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

Niels van der Laan[*†] Ward Romeijnders[*‡]

January 4, 2021

Abstract

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, and thus, e.g., do not require that the first-stage variables are binary. 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.

Keywords: stochastic programming, mixed-integer recourse, Benders' decomposition

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) models (Wallace and Ziemba 2005, Gassmann and Ziemba 2013), but these models are notoriously hard to solve (Dyer and Stougie 2006). Typically, MIR models are solved using decomposition algorithms inspired by Benders' decomposition (Benders 1962, Küçükyavuz and Sen 2017). 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 paper, we develop a *tractable* 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* 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 by Van Slyke and Wets (1969), 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

[*]Department of Operations, University of Groningen, Groningen, The Netherlands
[†]n.van.der.laan@rug.nl
[‡]w.romeijnders@rug.nl

if the first-stage decision variables are pure binary. In fact, there is an array of decomposition algorithms developed for this special case (Laporte and Louveaux 1993, Sherali and Fraticelli 2002, Sen and Higle 2005, Sen and Sherali 2006, Ntaimo and Tanner 2008, Ntaimo 2010, 2013, Gade et al. 2014, Angulo et al. 2016, Qi and Sen 2017, Zou et al. 2019). 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 Carøe and Schultz (1999) and Sherali and Zhu (2006). A notable exception is the algorithm by Zhang and Küçükyavuz (2014) 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 (Carøe and Schultz 1999, Ahmed et al. 2004, Sherali and Zhu 2006), or they introduce auxiliary first-stage integer decision variables (Carøe and Tind 1998, Ahmed et al. 2020).

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, we exploit the definition of the current outer approximation to update the outer approximation from one iteration to the next. 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 *scaling* the non-linear optimality cuts, we obtain linear cuts for the *expected* second-stage cost function, which we refer to as scaled cuts.

We are able to efficiently compute our scaled cuts, by exploiting ideas from robust optimization and deterministic mixed-integer programming. Moreover, we prove that scaled cuts are able to 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 tractable. 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 an investment planning problem (IPP) by Schultz et al. (1998) and the DCAP problem instances (Ahmed and Garcia 2003) from SIPLIB (Ahmed et al. 2015), as well as variants of both problems. Our results show that scaled cuts outperform traditional optimality cuts, in the sense that we are able to significantly reduce the optimality gap at the root node of the Benders' master problem. Indeed, on the IPP instances and the DCAP instances, we respectively achieve an average 92% and 51% reduction of the root node gap compared to traditional optimality cuts, and, moreover, we achieve a zero root node 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 tractable Benders' decomposition algorithm which solves 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 paper is organized as follows. In Section 2, we formally introduce MIR models and review solution approaches. Next, we introduce scaled cuts and develop our Benders' decomposition algorithm in Section 3, and we describe several strategies to compute scaled cuts in Section 4. Section 5 concerns the proof of convergence of the scaled cuts. We report on our numerical experiments in Section 6, and we conclude in Section 7.

Notation: Throughout, $\mathrm{conv}(A)$ denotes the convex hull of a set A. For a function $f : A \mapsto \mathbb{R} \cup \{\infty\}$, we define its convex envelope $\mathrm{co}(f) : \mathrm{conv}(A) \mapsto \mathbb{R}$ and its closed convex envelope $\overline{\mathrm{co}}(f) : \mathrm{conv}(A) \mapsto \mathbb{R}$ as the pointwise supremum of all convex, respectively affine, functions majorized by f. In addition, we define $\mathrm{dom}(f) = \{x \in A : f(x) < \infty\}$. Finally, for any $B \subseteq A$, we denote by $\mathrm{epi}_{\mathrm{B}}(f)$ the epi-graph of f restricted to B, i.e., $\mathrm{epi}_{\mathrm{B}}(f) := \{(x, \theta) \in B \times \mathbb{R} : \theta \geq f(x)\}$, and we write $\mathrm{epi}(f) = \mathrm{epi}_{\mathrm{A}}(f)$.

2 Problem Description and Literature Review

2.1 Problem Description

Two-stage recourse models explicitly model parameter uncertainty by a random vector ω 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 ω, referred to as a scenario. We assume that the probability distribution of ω 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

$$\eta^* := \min_x \{c^\top x + \mathbb{E}_\omega[v_\omega(x)] : \; Ax = b, \; x \in \mathcal{X}\}, \tag{1}$$

where the second-stage costs $v_\omega(x)$ are defined as

$$v_\omega(x) := \min_y \{q_\omega^\top y : \; W_\omega y = h_\omega - T_\omega x, \; y \in \mathcal{Y}\}, \quad x \in X, \; \omega \in \Omega, \tag{2}$$

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 ω 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 $\mathbb{P}(\omega = \omega')$ is rational.

Assumption (A1) is known as relatively complete and sufficiently expensive recourse, and together with (A2) implies that $\mathbb{E}_\omega[v_\omega(x)]$ is finite for every $x \in X$. Furthermore, Assumption (A2) excludes the case where ω follows a continuous distribution. Nevertheless, continuous distributions are typically approximated by finite discrete distributions, e.g., using sample average approximation (Kleywegt et al. 2002). Finally, the assumptions in (A3) and (A4) guarantee that X is compact and $\bar{X} := \mathrm{conv}(X)$ is a polytope (Del Pia and Weismantel 2016, Theorem 1). In addition, by (A4) the second-stage cost functions v_ω, $\omega \in \Omega$ are lower semi-continuous (lsc) on $\bar{X}$ (Schultz 1995), and thus, using that $\bar{X}$ is compact, v_ω is bounded from below on $\bar{X}$ (Anger 1990, Theorem 3.7).

2.2 Benders' Decomposition for MIR Models

Benders' decomposition (Benders 1962) is widely used to solve MIR models, since it is able to exploit their underlying two-stage structure. A Benders' decomposition algorithm maintains an outer approximation $\hat{Q}_{\text{out}} : \bar{X} \mapsto \mathbb{R}$ of the expected second-stage cost function $Q(x) := \mathbb{E}_\omega[v_\omega(x)]$, i.e., $\hat{Q}_{\text{out}}(x) \le Q(x)\ \forall x \in X$. The corresponding relaxation of (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 (1). If, however, $\hat{Q}_{\text{out}}(\bar{x}) < Q(\bar{x})$, then the outer approximation is strengthened using an *optimality cut* for Q:

$$Q(x) \ge \alpha - \beta^\top x \quad \forall x \in X,$$

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 1. Throughout, we maintain a lower- and upper bound LB and UB on η^*, i.e., $LB \le \eta^* \le UB$.

Algorithm 1 Benders' Decomposition for MIR Models.

1: **Initialization**
2: $\hat{Q}_{\text{out}} \equiv L$, where $Q(x) \ge L\ \forall x \in X$.
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 $Q(x) \ge \alpha - \beta^\top x\ \forall x \in X$.
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., Zou et al. (2019), and thus the algorithm may stall. Therefore, in a practical implementation of Algorithm 1, 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 1 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 by Van Slyke and Wets (1969), which solves continuous recourse models, uses optimality cuts of the form

$$v_\omega(x) \geq \alpha_\omega - \beta_\omega^\top x \quad \forall x \in X,\ \omega \in \Omega, \tag{3}$$

by exploiting linear programming (LP) duality 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 (3) are aggregated to derive optimality cuts for Q. We review such generalizations in Section 2.2.1. However, if optimality cuts are computed by aggregating cuts of the form (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 3. Therefore, we propose a new family of optimality cuts which is suited for general MIR models in Section 3.1, and we use it to develop a modified Benders' decomposition in Section 3.2.

2.2.1 Generalizations to Mixed-Integer Recourse. The L-shaped method exploits that the expected second-stage cost function Q is convex polyhedral if the recourse is continuous, i.e., if $\mathcal{Y} = \mathbb{R}_+^{n_2}$. In contrast, if $\mathcal{Y}$ is a mixed-integer set, then Q is in general not convex, or even continuous, see, e.g., Schultz (1995). Therefore, the L-shaped method does not readily generalize to broader classes of MIR models. However, Laporte and Louveaux (1993) 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 exploit this result to develop the integer L-shaped algorithm for MIR models with $\mathcal{X} = \mathbb{B}^{n_1}$.

In fact, there exists a wide range of algorithms generalizing the L-shaped method to this special case, that typically use techniques for deterministic MIPs. For example, Sherali and Fraticelli (2002), Sen and Higle (2005), Ntaimo and Tanner (2008), Ntaimo (2010, 2013), Gade et al. (2014), and Qi and Sen (2017) 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 (2014) manage to generalize the approach based on Gomory cuts by Gade et al. (2014) 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, and Kim and Mehrotra (2015) use mixed-integer rounding cuts to derive tight continuous relaxations for a nurse scheduling problem with general mixed-integer decision variables and a totally unimodular recourse matrix.

In another direction, Sen and Sherali (2006) 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. (2019), 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.

In fact, there does not exist a *tractable* Benders' 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. Indeed, our Benders' decomposition algorithm generalizes the algorithms by Sherali and Fraticelli (2002), Sen and Higle (2005), Sen and Sherali (2006), Ntaimo and Tanner (2008), Ntaimo (2010, 2013), Gade et al. (2014), Zhang and Küçükyavuz (2014), and Qi and Sen (2017) to general MIR models.

The advantage of our method 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, the global branch-and-bound procedure by Ahmed et al. (2004), the dual decomposition approach by Carøe and Schultz (1999), and the decomposition-based branch-and-bound algorithm by Sherali and Zhu (2006) use spatial branching for convergence, and Carøe and Tind (1998) use auxiliary integer decision variables to capture non-convex terms in the master problem, exploiting general duality for MIPs. Similarly, the stochastic Lipschitz dynamic programming algorithm by Ahmed et al. (2020) introduces binary variables to include non-linear optimality cuts in the master problem.

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 $\text{co}(Q)$ of Q, so that we can solve the MIR model in (1) by replacing $Q(x)$ by $\text{co}(Q)(x)$ and the feasible region X by its convex hull $\bar{X}$. That is, the resulting convex relaxation of the original problem in (1), defined as

$$\hat{\eta} := \min_x \{c^\top x + \text{co}(Q)(x) : \ x \in \bar{X}\}, \tag{4}$$

satisfies $\hat{\eta} = \eta^*$, and moreover, if x^* is optimal in the original problem (1), then x^* is also optimal in (4), see, e.g., Proposition 2.4 in Tardella (2004).

In contrast, traditional Benders' decomposition algorithms for MIR models, see, e.g., Sherali and Fraticelli (2002), Sen and Higle (2005), and Gade et al. (2014), use optimality cuts which, in general, do not yield $\text{co}(Q)$. 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 $\mathbb{E}_\omega[\text{co}(v_\omega)]$. However, this expected value of the convex envelopes 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 $\mathbb{E}_\omega[\text{co}(v_\omega)(x)] \leq \text{co}(Q)(x)$, resulting in a *duality gap*, see also Carøe and Schultz (1999) and Boland et al. (2018). This gap is zero if $\mathcal{X} = \mathbb{B}^{n_1}$ (Zou et al. 2019, Theorem 1), but if $\mathcal{X}$ is a general mixed-integer set, then the duality gap may be positive, see Example 1.

Remark 1. In general, any family of linear optimality cuts for Q yields at most its *closed* convex envelope $\overline{\text{co}}(Q)$. However, since Q is lsc and X is compact, we have that $\overline{\text{co}}(Q) = \text{co}(Q)$ (Falk 1969, Theorem 2.2). Similarly, $\overline{\text{co}}(v_\omega) = \text{co}(v_\omega)$ for every $\omega \in \Omega$.

Example 1. Consider the expected second-stage cost function $Q(x) = \mathbb{E}_\omega[v_\omega(x)], \ x \in [0, 4]$, where

$$v_\omega(x) = \min_y \{2y : y \geq \omega - x, \ y \in \mathbb{Z}_+\}, \quad x \in [0, 4],$$

and ω 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., Louveaux and van der Vlerk (1993). 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 $\text{co}(v_{\omega_1})(x) = 2 \max\{0, \omega_1 - x, 3 - 2x\}$ and $\text{co}(v_{\omega_2})(x) = 2 \max\{0, \omega_2 - x\}$.

Figure 1 shows v_{ω_1} and v_{ω_2} and their convex envelopes as functions of x. Observe that the difference between $\text{co}(v_\omega)(x)$ and $v_\omega(x)$ in general not equal to zero, and that the values of x for which $\text{co}(v_\omega)(x) = v_\omega(x)$ are not the same for $\omega = \omega_1$ and $\omega = \omega_2$. This results in a positive duality gap between $\text{co}(Q)(x)$ and $\mathbb{E}_\omega[\text{co}(v_\omega)(x)]$, see Figure 2. For example, at $x = 1$, we have $\text{co}(Q)(1) = Q(1) = 4$, but $\mathbb{E}_\omega[\text{co}(v_\omega)(1)] = 3.5$, i.e., the duality gap at $x = 1$ is equal to $1/2$. $\diamondsuit$

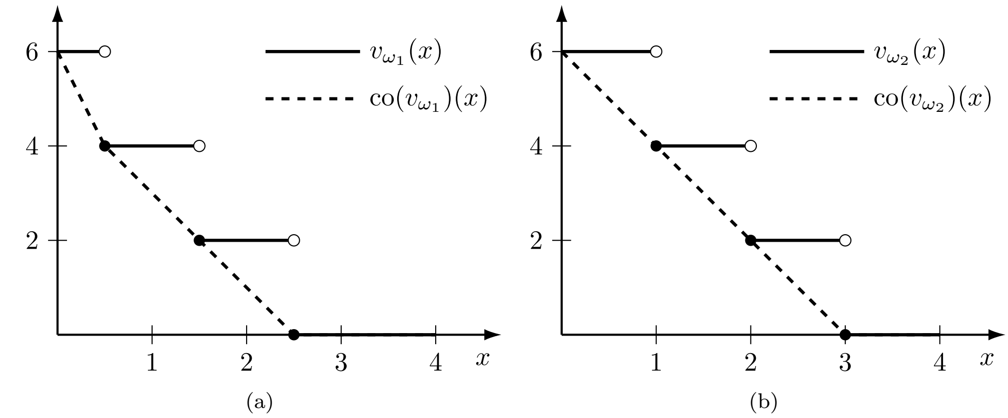

Figure 1: The Second-Stage Cost Functions v_{ω_1} and v_{ω_2} of Example 1 and Their Convex Envelopes.

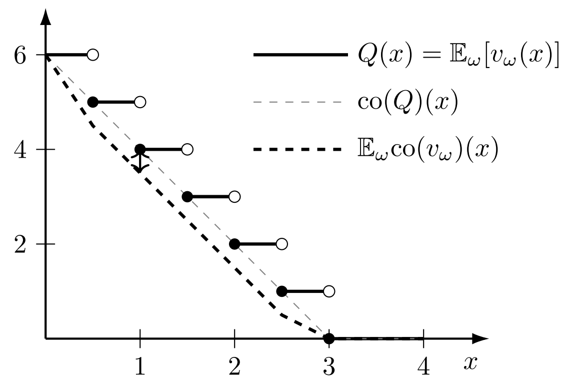

Figure 2: The Duality Gap for MIR Models: the difference between $\text{co}(Q)(x)$ and $\mathbb{E}_\omega \text{co}(v_\omega)(x)$ in Example 1 is in general non-negative, and equal to $1/2$ if, e.g., $x = 1$.

The duality gap illustrated in Example 1 may be closed using scaled cuts, which we derive in Section 3.1. Indeed, we show in Theorem 1 that they can be used to recover $\text{co}(Q)$. In Section 3.2, we use scaled cuts to develop a Benders' decomposition algorithm which solves MIR models with general mixed-integer variables.

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 is computationally tractable. 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 1 illustrates that the resulting cut

$$Q(x) \geq \mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top x \quad \forall x \in X, \tag{5}$$

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), and, in fact, they may be used to close the duality gap illustrated in Example 1.

More precisely, we consider cuts for v_ω, $\omega \in \Omega$, of the form

$$v_\omega(x) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega \phi(x) \quad \forall x \in X, \tag{6}$$

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. (2020) derive cuts of the form (6) using $\phi(x) = ||x - \bar{x}||$, where $\bar{x} \in \bar{X}$ and $|| \cdot ||$ 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 (6), which yields

$$Q(x) \geq \mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top x - \mathbb{E}_\omega \tau_\omega \phi(x) \quad \forall x \in X, \tag{7}$$

and if we use that $\phi(x) \leq Q(x)$ to obtain the following cut,

$$Q(x) \geq \frac{Q(x) + \mathbb{E}_\omega \tau_\omega \phi(x)}{1 + \mathbb{E}_\omega \tau_\omega} \geq \frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top x}{1 + \mathbb{E}_\omega \tau_\omega} \quad \forall x \in X.$$

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 (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 1, and in Example 2 we illustrate how to compute a scaled cut for the SIR model of Example 1. For technical reasons, we assume throughout that epi(ϕ) is a rational polyhedron; if ϕ satisfies this condition, we say that ϕ is a *rational convex polyhedral function.*

Definition 1 (scaled cuts)**.** Let $\phi : \bar{X} \mapsto \mathbb{R}$ be a rational convex polyhedral function 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 (6), i.e.,

$$\Pi_\omega(\phi) := \{(\alpha, \beta, \tau) :\ v_\omega(x) \geq \alpha - \beta^\top x - \tau \phi(x)\ \forall x \in X,\ \tau \geq 0\}.$$

Then, for any $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi)$, $\omega \in \Omega$, the optimality cut

$$Q(x) \geq \frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top x}{1 + \mathbb{E}_\omega \tau_\omega}, \quad \forall x \in X \tag{8}$$

is referred to as a *scaled cut.*

Example 2 (Example 1 continued)**.** Consider the SIR function Q of Example 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 $\bar{x}$, see Figure 3.

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 4a. We show the corresponding scaled cut $Q(x) \geq (8 - 3x)/2\ \forall x \in [0, 4]$ in Figure 4b. Figures 4a and 4b reveal the following geometric interpretation

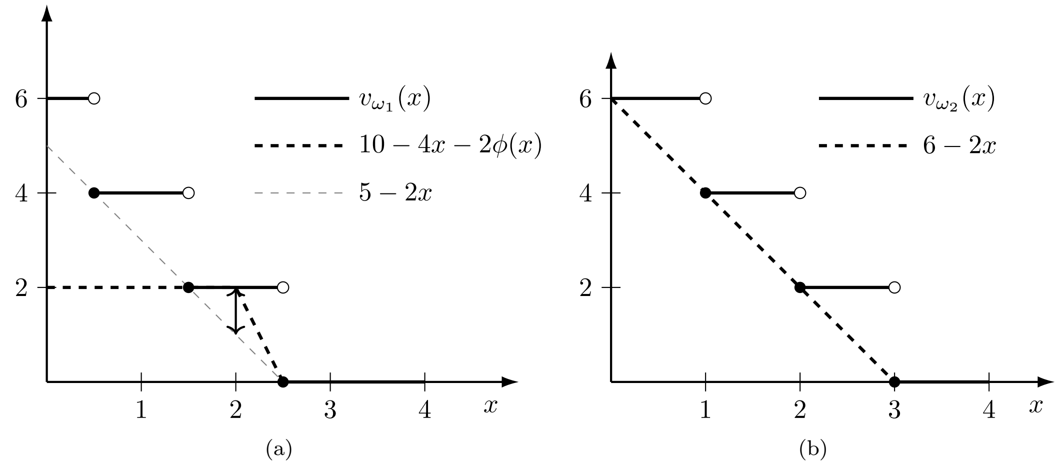

Figure 3: Illustration of the (Non-Linear) Cuts for the Second-Stage Cost Functions Derived in Example 2. The left figure displays the the second-stage cost function $v_\omega(x) = 2\lceil 2.5 - x\rceil^+$, and the non-linear cut $v_{\omega_1}(x) \geq 10 - 4x - 2\phi(x)\ \forall x \in [0, 4]$, where $\phi(x) = \max\{0, 4 - 2x\},\ x \in [0, 4]$. Observe that this cut is tight at, e.g., $\bar{x} = 2$, and strictly improves the best possible linear cut $v_{\omega_1}(x) \geq 5 - 2x\ \forall x \in [0, 4]$ at $\bar{x}$. The right figure displays the second-stage cost function $v_{\omega_2}(x) = 2\lceil 3 - x\rceil^+$ and the linear cut $v_{\omega_2}(x) \geq 6 - 2x$, which is tight at $\bar{x}$.

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

$$\frac{\alpha - \beta^\top x}{1 + \tau} = \phi(x) = \alpha - \beta^\top x - \tau\phi(x). \qquad \diamond$$

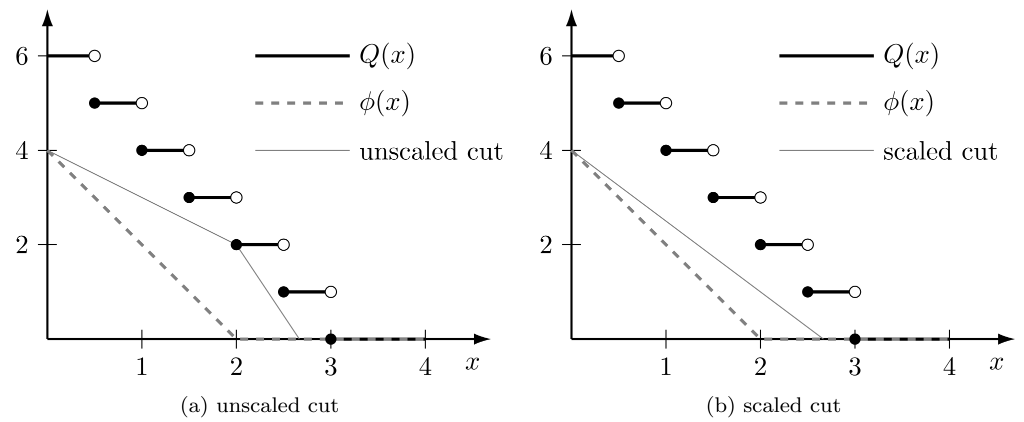

Figure 4: The left figure shows the unscaled cut $Q(x) \geq 8 - 3x - \phi(x)\ \forall x \in [0, 4]$ of Example 2, where $\phi(x) = \max\{0, 4 - 2x\},\ x \in [0, 4]$. The right figure shows the corresponding scaled cut $Q(x) \geq (8 - 3x)/2\ \forall x \in [0, 4]$.

In Example 2, the non-linear cuts for the non-convex second-stage costs functions v_ω are tight at $\bar{x}$. In Lemma 1, we derive general sufficient conditions for the cut-generating function ϕ so that such a tight non-linear cut of the form $v_\omega(x) \geq \alpha - \beta^\top x - \tau\phi(x)$ exists.

Lemma 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. □

An important implication of Lemma 1 is that if $\phi = \hat{Q}_{\text{out}}$, where $\hat{Q}_{\text{out}}$ is an outer approximation of Q, then there exists a scaled cut which improves $\hat{Q}_{\text{out}}$ at $\bar{x}$, if $(\bar{x}, \phi(\bar{x}))$ is an extreme point of $\text{conv}(\text{epi}_{\text{X}}(\phi))$ and $\hat{Q}_{\text{out}}(\bar{x}) < Q(\bar{x})$. Indeed, by Lemma 1, there exist cut coefficients $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi)$, $\omega \in \Omega$ such that the corresponding cut for v_ω is tight at $\bar{x}$, i.e., $v_\omega(\bar{x}) = \alpha_\omega - \beta_\omega^\top \bar{x} - \tau_\omega \phi(\bar{x})$, and thus the scaled cut in (8) improves $\hat{Q}_{\text{out}}$ in $\bar{x}$, since

$$\frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x}}{1 + \mathbb{E}_\omega \tau_\omega} = \frac{\mathbb{E}_\omega [v_\omega(\bar{x}) + \tau_\omega \phi(\bar{x})]}{1 + \mathbb{E}_\omega \tau_\omega} = \frac{Q(\bar{x}) + \mathbb{E}_\omega \tau_\omega \phi(\bar{x})}{1 + \mathbb{E}_\omega \tau_\omega} > \phi(\bar{x}) = \hat{Q}_{\text{out}}(\bar{x}),$$

where the inequality follows from $Q(\bar{x}) > \hat{Q}_{\text{out}}(\bar{x}) = \phi(\bar{x})$.

This suggests that we can use scaled cuts to iteratively improve outer approximations of Q. We formalize this intuition by showing that we can recover $\text{co}(Q)$ via 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 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 1.

Definition 2 (Scaled cut closure)**.** Let $\phi : \bar{X} \mapsto \mathbb{R}$ be a rational convex polyhedral function. Then, the scaled cut closure $\text{SCC}(\phi) : \bar{X} \mapsto \mathbb{R}$ of ϕ is defined as

$$\text{SCC}(\phi)(x) = \sup_{\alpha_\omega, \beta_\omega, \tau_\omega} \left\{ \frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top x}{1 + \mathbb{E}_\omega \tau_\omega} : \ (\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi) \ \forall \omega \in \Omega \right\}, \quad x \in \bar{X}.$$

The definition of the scaled cut closure implies that $\text{SCC}(\phi)$ can be described using *infinitely* many scaled cuts. It turns out, however, that $\text{SCC}(\phi)$ is convex polyhedral, see Proposition 1, i.e., $\text{SCC}(\phi)$ is the pointwise supremum of *finitely* many optimality cuts. Furthermore, if $\phi \leq Q$, then $\text{SCC}(\phi) \leq Q$, since the scaled cuts of Definition 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 convex approximations of MIR models by Romeijnders et al. (2016) and van der Laan and Romeijnders (2020). 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 1. *Let $\phi : \bar{X} \mapsto \mathbb{R}$ be a rational convex polyhedral function. Then, $\text{SCC}(\phi)$ is a rational convex polyhedral function.*

Proof. See appendix. □

Theorem 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 $\phi_{k+1} = \text{SCC}(\phi_k)$, $k \geq 0$. Then, ϕ_k converges uniformly to $\text{co}(\max\{\phi_0, Q\})$. In particular, if $\phi_0(x) \leq Q(x)\ \forall x \in X$, then $\phi_k \to \text{co}(Q)$.*

Proof. The proof is postponed to Section 5. □

Theorem 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 $\text{co}(Q)$ using scaled cuts, thereby solving the MIR model in (1). If, however, ϕ_0 is an inexact outer approximation obtained by solving a convex approximation of (1), then we may use scaled cuts to improve the quality of the resulting solution. Of course, in practice, a complete description of $\text{co}(Q)$ is typically not required to solve the MIR model in (1). Therefore, we use scaled cuts to develop an efficient Benders' decomposition algorithm for MIR models in Section 3.2.

3.2 Benders' Decomposition with Scaled Cuts

We propose a Benders' decomposition algorithm in which we iteratively construct tighter outer approximations of Q using scaled cuts. That is, we maintain an outer approximation $\hat{Q}_{\text{out}}$ of Q, and we solve the master problem

$$\eta^* = \min_x \{c^\top x + \hat{Q}_{\text{out}}(x) : \ x \in X\}, \tag{MP}$$

to obtain the current solution $\bar{x}$. If $\hat{Q}_{\text{out}}(\bar{x}) < Q(\bar{x})$, then we compute a scaled cut which improves $\hat{Q}_{\text{out}}$ at $\bar{x}$ using $\hat{Q}_{\text{out}}$ as a cut-generating function, i.e., we take $\phi = \hat{Q}_{\text{out}}$. Recall from Lemma 1 that such a scaled cut exists if (MP) returns an optimal solution $\bar{x}$ such that $(x, \phi(\bar{x}))$ is an extreme point of $\text{conv}(\text{epi}_X(\phi))$. In particular, then there exist cuts $v_\omega(x) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega \phi(x) \ \forall x \in X$, $\omega \in \Omega$, which are tight at $\bar{x}$, and thus 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 also tight at $\bar{x}$. In general, however, the resulting scaled cut

$$Q(x) \geq \frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top x}{1 + \mathbb{E}_\omega \tau_\omega} \quad \forall x \in X \tag{9}$$

is *not* tight at $\bar{x}$, unless $\mathbb{E}_\omega \tau_\omega = 0$, since otherwise

$$\frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top x}{1 + \mathbb{E}_\omega \tau_\omega} = \frac{\mathbb{E}_\omega [v_\omega(\bar{x}) + \tau_\omega \phi(\bar{x})]}{1 + \mathbb{E}_\omega \tau_\omega} = \frac{Q(\bar{x}) + \mathbb{E}_\omega \tau_\omega \phi(\bar{x})}{1 + \mathbb{E}_\omega \tau_\omega} < Q(\bar{x}),$$

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 (9) improves the outer approximation at $\bar{x}$. As a result, the scaled cut obtained by computing tight non-linear 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 order to compute the dominating scaled cut, we solve

$$\rho^* := \sup_{\alpha_\omega, \beta_\omega, \tau_\omega} \left\{ \frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x}}{1 + \mathbb{E}_\omega \tau_\omega} : \ (\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi) \ \forall \omega \in \Omega \right\}. \tag{10}$$

The optimization problem in (10) presents a significant challenge, since it features a non-linear objective function. A natural way to a 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

$$C(\rho) := \sup_{\alpha_\omega, \beta_\omega, \tau_\omega} \left\{ \mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x} - \rho(1 + \mathbb{E}_\omega \tau_\omega) : \ (\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi), \ \omega \in \Omega \right\}. \tag{11}$$

For arbitrary values of ρ, this linearised optimization problem merely represents an approximation of the one in (10). However, it turns out that if $C(\rho) = 0$, then $\rho = \rho^*$, i.e., ρ equals the optimal objective value in (11), and the optimal solutions of the optimization problems in (10) and (11) are the same. Indeed, if $C(\rho) = 0$, then $\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x} - \rho(1 + \mathbb{E}_\omega \tau_\omega) \leq 0$ for all $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi)$, $\omega \in \Omega$, and thus

$$\frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x}}{1 + \mathbb{E}_\omega \tau_\omega} \leq \rho \quad \forall (\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi), \omega \in \Omega. \tag{12}$$

Moreover, if the supremum in (11) is attained, then the optimal solution $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi)$, $\omega \in \Omega$, satisfies the inequality in (12) with equality, and thus $\rho = \rho^*$.

Instead of solving (10), we thus solve $C(\rho) = 0$ for ρ. Before explaining how we do so, we first introduce several properties of $C(\cdot)$ in Lemma 2 that we will exploit. In particular, we will use that $C(\cdot)$ is strictly decreasing, continuous and convex.

Lemma 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 (11) *is continuous, convex, and strictly decreasing on* $\text{dom}(C) = \{\rho : C(\rho) < \infty\}$,

(ii) the supremum in (11) *is attained if* $\rho \in \text{dom}(C)$*,*

(iii) for $\bar{\rho} \in \text{dom}(C)$*, a subgradient of* $C(\cdot)$ *at* $\bar{\rho}$ *is given by* $-(1 + \mathbb{E}_\omega \tau_\omega)$*, where* τ_ω*,* $\omega \in \Omega$*, correspond to an optimal solution of the problem in* (11) *with* $\rho = \bar{\rho}$*, and*

(iv) if $\bar{x} \in X$*, then* $\text{dom}(C) = [\phi(\bar{x}), \infty)$*.*

Proof. See appendix. □

Lemma 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 $\rho \in \text{dom}(C)$, leading to the following characterization of ρ^* in Lemma 3 that holds in general.

Lemma 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* (10) *satisfies*

$$\rho^* = \min_{\rho}\{\rho : \ C(\rho) \leq 0\}. \tag{13}$$

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. □

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

$$\rho_{k+1} = \rho_k + \frac{C(\rho_k)}{1 + \mathbb{E}_\omega \tau_\omega}, \tag{14}$$

where τ_ω, $\omega \in \Omega$, correspond to an optimal solution of the problem in (11) 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 (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 4 establishes that $\rho_k \to \rho^*$. To prove Lemma 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 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*,* $\omega \in \Omega$*, correspond to an optimal solution of the problem in* (11) *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. □

Based on Lemma 4, we propose to solve (10) 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 4 ensures that this algorithm is finitely convergent, and that on termination, $\rho_k \geq \rho^* - \delta$. Moreover, we note that $C(\rho)$ can be computed efficiently using the expression $C(\rho) = \mathbb{E}_\omega[C_\omega(\rho)]$, where

$$C_\omega(\rho) := \sup_{\alpha,\beta,\tau} \{\alpha - \beta^\top \bar{x} - \rho(1+\tau) : \ (\alpha,\beta,\tau) \in \Pi_\omega(\phi)\}. \tag{15}$$

That is, we exploit that the problem in (11) decomposes by scenario. Furthermore, we can efficiently parallelize our fixed point iteration algorithm by computing the quantities $C_\omega(\rho)$, $\omega \in \Omega$, in parallel. Finally, in Section 4, we describe several strategies for solving (15), which exploit that $\Pi_\omega(\phi)$ is a convex polyhedral set.

4 Computation of Scaled Cuts

In this section, we describe how to efficiently solve the problem in (15). This enables us to compute the dominating scaled cut at the current solution $\bar{x}$ via the fixed point iteration algorithm in Section 3.2. To solve (15), we exploit that $\Pi_\omega(\phi)$ is polyhedral, see Lemma 5. To derive this result, we recall that $\Pi_\omega(\phi)$ is the set of cut coefficients (α,β,τ) which define non-linear optimality cuts for the second-stage cost functions v_ω of the form

$$v_\omega(x) \geq \alpha - \beta^\top x - \tau\phi(x) \quad \forall x \in X. \tag{16}$$

We analyse cuts of the form (16) by exploiting that v_ω is a mixed-integer programming value function. In particular, note that for any $s \in \mathbb{R}$, we have $v_\omega(x) \geq s$ if and only if $q_\omega^\top y \geq s$ for every $y \in \mathcal{Y}$ such that $W_\omega y = h_\omega - T_\omega x$. If we assume, for the purpose of exposition, that $\phi \equiv 0$, then by similar reasoning, (α,β,τ) satisfies (16) if and only if $q_\omega^\top y \geq \alpha - \beta^\top x$ for every $(x,y) \in S_\omega := \{(x,y) \in X \times \mathcal{Y} : \ W_\omega y + T_\omega x = h_\omega\}$. In fact, we only need that $q_\omega^\top y^i \geq \alpha - \beta^\top x^i$ for each of the finitely many extreme points (x^i, y^i), $i = 1,\dots,k$, of $\mathrm{conv}(S_\omega)$. To see this, note that every $(x,y) \in S_\omega$ can be written as a convex combination of these extreme points, i.e., $(x,y) = \sum_{i=1}^k \lambda^i (x^i, y^i)$ for some $\lambda^i \geq 0$, $i = 1,\dots,k$, with $\sum_{i=1}^k \lambda^i = 1$, and thus

$$q_\omega^\top y = \sum_{i=1}^k \lambda^i q_\omega^\top y^i \geq \sum_{i=1}^k \lambda^i (\alpha - \beta^\top x^i) = \alpha - \beta^\top x$$

if $q_\omega^\top y^i \geq \alpha - \beta^\top x^i$ for every $i = 1,\dots,k$.

To derive a similar characterisation for the case where $\phi \not\equiv 0$, we first linearise the cut in (16) by noting that if $\tau \geq 0$, then (α,β,τ) satisfies (16) if and only if

$$v_\omega(x) \geq \alpha - \beta^\top x - \tau\theta \quad \forall (x,\theta) \in X \times \mathbb{R} \text{ such that } \theta \geq \phi(x). \tag{17}$$

In other words, we are able to derive non-linear cuts for v_ω in the x-space by deriving linear cuts for v_ω in the (x,θ)-space. Similar to the case where $\phi \equiv 0$, we have that (α,β,τ) satisfies (17) if and only if $q_\omega^\top y \geq \alpha - \beta^\top x - \tau\theta$ for every $(x,\theta,y) \in S_\omega^\phi$, where

$$S_\omega^\phi := \{(x,\theta,y) \in X \times \mathbb{R} \times \mathcal{Y} : \ \theta \geq \phi(x), \ W_\omega y = h_\omega - T_\omega x\}.$$

We are now in a position to state our representation result for $\Pi_\omega(\phi)$.

Lemma 5. *Let $\phi : \bar{X} \mapsto \mathbb{R}$ be a rational convex polyhedral function and consider $\Pi_\omega(\phi) = \{(\alpha,\beta,\tau) : \ v_\omega(x) \geq \alpha - \beta^\top x - \tau\phi(x) \ \forall x \in X, \ \tau \geq 0\}$. Then, $\Pi_\omega(\phi)$ is a rational polyhedron, and*

$$\Pi_\omega(\phi) = \{(\alpha,\beta,\tau) : \ q_\omega^\top y^i + \beta^\top x^i + \tau\theta^i \geq \alpha \ \forall i \in \{1,\dots,d\}, \ \tau \geq 0\}, \tag{18}$$

where $(x^i, \theta^i, y^i) \in S_\omega^\phi$, $i = 1,\dots,d$, denote the extreme points of $\mathrm{conv}(S_\omega^\phi)$.

Proof. Note that $(\alpha, \beta, \tau) \in \Pi_\omega(\phi)$ is equivalent to (17). Thus, using the definition of $v_\omega(x)$ and S_ω^ϕ, we have 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 $\text{conv}(S_\omega^\phi)$, we obtain that

$$\Pi_\omega(\phi) = \{(\alpha, \beta, \tau) : \; q_\omega^\top y + \beta^\top x + \tau\theta \geq \alpha \; \forall (x, \theta, y) \in \text{conv}(S_\omega^\phi)\}.$$

To obtain (18), observe that $\text{conv}(S_\omega^\phi)$ is a rational polyhedron (Del Pia and Weismantel 2016, Theorem 1) with one extreme direction, namely $(0, 1, 0)$, and finitely many extreme points. □

The expression in (18) reveals that

$$C_\omega(\rho) = \sup_{\alpha,\beta,\tau} \{\alpha - \beta^\top \bar{x} - \rho(1+\tau) : q_\omega^\top y^i + \beta^\top x^i + \tau\theta^i \geq \alpha \; \forall i \in \{1, \ldots, d\}, \; \tau \geq 0\}, \tag{19}$$

i.e., we can compute $C_\omega(\rho)$ by solving a linear programming problem if all extreme points of $\text{conv}(S_\omega^\phi)$ are known. In Section 4.1, we describe a row generation scheme for solving (19) by enumerating a sufficiently rich subset of the extreme points of $\text{conv}(S_\omega^\phi)$, and in Section 4.2, we solve the dual problem of (19) using cutting plane techniques.

4.1 A Row Generation Scheme

In general, the number of extreme points of $\text{conv}(S_\omega^\phi)$ may be very large, and in those cases directly solving the LP in (19) is computationally infeasible. Therefore, we propose a row generation scheme similar to approaches in robust optimization and disjunctive programming, see, e.g., Perregaard and Balas (2001), Zeng and Zhao (2013), and Georghiou et al. (2020). In this approach, we iteratively identify extreme points $(x^i, \theta^i, y^i) \in S_\omega^\phi$, $i = 1, \ldots, t$, and we solve the resulting *cut-generation master problem*

$$\max_{\alpha,\beta,\tau} \{\alpha - \beta^\top \bar{x} - \rho(1+\tau) : \; q_\omega^\top y^i + \beta^\top x^i + \tau\theta^i \geq \alpha \;\; \forall i \in \{1, \ldots, t\}, \; \tau \geq 0\}. \tag{CGMP}$$

We denote the optimal solution of (CGMP) by $(\alpha^t, \beta^t, \tau^t)$, and we attempt to identify a point $(x^{t+1}, \theta^{t+1}, y^{t+1}) \in S_\omega^\phi$ which violates the inequality $q_\omega^\top y + {\beta^t}^\top x + \tau^t\theta \geq \alpha^t$ by solving the *cut generation subproblem*

$$\nu^t := \min_{x,\theta,y} \{q_\omega^\top y + {\beta^t}^\top x + \tau^t\theta - \alpha^t : \; (x, \theta, y) \in S_\omega^\phi\}, \tag{CGSP}$$

which is a small-scale MIP. Note that $(\alpha^t, \beta^t, \tau^t)$ is feasible and thus optimal in (19) 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 $\text{conv}(S_\omega^\phi)$ 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 $\text{conv}(S_\omega^\phi)$. Indeed, since the objective function of (CGSP) is linear, it has an optimal solution which is an extreme point of $\text{conv}(S_\omega^\phi)$. Typically, only a small fraction of the total number of extreme points needs to be computed before the algorithm terminates. We summarize the row generation scheme in Algorithm 2.

Algorithm 2 Row Generation Scheme for Solving (15).

1: **Input:** $\bar{x} \in X$, cut-generating function $\phi : \bar{X} \mapsto \mathbb{R}$, $\rho \geq \phi(\bar{x})$, and tolerance level $\delta \geq 0$
2: **Initialization**
3: $\quad t = 1$ and $(x^1, \theta^1, y^1) = (\bar{x}, \rho, \bar{y})$, for an arbitrary $\bar{y} \in \{\mathcal{Y} : \; W_\omega y = h_\omega - T_\omega \bar{x}\}$.
4: **Iteration step**
5: $\quad$ Solve (CGMP) and update (CGSP) using optimal solution $(\alpha^t, \beta^t, \tau^t)$.
6: $\quad$ Solve (CGSP), denote optimal value by ν^t and optimal solution by $(x^{t+1}, \theta^{t+1}, y^{t+1})$.
7: $\quad$ Append constraint $q_\omega^\top y^{t+1} + \beta^\top x^{t+1} + \tau\theta^{t+1} \geq \alpha$ to (CGMP).
8: **Stopping criterion**
9: $\quad$ **if** $\nu^t \geq -\delta$ **then stop**: $(\alpha^t + \nu^t, \beta^t, \tau^t)$ is δ-optimal in (15)
10: $\quad$ **else**
11: $\qquad t \leftarrow t + 1$ and go to line 5
12: $\quad$ **end if**

In Algorithm 2, we initialize (CGMP) with the point $(\bar{x}, \rho, \bar{y}) \in S_\omega^\phi$ in order to ensure that (CGMP) is bounded. Note that Algorithm 2 can be implemented efficiently, since the problems in (CGMP) and (CGSP) are a small-scale LP and MIP, respectively. Furthermore, in the fixed point iteration algorithm in Section 3.2, we have to obtain $C_\omega(\rho)$ multiple times for different values of ρ, and thus we have to run Algorithm 2 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 of Algorithm 2 in subsequent runs. This is possible since the feasible region S_ω^ϕ of (CGSP) does not depend on ρ.

4.2 Convexification via Cutting Plane Techniques

The second approach we consider for solving the problem in (19) is to use cutting plane techniques to solve its dual LP, which we derive in Lemma 6 below. The advantage of this approach over the row generation scheme in Section 4.1 is that it only requires solving small-scale LPs, which is computationally less expensive, and thus it may be faster if not too many LPs need to be solved.

Lemma 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* (19)*. Then,*

$$C_\omega(\rho) = -\rho + \min_y \{q_\omega^\top y : (\bar{x}, \rho, y) \in \mathrm{conv}(S_\omega^\phi)\} \quad \forall \rho \in \mathrm{dom}(C_\omega). \tag{20}$$

Proof. We will show that the dual of (19) is given by the expression in (20), so that the result follows from strong LP duality. In particular, for arbitrary $\rho \in \mathrm{dom}(C_\omega)$, the dual of (19) 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 $\mathrm{conv}(S_\omega^\phi)$, the above is equivalent to

$$C_\omega(\rho) = -\rho + \min_{\theta, y} \left\{ q_\omega^\top y : \; (\bar{x}, \theta, y) \in \mathrm{conv}(S_\omega^\phi), \; \theta \leq \rho \right\}, \tag{21}$$

and (20) follows by noting that it is optimal to select $\theta = \rho$ in (21). $\square$

We solve the problem in (20) 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 $\mathrm{conv}(S_\omega^\phi)$, i.e.,

$$\mathrm{conv}(S_\omega^\phi) \subseteq \hat{S}_\omega^\phi := \{(x, \theta, y) : W_\omega y = h_\omega - T_\omega x, \; \hat{W}_\omega y \geq \hat{h}_\omega - \hat{T}_\omega x - r_\omega \theta\}. \tag{22}$$

In particular, we use these cutting planes to obtain the following relaxation of (20),

$$\begin{aligned}\hat{C}_\omega(\rho) &= -\rho + \min_y\{q_\omega^\top y : \ (\bar{x}, \rho, y) \in \hat{S}_\omega^\phi\} \\ &= -\rho + \min_y\{q_\omega^\top y : W_\omega y = h_\omega - T_\omega \bar{x}, \ \hat{W}_\omega y \geq \hat{h}_\omega - \hat{T}_\omega \bar{x} - r_\omega \rho\}. \end{aligned} \tag{23}$$

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 (23) reduces to the LP-relaxation of the second-stage subproblem. If the resulting solution $\bar{y}$ of this relaxation is such that $(\bar{x}, \rho, \bar{y}) \in \text{conv}(S_\omega^\phi)$, then we are done: $\bar{y}$ is optimal in (20) and $\hat{C}_\omega(\rho) = C_\omega(\rho)$. Otherwise, we derive a parametric cutting plane which separates $(\bar{x}, \rho, \bar{y})$ from $\text{conv}(S_\omega^\phi)$, after which we update $\hat{S}_\omega^\phi$ and resolve (23). Depending on the family of cutting planes that we use to recover $\text{conv}(S_\omega^\phi)$, this procedure is finitely convergent. In particular, if we use the Fenchel cuts by Boyd (1994), then the resulting algorithm is finitely convergent (Boyd 1995, Corollary 3.3). Before discussing further computational aspects of our cutting plane approach, Lemma 7 describes how we can retrieve an optimal solution (α, β, τ) of the *primal* problem in (19) if we have solved the *dual* problem in (20).

Lemma 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* (22)*. 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* (23)*, 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,*

$$(\alpha, \beta, \tau) := (\lambda_w^\top h_\omega + \pi_w^\top \hat{h}_\omega, \lambda_\omega^\top T_\omega + \pi_w^\top \hat{T}_\omega, \pi_\omega^\top r_\omega) \tag{24}$$

is feasible in (19)*, 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 (23), 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 (19) 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 $(x, \theta, y) \in S_\omega^\phi$, 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 (20) 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 $(\bar{x}, \rho, \bar{y}) \notin \text{conv}(S_\omega^\phi)$, then $(\bar{x}, \rho, \bar{y})$ does not satisfy the integer restrictions in S_ω^ϕ, and thus we can apply ideas from deterministic mixed-integer programming to generate specific types of cutting planes for S_ω^ϕ. For example, we outline how to generate (strengthened) lift-and-project (L&P) cuts in Section 4.2.1. Of course, it is also possible to generate other types of cutting planes, see, e.g., Balas and Jeroslow (1980) and Zhang and Küçükyavuz (2014) for Gomory mixed-integer (GMI) cuts, and Qi and Sen (2017) for multi-term disjunctive cuts. In the practical implementation of our cutting plane approach in Algorithm 3, we accommodate the case where the cutting planes do not converge finitely by stopping after a pre-specified number of iterations K, or if we are unable to cut away a fractional solution $(\bar{x}, \rho, \bar{y})$.

Algorithm 3 Cutting Plane Approach for Solving (19).

```
1: Input: x̄ ∈ X, cut-generating function φ : X̄ ↦ ℝ, and ρ ≥ φ(x̄), iteration limit K.
2: Initialization
3:     Let Ŵ_ω and T̂_ω denote empty matrices, ĥ_ω and r_ω denote empty vectors, and let k ← 0.
4: Iteration step
5:     Solve
```

$$\min_y \{q_\omega^\top y : \; W_\omega y = h_\omega - T_\omega \bar{x}, \; \hat{W}_\omega y \geq \hat{h}_\omega - \hat{T}_\omega \bar{x} - r_\omega \rho\},$$

store optimal solution $\bar{y}$, and dual multipliers λ_ω and π_ω.

6: Let $(\alpha, \beta, \tau) \leftarrow (\lambda_w^\top h_\omega + \pi_w^\top \hat{h}_\omega^k, \lambda_\omega^\top T_\omega + \pi_w^\top \hat{T}_\omega^k, \pi_\omega^\top r_\omega^k)$.

7: **Stopping criterion**

8: **if** $\bar{y}$ satisfies integer restrictions or if $k > K$ **then return** (α, β, τ).

9: **else**

10: Generate cutting plane $w^\top y + a^\top x + r\theta \geq s \; \forall (x, \theta, y) \in S_\omega^\phi$.

11: **if** $w^\top \bar{y} + a^\top \bar{x} + r\rho \geq s$ **then return** (α, β, τ).

12: **else**

13: Let $\hat{W}_\omega \leftarrow \begin{pmatrix} \hat{W}_\omega \\ w^\top \end{pmatrix}$, $\hat{T}_\omega \leftarrow \begin{pmatrix} \hat{T}_\omega \\ a^\top \end{pmatrix}$, $r_\omega \leftarrow \begin{pmatrix} r_\omega \\ r \end{pmatrix}$, and $\hat{h}_\omega \leftarrow \begin{pmatrix} \hat{h}_\omega \\ s \end{pmatrix}$.

14: $k \leftarrow k + 1$. Go to line 5.

15: **end if**

16: **end if**

Efficient implementations of Algorithm 3 are possible, since each iteration merely requires solving a small-scale LP. Furthermore, we may speed up the convergence of Algorithm 3 by adding multiple cutting planes to (23) in one iteration, e.g., by generating a round of GMI cuts. Finally, 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.

Remark 2. The decomposition algorithms for MIR models by Sherali and Fraticelli (2002), Sen and Higle (2005), Ntaimo and Tanner (2008), Ntaimo (2010, 2013), Gade et al. (2014), and Qi and Sen (2017) use cutting planes for the second-stage subproblems which depend only on x. These cutting planes are used to recover the convex hull of the set $\{(x, y) \in X \times \mathcal{Y} : \; W_\omega y = h_\omega - T_\omega x\}$, and the resulting continuous relaxation of $v_\omega(x)$ is guaranteed to be tight only if the first-stage variables are binary. Furthermore, the parametric Gomory cutting planes by Zhang and Küçükyavuz (2014) can be used to solve the second-stage subproblem if the first- and second-stage variables are pure integer. We are able to generalize these approaches to general mixed-integer variables by using cutting planes which depend parametrically on x and θ, where $\theta \geq \phi(x)$.

4.2.1 Lift-and-Project Cuts.

Suppose that $\bar{y}$ is a fractional solution of the LP in (23), i.e., $\bar{y}_i \notin \mathbb{Z}$ for some $i \in \{1, \ldots, p_2\}$. In order to generate an L&P cut which separates the point $(\bar{x}, \rho, \bar{y})$ from S_ω^ϕ, we denote by $\hat{S}_\omega^\phi$ the continuous relaxation of S_ω^ϕ defined by the cutting planes $\hat{W}_\omega y \geq \hat{h}_\omega - \hat{T}_\omega x - r_\omega \theta$, and we consider the disjunctive relaxation of S_ω^ϕ implied by the split disjunction $y_i \leq \lfloor \bar{y}_i \rfloor \vee y_i \geq \lceil \bar{y}_i \rceil$:

$$S_\omega^\phi \subseteq S_{\omega,\bar{y},i}^+ := \left\{(x, \theta, y) \in \hat{S}_\omega^\phi : \; y_i \leq \lfloor \bar{y}_i \rfloor \right\} \bigcup \left\{(x, \theta, y) \in \hat{S}_\omega^\phi : \; y_i \geq \lceil \bar{y}_i \rceil \right\}.$$

Next, we formulate a cut-generation LP (CGLP) which we use to recover $\mathrm{conv}(S_{\omega,\bar{y},i}^+)$ through cuts of the form $a^\top x + r\theta + w^\top y \geq s$. Without loss of generality, we may assume that there exist matrices C_ω^1 and C_ω^2, and vectors c_ω and d_ω such that $\hat{S}_\omega^\phi = \{(x, \theta, y) \in \mathbb{R}_+^{n_1+1+n_2} : \; C_\omega^1 x + c_\omega \theta + C_\omega^2 y \geq d_\omega\}$.

Then, the CGLP is given by

$$\begin{aligned}
&\min a^\top \bar{x} + r\rho + w^\top \bar{y} - s \\
&\quad \text{subject to} \\
&\quad a^\top - \lambda_i^\top C_\omega^1 \geq 0, && i = 1,2, \\
&\quad r^\top - \lambda_i^\top c_\omega \geq 0, && i = 1,2, \\
&\quad w^\top - \lambda_1^\top C_\omega^2 + \nu_1 e_i^\top \geq 0, \\
&\quad w^\top - \lambda_2^\top C_\omega^2 - \nu_2 e_i^\top \geq 0, \\
&\quad s - \lambda_1^\top d_\omega + \nu_1 \lfloor y_i \rfloor \leq 0, \\
&\quad s - \lambda_2^\top d_\omega - \nu_2 \lceil y_i \rceil \leq 0 \\
&\quad -\mathbf{1} \leq u \leq \mathbf{1},\ -1 \leq r \leq 1,\ -\mathbf{1} \leq w \leq \mathbf{1},\ -1 \leq s \leq 1, \\
&\quad \lambda_i \geq 0,\ \nu_i \geq 0, && i = 1,2,
\end{aligned} \tag{CGLP}$$

see, e.g., Balas and Perregaard (2003). Any feasible solution of (CGLP) corresponds to a valid cut for S_ω^ϕ of the form $a^\top x + r\theta + w^\top y \geq s$. Moreover, an optimal solution of (CGLP) corresponds to the deepest cut in the sense that the violation of the point $(\bar{x}, \rho, \bar{y})$ is maximized. Finally, it is possible to strengthen the resulting L&P cut in analogy to the procedure described in, e.g, Balas et al. (1996).

5 Proof of Convergence

In this section, we prove Theorem 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 $\phi_{k+1} = \mathrm{SCC}(\phi_k)$, $k \geq 0$, converges uniformly to $\mathrm{co}(\max\{\phi_0, Q\})$. For convenience, we recall that the scaled cut closure $\mathrm{SCC}(\phi)$ is defined as

$$\mathrm{SCC}(\phi)(x) = \sup_{\alpha_\omega, \beta_\omega, \tau_\omega} \left\{ \frac{\mathbb{E}_\omega[\alpha_\omega - \beta_\omega^\top x]}{1 + \mathbb{E}_\omega \tau_\omega} : (\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi)\ \forall \omega \in \Omega \right\}, \quad x \in \bar{X},$$

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 1 by showing, in Section 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 5.2, 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,

$$\begin{aligned}
\mathrm{SCC}(\phi)(x) &= \sup_{\tau_\omega \geq 0} \sup_{\alpha_\omega, \beta_\omega} \left\{ \frac{\mathbb{E}_\omega[\alpha_\omega - \beta_\omega^\top x]}{1 + \mathbb{E}_\omega \tau_\omega} :\ v_\omega(x') + \tau_\omega \phi(x') \geq \alpha_\omega - \beta_\omega^\top x'\ \forall x' \in X,\ \omega \in \Omega \right\} \\
&= \sup_{\tau_\omega \geq 0} \left\{ \frac{\mathbb{E}_\omega \overline{\mathrm{co}}(v_\omega + \tau_\omega \phi)(x)}{1 + \mathbb{E}_\omega \tau_\omega} \right\},
\end{aligned}$$

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 3.

Definition 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

$$(\mathbb{T}f)(x) = \sup_{\tau_\omega \geq 0} \left\{ \frac{\mathbb{E}_\omega \overline{\mathrm{co}}(v_\omega + \tau_\omega f)(x)}{1 + \mathbb{E}_\omega \tau_\omega} \right\}, \quad x \in \bar{X}, \quad f \in C(\bar{X}). \tag{25}$$

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 (25), $\mathbb{T}f$ 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 2 below that $\mathbb{T}f$ is continuous and bounded, i.e., $\mathbb{T}f \in C(\bar{X})$.

Theorem 2. *(Rockafellar 1970, 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 $k \geq 0$.

5.1 Uniform Convergence and Fixed Points

The main result of this section is Proposition 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 8.

Lemma 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 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 2, we obtain that ϕ_k converges *uniformly* to ϕ^* by showing that the pointwise limit ϕ^* is continuous. In addition, we exploit continuity of $\mathbb{T}$, see Lemma 9 below, to prove that ϕ^* is a fixed point of $\mathbb{T}$, i.e., $\mathbb{T}\phi^* = \phi^*$.

Lemma 9. *The mapping $\mathbb{T} : C(\bar{X}) \mapsto C(\bar{X})$ of Definition 3 is continuous on $C(\bar{X})$.*

Proof. See appendix. □

Proposition 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 (Rudin 1976, 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 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 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 9. □

5.2 Properties of Fixed Points of $\mathbb{T}$

By Proposition 2, ϕ_k converges uniformly to a fixed point of $\mathbb{T}$. We exploit this result to derive properties of the limit function ϕ^*. In particular, in Proposition 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 3, we need the following result.

Lemma 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. □

Intuitively, Lemma 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 $\tau_\omega \geq 0$. We may exploit this result to derive properties of fixed points of $\mathbb{T}$. For the purpose of exposition, assume that there exist $\tau_\omega \geq 0$ such that $\overline{\text{co}}(v_\omega + \tau_\omega f)(\bar{x}) = v_\omega(\bar{x}) + \tau_\omega f(\bar{x})$. Then,

$$(\mathbb{T}f)(\bar{x}) = \frac{\mathbb{E}_\omega[v_\omega(\bar{x}) + \tau_\omega f(\bar{x})]}{1 + \mathbb{E}_\omega \tau_\omega} = \frac{Q(\bar{x}) + \mathbb{E}_\omega \tau_\omega f(\bar{x})}{1 + \mathbb{E}_\omega \tau_\omega},$$

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 3 by formalizing this reasoning.

Proposition 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 Carathodory's theorem (Rockafellar and Wets 2009, 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 $\overline{\mathrm{co}}(\max\{\phi_0, Q\})$ to obtain the latter inequality.

We show that $f(\bar{x}) \geq \overline{\mathrm{co}}(\max\{\phi_0, Q\})(\bar{x})$ if $(\bar{x}, \bar{\theta})$ is an extreme point of 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 10,

$$\sup_{\tau_\omega \geq 0} \{\overline{\mathrm{co}}(v_\omega + \tau_\omega f)(\bar{x}) - \tau_\omega f(x)\} = \infty \quad \forall \omega \in \Omega.$$

It follows that there exists a $\tau_\omega \geq 0$ such that $\overline{\mathrm{co}}(v_\omega + \tau_\omega f)(\bar{x}) - \tau_\omega f(\bar{x}) > f(\bar{x})$. But then, for this choice of τ_ω, $\omega \in \Omega$,

$$(\mathbb{T}f)(\bar{x}) \geq \frac{\mathbb{E}_\omega \overline{\mathrm{co}}(v_\omega + \tau_\omega f)(\bar{x})}{1 + \mathbb{E}_w \tau_\omega} > \frac{\mathbb{E}_\omega[f(\bar{x}) + \tau_\omega f(\bar{x})]}{1 + \mathbb{E}_w \tau_\omega} = f(\bar{x}),$$

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 10 implies that there exist $\tau_\omega \geq 0$ such that

$$\overline{\mathrm{co}}(v_\omega + \tau_\omega f)(\bar{x}) - \tau_\omega f(\bar{x}) \geq v_\omega(\bar{x}) - \delta/2.$$

But then,

$$\begin{aligned}
(\mathbb{T}f)(\bar{x}) &\geq \frac{\mathbb{E}_\omega \overline{\mathrm{co}}(v_\omega + \tau_\omega f)(\bar{x})}{1 + \mathbb{E}_\omega \tau_\omega} \\
&\geq \frac{\mathbb{E}_\omega[v_\omega(\bar{x}) + \tau_\omega f(\bar{x}) - \delta/2]}{1 + \mathbb{E}_\omega \tau_\omega} \\
&= \frac{f(\bar{x}) + \delta/2 + \mathbb{E}_\omega[\tau_\omega f(\bar{x})]}{1 + \mathbb{E}_\omega(\tau_\omega)} > f(\bar{x}),
\end{aligned}$$

which contradicts $\mathbb{T}f = f$. □

We are now ready to prove Theorem 1.

Proof of Theorem 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 $\mathrm{co}(\max\{\phi_0, Q\})$. Proposition 2 implies that $\phi^* = \lim_{k \to \infty} \phi_k$ exists, and ϕ^* is a fixed point of $\mathbb{T}$. Moreover, using Lemma 8, we have that $\phi^* \leq \mathrm{co}(\max\{\phi_0, Q\})$, and monotonicity of ϕ_k implies that $\phi^* \geq \phi_0$. Thus, by Proposition 3, we have $\phi^* \geq \overline{\mathrm{co}}(\max\{\phi_0, Q\})$. Finally, since $\max\{\phi_0, Q\}$ is an lsc function defined on a compact domain, we have $\overline{\mathrm{co}}(\max\{\phi_0, Q\}) = \mathrm{co}(\max\{\phi_0, Q\})$ (Falk 1969, Theorem 2.2), and the result follows. □

6 Numerical Experiments

Theorem 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 1. Therefore, we assess the performance of scaled cuts on a range of problem instances, namely (variants of) an investment problem by Schultz et al. (1998), as well as (variants of) the DCAP problem instances by Ahmed and Garcia (2003) from SIPLIB (Ahmed et al. 2015), see Sections 6.3.2 and 6.3.3, respectively. In addition, in Section 6.3.1, we consider a problem instance by Carøe and Schultz (1999), 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 6.1, and in Section 6.2, we describe a cut-enhancement technique which we use to speed up the convergence of scaled cuts.

6.1 Setup of Numerical Experiments

In our numerical experiments, we compare scaled cuts to traditional optimality cuts in terms of bounds on the optimal value, solution quality, and running time. In particular, we consider traditional Benders' cuts of Van Slyke and Wets (1969), as well as the strengthened Benders' (SB) cuts and the Lagrangian (L) cuts of Zou et al. (2019). We compute L cuts using a row-generation scheme similar to Algorithm 2, with the additional restriction that $\tau = 0$ in (CGMP). Furthermore, we compare the different strategies for computing scaled cuts described in Section 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 the L&P cutting planes described in Section 4.2.1.

We assess solution quality by comparing the lower and upper bound, denoted by LB and UB, respectively, maintained during the Benders' decomposition, see Algorithm 1. In particular, we are interested in the relative optimality gap

$$\rho := \frac{UB - LB}{|LB|} \times 100\%,$$

and the relative LB and UB gaps, defined as $(\eta^* - LB)/|\eta^*| \times 100\%$ and $(UB - \eta^*)/|\eta^*| \times 100\%$, respectively, where η^* is the optimal value of the original MIR model. We expect that these gaps are smaller if we use scaled cuts, compared to traditional optimality cuts.

In our implementation of Algorithm 1, we use a warm start in which we solve the continuous relaxation of the original model using the L-shaped algorithm by Van Slyke and Wets (1969). Furthermore, we solve the master problem (MP) using branch-and-cut techniques if some of the first-stage decisions are integer. Interestingly, since scaled cuts can be used to recover $\text{co}(Q)$, a branch-and-cut scheme can converge to the optimal integer solution at the *root node*. Indeed, recall that instead of solving the original problem (1), we can equivalently solve its convex relaxation in (4). That is why we use a pure cutting plane approach to solve (MP), in which we use Fenchel cuts (Boyd 1994) to cut away non-integer solutions. In addition, for the larger DCAP problems, we also use a branch-and-cut scheme in which we add at most five Fenchel cuts to solve the nodal subproblems if the number of leaf nodes is less than eight. In this scheme, we maintain separate outer approximations for each nodal subproblem to speed up convergence, since they are potentially stronger than a global outer approximation.

In our experiments, we use parallelized implementations of all optimality cut computation routines, which exploit that the computations decompose by scenario. All our experiments are run on a machine with two Intel Xeon E5 2680v3 CPUs (24 cores @2.5GHz) and 128GB RAM using Gurobi 9.1.0; 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 1 if the outer approximation improves by less than ε, 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.

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 (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 1, we take

$$X' = \{x \in X : c^\top x + \hat{Q}_{\text{out}}(x) \leq UB\},$$

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 (1). Indeed, note that if $c^\top x + \hat{Q}_{\text{out}}(x) > UB$, then x is not optimal in (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 the cut-enhancement technique to speed up convergence of scaled cuts. 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^*. Note that we only use enhanced scaled cuts, since we expect them to outperform ordinary scaled cuts.

6.3 Results

6.3.1 The CS Instance. Carøe and Schultz (1999) 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 $h_\omega^s = \varepsilon^s$ and $h_\omega^{s+r/2} = 1/4 - \varepsilon^s$, 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 exploit parallelization 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 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 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 by Sherali and Fraticelli (2002). We report the results in Table 1.

Table 1: CS Instance.

Cut type	Lower bound (gap to $\eta^* = 0.2482$)	Upper bound (gap to $\eta^* = 0.2482$)	Cpu time	#Cuts (avg. cpu time)
Traditional cuts				
Benders	-0.0080 (103.21%)	0.7482 (201.48%)	0.168s	9 (0.003s)
SB	-0.0080 (103.21%)	0.7482 (201.48%)	0.364s	9 (0.021s)
L&P	0.0083 (96.64%)	0.7482 (201.48%)	0.255s	9 (0.011s)
L	0.0083 (96.64%)	0.7482 (201.48%)	0.384s	8 (0.027s)
Scaled cuts				
S-CP (GMI)	0.2482 (0.00%)	0.2488 (0.24%)	0.838s	33 (0.015s)
S-CP (L&P)	0.2482 (0.00%)	0.2484 (0.09%)	0.767s	17 (0.031s)
S-RG	0.2482 (0.00%)	0.2482 (0.00%)	3.233s	17 (0.176s)

What is immediately striking from Table 1 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 computing 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, as expected, 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.

6.3.2 Investment Planning Problems. Schultz et al. (1998) 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

$$\begin{aligned} v_\omega(x) = \min_{y\in\mathcal{Y}}\{-16y_1 - 19y_2 - 23y_3 - 28y_4 : \; & 2y_1 + 3y_2 + 4y_3 + 5y_4 \le h^1_\omega - x_1 \\ & 6y_1 + \; y_2 + 3y_3 + \; y_4 \le h^2_\omega - x_2\}, \end{aligned}$$

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., Ahmed et al. (2004), Ntaimo (2013), Gade et al. (2014), and Qi and Sen (2017). 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) 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 2 and 3, 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 2: Investment Planning Problems ($\mathcal{X} = \mathbb{Z}^2_+$): Root Node Gaps.

Instance			Root node gap (computation time)					
$\mathcal{Y}$	T_ω	$\lvert\Omega\rvert$	SB	SB*	L	L*	S-CP (L&P)	S-RG
$\mathbb{Z}^4_+$	I_2	121	3.87% (0s)	1.63% (0s)	1.00% (0s)	0.06% (0s)	0.00% (1s)	0.00% (1s)
		441	4.08% (1s)	1.27% (2s)	1.03% (1s)	0.00% (2s)	0.00% (124s)	0.00% (2s)
		10201	4.23% (23s)	1.28% (46s)	1.05% (35s)	0.00% (46s)	1.15% (3h)	0.00% (76s)
$\mathbb{Z}^4_+$	H	121	9.54% (0s)	8.90% (0s)	0.84% (1s)	0.00% (1s)	0.00% (20s)	0.00% (1s)
		441	9.04% (1s)	8.10% (1s)	2.09% (2s)	0.00% (3s)	0.75% (348s)	0.00% (5s)
		10201	7.94% (15s)	6.85% (22s)	2.40% (52s)	2.40% (51s)	0.88% (3h)	0.00% (521s)
$\{0,1\}^4$	I_2	121	6.73% (0s)	6.36% (0s)	2.90% (0s)	2.90% (0s)	0.00% (20s)	0.00% (3s)
		441	6.89% (1s)	6.30% (2s)	2.84% (1s)	2.84% (1s)	0.00% (132s)	0.00% (11s)
		10201	7.03% (23s)	6.45% (35s)	2.84% (30s)	1.38% (63s)	0.36% (3h)	0.00% (117s)
$\{0,1\}^4$	H	121	10.82% (0s)	10.39% (0s)	2.60% (1s)	1.61% (1s)	0.00% (19s)	0.00% (2s)
		441	10.39% (1s)	9.73% (1s)	3.36% (1s)	3.36% (1s)	0.00% (201s)	0.00% (23s)
		10201	9.64% (11s)	8.83% (29s)	2.64% (30s)	2.64% (32s)	0.00% (8169s)	0.00% (230s)

Table 3: Investment Planning Problems ($\mathcal{X} = \mathbb{R}^2_+$): Root Node Gaps.

Instance			Root node gap (computation time)					
$\mathcal{Y}$	T_ω	$\|\Omega\|$	SB	SB*	L	L*	S-CP (L&P)	S-RG
$\mathbb{Z}^4_+$	I_2	121	7.71% (0s)	4.13% (2s)	1.00% (1s)	0.00% (1s)	0.00% (10s)	0.00% (2s)
		441	7.78% (0s)	5.66% (5s)	2.37% (3s)	2.02% (5s)	0.62% (3h)	0.14% (239s)
		10201	9.47% (8s)	8.55% (82s)	3.97% (163s)	3.89% (152s)	3.13% (3h)	2.00% (3h)
$\mathbb{Z}^4_+$	H	121	9.58% (0s)	9.10% (0s)	0.94% (1s)	0.00% (2s)	0.00% (65s)	0.00% (5s)
		441	10.06% (1s)	9.41% (1s)	3.80% (5s)	3.51% (7s)	0.00% (1643s)	0.00% (21s)
		10201	9.66% (14s)	9.00% (43s)	4.16% (172s)	4.14% (206s)	3.13% (3h)	1.41% (3h)
$\{0,1\}^4$	I_2	121	10.44% (0s)	10.32% (0s)	3.26% (1s)	1.52% (1s)	0.00% (55s)	0.00% (4s)
		441	9.45% (1s)	9.37% (1s)	4.90% (3s)	4.88% (4s)	1.16% (3h)	1.15% (643s)
		10201	11.14% (20s)	11.07% (21s)	6.06% (88s)	6.04% (103s)	3.40% (3h)	2.86% (3h)
$\{0,1\}^4$	H	121	10.89% (0s)	10.50% (0s)	4.73% (1s)	4.72% (2s)	0.00% (1142s)	0.00% (23s)
		441	11.28% (0s)	10.94% (1s)	4.86% (5s)	4.85% (4s)	0.00% (916s)	0.00% (76s)
		10201	11.18% (10s)	10.80% (27s)	4.59% (128s)	4.56% (129s)	2.11% (3h)	1.50% (3h)

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 15%, or 1 percentage point, and the L* cuts improve over the L cuts on 19 out of 24 instances by approximately 45% on average, or 0.7 percentage points. In fact, on 6 instances, the enhanced L cuts were able to achieve a zero root node gap.

Second, we observe from Tables 2 and 3 that our scaled cuts are clearly outperform the traditional cuts: the S-RG cuts and the S-CP (L&P) cuts achieve a lower root node gap than the L cuts on 23 and 24 out of 24 instances, respectively. Furthermore, 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.

A comparison of the results in Tables 2 and 3 indicates that the instances with $X = \mathbb{R}^2_+$ are consistently harder to solve than the instances with $X = \mathbb{Z}^2_+$. For example, the S-RG cuts achieve a zero root node gap for all instances in Table 2, and half of the instances in Table 3. For the other instances in Table 3, the S-RG cuts reduce the average root node gap of the L cuts from 4.3% to 1.3%, which is a 70% reduction.

6.3.3 The DCAP Instances. The DCAP instances by Ahmed and Garcia (2003) 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 are larger than the investment planning problems, and differ in the number of resources, tasks, periods, and scenarios. Therefore, 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.

Furthermore, the first-stage problem has mixed-binary decision variables, which are used to model fixed set-up costs that we incur if we buy capacity. Thus, in order to investigate the performance of scaled cuts in the root node, we solve the master problem (MP) using a pure cutting plane approach, see the results in Table 4. In addition, we use the branch-and-cut scheme described in Section 6.1 to solve (MP), see Table 5.

Table 4: DCAP Instances: Root Node Gaps.

Instance	LB gap - UB gap (computation time)		
	SB*	SB*+L*	SB*+L*+S-RG
DCAP_233_200	26.78% - 3.68% (7s)	0.06% - 1.87% (542s)	0.01% - 0.17% (5416s)
DCAP_233_300	27.86% - 6.78% (17s)	0.08% - 0.28% (721s)	0.02% - 0.13% (6774s)
DCAP_233_500	30.24% - 11.12% (14s)	0.05% - 3.40% (906s)	0.01% - 0.34% (9188s)
DCAP_243_200	22.80% - 1.31% (12s)	0.08% - 0.75% (650s)	0.01% - 0.13% (7122s)
DCAP_243_300	22.69% - 1.39% (8s)	0.09% - 0.59% (1301s)	0.03% - 0.18% (3h)
DCAP_243_500	23.30% - 0.88% (13s)	0.09% - 0.56% (2362s)	0.02% - 0.25% (3h)
DCAP_332_200	44.53% - 51.42% (1s)	0.15% - 1.06% (319s)	0.06% - 1.06% (4343s)
DCAP_332_300	44.79% - 28.42% (1s)	0.20% - 0.20% (485s)	0.04% - 0.20% (3h)
DCAP_332_500	47.61% - 18.18% (1s)	0.12% - 0.55% (1122s)	0.08% - 0.55% (2862s)
DCAP_342_200	40.89% - 9.59% (6s)	0.13% - 9.15% (1045s)	0.05% - 2.12% (3h)
DCAP_342_300	40.92% - 7.15% (5s)	0.13% - 4.24% (1802s)	0.05% - 2.17% (3h)
DCAP_342_500	38.20% - 6.71% (18s)	0.10% - 2.61% (3272s)	0.04% - 1.50% (3h)

Table 5: DCAP Instances: Branch-and-Cut Scheme.

Instance	LB gap - UB gap (computation time)		
	SB*	SB*+L*	SB*+L*+S-RG
DCAP_233_200	26.78% - 3.68% (7s)	0.05% - 0.82% (190s)	0.00% - 0.00% (297s)
DCAP_233_300	27.87% - 6.78% (11s)	0.03% - 0.01% (281s)	0.00% - 0.01% (1993s)
DCAP_233_500	30.24% - 11.12% (12s)	0.04% - 0.77% (200s)	0.00% - 0.00% (1202s)
DCAP_243_200	22.80% - 1.31% (8s)	0.05% - 0.34% (264s)	0.01% - 0.13% (1726s)
DCAP_243_300	22.69% - 1.39% (11s)	0.09% - 0.25% (232s)	0.26% - 0.11% (3h)
DCAP_243_500	23.30% - 0.88% (13s)	0.07% - 0.46% (999s)	0.02% - 0.19% (3h)
DCAP_332_200	44.53% - 51.42% (1s)	0.30% - 1.15% (124s)	0.04% - 0.76% (4420s)
DCAP_332_300	44.79% - 28.42% (1s)	0.15% - 0.82% (197s)	0.04% - 1.27% (3h)
DCAP_332_500	47.61% - 18.18% (1s)	0.10% - 0.39% (617s)	0.03% - 0.48% (3h)
DCAP_342_200	40.74% - 9.59% (5s)	0.09% - 5.45% (258s)	0.10% - 4.57% (3h)
DCAP_342_300	40.92% - 7.15% (9s)	0.10% - 3.65% (329s)	0.00% - 0.17% (3h)
DCAP_342_500	38.22% - 6.71% (13s)	0.08% - 4.81% (410s)	0.02% - 0.73% (3h)

A first observation from the results in Tables 4 and 5 is that both cut hierarchies clearly outperform the stand-alone SB* cuts in terms of both LB and UB gaps. Moreover, similar as for the instances in Sections 6.3.1 and 6.3.2, our scaled cuts are able to significantly reduce the LB and UB gap in the root node. Indeed, compared to the SB*+L* cut hierarchy, including S-RG cuts in the hierarchy reduces the average LB and UB gaps by respectively 67% and 64%. However, for some instances, the UB gaps in the root node are still relatively large, but a comparison of the results in Tables 4 and 5 reveals that using a branch-and-cut scheme to solve (MP) further reduces the average UB gap on 9 out of 12 instances, and for these instances, the average UB gap decreases by roughly 61%. Closer inspection of the results in Table 4 reveals that the LB gaps achieved by the SB*+L* cut hierarchy in the root node are relatively small: they are at most 0.2%, and below 0.1% for 6 of the 12 instances. That is why we also consider relaxations of the original DCAP instances, for which we expect that the SB* and L* cuts perform less well. In particular, we relax the binary requirements in the first-stage problem, which comes down to assuming that there are no fixed set-up costs associated with buying capacity. We report the results in Table 6.

Table 6: DCAP Relaxations: Root Node Gaps.

Instance (LSDE gap)[b]	LB gap - UB gap[a] (computation time)		
	SB*	SB*+L*	SB*+L*+S-RG
DCAP_233_200	32.74% - 4.93% (2s)	0.68% - 1.62% (116s)	0.11% - 0.26% (3h)
DCAP_233_300 (0.05%)	30.70% - 7.50% (4s)	0.82% - 1.30% (189s)	0.19% - 0.82% (3h)
DCAP_233_500	31.26% - 12.61% (4s)	0.47% - 1.10% (248s)	0.08% - 0.09% (3h)
DCAP_243_200	21.50% - 0.60% (7s)	0.45% - 0.60% (134s)	0.05% - 0.15% (3h)
DCAP_243_300	22.06% - 0.80% (9s)	0.51% - 0.80% (173s)	0.16% - 0.52% (3h)
DCAP_243_500 (0.23%)	22.22% - 0.45% (14s)	0.41% - 0.45% (358s)	0.04% - 0.45% (3h)
DCAP_332_200 (0.02%)	42.22% - 57.03% (1s)	1.40% - 3.69% (87s)	0.34% - 2.69% (3h)
DCAP_332_300	42.83% - 33.15% (1s)	1.27% - 7.04% (93s)	0.46% - 4.94% (3h)
DCAP_332_500	46.93% - 19.58% (4s)	0.46% - 1.46% (421s)	0.20% - 1.46% (3h)
DCAP_342_200	32.96% - 1.24% (5s)	0.57% - 1.24% (98s)	0.12% - 1.24% (3h)
DCAP_342_300	36.81% - 2.43% (8s)	0.55% - 0.77% (166s)	0.18% - 0.77% (3h)
DCAP_342_500 (0.22%)	33.14% - 1.99% (13s)	0.49% - 1.65% (340s)	−0.04% - 1.65% (3h)

[a]The LB and UB gaps are computed using the best known lower and upper bound on η^*, respectively, obtained by solving the large-scale deterministic equivalent (LSDE) MIP using Gurobi with 24 threads, and a time limit of 12 hours.
[b]We report the LSDE gap if it exceeds 0.01%.

Similar as for the original DCAP instances, our cut hierarchies achieve significantly better LB and UB gaps compared to the benchmark SB* cuts. As we expected, however, the LB gaps achieved by the SB*+L* hierarchy are noticeably larger compared to the original DCAP instances. Nonetheless, in line with our previous findings, we are able to significantly reduce the LB gaps by also including S-RG cuts: we achieve a 78% reduction, on average. Moreover, on 7 out of 12 instances, we have found better incumbent solutions, and for these instances, the average UB gap is reduced by 41% from 2.3% to 1.4%.

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 of solution quality and the optimality gap at the root node of the Benders' master problem.

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

The proofs of Lemmas 1-4 and Proposition 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 and 6 in Section 4, respectively. The proofs of these lemmas are independent of the results in Section 3. The proofs of Lemmas 1-4 can be read in the same order as they appear in the main text. We only remark that the proof of Proposition 1 depends on Lemma 3 and is for this reason given after the proof of that lemma.

Proof of Lemma 1. We have to show that

$$\sup_{\alpha,\beta,\tau} \{\alpha - \beta^\top \bar{x} - \tau\phi(\bar{x}) : \ (\alpha,\beta,\tau) \in \Pi_\omega(\phi)\} = v_\omega(\bar{x}), \tag{26}$$

and that the supremum in (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 (15), which we repeat here for convenience,

$$C_\omega(\rho) = \sup_{\alpha,\beta,\tau} \{\alpha - \beta^\top \bar{x} - (1+\tau)\rho : \ (\alpha,\beta,\tau) \in \Pi_\omega(\phi)\}. \tag{27}$$

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 (27) with $\rho = \phi(\bar{x})$ is attained.

We first show that the problem in (27) is feasible and bounded, so that the corresponding supremum is attained, using the polyhedrality of $\Pi_\omega(\phi)$ from Lemma 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 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 $(x^i, \theta^i, y^i) \in S_\omega^\phi$, $i = 1, \ldots, d$, denote the extreme points of $\operatorname{conv}(S_\omega^\phi)$. Then, there exist $\lambda^i \geq 0$, $i = 1, \ldots, d$ 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 $(x^i, \theta^i) \in \operatorname{epi}_X(\phi)$, and using the assumption that $(\bar{x}, \phi(\bar{x}))$ is an extreme point of $\operatorname{conv}(\operatorname{epi}_X(\phi))$ 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 2. We first show that $C(\cdot)$ is convex and continuous on $\mathrm{dom}(C)$, and that the supremum in (11) is attained for all $\rho \in \mathrm{dom}(C)$. We use the expression $C(\rho) = \mathbb{E}_\omega[C_\omega(\rho)]$, where $C_\omega(\rho)$ is defined in (27), and we use the polyhedral representation of $\Pi_\omega(\phi)$ in Lemma 5 to obtain that

$$C_\omega(\rho) = \sup_{\alpha,\beta,\tau} \{\alpha - \beta^\top \bar{x} - \rho(1+\tau) : q_\omega^\top y^i + \beta^\top x^i + \tau\theta^i \geq \alpha \ \forall i \in \{1,\ldots,d\},\ \tau \geq 0\}, \tag{28}$$

where (x^i, θ^i, y^i), $i = 1,\ldots,d$, are the extreme points of $\mathrm{conv}(S_\omega^\phi)$. In particular, since the LP in (28) is feasible and bounded for all $\rho \in \mathrm{dom}(C_\omega)$, the corresponding supremum is attained, and the corresponding value function $C_\omega(\cdot)$ is convex and continuous on $\mathrm{dom}(C_\omega)$ for every $\omega \in \Omega$. It then follows from $C(\rho) = \mathbb{E}_\omega[C_\omega(\rho)]$ and

$$\mathrm{dom}(C) = \bigcap_{\omega\in\Omega} \mathrm{dom}(C_\omega), \tag{29}$$

that $C(\cdot)$ is convex and continuous on $\mathrm{dom}(C)$, and that the corresponding supremum is attained.

To see that $C(\cdot)$ is strictly decreasing on $\mathrm{dom}(C)$, fix $\rho_1, \rho_2 \in \mathrm{dom}(C)$ 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

$$C(\rho_2) = \mathbb{E}_\omega\alpha_\omega - \mathbb{E}_\omega\beta_\omega^\top \bar{x} - (1 + \mathbb{E}_\omega\tau_w)\rho_2,$$

and, using the definition of $C(\rho_1)$, we obtain

$$C(\rho_1) \geq \mathbb{E}_\omega\alpha_\omega - \mathbb{E}_\omega\beta_\omega^\top \bar{x} - (1 + \mathbb{E}_\omega\tau_w)\rho_1,$$

from which it follows that $C(\rho_1) > C(\rho_2)$, using that $\rho_1 < \rho_2$ and $\mathbb{E}_\omega\tau_w \geq 0$.

To prove (iii), fix $\bar{\rho} \in \mathrm{dom}(C)$, and denote an optimal solution of (11) with $\rho = \bar{\rho}$ by $(\alpha_\omega, \beta_\omega, \tau_\omega)$, $\omega \in \Omega$. 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 (11).

Finally, to show that $\mathrm{dom}(C) = [\phi(\bar{x}), \infty)$ if $\bar{x} \in X$, we will prove the slightly more general expression

$$\mathrm{dom}(C) = \{\rho : (\bar{x}, \rho) \in \mathrm{conv}(\mathrm{epi}_{\mathrm{X}}(\phi))\}, \tag{30}$$

for arbitrary $\bar{x} \in \bar{X}$, which reduces to $\mathrm{dom}(C) = [\phi(\bar{x}), \infty)$ if $\bar{x} \in X$, since then $(\bar{x}, \rho) \in \mathrm{conv}(\mathrm{epi}_{\mathrm{X}}(\phi))$ if and only if $\rho \geq \phi(\bar{x})$. We prove (30) from (29), by showing that $\mathrm{dom}(C_\omega) = \{\rho : (\bar{x}, \rho) \in \mathrm{conv}(\mathrm{epi}_{\mathrm{X}}(\phi))\}$ for every $\omega \in \Omega$. To do so, we use expression for the dual LP of (28) from Lemma 6, which we repeat here for convenience:

$$\min_y \{q_\omega^\top y :\ (\bar{x}, \rho, y) \in \mathrm{conv}(S_\omega^\phi)\}. \tag{31}$$

In particular, we show that the dual LP in (31) is bounded and feasible if and only if $(\bar{x}, \rho) \in \mathrm{conv}(\mathrm{epi}_{\mathrm{X}}(\phi))$. 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 $(\bar{x}, \rho) \in \mathrm{conv}(\mathrm{epi}_{\mathrm{X}}(\phi))$, suppose that $(\bar{x}, \rho) \in \mathrm{conv}(\mathrm{epi}_{\mathrm{X}}(\phi))$, i.e., there exist $\lambda^i \geq 0$, and $(x^i, \theta^i) \in \mathrm{epi}_{\mathrm{X}}(\phi)$, $i = 1,\ldots,d'$, 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

$$\left(\bar{x}, \rho, \sum_{i=1}^{d'} \lambda^i y^i\right) = \sum_{i=1}^{d'} \lambda^i (x^i, \theta^i, y^i) \in \mathrm{conv}(S_\omega^\phi),$$

and thus $y := \sum_{i=1}^{d'} \lambda^i y^i$ is feasible in (31). The converse claim can be proved in a similar way. □

Proof of Lemma 3. Using the definition of ρ^* in (10), we have

$$\begin{aligned}\rho^* =& \min_\rho \left\{ \rho :\ \rho \geq \frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x}}{1+\mathbb{E}_\omega \tau_\omega}\ \ \forall (\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi),\ \omega \in \Omega \right\}\\ =& \min_\rho \left\{ \rho :\ \mathbb{E}_\omega \alpha_\omega - \mathbb{E}_\omega \beta_\omega^\top \bar{x} - \rho(1+\mathbb{E}_\omega \tau_\omega) \leq 0\ \ \forall (\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi),\ \omega \in \Omega \right\},\end{aligned}$$

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 (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 2. □

Proof of Lemma 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 2, and ρ^* is the unique solution of $C(\rho) = 0$ by Lemma 3.

In order to prove that $C(\rho_k) \to 0$, we rewrite the updating rule in (14) as

$$\rho_{k+1} - \rho_k = \frac{C(\rho_k)}{1+\mathbb{E}_\omega \tau_{\omega,k}}, \tag{32}$$

in which we use the notation $\tau_{\omega,k}$ to emphasize that the value of τ_ω depends on the iteration, i.e., $\tau_{\omega,k}$ corresponds to an optimal solution of the problem in (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 (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 $\mathbb{E}_\omega \tau_{\omega,k} \leq \bar{\tau}$ for all $k \geq 1$.

We derive such a $\bar{\tau}$ by using that $\Pi_\omega(\phi)$ is polyhedral, see Lemma 5. In particular, let $(\alpha_\omega^i, \beta_\omega^i, \tau_\omega^i)$, $i = 1, \dots, d$, and $(\hat{\alpha}_\omega^j, \hat{\beta}_\omega^j, \hat{\tau}_\omega^j)$, $j = 1, \dots, r$ 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 (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 (27) with $\rho = \rho_k$, $k \geq 1$, is a convex combination of the extreme points $(\alpha_\omega^i, \beta_\omega^i, \tau_\omega^i)$, $i = 1, \dots, d$, of $\Pi_\omega(\phi)$. Hence, we can take $\bar{\tau} = \mathbb{E}_\omega[\max_{i=1,\dots,d}\{\tau_\omega^i\}]$.

Finally, we show that if $C(\rho_k) < \delta$, then $\rho_k \geq \rho^* - \delta$. To this end, let $(\alpha_{\omega,k}, \beta_{\omega,k}, \tau_{\omega,k}) \in \Pi_\omega(\phi)$, $\omega \in \Omega$, be such that $C(\rho_k) = \mathbb{E}_\omega \alpha_{\omega,k} - \mathbb{E}_\omega \beta_{\omega,k}^\top \bar{x} - (1 + \mathbb{E}_\omega \tau_{\omega,k})\rho_k$, and use the definition of $C(\rho^*)$ to obtain that

$$C(\rho^*) \leq \mathbb{E}_\omega \alpha_{\omega,k} - \mathbb{E}_\omega \beta_{\omega,k}^\top \bar{x} - (1 + \mathbb{E}_\omega \tau_{\omega,k})\rho^*.$$

It follows that

$$C(\rho_k) - C(\rho^*) \geq (1 + \mathbb{E}_\omega \tau_{\omega,k})(\rho^* - \rho_k) \geq \rho^* - \rho_k,$$

and we obtain $\rho^* - \rho^k \leq \delta$ by substituting $C(\rho^*) = 0$ and $C(\rho_k) \leq \delta$, as desired. □

Proof of Proposition 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 (10). By Lemma 3, we know that $\rho^* = \min_\rho\{\rho :\ C(\rho) \le 0\}$, where $C(\rho)$ is defined as in (11). In particular, $C(\rho^*) \le 0$, and we distinguish two cases: $C(\rho^*) = 0$, and $C(\rho^*) < 0$.

If $C(\rho^*) = 0$, then by Lemma 2, the optimal value ρ^* of the problem in (10) is attained by some $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\phi)$, $\omega \in \Omega$, where $(\alpha_\omega, \beta_\omega, \tau_\omega)$ attains the optimal value $C_\omega(\rho^*)$ of (27). Furthermore, since the feasible region $\Pi_\omega(\phi)$ of (27) is a rational polyhedron by Lemma 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 $\Pi_\omega(\phi)$. By definition of $\mathrm{SCC}(\phi)$, we have

$$\mathrm{SCC}(\phi)(x) \ge \frac{\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_w \beta_\omega^\top x}{1 + \mathbb{E}_\omega \tau_\omega} \quad \forall x \in \bar{X}. \tag{33}$$

and, in addition, $(\mathbb{E}_\omega \alpha_\omega - \mathbb{E}_w \beta_\omega^\top \bar{x})/(1+\mathbb{E}_\omega \tau_\omega) = \rho^* = \mathrm{SCC}(\phi)(\bar{x})$, i.e., the optimality cut in (33) is valid and tight at $\bar{x}$. Moreover, the cut in (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

$$\rho^* = \min_\rho\{\rho :\ C(\rho) < \infty\}. \tag{34}$$

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 $\rho^* = \min_\rho\{\rho :\ C(\rho) \le 0\}$, and thus the continuity of $C(\cdot)$ established in Lemma 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 (30), we obtain from (34) that

$$\rho^* = \min_\rho\{\rho :\ (\bar{x}, \rho) \in \mathrm{conv}(\mathrm{epi}_{\mathrm{X}}(\phi))\}, \tag{35}$$

and using similar reasoning for arbitrary $x \in \bar{X}$, we have that

$$\mathrm{SCC}(\phi)(x) \ge \min_\rho\{\rho :\ (\bar{x}, \rho) \in \mathrm{conv}(\mathrm{epi}_{\mathrm{X}}(\phi))\} \quad \forall x \in \bar{X}.$$

Thus, if we define E_ϕ as the set of cut coefficients which define valid inequalities for $\mathrm{conv}(\mathrm{epi}_{\mathrm{X}}(\phi))$, i.e.,

$$E_\phi := \{(\alpha, \beta) :\ \theta + \beta^\top x \ge \alpha\ \forall (x, \theta) \in \mathrm{conv}(\mathrm{epi}_{\mathrm{X}}(\phi))\},$$

then every $(\alpha, \beta) \in E_\phi$ defines an optimality cut of the form

$$\mathrm{SCC}(\phi)(x) \ge \alpha - \beta^\top x\ \forall x \in \bar{X}. \tag{36}$$

In addition, using (35) and that the function ϕ is convex polyhedral, we obtain

$$\rho^* = \max_{\alpha,\beta}\{\alpha - \beta^\top \bar{x} :\ (\alpha, \beta) \in E_\phi\}, \tag{37}$$

i.e., if (α, β) is optimal in (37), then the cut in (36) is tight at $\bar{x}$. Moreover, the maximum in (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 Theorem 1 in Del Pia and Weismantel (2016) to obtain that $\mathrm{conv}(\mathrm{epi}_{\mathrm{X}}(\phi))$ is a rational polyhedron, and note that

$$E_\phi = \{(\alpha, \beta) :\ \theta^i + \beta^\top x^i \ge \alpha,\ i = 1, \ldots, d\},$$

where (x^i, θ^i), $i = 1, \ldots, d$, denote the (rational) extreme points of $\mathrm{conv}(\mathrm{epi}_{\mathrm{X}}(\phi))$, see, e.g., Perregaard and Balas (2001). □

Proof of Lemma 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

$$\begin{aligned}
||\mathbb{T}f - \mathbb{T}g||_\infty &= \sup_{x\in\bar{X}} \left| \sup_{\tau_\omega \geq 0} \left\{ \frac{\mathbb{E}_\omega \overline{\text{co}}(v_\omega + \tau_\omega f)(x)}{1+\mathbb{E}_\omega \tau_\omega} \right\} - \sup_{\tau_\omega \geq 0} \left\{ \frac{\mathbb{E}_\omega \overline{\text{co}}(v_\omega + \tau_\omega g)(x)}{1+\mathbb{E}_\omega \tau_\omega} \right\} \right| \\
&\leq \sup_{x\in\bar{X}} \sup_{\tau_\omega \geq 0} \left| \left\{ \frac{\mathbb{E}_\omega \left[\overline{\text{co}}(v_\omega + \tau_\omega f)(x) - \overline{\text{co}}(v_\omega + \tau_\omega g)(x)\right]}{1+\mathbb{E}_\omega \tau_\omega} \right\} \right| \\
&\leq \sup_{\tau_\omega \geq 0} \left\{ \frac{\mathbb{E}_\omega ||\overline{\text{co}}(v_\omega + \tau_\omega f) - \overline{\text{co}}(v_\omega + \tau_\omega g)||_\infty}{1+\mathbb{E}_w \tau_\omega} \right\} \\
&\leq \sup_{\tau_\omega \geq 0} \left\{ \frac{\mathbb{E}_\omega \tau_\omega}{1+\mathbb{E}_w \tau_\omega} \right\} ||f-g||_\infty \\
&= ||f-g||_\infty,
\end{aligned}$$

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 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

$$v_\omega^+(x,\theta) \geq \alpha' - {\beta'}^\top x - \tau'\theta \quad \forall (x,\theta) \in \text{epi}(f), \tag{38}$$

and

$$\alpha' - {\beta'}^\top \bar{x} - \tau'\bar{\theta} \geq v_\omega(\bar{x}) - \varepsilon. \tag{39}$$

Note that (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, (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 epi(f) is a closed set. Analogous to Corollary 3.47 in Rockafellar and Wets (2009), 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 epi(f), we have $\text{co}(v_\omega^+)(\bar{x},\bar{\theta}) = v_\omega^+(\bar{x},\bar{\theta})$ (Tawarmalani and Sahinidis 2002, Corollary 3). Hence, $\overline{\text{co}}(v_\omega^+)(\bar{x},\bar{\theta}) = \text{co}(v_\omega^+)(\bar{x},\bar{\theta}) = v_\omega^+(\bar{x},\bar{\theta}) = v_\omega(\bar{x})$. □

Acknowledgements

The research of Ward Romeijnders has been supported by grant 451-17-034 4043 from The Netherlands Organisation for Scientific Research (NWO). Moreover, we would like to thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster.

References

Ahmed S, Cabral FG, da Costa BFP (2020) Stochastic lipschitz dynamic programming. *Mathematical Programming* 1–39.

Ahmed S, Garcia R (2003) Dynamic capacity acquisition and assignment under uncertainty. *Annals of Operations Research* 124(1-4):267–283.

Ahmed S, Garcia R, Kong N, Ntaimo L, Parija G, Qiu F, Sen S (2015) SIPLIB: A Stochastic Integer Programming Test Problem Library. `https://www2.isye.gatech.edu/~sahmed/siplib`.

Ahmed S, Tawarmalani M, Sahinidis NV (2004) A finite branch-and-bound algorithm for two-stage stochastic integer programs. *Mathematical Programming* 100:355–377.

Anger G (1990) *Inverse problems in differential equations*, volume 79 (Springer Science & Business Media).

Angulo G, Ahmed S, Dey SS (2016) Improving the integer l-shaped method. *INFORMS Journal on Computing* 28(3):483–499.

Balas E, Ceria S, Cornuéjols G (1996) Mixed 0-1 programming by lift-and-project in a branch-and-cut framework. *Management Science* 42(9):1229–1246.

Balas E, Jeroslow RG (1980) Strengthening cuts for mixed integer programs. *European Journal of Operational Research* 4(4):224–234.

Balas E, Perregaard M (2003) 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.

Benders JF (1962) Partitioning procedures for solving mixed-variables programming problems. *Numerische mathematik* 4(1):238–252.

Boland N, Christiansen J, Dandurand B, Eberhard A, Linderoth J, Luedtke J, Oliveira F (2018) 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.

Boyd EA (1994) Fenchel cutting planes for integer programs. *Operations Research* 42(1):53–64.

Boyd EA (1995) On the convergence of Fenchel cutting planes in mixed-integer programming. *SIAM Journal on Optimization* 5(2):421–435.

Carøe CC, Schultz R (1999) Dual decomposition in stochastic integer programming. *Operations Research Letters* 24(1-2):37–45.

Carøe CC, Tind J (1998) L-shaped decomposition of two-stage stochastic programs with integer recourse. *Mathematical Programming* 83(1):451–464.

Del Pia A, Weismantel R (2016) Relaxations of mixed integer sets from lattice-free polyhedra. *Annals of Operations Research* 240(1):95–117.

Dyer M, Stougie L (2006) Computational complexity of stochastic programming problems. *Mathematical Programming* 106(3):423–432.

Falk JE (1969) Lagrange multipliers and nonconvex programs. *SIAM Journal on Control* 7(4):534–545.

Gade D, Küçükyavuz S, Sen S (2014) Decomposition algorithms with parametric Gomory cuts for two-stage stochastic integer programs. *Mathematical Programming* 144(1-2):39–64.

Gassmann HI, Ziemba WT, eds. (2013) *Stochastic Programming: Applications in Finance, Energy, Planning and Logistics*, volume 4 of *World Scientific Series in Finance* (World Scientific).

Georghiou A, Tsoukalas A, Wiesemann W (2020) A primal–dual lifting scheme for two-stage robust optimization. *Operations Research* 68(2):572–590.

Kim K, Mehrotra S (2015) A two-stage stochastic integer programming approach to integrated staffing and scheduling with application to nurse management. *Operations Research* 63(6):1431–1451.

Kleywegt AJ, Shapiro A, Homem-de Mello T (2002) The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization* 12(2):479–502.

Küçükyavuz S, Sen S (2017) An introduction to two-stage stochastic mixed-integer programming. *INFORMS TutORials in Operations Research*, 1–27.

Laporte G, Louveaux FV (1993) The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters* 13(3):133–142.

Louveaux FV, van der Vlerk MH (1993) Stochastic programming with simple integer recourse. *Mathematical Programming* 61(1-3):301–325.

Ntaimo L (2010) Disjunctive decomposition for two-stage stochastic mixed-binary programs with random recourse. *Operations Research* 58(1):229–243.

Ntaimo L (2013) Fenchel decomposition for stochastic mixed-integer programming. *Journal of Global Optimization* 55(1):141–163.

Ntaimo L, Tanner MW (2008) Computations with disjunctive cuts for two-stage stochastic mixed 0-1 integer programs. *Journal of Global Optimization* 41(3):365–384.

Perregaard M, Balas E (2001) Generating cuts from multiple-term disjunctions. *International Conference on Integer Programming and Combinatorial Optimization*, 348–360 (Springer).

Qi Y, Sen S (2017) 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.

Rockafellar RT (1970) *Convex analysis*. Number 28 (Princeton university press).

Rockafellar RT, Wets RJB (2009) *Variational analysis*, volume 317 (Springer).

Romeijnders W, Schultz R, van der Vlerk MH, Klein Haneveld WK (2016) A convex approximation for two-stage mixed-integer recourse models with a uniform error bound. *SIAM Journal on Optimization* 26(1):426–447.

Rudin W (1976) *Principles of mathematical analysis* (McGraw-Hill New York).

Schultz R (1995) On structure and stability in stochastic programs with random technology matrix and complete integer recourse. *Mathematical Programming* 70(1-3):73–89.

Schultz R, Stougie L, van der Vlerk MH (1998) Solving stochastic programs with integer recourse by enumeration: A framework using Gröbner basis reductions. *Mathematical Programming* 83(1-3):229–252.

Sen S, Higle JL (2005) The C^3 theorem and a D^2 algorithm for large scale stochastic mixed-integer programming: Set convexification. *Mathematical Programming* 104(1):1–20.

Sen S, Sherali HD (2006) Decomposition with branch-and-cut approaches for two-stage stochastic mixed-integer programming. *Mathematical Programming* 106(2):203–223.

Sherali HD, Fraticelli BMP (2002) 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.

Sherali HD, Zhu X (2006) On solving discrete two-stage stochastic programs having mixed-integer first-and second-stage variables. *Mathematical Programming* 108(2-3):597–616.

Tardella F (2004) On the existence of polyhedral convex envelopes. Floudas CA, Pardalos PM, eds., *Frontiers in global optimization*, 563–573 (Springer).

Tawarmalani M, Sahinidis NV (2002) Convex extensions and envelopes of lower semi-continuous functions. *Mathematical Programming* 93(2):247–263.

van der Laan N, Romeijnders W (2020) A loose Benders decomposition algorithm for approximating two-stage mixed-integer recourse models. *Mathematical Programming* 1–34.

Van Slyke RM, Wets R (1969) L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics* 17(4):638–663.

Wallace SW, Ziemba WT, eds. (2005) *Applications of Stochastic Programming*. MPS-SIAM Series on Optimization.

Zeng B, Zhao L (2013) Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters* 41(5):457–461.

Zhang M, Küçükyavuz S (2014) Finitely convergent decomposition algorithms for two-stage stochastic pure integer programs. *SIAM Journal on Optimization* 24(4):1933–1951.

Zou J, Ahmed S, Sun XA (2019) Stochastic dual dynamic integer programming. *Mathematical Programming* 175(1-2):461–502.