations. Overall, they yield good candidate solutions, and in line with our performance guarantees, we obtain near-optimal solutions if the variability of the random parameters in the model is large. Their

performance is less good, however, if this variability is small. The reason is that the original MIR model is then highly non-convex and thus the approximation error of our convex approximations is large.

That is why, in Chapter 5, we propose a Benders' decomposition algorithm which solves general two-stage MIR models, even if the underlying recourse function Q is highly non-convex. In order to achieve this, we derive a new family of optimality cuts for Q. These *scaled cuts* can be applied recursively, similar to cutting plane techniques for deterministic MIPs. In fact, we prove that a recursive application of our scaled cuts yields the convex envelope of Q in the limit. That is, the scaled cuts define an *exact* relaxation of the original model, and thus they can be used to obtain the optimal solution, in general. Indeed, numerical experiments confirm the superior performance of Benders' decomposition with scaled cuts compared to traditional optimality cuts.

Finally, in Chapter 6, we conclude by summarizing our findings and discussing possible avenues for future research.

List of manuscripts

The chapters of this thesis are based on the following journal publications.

Chapter 2:

N. van der Laan, W. Romeijnders, and M. H. van der Vlerk. Higher-order total variation bounds for expectations of periodic functions and simple integer recourse approximations. *Computational Management Science*, 15(3-4):325–349, 2018.

Chapter 3:

W. Romeijnders and N. van der Laan. Pseudo-valid cutting planes for two-stage mixed-integer stochastic programs with right-hand-side uncertainty. *Operations Research*, 68(4):1199–1217, 2020.

Chapter 4:

N. van der Laan and W. Romeijnders. A loose Benders decomposition algorithm for approximating two-stage mixed-integer recourse models. *Mathematical Programming*, 2020. DOI: 10.1007/s10107-020-01559-1.

Chapter 5:

N. van der Laan and W. Romeijnders. A converging Benders' decomposition algorithm for two-stage mixed-integer recourse models. *Accepted for publication in Operations Research*, 2021. Available from Optimization Online.

Chapter 2

Higher-order total variation bounds for expectations of periodic functions and simple integer recourse approximations

We derive bounds on the expectation of a class of periodic functions using the total variations of higher-order derivatives of the underlying probability density function. These bounds are a strict improvement over those of Romeijnders et al. [61], and we use them to derive error bounds for convex approximations of simple integer recourse models. In fact, we obtain a hierarchy of error bounds that become tighter if the total variations of additional higher-order derivatives are taken into account. Moreover, each error bound decreases if these total variations become smaller. The improved bounds may be used to derive tighter error bounds for convex approximations of more general recourse models involving integer decision variables.

This chapter is based on the journal publication [79].

2.1 Introduction

Consider the two-stage recourse model with random right-hand side

$$\eta^* := \min_x \left\{ cx + Q(z) : Ax \geq b, z = Tx, x \in X \subset \mathbb{R}^{n_1}_+ \right\}, \tag{2.1}$$

where the recourse function Q is defined for the tender variables $z \in \mathbb{R}^m$ as

$$Q(z) := \mathbb{E}_\omega \left[v(\omega - z) \right],$$

and the value function v is defined for $s \in \mathbb{R}^m$ as

$$v(s) := \min_y \left\{ qy : Wy \geq s, y \in Y \subset \mathbb{R}^{n_2}_+ \right\}.$$

This model describes a two-stage decision problem under uncertainty. The uncertainty arises from the random vector ω of which the distribution is known. In the first stage, a decision x has to be made while the realization of ω is not yet available, whereas in the second stage, the realization of ω is known and recourse actions y can be taken to repair infeasibilities of the random constraints $Tx \geq \omega$. The model is called an integer recourse model if $Y = \mathbb{Z}^{n_2}_+$. If in addition $W = I_m$, then the model is referred to as a simple integer recourse (SIR) model. More general formulations of (2.1) exist, with uncertainty in the technology matrix T and cost parameters q, see, e.g., the textbooks [16, 71].

Throughout this chapter, we make the following assumptions, which guarantee that Q is finite everywhere.

(i) The recourse is complete: for all $s \in \mathbb{R}^m$, there exists a $y \in Y$ such that $Wy \geq s$, so that $v(s) < \infty$.

(ii) The recourse is sufficiently expensive; $v(s) > -\infty$ for all $s \in \mathbb{R}^m$.

(iii) $\mathbb{E}[|\omega_i|]$ is finite for all $i = 1, \ldots, m$.

Recourse models are highly relevant in practice, as demonstrated by numerous applications in problems where uncertainty plays a role. Areas of application include energy, telecommunication, production planning, and environmental control see, e.g., [32, 86]. Furthermore, integrality restrictions on the recourse actions arise naturally when modelling real-life situations, for example to capture on/off decisions or batch size restrictions.

Unfortunately, solving integer recourse problems is generally time-consuming and practically impossible, because the recourse function Q is in general non-convex (Rinnooy Kan and Stougie [54]). Traditional solution methods typically combine ideas from deterministic mixed-integer programming and stochastic continuous programming, see, e.g., [4, 24, 31, 49, 67, 44, 69, 88], and the survey papers [58, 66, 68].

However, in this chapter we focus on an alternative solution methodology introduced by van der Vlerk [80]. His approach is to approximate the non-convex recourse function Q by a convex approximation $\hat{Q}$, obtaining an approximating model for (2.1). The advantage is that the approximating model can be solved efficiently using known convex optimization techniques to obtain an approximate solution $(\hat{x}, \hat{z})$ for (2.1).

In the literature, convex approximations are typically derived by simultaneously modifying the underlying recourse structure (Y, q, W) and the distribution of the random vector ω. For example, Klein Haneveld et al. [41] propose the so-called α-approximations for SIR models, van der Vlerk [81] studies a class of convex approximations for the general integer case, and Romeijnders et al. [61] propose a convex approximation, the shifted LP-relaxation, for integer recourse problems with a totally unimodular (TU) recourse matrix W. The latter approximation is generalized to the general mixed-integer recourse case by Romeijnders et al. [60].

To guarantee the performance of the approximating solution $(\hat{x}, \hat{z})$ in the original model (1), Romeijnders et al. [59] show that for every approximation $\hat{Q}$,

$$c\hat{x} + Q(\hat{z}) - \eta^* \leq 2\,||\hat{Q} - Q||_\infty := 2 \sup_{z \in \mathbb{R}} |\hat{Q}(z) - Q(z)|.$$

That is, the absolute optimality gap of $(\hat{x}, \hat{z})$ is at most $2\|\hat{Q} - Q\|_\infty$. For this reason, bounds on $\|\hat{Q} - Q\|_\infty$ are used to guarantee the quality of the approximating solution $(\hat{x}, \hat{z})$. Such error bounds are derived by Klein Haneveld et al. [41] and by Romeijnders et al. [59] for various types of convex approximations. They are expressed in terms of the total variations of the marginal probability density functions (pdf) of the random right-hand side vector ω. For general mixed-integer and TU integer recourse models, Romeijnders et al. [60] and Romeijnders et al. [61], respectively, derive similar error bounds by making use of periodicity in the difference of the underlying value functions of the recourse function Q and its convex approximation $\hat{Q}$. In this way, bounds on $\|\hat{Q} - Q\|_\infty$ are obtained by deriving total variation bounds on the expectation of periodic functions. Since we also exploit this

relationship between expectations of periodic functions and the difference between $\hat{Q}$ and Q, we explain this relationship in more detail in Section 2.2.

Romeijnders et al. [61] use worst-case analysis to prove that their error bounds are tight for certain piecewise constant pdf. For other pdf, there may be a considerable difference between the error bound and the actual error as shown by numerical experiments on a fleet allocation and routing problem and on an investment problem in stochastic activity networks (Romeijnders et al. [62]). Motivated by these observations, we improve the erorr bounds of Romeijnders et al. [61] by using information on the higher-order derivatives of the underlying pdf of the random variables in the model. To be specific, we use that the total variation of the higher-order derivatives of the underlying pdf can be used to improve the error bound. The intuition behind our approach is that by imposing restrictions on the higher-order derivatives of the underlying pdf we are able to exclude the piecewise constant pdf, which have jump discontinuities.

To obtain these error bounds we improve the existing bounds on the expectation of periodic functions by Romeijnders et al. [61] using higher-order total variations, i.e., total variations of the higher-order derivatives of the underlying marginal pdfs. This is the main contribution of this chapter since these bounds may be used to improve error bounds of convex approximations for mixed-integer recourse models in general. We illustrate their potential by improving error bounds for the shifted LP-relaxation approximation of SIR models by Romeijnders et al. [61]. The improved error bounds decrease with the total variations of the underlying pdf and its higher-order derivatives.

The remainder of this chapter is organized as follows. Section 2.2 describes in more detail the relationship between error bounds for the shifted LP-relaxation of integer recourse models and bounds on the expectation of periodic functions. In Section 2.3, we improve bounds on the expectation of a class of periodically monotone functions, and in Section 2.4 we illustrate how these results can be applied in the setting of SIR models. In Section 2.5 we conclude and summarize our results.

2.2 Total variation error bounds

In this section, we describe parts of the procedure employed by Romeijnders et al. [61] to derive an error bound for the so-called *shifted LP-relaxation approximation* $\hat{Q}$ of Q for TU integer recourse models. The idea behind this approximation is to simultaneously relax the integrality restrictions in the model and to perturb the

random right-hand side vector ω.

Definition 2.1. The *shifted LP-relaxation* $\hat{Q}$ of the mixed-integer recourse function Q is defined as

$$\hat{Q}(z) := \mathbb{E}_\omega \left[\min_y \left\{ qy : Wy \geq \omega + \frac{1}{2} e_m - z, y \in \mathbb{R}_+^{n_2} \right\} \right], \quad z \in \mathbb{R}^m,$$

where e_m is the m-dimensional all-one vector.

The shifted LP-relaxation is a special case of the convex approximation for general two-stage mixed-integer recourse models by Romeijnders et al. [60]. Here, we focus on the one-dimensional SIR case. Partly this is for simplicity, but we are also inspired by the fact that Romeijnders et al. [61] derive an error bound for the TU integer recourse case using one-dimensional results. Setting $q = 1$ and $W = 1$, the integer recourse function Q reduces to

$$Q(z) = \mathbb{E}_\omega \left[\lceil \omega - z \rceil^+ \right], \quad z \in \mathbb{R},$$

where $\lceil \cdot \rceil$ denotes the round-up function and $\lceil s \rceil^+ := \max\{0, \lceil s \rceil\}$, $s \in \mathbb{R}$. The shifted LP-relaxation then becomes

$$\hat{Q}(z) = \mathbb{E}_\omega \left[(\omega - z + 1/2)^+ \right], \quad z \in \mathbb{R}.$$

Since $||\hat{Q} - Q||_\infty$ is the quantity of interest, consider

$$\sup_{z \in \mathbb{R}} \left| \hat{Q}(z) - Q(z) \right| = \sup_{z \in \mathbb{R}} \left| \mathbb{E} \left[\psi \left(\omega - z + 1/2 \right) \right] \right|, \tag{2.2}$$

with underlying difference function $\psi(t) := (t)^+ - \lceil t - 1/2 \rceil^+$, see Figure 2.1.

If we ignore the positive part operators and define

$$\varphi(t) := t - \lceil t - 1/2 \rceil, \tag{2.3}$$

then φ is a periodic function with period $p = 1$ and mean value $\nu = p^{-1} \int_0^p \varphi(t) dt = 0$. The function ψ, however, is only half-periodic, that is

$$\psi(t) = \begin{cases} 0, & t \leq 0, \\ \varphi(t), & t \geq 0. \end{cases} \tag{2.4}$$

Romeijnders et al. [61] use this property to derive error bounds for the shifted

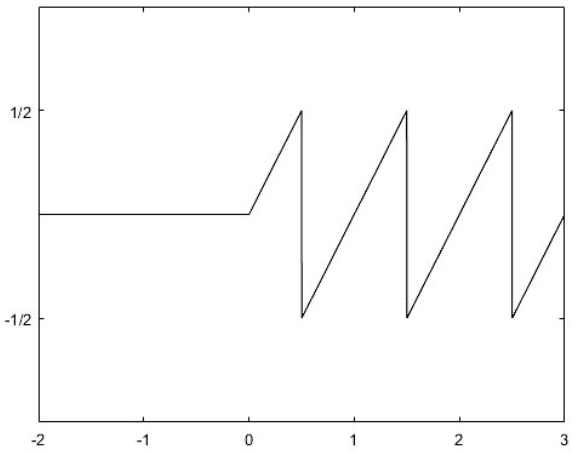


Figure 2.1. The difference function ψ.

LP-relaxation. Similarly, Romeijnders et al. [60] make use of asymptotic periodicity results for mixed-integer programs to find error error bounds for the general mixed-integer recourse case.

The error bounds are based on a *worst-case approach* with respect to the total variation $|\Delta|f$ of the underlying pdf f of ω, since the expectation in (2.2) is intractable in general. The following definition of total variation is taken directly from Romeijnders et al. [61].

Definition 2.2. Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function, and let $I \subset \mathbb{R}$ be an interval. Let $\Pi(I)$ denote the set of all finite ordered sets $P = \{x_1, \dots, x_{N+1}\}$ with $x_1 < \cdots < x_{N+1}$ in I. Then, the *total variation* of f on I, denoted $|\Delta|f(I)$, is defined as

$$|\Delta|f(I) = \sup_{P \in \Pi(I)} V_f(P),$$

where

$$V_f(P) = \sum_{i=1}^{N} |f(x_{i+1}) - f(x_i)|.$$

Write $|\Delta|f := |\Delta|f(\mathbb{R})$. A function f is said to be of bounded variation if $|\Delta|f$ is finite.

The worst-case approach adopted by Romeijnders et al. [61] is to derive a bound,

for any bounded periodic function φ and $B \in \mathbb{R}$ with $B > 0$, on

$$M(\varphi, B) := \sup_{f \in \mathcal{F}} \left\{ \mathbb{E}_f[\varphi(\omega)] : |\Delta| f \leq B \right\},$$

where $\mathcal{F}$ denotes the set of all continuous pdf of bounded variation. This bound is used to derive an error bound on $\|\hat{Q} - Q\|_\infty$ in (2.2), presented in Theorem 2.1 below. We will improve the bounds on the expectation of periodic and half-periodic functions, which directly leads to an improved error bound for SIR models.

Theorem 2.1. *Consider the simple integer recourse function Q defined as*

$$Q(z) = \mathbb{E}_\omega \left[\lceil \omega - z \rceil^+ \right], \quad z \in \mathbb{R},$$

and its shifted LP-relaxation approximation $\hat{Q}$, defined as

$$\hat{Q}(z) = \mathbb{E}_\omega \left[\left(\omega - z + \frac{1}{2} \right)^+ \right], \quad z \in \mathbb{R},$$

where ω is a continuous random variable with pdf f of bounded variation. Then

$$\sup_{z \in \mathbb{R}} \left| \hat{Q}(z) - Q(z) \right| \leq \frac{1}{2} h(|\Delta| f),$$

where

$$h(t) = \begin{cases} \frac{t}{8}, & 0 < t \leq 4, \\ 1 - \frac{2}{t}, & t \geq 4. \end{cases}$$

Proof. See Romeijnders et al. [61]. ■

Romeijnders et al. [61] show that there exist piecewise constant pdf for which the error bound in Theorem 1 is tight. For example, for $B \geq 4$, the pdf $\hat{f} : \mathbb{R} \to \mathbb{R}$ defined as

$$\hat{f}(x) = \begin{cases} \frac{B}{2}, & 0 < x \leq \frac{2}{B}, \\ 0, & \text{otherwise}, \end{cases}$$

satisfies $|\Delta|\hat{f} = B$ and

$$\sup_{z\in\mathbb{R}} \left|\hat{Q}(z) - Q(z)\right| = \frac{1}{2}h(|\Delta|\hat{f}).$$

Remark 2.1. Observe that $\hat{f}$ can be interpreted as a pdf with $|\Delta|\hat{f}' = +\infty$.

In the next sections we will assume that f is continuously differentiable and that its derivative f' is of bounded variation. In this way, we exclude piecewise constant densities such as $\hat{f}$. By using the higher-order total variations of f we are able to derive tighter bounds.

2.3 Improving bounds on the expectation of periodic functions

In this section, we improve the bounds on the expectation of periodic functions derived by Romeijnders et al. [61] for the class of point-symmetric periodic functions (see Definition 2.7). Next to that, we improve bounds on the expectation of *half-periodic* functions $\psi : \mathbb{R} \to \mathbb{R}$, which are of the form

$$\psi(x) = \begin{cases} 0, & x < 0, \\ \varphi(x), & x \geq 0, \end{cases} \tag{2.5}$$

where φ is a point-symmetric periodic function.

The organization of this section is as follows. In Section 2.3.1, we consider packed densities, a concept introduced by Romeijnders et al. [61] which we generalize to higher-order derivatives. Next, in Section 2.3.2, we introduce *point-symmetric periodic functions*, the class of periodic functions for which we are able to derive results. In Section 2.3.3 we use the concept of bound propagation to derive a hierarchy of total variation bounds. In Section 2.3.4, we present bounds on the expectation of periodic and half-periodic functions.

2.3.1 Higher-order derivatives of packed densities

The main contribution of this chapter is to generalize the results by Romeijnders et al. [61] to higher-order derivatives. The key insight here is that including information on the total variation and the maximum norm of higher-order derivatives of the underlying pdf leads to improved expectation bounds. We generalize

several definitions by Romeijnders et al. [61] to allow for these elements to be included in our analysis. They consider the set $\mathcal{F}$ containing all one-dimensional pdf of bounded variation. In Definition 2.3, we introduce the sets $\mathcal{F}_n$, $n \in \mathbb{N}$, which satisfy the relationship $\mathcal{F} \supset \mathcal{F}_0 \supset \mathcal{F}_1 \supset \dots$.

Definition 2.3. Let $\mathcal{F}_n$ denote the set of one-dimensional pdf f, such that the first n derivatives of f exist, are continuous, and are of bounded variation. Denote the k-th derivative of f by $f^{(k)}$ and write $f = f^{(0)}$.

Definition 2.4. For all bounded integrable functions $\varphi : \mathbb{R} \to \mathbb{R}$ and positive constants $B := (B_0, \dots, B_n)$ and $C := (C_0, \dots, C_n)$ define M_n as

$$M_n(\varphi, B, C) := \sup_{f \in \mathcal{F}_n} \Bigg\{ \Big| \mathbb{E}_f[\varphi(\omega)] \Big| : |\Delta| f^{(k)} \leq B_k, \\ ||f^{(k)}||_\infty \leq C_k, \text{ for } k = 0, \dots, n \Bigg\},$$

where $||\cdot||_\infty$ denotes the maximum norm defined as

$$||f||_\infty := \max_{t \in \mathbb{R}} |f(t)|.$$

Remark 2.2. In Definition 2.4, we could have suppressed the constants $C_0, \dots, C_n$, since an upper bound B_k on $|\Delta| f^{(k)}$ directly implies an upper bound $\frac{B_k}{2}$ on $||f^{(k)}||_\infty$. However, these bounds may be larger than C_k so that by including these constants we may obtain tighter bounds.

The main goal of this section is to derive a bound on $M_n(\varphi, B, C)$. This is useful, because such a bound can be used to derive performance guarantees for convex approximations of mixed-integer recourse functions. We derive such bounds for periodic and half-periodic functions φ. The key objects of our analysis are packed densities, introduced by Romeijnders et al. [61], which are functions defined on $[0, p]$ such that for either periodic of half-periodic functions ϕ with period p,

$$\mathbb{E}_f[\varphi(\omega)] = \int_0^p \varphi(x) f_p(x) dx.$$

For periodic functions φ, Romeijnders et al. [61] define this packed density as

$$f_p(x) := \sum_{k \in \mathbb{Z}} f(x + pk), \quad x \in [0, p].$$

Intuitively, $f_p(x)$ represents the sum of the densities corresponding to $\varphi(x)$ and

$\varphi(x+pk)$ for every $k \in \mathbb{Z} \setminus \{0\}$. For half-periodic functions φ, however, we require an alternative packed density, which we call the *half-packed density*. This packed density does not include the values of $f(x+pk)$ for $k < 0$, since φ is only half-periodic and thus $\varphi(x+pk) = 0$ for $k < 0$. Definition 2.5 below contains the definitions of both the original and the new packed density.

Definition 2.5. For all $f \in \mathcal{F}_0$ and $p \in \mathbb{R}$ with $p > 0$, define the *classical packed density* $f_p : [0,p] \to \mathbb{R}$ as

$$f_p(x) := \sum_{k\in\mathbb{Z}} f(x+pk), \quad x \in [0,p]$$

Further, define the *half-packed density* $\hat{f}_p : [0,p] \to \mathbb{R}$ as

$$\hat{f}_p(x) = \sum_{k=0}^{\infty} f(x+pk), \quad x \in [0,p].$$

Remark 2.3. Note that the half-packed density does not integrate to one in general, as opposed to the classical packed density. As a result, the half-packed density cannot be interpreted as a pdf.

Remark 2.4. We define the (higher-order) derivatives of the classical packed density and the half-packed density at the endpoints of the closed interval $[0,p]$ by their one-sided derivatives, provided that they exist.

Romeijnders et al. [61] show that the total variation of f can be used to bound the total variation of f_p on $[0,p]$. Interestingly, similar bounds can be derived for the total variation of half-packed densities $\hat{f}_p$ on $[0,p]$. Moreover, similar bounds also hold for higher-order derivatives $f^{(k)}, k = 1,\dots,n$, of f.

Lemma 2.1. *Let $n \in \mathbb{N}$ be given and let $f \in \mathcal{F}_n$. Consider its corresponding classical packed density f_p and half-packed density $\hat{f}_p$, as in Definition 2.5. Then, for $k = 0,\dots,n$,*

(i) $f_p^{(k)}$ *and* $\hat{f}_p^{(k)}$ *exist and are continuous on* $[0,p]$,

(ii) $f_p^{(k)}(0) = f_p^{(k)}(p)$ *and* $\hat{f}_p^{(k)}(0) = \hat{f}_p^{(k)}(p) + f^{(k)}(0)$, *and*

(iii) $|\Delta| f_p^{(k)}([0,p]) \leq |\Delta| f^{(k)}$ *and* $|\Delta| \hat{f}_p^{(k)}([0,p]) \leq |\Delta| f^{(k)} - |f^{(k)}(0)|$.

Moreover, for all bounded integrable periodic functions φ with period p, and half-periodic

functions ψ of the form

$$\psi(x) = \begin{cases} 0, & x < 0, \\ \varphi(x), & x \geq 0, \end{cases}$$

(iv) $\mathbb{E}_f[\varphi(\omega)] = \int_0^p \varphi(x) f_p(x) dx$ *and* $\mathbb{E}_f[\psi(\omega)] = \int_0^p \varphi(x) \hat{f}_p(x) dx$.

Proof. We refer to the proof of corresponding properties involving packed densities in Romeijnders et al. [61]. ■

Observe that the properties involving the classical packed density generalize readily to higher-order derivatives. With respect to the half-packed density, notice the additional terms $f^{(k)}(0)$ and $-|f^{(k)}(0)|$ in properties (ii) and (iii), respectively. These terms result from the fact that the half-packed density only sums over the non-negative integers.

In our analysis in the next sections, we will initially not make a distinction between the classical packed density and the half-packed density. This is possible because they belong to the broader class of packed densities, which we introduce in Definition 2.6.

Definition 2.6. For a given $n \in \mathbb{N}$ and $p \in \mathbb{R}$ with $p > 0$, let $g : [0, p] \to \mathbb{R}$ be an n times continuously differentiable function such that its first n derivatives are of bounded variation on the interval $[0, p]$. Assume $\gamma_k := g^{(k)}(p) - g^{(k)}(0)$ is bounded for $k = 0, \ldots, n$. Then g is referred to as a packed density with discontinuities $\gamma = (\gamma_0, \ldots, \gamma_n)$.

Indeed, from Lemma 2.1 it follows directly that both the classical and the half-packed density are packed densities with discontinuities $\gamma = (0, \ldots, 0)$ and $(f^{(0)}(0), \ldots, f^{(n)}(0))$, respectively. Considering property (iv) in Lemma 2.1, we are interested in the quantity

$$D(\varphi, g) := \int_0^p \varphi(x) g(x) dx, \tag{2.6}$$

for a periodic function φ and a packed density g of Definition 2.6. It equals the expectation of a periodic function if g is a classical packed density, in which case the discontinuities corresponding to g are $\gamma = (0, \ldots, 0)$. Furthermore, for φ defined as in (2.3), the quantity $D(\varphi, g)$ is equal to the approximation error of the shifted-LP relaxation of a simple integer recourse model if g is a half-packed density, $\hat{f}_p$ of Definition 5, with $\gamma = (f^{(0)}(0), \ldots, f^{(n)}(0))$.

2.3.2 Point-symmetric periodic functions

We restrict our attention to *point-symmetric periodic functions*, a concept we introduce in Definition 2.7, since the underlying periodic function that arises when studying the shifted LP-relaxation for TU integer recourse models is point-symmetric periodic with period $p = 1$. However, our results apply to all point-symmetric periodic functions φ with period $p > 0$.

Definition 2.7. A periodic function $\varphi : \mathbb{R} \to \mathbb{R}$ with period p satisfying

$$\varphi(x) = -\varphi(p - x),$$

or equivalently

$$\varphi(x) = \frac{1}{2}(\varphi(x) - \varphi(p - x)),$$

for all $x \in [0, p]$ is said to be *point-symmetric periodic*, or a PSP(p) function.

For a packed density g of Definition 2.6 and a PSP(p) function φ, we are able to derive an upper bound on $D(\varphi, g)$. The first step we take in obtaining such a bound is to prove in Lemma 2.2 that there exists a function g_p that is symmetric in the same sense as φ and carries all relevant information of g for computing $D(\varphi, g)$. Moreover, the total variation of g_p on $[0, p]$ does not exceed that of g on $[0, p]$. Next, in Lemma 2.3, we obtain bounds on higher-order derivatives of g_p by combining the symmetry of g_p with the mean-value theorem. In Section 2.3.3, we translate these bounds into a single bound on g_p, a process we refer to as *bound propagation*.

Lemma 2.2. *Let $n \in \mathbb{N}$ and $p \in \mathbb{R}$ with $p > 0$ be given. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a PSP(p) function. For a packed density g of Definition 2.6, define $g_p : [0, p] \to \mathbb{R}$ as*

$$g_p(x) := \frac{1}{2}(g(x) - g(p - x)), \quad x \in [0, p]. \tag{2.7}$$

Then,

(i) $D(\varphi, g) = \int_0^p \varphi(x) g_p(x) dx = 2 \int_0^{p/2} \varphi(x) g_p(x) dx$

and for all integers k satisfying $0 \leq k \leq n$,

(ii) $|\Delta| g_p^{(k)}([0, p]) \leq |\Delta| g^{(k)}([0, p])$.

Proof. For property (i), note that it follows from a substitution that

$$\int_0^p \varphi(x)g(x)dx = \int_0^p \varphi(p-x)g(p-x)dx = -\int_0^p \varphi(x)g(p-x)dx,$$

where the latter equality holds since φ is a PSP(p) function. Using (iv) in Lemma 2.1 we have

$$\begin{aligned} D(\varphi, g) &= \int_0^p \varphi(x)g(x)dx \\ &= \frac{1}{2}\int_0^p \varphi(x)g(x)dx + \frac{1}{2}\int_0^p \varphi(x)g(x)dx \\ &= \frac{1}{2}\int_0^p \varphi(x)g(x) - \frac{1}{2}\int_0^p \varphi(x)g(p-x))dx \\ &= \int_0^p \varphi(x)g_p(x)dx, \end{aligned}$$

and thus the first equality in (i) holds. To prove the second equality in (i), note that

$$\int_0^p \varphi(x)g_p(x)dx = \int_0^{p/2} \varphi(x)g_p(x)dx + \int_{p/2}^p \varphi(x)g_p(x)dx,$$

where the second term on the right-hand side can be rewritten as

$$\int_0^{p/2} \varphi(p-x)g_p(p-x)dx = \int_0^{p/2} \varphi(x)g_p(x)dx.$$

To prove (ii) we make use of standard properties the total variation of functions. Define $l : \mathbb{R} \to \mathbb{R}$ as $l(x) := g(p-x)$. We have

$$\begin{aligned} |\Delta| g_p^{(k)}([0,p]) &= |\Delta| \left(\frac{1}{2}\left(g^{(k)} + (-1)^k l^{(k)} \right) \right)([0,p]) \\ &= \frac{1}{2}|\Delta|(g^{(k)} + (-1)^k l^{(k)})([0,p]) \\ &\leq \frac{1}{2}\left(|\Delta| g^{(k)}([0,p]) + |\Delta| l^{(k)}([0,p]) \right) \\ &= \frac{1}{2}\left(|\Delta| g^{(k)}([0,p]) + |\Delta| g^{(k)}([0,p]) \right) \\ &= |\Delta| g^{(k)}([0,p]), \end{aligned}$$

where the second equality holds since for every closed interval $I \subset \mathbb{R}$, $|\Delta|(\alpha f)(I) = |\alpha||\Delta| f(I)$, for scalar α and f of bounded variation, and where the inequality follows from $|\Delta|(f+g)(I) \leq |\Delta| f(I) + |\Delta| g(I)$ for f and g of bounded variation. ■

We now state and prove a number of properties of g_p that we will use to prove the results in Sections 2.3.3 and 2.3.4. We make a distinction between odd and even-order derivatives of g_p, because they share different properties.

Lemma 2.3. *For a given $n \in \mathbb{N}$ and $p \in \mathbb{R}$ with $p > 0$, let $g : [0,p] \to \mathbb{R}$ be a packed density with discontinuities $\gamma_k := g^{(k)}(0) - g^{(k)}(p)$, $k = 0, \ldots, n$. Define $g_p : [0,p] \to \mathbb{R}$ as $g_p(x) := \frac{1}{2}(g(x) - g(p-x))$. Then, for all even k, $0 \leq k \leq n$,*

(i) $g_p^{(k)}(p/2) = 0$,

(ii) $g_p^{(k)}(0) = -g_p^{(k)}(p) = \dfrac{\gamma_k}{2}$, *and*

(iii) $\left|g_p^{(k)}(x)\right| \leq w_k := \dfrac{|\gamma_k| + |\Delta| g^{(k)}([0,p])}{4}$ *for all $x \in [0,p]$.*

Furthermore, for all odd k, $1 \leq k \leq n$,

(iv) there exist $x_k \in (0, p/2]$, such that $g_p^{(k)}(x_k) = -\frac{\gamma_{k-1}}{p}$, and

(v) $\left|g_p^{(k)}(x)\right| \leq w_k := \dfrac{|\gamma_{k-1}|}{p} + \dfrac{|\Delta| g^{(k)}([0,p])}{2}$ *for all $x \in [0,p]$.*

Proof. Note that from the definition of g_p,

$$g_p^{(k)}(x) = \frac{1}{2}\left(g^{(k)}(x) + (-1)^{k+1} g^{(k)}(p-x)\right), \quad k = 0, \ldots, n, \quad x \in \mathbb{R},$$

so that,

$$g_p^{(k)}(x) = (-1)^{k+1} g_p^{(k)}(p-x). \tag{2.8}$$

Properties (i) and (ii) then follow directly.

To prove property (iv) we make use of the mean-value theorem and property (ii). Together they imply that there exists a $c_k \in (0,p)$ such that

$$g_p^{(k)}(c_k) = \frac{-\frac{\gamma_{k-1}}{2} - \frac{\gamma_{k-1}}{2}}{p-0} = -\frac{\gamma_{k-1}}{p}.$$

The existence of $x_k \in (0, p/2]$ as in property (iv) then follows from the symmetry of odd-order derivatives, see (2.8).

We will prove (iii) by contradiction using property (ii) in Lemma 2.2. Fix an even k and assume for contradiction that $\left|g_p^{(k)}(t^*)\right| > w_k$ for some $t^* \in [0, p/2)$.

This implies that

$$\left|g_p^{(k)}(t^*)\right| > \frac{|\gamma_k|}{2},$$

because, as a consequence of (ii), $|\Delta|g^{(k)}([0,p]) \geq |\gamma_k|$.

Consider the partition $P = \{0, t^*, p - t^*, p\}$. Following the notation introduced in Definition 2.2, we have for this partition,

$$V_{g_p^{(k)}}(P) = 4\left|g_p^{(k)}(t^*)\right| \pm |\gamma_k| > 4w_k \pm |\gamma_k| \geq |\Delta|g^{(k)}([0,p]),$$

where the equality is due to equation (2.8) and property (ii). We thus find a contradiction with Lemma 2.2, property (ii).

Property (v) follows in a similar fashion, since by combining (iv) with (2.8) it is possible to arrive at a contradiction with property (ii) in Lemma 2.2 if (v) does not hold. ■

2.3.3 Bound propagation

In this section, we use the bounds on $g_p^{(k)}, k = 0, \ldots, n$, in Lemma 2.3 to derive a single tighter bound on g_p as defined in (2.7) for a given packed density g with known discontinuities $(\gamma_0, \ldots, \gamma_n)$. We use information on the total variation of g and its higher-order derivatives to bound g_p. Clearly, we obtain tighter bounds if we include information on more higher-order derivatives of g.

For illustration, suppose that information is available on g and $g^{(1)}$. Given a bound on $|\Delta|g$, property (iii) in Lemma 2.3 directly yields a uniform bound on $|g_p(x)|$ for $x \in [0,p]$, denoted w_0. An additional bound on $|\Delta|g^{(1)}$ yields a bound on $|g_p^{(1)}(x)|$, denoted w_1, using property (v) in Lemma 2.3. The latter bound can be used to improve the bound on $|g_p(x)|$ itself, using the concept of bound propagation, which we will now demonstrate.

Let $q(x)$ be such that $\left|g_p^{(1)}(x)\right| \leq q(x)$ for $x \in [0, p/2]$. Using that $g_p\,(p/2) = 0$ according to property (i) in Lemma 2.3, and that

$$\begin{aligned}
\left|g_p(x)\right| &= \left|g_p(x)\right| - \left|g_p\,(p/2)\right| \\
&\leq \left|g_p(x) - g_p\,(p/2)\right| \\
&\leq \int_x^{p/2} \left|g_p^{(1)}(s)\right| ds \\
&\leq \int_x^{p/2} q(s) ds,
\end{aligned}$$

we obtain an upper bound on $|g_p(x)|$. Similarly we can extrapolate around $x = 0$ and use the fact that $g_p(0) = \frac{\gamma_0}{2}$ to obtain

$$|g_p(x)| \leq \frac{|\gamma_0|}{2} + \int_0^x q(s)ds.$$

This means that we can transform an upper bound $q(x)$ on $|g_p^{(1)}(x)|$ for all $x \in [0, p/2]$ into an upper bound on $|g_p(x)|$ for every $x \in [0, p/2]$. We formalize this transformation by defining an operator T that maps the upper bound q on $|g_p^{(1)}|$ into an upper bound Tq on $|g_p|$. This operator is defined as,

$$(Tq)(x) := \min\left\{w_0, \frac{|\gamma_0|}{2} + \int_0^x q(s)ds, \int_x^{p/2} q(s)ds\right\}, \quad x \in [0, p/2],$$

and includes the uniform upper bound w_0 on $|g_p|$ as defined in Lemma 3 (iii). From the analysis above it follows that $|g_p(x)| \leq (Tq)(x)$ for $x \in [0, p/2]$. Note that $(Tq)(x)$ provides a tighter bound than the bound based on $|\Delta|g$ alone, which is given by w_0.

We generalize this idea to higher-order derivatives by introducing appropriate operators in Definition 2.8. The idea is that a bound q on $|g_p^n|$ may be propogated in dynamic programming fashion to obtain a bound $T \cdots Tq$ on $|g_p|$. In our case, however, we have to define two operators T^1 and T^2 since the process of bound propagation is different for odd and even higher-order derivatives. In Definition 2.9, we use these operators to define functions q_0^n, $n \in \mathbb{N}$, which represent a bound on $\left|g_p(x)\right|$ based on information on the first n derivatives of g. Intuitively, including more derivatives should lead to sharper bounds. This intuition is confirmed in Corollary 2.1.

Definition 2.8. For a given $p > 0$, a function $q : [0, p/2] \to \mathbb{R}$, and parameters $w > 0$ and $\gamma \in \mathbb{R}$, define the operators $T^1_{w,\gamma}$ and $T^2_{w,\gamma}$ by

$$(T^1_{w,\gamma}q)(x) := \min\left\{w, \frac{|\gamma|}{p} + \max\left\{\int_0^x q(s)ds, \int_x^{p/2} q(s)ds\right\}\right\}, \quad x \in [0, p/2],$$

and

$$(T^2_{w,\gamma}q)(x) := \min\left\{w, \frac{|\gamma|}{2} + \int_0^x q(s)ds, \int_x^{p/2} q(s)ds\right\}, \quad x \in [0, p/2].$$

Definition 2.9. Let $n \in \mathbb{N}$ and $p > 0$ be given. Let $w_k > 0$ and $\gamma_k \in \mathbb{R}$, $k = 0, \ldots, n$, be given constants. Define $q_n^n : [0, p/2] \to \mathbb{R}$ as $q_n^n(x) := w_n$. For $k < n$, define

$q_k^n : [0, p/2] \to \mathbb{R}$ using backward recursion as

$$q_k^n := \begin{cases} T^1_{w_k,\gamma_{k-1}} q_{k+1}^n, & \text{if } k \text{ is odd,} \\ T^2_{w_k,\gamma_k} q_{k+1}^n, & \text{if } k \text{ is even.} \end{cases}$$

Before we are ready to prove that q_0^n yields a bound on $|g_p|$, we need some elementary properties of the operators introduced in Definition 2.8. In Lemma 2.4, we prove that T^j, $j = 1, 2$, are non-negative and monotone operators.

Lemma 2.4. *Let $p > 0$ be given. Let $q, \tilde{q} : [0, p/2] \to \mathbb{R}$ be non-negative functions such that $\bar{q} \geq q$. Then for all $w > 0$ and $\gamma \in \mathbb{R}$,*

(i) $T^j_{w,\gamma} q$ is a non-negative function, $j = 1, 2$, and

(ii) $T^j_{w,\gamma} \bar{q} \geq T^i_{w,\gamma} q$, $j = 1, 2$.

Proof. Property (i) follows directly from the non-negativity of q. Property (ii) is a direct consequence of $\bar{q} \geq q$. ■

Proposition 2.1. *Let $n \in \mathbb{N}$ and $p > 0$ be given. Let $g : [0, p] \to \mathbb{R}$ be a packed density with discontinuities $(\gamma_0, \ldots, \gamma_n)$. Define $g_p : [0, p] \to \mathbb{R}$ as $g_p(x) := \frac{1}{2}(g(x) - g(p - x))$. For $k = 0, \ldots, n$, define*

$$w_k := \begin{cases} \frac{|\gamma_{k-1}|}{p} + \frac{|\Delta| g^{(k)}([0,p])}{2}, & \text{if } k \text{ is odd,} \\ \frac{|\gamma_k| + |\Delta| g^{(k)}([0,p])}{4}, & \text{if } k \text{ is even.} \end{cases},$$

so that w_k denotes a uniform bound on $\left| g_p^{(k)} \right|$. Define q_k^n, $k = 0, \ldots, n$, as in Definition 2.9. Then,

$$\left| g_p(x) \right| \leq q_0^n(x), \quad x \in [0, p/2]\,.$$

Proof. The non-negativity of T^j, $j = 1, 2$, implies that q_k^n, $k = 0, \ldots, n$, are non-negative functions. We prove the stronger claim

$$|g_p^{(k)}(x)| \leq q_k^n(x), \tag{2.9}$$

for all $x \in [0, p/2]$, for $k = 0, \ldots, n$ using backward induction.

For $k = n$, the inequality in (2.9) follows directly from property (iii) and (iv) in Lemma 2.3. For the induction step, assume (2.9) holds for $k = m + 1$, with

$0 \leq m < n$. We consider odd and even m separately. For even m, we have to show that

$$\left| g_p^{(m)}(x) \right| \leq (T^2_{w_m,\gamma_m} q^n_{m+1})(x), \quad x \in [0, p/2]\,.$$

Since

$$(T^2_{w_m,\gamma_m} q^n_{m+1})(x) = \min\left\{ w_m, \frac{|\gamma_m|}{2} + \int_0^x q^n_{m+1}(s)ds, \int_x^{p/2} q^n_{m+1}(s)ds \right\},$$

it suffices to show that for all $x \in [0, p/2]$,

$$\left| g_p^{(m)}(x) \right| \leq w_m, \tag{2.10}$$

$$\left| g_p^{(m)}(x) \right| \leq \frac{|\gamma_m|}{2} + \int_0^x q^n_{m+1}(s)ds, \tag{2.11}$$

and

$$\left| g_p^{(m)}(x) \right| \leq \int_x^{p/2} q^n_{m+1}(s)ds. \tag{2.12}$$

The induction step for even m then follows by combining (2.10), (2.11), and (2.12).

The inequality in (2.10) is due to property (iii) in Lemma 2.3. To prove (2.11), we use that

$$\left| g_p^{(m)}(x) \right| - \left| g_p^{(m)}(0) \right| \leq \left| g_p^{(m)}(x) - g_p^{(m)}(0) \right| = \left| \int_0^x g_p^{(m+1)}(s)ds \right|,$$

by the triangle inequality. Using the induction hypothesis and the fact that q^n_{m+1} is a non-negative function, we obtain

$$\left| g_p^{(m)}(x) \right| - \left| g_p^{(m)}(0) \right| \leq \int_0^x q^n_{m+1}(s)ds.$$

Then, (2.11) follows by inserting $g_p^{(m)}(0) = \frac{\gamma_m}{2}$ by Lemma 2.3 (ii), and rearranging terms. The inequality in (2.12) can be proved in a similar manner as (2.11) by applying the triangle inequality to $\left| g_p^{(m)}(p/2) - g_p^{(m)}(x) \right|$. We conclude that the induction step holds for even m.

For odd m, we have to show that

$$\left|g_p^{(m)}(x)\right| \leq (T^1_{w_m,\gamma_{m-1}} q^n_{m+1})(x), \quad x \in [0, p/2]\,.$$

By definition,

$$\begin{aligned}(T^1_{w_m,\gamma_{m-1}} q^n_{m+1})(x) \\ = \min\left\{w_m, \frac{|\gamma_{m-1}|}{p} + \max\left\{\int_0^x q^n_{m+1}(s)ds, \int_x^{p/2} q^n_{m+1}(s)ds\right\}\right\},\end{aligned}$$

so that it suffices to show for all $x \in [0, p/2]$ that

$$\left|g_p^{(m)}(x)\right| \leq w_m, \tag{2.13}$$

and

$$\left|g_p^{(m)}(x)\right| \leq \frac{|\gamma_{m-1}|}{p} + \max\left\{\int_0^x q^n_{m+1}(s)ds, \int_x^{p/2} q^n_{m+1}(s)ds\right\}. \tag{2.14}$$

The inequality in (2.13) follows directly from property (v) in Lemma 2.3. To prove (2.14) we make use of in Lemma 2.3 (iv), which states that there exists $x_m \in (0, p/2]$ such that $g_p^{(m)}(x_m) = -\frac{\gamma_{m-1}}{p}$. By the triangle inequality,

$$\left|g_p^{(m)}(x)\right| - \left|g_p^{(m)}(x_m)\right| \leq \left|g_p^{(m)}(x) - g_p^{(m)}(x_m)\right| = \left|\int_{x_m}^x g_p^{(m+1)}(s)ds\right|.$$

Together with the induction hypothesis, this yields

$$\left|g_p^{(m)}(x)\right| - \left|g_p^{(m)}(x_m)\right| \leq \left|\int_{x_m}^x q^n_{m+1}(s)ds\right|.$$

To arrive at (2.14), we make the following observation

$$\begin{aligned}\left|\int_{x_m}^x q^n_{m+1}(s)ds\right| &= \max\left\{\int_{x_m}^x q^n_{m+1}(s)ds, \int_x^{x_m} q^n_{m+1}(s)ds\right\} \\ &\leq \max\left\{\int_0^x q^n_{m+1}(s)ds, \int_x^{p/2} q^n_{m+1}(s)ds\right\},\end{aligned}$$

where the inequality holds since $x_m \in (0, p/2]$ and q^n_{m+1} is non-negative. This completes the induction step for odd m, and the proof of (2.9). ■

We now present a corollary of Lemma 2.4, which states the intuitive result that

the bounds on g_p become sharper if more higher-order derivatives are included.

Corollary 2.1. *Let $n \in \mathbb{N}$ and $p > 0$ be given. Let γ_k and $w_k > 0$, $k = 0, \ldots, n$, be given constants. Define q_0^l, $l = 0, \ldots, n$, as in Definition 2.9. Then,*

$$q_0^0 \geq q_0^1 \geq \cdots \geq q_0^n.$$

Proof. Fix an $l \in \{1, \ldots, n\}$. We will use backward induction to prove that $q_k^{l-1} \geq q_k^l$ for $k = 0, \ldots, l-1$. The claim then follows by setting $k = 0$.

It follows directly from the definition of q_k^n that $q_{l-1}^{l-1} \geq q_{l-1}^l$. For the induction step, suppose that $q_k^{l-1} \geq q_k^l$ for some k, $0 < k \leq l-1$. It follows from the monotonicity of T^j, $j = 1, 2$, that $q_{k-1}^{l-1} \geq q_{k-1}^l$, completing the proof. ■

2.3.4 Error bounds

We are now ready to state the main results of this section. In Theorem 2.2, we formulate an improved bound on the expectation of PSP(p) functions and Theorem 2.3 states the bound on the expectation of half-periodic functions where the underlying periodic function is PSP(p).

Theorem 2.2. *Let $n \in \mathbb{N}$ and $p > 0$ be given. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a PSP(p) function and let $B = (B_0, \ldots, B_n)$ and $C = (C_0, \ldots, C_n)$ be positive constants. Then, for M_n defined as*

$$M_n(\varphi, B, C) := \sup_{f \in \mathcal{F}_n} \left\{ \left| \mathbb{E}_f[\varphi(\omega)] \right| : |\Delta| f^{(k)} \leq B_k, \right.$$
$$\left. ||f^{(k)}||_\infty \leq C_k, \text{ for } k = 0, \ldots, n \right\},$$

there holds

$$M_n(\varphi, B, C) \leq 2 \int_0^{p/2} |\varphi(x)| \, q_0^n(x) dx,$$

where q_0^n is defined as in Definition 2.9, with $\gamma_k = 0$, $k = 0, \ldots, n$, and

$$w_k = \begin{cases} \frac{B_k}{2}, & \text{if } k \text{ is odd,} \\ \frac{B_k}{4}, & \text{if } k \text{ is even.} \end{cases}$$

Proof. For a given PSP(p) function φ and pdf $f \in \mathcal{F}_n$, we know from Lemma 2.1

that

$$\left|\mathbb{E}_f[\varphi(\omega)]\right| = \left|\int_0^p \varphi(x) f_p(x) dx\right|,$$

where f_p denotes the classical packed density corresponding to f. The function f_p is a packed density with discontinuities $\gamma = (0, \ldots, 0)$. Lemma 2.2 informs us that

$$\left|\mathbb{E}_f[\varphi(\omega)]\right| = 2\left|\int_0^{p/2} \varphi(x) g_p(x) dx\right|,$$

where $g_p(x) = \frac{1}{2}(f_p(x) - f_p(p-x))$. This implies

$$\left|\mathbb{E}_f[\varphi(\omega)]\right| \leq 2\int_0^{p/2} |\varphi(x)||g_p(x)| dx.$$

By Lemma 2.1, property (iii), we know that B_k is an upper bound on $|\Delta| f_p^k([0,p])$, $k = 0, \ldots, n$. Combining Proposition 2.1 with the monotonicity of the operators T^j, $j = 1, 2$, yields that q_0^n is an upper bound on $|g_p|$, completing the proof. ■

Remark 2.5. Note that the bound presented in Theorem 2.2 is independent of C. This is because the discontinuities of the classical packed density are zero.

Theorem 2.3. *Let $n \in \mathbb{N}$ and $p > 0$ be given. Let $\psi : \mathbb{R} \to \mathbb{R}$ be a half-periodic function defined as*

$$\psi(x) := \begin{cases} 0, & x < 0, \\ \varphi(x), & x \geq 0, \end{cases}$$

where $\varphi : \mathbb{R} \to \mathbb{R}$ is a PSP(p) function. Let $B = (B_0, \ldots, B_n)$ and $C = (C_0, \ldots, C_n)$ be positive constants. Then, for M_n defined as

$$M_n(\psi, B, C) := \sup_{f \in \mathcal{F}_n} \left\{ \left|\mathbb{E}_f[\psi(\omega)]\right| : |\Delta| f^{(k)} \leq B_k, \right. \\ \left. ||f^{(k)}||_\infty \leq C_k, \text{ for } k = 0, \ldots, n \right\},$$

there holds

$$M_n(\psi, B, C) \leq 2\int_0^{p/2} |\varphi(x)|\, q_0^n(x) dx,$$

where q_0^n is defined as in Definition 2.9, with $\gamma_k = C_k$, $k = 0, \ldots, n$, and

$$w_k = \begin{cases} \frac{C_{k-1}}{p} + \frac{B_k}{2}, & \text{if } k \text{ is odd,} \\ \frac{B_k}{4}, & \text{if } k \text{ is even.} \end{cases}$$

Proof. For a given half-periodic function ψ with underlying PSP(p) function φ and pdf $f \in \mathcal{F}_n$, we know from Lemma 2.1 that

$$\left| \mathbb{E}_f[\psi(\omega)] \right| = \left| \int_0^p \varphi(x) \hat{f}_p(x) dx \right|,$$

where $\hat{f}_p$ denotes the half-packed density corresponding to f. The function $\hat{f}_p$ is a packed density with discontinuities $\gamma = (f^{(0)}(0), \ldots, f^{(n)}(0))$. Lemma 2.2 informs us that

$$\left| \mathbb{E}_f[\psi(\omega)] \right| = 2 \left| \int_0^{p/2} \varphi(x) g_p(x) dx \right|,$$

where $g_p(x) = \frac{1}{2}(\hat{f}_p(x) - \hat{f}_p(p-x))$. This implies

$$\left| \mathbb{E}_f[\psi(\omega)] \right| \leq 2 \int_0^{p/2} |\varphi(x)||g_p(x)| dx.$$

By Lemma 2.1, property (iii), we know that $B_k - |f^{(k)}(0)|$ is an upper bound on $|\Delta| \hat{f}_p^k([0,p])$, $k = 0, \ldots, n$. Furthermore, $|\gamma_k| = |f^{(k)}(0)| \leq C_k$. Combining Proposition 2.1 with the monotonicity of the operators T^j, $j = 1, 2$, yields that q_0^n is an upper bound on $|g_p|$, and the result follows. ∎

2.4 Applications and examples

In this section, we apply the results in Section 2.3 to specific functions. As in Section 2.3, we make a distinction between periodic and half-periodic functions. First, we apply Theorem 2.2 to a particular PSP(p) function. Second, we derive an improved error bound for the shifted LP-relaxation approximation of one-dimensional SIR models. Since the underlying difference function is half-periodic, we can use Theorem 2.3 to derive such a bound. This error bound can be generalized to higher-dimensional SIR models. Finally, we conduct numerical experiments to compare the performance of the error bound by Romeijnders et al. [61] and the improved bound derived in this section.

2.4.1 Point symmetric periodic functions

Here, we derive a bound on the expectation of the underlying periodic function φ for the shifted LP-relaxation for SIR models, defined as $\varphi(x) := x - \lceil x - \frac{1}{2} \rceil$. Note that φ is a PSP(1) function and that $|\varphi(x)| = x$ for $x \in [0, \frac{1}{2}]$. In the notation of Theorem 2.2, take $n = 1$, which is to say that we make use of the total variation of the underlying pdf and its first derivative. Furthermore, $q_1^1(x) = \frac{B_1}{2}$ and

$$q_0^1(x) = \min\left\{ \frac{B_1}{2}x, \frac{B_0}{4}, \frac{B_1}{2}\left(\frac{1}{2} - x\right) \right\}$$

Applying Theorem 2.2 yields

$$M_1(\varphi, B_0, B_1, C_0, C_1) \leq \int_0^{\frac{1}{2}} x \min\left\{ B_1 x, \frac{B_0}{2}, B_1\left(\frac{1}{2} - x\right) \right\} dx.$$

Using this, we find that for any $f \in \mathcal{F}_1$

$$\left| \mathbb{E}_f[\varphi(\omega)] \right| \leq k(|\Delta|f, |\Delta|f') := \begin{cases} \frac{|\Delta|f}{16}\left(1 - \frac{|\Delta|f}{|\Delta|f'}\right), & \frac{|\Delta|f}{|\Delta|f'} < \frac{1}{2}, \\ \frac{|\Delta|f'}{64}, & \frac{|\Delta|f}{|\Delta|f'} \geq \frac{1}{2}. \end{cases}$$

Since this bound holds for all $f \in \mathcal{F}_1$, it follows that if ω has pdf $f \in \mathcal{F}_1$, then

$$\sup_{z \in \mathbb{R}} \mathbb{E}_\omega[\varphi(\omega - z)] \leq k(|\Delta|f, |\Delta|f').$$

This result is an improvement of the bound by Romeijnders et al. [61], who showed that

$$\left| \mathbb{E}_f[\varphi(\omega)] \right| \leq \frac{1}{2} h(|\Delta|f) = \begin{cases} \frac{|\Delta|f}{16}, & 0 < |\Delta|f \leq 4, \\ \frac{1}{2} - \frac{1}{|\Delta|f}, & |\Delta|f \geq 4. \end{cases}$$

Note that the improvement is large if $|\Delta|f'$ is small relative to $|\Delta|f$. We now apply this result to a range of specific pdf's.

Example 2.1. Let f denote the pdf of a normally distributed random variable ω with variance σ^2. Then,

$$|\Delta|f = \sigma^{-1}\sqrt{2/\pi}, \quad \text{and } |\Delta|f' = \sigma^{-2}\sqrt{\frac{8}{\pi e}}.$$

Observe that the ratio

$$\frac{|\Delta|f}{|\Delta|f'} = \sigma\frac{\sqrt{e}}{2}$$

increases linearly in σ. This implies that the improvement over the original total variation bound is unbounded and increases with σ. We have

$$\left|\mathbb{E}_f[\varphi(\omega)]\right| \leq k(|\Delta|f, |\Delta|f') = \begin{cases} \frac{1}{16}\sqrt{\frac{2}{\pi}}\frac{1}{\sigma}\left(1 - \frac{\sqrt{e}}{2}\sigma\right), & \text{if } \sigma < \frac{1}{\sqrt{e}}, \\ \frac{1}{32}\sqrt{\frac{2}{\pi e}}\frac{1}{\sigma^2}, & \text{if } \sigma \geq \frac{1}{\sqrt{e}}. \end{cases}$$

Figure 2.2 shows the true value of $\sup_{z\in\mathbb{R}}\{\mathbb{E}_\omega[\varphi(\omega - z)]\}$, the upper bound $\frac{1}{2}h(|\Delta|f)$ derived by Romeijnders et al. [61], and the upper bound $k(|\Delta|f, |\Delta|f')$ based on M_1. Moreover, we include the upper bound based on M_2. The latter bound is tractable, but an analytical expression is cumbersome and does not yield further insights and is therefore omitted. Note that the bounds based on M_1 and M_2 provide a tighter bound compared to $\frac{1}{2}h(|\Delta|f)$ for large values of σ. ◊

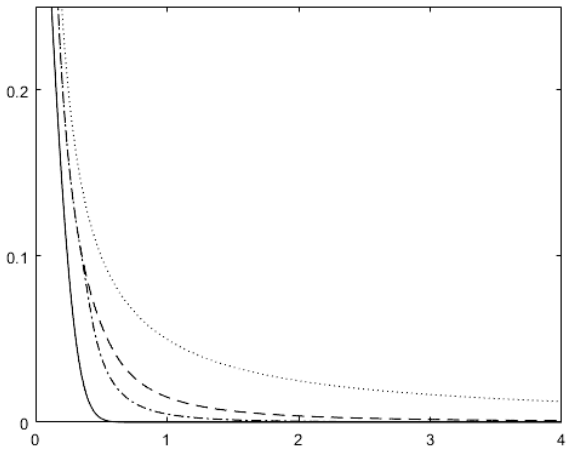


Figure 2.2. The true value of $\sup_{z\in\mathbb{R}}\{\mathbb{E}_\omega[\varphi(\omega - z)]\}$ (solid), and the upper bounds based on [61] (dotted), M_1 (dashed), and M_2 (dash-dotted) as a function of σ, where ω follows a normal distribution with variance σ^2.

Note that our results only apply to continuously differentiable pdf, nevertheless, for continuous pdf f that are not continuously differentiable we can still apply our results by considering the right derivative of f, denoted f'_+, instead of the derivative of f. The reason for this lies in the fact that there exists a continuously

differentiable approximation $\tilde{f}$ of f such that $|\Delta|f = |\Delta|\tilde{f}$, $|\Delta|f'_+ = |\Delta|\tilde{f}'$, and $\mathbb{E}_f[\varphi(\omega - z)]$ is arbitrarily close to $\mathbb{E}_{\tilde{f}}[\varphi(\omega - z)]$. We illustrate this in the next example.

Example 2.2. Suppose that ω follows a triangular distribution with support $[a, b]$ and mode m, $a < m < b$. See Figure 2.3 for illustration.

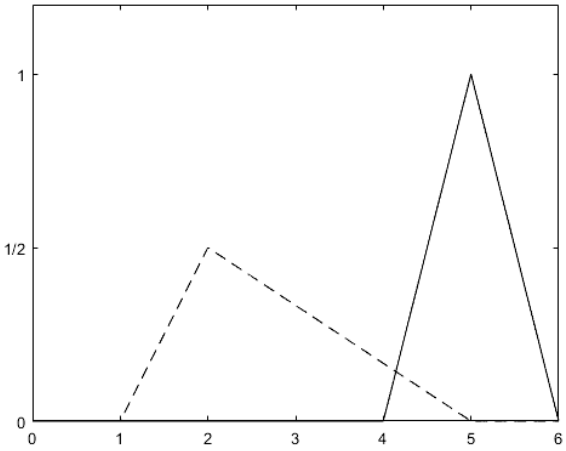


Figure 2.3. The probability density function of a random variable following a triangular distribution with support $[4, 6]$ and mode 5 (solid); and support $[1, 5]$ and mode 2 (dashed).

Denote the pdf of ω by f. Write $d := b - a$ and note that $f(m) = \frac{2}{d}$, so that,

$$|\Delta|f = \frac{4}{d},$$

furthermore,

$$|\Delta|f'_+ = \frac{4}{(m-a)(b-m)}.$$

This leads to

$$\left|\mathbb{E}_\omega[\varphi(\omega)]\right| \leq \begin{cases} \frac{1}{4d}\left(1 - \frac{(m-a)(b-m)}{d}\right), & \text{if } (m-a)(b-m) < \frac{1}{2}d, \\ \frac{1}{16(m-a)(b-m)}, & \text{if } (m-a)(b-m) \geq \frac{1}{2}d. \end{cases}$$

Note that $|\Delta|f$ is independent of the mode m, however, changing m does affect $|\Delta|f'_+$. In fact, for given a and b, $|\Delta|f'_+$ is minimized by $m = \frac{1}{2}(a + b)$. Hence, as the

mode is closer to the midpoint of the support of ω, $|\Delta|f'_+$ is smaller and we obtain tighter bounds. $\Diamond$

2.4.2 Error bound shifted-LP relaxation

In this section, we consider the underlying difference function ψ, defined in (2.4), for the shifted LP-relaxation for SIR models. Recall that bounds on the expectation of ψ can be used directly to derive error bounds for the shifted LP-relaxation for SIR models.

We apply Theorem 2.3 with $n = 1$ to derive a such a bound. We have $\varphi(x) = x - \left\lceil x - \frac{1}{2} \right\rceil$. Like in Section 2.4.1, $|\varphi(x)| = x$ for all $x \in [0, \frac{1}{2}]$. Note that $q_1^1(x) = C_0 + \frac{B_1}{2} =: D$ and that

$$q_0^1(x) = \min\left\{ \frac{B_0}{4}, \frac{C_0}{2} + Dx, D\left(\frac{1}{2} - x\right) \right\}.$$

Hence, by Theorem 2.3,

$$M_1(\psi, B_0, B_1, C_0, C_1) \leq \int_0^{\frac{1}{2}} x \min\left\{ \frac{B_0}{2}, C_0 + Dx, D\left(\frac{1}{2} - x\right) \right\} dx,$$

where $D = 2C_0 + B_1$. Note that for any pdf f,

$$||f||_\infty \leq \frac{|\Delta|f}{2},$$

where equality holds if f is unimodal. For this reason, we take $C_0 = \frac{B_0}{2}$, so that the resulting bound holds for all pdf $f \in \mathcal{F}_1$. For unimodal pdf's the resulting bound is the tightest bound we can provide, whereas tighter bounds can be derived for non-unimodal pdf. Simple computations yield

$$M_1(\psi, B_0, B_1, \frac{B_0}{2}, C_1) \leq \frac{B_0}{16} S(B_0, B_1),$$

where

$$S(x, y) := \left(1 - \frac{x\,(2x + 3y)}{3\,(x + y)^2}\right) \in (1/3, 1),$$

for positive x and y. We thus have that for any $f \in \mathcal{F}_1$,

$$\left|\mathbb{E}_f[\psi(\omega)]\right| \leq \frac{|\Delta|f}{16} S(|\Delta|f, |\Delta|f').$$

The original bound by Romeijnders et al. [61] is given by $\frac{|\Delta|f}{16}$, so $S(|\Delta|f, |\Delta|f')$ represents the improvement over their results. Note that $S(|\Delta|f, |\Delta|f') \to \frac{1}{3}$ as $\frac{|\Delta|f'}{|\Delta|f} \to 0$, which is to say that the improvement factor converges to three. We return to this fact in the following example, where we apply our results to a normally distributed random variable.

Example 2.3. We numerically evaluate $||\hat{Q} - Q||_\infty$, where $\hat{Q}$ denotes the shifted-LP relaxation by Romeijnders et al. [61] as in Definition 2.1, in the one-dimensional SIR case where ω follows a normal distribution with arbitrary mean and variance σ^2, for $\sigma \in [0.2, 4]$. We compare the actual error to the error bound based on [61] and the improved error bounds based on M_1 and M_2. We omit the analytical expression for the bound based on M_2, which is tractable but cumbersome.

Denote the pdf of ω by f. To compute the (improved) error bounds, we make use of the expressions for $|\Delta|f$ and $|\Delta|f'$ found in Example 2.1. It follows from these expressions that

$$\frac{|\Delta|f'}{|\Delta|f} \to 0$$

as $\sigma \to \infty$, which implies that the improvement factor over the original total variation bound converges to three for large values of σ. Figure 2.4 shows the results of this numerical experiment. Note that the improvement over the original error bound increases with σ. $\Diamond$

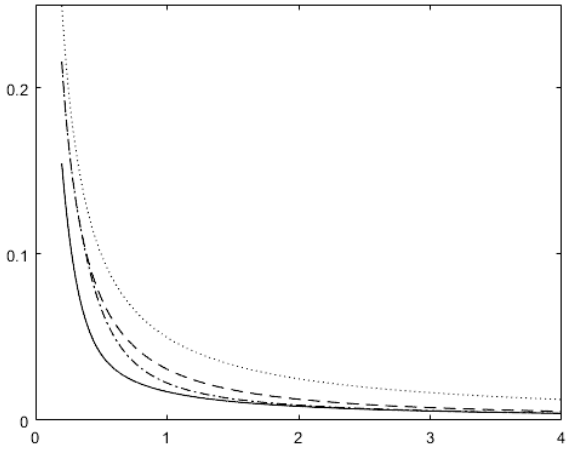


Figure 2.4. The true value of $||\hat{Q} - Q||_\infty$ (solid), and the upper bounds based on [61] (dotted), M_1 (dashed), and M_2 (dash-dotted) in the one-dimensional SIR case as a function of σ, where ω follows a normal distribution with variance σ^2.

2.5 Conclusion

We consider existing convex approximations for two-stage mixed-integer recourse models. We construct a hierarchy of bounds on the expectation of periodic and half-periodic functions using total variations of higher-order derivatives of the underlying probability density function. We use these results to derive improved error bounds for the shifted LP-relaxation of simple integer recourse models. Moreover, the results presented here may be used to improve error bounds of convex approximations for general mixed-integer recourse model approximations.

There are multiple directions for future research. One extension is to generalize our results to a to a higher-dimensional setting. A first step in this direction may be to consider totally unimodular integer recourse models. Another avenue is to apply the results to a particular application of integer recourse models. Finally, our results may be extended to a larger class of periodic functions, which may be useful for other types of convex approximations.

Chapter 3

Pseudo-valid cutting planes for two-stage mixed-integer stochastic programs with right-hand side uncertainty

We propose a novel way of applying cutting plane techniques to two-stage mixed-integer stochastic programs with uncertainty in the right-hand side. Instead of using cutting planes that are always valid, our idea is to apply pseudo-valid cutting planes to the second-stage feasible regions that may cut away feasible integer second-stage solutions for some scenarios and may be overly conservative for others. The advantage is that it allows us to use cutting planes that are affine in the first-stage decision variables, so that the approximation is convex, and can be solved efficiently using techniques from convex optimization. We derive tight performance guarantees for using particular types of pseudo-valid cutting planes for simple integer recourse models. Moreover, we show in general that using pseudo-valid cutting planes leads to good first-stage solutions if the total variations of the one-dimensional conditional probability density functions of the random variables in the model converge to zero.

This chapter is based on the journal publication [57].

3.1 Introduction

Many practical problems under uncertainty in, e.g., energy, finance, logistics, and healthcare involve integer decision variables. Such problems can be modelled as mixed-integer stochastic programs (MISPs), but are notoriously difficult to solve. In this chapter, we do not attempt to solve these problems exactly. Instead, we introduce a novel approach to approximately solve two-stage MISPs with uncertainty in the right-hand side, and we derive performance guarantees for the resulting approximating solutions.

Traditional solution methods for MISPs combine solution approaches for continuous stochastic programs and deterministic mixed-integer programs (MIPs). See, e.g., Ahmed et al. [4] for branch-and-bound, Sen and Higle [69] and Ntaimo [48] for disjunctive decomposition, Carøe and Schultz [24] for dual decomposition, Laporte and Louveaux [44] for the integer L-shaped method, and Zhang and Küçükyavuz [88] for cutting plane techniques. All these solution methods aim at finding the exact optimal solution for MISPs, but generally have difficulties scaling up to solve large problem instances. This is not surprising, since contrary to their continuous counterparts, these MISPs are non-convex in general [54]. This means that efficient techniques from convex optimization cannot be used to solve these problems.

Based on this observation and inspired by the success of cutting plane techniques for deterministic MIPs [47], we propose to use *cutting planes* to solve two-stage MISPs. However, we will use them in a fundamentally different way than in existing methods for both deterministic and stochastic MIPs. Instead of using exact cutting planes that are always valid, we propose to use *pseudo-valid* cutting planes for the second-stage feasible regions in such a way that the approximating problem remains convex in the first-stage decision variables, and thus efficient convex optimization techniques can be used to solve the approximation.

The disadvantage of using pseudo-valid cutting planes is that they may cut away part of the second-stage feasible region or that they may be overly conservative, so that we significantly over- or underestimate the second-stage costs, respectively. However, for MISPs this may be justified since our aim is not to find the exact and complete characterization of the integer hulls of the second-stage feasible regions, but rather to obtain good first-stage decisions. In fact, one of our main contributions is that we obtain good or even near-optimal *first-stage* decisions for pseudo-valid cutting plane approximations in which the cutting planes only need to be valid for a grid of first-stage decisions.

For simple integer recourse (SIR) models, a special type of MISP, our pseudo-

valid cutting plane approximation turns out to be equivalent to convex α-approximations, derived by Klein Haneveld et al. [41] from a completely different perspective. By reinterpreting these α-approximations using pseudo-valid cutting planes, we connect two existing solution methodologies for MISPs that use convex approximations and exact cutting planes, respectively. Moreover, this reinterpretation allows us to apply existing performance guarantees derived in Romeijnders et al. [59] for α-approximations to pseudo-valid cutting plane techniques for SIR models. Furthermore, we use results from Romeijnders et al. [60] to show for general MISPs with continuously distributed random right-hand side parameters that the error of using pseudo-valid cutting plane approximations converges to zero if all total variations of the one-dimensional conditional probability density functions of these random parameters converge to zero.

Contrary to the convex approximation derived in [60], our pseudo-valid cutting planes are suitable for solving large-scale MISPs. Indeed, in a numerical case study we solve large nurse scheduling problems using tight pseudo-valid cutting planes and obtain first-stage decisions that are close to optimal. Moreover, we derive pseudo-valid mixed-integer Gomory cuts for general two-stage MISPs with right-hand side uncertainty.

Summarizing, the main contributions of this chapter are as follows.

- We propose a novel solution approach for two-stage MISPs with uncertainty in the right-hand side by applying pseudo-valid cutting planes to the second-stage feasible regions.
- We reinterpret α-approximations for SIR models as pseudo-valid cutting plane approximations, connecting two existing solution methodologies for MISPs, and yielding a tight error bound for applying pseudo-valid cutting planes to SIR models.
- We derive a performance guarantee for applying pseudo-valid cutting planes to general MISPs with continuous distributions and uncertainty in the right-hand side.
- We derive pseudo-valid mixed-integer Gomory cuts for general MISPs and derive tight pseudo-valid cutting planes for a nurse scheduling problem.
- We carry out numerical experiments on a nurse scheduling problem and show that we obtain good first-stage decisions when using pseudo-valid cutting planes.

The remainder of this chapter is organized as follows. In Section 3.2 we define MISPs and explain our pseudo-valid cutting plane approach. In Section 3.3, we reinterpret α-approximations for SIR models using pseudo-valid cutting planes, and in Section 3.4 we derive an error bound for pseudo-valid cutting plane approximations for general MISPs. In Section 3.5, we derive pseudo-valid mixed-integer Gomory cuts, and apply tight pseudo-valid cutting planes to a nurse scheduling problem. We present numerical experiments for the latter type of cutting planes in Section 3.6, and we end with a discussion in Section 3.7.

3.2 Problem definition and solution approach

3.2.1 Problem definition

Two-stage MISPs can be interpreted as hierarchical planning problems. In the first stage, decisions x have to be made before some random parameters ω are known, whereas in the second stage, decisions y are made after the realizations of these random parameters ω are revealed. We assume that the probability distribution of ω is known, with F denoting the cumulative distribution function and Ω the support of ω. The MISPs that we consider are defined as

$$\eta^* := \min_{x,z} \Big\{ c^\top x + Q(z) : \; Ax = b, \; z = Tx, \; x \in X \Big\}, \tag{3.1}$$

where $z = Tx \in \mathbb{R}^m$ represent tender variables. Moreover, the *expected value function* Q represents the expected second-stage costs

$$Q(z) := \mathbb{E}_\omega[v(\omega, z)], \quad z \in \mathbb{R}^m, \tag{3.2}$$

where the *second-stage value function* v is defined as

$$v(\omega, z) := \min_y \Big\{ q^\top y : \; Wy = \omega - z, \; y \in Y \Big\}, \quad \omega \in \Omega, \, z \in \mathbb{R}^m. \tag{3.3}$$

The second-stage decisions y are also called *recourse actions*. Indeed, if $Tx = \omega$ represents random goal constraints, then the second-stage optimization problem v models all possible recourse actions y, and their corresponding costs, to compensate for infeasibilities of these goal constraints. Observe that we only consider randomness in the right-hand side of these goal constraints. Moreover, we assume that at least some of the second-stage decision variables y_i are restricted to be in-

teger. This is captured by the feasible regions $X \subset \mathbb{R}^{n_1}_+$ and $Y \subset \mathbb{R}^{n_2}_+$ that may impose integrality restrictions on the first- and second-stage decision variables, respectively.

Throughout this chapter we make the following assumptions. The first is often referred to as the *complete recourse* assumption, meaning that there always exists a feasible recourse action y, ensuring that $v(\omega, z) < +\infty$ for all $\omega \in \Omega$ and $z \in \mathbb{R}^m$. The second is equivalent to the dual feasible region of the LP-relaxation of v being non-empty, implying that $v(\omega, z) > -\infty$ for all $\omega \in \Omega$ and $z \in \mathbb{R}^m$. Together with the third assumption, these assumptions guarantee that $Q(z)$ is finite for every $z \in \mathbb{R}^m$.

Assumption 3.1. We assume that

- there exists $y \in Y$ such that $Wy = \omega - z$ for every $\omega \in \Omega$ and $z \in \mathbb{R}^m$,
- there exists $\lambda \in \mathbb{R}^m$ such that $W^\top \lambda \leq q$, and
- $\mathbb{E}_\omega[|\omega_i|] < +\infty$, for all $i = 1, \ldots, m$.

3.2.2 Novel solution approach: pseudo-valid cutting planes

To solve the MISP defined in (3.1), we propose to relax the integrality restrictions on the second-stage decision variables y and to add *pseudo-valid* cutting planes to the second-stage feasible region

$$Y(\omega, z) := \Big\{ y \in Y : \; Wy = \omega - z \Big\}.$$

In particular, we assume that the cutting planes are of the form $\hat{W}(\omega)y \geq \hat{h}(\omega) - \hat{T}(\omega)z$, so that they are affine in the tender variables z.

Definition 3.1. Consider the second-stage value function v defined in (3.3). Then, we call $\hat{v}$ an *affine cutting plane approximation* of v if it is of the form

$$\hat{v}(\omega, z) = \min_y \{ q^\top y : Wy = \omega - z, \\ \hat{W}(\omega)y \geq \hat{h}(\omega) - \hat{T}(\omega)z, \; y \in \mathbb{R}^{n_2}_+ \}, \quad \omega \in \Omega, \, z \in \mathbb{R}^m.$$

Moreover, we define the affine cutting plane approximation $\hat{Q}$ of the expected value function Q, defined in (3.2), as $\hat{Q}(z) := \mathbb{E}_\omega[\hat{v}(\omega, z)]$, $z \in \mathbb{R}^m$.

The main reason we use cutting planes that are *affine in* z is that the approximating value function $\hat{v}(\omega, z)$ with feasible region

$$\hat{Y}(\omega, z) := \left\{ y \in \mathbb{R}^{n_2}_+ : \begin{array}{l} Wy = \omega - z \\ \hat{W}(\omega)y \geq \hat{h}(\omega) - \hat{T}(\omega)z \end{array} \right\}$$

is *convex* in z for every fixed $\omega \in \Omega$, and thus the corresponding approximating expected value function $\hat{Q}$ is convex. This means that the MISP in (3.1) with Q replaced by $\hat{Q}$ is much easier to solve than the original MISP.

Observation 1. Consider the affine cutting plane approximations $\hat{v}$ and $\hat{Q}$ of Definition 3.1. Then, $\hat{Q}$ is convex, and $\hat{v}(\omega, z)$ is convex in z for every fixed $\omega \in \Omega$.

Obviously, we cannot expect to obtain a good approximation $\hat{Q}$ if we arbitrarily add invalid affine cutting planes. That is why we restrict our attention to so-called *pseudo-valid cutting planes*. Such cutting planes are valid on a grid of points z, but may be invalid for other values of z.

Definition 3.2. Consider the second-stage value function v defined in (3.3). Then, the cutting planes $\hat{W}(\omega)y \geq \hat{h}(\omega) - \hat{T}(\omega)z$ are called *pseudo-valid* for v if there exist $\alpha \in \mathbb{R}^m$ and $\beta \in \mathbb{Z}^m$ such that for all $z \in \alpha + \beta\mathbb{Z}^m$ and for all $\omega \in \Omega$,

$$\left\{ y \in \mathbb{Z}^{p_2}_+ \times \mathbb{R}^{n_2-p_2}_+ : Wy = \omega - z \right\} \subset \left\{ y \in \mathbb{R}^{n_2}_+ : Wy = \omega - z, \right. \\ \left. \hat{W}(\omega)y \geq \hat{h}(\omega) - \hat{T}(\omega)z \right\}. \quad (3.4)$$

If, in addition, $\hat{v}(\omega, z) = v(\omega, z)$ for all $z \in \alpha + \beta\mathbb{Z}^m$ and for all $\omega \in \Omega$, then we call the pseudo-valid cutting planes *tight*.

Remark 3.1. With slight abuse of notation we use $\alpha + \beta\mathbb{Z}^m$ to represent the grid of points

$$\alpha + \beta\mathbb{Z}^m := \left\{ (\alpha_1 + \beta_1 l_1, \ldots, \alpha_m + \beta_m l_m) : \; l \in \mathbb{Z}^m \right\}.$$

There are different classes of exact cutting planes that can be applied as pseudo-valid cutting planes to two-stage MISPs. For example, in Section 3.5 we derive pseudo-valid mixed-integer Gomory cuts and tight pseudo-valid cutting planes for a nurse scheduling problem. However, the main focus of this chapter is not on how to obtain the pseudo-valid cuts. Instead, we assume that they are given or can be

iteratively generated by an algorithm, and we consider the performance of using such pseudo-valid cutting planes.

The performance of these pseudo-valid cutting planes may be surprisingly good, even if they cut away feasible integer second-stage solutions or admit second-stage solutions outside the integer hull $\bar{Y}(\omega, z)$ of the second-stage feasible region $Y(\omega, z)$; in these cases, $\hat{v}(\omega, z)$ may significantly over- or underestimate $v(\omega, z)$, respectively. However, to obtain good first-stage decisions x, we do not require $\hat{v}(\omega, z)$ to be a good approximation of $v(\omega, z)$ for *every* $\omega \in \Omega$ and $z \in \mathbb{R}^m$, but merely require $\hat{v}(\omega, z)$ to be a good approximation of $v(\omega, z)$ *on average* for every $z \in \mathbb{R}^m$. This explains why applying pseudo-valid cutting planes may work for stochastic MIPs but not for deterministic MIPs.

Using a one-dimensional example, we illustrate a class of pseudo-valid cutting planes.

Example 3.1. Consider a special case of the second-stage value function defined in (3.3), given by

$$\begin{aligned} v(\omega, z) = \min_{y, u_1, u_2} \quad & qy + ru_1 + ru_2 \\ \text{s.t.} \quad & y - u_1 + u_2 = \omega - z \\ & y \in \mathbb{Z}_+,\ u_1, u_2 \in \mathbb{R}_+, \end{aligned} \tag{3.5}$$

where $0 < q < r$. By rewriting the equality in (3.5) as $u_2 = \omega - z - y + u_1$, we can eliminate the variable u_2 from the second-stage value function to obtain

$$\begin{aligned} v(\omega, z) = r(\omega - z) + \min_{y, u_1} \quad & (q - r)y + 2ru_1 \\ \text{s.t.} \quad & y - u_1 \leq \omega - z \\ & y \in \mathbb{Z}_+,\ u_1 \in \mathbb{R}_+. \end{aligned} \tag{3.6}$$

Since the minimization problem in (3.6) only has two decision variables, y and u_1, we can graphically depict its feasible region $Y(\omega, z)$. The left panel in Figure 3.1 shows this feasible region for $\omega = 2.5$ and $z = 1$, and also depicts the feasible region of the LP-relaxation of $v(\omega, z)$. Clearly, the latter is larger than the integer hull $\bar{Y}(\omega, z)$ of $Y(\omega, z)$.

It is well known that the integer hull $\bar{Y}(\omega, z)$ can be obtained by adding a mixed-integer rounding (MIR) inequality, so that for every $\omega \in \Omega$ and $z \in \mathbb{R}$, the integer

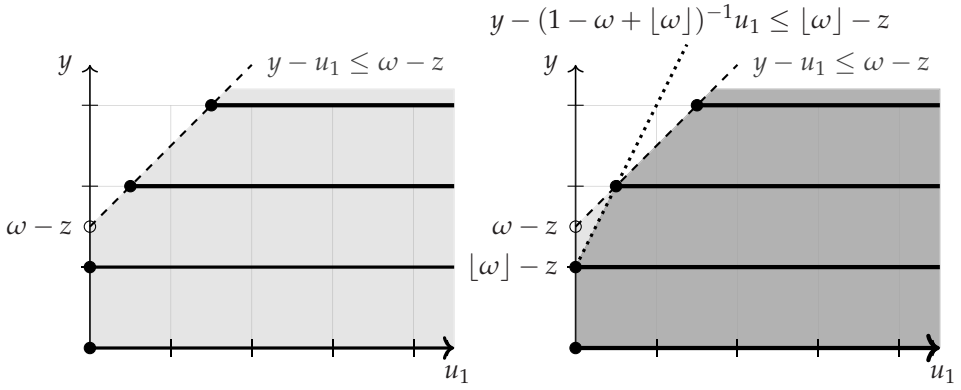


Figure 3.1. Illustration of the feasible region of $v(\omega, z)$ of Example 3.1 with $\omega = 2.5$ and $z = 1$. The feasible region $Y(\omega, z)$ is represented by the black dots and the thick black lines. In the left panel the shaded region corresponds to the feasible region of the LP-relaxation of v, whereas in the right panel, the MIR inequality is added, and the dark shaded region represents the integer hull $\bar{Y}(\omega, z)$ of the feasible region $Y(\omega, z)$ of $v(\omega, z)$.

hull $\bar{Y}(\omega, z)$ equals

$$\bar{Y}(\omega, z) := \Big\{(y, u_1) \in \mathbb{R}^2_+ : y - u_1 \leq \omega - z, \\ y - \frac{1}{1 - (\omega - z) + \lfloor \omega - z \rfloor} u_1 \leq \lfloor \omega - z \rfloor \Big\}.$$

The right panel in Figure 3.1 shows $\bar{Y}(\omega, z)$ and this MIR inequality.

Observe that the MIR inequality is not affine in z, which means that it will be hard to use for optimization purposes. However, if $z \in \mathbb{Z}$, then it reduces to

$$y - \frac{1}{1 - \omega + \lfloor \omega \rfloor} u_1 \leq \lfloor \omega \rfloor - z, \tag{3.7}$$

which means it is of the form of the affine cutting planes in Definition 3.1. Thus, a natural idea is to use the cutting planes in (3.7), also when $z \notin \mathbb{Z}$. In the latter case the cutting planes are not always valid. However, it is not hard to show that these cutting planes are *pseudo-valid* and *tight*.

Figure 3.2 shows the approximating feasible region

$$\hat{Y}(\omega, z) = \Big\{(y, u_1) \in \mathbb{R}^2_+ : \; y - u_1 \leq \omega - z, \; y - \frac{1}{1 - \omega + \lfloor \omega \rfloor} u_1 \leq \lfloor \omega \rfloor - z \Big\},$$

for $z = 0.5$ and $\omega = 1.5, 1.75, 2, 2.25$. We observe that for $\omega = 2$, the approximating

MIR inequality coincides with the constraint $y - u \leq \omega - z$, so that $\hat{Y}(\omega, z)$ is equal to the feasible region of the LP-relaxation of $v(\omega, z)$, and thus admits solutions outside the integer hull $\bar{Y}(\omega, z)$. For $\omega = 1.5$, on the other hand, the approximating MIR inequality cuts away feasible integer solutions. For $\omega = 1.75$ and $\omega = 2.25$ we see a combination of both.

In Section 3.5.2 we numerically assess the performance of the pseudo-valid cutting plane approximation

$$\hat{v}(\omega, z) := r(\omega - z) + \min_{y,u_1} \Big\{ (q - r)y + 2ru_1 : \ (y, u_1) \in \hat{Y}(\omega, z) \Big\}, \quad \omega \in \Omega,\, z \in \mathbb{R}, \quad (3.8)$$

and show that for a normally distributed random variable $\omega \sim N(\mu, \sigma^2)$, $\hat{Q}$ is a good approximation of Q for medium to large values of the standard deviation σ. $\Diamond$

3.3 Pseudo-valid cutting planes for simple integer recourse models

In this section we show that existing convex approximations for simple integer recourse (SIR) models can be interpreted as pseudo-valid cutting plane approximations. SIR models are introduced by Louveaux and Van der Vlerk [45], and can be considered the most simple version of a MISP as defined in (3.1). For ease of exposition, we consider here the one-sided and one-dimensional version of SIR, where the second-stage value function v is defined as

$$v(\omega, z) = \min_{y} \Big\{ qy : \ y \geq \omega - z, \ y \in \mathbb{Z}_+ \Big\}, \quad \omega \in \Omega,\, z \in \mathbb{R}.$$

Observe that we can derive a closed-form expression for v since for every $\omega \in \Omega$ and $z \in \mathbb{R}$, the optimal solution is $y^* = \lceil \omega - z \rceil^+ := \max\{0, \lceil \omega - z \rceil\}$, and thus $v(\omega, z) = q \lceil \omega - z \rceil^+$. Clearly, $v(\omega, z)$ is a non-convex function of z because of the round-up operator.

We, however, focus on the feasible region $Y(\omega, z) = \{y \in \mathbb{Z}_+ : y \geq \omega - z\}$ and its integer hull

$$\bar{Y}(\omega, z) = \Big\{ y \in \mathbb{R}_+ : y \geq \lceil \omega - z \rceil \Big\}, \quad \omega \in \Omega,\, z \in \mathbb{R}.$$

Here, the cutting plane $y \geq \lceil \omega - z \rceil$ makes the original constraint $y \geq \omega - z$ re-

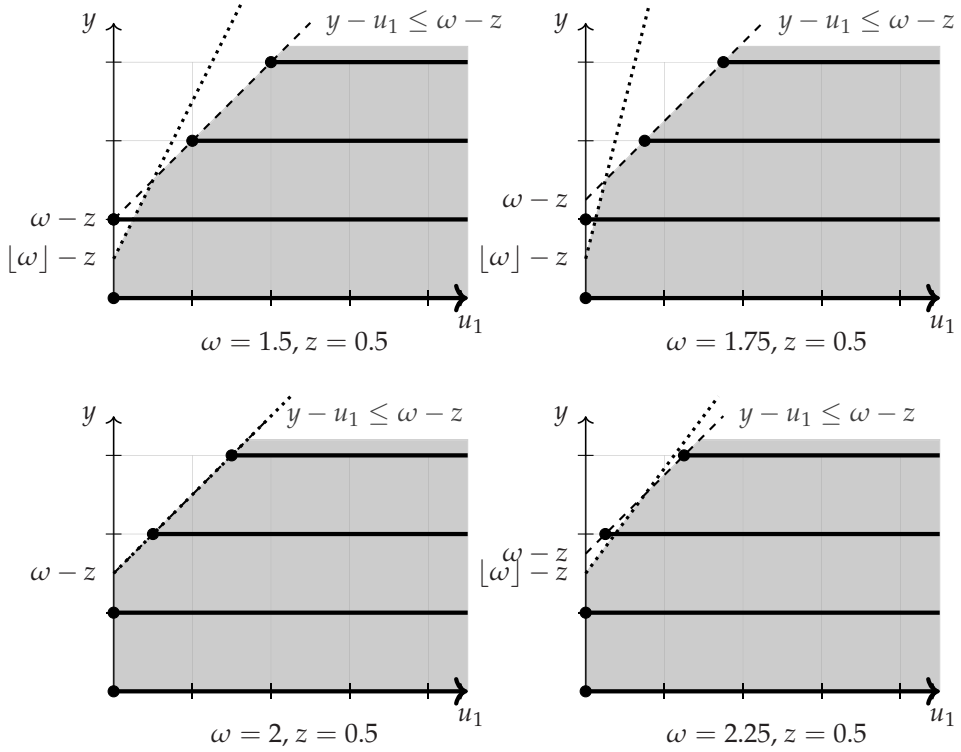


Figure 3.2. Illustration of the feasible region of $v(\omega, z)$ of Example 3.1 with $z = 0.5$ and $\omega = 1.5, 1.75, 2$, and 2.25. The feasible region $Y(\omega, z)$ is represented by the black dots and the thick black lines. The dotted line represents the pseudo-valid MIR inequality defined in (3.7), and the shaded regions the approximating feasible region $\hat{Y}(\omega, z)$.

dundant. Similar to Example 3.1, this exact cutting plane is not affine in z and thus not suitable for optimization purposes. However, if $z \in \mathbb{Z}$, then the cutting plane is equivalent to $y \geq \lceil \omega \rceil - z$, which we can use as a pseudo-valid cutting plane for $z \notin \mathbb{Z}$. In fact, $y \geq \lceil \omega \rceil - z$ is a *tight* pseudo-valid cutting plane. We define a family of tight pseudo-valid cutting plane approximations $\hat{v}_\alpha$, each of them using the cutting plane $y \geq \lceil \omega - \alpha \rceil + \alpha - z$ that is exact for $z \in \alpha + \mathbb{Z}$.

Definition 3.3. For every $\alpha \in \mathbb{R}$, define the pseudo-valid cutting plane approximation $\hat{v}_\alpha$ for the SIR second-stage value function v as

$$\begin{aligned}\hat{v}_\alpha(\omega, z) &= \min_y \left\{ qy : \; y \geq \lceil \omega - \alpha \rceil + \alpha - z, \; y \in \mathbb{R}_+ \right\} \\ &= q\Big(\lceil \omega - \alpha \rceil + \alpha - z \Big)^+, \quad \omega \in \Omega, z \in \mathbb{R}.\end{aligned}$$

Moreover, define the corresponding pseudo-valid cutting plane approximation $\hat{Q}_\alpha$ for the SIR expected value function Q as $\hat{Q}_\alpha(z) = q\mathbb{E}_\omega[(\lceil \omega - \alpha \rceil + \alpha - z)^+], z \in \mathbb{R}$.

Surprisingly, the pseudo-valid cutting plane approximation $\hat{Q}_\alpha$ equals the α-approximations of Klein Haneveld et al. [41], derived from a completely different perspective. They first identify all probability distributions of ω for which the expected value function Q is convex. This turns out to be all continuous distributions with probability density function f satisfying $f(s) = G(s+1) - G(s)$, $s \in \mathbb{R}$, for some cumulative distribution function G with finite mean. For all other distributions, they use this condition to generate an approximating density function $\hat{f}$, resulting in a convex approximation $\hat{Q}$ of Q. Selecting $G(s+1) = F(\lceil s - \alpha \rceil + \alpha)$, $s \in \mathbb{R}$, yields the α-approximation $\hat{Q}_\alpha(z) := q\mathbb{E}_\omega[(\lceil \omega - \alpha \rceil + \alpha - z)^+]$, $z \in \mathbb{R}$, equivalent to the pseudo-valid cutting plane approximation of Definition 3.3.

In this section, we reinterpret $\hat{Q}_\alpha$ as a pseudo-valid cutting plane approximation, connecting the convex approximation solution philosophy, introduced by Van der Vlerk [80] and continued in among others [41, 59, 60, 61, 82, 83] and in Chapter 2, with exact cutting plane techniques for MISPs, studied in, e.g., [13, 20, 25, 31, 35, 50, 53, 70, 73, 88]. This is particularly relevant, since performance guarantees are available for using convex approximations that may be used for pseudo-valid cutting plane approximations. In fact, for SIR models, Romeijnders et al. [59] derive an upper bound on $\|Q - \hat{Q}_\alpha\|_\infty := \sup_{z \in \mathbb{R}} |Q(z) - \hat{Q}_\alpha(z)|$ for every $\alpha \in \mathbb{R}$, that depends on the *total variation* of the probability density function f of the random variable ω.

Definition 3.4. Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function and let $I \subset \mathbb{R}$ be an interval. Let $\Pi(I)$ denote the set of all finite ordered sets $P = \{x_1, \ldots, x_{N+1}\}$ with $x_1 < \cdots < x_{N+1}$ in I. Then, the *total variation* of f on I, denoted $|\Delta|f(I)$, is defined as

$$|\Delta|f(I) = \sup_{P \in \Pi(I)} V_f(P),$$

where $V_f(P) = \sum_{i=1}^{N} |f(x_{i+1}) - f(x_i)|$. We write $|\Delta|f := |\Delta|f(\mathbb{R})$.

Theorem 3.1. *Consider the SIR expected value function $Q(z) = q\mathbb{E}_\omega[\lceil \omega - z \rceil^+]$, $z \in \mathbb{R}$, and its pseudo-valid cutting plane approximation $\hat{Q}_\alpha(z) = q\mathbb{E}_\omega[(\lceil \omega - \alpha \rceil - \alpha - z)^+]$, $z \in \mathbb{R}$, for $\alpha \in \mathbb{R}$. Then, for every continuous random variable ω with probability density function f, we have*

$$\|Q - \hat{Q}_\alpha\|_\infty \leq qh(|\Delta|f),$$

where $h : [0, \infty) \mapsto \mathbb{R}$ is defined as

$$h(|\Delta|f) = \begin{cases} |\Delta|f/8, & |\Delta|f \leq 4, \\ 1 - 2/|\Delta|f, & |\Delta|f \geq 4. \end{cases}$$

Proof. See Romeijnders et al. [59]. ■

Remark 3.2. Romeijnders et. al [59] actually derive an error bound for the α-approximations of Van der Vlerk [83] for integer recourse models with totally unimodular recourse matrix W. Also these α-approximations can be interpreted as tight pseudo-valid cutting plane approximations.

Theorem 3.1 shows that the error bound is smaller for lower values of the total variation $|\Delta|f$ of the probability density function f of ω. For unimodal density functions, such as the normal density function in Example 3.2 below, this total variation $|\Delta|f$ is small if the variance of the random variable ω is large. Thus, in general we conclude from Theorem 3.1 that the larger the variability of the random variable in the model, the better the pseudo-valid cutting plane approximation.

Example 3.2. Let ω be a normal random variable with mean μ and standard deviation σ. Then, the probability density function f of ω is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big\{ - \frac{(x-\mu)^2}{2\sigma^2} \Big\}, \quad x \in \mathbb{R},$$

which is unimodal with mode μ, and thus has total variation $|\Delta|f = 2f(\mu) = \sigma^{-1}\sqrt{2/\pi}$. Hence, if the standard deviation σ increases, then the total variation $|\Delta|f$ of f decreases, and thus the upper bound on $\|Q - \hat{Q}_\alpha\|_\infty$ in Theorem 3.1 decreases. In other words, if the standard deviation is large, then $\hat{Q}_\alpha$ is a close approximation of Q, and thus the resulting approximating first-stage decision $\hat{x}_\alpha$ will be good. $\Diamond$

3.4 Pseudo-valid cutting plane approximations for general MISPs

In this section we consider tight pseudo-valid cutting plane approximations $\hat{Q}$ for general MISPs, and we derive an upper bound on $\|Q - \hat{Q}\|_\infty$ for the case that the random right-hand side vector is continuously distributed . We use $\|Q - \hat{Q}\|_\infty$ to measure the error of the approximation $\hat{Q}$, since it can be used to bound the optimality gap, $c^\top \hat{x} + Q(\hat{x}) - \eta^*$, of the approximating solution $\hat{x}$ obtained by solving (3.1) with Q replaced by $\hat{Q}$. Indeed, Romeijnders et al. [59] show that for any approximation $\hat{Q}$, we have

$$c^\top \hat{x} + Q(\hat{x}) - \eta^* \leq 2\|Q - \hat{Q}\|_\infty.$$

We prove that for tight pseudo-valid cutting plane approximations $\hat{Q}$, the error $\|Q - \hat{Q}\|_\infty$ vanishes if the total variations of the one-dimensional conditional probability density functions of the random vector ω in the model go to zero. For example, for normally distributed ω this means that the solutions obtained by using tight pseudo-valid cutting planes are good if the variance of ω is large enough. The final result is Theorem 3.2, which is conveniently stated here below. If the pseudo-valid cuts are not tight, then we use $\hat{Q}$ to derive a lower bound for Q in Corollary 3.1.

Definition 3.5. For every $i = 1, \ldots, m$ and $t \in \mathbb{R}^m$, we let $t_{-i} \in \mathbb{R}^{m-1}$ denote the vector t without its i-th component.

Definition 3.6. For every $i = 1, \ldots, m$ and $t_{-i} \in \mathbb{R}^{m-1}$, define the i-th *conditional density function* $f_i(\cdot|t_{-i})$ of the m-dimensional joint pdf f as

$$f_i(t_i|t_{-i}) = \begin{cases} \dfrac{f(t)}{f_{-i}(t_{-i})}, & f_{-i}(t_{-i}) > 0, \\ 0, & f_{-i}(t_{-i}) = 0, \end{cases}$$

where f_{-i} represents the joint density function of ω_{-i}, the random vector obtained by removing the i-th element of ω.

Definition 3.7. Let $\mathcal{H}^m$ denote the set of all m-dimensional joint pdfs f whose conditional density functions $f_i(\cdot|t_{-i})$ are of bounded variation.

Theorem 3.2. *Consider the mixed-integer recourse function Q and its* tight pseudo-valid cutting plane approximation $\hat{Q}$. *Then, under Assumption 3.1, there exists a constant $C \in \mathbb{R}$ with $C > 0$ such that for all ω with pdf $f \in \mathcal{H}^m$,*

$$\|Q - \hat{Q}\|_\infty \leq C \sum_{i=1}^{m} \mathbb{E}_{\omega_{-i}} \Big[|\Delta| f_i(\cdot|\omega_{-i}) \Big].$$

If all components of the random vector ω are independent, then $|\Delta| f_i(\cdot|\omega_{-i}) = |\Delta| f_i$ for every $i = 1, \ldots, m$ and $\omega_{-i} \in \mathbb{R}^{m-1}$, and thus the error bound in Theorem 3.2 reduces to $C \sum_{i=1}^{m} |\Delta| f_i$, where $|\Delta| f_i$ is the total variation of the *marginal* probability density function f_i of ω_i, $i = 1, \ldots, m$. Hence, if the components of ω are independent and normally distributed, then the error bound in Theorem 3.2 is small if the standard deviations of the components of ω are large, see Example 3.2.

In the next corollary, we derive a lower bound for Q that holds irrespective of whether the pseudo-valid cuts are tight or not.

Corollary 3.1. *Consider the mixed-integer recourse function Q and its* pseudo-valid cutting plane approximation $\hat{Q}$. *Then, under Assumption 3.1, there exists a constant $C \in \mathbb{R}$ with $C > 0$ such that for all ω with pdf $f \in \mathcal{H}^m$ and $z \in \mathbb{R}^m$,*

$$Q(z) \geq \hat{Q}(z) - C \sum_{i=1}^{m} \mathbb{E}_{\omega_{-i}} \Big[|\Delta| f_i(\cdot|\omega_{-i}) \Big].$$

Since $\hat{Q}$ is typically strictly larger than the LP-relaxation Q_{LP} of Q, it follows that Corollary 3.1 defines a better lower bound on Q than Q_{LP}, at least if the second right-hand side term vanishes. Such a better lower bound may be useful in, e.g., a branch-and-bound algorithm if some of the first-stage decision variables are integer or binary.

Remark 3.3. Theorem 3.2 and Corollary 3.1 provide a performance guarantee for pseudo-valid cutting plane approximations of MISPs with continuous distributions only. However, for practical computations these continuous distributions are typically discretized, e.g., using a sample average approximation (SAA) [42], to deal

with the high-dimensional integrals in the expected value function. Since $\hat{Q}$ is essentially a *continuous* stochastic program, the additional error that we incur by discretizing continuous distributions is thus the same as for standard continuous stochastic programs with continuous distributions. Moreover, if the discretization is fine enough, then the discretized version of $\hat{Q}$ will be sufficiently close to the original pseudo-valid cutting plane approximation $\hat{Q}$ with continuous distributions [37].

The proofs of Theorem 3.2 and Corollary 3.1 are postponed to Section 3.4.4. First, however, we discuss preliminary results required for these proofs. In particular, in Section 3.4.1 we review properties of the mixed-integer value function $v(\omega, z)$, in Section 3.4.2 we show that a tight pseudo-valid cutting plane approximation $\hat{v}(\omega, z)$ is affine in z on parts of its domain, and in Section 3.4.3 we derive bounds on $\hat{v}$. The proofs of our auxiliary lemmas and propositions in these sections are postponed to Appendix 3.A.

3.4.1 Properties of mixed-integer value functions

Let B be a dual feasible basis matrix of the LP-relaxation v_{LP} of v. Then, we can rewrite v_{LP} as

$$\begin{aligned} v_{LP}(\omega, z) = \min_{y_B, y_N} \quad & q_B^\top y_B + q_N^\top y_N \\ \text{s.t.} \quad & B y_B + N y_N = \omega - z \\ & y_B \in \mathbb{R}^m_+, \; y_N \in \mathbb{R}^{n_2 - m}_+, \end{aligned} \tag{3.9}$$

where y_B denote the basic variables and y_N the non-basic variables. Using the equality in (3.9) to solve for the basic variables y_B, we obtain the equivalent representation

$$\begin{aligned} v_{LP}(\omega, z) = q_B^\top B^{-1}(\omega - z) + \min_{y_N} \quad & \bar{q}_N^\top y_N \\ \text{s.t.} \quad & B^{-1}(\omega - z) - B^{-1} N y_N \geq 0 \\ & y_N \in \mathbb{R}^{n_2 - m}_+, \end{aligned} \tag{3.10}$$

with reduced costs $\bar{q}_N^\top := q_N^\top - q_B^\top B^{-1} N \geq 0$. Obviously, it is optimal to select the non-basic variables y_N equal to zero in the minimization problem in (3.10) if $B^{-1}(\omega - z) \geq 0$. The latter condition can conveniently be rewritten as $\omega - z \in \Lambda$, where the simplicial cone Λ is defined as $\Lambda := \{t \in \mathbb{R}^m : B^{-1} t \geq 0\}$. Thus, if

$\omega - z \in \Lambda$, then

$$v_{LP}(\omega, z) = q_B^\top B^{-1}(\omega - z).$$

This result holds for every dual feasible basis matrix B. In fact, the basis decomposition theorem of Walkup and Wets [85] shows that there exist basis matrices B_k and corresponding simplicial cones $\Lambda^k := \{t \in \mathbb{R}^m : B_k^{-1} t \geq 0\}$, $k = 1, \ldots, K$, such that these cones Λ^k cover $\mathbb{R}^m$, the interiors of these cones Λ^k are mutually disjoint, and $v_{LP}(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z)$ for $\omega - z \in \Lambda^k$ for every $k = 1, \ldots, K$.

Romeijnders et al. [60] prove a similar result for the *mixed-integer* value function v, involving the same basis matrices B_k and simplicial cones Λ^k, $k = 1, \ldots, K$. They show that there exist distances $d_k \geq 0$ such that if $\omega - z \in \Lambda^k$ and $\omega - z$ has at least Euclidean distance d_k to the boundary of Λ^k, then

$$v(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - z),$$

where ψ^k is a B_k-periodic function, see Definition 3.8 below. The first term is the same as the LP-relaxation v_{LP}, and thus the second term can be interpreted as the additional costs of having integer variables instead of continuous ones. Theorem 3.3 summarizes these results.

Definition 3.8. Let $B \in \mathbb{Z}^{m \times m}$ be an integer matrix. Then, a function $\psi : \mathbb{R}^m \mapsto \mathbb{R}$ is called B-periodic if and only if $\psi(z) = \psi(z + Bl)$ for every $z \in \mathbb{R}^m$ and $l \in \mathbb{Z}^m$.

Definition 3.9. Let $\Lambda \subset \mathbb{R}^m$ be a closed convex cone and let $d \in \mathbb{R}$ with $d > 0$ be given. Then, we define $\Lambda(d)$ as

$$\Lambda(d) := \Big\{ s \in \Lambda : \mathcal{B}(s, d) \subset \Lambda \Big\},$$

where $\mathcal{B}(s, d) := \{t \in \mathbb{R}^m : \|t - s\| \leq d\}$ is the closed ball centered at s with radius d. We can interpret $\Lambda(d)$ as the set of points in Λ with at least Euclidean distance d to the boundary of Λ.

Theorem 3.3. *Consider the mixed-integer value function*

$$v(\omega, z) = \min \Big\{ q^\top y : Wy = \omega - z,\ y \in \mathbb{Z}_+^{p_2} \times \mathbb{R}_+^{n_2 - p_2} \Big\}, \quad z \in \mathbb{R}^m,$$

where W is an integer matrix, and $v(\omega, z)$ is finite for all $\omega \in \Omega$ and $z \in \mathbb{R}^m$ by Assumption 3.1. Then, there exist dual feasible basis matrices B_k of v_{LP}, $k = 1, \ldots, K$, simplicial

cones $\Lambda^k := \{t \in \mathbb{R}^m : B_k^{-1}t \geq 0\}$*, distances* $d_k \geq 0$*, and bounded* B_k*-periodic functions* ψ^k *such that*

- $\bigcup_{k=1}^{K} \Lambda^k = \mathbb{R}^m$,
- $(\operatorname{int}\Lambda^k) \cap (\operatorname{int}\Lambda^l) = \emptyset$ *for every* $k, l \in \{1, \ldots, K\}$ *with* $k \neq l$*, and*
- $v(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - z)$ *for every* $\omega - z \in \Lambda^k(d_k)$.

Proof. See [60]. ■

3.4.2 Linearity regions of pseudo-valid cutting plane approximations

Consider a tight pseudo-valid cutting plane approximation $\hat{v}(\omega, z)$. Moreover, let $k = 1, \ldots, K$ be given and consider a fixed $\omega \in \Omega$. Theorem 3.3 shows that for all $z \in \mathbb{R}^m$ with $\omega - z \in \Lambda^k(d_k)$, i.e., for all $z \in \omega - \Lambda^k(d_k)$, the mixed-integer value function v is given by

$$v(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - z).$$

Since ψ^k is B_k-periodic there exist values of β for which $\psi^k(\omega - z) = \psi^k(\omega - \alpha)$ for all $z \in \alpha + \beta\mathbb{Z}^m$; see the proof of Proposition 3.1. Let β be such a value. Then, for all $z \in \omega - \Lambda^k(d_k)$ and $z \in \alpha + \beta\mathbb{Z}^m$, we have $\hat{v}(\omega, z) = v(\omega, z)$, and thus the tight pseudo-valid cutting plane approximation $\hat{v}$ equals

$$\hat{v}(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha). \tag{3.11}$$

Thus, for a fixed $\omega \in \Omega$, the tight pseudo-valid cutting plane approximation $\hat{v}(\omega, z)$ is affine in z over a grid of points in $\omega - \Lambda^k(d_k)$. Since $\hat{v}(\omega, z)$ is convex in z, we intuitively expect $\hat{v}(\omega, z)$ to satisfy (3.11) for points outside the grid in $\omega - \Lambda^k(d_k)$ as well. Lemma 3.1 confirms our intuition.

Lemma 3.1. *Let* $v : \mathbb{R}^m \mapsto \mathbb{R}$ *be a convex function and let* $C \subset \mathbb{R}^m$ *be a compact convex set with extreme points* $z^j \in C$*,* $j = 1, \ldots, J$*, and interior point* $z^0 \in C$*. Suppose that there exist* $a \in \mathbb{R}^m$ *and* $b \in \mathbb{R}$ *such that* $v(z^j) = a^\top z^j + b$ *for all* $j = 0, \ldots, J$*. Then,* $v(z) = a^\top z + b$ *for all* $z \in C$.

To apply Lemma 3.1 to $\hat{v}(\omega, z)$ we introduce hyperrectangles $C^l(\alpha, \beta)$ that have extreme points on the grid $\alpha + \beta\mathbb{Z}^m$.

Definition 3.10. Let $\alpha \in \mathbb{R}^m$ and $\beta \in \mathbb{R}^m$ be given. For every $l \in \mathbb{Z}^m$, we define the hyperrectangle $C^l(\alpha, \beta)$ as

$$C^l(\alpha, \beta) := \prod_{i=1}^{m} \Big[\alpha_i + \beta_i(l_i - 1), \alpha_i + \beta_i(l_i + 1)\Big].$$

For every value of $\alpha, \beta \in \mathbb{R}^m$ and $l \in \mathbb{Z}^m$, the hyperrectangle $C^l(\alpha, \beta) \subset \mathbb{R}^m$ is convex. Moreover, all its extreme points and the interior point $(\alpha_1 + \beta_1 l_1, \ldots, \alpha_m + \beta_m l_m)$ are on the grid $\alpha + \beta\mathbb{Z}^m$. Thus, if $C^l(\alpha, \beta) \subset \omega - \Lambda^k(d_k)$, then we can apply Lemma 3.1 to $\hat{v}(\omega, \cdot)$ with $C := C^l(\alpha, \beta)$ to conclude that $\hat{v}(\omega, z)$ satisfies (3.11) for all $z \in C^l(\alpha, \beta)$, and thus $\hat{v}(\omega, z)$ is affine in z over $C^l(\alpha, \beta)$. Applying Lemma 3.1 for all $C^l(\alpha, \beta)$ that are completely contained in $\omega - \Lambda^k(d_k)$, we can show that $\hat{v}(\omega, z)$ is affine in z over at least $\omega - \Lambda^k(d_k + 2\|\beta\|)$. This is true since the *diameter* of $C^l(\alpha, \beta)$ is $2\|\beta\|$, and $\Lambda^k(d_k + 2\|\beta\|)$ represents all points in Λ^k with at least Euclidean distance $d_k + 2\|\beta\|$ to the boundary of Λ^k. Thus, for every $z \in \omega - \Lambda^k(d_k + 2\|\beta\|)$ there exists a hyperrectangle $C^l(\alpha, \beta) \subset \omega - \Lambda^k(d_k)$ that contains z. Here, the diameter of $C^l(\alpha, \beta)$ is defined as

$$\max_{z_1, z_2} \Big\{ \|z_1 - z_2\| : z_1, z_2 \in C^l(\alpha, \beta) \Big\} = 2\|\beta\|.$$

Proposition 3.1 shows all *linearity regions* of $\hat{v}(\omega, z)$ for fixed $\omega \in \Omega$. These are subsets of the domain of $\hat{v}(\omega, \cdot)$ on which $\hat{v}(\omega, z)$ is affine in z.

Proposition 3.1. *Consider a tight pseudo-valid cutting plane approximation $\hat{v}(\omega, z)$ as defined in Definition 3.2, and let Λ^k, $k = 1, \ldots, K$, denote the simplicial cones from Theorem 3.3. Then, under Assumption 3.1, for every $k = 1, \ldots, K$, there exists a distance $d'_k \geq 0$ such that if $\omega - z \in \Lambda^k(d'_k)$, then*

$$\hat{v}(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha). \tag{3.12}$$

3.4.3 Bounds on the value function of a pseudo-valid cutting plane approximation

Proposition 3.1 defines $\hat{v}(\omega, z)$ on the linearity regions $\Lambda^k(d'_k)$. In fact, on these linearity regions, $v(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - z)$ and $\hat{v}(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega -$

$z) + \psi^k(\omega - \alpha)$, so that the difference between the two equals

$$v(\omega, z) - \hat{v}(\omega, z) = \psi^k(\omega - z) - \psi^k(\omega - \alpha), \quad z \in \omega - \Lambda^k(d'_k).$$

This difference is B_k-periodic and bounded, since ψ^k is a bounded B_k-periodic function by Theorem 3.3. These properties will be exploited to derive an error bound for tight pseudo-valid cutting plane approximations $\hat{Q}$ in Section 3.4.4.

Outside the linearity regions, i.e., on $\mathcal{N} := \mathbb{R}^m \setminus \bigcup_{k=1}^{K} \Lambda^k(d'_k)$, we cannot prove such properties for $v(\omega, z)$ and $\hat{v}(\omega, z)$. However, we can show that the difference between the two is bounded. That is, there exists $R \in \mathbb{R}$ such that

$$\|v - \hat{v}\|_\infty := \sup_{\omega, z} |v(\omega, z) - \hat{v}(\omega, z)| \leq R.$$

To prove this result we use that $\mathcal{N}$ can be covered by finitely many hyperslices H_j, $j \in \mathcal{J}$, see [60].

Definition 3.11. Let $\delta > 0$ and normal vector $a \in \mathbb{R}^m \setminus \{0\}$ be given. Then, the *hyperslice* $H(a, \delta)$ is defined as

$$H(a, \delta) := \{z \in \mathbb{R}^m : 0 \leq a^\top z \leq \delta\}.$$

However, before we derive an upper bound on $\|v - \hat{v}\|_\infty$, we first derive a lower bound and upper bound on the value function $\hat{v}(\omega, z)$ of the tight pseudo-valid cutting plane approximation. The lower bound follows directly from Proposition 3.1 and the fact that $\hat{v}(\omega, z)$ is convex in z for every fixed $\omega \in \Omega$.

Lemma 3.2. *Consider a tight pseudo-valid cutting plane approximation $\hat{v}(\omega, z)$ as defined in Definition 3.2. Then, under Assumption 3.1, we have for every $\omega \in \Omega$ and $z \in \mathbb{R}^m$ that*

$$\hat{v}(\omega, z) \geq \max_{k=1,\ldots,K} \left\{ q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha) \right\}.$$

The lower bound of $\hat{v}(\omega, z)$ in Lemma 3.2 is not only valid on the linearity regions of $\hat{v}(\omega, \cdot)$, but also on $\omega - \mathcal{N}$. We will show that the difference between $\hat{v}(\omega, z)$ and this lower bound is bounded. Again, we use the fact that $\hat{v}(\omega, z)$ is convex in z for every fixed $\omega \in \Omega$.

Lemma 3.3. *Consider a tight pseudo-valid cutting plane approximation $\hat{v}(\omega, z)$ as defined*

in Definition 3.2. Then, under Assumption 3.1, there exists $R' \in \mathbb{R}$ such that

$$\hat{v}(\omega, z) - \max_{k=1,\dots,K} \left\{ q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha) \right\} \leq R'. \tag{3.13}$$

Now we are ready to prove an upper bound on $\|v - \hat{v}\|_\infty$. The idea of the proof is that we can use Lemma 3.2 and 3.3 to bound $\|\hat{v} - v_{LP}\|_\infty$, where the LP-relaxation $v_{LP}(\omega, z)$ of $v(\omega, z)$ is equal to

$$v_{LP}(\omega, z) = \max_{k=1,\dots,K} \left\{ q_{B_k}^\top B_k^{-1}(\omega - z) \right\}, \tag{3.14}$$

and the maximum difference between v and v_{LP} is known.

Proposition 3.2. *Consider a tight pseudo-valid cutting plane approximation $\hat{v}(\omega, z)$ as defined in Definition 3.2. Then, under Assumption 3.1, there exists $R \in \mathbb{R}$ such that*

$$\|v - \hat{v}\|_\infty \leq R.$$

3.4.4 Proof of error bound

In this section we give the proofs of Theorem 3.2 and Corollary 3.1. Whereas the focus in Sections 3.4.2 and 3.4.3 was on $\hat{v}(\omega, z)$ as a function of z for fixed $\omega \in \Omega$, we now consider the difference $v(\omega, z) - \hat{v}(\omega, z)$ as a function of ω for fixed $z \in \mathbb{R}^m$. This is because $Q(z) - \hat{Q}(z) = \mathbb{E}_\omega[v(\omega, z) - \hat{v}(\omega, z)]$, $z \in \mathbb{R}^m$, and thus $v(\omega, z) - \hat{v}(\omega, z)$ can be interpreted as the underlying difference function for fixed $z \in \mathbb{R}^m$. Based on Propositions 3.1 and 3.2, we know for tight pseudo-valid cutting plane approximations $\hat{v}(\omega, z)$ that for $\omega \in z + \Lambda^k(d'_k)$, $k = 1, \dots, K$,

$$v(\omega, z) - \hat{v}(\omega, z) = \psi^k(\omega - z) - \psi^k(\omega - \alpha),$$

and for $\omega \in z + \mathcal{N}$,

$$\left| v(\omega, z) - \hat{v}(\omega, z) \right| \leq R.$$

We will use these two main properties to derive an upper bound for $\|Q - \hat{Q}\|_\infty$ that depends on the total variations of the one-dimensional conditional probability density functions of the random variables in the model.

Proof of Theorem 3.2. Combining Theorem 3.3 and Proposition 3.1, there exist basis matrices B_k, corresponding simplicial cones Λ^k, distances $d'_k \geq 0$, and bounded B_k-periodic functions ψ^k such that for $\omega - z \in \Lambda^k(d'_k)$,

$$v(\omega, z) - \hat{v}(\omega, z) = \psi^k(\omega - z) - \psi^k(\omega - \alpha).$$

Moreover, by Proposition 3.2, there exists $R \in \mathbb{R}$ such that $\|v - \hat{v}\|_\infty \leq R$.

Fix $z \in \mathbb{R}^m$ and consider the difference $v(\omega, z) - \hat{v}(\omega, z)$ as a function of ω. We will use that for $\omega \in z + \Lambda^k(d'_k)$, this difference is B_k-periodic, and for $\omega \in z + \mathcal{N}$, it is bounded by R. In fact, using Theorems 4.6 and 4.13 in [60] we can show that there exist constants $D > 0$ and $C'_k > 0$, $k = 1, \ldots, K$, such that

$$\mathbb{P}\{\omega \in z + \mathcal{N}\} \leq D \sum_{i=1}^{m} \mathbb{E}_{\omega_{-i}}\Big[|\Delta| f_i(\cdot|\omega_{-i})\Big], \tag{3.15}$$

and for every $k = 1, \ldots, K$,

$$\left| \int_{z+\Lambda^k(d'_k)} (\psi^k(t - z) - \psi^k(t - \alpha)) f(t) dt \right| \leq C'_k \sum_{i=1}^{m} \mathbb{E}_{\omega_{-i}}\Big[|\Delta| f_i(\cdot|\omega_{-i})\Big]. \tag{3.16}$$

Then,

$$\begin{aligned}
|Q(z) - \hat{Q}(z)| &= \left| \int_{\mathbb{R}^m} (v(t, z) - \hat{v}(t, z)) f(t) dt \right| \\
&\leq \left| \int_{z+\mathcal{N}} (v(t, z) - \hat{v}(t, z)) f(t) dt \right| \\
&\qquad + \sum_{k=1}^{K} \left| \int_{z+\Lambda^k(d'_k)} (v(t, z) - \hat{v}(t, z)) f(t) dt \right| \\
&\leq R\mathbb{P}\{\omega \in z + \mathcal{N}\} + \sum_{k=1}^{K} \left| \int_{z+\Lambda^k(d'_k)} (v(t, z) - \hat{v}(t, z)) f(t) dt \right|.
\end{aligned}$$

Applying the bound in (3.15) to the first term and the bounds in (3.16) to the second term, we obtain

$$\begin{aligned}
|Q(z) - \hat{Q}(z)| &\leq RD \sum_{i=1}^{m} \mathbb{E}_{\omega_{-i}}\Big[|\Delta| f_i(\cdot|\omega_{-i})\Big] + \sum_{k=1}^{K} C'_k \sum_{i=1}^{m} \mathbb{E}_{\omega_{-i}}\Big[|\Delta| f_i(\cdot|\omega_{-i})\Big] \\
&= C \sum_{i=1}^{m} \mathbb{E}_{\omega_{-i}}\Big[|\Delta| f_i(\cdot|\omega_{-i})\Big],
\end{aligned}$$

where the constant C is defined as $C := RD + \sum_{k=1}^{K} C'_k$. ■

Proof of Corollary 3.1. Since (3.4) holds for the pseudo-valid cutting plane approximation $\hat{v}(\omega, z)$, it follows immediately that there exist $\alpha \in \mathbb{R}^m$ and $\beta \in \mathbb{Z}^m$ such that for all $z \in \alpha + \beta\mathbb{Z}^m$ and $\omega \in \Omega$,

$$\hat{v}(\omega, z) \leq v(\omega, z).$$

This implies for example that (3.12) holds with "$\leq$" instead of "$=$". Thus, analogously to the proof of Theorem 3.2 we are able to show that there exists a constant $C \in \mathbb{R}$ with $C > 0$ such that for all ω with pdf $f \in \mathcal{H}^m$, and for all $z \in \mathbb{R}^m$,

$$\hat{Q}(z) - Q(z) \leq C \sum_{i=1}^{m} \mathbb{E}_{\omega_{-i}}\Big[|\Delta| f_i(\cdot|\omega_{-i})\Big]. \tag{3.17}$$

Notice that this upper bound does not necessarily apply to $Q(z) - \hat{Q}(z)$ since the pseudo-valid cuts are not necessarily tight. The claim follows by rearranging terms in (3.17). ■

3.5 Examples of pseudo-valid cutting planes

In this section we consider examples of pseudo-valid cutting plane approximations. We derive pseudo-valid mixed-integer Gomory cuts in Section 3.5.1, and a tight pseudo-valid cutting plane approximation for a nurse scheduling problem in Section 3.5.2.

3.5.1 Pseudo-valid mixed-integer Gomory cuts

Consider the second-stage value function

$$v(\omega, z) := \min_{y_B, y_N} \Big\{ q_B^\top y_B + q_N^\top y_N : By_B + Ny_N = \omega - z, \\ y_B \in Y_B,\, y_N \in Y_N \Big\}, \quad \omega \in \Omega,\, z \in \mathbb{R}^m,$$

where similar as in Section 3.4.1, we let B denote a dual feasible basis matrix of the LP-relaxation of v. Multiplying the equality constraint in $v(\omega, z)$ by $e_i^\top B^{-1}$, where e_i is the i-th unit vector, we obtain

$$y_{B_i} + e_i^\top B^{-1} N y_N = e_i^\top B^{-1}(\omega - z), \tag{3.18}$$

where y_{B_i} denotes the i-th basic variable. Let $\bar{w}_{ij}$ denote the j-th component of the vector $e_i^\top B^{-1}N$, let y_{N_j} denote the j-th non-basic variable, and let $r_i(\omega, z) := e_i^\top B^{-1}(\omega - z) - \lfloor e_i^\top B^{-1}(\omega - z) \rfloor$. If the i-th basic variable y_{B_i} is restricted to be integer, then we can derive from (3.18) the exact mixed-integer Gomory cut

$$\sum_{j \in J_1} \min \left\{ \frac{\bar{w}_{ij} - \lfloor \bar{w}_{ij} \rfloor}{r_i(\omega, z)}, \frac{1 - \bar{w}_{ij} + \lfloor \bar{w}_{ij} \rfloor}{1 - r_i(\omega, z)} \right\} y_{N_j} + \sum_{j \in J_2} \max \left\{ \frac{\bar{w}_{ij}}{r_i(\omega, z)}, \frac{-\bar{w}_{ij}}{1 - r_i(\omega, z)} \right\} y_{N_j} \geq 1, \quad (3.19)$$

where J_1 denotes the index set of integer non-basic variables y_{N_j} and J_2 the index set of continuous non-basic variables y_{N_j}; see, e.g., [9].

Obviously, the exact mixed-integer Gomory cut in (3.19) is not affine in z, among others since $r_i(\omega, z)$ is not affine in z. However, if $z \in \beta\mathbb{Z}^m$, where $\beta := |\det(B)|e$ with e the m-dimensional all-one vector, then under the assumption that W is integer, we can show that $r_i(\omega, z) = r_i(\omega, 0)$, and thus the mixed-integer Gomory cut in (3.19) does not depend on z. This is true, since for such z, we have $e_i^\top B^{-1} z = e_i^\top \Big(\det(B)^{-1} \mathrm{adj}(B) \Big) z \in \mathbb{Z}$, and thus

$$\begin{aligned} r_i(\omega, z) &= e_i^\top B^{-1}(\omega - z) - \left\lfloor e_i^\top B^{-1}(\omega - z) \right\rfloor \\ &= e_i^\top B^{-1}\omega - \left\lfloor e_i^\top B^{-1}\omega \right\rfloor = r_i(\omega, 0). \end{aligned}$$

Similarly, if $z \in \alpha + \beta\mathbb{Z}^m$ with $\beta := |\det(B)|e$, then $r_i(\omega, z) = r_i(\omega, \alpha)$. Thus, replacing $r_i(\omega, z)$ by $r_i(\omega, \alpha)$ in (3.19) yields an approximate mixed-integer Gomory cut that does not depend on z and is valid for all ω and for all z on a grid of points $\alpha + \beta\mathbb{Z}^m$. Hence, the approximate mixed-integer Gomory cut

$$\sum_{j \in J_1} \min \left\{ \frac{\bar{w}_{ij} - \lfloor \bar{w}_{ij} \rfloor}{r_i(\omega, \alpha)}, \frac{1 - \bar{w}_{ij} + \lfloor \bar{w}_{ij} \rfloor}{1 - r_i(\omega, \alpha)} \right\} y_{N_j} + \sum_{j \in J_2} \max \left\{ \frac{\bar{w}_{ij}}{r_i(\omega, \alpha)}, \frac{-\bar{w}_{ij}}{1 - r_i(\omega, \alpha)} \right\} y_{N_j} \geq 1, \quad (3.20)$$

is pseudo-valid and thus satisfies the assumptions of Corollary 3.1. This implies that we can use these pseudo-valid mixed-integer Gomory cuts to obtain a strictly better (approximate) lower bound of Q than Q_{LP}, as we illustrate in Example 3.3 below.

Example 3.3. Consider an adjusted version of the MISP in [67], given by

$$\min_{x\in X} \; -\frac{3}{2}x_1 - 4x_2 + Q(x),$$

where $X = [0,5]^2$ and $Q(x) = \mathbb{E}_\omega[v(\omega, x)]$ with

$$\begin{aligned} v(w,x) = \min_{y,u,v} \quad & -16y_1 - 19y_2 - 23y_3 - 28y_4 + 50v_1 + 50v_2 \\ \text{s.t.} \quad & 2y_1 + 3y_2 + 4y_3 + 5y_4 + u_1 - v_1 = \omega_1 - x_1 \\ & 6y_1 + \; y_2 + 3y_3 + 2y_4 + u_2 - v_2 = \omega_2 - x_2 \\ & y \in \mathbb{Z}^4_+, \quad u, v \in \mathbb{R}^2_+. \end{aligned}$$

Similar as in [67], the random vector ω follows a discrete uniform distribution on $\{5, 5.5, \ldots, 15\} \times \{5, 5.5, \ldots, 15\}$, and thus the support Ω of ω consists of 441 scenarios. Different variants of this problem are considered in, e.g., [4, 24, 31, 53, 73].

We use (3.20) to derive pseudo-valid mixed-integer Gomory cuts for Q. To be precise, we derive pseudo-valid mixed-integer Gomory cuts from those dual feasible basis matrices B that are optimal for the LP-relaxation v_{LP} of v for at least one pair of $\omega \in \Omega$ and $x \in X$. These basis matrices turn out to be

$$B_1 = \begin{pmatrix} 2 & 3 \\ 6 & 1 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 2 & 0 \\ 6 & 1 \end{pmatrix}, \quad \text{and } B_3 = \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix}.$$

Since both columns of B_1 and only the first column of B_2 and B_3 correspond to integer second-stage variables, we derive four pseudo-valid mixed-integer Gomory cuts: two corresponding to B_1 and one corresponding to B_2 and B_3.

Since the first-stage decision vector x is two-dimensional, we are able to depict both Q and Q_{LP} graphically in Figure 3.3. Moreover, we also show four pseudo-valid cutting plane approximations $\hat{Q}$, obtained by iteratively adding an additional pseudo-valid mixed-integer Gomory cut. From Figure 3.3 we observe that already by adding a single pseudo-valid mixed-integer Gomory cut we significantly improve the LP-relaxation lower bound Q_{LP} of Q. Moreover, this lower bound improves further if we add additional pseudo-valid mixed-integer Gomory cuts. Notice that $\hat{Q}$ is indeed an *approximate* lower bound, since for $\hat{Q}$ with three or four cuts, the function $\hat{Q}$ may exceed the original expected value function Q. However, Corollary 3.1 guarantees that $\sup_{x\in X}\{\hat{Q}(x) - Q(x)\}$ cannot be too large.

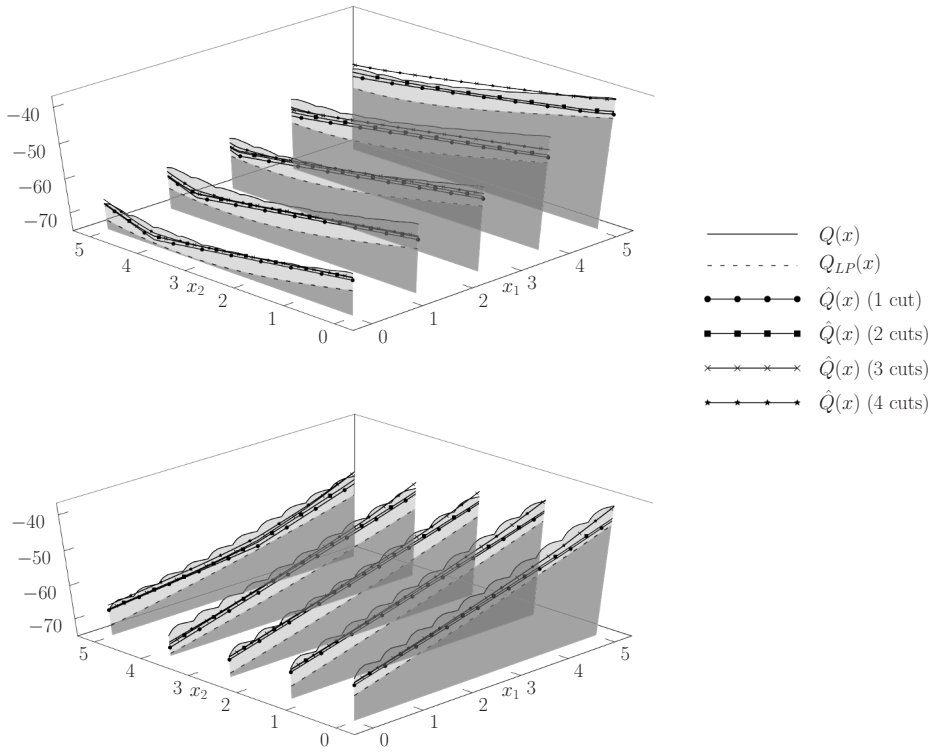


Figure 3.3. The expected value function Q of Example 3.3, its LP-relaxation Q_{LP}, and four pseudo-valid cutting plane approximations $\hat{Q}$ with a different number of pseudo-valid mixed-integer Gomory cuts.

Finally, in Figure 3.4 we show the approximation error

$$\|Q - \hat{Q}\|_\infty := \sup_{x \in X} |Q(x) - \hat{Q}(x)|$$

for the four pseudo-valid cutting plane approximations $\hat{Q}$ with respect to the integrality gap $\|Q - Q_{LP}\|_\infty$ of Q_{LP}. We observe that this integrality gap reduces by almost 20% if we add a single pseudo-valid mixed-integer Gomory cut. Moreover, by adding all four cuts, the approximation error of $\hat{Q}$ reduces to approximately 50% of that of Q_{LP}. ◊

3.5.2 Nurse scheduling problem

In this section we apply tight pseudo-valid cutting planes to a nurse scheduling problem, introduced by Kim and Mehrotra [40]. In this problem, a regular work

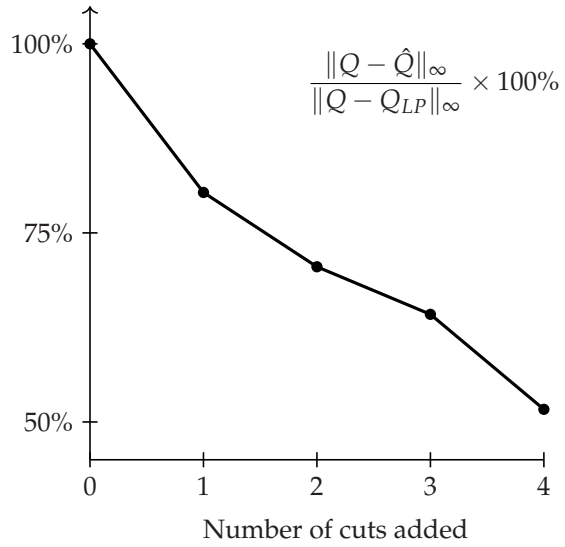


Figure 3.4. The relative approximation errors of the pseudo-valid cutting plane approximations $\hat{Q}$ of Example 3.3.

schedule for the nurses is determined in the first stage, resulting in a number z_t of available nurses per time period $t = 1, \ldots, T$. This regular work schedule is determined before the random demand ω_t for nurses per time period is known. Thus, it may turn out that we have a shortage or surplus of nurses in some of the time periods. In this case, it is possible to add or subtract nurse shifts, consisting of several consecutive time periods, after the demands ω_t are known. Moreover, we penalize any remaining nurse shortages and nurse surpluses using unit penalty costs per time period. The corresponding second-stage value function v is given by

$$\begin{aligned} v(\omega, z) = \min_{y, u_1, u_2} \quad & q^\top y + r_1^\top u_1 + r_2^\top u_2 \\ \text{s.t.} \quad & Wy - u_1 + u_2 = \omega - z \\ & y \in \mathbb{Z}_+^{n_2},\ u_1, u_2 \in \mathbb{R}_+^T, \end{aligned} \tag{3.21}$$

where $y \in \mathbb{Z}_+^{n_2}$ represents the possibility to add or subtract nurse shifts, and W is a $\{-1, 0, 1\}$-matrix, modelling which time periods are contained in which shift. Kim and Mehrotra [40] show that W is a totally unimodular matrix. Moreover, they show that if $z \in \mathbb{Z}^T$, then the cutting planes $Wy - \hat{D}(\omega)u_1 \leq \lfloor \omega \rfloor - z$, with $\hat{D}(\omega)$

the diagonal matrix with t-th diagonal component $\hat{D}_{tt}(\omega)$ equal to

$$\hat{D}_{tt}(\omega) = \frac{1}{1 - \omega_t + \lfloor \omega_t \rfloor}, \quad t = 1, \ldots, T,$$

are valid for all $\omega \in \Omega$. In particular, combined with the constraints $Wy - u_1 + u_2 = \omega - z$, they completely define the integer hull $\bar{Y}(\omega, z)$ of the feasible region $Y(\omega, z)$ of $v(\omega, z)$. That is, for every $\omega \in \Omega$ and $z \in \mathbb{Z}^T$,

$$\bar{Y}(\omega, z) = \Big\{ (y, u_1, u_2) \in \mathbb{R}_+^{n_2+2T} : Wy - u_1 + u_2 = \omega - z, \\ Wy - \hat{D}(\omega) u_1 \leq \lfloor \omega \rfloor - z \Big\}.$$

If we assume, contrary to [40], that z is not necessarily integral, then we may use the cutting planes $Wy - \hat{D}(\omega)u_1 \leq \lfloor \omega \rfloor - z$ to derive the tight pseudo-valid cutting plane approximation

$$\begin{aligned} \hat{v}(\omega, z) = \min_{y, u_1, u_2} \quad & q^\top y + r_1^\top u_1 + r_2^\top u_2 \\ \text{s.t.} \quad & Wy - u_1 + u_2 = \omega - z \\ & Wy - \hat{D}(\omega) u_1 \leq \lfloor \omega \rfloor - z \\ & y \in \mathbb{R}_+^{n_2}, \; u_1, u_2 \in \mathbb{R}_+^T. \end{aligned} \tag{3.22}$$

Since the cutting planes are valid for all $z \in \alpha + \beta\mathbb{Z}^m$, with $\alpha = 0$ and $\beta = e$, and for all $\omega \in \Omega$, they are indeed pseudo-valid and tight. Hence, by Theorem 3.2 the error of the corresponding tight pseudo-valid cutting plane approximation $\hat{Q}$ converges to zero if all total variations of one-dimensional conditional pdf of ω converge to zero. In Example 3.4 below, we numerically show the actual performance of this approximation for the one-dimensional second-stage value function of Example 3.1 in Section 3.2, which can be considered a special case of (3.21).

Example 3.4. Consider the second-stage value function $v(\omega, z)$ of Example 3.1,

$$\begin{aligned} v(\omega, z) = r(\omega - z) + \min_{y, u_1} \quad & (q - r)y + 2ru_1 \\ \text{s.t.} \quad & y - u_1 \leq \omega - z \\ & y \in \mathbb{Z}_+, \; u_1 \in \mathbb{R}_+, \end{aligned}$$

and its tight pseudo-valid cutting plane approximation defined in (3.8). Let ω be a normal random variable with mean μ and standard deviation σ. Then, as shown

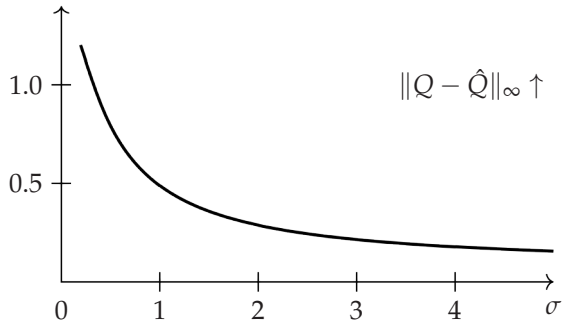


Figure 3.5. The maximum difference between Q and its pseudo-valid cutting plane approximation $\hat{Q}$ of Example 3.4, with $q = 1$ and $r = 2$, as a function of the standard deviation σ of a normal random variable ω.

in Example 3.2, the total variation $|\Delta|f$ of the probability density function f of ω equals $|\Delta|f = \sigma^{-1}\sqrt{2/\pi}$. Figure 3.5 shows $\|Q - \hat{Q}\|_\infty$, the maximum difference between the expected value function Q and its tight pseudo-valid cutting plane approximation $\hat{Q}$, as a function of the standard deviation σ for $q = 1$ and $r = 2$. We observe that this difference decreases if σ increases. This is in line with Theorem 3.2, since the total variation $|\Delta|f$ of a normal probability density function f decreases if the standard deviation σ increases. $\Diamond$

3.6 Numerical case study

In this section we evaluate the performance of the tight pseudo-valid cutting planes derived for the nurse scheduling problem in Section 3.5.2. This nurse scheduling problem is adapted from [40] so that all first-stage decision variables are continuous instead of integer. For small problem instances we compute the optimality gap of our pseudo-valid cutting plane approximation, and for larger problem instances we compare our approximation with two benchmark approximations. All computations are carried out on a single Intel Xeon E5 2680v3 core (2.5 GHz).

3.6.1 Experimental design

Similar as in Kim and Mehrotra [40], the regular nurse schedule that we have to determine is a weekly repeating nurse scheduling for a 12-week period in which nurses work in shifts of 8 or 12 hours. Moreover, there are three possible employment types for nurses: full-time, part-time, and casual. See [40] for the details.

The normalized costs for regular nurses are 1 per hour. In the second-stage, nurse shifts can be cancelled for free, but they can only be added at a cost of 1.5 per hour. Moreover, the unit penalty costs r_1 and r_2 from (3.21) are 0 and 10^4 for each time period, respectively. In this way, demand for nurses should be met either by the regular workforce or by adding nurse shifts in the second-stage to avoid high penalty costs.

The demand for nurses ω_t during time period t is determined by the patient volume ξ_t and the nurse-to-patient ratio. Similar as in [40], we use a nurse-to-patient ratio of 1:4 from 8AM to 4PM and 1:5 for the remainder of the day. For the distribution of the patient volume ξ_t, we differ from [11], and consider both normal and uniform distributions with mean $\mu_t := \mathbb{E}[\xi_t] \in \{25, 75, 125\}$ and standard deviation $\sigma_t \in \{0.5, 1, 2, 5, 10, 19\}$. We assume that all ξ_t are identically and independently distributed. We include $\mu_t = 125$ and $\sigma_t = 19$ since Kim and Mehrotra [40] report these statistics for actual patient volume data. Moreover, we note that in our experiments we truncate the patient volume if necessary, so that it is always nonnegative.

We test the performance of our tight pseudo-valid cutting plane approximation on both small and large problem instances of this nurse scheduling problem. In the small problem instances a 3-day regular nurse schedule has to be determined, taking into account 100 scenarios for future patient volume, sampled from the abovementioned distributions. The large problem instances correspond to the 12-week nurse scheduling problem with 1000 scenarios. These problems instances have the same size as those in [40].

3.6.2 Solution methods

For both small and large problem instances we solve the pseudo-valid cutting plane approximation from (3.22). We do so using a standard L-shaped algorithm [84], yielding the approximating solution $\hat{x}_0$. The solution is called $\hat{x}_0$ since the pseudo-valid cutting planes in (3.22) can be considered a special case of the family of pseudo-valid cutting planes

$$Wy - \hat{D}(\omega - \alpha)u_1 \leq \lfloor \omega - \alpha \rfloor + \alpha - z,$$

with $\alpha = 0$. We, however, also compute the pseudo-valid cutting plane solutions $\hat{x}_\alpha$ for 100 different values of α, and let $\hat{x}^*$ denote the best solution among those. The different values for α that we use are $\alpha = 0.01ke$, $k = 1, \ldots, 100$, where e is the

all-one vector.

For the small nurse scheduling problems we solve their large-scale deterministic equivalent MIP, and we let x^* denote the best found solution within 12 hours of computation time. We also obtain the MIP optimality gap corresponding to x^*. If this gap is zero, then x^* is the optimal solution. Alternatively, we could have used a state-of-the-art solution method, such as, e.g., [53], to obtain the exact optimal solution. However, the MIP optimality gaps that we obtain by solving the large-scale deterministic equivalent MIP are relatively small already.

The larger problem instances cannot be solved to optimality within reasonable time limits. For these instances we compute two benchmark approximations. The first is the LP-relaxation in which the integrality conditions on the second-stage variables y are relaxed, and the second is the expected-value model in which the random patient volumes are replaced by their mean values. We denote the resulting solutions by x_{LP} and x_μ, respectively. The solution of the LP-relaxation is obtained using a similar L-shaped algorithm as for the pseudo-valid cutting plane approximation. The expected-value model is a relatively small MIP that can easily be solved to optimality. For each problem instance, both small and large, we dedicate at most 12 hours of computation time to obtaining each candidate solution.

3.6.3 Performance measures

For the small problem instances we report in-sample costs, out-of-sample costs, and running times of $\hat{x}_0$ and $\hat{x}^*$. The in-sample costs are based on the 100 generated scenarios for that problem instance, whereas the out-of-sample costs are based on a new sample of 10^6 scenarios. We also report the relative difference (in %) with the in-sample and out-of-sample costs of x^*. The in-sample difference represents the optimality gap in the 100-scenario problem. However, since not all problem instances could be solved to optimality, we use the MIP optimality gap of x^* to compute lower and upper bounds on the optimality gaps of $\hat{x}_0$ and $\hat{x}^*$. For the larger problem instances we do not report these optimality gaps, but only the normalized in-sample costs of the pseudo-valid cutting plane solutions and the benchmark solutions.

The running times for $\hat{x}^*$ for the small problem instances represent the time required to solve the 100 pseudo-valid cutting plane approximations in series. For the large problem instances, however, we report the maximum computing time over all 100 approximations. This is the time required to obtain $\hat{x}^*$ if all pseudo-valid cutting plane approximations are solved in parallel. We actually used parallel

computing over 24 cores to obtain $\hat{x}^*$.

3.6.4 Numerical results

Table 3.1 shows the results for the small problem instances. We observe that the overall performance of the pseudo-valid cutting plane solutions $\hat{x}_0$ and $\hat{x}^*$ is very good, both for the normal and uniform distribution. In fact, in most of the problem instances the in-sample optimality gap for $\hat{x}^*$ is below 1%. The optimality gaps are smaller for larger values of the standard deviation σ. This is in line with Theorem 3.2 since the total variation of both the normal and uniform probability density function is smaller if σ is larger.

For small values of the standard deviation, i.e., $\sigma \leq 2$, and for $\mu = 25$, we observe that the optimality gap of $\hat{x}_0$ can be quite high. In these cases we do not recommend to use this pseudo-valid cutting plane approximation. It turns out, however, that the optimality gap can be significantly reduced, in particular when σ is small, by considering multiple pseudo-valid cutting plane approximations with different values for α, and to select the best among them. Indeed, the in-sample optimality gap of $\hat{x}^*$ never exceeds 4%.

Compared to the optimal in-sample solution x^*, the out-of-sample performance of the pseudo-valid cutting plane solutions $\hat{x}_0$ and $\hat{x}^*$ is better than their in-sample performance. In fact, in some cases the out-of-sample costs of $\hat{x}^*$ are actually lower than those for x^*. In general, we conclude that for larger values of the standard deviation, i.e., $\sigma \geq 5$, the out-of-sample performance of $\hat{x}^*$ is equivalent to that of x^*.

Finally, we note from Table 3.1 that the pseudo-valid cutting plane solutions are obtained extremely fast for the small problem instances. Indeed, the running times for $\hat{x}_0$ are below 10 seconds, and to obtain $\hat{x}^*$, the best of 100 different pseudo-valid cutting plane solutions, takes at most 16 minutes. In contrast, the optimal solution x^* could not be computed within 12 hours for most of the small problem instances.

Table 3.2 shows the results for the large problem instances. We observe that both pseudo-valid cutting plane solutions $\hat{x}_0$ and $\hat{x}^*$ consistently outperform the benchmark solutions x_{LP} and x_μ, both for the normal and uniform distribution. Similar as for the small problem instances, $\hat{x}^*$ can be significantly better than $\hat{x}_0$, in particular for small values of the standard deviation σ. For large values of σ, we observe that the solution x_{LP} of the LP-relaxation is almost as good as the pseudo-valid cutting plane solutions $\hat{x}^*$ and $\hat{x}_0$. Since $\hat{x}^*$ and $\hat{x}_0$ are close to optimal for the small problem instances in these cases, we conjecture that $\hat{x}^*$, $\hat{x}_0$, and x_{LP} are all

Table 3.1. Numerical results for the small nurse scheduling problems.

		Running times (seconds)		In-sample costs (percentage gap)		Out of sample costs (percentage gap)	
μ	σ	$\hat{x}_0$	$\hat{x}^*$	$\hat{x}_0$	$\hat{x}^*$	$\hat{x}_0$	$\hat{x}^*$
Normally distributed patient volume							
25	0.5	1.9	124.3	456 (10.9% - 11.0%)	425 (3.4% - 3.5%)	456 (9.9%)	426 (2.7%)
25	1.0	1.6	177.4	456 (5.9% - 6.4%)	443 (2.9% - 3.3%)	456 (4.8%)	444 (2.0%)
25	2.0	1.8	246.2	474 (1.8% - 3.2%)	473 (1.7% - 3.0%)	474 (0.7%)	475 (0.9%)
25	5.0	3.3	401.9	558 (1.3% - 3.6%)	555 (0.9% - 3.1%)	559 (0.5%)	557 (0.1%)
25	10.0	3.9	415.3	686 (0.6% - 2.8%)	683 (0.2% - 2.4%)	688 (-0.1%)	688 (-0.1%)
25	19.0	4.2	434.7	919 (0.4% - 2.0%)	917 (0.2% - 1.8%)	924 (0.0%)	923 (-0.2%)
75	0.5	2.0	154.5	1230 (3.2% - 3.3%)	1196 (0.4% - 0.4%)	1230 (2.9%)	1197 (0.2%)
75	1.0	2.9	253.8	1248 (3.1% - 3.3%)	1214 (0.3% - 0.5%)	1248 (2.7%)	1216 (0.1%)
75	2.0	2.1	302.9	1252 (0.5% - 1.0%)	1249 (0.3% - 0.8%)	1253 (0.2%)	1250 (-0.1%)
75	5.0	5.6	517.8	1337 (0.5% - 1.4%)	1334 (0.3% - 1.2%)	1340 (0.2%)	1337 (0.0%)
75	10.0	8.0	676.8	1466 (0.5% - 1.3%)	1464 (0.3% - 1.2%)	1471 (0.2%)	1468 (0.0%)
75	19.0	6.9	713.2	1698 (0.1% - 1.1%)	1697 (0.0% - 1.0%)	1705 (0.0%)	1705 (0.0%)
125	0.5	2.0	158.2	2016 (2.3% - 2.3%)	1985 (0.7% - 0.7%)	2016 (2.1%)	1986 (0.6%)
125	1.0	1.7	242.2	2016 (1.3% - 1.4%)	2003 (0.6% - 0.7%)	2016 (1.0%)	2004 (0.4%)
125	2.0	2.6	337.3	2034 (0.4% - 0.7%)	2033 (0.4% - 0.7%)	2034 (0.2%)	2035 (0.2%)
125	5.0	5.4	572.7	2118 (0.3% - 0.9%)	2115 (0.2% - 0.8%)	2119 (0.1%)	2118 (0.1%)
125	10.0	7.8	807.3	2246 (0.3% - 0.8%)	2243 (0.1% - 0.7%)	2248 (0.0%)	2248 (0.0%)
125	19.0	9.6	915.0	2478 (0.1% - 0.7%)	2477 (0.0% - 0.7%)	2486 (0.0%)	2485 (0.0%)
Uniformly distributed patient volume							
25	0.5	1.3	112.0	456 (13.0% - 13.0%)	418 (3.6% - 3.6%)	456 (12.7%)	418 (3.3%)
25	1.0	1.8	173.2	456 (9.4% - 9.4%)	433 (3.9% - 3.9%)	456 (9.0%)	433 (3.5%)
25	2.0	1.7	196.4	472 (6.5% - 6.5%)	458 (3.3% - 3.3%)	471 (5.7%)	458 (2.8%)
25	5.0	4.2	335.9	547 (4.4% - 4.8%)	533 (1.6% - 2.0%)	548 (3.9%)	533 (1.1%)
25	10.0	4.8	410.0	663 (1.9% - 3.4%)	652 (0.3% - 1.7%)	664 (1.1%)	656 (-0.1%)
25	19.0	6.7	513.4	868 (0.4% - 2.1%)	866 (0.1% - 1.8%)	875 (0.2%)	871 (-0.3%)
75	0.5	1.6	128.6	1224 (3.4% - 3.4%)	1189 (0.5% - 0.5%)	1224 (3.3%)	1189 (0.4%)
75	1.0	2.2	187.6	1248 (4.3% - 4.3%)	1201 (0.4% - 0.4%)	1248 (4.1%)	1201 (0.2%)
75	2.0	1.8	264.6	1248 (2.0% - 2.0%)	1226 (0.2% - 0.2%)	1248 (1.8%)	1226 (0.0%)
75	5.0	3.3	385.6	1320 (1.2% - 1.3%)	1311 (0.5% - 0.6%)	1320 (1.0%)	1311 (0.3%)
75	10.0	6.2	602.6	1441 (0.8% - 1.4%)	1436 (0.4% - 1.0%)	1444 (0.5%)	1438 (0.1%)
75	19.0	7.7	720.4	1649 (0.4% - 1.1%)	1647 (0.2% - 1.0%)	1654 (0.1%)	1653 (0.0%)
125	0.5	1.8	127.7	2016 (2.7% - 2.7%)	1978 (0.7% - 0.7%)	2016 (2.6%)	1978 (0.7%)
125	1.0	1.6	196.6	2016 (2.0% - 2.0%)	1993 (0.8% - 0.8%)	2016 (1.9%)	1993 (0.7%)
125	2.0	2.4	269.9	2032 (1.4% - 1.4%)	2018 (0.7% - 0.7%)	2031 (1.3%)	2018 (0.6%)
125	5.0	4.8	482.8	2107 (1.1% - 1.2%)	2093 (0.4% - 0.5%)	2108 (1.0%)	2093 (0.3%)
125	10.0	6.8	614.4	2222 (0.6% - 1.0%)	2212 (0.1% - 0.5%)	2225 (0.4%)	2216 (0.0%)
125	19.0	6.9	774.4	2429 (0.3% - 0.7%)	2427 (0.2% - 0.7%)	2435 (0.1%)	2432 (0.0%)

close to optimal in the corresponding large problem instances as well.

With respect to the running times we observe that the pseudo-valid cutting plane approximation can typically be solved much faster than the LP-relaxation. We emphasize that both solution methods are implemented using the same standard version of the L-shaped algorithm. Further analysis of the computational results shows that computing x_{LP} requires much more master iterations than computing $\hat{x}_0$, explaining the surprising difference in running times.

3.7 Discussion

We consider a new solution method for solving two-stage mixed-integer stochastic programs (MISPs) with uncertainty in the right-hand side. Instead of applying exact cuts to the second-stage feasible regions that are always valid, we propose to use *pseudo-valid* cutting planes that are affine in the first-stage decision variables. The advantage is that the approximating problem, that uses these pseudo-valid cuts, is convex, and thus it is much easier to solve than the original MISP.

For simple integer recourse (SIR) models, we show that the α-approximations of Klein Haneveld et al. [41] can be interpreted as pseudo-valid cutting plane approximations. A direct consequence of this result is that we obtain an error bound on the quality of the solution obtained using pseudo-valid cutting planes for SIR models. For general MISPs we derive a similar bound for so-called tight pseudo-valid cutting plane approximations. We show that the error of using such approximations converges to zero if the total variations of the one-dimensional conditional probability density functions in the model converge to zero.

For general MISPs we also derive pseudo-valid mixed-integer Gomory cuts, and for a nurse scheduling problem we derive pseudo-valid cutting planes that are tight. Numerical experiments for the latter type of cutting planes show that we are able to find good first-stage decisions, also for large-scale problems. In line with the error bound that we derive, the performance of the pseudo-valid cutting planes is better if the standard deviations of the random variables in the model are larger.

A direction for future research is to derive problem-specific pseudo-valid cutting planes for applications of two-stage MISPs. Moreover, tighter error bounds may be derived for these problem-specific cutting plane approximations using the special structure of the problems, similar as for simple integer recourse models. Alternatively, error bounds may be derived for pseudo-valid cutting planes for two-stage MISPs with discretely distributed random right-hand side vectors. Another

Table 3.2. Numerical results for the large nurse scheduling problems.

		Running times (seconds)				In-sample costs (normalized)			
μ	σ	$\hat{x}^*$	$\hat{x}_0$	x_{LP}	x_μ	$\hat{x}^*$	$\hat{x}_0$	x_{LP}	x_μ
Normally distributed patient volume									
25	0.5	4942	1063	11289	85	100.0	107.2	110.8	116.0
25	1.0	6019	1130	11101	82	100.0	102.9	108.0	111.4
25	2.0	7791	1998	12285	87	100.0	100.0	104.1	105.7
25	5.0	8937	5672	13340	89	100.0	100.3	100.8	105.4
25	10.0	10190	7626	12388	84	100.0	100.0	100.3	106.7
25	19.0	12407	12407	13123	87	100.0	100.1	100.1	107.6
75	0.5	4907	1275	10720	85	100.0	102.7	104.5	106.4
75	1.0	8471	2307	11394	85	100.0	102.7	103.6	104.8
75	2.0	7151	1853	12722	78	100.0	100.3	102.0	102.5
75	5.0	9231	7153	12385	90	100.0	100.3	100.4	102.4
75	10.0	14060	9105	13395	83	100.0	100.1	100.1	103.1
75	19.0	22212	17905	13866	85	100.0	100.1	100.1	104.4
125	0.5	5168	1112	11160	88	100.0	101.5	102.3	103.4
125	1.0	7091	1137	12254	84	100.0	100.6	101.8	102.5
125	2.0	7718	2070	13476	92	100.0	100.0	101.0	101.3
125	5.0	7529	6909	12279	85	100.0	100.1	100.2	101.4
125	10.0	10768	7081	13650	88	100.0	100.0	100.1	102.0
125	19.0	13524	11636	13521	85	100.0	100.0	100.1	103.0
Uniformly distributed patient volume									
25	0.5	5346	1210	11183	85	100.0	109.1	113.0	118.2
25	1.0	5507	1017	10797	77	100.0	105.3	111.1	114.1
25	2.0	6827	1722	10854	82	100.0	102.8	108.8	107.9
25	5.0	7719	5891	12549	89	100.0	102.8	103.6	109.3
25	10.0	9591	6075	13206	89	100.0	101.6	101.2	110.6
25	19.0	13995	10704	13572	72	100.0	100.6	100.4	111.7
75	0.5	5223	1093	10500	94	100.0	102.9	105.3	107.1
75	1.0	5617	1837	11594	86	100.0	103.9	105.0	106.1
75	2.0	7215	1278	11497	82	100.0	101.8	104.3	103.9
75	5.0	9031	3717	13427	76	100.0	100.7	101.6	103.9
75	10.0	10096	6199	12944	82	100.0	100.4	100.4	104.7
75	19.0	11780	10728	14084	89	100.0	100.2	100.2	106.3
125	0.5	5520	1027	10944	90	100.0	101.9	102.7	103.8
125	1.0	5636	1072	11824	86	100.0	101.1	102.4	103.1
125	2.0	6980	2030	12666	91	100.0	100.6	102.0	101.8
125	5.0	8013	5897	13650	86	100.0	100.7	100.9	102.4
125	10.0	8627	6290	13286	95	100.0	100.5	100.4	103.1
125	19.0	13637	8684	13872	86	100.0	100.2	100.1	104.4

future research direction is to combine exact and pseudo-valid cutting planes to obtain more accurate approximations at the expense of increasing the computational effort of solving the approximation.

Appendix 3.A Postponed proofs

Proof of Lemma 3.1. Since C is a compact convex set, every $z \in C$ can be written as a convex combination of its extreme points:

$$z = \sum_{j=1}^{J} \mu_j z^j,$$

with $\sum_{j=1}^{J} \mu_j = 1$, and $\mu_j \geq 0, j = 1, \ldots, J$. Since v is convex, this implies that for all $z \in C$

$$v(z) = v\Big(\sum_{j=1}^{J} \mu_j z^j\Big) \leq \sum_{j=1}^{J} \mu_j v(z^j) = \sum_{j=1}^{J} \mu_j (a^\top z^j + b) = a^\top z + b. \tag{3.23}$$

To prove that also $v(z) \geq a^\top z + b$ for all $z \in C$, assume for contradiction that there exists $\bar{z} \in C$ such that $v(\bar{z}) < a^\top \bar{z} + b$. Since C is convex and z^0 is an interior point of C there exists $\epsilon > 0$ such that $\hat{z} := z^0 + \epsilon(z^0 - \bar{z}) \in C$. This point $\hat{z}$ is defined in such a way that z^0 can be written as a convex combination of $\bar{z}$ and $\hat{z}$:

$$z^0 = \frac{1}{1+\epsilon}\hat{z} + \frac{\epsilon}{1+\epsilon}\bar{z}.$$

Since v is convex, this implies that

$$\begin{aligned} v(z^0) &\leq \frac{1}{1+\epsilon} v(\hat{z}) + \frac{\epsilon}{1+\epsilon} v(\bar{z}) \\ &< \frac{1}{1+\epsilon}(a^\top \hat{z} + b) + \frac{\epsilon}{1+\epsilon}(a^\top \bar{z} + b) = a^\top z^0 + b, \end{aligned} \tag{3.24}$$

where we use that $v(\bar{z}) < a^\top \bar{z} + b$ by assumption and $v(\hat{z}) \leq a^\top \hat{z} + b$ by (3.23). Since (3.24) contradicts the assumption that $v(z^0) = a^\top z^0 + b$, we conclude that $v(z) = a^\top z + b$ for all $z \in C$. ■

Proof of Proposition 3.1. Since $\hat{v}(\omega, z) = v(\omega, z)$ for all $\omega \in \Omega$ and $z \in \alpha + \beta\mathbb{Z}^m$, it follows from Theorem 3.3 that for every $k = 1, \ldots, K$, there exists d_k such that for

all $\omega \in \Omega$ and $z \in \alpha + \beta\mathbb{Z}^m$ with $\omega - z \in \Lambda^k(d_k)$,

$$\hat{v}(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - z).$$

Fix $\omega \in \Omega$. Then, $\psi^k(\omega - z)$ is B_k-periodic in z and thus $\psi^k(\omega - z) = \psi^k(\omega - z - \det(B_k)l)$ for every $l \in \mathbb{Z}^m$ by Lemma 4.8 in [60]. Define $\delta^k := \det(B_k)\beta \in \mathbb{Z}^m$ and let $l \in \mathbb{Z}^m$ be given, and consider the hyperrectangle $C^l(\alpha, \delta^k)$. Let z^j, $j = 1, \ldots, J$, denote its extreme points and let $z^0 := (\alpha_1 + \delta_1^k l_1, \ldots, \alpha_m + \delta_m^k l_m)$ be an interior point. If $C^l(\alpha, \delta^k) \subset \omega - \Lambda^k(d_k)$, then we can apply Lemma 3.1 with $a := -q_{B_k}^\top B_k^{-1}$, $b := q_{B_k}^\top B_k^{-1}\omega + \psi^k(\omega - \alpha)$, and $C := C^l(\alpha, \delta^k)$ to conclude that $\hat{v}(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha)$ for all $z \in C^l(\alpha, \delta^k)$. Since the diameter of $C^l(\alpha, \delta^k)$ is $2\|\delta^k\|$, we conclude that the result holds for all $\omega - z \in \Lambda^k(d_k + 2\|\delta^k\|)$. Indeed, $\omega - z$ will be in $\omega - C^l(\alpha, \delta^k)$ for some $l \in \mathbb{Z}^m$. The claim now follows by defining $d_k' := d_k + 2\|\delta^k\|$. ■

Proof of Lemma 3.2. Fix $\omega \in \Omega$. Then, by Proposition 3.1 it follows that for every $k = 1, \ldots, K$,

$$\hat{v}(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha), \quad z \in \omega - \Lambda^k(d_k').$$

Since $\hat{v}(\omega, z)$ is convex in z, and affine on $\omega - \Lambda^k(d_k')$, we can derive a subgradient inequality for each $k = 1, \ldots, K$:

$$\hat{v}(\omega, z) \geq q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha), \quad z \in \mathbb{R}^m.$$

Combining these inequalities over all $k = 1, \ldots, K$, yields the desired result. ■

Proof of Lemma 3.3. Fix $\omega \in \Omega$. By Proposition 3.1, there exist distances $d_k' \geq 0$ such that for every $k = 1, \ldots, K$,

$$\hat{v}(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha), \quad z \in \omega - \Lambda^k(d_k').$$

Thus, on the linearity regions $\omega - \Lambda^k(d_k')$, $\hat{v}(\omega, z)$ equals its lower bound from Lemma 3.2. Therefore, we only have to show (3.13) for $z \in \omega - \mathcal{N}$. To this end, let $z \in \omega - \mathcal{N}$ be given. Since $\mathcal{N}$ can be covered by finitely many hyperslices, there exist $a_j \in \mathbb{R}^m \backslash \{0\}$ and $\delta_j > 0$, $j \in \mathcal{J}$, such that

$$\mathcal{N} \subset \bigcup_{j \in \mathcal{J}} H_j,$$

where $H_j := H(a_j, \delta_j)$, $j \in \mathcal{J}$. We will construct points z_1 and z_2 in the linearity regions of $\hat{v}(\omega, \cdot)$, so that z is a convex combination of z_1 and z_2. Then, we can use that $\hat{v}(\omega, z)$ is convex in z to derive an upper bound on $\hat{v}(\omega, z)$ in terms of $\hat{v}(\omega, z_1)$ and $\hat{v}(\omega, z_2)$. Since z_1 and z_2 are in the linearity regions, these values are known.

To construct such z_1 and z_2, let $d \in \mathbb{R}^m \backslash \{0\}$ be a direction of unit length not parallel to any of the hyperslices H_j, and thus not orthogonal to any of the normal vectors a_j, $j \in \mathcal{J}$. Then, $a_j^\top d \neq 0$, $j \in \mathcal{J}$, and $\|d\| = 1$. We consider the line through z with direction d and define the halflines L_1 and L_2 as

$$L_1 := \{z + \mu d : \mu \in \mathbb{R}_+\} \quad \text{and} \quad L_2 := \{z - \mu d : \mu \in \mathbb{R}_+\}.$$

Since the direction d is not parallel to any of the hyperslices, we have $L_1 \not\subset \bigcup_{j\in\mathcal{J}} H_j$ and $L_2 \not\subset \bigcup_{j\in\mathcal{J}} H_j$, and thus $L_i \cap (\omega - \bigcup_{k=1}^K \Lambda^k(d'_k)) \neq \emptyset$, $i = 1, 2$. This means that it is possible to select $z^1, z^2 \in \omega - \bigcup_{k=1}^K \Lambda^k(d'_k)$ on L_1 and L_2, respectively, with minimal distance to z:

$$z^i := \arg\min_{z'} \Big\{ \|z - z'\| : z' \in L_i \cap \Big(\omega - \bigcup_{k=1}^K \Lambda^k(d'_k)\Big) \Big\}, \quad i = 1, 2.$$

Since z is on the line segment between z^1 to z^2, we can write z as a convex combination $z = \mu z^1 + (1-\mu) z^2$ of z^1 and z^2 with $\mu \in [0, 1]$. We will use the convexity of $\hat{v}(\omega, \cdot)$ to derive an upper bound on $\hat{v}(\omega, z)$. Here, we will assume without loss of generality that $z^1 \in \omega - \Lambda^{k_1}(d'_{k_1})$ and $z^2 \in \omega - \Lambda^{k_2}(d'_{k_2})$ with $k_1, k_2 \in \{1, \ldots, K\}$. We obtain

$$\begin{aligned}
\hat{v}(\omega, z) &\leq \mu \hat{v}(\omega, z^1) + (1-\mu)\hat{v}(\omega, z^2) \\
&= \mu \Big(q_{B_{k_1}}^\top B_{k_1}^{-1}(\omega - z^1) + \psi^{k_1}(\omega - \alpha) \Big) \\
&\quad + (1-\mu) \Big(q_{B_{k_2}}^\top B_{k_2}^{-1}(\omega - z^2) + \psi^{k_2}(\omega - \alpha) \Big).
\end{aligned}$$

To obtain the bound in (3.13) on the difference between $\hat{v}(\omega, z)$ and its lower bound, we subtract this lower bound from both the left- and right-hand side of the inequality above. Defining $k^* := \arg\max_{k=1,\ldots,K} \{q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha)\}$, the difference between $\hat{v}(\omega, z)$ and its lower bound can then be bounded by

$$\begin{aligned}
&\mu \Big(q_{B_{k_1}}^\top B_{k_1}^{-1}(\omega - z^1) + \psi^{k_1}(\omega - \alpha) - q_{B_{k^*}}^\top B_{k^*}^{-1}(\omega - z) - \psi^{k^*}(\omega - \alpha) \Big) \\
&+ (1-\mu) \Big(q_{B_{k_2}}^\top B_{k_2}^{-1}(\omega - z^2) + \psi^{k_2}(\omega - \alpha) - q_{B_{k^*}}^\top B_{k^*}^{-1}(\omega - z) - \psi^{k^*}(\omega - \alpha) \Big).
\end{aligned}$$

Since k_1 and k_2 are not necessarily the maximizing index for

$$\max_{k=1,\dots,K}\{q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha)\},$$

we may replace k^* by k_1 and k_2, respectively, to obtain after straightforward simplifications,

$$\hat{v}(\omega, z) - \max_{k=1,\dots,K}\left\{q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha)\right\} \le \mu q_{B_{k_1}}^\top B_{k_1}^{-1}(z - z^1) + (1-\mu) q_{B_{k_2}}^\top B_{k_2}^{-1}(z - z^2).$$

We will bound the right-hand side in terms of the distance $\|z^1 - z^2\|$ between z^1 and z^2. Here, we use $\lambda_i^* := \max_{k=1,\dots,K} |q_{B_k}^\top (B_k)^{-1} e_i|$, for $i = 1, \dots, m$, where e_i is the i-th unit vector. We have

$$\begin{aligned}
\hat{v}(\omega, z) - \max_{k=1,\dots,K}\left\{q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha)\right\} &\le \mu \sum_{i=1}^m \lambda_i^* |z_i^1 - z_i| \\
&\quad + (1-\mu) \sum_{i=1}^m \lambda_i^* |z_i^2 - z_i| \\
&\le \mu \sum_{i=1}^m \lambda_i^* \|z^1 - z\| \\
&\quad + (1-\mu) \sum_{i=1}^m \lambda_i^* \|z^2 - z\| \\
&\le \|z^1 - z^2\| \sum_{i=1}^m \lambda_i^*, \qquad (3.25)
\end{aligned}$$

where the last inequality holds since z is on the line segment between z_1 to z_2, and thus $\|z^1 - z\| \le \|z^1 - z^2\|$ and $\|z^2 - z\| \le \|z^1 - z^2\|$.

It remains to derive an upper bound on $\|z^1 - z^2\|$. To do so, observe that $\omega - z$ is on the line segment $\omega - L$ between $\omega - z^1$ and $\omega - z^2$. Moreover, this line segment is completely contained in the union of the hyperslices H_j, $j \in \mathcal{J}$. Hence, letting $\mathcal{L}(L)$ denote the length of line segment L, we have

$$\begin{aligned}
\|z^1 - z^2\| &= \|(\omega - z^1) - (\omega - z^2)\| \\
&= \mathcal{L}\Big((\omega - L) \cap (\bigcup_{j \in \mathcal{J}} H_j)\Big) \\
&= \mathcal{L}\Big(\bigcup_{j \in \mathcal{J}} ((\omega - L) \cap H_j)\Big) \le \sum_{j \in \mathcal{J}} \mathcal{L}\Big((\omega - L) \cap H_j\Big).
\end{aligned}$$

To find $\mathcal{L}((\omega - L) \cap H_j)$, observe that $\hat{z} \in L$ satisfies $\hat{z} = z - \hat{\mu}d$ for some $\hat{\mu} \in \mathbb{R}$. Moreover, $\omega - \hat{z} \in H_j := H(a_j, \delta_j)$ if $0 \leq a_j^\top(\omega - z + \hat{\mu}d) \leq \delta_j$, or equivalently if

$$\begin{cases} \dfrac{-a_j^\top(\omega - z)}{a_j^\top d} =: \underline{\mu} \leq \hat{\mu} \leq \overline{\mu} := \dfrac{\delta_j - a_j^\top(\omega - z)}{a_j^\top d}, & \text{if } a_j^\top d > 0, \\ \dfrac{\delta_j - a_j^\top(\omega - z)}{a_j^\top d} =: \underline{\mu} \leq \hat{\mu} \leq \overline{\mu} := \dfrac{-a_j^\top(\omega - z)}{a_j^\top d}, & \text{if } a_j^\top d < 0. \end{cases}$$

Then, $\mathcal{L}((\omega - L) \cap H_j) = (\overline{\mu} - \underline{\mu})\|d\| = \frac{\delta_j}{|a_j^\top d|}$, where we use that $\|d\| = 1$. Thus, by defining

$$R' := \Big(\sum_{i=1}^{m} \lambda_i^*\Big)\Big(\sum_{j \in \mathcal{J}} \frac{\delta_j}{|a_j^\top d|}\Big),$$

the claim follows from combining $\|z^1 - z^2\| \leq \sum_{j \in \mathcal{J}} \frac{\delta_j}{|a_j^\top d|}$ and (3.25). ■

Proof of Proposition 3.2. Consider the LP-relaxation $v_{LP}(\omega, z)$ of $v(\omega, z)$ as defined in (3.14). Then, by, e.g., [19] and [27], there exists R'' such that $\|v - v_{LP}\|_\infty \leq R''$. Moreover, by combining Lemma 3.2 and 3.3, we conclude that $\|v_{LP} - \hat{v}\|_\infty \leq R' + \max_{k=1,\dots,K} \sup_{s \in \mathbb{R}^m} |\psi^k(s)|$. If we define $R := R'' + R' + \max_{k=1,\dots,K} \sup_{s \in \mathbb{R}^m} |\psi^k(s)|$, then

$$\|v - \hat{v}\| \leq \|v - v_{LP}\| + \|v_{LP} - \hat{v}\| \leq R'' + R' + \max_{k=1,\dots,K} \sup_{s \in \mathbb{R}^m} |\psi^k(s)| =: R,$$

where the first inequality follows from the triangle inequality. Moreover, $\sup_{s \in \mathbb{R}^m} |\psi^k(s)|$ is finite for every $k = 1, \dots, K$, since by Theorem 3.3 the function ψ^k is bounded for every $k = 1, \dots, K$. ■

Chapter 4

A loose Benders' decomposition algorithm for approximating two-stage mixed-integer recourse models

We propose a new class of convex approximations for two-stage mixed-integer recourse models, the so-called generalized alpha-approximations. The advantage of these convex approximations over existing ones is that they are more suitable for efficient computations. Indeed, we construct a loose Benders' decomposition algorithm that solves large problem instances in reasonable time. To guarantee the performance of the resulting solution, we derive corresponding error bounds that depend on the total variations of the probability density functions of the random variables in the model. The error bounds converge to zero if these total variations converge to zero. We empirically assess our solution method on several test instances, including the SIZES and SSLP instances from SIPLIB. We show that our method finds near-optimal solutions if the variability of the random parameters in the model is large. Moreover, our method outperforms existing methods in terms of computation time, especially for large problem instances.

This chapter is based on the journal publication [77].

4.1 Introduction

Consider the two-stage mixed-integer recourse model with random right-hand side

$$\eta^* := \min_x \left\{ cx + Q(x) : \quad Ax = b,\ x \in X \subseteq \mathbb{R}^{n_1}_+ \right\}, \tag{4.1}$$

where the recourse function Q is defined as

$$Q(x) := \mathbb{E}_\omega \left[\min_y \left\{ qy : \quad Wy = \omega - Tx,\ y \in Y \subseteq \mathbb{R}^{n_2}_+ \right\} \right], \quad x \in X. \tag{4.2}$$

This model represents a two-stage decision problem under uncertainty. In the first stage, a decision x has to be made here-and-now, subject to deterministic constraints $Ax = b$ and random goal constraints $Tx = \omega$. Here, ω is a continuous or discrete random vector whose probability distribution is known. In the second stage, the realization of ω becomes known and any infeasibilities with respect to $Tx = \omega$ have to be repaired. This is modelled by the second-stage problem

$$v(\omega, x) := \min_y \left\{ qy : \quad Wy = \omega - Tx,\ y \in Y \subseteq \mathbb{R}^{n_2}_+ \right\}. \tag{4.3}$$

The objective in this two-stage recourse model is to minimize the sum of immediate costs cx and expected second-stage costs $Q(x) = \mathbb{E}_\omega[v(\omega, x)]$, $x \in X$.

Frequently, integrality restrictions are imposed on the first- and second-stage decisions. That is, X and Y are of the form $X = \mathbb{Z}^{p_1}_+ \times \mathbb{R}^{n_1-p_1}_+$ and $Y = \mathbb{Z}^{p_2}_+ \times \mathbb{R}^{n_2-p_2}_+$. Such restrictions arise naturally when modelling real-life problems, for example to model on/off decisions or batch size restrictions. The resulting model is called a mixed-integer recourse (MIR) model. Such models have many practical applications in for example energy, telecommunication, production planning, and environmental control, see e.g. [32] and [86].

While MIR models are highly relevant in practice, they are notoriously difficult to solve. The reason is that Q is in general non-convex if integrality restrictions are imposed on the second-stage decision variables y, see [54]. Therefore, standard techniques for convex optimization cannot be used to solve these models. In contrast, if $Y = \mathbb{R}^{n_2}_+$, then Q is convex and efficient solution methods are available, most notably the L-shaped method in [84] and variants thereof.

Because of the non-convexity of Q, traditional solution methods for MIR models typically combine ideas from deterministic mixed-integer programming and

stochastic continuous programming, see e.g. [13, 24, 31, 44, 49, 53, 67, 69, 88], and the survey papers [58, 66, 68]. We, however, use a fundamentally different approach to deal with the non-convex recourse function Q. Instead of solving the original MIR model in (4.1), we solve an approximating problem in which Q is replaced by a *convex approximation* $\hat{Q}$ of Q. This convex approximation $\hat{Q}$ does not have to be a lower bound of Q. The resulting approximating optimization problem is given by

$$\hat{\eta} := \min_x \left\{ cx + \hat{Q}(x) : \quad Ax = b,\ x \in X \right\}. \tag{4.4}$$

Because $\hat{Q}$ is convex, we can use efficient techniques from convex optimization to solve the optimization problem in (4). Indeed, this optimization problem is convex if all first-stage variables $x \in X$ are continuous, and it is a MIP with a convex objective if some of these first-stage variables are integer. Thus, compared to traditional solution methods, we are able to solve similar-sized problems much faster, and solve larger problem instances. In fact, we demonstrate this in our numerical experiments on problem instances available from [67] and SIPLIB [5], and on randomly generated problem instances.

Obviously, the optimal solution $\hat{x}$ of the approximating problem in (4.4) is not necessarily optimal for the original MIR model in (4.1). That is why we guarantee the quality of the approximating solution $\hat{x}$, by deriving an error bound on

$$||Q - \hat{Q}||_\infty := \sup_x |Q(x) - \hat{Q}(x)|.$$

This error bound directly gives us an upper bound on the optimality gap of $\hat{x}$:

$$c\hat{x} + Q(\hat{x}) - \eta^* \leq 2||Q - \hat{Q}||_\infty,$$

see [59].

Convex approximations and corresponding error bounds have been derived for many different classes of models. The idea to use convex approximations $\hat{Q}$ for the non-convex mixed-integer recourse function Q dates back to [80], in which van der Vlerk proposes to use α-approximations for the special case of simple integer recourse (SIR) models. These α-approximations are obtained by perturbing the probability distribution of the random vector ω. Klein Haneveld et al. [41] derive an error bound for the α-approximations that depends on the total variations of the marginal density functions of the random variables in the SIR model.

More convex approximations have been described for more general classes of problems. For example, in [83] and [81], van der Vlerk extends the α-approximations to a class of MIR models with a single recourse constraint, and to integer recourse models with a totally unimodular (TU) recourse matrix, respectively. Furthermore, Romeijnders et al. [61] derive an error bound for the latter approximation, and they derive a tighter error bound for the *shifted LP-relaxation* approximation for the same class of problems. The quality of the convex approximations for TU integer recourse models is assessed empirically in [62], and it turns out that they perform well if the variability of the random parameters in the model is large enough.

Romeijnders et al. [60] generalize the shifted LP-relaxation to general two-stage MIR models, and they derive a corresponding *asymptotic error bound*. This error bound converges to zero if the variability of the random parameters in the model increases. In Chapter 3, we derive similar error bounds for convex approximations that fit into a specific framework. In this framework, convex approximations are defined using *pseudo-valid cutting planes* for the second-stage feasible regions. Their idea is to only use cutting planes which are affine in the first-stage decision variables, so that the corresponding expected value function is convex. In general, the approximations are not exact, since pseudo-valid cutting planes cut away feasible solutions and are allowed to be overly conservative. Nevertheless, a similar error bound as for the shifted LP-relaxation has been derived if the pseudo-valid cutting planes are *tight*. That is, if they are exact on an entire grid of first-stage solutions.

Both the shifted LP-relaxation of [60] and the cutting plane framework of Chapter 3, however, cannot be applied directly to efficiently solve MIR models in general. This is because the shifted LP-relaxation is very difficult to compute in general, as discussed in [60], and furthermore, the asymptotic error bound in the cutting plane framework of Chapter 3 only applies for tight pseudo-valid cutting planes, which are only available in special cases, e.g. for SIR models. That is why we propose an alternative class of convex approximations for general two-stage MIR models, the so-called generalized α-approximations. They are derived by exploiting properties of Gomory relaxations [34] of the second-stage mixed-integer programming problems.

Contrary to the shifted LP-relaxation, the generalized α-approximations, denoted by $\hat{Q}_\alpha$, can be solved efficiently. In fact, we develop a so-called loose Benders' decomposition algorithm to solve the approximating model in (4.4) with $\hat{Q} = \hat{Q}_\alpha$. Our Benders' decomposition is called loose, because we derive optimality cuts for $\hat{Q}_\alpha$ which are in general not tight at the current solution. While these loose

optimality cuts are in general not sufficient to find the optimal solution $\hat{x}$ of the approximating model in (4.4), we prove that they are tight enough in the sense that a similar performance guarantee applies to the solution obtained by the loose Benders' decomposition as to $\hat{x}$.

Summarizing, our main contributions are as follows.

- We propose a new class of convex approximations for general two-stage MIR models, which are based on Gomory relaxations of the second-stage problems. These generalized α-approximations can be solved efficiently, and a similar error bound as for the shifted LP-relaxation applies to the generalized α-approximations.
- We derive a loose Benders' decomposition algorithm to (approximately) solve the approximating model with the generalized α-approximations. This is the first efficient algorithm for solving non-trivial convex approximations of general two-stage MIR models.
- We prove that the solution obtained by the loose Benders' decomposition algorithm has a similar performance guarantee as the exact solution to the generalized α-approximations.
- We carry out extensive numerical experiments on 41 test instances from the literature and on 240 randomly generated instances, and show that using our loose Benders' decomposition algorithm we obtain good solutions within reasonable time, also for large problem instances, in particular when the variability of the random parameters in the model is large.

The remainder of this chapter is organized as follows. In Section 4.2, we discuss the shifted LP-relaxation of [60] and its corresponding error bound. In Section 4.3, we present the generalized α-approximations and an efficient algorithm for solving the corresponding approximating problem. Section 4.4 contains the proof of the performance guarantee for the loose Benders' decomposition algorithm. In Section 4.5, we report on numerical experiments to evaluate the performance of our algorithm. Finally, we conclude in Section 4.6.

Throughout, we make the following assumptions. Assumptions (A2)-(A4) guarantee that $Q(x)$ is finite for all $x \in X$ such that $Ax = b$.

(A1) The first-stage feasible region $\mathcal{X} := \{x \in X : Ax = b\}$ is bounded.

(A2) The recourse is relatively complete: for all $\omega \in \mathbb{R}^m$ and $x \in \mathcal{X}$, there exists a $y \in Y$ such that $Wy = \omega - Tx$, so that $v(\omega, x) < \infty$.

(A3) The recourse is sufficiently expensive: $v(\omega, x) > -\infty$ for all $\omega \in \mathbb{R}^m$ and $x \in \mathcal{X}$.

(A4) $\mathbb{E}[|\omega_i|]$ is finite for all $i = 1, \ldots, m$.

(A5) The recourse matrix W is integer.

4.2 Existing convex approximations of mixed-integer recourse functions

In this section, we review the shifted LP-relaxation approximation of [60] and its corresponding error bound. First, however, we review results on asymptotic periodicity in mixed-integer linear programming in Section 4.2.1. We do so since these results are not only used to derive the shifted LP-relaxation in Section 4.2.2, but also to derive the generalized α-approximations in Section 4.3.1.

4.2.1 Asymptotic periodicity in mixed-integer programming

In order to derive convex approximations of the MIR function Q we analyze the value function v of the second-stage problem, defined as

$$v(\omega, x) = \min_y \left\{ qy : \quad Wy = \omega - Tx,\ y \in \mathbb{Z}_+^{p_2} \times \mathbb{R}_+^{n_2 - p_2} \right\}. \tag{4.5}$$

In particular, LP-duality implies that the LP-relaxation $v_{\mathrm{LP}}(\omega, x)$ of $v(\omega, x)$ is polyhedral in the right-hand side vector $\omega - Tx$:

$$v_{\mathrm{LP}}(\omega, x) = \max_{k=1,\ldots,K} \lambda^k(\omega - Tx),$$

where $\lambda^k, k = 1 \ldots, K$, are the extreme points of the dual feasible region $\{\lambda : \lambda W \leq q\}$. Romeijnders et al. [60] derive a similar characterization of $v(\omega, x)$ in terms of linear and periodic functions by exploiting so-called Gomory relaxations. We briefly discuss these Gomory relaxation, before stating the characterization of $v(\omega, x)$ in Lemma 4.1.

The Gomory relaxation of $v(\omega, x)$ is defined for any dual feasible basis matrix of $v_{\mathrm{LP}}(\omega, x)$. Let B denote such a matrix and let N be such that $W \equiv \begin{pmatrix} B & N \end{pmatrix}$, meaning equality up to a permutation of the columns. Let y_B and y_N denote the second-stage variables corresponding to the columns in B and N, respectively, and

q_B and q_N their corresponding cost parameters. The Gomory relaxation v_B is obtained by *relaxing the non-negativity constraints* of the basic variables y_B. Romeijnders et al. [60] derive the following expression for v_B:

$$v_B(\omega, x) = q_B B^{-1}(\omega - Tx) + \psi_B(\omega - Tx),$$

where

$$\psi_B(s) := \min\Big\{\bar{q}_N y_N : \quad B^{-1}(s - Ny_N) \in \mathbb{Z}^{p_B} \times \mathbb{R}^{n_B}, \\ y_N \in \mathbb{Z}_+^{p_N} \times \mathbb{R}_+^{n_N}\Big\}, \quad s \in \mathbb{R}^m, \tag{4.6}$$

and $\bar{q}_N = q_N - q_B(B^{-1})N$. Moreover, they show that

$$v(\omega, x) = v_B(\omega, x) = q_B B^{-1}(\omega - Tx) + \psi_B(\omega - Tx)$$

if $\omega - Tx \in \Lambda := \{t : B^{-1}t \geq 0\}$, and if the distance of $\omega - Tx$ to the boundary of Λ is sufficiently large, see Definition 4.1. We say that $v(\omega, x)$ is asymptotically periodic, since ψ_B is a B-periodic function, i.e., $\psi_B(s + Bl) = \psi_B(s)$ for every $s \in \mathbb{R}^m$ and $l \in \mathbb{Z}^m$, see [60].

Definition 4.1. Let $\Lambda \subset \mathbb{R}^m$ be a closed convex set and let $d \in \mathbb{R}$ with $d > 0$ be given. Then, we define $\Lambda(d)$ as

$$\Lambda(d) := \{s \in \Lambda : \mathcal{B}(s, d) \subset \Lambda\},$$

where $\mathcal{B}(s, d) := \{t \in \mathbb{R}^m : ||t - s||_2 \leq d\}$ is the closed ball centered at s with radius d. We can interpret $\Lambda(d)$ as the set of points in Λ with at least Euclidean distance d to the boundary of Λ.

Lemma 4.1. [60, Theorem 2.9] *Consider the mixed-integer programming problem*

$$v(\omega, x) := \min_y \Big\{qy : \quad Wy = \omega - Tx,\ y \in \mathbb{Z}_+^{p_2} \times \mathbb{R}_+^{n_2 - p_2}\Big\},$$

where W is an integer matrix, and $v(\omega, x)$ is finite for all $\omega \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$. Then, there exist dual feasible basis matrices B^k of v_{LP}, $k = 1, \ldots, K$, closed convex polyhedral cones $\Lambda^k := \{t \in \mathbb{R}^m : (B^k)^{-1}t \geq 0\}$, distances d_k, B^k-periodic functions ψ^k, and constants w_k such that we have the following:

(i) $\bigcup_{k=1}^{K} \Lambda^k = \mathbb{R}^m$.

(ii) $(\text{int}\,\Lambda^k) \cap (\text{int}\,\Lambda^l) = \emptyset$ *for every* $k, l \in \{1, \ldots, K\}$ *with* $k \neq l$.

(iii) If $\omega - Tx \in \Lambda^k(d_k)$, *then*

$$v(\omega, x) = v_{LP}(\omega, x) + \psi^k(\omega - Tx) = q_{B^k}(B^k)^{-1}(\omega - Tx) + \psi^k(\omega - Tx),$$

where $\psi^l \equiv \psi^k$ *if* $q_{B^k}(B^k)^{-1} = q_{B^l}(B^l)^{-1}$.

(iv) $0 \leq \psi^k(\omega - Tx) \leq w_k$ *for all* $\omega \in \mathbb{R}^m$ *and* $x \in \mathbb{R}^{n_1}, \quad k = 1, \ldots, K$.

Lemma 4.1 (iii) shows that if $\omega - Tx \in \Lambda^k(d_k)$ for some $k = 1, \ldots, K$, then $v(\omega, x)$ is equal to the sum of the LP-relaxation $v_{\text{LP}}(\omega, x)$ and $\psi^k(\omega - Tx)$. Hence, $\psi^k(\omega - Tx)$ can be interpreted as the additional costs resulting from the integrality restrictions on the decision variables y. For a discussion on how to obtain d_k or how to represent $\Lambda^k(d_k)$ using a system of linear inequalities we refer to [60].

4.2.2 The shifted LP-relaxation approximation

Lemma 4.1 shows why the second-stage value function is not convex in x. On regions of its domain it is the sum of a linear function $q_{B^k}(B^k)^{-1}(\omega - Tx)$ and a periodic function $\psi^k(\omega - Tx)$. Clearly the periodic part is causing v to be non-convex. That is why the shifted LP-relaxation is obtained by replacing this periodic part $\psi^k(\omega - Tx)$ by a constant Γ_k for every $k = 1, \ldots, K$, with Γ_k defined as

$$\Gamma_k := p_k^{-m} \int_0^{p_k} \cdots \int_0^{p_k} \psi^k(x) dx_1 \ldots dx_m, \tag{4.7}$$

where $p_k = |\det B_k|$. The K constants Γ_k can be interpreted as the averages of the periodic functions ψ^k. The shifted LP-relaxation approximation is obtained by taking the pointwise maximum over all dual feasible basis matrices $B^k, k = 1, \ldots, K$.

Definition 4.2. Define the *shifted LP-relaxation approximation* $\tilde{Q}$ of the MIR function Q as $\tilde{Q}(x) = \mathbb{E}_\omega[\tilde{v}(\omega, x)]$, where

$$\tilde{v}(\omega, x) := \max_{k=1,\ldots,K} \{q_{B^k}(B^k)^{-1}(\omega - Tx) + \Gamma_k\},$$

where B^k, $k = 1, \ldots, K$, are the dual feasible basis matrices of Lemma 4.1, and Γ_k is defined in (4.7).

Romeijnders et al. [60] derive a total variation error bound on the approximation error $||Q - \tilde{Q}||_\infty$ of the shifted LP-relaxation $\tilde{Q}$. This error bound is expressed in terms of the *total variations* of the one-dimensional conditional probability density functions (pdf) of the random vector ω.

Definition 4.3. Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function, and let $I \subset \mathbb{R}$ be an interval. Let $\Pi(I)$ denote the set of all finite ordered sets $P = \{x_1, \ldots, x_{N+1}\}$ with $x_1 < \cdots < x_{N+1}$ in I. Then, the *total variation* of f on I, denoted by $|\Delta|f(I)$, is defined as

$$|\Delta|f(I) = \sup_{P \in \Pi(I)} V_f(P),$$

where

$$V_f(P) = \sum_{i=1}^{N} |f(x_{i+1}) - f(x_i)|.$$

We write $|\Delta|f := |\Delta|f(\mathbb{R})$.

Definition 4.4. For every $i = 1, \ldots, m$ and $x_{-i} \in \mathbb{R}^{m-1}$, define the i-th conditional density function $f_i(\cdot|x_{-i})$ of the m-dimensional joint pdf f as

$$f_i(x_i|x_{-i}) = \frac{f(x)}{f_{-i}(x_{-i})},$$

where f_{-i} is the joint pdf of the $(m-1)$-dimensional random vector ω_{-i}, which is equal to ω without its i-th component. Define $\mathcal{H}^m$ as the set of all m-dimensional joint pdf f such that $f_i(\cdot|x_{-i})$ is of bounded variation for all $i = 1, \ldots, m$ and $x_{-i} \in \mathbb{R}^{m-1}$.

Theorem 4.1. [60, Theorem 5.1] *There exists a constant $C > 0$ such that for every continuous random vector ω with joint pdf $f \in \mathcal{H}^m$,*

$$\sup_x |Q(x) - \tilde{Q}(x)| \leq C \sum_{i=1}^{m} \mathbb{E}_{\omega_{-i}} \left[|\Delta| f_i(\cdot|\omega_{-i}) \right]. \tag{4.8}$$

In general, the error bound in Theorem 4.1 may be large, in particular since the constant $C > 0$ may be large. Nevertheless, the theorem shows that if the

total variations of the one-dimensional conditional pdf are small, then $\tilde{Q}$ is a good approximation of Q. For example, if the components ω_i of ω follow independent normal distributions, with mean μ_i and variance σ_i^2, for $i = 1, \ldots, m$, then the error bound in (4.8) simplifies to $C' \sum_{i=1}^{m} \sigma_i^{-1}$ for some $C' > 0$; see Example 3.2 for details. Observe that the error bound goes to zero if $\sigma_i \to \infty$ for all $i = 1, \ldots, m$. Thus, the error of using the shifted LP-relaxation approximation decreases if the variability of the random parameters in the model increases.

4.3 Loose Benders' decomposition algorithm for two-stage mixed-integer recourse models

4.3.1 Generalized α-approximations

To derive the generalized α-approximations $\hat{Q}_\alpha$, we first derive a convex approximation $\hat{v}_\alpha$ of the second-stage value function v defined in (4.5), and we define $\hat{Q}_\alpha(x) := \mathbb{E}_\omega[\hat{v}_\alpha(\omega, x)]$. Similar as for the shifted LP-relaxation $\tilde{v}$, we use Lemma 4.1, i.e., for $\omega - Tx \in \Lambda^k(d_k)$,

$$v(\omega, x) = v_{B^k}(\omega, x) = \; q_{B^k}(B^k)^{-1}(\omega - Tx) + \psi^k(\omega - Tx).$$

However, instead of replacing $\psi^k(\omega - Tx)$ by its average Γ_k, as is done for the shifted LP-relaxation, we replace $\psi^k(\omega - Tx)$ by $\psi^k(\omega - \alpha)$ for some $\alpha \in \mathbb{R}^m$ to obtain the generalized α-approximation $\hat{v}_\alpha^k$ of v_{B^k} defined as

$$\hat{v}_\alpha^k(\omega, x) = q_{B^k}(B^k)^{-1}(\omega - Tx) + \psi^k(\omega - \alpha),$$

and we define the generalized α-approximation $\hat{v}_\alpha$ of v as

$$\hat{v}_\alpha(\omega, x) = \max_k \hat{v}_\alpha^k(\omega, x).$$

The difference compared to the shifted LP-relaxation seems small: the constants Γ_k are replaced by $\psi^k(\omega - \alpha)$. From a computational point of view, however, this difference is significant. This is because the constants Γ_k are the *averages* over ψ^k, and in general need to be obtained by computing a multi-dimensional integral of a mixed-integer value function. For a fixed ω and α, however, the value of $\psi^k(\omega - \alpha)$ is obtained by solving *a single* mixed-integer programming problem of the same size as the second-stage problem. In fact, we need to solve the Gomory

relaxation discussed in Section 4.2.1, which can be done in polynomial time if all second-stage variables are integer [34].

Definition 4.5. For $\alpha \in \mathbb{R}^m$, we define the *generalized α-approximation* $\hat{Q}_\alpha$ of Q as

$$\hat{Q}_\alpha(x) := \mathbb{E}_\omega \left[\max_{k=1,\dots,K} \{\lambda^k(\omega - Tx) + \psi^k(\omega - \alpha)\} \right], \quad x \in \mathbb{R}^{n_1},$$

with $\lambda^k := q_{B^k}(B^k)^{-1}$ and $\psi^k := \psi_{B^k}$, where B^k, $k = 1, \dots, K$, are the dual feasible basis matrices of Lemma 4.1.

Remark 4.1. The generalized α-approximations are a generalization of the α-approximations defined for *TU integer recourse models* [61], which arise if $y \in \mathbb{Z}^p_+$ and if the second-stage constraints are of the form $Wy \geq \omega - Tx$, where W is a TU matrix. Indeed, for these models, we have $\psi^k(s) = \lambda^k(\lceil s \rceil - s)$, see [60], and thus

$$\hat{Q}_\alpha(x) = \mathbb{E}_\omega \left[\max_{k=1,\dots,K} \lambda^k(\lceil \omega - \alpha \rceil + \alpha - Tx) \right],$$

which is the expression for the α-approximations for TU integer recourse models in [61].

It turns out that we can derive a similar error bound for the generalized α-approximations as for the shifted LP-relaxation.

Theorem 4.2. *There exists a constant $C > 0$, such that for every $\alpha \in \mathbb{R}^m$ and for every continuous random vector ω with joint pdf $f \in \mathcal{H}^m$,*

$$\sup_x |Q(x) - \hat{Q}_\alpha(x)| \leq C \sum_{i=1}^m \mathbb{E}_{\omega_{-i}} \left[|\Delta| f_i(\cdot|\omega_{-i}) \right].$$

Proof. See Appendix 4.A. ■

Theorem 4.2 states that the approximation error of $\hat{Q}_\alpha$ goes to zero as the total variations $\mathbb{E}_{\omega_{-i}} [|\Delta| f_i(\cdot|\omega_{-i})]$, $i = 1, \dots, m$, all go to zero. This error bound is independent of α, i.e., $\hat{Q}_\alpha$ is a good approximation of the true MIR function Q for any $\alpha \in \mathbb{R}^m$.

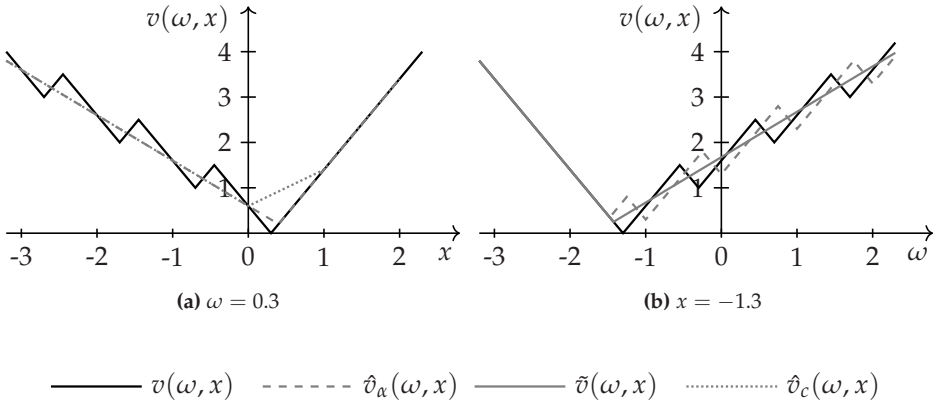
An interesting difference between the generalized α-approximations and the shifted LP-relaxation of Section 4.2.2 is that the approximating value function $\hat{v}_\alpha$ is not convex in ω for fixed $x \in \mathbb{R}^{n_1}$. Indeed, $\hat{v}_\alpha$ is only convex in x for every fixed ω, but this is sufficient to guarantee that the generalized α-approximation $\hat{Q}_\alpha$ is convex. In contrast, the value function $\tilde{v}$ of the shifted LP-relaxation is convex

in both ω and x. We illustrate these properties in Example 4.1. In this example, we also illustrate the difference between the generalized α-approximations and the pseudo-valid cutting plane approximation of Chapter 3.

Example 4.1. Consider the second-stage mixed-integer value function v defined as

$$v(\omega, x) := \min\{y_1 + 2y_2 + 2y_3 : \ y_1 + y_2 - y_3 = \omega - x, \ y_1 \in \mathbb{Z}_+, \ y_2, \ y_3 \in \mathbb{R}_+\}.$$

We use this example, since it also appears in [60] and in Chapter 3, allowing us to compare the generalized α-approximations to the shifted LP-relaxation and a pseudo-valid cutting plane approximation.



(a) $\omega = 0.3$ **(b)** $x = -1.3$

Figure 4.1. The value function v and the approximating value functions $\tilde{v}$, $\hat{v}_c$, and $\hat{v}_\alpha$, where $\alpha = 0$. The left figure shows $v(\omega, x)$, $\hat{v}_\alpha(\omega, x)$, and $\hat{v}_c(\omega, x)$, as a function of x, whereas the right figure shows $v(\omega, x)$, $\hat{v}_\alpha(\omega, x)$, and $\tilde{v}(\omega, x)$ as a function of ω.

The LP-relaxation of v has two dual feasible basis matrices $B^1 = [-1]$ and $B^2 = [1]$. Thus, $K = 2$, and straightforward computations yield $\lambda^1 = -2$, $\lambda^2 = 1$, $\psi^1 \equiv 0$, and for every $s \in \mathbb{R}$,

$$\psi^2(s) = \begin{cases} s - \lfloor s \rfloor, & \text{if } s - \lfloor s \rfloor \leq 3/4, \\ 3 - 3(s - \lfloor s \rfloor), & \text{if } s - \lfloor s \rfloor \geq 3/4. \end{cases}$$

It follows that $v_{\mathrm{LP}}(\omega, x) = \max\{-2(\omega - x), \omega - x\}$. Moreover, for every $\alpha \in \mathbb{R}$, the approximating value function is defined as

$$\hat{v}_\alpha(\omega, x) = \max\{-2(\omega - x), \omega - x + \psi^2(\omega - \alpha)\}.$$

In contrast, the value function $\tilde{v}$ of the shifted LP-relaxation equals

$$\tilde{v}(\omega, x) = \max\{-2(\omega - x), \omega - x + 3/8\},$$

since $\Gamma^2 := \int_0^1 \psi^2(s)ds = 3/8$. Finally, the tight pseudo-valid cutting plane approximation $\hat{v}_c$ is given by

$$\hat{v}_c(\omega, x) = \min_{y \in \mathbb{R}^3_+} \{y_1 + 2y_2 + 2y_3 : y_1 + y_2 - y_3 = \omega - x, \\ y_1 - \frac{1}{1 + \lfloor\omega\rfloor - \omega} y_3 \leq \lfloor\omega\rfloor - x\}.$$

Figures 4.1a and 4.1b show $v(\omega, x)$, $\hat{v}_\alpha(\omega, x)$, $\tilde{v}(\omega, x)$, and $\hat{v}_c(\omega, x)$ as a function of x and ω, respectively. They illustrate that $\hat{v}_\alpha$ is convex in x for fixed ω, but not convex in ω for fixed x, respectively. Moreover, Figure 4.1b illustrates a key difference between $\tilde{v}(\omega, x)$ and $\hat{v}_\alpha(\omega, x)$. Whereas $\tilde{v}(\omega, x)$ is obtained by replacing the periodic part $\psi^2(\omega - x)$ of $v(\omega, x)$ by its average Γ^2, we obtain $\hat{v}_\alpha(\omega, x)$ by *shifting* $\psi^2(\omega - x)$ towards $\psi^2(\omega - \alpha)$.

Finally, Figure 4.1a shows that $\hat{v}_\alpha(\omega, x) = \hat{v}_c(\omega, x)$ for a range of x values. However, the difference $\hat{v}_c(\omega, x) - \hat{v}_\alpha(\omega, x)$ is relatively large if $x \in (-0.3, 0.7)$. This is true in general for tight pseudo-valid cutting plane approximations: there always exist $\hat{d}_k$ such that $\hat{v}_\alpha(\omega, x) = \hat{v}_c(\omega, x)$ if $\omega - Tx \in \Lambda^k(\hat{d}_k)$. However, the two approximations may differ if $\omega - Tx \notin \bigcup_k \Lambda^k(\hat{d}_k)$. ◊

4.3.2 Benders' decomposition for the generalized α-approximations

We solve the approximating problem

$$\hat{\eta}_\alpha := \min_x \{cx + \hat{Q}_\alpha(x) : \quad Ax = b,\ x \in X\}, \tag{4.9}$$

using an SAA of $\hat{Q}_\alpha$. Given a sample $\{\omega^1, \ldots, \omega^S\}$ of size S from the distribution of ω, the SAA $\hat{Q}^S_\alpha$ of $\hat{Q}_\alpha$ is defined as

$$\hat{Q}^S_\alpha(x) := \frac{1}{S} \sum_{s=1}^{S} \hat{v}_\alpha(\omega^s, x), \quad x \in X, \tag{4.10}$$

and the corresponding SAA problem as

$$\hat{\eta}^S_\alpha := \min_x \{cx + \hat{Q}^S_\alpha(x) : \quad Ax = b,\ x \in X\}. \tag{4.11}$$

Since $\hat{Q}_\alpha^S$ is convex, we can solve the SAA problem using Benders' decomposition [15], i.e., using an L-shaped algorithm [84]. In iteration τ of this algorithm, we solve (4.11) with $\hat{Q}_\alpha^S$ replaced by a convex polyhedral *outer approximation* $\hat{Q}_{\text{out}}^\tau \leq \hat{Q}_\alpha^S$. This problem is called the *master problem*

$$\hat{\eta}_\tau := \min_x \{cx + \hat{Q}_{\text{out}}^\tau(x) : \quad Ax = b,\ x \in X\}, \tag{4.12}$$

and its optimal solution x^τ is referred to as the *current solution* at iteration τ. If $\hat{Q}_{\text{out}}^\tau(x^\tau) = \hat{Q}_\alpha^S(x^\tau)$, then the current solution x^τ is optimal for the original SAA problem in (4.11). If not, then we strengthen the outer approximation using an *optimality cut* for $\hat{Q}_\alpha^S$ of the form

$$\hat{Q}_\alpha^S(x) \geq \beta_{\tau+1}x + \delta_{\tau+1}, \quad \forall x \in X,$$

which is tight at x^τ, i.e., $\hat{Q}_\alpha^S(x^\tau) = \beta_{\tau+1}x^\tau + \delta_{\tau+1}$. The outer approximation $\hat{Q}_{\text{out}}^\tau$ in iteration τ is the pointwise maximum over all optimality cuts derived in previous iterations:

$$\hat{Q}_{\text{out}}^\tau(x) = \max_{t=1,\ldots,\tau} \{\beta_t x + \delta_t\}, \quad x \in X.$$

The challenge for the generalized α-approximations, however, is to compute tight optimality cuts of $\hat{Q}_\alpha^S$. Since

$$\hat{Q}_\alpha^S(x^\tau) = \frac{1}{S}\sum_{s=1}^{S} \max_{k=1,\ldots,K} \{\lambda^k(\omega^s - Tx^\tau) + \psi^k(\omega^s - \alpha),\} \quad x \in X,$$

such tight optimality cuts can be obtained by identifying for each scenario s the maximizing index k_s^τ at x^τ, defined as

$$k_s^\tau \in \arg\max_{k=1,\ldots,K} \{\lambda^k(\omega^s - Tx^\tau) + \psi^k(\omega^s - \alpha)\}, \quad s = 1, \ldots, S. \tag{4.13}$$

However, this is computationally too expensive, since we need to compute $\psi^k(\omega^s - \alpha)$ for all $k = 1, \ldots, K$, and K grows *exponentially* in the size of the second-stage problem. Indeed, K is at least as large as the number of dual vertices of the feasible region of the LP-relaxation v_{LP} of v, of which there are exponentially many.

4.3.3 Loose Benders' decomposition for the generalized α-approximations

To overcome this computational challenge we propose to use approximate indices $\hat{k}_s^\tau$ defined as

$$\hat{k}_s^\tau \in \arg\max_{k=1,\dots,K} \lambda^k(\omega^s - Tx^\tau), \quad s = 1,\dots,S. \tag{4.14}$$

These indices are *computationally tractable* since they correspond to the optimal basis matrix index of the LP-relaxation $v_{\text{LP}}(\omega^s, x^\tau)$, defined in (4.5). Hence, they can be obtained by solving a single LP. Moreover, if the values of $\psi^k(\omega^s - \alpha)$ are equal for all $k = 1,\dots,K$, then the indices $\hat{k}_s^\tau$ are optimal in (4.13), i.e., $\hat{k}_s^\tau = k_s^\tau$. In general however, the indices k_s are suboptimal in (4.13), leading to the following definition.

Definition 4.6. Let x^τ be given and let $\hat{k}_s^\tau$, $s = 1,\dots,S$, be as in (4.14). We define the *loose optimality cut* for $\hat{Q}_\alpha^S$ at x^τ as

$$\hat{Q}_\alpha^S(x) \geq \frac{1}{S}\sum_{s=1}^{S} \lambda^{\hat{k}_s^\tau}(\omega^s - Tx) + \psi^{\hat{k}_s^\tau}(\omega^s - \alpha), \quad x \in X.$$

We use these loose optimality cuts in our loose Benders' decomposition algorithm, LBDA(α). In this algorithm the outer approximation $\hat{Q}_{\text{out}}^\tau$ of $\hat{Q}_\alpha^S$ is defined using our loose optimality cuts. The algorithm terminates with tolerance level $\varepsilon \geq 0$ if $\hat{Q}_{\text{out}}^\tau(x^\tau) \geq \hat{Q}_{\text{out}}^{\tau+1}(x^\tau) - \varepsilon$; then, LBDA($\alpha$) reports $\hat{x}_\alpha = x^\tau$ as solution.

A full description of LBDA(α) is given below. Note that the algorithm requires a lower bound L on $\hat{Q}_\alpha^S$ such that $\hat{Q}_\alpha^S(x) \geq L$ for all $x \in \mathcal{X}$. Such a lower bound L exists because of Assumptions (A1)-(A4).

Loose Benders' decomposition algorithm (LBDA(α))

1: **Inputs** Parameters: A, b, c, T, q, W. Distribution of ω. Lower bound L on $\hat{Q}_\alpha^S$. Shift parameter α. Tolerance ε. Sample size S.
2: **Output** Near-optimal solution $\hat{x}_\alpha$.
3: **Initialization**
4: Initialize $\tau = 0$ and $\hat{Q}_{\text{out}}^\tau \equiv L$.
5: Obtain a sample $\{\omega^1,\dots,\omega^S\}$ of size S from the distribution of ω.
6: **Iteration step**
7: Solve $\min_x\{cx + \hat{Q}_{\text{out}}^\tau(x) : Ax = b,\ x \in X\}$.
8: Denote the optimal solution by x^τ.
9: **for** $s = 1,\dots,S$ **do**

10: Solve $v_{\mathrm{LP}}(\omega^s, x^\tau)$.
11: Denote the optimal basis matrix index by $\hat{k}_s^\tau$.
12: Solve $v_{B^{\hat{k}_s^\tau}}(\omega^s - \alpha, 0)$.
13: Compute $\psi^{\hat{k}_s^\tau}(\omega^s - \alpha) = v_{B^{\hat{k}_s^\tau}}(\omega^s - \alpha, 0) - \lambda^{\hat{k}_s^\tau}(\omega^s - \alpha)$.
14: **end for**
15: $\beta_{\tau+1} \leftarrow -\frac{1}{S}\sum_{s=1}^{S} \lambda^{\hat{k}_s^\tau} T$.
16: $\delta_{\tau+1} \leftarrow \frac{1}{S}\sum_{s=1}^{S} \left\{\lambda^{\hat{k}_s^\tau}\omega^s + \psi^{\hat{k}_s^\tau}(\omega^s - \alpha)\right\}$.
17: $\hat{Q}_{\mathrm{out}}^{\tau+1}(x) := \max\{\hat{Q}_{\mathrm{out}}^{\tau}(x), \beta_{\tau+1}x + \delta_{\tau+1}\}$.
18: **Stopping criterion**
19: **if** $\hat{Q}_{\mathrm{out}}^{\tau}(x^\tau) \geq \hat{Q}_{\mathrm{out}}^{\tau+1}(x^\tau) - \varepsilon$ **then**
20: **return** $\hat{x}_\alpha := x^\tau$.
21: **stop**.
22: **else**
23: $\tau \leftarrow \tau + 1$. Go to line 7.
24: **end if**

The solution $\hat{x}_\alpha$ of LBDA(α) is not necessarily ε-optimal for the SAA problem in (4.11), since LBDA(α) uses the loose optimality cuts of Definition 4.6. If the optimality cuts were tight, then $\hat{Q}_{\mathrm{out}}^{\tau+1}(x^\tau) = \hat{Q}_\alpha^S(x^\tau)$ for every iteration τ, and thus the algorithm would terminate if $\hat{Q}_{\mathrm{out}}^{\tau}(x^\tau) \geq \hat{Q}_\alpha^S(x^\tau) - \varepsilon$. We, however, show that our loose optimality cuts are *asymptotically tight* at x^τ: the difference $\hat{Q}_{\mathrm{out}}^{\tau+1}(x^\tau) - \hat{Q}_\alpha^S(x^\tau)$ converges to zero if the total variations of the one-dimensional conditional pdf of the random vector ω go to zero and as $S \to \infty$, see Proposition 4.1. In this case, we are able to prove that the LBDA(α) solution $\hat{x}_\alpha$ is near-optimal; see Theorem 3 below for a bound on the optimality gap $c\hat{x}_\alpha + Q(\hat{x}_\alpha) - \eta^*$. This performance guarantee is independent of α, and thus it applies if we take, e.g., $\alpha = 0$. We further explore selection of α in Section 4.5. The proof of the performance guarantee is postponed to Section 4.4.2.

Theorem 4.3. *Consider the two-stage mixed-integer recourse model*

$$\eta^* = \min_x \{cx + Q(x) : \quad Ax = b,\ x \in X\}.$$

Let $\hat{x}_\alpha$ denote the solution by LBDA(α) with tolerance ε and sample size S. Then, there exists a constant $C > 0$ such that for every continuous random vector ω with joint pdf

$f \in \mathcal{H}^m$

$$c\hat{x}_\alpha + Q(\hat{x}_\alpha) - \eta^* \leq \varepsilon + C \sum_{i=1}^{m} \mathbb{E}_{\omega_{-i}} \left[|\Delta| f_i(\cdot|\omega_{-i}) \right],$$

w.p. 1 *as* $S \to \infty$.

Theorem 4.3 implies that the optimality gap of $\hat{x}_\alpha$ converges to the prespecified tolerance ε as the total variations of the underlying one-dimensional conditional pdf go to zero.

4.3.4 Implementation details of LBDA(α)

LBDA(α) can be implemented efficiently if the input size of the second-stage problem $v(\omega, x)$ is moderate. During each iteration τ, we have to solve the LP-relaxation $v_{\text{LP}}(\omega^s, x^\tau)$ and the Gomory relaxation $v_{B^{k_s^\tau}}(\omega^s - \alpha, 0)$ for each scenario $s = 1, \ldots, S$, in order to generate a loose optimality cut. If the input size of $v(\omega, x)$ is not too large, then $v_{\text{LP}}(\omega^s, x^\tau)$ and $v_{B^{k_s^\tau}}(\omega^s - \alpha, 0)$ can be solved in reasonable time using standard LP and MIP solvers, respectively. Moreover, the master problem can be solved efficiently using a standard LP solver, or MIP solver if some of the first-stage decision variables are integer. Improved implementations of LBDA(α) using a multicut approach [17] and regularization techniques [64] are possible. Furthermore, the subproblems $v_{B^{k_s^\tau}}(\omega^s - \alpha, 0)$, $s = 1, \ldots, S$ can be solved in parallel, and it may be beneficial to solve these subproblems inexactly in the first phase of the algorithm.

In Section 4.5, we exploit that LBDA(α) can be run multiple times, using different values of α. The performance guarantee of LBDA(α) is independent of α and thus applies to every candidate solution that we obtain in this way. Moreover, we use in- and out-of-sample evaluation to select the best candidate solution. In our numerical experiments, we investigate several schemes for selecting the values of α. Running LBDA(α) multiple times can be done very efficiently by using parallelization and a *common warm start*. That is, we first apply the L-shaped algorithm of [84] to the LP-relaxation of the original problem (4.1), in which the integer restrictions on the second-stage variables y are relaxed. Next, we run LBDA(α) for each value of α, and we keep the optimality cuts generated for the LP-relaxation Q_{LP} of Q, which are also valid for $\hat{Q}_\alpha$, for any value of α.

4.4 Performance guarantee of LBDA(α)

4.4.1 Convergence of sampling and loose optimality cuts

The performance guarantee of LBDA(α) does not follow directly from our error bound for the generalized α-approximations. The reason is that LBDA(α) uses sampling and the loose optimality cuts of Definition 4.6 to solve the corresponding approximating problem. We consider these aspects in Sections 4.4.1.1 and 4.4.1.2, respectively. In particular, we prove consistency of the SAA and asymptotic tightness of our loose optimality cuts.

4.4.1.1 Consistency of the sample average approximation

Intuitively, the SAA becomes better as the sample size S increases. Indeed, we show in Lemma 4.2 that the SAA $\hat{Q}_\alpha^S$ of $\hat{Q}_\alpha$ converges uniformly to $\hat{Q}_\alpha$ on $\mathcal{X}$ w.p. 1 as $S \to \infty$.

Lemma 4.2. *Consider the generalized α-approximation $\hat{Q}_\alpha$ and its sample average approximation $\hat{Q}_\alpha^S$. Then,*

$$\sup_{x \in \mathcal{X}} \left| \hat{Q}_\alpha(x) - \hat{Q}_\alpha^S(x) \right| \to 0$$

w.p. 1 as $S \to \infty$, where $\mathcal{X} = \{x \in \mathbb{R}_+^{n_1} : Ax \leq b\}$.

Proof. See Appendix 4.A. ■

Corollary 4.1. $\sup_{x \in \mathcal{X}} |Q(x) - \hat{Q}_\alpha^S(x)| \leq C \sum_{i=1}^{m} \mathbb{E}_{\omega_{-i}} \left[|\Delta| f_i(\cdot|\omega_{-i}) \right]$ *w.p. 1 as $S \to \infty$.*

Proof. Since

$$\sup_{x \in \mathcal{X}} |Q(x) - \hat{Q}_\alpha^S(x)| \leq \sup_{x \in \mathcal{X}} |Q(x) - \hat{Q}_\alpha(x)| + \sup_{x \in \mathcal{X}} |\hat{Q}_\alpha(x) - \hat{Q}_\alpha^S(x)|,$$

the result follows directly by combining Theorem 4.2 and Lemma 4.2. ■

Corollary 4.1 implies a similar error bound on the difference between the optimal values η^* and $\hat{\eta}_\alpha^S$ of the original MIR problem (4.1) and the SAA problem (4.11), respectively, see Corollary 4.2.

Corollary 4.2. *Consider the optimal values η^* and $\hat{\eta}^S_\alpha$ of the MIR problem* (4.1) *and the SAA problem* (4.11)*. Then,*

$$\hat{\eta}^S_\alpha - \eta^* \leq C \sum_{i=1}^{m} \mathbb{E}_{\omega_{-i}} \left[|\Delta| f_i(\cdot|\omega_{-i}) \right]$$

w.p. 1 *as* $S \to \infty$.

Proof. Since the optimal solution x^* of (4.1) is feasible but not necessarily optimal in (4.11), we have

$$\hat{\eta}^S_\alpha - \eta^* \leq cx^* + \hat{Q}^S_\alpha(x^*) - \eta^* = \hat{Q}^S_\alpha(x^*) - Q(x^*) \leq \sup_{x \in \mathcal{X}} |Q(x) - \hat{Q}^S_\alpha(x)|.$$

The result follows directly from Corollary 4.1. ■

4.4.1.2 Asymptotic tightness of loose optimality cuts

If we derive a loose optimality cut $\hat{Q}^S_\alpha(x) \geq \beta_{\tau+1}x + \delta_{\tau+1}$ for $\hat{Q}^S_\alpha$ at x^τ in LBDA(α), then the gap $\hat{Q}^S_\alpha(x^\tau) - (\beta_{\tau+1}x^\tau + \delta_{\tau+1})$ may be positive, since the cut is not necessarily tight at x^τ. However, we derive a bound on this gap by considering the index function

$$\hat{k}(\omega, x) \in \arg\max_{k=1,\ldots,K} \lambda^k(\omega - Tx), \quad \omega \in \Omega,\ x \in X.$$

Note that $\hat{k}(\omega^s, x^\tau) = \hat{k}^\tau_s$, i.e., the index function $\hat{k}(\omega, x)$ is used to derive our loose optimality cuts. Based on this index function, we define the lower bounding function L^S_α of $\hat{Q}^S_\alpha$, given by

$$L^S_\alpha(x) := \frac{1}{S} \sum_{s=1}^{S} \left[\lambda^{\hat{k}(\omega^s,x)}(\omega^s - Tx) + \psi^{\hat{k}(\omega^s,x)}(\omega^s - \alpha) \right], \quad x \in X, \tag{4.15}$$

and we note that $L^S_\alpha(x^\tau)$ is the value of the loose optimality cut at x^τ, i.e., $L^S_\alpha(x^\tau) = \beta_{\tau+1}x^\tau + \delta_{\tau+1}$. The function L^S_α is a lower bound of $\hat{Q}^S_\alpha$ since its corresponding value function, defined as

$$\nu_\alpha(\omega, x) := \lambda^{\hat{k}(\omega,x)}(\omega - Tx) + \psi^{\hat{k}(\omega,x)}(\omega - \alpha), \quad x \in X, \tag{4.16}$$

is a lower bound of $\hat{v}_\alpha(\omega, x)$ for all $\omega \in \Omega$ and $x \in X$.

Proposition 4.1 contains a uniform total variation bound on the difference be-

tween L_α^S and $\hat{Q}_\alpha^S$. In order to derive this result, we first analyze the difference between ν_α and $\hat{v}_\alpha$. It turns out that ν_α equals $\hat{v}_\alpha$ on large parts of the domain, see Lemma 4.3.

Lemma 4.3. *Consider the value functions ν_α and $\hat{v}_\alpha$, with ν_α defined in* (4.16) *and $\hat{v}_\alpha$ defined as*

$$\hat{v}_\alpha(\omega, x) = \max_{k=1,\ldots,K} \left\{ \lambda^k(\omega - Tx) + \psi^k(\omega - \alpha) \right\}. \tag{4.17}$$

Let Λ^k, $k = 1, \ldots, K$, denote the closed convex cones from Lemma 4.1. Then, there exist vectors $\sigma_k \in \Lambda^k$ and a constant $R > 0$ such that

(i) $0 \leq \hat{v}_\alpha(\omega, x) - \nu_\alpha(\omega, x) \leq R$ *for all* $\omega \in \mathbb{R}^m$ *and* $x \in \mathbb{R}^{n_1}$, *and*

(ii) $\omega - Tx \in \sigma_k + \Lambda^k \implies \nu_\alpha(\omega, x) = \hat{v}_\alpha(\omega, x) = \lambda^k(\omega - Tx) + \psi^k(\omega - \alpha)$.

Proof. See Appendix 4.A. ■

Proposition 4.1. *Consider the SAA of the generalized α-approximation $\hat{Q}_\alpha^S$ and its lower bound L_α^S, defined in* (4.15). *There exists a constant $C > 0$ such that for every continuous random vector ω with joint pdf $f \in \mathcal{H}^m$,*

$$\sup_{x \in \mathcal{X}} \left| \hat{Q}_\alpha^S(x) - L_\alpha^S(x) \right| \leq C \sum_{i=1}^{m} \mathbb{E}_{\omega_{-i}} \left[|\Delta| f_i(\cdot | \omega_{-i}) \right]$$

w.p. 1 as $S \to \infty$.

Proof. Define $\Delta(\omega, x) := \hat{v}_\alpha(\omega, x) - \nu_\alpha(\omega, x)$, so that

$$\hat{Q}_\alpha^S(x) - L_\alpha^S(x) = \frac{1}{S} \sum_{s=1}^{S} \Delta(\omega^s, x).$$

We derive an upper bound on $\Delta(\omega, x)$, independent of x. We then apply the strong law of large numbers (SLLN) to obtain the desired result.

If we define $\mathcal{M} = \cup_{k=1}^K (\sigma_k + \Lambda_k)$, where σ_k and Λ^k, $k = 1, \ldots, K$, are the vectors and closed convex cones from Lemma 4.3, then

$$\Delta(\omega, x) \leq \xi(\omega, x) := \begin{cases} R, & \text{if } \omega - Tx \notin \mathcal{M}, \\ 0, & \text{if } \omega - Tx \in \mathcal{M}, \end{cases}$$

where R is the upper bound on $\hat{v}_\alpha(\omega, x) - \nu_\alpha(\omega, x)$ from Lemma 4.3. Moreover, we can derive a bound on $\mathbb{P}[\omega - Tx \in \mathcal{M}]$. Unfortunately, the random variable $\xi(\omega, x)$

depends on x. Therefore, we cannot apply the SLLN to

$$\sup_{x\in\mathcal{X}} \frac{1}{S}\sum_{s=1}^{S} \xi(\omega^s, x)$$

if $\mathcal{X}$ is infinite. To resolve this, we use that $\mathcal{X}$ is bounded. Let D denote the diameter of $T\mathcal{X}$, i.e., $||Tx - Tx'|| \leq D$ for all $x, x' \in \mathcal{X}$. Define $\mathcal{M}' \subset \mathcal{M}$ as $\mathcal{M}' := \bigcup_{k=1}^{K}(\sigma_k + \Lambda^k)(D)$. Fix an arbitrary $\bar{x} \in X$. Note that for all $x \in \mathcal{X}$,

$$\begin{aligned}\omega - T\bar{x} \in \mathcal{M}' &\implies \exists k : \omega - T\bar{x} \in (\sigma_k + \Lambda^k)(D)\\ &\implies \exists k : \omega - Tx \in (\sigma_k + \Lambda^k) \implies \omega - Tx \in \mathcal{M}.\end{aligned}$$

We obtain

$$\Delta(\omega, x) \leq \bar{\xi}(\omega) := \begin{cases} R & \text{if } \omega - T\bar{x} \notin \mathcal{M}' \\ 0 & \text{if } \omega - T\bar{x} \in \mathcal{M}' \end{cases}$$

Note that $\bar{\xi}(\omega)$ only depends on a fixed $x^* \in \mathcal{X}$ and is independent of x. By the SLLN,

$$\frac{1}{S}\sum_{s=1}^{S} \bar{\xi}(\omega^s) \to R\mathbb{P}[\omega - T\bar{x} \notin \mathcal{M}'],$$

w.p. 1 as $S \to \infty$.

By [60, Lemma 3.9], $\mathbb{R}^m \setminus \mathcal{M}'$ can be covered by finitely many hyperslices, that is,

$$\mathbb{R}^m \setminus \mathcal{M}' \subset \bigcup_{j=1}^{J} H_j,$$

where the hyperslices H_j are defined as

$$H_j := \{x \in \mathbb{R}^m : 0 \leq a_j^T x \leq \delta_j\},$$

for some a_j^T and δ_j. By [60, Theorem 4.6], there exists a constant $\beta > 0$ such that

$$\mathbb{P}[\omega - Tx^* \notin \mathcal{M}'] \leq \beta \sum_{i=1}^{m} \mathbb{E}_{\omega_{-i}}\left[|\Delta| f_i(\cdot|\omega_{-i})\right].$$

The result now follows from

$$\begin{aligned}\sup_{x\in\mathcal{X}}\left|\hat{Q}_\alpha^S(x)-L_\alpha^S(x)\right| &= \sup_{x\in\mathcal{X}}\frac{1}{S}\sum_{s=1}^{S}\Delta(\omega^s,x)\\ &\leq \frac{1}{S}\sum_{s=1}^{S}\bar{\xi}(\omega^s)\to R\mathbb{P}[\omega-Tx^*\notin\mathcal{M}']\\ &\leq C\sum_{i=1}^{m}\mathbb{E}_{\omega_{-i}}\left[|\Delta|f_i(\cdot|\omega_{-i})\right],\end{aligned}$$

w.p. 1 as $S\to\infty$, where $C=R\beta$. ■

Proposition 4.1 shows that our loose optimality cuts are asymptotically tight, since for every iteration τ in LBDA(α) the loose optimality cut for $\hat{Q}_\alpha^S$ at x^τ is tight for L_α^S at x^τ. Hence, if the underlying total variations are small enough and if S is large enough, then the loose optimality cut is nearly tight for $\hat{Q}_\alpha^S$ at x^τ.

4.4.2 Error bound on the optimality gap of LBDA(α)

We are now ready to prove the performance guarantee of LBDA(α) in Theorem 4.3. Intuitively, we are able to derive this bound since (i) our loose optimality cuts are asymptotically tight, and thus $\hat{x}_\alpha$ is near-optimal for the SAA problem with $\hat{Q}_\alpha^S$, (ii) the SAA $\hat{Q}_\alpha^S$ converges uniformly to $\hat{Q}_\alpha$, and (iii) Theorem 4.2 contains a uniform error on the difference between $\hat{Q}_\alpha$ and Q.

Proof. of Theorem 4.3 Let $\hat{x}_\alpha$ denote the solution returned by LBDA(α). Since $\hat{x}_\alpha := x^\tau$ is the current solution in the final iteration τ of the algorithm, it follows that $\hat{x}_\alpha$ is a minimizer of

$$\min_x\{cx+\hat{Q}_{\text{out}}^\tau(x):\quad Ax=b, x\in X\}.$$

Moreover, since $\hat{Q}_{\text{out}}^\tau\leq\hat{Q}_\alpha^S$, it follows that $c\hat{x}_\alpha+\hat{Q}_{\text{out}}^\tau(\hat{x}_\alpha)\leq\hat{\eta}_\alpha^S$, and thus, by rearranging terms and adding $\hat{Q}_\alpha^S$ on both sides,

$$c\hat{x}_\alpha+\hat{Q}_\alpha^S(\hat{x}_\alpha)-\hat{\eta}_\alpha^S\leq\hat{Q}_\alpha^S(\hat{x}_\alpha)-\hat{Q}_{\text{out}}^\tau(\hat{x}_\alpha). \tag{4.18}$$

The right-hand side of (4.18) represents an upper bound on the optimality gap of $\hat{x}_\alpha$ in the SAA problem (4.11). The termination criterion of LBDA(α) guarantees that

this upper bound is not too large. Indeed, at termination it holds that

$$\hat{Q}^{\tau}_{\text{out}}(\hat{x}_\alpha) \geq \hat{Q}^{\tau+1}_{\text{out}}(\hat{x}_\alpha) - \varepsilon \geq L^S_\alpha(\hat{x}_\alpha) - \varepsilon,$$

and thus the upper bound in (4.18) reduces to

$$\begin{aligned} c\hat{x}_\alpha + \hat{Q}^S_\alpha(\hat{x}_\alpha) - \hat{\eta}^S_\alpha &\leq \hat{Q}^S_\alpha(\hat{x}_\alpha) - L^S_\alpha(\hat{x}_\alpha) + \varepsilon \\ &\leq \sup_{x\in\mathcal{X}} |\hat{Q}^S_\alpha(x) - L^S_\alpha(x)| + \varepsilon. \end{aligned} \tag{4.19}$$

In the end, however, we are not interested in the optimality gap of $\hat{x}_\alpha$ in the SAA problem (4.11) but in the optimality gap

$$c\hat{x}_\alpha + Q(\hat{x}_\alpha) - \eta^*$$

of $\hat{x}_\alpha$ in the original MIR problem (4.1). Adding and subtracting both $\hat{Q}^S_\alpha(\hat{x}_\alpha)$ and $\hat{\eta}^S_\alpha$, and using (4.19), yields

$$\begin{aligned} c\hat{x}_\alpha + Q(\hat{x}_\alpha) - \eta^* &= (c\hat{x}_\alpha + Q(\hat{x}_\alpha) - \hat{\eta}^S_\alpha) + (Q(\hat{x}_\alpha) - \hat{Q}^S_\alpha(\hat{x}_\alpha)) + (\hat{\eta}^S_\alpha - \eta^*) \\ &\leq \sup_{x\in\mathcal{X}} |\hat{Q}^S_\alpha(x) - L^S_\alpha(x)| + \varepsilon + \sup_{x\in\mathcal{X}} |Q(x) - \hat{Q}^S_\alpha(x)| + \hat{\eta}^S_\alpha - \eta^*. \end{aligned}$$

Applying Corollaries 4.1 and 4.2, and Proposition 4.1, we conclude that there exists a constant $C > 0$ such that

$$c\hat{x}_\alpha + Q(\hat{x}_\alpha) - \eta^* \leq C \sum_{i=1}^{m} \mathbb{E}_{\omega_{-i}} [|\Delta| f_i(\cdot|\omega_{-i})] + \varepsilon,$$

w.p. 1 as $S \to \infty$. ■

The performance guarantee for LBDA(α) in Theorem 4.3 is a worst-case bound. For many problem instances, the actual performance may be much better. In Section 4.5, we assess the performance of LBDA(α) empirically on a wide range of test instances.

4.5 Numerical experiments

We test the performance of LBDA(α) on problem instances from the literature and on randomly generated instances, see Sections 4.5.2 and 4.5.3, respectively. In particular, we consider (variations of) an investment planning problem in [67], and

two classes of problem instances available from SIPLIB [5], namely the SIZES problem [39] and the stochastic server location problem (SSLP) [50]. First, however, we describe the set-up of our numerical experiments in Section 4.5.1.

4.5.1 Set-up of numerical experiments

In our numerical experiments, we compare LBDA(α) to several benchmark methods, in terms of costs, relative optimality gaps, and computations times. Since the performance of LBDA(α) depends on α, we investigate four different approaches to select α.

First, we take α equal to the zero vector. Second, we take $\alpha = \alpha^* := Tx^*$, where x^* is the optimal solution of the original problem. Obviously, for large problem instances x^* is unknown, however, we expect that α^* is a good choice of α since the generalized α-approximations are obtained by replacing Tx by α in the Gomory relaxations, and thus $\hat{Q}_{\alpha^*}$ is a good approximation of Q near the true optimal solution x^*. We test this for the smaller problem instances for which x^* is known.

Since α^*, however, typically cannot be computed for larger problem instances, we also propose an iterative scheme, in which we first obtain $\hat{x}_{\alpha_0}$ by running LBDA(α_0), where α_0 is the zero vector. Next, we run LBDA(α_1), where $\alpha_1 := T\hat{x}_{\alpha_0}$. We extend this scheme to 100 iterations by recursively defining $\alpha_{k+1} := T\hat{x}_{\alpha_k}$, $k = 1, \ldots, 100$. We then select the best value of α in terms of expected costs, denoted by $\alpha^{\#}$. Finally, we apply LBDA(α) multiple times using 100 different values of α, drawn from an multivariate uniform distribution on $[0, 100]^m$, and we denote the best value of α in terms of the expected costs by α^+.

In order to compare these approaches, note that the expected costs $cx + Q(x)$ of a candidate solution x can be computed exactly if the random vector ω has a finite number of realizations, as is the case for the SIZES, SSLP, and investment planning problems that we consider. Therefore, if the optimal value η^* is known, then the relative optimality gap $\rho(x)$, defined as

$$\rho(x) = \frac{cx + Q(x) - \eta^*}{|\eta^*|} * 100\%,$$

can be computed exactly, and otherwise bounds on $\rho(x)$ can be computed.

In contrast, for our randomly generated instances, we assume that ω is continuously distributed. For these instances, we use the multiple replications procedure (MRP) [46] with Latin hypercube sampling [14] to obtain 95% confidence upper bounds on $\rho(x)$. Moreover, we compare the performance of LBDA(α) to a range of

benchmark solutions, using out-of-sample estimation of $cx + Q(x)$, with a sample size of 10^5, which guarantees that the standard errors of our results are sufficiently small. The benchmark and LBDA(α) solutions are computed using a sample of size $S = 1000$.

The first benchmark solution $\bar{x}_S$ is obtained by solving the deterministic equivalent formulation (DEF) of the corresponding SAA of the original problem (4.1). The DEF is a large-scale MIP, which, typically, cannot be solved in reasonable time by standard MIP solvers. Hence we also solve the DEF using a smaller sample size $S' = 100$, resulting in the second benchmark solution $\bar{x}_{S'}$.

We obtain three additional benchmark solutions by solving the generalized α-approximations exactly for $\alpha = 0$, $\alpha = \alpha^*$ and $\alpha = \alpha^+$, that is, we find the optimal solution x^*_α of the approximating problem (4.4) with $\hat{Q} = \hat{Q}^S_\alpha$. We do so by solving the approximation second-stage problems

$$\max_{k=1,\dots,K} \{\lambda^k(\omega - Tx) + \psi^k(\omega - \alpha)\}$$

by enumeration over $k = 1, \dots, K$. For this reason, x^*_α can only be computed in reasonable time for small problem instances.

Finally, we consider two trivial benchmark solutions, which we expect to outperform significantly. First, we relax the integer restrictions on the second-stage variables in the SAA of the original problem (4.1), resulting in the benchmark solution x_{LP}. Second, we solve the Jensen approximation, which replaces the distribution of ω by a degenerate distribution at $\mu = \mathbb{E}_\omega[\omega]$, and denote the optimal solution by x_μ.

We run our experiments on a single Intel Xeon E5 2680v3 core @2.5GHz with Gurobi 7.0.2. To ensure a fair comparison between solutions, we use common random numbers where possible and we limit the computation time of each algorithm to two hours.

4.5.2 Test instances from the literature

4.5.2.1 The SIZES problem

We first consider all instances of the SIZES test problem suite [39] from SIPLIB. These instances have mixed-binary variables in both stages, and differ in the number of scenarios, namely 3, 5, and 10. The DEF of the largest instance has 341 constraints and 825 variables, of which 110 are binary. We refer to [5] for further details. In Table 4.1, we report the outcomes of LBDA(α) and solving the DEF.

Table 4.1. The SIZES problem.

	Computation time (s) (optimality gap)				
	DEF	LBDA(α)			
Instance	x^*	$\alpha = 0$	$\alpha = \alpha^*$	$\alpha = \alpha^\#$	$\alpha = \alpha^+$
SIZES3	0.2 (0.0%)	0.2 (0.35%)	0.3 (0.20%)	24.8 (0.32%)	23.8 (0.20%)
SIZES5	1.5 (0.0%)	0.3 (0.44%)	0.6 (0.26%)	34.0 (0.35%)	34.0 (0.14%)
SIZES10	312.1 (0.0%)	0.7 (0.47%)	0.9 (0.08%)	68.6 (0.10%)	53.8 (0.11%)

We observe from Table 4.1 that LBDA(α) performs very well for all choices of α. Indeed, on every instance, the optimality gaps of all LBDA(α) solutions are below 0.5%, and below 0.2% for $\alpha = \alpha^+$. Moreover, LBDA(α) runs very fast for all instances: for a single value of α, the computation time of LBDA(α) is always below one second.

Another observation is that LBDA($\alpha^\#$) and LBDA(α^+) consistently outperform LBDA(0), at the expense of additional computation time. Nevertheless, the computation times of LBDA(α) for $\alpha^\#$ and α^+ are still moderate, and they scale much better to larger instances than solving the DEF. In particular, the time taken to solve the DEF grows exponentially as the sample size S increases, whereas the computation times of LBDA(α) are approximately linear in S. Finally, LBDA(α) performs very well if $\alpha = \alpha^*$. However, since α^* is not known in practice, it is useful to observe that LBDA(α^+) achieves similar performance.

4.5.2.2 The stochastic server location problem

The SSLP instances are more challenging in terms of input size than the SIZES instances. Indeed, the DEF of the largest instance has over 1,000,000 binary decision variables and 120,000 constraints. Their first- and second-stage problems are pure binary and mixed-binary, respectively, and ω follows a discrete distribution. A full problem description can be found in [50], in which the instances are solved using the D^2 algorithm of [69]. See also [1] and [36] for more recent computational experiments on these test instances using other exact approaches. In Table 2 we report the best known running time for each SSLP instance over all these exact approaches, along with the outcomes of LBDA(α).

Strikingly, LBDA(α) was able to solve all instances to optimality for $\alpha = \alpha^+$ and $\alpha = \alpha^*$. Moreover, LBDA(0) and LBDA($\alpha^\#$) solved all instances except for SSLP_15_45_5, on which both achieved an optimality gap of 0.45%. In terms of computation time, LBDA(0) is clearly preferred to LBDA($\alpha^\#$) and LBDA(α^+), while

Table 4.2. The stochastic server location problem.

	Computation time (s) (optimality gap)				
	Exact approaches	LBDA(α)			
Instance	x^*	$\alpha = 0$	$\alpha = \alpha^*$	$\alpha = \alpha^\#$	$\alpha = \alpha^+$
SSLP_5_25_50	0.5[1] (0.0%)	0.2 (0.0%)	0.2 (0.0%)	2.8 (0.0%)	2.8 (0.0%)
SSLP_5_25_100	1.0[1] (0.0%)	0.2 (0.0%)	0.2 (0.0%)	5.3 (0.0%)	5.3 (0.0%)
SSLP_15_45_5	4.3[2] (0.0%)	1.2 (0.45%)	2.1 (0.0%)	79.3 (0.45%)	13.5 (0.0%)
SSLP_15_45_10	35.3[2] (0.0%)	2.9 (0.0%)	2.7 (0.0%)	101.5 (0.0%)	10.8 (0.0%)
SSLP_15_45_15	92.7[2] (0.0%)	4.6 (0.0%)	5.6 (0.0%)	270.7 (0.0%)	105.4 (0.0%)
SSLP_10_50_50	65.7[2] (0.0%)	0.6 (0.0%)	1.1 (0.0%)	102.0 (0.0%)	48.3 (0.0%)
SSLP_10_50_100	153.3[2] (0.0%)	1.9 (0.0%)	1.9 (0.0%)	137.0 (0.0%)	121.5 (0.0%)
SSLP_10_50_500	1033.0[3] (0.0%)	8.7 (0.0%)	8.6 (0.0%)	525.7 (0.0%)	482.1 (0.0%)
SSLP_10_50_1000	2691.0[2] (0.0%)	13.5 (0.0%)	58.7 (0.0%)	5529.5 (0.0%)	903.5 (0.0%)
SSLP_10_50_2000	4952.0[2] (0.0%)	55.7 (0.0%)	73.6 (0.0%)	6020.8 (0.0%)	1698.9 (0.0%)

[1]See [50]. [2]See [1]. [3]See [36].

achieving similar results. Finally, although directly comparing LBDA(α) to the other approaches is not completely fair since the algorithms were run on different machines, it is clear that LBDA(α) is generally faster than exact approaches. Indeed, LBDA(0) solved all instances in under one minute, and eight out of ten instances were solved in less than ten seconds, whereas the fastest exact approach required at least one minute for six out of ten instances, and over one hour for the largest instance.

4.5.2.3 An investment planning problem

We consider the following problem by Schultz et al. [67],

$$\min_x \{-3/2x_1 - 4x_2 + \mathbb{E}_\omega[v(\omega, x)] : \quad x \in [0,5]^2\},$$

where

$$v(\omega, x) = \min_{y \in Y} \{-16y_1 - 19y_2 - 23y_3 - 28y_4 : 2y_1 + 3y_2 + 4y_3 + 5y_4 \leq \omega_1 - x_1$$
$$6y_1 + \ y_2 + 3y_3 + \ y_4 \leq \omega_2 - x_2\},$$

and where the second-stage decision variables are binary, i.e., $Y = \{0,1\}^4$, and the random vector $\omega = (\omega_1, \omega_2)$ follows a discrete distribution which assigns equal probabilities to $S = 441$ equidistant lattice points of $[5,15]^2$. Schultz et al. consider

Table 4.3. An investment planning problem.

Instance			Computation time (s) (gap to $\hat{x}^*$)				
			DEF	LBDA(α)			
Y	T	S	$\hat{x}^*$	$\alpha = 0$	$\alpha = T\hat{x}^*$	$\alpha = \alpha^\#$	$\alpha = \alpha^+$
$\mathbb{Z}^4_+$	H	4	0	0.0 (0.0%)	0.0 (0.0%)	0.4 (0.0%)	0.4 (0.0%)
		9	0	0.0 (12.9%)	0.0 (0.0%)	0.7 (2.2%)	0.7 (0.1%)
		36	0.5	0.0 (5.0%)	0.0 (5.8%)	3.0 (0.5%)	2.9 (0.0%)
		121	8.2	0.1 (4.0%)	0.1 (4.0%)	7.6 (4.0%)	7.7 (4.0%)
		441	7200[1]	0.3 (3.1%)	0.2 (3.1%)	27.6 (3.1%)	27.8 (3.1%)
		1681	7200[1]	0.9 (0.0%)	0.9 (0.0%)	104.4 (0.0%)	105.6 (0.0%)
		10201	7397[1]	5.5 (-0.2%)	5.4 (-0.2%)	631.8 (-0.2%)	639.1 (-0.2%)
$\mathbb{Z}^4_+$	I_2	4	0	0.0 (0.0%)	0.0 (0.0%)	0.3 (0.0%)	0.3 (0.0%)
		9	0	0.0 (0.9%)	0.0 (1.0%)	0.8 (0.9%)	0.8 (0.0%)
		36	0	0.0 (4.3%)	0.0 (0.9%)	2.9 (0.9%)	2.9 (0.0%)
		121	0.5	0.1 (4.0%)	0.1 (2.8%)	9.8 (0.0%)	9.9 (0.0%)
		441	20.5	0.3 (0.9%)	0.3 (1.5%)	36.5 (0.2%)	36.1 (0.0%)
		1681	7200[1]	1.3 (0.7%)	1.2 (0.1%)	136.1 (0.0%)	138.2 (0.0%)
		10201	7269[1]	8.2 (0.3%)	7.9 (0.0%)	864.5 (-0.1%)	883.0 (-0.1%)
$\{0,1\}^4$	H	4	0	0.0 (8.8%)	0.0 (8.8%)	0.2 (8.8%)	0.2 (8.8%)
		9	0	0.0 (5.0%)	0.0 (5.0%)	0.4 (5.0%)	0.5 (5.0%)
		36	0.7	0.0 (1.6%)	0.0 (1.6%)	1.9 (1.6%)	2.0 (1.6%)
		121	32.5	0.1 (1.8%)	0.1 (1.8%)	6.7 (1.8%)	6.7 (1.8%)
		441	7200[1]	0.3 (2.1%)	0.3 (2.1%)	24.2 (2.1%)	24.3 (2.1%)
		1681	7200[1]	0.9 (0.3%)	0.9 (0.3%)	92.0 (0.3%)	91.9 (0.3%)
		10201	7204[1]	5.0 (0.1%)	5.1 (0.1%)	554.5 (0.1%)	553.4 (0.1%)
$\{0,1\}^4$	I_2	4	0	0.0 (1.9%)	0.0 (14.8%)	0.3 (1.9%)	0.3 (0.1%)
		9	0	0.0 (1.4%)	0.0 (0.0%)	0.6 (1.4%)	0.7 (0.1%)
		36	0	0.0 (5.0%)	0.0 (2.6%)	3.4 (2.0%)	3.3 (1.0%)
		121	0.9	0.1 (2.1%)	0.1 (4.7%)	10.7 (1.5%)	11.1 (0.4%)
		441	118.3	0.4 (2.5%)	0.4 (0.3%)	41.6 (1.3%)	42.5 (0.4%)
		1681	7204[1]	1.8 (0.9%)	1.8 (1.1%)	162.1 (0.9%)	166.5 (0.3%)
		10201	7293[1]	10.5 (1.1%)	10.9 (0.9%)	1032.7 (0.8%)	1055.1 (0.0%)

[1] DEF could not be solved in time: $\hat{x}^*$ denotes best solution found by Gurobi.

a second variant of this problem by choosing the technology matrix T as

$$T = H := \begin{pmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix},$$

whereas in the original formulation, T is the identity matrix I_2. For both variants, we consider $S \in \{4, 9, 36, 121, 441, 1681, 10201\}$ and $Y = \mathbb{Z}^4_+$, in addition to $Y = \{0,1\}^4$, as is done in [53]. For each of the resulting 28 instances, Table 4.3 shows the results of LBDA(α) and solving the DEF. Note that if Gurobi could not solve an instance within two hours, then we report the gaps to $\hat{x}^*$, the best solution that Gurobi was able to find.

Overall, LBDA(α) performs well on the instances in Table 4.3. In particular, for $\alpha = \alpha^{\#}$ and $\alpha = \alpha^{+}$, LBDA(α) achieves gaps that are below 2% on 22 and 23 out of 28 instances, respectively. In general, LBDA(α) achieves better results if S is larger. For example, if $Y = \{0,1\}^4$ and $T = H$, then the gaps are strictly decreasing in S. This is in line with the performance guarantee of LBDA(α) in Theorem 4.3: if S is larger, then the distributions of ω_1 and ω_2 more closely resemble a continuous uniform distribution on $[5,15]$, which has small total variation, i.e., the error bound in Theorem 4.3 is small.

On some instances, LBDA(α) does not consistently perform well for all values of α. For example, if $Y = \mathbb{Z}^4_+$, $T = H$, and $S = 9$, then LBDA(α^+) and LBDA(0) achieve gaps of 0.1% and 12.9%, respectively. Furthermore, the instances with $Y = \{0,1\}^4$, $T = H$, and $S \in \{4,9\}$ turn out to be difficult instances for LBDA(α): for all choices of α, the resulting gaps are 8.8% and 5%, respectively.

In fact, for every choice of α, LBDA(α) achieves identical gaps if $Y = \{0,1\}^4$ and $T = H$, but the gaps are much smaller if $S \geq 36$, e.g., if $S = 10201$, then the gaps are 0.1%. In contrast, there are large differences between the different choices of α if $Y = \{0,1\}^4$ and $T = I$. For these instances, $\alpha = \alpha^{\#}$ and $\alpha = \alpha^{+}$ outperform the other choices of α. However, similar as for the SIZES instances in Section 4.5.2.1, there is a trade-off between performance and computation times, since LBDA(α^+) and LBDA($\alpha^{\#}$) require 100 LBDA(α) runs, which is computationally more demanding than LBDA(0). Nevertheless, on every instance, the computation times of LBDA(α) for $\alpha \in \{\alpha^{\#}, \alpha^{+}\}$ are below 20 minutes, and below 3 minutes if $S \leq 1681$.

4.5.3 Randomly generated test instances

We generate random MIR problems of the form (4.1), with $X = \mathbb{R}^{n_1}_+$ and

$$Q(x) = \mathbb{E}_\omega \left[\min_y \left\{ qy : \quad Wy \geq \omega - Tx,\ y \in \mathbb{Z}^p_+ \right\} \right], \quad x \in X.$$

In addition, we assume that the components of the random vector $\omega \in \mathbb{R}^m$ follow independent normal distributions with mean 10 and standard deviation $\sigma \in \{0.1, 0.5, 1, 2, 4, 10\}$. The parameters c, q, T, and W are fixed, and their elements are drawn from discrete uniform distributions with supports contained in $[1,5]$, $[5,10]$, $[1,6]$, and $[1,6]$, respectively.

The reason that we consider multiple values of the standard deviation σ is that Theorem 4.3 implies that LBDA(α) performs better as σ increases. This is because

the total variations of the one-dimensional conditional pdf are small if σ is large, and thus the error bound on the optimality gap achieved by LBDA(α) is also small.

To prevent noise in the outcomes of our experiments, we compute the average optimality gaps, costs, and computation times over 20 randomly generated test instances for each value of σ. We consider test instances of two different sizes, namely $n_1 = 10$, $p = 5$, $m = 5$ (small), and $n_1 = 100$, $p = 40$, $m = 20$ (large). Tables 4.4 and 4.5 display the results for the small and large versions, respectively.

From these results, we observe that LBDA(α) clearly outperforms the sampling solutions in terms of computation time and scalability to larger problem instances. In particular, we observe that the computation time of LBDA(α) is of the same order of magnitude as that of x_{LP}, while it performs significantly better in terms of optimality gaps and out-of-sample estimated expected costs. Undeniably, our results indicate that LBDA(α) can be implemented very efficiently and that it can handle large MIR problem instances.

In line with the performance guarantee in Theorem 4.3, LBDA(α) performs better for larger values of σ. For example, on the small instances, the optimality gaps achieved by LBDA(0) are strictly decreasing in σ, and for $\sigma = 10.0$, LBDA(α^+) outperforms the sampling solution $\bar{x}_S$. A similar observation is true for the optimal solution of the generalized α-approximations $\hat{Q}_\alpha$, as we would expect based on the error bound for $\hat{Q}_\alpha$ in Theorem 4.2. Observe, however, that even for large values of σ, the optimality gaps reported in Table 4.4 for LBDA(α) are relatively large, i.e., around 3-4%, and that the optimality gaps achieved by e.g. LBDA(α^+) are not strictly decreasing in σ. The reason is that the actual optimality gaps are likely much smaller than the upper bounds reported in Table 4.4. This is because they are computed using the MRP, which relies on solving multiple SAAs of the original problem. Since, however, Gurobi has difficulties solving the DEFs of these SAAs in reasonable time, especially for large values of σ, the bounds obtained using the MRP are typically not sharp.

Interestingly, x_{LP} also performs better as σ increases. An explanation is that our error bound for the generalized α-approximation implies that the MIR function Q becomes closer to a convex function as σ increases. Thus, since the LP-relaxation of Q is a convex lower bound of Q, its approximation error is expected to become smaller as σ increases. Note however, that unlike the generalized α-approximations, the approximation error of the LP-relaxation does not go to zero. Indeed, based on our results, LBDA(α) is clearly preferred to x_{LP}, since LBDA(α) consistently outperforms x_{LP}, at the expense of very little additional computation time.

Table 4.4. Randomly generated test instances ($n_1 = 10$, $p = 5$, $m = 5$).

	Computation time (s) (optimality gap)										
	Benchmarks				x^*_α			LBDA(α)			
σ	$\bar{x}_S$	$\bar{x}_{S'}$	x_μ	x_{LP}	$\alpha = 0$	$\alpha = T\bar{x}_S$	$\alpha = \alpha^+$	$\alpha = 0$	$\alpha = T\bar{x}_S$	$\alpha = \alpha^\#$	$\alpha = \alpha^+$
0.1	677.8 (0.3%)	0.1 (2.6%)	0.0 (102.5%)	2.3 (37.0%)	169.3 (28.3%)	83.2 (0.3%)	86.3 (11.8%)	4.0 (24.5%)	3.8 (0.4%)	389.4 (15.0%)	358.3 (3.2%)
0.5	7125.0[2] (1.1%)	0.3 (2.3%)	0.0 (84.8%)	1.8 (28.1%)	169.6 (17.0%)	106.3 (1.1%)	96.6 (7.9%)	4.4 (12.1%)	2.7 (1.6%)	206.8 (5.8%)	200.9 (1.7%)
1.0	7200[1] (1.9%)	1.0 (2.8%)	0.0 (71.9%)	1.7 (21.2%)	176.4 (10.2%)	121.5 (2.4%)	104.9 (7.0%)	2.9 (9.0%)	2.9 (4.5%)	351.5 (4.9%)	313.0 (2.3%)
2.0	7200[1] (3.0%)	368.6[4] (4.1%)	0.0 (60.0%)	1.9 (13.8%)	173.1 (5.6%)	150.7 (3.3%)	128.5 (4.9%)	3.3 (7.7%)	3.4 (5.0%)	277.5 (4.5%)	243.6 (3.6%)
4.0	7200[1] (4.1%)	436.9[4] (4.6%)	0.0 (65.2%)	2.0 (8.4%)	180.0 (4.8%)	125.0 (4.1%)	94.0 (4.6%)	3.4 (6.7%)	3.5 (5.2%)	303.6 (5.0%)	275.5 (4.2%)
10.0	7200[1] (3.3%)	1119.2[3] (4.0%)	0.0 (87.6%)	2.3 (4.2%)	173.8 (3.3%)	143.8 (3.3%)	126.8 (3.3%)	3.8 (3.7%)	4.1 (3.5%)	344.9 (3.3%)	303.3 (3.2%)

[1] 0/20 instances solved in time.
[2] 1/20 instances solved in time.
[3] 17/20 instances solved in time.
[4] 19/20 instances solved in time.

Table 4.5. Randomly generated test instances ($n_1 = 100$, $p = 40$, $m = 20$).

	Computation time (s) (gap to $\bar{x}_S$)							
	Benchmarks				LBDA(α)			
σ	$\bar{x}_S$	$\bar{x}_{S'}$	x_μ	x_{LP}	$\alpha = 0$	$\alpha = T\bar{x}_S$	$\alpha = \alpha^\#$	$\alpha = \alpha^+$
0.1	1519.9	1.7 (13.0%)	0.0 (142.8%)	25.4 (42.7%)	34.3 (31.7%)	31.2 (0.4%)	995.4 (6.7%)	1298.2 (8.2%)
0.5	7200[1]	3.9 (9.9%)	0.0 (116.6%)	26.8 (33.3%)	33.4 (22.3%)	32.4 (2.0%)	1196.5 (4.0%)	1131.1 (5.3%)
1.0	7200[1]	4.9 (8.0%)	0.0 (93.0%)	25.6 (25.0%)	31.2 (11.4%)	31.9 (3.0%)	1008.7 (4.1%)	897.4 (4.1%)
2.0	7200[1]	38.5 (5.7%)	0.0 (85.6%)	23.6 (14.9%)	30.3 (6.2%)	29.2 (4.1%)	1062.9 (3.2%)	933.1 (2.8%)
4.0	7200[1]	165.5 (4.0%)	0.0 (105.4%)	20.7 (6.3%)	28.9 (3.1%)	28.9 (2.2%)	1244.3 (1.6%)	1111.8 (1.5%)
10.0	7200[1]	5692.5[2] (2.5%)	0.0 (132.6%)	18.4 (1.3%)	29.0 (1.0%)	28.9 (0.5%)	1502.4 (0.3%)	1383.1 (0.3%)

[1] 0/20 instances solved in time.
[2] 7/20 instances solved in time.

Furthermore, the results in Tables 4.4 and 4.5 indicate that $\alpha = T\bar{x}_S$ is a good choice for LBDA(α). Indeed, if $\alpha = T\bar{x}_S$, then LBDA(α) and x^*_α perform similar to $\bar{x}_S$. For example, on the small instances, they achieve optimality gaps that are on average within 0.2% and 1.1% of $\bar{x}_S$, respectively. However, since $\bar{x}_S$ is difficult to compute in practice, the use of LBDA($T\bar{x}_S$) is limited, whereas LBDA(α^+) and LBDA($\alpha^\#$) can be applied directly. Similar as for the instances in Section 4.5.2.1 and 4.5.2.3, they achieve significantly better results than LBDA(0), at the expense of higher computation times. In particular, on the large instances, LBDA($\alpha^\#$) performs 2 to 5 times as well as LBDA(0), and on the small instances, LBDA(α^+) achieves optimality gaps that are within 0.6% of $\bar{x}_S$ for $\sigma \geq 0.5$. While the computation times of LBDA(α) increase for $\alpha = \alpha^+$ and $\alpha = \alpha^\#$ compared to $\alpha = 0$, they remain manageable: even the large instances are solved within 26 minutes.

Finally, observe from Table 4.4 that LBDA(α) generally achieves better or similar performance as x^*_α. In other words, the fact that LBDA(α) uses loose optimality cuts to solve the generalized α-approximations has no negative effect on the solution quality.

4.6 Conclusion

We consider two-stage mixed-integer recourse models with random right-hand side. Due to non-convexity of the recourse function, such models are extremely difficult to solve. We develop a tractable approximating model by using convex approximations of the recourse function. In particular, we propose a new class of convex approximations, the so-called generalized α-approximations, and we derive a corresponding error bound on the difference between these approximations and the true recourse function. In addition, we show that this error bound is small if the variability of the random parameters in the model is large. More precisely, the error bound for the generalized α-approximations goes to zero as the total variations of the one-dimensional conditional probability density functions of the random right-hand side vector in the model go to zero.

The advantage of the generalized α-approximations over existing convex approximations is that it can be solved efficiently. In fact, we describe a loose Benders' decomposition algorithm, LBDA(α), which efficiently solves the corresponding approximating model. The quality of the candidate solution $\hat{x}_\alpha$ generated by LBDA(α) in the original model is guaranteed by Theorem 4.3, which states an upper bound on the optimality gap of $\hat{x}_\alpha$. This performance guarantee is similar to

the error bound we prove for the generalized α-approximations. Indeed, we show that the optimality gap of $\hat{x}_\alpha$ is small if the variability of the random parameters in the model is large.

In addition to this theoretical guarantee on the solution quality, we empirically assess LBDA(α) on a range of test instances. In particular, we consider the SIZES and SSLP instances from SIPLIB, an investment planning problem by [67], and randomly generated instances. We find that LBDA(α) performs well in terms of computation times, scalability to larger problem instances, and solution quality. In particular, LBDA(α) is able to solve larger instances than traditional sampling techniques and its computation times scale more favourably in the input size of the instances. In terms of solution quality, LBDA(α) solves the SIZES and SSLP instances to near optimality and generally performs very well on the investment planning instances. Moreover, on the randomly generated instances, LBDA(α) performs similar to traditional sampling techniques and achieves small optimality gaps if the variability of the random parameters in the model is medium to large.

One avenue for future research is to derive sharper theoretical error bounds for the generalized α-approximations. While Theorem 4.3 provides conditions under which our solution method performs well, the quantitative error bound cannot be computed, as it depends on an unknown and potentially large constant C. A sharp tractable error bound would be an improvement over our current results. Another avenue is the extension of our solution method to more general mixed-integer recourse models, for example by allowing for randomness in the second-stage cost coefficients q, technology matrix T, or recourse matrix W.

Appendix 4.A Postponed proofs

Proof of Theorem 4.2. Our proof is similar to, e.g., [60, Theorem 5.1] and Theorem 3.2. Here, we point out the differences. In particular, we show that there exist vectors σ_k, $k = 1, \dots, K$ and a constant $R > 0$, such that

(i) if $\omega - Tx \in \sigma_k + \Lambda^k$, then $v(\omega, x) - \hat{v}_\alpha(\omega, x)$ is zero-mean B^k-periodic, and

(ii) $|v(\omega, x) - \hat{v}_\alpha(\omega, x)| \leq R$,

where Λ^k, $k = 1, \dots, K$, are the closed convex cones from Lemma 4.1. Property (i) follows from Lemma 4.1 (iii) and Lemma 4.3 (ii), and the fact that $\psi^k(\omega - Tx) - \psi^k(\omega - \alpha)$ is zero-mean B^k-periodic.

In order to prove (ii), we use a similar argument as in [60, Lemma 3.6] and Proposition 3.2. It suffices to show that there exists a constant R' such that $|\hat{v}_\alpha(\omega, x) - v_{\mathrm{LP}}(\omega, x)| \leq R'$. We can take $R' = \max_k w_k$, where w_k are the upper bounds on ψ^k from Lemma 4.1, $k = 1, \ldots, K$. ■

Proof of Lemma 4.2. Our line of proof is based on [3]. For any $\nu > 0$, consider a finite set $\mathcal{X}_\nu$ such that for all $x \in \mathcal{X}$, there exists an $x' \in \mathcal{X}_\nu$ such that $||x - x'|| \leq \nu$. Such a set $\mathcal{X}_\nu$ exist due to Assumption (A1). Let $x \in \mathcal{X}$ be given and let $x' \in \mathcal{X}_\nu$ be such that $||x - x'|| \leq \nu$. Note that

$$|\hat{Q}_\alpha(x) - \hat{Q}^S_\alpha(x)| \leq |\hat{Q}_\alpha(x) - \hat{Q}_\alpha(x')| + |\hat{Q}_\alpha(x') - \hat{Q}^S_\alpha(x')| + |\hat{Q}^S_\alpha(x') - \hat{Q}^S_\alpha(x)|. \quad (4.20)$$

The first and third term on the right-hand side of (4.20) can be bounded by noting that both $\hat{Q}_\alpha$ and $\hat{Q}^S_\alpha$ are Lipschitz continuous. Denote Lipschitz constants of $\hat{Q}_\alpha$ and $\hat{Q}^S_\alpha$ by L_1 and L_2, respectively. We obtain

$$|\hat{Q}_\alpha(x) - \hat{Q}^S_\alpha(x)| \leq (L_1 + L_2)\nu + |\hat{Q}_\alpha(x') - \hat{Q}^S_\alpha(x')|,$$

which gives

$$\sup_{x \in X} \left|\hat{Q}_\alpha(x) - \hat{Q}^S_\alpha(x)\right| \leq (L_1 + L_2)\nu + \sup_{x' \in X_\nu} \left|\hat{Q}_\alpha(x') - \hat{Q}^S_\alpha(x')\right|.$$

The first term $(L_1 + L_2)\nu$ can be made arbitrarily small by letting $\nu \to 0$. The result follows, because for fixed ν, the second term $\sup_{x' \in X_\nu} \left|\hat{Q}_\alpha(x') - \hat{Q}^S_\alpha(x')\right|$ goes to zero w.p. 1 as $S \to \infty$. To see this, fix any $x' \in X_\nu$, and consider the random variable

$$\xi := \max_{k=1,\ldots,K} \{\lambda^k(\omega^s - Tx') + \psi^k(\omega^s - \alpha)\}.$$

Thus, by the SLLN, $\hat{Q}^S_\alpha(x') \to \hat{Q}_\alpha(x')$ w.p. 1 as $S \to \infty$. We can apply the SLLN since $\mathbb{E}[\xi]$ exists and is finite by Assumptions (A2)-(A4). The result follows because $\mathcal{X}_\nu$ is finite. ■

Proof of Lemma 4.3. It follows from the definitions of $\hat{v}_\alpha$ and ν_α that $\hat{v}_\alpha \geq \nu_\alpha$. Moreover, we can take $R = \max_k w_k$, where w_k are the upper bounds on ψ^k, $k = 1, \ldots, K$, from Lemma 4.1.

To prove (ii), note that by the Basis Decomposition Theorem in [85], there exist basis matrices B^k, $k = 1, \ldots, K$, and closed convex cones $\Lambda^k = \{t : (B^k)^{-1}t \geq 0\}$ such that $\omega - Tx \in \Lambda^k$ implies $k(\omega, x) = k$, and thus

$$v_\alpha(\omega, x) = \lambda^k(\omega - Tx) + \psi^k(\omega - \alpha). \tag{4.21}$$

It remains to show that there exists $\sigma_k \in \Lambda^k$ such that $\hat{v}_\alpha(\omega, x) = \lambda^k(\omega - Tx) + \psi^k(\omega - \alpha)$ if $\omega - Tx \in \sigma_k + \Lambda^k$. Fix arbitrary $l \in \{1, \ldots, K\}$. It suffices to show that there exist $\sigma_{kl} \in \Lambda^k$ such that $\omega - Tx \in \sigma_{kl} + \Lambda^k$ implies that

$$\lambda^k(\omega - Tx) + \psi^k(\omega - \alpha) \geq \lambda^l(\omega - Tx) + \psi^l(\omega - \alpha). \tag{4.22}$$

This is because

$$\bigcap_{l=1}^{K} (\sigma_{kl} + \Lambda^k) = \sigma_k + \Lambda^k$$

for some $\sigma_k \in \Lambda^k$. Hence, if $\omega - Tx \in \sigma_k + \Lambda^k$, then

$$\hat{v}_\alpha(\omega, x) = \max_{k=1,\ldots,K} \left\{ \lambda^k(\omega - Tx) + \psi^k(\omega - \alpha) \right\} = \lambda^k(\omega - Tx) + \psi^k(\omega - \alpha).$$

To prove (4.22), note that if $\lambda^k = \lambda^l$, then by Lemma 4.1 (iii), $\psi^k(\omega - \alpha) = \psi^l(\omega - \alpha)$, so that (4.22) holds with equality. If $\lambda^k \neq \lambda^l$, then $\lambda^k s > \lambda^l s$ for any $s \in \text{int}(\Lambda^k)$. For sufficiently large $\gamma > 0$, we thus have

$$\gamma(\lambda^k s - \lambda^l s) \geq w_l.$$

If we take $\sigma_{kl} = \gamma s$, then (4.22) holds by observing that $\psi^k(\omega - \alpha) \geq 0$ and $\psi^l(\omega - \alpha) \leq w_l$. ■

Chapter 5

A converging Benders' decomposition algorithm for two-stage mixed-integer recourse models

We propose a new solution method for two-stage mixed-integer recourse models. In contrast to existing approaches, we can handle general mixed-integer variables in both stages. Our solution method is a Benders' decomposition, in which we iteratively construct tighter approximations of the expected second-stage cost function using a new family of optimality cuts. We derive these optimality cuts by parametrically solving extended formulations of the second-stage problems using deterministic mixed-integer programming techniques. We establish convergence by proving that the optimality cuts recover the convex envelope of the expected second-stage cost function. Finally, we demonstrate the potential of our approach by conducting numerical experiments on several investment planning and capacity expansion problems.

5.1 Introduction

Frequently, practical problems in, e.g., healthcare, energy, manufacturing, and logistics involve both uncertainty and integer decision variables. A powerful modelling tool for such problems is the class of two-stage mixed-integer recourse (MIR)

This chapter is based on the journal publication [78].

models [86, 32], but these models are notoriously hard to solve [29]. Typically, MIR models are solved using decomposition algorithms inspired by Benders' decomposition [15, 43]. However, existing decomposition approaches can only handle special cases of MIR models, or they are not attractive from a computational point of view. In this chapter, we develop a computationally efficient Benders' decomposition algorithm which solves *general* two-stage MIR models. In order to achieve this, we propose a new family of optimality cuts for MIR models, i.e., supporting hyperplanes which describe the expected second-stage cost function. The advantage of our so-called *scaled cuts* over existing optimality cuts is twofold. First, we prove that scaled cuts can be used to recover the convex envelope of the expected second-stage cost function in general, i.e., we do not require assumptions on the first- and second-stage decision variables. Second, scaled cuts can be computed efficiently using state-of-the-art techniques for deterministic mixed-integer programs (MIPs).

In a decomposition algorithm, optimality cuts are used to iteratively construct tighter *outer approximations* of the expected second-stage cost function. A prime example is the L-shaped method in [84], which efficiently solves continuous recourse models. We, however, focus on MIR models with mixed-integer second-stage decisions, which are much harder to solve, since the expected mixed-integer second-stage cost function is non-convex, and thus the rich toolbox of convex optimization cannot be used. It turns out that this difficulty is mitigated if the first-stage decision variables are pure binary. In fact, there is an array of decomposition algorithms developed for this special case [8, 31, 44, 48, 49, 51, 53, 69, 70, 72, 89]. However, these algorithms suffer from a positive *duality gap* when applied to MIR models with general mixed-integer first-stage variables, since they use optimality cuts which are in general not tight, see [24] and [73]. A notable exception is the algorithm in [88] for MIR models with pure integer first- and second-stage decision variables, but their approach does not apply to general mixed-integer variables. Existing solution methods for general MIR models are of limited practical use, since they branch on continuous first-stage variables [4, 24, 73], or they introduce auxiliary first-stage integer decision variables [6, 26].

Similar as in traditional decomposition algorithms for MIR models, we iteratively improve an outer approximation of the expected second-stage cost function. In contrast to traditional approaches, however, the *optimality cuts* that are used to update the outer approximation from one iteration to the next depend on the current outer approximation. More precisely, we propose a *recursive* scheme to update

the outer approximation, in which we solve extended formulations of the second-stage subproblems, whose definitions depend on the outer approximation in the current iteration. In this way, we derive non-linear optimality cuts for the non-convex second-stage cost functions, which we use to improve the current outer approximation. The problem is, of course, that non-linear optimality cuts introduce non-convexities in the master problem, which presents computational challenges. However, by transforming the non-linear optimality cuts, we obtain linear cuts for the *expected* second-stage cost function, which we refer to as scaled cuts.

We show that the coefficients of our scaled cuts are the optimal solutions of a static robust optimization problem with a mixed-integer uncertainty set, and use efficient row generation and cutting plane techniques to solve them, similar as in robust optimization and deterministic mixed-integer programming, respectively. Moreover, we prove that scaled cuts recover the convex hull of the expected second-stage cost function. In particular, we consider the *scaled cut closure* of a given outer approximation, defined as the pointwise supremum of all scaled cuts that we can compute using the current outer approximation, and we prove that the sequence of outer approximations defined by recursively computing the scaled cut closure converges to the convex envelope of the expected second-stage cost function. In addition, we prove that the scaled cut closure of a convex polyhedral outer approximation remains convex polyhedral. In other words, the scaled cut closure can be described using *finitely* many scaled cuts.

We use scaled cuts to develop a Benders' decomposition algorithm which solves two-stage MIR models with general mixed-integer variables in both stages. In this way, we close the duality gap of traditional optimality cuts. Since scaled cuts are linear in the first-stage decision variables, our Benders' decomposition algorithm is computationally efficient. In particular, we do not introduce auxiliary variables or require spatial branching of the first-stage feasible region for convergence. We do use a novel cut-enhancement technique to speed up convergence of the scaled cuts. The idea is to use the current outer approximation to identify solutions that cannot be optimal. Doing so allows us to construct stronger scaled cuts that do not have to be valid for these suboptimal solutions. We empirically test the quality of scaled cuts by conducting numerical experiments on variants of an investment planning problem (IPP) by Schultz et al. [67] and the DCAP problem instances [2] from SIPLIB [5]. Our results show that scaled cuts outperform traditional optimality cuts: they are able to significantly reduce the relative optimality gap between the best available lower and upper bound of traditional Benders' decomposition

algorithms. Indeed, on the IPP instances and the DCAP instances, we respectively achieve an average 93% and 45% reduction of this optimality gap compared to traditional optimality cuts, and, moreover, we achieve a zero optimality gap on 18 out of 24 IPP instances.

Summarizing, our main contributions are the following.

- We derive a new family of optimality cuts for MIR models, the scaled cuts, and we propose efficient strategies to compute these cuts.
- Using these scaled cuts, we develop a computationally efficient Benders' decomposition algorithm for MIR models with general mixed-integer variables in both stages.
- We prove that scaled cuts can be used to recover the convex envelope of the expected second-stage cost function.
- We propose an optimality cut-enhancement technique, which we use to speed up convergence of scaled cuts and to reduce the duality gap of traditional cuts.
- We conduct numerical experiments to test our scaled cuts, and we show that our (enhanced) scaled cuts can be used to close or significantly reduce the duality gap of traditional optimality cuts.

The remainder of this chapter is organized as follows. In Section 5.2, we formally introduce MIR models and review solution approaches. Next, we introduce scaled cuts and develop our Benders' decomposition algorithm in Section 5.3, and we describe several strategies to compute scaled cuts in Section 5.4. Section 5.5 concerns the proof of convergence of the scaled cuts. We report on our numerical experiments in Section 5.6, and we conclude in Section 5.7.

Notation: Throughout, $\text{conv}(U)$ denotes the convex hull of a set U. For a function $f : U \mapsto \mathbb{R} \cup \{\infty\}$, we define its convex envelope $\text{co}(f) : \text{conv}(U) \mapsto \mathbb{R}$ and its closed convex envelope $\overline{\text{co}}(f) : \text{conv}(U) \mapsto \mathbb{R}$ as the pointwise supremum of all convex, respectively affine, functions majorized by f. In addition, we define $\text{dom}(f) = \{x \in U : f(x) < \infty\}$. Finally, for any $U' \subseteq U$, we denote by $\text{epi}_{U'}(f)$ the epi-graph of f restricted to U', i.e., $\text{epi}_{U'}(f) := \{(x, \theta) \in U' \times \mathbb{R} : \ \theta \geq f(x)\}$, and we write $\text{epi}(f) = \text{epi}_U(f)$.

5.2 Problem description and literature review

5.2.1 Problem description

Two-stage recourse models explicitly model parameter uncertainty by a random vector $\boldsymbol{\omega}$ whose realization is unknown when a first-stage decision x has to be made. In contrast, the second-stage decision vector y is allowed to depend on the realization of $\boldsymbol{\omega}$, denoted ω, referred to as a scenario. We assume that the probability distribution of $\boldsymbol{\omega}$ is known, and we denote its support by Ω. A possible interpretation is that the first-stage decision corresponds to a long-term, strategic decision, concerning, e.g., facility location or investment planning, whereas the second-stage decisions are short-term in nature, corresponding to, e.g., routing adjustments or reordering decisions. We consider two-stage recourse models of the form

$$\eta^* := \min_x \{c^\top x + \mathbb{E}_{\boldsymbol{\omega}}[v_{\boldsymbol{\omega}}(x)] : \ Ax = b, \ x \in \mathcal{X}\}, \tag{5.1}$$

where the second-stage costs $v_\omega(x)$ are defined as

$$v_\omega(x) := \min_y \{q_\omega^\top y : \ W_\omega y = h_\omega - T_\omega x, \ y \in \mathcal{Y}\}, \quad x \in X, \ \omega \in \Omega, \tag{5.2}$$

and $v_\omega(x) = \infty$ if $x \notin X$, where $X := \{x \in \mathcal{X} : \ Ax = b\}$. Note that we consider randomness in all data elements of the second-stage problem. Furthermore, the sets $\mathcal{X}$ and $\mathcal{Y}$ may impose integer restrictions on the first- and second-stage decision variables, i.e., $\mathcal{X} = \mathbb{Z}_+^{p_1} \times \mathbb{R}_+^{n_1 - p_1}$ and $\mathcal{Y} = \mathbb{Z}_+^{p_2} \times \mathbb{R}_+^{n_2 - p_2}$. The resulting model is called a two-stage mixed-integer recourse model.

Throughout, we make the following assumptions.

(A1) For every $\omega \in \Omega$ and $x \in X$, we have $-\infty < v_\omega(x) < \infty$.

(A2) The support Ω of $\boldsymbol{\omega}$ is finite.

(A3) The first-stage feasible region X is non-empty and bounded.

(A4) The components of A, b, and W_ω, $\omega \in \Omega$ are rational, and for every $\omega \in \Omega$, the probability $\mathbb{P}(\boldsymbol{\omega} = \omega)$ is rational.

Assumption (A1) is known as relatively complete and sufficiently expensive recourse, and together with (A2) implies that $\mathbb{E}_{\boldsymbol{\omega}}[v_{\boldsymbol{\omega}}(x)]$ is finite for every $x \in X$. Furthermore, Assumption (A2) excludes the case where $\boldsymbol{\omega}$ follows a continuous

distribution. Nevertheless, continuous distributions are typically approximated by finite discrete distributions, e.g., using sample average approximation [42]. Finally, the assumptions in (A3) and (A4) guarantee that X is compact and $\bar{X} := \text{conv}(X)$ is a polytope [28, Theorem 1]. In addition, by (A4) the second-stage cost functions v_ω, $\omega \in \Omega$ are lower semi-continuous (lsc) on $\bar{X}$ [65], and thus, using that $\bar{X}$ is compact, v_ω is bounded from below on $\bar{X}$ [7, Theorem 3.7].

5.2.2 Decomposition for MIR models

If the number of scenarios $|\Omega|$ is large, then the MIR model in (5.1) is very difficult to solve directly due to its size. That is why Benders' decomposition [15] is widely used to solve MIR models, because it decomposes (5.1) into many much smaller subproblems. A Benders' decomposition algorithm maintains an outer approximation $\hat{Q}_{\text{out}} : \bar{X} \mapsto \mathbb{R}$ of the expected second-stage cost function $Q(x) := \mathbb{E}_{\boldsymbol{\omega}}[v_{\boldsymbol{\omega}}(x)]$, i.e., $\hat{Q}_{\text{out}}(x) \leq Q(x)\ \forall x \in X$. The corresponding relaxation of (5.1) defined as

$$\min_x \{c^\top x + \hat{Q}_{\text{out}}(x) :\ x \in X\} \tag{MP}$$

is referred to as the *master problem*, and an optimal solution $\bar{x}$ of (MP) is known as the *current solution*. Typically, $\hat{Q}_{\text{out}}$ is convex polyhedral, and thus (MP) can be solved efficiently. Note that if $\hat{Q}_{\text{out}}(\bar{x}) = Q(\bar{x})$, then $\bar{x}$ is also optimal in the original problem (5.1). If, however, $\hat{Q}_{\text{out}}(\bar{x}) < Q(\bar{x})$, then the outer approximation is strengthened using an *optimality cut* for Q:

$$Q(x) \geq \alpha - \beta^\top x \quad \forall x \in X,$$

which is such that $\alpha - \beta^\top \bar{x} > \hat{Q}_{\text{out}}(\bar{x})$, i.e., the outer approximation is strictly improved at $\bar{x}$. Next, the master problem (MP) is resolved using the strengthened outer approximation. We summarize Benders' decomposition for MIR models in Algorithm 2. Throughout, we maintain a lower- and upper bound LB and UB on η^*, i.e., $LB \leq \eta^* \leq UB$.

Algorithm 2 Benders' decomposition for MIR models.

1: **Initialization**
2: $\hat{Q}_{\text{out}} \equiv L$, where $Q(x) \geq L\ \forall x \in X$.
3: $LB \leftarrow -\infty$, $UB \leftarrow \infty$.
4: **Iteration step**
5: Solve (MP), denote optimal solution by $\bar{x}$ (current solution).

6: $LB \leftarrow c^\top \bar{x} + \hat{Q}_{\text{out}}(\bar{x})$.
7: $UB \leftarrow \min\left\{c^\top \bar{x} + Q(\bar{x}), UB\right\}$
8: Compute optimality cut $Q(x) \geq \alpha - \beta^\top x \; \forall x \in X$.
9: **Stopping criterion**
10: **if** $UB - LB < \varepsilon$ **then stop**: **return** $\bar{x}$
11: **else**
12: Add optimality cut to (MP):

$$\hat{Q}_{\text{out}}(x) \leftarrow \max\left\{\hat{Q}_{\text{out}}(x), \alpha - \beta^\top x\right\}, \quad x \in X.$$

13: Go to line 5.
14: **end if**

Typically, optimality cuts are tight for Q at the current solution $\bar{x}$, i.e., $\alpha - \beta^\top \bar{x} = Q(\bar{x})$, which ensures that the outer approximation strictly improves at $\bar{x}$, and as a result, we find a different solution in the next iteration. In some cases, however, the optimality cuts are not tight at $\bar{x}$, see, e.g., [89], and thus the algorithm may stall. Therefore, in a practical implementation of Algorithm 2, we stop in these cases if the outer approximation improves by less than ε at $\bar{x}$, i.e., if $\hat{Q}_{\text{out}}(\bar{x}) > \alpha - \beta^\top \bar{x} - \varepsilon$, and on termination, we return the best incumbent solution that we encountered during the algorithm.

An important observation is that Algorithm 2 allows for decomposition by scenario: optimality cuts for Q can be computed by aggregating optimality cuts for the second-stage cost functions v_ω, $\omega \in \Omega$. For example, the L-shaped method in [84], which solves continuous recourse models, uses optimality cuts of the form

$$v_\omega(x) \geq \alpha_\omega - \beta_\omega^\top x \quad \forall x \in X, \; \omega \in \Omega, \tag{5.3}$$

where α_ω and β_ω depend on the dual multipliers of the second-stage subproblems. Taking expectations then yields the optimality cut $Q(x) \geq \mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top x \; \forall x \in X$. In fact, Benders' decomposition algorithms that generalize the L-shaped method to more general classes of MIR models typically use the same strategy to compute optimality cuts, i.e., cuts of the form (5.3) are aggregated to derive optimality cuts for Q. We review such generalizations in Section 5.2.2.1. However, if optimality cuts are computed by aggregating cuts of the form (5.3), then the resulting Benders' decomposition algorithm is not able to solve MIR models with general mixed-integer variables in both stages, as we explain in Section 5.3. Therefore, in Section 5.3.1, we

develop a modified Benders' decomposition algorithm by proposing a new family of optimality cuts which is suited for general MIR models.

5.2.2.1 Generalizations to mixed-integer recourse

If the recourse is continuous, i.e., if $\mathcal{Y} = \mathbb{R}^{n_2}_+$, then the L-shaped method converges to the optimal solution, since the expected second-stage cost function Q is a convex polyhedral function. In contrast, if $\mathcal{Y}$ is a mixed-integer set, then Q is in general not convex, or even continuous, see, e.g., [65]. Therefore, the L-shaped method does not readily generalize to broader classes of MIR models. However, Laporte and Louveaux [44] show that if the first-stage decisions are binary, i.e., if $\mathcal{X} = \mathbb{B}^{n_1}$, then there exists a finite family of optimality cuts which describe Q. In other words, there exists a convex polyhedral outer approximation $\hat{Q}_{\text{out}}$ of Q defined on $\bar{X}$ such that $\hat{Q}_{\text{out}}(x) = Q(x)\ \forall x \in X$. They use this result to develop the integer L-shaped algorithm for MIR models with $\mathcal{X} = \mathbb{B}^{n_1}$, see also [8].

In fact, there exists a wide range of algorithms generalizing the L-shaped method to the special case where $\mathcal{X} = \mathbb{B}^{n_1}$, that typically use techniques for deterministic MIPs. For example, the algorithms in [31, 48, 49, 51, 69, 72] use cutting planes to derive strong continuous relaxations of the second-stage subproblems. These *parametric* cutting planes depend linearly on the first-stage decision vector x, and thus they can be re-used in subsequent iterations. Moreover, since the resulting relaxation of the second-stage problem is continuous, LP-duality can be used to derive optimality cuts for the second-stage cost functions. In general, convergence of these methods is only guaranteed if $\mathcal{X} = \mathbb{B}^{n_1}$, since this condition ensures that the continuous relaxations defined by the parametric cutting planes are tight. However, Zhang and Küçükyavuz [88] manage to generalize the approach based on Gomory cuts in [31] to pure integer MIR models, i.e., $\mathcal{X} = \mathbb{Z}^{n_1}_+$ and $\mathcal{Y} = \mathbb{Z}^{n_2}_+$, by identifying feasible basis matrices of the extended formulation. Moreover, for the case where $\mathcal{Y}$ is a mixed-integer set, Kim and Mehrotra [40] use mixed-integer rounding cuts to derive tight continuous relaxations for a nurse scheduling problem with a totally unimodular recourse matrix, and Bansal et al. [13] derive special cases where the second-stage feasible regions can be convexified via parametric cutting planes.

In another direction, Sen and Sherali [70] use branch-and-bound for MIPs to obtain a disjunctive characterization of the second-stage cost functions v_ω, $\omega \in \Omega$. They then use techniques from disjunctive programming to construct a convex relaxation of v_ω, and show that their approximation is exact if x is an extreme point of $\bar{X}$. As a consequence, the resulting D^2-BAC algorithm solves two-stage MIR

models with $\mathcal{X} = \mathbb{B}^{n_1}$. A different approach is taken by Zou et al. [89], who develop the SDDiP algorithm for multi-stage MIR models with binary state variables. They construct tight lower-bounding approximation of the second-stage cost functions using *Lagrangian cuts*, which are computed by solving Lagrangian relaxations of specific reformulations of the second-stage subproblems. However, Lagrangian cuts are not tight in case of general mixed-integer state variables.

An alternative to Benders' decomposition is the dual decomposition algorithm of Carøe and Schultz [24], which solves the Lagrangian relaxation of the MIR model in (5.1) obtained by introducing copies of the first-stage variables for each scenario, and by relaxing the *non-anticipativity* constraints on the copy variables. Because the Lagrangian dual bound is in general not tight, Carøe and Schultz propose a spatial branch-and-bound search, where the first-stage feasible region is partitioned into smaller subsets in order to determine good global lower and upper bounds. Similarly, Ahmed et al. [4] and Trapp et al. [76] develop global branch-and-bound algorithms for the special case where the second-stage problem is pure integer and the technology matrix T_ω is fixed, by analysing level sets of the second-stage cost functions v_ω. In addition, spatial branching strategies have also been applied in Benders' decomposition to ensure convergence in the case where $\mathcal{X}$ is a mixed-integer set. See, e.g., Qi and Sen [53] and Sherali and Zhu [73], who use parametric cutting planes to approximate the second-stage cost functions v_ω at the nodes of the first-stage branch-and-bound tree, and show that if the partition is sufficiently fine, then the resulting algorithms converge to the global optimal solution. In general, however, the practical use of these approaches is limited, since the spatial branch-and-bound tree may grow very large, if we have to branch on continuous first-stage variables.

In fact, there does not exist a *computationally efficient* decomposition algorithm for two-stage MIR models with general mixed-integer variables in both stages. We provide this missing link by proposing scaled cuts for MIR models. The advantage of the resulting Benders' decomposition algorithm compared to existing solution methods for general MIR models is that we do not use spatial branching of the first-stage feasible region or auxiliary integer variables for convergence. In contrast, Carøe and Tind [26] use auxiliary integer decision variables to capture non-convex terms in the master problem, which are determined using general duality for MIPs. Similarly, the stochastic Lipschitz dynamic programming algorithm in [6] introduces binary variables to include non-linear optimality cuts in the master problem.

5.3 Benders' decomposition for general MIR models

In this section, we introduce our family of linear optimality cuts for the expected second-stage cost function Q. Using these so-called *scaled cuts*, we are able to recover the convex envelope $\mathrm{co}(Q)$ of Q, so that we can solve the MIR model in (5.1) by replacing $Q(x)$ by $\mathrm{co}(Q)(x)$ and the feasible region X by its convex hull $\bar{X}$. That is, the resulting convex relaxation of the original problem in (5.1), defined as

$$\hat{\eta} := \min_x \{c^\top x + \mathrm{co}(Q)(x) : \ x \in \bar{X}\}, \tag{5.4}$$

satisfies $\hat{\eta} = \eta^*$, and moreover, if x^* is optimal in the original problem (5.1), then x^* is also optimal in (5.4), see, e.g., [74, Proposition 2.4].

In contrast, traditional Benders' decomposition algorithms for MIR models, see, e.g., [31, 69, 72], use optimality cuts which, in general, do not yield $\mathrm{co}(Q)$. More precisely, if we compute optimality cuts for Q by aggregating linear cuts $v_\omega(x) \geq \alpha_\omega - \beta_\omega^\top x \ \forall x \in X$ for the second-stage cost functions, then we obtain at most $\mathbb{E}_{\boldsymbol{\omega}}[\mathrm{co}(v_{\boldsymbol{\omega}})]$. However, this expected value of the convex envelope of the second-stage cost functions v_ω is not the same as the convex envelope of the expected second-stage cost function Q. In fact, in general $\mathbb{E}_{\boldsymbol{\omega}}[\mathrm{co}(v_{\boldsymbol{\omega}})(x)] \leq \mathrm{co}(Q)(x)$, resulting in a *duality gap*, see also [21, 24]. This gap is zero if $\mathcal{X} = \mathbb{B}^{n_1}$ [89, Theorem 1], but if $\mathcal{X}$ is a general mixed-integer set, then the duality gap may be positive, see Example 5.1.

Remark 5.1. In general, any family of linear optimality cuts for Q yields at most its *closed* convex envelope $\overline{\mathrm{co}}(Q)$. However, since Q is lsc and X is compact, we have that $\overline{\mathrm{co}}(Q) = \mathrm{co}(Q)$ [30, Theorem 2.2]. Similarly, $\overline{\mathrm{co}}(v_\omega) = \mathrm{co}(v_\omega)$ for every $\omega \in \Omega$.

Example 5.1. Consider the expected second-stage cost function $Q(x) = \mathbb{E}_{\boldsymbol{\omega}}[v_{\boldsymbol{\omega}}(x)]$, $x \in [0, 4]$, where

$$v_\omega(x) = \min_y \{2y : y \geq \omega - x, \ y \in \mathbb{Z}_+\}, \quad x \in [0, 4],$$

and $\boldsymbol{\omega}$ is discretely distributed with mass points $\omega_1 = 2.5$ and $\omega_2 = 3$, both with probability $1/2$. The function Q is known as a simple integer recourse (SIR) function, see, e.g., [45]. For a given ω and x, the optimal second-stage decision y is the smallest non-negative integer such that $y \geq \omega - x$, denoted by $\lceil \omega - x \rceil^+$, and thus $v_\omega(x) = 2\lceil \omega - x \rceil^+$. Furthermore, straightforward computations yield $\mathrm{co}(v_{\omega_1})(x) = 2\max\{0, 2.5 - x, 3 - 2x\}$ and $\mathrm{co}(v_{\omega_2})(x) = 2\max\{0, 3 - x\}$.

In general, the difference between $\text{co}(v_\omega)(x)$ and $v_\omega(x)$ not equal to zero, and, moreover, the values of x for which $\text{co}(v_\omega)(x) = v_\omega(x)$ are not the same for $\omega = \omega_1$ and $\omega = \omega_2$. Indeed, straightforward computations yield that $\text{co}(v_{\omega_1})(x) = v_{\omega_1}(x)$ only if $x \in \{0, 1/2, 3/2\}$ or $x \geq 5/2$, and $\text{co}(v_{\omega_2})(x) = v_{\omega_2}(x)$ if $x \in \{0, 1, 2\}$ or $x \geq 3$. This results in a positive duality gap between $\text{co}(Q)(x)$ and $\mathbb{E}_{\boldsymbol{\omega}}[\text{co}(v_{\boldsymbol{\omega}})(x)]$, see Figure 5.1. For example, at $x = 1$, we have $\text{co}(Q)(1) = Q(1) = 4$, but $\mathbb{E}_{\boldsymbol{\omega}}[\text{co}(v_{\boldsymbol{\omega}})(1)] = 3.5$, i.e., the duality gap at $x = 1$ is equal to $1/2$. $\Diamond$

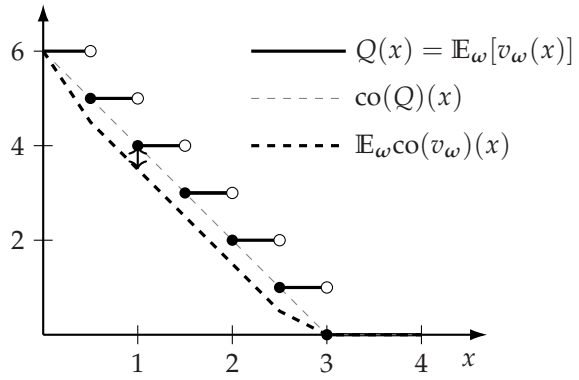


Figure 5.1. The duality gap for MIR models: the difference between $\text{co}(Q)(x)$ and $\mathbb{E}_{\boldsymbol{\omega}}\text{co}(v_{\boldsymbol{\omega}})(x)$ in Example 5.1 is in general non-negative, and equal to $1/2$ if, e.g., $x = 1$.

The duality gap illustrated in Example 5.1 may be closed using scaled cuts, which we derive in Section 5.3.1. Indeed, in Section 5.3.2, we show that scaled cuts can be used to recover $\text{co}(Q)$, see Theorem 5.1, and we use them to develop a Benders' decomposition algorithm which solves MIR models with general mixed-integer variables.

5.3.1 Scaled cuts for MIR models

We approximate the expected second-stage cost function Q using linear optimality cuts, in order to ensure that the master problem can be solved efficiently. Evidently, we may obtain such cuts by aggregating linear optimality cuts for the second-stage cost functions of the form $v_\omega(x) \geq \alpha_\omega - \beta_\omega^\top x \ \forall x \in X$, but Example 5.1 illustrates that the resulting cut

$$Q(x) \geq \mathbb{E}_{\boldsymbol{\omega}}\alpha_{\boldsymbol{\omega}} - \mathbb{E}_{\boldsymbol{\omega}}\beta_{\boldsymbol{\omega}}^\top x \quad \forall x \in X, \tag{5.5}$$

is in general not tight. Instead, we may use non-linear cuts to construct tight non-convex approximations of v_ω and Q, but the resulting master problem is highly non-convex, and thus solving it is in general not realistic from a computational point of view. That is why we propose to use non-linear optimality cuts for v_ω, $\omega \in \Omega$, and we transform these cuts into linear cuts for Q, thereby maintaining a tractable master problem. The resulting scaled cuts generally yield stronger outer approximations than cuts of the form (5.5), and, in fact, they may be used to close the duality gap illustrated in Example 5.1.

More precisely, we consider cuts for v_ω, $\omega \in \Omega$, of the form

$$v_\omega(x) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega \phi(x) \quad \forall x \in X, \tag{5.6}$$

where $\phi : \bar{X} \mapsto \mathbb{R}$ is a convex polyhedral function, referred to as a cut-generating function, and $\tau_\omega \geq 0$. For example, Ahmed et al. [6] derive cuts of the form (5.6) using $\phi(x) = ||x - \bar{x}||$, where $||\cdot||$ is a norm on $\mathbb{R}^{n_1}$. We, however, propose to use $\phi = \hat{Q}_{\text{out}}$, where $\hat{Q}_{\text{out}}$ is a convex polyhedral outer approximation of Q, i.e., $\hat{Q}_{\text{out}}(x) \leq Q(x)\ \forall x \in X$. The advantage of using $\phi = \hat{Q}_{\text{out}}$ becomes clear if we take expectations on both sides of (5.6), yielding

$$Q(x) \geq \mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top x - \mathbb{E}_\omega \tau_\omega \phi(x) \quad \forall x \in X, \tag{5.7}$$

and use $\phi(x) \leq Q(x)$ and $E_\omega \tau_\omega \geq 0$ to obtain that $Q(x) \geq \mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top x - \mathbb{E}_\omega \tau_\omega Q(x)$, which we rearrange as

$$Q(x) \geq \frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top x}{1 + \mathbb{E}_\omega \tau_\omega} \quad \forall x \in X.$$

In particular, this so-called *scaled cut* is linear in the first-stage decision vector x, and is therefore suitable for efficient computations, whereas the cut in (5.7) introduces non-linear, non-convex terms in the master problem, which is undesirable from a computational point of view.

We formally introduce scaled cuts in Definition 5.1, and in Example 5.2 we illustrate how to compute a scaled cut for the SIR model of Example 5.1. Furthermore, in Section 5.4, we explain how to compute the cut coefficients α_ω, β_ω, and τ_ω in (5.6) if v_ω is a general mixed-integer second-stage cost function. For technical reasons, we assume throughout that $\text{epi}(\phi)$ is a rational polyhedron; if ϕ satisfies this condition, we say that ϕ is a *rational convex polyhedral function*.

Definition 5.1 (scaled cuts)**.** Let $\phi : \bar{X} \mapsto \mathbb{R}$ be a rational convex polyhedral func-

tion such that $\phi(x) \leq Q(x)\ \forall x \in X$, and denote by $\Pi_\omega(\phi)$ the set of cut coefficients which define optimality cuts of the form (5.6), i.e.,

$$\Pi_\omega(\phi) := \{(\alpha, \beta, \tau) :\ v_\omega(x) \geq \alpha - \beta^\top x - \tau\phi(x)\ \forall x \in X,\ \tau \geq 0\},\ \omega \in \Omega.$$

If $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi)$ for every $\omega \in \Omega$, then the optimality cut

$$Q(x) \geq \frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top x}{1 + \mathbb{E}_\omega \tau_\omega} \quad \forall x \in X \tag{5.8}$$

is referred to as a *scaled cut*.

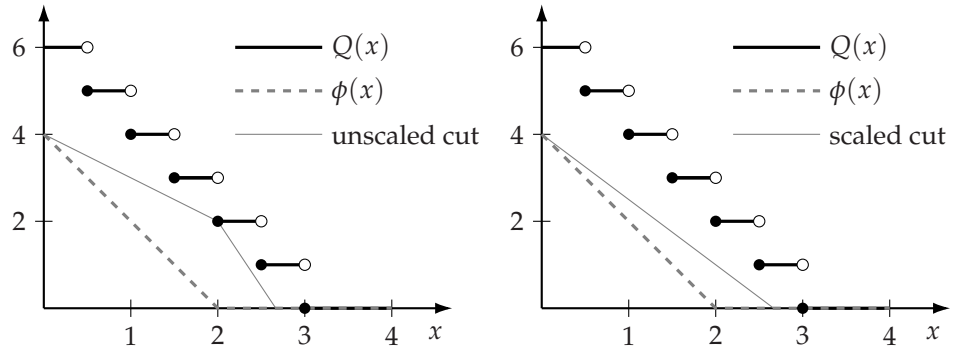
Example 5.2 (Example 5.1 continued)**.** Consider the SIR function Q of Example 5.1. Note that $Q(x) \geq 0$ and $Q(x) \geq 4 - 2x$ for every $x \in [0,4]$, and thus an outer approximation of Q is given by $\hat{Q}_{\text{out}}(x) = \max\{0, 4-2x\},\ x \in [0,4]$. Therefore, we can use $\phi = \hat{Q}_{\text{out}}$ as a cut-generating function to derive a scaled cut for Q at, e.g., $\bar{x} = 2$. To this end, we compute cuts of the form $v_\omega(x) \geq \alpha - \beta x - \tau\phi(x)\ \forall x \in [0,4]$, $\omega \in \{\omega_1, \omega_2\}$, which are tight at $\bar{x}$. In particular, it is easy to verify that the cuts $v_{\omega_1}(x) \geq 10 - 4x - 2\phi(x)\ \forall x \in [0,4]$, and $v_{\omega_2}(x) \geq 6 - 2x\ \forall x \in [0,4]$ are tight at $\bar{x}$.

Since the cuts for v_{ω_1} and v_{ω_2} are tight at $\bar{x}$, the resulting *unscaled cut*

$$Q(x) \geq 1/2(10 - 4x - 2\phi(x)) + 1/2(6 - 2x) = 8 - 3x - \phi(x) \quad \forall x \in [0,4],$$

is also tight at $\bar{x} = 2$, see Figure 5.2a. We show the corresponding scaled cut $Q(x) \geq (8-3x)/2\ \forall x \in [0,4]$ in Figure 5.2b. Figures 5.2a and 5.2b reveal the following geometric interpretation of scaled cuts: they pass through those points where the cut-generating function $\phi(x)$ and the unscaled cut $\alpha - \beta^\top x - \tau\phi(x)$ intersect. Indeed, if x is such that $\phi(x) = \alpha - \beta^\top x - \tau\phi(x)$, then

$$\frac{\alpha - \beta^\top x}{1 + \tau} = \phi(x) = \alpha - \beta^\top x - \tau\phi(x). \qquad \Diamond$$



(a) unscaled cut **(b)** scaled cut

Figure 5.2. The left figure shows the unscaled cut $Q(x) \geq 8 - 3x - \phi(x)\ \forall x \in [0,4]$ of Example 5.2, where $\phi(x) = \max\{0, 4-2x\},\ x \in [0,4]$. The right figure shows the corresponding scaled cut $Q(x) \geq (8-3x)/2\ \forall x \in [0,4]$.

5.3.2 Benders' decomposition with scaled cuts

We propose to solve MIR models using a Benders' decomposition with scaled cuts, where in every iteration, we use the current outer approximation $\hat{Q}_{\text{out}}$ as the cut-generating function, i.e., $\phi = \hat{Q}_{\text{out}}$. Thus, we use a different cut-generating function in every iteration of our Benders' decomposition, since we update the outer approximation from one iteration to the next.

We will show that a sufficient condition for the outer approximation $\hat{Q}_{\text{out}}$ to improve at the current solution $\bar{x}$ of the Benders' decomposition algorithm is that the non-linear cuts of the non-convex second-stage cost functions v_ω are tight at $\bar{x}$, similar as in Example 5.2. In Lemma 5.1, we derive general sufficient conditions for the cut-generating function ϕ so that such a tight non-linear cut exists.

Lemma 5.1. *Let $\bar{x} \in X$ be given, and let $\phi : \bar{X} \mapsto \mathbb{R}$ be a rational convex polyhedral function. If $(\bar{x}, \phi(\bar{x}))$ is an extreme point of $\text{conv}(\text{epi}_X(\phi))$, then there exist α, β, and $\tau \geq 0$ such that the optimality cut $v_\omega(x) \geq \alpha - \beta^\top x - \tau\phi(x)\ \forall x \in X$ is tight at $\bar{x}$, i.e., $v_\omega(\bar{x}) = \alpha - \beta^\top \bar{x} - \tau\phi(\bar{x})$.*

Proof. See Appendix 5.A. ■

To see how we can use scaled cuts to iteratively improve outer approximations in our Benders' decomposition algorithm, consider the master problem

$$\eta^* = \min_x \{c^\top x + \hat{Q}_{\text{out}}(x) :\ x \in X\}, \tag{MP}$$

and note that (MP) has an optimal current solution $\bar{x}$ such that $(\bar{x}, \hat{Q}_{\text{out}}(\bar{x}))$ is an extreme point of $\text{conv}(\text{epi}_X(\hat{Q}_{\text{out}}))$. Hence, if $\phi = \hat{Q}_{\text{out}}$, then by Lemma 5.1, there exist cut coefficients $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi)$, such that the corresponding cut for v_ω is tight at $\bar{x}$, i.e., $v_\omega(\bar{x}) = \alpha_\omega - \beta_\omega^\top \bar{x} - \tau_\omega \phi(\bar{x})$, $\omega \in \Omega$, and thus, if we assume that $\hat{Q}_{\text{out}}(\bar{x}) < Q(\bar{x})$, then the scaled cut in (5.8) improves $\hat{Q}_{\text{out}}$ in $\bar{x}$, since

$$\begin{aligned}\frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x}}{1 + \mathbb{E}_\omega \tau_\omega} &= \frac{\mathbb{E}_\omega [v_\omega(\bar{x}) + \tau_\omega \phi(\bar{x})]}{1 + \mathbb{E}_\omega \tau_\omega} \\ &= \frac{Q(\bar{x}) + \mathbb{E}_\omega \tau_\omega \phi(\bar{x})}{1 + \mathbb{E}_\omega \tau_\omega} > \phi(\bar{x}) = \hat{Q}_{\text{out}}(\bar{x}),\end{aligned}$$

where the inequality follows from $Q(\bar{x}) > \hat{Q}_{\text{out}}(\bar{x}) = \phi(\bar{x})$.

Moreover, it follows from $v_\omega(\bar{x}) = \alpha_\omega - \beta_\omega^\top \bar{x} - \tau_\omega \phi(\bar{x})$ $\forall \omega \in \Omega$ that the unscaled cut $Q(x) \geq \mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top x - \mathbb{E}_\omega \tau_\omega \phi(x)$ is tight at $\bar{x}$. Note that if $\mathbb{E}_\omega \tau_\omega = 0$, then the resulting scaled cut in (5.8) coincides with both the unscaled cut and the traditional cut in (5.5). In general, however, it is not true that $\mathbb{E}_\omega \tau_\omega = 0$, see Example 5.1, and if instead $\mathbb{E}_\omega \tau_\omega > 0$, then the scaled cut in (5.8) is *not* tight at $\bar{x}$, unless $\hat{Q}_{\text{out}}(\bar{x}) = Q(\bar{x})$, since

$$\frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x}}{1 + \mathbb{E}_\omega \tau_\omega} = \frac{Q(\bar{x}) + \mathbb{E}_\omega \tau_\omega \phi(\bar{x})}{1 + \mathbb{E}_\omega \tau_\omega} < Q(\bar{x}),$$

where the inequality is due to $\mathbb{E}_\omega \tau_\omega > 0$ and $\phi(\bar{x}) = \hat{Q}_{\text{out}}(\bar{x}) < Q(\bar{x})$. In fact, the larger the *scaling factor* $\mathbb{E}_\omega \tau_\omega$, the less the scaled cut in (5.8) improves the outer approximation at $\bar{x}$. As a result, the scaled cut obtained by computing tight nonlinear cuts for v_ω is not necessarily the *dominating* scaled cut, i.e., the scaled cut which yields the most improvement of $\hat{Q}_{\text{out}}$ at $\bar{x}$. In our Benders' decomposition algorithm, however, we will use such dominating scaled cuts to iteratively improve the outer approximation. Therefore, in Section 5.4, we show how to compute such cuts by solving

$$\rho^* := \sup_{\alpha_\omega, \beta_\omega, \tau_\omega} \left\{ \frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x}}{1 + \mathbb{E}_\omega \tau_\omega} : \ (\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi) \ \forall \omega \in \Omega \right\}. \tag{5.9}$$

First, however, we state our main result: we can recover $\text{co}(Q)$ via dominating scaled cuts. In particular, we define the scaled cut closure of a cut-generating function ϕ as the pointwise supremum of all scaled cuts corresponding to ϕ, see Definition 5.2, and we show that the sequence of outer approximations obtained by recursively computing the scaled cut closure converges uniformly to $\text{co}(Q)$, see

Theorem 5.1.

Definition 5.2 (Scaled cut closure)**.** Let $\phi : \bar{X} \mapsto \mathbb{R}$ be a rational convex polyhedral function. Then, the scaled cut closure $\mathrm{SCC}(\phi) : \bar{X} \mapsto \mathbb{R}$ of ϕ is defined as

$$\mathrm{SCC}(\phi)(x) := \sup_{\alpha_\omega, \beta_\omega, \tau_\omega} \left\{ \frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top x}{1 + \mathbb{E}_\omega \tau_\omega} : (\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi) \ \forall \omega \in \Omega \right\}, \quad x \in \bar{X}.$$

The definition of the scaled cut closure implies that $\mathrm{SCC}(\phi)$ can be described using *infinitely* many scaled cuts. It turns out, however, that $\mathrm{SCC}(\phi)$ is convex polyhedral, see Proposition 5.1, i.e., $\mathrm{SCC}(\phi)$ is the pointwise supremum of *finitely* many optimality cuts. Furthermore, if $\phi \leq Q$, then $\mathrm{SCC}(\phi) \leq Q$, since the scaled cuts of Definition 5.1 are valid if $\phi \leq Q$. However, the scaled cut closure of ϕ is defined for an arbitrary convex polyhedral function ϕ, i.e., we do not require that $\phi \leq Q$. This is because we may compute scaled cuts using an *inexact* outer approximation of Q, obtained, e.g., by solving the convex approximations of MIR models proposed in [60] and in Chapter 4. In fact, we prove that for an arbitrary convex polyhedral approximation ϕ_0 of Q, scaled cuts are able to recover the convex envelope of $\max\{\phi_0, Q\}$.

Proposition 5.1. *Let $\phi : \bar{X} \mapsto \mathbb{R}$ be a rational convex polyhedral function. Then,* $\mathrm{SCC}(\phi)$ *is a rational convex polyhedral function.*

Proof. See Appendix 5.A. ■

Theorem 5.1. *Let $\phi_0 : \bar{X} \mapsto \mathbb{R}$ be a rational convex polyhedral function. Recursively define the sequence $\{\phi_k\}_{k \geq 0}$ as $\phi_{k+1} = \mathrm{SCC}(\phi_k)$, $k \geq 0$. Then, ϕ_k converges uniformly to* $\mathrm{co}(\max\{\phi_0, Q\})$*. In particular, if $\phi_0(x) \leq Q(x)$ $\forall x \in X$, then $\phi_k \to \mathrm{co}(Q)$.*

Proof. The proof is postponed to Section 5.5. ■

Theorem 5.1 implies that if ϕ_0 is defined as, e.g., a trivial lower bound of Q, or the LP-relaxation of Q, obtained by relaxing the integer restrictions on the second-stage decision variables y, then we can recover $\mathrm{co}(Q)$ using scaled cuts, thereby solving the MIR model in (5.1). If, however, ϕ_0 is an inexact outer approximation obtained by solving a convex approximation of (5.1), then we may use scaled cuts to improve the quality of the resulting solution.

5.4 Computation of dominating scaled cuts

In this section, we describe how to solve the optimization problem in (5.9) for computing a dominating scaled cut at $\bar{x}$. Note that solving (5.9) presents a significant challenge, since it features a non-linear objective function. A natural way to address this challenge is to linearise the objective function by introducing a penalty parameter ρ, penalizing large vales of $1 + \mathbb{E}_\omega \tau_\omega$, yielding

$$C(\rho) := \sup_{\alpha_\omega, \beta_\omega, \tau_\omega} \big\{ \mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x} - \rho(1 + \mathbb{E}_\omega \tau_\omega) : \\ (\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi),\ \omega \in \Omega \big\}. \quad (5.10)$$

For arbitrary values of ρ, this linearised optimization problem merely represents an approximation of (5.9). However, we claim that if $C(\rho) = 0$ and if the supremum in (5.10) is attained by some $(\alpha_\omega, \beta_\omega, \tau_\omega)$, i.e.,

$$\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x} - \rho(1 + \mathbb{E}_\omega \tau_\omega) = 0, \quad (5.11)$$

then $\rho = \rho^*$, and $(\alpha_\omega, \beta_\omega, \tau_\omega)$ is optimal in (5.9). To see this, note that if $C(\rho) = 0$, then from (5.10), we have that $\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x} - \rho(1 + \mathbb{E}_\omega \tau_\omega) \leq 0$ if $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi)$ for every $\omega \in \Omega$, and thus

$$\frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x}}{1 + \mathbb{E}_\omega \tau_\omega} \leq \rho. \quad (5.12)$$

Since the expression on the left-hand side of (5.12) is the objective function of the problem in (5.9), an upper bound on its optimal value is given by ρ, and moreover, we obtain from (5.11) that $(\alpha_\omega, \beta_\omega, \tau_\omega)$ attains this upper bound.

Instead of solving (5.9), we thus find ρ such that $C(\rho) = 0$, and we solve the corresponding linearised problem in (5.10) to obtain the dominating scaled cut parameters $(\alpha_\omega, \beta_\omega, \tau_\omega)$. In Section 5.4.1, we describe how to efficiently solve $C(\rho) = 0$ for ρ using a *fixed point iteration* algorithm.

5.4.1 Fixed point iteration algorithm

In particular, we take advantage of several properties of $C(\cdot)$ in Lemma 5.2 below, namely that $C(\cdot)$ is strictly decreasing, continuous and convex.

Lemma 5.2. *Let $\bar{x} \in \bar{X}$ be given and let $\phi : \bar{X} \mapsto \mathbb{R}$ be a rational convex polyhedral function. Then,*

(i) the value function $C(\cdot)$ defined in (5.10) *is continuous, convex, and strictly decreasing on* $\mathrm{dom}(C) = \{\rho : C(\rho) < \infty\}$,

(ii) the supremum in (5.10) *is attained if* $\rho \in \mathrm{dom}(C)$,

(iii) for $\bar{\rho} \in \mathrm{dom}(C)$, *a subgradient of* $C(\cdot)$ *at* $\bar{\rho}$ *is given by* $-(1 + \mathbb{E}_\omega \tau_\omega)$, *where* τ_ω *corresponds to an optimal solution of the problem in* (5.10) *with* $\rho = \bar{\rho}$, *and*

(iv) if $\bar{x} \in X$, *then* $\mathrm{dom}(C) = [\phi(\bar{x}), \infty)$.

Proof. See Appendix 5.A. ■

Lemma 5.2 shows that if the penalty parameter ρ is not large enough, i.e., if $\bar{x} \in X$ and $\rho < \phi(\bar{x})$, then we have $C(\rho) = \infty$. Typically, for $\rho = \phi(\bar{x})$, we have $C(\rho) > 0$ and then $C(\cdot)$ continuously decreases until $C(\rho) = 0$ for $\rho = \rho^*$. There are, however, exceptions for which $C(\rho) < 0$ for all $\rho \in \mathrm{dom}(C)$, leading to the following characterization of ρ^* in Lemma 5.3 that holds in general.

Lemma 5.3. *Let* $\bar{x} \in \bar{X}$ *be given and let* $\phi : \bar{X} \mapsto \mathbb{R}$ *be a rational convex polyhedral function. Then, the optimal value* ρ^* *of the problem in* (5.9) *satisfies*

$$\rho^* = \min_{\rho} \{\rho : \ C(\rho) \leq 0\}. \tag{5.13}$$

In particular, if $\bar{x} \in X$ *and* $\rho^* > \phi(\bar{x})$, *then* ρ^* *is the unique solution of* $C(\rho) = 0$.

Proof. See Appendix 5.A. ■

To compute the dominating scaled cut parameters for a given $\bar{x} \in X$ in our Benders' decomposition, we use an iterative approach to obtain ρ^*. First we compute $C(\rho_0)$ for $\rho_0 = \phi(\bar{x})$. If $C(\rho_0) \leq 0$, then we can stop: $\rho^* = \rho_0$. Otherwise, we conclude that ρ_0 is a lower bound for ρ^*, i.e., $\rho_0 < \rho^*$, since $C(\cdot)$ is strictly decreasing. However, since $C(\cdot)$ is convex we can immediately derive a better lower bound for ρ^* without any additional computations. This lower bound, denoted ρ_1, is the value of ρ for which the right-hand side of the subgradient inequality

$$C(\rho) \geq C(\rho_0) - (1 + \mathbb{E}_\omega \tau_\omega)(\rho - \rho_0) \quad \forall \rho \in \mathbb{R}.$$

equals 0. That is, $\rho_1 = \rho_0 + C(\rho_0)/(1 + \mathbb{E}_\omega \tau_\omega)$. Note that $\rho_1 > \rho_0$, since $C(\rho_0) > 0$ and $1 + \mathbb{E}_\omega \tau_\omega > 0$.

In general, we iteratively compute ρ_k, $k \geq 0$, using the updating rule

$$\rho_{k+1} = \rho_k + \frac{C(\rho_k)}{1 + \mathbb{E}_\omega \tau_\omega}, \tag{5.14}$$

where τ_ω corresponds to an optimal solution of the problem in (5.10) with $\rho = \rho_k$. It follows from convexity of $C(\cdot)$ that the resulting sequence $\{\rho_k\}_{k\geq 0}$ is non-decreasing. To see this, substitute $\rho = \rho_{k+1}$ in the subgradient inequality

$$C(\rho) \geq C(\rho_k) - (1 + \mathbb{E}_\omega \tau_\omega)(\rho - \rho_k)$$

to obtain $C(\rho_{k+1}) \geq 0$, and use the updating rule in (5.14). An additional consequence of $C(\rho_{k+1}) \geq 0$ is that $\{\rho_k\}_{k\geq 0}$ is bounded from above by ρ^*. In fact, Lemma 5.4 establishes that $\rho_k \to \rho^*$. To prove Lemma 5.4, we need the technical assumption that $C(\rho_0) > 0$; recall that if $C(\rho_0) \leq 0$, then we are done, since then $\rho^* = \rho_0$.

Lemma 5.4. *Let $\bar{x} \in X$ be given and let $\phi : \bar{X} \mapsto \mathbb{R}$ be a rational convex polyhedral function. Let $\rho_0 = \phi(\bar{x})$, and assume that $C(\rho_0) > 0$. Recursively define $\rho_{k+1} = \rho_k + C(\rho_k)/(1 + \mathbb{E}_\omega \tau_\omega)$, $k \geq 0$, where τ_ω corresponds to an optimal solution of the problem in* (5.10) *with $\rho = \rho_k$. Then, the resulting sequence $\{\rho_k\}_{k\geq 0}$ is such that $\rho_k \to \rho^*$, $C(\rho_k) \to 0$, and if $C(\rho_k) < \delta$, then $\rho_k \geq \rho^* - \delta$.*

Proof. See Appendix 5.A. ■

Based on Lemma 5.4, we propose to solve (5.9) using a fixed point iteration algorithm, in which we iteratively construct the sequence $\{\rho_k\}_{k\geq 0}$, and we stop if $C(\rho_k) < \delta$. Lemma 5.4 ensures that this algorithm is finitely convergent, and that on termination, $\rho_k \geq \rho^* - \delta$. Moreover, $C(\rho)$ can be computed efficiently using the expression $C(\rho) = \mathbb{E}_\omega[C_\omega(\rho)]$, where

$$C_\omega(\rho) := \sup_{\alpha,\beta,\tau} \{\alpha - \beta^\top \bar{x} - \rho(1+\tau) : \ (\alpha, \beta, \tau) \in \Pi_\omega(\phi)\}, \ \omega \in \Omega. \tag{5.15}$$

Thus, we can efficiently parallelize our fixed point iteration algorithm by computing the quantities $C_\omega(\rho)$, $\omega \in \Omega$, in parallel. In Sections 5.4.1.1 and 5.4.1.2, we describe a primal and dual algorithm to solve (5.15), respectively. To this end, we first show that $\Pi_\omega(\phi)$ is polyhedral, see Lemma 5.5, and thus (5.15) is equivalent to a linear programming problem.

Lemma 5.5. *Let $\phi : \bar{X} \mapsto \mathbb{R}$ be a rational convex polyhedral function and consider $\Pi_\omega(\phi) = \{(\alpha, \beta, \tau) : \ v_\omega(x) \geq \alpha - \beta^\top x - \tau\phi(x) \ \forall x \in X, \ \tau \geq 0\}$. Define*

$$S^\phi_\omega := \{(x, \theta, y) \in X \times \mathbb{R} \times \mathcal{Y} : \ \theta \geq \phi(x), \ W_\omega y = h_\omega - T_\omega x\}.$$

Then, $\Pi_\omega(\phi)$ is a rational polyhedron, and

$$\Pi_\omega(\phi) = \{(\alpha, \beta, \tau) : \ q_\omega^\top y^i + \beta^\top x^i + \tau\theta^i \geq \alpha \ \forall i \in \{1, \dots, d\}, \ \tau \geq 0\}, \tag{5.16}$$

where $(x^i, \theta^i, y^i) \in S_\omega^\phi$, $i = 1, \dots, d$, denote the extreme points of $\operatorname{conv}(S_\omega^\phi)$.

Proof. Recall that $\Pi_\omega(\phi)$ is the set of cut coefficients (α, β, τ) with $\tau \geq 0$ which define non-linear optimality cuts for the second-stage cost functions v_ω of the form $v_\omega(x) \geq \alpha - \beta^\top x - \tau\phi(x) \ \forall x \in X$. Since $\tau \geq 0$, the latter condition is equivalent to

$$v_\omega(x) \geq \alpha - \beta^\top x - \tau\theta \ \forall (x, \theta) \in X \times \mathbb{R} \text{ such that } \theta \geq \phi(x) \tag{5.17}$$

By substituting the definition of v_ω into (5.17), we obtain that $(\alpha, \beta, \tau) \in \Pi_\omega(\phi)$ if and only if $q_\omega^\top y + \beta^\top x + \tau\theta \geq \alpha$ for every $(x, \theta, y) \in S_\omega^\phi$. Because the latter inequality is also valid for $\operatorname{conv}(S_\omega^\phi)$, we obtain that

$$\Pi_\omega(\phi) = \{(\alpha, \beta, \tau) : \ q_\omega^\top y + \beta^\top x + \tau\theta \geq \alpha \ \forall (x, \theta, y) \in \operatorname{conv}(S_\omega^\phi)\}.$$

To obtain (5.16), observe that $\operatorname{conv}(S_\omega^\phi)$ is a rational polyhedron [28, Theorem 1] with one extreme direction, namely $(0, 1, 0)$, and finitely many extreme points. ■

The expression in (5.16) reveals that

$$C_\omega(\rho) = \sup_{\alpha,\beta,\tau} \{\alpha - \beta^\top \bar{x} - \rho(1 + \tau) : \\ q_\omega^\top y^i + \beta^\top x^i + \tau\theta^i \geq \alpha \ \forall i \in \{1, \dots, d\}, \ \tau \geq 0\}, \tag{5.18}$$

i.e., we can compute $C_\omega(\rho)$ by solving a linear programming problem if all extreme points of $\operatorname{conv}(S_\omega^\phi)$ are known. In Section 5.4.1.1, we describe a row generation scheme for solving (5.18) by enumerating a sufficiently rich subset of the extreme points of $\operatorname{conv}(S_\omega^\phi)$, and in Section 5.4.1.2, we solve the dual problem of (5.18) using cutting plane techniques.

5.4.1.1 A row generation scheme

In general, the number of extreme points of $\operatorname{conv}(S_\omega^\phi)$ may be very large, and in those cases directly solving the LP in (5.18) is computationally infeasible. Therefore, we propose a row generation scheme similar to approaches in robust optimization

and disjunctive programming, see, e.g., [33, 52, 87]. In this approach, we iteratively identify extreme points $(x^i, \theta^i, y^i) \in S^\phi_\omega$, $i = 1, \ldots, t$, and we solve the resulting *cut-generation master problem*

$$\max_{\alpha,\beta,\tau}\{\alpha - \beta^\top \bar{x} - \rho(1+\tau) : \\ q_\omega^\top y^i + \beta^\top x^i + \tau\theta^i \geq \alpha \ \ \forall i \in \{1, \ldots, t\},\ \tau \geq 0\}. \quad \text{(CGMP)}$$

We denote the optimal solution of (CGMP) by $(\alpha^t, \beta^t, \tau^t)$, and we attempt to identify a point $(x^{t+1}, \theta^{t+1}, y^{t+1}) \in S^\phi_\omega$ which violates the inequality $q_\omega^\top y + {\beta^t}^\top x + \tau^t\theta \geq \alpha^t$ by solving the *cut generation subproblem*

$$\nu^t := \min_{x,\theta,y}\{q_\omega^\top y + {\beta^t}^\top x + \tau^t\theta - \alpha^t : \ (x, \theta, y) \in S^\phi_\omega\}, \quad \text{(CGSP)}$$

which is a small-scale MIP. Note that $(\alpha^t, \beta^t, \tau^t)$ is feasible and thus optimal in (5.18) if and only if $\nu^t \geq 0$. If $\nu^t < 0$, then we consider an optimal solution $(x^{t+1}, \theta^{t+1}, y^{t+1})$ of (CGSP) and use it to strengthen (CGMP), i.e., we add the constraint $q_\omega^\top y^{t+1} + \beta^\top x^{t+1} + \tau\theta^{t+1} \geq \alpha$ to (CGMP) and resolve (CGMP). Since $\text{conv}(S^\phi_\omega)$ has finitely many extreme points, finite termination of the row generation scheme is guaranteed if (CGSP) returns an optimal solution $(x^{t+1}, \theta^{t+1}, y^{t+1})$ which is an extreme point of $\text{conv}(S^\phi_\omega)$. Indeed, since the objective function of (CGSP) is linear, it has an optimal solution which is an extreme point of $\text{conv}(S^\phi_\omega)$. Finally, in order to ensure that (CGMP) is bounded, we choose $(x^1, \theta^1, y^1) = (\bar{x}, \rho, \bar{y})$, for some arbitrary $\bar{y} \in \{\mathcal{Y} : \ W_\omega y = h_\omega - T_\omega \bar{x}\}$.

Note that in each iteration of the row generation scheme, we have to solve the small-scale LP and MIP in (CGMP) and (CGSP), respectively, and thus the computation time strongly depends on the number of iterations required for convergence. In the worst case, the number of iterations equals the number of extreme points of $\text{conv}(S^\phi_\omega)$, which can be exponentially large. In general, however, we do not expect that an exhaustive exhaustive enumeration of all extreme points is required, and in fact, our numerical experiments indicate that, typically, only a small fraction of the total number of extreme points needs to be computed before the algorithm terminates. Furthermore, in our fixed point iteration algorithm, we have to obtain $C_\omega(\rho)$ multiple times for different values of ρ, and thus we have to run the row generation scheme repeatedly. This can be done efficiently by implementing a warm start for the row generation scheme, in which we reuse the points (x^i, θ^i, y^i) identified during one run in subsequent runs. This is possible since the feasible region S^ϕ_ω

of (CGSP) does not depend on ρ.

5.4.1.2 Convexification via cutting plane techniques

The second approach we consider for solving the problem in (5.18) is to use cutting plane techniques to solve its dual LP, which we derive in Lemma 5.6 below. The advantage of this approach over the row generation scheme in Section 5.4.1.1 is that it only requires solving small-scale LPs. In general, however, the number of LPs that need to be solved may be exponentially large, similar to cutting plane algorithms for deterministic MIPs.

Lemma 5.6. *Let $\phi : \bar{X} \mapsto \mathbb{R}$ be a rational convex polyhedral function, let $\bar{x} \in \bar{X}$ be given, and consider the value function $C_\omega(\rho)$ defined in* (5.18)*. Then,*

$$C_\omega(\rho) = -\rho + \min_y \{q_\omega^\top y : (\bar{x}, \rho, y) \in \text{conv}(S_\omega^\phi)\} \quad \forall \rho \in \text{dom}(C_\omega). \tag{5.19}$$

Proof. We will show that the dual of (5.18) is given by the expression in (5.19), so that the result follows from strong LP duality. In particular, for arbitrary $\rho \in \text{dom}(C_\omega)$, the dual of (5.18) is given by

$$C_\omega(\rho) = -\rho + \min_{\lambda^i \geq 0} \left\{ \sum_{i=1}^d \lambda^i q_\omega^\top y^i : \sum_{i=1}^d \lambda^i = 1, \sum_{i=1}^d \lambda^i x^i = \bar{x}, \sum_{i=1}^d \lambda^i \theta^i \leq \rho \right\}.$$

Since (x^i, θ^i, y^i), $i = 1, \ldots, d$, are the extreme points of $\text{conv}(S_\omega^\phi)$, the above is equivalent to

$$C_\omega(\rho) = -\rho + \min_{\theta, y} \left\{ q_\omega^\top y : \ (\bar{x}, \theta, y) \in \text{conv}(S_\omega^\phi), \ \theta \leq \rho \right\}, \tag{5.20}$$

and (5.19) follows by noting that it is optimal to select $\theta = \rho$ in (5.20). ■

The problem in (5.19) cannot be solved directly by using an off-the-shelf MIP solver, since the constraint $(\bar{x}, \rho, y) \in \text{conv}(S_\omega^\phi)$ depends *parametrically* on $\bar{x}$ and ρ. That is why we solve the problem in (5.19) by using *parametric* cutting planes of the form $\hat{W}_\omega y \geq \hat{h}_\omega - \hat{T}_\omega x - r_\omega \theta$ to recover $\text{conv}(S_\omega^\phi)$, i.e.,

$$\begin{aligned} \text{conv}(S_\omega^\phi) \subseteq \hat{S}_\omega^\phi := \{(x, \theta, y) : \ & W_\omega y = h_\omega - T_\omega x, \\ & \hat{W}_\omega y \geq \hat{h}_\omega - \hat{T}_\omega x - r_\omega \theta\}. \end{aligned} \tag{5.21}$$

In particular, we use these cutting planes to obtain the following relaxation of (5.19),

$$\begin{aligned}\hat{C}_\omega(\rho) &= -\rho + \min_y\{q_\omega^\top y :\ (\bar{x},\rho,y) \in \hat{S}_\omega^\phi\}\\ &= -\rho + \min_y\{q_\omega^\top y : W_\omega y = h_\omega - T_\omega \bar{x},\ \hat{W}_\omega y \geq \hat{h}_\omega - \hat{T}_\omega \bar{x} - r_\omega \rho\}.\end{aligned} \tag{5.22}$$

Initially, the collection of cutting planes $\hat{W}_\omega y \geq \hat{h}_\omega - \hat{T}_\omega x - r_\omega \theta$ is empty, and the relaxation in (5.22) reduces to the LP-relaxation of the second-stage subproblem. If the resulting solution $\bar{y}$ of this relaxation is such that $(\bar{x},\rho,\bar{y}) \in \mathrm{conv}(S_\omega^\phi)$, then we are done: $\bar{y}$ is optimal in (5.19) and $\hat{C}_\omega(\rho) = C_\omega(\rho)$. Otherwise, we derive a parametric cutting plane which separates $(\bar{x},\rho,\bar{y})$ from $\mathrm{conv}(S_\omega^\phi)$, after which we update $\hat{S}_\omega^\phi$ and resolve (5.22). Depending on the family of cutting planes that we use to recover $\mathrm{conv}(S_\omega^\phi)$, this procedure is finitely convergent. In particular, if we use the Fenchel cuts described in [22], then the resulting algorithm is finitely convergent [23, Corollary 3.3]. Before discussing further computational aspects of our cutting plane approach, Lemma 5.7 describes how we can retrieve an optimal solution (α,β,τ) of the *primal* problem in (5.18) if we have solved the *dual* problem in (5.19).

Lemma 5.7. *Let $\phi : \bar{X} \mapsto \mathbb{R}$ be a rational convex polyhedral function, and suppose that the cutting planes $\hat{W}_\omega y \geq \hat{h}_\omega - \hat{T}_\omega x - r_\omega \theta$ satisfy* (5.21)*. Let $\bar{x} \in X$ and $\rho \geq \phi(\bar{x})$ be given, and consider the cutting plane relaxation $\hat{C}_\omega(\rho)$ defined in* (5.22)*, and denote by λ_ω and π_w optimal dual multipliers corresponding to the constraints $W_\omega y = h_\omega - T_w \bar{x}$ and $\hat{W}_\omega y \geq \hat{h}_\omega - \hat{T}_\omega \bar{x} - r_\omega \rho$, respectively. Then,*

$$(\alpha,\beta,\tau) := (\lambda_w^\top h_\omega + \pi_w^\top \hat{h}_\omega, \lambda_\omega^\top T_\omega + \pi_w^\top \hat{T}_\omega, \pi_\omega^\top r_\omega) \tag{5.23}$$

is feasible in (5.18)*, and $\hat{C}_\omega(\rho) = \alpha - \beta^\top \bar{x} - (1+\tau)\rho$.*

Proof. Since λ_ω and π_w are optimal dual multipliers of (5.22), strong LP duality implies that $\hat{C}_\omega(\rho) = -\rho + \lambda_\omega^\top(h_\omega - T_\omega x) + \pi_\omega^\top(\hat{h}_\omega - \hat{T}_\omega x - r_\omega \theta)$ and it follows from the definition of (α,β,τ) that $\hat{C}_\omega(\rho) = \alpha - \beta^\top \bar{x} - (1+\tau)\rho$.

Moreover, we prove that (α,β,τ) is feasible in (5.18) by showing that $q_\omega^\top y + \beta^\top x + \tau\theta \geq \alpha$ for every $(x,\theta,y) \in S_\omega^\phi$. Indeed, for arbitrary $(x,\theta,y) \in S_\omega^\phi$, we have

$$\begin{aligned}\alpha - \beta^\top x - \tau\theta &= \lambda_\omega^\top(h_\omega - T_\omega x) + \pi_\omega^\top(\hat{h}_\omega - \hat{T}_\omega x - r_\omega \theta)\\ &\leq \lambda_\omega^\top W_\omega y + \pi_\omega^\top \hat{W}_\omega y \leq q_\omega^\top y,\end{aligned}$$

where the first inequality is due to $\pi_\omega \geq 0$ and $(x, \theta, y) \in S^\phi_\omega$, so that $W_\omega y = h_\omega - T_\omega x$ and $\hat{W}_\omega y \geq \hat{h}_\omega - \hat{T}_\omega x - r_\omega \theta$, and the latter inequality follows from dual feasibility and $y \geq 0$. ■

As mentioned earlier, it is possible to solve the problem in (5.19) in finitely many iterations using Fenchel cuts. In practice, however, computing these Fenchel cuts takes significant time. That is why it may be advantageous to use other parametric cutting planes that can be computed faster, but do not necessarily converge in a finite number of iterations. To generate such cutting planes, note that if $(\bar{x}, \rho, \bar{y}) \notin \text{conv}(S^\phi_\omega)$, then $(\bar{x}, \rho, \bar{y})$ does not satisfy the integer restrictions in S^ϕ_ω, and thus we can apply ideas from deterministic mixed-integer programming to generate specific types of cutting planes for S^ϕ_ω. See, e.g., [10, 11] for (strengthened) lift-and-project cuts, [9, 88] for Gomory mixed-integer (GMI) cuts, and [53] for multi-term disjunctive cuts.

In our practical implementation of the cutting plane approach, we accommodate the case where the cutting planes do not converge finitely by stopping after a pre-specified number of iterations, or if we are unable to cut away a fractional solution $(\bar{x}, \rho, \bar{y})$. Efficient implementations are possible, since each iteration merely requires solving a small-scale LP, and we may speed up convergence by adding multiple cutting planes to (5.22) in one iteration, e.g., by generating a round of GMI cuts. Moreover, since the cutting planes that we use depend parametrically on x and θ, they can be reused in subsequent iterations of the Benders' decomposition algorithm and the fixed point iteration algorithm.

5.5 Proof of convergence

In this section, we prove Theorem 5.1. That is, we show that for any convex polyhedral function $\phi_0 : \bar{X} \mapsto \mathbb{R}$, the sequence $\{\phi_k\}_{k \geq 0}$ defined recursively as $\phi_{k+1} = \text{SCC}(\phi_k)$, $k \geq 0$, converges uniformly to $\text{co}(\max\{\phi_0, Q\})$. For convenience, we recall that the scaled cut closure $\text{SCC}(\phi)$ is defined as

$$\text{SCC}(\phi)(x) = \sup_{\alpha_\omega, \beta_\omega, \tau_\omega} \left\{ \frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top x}{1 + \mathbb{E}_\omega \tau_\omega} : \right.$$
$$\left. (\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi) \ \forall \omega \in \Omega \right\}, \quad x \in \bar{X},$$

where $\Pi_\omega(\phi) := \{(\alpha, \beta, \tau) : \ v_\omega(x) \geq \alpha - \beta^\top x - \tau\phi(x) \ \forall x \in X, \ \tau \geq 0\}$. We prove Theorem 5.1 by showing, in Section 5.5.1, that ϕ_k converges to a limit function ϕ^* satisfying $\mathrm{SCC}(\phi^*) = \phi^*$, i.e., ϕ^* is a fixed point of the scaled cut closure operation. Next, in Section 5.5.2, we show that such a fixed point must satisfy $\phi^* = \mathrm{co}(\max\{\phi_0, Q\})$, which completes the proof.

In order to obtain these results, we derive an alternative expression for $\mathrm{SCC}(\phi)$, as follows,

$$\begin{aligned}\mathrm{SCC}(\phi)(x) &= \sup_{\tau_\omega \geq 0} \sup_{\alpha_\omega, \beta_\omega} \left\{ \frac{\mathbb{E}_\omega[\alpha_\omega - \beta_\omega^\top x]}{1 + \mathbb{E}_\omega \tau_\omega} : \right. \\ &\qquad \left. v_\omega(x') + \tau_\omega \phi(x') \geq \alpha_\omega - \beta_\omega^\top x' \ \forall x' \in X, \ \omega \in \Omega \right\} \\ &= \sup_{\tau_\omega \geq 0} \left\{ \frac{\mathbb{E}_\omega \overline{\mathrm{co}}(v_\omega + \tau_\omega \phi)(x)}{1 + \mathbb{E}_\omega \tau_\omega} \right\},\end{aligned}$$

where the latter equality follows directly from the definition of the closed convex envelope. We use this expression to define a mapping $\mathbb{T}$ defined on the space of continuous bounded functions, which is such that $\mathbb{T}\phi = \mathrm{SCC}(\phi)$, see Definition 5.3.

Definition 5.3. Consider the space $C(\bar{X})$ of continuous bounded functions mapping from $\bar{X}$ to $\mathbb{R}$, equipped with the metric d, defined as

$$d(f, g) := ||f - g||_\infty = \sup_{x \in \bar{X}} |f(x) - g(x)|, \quad f, g \in C(\bar{X}),$$

and define $\mathbb{T} : C(\bar{X}) \mapsto C(\bar{X})$ as

$$(\mathbb{T}f)(x) = \sup_{\tau_\omega \geq 0} \left\{ \frac{\mathbb{E}_\omega \overline{\mathrm{co}}(v_\omega + \tau_\omega f)(x)}{1 + \mathbb{E}_\omega \tau_\omega} \right\}, \quad x \in \bar{X}, \quad f \in C(\bar{X}). \tag{5.24}$$

In order to see that $\mathbb{T}$ maps into $C(\bar{X})$, i.e., $\mathbb{T}f \in C(\bar{X})$ for every $f \in C(\bar{X})$, note that by (5.24), $\mathbb{T}f$ is the pointwise supremum of convex lsc functions, and thus $\mathbb{T}f$ is convex and lsc. Furthermore, since $\bar{X}$ is a compact polyhedral set, it follows from Theorem 5.2 below that $\mathbb{T}f$ is continuous and bounded, i.e., $\mathbb{T}f \in C(\bar{X})$.

Theorem 5.2. *[55, Theorem 10.2] If $f : D \mapsto \mathbb{R}$ is a convex lsc function defined on a convex polyhedral domain D, then f is continuous on D.*

Since $\mathbb{T}\phi = \mathrm{SCC}(\phi)$, we can also define the sequence $\{\phi_k\}_{k \geq 0}$ in terms of $\mathbb{T}$. That is, for a given $\phi_0 \in C(\bar{X})$ such that ϕ_0 is convex, we define $\phi_{k+1} := \mathbb{T}\phi_k$, $k \geq 0$. Since $\mathbb{T}$ maps into $C(\bar{X})$, it follows that $\phi_{k+1} = \mathbb{T}\phi_k \in C(\bar{X})$ for every $k \geq 0$, and

thus ϕ_k is well-defined for every $k \geq 0$. In addition, ϕ_k is convex for every $k \geq 0$.

5.5.1 Uniform convergence and fixed points

The main result of this section is Proposition 5.2, which states that ϕ_k converges uniformly to a fixed point of $\mathbb{T}$. In order to prove it, we derive several properties of the sequence $\{\phi_k\}_{k\geq 0}$ in Lemma 5.8.

Lemma 5.8. *Let $\phi_0 \in C(\bar{X})$ be a convex function, and consider the sequence $\{\phi_k\}_{k\geq 0} \subseteq C(\bar{X})$ defined by $\phi_{k+1} := \mathbb{T}\phi_k$, $k \geq 0$. Then, ϕ_k is monotone increasing in k, i.e., $\phi_{k+1} \geq \phi_k$ for every $k \geq 0$, and, moreover, $\phi_k \leq \mathrm{co}(\max\{\phi_0, Q\})$ for every $k \geq 0$.*

Proof. We prove monotonicity of ϕ_k by showing that $\mathbb{T}f \geq f$ for every convex $f \in C(\bar{X})$. Indeed, if $f \in C(\bar{X})$ is convex, then

$$\mathbb{T}f \geq \sup_{\tau_\omega \geq 0} \left\{ \frac{\mathbb{E}_\omega[\overline{\mathrm{co}}(v_\omega) + \tau_\omega \overline{\mathrm{co}}(f)]}{1 + \mathbb{E}_\omega \tau_\omega} \right\} \geq \overline{\mathrm{co}}(f) = f,$$

where the second inequality follows by letting $\tau_\omega \to \infty$ for every $\omega \in \Omega$.

Next, we prove by induction that $\phi_k \leq \mathrm{co}(\max\{\phi_0, Q\})$ for every $k \geq 0$. Note that $\phi_0 \leq \mathrm{co}(\max\{\phi_0, Q\})$ follows directly from convexity of ϕ_0. Next, we fix arbitrary $k \geq 0$, and we assume that $\phi_k \leq \mathrm{co}(\max\{\phi_0, Q\})$, so that

$$\phi_k(x) \leq \max\{\phi_0(x), Q(x)\} \ \forall x \in X.$$

Then, for every $x \in X$,

$$\begin{aligned}
\phi_{k+1}(x) &= (\mathbb{T}\phi_k)(x) \\
&\leq \sup_{\tau_\omega \geq 0} \left\{ \frac{\mathbb{E}_\omega[v_\omega(x) + \tau_\omega \phi_k(x)]}{1 + \mathbb{E}_\omega \tau_\omega} \right\} \\
&\leq \sup_{\tau_\omega \geq 0} \left\{ \frac{Q(x) + \mathbb{E}_\omega \tau_\omega \phi_k(x)}{1 + \mathbb{E}_\omega \tau_\omega} \right\}, \\
&\leq \sup_{\tau_\omega \geq 0} \left\{ \frac{\max\{\phi_0(x), Q(x)\} + \mathbb{E}_\omega \tau_\omega \max\{\phi_0(x), Q(x)\}}{1 + \mathbb{E}_\omega \tau_\omega} \right\} \\
&= \max\{\phi_0(x), Q(x)\}.
\end{aligned}$$

Hence, $\phi_{k+1} \leq \mathrm{co}(\max\{\phi_0, Q\})$, since ϕ_{k+1} is a convex function majorized by $\max\{\phi_0, Q\}$. ∎

Since the sequence $\{\phi_k\}_{k\geq 0}$ is monotone increasing and bounded, ϕ_k converges

pointwise to some limit function. Indeed, for every $x \in \bar{X}$, the real-valued sequence $\{\phi_k(x)\}_{k\geq 0}$ is monotone increasing and bounded, and thus convergent. Therefore, we may define ϕ^* as the pointwise limit of ϕ_k, i.e., $\phi^*(x) := \lim_{k\to\infty} \phi_k(x),\ x \in \bar{X}$. We, however, need a stronger type of convergence than pointwise convergence for the proof of Theorem 5.1, namely uniform convergence: ϕ_k converges uniformly to ϕ^* if for every $\varepsilon > 0$, there exists a $K \geq 0$ such that $||\phi_k - \phi^*||_\infty \leq \varepsilon\ \forall k \geq K$. In Proposition 5.2, we obtain that ϕ_k converges *uniformly* to ϕ^* by showing that the pointwise limit ϕ^* is continuous. In addition, it then follows from continuity of $\mathbb{T}$, see Lemma 5.9 below, that ϕ^* is a fixed point of $\mathbb{T}$, i.e., $\mathbb{T}\phi^* = \phi^*$.

Lemma 5.9. *The mapping $\mathbb{T} : C(\bar{X}) \mapsto C(\bar{X})$ of Definition 5.3 is continuous on $C(\bar{X})$.*

Proof. See Appendix 5.A. ■

Proposition 5.2. *Let $\phi_0 \in C(\bar{X})$ be a convex function. Then, the sequence $\{\phi_k\}_{k\geq 0}$ defined by $\phi_{k+1} = \mathbb{T}\phi_k$, $k \geq 0$, converges uniformly to its pointwise limit ϕ^*. Moreover, ϕ^* is convex and continuous, and ϕ^* is a fixed point of $\mathbb{T}$, i.e., $\mathbb{T}\phi^* = \phi^*$.*

Proof. Dini's theorem [63, Theorem 7.13] states that if a monotone increasing sequence of continuous functions converges pointwise to a continuous function, then the convergence is uniform. Therefore, it suffices to show that ϕ^* is continuous in order to establish that ϕ_k converges uniformly to ϕ^*. We prove that ϕ^* is continuous by noting that monotonicity of ϕ_k, see Lemma 5.8, implies that $\phi^*(x) = \sup_{k\geq 0} \phi_k(x)$, i.e., ϕ^* is the pointwise supremum of convex continuous functions. It follows that ϕ^* is convex and lsc, and thus, using Theorem 5.2, ϕ^* is continuous. In order to see that ϕ^* is a fixed point of $\mathbb{T}$, note that

$$\mathbb{T}\phi^* = \mathbb{T} \lim_{k\to\infty} \phi_k = \lim_{k\to\infty} \mathbb{T}\phi_k = \lim_{k\to\infty} \phi_{k+1} = \phi^*,$$

where the second equality follows from the continuity of $\mathbb{T}$ in Lemma 5.9. ■

5.5.2 Properties of fixed points of $\mathbb{T}$

By Proposition 5.2, ϕ_k converges uniformly to a fixed point of $\mathbb{T}$. Thus, we may derive properties of ϕ^* by deriving properties of fixed points of $\mathbb{T}$. In particular, in Proposition 5.3, we show that any convex fixed point f of $\mathbb{T}$ is such that $f \geq \overline{\text{co}}(Q)$. In order to prove Proposition 5.3, we need the following result.

Lemma 5.10. *Assume that $f \in C(\bar{X})$ is convex. If $(\bar{x}, \bar{\theta}) = (\bar{x}, f(\bar{x}))$ is an extreme point of $\text{epi}(f) = \{(x, \theta) \in \bar{X} \times \mathbb{R} :\ \theta \geq f(x)\}$, then $\sup_{\tau_\omega \geq 0}\{\overline{\text{co}}(v_\omega + \tau_\omega f)(\bar{x}) - \tau_\omega f(\bar{x})\} \geq v_\omega(\bar{x})$ for every $\omega \in \Omega$.*

Proof. See Appendix 5.A. ■

Intuitively, Lemma 5.10 says that if $\bar{x}$ corresponds to an extreme point of $\text{epi}(f)$, then the gap between $v_\omega(x) + \tau_\omega f(x)$ and $\overline{\text{co}}(v_\omega + \tau_\omega f)(\bar{x})$ can be made arbitrarily small by choosing appropriate $\tau_\omega \geq 0$. To see how we can use this result to derive properties of fixed points of $\mathbb{T}$, assume for the purpose of exposition that there exists a $\tau_{\boldsymbol{\omega}}$ such that $\overline{\text{co}}(v_\omega + \tau_\omega f)(\bar{x}) = v_\omega(\bar{x}) + \tau_\omega f(\bar{x})$ for every $\omega \in \Omega$. Then,

$$(\mathbb{T}f)(\bar{x}) = \frac{\mathbb{E}_{\boldsymbol{\omega}}[v_{\boldsymbol{\omega}}(\bar{x}) + \tau_{\boldsymbol{\omega}} f(\bar{x})]}{1 + \mathbb{E}_{\boldsymbol{\omega}} \tau_{\boldsymbol{\omega}}} = \frac{Q(\bar{x}) + \mathbb{E}_{\boldsymbol{\omega}} \tau_{\boldsymbol{\omega}} f(\bar{x})}{1 + \mathbb{E}_{\boldsymbol{\omega}} \tau_{\boldsymbol{\omega}}},$$

which reveals that, unless $f(\bar{x}) \geq Q(\bar{x})$, we have $\mathbb{T}f(\bar{x}) > f(\bar{x})$, i.e., f is not a fixed point of $\mathbb{T}$. We prove Proposition 5.3 by formalizing this reasoning.

Proposition 5.3. *Let $\phi_0 \in C(\bar{X})$ be given. Assume that $f \in C(\bar{X})$ is convex and $f \geq \phi_0$. If f is a fixed point of $\mathbb{T}$, i.e. if $\mathbb{T}f = f$, then $f \geq \overline{\text{co}}(\max\{\phi_0, Q\})$.*

Proof. We will show that for every extreme point $(\bar{x}, f(\bar{x}))$ of $\text{epi}(f)$, we have $\bar{\theta} = f(\bar{x}) \geq \overline{\text{co}}(\max\{\phi_0, Q\})(\bar{x})$. This suffices to prove $f(x) \geq \overline{\text{co}}(Q)(x)\ \forall x \in \bar{X}$, since Carathéodory's theorem [56, Theorem 2.29] implies that, for arbitrary $x \in \bar{X}$, the point $(x, f(x)) \in \text{epi}(f)$ can be written as a convex combination of $n_1 + 2$ extreme points of $\text{epi}(f)$, i.e.,

$$(x, f(x)) = \sum_{i=1}^{n_1+2} \lambda^i (x^i, f(x^i)),$$

where $\sum_{i=1}^{n_1+2} \lambda^i = 1$, $\lambda^i \geq 0$, and $(x^i, f(x^i))$ is an extreme point of $\text{epi}(f)$, $i = 1, \ldots, n_1 + 2$, and thus

$$f(x) = \sum_{i=1}^{n_1+2} \lambda^i f(x^i) \geq \sum_{i=1}^{n_1+2} \lambda^i \overline{\text{co}}(\max\{\phi_0, Q\})(x^i) \geq \overline{\text{co}}(\max\{\phi_0, Q\})(x),$$

where we used convexity of $\overline{\text{co}}(\max\{\phi_0, Q\})$ to obtain the latter inequality.

We show that $f(\bar{x}) \geq \overline{\text{co}}(\max\{\phi_0, Q\})(\bar{x})$ if $(\bar{x}, \bar{\theta})$ is an extreme point of $\text{epi}(f)$ by proving that (i) $\bar{x} \in X$, and (ii) $f(\bar{x}) \geq \max\{\phi_0(\bar{x}), Q(\bar{x})\}$ if $\bar{x} \in X$. We prove these claims by contradiction. First, suppose that $\bar{x} \notin X$. Then, $v_\omega(\bar{x}) = \infty$ for every $\omega \in \Omega$, and thus, by Lemma 5.10,

$$\sup_{\tau_\omega \geq 0} \{\overline{\text{co}}(v_\omega + \tau_\omega f)(\bar{x}) - \tau_\omega f(x)\} = \infty \quad \forall \omega \in \Omega.$$

It follows that for every $\omega \in \Omega$, there exists a $\tau_\omega \geq 0$ such that $\overline{\text{co}}(v_\omega + \tau_\omega f)(\bar{x}) - \tau_\omega f(\bar{x}) > f(\bar{x})$. But then, for this choice of τ_ω,

$$(\mathbb{T}f)(\bar{x}) \geq \frac{\mathbb{E}_\omega \overline{\text{co}}(v_\omega + \tau_\omega f)(\bar{x})}{1 + \mathbb{E}_\omega \tau_\omega} > \frac{\mathbb{E}_\omega [f(\bar{x}) + \tau_\omega f(\bar{x})]}{1 + \mathbb{E}_\omega \tau_\omega} = f(\bar{x}),$$

which is a contradiction, since $\mathbb{T}f = f$.

Next, suppose that $\bar{x} \in X$, but $f(\bar{x}) < \max\{\phi_0(\bar{x}), Q(\bar{x})\}$. Since, by assumption, $f(x) \geq \phi_0(x)$, it must be that $f(\bar{x}) < Q(\bar{x})$. Let $\delta = Q(\bar{x}) - f(\bar{x}) > 0$, and note that Lemma 5.10 implies that for every $\omega \in \Omega$, there exists a $\tau_\omega \geq 0$ such that

$$\overline{\text{co}}(v_\omega + \tau_\omega f)(\bar{x}) - \tau_\omega f(\bar{x}) \geq v_\omega(\bar{x}) - \delta/2.$$

But then,

$$\begin{aligned}(\mathbb{T}f)(\bar{x}) &\geq \frac{\mathbb{E}_\omega \overline{\text{co}}(v_\omega + \tau_\omega f)(\bar{x})}{1 + \mathbb{E}_\omega \tau_\omega} \\ &\geq \frac{\mathbb{E}_\omega [v_\omega(\bar{x}) + \tau_\omega f(\bar{x}) - \delta/2]}{1 + \mathbb{E}_\omega \tau_\omega} \\ &= \frac{f(\bar{x}) + \delta/2 + \mathbb{E}_\omega [\tau_\omega f(\bar{x})]}{1 + \mathbb{E}_\omega \tau_\omega} > f(\bar{x}),\end{aligned}$$

which contradicts $\mathbb{T}f = f$. ■

We are now ready to prove Theorem 5.1.

Proof of Theorem 5.1. It suffices to prove that for any convex $\phi_0 \in C(\bar{X})$ the sequence $\{\phi_k\}_{k\geq 0}$ defined by $\phi_{k+1} = \mathbb{T}\phi_k$, $k \geq 0$, converges uniformly to $\text{co}(\max\{\phi_0, Q\})$. Proposition 5.2 implies that $\phi^* = \lim_{k\to\infty} \phi_k$ exists, and ϕ^* is a fixed point of $\mathbb{T}$. Moreover, using Lemma 5.8, we have that $\phi^* \leq \text{co}(\max\{\phi_0, Q\})$, and monotonicity of ϕ_k implies that $\phi^* \geq \phi_0$. Thus, by Proposition 5.3, we have $\phi^* \geq \overline{\text{co}}(\max\{\phi_0, Q\})$. Finally, since $\max\{\phi_0, Q\}$ is an lsc function defined on a compact domain, we have $\overline{\text{co}}(\max\{\phi_0, Q\}) = \text{co}(\max\{\phi_0, Q\})$ [30, Theorem 2.2], and the result follows. ■

5.6 Numerical experiments

Theorem 5.1 states that our scaled cuts can be used to recover the convex envelope of the expected second-stage cost function by recursively computing the scaled cut closure, and thus they can be used to solve general MIR models. Of course, in practice, we do not compute the full scaled cut closure, but we strengthen the

outer approximation using a single (dominating) scaled cut in every iteration of our Benders' decomposition, in line with Algorithm 2. Therefore, we assess the performance of scaled cuts on a range of problem instances, namely (variants of) an investment problem by Schultz et al. [67] in Sections 5.6.3.2 and 5.6.3.3, as well as the DCAP problem instances by Ahmed and Garcia [2] from SIPLIB [5] in Section 5.6.3.4. In addition, in Section 5.6.3.1, we consider a problem instance by Carøe and Schultz [24], to which we refer as the CS instance, which is known to have a relatively large duality gap. Before we discuss our results, we first describe the setup of our numerical experiments in Section 5.6.1, and in Section 5.6.2, we describe a cut-enhancement technique which we use to speed up the convergence of scaled cuts.

5.6.1 Setup of numerical experiments

In our numerical experiments, we compare our Benders' decomposition algorithm with scaled cuts to traditional solution approaches in terms of bounds on the optimal value, solution quality, and running time. In particular, we use the Lagrangian (L) cuts of Zou et al. [89] as a benchmark, since the best-case performance of the cutting plane algorithms in, e.g., [31] and [53] is the same as the performance of L cuts, and similarly, the Lagrangian dual bound obtained via dual decomposition [24] is at most the lower bound obtained by L cuts. We compute L cuts using the row-generation scheme described in Section 5.4.1.1, with the additional restriction that $\tau = 0$ in (CGMP). In addition, we benchmark our scaled cuts against the so-called strengthened Benders' (SB) cuts [89], which can be computed faster than L cuts, but generally yield weaker lower bounds. Finally, we compare the different strategies for computing scaled cuts described in Section 5.4, i.e., we consider scaled cuts obtained using row generation (S-RG cuts) and cutting plane techniques (S-CP cuts). For the S-CP cuts, we solve the second-stage subproblems using both GMI cutting planes as well as L&P cutting planes.

We compare the different types of optimality cuts in terms of the relative optimality gap

$$\frac{UB - LB}{|LB|} \times 100\%, \tag{5.25}$$

where LB and UB denote the best known lower and upper bound on termination of the Benders' decomposition in Algorithm 2. In our base implementation of Algorithm 2, we use a pure cutting plane approach to solve the master problem (MP).

That is, we maintain a single relaxation of (MP), which we update from one iteration of our Benders' decomposition to the next by adding a single dominating scaled cut, and we use Fenchel cuts [22] to cut away non-integer first-stage solutions. In this case, we refer to the optimality gap in (5.25) as the *root node gap*. Since scaled cuts asymptotically yield $\mathrm{co}(Q)$, we expect that the root node gap converges to zero, i.e., that we obtain the optimal solution without branching on the first-stage decision variables.

In our experiments, we use parallelized implementations of all optimality cut computation routines, by solving the subproblems for each scenario in parallel. All our experiments are run on a machine with two Intel Xeon E5 2680v3 CPUs (24 cores @2.5GHz) and 128GB RAM using Gurobi 9.1.1; computation time is limited to three hours. Furthermore, the tolerance levels ε and δ in the Benders' decomposition and the fixed point iteration algorithm are set to 10^{-4}, unless mentioned otherwise. Finally, in order to prevent numerical instability, we stop Algorithm 2 if the outer approximation improves by less than ε; we terminate our cutting plane approach for computing $C_\omega(\rho)$ if more than 25 iterations are required for convergence; and in our row generation scheme for computing S-RG and L cuts, we restrict the absolute value of the cut coefficients (α, β, τ) in (CGMP) to be at most 10^8.

5.6.2 Cut-enhancement technique

The main idea of our cut-enhancement technique is to derive cuts which are only valid on a subset X' of the first-stage feasible region X. That is, we consider optimality cuts of the form

$$Q(x) \geq \alpha - \beta^\top x \quad \forall x \in X' \subseteq X.$$

Clearly, these optimality cuts are in general at least as strong as cuts which are valid for every $x \in X$. However, the resulting algorithm is only correct if the optimal solution x^* of the MIR model in (5.1) is contained in X'. Thus, in the definition of X', we may exclude feasible solutions which cannot be optimal. In particular, in our Benders' decomposition for MIR models, see Algorithm 2, we take

$$X' = \{x \in X : c^\top x + \hat{Q}_{\mathrm{out}}(x) \leq UB\},$$

where $\hat{Q}_{\text{out}}$ is the current outer approximation of Q, and UB is the best known upper bound on the optimal value η^* of the MIR model in (5.1). Indeed, note that if $c^\top x + \hat{Q}_{\text{out}}(x) > UB$, then x is not optimal in (5.1), since otherwise $Q(x) \geq \hat{Q}_{\text{out}}(x)$ and thus $c^\top x + Q(x) > UB$. In an alternative implementation, we may use a heuristic approach to obtain a candidate solution and a corresponding upper bound on η^*. Finally, note that the constraint $c^\top x + \hat{Q}_{\text{out}}(x) \leq UB$ is polyhedral if $\hat{Q}_{\text{out}}$ is a convex polyhedral function, which ensures that our enhancement technique is computationally feasible.

In our experiments, we use enhanced scaled cuts, referred to as S-RG* cuts to speed up convergence of our Benders' decomposition. In addition, we assess the effect of computing enhanced L and SB cuts, referred to as SB* and L* cuts, respectively, by comparing them to their unenhanced counterparts in terms of the resulting bounds on η^*. We refer to Table 5.1 for an overview of cut type abbreviations.

Table 5.1. Optimailty cut type abbreviations.

Abbreviation[a]	Optimality cut type
L	Lagrangian cut [89]
SB	Strengthened Benders' cut [89]
S-RG	Dominating scaled cut (row generation)
S-CP (L&P)	Dominating scaled cut (lift-and-project cutting planes)
S-CP (GMI)	Domination scaled cut (Gomory mixed-integer cuts)

[a]Starred abbreviations of cut types, e.g., L* cuts, refer to their enhanced counterparts.

5.6.3 Results

5.6.3.1 The CS instance

Carøe and Schultz [24] describe a set of MIR problem instances for which the duality gap is at least 1/16. These instances are defined as

$$\eta^* = \min_{0 \leq x \leq 1} \left\{ 3x + \mathbb{E}_\omega \left[\min_{y \in \{0,1\}} \{-2y : \ -1/2y \geq h_\omega - x\} \right] \right\},$$

where h_ω follows a discrete symmetric uniform distribution with r realizations for some even r; the realizations of h_ω are given by $h^s_\omega = \varepsilon^s$ and $h^{s+r/2}_\omega = 1/4 - \varepsilon^s$, where $\varepsilon^s \in (0, 1/32)$, $s = 1, \ldots, r/2$, are all distinct. We choose $r = 100$, and $\varepsilon^s = \Delta s$, $s = 1, \ldots, r/2$, where $\Delta = \frac{1/32}{1+r/2}$.

Since the input size of the CS instance is relatively small, we do not use a paral-

lelized implementation to compute these cuts in order to avoid overhead, we use a tolerance level $\varepsilon = 10^{-6}$, and we do not use a warm start with Benders' cuts as described in Section 5.6.1, to ensure that the difference in outcomes can be attributed completely to the different cut types. In our experiments, we compare the different types of optimality cuts mentioned in Section 5.6.1. For comparison, we do not only compute the S-CP (L&P) cuts, but we also compute the traditional counterpart of these cuts, obtained by solving the second-stage problem using L&P cuts, as described in [72]. In Table 5.2, we report the resulting lower and upper bounds on η^*, and the computation time and number of optimality cuts required for convergence of our Benders' decomposition.

Table 5.2. CS instance.

Cut type	Lower bound (gap to $\eta^* = 0.2482$)	Upper bound (gap to $\eta^* = 0.2482$)	Computation time	#cuts (average cut generation time)
Traditional cuts				
Benders	-0.0080 (103.21%)	0.7482 (201.48%)	0.166s	9 (0.003s)
SB	-0.0080 (103.21%)	0.7482 (201.48%)	0.344s	9 (0.020s)
Lift-and-project	0.0083 (96.64%)	0.7482 (201.48%)	0.254s	9 (0.011s)
L	0.0083 (96.64%)	0.7482 (201.48%)	0.372s	8 (0.027s)
Scaled cuts				
S-CP* (GMI)	0.2482 (0.00%)	0.2488 (0.24%)	0.830s	33 (0.015s)
S-CP* (L&P)	0.2482 (0.00%)	0.2484 (0.09%)	0.771s	17 (0.031s)
S-RG*	0.2482 (0.00%)	0.2482 (0.00%)	3.331s	17 (0.182s)

What is immediately striking from Table 5.2 is that the scaled cuts are able to completely close the duality gap of traditional cuts, which is relatively large for the CS instance. In particular, the LB gap of all types of scaled cuts is zero, whereas traditional cuts have LB gaps of around 100%. In other words, the quality of the lower bound obtained using traditional cuts is very poor, and can be significantly improved using scaled cuts. Similarly, the UB gap, which measures the quality of the incumbent solution, is over 200% if we use traditional cuts, and can be reduced to zero using S-RG* cuts. We are not able to find the optimal solution using the S-CP* cuts, but the resulting gaps are very small (less than 0.25%) compared to traditional cuts.

In terms of computation time, we observe that generating scaled cuts generally requires more time compared to their traditional counterparts. For example, the average computation time per cut of S-CP* (L&P) cuts compared to traditional L&P

cuts has roughly tripled, and S-RG* cuts take over six times as long to compute as L cuts. Finally, we observe that the row generation scheme for computing scaled cuts requires significantly more time than the S-CP* cuts. However, we recall that in general, stronger performance guarantees are available for the S-RG* cuts, since the row generation scheme computes $C_\omega(\rho)$ exactly whereas the cutting plane techniques may only yield a lower bound. This is reflected by the non-zero UB gap of the S-CP* cuts.

5.6.3.2 Small investment planning problems

Schultz et al. [67] consider the following investment planning problem

$$\min_{x\in\mathcal{X}} \left\{ -3/2x_1 - 4x_2 + \mathbb{E}_\omega[v_\omega(x)] : \quad x \in [0,5]^2 \right\},$$

where $\mathcal{X} = \mathbb{R}^2$, and

$$v_\omega(x) = \min_{y\in\mathcal{Y}}\{-16y_1-19y_2-23y_3-28y_4 : 2y_1+3y_2+4y_3+5y_4 \le h^1_\omega - x_1 \\ 6y_1+ \ y_2+3y_3+2y_4 \le h^2_\omega - x_2\},$$

where $\mathcal{Y} = \{0,1\}^4$; the random variables h^1_ω and h^2_ω follow independent discrete uniform distributions on $\{5, 5.5, \ldots, 15\}$. This problem, and variants thereof are frequently used as benchmark instances in the literature, see, e.g., [4, 31, 49, 53]. The variants we consider are obtained by setting $\mathcal{X} = \mathbb{Z}^2_+$, as well as $\mathcal{Y} = Z^4_+$. In another variant, the technology matrix is given by

$$T_\omega = H := \begin{pmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix},$$

whereas in the original problem, $T_\omega = I_2$. Finally, we vary the distribution of h_ω by letting h^1_ω and h^2_ω follow independent discrete uniform distributions on S equidistant lattice points of the interval $[5, 15]$, so that $|\Omega| = S^2$. Note that in the original problem, $S = 21$, we additionally consider $S = 11$ and $S = 101$. For the resulting 24 instances, we compare the SB and L cuts to their enhanced counterparts and to the S-CP* (L&P), S-RG, and S-RG* cuts. We report the results for the instances with $\mathcal{X} = \mathbb{Z}^2_+$ and $\mathcal{X} = \mathbb{R}^2_+$ in Tables 5.3 and 5.4, respectively. We do not report results for S-CP* (GMI) cuts, since similar as for the CS instance, they perform worse than the S-CP* (L&P) cuts.

Table 5.3. Investment planning problems ($\mathcal{X} = \mathbb{Z}^2_+$): root node gaps.

Instance			Root node gap (computation time[a])						
$\mathcal{Y}$	T_ω	$\lvert\Omega\rvert$	SB	SB*	L	L*	S-CP* (L&P)	S-RG	S-RG*
$\mathbb{Z}^4_+$	I_2	121	3.87% (0s)	1.63% (0s)	1.00% (0s)	0.42% (0s)	0.00% (1s)	0.00% (2s)	0.00% (1s)
		441	4.08% (1s)	1.27% (2s)	1.03% (1s)	0.00% (1s)	0.01% (146s)	0.00% (10s)	0.00% (2s)
		10201	4.23% (18s)	1.28% (34s)	1.05% (27s)	0.59% (33s)	1.13% (3h)	0.00% (172s)	0.00% (111s)
$\mathbb{Z}^4_+$	H	121	9.54% (0s)	8.90% (0s)	0.84% (1s)	0.00% (1s)	0.00% (17s)	0.00% (2s)	0.00% (1s)
		441	9.04% (1s)	8.10% (1s)	2.09% (2s)	0.00% (3s)	0.00% (538s)	0.00% (25s)	0.00% (4s)
		10201	7.94% (13s)	6.85% (18s)	2.40% (44s)	2.40% (42s)	0.79% (3h)	0.00% (687s)	0.00% (417s)
$\{0,1\}^4$	I_2	121	6.73% (0s)	6.36% (0s)	2.90% (0s)	2.90% (0s)	0.00% (9s)	0.00% (12s)	0.00% (6s)
		441	6.89% (1s)	6.30% (1s)	2.84% (1s)	2.84% (1s)	0.00% (78s)	0.00% (22s)	0.00% (11s)
		10201	7.03% (17s)	6.45% (26s)	2.84% (22s)	1.38% (44s)	0.73% (3h)	0.00% (376s)	0.00% (115s)
$\{0,1\}^4$	H	121	10.82% (0s)	10.39% (0s)	2.60% (1s)	1.61% (1s)	0.00% (14s)	0.00% (6s)	0.00% (1s)
		441	10.39% (0s)	9.73% (1s)	3.36% (1s)	3.36% (1s)	0.00% (249s)	0.00% (28s)	0.00% (21s)
		10201	9.64% (9s)	8.83% (18s)	2.64% (23s)	2.64% (24s)	0.00% (4,835s)	0.00% (326s)	0.00% (190s)

[a] We report 3h if the time limit of three hours is exceeded.

Table 5.4. Investment planning problems ($\mathcal{X} = \mathbb{R}^2_+$): root node gaps.

Instance			Root node gap (computation time[a])						
$\mathcal{Y}$	T_ω	$\lvert\Omega\rvert$	SB	SB*	L	L*	S-CP* (L&P)	S-RG	S-RG*
$\mathbb{Z}^4_+$	I_2	121	7.71% (0s)	4.13% (2s)	1.00% (1s)	0.00% (1s)	0.00% (13s)	0.00% (6s)	0.00% (2s)
		441	7.78% (0s)	5.66% (4s)	2.37% (3s)	2.02% (5s)	0.46% (3h)	0.52% (157s)	0.15% (555s)
		10201	9.47% (7s)	8.55% (67s)	3.97% (137s)	3.89% (128s)	3.13% (3h)	1.85% (3h)	1.85% (3h)
$\mathbb{Z}^4_+$	H	121	9.58% (0s)	9.10% (0s)	0.94% (1s)	0.00% (2s)	0.00% (65s)	0.00% (7s)	0.00% (4s)
		441	10.06% (1s)	9.41% (1s)	3.80% (5s)	3.51% (7s)	0.00% (1,504s)	0.00% (62s)	0.00% (21s)
		10201	9.66% (12s)	9.00% (34s)	4.16% (148s)	4.14% (191s)	3.03% (3h)	1.59% (3h)	1.39% (3h)
$\{0,1\}^4$	I_2	121	10.44% (0s)	10.32% (0s)	3.26% (1s)	1.52% (1s)	0.00% (47s)	0.00% (42s)	0.00% (4s)
		441	9.45% (1s)	9.37% (1s)	4.90% (3s)	4.88% (4s)	1.12% (3h)	0.33% (5,685s)	0.45% (3,467s)
		10201	11.14% (16s)	11.07% (17s)	6.06% (80s)	6.04% (85s)	3.45% (3h)	3.09% (3h)	2.87% (3h)
$\{0,1\}^4$	H	121	10.89% (0s)	10.50% (0s)	4.73% (1s)	4.72% (2s)	0.01% (151s)	0.00% (65s)	0.00% (12s)
		441	11.28% (0s)	10.94% (1s)	4.86% (4s)	4.85% (4s)	0.00% (282s)	0.00% (124s)	0.00% (71s)
		10201	11.18% (9s)	10.80% (21s)	4.59% (105s)	4.56% (103s)	2.16% (3h)	1.53% (3h)	1.24% (3h)

[a]We report 3h if the time limit of three hours is exceeded.

There are several interesting observations to make from these results. First, observe that our enhanced cuts are able to significantly reduce the root node gaps for both SB and L cuts, at the expense of very little computational overhead. Indeed, the SB* cuts reduce the root node gap compared to the SB cuts by an average of roughly 16%, or 1 percentage point, and the L* cuts improve over the L cuts on 19 out of 24 instances by approximately 40% on average, or 0.6 percentage points. In fact, on 5 instances, the enhanced L cuts were able to achieve a zero root node gap.

Second, we observe from Tables 5.3 and 5.4 that our scaled cuts clearly outperform the traditional cuts, see also Figure 5.3 for a direct comparison of S-RG* cuts to L cuts. For example, both the S-RG and S-RG* cuts achieve a zero root node gap for all instances in Table 5.3 and half of the instances in Table 5.4. For the other instances in Table 5.4, the S-RG and S-RG* cuts strictly reduce the root node gap of the L cuts, by an average of 66% and 71%, respectively. Overall, we prefer the S-RG* cuts over their unenhanced counterparts, because, in general, they are considerably faster and perform slightly better in terms of root node gaps. For example, the S-RG* cuts are on average almost three times faster than the S-RG cuts for the instances in Table 5.3. Finally, the S-CP* (L&P) cuts outperform the L cuts on 23 out of 24 instances, reducing the average root node gap from 2.9% to 0.7%, and achieving a zero root node gap on 13 instances. However, a head-to-head comparison reveals that the S-RG* cuts are strictly preferred over the S-CP* cuts, because the S-RG* cuts perform at least as well in terms of both the root node gap and computation time.

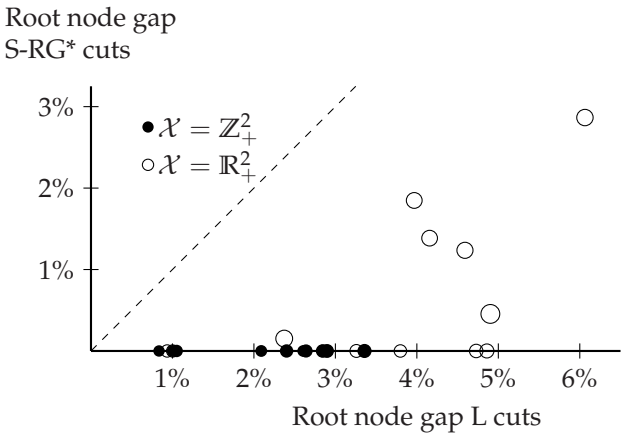


Figure 5.3. Investment planning problems: root node gap of L and S-RG* cuts. Point size is proportional to the logarithm of the computation time ratio of S-RG* cuts relative to L cuts.

Clearly, the tighter gaps of the S-RG* cuts compared to the L cuts are achieved at the cost of higher computation times. Indeed, the S-RG* cuts are not able to solve the four instances with $\mathcal{X} = \mathbb{R}^2_+$ and $S = 10201$ within the time limit of three hours, as opposed to the L cuts, and for the remaining instances, the computation time with S-RG* cuts is typically well over tenfold the computation time with L cuts. The reason is that, on average, the number of S-RG* cuts required for convergence is roughly ten times the number of L cuts, and, in addition, generating a single S-RG* cut typically takes five times as long as a single L cut. To analyse this further, consider Figure 5.4, which breaks down the computation time for generating a single dominating S-RG* cut.

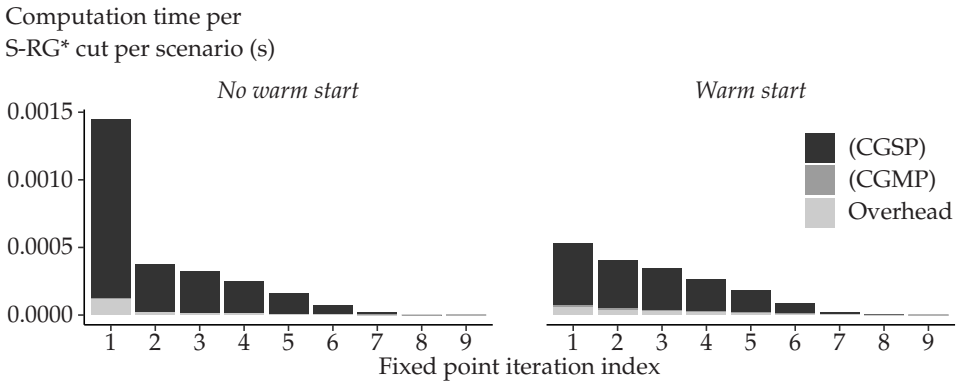


Figure 5.4. Breakdown of the computation time required to generate a single dominating S-RG* cut for the investment planning problems (averaged over the instances). The left figure shows the computation times if the extreme points of $\text{conv}(S^{\phi}_{\omega})$ identified during previous runs of the fixed point iteration algorithm are discarded, whereas for the right figure, we use those points to warm-start the row-generation scheme in subsequent runs of the fixed point iteration algorithm.

First, we observe from Figure 5.4 that most of the computation time for generating S-RG* cuts is devoted to solving (CGSP), and that in comparison, the time spent solving (CGMP) is negligible. This is in line with our expectations, since (CGSP) and (CGMP) are a small-scale MIP and LP, respectively, and they have to be solved an equal number of times during a run of the row-generation scheme. Second, the fixed point iteration algorithm spends a significant fraction of the computation time in the first iteration, and the computation time per iteration strictly diminishes with the iteration index. The reason is that extreme points of $\text{conv}(S^{\phi}_{\omega})$ identified during an iteration of the fixed point iteration algorithm are reused in subsequent itera-

tions, and thus we have to solve the subproblems (CGMP) and (CGSP) less often. In fact, Figure 5.4 shows that if we reuse the extreme points of $\text{conv}(S_\omega^\phi)$ identified during a *run* of the fixed-point iteration algorithm in subsequent runs, then we are able to compute S-RG* cuts faster, in particular by reducing the time spent in the first iteration.

5.6.3.3 Large investment planning problems

In order to assess the impact of computation times on the effectiveness of scaled cuts, we consider larger versions of the original IPP instance. We refer to Appendix 5.B for the construction of these instances, which vary in the input size of the first and second-stage subproblems, the number of scenarios, and the integer restrictions on the decision variables. In particular, we consider $n_1 \in \{2, 6, 10\}$, $n_2 \in \{10, 20\}$, $m_2 \in \{5, 10\}$, and $|\Omega| \in \{100, 500\}$. As before, $\mathcal{X}$ is either $\mathbb{Z}_+^{n_1}$ or $\mathbb{R}_+^{n_1}$ and $\mathcal{Y}$ is either $\{0,1\}^{n_2}$ or $\mathbb{Z}_+^{n_2}$, resulting in a total of 96 instances by examining all possible combinations. For these instances, we compare the S-RG* cuts to L cuts in terms of the resulting root node gap and computation time, see Figure 5.5.

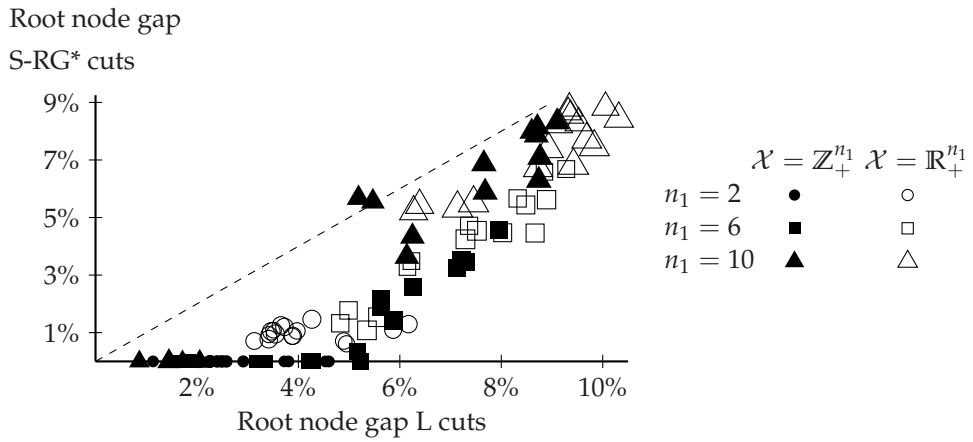


Figure 5.5. Large investment planning problems: root node gap of L and S-RG* cuts. Point size is proportional to the logarithm of the computation time ratio of S-RG* cuts relative to L cuts.

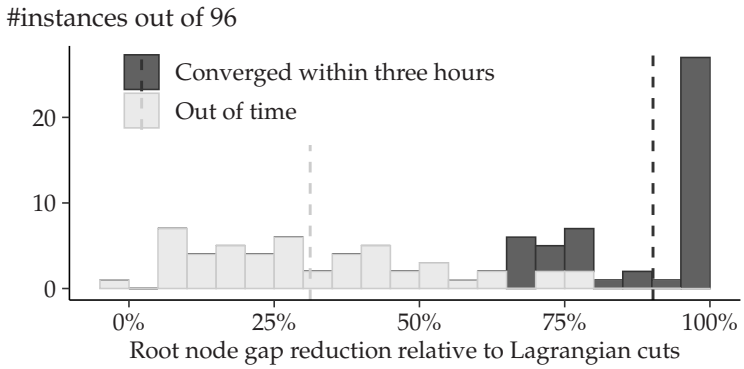


Figure 5.6. Performance of S-RG* cuts on the large investment planning problem instances, measured by the resulting reduction of the root node gap of L cuts.

First, observe from Figure 5.5 that the S-RG* cuts are able to significantly reduce the root node gap of L cuts. Indeed, a closer inspection reveals that the average reduction is almost 60%, and that we achieve a zero root node gap on 25 out of 96 instances. Second, the instances with $\mathcal{X} = \mathbb{R}^{n_1}_+$ are consistently harder to solve compared to $\mathcal{X} = \mathbb{Z}^{n_1}_+$: for the former instances, the S-RG* cuts reduce the root node gap of L cuts from 6.7% to 4.1%, compared to a reduction from 4.6% to 2.1% if $\mathcal{X} = \mathbb{Z}^{n_1}_+$. Third, the effectiveness of scaled cuts relative to L cuts is less convincing if $n_1 = 10$ compared to $n_1 \in \{2, 6\}$. An explanation suggested by Figure 5.5 is that solving the instances with $n_1 = 10$ using S-RG* cuts requires significantly more computation time than L cuts. In fact, our Benders' decomposition was able to converge within three hours of computation time for all 32 instances with $n_1 = 2$, but only 9 and 4 of the instances with $n_1 = 6$ and $n_1 = 10$, respectively. In addition, Figure 5.6 clearly shows that if the S-RG* cuts were able to converge, then they achieved a much larger reduction of the root node gap of L cuts compared to the instances for which the S-RG* cuts timed out.

An explanation for these notable differences in computation times is offered by Figure 5.7, which shows the number of row generation iterations required to generate a dominating scaled cut, i.e., the number of extreme points of $\text{conv}(S^{\phi}_{\omega})$ identified during a run of the row-generation scheme. Overall, the number of extreme points required for convergence is modest, e.g., even for the relatively large instances with $(n_1, n_2, m_2) = (6, 20, 10)$, no more than 30 points are required for convergence on average. In fact, the number of extreme points is relatively stable as the input size of the second-stage problem varies, but depends strongly on the number of first-stage variables n_1, which reveals why the instances with $n_1 = 10$

require more computation time compared to $n_1 \in \{2, 6\}$. Finally, Figure 5.7 clearly shows that the row generation scheme requires more iterations if $\mathcal{X} = \mathbb{R}_+^{n_1}$ than if $\mathcal{X} = \mathbb{Z}_+^{n_1}$, which is in line with our earlier observation that the former instances are harder to solve.

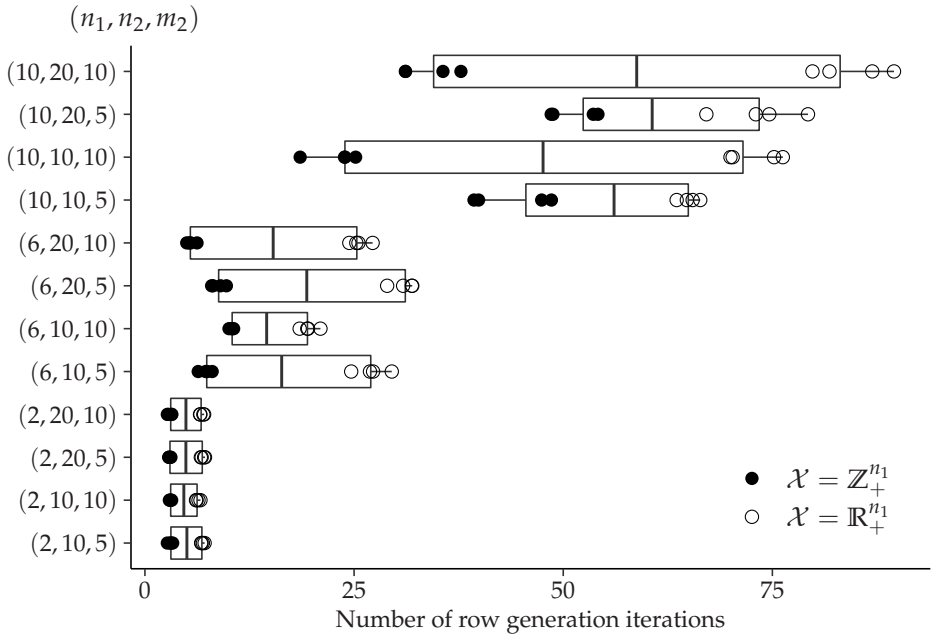


Figure 5.7. Boxplots of the average number of row generation scheme iterations per scenario for generating a single dominating scaled cut for the large IPP instances. The boxplot for each input size (n_1, n_2, m_2) corresponds to eight IPP instances (indicated by dots), with $\mathcal{X} \in \{\mathbb{Z}_+^{n_1}, \mathbb{R}_+^{n_1}\}$, $\mathcal{Y} \in \{\{0,1\}^{n_2}, \mathbb{Z}_+^{n_2}\}$, and $|\Omega| \in \{100, 500\}$.

5.6.3.4 The DCAP instances

The DCAP instances in [2] concern a multi-period capacity planning problem, in which the first-stage decisions pertain to buying resource capacity, and the second-stage problem is to assign these resources to a set of tasks. In addition, task processing requirements are uncertain, which translates to randomness in the recourse matrix W_ω. The instances are larger than the investment planning problem in [67], and differ in the number of resources, tasks, periods, and scenarios, see Table 5.5. Furthermore, they have mixed-binary first-stage decision variables, which are used to model fixed set-up costs that we incur if we buy capacity. As a result, the Benders' master problem (MP) is a MIP, which we solve the using a pure cutting plane approach, similar to the IPP instances. In addition, however, we solve (MP) using

a branch-and-cut algorithm in which we add at most five Fenchel cuts to solve the nodal subproblems if the number of leaf nodes is less than eight. Furthermore, we maintain separate outer approximations for each nodal subproblem to speed up convergence, since they are potentially stronger than a global outer approximation.

In the resulting Benders' decomposition algorithms, we solve the DCAP instances by adding optimality cuts according to cut hierarchies: we first exhaust lower-level optimality cuts before using higher-level cuts. For example, in every iteration of our Benders' decomposition, we first use SB* cuts to improve the outer approximation, and if this fails, we resort to L* cuts. We use the notation SB*+L* to denote this specific cut hierarchy. In addition, we consider the cut hierarchy SB*+L*+S-RG*, and we benchmark both hierarchies against stand-alone SB* cuts, see Table 5.6. Finally, we do not include S-CP* cuts in the comparison, because our results indicate that they are consistently outperformed by the S-RG* cuts, similar as for the IPP instances.

Table 5.5. DCAP instances: input size of large-scale deterministic equivalent model.

Instance	#scenarios	#constraints	#variables	#binary variables	#nonzero constraint coefficients
DCAP_233_200	200	3,006	5,412	5,406	11,412
DCAP_233_300	300	4,506	8,112	8,106	17,112
DCAP_233_500	500	7,506	13,512	13,506	28,512
DCAP_243_200	200	3,606	7,212	7,206	14,412
DCAP_243_300	300	5,406	10,812	10,806	21,612
DCAP_243_500	500	9,006	18,012	18,006	36,012
DCAP_332_200	200	2,406	4,812	4,806	10,212
DCAP_332_300	300	3,606	7,212	7,206	15,312
DCAP_332_500	500	6,006	12,012	12,006	25,512
DCAP_342_200	200	2,806	6,412	6,406	13,012
DCAP_342_300	300	4,206	9,612	9,606	19,512
DCAP_342_500	500	7,006	16,012	16,006	32,512

Table 5.6. DCAP instances: optimality gaps

	LB gap - UB gap[a] (computation time[b]).			
	MP: Fenchel cutting planes			MP: Branch-and-cut
Instance	SB*	SB*+L*	SB*+L*+S-RG*	SB*+L*+S-RG*
DCAP_233_200	26.78% - 3.68% (7s)	0.12% - 2.49% (359s)	0.01% - 0.45% (2,238s)	0.00% - 0.00% (617s)
DCAP_233_300	27.86% - 6.78% (10s)	0.10% - 0.20% (435s)	0.02% - 0.13% (9,468s)	0.02% - 0.11% (1,177s)
DCAP_233_500	30.24% - 11.12% (8s)	0.04% - 0.46% (551s)	0.00% - 0.07% (1,551s)	0.00% - 0.00% (527s)
DCAP_243_200	22.80% - 1.31% (4s)	0.04% - 0.15% (943s)	0.01% - 0.15% (2,160s)	0.00% - 0.14% (2,102s)
DCAP_243_300	22.69% - 1.39% (8s)	0.10% - 0.54% (552s)	0.03% - 0.37% (3h)	0.03% - 0.26% (8,442s)
DCAP_243_500	23.30% - 0.88% (12s)	0.12% - 0.47% (4,252s)	0.04% - 0.24% (3h)	0.02% - 0.15% (5,885s)
DCAP_332_200	44.53% - 51.42% (2s)	0.21% - 0.59% (261s)	0.06% - 0.38% (3h)	0.02% - 0.39% (3h)
DCAP_332_300	44.79% - 28.42% (1s)	0.18% - 0.25% (793s)	0.06% - 0.25% (3h)	0.04% - 0.72% (3h)
DCAP_332_500	47.61% - 18.18% (1s)	0.13% - 0.52% (1,306s)	0.08% - 0.52% (2,236s)	0.47% - 0.43% (3h)
DCAP_342_200	40.89% - 9.59% (6s)	0.12% - 7.21% (1,382s)	0.07% - 3.69% (3,535s)	0.12% - 1.76% (3h)
DCAP_342_300	40.92% - 7.15% (5s)	0.11% - 3.64% (1,958s)	0.04% - 2.33% (7,115s)	0.22% - 0.26% (3h)
DCAP_342_500	38.20% - 6.71% (14s)	0.09% - 3.05% (3,390s)	0.05% - 1.55% (7,087s)	0.03% - 0.29% (3h)

[a] The LB and UB gaps are defined as $(\eta^* - LB)/|\eta^*| \times 100\%$ and $(UB - \eta^*)/|\eta^*| \times 100\%$, respectively.
[b] We report 3h if the time limit of three hours is exceeded.

We observe from the results in Table 5.6 that both cut hierarchies clearly outperform the stand-alone SB* cuts, and that our scaled cuts yield are able to reduce the LB and UB gaps of traditional cuts. Indeed, if we solve (MP) using a pure cutting plane approach, then the average LB gap of SB* cuts reduces from 34% to 0.4% by including LR* cuts in the hierarchy, and further reduces to 0.1% if we also include S-RG* cuts. In addition, we are able to find better incumbent solutions: the average UB gap resulting from SB* cuts reduces from 12% to 1.6% and 0.8% if we additionally include LR* and S-RG* cuts in the hierarchy, respectively. A direct comparison of the pure cutting planes approach to the branch-and-cut approach for solving (MP) indicates that we typically obtain better candidate solutions if we solve the master problem using a B&C approach: the UB gap improves on 9 out of 12 instances compared to a pure cutting plane approach, and the average UB gap reduces from over 0.8% to less than 0.4%. In terms of computation time, however, neither approach strictly outperforms the other. For example, the branch-and-cut approach converges faster on the DCAP_233 and DCAP_243 instances, but, in contrast to the cutting plane approach, is not able to solve the DCAP_342 instances within three hours of computation time.

5.7 Conclusion

We propose a new family of optimality cuts which can be used to solve general two-stage mixed-integer recourse (MIR) models. These so-called *scaled cuts* are derived by solving extended formulations of the second-stage subproblems. In contrast to existing optimality cuts, scaled cuts can be used to recover the convex envelope of the expected second-stage cost function *in general*. That is, we allow for general mixed-integer decision variables in both stages, and we do not make restrictive assumptions regarding the uncertain parameters in the model, e.g., we do not require that the problem exhibits fixed recourse. We describe efficient primal and dual subroutines for computing our scaled cuts, which are based on vertex enumeration and cutting planes techniques, respectively, and we propose a novel cut-enhancement technique to accelerate the convergence of our scaled cuts. To demonstrate the effectiveness of the (enhanced) scaled cuts, we solve a number of MIR problem instances from the literature, and we find that we are able to improve significantly over existing optimality cuts in terms the resulting optimality gap at the expense of additional computation time.

One avenue for future research is the extension to multi-stage MIR models and to problems with non-linear cost functions, such as quadratic or conic MIR models. An alternative direction is to compute scaled cuts using inexact lower bounds for the expected second-stage cost function, which can be obtained by solving convex approximations of the original MIR model. Typically, such inexact lower bounds are relatively inexpensive to generate, and thus they may be used to speed up the convergence of our scaled cuts.

Appendix 5.A Postponed proofs

The proofs of Lemmas 5.1-5.4 and Proposition 5.1 are not only postponed to the appendix for ease of presentation, they also depend on the characterizations of the set $\Pi_\omega(\phi)$ and the function $C_\omega(\rho)$ in Lemmas 5.5 and 5.6 in Section 5.4, respectively. The proofs of these lemmas are independent of the results in Section 5.3. The proofs of Lemmas 5.1-5.4 can be read in the same order as they appear in the main text. We only remark that the proof of Proposition 5.1 depends on Lemma 5.3 and is for this reason given after the proof of that lemma.

Proof of Lemma 5.1. We have to show that

$$\sup_{\alpha,\beta,\tau}\{\alpha - \beta^\top \bar{x} - \tau\phi(\bar{x}) : \ (\alpha,\beta,\tau) \in \Pi_\omega(\phi)\} = v_\omega(\bar{x}), \tag{5.26}$$

and that the supremum in (5.26) is attained by some $(\alpha,\beta,\tau) \in \Pi_\omega(\phi)$. In the proof, we will use the definition of $C_\omega(\rho)$ in (5.15), which we repeat here for convenience,

$$C_\omega(\rho) = \sup_{\alpha,\beta,\tau}\{\alpha - \beta^\top \bar{x} - (1+\tau)\rho : \ (\alpha,\beta,\tau) \in \Pi_\omega(\phi)\}. \tag{5.27}$$

In particular, it also suffices to show that $C_\omega(\phi(\bar{x})) = -\phi(\bar{x}) + v_\omega(\bar{x})$, and that the supremum in (5.27) with $\rho = \phi(\bar{x})$ is attained.

We first show that the problem in (5.27) is feasible and bounded, so that the corresponding supremum is attained, using the polyhedrality of $\Pi_\omega(\phi)$ from Lemma 5.5. Feasibility follows from the fact that v_ω is bounded from below, which is a consequence of Assumptions (A3) and (A4). Boundedness follows from the definition of $\Pi_\omega(\phi)$, which implies that

$$C_\omega(\phi(\bar{x})) \leq -\phi(\bar{x}) + v_w(\bar{x}) < \infty,$$

where the latter inequality follows from $\bar{x} \in X$ and Assumption (A1). In the remainder of the proof, we show that $C_\omega(\phi(\bar{x})) \geq -\phi(\bar{x}) + v_\omega(\bar{x})$.

In particular, we use the dual representation of $C_\omega(\rho)$ in Lemma 5.6 to obtain that

$$C_\omega(\phi(\bar{x})) = -\phi(\bar{x}) + \min_y\{q_\omega^\top y : \ (\bar{x},\phi(\bar{x}),y) \in \operatorname{conv}(S_\omega^\phi)\},$$

where

$$S_\omega^\phi := \{(x,\theta,y) \in X \times \mathbb{R} \times \mathcal{Y} : \ \theta \geq \phi(x), \ W_\omega y = h_\omega - T_\omega x\},$$

and we show that $q_\omega^\top y \geq v_\omega(\bar{x})$ for every y such that $(\bar{x},\phi(\bar{x}),y) \in \operatorname{conv}(S_\omega^\phi)$. Fix such y arbitrarily, and let $(x^i,\theta^i,y^i) \in S_\omega^\phi$, $i = 1,\dots,d$, denote the extreme points of $\operatorname{conv}(S_\omega^\phi)$. Then, there exist $\lambda^i \geq 0$, $i = 1,\dots,d$ and $\mu_1 \geq 0$, for which

$$(\bar{x},\phi(\bar{x}),y) = \sum_{i=1}^d \lambda^i(x^i,\theta^i,y^i) + (0,\mu_1,0),$$

and $\sum_{i=1}^{d} \lambda^i = 1$. Noting that $(x^i, \theta^i) \in \text{epi}_X(\phi)$, and using the assumption that $(\bar{x}, \phi(\bar{x}))$ is an extreme point of $\text{conv}(\text{epi}_X(\phi))$ it follows that $(x^i, \theta^i) = (\bar{x}, \phi(\bar{x}))$ for every $i = 1, \ldots, d$. The desired inequality then follows:

$$q_\omega^\top y = \sum_{i=1}^{d} \lambda^i q_\omega^\top y^i \geq \sum_{i=1}^{d} \lambda^i v_\omega(x^i) = \sum_{i=1}^{d} \lambda^i v_\omega(\bar{x}) = v_\omega(\bar{x}),$$

where the inequality follows from feasibility of y^i in $v_\omega(x^i) = \min_{y \in \mathcal{Y}}\{q_\omega^\top y : W_\omega y = h_\omega - T_\omega x^i\}$. ■

Proof of Lemma 5.2. We first show that $C(\cdot)$ is convex and continuous on $\text{dom}(C)$, and that the supremum in (5.10) is attained for all $\rho \in \text{dom}(C)$. We use the expression $C(\rho) = \mathbb{E}_{\boldsymbol{\omega}}[C_{\boldsymbol{\omega}}(\rho)]$, where $C_\omega(\rho)$ is defined in (5.27), and we use the polyhedral representation of $\Pi_\omega(\phi)$ in Lemma 5.5 to obtain that

$$C_\omega(\rho) = \sup_{\alpha,\beta,\tau}\{\alpha - \beta^\top \bar{x} - \rho(1+\tau) : \\ q_\omega^\top y^i + \beta^\top x^i + \tau\theta^i \geq \alpha \ \forall i \in \{1, \ldots, d\},\ \tau \geq 0\}, \quad (5.28)$$

where (x^i, θ^i, y^i), $i = 1, \ldots, d$, are the extreme points of $\text{conv}(S_\omega^\phi)$. In particular, since the LP in (5.28) is feasible and bounded for all $\rho \in \text{dom}(C_\omega)$, the corresponding supremum is attained, and the corresponding value function $C_\omega(\cdot)$ is convex and continuous on $\text{dom}(C_\omega)$ for every $\omega \in \Omega$. It then follows from $C(\rho) = \mathbb{E}_{\boldsymbol{\omega}}[C_{\boldsymbol{\omega}}(\rho)]$ and

$$\text{dom}(C) = \bigcap_{\omega \in \Omega} \text{dom}(C_\omega), \quad (5.29)$$

that $C(\cdot)$ is convex and continuous on $\text{dom}(C)$, and that the corresponding supremum is attained.

To see that $C(\cdot)$ is strictly decreasing on $\text{dom}(C)$, fix $\rho_1, \rho_2 \in \text{dom}(C)$ such that $\rho_1 < \rho_2$. We know that there exist $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi)$, $\omega \in \Omega$, such that

$$C(\rho_2) = \mathbb{E}_{\boldsymbol{\omega}}\alpha_{\boldsymbol{\omega}} - \mathbb{E}_{\boldsymbol{\omega}}\beta_{\boldsymbol{\omega}}^\top \bar{x} - (1 + \mathbb{E}_{\boldsymbol{\omega}}\tau_{\boldsymbol{\omega}})\rho_2,$$

and, using the definition of $C(\rho_1)$, we obtain

$$C(\rho_1) \geq \mathbb{E}_{\boldsymbol{\omega}}\alpha_{\boldsymbol{\omega}} - \mathbb{E}_{\boldsymbol{\omega}}\beta_{\boldsymbol{\omega}}^\top \bar{x} - (1 + \mathbb{E}_{\boldsymbol{\omega}}\tau_{\boldsymbol{\omega}})\rho_1,$$

from which it follows that $C(\rho_1) > C(\rho_2)$, using that $\rho_1 < \rho_2$ and $\mathbb{E}_\omega \tau_\omega \geq 0$.

To prove (iii), fix $\bar{\rho} \in \text{dom}(C)$, and let $(\alpha_\omega, \beta_\omega, \tau_\omega)$ be an optimal solution of (5.10) with $\rho = \bar{\rho}$. We have to show that

$$C(\rho) \geq C(\bar{\rho}) - (1 + \mathbb{E}_\omega \tau_\omega)(\rho - \bar{\rho}) \quad \forall \rho \in \mathbb{R}.$$

This follows directly by substituting $C(\bar{\rho}) = \mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x} - \bar{\rho}(1 + \mathbb{E}_\omega \tau_\omega)$ and using the definition of $C(\rho)$ in (5.10).

Finally, to show that $\text{dom}(C) = [\phi(\bar{x}), \infty)$ if $\bar{x} \in X$, we will prove the slightly more general expression

$$\text{dom}(C) = \{\rho : (\bar{x}, \rho) \in \text{conv}(\text{epi}_X(\phi))\}, \tag{5.30}$$

for arbitrary $\bar{x} \in \bar{X}$, which reduces to $\text{dom}(C) = [\phi(\bar{x}), \infty)$ if $\bar{x} \in X$, since then $(\bar{x}, \rho) \in \text{conv}(\text{epi}_X(\phi))$ if and only if $\rho \geq \phi(\bar{x})$. We prove (5.30) from (5.29), by showing that $\text{dom}(C_\omega) = \{\rho : (\bar{x}, \rho) \in \text{conv}(\text{epi}_X(\phi))\}$ for every $\omega \in \Omega$. To do so, we use expression for the dual LP of (5.28) from Lemma 5.6, which we repeat here for convenience:

$$\min_y \{q_\omega^\top y : \ (\bar{x}, \rho, y) \in \text{conv}(S_\omega^\phi)\}. \tag{5.31}$$

In particular, we show that the dual LP in (5.31) is bounded and feasible if and only if $(\bar{x}, \rho) \in \text{conv}(\text{epi}_X(\phi))$. In fact, the dual LP is bounded for all $\rho \in \mathbb{R}$ as a consequence of Assumption (A1). To see that the dual problem is feasible if and only if $(\bar{x}, \rho) \in \text{conv}(\text{epi}_X(\phi))$, suppose that $(\bar{x}, \rho) \in \text{conv}(\text{epi}_X(\phi))$, i.e., there exist $\lambda^i \geq 0$, and $(x^i, \theta^i) \in \text{epi}_X(\phi)$, $i = 1, \ldots, d'$, such that $\sum_{i=1}^{d'} \lambda^i = 1$ and

$$(\bar{x}, \rho) = \sum_{i=1}^{d'} \lambda^i (x^i, \theta^i).$$

It follows from Assumption (A1) that there exist y^i such that $(x^i, \theta^i, y^i) \in S_\omega^\phi$, $i = 1, \ldots, d'$, and as a result

$$\left(\bar{x}, \rho, \sum_{i=1}^{d'} \lambda^i y^i\right) = \sum_{i=1}^{d'} \lambda^i (x^i, \theta^i, y^i) \in \text{conv}(S_\omega^\phi),$$

and thus $y := \sum_{i=1}^{d'} \lambda^i y^i$ is feasible in (5.31). The converse claim can be proved in a similar way. ■

Proof of Lemma 5.3. Using the definition of ρ^* in (5.9), we have

$$\begin{aligned}\rho^* &= \min_{\rho}\left\{\rho:\ \rho \geq \frac{\mathbb{E}_{\omega}\alpha_{\omega} - \mathbb{E}_{\omega}\beta_{\omega}^{\top}\bar{x}}{1+\mathbb{E}_{\omega}\tau_{\omega}}\ \forall(\alpha_{\omega},\beta_{\omega},\tau_{\omega})\in\Pi_{\omega}(\phi),\ \omega\in\Omega\right\}\\ &= \min_{\rho}\big\{\rho:\ \mathbb{E}_{\omega}\alpha_{\omega} - \mathbb{E}_{\omega}\beta_{\omega}^{\top}\bar{x} - \rho(1+\mathbb{E}_{\omega}\tau_{\omega}) \leq 0\ \forall(\alpha_{\omega},\beta_{\omega},\tau_{\omega})\in\\ &\qquad\qquad \Pi_{\omega}(\phi),\ \omega\in\Omega\big\},\end{aligned}$$

and using the definition of $C(\rho)$, we obtain $\rho^* = \min_{\rho}\{\rho:\ C(\rho)\leq 0\}$. Suppose now that $\bar{x}\in X$ and $\rho^* > \phi(\bar{x})$. It follows from $\rho^* > \phi(\bar{x})$ and (5.13) that $C(\phi(\bar{x}))> 0$, and thus ρ^* is the unique solution of $C(\rho)=0$, since $C(\cdot)$ is continuous and strictly decreasing on $\mathrm{dom}(C)=[\phi(\bar{x}),\infty)$, see Lemma 5.2. ■

Proof of Lemma 5.4. We first show that $C(\rho_k)\to 0$, which suffices to show that $\rho_k\to\rho^*$, since $C(\cdot)$ is continuous by Lemma 5.2, and ρ^* is the unique solution of $C(\rho)=0$ by Lemma 5.3.

In order to prove that $C(\rho_k)\to 0$, we rewrite the updating rule in (5.14) as

$$\rho_{k+1} - \rho_k = \frac{C(\rho_k)}{1+\mathbb{E}_{\omega}\tau_{\omega,k}}, \tag{5.32}$$

in which we use the notation $\tau_{\omega,k}$ to emphasize that τ_{ω} depends on the iteration, i.e., $\tau_{\omega,k}$ corresponds to an optimal solution of the problem in (5.27) with $\rho=\rho_k$. By construction, the sequence $\{\rho_k\}$ is non-decreasing and bounded, and thus convergent. Therefore, taking limits on both sides of (5.32) yields

$$0 = \lim_{k\to\infty}\frac{C(\rho_k)}{1+\mathbb{E}_{\omega}\tau_{\omega,k}},$$

and thus, we have to show that $1+\mathbb{E}_{\omega}\tau_{\omega,k}$ is eventually bounded over k. That is, it suffices to show that there exists a $\bar{\tau}$ such that $\tau_{\omega,k}\leq\bar{\tau}$ for all $k\geq 1$ and $\omega\in\Omega$.

We derive such a $\bar{\tau}$ by using that $\Pi_{\omega}(\phi)$ is polyhedral, see Lemma 5.5. In particular, let $(\alpha_{\omega}^i,\beta_{\omega}^i,\tau_{\omega}^i)$, $i=1,\ldots,d$, and $(\hat{\alpha}_{\omega}^j,\hat{\beta}_{\omega}^j,\hat{\tau}_{\omega}^j)$, $j=1,\ldots,r$ denote the extreme points and directions of $\Pi_{\omega}(\phi)$, respectively, $\omega\in\Omega$. Since $C_{\omega}(\rho_0)<\infty$ it must be that $\hat{\alpha}_{\omega}^j - \hat{\beta}_{\omega}^{j\top}\bar{x} - \rho_0(1+\hat{\tau}_{\omega}^j)\leq 0$, since otherwise it would be possible to improve the objective in (5.27) with $\rho=\rho_0$ without bound. Furthermore, we have that $\rho_k>\rho_0$ for every $k\geq 1$, since $\{\rho_k\}_{k\geq 0}$ is increasing, and $\rho_1>\rho_0$ by the assumption that $C(\rho_0)>0$. It follows that $\hat{\alpha}_{\omega}^j - \hat{\beta}_{\omega}^{j\top}\bar{x} - \rho_k(1+\hat{\tau}_{\omega}^j) < 0$ for every $k\geq 1$, and thus any optimal solution of the problem in (5.27) with $\rho=\rho_k$, $k\geq 1$, is a convex combination of the extreme points $(\alpha_{\omega}^i,\beta_{\omega}^i,\tau_{\omega}^i)$, $i=1,\ldots,d$, of $\Pi_{\omega}(\phi)$. Hence, we

can take $\bar{\tau} = \max_{\omega\in\Omega} \max_{i=1,\dots,d} \tau^i_\omega$.

Finally, we show that if $C(\rho_k) < \delta$, then $\rho_k \geq \rho^* - \delta$. To this end, let $(\alpha_\omega, \beta_\omega, \tau_\omega)$ be such that $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi)$ for every $\omega \in \Omega$, and $C(\rho^*) = \mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x} - (1 + \mathbb{E}_\omega \tau_\omega)\rho^*$, and use the definition of $C(\rho_k)$ to obtain that

$$C(\rho_k) \geq \mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x} - (1 + \mathbb{E}_\omega \tau_\omega)\rho_k.$$

It follows that

$$C(\rho_k) - C(\rho^*) \geq (1 + \mathbb{E}_\omega \tau_\omega)(\rho^* - \rho_k) \geq \rho^* - \rho_k,$$

and we obtain $\rho^* - \rho_k \leq \delta$ by substituting $C(\rho^*) = 0$ and $C(\rho_k) \leq \delta$, as desired. ■

Proof of Proposition 5.1. We will show that there exists a *finite* collection of optimality cuts

$$\mathrm{SCC}(\phi)(x) \geq \alpha_k - \beta_k^\top x \quad \forall x \in \bar{X}, \quad k = 1, \dots, K,$$

defined by rational data, which completely describe $\mathrm{SCC}(\phi)$, i.e., for every $\bar{x} \in \bar{X}$, there exist rational α_k and β_k, such that $\mathrm{SCC}(\phi)(\bar{x}) = \alpha_k - \beta_k^\top \bar{x}$. To this end, fix arbitrary $\bar{x} \in \bar{X}$, and note that $\mathrm{SCC}(\phi)(\bar{x}) = \rho^*$, where ρ^* is the optimal value of the problem in (5.9). By Lemma 5.3, we know that $\rho^* = \min_\rho\{\rho : \ C(\rho) \leq 0\}$, where $C(\rho)$ is defined as in (5.10). In particular, $C(\rho^*) \leq 0$, and we distinguish two cases: $C(\rho^*) = 0$, and $C(\rho^*) < 0$.

If $C(\rho^*) = 0$, then by Lemma 5.2, the optimal value ρ^* of the problem in (5.9) is attained by some $(\alpha_\omega, \beta_\omega, \tau_\omega)$, where for every $\omega \in \Omega$, $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi)$ attains the optimal value $C_\omega(\rho^*)$ of (5.27). Furthermore, since the feasible region $\Pi_\omega(\phi)$ of (5.27) is a rational polyhedron by Lemma 5.5, and the objective function is linear, it follows that the optimal value $C_\omega(\rho^*)$ is in fact attained by one of the finitely many rational extreme points of $\Pi_\omega(\phi)$, and thus we assume, without loss of generality, that $(\alpha_\omega, \beta_\omega, \tau_\omega)$ is a rational extreme point of $\Pi_\omega(\phi)$, $\omega \in \Omega$. By definition of $\mathrm{SCC}(\phi)$, we have

$$\mathrm{SCC}(\phi)(x) \geq \frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top x}{1 + \mathbb{E}_\omega \tau_\omega} \quad \forall x \in \bar{X}. \tag{5.33}$$

and, in addition, $(\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x})/(1 + \mathbb{E}_\omega \tau_\omega) = \rho^* = \mathrm{SCC}(\phi)(\bar{x})$, i.e., the optimality cut in (5.33) is valid and tight at $\bar{x}$. Moreover, the cut in (5.33) corresponds to one of the finitely many combinations of rational extreme points of $\Pi_\omega(\phi)$, $\omega \in \Omega$.

In order to analyse the case where $C(\rho^*) < 0$, we first show that

$$\rho^* = \min_{\rho}\{\rho : \ C(\rho) < \infty\}. \tag{5.34}$$

To see this, suppose for contradiction that there exists a $\rho' < \rho^*$ such that $C(\rho') < \infty$. But then, it must be that $C(\rho') > 0$, since $\rho^* = \min_\rho\{\rho : \ C(\rho) \leq 0\}$, and thus the continuity of $C(\cdot)$ established in Lemma 5.2 implies that there exists a $\rho'' \in (\rho', \rho^*)$ such that $C(\rho'') = 0$, which is a contradiction. Then, using the expression for $\mathrm{dom}(C)$ in (5.30), we obtain from (5.34) that

$$\rho^* = \min_{\rho}\{\rho : \ (\bar{x}, \rho) \in \mathrm{conv}(\mathrm{epi}_X(\phi))\}, \tag{5.35}$$

and using similar reasoning for arbitrary $x \in \bar{X}$, we have that

$$\mathrm{SCC}(\phi)(x) \geq \min_{\rho}\{\rho : \ (\bar{x}, \rho) \in \mathrm{conv}(\mathrm{epi}_X(\phi))\} \quad \forall x \in \bar{X}.$$

Thus, if we define E_ϕ as the set of cut coefficients which define valid inequalities for $\mathrm{conv}(\mathrm{epi}_X(\phi))$, i.e.,

$$E_\phi := \{(\alpha, \beta) : \ \theta + \beta^\top x \geq \alpha \ \forall (x, \theta) \in \mathrm{conv}(\mathrm{epi}_X(\phi))\},$$

then every $(\alpha, \beta) \in E_\phi$ defines an optimality cut of the form

$$\mathrm{SCC}(\phi)(x) \geq \alpha - \beta^\top x \ \forall x \in \bar{X}. \tag{5.36}$$

In addition, using (5.35) and that the function ϕ is convex polyhedral, we obtain

$$\rho^* = \max_{\alpha,\beta}\{\alpha - \beta^\top \bar{x} : \ (\alpha, \beta) \in E_\phi\}, \tag{5.37}$$

i.e., if (α, β) is optimal in (5.37), then the cut in (5.36) is tight at $\bar{x}$. Moreover, the maximum in (5.37) is attained by one of the finitely many rational extreme points of E_ϕ, since E_ϕ is a rational polyhedron. To see this, use [28, Theorem 1] to obtain that $\mathrm{conv}(\mathrm{epi}_X(\phi))$ is a rational polyhedron, and note that

$$E_\phi = \{(\alpha, \beta) : \ \theta^i + \beta^\top x^i \geq \alpha, \ i = 1, \ldots, d\},$$

where (x^i, θ^i), $i = 1, \ldots, d$, denote the (rational) extreme points of $\mathrm{conv}(\mathrm{epi}_X(\phi))$, see, e.g., [52]. ■

Proof of Lemma 5.9. We show that $\mathbb{T}$ is Lipschitz continuous with Lipschitz constant equal to 1, i.e., $||\mathbb{T}f - \mathbb{T}g||_\infty \leq ||f-g||_\infty$ for all $f, g \in C(\bar{X})$. Indeed, for arbitrary $f, g \in C(\bar{X})$, we have

$$\begin{aligned}
||\mathbb{T}f - \mathbb{T}g||_\infty &= \sup_{x\in\bar{X}} \Bigg| \sup_{\tau_\omega \geq 0} \left\{ \frac{\mathbb{E}_\omega \overline{\text{co}}(v_\omega + \tau_\omega f)(x)}{1+\mathbb{E}_\omega \tau_\omega} \right\} \\
&\qquad - \sup_{\tau_\omega \geq 0} \left\{ \frac{\mathbb{E}_\omega \overline{\text{co}}(v_\omega + \tau_\omega g)(x)}{1+\mathbb{E}_\omega \tau_\omega} \right\} \Bigg| \\
&\leq \sup_{x\in\bar{X}} \sup_{\tau_\omega \geq 0} \left| \left\{ \frac{\mathbb{E}_\omega \left[\overline{\text{co}}(v_\omega + \tau_\omega f)(x) - \overline{\text{co}}(v_\omega + \tau_\omega g)(x)\right]}{1+\mathbb{E}_\omega \tau_\omega} \right\} \right| \\
&\leq \sup_{\tau_\omega \geq 0} \left\{ \frac{\mathbb{E}_\omega ||\overline{\text{co}}(v_\omega + \tau_\omega f) - \overline{\text{co}}(v_\omega + \tau_\omega g)||_\infty}{1+\mathbb{E}_w \tau_\omega} \right\} \\
&\leq \sup_{\tau_\omega \geq 0} \left\{ \frac{\mathbb{E}_\omega \tau_\omega}{1+\mathbb{E}_w \tau_\omega} \right\} ||f-g||_\infty \\
&= ||f-g||_\infty,
\end{aligned}$$

where the final inequality follows from the fact that $||\overline{\text{co}}(f) - \overline{\text{co}}(g)||_\infty \leq ||f-g||_\infty$. To see this, let $\delta = ||f-g||_\infty$, and fix $x \in \bar{X}$. Note that

$$\overline{\text{co}}(f)(x) - \delta \leq f(x) - \delta \leq g(x).$$

Because $\overline{\text{co}}(f) - \delta$ is convex and lsc, it follows that

$$\overline{\text{co}}(f)(x) - \delta \leq \overline{\text{co}}(g)(x).$$

Analogously, we can show that $\overline{\text{co}}(g)(x) - \delta \leq \overline{\text{co}}(f)(x)$, and the result follows. ■

Proof of Lemma 5.10. We have to prove that for every $\varepsilon > 0$ there exists $\tau_\omega \geq 0$ such that $\overline{\text{co}}(v_\omega + \tau f)(\bar{x}) \geq v_\omega(\bar{x}) + \tau f(\bar{x}) - \varepsilon$. We will do so by showing that there exist α, β, and $\tau \geq 0$ such that (i) $v_\omega(x) + \tau f(x) \geq \alpha - \beta^\top x \ \forall x \in X$, and (ii) $\alpha - \beta^\top \bar{x} \geq v_\omega(\bar{x}) + \tau f(\bar{x}) - \varepsilon$. The claim then follows by letting $\varepsilon \to 0$.

Let $\varepsilon > 0$ be given and define $v_\omega^+ : \text{epi}(f) \mapsto \mathbb{R}$ as $v_\omega^+(x,\theta) = v_\omega(x),\ (x,\theta) \in \text{epi}(f)$. We prove that α, β, and $\tau \geq 0$ satisfying (i) and (ii) exist by showing that $\overline{\text{co}}(v_\omega^+)(\bar{x},\bar{\theta}) = v_\omega(\bar{x})$. Then, by definition of $\overline{\text{co}}(v_\omega^+)$, there exist α', β', and τ' such that

$$v_\omega^+(x,\theta) \geq \alpha' - {\beta'}^\top x - \tau'\theta \quad \forall (x,\theta) \in \text{epi}(f), \tag{5.38}$$

and

$$\alpha' - {\beta'}^\top \bar{x} - \tau' \bar{\theta} \geq v_\omega(\bar{x}) - \varepsilon. \tag{5.39}$$

Note that (5.38) implies that $v_\omega(x) \geq \alpha' - {\beta'}^\top x - \tau' f(x)\ \forall x \in X$ since $v_\omega^+(x,\theta) = v_\omega(x)$ and $(x, f(x)) \in \text{epi}(f)\ \forall x \in X$. In addition, (5.39) implies that, $\alpha' - {\beta'}^\top \bar{x} - \tau' f(\bar{x}) \geq v_\omega(\bar{x}) - \varepsilon$, since $\bar{\theta} = f(\bar{x})$. Thus, we may take $(\alpha', \beta', \tau') = (\alpha, \beta, \tau)$.

It remains to show that $\overline{\text{co}}(v_\omega^+)(\bar{x}, \bar{\theta}) = v_\omega(\bar{x})$. Note that v_ω^+ is lsc, since v_ω is lsc and $\text{epi}(f)$ is a closed set. Analogous to [56, Corollary 3.47], it follows that $\text{co}(v_\omega^+)$ is lsc, and as a result, $\text{co}(v_\omega^+) = \overline{\text{co}}(v_\omega^+)$. In addition, since $(\bar{x}, \bar{\theta})$ is an extreme point of $\text{epi}(f)$, we have $\text{co}(v_\omega^+)(\bar{x}, \bar{\theta}) = v_\omega^+(\bar{x}, \bar{\theta})$ [75, Corollary 3]. Hence, $\overline{\text{co}}(v_\omega^+)(\bar{x}, \bar{\theta}) = \text{co}(v_\omega^+)(\bar{x}, \bar{\theta}) = v_\omega^+(\bar{x}, \bar{\theta}) = v_\omega(\bar{x})$. ■

Appendix 5.B Construction of large IPP instances

The large IPP instances that we consider are of the form

$$\min_{x \in \mathcal{X}} \left\{ c^\top x + \mathbb{E}_{\boldsymbol{\omega}} \left[\min_{y \in \mathcal{Y}} \{ q_{\boldsymbol{\omega}}^\top y :\ W_{\boldsymbol{\omega}} y \leq h_{\boldsymbol{\omega}} - T_{\boldsymbol{\omega}} x,\ y \in \mathcal{Y} \} \right] :\ x \in [0,5]^{n_1} \right\}, \tag{5.40}$$

where $\mathcal{X}$ is either $\mathbb{R}_+^{n_1}$ or $\mathbb{Z}_+^{n_1}$ and $\mathcal{Y}$ is either $\{0,1\}^{n_2}$ or $\mathbb{Z}_+^{n_2}$. Furthermore, the components of the cost vector c and the distributions of the stochastic elements $(h_{\boldsymbol{\omega}}, T_{\boldsymbol{\omega}}, q_{\boldsymbol{\omega}}, W_{\boldsymbol{\omega}})$ are chosen in such a way that we obtain balanced problems with a non-trivial solution. In particular, the elements of c are evenly spaced on the interval $[-5,-1]$, i.e., the i-th element of c is given by $c_i = -1 - 4\frac{i-1}{n_1-1}$; the right-hand side vector $h_{\boldsymbol{\omega}}$ follows a uniform continuous distribution on $[5,15]$; and the elements of the second-stage cost vector $q_{\boldsymbol{\omega}}$ and the recourse matrix $W_{\boldsymbol{\omega}}$ follow uniform discrete distributions on $-5n_1 + [-10,0]$ and $[1,6]$, respectively. Finally, the technology matrix $T_{\boldsymbol{\omega}}$ follows a uniform two-point distribution where the outcomes $T_{\boldsymbol{\omega}}^1$ and $T_{\boldsymbol{\omega}}^2$ are $m_2 \times n_1$ matrices, whose (i,j)-th elements are given by

$$T_{\boldsymbol{\omega},i,j}^1 = \begin{cases} 1/\alpha, & \text{if } \min\{m_2 - n_1, 0\} \leq i - j \leq \max\{m_2 - n_1, 0\} \\ 0, & \text{otherwise,} \end{cases}$$

and

$$T^2_{\omega,i,j} = \begin{cases} 2/\beta, & \text{if } \min\{m_2 - n_1, 0\} \leq i - j \leq \max\{m_2 - n_1, 0\} \\ 1/\beta, & \text{otherwise,} \end{cases}$$

where

$$\alpha = 1 + \min\{n_1 - 1, |m_2 - n_1|, \max\{n_1 - m_2, 0\} + i - 1, \max\{m_2, n_1\} - i\},$$

and $\beta = \alpha + m_2$. For example, if $n_1 = m_2$, then T^1_ω is the identity matrix, and T^2_ω is the square matrix such that the off-diagonal and diagonal elements are given by $1/(m_2 + 1)$ and $2/(m_2 + 1)$, respectively. This choice ensures that the rows of T_ω sum to one, so that the resulting large IPP instances satisfy the complete recourse assumption in (A1).

In our experiments, we draw a sample of size $S \in \{100, 500\}$ from the joint distribution of $(h_\omega, T_\omega, q_\omega, W_\omega)$ and we solve the resulting sample average approximation of (5.40), which is obtained by replacing the distribution of $(h_\omega, T_\omega, q_\omega, W_\omega)$ with the empirical distribution of the sample.

Chapter 6

Summary and conclusion

A wide range of practical decision problems in, e.g., manufacturing, inventory management, and environmental control involve uncertainty. We can model such problems as two-stage mixed-integer recourse (MIR) models, which explicitly capture parameter uncertainty, and permit integer decision variables for sensible modelling. We develop solution methods for such two-stage MIR models, which are of the form

$$\min_x \{c^\top x + \mathbb{E}_\omega[v(\omega, x)] : \ Ax \geq b, \ x \in X\},$$

where $X \subset \mathbb{R}^n$ imposes possible non-negativity and integer restrictions on the decision vector x, and the second-stage cost function v is defined as

$$v(\omega, x) := \min_y \left\{ q^\top y : \ Wy \geq h(\omega) - T(\omega)x, \ y \in \mathbb{Z}_+^{p_2} \times \mathbb{R}_+^{n_2 - p_2} \right\},$$
$$\omega \in \Omega, \ x \in \mathbb{R}^n.$$

Briefly, the problem is to choose a first-stage decision x such that the resulting total expected costs, consisting of the immediate costs $c^\top x$ and the expected second-stage costs $Q(x) = \mathbb{E}_\omega[v(\omega, x)]$, are minimized. The second-stage costs $v(\omega, x)$ are defined in terms of a mixed-integer program whose parameters depend on the first-stage decision x and the random vector ω, which explicitly models parameter uncertainty. The main difficulty in solving two-stage MIR models is that the *recourse function* Q is in general non-convex, and thus we cannot directly use efficient convex optimization techniques. We overcome this difficulty by using *convexification*. That is, we derive a convex approximating model that lends itself well to

effective decomposition strategies and thus is easier to solve. Indeed, we develop computationally efficient decomposition algorithms that solve the approximating model, and moreover, we derive performance guarantees to obtain *provably* good, or even near-optimal, solutions of the original model.

In fact, we describe two *complementary* convexification approaches. Our first approach is to obtain an approximating model by replacing the recourse function Q by a convex approximation $\hat{Q}$ that closely approximates Q. That is, we construct $\hat{Q}$ such that the approximation error $\sup_x |Q(x) - \hat{Q}(x)|$ is small, and in addition, we derive a corresponding error bound. In this way, we guarantee the quality of the solution $\hat{x}$ obtained by solving the corresponding approximating model. Indeed, an error bound on the approximation error directly translates into a performance guarantee for $\hat{x}$.

If the recourse function Q is highly non-convex, however, then it does not admit a close approximation by a convex function, motivating our complementary convexification approach. The idea is to solve a convex relaxation of the original model which is defined in terms of optimality cuts that underestimate Q. In particular, we derive a family of optimality cuts whose closure asymptotically converges to the convex envelope of Q. The advantage of using such optimality cuts is that the relaxation defined in terms of this convex envelope is *exact*: we are able to obtain the optimal solution of two-stage MIR models, in general. In the remainder, we describe the main findings of each chapter separately, and subsequently, we discuss our main contributions and directions for future research.

In Chapter 2, we consider the so-called shifted LP-relaxation approximation derived by Romeijnders et al. [61] for the special class of simple integer recourse (SIR) models. In particular, we derive a *hierarchy* of corresponding error bounds, which depend on the *total variations* $|\Delta|f$ and $|\Delta|f^{(k)}$, $k \geq 1$, of the probability density function (pdf) f of the right-hand side vector $h(\omega)$, and its higher-order derivatives $f^{(k)}$. To be specific, we obtain sequentially tighter error bounds as we include more higher-order derivatives of f. In this way, we provide sharper performance guarantees for the solution obtained by solving the shifted LP-relaxation. Interestingly, we strictly improve the original error bound in [61], which depends only on $|\Delta|f$. For example, we are able to improve it by up to a factor three if we also include the total variation $|\Delta|f'$ of the first derivative f' of the pdf f. Moreover, this improvement factor approaches three if the ratio $|\Delta|f'/|\Delta|f$ is close to zero. This condition holds if, e.g., f is the density of a normally distributed random variable whose standard deviation is large.

In Chapter 3, we construct new convex approximations of the recourse function Q for general two-stage MIR models with a fixed technology matrix $T(\omega)$. The idea is to approximate the second-stage feasible regions by using *pseudo-valid* cutting planes, which are affine in the first-stage decisions x. The rationale of using affine cutting planes is that the corresponding approximating model can be solved efficiently, because the resulting approximation $\hat{v}(\omega, x)$ of $v(\omega, x)$ is convex, and thus $\hat{Q}(x) := \mathbb{E}_\omega[\hat{v}(\omega, x)]$ is convex as well. The drawback is that, in general, the second-stage cost approximation $\hat{v}(\omega, x)$ is not exact, because pseudo-valid cutting planes may either cut away feasible solutions, or they are potentially overly conservative. Our philosophy is, however, to use pseudo-valid cutting planes such that the approximation $\hat{v}(\omega, x)$ is good *on average*. Indeed, we are able to derive an error bound on the approximation error $\sup_x |Q(x) - \hat{Q}(x)|$ if the pseudo-valid cutting planes are *tight*, i.e., if they are exact on a grid of first-stage decisions. In particular, we show that the pseudo-valid cutting plane approximation $\hat{Q}$ closely approximates Q if the total variations of the underlying one-dimensional conditional pdf are small. Moreover, we derive tight pseudo-valid cutting planes for SIR models as well as for a nurse scheduling problem, and we use them to demonstrate the applicability of our approach. Indeed, we obtain good approximate solutions for an array of nurse scheduling problem instances, and, in line with our error bound, the pseudo-valid cutting planes yield better results if the variability of the random parameters in the model is larger.

In Chapter 4, we propose an alternative type of convex approximations, the so-called generalized α-approximations, and we derive a corresponding error bound, which is similar to the one in Chapter 3. A key feature of the generalized α-approximations $\hat{Q}_\alpha$ is that the computations required for a corresponding Benders' decomposition algorithm can be carried out efficiently. Indeed, we derive a family of *loose* optimality cuts, i.e., supporting hyperplanes which define a lower bound for $\hat{Q}_\alpha$, that can be computed fast, and we use them to develop the *loose Benders' decomposition algorithm* (LBDA). Our optimality cuts are called loose, because the resulting polyhedral lower bound is in general not tight for the generalized α-approximations, and as a result, the solution $\hat{x}$ identified by LBDA is not necessarily optimal in the approximating model corresponding to $\hat{Q}_\alpha$. Nevertheless, we show that our error bound for $\hat{Q}_\alpha$ carries over to the performance of $\hat{x}$ in the original two-stage MIR model. Indeed, numerical experiments confirm that we find good solutions if the underlying one-dimensional conditional pdf have small total variations.

In Chapter 5, we propose a family of optimality cuts for Q that define a corresponding convex polyhedral lower bound, and we use these *scaled cuts* to develop a Benders' decomposition algorithm that solves general two-stage MIR models. The main insight is that *non-affine* cuts for the second-stage cost functions $v(\omega, x)$ can be transformed into *affine* cuts for $Q(x) = \mathbb{E}_\omega[v(\omega, x)]$, which enables efficient decomposition. In particular, if ϕ is an arbitrary lower bound of Q, i.e., if $\phi \leq Q$, then non-affine cuts of the form

$$v(\omega, x) \geq \alpha(\omega) - \beta(\omega)^\top x - \tau(\omega)\phi(x) \quad \forall \omega \in \Omega, x \in \mathbb{R}^n,$$

where $\tau(\omega) \geq 0 \; \forall \omega \in \Omega$, yield the following scaled cut

$$Q(x) \geq \frac{\mathbb{E}_\omega[\alpha(\omega) - \beta(\omega)^\top x]}{1 + \mathbb{E}_\omega \tau(\omega)} \quad \forall x \in \mathbb{R}^n,$$

which is affine in x. We show that a recursive application of our scaled cuts, where ϕ is defined as the pointwise maximum of previously computed optimality cuts, yields the convex envelope of Q in the limit, i.e., the family of scaled cuts is sufficiently rich to identify the optimal solution. Indeed, numerical experiments indicate that we are able to obtain provably good solutions and that we consistently outperform benchmark solution methods in terms of solution quality.

Discussion and future research directions

The recurring theme of Chapters 2-5 is convexification: we propose approximate and exact techniques for convexifying the recourse function Q that are complementary. Indeed, the exact convexification techniques of Chapter 5 are well-suited if Q is highly non-convex, whereas the use of convex approximations is justified if Q can be approximated closely by a convex function. One limitation is that, a priori, it is not clear which solution method is best-suited to solve a given MIR problem instance. Therefore, an opportunity for future research is to combine the approximate and exact convexification techniques. For example, we may use the scaled cuts of Chapter 5 to reduce the approximation error of the convex approximations that we develop in Chapters 3 and 4. As a result, we potentially find better approximate solutions and correspondingly, we obtain stronger performance guarantees. Alternatively, we may compute our scaled cuts by using an approximate lower bound of Q that we obtain via approximate convexification techniques. By doing so, we speed up convergence of the scaled cuts, at the expense of introducing

approximation error, and thus the resulting Benders' decomposition may converge to suboptimal solutions.

Another approach is to determine a priori which convexification strategy is appropriate by deriving *sharp, tractable* error bounds for the convex approximations of Chapters 3 and 4: if such error bounds are small, then Q admits a close convex approximation, whereas otherwise Q is likely highly non-convex, justifying the use of exact convexification. In fact, this links to a second limitation of the results in this thesis: the error bounds that we derive in Chapters 3 and 4 are in general intractable, and they significantly overstate the actual approximation error.

In this respect, one possible direction is to extend the error bounds of Chapter 2, which apply to one-dimensional SIR model approximations, to the convex approximations of general MIR models of Chapters 3 and 4. The key insight of Chapter 2 is that the error bound in [61] for the shifted LP-relaxation of SIR models can be improved by using the total variations of the higher-order derivatives of the underlying pdf. In particular, we are able to derive stronger bounds on the expectation of special types of periodic functions, which directly translate into error bounds for the shifted LP-relaxation. Interestingly, the error bounds that we derive for the pseudo-valid cutting plane approximation and the generalized α-approximation of Chapters 3 and 4, respectively, are also obtained by proving that the underlying difference function is asymptotically periodic. Thus, it remains to investigate if the higher-order total variation bounds on the expectation of periodic functions in Chapter 2 can be generalized to the more general periodic functions that arise as difference functions in Chapters 3 and 4.

Another avenue for future research is to construct a tight family of pseudo-valid cutting planes for general MIR models. In Chapter 3, we derive such cutting planes for a nurse scheduling problem, as well as the special class of SIR models, but they are not available in general. In this sense, the loose Benders' decomposition algorithm of Chapter 4 complements the pseudo-valid cutting plane methodology, as we can use LBDA if tight pseudo-valid cutting planes are not available. Thus, a direct comparison of the two approaches would be interesting.

Finally, it is worthwhile to investigate if the proposed convexification techniques can be extended to more general types of MIR models, such as multi-stage models, or models that feature non-linear, e.g., conic or quadratic, cost functions.

Bibliography

[1] S. Ahmed. A scenario decomposition algorithm for 0–1 stochastic programs. *Operations Research Letters*, 41(6):565–569, 2013.

[2] S. Ahmed and R. Garcia. Dynamic capacity acquisition and assignment under uncertainty. *Annals of Operations Research*, 124(1-4):267–283, 2003.

[3] S. Ahmed and A. Shapiro. The sample average approximation method for stochastic programs with integer recourse. 2002. Available from Optimization Online.

[4] S. Ahmed, M. Tawarmalani, and N. V. Sahinidis. A finite branch-and-bound algorithm for two-stage stochastic integer programs. *Mathematical Programming*, 100(2):355–377, 2004.

[5] S Ahmed, R. Garcia, N. Kong, L. Ntaimo, G. Parija, F. Qiu, and S. Sen. SIPLIB: A Stochastic Integer Programming Test Problem Library. `https://www2.isye.gatech.edu/~sahmed/siplib`, 2015.

[6] S. Ahmed, F. G. Cabral, and B. F. P. da Costa. Stochastic Lipschitz dynamic programming. *Mathematical Programming*, 2020. DOI: 10.1007/s10107-020-01569-z.

[7] G. Anger. *Inverse problems in differential equations*. Akademie-Verlag, 1990.

[8] G. Angulo, S. Ahmed, and S. S. Dey. Improving the integer L-shaped method. *INFORMS Journal on Computing*, 28(3):483–499, 2016.

[9] E. Balas and R. G. Jeroslow. Strengthening cuts for mixed integer programs. *European Journal of Operational Research*, 4(4):224–234, 1980.

[10] E. Balas and M. Perregaard. A precise correspondence between lift-and-project cuts, simple disjunctive cuts, and mixed integer Gomory cuts for 0-1 programming. *Mathematical Programming*, 94(2-3):221–245, 2003.

[11] E. Balas, S. Ceria, and G. Cornuéjols. Mixed 0-1 programming by lift-and-project in a branch-and-cut framework. *Management Science*, 42(9):1229–1246, 1996.

[12] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer. *Non-linear parametric optimization*. Springer, 1982.

[13] M. Bansal, K.-L. Huang, and S. Mehrotra. Tight second stage formulations in two-stage stochastic mixed integer programs. *SIAM Journal on Optimization*, 28(1):788–819, 2018.

[14] G. Bayraksan and D. P. Morton. Assessing solution quality in stochastic programs. *Mathematical Programming*, 108(2-3):495–514, 2006.

[15] J. F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4(1):238–252, 1962.

[16] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, New York, 1997.

[17] J. R. Birge and F. V. Louveaux. A multicut algorithm for two-stage stochastic linear programs. *European Journal of Operational Research*, 34(3):384–392, 1988.

[18] C. E. Blair and R. G. Jeroslow. The value function of a mixed integer program: I. *Discrete Mathematics*, 19(2):121–138, 1977.

[19] C. E. Blair and R. G. Jeroslow. The value function of a mixed integer program II. *Discrete Mathematics*, 25(1):7–19, 1979.

[20] M. Bodur, S. Dash, O. Günlük, and J. Luedtke. Strengthened Benders cuts for stochastic integer programs with continuous recourse. *INFORMS Journal on Computing*, 29(1):77–91, 2017.

[21] N. Boland, J. Christiansen, B. Dandurand, A. Eberhard, J. Linderoth, J. Luedtke, and F. Oliveira. Combining progressive hedging with a Frank–Wolfe method to compute Lagrangian dual bounds in stochastic mixed-integer programming. *SIAM Journal on Optimization*, 28(2):1312–1336, 2018.

[22] E. A. Boyd. Fenchel cutting planes for integer programs. *Operations Research*, 42(1):53–64, 1994.

[23] E. A. Boyd. On the convergence of Fenchel cutting planes in mixed-integer programming. *SIAM Journal on Optimization*, 5(2):421–435, 1995.

[24] C. C. Carøe and R. Schultz. Dual decomposition in stochastic integer programming. *Operations Research Letters*, 24(1-2):37–45, 1999.

[25] C. C. Carøe and J. Tind. A cutting-plane approach to mixed 0–1 stochastic integer programs. *European Journal of Operational Research*, 101(2):306–316, 1997.

[26] C. C. Carøe and J. Tind. L-shaped decomposition of two-stage stochastic programs with integer recourse. *Mathematical Programming*, 83(1-3):451–464, 1998.

[27] W. Cook, A.M.H. Gerards, A. Schrijver, and É. Tardos. Sensitivity theorems in integer linear programming. *Mathematical Programming*, 34(3):251–264, 1986.

[28] A. Del Pia and R. Weismantel. Relaxations of mixed integer sets from lattice-free polyhedra. *Annals of Operations Research*, 240(1):95–117, 2016.

[29] M. Dyer and L. Stougie. Computational complexity of stochastic programming problems. *Mathematical Programming*, 106(3):423–432, 2006.

[30] J. E. Falk. Lagrange multipliers and nonconvex programs. *SIAM Journal on Control*, 7(4):534–545, 1969.

[31] D. Gade, S. Küçükyavuz, and S. Sen. Decomposition algorithms with parametric Gomory cuts for two-stage stochastic integer programs. *Mathematical Programming*, 144(1-2):39–64, 2014.

[32] H. I. Gassmann and W. T. Ziemba, editors. *Stochastic Programming: Applications in Finance, Energy, Planning and Logistics*, volume 4 of *World Scientific Series in Finance*. World Scientific, 2013.

[33] A. Georghiou, A. Tsoukalas, and W. Wiesemann. A primal–dual lifting scheme for two-stage robust optimization. *Operations Research*, 68(2):572–590, 2020.

[34] R. E. Gomory. Outline of an algorithm for integer solutions to linear programs. *Bulletin of the American Mathematical Society*, 64(5):275–278, 1958.

[35] Y. Guan, S. Ahmed, and G.L. Nemhauser. Cutting planes for multistage stochastic integer programs. *Operations Research*, 57(2):287–298, 2009.

[36] G. Guo, G. Hackebeil, S. M. Ryan, J.-P. Watson, and D. L. Woodruff. Integration of progressive hedging and dual decomposition in stochastic integer programs. *Operations Research Letters*, 43(3):311–316, 2015.

[37] Y. Han and Z. Chen. Quantitative stability of full random two-stage stochastic programs with recourse. *Optimization Letters*, 9(6):1075–1090, 2015.

[38] G. A. Hanasusanto, D. Kuhn, and W. Wiesemann. A comment on "computational complexity of stochastic programming problems". *Mathematical Programming*, 159(1-2):557–569, 2016.

[39] S. Jorjani, C. H. Scott, and D. L. Woodruff. Selection of an optimal subset of sizes. *International Journal of Production Research*, 37(16):3697–3710, 1999.

[40] K. Kim and S. Mehrotra. A two-stage stochastic integer programming approach to integrated staffing and scheduling with application to nurse management. *Operations Research*, 63(6):1431–1451, 2015.

[41] W. K. Klein Haneveld, L. Stougie, and M. H. van der Vlerk. Simple integer recourse models: convexity and convex approximations. *Mathematical Programming*, 108(2-3):435–473, 2006.

[42] A. J. Kleywegt, A. Shapiro, and T. Homem-de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002.

[43] S. Küçükyavuz and S. Sen. An introduction to two-stage stochastic mixed-integer programming. In *INFORMS TutORials in Operations Research*, pages 1–27. INFORMS, 2017.

[44] G. Laporte and F. V. Louveaux. The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters*, 13(3): 133–142, 1993.

[45] F.V. Louveaux and M.H. van der Vlerk. Stochastic programming with simple integer recourse. *Mathematical Programming*, 61(1-3):301–325, 1993.

[46] W. K. Mak, D. P. Morton, and R. K. Wood. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(1-2):47–56, 1999.

[47] H. Marchand, A. Martin, R. Weismantel, and L. Wolsey. Cutting planes in integer and mixed integer programming. *Discrete Applied Mathematics*, 123(1-3):397–446, 2002.

[48] L. Ntaimo. Disjunctive decomposition for two-stage stochastic mixed-binary programs with random recourse. *Operations Research*, 58(1):229–243, 2010.

[49] L. Ntaimo. Fenchel decomposition for stochastic mixed-integer programming. *Journal of Global Optimization*, 55(1):141–163, 2013.

[50] L. Ntaimo and S. Sen. The million-variable "march" for stochastic combinatorial optimization. *Journal of Global Optimization*, 32(3):385–400, 2005.

[51] L. Ntaimo and M. W. Tanner. Computations with disjunctive cuts for two-stage stochastic mixed 0-1 integer programs. *Journal of Global Optimization*, 41 (3):365–384, 2008.

[52] M. Perregaard and E. Balas. Generating cuts from multiple-term disjunctions. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 348–360. Springer, 2001.

[53] Y. Qi and S. Sen. The ancestral Benders' cutting plane algorithm with multi-term disjunctions for mixed-integer recourse decisions in stochastic programming. *Mathematical Programming*, 161(1-2):193–235, 2017.

[54] A. H. G. Rinnooy Kan and L. Stougie. Stochastic integer programming. In Yu. Ermoliev and R. J.-B. Wets, editors, *Numerical Techniques for Stochastic Optimization*, volume 10 of *Springer Series in Computational Mathematics*, pages 201–213. Springer, 1988.

[55] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[56] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften*. Springer, 2009.

[57] W. Romeijnders and N. van der Laan. Pseudo-valid cutting planes for two-stage mixed-integer stochastic programs with right-hand-side uncertainty. *Operations Research*, 68(4):1199–1217, 2020.

[58] W. Romeijnders, L. Stougie, and M. H. van der Vlerk. Approximation in two-stage stochastic integer programming. *Surveys in Operations Research and Management Science*, 19(1):17–33, 2014.

[59] W. Romeijnders, M. H. van der Vlerk, and W. K. Klein Haneveld. Convex approximations of totally unimodular integer recourse models: A uniform error bound. *SIAM Journal on Optimization*, 25(1):130–158, 2015.

[60] W. Romeijnders, R. Schultz, M. H. van der Vlerk, and W. K. Klein Haneveld. A convex approximation for two-stage mixed-integer recourse models with a uniform error bound. *SIAM Journal on Optimization*, 26(1):426–447, 2016.

[61] W. Romeijnders, M. H. van der Vlerk, and W. K. Klein Haneveld. Total variation bounds on the expectation of periodic functions with applications to recourse approximations. *Mathematical Programming*, 157(1):3–46, 2016.

[62] W. Romeijnders, D. Morton, and M. H. van der Vlerk. Assessing the quality of convex approximations for two-stage totally unimodular integer recourse models. *INFORMS Journal on Computing*, 29(2):211–231, 2017.

[63] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill New York, 1976.

[64] A. Ruszczyński. A regularized decomposition method for minimizing a sum of polyhedral functions. *Mathematical Programming*, 35(3):309–333, 1986.

[65] R. Schultz. On structure and stability in stochastic programs with random technology matrix and complete integer recourse. *Mathematical Programming*, 70(1-3):73–89, 1995.

[66] R. Schultz. Stochastic programming with integer variables. *Mathematical Programming*, 97(1-2):285–309, 2003.

[67] R. Schultz, L. Stougie, and M. H. van der Vlerk. Solving stochastic programs with integer recourse by enumeration: A framework using Gröbner basis reductions. *Mathematical Programming*, 83(1-3):229–252, 1998.

[68] S. Sen. Algorithms for stochastic mixed-integer programming models. *Handbooks in Operations Research and Management Science*, 12:515–558, 2005.

[69] S. Sen and J. L. Higle. The C^3 theorem and a D^2 algorithm for large scale stochastic mixed-integer programming: Set convexification. *Mathematical Programming*, 104(1):1–20, 2005.

[70] S. Sen and H. D. Sherali. Decomposition with branch-and-cut approaches for two-stage stochastic mixed-integer programming. *Mathematical Programming*, 106(2):203–223, 2006.

[71] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia, 2009.

[72] H. D. Sherali and B. M. P. Fraticelli. A modification of Benders' decomposition algorithm for discrete subproblems: An approach for stochastic programs with integer recourse. *Journal of Global Optimization*, 22(1-4):319–342, 2002.

[73] H. D. Sherali and X. Zhu. On solving discrete two-stage stochastic programs having mixed-integer first-and second-stage variables. *Mathematical Programming*, 108(2-3):597–616, 2006.

[74] F. Tardella. On the existence of polyhedral convex envelopes. In C. A. Floudas and P. M. Pardalos, editors, *Frontiers in Global Optimization*, pages 563–573. Springer, 2004.

[75] M. Tawarmalani and N. V. Sahinidis. Convex extensions and envelopes of lower semi-continuous functions. *Mathematical Programming*, 93(2):247–263, 2002.

[76] A. C. Trapp, O. A. Prokopyev, and A. J. Schaefer. On a level-set characterization of the value function of an integer program and its application to stochastic programming. *Operations Research*, 61(2):498–511, 2013.

[77] N. van der Laan and W. Romeijnders. A loose Benders decomposition algorithm for approximating two-stage mixed-integer recourse models. *Mathematical Programming*, 2020. DOI: 10.1007/s10107-020-01559-1.

[78] N. van der Laan and W. Romeijnders. A converging Benders' decomposition algorithm for two-stage mixed-integer recourse models. *Accepted for publication in Operations Research*, 2021. Available from Optimization Online.

[79] N. van der Laan, W. Romeijnders, and M. H. van der Vlerk. Higher-order total variation bounds for expectations of periodic functions and simple integer recourse approximations. *Computational Management Science*, 15(3-4):325–349, 2018.

[80] M. H. van der Vlerk. *Stochastic Programming with Integer Recourse*. PhD thesis, University of Groningen, The Netherlands, 1995.

[81] M. H. van der Vlerk. Convex approximations for complete integer recourse models. *Mathematical Programming*, 99(2):297–310, 2004.

[82] M. H. van der Vlerk. On multiple simple recourse models. *Mathematical Methods of Operations Research*, 62(2):225–242, 2005.

[83] M. H. Van der Vlerk. Convex approximations for a class of mixed-integer recourse models. *Annals of Operations Research*, 177(1):139–150, 2010.

[84] R. M. van Slyke and R. Wets. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17(4):638–663, 1969.

[85] D. W. Walkup and R. J.-B. Wets. Lifting projections of convex polyhedra. *Pacific Journal of Mathematics*, 28(2):465–475, 1969.

[86] S. W. Wallace and W. T. Ziemba, editors. *Applications of Stochastic Programming*. MPS-SIAM Series on Optimization. 2005.

[87] B. Zeng and L. Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41(5): 457–461, 2013.

[88] M. Zhang and S. Küçükyavuz. Finitely convergent decomposition algorithms for two-stage stochastic pure integer programs. *SIAM Journal on Optimization*, 24(4):1933–1951, 2014.

[89] J. Zou, S. Ahmed, and X. A. Sun. Stochastic dual dynamic integer programming. *Mathematical Programming*, 175(1-2):461–502, 2019.

Samenvatting en conclusie (summary and conclusion in Dutch)

Een breed scala aan praktische beslissingsproblemen met betrekking tot, bijvoorbeeld, productie, voorraadbeheer, en milieubeheer brengt onzekerheid met zich mee. Dergelijke problemen kunnen worden gemodelleerd als twee-stadia gemengd geheeltallige recourse (mixed-integer recourse, MIR) modellen, die expliciet parameteronzekerheid modelleren en geheeltallige beslissingsvariabelen toestaan om tot een steekhoudend model te komen. We ontwikkelen oplossingsmethoden voor dergelijke twee-stadia MIR modellen, die geformuleerd zijn als

$$\min_x \{c^\top x + \mathbb{E}_\omega[v(\omega, x)] : \ Ax \geq b, \ x \in X\},$$

waar $X \subset \mathbb{R}^n$ niet-negativiteits- en geheeltalligheidsrestricties kan opleggen op de beslissingsvector x, en waar de tweedestadiumkostenfunctie v is gedefinieerd als

$$v(\omega, x) := \min_y \left\{ q^\top y : \ Wy \geq h(\omega) - T(\omega)x, \ y \in \mathbb{Z}_+^{p_2} \times \mathbb{R}_+^{n_2 - p_2} \right\}, \\ \omega \in \Omega, \ x \in \mathbb{R}^n.$$

In het kort is het probleem om de eerstestadiumbeslissing x zodanig te kiezen dat de resulterende totale verwachte kosten, bestaande uit de directe kosten $c^\top x$ en de verwachte tweedestadiumkosten $Q(x) = \mathbb{E}_\omega[v(\omega, x)]$, worden geminimaliseerd. De tweedestadiumkosten $v(\omega, x)$ zijn gedefinieerd in termen van een gemengd geheeltallig programmeringsprobleem, waarvan de parameters afhangen van de eerstestadiumbeslissing x en de stochastische vector ω, die expliciet para-

meteronzekerheid modelleert. De voornaamste moeilijkheid in het oplossen van twee-stadia MIR modellen is dat de *recourse functie* Q in het algemeen niet convex is, waardoor we niet direct efficiënte technieken van convexe optimalisatie kunnen toepassen. We omzeilen deze moeilijkheid door middel van *convexificatie*. Dat wil zeggen, we leiden een convex benaderingsmodel af dat zich goed leent voor effectieve decompositietechnieken en dus makkelijker op te lossen is. Dit stelt ons in staat om computationeel efficiënte decompositie-algoritmes te ontwikkelen die het benaderingsmodel oplossen. Bovendien leiden we kwaliteitsgaranties af om zo *aantoonbaar* goede, of zelfs bijna-optimale, oplossingen te verkrijgen voor het oorspronkelijke model.

In feite beschrijven we twee complementaire aanpakken gebaseerd op convexificatie. Onze eerste aanpak is om een benaderingsmodel te verkrijgen door de recourse functie Q te vervangen door een convexe benadering $\hat{Q}$ die Q dicht benadert. Met andere woorden, we construeren $\hat{Q}$ zodanig dat de benaderingsfout $\sup_x |Q(x) - \hat{Q}(x)|$ klein is en we leiden een bijbehorende bovengrens af. Op deze manier garanderen we de kwaliteit van de oplossing $\hat{x}$ die verkregen is door het benaderingsmodel op te lossen. Een bovengrens op de benaderingsfout vertaalt zich inderdaad direct naar een kwaliteitsgarantie voor $\hat{x}$.

Als de recourse functie Q echter sterk niet-convex is, dan kan Q niet goed benaderd worden door een convexe functie, hetgeen onze complementaire aanpak motiveert. Het idee is om een convexe relaxatie van het oorspronkelijke model op te lossen, die gedefinieerd is in termen van *optimaliteitssneden* (optimality cuts) die Q van beneden begrenzen. Om precies te zijn, we leiden een familie van optimaliteitssneden af waarvan de afsluiting, oftewel het puntsgewijze maximum, in asymptotische zin convergeert naar de convexe omhulling van Q. Het voordeel van deze optimaliteitssneden is dat de relaxatie gedefinieerd in termen van deze convexe omhulling *exact* is, wat ons in staat stelt om de optimale oplossing te verkrijgen van algemene twee-stadia MIR modellen. In het resterende gedeelte beschrijven we de voornaamste bevindingen van elk hoofdstuk afzonderlijk, om daarna onze hoofdbijdragen en richtingen voor toekomstig onderzoek te bespreken.

In Hoofdstuk 2 beschouwen we de zogenoemde verschoven LP-relaxatiebenadering (shifted LP-relaxation approximation) afgeleid door Romeijnders et al. [61] voor de speciale klasse van enkelvoudig geheeltallige recourse (simple integer recourse, SIR) modellen. In het bijzonder, we leiden een hiërarchie van bijbehorende foutbegrenzingen af, die afhangen van de totale variaties $|\Delta|f$ en $|\Delta|f^{(k)}$, $k \geq 1$, van de kansdichtheidsfunctie f van het rechterlid $h(\omega)$, en de hogere-orde afgeleiden $f^{(k)}$ van f. Om specifiek te zijn, we verkrijgen successievelijk sterkere foutbegrenzingen naarmate we meer hogere orde afgeleiden van f meenemen. Op die manier

leiden we scherpere kwaliteitsgaranties af voor de oplossing verkregen door de verschoven LP-relaxatie op te lossen. Een interessante observatie is dat we de oorspronkelijke foutbegrenzing in [61], die enkel afhangt van $|\Delta|f$, strikt verbeteren. Bijvoorbeeld, onze foutbegrenzing is tot een factor drie kleiner als we ook de totale variatie $|\Delta|f'$ van de eerste afgeleide f' van f meenemen. Bovendien nadert deze verbeteringsfactor drie naarmate de verhouding $|\Delta|f'/|\Delta|f$ nagenoeg nul is. Aan deze voorwaarde is voldaan als, bijvoorbeeld, f de kansdichtheidsfunctie is van een normaal verdeelde stochast met grote standaardafwijking.

In Hoofdstuk 3 construeren we nieuwe convexe benaderingen van de recourse functie Q voor algemene twee-stadia MIR modellen waarvan de technologie matrix $T(\omega)$ deterministisch is. Het idee is om het toegelaten gebied van het tweedestadiumprobleem af te schatten door middel van pseudovalide sneden (cutting planes), die affien zijn in de eerstestadiumbeslissingen x. De onderliggende redenering is dat affiene sneden leiden tot een benaderingsmodel dat efficiënt kan worden opgelost, omdat de resulterende benadering $\hat{v}(\omega, x)$ van $v(\omega, x)$ convex is, en daarmee ook $\hat{Q}(x) = \mathbb{E}_\omega[\hat{v}(\omega, x)]$. Het nadeel is dat de benadering $\hat{v}(\omega, x)$ van de tweedestadiumkosten in het algemeen niet exact is, omdat pseudovalide sneden mogelijk toegelaten oplossingen wegsnijden, of overmatig conservatief kunnen zijn. Onze gedachtegang is, echter, om pseudovalide sneden te gebruiken zodanig dat de benadering $\hat{v}(\omega, x)$ *gemiddeld genomen* goed is. Hierdoor zijn we in staat om een bovengrens af te leiden op de benaderingsfout $\sup_x |Q(x) - \hat{Q}(x)|$, mits de pseudovalide sneden *scherp* zijn, oftewel als ze exact zijn op een rooster van eerstestadiumbeslissingen. In het bijzonder laten we zien dat de benadering $\hat{Q}$ gebaseerd op pseudovalide sneden een goede benadering is van Q als de totale variaties van de onderliggende ééndimensionale voorwaardelijke kansdichtheidsfuncties klein zijn. Bovendien leiden we pseudovalide sneden af voor SIR modellen en voor een probleem aangaande het roosteren van verplegers, en we gebruiken ze om de toepasbaarheid van onze aanpak aan te tonen. We verkrijgen inderdaad benaderingsoplossingen van hoge kwaliteit voor een scala aan instanties van het roosterprobleem. Daarnaast observeren we dat, in lijn met onze foutbegrenzing, de pseudovalide sneden beter werken als de variabiliteit van de stochastische parameters in het probleem groter is.

In Hoofdstuk 4 stellen we een alternatieve convexe benadering voor, de zogenoemde gegeneraliseerde α-benaderingen, en leiden we een bijbehorende foutbegrenzing af, die vergelijkbaar is met de foutbegrenzing uit Hoofdstuk 3. Een belangrijke eigenschap van de gegeneraliseerde α-benaderingen $\hat{Q}_\alpha$ is dat de berekeningen die nodig zijn in een bijbehorend Benders' decompositie-algoritme efficiënt kunnen worden uitgevoerd. Dit blijkt uit het feit dat we een familie van snel uit-

rekenbare *losse* optimaliteitssneden afleiden, oftewel ondersteunende hypervlakken die een ondergrens definiëren voor $\hat{Q}_\alpha$, en dat we die gebruiken om het *losse Benders' decompositie-algoritme* (LBDA) te ontwikkelen. We noemen onze optimaliteitssneden los, omdat de resulterende polyhedrale ondergrens voor de gegeneraliseerde α-benaderingen $\hat{Q}_\alpha$ in het algemeen niet scherp is. Als gevolg daarvan is de oplossing $\hat{x}$ niet per se optimaal in het bijbehorende benaderingsmodel. Desondanks laten we zien dat onze foutbegrenzing voor $\hat{Q}_\alpha$ ook geldt voor de kwaliteit van $\hat{x}$ in het oorspronkelijke model. Numerieke experimenten bevestigen dat we inderdaad goede oplossingen verkrijgen als de totale variaties van de onderliggende ééndimensionale voorwaardelijke kansdichtheidsfuncties klein zijn.

In Hoofdstuk 5 beschrijven we een familie van optimaliteitssneden die een convex polyhedrale ondergrens voor Q definiëren, en we gebruiken deze *geschaalde sneden* (scaled cuts) om een Benders' decompositie-algoritme te ontwikkelen dat algemene twee-stadia MIR modellen oplost. Het belangrijkste inzicht is dat *niet-affiene* cuts voor de tweedestadiumkostenfuncties $v(\omega, x)$ vertaald kunnen worden naar *affiene* cuts voor $Q(x) = \mathbb{E}_\omega[v(\omega, x)]$, hetgeen efficiënte decompositie mogelijk maakt. In het bijzonder, als ϕ een willekeurige ondergrens is voor Q, oftewel als $\phi \leq Q$, dan verkrijgen we uit niet-affiene cuts van de vorm

$$v(\omega, x) \geq \alpha(\omega) - \beta(\omega)^\top x - \tau(\omega)\phi(x) \quad \forall \omega \in \Omega, x \in \mathbb{R}^n,$$

waar $\tau(\omega) \geq 0 \; \forall \omega \in \Omega$, de volgende scaled cut

$$Q(x) \geq \frac{\mathbb{E}_\omega[\alpha(\omega) - \beta(\omega)^\top x]}{1 + \mathbb{E}_\omega \tau(\omega)} \quad \forall x \in \mathbb{R}^n,$$

welke affien is in x. We bewijzen dat een recursieve toepassing van onze geschaalde sneden, waar ϕ gedefinieerd is als het puntsgewijze maximum van eerder berekende geschaalde sneden, in de limiet convergeert naar de convexe omhulling van Q. Met andere woorden, de familie van geschaalde sneden is voldoende rijk om de optimale oplossing te identificeren. Numerieke experimenten laten inderdaad zien dat we aantoonbaar goede oplossingen verkrijgen, en dat we systematisch betere resultaten behalen vergeleken met benchmarkmethodes in termen van oplossingskwaliteit.

Discussie en richtingen voor vervolgonderzoek

Het terugkerende thema van Hoofdstukken 2-5 is convexificatie: we beschrijven exacte en inexacte technieken om de recourse functie Q te convexificeren, die com-

plementair zijn. In het bijzonder zijn de exacte convexificatietechnieken uit Hoofdstuk 5 geschikt als Q sterk niet-convex is, waar het gebruik van convexe benaderingen gerechtvaardigd is als Q dicht kan worden benaderd door een convexe functie. Een beperking is dat het a priori niet duidelijk is welke oplossingsmethode het meest geschikt om een gegeven MIR probleeminstantie op te lossen. Een richting voor vervolgonderzoek is daarom om de exacte en inexacte technieken te combineren. Bijvoorbeeld, we kunnen de geschaalde sneden van Hoofdstuk 5 gebruiken om de benaderingsfout van de convexe benaderingen uit Hoofdstukken 3 en 4 te verkleinen. Hierdoor vinden we potentieel betere benaderingsoplossingen en daarmee verkrijgen we betere bijbehorende kwaliteitsgaranties. Een alternatief is om onze geschaalde sneden te berekenen met een inexacte ondergrens voor Q die verkregen is door middel van inexacte benaderingstechnieken. Op deze manier versnellen we de convergentie van de geschaalde sneden, ten koste van het meebrengen van een eventuele benaderingsfout, waardoor de resulterende Benders' decompositie naar een suboptimale oplossing kan convergeren.

Een andere aanpak is om a priori te bepalen welke convexificatiestrategie het meest geschikt is door scherpe, uitrekenbare foutbegrenzingen af te leiden voor de convexe benaderingen uit Hoofdstukken 3 en 4. Inderdaad, als dergelijke foutbegrenzingen klein zijn, dan laat Q een goede convexe benadering toe, waar anders Q waarschijnlijk sterk niet-convex is, hetgeen exacte convexificatietechnieken rechtvaardigt. Dit relateert aan een andere beperking van de resultaten in dit proefschrift, namelijk dat de foutbegrenzingen op de benaderingsfout die we afleiden in Hoofdstukken 3 en 4 in het algemeen niet uit te rekenen zijn, en de daadwerkelijke benaderingsfout significant overschatten.

In dit opzicht is een mogelijke richting om de foutbegrenzing uit Hoofdstuk 2, die van toepassing zijn op ééndimensionale SIR modellen, te generaliseren naar de convexe benaderingen van algemene MIR modellen uit Hoofdstukken 3 en 4. Het belangrijkste inzicht van Hoofdstuk 2 is dat de foutbegrenzing in [61] voor de verschoven LP-relaxatiebenadering van SIR modellen verbeterd kan worden door de totale variaties van hogere order afgeleiden van de onderliggende kansdichtheidsfuncties te gebruiken. In het bijzonder, we zijn in staat om sterkere grenzen af te leiden op de verwachting van speciale soorten periodieke functies, die direct vertaald kunnen worden naar foutbegrenzingen voor de verschoven LP-relaxatie. Een interessante observatie is dat de foutbegrenzingen die we afleiden voor de benadering gebaseerd op pseudovalide sneden en de gegeneraliseerde α-benaderingen uit Hoofdstukken 3 en 4 ook zijn verkregen door te bewijzen dat de onderliggende verschilfunctie asymptotisch periodiek is. Een relevante open vraag is daarom of de grenzen op de verwachting van periodieke functies in termen van hogere orde

totale variaties uit Hoofdstuk 2 gegeneraliseerd kunnen worden naar de meer algemene periodieke functies die naar voren komen als verschilfuncties in Hoofdstukken 3 en 4.

Een andere richting voor vervolgonderzoek is om een familie van scherpe pseudovalide sneden te construeren voor algemene MIR modellen. In Hoofdstuk 3 leiden we dergelijke sneden af voor een probleem aangaande het roosteren van verplegers, alsmede voor de speciale klasse van SIR modellen, maar ze zijn niet voorhanden in het algemene geval. In dit opzicht complementeert het losse Benders' decompositie-algoritme uit Hoofdstuk 4 de methodologie van pseudovalide sneden, daar we LBDA kunnen gebruiken als pseudovalide sneden niet voorhanden zijn. Een directe vergelijking van de twee aanpakken zou daarom interessant zijn.

Ten slotte is het de moeite waard om te onderzoeken of de voorgestelde convexificatietechnieken uitgebreid kunnen worden naar meer algemene klassen van MIR modellen, zoals multi-stadia modellen, of modellen met niet-lineare, bijvoorbeeld conische of kwadratische, kostenfuncties.

Acknowledgements

This thesis marks the end of a three-year period which I spent as a PhD candidate at the University of Groningen. There are a number of people that I would like to thank for supporting me during that period, and for contributing to the chapters of this thesis.

I will always be grateful to my supervisors Ward Romeijnders and Ruud Teunter. Ward sparked my interest in research at an early stage of my studies, and since that moment, he has guided me for over six years during various courses, research assistantships, and finally, my PhD trajectory. During that trajectory, he has been an excellent supervisor. I have always greatly enjoyed our discussions about research, education, and other related matters, and I hope we can continue them in the future as well. I am grateful to Ruud for his outstanding supervision during the past three years, for keeping an eye on the bigger picture, and for taking an interest in stochastic optimization.

I would like to express my sincere gratitude towards the members of my Assessment committee, consisting of Professors Kees Jan Roodbergen, Rüdiger Schultz, and Marc Uetz for the time and effort they spent reading this thesis. I am also indebted to Professor Wim Klein Haneveld, who read and commented on the manuscripts that were previous versions of the chapters that comprise this thesis. I am grateful to Ward for passing on these comments to me.

I am grateful to all colleagues and fellow students that I have met during my time at the University of Groningen. My officemate Ruben van Beesten deserves a special mention here, for agreeing to be my paranymph, together with my brother Lars. Moreover, Ruben is probably one of the few persons who understood quite clearly the topic of my research, as well as the inherent difficulties of explaining it to others. Indeed, I have probably never once been able to clearly explain the topic of my research to my family and friends, who nonetheless have been a continuous support. My thanks thus goes out all of them, in particular my parents Jan and Anja, grandparents Albert and Niesje, and siblings Kirsten and Lars. I should not

leave out our deeply loved dog Floor, for being an excellent listener. Finally, I am eternally grateful to Marjolein for being there from start to finish, and beyond.

Niels van der Laan