

On the ReLU Lagrangian Cuts for Stochastic Mixed Integer Programming

Haoyun Deng

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA,
hdeng81@gatech.edu

Weijun Xie

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA,
wxie@gatech.edu

We study stochastic mixed integer programs where both first-stage and recourse decisions can be mixed integers. A new family of Lagrangian cuts, termed “ReLU Lagrangian cuts,” is introduced by reformulating the nonanticipativity constraints using ReLU functions. These cuts can be integrated into scenario decomposition algorithms. Unlike the ordinary Lagrangian cuts, we prove that the inclusion of ReLU Lagrangian cuts is sufficient to solve the original stochastic mixed integer programs to optimality. Without solving the Lagrangian dual problems, we derive closed-form expressions for these cuts. Furthermore, to speed up the cut-generating procedures, we introduce linear programming-based techniques to enhance the cut coefficients. Numerical studies demonstrate the effectiveness of the proposed cuts compared to existing methods.

Key words: Stochastic Mixed Integer Programming; ReLU Lagrangian Cuts; Cutting Plane; L-shaped Cuts.

1. Introduction

Consider a two-stage Stochastic Mixed Integer Program (SMIP) with finite support of the form:

$$\min_{\mathbf{x} \in \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2}} \left\{ \mathbf{c}^\top \mathbf{x} + \sum_{s \in [N]} p_s \mathcal{Q}_s(\mathbf{x}) : \mathbf{A}\mathbf{x} \geq \mathbf{b} \right\}, \quad (1a)$$

where the local recourse function is defined as

$$\mathcal{Q}_s(\mathbf{x}) = \min_{\mathbf{y} \in \mathbb{Z}^{m_1} \times \mathbb{R}^{m_2}} \{ (\mathbf{q}^s)^\top \mathbf{y} : \mathbf{W}^s \mathbf{y} \geq \mathbf{h}^s - \mathbf{T}^s \mathbf{x} \}. \quad (1b)$$

Here, we let \mathbf{x} and \mathbf{y} represent the decisions in the first and second stages, respectively. In the objective function of the first-stage problem (1a), $\mathbf{c}^\top \mathbf{x}$ denotes the first-stage cost, and $\sum_{s \in [N]} p_s \mathcal{Q}_s(\mathbf{x})$ is the expected recourse function that takes the expectation of the second-stage cost (i.e., local recourse function) over finite support. Given a first-stage decision, the local recourse function in scenario s , denoted by $\mathcal{Q}_s(\mathbf{x})$, is determined by a mixed integer program (1b). In this second-stage problem, $\mathbf{q}^s, \mathbf{W}^s, \mathbf{T}^s, \mathbf{h}^s$ are the realizations of random parameters in scenario s , which are assumed to be rational (Louveaux and Schultz 2003). For notational convenience, we let $n = n_1 + n_2$ and $m = m_1 + m_2$. In addition, we make the following assumptions throughout the paper:

ASSUMPTION 1. *The SMIP (1a) has a relatively complete recourse, that is, for every feasible first-stage decision \mathbf{x} in (1a), the second-stage problem (1b) is feasible.*

ASSUMPTION 2. *There exists a universal lower bound L for local recourse functions.*

ASSUMPTION 3. *The feasible region of the first-stage problem is nonempty and bounded.*

Assumption 1 ensures that the conditions of the fundamental theorem of integer programming Schrijver (1998) hold. Assumptions 1 - 3 are standard in an SMIP setting Louveaux and Schultz (2003), which together imply that the SMIP (1a) is always feasible, and its optimal value is bounded from below. According to Assumption 3, we can shift the feasible region such that all variables are nonnegative, i.e., $\bar{X} \subseteq \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2} \cap ([0, B_1] \times \cdots \times [0, B_n])$.

1.1. Related literature

A SMIP is a widely used modeling paradigm for sequential decision-making under uncertainty, where decisions are made as uncertain parameters are revealed over time. However, solving the SMIP is usually computationally challenging. A common approach to addressing this challenge involves decomposing an SMIP into smaller scenario-based subproblems, which can be optimized independently and subsequently provide valid inequalities to the master problem. To accelerate this decomposition method, valid inequalities for the epigraph of the expected recourse function are iteratively added to enhance the solution quality of the master problem. These inequalities are commonly referred to as “cuts” in the literature.

Benders decomposition (Benders 1962, Van Slyke and Wets 1969) is one of the most classical methods, where the linear programming (LP) relaxation of subproblems is solved to generate cuts. This approach is only guaranteed to solve SMIP models with continuous second-stage decisions to optimality. Therefore, Benders cuts can only recover the epigraph of LP relaxation of a given local recourse function, since there exists an integrality gap between a local recourse problem and its LP relaxation in each scenario. To resolve this issue and improve the performance of Benders decomposition, one earlier effort is to obtain integer L-shaped cuts (Laporte and Louveaux 1993) for purely binary first-stage decisions. Angulo et al. (2016) improve this method by introducing an alternative cut-generating strategy and deriving new L-shaped cuts based on the explored recourse function values. In addition, disjunctive cuts (Sen and Higle 2005, Sen and Sherali 2006), Fenchel cuts (Gade et al. 2014), and Gomory cuts (Gade et al. 2014) have also been studied to solve an SMIP with purely binary first-stage variables. However, these methods struggle to handle problems with general mixed integer first-stage variables, even when the underlying distribution contains only one scenario.

Building on Benders decomposition, Zou et al. (2019) propose Lagrangian cuts and strengthened Benders cuts, which can be viewed as special cases of the Benders dual decomposition (BDD) framework later introduced by Rahmaniani et al. (2020). Unlike traditional Benders decomposition, which takes the duals of the LP relaxations of the local recourse functions, the BDD method introduces a copy of the first-stage variables in each subproblem, and enforces nonanticipativity constraints to ensure these copies are equal, similar to the dual decomposition method (Carøe and Schultz 1999). The key difference from Benders decomposition is that the BDD method retains the integrality of second-stage variables when taking the Lagrangian dual with respect to the nonanticipativity constraints. Strengthened Benders cuts are derived by optimizing Lagrangian functions, with Lagrangian dual multipliers equal to the Bender cuts coefficients. These cuts are parallel to the original Benders cuts but are shifted upward to have higher vertical intercepts by enforcing the integrality constraints in the inner minimization subproblems of the Lagrangian duals. For the Lagrangian cuts, the cut coefficients are optimal dual multipliers. This family of cuts is exact for binary first-stage variables, which guarantees convergence of the cutting plane method to solve the SMIP problem to optimality. It has also been shown to be sufficient to recover the convex hulls of epigraphs of local recourse functions (Chen and Luedtke 2022). However, generating such cuts can be computationally demanding. Existing methods rely on first-order approaches and require solving multiple MIPs that correspond to the inner minimization subproblems. Moreover, the Lagrangian dual problem may have multiple optimal solutions, though not all are effective Lagrangian cuts. For instance, Bansal and Küçükyavuz (2024) demonstrate that the integer L-shaped cut—known for its weak global approximation and resulting slow convergence—is, in fact, a Lagrangian cut. The existing literature improves the implementation of Lagrangian cuts primarily in three areas: accelerating the solution procedure of Lagrangian dual problems, selecting cuts with specific desirable properties, and improving the decomposition and cutting plane framework by incorporating additional cut families.

Rahmaniani et al. (2020) propose a three-phase implementation strategy in which Lagrangian cuts are generated at the final stage by heuristically solving Lagrangian duals using an inner approximation. They also suggest partially relaxing integrality constraints or fixing certain variables in subproblems and solving the Lagrangian dual to ϵ -optimality. Chen and Luedtke (2022) provide a new formulation that can be used to derive both optimality and feasibility cuts, with the cut coefficients restricted to the span of previous Benders cuts' coefficients under certain normalization. Recently, in the context of multistage SIMPs, Bansal and Küçükyavuz (2024) conduct computational studies on the alternating cut procedure for Lagrangian cuts and other valid cut families to improve overall performance on multistage SMIP problems. Füllner et al. (2024b) generate tight Lagrangian cuts with bounded coefficients using regularized local recourse problems, and Füllner

et al. (2024a) extend existing concepts in Benders decomposition literature to derive facet-defining, Pareto-optimal, or deep cuts using proper normalization techniques.

However, linear cuts, such as Lagrangian cuts, can at most recover the convex envelopes of local recourse functions. In addition, the expectation of the convex envelopes of the local recourse functions is not greater than, and can be strictly less than, the convex envelope of the expected recourse function in certain first-stage decisions (van der Laan and Romeijnders 2023). To address this issue, Zou et al. (2019) propose to approximate SMIP problems by stochastic integer programs with purely binary first-stage decisions and prove that under some assumptions, such as Lipschitz continuity, the number of binary variables required can be bounded based on the desired precision of the optimal solution. Other methods directly solve the SMIP problems. For example, Ahmed et al. (2022) introduce reverse norm cuts, which leverage the Lipschitz continuity of local recourse functions, and augmented Lagrangian cuts. The tightness of the reverse norm cuts is ensured by the strong duality of the augmented Lagrangian duals, as shown in Feizollahi et al. (2017). van der Laan and Romeijnders (2023) derive scale cuts for the expected recourse function and recently extend them on multistage SMIP problems Romeijnders and van der Laan (2024).

1.2. Summary of contributions

The main contributions of this work are summarized below.

- 1) We introduce a new family of nonlinear cuts, referred to as “ReLU Lagrangian cuts.” These cuts are effective for solving a two-stage SMIP that involve general mixed integer decisions in the first stage. Through establishing a strong duality theory, we show that the ReLU Lagrangian cuts are tight, enabling the recovery of the epigraphs for both the local recourse functions and the expected recourse function.
- 2) We compare ReLU Lagrangian cuts with existing cut families from two perspectives: (i) ReLU Lagrangian cuts provide outer approximations of local recourse epigraphs and their convex hulls that are at least as strong as those from existing methods, requiring fewer iterations before the cutting plane method terminates, and (ii) since existing cuts are special ReLU Lagrangian cuts, they can serve as starting points for generating the strong ReLU Lagrangian cuts.
- 3) We show the equivalence between traditional Lagrangian cuts and ReLU Lagrangian cuts for purely binary first-stage decisions. We propose a cut generation scheme that begins with integer L-shaped cuts and strengthens them by solving LPs. This approach overcomes a limitation in the existing literature, where obtaining ideal coefficients for Lagrangian cuts often requires solving multiple MILPs within an iterative procedure.
- 4) For pure integer first-stage decisions, we theoretically compare two alternative approaches for generating initial ReLU Lagrangian cuts: (i) focusing on the original space to generalize integer

L-shaped cuts, and (ii) using the binary expansion technique from Zou et al. (2019). For mixed-integer first-stage decisions, we propose both binary search and closed-form methods to derive the initial ReLU Lagrangian cuts. In both cases, the cuts are further strengthened using an LP-based approach.

Organization. In Section 2, we review the ordinary Lagrangian cuts and introduce the ReLU Lagrangian cuts. In Section 3 and Section 4, we study the properties of ReLU Lagrangian cuts for SMIPs with purely binary and general mixed integer first-stage decisions, respectively. Section 5 shows the numerical evidence of the effectiveness of the proposed ReLU Lagrangian cuts, and Section 6 concludes the paper.

Notation. For a given function f defined with domain S , let $\text{epi}_S(f) := \{(\mathbf{x}, \theta) \in S \times \mathbb{R} : \theta \geq f(\mathbf{x})\}$ denote its epigraph and $\text{conv}(\text{epi}_S(f))$ denote the convex hull of its epigraph. Let us define the sets $\bar{X} = \{\mathbf{x} \in \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2} : \mathbf{A}\mathbf{x} \geq \mathbf{b}\}$, $X = \{\mathbf{x} \in \{0, 1\}^n : \mathbf{A}\mathbf{x} \geq \mathbf{b}\}$, and $X^{LP} = \{\mathbf{x} \in [0, 1]^n : \mathbf{A}\mathbf{x} \geq \mathbf{b}\}$. Given a positive integer τ and a nonnegative integer $\ell \leq \tau$, we let $[\tau] = \{1, \dots, \tau\}$ and $[\ell, \tau] = \{\ell, \ell + 1, \dots, \tau\}$. A variable is bold when it is a vector.

2. ReLU Lagrangian Cuts

In this section, we introduce a new family of cuts, termed ReLU Lagrangian cuts, which generalize the ordinary Lagrangian cuts (Zou et al. 2019). Throughout this paper, unless otherwise specified, each cut is a local cut, that is, it is derived based on a given scenario $s \in [N]$.

2.1. Preliminary: ordinary Lagrangian cuts

Lagrangian cuts, as introduced in Zou et al. (2019), are derived by dualizing the nonanticipativity constraints. Namely, for a given feasible first-stage decision $\hat{\mathbf{x}} \in \bar{X}$ and a scenario $s \in [N]$, we have

$$\mathcal{Q}_s(\hat{\mathbf{x}}) = \inf_{\mathbf{x}} \{ \mathcal{Q}_s(\mathbf{x}) : \mathbf{x} = \hat{\mathbf{x}}, \mathbf{x} \in \bar{X} \} = \inf_{\mathbf{x}} \{ \theta : (\mathbf{x}, \theta) \in \text{epi}_{\bar{X}}(\mathcal{Q}_s), \mathbf{x} = \hat{\mathbf{x}} \}, \quad (2)$$

where the epigraphical set

$$\begin{aligned} \text{epi}_{\bar{X}}(\mathcal{Q}_s) &= \left\{ (\mathbf{x}, \theta) : \theta \geq \min_{\mathbf{y}} \{ (\mathbf{q}^s)^\top \mathbf{y} : \mathbf{T}^s \mathbf{x} + \mathbf{W}^s \mathbf{y} \geq \mathbf{h}^s, \mathbf{y} \in \mathbb{Z}^{m_1} \times \mathbb{R}^{m_2} \}, \mathbf{x} \in \bar{X} \right\} \\ &= \{ (\mathbf{x}, \theta) : \exists \mathbf{y} \in \mathbb{Z}^{m_1} \times \mathbb{R}^{m_2}, \theta \geq (\mathbf{q}^s)^\top \mathbf{y}, \mathbf{T}^s \mathbf{x} + \mathbf{W}^s \mathbf{y} \geq \mathbf{h}^s, \mathbf{x} \in \bar{X} \}. \end{aligned}$$

We observe that set $\text{epi}_{\bar{X}}(\mathcal{Q}_s)$ can be described by linear inequalities and integrality constraints. Therefore, its convex hull $\text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s))$ is a polyhedron. Following theorem 1 from Geoffrion (1974), We further obtain that

$$\mathcal{Q}_s(\hat{\mathbf{x}}) \geq \min_{\mathbf{x}} \{ \theta : (\mathbf{x}, \theta) \in \text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s)), \mathbf{x} = \hat{\mathbf{x}} \} \quad (3a)$$

$$= \max_{\boldsymbol{\pi}} \left\{ \min_{\mathbf{x}} \{ \mathcal{Q}_s(\mathbf{x}) + \boldsymbol{\pi}^\top (\hat{\mathbf{x}} - \mathbf{x}) : \mathbf{x} \in \bar{X} \} \right\}, \quad (3b)$$

where the equality follows from the strong duality of linear programming and the fact that minimizing a linear function over a set is equivalent to minimizing over its convex hull. Here, problem (3b) is a Lagrangian dual of the equivalent formulation (2) for a given first-stage decision. The LP (3a) is the primal characterization of the Lagrangian dual. This LP is feasible and bounded since $\hat{\mathbf{x}} \in \bar{X}$ and $\mathcal{Q}_s(\mathbf{x}) \geq L$ for all $\mathbf{x} \in \bar{X}$. Thus, there always exists an optimal dual multiplier $\boldsymbol{\pi}$ to (3b). Zou et al. (2019) further prove that the inequality of (3a) holds at equality when the first-stage variables are binary and introduce the following Lagrangian cuts.

DEFINITION 1 (LAGRANGIAN CUTS). Given a feasible first-stage decision $\hat{\mathbf{x}} \in \bar{X}$, let $\hat{\boldsymbol{\pi}}$ be optimal to the outer maximization problem of (3b). A Lagrangian cut takes the form

$$\theta \geq \mathcal{L}_s(\hat{\boldsymbol{\pi}}; \hat{\mathbf{x}}) + \hat{\boldsymbol{\pi}}^\top (\mathbf{x} - \hat{\mathbf{x}}), \quad (4)$$

where $\mathcal{L}_s(\boldsymbol{\pi}; \hat{\mathbf{x}}) := \min_{\mathbf{x}} \{ \mathcal{Q}_s(\mathbf{x}) + \boldsymbol{\pi}^\top (\hat{\mathbf{x}} - \mathbf{x}) : \mathbf{x} \in \bar{X} \}$.

By preserving the integrality constraints in each local recourse problem, the Lagrangian dual (3b) yields a stronger lower bound for the local recourse function value compared to the bound obtained from the LP relaxation used to derive a Benders cut. Under certain conditions, this lower bound is exact, and the resulting cut is referred to as a tight cut.

DEFINITION 2 (TIGHT CUTS). A cut generated for a function f at the incumbent solution $\hat{\mathbf{x}}$ is tight if the cut's corresponding hyperplane passes through the point $(\hat{\mathbf{x}}, f(\hat{\mathbf{x}}))$.

It is known (Zou et al. 2019) that when the first-stage decision variables are binary, the Lagrangian cuts are tight. However, this property does not extend to general mixed integer first-stage stochastic programs. In fact, the primal characterizations (3a) of the Lagrangian dual also define the convex envelopes of the local recourse functions, as formally proved in Füllner et al. (2024b). Furthermore, Chen and Luedtke (2022) showed that Lagrangian cuts are sufficient to describe the convex hulls of epigraphs of the local recourse functions, which we refer to as the “local convex hulls.”

PROPOSITION 1. [Theorem 3.9, Füllner et al. 2024b] *The primal characterization*

$$\inf_{\mathbf{x}} \{ \theta : (\mathbf{x}, \theta) \in \text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s)), \mathbf{x} = \hat{\mathbf{x}} \} = \text{co}(\mathcal{Q}_s(\hat{\mathbf{x}})),$$

where the convex envelope of \mathcal{Q}_s , denoted by $\text{co}(\mathcal{Q}_s) : \text{conv}(\bar{X}) \rightarrow \mathbb{R}$, is defined as $\text{co}(\mathcal{Q}_s)(\mathbf{x}) = \sup \{ g(\mathbf{x}) : g \text{ is convex and } g(\mathbf{z}) \leq \mathcal{Q}_s(\mathbf{z}), \forall \mathbf{z} \in \bar{X} \}$.

PROPOSITION 2. [Proposition 2, Theorem 3, Chen and Luedtke 2022] *Let us define $\mathcal{D}_{\hat{\mathbf{x}}}^s = \{ \boldsymbol{\pi} : \boldsymbol{\pi} \text{ is optimal to (3b) \}$ for any feasible first-stage decision $\hat{\mathbf{x}} \in \bar{X}$. Then, the local convex hull $\text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s)) = \{ (\mathbf{x}, \theta) \in \text{conv}(\bar{X}) \times \mathbb{R} : \theta \geq \mathcal{L}_s(\boldsymbol{\pi}; \hat{\mathbf{x}}) + \boldsymbol{\pi}^\top (\mathbf{x} - \hat{\mathbf{x}}), \forall \hat{\mathbf{x}} \in \bar{X}, \boldsymbol{\pi} \in \mathcal{D}_{\hat{\mathbf{x}}}^s \}$.*

The following corollaries provide insights into the properties of Lagrangian cuts.

COROLLARY 1. *Given a first-stage solution $\hat{\mathbf{x}} \in \bar{X}$, a dual multiplier $\boldsymbol{\pi} \in \mathcal{D}_{\hat{\mathbf{x}}}^s$ if and only if it satisfies the optimality condition:*

$$\mathcal{L}_s(\boldsymbol{\pi}; \hat{\mathbf{x}}) = \min_{\mathbf{x}} \{ \mathcal{Q}_s(\mathbf{x}) + \boldsymbol{\pi}^\top (\hat{\mathbf{x}} - \mathbf{x}) : \mathbf{x} \in \bar{X} \} \geq \text{co}(\mathcal{Q}_s)(\hat{\mathbf{x}}). \quad (5)$$

The result in Corollary 1 follows directly from Proposition 1 and the strong duality between (3a) and (3b).

COROLLARY 2. *A Lagrangian cut generated at $\hat{\mathbf{x}} \in \bar{X}$ is tight if and only if there exists $\boldsymbol{\alpha} \in \mathbb{R}^n$, $\boldsymbol{\alpha} \neq \mathbf{0}$, such that $\boldsymbol{\alpha}^\top \mathbf{x} + \theta \geq \boldsymbol{\alpha}^\top \hat{\mathbf{x}} + \mathcal{Q}_s(\hat{\mathbf{x}})$ for all $(\mathbf{x}, \theta) \in \text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s))$.*

Proof. See Appendix A.1. □

When $(\hat{\mathbf{x}}, \mathcal{Q}_s(\hat{\mathbf{x}}))$ is an extreme point of $\text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s))$, such a supporting hyperplane always exists, and a Lagrangian cut derived at this point is tight. The following corollary shows that a local convex hull can be characterized using tight Lagrangian cuts generated at its extreme points. Since the set $\text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s))$ is a polyhedron and \bar{X} is nonempty and bounded, we have that

COROLLARY 3. *Set $\text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s)) = \text{conv}\{(\mathbf{x}^k, \mathcal{Q}_s(\mathbf{x}^k)) : k \in K\} + \text{cone}\{(\mathbf{0}, 1)\}$, where $\{(\mathbf{x}^k, \mathcal{Q}_s(\mathbf{x}^k))\}_{k \in K}$ are the extreme points of $\text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s))$. In addition, the local convex hull can be represented as $\text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s)) = \{(\mathbf{x}, \theta) \in \text{conv}(\bar{X}) \times \mathbb{R} : \theta \geq \mathcal{Q}_s(\mathbf{x}^k) + \boldsymbol{\pi}^\top (\mathbf{x} - \mathbf{x}^k), \forall k \in K, \boldsymbol{\pi} \in \mathcal{D}_{\mathbf{x}^k}^s\}$.*

The above results show that by generating facet-defining Lagrangian cuts, we can efficiently recover the local convex hulls. Taking the average of the Lagrangian cuts over all scenarios, we obtain a cut that is valid for the expected recourse function. This cut is tight for the expected recourse function if and only if the cut for each individual scenario is tight. However, due to the linearity of a Lagrangian cut, tightness is achieved only if the incumbent solution used to construct the Lagrangian cut is separable from the local convex hull $\text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s))$ for each $s \in [N]$. Thus, this approach may yield only a lower estimate of the expected recourse function rather than an exact characterization, resulting in only a lower bound of (1a) (see the example below).

EXAMPLE 1. Consider a two stage problem $\min\{-x + \mathcal{Q}(x) : x \in \{0, 1, 2\}\}$, where $\mathcal{Q}(x) = \frac{1}{2}\mathcal{Q}_1(x) + \frac{1}{2}\mathcal{Q}_2(x)$ and local recourse functions are given by $\mathcal{Q}_1(x) = \min\{y : y \geq \frac{1}{2}x + 1, y \in \mathbb{Z}^+\}$ and $\mathcal{Q}_2(x) = \min\{y : y \geq 2x - 1, y \in \mathbb{Z}^+\}$. The values of the local recourse functions are $\mathcal{Q}_1(0) = 1, \mathcal{Q}_1(1) = 2, \mathcal{Q}_1(2) = 2$ (see Figure 1(a)) and $\mathcal{Q}_2(0) = 0, \mathcal{Q}_2(1) = 1, \mathcal{Q}_2(2) = 3$ (see Figure 1(b)). Thus, the expected recourse function values are $\mathcal{Q}(0) = \frac{1}{2}, \mathcal{Q}(1) = \frac{3}{2}, \mathcal{Q}(2) = \frac{5}{2}$. At $x = 1$, the optimal Lagrangian dual values yield $\text{co}(\mathcal{Q}_1)(1) = \frac{3}{2}$ and $\text{co}(\mathcal{Q}_2)(1) = 1$, resulting in an outer approximation of the expected recourse function: $\sum_{s \in [2]} p_s \text{co}(\mathcal{Q}_s)(1) = \frac{5}{4}$. However, as depicted in Figure 1(c), $\text{co}(\mathcal{Q})(1) = \mathcal{Q}(1) = \frac{3}{2}$. In fact, solving the problem solely using Lagrangian cuts provides only a

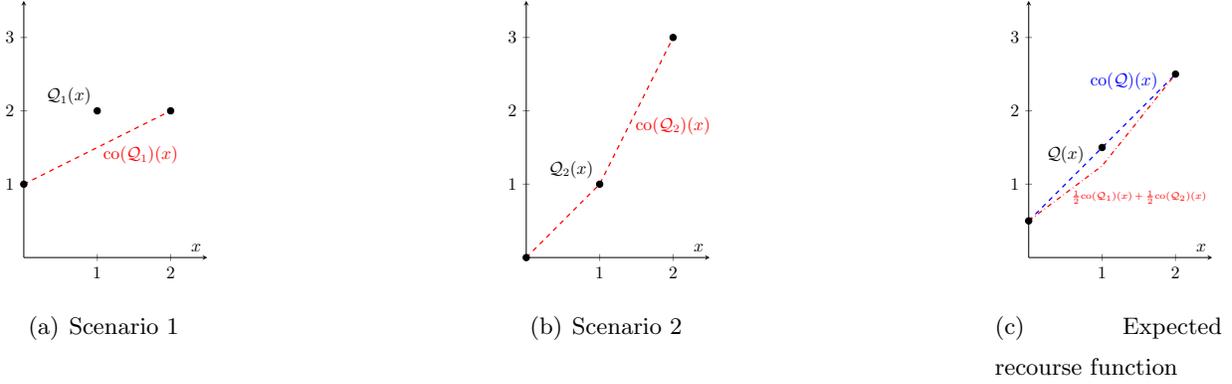


Figure 1 The illustration of Example 1

lower bound of $\frac{1}{4}$ with the solution $x = 1$. However, the optimal value of this two-stage problem is $\frac{1}{2}$, achieved at points $x = 0, 1, 2$. \diamond

This example indicates that, for a general mixed integer stochastic program, the convex envelope of the expected recourse function may be strictly greater than the expected convex envelope of the local recourse functions. However, to solve (1a) to optimality, we need an approximation that is at least as strong as the convex envelope of the expected recourse function. Unfortunately, Lagrangian cuts are often insufficient for general mixed integer first-stage decisions.

To address this issue, one approach is to derive linear cuts directly for the expected recourse function, such as the scaled cuts proposed in van der Laan and Romeijnders (2023), which cannot be computed using scenario decomposition algorithms. Another approach is to develop nonlinear cuts that recover the epigraphs of local recourse functions directly instead of their convex hulls, such as the reverse norm cuts and augmented Lagrangian cuts introduced in Ahmed et al. (2022). We will derive stronger cuts that remain tight for general mixed-integer first-stage decisions while preserving the effectiveness of Lagrangian cuts in representing local convex hulls.

2.2. ReLU Lagrangian cuts

In this section, we introduce the ReLU Lagrangian cuts. Using strong duality, we prove that the cut generated at any feasible first-stage decision is tight, directly recovering the epigraphs of both the local recourse function and its expectation. Unlike ordinary Lagrangian cuts, where nonanticipativity constraints are linear and only linear cuts are produced that are valid for the local convex hull, we study new nonanticipativity constraints using the ReLU function.

For a given feasible first-stage decision $\hat{\mathbf{x}} \in \bar{X}$, its local recourse function value can be obtained by solving

$$\mathcal{Q}_s(\hat{\mathbf{x}}) = \min_{\mathbf{x} \in \bar{X}} \{ \mathcal{Q}_s(\hat{\mathbf{x}}) : (x_i - \hat{x}_i)^+ = 0, (x_i - \hat{x}_i)^- = 0, \forall i \in [n] \}, \quad (6)$$

where we define two ReLU functions as $(x_i - \hat{x}_i)^+ = \max\{x_i - \hat{x}_i, 0\}$ and $(x_i - \hat{x}_i)^- = \max\{\hat{x}_i - x_i, 0\}$. Taking the dual of this problem with respect to the nonanticipativity constraints, we have

$$\underline{Q}_s(\mathbf{x}) = \sup_{\boldsymbol{\pi}^+, \boldsymbol{\pi}^- \in \mathbb{R}^n} \mathcal{L}_s(\boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}}), \quad (7)$$

where

$$\mathcal{L}_s(\mathbf{x}, \boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}}) := \inf_{\mathbf{x} \in \bar{X}} Q_s(\mathbf{x}) - \sum_{i \in [n]} \pi_i^+ (x_i - \hat{x}_i)^+ - \sum_{i \in [n]} \pi_i^- (x_i - \hat{x}_i)^-. \quad (8)$$

The difference between this ReLU Lagrangian dual (6) and the ordinary Lagrangian dual (3b) lies in the ReLU functions, which are nonlinear and can be either linearized or represented by introducing extra binary variables. We can prove that the strong duality holds for the ReLU Lagrangian dual (6), i.e., $Q_s(\hat{\mathbf{x}}) = \underline{Q}_s(\hat{\mathbf{x}})$. As a side product, we can derive optimal dual multipliers in closed form.

THEOREM 1. *Under Assumptions 1, 2 and 3, $\underline{Q}_s(\hat{\mathbf{x}}) = Q_s(\hat{\mathbf{x}})$. Moreover, when*

$$\rho^* \geq \frac{Q_s(\hat{\mathbf{x}}) - L}{d}, \quad (9)$$

$(-\mathbf{1}\rho^*, -\mathbf{1}\rho^*)$ is optimal to (7), where $d = \min\{\|\mathbf{x}^k - \hat{\mathbf{x}}\|_1 : k \in K, \mathbf{x}^k \neq \hat{\mathbf{x}}\}$, and $\{(\mathbf{x}^k, \mathbf{y}^k, \omega_k)\}_{k \in K}$ are extreme points of the set $\text{conv}\{(\mathbf{x}, \mathbf{y}, \omega) : \mathbf{T}^s \mathbf{x} + \mathbf{W}^s \mathbf{y} \geq \mathbf{h}^s, \mathbf{x} \in \bar{X}, \mathbf{y} \in \mathbb{Z}^{m_1} \times \mathbb{R}^{m_2}, \|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \omega\}$.

Proof. See Appendix A.2. \square

The above theorem shows that for any feasible first-stage decision, strong duality must hold. Besides, there also exists a finite dual optimal solution that attains the dual optimality. This ensures that tight cuts can be generated at any feasible point, a fundamental difference between the ReLU Lagrangian and the ordinary Lagrangian cuts. Before formally defining the ReLU Lagrangian cuts, we present the following corollaries, which are useful in subsequent discussions.

COROLLARY 4. *For any bounded set S such that $S \supseteq \bar{X}$, if $(\boldsymbol{\pi}^{+*}, \boldsymbol{\pi}^{-*})$ is optimal to $\sup_{\boldsymbol{\pi}^+, \boldsymbol{\pi}^- \in \mathbb{R}^n} \inf_{\mathbf{x} \in S} L_s(\mathbf{x}, \boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}})$, then it is also optimal to $\sup_{\boldsymbol{\pi}^+, \boldsymbol{\pi}^- \in \mathbb{R}^n} \inf_{\mathbf{x} \in \bar{X}} L_s(\mathbf{x}, \boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}})$, where*

$$L_s(\mathbf{x}, \boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}}) := Q_s(\mathbf{x}) - \sum_{i \in [n]} \pi_i^+ (x_i - \hat{x}_i)^+ - \sum_{i \in [n]} \pi_i^- (x_i - \hat{x}_i)^-.$$

Proof. See Appendix A.3

The following corollary provides an optimality condition for a dual optimal solution.

COROLLARY 5. *For a given feasible first-stage decision $\hat{\mathbf{x}} \in \bar{X}$, $(\hat{\boldsymbol{\pi}}^+, \hat{\boldsymbol{\pi}}^-)$ is optimal to (7) if and only if*

$$Q_s(\hat{\mathbf{x}}) \leq Q_s(\mathbf{x}) - \sum_{i \in [n]} \hat{\pi}_i^+ (x_i - \hat{x}_i)^+ - \sum_{i \in [n]} \hat{\pi}_i^- (x_i - \hat{x}_i)^-, \quad (10)$$

for all $x \in \bar{X}$, or equivalently,

$$\theta \geq Q_s(\mathbf{x}) \geq Q_s(\hat{\mathbf{x}}) + \sum_{i \in [n]} \hat{\pi}_i^+ (x_i - \hat{x}_i)^+ + \sum_{i \in [n]} \hat{\pi}_i^- (x_i - \hat{x}_i)^-,$$

for all (\mathbf{x}, θ) in the epigraph of Q_s .

Proof. See Appendix A.4. □

Now we are ready to formally define the ReLU Lagrangian cuts.

DEFINITION 3 (ReLU LAGRANGIAN CUTS). For a given feasible first-stage decision $\hat{\mathbf{x}} \in \bar{X}$, let $(\hat{\boldsymbol{\pi}}^+, \hat{\boldsymbol{\pi}}^-)$ be an optimal solution to the dual problem (7). The following cut is valid for the epigraph of \mathcal{Q}_s :

$$\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) + \sum_{i \in [n]} \hat{\pi}_i^+ (x_i - \hat{x}_i)^+ + \sum_{i \in [n]} \hat{\pi}_i^- (x_i - \hat{x}_i)^-. \quad (11)$$

When the first-stage decision is purely binary, that is, when $\bar{X} = X$, we have $x_i - \hat{x}_i \geq 0$ when $\hat{x}_i = 0$ and $x_i - \hat{x}_i \leq 0$ when $\hat{x}_i = 1$. In this case, given a binary $\hat{\mathbf{x}} \in X$, the ReLU Lagrangian cut simplifies to the following linear inequality:

$$\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) + \sum_{i \notin I_{\hat{\mathbf{x}}}} \hat{\pi}_i^+ (x_i - \hat{x}_i) - \sum_{i \in I_{\hat{\mathbf{x}}}} \hat{\pi}_i^- (x_i - \hat{x}_i), \quad (12)$$

where $I_{\hat{\mathbf{x}}} := \{i \in [n] : \hat{x}_i = 1\}$. However, when the first-stage decision is mixed integer, cut (11) is generally nonlinear. To incorporate this nonlinear cut effectively into the master problem when using the cutting plane method, we add the following constraints:

$$\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) + \sum_{i \in [n]} \hat{\pi}_i^+ \omega_i^+ + \sum_{i \in [n]} \hat{\pi}_i^- \omega_i^-, \quad (13a)$$

$$\omega_i^+ - \omega_i^- = x_i - \hat{x}_i, 0 \leq \omega_i^+ \leq (B_i - \hat{x}_i)z_i, 0 \leq \omega_i^- \leq \hat{x}_i(1 - z_i), \forall i \in [n], \quad (13b)$$

$$\mathbf{z} \in \{0, 1\}^n. \quad (13c)$$

The system of inequalities in (13) is tight in the sense that its continuous relaxation recovers the convex hull of a relaxed local epigraphical set. To be specific, let us define $\bar{B} = \times_{i \in [n]} [0, B_i] \cap \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2}$ as a relaxed domain of the first stage decision and a mixed integer set $S_1 = \{(\mathbf{x}, \theta) \in \bar{B} \times \mathbb{R} : (12)\}$ that consists of the relaxed domain set and a ReLU Lagrangian cut. The following proposition summarizes this result.

PROPOSITION 3. *Given a first-stage decision $\hat{\mathbf{x}} \in \bar{X}$, we have*

$$\text{conv}(S_1) = \{(\mathbf{x}, \theta) \in \times_{i \in [n]} [0, B_i] \times \mathbb{R} : \exists (\boldsymbol{\omega}^+, \boldsymbol{\omega}^-, \mathbf{z}) \in \mathbb{R}^n \times \mathbb{R}^n \times [0, 1]^n, (13a), (13b)\}. \quad (14)$$

Proof. See Appendix A.5. □

To obtain a dual optimal solution and derive a ReLU Lagrangian cut, ?? suggests selecting a sufficiently large ρ . However, it also indicates that as ρ approaches infinity, the resulting cut may become arbitrarily weak. In the following sections, we will show how to appropriately select ρ by leveraging existing families of cuts that, while tight, may be weak, and how to strengthen these cuts. Additionally, Corollary 4 presents an alternative approach: rather than solving (7) directly, we may solve the dual problem with an expanded first-stage feasible region. To verify whether an inequality qualifies as a ReLU Lagrangian cut, Corollary 5 provides a criterion: any inequality valid for the epigraph of a local recourse function in the form (11) is a ReLU Lagrangian cut at $\hat{\mathbf{x}}$.

2.3. Related cut families

In this subsection, we establish that ReLU Lagrangian cuts include all tight Lagrangian cuts, therefore retaining their effectiveness in describing the local convex hull. Moreover, we show that certain nonlinear cuts—such as reverse norm cuts and augmented Lagrangian cuts (Ahmed et al. 2022), as well as integer L-shaped cuts (Laporte and Louveaux 1993)—are special cases of ReLU Lagrangian cuts. We extend the concept of L-shaped cuts to derive new cuts for purely integer first-stage decisions, termed “ Λ -shaped cuts.” A comparison of ReLU Lagrangian cuts with existing cut families demonstrates their advantages and necessity in accurately describing both local convex hulls and local epigraphs.

We define the set of admissible ReLU Lagrangian cut coefficients.

DEFINITION 4 (ADMISSIBLE RELU LAGRANGIAN CUT COEFFICIENTS). For a given $\hat{\mathbf{x}} \in \bar{X}$, let set $\Pi_{\hat{\mathbf{x}}}^s := \{(\boldsymbol{\pi}^+, \boldsymbol{\pi}^-) \in \mathbb{R}^{2n} : (\boldsymbol{\pi}^+, \boldsymbol{\pi}^-) \text{ is optimal to (7)}\}$ denote all optimal solutions to the ReLU Lagrangian dual problem (7).

Our first result shows that

PROPOSITION 4. *Any tight Lagrangian cut is a ReLU Lagrangian cut.*

Proof. See Appendix A.6. □

As shown in Corollary 3, the local convex hull can be recovered using Lagrangian cuts that are tight at its extreme points. By this proposition, these cuts are also ReLU Lagrangian cuts.

COROLLARY 6. *For a given $\hat{\mathbf{x}} \in \bar{X}$, the following result must hold: $\{(\mathbf{x}, \theta) \in \text{conv}(\bar{X}) \times \mathbb{R} : \theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) - \sum_{i \in [n]} \pi_i^+ (x_i - \hat{x}_i)^+ - \sum_{i \in [n]} \pi_i^- (x_i - \hat{x}_i)^-, \forall \hat{\mathbf{x}} \in \bar{X}, (\boldsymbol{\pi}^+, \boldsymbol{\pi}^-) \in \Pi_{\hat{\mathbf{x}}}^s\} \subseteq \text{conv}(\text{epi } \mathcal{Q}_s)$.*

This local convex hull description property demonstrates the strength of the ReLU Lagrangian cuts, distinguishing them from two other cuts: reverse norm cuts and the integer L-shaped cuts.

DEFINITION 5 (REVERSE NORM CUTS, AHMED ET AL. 2022). If the local recourse function \mathcal{Q}_s is Lipschitz continuous with Lipschitz constant ρ under the L_1 -norm, i.e., $|\mathcal{Q}_s(\mathbf{x}) - \mathcal{Q}_s(\mathbf{x}')| \leq \rho \|\mathbf{x} - \mathbf{x}'\|_1$, for all $\mathbf{x}, \mathbf{x}' \in \bar{X}$, then given $\hat{\mathbf{x}} \in \bar{X}$, we can derive a cut:

$$\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) - \rho \|\mathbf{x} - \hat{\mathbf{x}}\|_1.$$

PROPOSITION 5. *Reverse norm cuts are ReLU Lagrangian cuts.*

Although reverse norm cuts can be applied to general mixed integer first-stage decisions, they are restricted by the requirement that local recourse functions must be Lipschitz continuous—a condition that is often difficult to satisfy when part of second-stage decision variables is discrete. The following integer L-shaped cuts can be derived using only the current recourse function value and a lower bound of the recourse function. It is applicable when the first-stage variables are binary.

DEFINITION 6 (INTEGER L-SHAPED CUTS, LAPORTE AND LOUVEAUX 1993). Let L be a valid lower bound of the local recourse function \mathcal{Q}_s defined in (1b). Given a feasible first-stage decision $\hat{\mathbf{x}} \in X$, an integer L-shaped cut generated at $\hat{\mathbf{x}}$ admits the form

$$\theta \geq (\mathcal{Q}_s(\hat{\mathbf{x}}) - L) \left(\sum_{i \in I_{\hat{\mathbf{x}}}} x_i - \sum_{i \notin I_{\hat{\mathbf{x}}}} x_i \right) - (\mathcal{Q}_s(\hat{\mathbf{x}}) - L) (|I| - 1) + L,$$

where $I_{\hat{\mathbf{x}}} := \{i \in [n] : \hat{x}_i = 1\}$.

We can equivalently write this cut as $\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) - (\mathcal{Q}_s(\hat{\mathbf{x}}) - L) \|\mathbf{x} - \hat{\mathbf{x}}\|_1$. Note that in the closed-form expression for the optimal dual solution (9), the distance d under L_1 -norm is at least one for integer first-stage decisions. This coincides with the coefficients of L-shaped cuts. Similarly, when the first-stage feasible region is purely integer, we generalize the idea of L-shaped cuts to derive the Λ -shaped cuts.

DEFINITION 7 (Λ -SHAPED CUTS). When the first-stage feasible region $\bar{X} \subseteq \mathbb{Z}^n$, given $\hat{\mathbf{x}} \in \bar{X}$ and a lower bound L of the recourse function, we can derive a cut

$$\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) - (\mathcal{Q}_s(\hat{\mathbf{x}}) - L) \|\mathbf{x} - \hat{\mathbf{x}}\|_1. \quad (15)$$

PROPOSITION 6. *L-shaped cuts and Λ -shaped cuts are ReLU Lagrangian cuts.*

The cuts discussed above either cannot be applied to general SMIPs or impose specific requirements on local recourse functions. Due to the symmetry of the ℓ_1 -norm, these cuts fail to recover local convex hulls, resulting in weaker cuts that may not effectively enhance the outer approximation of the expected epigraph during the solution procedure. The following example demonstrates the insufficiency of reverse norm cuts and Λ -shaped cuts in describing the local convex hull.

EXAMPLE 2. Given set $\bar{X} = \{(0,0)^\top, (0,1)^\top, (2,1)^\top, (0,3)^\top, (2,3)^\top, (1,4)^\top, (1,2)^\top\}$, consider a local recourse function given by $\mathcal{Q}_s(0,0) = 0$, $\mathcal{Q}_s(0,1) = \mathcal{Q}_s(2,1) = 1$, $\mathcal{Q}_s(0,3) = \mathcal{Q}_s(2,3) = 3$, $\mathcal{Q}_s(1,4) = 4$ and $\mathcal{Q}_s(1,2) = -10$. It is clear that the Lipschitz constant $\rho \geq \frac{3}{10}$. The strongest reverse norm cut we can derive is $\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) - \frac{3}{10} \|\mathbf{x} - \hat{\mathbf{x}}\|_1$ for all $\hat{\mathbf{x}} \in \bar{X}$.

Similarly, to derive Λ -shaped cuts, we note that $L = -10$ is the best lower bound of the local recourse function \mathcal{Q}_s . Then, we can derive cuts $\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) - (\mathcal{Q}_s(\hat{\mathbf{x}}) + 10) \|\mathbf{x} - \hat{\mathbf{x}}\|_1$ for all $\hat{\mathbf{x}} \in \bar{X}$.

It is easy to check that $x_1 = 1.5, x_2 = 0.5, \theta = 0$ is valid for all reverse norm cuts and Λ -shaped cuts. However, it is not in $\text{conv}(\text{epi}(\mathcal{Q}_s))$ since $\text{co}(\mathcal{Q}_s)(1.5, 0.5) = 0.5$. \diamond

The following family of cuts subsumes all previously discussed types and provides a stronger outer approximation of local convex hulls, which is also a special case of ReLU Lagrangian cut.

DEFINITION 8 (AUGMENTED LAGRANGIAN CUT AHMED ET AL. 2022). Given $\hat{\mathbf{x}} \in \bar{X}$, $\boldsymbol{\pi} \in \mathbb{R}^n$, $\rho \geq 0$, let $L_A^s(\boldsymbol{\pi}, \rho; \hat{\mathbf{x}}) = \min_{\mathbf{x}} \{ \mathcal{Q}_s(\mathbf{x}) + \boldsymbol{\pi}^\top(\hat{\mathbf{x}} - \mathbf{x}) + \rho \|\hat{\mathbf{x}} - \mathbf{x}\|_1 : \mathbf{x} \in \bar{X} \}$. The following augmented Lagrangian cut is valid for the epigraph of \mathcal{Q}_s :

$$\theta \geq L_A^s(\boldsymbol{\pi}, \rho; \hat{\mathbf{x}}) + \boldsymbol{\pi}^\top(\mathbf{x} - \hat{\mathbf{x}}) - \rho \|\mathbf{x} - \hat{\mathbf{x}}\|_1.$$

Given a tight augmented Lagrangian cut

$$\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) + \boldsymbol{\pi}^\top(\mathbf{x} - \hat{\mathbf{x}}) - \rho \|\mathbf{x} - \hat{\mathbf{x}}\|_1, \quad (16)$$

($\boldsymbol{\pi} = \mathbf{0}$ when it is a reverse norm cut), we can equivalently write it as

$$\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) + \sum_{i \in [n]} (\pi_i - \rho)(x_i - \hat{x}_i)^+ + \sum_{i \in [n]} (-\pi_i - \rho)(x_i - \hat{x}_i)^-. \quad (17)$$

Then, according to Corollary 5, we have

PROPOSITION 7. *Tight augmented Lagrangian cuts are ReLU Lagrangian cuts.*

Augmented Lagrangian cuts are sufficient to describe local convex hulls, as they reduce to ordinary Lagrangian cuts when penalty terms are omitted. However, ReLU Lagrangian cuts generally provide stronger approximations of local epigraphs, allowing for fewer cuts in the solution procedure. This phenomenon is illustrated in the following example.

EXAMPLE 3. Consider a local recourse function given by $\mathcal{Q}_s(0, 1) = 3$, $\mathcal{Q}_s(0, 2) = 2$, $\mathcal{Q}_s(0, 3) = 1$, $\mathcal{Q}_s(1, 0) = 5$, $\mathcal{Q}_s(1, 1) = \frac{15}{2}$, $\mathcal{Q}_s(1, 2) = 10$, $\mathcal{Q}_s(1, 3) = \frac{11}{2}$, $\mathcal{Q}_s(1, 4) = 1$, $\mathcal{Q}_s(2, 1) = 5$, $\mathcal{Q}_s(2, 2) = 4$, $\mathcal{Q}_s(2, 3) = 3$. At $\hat{\mathbf{x}} = (1, 2)^\top$, we can derive a ReLU Lagrangian cut

$$\theta \geq 10 - 6(x - 1)^+ - 8(x - 1)^- - \frac{9}{2}(x - 2)^+ - \frac{5}{2}(x - 2)^-. \quad (18)$$

As shown in Figure 2(a), this cut is tight at points $(0, 2, \mathcal{Q}_s(0, 2))^\top$, $(1, 0, \mathcal{Q}_s(1, 0))^\top$, $(1, 1, \mathcal{Q}_s(1, 1))^\top$, $(1, 2, \mathcal{Q}_s(1, 2))^\top$, $(1, 3, \mathcal{Q}_s(1, 3))^\top$, $(1, 4, \mathcal{Q}_s(1, 4))^\top$ and $(2, 2, \mathcal{Q}_s(2, 2))^\top$ of the epigraph. To describe the epigraph, we need add one more cut $\theta \geq 3 + (x_1 - 0)^+ - (x_2 - 1)^+ + (x_2 - 1)^-$ generated at $\hat{\mathbf{x}} = (0, 1)^\top$.

To describe the local epigraph with augmented Lagrangian cuts, one can check by enumeration that at least three cuts are required. At $\mathbf{x} = (1, 2)^\top$, cut (18) cannot be expressed as an augmented Lagrangian cut since the system of linear equations $\pi_1 + \rho = -6$, $-\pi_1 + \rho = -8$, $\pi_2 + \rho = -\frac{9}{2}$, $-\pi_2 + \rho = -\frac{5}{2}$ has no solution. In fact, at least two augmented Lagrangian cuts, for example, the following two cuts

$$\theta \geq 10 + (x_1 - 1) + \frac{5}{2}(x_2 - 2) - 7\|\mathbf{x} - \hat{\mathbf{x}}\|_1, \quad (19a)$$

$$\theta \geq 10 + (x_1 - 1) - \frac{9}{2}(x_2 - 2) - 7\|\mathbf{x} - \hat{\mathbf{x}}\|_1, \quad (19b)$$

are required to separate $(1, 0, \mathcal{Q}_s(1, 0))^\top$, $(1, 1, \mathcal{Q}_s(1, 1))^\top$, $(1, 2, \mathcal{Q}_s(1, 2))^\top$, $(1, 3, \mathcal{Q}_s(1, 3))^\top$ and $(1, 4, \mathcal{Q}_s(1, 4))^\top$ from the epigraph, as shown in Figure 2(b). \diamond

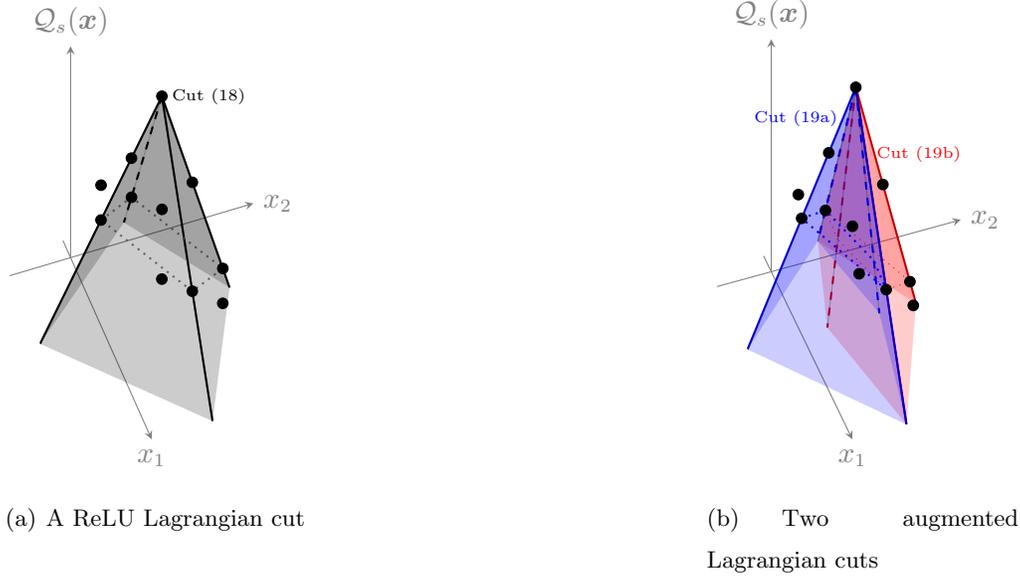


Figure 2 The illustration of Example 3.

We will further explain in Section 4.2 that the cut (18) is the strongest possible ReLU Lagrangian cut generated at $\hat{x} = (1, 2)^\top$ by analyzing the structure of the set $\Pi_{\hat{x}}^s$. Intuitively, from (17), a ReLU Lagrangian cut reduces to an augmented Lagrangian cut if and only if $\pi_i^+ = \pi_i - \rho$, $\pi_i^- = -\pi_i - \rho$ for all $i \in [n]$ and $\rho \geq 0$, which is equivalent to say $\pi_i^+ + \pi_i^- = \pi_j^+ + \pi_j^- \leq 0$ for all $i, j \in [n]$. This requirement restricts our choice of cut coefficients in $\Pi_{\hat{x}}^s$.

It is also worth noting that although both augmented Lagrangian cuts and ReLU Lagrangian cuts can describe the local convex hulls, they fail to describe the convex hull of the epigraph of the expected recourse function, as demonstrated in the following example:

EXAMPLE 4. Let $X = \{0, 1, 2, 3\}$ and define the local recourse functions as $\mathcal{Q}_1(x) = 0$ if $x \in \{0, 3\}$, $\mathcal{Q}_1(x) = 1$ if $x \in \{1, 2\}$, and $\mathcal{Q}_2(x) = 4$ if $x \in \{0, 3\}$, $\mathcal{Q}_2(x) = 1$ if $x \in \{1, 2\}$. The expected recourse function is then given by $\mathcal{Q}(x) = 2$ if $x \in \{0, 3\}$, and $\mathcal{Q}(x) = 1$ if $x \in \{1, 2\}$. For scenario 1, the strongest ReLU Lagrangian cut we can derive are $\theta \geq 0$ at $x = 0$ and $x = 3$, $\theta \geq 1 - (x - 1)^- - \frac{1}{2}(x - 1)^+$ at $x = 1$, and $\theta \geq 1 - \frac{1}{2}(x - 2)^- - (x - 2)^+$ at $x = 2$. For scenario 2, the strongest ReLU Lagrangian cuts we can derive are $\theta \geq 4 - 3(x - 0)^+$ at $x = 0$, $\theta \geq 1 + 3(x - 1)^-$ at $x = 1$, $\theta \geq 1 + 3(x - 2)^+$ at $x = 2$, and $\theta \geq 4 - 3(x - 3)^-$ at $x = 3$. Combining the two scenarios, we have $\theta \geq 2 - \frac{3}{2}(x - 0)^+$ at $x = 0$, $\theta \geq 1 + (x - 1)^- - \frac{1}{4}(x - 1)^+$ at $x = 1$, $\theta \geq 1 - \frac{1}{4}(x - 2)^- + (x - 2)^+$ at $x = 2$, and $\theta \geq 2 - \frac{3}{2}(x - 3)^-$ at $x = 3$. In Figure 3, the black dots represent the local and expected recourse functions, while the dash lines represent ReLU Lagrangian cuts. The gray areas depict the convex hulls of the epigraphs, and the shadowed areas are outer approximations derived by ReLU Lagrangian cuts. We can observe that for the expected recourse function, the point $(\frac{3}{2}, \frac{7}{8})^\top$ is contained in the outer approximation shaped by ReLU Lagrangian cuts, while it is not in the convex hull of the epigraph. \diamond

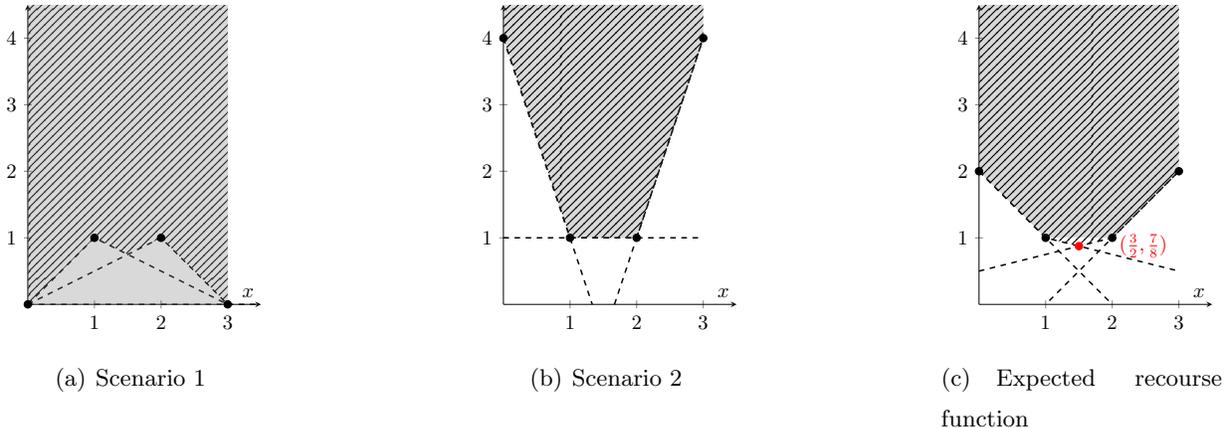


Figure 3 The illustration of Example 4.

This example also demonstrates that by adding the ReLU Lagrangian cuts, the outer approximation eventually converges to the epigraph of the expected recourse function when restricting the first-stage variables to their feasible region. That is, we have the following equalities:

$$v^* = \min_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} + \theta : (\mathbf{x}, \theta) \in \text{epi}_{\bar{X}}(\mathcal{Q}) \} = \min_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} + \theta : \mathbf{x} \in \bar{X}, (\mathbf{x}, \theta) \in E_R \},$$

where $E_R = \{ (\mathbf{x}, \theta) \in \mathbb{R}^n \times \mathbb{R} : \theta \geq \mathcal{Q}(\hat{\mathbf{x}}) - \sum_{i \in [N]} \pi_i^+(x_i - \hat{x}_i)^+ - \sum_{i \in [n]} \pi_i^-(x_i - \hat{x}_i)^-, \forall \hat{\mathbf{x}} \in \bar{X}, (\boldsymbol{\pi}^+, \boldsymbol{\pi}^-) \in \sum_{s \in [N]} p_s \Pi_{\hat{\mathbf{x}}}^s \}$. This is different from Lagrangian cuts, where the outer approximation only provides a lower bound on the optimal value.

The integrality constraints in the first stage of Example 4 are essential to obtain the optimal solution. It is insufficient to recover the convex hull of the epigraph by extending the feasible first-stage region to its convex hull and adding all ReLU Lagrangian cuts generated at feasible decisions. Specifically, since $\text{conv}\{(\mathbf{x}, \theta) \in \text{conv}(\bar{X}) \times \mathbb{R} : (\mathbf{x}, \theta) \in E_R\} \supseteq \text{conv}(\text{epi}_{\bar{X}} \mathcal{Q})$, we must have

$$\begin{aligned} v^* &= \min_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} + \theta : \mathbf{x} \in \bar{X}, (\mathbf{x}, \theta) \in E_R \} \geq \min_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} + \theta : \mathbf{x} \in \text{conv}(\bar{X}), (\mathbf{x}, \theta) \in E_R \} \\ &\geq \min_{\mathbf{x}} \{ \mathbf{c}^\top \mathbf{x} + \theta : \mathbf{x} \in \bar{X}^{LP}, (\mathbf{x}, \theta) \in E_R \}. \end{aligned}$$

Therefore, when implementing these cuts, maintaining the integrality constraints in the first stage ensures that we achieve the optimal value v^* .

Finally, we note that there is a more straightforward way to obtain a ReLU Lagrangian cut in practice than by directly solving the Lagrangian dual (7). This observation motivates our strategy of initially generating a valid, albeit potentially weak, cut at a low computational cost and subsequently refining it into a stronger cut. The methodology and implementation details will be elaborated upon in the following two sections.

3. Purely Binary First-Stage Decisions

Existing literature generates Lagrangian cuts by solving a sequence of optimization problems (see, e.g., Zou et al. 2019, Chen and Luedtke 2022), which can be computationally demanding. In this

section, we propose a simple and effective method for generating ReLU Lagrangian cuts when the first-stage decision variables are purely binary.

3.1. Dual optimal solution set $\Pi_{\hat{\mathbf{x}}}^s$

We first analyze the structure of $\Pi_{\hat{\mathbf{x}}}^s$, the set of all optimal solutions to the Lagrangian dual (7).

PROPOSITION 8. *Given scenario $s \in [N]$, the set $\Pi_{\hat{\mathbf{x}}}^s$ is a polyhedron.*

Proof. See Appendix A.7. □

This result holds for general mixed integer first-stage decisions. However, when the first-stage decision variables are purely binary, the set $\Pi_{\hat{\mathbf{x}}}^s$ is a non-pointed polyhedron. Intuitively, we observe that for any $(\boldsymbol{\pi}^+, \boldsymbol{\pi}^-) \in \Pi_{\hat{\mathbf{x}}}^s$, a ReLU Lagrangian cut (11) is equivalent to a linear cut (12). This linear cut is uniquely determined by the entries π_i^+ for $i \notin I_{\hat{\mathbf{x}}}$ and π_i^- for $i \in I_{\hat{\mathbf{x}}}$, while the remaining entries in the ReLU Lagrangian cut can take arbitrary values.

LEMMA 1. *Given scenario $s \in [N]$, when the first-stage feasible decision is purely binary, the recession cone of $\Pi_{\hat{\mathbf{x}}}^s$ contains $\text{lin}\{(\mathbf{e}_i, \mathbf{0})\}_{i \in I_{\hat{\mathbf{x}}}} + \text{lin}\{(\mathbf{0}, \mathbf{e}_i)\}_{i \notin I_{\hat{\mathbf{x}}}}$.*

Proof. See Appendix A.8. □

Given this lemma, we focus on the restriction of $\Pi_{\hat{\mathbf{x}}}^s$ to the orthogonal complement of the linear subspace $\text{lin}\{(\mathbf{e}_i, \mathbf{0})\}_{i \in I_{\hat{\mathbf{x}}}} + \text{lin}\{(\mathbf{0}, \mathbf{e}_i)\}_{i \notin I_{\hat{\mathbf{x}}}}$. Projecting out the entries that are fixed to be zero, we consider the set

$$\begin{aligned} & \text{Proj}_{([n] \setminus I_{\hat{\mathbf{x}}}, I_{\hat{\mathbf{x}}})} \Pi_{\hat{\mathbf{x}}}^s \Big|_{\text{lin}\{(\mathbf{e}_i, \mathbf{0})\}_{i \in I_{\hat{\mathbf{x}}}} + \text{lin}\{(\mathbf{0}, \mathbf{e}_i)\}_{i \notin I_{\hat{\mathbf{x}}}}} \\ & := \{ \boldsymbol{\pi} \in \mathbb{R}^n : \exists (\boldsymbol{\pi}^+, \boldsymbol{\pi}^-) \in \Pi_{\hat{\mathbf{x}}}^s : \pi_i = \pi_i^+, \forall i \notin I_{\hat{\mathbf{x}}}, \pi_i = \pi_i^-, \forall i \in I_{\hat{\mathbf{x}}} \} = \text{diag}(\boldsymbol{\chi}) \mathcal{D}_{\hat{\mathbf{x}}}^s, \end{aligned}$$

where we let $\chi_i = 2\hat{x}_i - 1$ for each $i \in [n]$ and $\text{diag}(\boldsymbol{\chi})$ is the diagonalization of the vector $\boldsymbol{\chi}$. Note that the diagonal matrix $\text{diag}(\boldsymbol{\chi})$ is symmetric and orthonormal. In Proposition 2, we define $\mathcal{D}_{\hat{\mathbf{x}}}^s$ as the set of all optimal solutions to the Lagrangian dual (3b) for the ordinary Lagrangian cuts. Thus, ReLU Lagrangian cuts coincide with ordinary Lagrangian cuts for purely binary variables. The orthonormal transformation $\text{diag}(\boldsymbol{\chi})$ preserves the polyhedral structure— for example, the extreme points and the extreme rays— of a polyhedron. Although Lagrangian cuts are sufficient to characterize the epigraphs of local recourse functions, many alternative Lagrangian cuts exist since the set $\mathcal{D}_{\hat{\mathbf{x}}}^s$ may not be a singleton. Our goal is to identify the strongest cuts that can reconstruct local convex hulls. This motivates us to seek facet-defining cuts of the local convex hulls.

DEFINITION 9. A cut of the form $\theta \geq v + \boldsymbol{\pi}^\top \mathbf{x}$ is facet-defining for $\text{conv}(\text{epi}_{\bar{\mathbf{x}}}(\mathcal{Q}^s))$ if it is valid for $\text{conv}(\text{epi}_{\bar{\mathbf{x}}}(\mathcal{Q}^s))$ and the corresponding hyperplane $\theta = v + \boldsymbol{\pi}^\top \mathbf{x}$ defines a facet of $\text{conv}(\text{epi}_{\bar{\mathbf{x}}}(\mathcal{Q}^s))$.

The following theorem demonstrates that the optimal solutions to the outer supremum of (3b) form a polyhedron, with facet-defining cuts corresponding to its extreme points.

THEOREM 2. *The following properties hold for the set $\mathcal{D}_{\hat{\mathbf{x}}}^s$ of dual optimal solutions and Lagrangian cuts generated at $\hat{\mathbf{x}} \in \bar{X}$: (i) The set $\mathcal{D}_{\hat{\mathbf{x}}}^s$ is a polyhedron; (ii) A Lagrangian cut $\theta \geq \text{co}(\mathcal{Q}_s)(\hat{\mathbf{x}}) + \boldsymbol{\pi}^\top(\mathbf{x} - \hat{\mathbf{x}})$ is facet-defining if and only if $\boldsymbol{\pi}$ is an extreme point of $\mathcal{D}_{\hat{\mathbf{x}}}^s$; (iii) The recession cone of $\mathcal{D}_{\hat{\mathbf{x}}}^s$ is the normal cone of \bar{X} at $\hat{\mathbf{x}}$; and (iv) The local convex hull can be described using the facet-defining Lagrangian cuts:*

$$\text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s)) = \{(\mathbf{x}, \theta) \in \text{conv}(\bar{X}) \times \mathbb{R} : \theta \geq \text{co}(\mathcal{Q}_s)(\hat{\mathbf{x}}) + \boldsymbol{\pi}^\top(\mathbf{x} - \hat{\mathbf{x}}), \forall \hat{\mathbf{x}} \in \bar{X}, \boldsymbol{\pi} \in \text{ext}(\mathcal{D}_{\hat{\mathbf{x}}}^s)\}.$$

Proof. See Appendix A.9. □

This result also suggests that as the cut coefficients shift in the direction of the recession cone of set $\mathcal{D}_{\hat{\mathbf{x}}}^s$, the resulting Lagrangian cuts may serve as valid inequalities only for $\text{conv}(\bar{X})$. However, these cuts do not contribute to the outer approximations of local epigraphs and should be avoided.

3.2. Cut strengthening: theoretical foundation

According to Proposition 6, we know that any integer L-shaped cut is a special ReLU Lagrangian cut, which is known to be weak (Zou et al. 2019). In this subsection, we present a method for deriving a stronger ReLU Lagrangian cut by strengthening the integer L-shaped cut coefficients.

Let us define $\hat{\boldsymbol{\pi}} = (L - \mathcal{Q}_s(\hat{\mathbf{x}}))\boldsymbol{\chi}$. We aim to find a proper $\boldsymbol{\eta}$ such that the strengthened cut:

$$\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) + (\hat{\boldsymbol{\pi}} + \boldsymbol{\eta})^\top(\mathbf{x} - \hat{\mathbf{x}}), \quad (20)$$

can be close to a non-trivial facet of $\text{conv}(\text{epi}_X(\mathcal{Q}_s))$. By Theorem 2, we know that (20) is a valid ReLU Lagrangian cut if and only if $\hat{\boldsymbol{\pi}} + \boldsymbol{\eta} \in \mathcal{D}_{\hat{\mathbf{x}}}^s$. Therefore, it suffices to find $\boldsymbol{\eta}$ such that $\hat{\boldsymbol{\pi}} + \boldsymbol{\eta}$ is an extreme point of $\mathcal{D}_{\hat{\mathbf{x}}}^s$. To identify an extreme point of a polyhedron, a natural approach is to solve a linear program:

$$\min \{\mathbf{a}^\top \boldsymbol{\eta} : \hat{\boldsymbol{\pi}} + \boldsymbol{\eta} \in \mathcal{D}_{\hat{\mathbf{x}}}^s\}. \quad (\text{Stren})$$

Note that the feasible region of this problem, denoted by $\mathcal{D}_{\hat{\mathbf{x}}}^s - \hat{\boldsymbol{\pi}} := \{\boldsymbol{\eta} : \mathcal{Q}_s(\hat{\mathbf{x}}) \leq \mathcal{Q}_s(\mathbf{x}) + (\hat{\boldsymbol{\pi}} + \boldsymbol{\eta})^\top(\hat{\mathbf{x}} - \mathbf{x}), \forall \mathbf{x} \in X\} = \{\boldsymbol{\eta} : (\hat{\mathbf{x}} - \mathbf{x})^\top \boldsymbol{\eta} \geq \mathcal{Q}_s(\hat{\mathbf{x}}) + \hat{\boldsymbol{\pi}}^\top(\hat{\mathbf{x}} - \mathbf{x}) - \mathcal{Q}_s(\mathbf{x}), \forall \mathbf{x} \in X\}$ is unbounded according to Theorem 2. We illustrate this with the following example:

EXAMPLE 5. Consider the recourse function $\mathcal{Q}(x_1, x_2) = \min\{2y_1 + 2y_2 : 0.2y_1 + y_2 + x_1 + 0.5x_2 \geq 2.4, \mathbf{y} \in \{0, 1, 2\}^2\}$, where $\mathbf{x} \in \{0, 1\}^2$. This function takes the values $\mathcal{Q}(0, 0) = 8$, $\mathcal{Q}(1, 0) = 4$, $\mathcal{Q}(0, 1) = 4$ and $\mathcal{Q}(1, 1) = 2$. Let $\hat{\mathbf{x}} = (1, 0)^\top$ and $L = 0$ be a lower bound of \mathcal{Q} . We can derive an L-shaped cut $\theta \geq 4 + 4(x_1 - 1) - 4x_2$. To enhance this cut, we consider the feasible region of

the strengthening problem (Stren), given by $\mathcal{D}_{\hat{\mathbf{x}}}^s - \hat{\boldsymbol{\pi}} = \{\boldsymbol{\eta} : \eta_1 \geq -8, \eta_1 - \eta_2 \geq -8, \eta_2 \leq 2\}$ (see Figure 4(a)). It has two extreme points $(-8, 0)^\top$ and $(-6, 2)^\top$, corresponding to two facet-defining cuts:

$$\theta \geq 4 - 4(x_1 - 1) - 4x_2, \quad (21a)$$

passing through $(0, 0, 8)^\top$, $(1, 0, 4)^\top$ and $(0, 1, 4)^\top$, and

$$\theta \geq 4 - 2(x_1 - 1) - 2x_2, \quad (21b)$$

passing through $(1, 0, 4)^\top$, $(0, 1, 4)^\top$ and $(1, 1, 2)^\top$. Meanwhile, $\mathcal{D}_{\hat{\mathbf{x}}}^s - \hat{\boldsymbol{\pi}}$ is unbounded with extreme rays $(1, 0)^\top$ and $(0, -1)^\top$. The strengthened cut $\theta \geq 4 + (4 + \eta_1)(x_1 - 1) + (-4 + \eta_2)(x_2 - 0)$ approaches $x_1 \leq 1$ as $\eta_1 \rightarrow \infty$ and $x_2 \geq 0$ as $\eta_2 \rightarrow -\infty$. \diamond

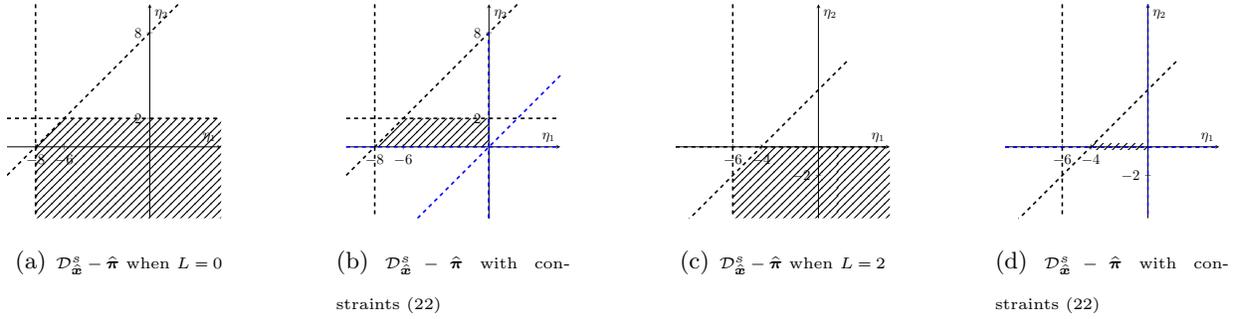


Figure 4 The illustration of Examples 5 and 6

We propose two strategies to prevent the unboundedness of the strengthening problem (Stren).

• **Strategy 1: Introduce additional constraints to bound the feasible region $F_{\boldsymbol{\eta}}$**

When set X is full-dimensional¹, we can restrict $\boldsymbol{\eta}$ to lie in the reverse direction of the normal cone of X . Specifically, the constraints we add are

$$(\hat{\mathbf{x}} - \mathbf{x})^\top \boldsymbol{\eta} \leq 0, \quad \forall \mathbf{x} \in X. \quad (22)$$

In this way, the recession cone of the feasible region becomes $\{\mathbf{d} : (\hat{\mathbf{x}} - \mathbf{x})^\top \mathbf{d} \geq 0, (\hat{\mathbf{x}} - \mathbf{x})^\top \mathbf{d} \leq 0, \forall \mathbf{x} \in X\} = \{\mathbf{d} : (\hat{\mathbf{x}} - \mathbf{x})^\top \mathbf{d} = 0, \forall \mathbf{x} \in X\} = \{\mathbf{0}\}$, where the last equality follows from the full dimensionality of X . However, this approach may eliminate extreme points that satisfy $(\hat{\mathbf{x}} - \mathbf{x})^\top \boldsymbol{\eta} > 0$ for some $\mathbf{x} \in X$. In general, we can add box constraints:

$$\underline{M}_i \leq \eta_i \leq \overline{M}_i, \quad i \in [n]. \quad (23)$$

One possibility is to ensure that $\text{diag}(\boldsymbol{\chi})^{-1} \boldsymbol{\eta} \geq \mathbf{0}$ and $\text{diag}(\boldsymbol{\chi})^{-1} \boldsymbol{\eta} \leq (\mathcal{Q}_s(\hat{\mathbf{x}}) - L)\mathbf{1}$, which implies that the absolute values of the entries of $\boldsymbol{\eta}$ are in the range $[0, \mathcal{Q}_s(\hat{\mathbf{x}}) - L]$. When $|\underline{M}_i|$ and $|\overline{M}_i|$ are large enough, we can retain all extreme points of the original set $\mathcal{D}_{\hat{\mathbf{x}}}^s - \hat{\boldsymbol{\pi}}$. A drawback of this strategy is that additional constraints can introduce extra extreme points.

¹If set X is not full-dimensional, we can choose $\boldsymbol{\eta}$ to be in the null space of affine hull of set X .

EXAMPLE 6. In Example 5, at $\hat{\mathbf{x}} = (1, 0)^\top$, adding constraints (22) to $\mathcal{D}_{\hat{\mathbf{x}}}^s - \hat{\boldsymbol{\pi}}$, the feasible region of (Stren) becomes $\{\boldsymbol{\eta} : -8 \leq \eta_1 \leq 0, -8 \leq \eta_1 - \eta_2 \leq 0, 0 \leq \eta_2 \leq 2\}$. It has the original extreme points and includes two new extreme points $(0, 0)^\top$ and $(0, 2)^\top$ (see Figure 4(b)). When adding constraints $\eta_1 \leq 0, \eta_2 \geq 0$ to bound the entries of $\boldsymbol{\eta}$, we get the same result.

If we take $L = 2$ as the lower bound of the recourse function \mathcal{Q} , the L-shaped cut becomes $\theta \geq 4 + 2(x_1 - 1) - 2x_2$, and $\mathcal{D}_{\hat{\mathbf{x}}}^s - \hat{\boldsymbol{\pi}} = \{\boldsymbol{\eta} : \eta_1 \geq -6, \eta_1 - \eta_2 \geq -4, \eta_2 \leq 0\}$. It has extreme points $(-6, -2)^\top$ and $(-4, 0)^\top$ (see Figure 4(c)). By adding constraints (22), the feasible region of (Stren) becomes $\{\boldsymbol{\eta} : -4 \leq \eta_1 \leq 0, \eta_2 = 0\}$, which excludes the extreme point $(-6, -2)^\top$ (see Figure 4(d)). If we bound the entries of $\boldsymbol{\eta}$ by adding upper bounds and lower bounds with sufficiently large absolute values, for example, $|\eta_1| \leq 6, |\eta_2| \leq 6$, the feasible region of (Stren) becomes $F'_\eta = \{-6 \leq \eta_1 \leq 6, -6 \leq \eta_2 \leq 0, \eta_1 - \eta_2 \geq 4\}$. In this case, we have both extreme points but also introduce three new extreme points $(6, 0)^\top, (6, -6)^\top$ and $(-6, -6)^\top$. They correspond to cuts $\theta \geq 4 + 8(x_1 - 1) - 2x_2, \theta \geq 4 + 8(x_1 - 1) - 8x_2$, and $\theta \geq 4 - 4(x_1 - 1) - 8x_2$, which are dominated by cuts (21a), (21b). \diamond

- **Strategy 2: Choose an appropriate coefficient \mathbf{a} of the objective function**

We restrict the objective coefficient to lie in a reverse direction of the tangent cone of $\text{conv}(X)$ at $\hat{\mathbf{x}}$, i.e., $-\mathbf{a} \in \mathcal{T}_{\text{conv}(X)}(\hat{\mathbf{x}}) := \text{cone}\{\mathbf{x} - \hat{\mathbf{x}} : \mathbf{x} \in X\}$.

PROPOSITION 9. *The strengthening problem (Stren) is bounded if and only if $-\mathbf{a} \in \mathcal{T}_{\text{conv}(X)}(\hat{\mathbf{x}})$.*

Proof. See Appendix A.10. \square

A practical choice is to let $\mathbf{a} = \sum_{i \in I} (\hat{\mathbf{x}} - \mathbf{x}^i)$, where (\mathbf{x}^i) 's are all explored first-stage decisions.

EXAMPLE 7. In Example 5, since $\hat{x}_1 = 1$ and $\hat{x}_2 = 0$, we let $a_1 = -1$ and $a_2 = 1$. The strengthening problem (Stren) then becomes $\min\{\eta_1 - \eta_2 : \eta_1 \geq -8, \eta_1 - \eta_2 \geq -8, \eta_2 \leq 2\}$. Both extreme points $(-8, 0)^\top$ and $(-6, 2)^\top$ and their convex combinations are optimal. \diamond

3.3. Cut strengthening – practical implementations based on LP relaxation

In this subsection, we integrate Strategies 1 and 2 within a practical framework to efficiently strengthen the ReLU Lagrangian cuts based on L-shaped ones. Although the feasible region of (Stren) is a polyhedron, it is generally NP-hard to separate from it. As an alternative, we replace the set X with its LP relaxation, X^{LP} . Thus, rather than solving the original strengthening problem

$$\min_{\boldsymbol{\eta}} \left\{ \mathbf{a}^\top \boldsymbol{\eta} : \min_{\mathbf{x}} \left\{ \mathcal{Q}_s(\mathbf{x}) + (\hat{\boldsymbol{\pi}} + \boldsymbol{\eta})^\top (\hat{\mathbf{x}} - \mathbf{x}) : \mathbf{x} \in X \right\} \geq \mathcal{Q}_s(\hat{\mathbf{x}}) \right\},$$

we consider

$$\min_{\boldsymbol{\eta}} \left\{ \mathbf{a}^\top \boldsymbol{\eta} : \min_{\mathbf{x}} \left\{ \mathcal{Q}_s^{LP}(\mathbf{x}) + (\hat{\boldsymbol{\pi}} + \boldsymbol{\eta})^\top (\hat{\mathbf{x}} - \mathbf{x}) : \mathbf{x} \in X^{LP} \right\} \geq \mathcal{Q}_s(\hat{\mathbf{x}}) \right\}, \quad (24)$$

where \mathcal{Q}_s^{LP} is defined by solving the LP relaxation of the local recourse problem (1b).

PROPOSITION 10. *Any feasible solution of (24) is also feasible for (Stren).*

Proof. See Appendix A.11. \square

This proposition shows that by relaxing the inner minimization problem, we can tighten the feasible region of the strengthening problem. When $\mathcal{Q}_s^{LP}(\hat{\mathbf{x}}) < \mathcal{Q}_s(\hat{\mathbf{x}})$, we have $\min_{\mathbf{x} \in X^{LP}} \mathcal{Q}_s^{LP}(\mathbf{x}) + (\hat{\boldsymbol{\pi}} + \boldsymbol{\eta})^\top (\hat{\mathbf{x}} - \mathbf{x}) \leq \mathcal{Q}_s^{LP}(\hat{\mathbf{x}}) < \mathcal{Q}_s(\hat{\mathbf{x}})$ for any $\boldsymbol{\eta} \in \mathbb{R}^n$. Hence, it is possible that the resulting formulation (24) is infeasible. In this case, we will improve the formulation (24).

EXAMPLE 8. Consider the local recourse function $\mathcal{Q}_1(x) = \min\{y_1 + y_2 : 2y_1 + y_2 \geq 3x + 2, 0 \leq y_1 \leq 2, 0 \leq y_2 \leq 3, \mathbf{y} \in \mathbb{Z}^2\}$. It is easy to check that $\mathcal{Q}_1(1) = \mathcal{Q}_1^{LP}(1) = 3$. Therefore, when $\hat{x} = 1$, the relaxed strengthening problem (24) is feasible. However, consider another recourse function: $\mathcal{Q}_2(x) = \min\{y_1 + y_2 : 2y_1 + y_2 \geq 3x, 0 \leq y_1 \leq 2, 0 \leq y_2 \leq 3, \mathbf{y} \in \mathbb{Z}^2\}$. In this case $\mathcal{Q}_2(1) = 2$ while $\mathcal{Q}_2^{LP}(1) = \frac{3}{2}$. Thus, the relaxed strengthening problem (24) becomes infeasible. \diamond

To address the feasibility issue, we incorporate the following no-good cut into the inner minimization problem of the formulation (24):

$$\boldsymbol{\chi}^\top (\hat{\mathbf{x}} - \mathbf{x}) \geq 1,$$

where we recall that $\chi_i = 2\hat{x}_i - 1$ for each $i \in [n]$. Next, we show that the strengthening problem (24) is always feasible after incorporating the no-good cut.

PROPOSITION 11. *The set*

$$F_s^{LP} := \left\{ \boldsymbol{\eta} : \min_{\mathbf{x}} \left\{ \mathcal{Q}_s^{LP}(\mathbf{x}) + (\hat{\boldsymbol{\pi}} + \boldsymbol{\eta})^\top (\hat{\mathbf{x}} - \mathbf{x}) : \mathbf{x} \in \tilde{X} \right\} \geq \mathcal{Q}_s(\hat{\mathbf{x}}) \right\} \quad (25)$$

is always nonempty, where $\tilde{X} := \{\mathbf{x} \in X^{LP} : \boldsymbol{\chi}^\top (\hat{\mathbf{x}} - \mathbf{x}) \geq 1\}$.

Proof. See Appendix A.12. \square

The relaxed strengthening problem is then given by

$$\min_{\boldsymbol{\eta}} \left\{ \mathbf{a}^\top \boldsymbol{\eta} : \mathcal{Q}_s(\hat{\mathbf{x}}) \leq \min_{\mathbf{x}, \mathbf{y}} \left\{ \mathbf{q}^\top \mathbf{y} + (\hat{\boldsymbol{\pi}} + \boldsymbol{\eta})^\top (\hat{\mathbf{x}} - \mathbf{x}) : \mathbf{A}\mathbf{x} \geq \mathbf{b}, \mathbf{T}^s \mathbf{x} + \mathbf{W}^s \mathbf{y} \geq \mathbf{h}^s, \boldsymbol{\chi}^\top (\hat{\mathbf{x}} - \mathbf{x}) \geq 1, 0 \leq \mathbf{x} \leq \mathbf{1} \right\} \right\}.$$

Taking the dual of the minimization problem on the right-hand side of the constraint, we have

$$\min_{\boldsymbol{\eta}} \left\{ \mathbf{a}^\top \boldsymbol{\eta} : \mathcal{Q}_s(\hat{\mathbf{x}}) \leq (\hat{\boldsymbol{\pi}} + \boldsymbol{\eta})^\top \hat{\mathbf{x}} + \max_{\boldsymbol{\tau}, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\omega}} \left\{ \mathbf{b}^\top \boldsymbol{\tau} + (\mathbf{h}^s)^\top \boldsymbol{\sigma} + (1 - \sum_{i \in [n]} \hat{x}_i) \boldsymbol{\rho} + \mathbf{1}^\top \boldsymbol{\omega} : \right. \right. \\ \left. \left. \mathbf{A}^\top \boldsymbol{\tau} + (\mathbf{T}^s)^\top \boldsymbol{\sigma} - \boldsymbol{\chi} \boldsymbol{\rho} + \boldsymbol{\omega} \leq -(\hat{\boldsymbol{\pi}} + \boldsymbol{\eta}), (\mathbf{W}^s)^\top \boldsymbol{\sigma} \leq \mathbf{q}, \boldsymbol{\tau}, \boldsymbol{\sigma}, \boldsymbol{\rho} \geq 0, \boldsymbol{\omega} \leq 0 \right\} \right\}.$$

Replacing the maximization operator with existence one, we obtain an equivalent strengthening linear program:

$$\min_{\boldsymbol{\eta}, \boldsymbol{\tau}, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\omega}} \left\{ \mathbf{a}^\top \boldsymbol{\eta} : \hat{\mathbf{x}}^\top \boldsymbol{\eta} + \mathbf{b}^\top \boldsymbol{\tau} + (\mathbf{h}^s)^\top \boldsymbol{\sigma} + (1 - \sum_{i \in [n]} \hat{x}_i) \boldsymbol{\rho} + \mathbf{1}^\top \boldsymbol{\omega} \geq \mathcal{Q}_s(\hat{\mathbf{x}}) + \hat{\mathbf{x}}^\top \hat{\boldsymbol{\pi}}, \right. \\ \left. \mathbf{A}^\top \boldsymbol{\tau} + (\mathbf{T}^s)^\top \boldsymbol{\sigma} + \boldsymbol{\chi} \boldsymbol{\rho} + \boldsymbol{\omega} + \boldsymbol{\eta} \leq -\hat{\boldsymbol{\pi}}, (\mathbf{W}^s)^\top \boldsymbol{\sigma} \leq \mathbf{q}, \boldsymbol{\tau}, \boldsymbol{\sigma}, \boldsymbol{\rho} \geq 0, \boldsymbol{\omega} \leq 0 \right\}. \quad (\text{Stren LP})$$

This LP-based formulation can also be unbounded. Similar to the analysis of (Stren), the feasible region for $\boldsymbol{\eta}$ is unbounded, with its recession cone being the normal cone of $\text{conv}(\tilde{X})$ at point $\hat{\boldsymbol{x}}$. As stated in Proposition 10, the feasible region of the LP-based formulation is a subset of that of the original strengthening problem. Therefore, the two strategies introduced in Section 3.2 can effectively address the unboundedness issue here. Specifically, in Strategy 1, the additional constraints that ensure the boundedness of the feasible region in (Stren) are also sufficient to guarantee that the feasible region of (Stren LP) is bounded. When restricting $\boldsymbol{\eta}$ in the reverse direction of the recession cone, we add the following constraint to (Stren LP):

$$\min_{\boldsymbol{x}} \{ \boldsymbol{x}^\top \boldsymbol{\eta} : \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}, \boldsymbol{\chi}^\top (\hat{\boldsymbol{x}} - \boldsymbol{x}) \geq 1, \mathbf{0} \leq \boldsymbol{x} \leq \mathbf{1} \} \geq \hat{\boldsymbol{x}}^\top \boldsymbol{\eta},$$

which is different from (22) since we replace set X^{LP} by a smaller set \tilde{X} . Similarly, we can take the dual of the minimization problem on the left-hand side to make it compatible with the LP. Then, it is equivalent to adding the following constraints to (Stren LP):

$$\boldsymbol{b}^\top \boldsymbol{\tau}_1 + (1 - \sum_{i \in [n]} \hat{x}_i) \rho_1 + \mathbf{1}^\top \boldsymbol{\omega}_1 - \hat{\boldsymbol{x}}^\top \boldsymbol{\eta} \geq 0, \boldsymbol{A}^\top \boldsymbol{\tau}_1 + \boldsymbol{\chi} \rho_1 + \boldsymbol{\omega}_1 - \boldsymbol{\eta} \leq \mathbf{0}, \boldsymbol{\tau}_1, \rho_1 \geq 0, \boldsymbol{\omega} \leq \mathbf{0}.$$

In Strategy 2, for any objective coefficient \boldsymbol{a} such that $-\boldsymbol{a} \in \mathcal{T}_{\text{conv}(X)}(\hat{\boldsymbol{x}})$, we have $-\boldsymbol{a} \in \mathcal{T}_{\text{conv}(\tilde{X})}(\hat{\boldsymbol{x}})$ as $X \setminus \{\hat{\boldsymbol{x}}\} \subseteq \tilde{X}$. Thus, the LP-based formulation (Stren) is bounded according to Proposition 9.

The next example illustrates the importance of the no-good cuts.

EXAMPLE 9. Recall in Example 5, $\mathcal{Q}(x_1, x_2) = \min_{\boldsymbol{y}} \{2y_1 + 2y_2 : 0.2y_1 + y_2 + x_1 + 0.5x_2 \geq 2.4, \boldsymbol{y} \in \{0, 1, 2\}^2\}$, $\boldsymbol{x} \in \{0, 1\}^2$. At $\hat{\boldsymbol{x}} = (1, 1)$, we can generate an L-shaped cut $\theta \geq 2 + 2(x_1 - 1) + 2(x_2 - 1)$. To strengthen this cut, we solve the following problem using Strategy 2 to avoid unboundedness:

$$\min_{\boldsymbol{\eta}} \left\{ \eta_1 + \eta_2 : \min_{\boldsymbol{x} \in [0,1]^2} \{ \mathcal{Q}^{LP}(\boldsymbol{x}) + (2 + \eta_1)(1 - x_1) + (2 + \eta_2)(1 - x_2) \} \geq 2 \right\}, \quad (26)$$

which is equivalent to

$$\begin{aligned} \min_{\boldsymbol{\eta}} \{ \eta_1 + \eta_2 : 4 + \eta_1 + \eta_2 + 2.4\alpha + 2\beta_1 + 2\beta_2 + \gamma_1 + \gamma_2 \geq 2, 0.2\alpha + \beta_1 \leq 2, \alpha + \beta_2 \leq 2, \\ \alpha + \gamma_1 \leq -(2 + \eta_1), 0.5\alpha + \gamma_2 \leq -(2 + \eta_2), \alpha \geq 0, \beta \leq 0, \gamma \leq 0 \}. \end{aligned}$$

This problem is infeasible. However, if we add a constraint $(1 - x_1) + (1 - x_2) \geq 1$ to the minimization problem in (26) the problem becomes feasible with optimal solution $(-3.8, -2.8)$. In this way, we obtain a stronger Lagrangian cut $\theta \geq 2 - 1.8(x_1 - 1) - 0.8(x_2 - 1)$. \diamond

To further enhance the relaxed strengthening problem (24), we can incorporate additional valid inequalities of $\text{conv}\{(\boldsymbol{x}, \theta) \in \text{epi}_X(\mathcal{Q}_s) : \boldsymbol{x} \neq \hat{\boldsymbol{x}}\}$ into the inner minimization problem, similar to the no-good cuts. For example, we can add the following objective cuts.

DEFINITION 10 (OBJECTIVE CUTS). Given scenario $s \in [N]$, let L_s be a lower bound of Q_s . Then, we can introduce the following objective cuts to the inner minimization problem of (Stren LP):

$$(\mathbf{q}^s)^\top \mathbf{y} \geq L_s,$$

where we assume that $L_s \geq L$. In fact, this lower bound L_s can be obtained by solving the scenario problem $L_s := \min_{\mathbf{x} \in X} Q_s(\mathbf{x})$.

3.4. Implementation details

This subsection details the implementations of the strengthened Lagrangian cut for stochastic integer programs with purely binary first-stage decisions within the framework of the basic cutting plane method (see Algorithm 1). This procedure is also suitable for the branch-and-cut algorithm.

Algorithm 1 ReLU Lagrangian cuts for binary first-stage decisions

- 1: **Input:** Master problem: $v^* = \min\{\mathbf{c}^\top \mathbf{x} + \theta : \mathbf{A}\mathbf{x} \geq \mathbf{b}, \mathbf{x} \in \bar{X}\}$ and subproblems (1b)
 - 2: **Output:** Optimal solution x^*
 - 3: **Initialize:** $lb \leftarrow -\infty$, $ub \leftarrow +\infty$, $i \leftarrow 0$
 - 4: Analyze the master problem and its LP relaxation to select an appropriate strategy for avoiding unboundedness
 - 5: **while** stopping criterion not met **do**
 - 6: Solve the master problem, obtain an optimal solution $\hat{\mathbf{x}}$, and set $lb \leftarrow v^*$
 - 7: **for** $s \in [N]$ **do**
 - 8: Solve the subproblem to get $Q_s(\hat{\mathbf{x}})$, and generate the L-shaped coefficients $\boldsymbol{\pi}_s^i$
 - 9: Update the strengthening problem (Stren LP) and solve it
 - 10: **if** (Stren LP) is optimal **then**
 - 11: Let $\hat{\boldsymbol{\eta}}_s^i$ be an optimal solution, and $\boldsymbol{\pi}_s^i \leftarrow \boldsymbol{\pi}_s^i + \hat{\boldsymbol{\eta}}_s^i$
 - 12: Add the Lagrangian cut $\theta \geq p_s(Q_s(\hat{\mathbf{x}}) + (\boldsymbol{\pi}_s^i)^\top \mathbf{x})$ to the master problem
 - 13: $ub \leftarrow \min\{ub, \mathbf{c}^\top \hat{\mathbf{x}} + \sum_{s \in [N]} p_s Q_s(\hat{\mathbf{x}})\}$ and $i \leftarrow i + 1$
-

Given that the feasible region is unbounded for any binary $\hat{\mathbf{x}}$, it is important to ensure that the strengthening problem (Stren LP) is bounded prior to the cut-strengthening procedure. We seek an approach, potentially combining both strategies introduced in Section 3.2, that preserves as many extreme points of the original feasible region as possible, as they correspond to cuts that are closer to being facet-defining. Since both strategies rely on an incumbent solution to determine the additional constraints and the objective function, we also aim to minimize the necessary updates

to the strengthening problem in each iteration. First, we try to select appropriate objective coefficients following Strategy 2 without adjusting the feasible region. However, in practice, tracking the tangent cone at the incumbent solution can be challenging. It is desirable to add constraints and choose the objective coefficients to guarantee that the strengthening problem is bounded while its optimal solution effectively improves the cut. One approach is to set the objective coefficients $\mathbf{a} = \boldsymbol{\chi}$ and enforce $\text{diag}(\boldsymbol{\chi})^{-1}\boldsymbol{\eta} \geq \mathbf{0}$. In this way, we have $\boldsymbol{\eta}^\top(\mathbf{x} - \hat{\mathbf{x}}) = (\text{diag}(\boldsymbol{\chi})^{-1}\boldsymbol{\eta})^\top \text{diag}(\boldsymbol{\chi})(\mathbf{x} - \hat{\mathbf{x}}) \geq 0$. If the strengthening problem is optimal, the resulting cut (20), though not necessarily facet-defining, will dominate the original L-shaped cut. If the problem remains unbounded, we add constraints $\text{diag}(\boldsymbol{\chi})^{-1}\boldsymbol{\eta} \leq (\mathcal{Q}_s(\hat{\mathbf{x}}) - L)\mathbf{1}$ to it and solve the problem again, as described in Strategy 1.

At each iteration, we need to update the optimality condition in the strengthening problem based on the current first-stage decision $\hat{\mathbf{x}}$ and its local recourse function value $\mathcal{Q}_s(\hat{\mathbf{x}})$. In addition, depending on which strategy is used to avoid unboundedness, we may need to adjust the objective coefficients according to $\hat{\mathbf{x}}$ or random coefficients and set limits for the entries of $\boldsymbol{\eta}$ according to the L-shaped cuts' coefficients. Although we address the infeasibility of the strengthening problem caused by the integrality gap by adding a no-good cut to the inner minimization problem, the additional constraints introduced in Strategy 1 to prevent unboundedness may still render the problem infeasible. In this case, the L-shaped cuts serve as alternative ReLU Lagrangian cuts.

We can also combine them with other cuts that help accelerate the algorithm. For example, in Zou et al. (2019), the Lagrangian cuts can be implemented together with the Benders cuts and the strengthened Benders cuts. These cuts may not be tight but can improve the outer approximation of the expected epigraph.

4. General Mixed Integer First-Stage Decisions

In this section, we study ReLU Lagrangian cuts within the mixed-integer first-stage decision setting. Following the strategy outlined in Section 3, we begin with an easily obtainable valid cut. For stochastic programs with purely integer first-stage decisions, we convert them into models with binary first-stage variables and then apply the methods from Section 3. For programs with mixed-integer first-stage decisions, we derive cuts analogous to the reverse norm cuts. In both cases, as in Section 3, the cuts can be strengthened based on dual optimality conditions.

4.1. General integer first-stage decisions: binarization vs. non-binarization

According to Zou et al. (2019), Lagrangian cuts can be applied to general integer first-stage decisions by binarizing integer variables. In this subsection, we show that while the ReLU Lagrangian cuts can effectively solve the problem without reformulating the original problem, binarization

provides additional benefits. Specifically, it not only reduces the number of variables required to linearize the cuts but also achieves stronger outer approximations of the local epigraphs. Throughout this subsection, we assume that $\bar{X} \subseteq \mathbb{Z}^n$.

For any given integer $\hat{\boldsymbol{x}} \in \bar{X}$, we can construct a Λ -shaped cut (15) as an initial ReLU Lagrangian cut. Alternatively, we can first binarize the entries of \boldsymbol{x} . Let $N_i = \lfloor \log_2 B_i \rfloor$ and represent x_i as $x_i = \sum_{j \in [0, N_i]} 2^j \delta_j^i$ with $\delta^i \in \{0, 1\}^{N_i+1}$ for each $i \in [n]$. In the binarized space, we can then derive an L-shaped cut as an initial cut:

$$\theta \geq \mathcal{Q}_s(\hat{\boldsymbol{x}}) - (\mathcal{Q}_s(\hat{\boldsymbol{x}}) - L) \sum_{i \in [n]} \|\boldsymbol{\delta}^i - \hat{\boldsymbol{\delta}}^i\|_1. \quad (27)$$

We compare cuts (15) and (27) from two perspectives. First, considering the additional variables introduced: in (15), for each $i \in [n]$ and $\hat{x}_i \in [1, B_i - 1]$, as $|x_i - \hat{x}_i| = (x_i - \hat{x}_i)^+ + (x_i - \hat{x}_i)^-$, we need a binary variable to linearize the terms $(x_i - \hat{x}_i)^+$ and $(x_i - \hat{x}_i)^-$ as shown in the constraint system (13). This results in up to $\sum_{i \in [n]} (B_i - 1)$ additional binary variables. In contrast, it takes $\sum_{i \in [n]} (1 + \lfloor \log_2 B_i \rfloor)$ binary variables to represent the first-stage decisions.

More importantly, we show that the L-shaped cuts derived in the binarized space perform better by proving the inclusion of the convex hulls of the subsystems described by two types of initial cuts.

PROPOSITION 12. *Given $\hat{\boldsymbol{x}} \in \bar{X}$, let us define $E_\Lambda^s = \{(\boldsymbol{x}, \theta) \in \mathcal{B} \times \mathbb{R} : (15)\}$ and*

$$E_L^s = \left\{ (\boldsymbol{x}, \theta) \in \mathcal{B} \times \mathbb{R} : \exists \boldsymbol{\delta} \in \{0, 1\}^{N_1+1} \times \dots \times \{0, 1\}^{N_n+1}, x_i = \sum_{j \in [0, N_i]} 2^j \delta_j^i, \forall i \in [n], (27) \right\},$$

where $\mathcal{B} := \times_{i \in [n]} [0, B_i]$. If $B_i \geq 3$ for all $i \in [n]$, then $\text{conv}(E_\Lambda^s) \supseteq \text{conv}(E_L^s)$.

Proof. See Appendix A.13. □

When $B_i \leq 2$, the inclusion of the two sets may not hold, as shown in the following example.

EXAMPLE 10. Consider the local recourse function $\mathcal{Q}_s(x) = \min\{y : y \geq x, y \in \mathbb{Z}\}$ for $x \in \{0, 1, 2\}$. Let $L = 0$. When $\hat{x} = 1$, we can derive a Λ -shaped cut $\theta \geq 1 + (0 - 1)|x - \hat{x}|$. For this cut, we have $\text{conv}(E_\Lambda^s) = \{(x, \theta) : \theta \geq 1 - (\omega^+ + \omega^-), \omega^+ - \omega^- = x - \hat{x}, 0 \leq \omega^+ \leq (2 - 1)z, 0 \leq \omega^- \leq 1 - z, z \in [0, 1]\}$.

It is easy to check that $\theta \geq 1 - \max\{\hat{x}, B - \hat{x}\} = 0$ when $x = 2$. Meanwhile, binarizing x as $x = \delta_0 + 2\delta_1$, we can derive a cut $\theta \geq 1 + (0 - 1)(|\delta_0 - 1| + |\delta_1 - 0|)$. When $x = 2$, we have $\theta \geq -1$. There exists a point $(2, -1) \in \text{conv}(E_L^s)$ that is not in $\text{conv}(E_\Lambda^s)$. ◇

This example indicates that binarization generally provides stronger initial cuts prior to the cut-strengthening procedure. It is more effective as fewer additional binary variables are required. In addition, the strengthening strategy introduced in the previous section can be applied directly after binarization.

We close this subsection by providing the convex hull of the set E_L^s .

PROPOSITION 13. Suppose that integer $B_i = 2^{j_{1i}} + \dots + 2^{j_{\ell i}} + 2^{N_i}$, and set $J_i = \{j_{1i}, \dots, j_{\ell i}, N_i\}$ for each $i \in [n]$. Then,

$$\text{conv}(E_L^s) = \left\{ \begin{array}{l} \exists \delta \in [0, 1]^{N_1+1} \times \dots \times [0, 1]^{N_n+1}, (27), \\ (\mathbf{x}, \theta) \in \times_{i \in [n]} [0, B_i] \times \mathbb{R} : \\ \quad x_i = \sum_{j \in [0, N_i]} 2^j \delta_j^i \leq B_i, \forall i \in [n], \\ \quad \delta_r^i + \sum_{\tau \in J_{ir}} \delta_\tau^i \leq |J_{ir}|, \forall r \in \{0, 1, \dots, N_i\} \setminus J_i, \forall i \in [n] \end{array} \right\},$$

where we let $J_{ir} = \{\ell \in J_i : \ell > r\}$.

Proof. See Appendix A.14. □

4.2. Mixed integer first-stage decisions

For mixed integer first-stage decisions, according to Zou et al. (2019), binarization may still remain an option and the number of binary variables required can be bounded in terms of the desired approximation accuracy. However, this approach may introduce an excessive number of additional binary variables and potentially fail to achieve an exact optimal solution due to binarization. Given these issues, we propose solving the problem directly using ReLU Lagrangian cuts.

Similar to the L-shaped and Λ -shaped cuts used for purely binary and integer variables, we develop an initial cut to begin with. According to Theorem 1, for any feasible first-stage decisions $\hat{\mathbf{x}}$, there exists a $\rho^* > 0$ such that $(-\rho^* \mathbf{1}, -\rho^* \mathbf{1})$ is optimal for the dual problem. This allows us to derive a cut of the form $\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) - \rho^* \|\mathbf{x} - \hat{\mathbf{x}}\|_1$. If the recourse function is Lipschitz continuous, we use the Lipschitz constant as ρ^* and derive a reverse norm cut. Otherwise, ρ^* may need to be adjusted based on the given $\hat{\mathbf{x}}$, which can be determined using binary search.

When the second-stage feasible region has a special structure, such as a knapsack-constrained set, we can determine a valid ρ^* by following the procedure described in the proof of Theorem 1. First, we relax the first-stage feasible region to a larger set S with fewer constraints, facilitating the identification of extreme points $\{(\mathbf{x}^k, \mathbf{y}^k, \omega_k)\}_{k \in K}$ of $\text{conv}(F)$, where $F = \{(\mathbf{x}, \mathbf{y}, \omega) \in \mathbb{R}^n \times \mathbb{Z}^{m_1} \times \mathbb{R}^{m_2} \times \mathbb{R} : \mathbf{T}^s \mathbf{x} + \mathbf{W}^s \mathbf{y} \geq \mathbf{h}^s, \mathbf{x} \in S\}$. We then compute the distance $d \leq \min\{\|\mathbf{x}^k - \hat{\mathbf{x}}\|_1 : k \in K, \mathbf{x}^k \neq \hat{\mathbf{x}}\}$ and let

$$\rho^* = \frac{\mathcal{Q}_s(\hat{\mathbf{x}}) - L}{d}. \quad (28)$$

Letting $\boldsymbol{\pi}^+ = \boldsymbol{\pi}^- = -\rho^* \mathbf{1}$, we find an optimal solution to $\sup_{\boldsymbol{\pi}^+, \boldsymbol{\pi}^- \in \mathbb{R}^n} \inf_{\mathbf{x} \in S} L_s(\mathbf{x}, \boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}})$. According to Corollary 4, it is also optimal for the original dual problem (7). In this way, we can derive an initial ReLU Lagrangian cut:

$$\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) - \rho^* \left(\sum_{i \in [n]} (x_i - \hat{x}_i)^+ + \sum_{i \in [n]} (x_i - \hat{x}_i)^- \right) = \mathcal{Q}_s(\hat{\mathbf{x}}) - \rho^* \|\mathbf{x} - \hat{\mathbf{x}}\|_1. \quad (29)$$

The following example illustrates how to find such ρ^* using this approach.

EXAMPLE 11. Let $Q(\mathbf{x}) = \min\{y : y \geq x_1 + x_2, y \in \mathbb{Z}\}$, where the domain $\bar{X} = \{\mathbf{x} \in \mathbb{Z} \times \mathbb{R} : 0 \leq x_1 \leq 2, 0 \leq x_2 \leq 2\}$, $\hat{\mathbf{x}} = (1, 1)^\top$ and $L = 0$. We relax \bar{X} to $S = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq 2\}$. Then, we have $F = \{(\mathbf{x}, y, \omega) \in \mathbb{R}^2 \times \mathbb{Z} \times \mathbb{R} : \|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq 2, \|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \omega, y \geq x_1 + x_2\} = \{(\mathbf{x}, y, \omega) \in \mathbb{R}^2 \times \mathbb{Z} \times \mathbb{R} : x_1 + x_2 - 2 \leq \omega, x_1 - x_2 \leq \omega, x_2 - x_1 \leq \omega, -x_1 - x_2 + 2 \leq \omega, 0 \leq \omega \leq 2, y \geq x_1 + x_2\}$. This is an integral polyhedron, i.e., $\text{conv}(F) = \{(\mathbf{x}, y, \omega) : x_1 + x_2 - 2 \leq \omega, x_1 - x_2 \leq \omega, x_2 - x_1 \leq \omega, -x_1 - x_2 + 2 \leq \omega, 0 \leq \omega \leq 2, y \geq x_1 + x_2\}$ with extreme points $(-1, 1, 0, 2)^\top, (1, -1, 0, 2)^\top, (1, 3, 4, 2)^\top, (3, 1, 4, 2)^\top$ and $(1, 1, 2, 0)^\top$. Then $d = \min\{\|\mathbf{x}^k - \hat{\mathbf{x}}\|_1 : k \in K, \mathbf{x}^k \neq \hat{\mathbf{x}}\} = 2$. We let $\rho = \frac{2(2-0)}{2} = 2$. The initial ReLU Lagrangian cut we derive is $\theta \geq 2 - 2\|\mathbf{x} - \hat{\mathbf{x}}\|_1$. \diamond

To strengthen this initial ReLU cut, we lift the first-stage feasible region \bar{X} to a higher-dimensional space by considering the set $\Omega_{\hat{\mathbf{x}}} := \{(\boldsymbol{\omega}^+, \boldsymbol{\omega}^-) : (13b), (13c), \mathbf{x} \in \bar{X}\}$. Then, we redefine the recourse function $Q_s(\boldsymbol{\omega}^+, \boldsymbol{\omega}^-) := Q_s(\hat{\mathbf{x}} + \boldsymbol{\omega}^+ - \boldsymbol{\omega}^-)$. With this transformation, the ReLU Lagrangian cuts generated at $\hat{\mathbf{x}}$ become Lagrangian cuts generated at $(\mathbf{0}, \mathbf{0})$ for the lifted epigraphical set $\text{epi}_{\Omega_{\hat{\mathbf{x}}}}(Q_s)$. Therefore, the properties of the Lagrangian cuts discussed in Section 2 can be directly applied to the transformed problem.

Based on Theorem 2, we can define a ‘‘facet-defining’’ cut generated at $\hat{\mathbf{x}}$ whose coefficients correspond to an extreme point of the set $\Pi_{\hat{\mathbf{x}}}^s$ (recall that set $\Pi_{\hat{\mathbf{x}}}^s$ is defined in Definition 4). It also implies that in the lifted space, the number of affinely independent points at which a ReLU Lagrangian cut is tight is bounded by the dimension of the lifted feasible region $\Omega_{\hat{\mathbf{x}}}$. This further explains why certain ReLU Lagrangian cuts can be stronger than augmented Lagrangian cuts, where the latter, due to the symmetry of the ℓ_1 -norm, may fail to be ‘‘facet-defining’’ in the lifted space.

EXAMPLE 12. In Example 3, let $\hat{\mathbf{x}} = (1, 2)^\top$. We have $(1, 2)^\top = \hat{\mathbf{x}} + (0, 0)^\top - (0, 0)^\top$, $(1, 0)^\top = \hat{\mathbf{x}} + (0, 0)^\top - (0, 2)^\top$, $(0, 1)^\top = \hat{\mathbf{x}} + (0, 0)^\top - (1, 1)^\top$, $(0, 2)^\top = \hat{\mathbf{x}} + (0, 0)^\top - (1, 0)^\top$, $(0, 3)^\top = \hat{\mathbf{x}} + (0, 1)^\top - (1, 0)^\top$, $(1, 4)^\top = \hat{\mathbf{x}} + (0, 2)^\top - (0, 0)^\top$, $(2, 3)^\top = \hat{\mathbf{x}} + (1, 1)^\top - (0, 0)^\top$, $(2, 2)^\top = \hat{\mathbf{x}} + (1, 0)^\top - (0, 0)^\top$, $(2, 1)^\top = \hat{\mathbf{x}} + (1, 0)^\top - (0, 1)^\top$. The first-stage feasible region can be transformed to the set $\Omega_{\hat{\mathbf{x}}} = \{(0, 0, 0, 0)^\top, (0, 0, 0, 2)^\top, (0, 0, 1, 1)^\top, (0, 0, 1, 0)^\top, (0, 1, 1, 0)^\top, (0, 2, 0, 0)^\top, (1, 1, 0, 0)^\top, (1, 0, 0, 0)^\top, (1, 0, 0, 1)^\top\}$ in the lifted space. The cut (18) is facet-defining as it passes through five points $(0, 0, 0, 0, 10)^\top, (0, 0, 0, 2, 5)^\top, (0, 0, 1, 0, 2)^\top, (0, 2, 0, 0, 1)^\top, (1, 0, 0, 0, 4)^\top$ that are affinely independent in the convex hull of the epigraph that is 5-dimensional. Meanwhile, $\Pi_{\hat{\mathbf{x}}}^s = \{(\pi_1^+, \pi_2^+, \pi_1^-, \pi_2^-) : \pi_1^- + \pi_2^- \leq -7, \pi_1^- \leq -8, \pi_2^+ + \pi_1^- \leq -9, 2\pi_2^- \leq -5, \pi_2^- \leq -\frac{5}{2}, \pi_2^+ \leq -\frac{9}{2}, 2\pi_2^+ \leq -9, -\pi_2^- \leq -5, \pi_1^+ \leq -6, \pi_1^+ + \pi_2^+ \leq -7\} = \{(\pi_1^+, \pi_2^+, \pi_1^-, \pi_2^-) : \pi_1^+ \leq -6, \pi_2^+ \leq -\frac{9}{2}, \pi_1^- \leq -8, \pi_2^- \leq \frac{5}{2}\}$. The vector of the coefficients of (18) is the only extreme point of $\Pi_{\hat{\mathbf{x}}}^s$.

For augmented Lagrangian cuts, the cut coefficients are selected from a more restricted set $\Pi_{\hat{\mathbf{x}}}^s \cap \{(\pi_1^+, \pi_2^+, \pi_1^-, \pi_2^-) : \pi_1^+ + \pi_1^- = \pi_2^+ + \pi_2^- \leq 0\}$. It can be shown that the extreme points of set

$\Pi_{\hat{\mathbf{x}}}^s$ are eliminated by the extra constraints, and the restricted set contains two new extreme points that correspond to the coefficients of cuts (19a) and (19b). These new extreme points satisfy three linearly independent constraints in $\Pi_{\hat{\mathbf{x}}}^s$ with equality, and the corresponding cuts are tight at four affinely independent points in the convex hull of the epigraph. \diamond

This example suggests that the cut coefficients $(\boldsymbol{\pi}^+, \boldsymbol{\pi}^-)$ of one of the strongest augmented Lagrangian cuts have $n - 1$ linearly independent equality constraints from the set $\{(\boldsymbol{\pi}^+, \boldsymbol{\pi}^-) : \pi_i^+ + \pi_i^- = \pi_j^+ + \pi_j^-, \forall i, j \in [n]\}$ and require $n + 1$ linearly independent inequalities from $\Pi_{\hat{\mathbf{x}}}^s$ that can be active for the cut coefficients $(\boldsymbol{\pi}^+, \boldsymbol{\pi}^-)$. Therefore, the number of linearly independent points in the epigraph at which this augmented Lagrangian cut is tight is at most $n + 1$, which is clearly not facet-defining. Hence, more augmented Lagrangian cuts may be needed to fully recover the epigraph, compared to ReLU Lagrangian cuts.

Similar to the approach applied to purely binary first-stage decisions, we can enhance the initial cuts (29) using the following strengthening problem in the lifted space:

$$\max_{\boldsymbol{\eta}^+, \boldsymbol{\eta}^-} \left\{ (\mathbf{a}^+)^{\top} \boldsymbol{\eta}^+ + (\mathbf{a}^-)^{\top} \boldsymbol{\eta}^- : \mathcal{Q}_s(\hat{\mathbf{x}}) \leq \min_{\mathbf{x} \in \bar{X}} \mathcal{Q}_s(\mathbf{x}) + \sum_{i \in [n]} (\rho^* - \eta_i^+) (x_i - \hat{x}_i)^+ + \sum_{i \in [n]} (\rho^* - \eta_i^-) (x_i - \hat{x}_i)^- \right\},$$

which is equivalent to

$$\begin{aligned} \max_{\boldsymbol{\eta}^+, \boldsymbol{\eta}^-} \left\{ (\mathbf{a}^+)^{\top} \boldsymbol{\eta}^+ + (\mathbf{a}^-)^{\top} \boldsymbol{\eta}^- : \mathcal{Q}_s(\hat{\mathbf{x}}) \leq \min_{\mathbf{x}, \mathbf{y}, \boldsymbol{\omega}^+, \boldsymbol{\omega}^-, \mathbf{z}} \left\{ (\mathbf{q}^s)^{\top} \mathbf{y} + \sum_{i \in [n]} (\rho^* - \eta_i^+) \omega_i^+ + \sum_{i \in [n]} (\rho^* - \eta_i^-) \omega_i^- : \right. \right. \\ \left. \left. \mathbf{A}\mathbf{x} \geq \mathbf{b}, \mathbf{T}^s \mathbf{x} + \mathbf{W}^s \mathbf{y} \geq \mathbf{h}^s, \omega_i^+ - \omega_i^- = x_i - \hat{x}_i, \forall i \in [n], 0 \leq \omega_i^+ \leq (B_i - \hat{x}_i) z_i, \forall i \in [n], \right. \right. \\ \left. \left. 0 \leq \omega_i^- \leq \hat{x}_i (1 - z_i), \forall i \in [n], \mathbf{x} \in \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2}, \mathbf{y} \in \mathbb{Z}^{m_1} \times \mathbb{R}^{m_2}, \mathbf{z} \in \{0, 1\}^n \right\} \right\}. \end{aligned}$$

Relaxing the integrality constraints and taking the dual of the inner minimization problem yields

$$\begin{aligned} \max_{\boldsymbol{\eta}^+, \boldsymbol{\eta}^-} \left\{ (\mathbf{a}^+)^{\top} \boldsymbol{\eta}^+ + (\mathbf{a}^-)^{\top} \boldsymbol{\eta}^- : \mathbf{b}^{\top} \boldsymbol{\tau} + (\mathbf{h}^s)^{\top} \boldsymbol{\sigma} - \hat{\mathbf{x}}^{\top} \boldsymbol{\gamma} + \hat{\mathbf{x}}^{\top} \boldsymbol{\psi} + \mathbf{1}^{\top} \boldsymbol{\kappa} \geq \mathcal{Q}_s(\hat{\mathbf{x}}), \right. \\ \left. \mathbf{A}^{\top} \boldsymbol{\tau} + (\mathbf{T}^s)^{\top} \boldsymbol{\sigma} - \boldsymbol{\gamma} = \mathbf{0}, (\mathbf{W}^s)^{\top} \boldsymbol{\sigma} = \mathbf{q}^s, \boldsymbol{\gamma} + \boldsymbol{\phi} + \boldsymbol{\eta}^+ \leq \rho^* \mathbf{1}, \boldsymbol{\gamma} + \boldsymbol{\psi} + \boldsymbol{\eta}^- \leq \rho^* \mathbf{1}, \right. \\ \left. (\hat{x}_i - B_i) \phi_i + \hat{x}_i \psi_i + \kappa_i \leq 0, \forall i \in [n], \boldsymbol{\tau}, \boldsymbol{\sigma} \geq \mathbf{0}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\kappa} \leq \mathbf{0} \right\}. \end{aligned} \quad (30)$$

If this problem has an optimal solution with $\boldsymbol{\eta}^+ = \hat{\boldsymbol{\eta}}^+$ and $\boldsymbol{\eta}^- = \hat{\boldsymbol{\eta}}^-$, we can derive a stronger cut

$$\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) + \sum_{i \in [n]} (\hat{\eta}_i^+ - \rho^*) (x_i - \hat{x}_i)^+ + \sum_{i \in [n]} (\hat{\eta}_i^- - \rho^*) (x_i - \hat{x}_i)^-.$$

The strategies described in Section 3.2 can be used to address the unboundedness. Other techniques, such as objective cuts, can also be used to improve the LP-based strengthening problem. Below is an illustrative example.

EXAMPLE 13. We use the strengthening problem to improve the ReLU Lagrangian cuts derived in Example 11. We solve the following relaxed strengthening problem:

$$\min_{\eta^+, \eta^-} \left\{ \sum_{i \in [2]} \eta_i^+ + \sum_{i \in [2]} \eta_i^- : 2 \leq \min_{x, y} \left\{ y + \sum_{i \in [2]} (2 - \eta_i^+) (x_i - \hat{x}_i)^+ + \sum_{i \in [2]} (2 - \eta_i^-) (x_i - \hat{x}_i)^- : y \geq x_1 + x_2, \right. \right. \\ \left. \left. 0 \leq x_1, \leq 2, 0 \leq x_2 \leq 2 \right\} \right\},$$

which is equivalent to

$$\min_{\eta^+, \eta^-} \left\{ \sum_{i \in [2]} \eta_i^+ + \sum_{i \in [2]} \eta_i^- : 2\tau_1 + 2\tau_2 - \gamma_1 - \gamma_2 + \psi_1 + \psi_2 + \kappa_1 + \kappa_2 \geq 2, \tau_i - \sigma - \gamma_i \leq 0, \forall i \in [2], \right. \\ \left. \sigma = 1, \gamma_i + \phi_i + \eta_i^- \leq 2, \forall i \in [2], -\gamma_i + \psi_i + \eta_i^- \leq 2, \forall i \in [2], -\phi_i + \psi_i + \kappa_i \leq 0, \forall i \in [2], \right. \\ \left. \sigma \geq 0, \tau, \phi, \psi, \kappa \leq \mathbf{0} \right\}.$$

The optimal solution is $\hat{\eta}^+ = (3, 3)^\top$ and $\hat{\eta}^- = (1, 1)^\top$, which corresponds to the cut $\theta \geq 2 + \sum_{i \in [n]} (x_i - \hat{x}_i)^+ - \sum_{i \in [n]} (x_i - \hat{x}_i)^-$. This, in fact, is a facet-defining cut in the lifted space. \diamond

5. Numerical Experiments

In this section, we compare the performance of ReLU Lagrangian cuts and other existing cut families in the literature through numerical studies on two-stage and multistage models. All numerical experiments are conducted in Python using Gurobi version 11.0.1 on virtualized Intel Xeon Cascade Lake CPUs running at 2.9 GHz, with 61 GB RAM, under a Linux operating system.

5.1. Two-stage models

We test the performance of four cut combinations for solving two-stage SMIPs with purely binary first-stage decisions and mixed integer second-stage decisions. The four cut combinations are the integer L-shaped cut (L), the Benders cut combined with the integer L-shaped cut (B), the strengthened Benders cut combined with the integer L-shaped cut (SB), and the ReLU Lagrangian cut (R), generated following the procedure described in Section 3.

When implementing the cuts, we begin by adding Benders cuts to solve the LP relaxation to optimality at the root node. Next, we use Gurobi's lazy callback function to solve the problem through a combination of cuts within the branch-and-cut framework.

Experiment 1. In this experiment, we consider the stochastic server location problem (SSLP) (Ntamo and Sen 2005) formulated as:

$$\min_{\mathbf{x}} \left\{ \sum_{j \in J} c_j x_j + \frac{1}{N} \sum_{s \in [N]} \mathcal{Q}_s(\mathbf{x}) : \sum_{j \in J} x_j \leq v, x_j \in \{0, 1\}, \forall j \in J \right\}$$

Algorithm 2 ReLU Lagrangian cuts for two-stage SMIP

1: **Input:** Master problem: $v^* = \min_{\mathbf{x}, \theta} \{ \mathbf{c}^\top \mathbf{x} + \theta : \mathbf{A}\mathbf{x} \geq \mathbf{b}, \mathbf{x} \in \bar{X} \}$ and subproblems (1b)

2: **Output:** Optimal solution x^*

3: **Initialize:** $lb \leftarrow -\infty, ub \leftarrow +\infty, iter \leftarrow 0$

4: **if** $\bar{X} \subseteq \{0, 1\}^n$ **then**

5: Solve it using Algorithm 1

6: **else if** $\bar{X} \subseteq \mathbb{Z}^n$ **then**

7: Binarize the first-stage variables and solve it using Algorithm 1

8: **else**

9: **while** stopping criterion not met **do**

10: Solve the master problem, obtain an optimal solution $\hat{\mathbf{x}}$, and set $lb \leftarrow v^*$

11: **for** $s \in [N]$ **do**

12: Solve the subproblem to get $\mathcal{Q}_s(\hat{\mathbf{x}})$, compute ρ^* using binary search or the closed form (28), and update the strengthening problem (30) and solve it

13: **if** (Stren LP) is optimal **then**

14: Let $(\hat{\boldsymbol{\eta}}_s^+, \hat{\boldsymbol{\eta}}_s^-)$ be an optimal solution. $(\boldsymbol{\pi}_s^+)^{iter} \leftarrow \hat{\boldsymbol{\eta}}_s^+ - \rho^* \mathbf{1}, (\boldsymbol{\pi}_s^-)^{iter} \leftarrow \hat{\boldsymbol{\eta}}_s^- - \rho^* \mathbf{1}$

15: **else**

16: $(\boldsymbol{\pi}_s^+)^{iter} \leftarrow -\rho^* \mathbf{1}, (\boldsymbol{\pi}_s^-)^{iter} \leftarrow -\rho^* \mathbf{1}$

17: $q_{iter} \leftarrow \sum_{s \in [N]} \mathcal{Q}_s(\hat{\mathbf{x}}), (\boldsymbol{\pi}^+)^{iter} \leftarrow \sum_{s \in [N]} (\boldsymbol{\pi}_s^+)^{iter}, (\boldsymbol{\pi}^-)^{iter} \leftarrow \sum_{s \in [N]} (\boldsymbol{\pi}_s^-)^{iter}$

18: Add the following constraints to the master problem

$$\theta \geq q_{iter} + \sum_{i \in [n]} (\boldsymbol{\pi}_i^+)^{iter} \omega_i^+(\hat{x}_i) + \sum_{i \in [n]} (\boldsymbol{\pi}_i^-)^{iter} \omega_i^-(\hat{x}_i), \omega_i^+(\hat{x}_i) - \omega_i^-(\hat{x}_i) = x_i - \hat{x}_i,$$

$$0 \leq \omega_i^+(\hat{x}_i) \leq (B_i - \hat{x}_i) z_i(\hat{x}_i), 0 \leq \omega_i^-(\hat{x}_i) \leq \hat{x}_i (1 - z_i(\hat{x}_i)), z_i(\hat{x}_i) \in \{0, 1\}, \forall i \in [n].$$

19: $ub \leftarrow \min \{ ub, \mathbf{c}^\top \hat{\mathbf{x}} + q_{iter} \}, iter \leftarrow iter + 1$

where for each scenario $s \in [N]$, the local recourse function is defined as

$$\mathcal{Q}_s(\mathbf{x}) := \min_{\mathbf{y}} \sum_{j \in J} q_{0j} y_{0j} - \sum_{i \in I} \sum_{j \in J} q_{ij} y_{ij}$$

$$\text{s.t. } \sum_{i \in I} d_{ij} y_{ij} - y_{0j} \leq u x_j, \sum_{j \in J} y_{ij} = h_i^s, y_{ij} \in \{0, 1\}, y_{0j} \geq 0, \quad \forall i \in I, j \in J.$$

The model parameters are generated as follows: For the first stage, the upper bound v on the total number of servers is set to $\lceil |J|/3 \rceil$, the cost c_i of locating a server at the location j follows a discrete uniform distribution $U[40, 80]$. For the second stage, the demand d_{ij} of client $i \in I$ from the server located at $j \in J$ and the revenue q_{ij} from client i served by the server at location j , both are drawn from a discrete uniform distribution $U[0, 25]$. The server capacity u is 50. The overflow rate q_{0j} for

server $j \in J$ is 1000. Client availability \tilde{h}_i is stochastic and follows a Bernoulli distribution with a success rate 0.7.

We perform numerical studies on SSLP with 20 and 30 locations and clients, considering 10, 50, 100, and 200 scenarios. The results are displayed in Table 1. The algorithm terminates when the gap, defined as the difference between the best upper and lower bounds divided by absolute value of the best lower bound, is less than 0.01% within a time limit of 1 hour.

Columns 1 and 2 of Table 1 represent the total number of locations and clients, respectively. Column 3 shows the number of scenarios, N . Columns 4 and 5 are the best lower and upper bounds obtained from Gurobi. Column 6 reports the gap. Columns 7 and 8 contain the solution time and the number of nodes explored during the branch-and-cut procedure, respectively.

Our results show that all four methods solve all the instances within the time limit. The ReLU Lagrangian cuts achieve the shortest solution time. Additionally, this method significantly reduces the number of nodes explored. For the other methods, the number of nodes explored is similar. This can be because methods L, B, and SB rely heavily on the integer L-shaped cuts, which are naive ReLU Lagrangian cuts. This implies that properly strengthening ReLU Lagrangian cuts can be beneficial in solving two-stage stochastic integer programs.

Table 1 Numerical results for SSLP instances

$ J $	$ I $	N	cut	lb	ub	Gap (%)	time (s)	node	$ J $	$ I $	N	cut	lb	ub	Gap (%)	time (s)	node
20	20	10	L	-25.4	-25.4	0.0	12.6	26	20	20	100	L	-26.2	-26.2	0.0	221.3	76
			B	-25.4	-25.4	0.0	16.1	52				B	-26.2	-26.2	0.0	243.8	75
			R	-25.4	-25.4	0.0	2.8	7				R	-26.2	-26.2	0.0	143.5	48
			SB	-25.4	-25.4	0.0	16.4	52				SB	-26.2	-26.2	0.0	260.2	75
30	30	10	L	-34.4	-34.4	0.0	40.8	67	30	30	100	L	-36.0	-36.0	0.0	1280.7	116
			B	-34.4	-34.4	0.0	31.3	63				B	-36.0	-36.0	0.0	839.9	93
			R	-34.4	-34.4	0.0	13.3	14				R	-36.0	-36.0	0.0	328.6	27
			SB	-34.4	-34.4	0.0	31.7	63				SB	-36.0	-36.0	0.0	805.3	93
20	20	50	L	-26.4	-26.4	0.0	122.5	80	20	20	200	L	-26.0	-26.0	0.0	308.8	79
			B	-26.4	-26.4	0.0	125.2	125				B	-26.0	-26.0	0.0	499.3	93
			R	-26.4	-26.4	0.0	73.2	65				R	-26.0	-26.0	0.0	239.3	63
			SB	-26.4	-26.4	0.0	158.8	125				SB	-26.0	-26.0	0.0	497.7	93
30	30	50	L	-35.6	-35.6	0.0	325.1	124	30	30	200	L	-36.1	-36.1	0.0	1717.6	156
			B	-35.6	-35.6	0.0	416.5	122				B	-36.1	-36.1	0.0	2086.4	126
			R	-35.6	-35.6	0.0	217.6	37				R	-36.1	-36.1	0.0	591.1	24
			SB	-35.6	-35.6	0.0	433.3	122				SB	-36.1	-36.1	0.0	2252.1	126

Experiment 2. To further evaluate the ReLU Lagrangian cuts, we consider the following stochastic multiple-resource-constrained scheduling problem (SMRCSP) (Keller and Bayraksan 2009):

$$\begin{aligned}
\min_{\mathbf{x}, \mathbf{z}} \quad & \sum_{j \in J} \sum_{t \in [T-p_j+1]} c_{jt} x_{jt} + \sum_{k \in [K]} \sum_{t \in [T_0]} b_k z_{tk} + \frac{1}{N} \sum_{s \in [N]} Q_s(\mathbf{x}) \\
\text{s.t.} \quad & \sum_{t \in [T-p_j+1]} x_{jt} = 1, \quad \forall j \in J,
\end{aligned}$$

$$\begin{aligned} \sum_{j \in J} \sum_{\tau \in S(j,t)} r_{jk} x_{j\tau} - M_{tk} z_{tk} &\leq R_k, \quad \forall t \in [T_0], k \in [K], \\ x_{jt} &\in \{0, 1\}, \quad \forall j \in J, t \in [T - p_j + 1], \quad z_{tk} \in \{0, 1\}, \quad \forall t \in [T_0], k \in [K], \end{aligned}$$

where for each scenario $s \in [N]$, the local recourse function is

$$\begin{aligned} \mathcal{Q}_s(\mathbf{x}) := \min_{\mathbf{y}, \mathbf{u}} & \sum_{j \in J_B} \sum_{t=T_0+1}^{T+T_0-p_j+1} c_{jt} y_{jt} + \sum_{k \in [K]} \sum_{t=T_0+1}^{T+T_0} b_k u_{tk} \\ \text{s.t.} & \sum_{t=T_0+1}^{T+T_0-p_j+1} y_{jt} = a_j^s, \quad \forall j \in J_B, \\ & \sum_{j \in J_B} \sum_{\tau \in S_B(j,t)} r_{jk} y_{j\tau} - M_{tk} u_{tk} \leq R_k - \sum_{j \in J} \sum_{\tau \in S(j,t)} r_{jk} x_{j\tau}, \quad \forall t \in [T_0 + 1, T + T_0], k \in [K], \\ & y_{jt} \in \{0, 1\}, \quad \forall j \in J_B, t \in [T_0 + 1, T + T_0 - p_j + 1], \\ & u_{tk} \in \{0, 1\}, \quad \forall t \in [T_0 + 1, T + T_0], k \in [K], \end{aligned}$$

and $S(j, t) = [\max\{1, t - p_j + 1\}, \min\{t, T - p_j + 1\}]$, $S_B(j, t) = [\max\{T_0 + 1, t - p_j + 1\}, \min\{t, T + T_0 - p_j + 1\}]$. In SMRCS, we impose a fixed cost for expansions when the required resources exceed the amount available in a given time period, i.e., z_{tk} and u_{tk} are binary variables. Compared with SSLP with continuous penalties, the integrality gap of SMRCSP is often large due to big M coefficients.

The parameters are generated as follows. The time period T_0 , at which we learn the accepted job bids, is set to $[0.25T_0]$, the processing time p_j of job j is generated from a discrete uniform distribution $U[1, T]$, and the cost c_{jt} of starting job j in period t is set to completion time $t + p_j - 1$. For each period, the cost b_k of temporary expansion follows a discrete uniform distribution $U[10, 20]$, the amount of resource from class k consumed by job j (i.e., r_{jk}) follows a discrete uniform distribution $U[1, 5]$. The resource capacity $R_k = \frac{\bar{p} \bar{r}_k |J| + 0.75 \bar{p}_B \bar{r}_{Bk} |J_B|}{(T + T_0) \rho}$, where $\bar{p} = \frac{1}{|J|} \sum_{j \in J} p_j$, $\bar{p}_B = \frac{1}{|J_B|} \sum_{j \in J_B} p_j$, $\bar{r}_k = \frac{1}{|J|} \sum_{j \in J} r_{jk}$, $\bar{r}_{Bk} = \frac{1}{|J_B|} \sum_{j \in J_B} r_{jk}$, and ρ follows a continuous uniform distribution $\mathcal{U}[0.5, 1.2]$. The big Ms are set to $(|J| + |J_B|)T$. The indicator a_j of whether the bid on job j is accepted or not is stochastic and follows a Bernoulli distribution with a success rate 0.75. In our numerical study, we set the number of known jobs $|J|$ to 5 and 7, and the number of jobs available for bidding $|J_b|$ to 5 and 10. The number of time intervals T is set to 10, and we consider $N = 10$ and $N = 100$ scenarios for the second stage. The results are shown in Table 2. The stopping criterion and the meanings of the other columns are consistent with those presented in Table 1.

In Table 2, we see that only ReLU Lagrangian cuts can solve all instances within the time limit. In general, they close the optimality gap more quickly, and require fewer nodes to reach an optimal solution compared to integer L-shaped cuts and Benders cuts. While strengthened Benders cuts outperform standard Benders cuts and require fewer nodes to explore than ReLU Lagrangian cuts

Table 2 Numerical results for SMRCSP instances

$ J $	$ J_b $	T	N	cut	lb	ub	Gap (%)	time (s)	node	$ J $	$ J_b $	T	N	cut	lb	ub	Gap (%)	time (s)	node
5	5	10	10	L	243.1	243.1	0.0	5.7	1570	7	5	10	10	L	288.1	288.1	0.0	859.0	181689
				B	243.1	243.1	0.0	7.0	1794					B	288.1	288.1	0.0	1064.5	186444
				R	243.1	243.1	0.0	3.7	428					R	288.1	288.1	0.0	542.1	51487
				SB	243.1	243.1	0.0	8.4	422					SB	288.1	288.1	0.0	1281.0	45846
5	5	10	100	L	256.1	256.1	0.0	47.7	1574	7	5	10	100	L	130.2	285.0	118.9	3600.0	96382
				B	256.1	256.1	0.0	56.3	1717					B	125.3	285.0	127.4	3600.0	78561
				R	256.1	256.1	0.0	30.7	392					R	285.0	285.0	0.0	3099.8	35840
				SB	256.1	256.1	0.0	70.3	390					SB	277.6	285.0	2.7	3600.0	11761
5	10	10	10	L	135.6	135.6	0.0	36.8	1215	7	10	10	10	L	115.4	143.1	24.0	3600.0	30361
				B	135.6	135.6	0.0	38.7	1349					B	117.2	143.1	22.1	3600.0	30260
				R	135.6	135.6	0.0	0.4	1					R	143.1	143.1	0.0	48.1	737
				SB	135.6	135.6	0.0	2.2	23					SB	143.1	143.1	0.0	40.5	276
5	10	10	100	L	139.9	139.9	0.0	321.3	1248	7	10	10	100	L	112.3	157.2	40.0	3600.0	1936
				B	139.9	139.9	0.0	332.1	1337					B	112.7	157.2	39.4	3600.0	2336
				R	139.9	139.9	0.0	3.7	1					R	157.2	157.2	0.0	240.6	293
				SB	139.9	139.9	0.0	16.6	11					SB	157.2	157.2	0.0	644.1	224

in most instances, they still underperform relative to ReLU Lagrangian cuts. This may be due to the need to solve an MIP to obtain a strengthened Benders cut.

Experiment 3. Finally, we consider the following dynamic capacity acquisition and allocation problem (DCAP) adapted from Ahmed and Garcia (2003) with mixed integer first-stage decisions:

$$\min_{\mathbf{x}, \mathbf{u}} \left\{ \sum_{t \in T} \sum_{i \in I} (\alpha_{it} x_{it} + \beta_{it} u_{it}) + \frac{1}{N} \sum_{s \in [N]} \mathcal{Q}_s(\mathbf{x}) : x_{it} \leq b_{it} u_{it}, x_{it} \in [0, b_{it}], u_{it} \in \{0, 1\}, \forall i \in I, t \in T \right\},$$

where the local recourse functions

$$\begin{aligned} \mathcal{Q}_s(\mathbf{x}) &:= \min_{\mathbf{y}} \sum_{t \in T} \sum_{i \in I} \sum_{j \in J} c_{ijt}^s y_{ijt} + \sum_{t \in T} \sum_{i \in I} p_{it} y_{it}^0 \\ \text{s.t.} \quad &\sum_{j \in J} d_{jt}^s y_{ijt} - y_{it}^0 \leq \sum_{\tau \in [t]} x_{i\tau}, \quad \forall i \in I, t \in T, \\ &\sum_{i \in I} y_{ijt} = 1, y_{it}^0 \geq 0, y_{ijt} \in \{0, 1\}, \quad \forall i \in I, j \in J, t \in T. \end{aligned}$$

In this experiment, the parameters are generated as follows. In the first stage, the capacity expansion cost of acquiring resource i in period t consists of two parts, where the variable cost α_{it} and the fixed cost β_{it} are drawn from discrete uniform distributions $U[20, 40]$ and $U[50, 70]$, respectively. The maximum number of units of resource i that can be required at time t , denoted as b_{it} , is set to 50. In the second stage, the cost \tilde{c}_{ijt} of processing task j using resource i in period t , as well as the processing requirement \tilde{d}_{jk} for task j in period t , are stochastic. They follow discrete uniform distributions $U[40, 80]$ and $U[1, 10]$, respectively. To ensure relatively complete recourse, we add a penalty term for overflow, where the penalty rate p_{it} is set to 1000.

We compare the performance of Augmented Lagrangian cuts (AL) given by (16) with $\boldsymbol{\pi} = \mathbf{0}$ and ReLU Lagrangian cuts (R) obtained by improving these augmented Lagrangian cuts through the cut-strengthening procedure described in Section 4. When solving the instances, we first add Benders cuts to solve the LP relaxation to optimality. Then, we maintain the integrality constraints

in the first stage while taking the LP relaxation of the second stage and solve this relaxed problem to optimality using Benders cuts. At each iteration, we add a strengthened Benders cut with a nonlinear cut (AL or R) until the gap falls below $0.1\%^2$ or the time limit of one hour is reached.

The results are shown in Table 3. Columns 1, 2, and 3 represent the total number of resource types, tasks, and periods, respectively. Column 4 shows the number of scenarios. The iteration column lists the number of iterations used to add nonlinear cuts, excluding those for solving the relaxed problem with Benders cuts. As shown by the results, our cut-strengthening procedure leads to significant improvements over augmented Lagrangian cuts. For the instances in the first two rows, it significantly reduces the number of cuts required to solve the problem and shortens the solution time. For the larger instances, although both methods reach the time limit, ReLU Lagrangian cuts still achieve much smaller gaps.

Table 3 Numerical results for DCAP instances

$ I $	$ J $	$ T $	N	cut	lb	ub	Gap (%)	iteration	time (s)	$ I $	$ J $	$ T $	N	cut	lb	ub	Gap (%)	iteration	time (s)
2	2	4	10	AL	1014.2	1015.2	<0.1	998	3478.8	2	2	4	100	AL	1047.0	1048.0	<0.1	629	1000.3
				R	1014.9	1015.4	<0.1	41	6.7					R	1047.4	1048.4	<0.1	22	24.6
2	3	4	10	AL	2134.0	2280.7	6.9	521	3600.0	2	3	4	100	AL	1976.1	2092.5	5.9	438	3600.0
				R	2257.5	2259.3	<0.1	89	35.1					R	2067.3	2069.3	<0.1	381	3289.8
3	4	5	10	AL	2189.1	2366.1	8.1	324	3600.0	3	4	5	100	AL	2307.6	2469.6	7.0	260	3600.0
				R	2218.6	2247.9	1.3	361	3600.0					R	2340.6	2388.6	2.1	291	3600.0
4	5	6	10	AL	2994.7	3192.0	6.6	509	3600.0	4	5	6	100	AL	3005.0	3170.8	5.5	307	3600.0
				R	3009.7	3192.0	6.1	393	3600.0					R	3020.3	3113.6	3.1	255	3600.0

Based on the numerical experiments above, we conclude the advantages of ReLU Lagrangian cuts in solving two-stage SMIPs as follows: (i) For binary first-stage decisions, they can replace the combination of Benders and integer L-shaped cuts. By improving the integer L-shaped cuts through LP-based strengthening problems, ReLU Lagrangian cuts inherit the strength of Benders cuts with the tightness of integer L-shaped cuts into a single cut, making the master problem easier to solve compared to adding both Benders and integer L-shaped cuts simultaneously. (ii) Similar to strengthened Benders cuts, they provide a stronger outer approximation than the epigraph of the continuous relaxation of local recourse functions. The efficiency of the ReLU Lagrangian cuts is possibly due to incorporating valid inequalities of local convex hulls, including no-good and objective cuts, to the LP-based strengthening problems. (iii) For mixed integer first-stage decisions, they improve augmented Lagrangian cuts at a low computational cost and achieve better performance. And (iv) by implementing an appropriate strategy to prevent unboundedness in the strengthening problems, the resulting cuts accelerate the master problem and mitigate potential numerical issues arising from large coefficients in the integer L-shaped cuts or augmented Lagrangian cuts.

² Since this is an iterative solution procedure, we choose a slightly worse stopping criterion for the gap. If we choose the gap to be 0.01% , in most cases, the solution time will exceed an hour, but ReLU Lagrangian cuts still perform better than augmented Lagrangian cuts.

5.2. A multistage model

As ReLU Lagrangian cuts can be integrated into the SDDiP framework, we evaluate their performance on a multistage model from Zou et al. (2019)– the airline revenue management problem. In this numerical experiment, we compare different combinations of nontight and tight cuts. Nontight cuts include Benders cuts (B), strengthened Benders cuts (SB), and improved Benders cuts (IB), obtained by applying our cut-strengthening procedure to Benders cuts. Tight cuts are selected from either integer L-shaped cuts (L) or ReLU Lagrangian cuts (R).

Experiment 4. Let us consider the following airline revenue management problem (ARM):

$$\begin{aligned} \max \quad & \sum_{t \in [T]} [(\mathbf{f}^b)^\top \mathbf{b}_t - (\mathbf{f}^c)^\top \mathbf{c}_t] \\ \text{s.t.} \quad & \mathbf{B}_t = \mathbf{B}_{t-1} + \mathbf{b}_t, \mathbf{C}_t = \mathbf{C}_{t-1} + \mathbf{c}_t, \mathbf{C}_t = \lceil \mathbf{\Gamma} \mathbf{B}_t + 0.5 \rceil, \mathbf{b}_t \leq \mathbf{d}_t, \mathbf{A}(\mathbf{B}_t - \mathbf{C}_t) \leq \mathbf{R}, \quad \forall t \in [T], \\ & \mathbf{B}_0 = \bar{\mathbf{B}}_0, \mathbf{C}_0 = \bar{\mathbf{C}}_0, \mathbf{B}_t, \mathbf{C}_t, \mathbf{b}_t, \mathbf{c}_t \in \mathbb{Z}_+^m, \quad \forall t \in [T]. \end{aligned}$$

The parameters are primarily based on the descriptions in Zou et al. (2019) and Möller et al. (2008) with slightly different origin-destination (OD) pairs and fare classes. Specifically, we suppose that there are 4 OD pairs, each having 2 classes, and each class offers 2 fare prices. Therefore, the total number of classes n is 8, and the total number of ticket types m is 12. The vector \mathbf{f}^b representing the ticket prices is $(500, 340, 200, 100, 500, 340, 200, 100, 800, 540, 320, 160, 800, 540, 320, 160)^\top$, and the vector \mathbf{f}^c representing the refund amount for cancellations is set as $0.8\mathbf{f}^b$. The matrix $\mathbf{\Gamma}$ is a diagonal matrix, where the diagonal entries represent the cancellation rates. The cancellation rate for the first four types of tickets is 0.15, for the next eight types, it is 0.1, and for the last four types, it is 0.05. The matrix \mathbf{A} is an $n \times m$ indicator matrix, where $A_{ij} = 1$ if ticket type i belongs to class j , and 0 otherwise. The seating capacity for the first class is 24, and for the economy class, it is 120. Thus, the vector \mathbf{R} representing seat capacities is set to $(24, 120, \dots, 24, 120)^\top$. Both the initial bookings vector $\bar{\mathbf{B}}_0$ and the initial cancellations vector $\bar{\mathbf{C}}_0$ are zero vectors.

The demand vector \mathbf{d}_t at stage t is stochastic. For each ticket type i , the total number of bookings follows a gamma distribution $G(p_i, g_i)$. These bookings are distributed over 60 days according to a beta distribution $B(\alpha_i, \beta_i)$, and the bookings for the 60 days are then assigned to the stages T . We repeat the sampling procedure 50 times and for each stage $t \in [T]$, the demand vector \mathbf{d}_t is drawn from these samples independently. Note that we suppose that $\mathbf{p} = (3.0, 3.0, 70.0/3, 52.0, 2.0, 2.0, 35.0/3, 26.0, 2.0, 2.0, 17.5, 39.0, 3.0, 3.0, 35.0, 78.0)^\top$, $\mathbf{g} = (1.5, 1.5, 1.2, 1.0, \dots, 1.5, 1.5, 1.2, 1.0)^\top$, $\boldsymbol{\alpha} = (12.0, 8.0, 6.0, 2.0, \dots, 12.0, 8.0, 6.0, 2.0)^\top$, and $\boldsymbol{\beta} = (1.5, 2.0, 2.0, 4.0, \dots, 1.5, 2.0, 2.0, 4.0)$.

When implementing the SDDiP algorithm, during the forward step, we sample five paths and use the results from all five paths to generate cuts in the backward step. The statistical lower

Table 4 Numerical results for the multistage ARM instances

stage	scen- ario	cut	(stat) lb	(best) ub	gap (%)	iter- ation	stage	scen- ario	cut	(stat) lb	(best) ub	gap (%)	iter- ation
4	3	B+L	64925.9	66524.6	2.46	249	4	5	B+L	65276.6	66737.9	2.24	211
		B+R	64887.9	66407.6	2.34	298			B+R	65172.6	66705.1	2.35	253
		SB+L	65318.6	66530.5	1.86	174			SB+L	65278.0	66809.0	2.35	135
		SB+R	65319.8	66494.1	1.80	184			SB+R	65516.9	66798.3	1.96	144
		IB+L	65041.2	66591.0	2.38	197			IB+L	65125.8	66796.8	2.57	170
		IB+R	64942.7	66577.4	2.52	215			IB+R	65193.0	66782.5	2.44	183
6	3	B+L	60744.3	62603.0	3.06	188	6	5	B+L	64148.0	65722.4	2.45	144
		B+R	60745.6	62564.9	3.00	225			B+R	64162.7	65721.6	2.43	157
		SB+L	60914.9	62670.7	2.88	127			SB+L	64216.3	65770.9	2.42	88
		SB+R	61083.7	62648.7	2.56	140			SB+R	64224.7	65763.3	2.40	97
		IB+L	60685.0	62629.7	3.20	161			IB+L	64171.8	65753.3	2.46	120
		IB+R	60731.7	62652.4	3.16	173			IB+R	64208.4	65746.0	2.39	130
8	3	B+L	63582.7	65131.5	2.44	155	8	5	B+L	66238.6	68181.7	2.93	125
		B+R	63593.9	65120.8	2.40	176			B+R	66224.2	68173.5	2.94	140
		SB+L	63949.0	65165.7	1.90	96			SB+L	66165.9	68090.1	2.91	73
		SB+R	63943.6	65156.2	1.90	108			SB+R	66251.7	68086.7	2.77	82
		IB+L	63667.2	65154.9	2.34	132			IB+L	66182.9	68191.5	3.03	107
		IB+R	63859.5	65140.2	2.01	144			IB+R	66199.0	68166.6	2.97	120
10	3	B+L	63615.3	65037.2	2.24	136	10	5	B+L	66954.7	68544.1	2.37	111
		B+R	63607.9	65027.4	2.23	149			B+R	66928.5	68518.2	2.38	124
		SB+L	63652.5	65070.0	2.23	84			SB+L	66860.0	68576.4	2.57	64
		SB+R	63643.1	65065.9	2.24	92			SB+R	66788.5	68555.4	2.65	69
		IB+L	63618.5	65052.0	2.25	117			IB+L	66919.6	68531.1	2.41	94
		IB+R	63617.5	65043.1	2.24	130			IB+R	66729.2	68527.9	2.70	101

bounds are obtained by computing a 95% confidence value using 1,500 sampled paths after the algorithm terminates. As shown by the results in Table 4, any cut combination involving ReLU Lagrangian cuts generally outperforms its counterpart with integer L-shaped cuts. As indicated in the iteration column, ReLU Lagrangian cuts also enable more iterations to be solved within the same time frame. This may be because by using Strategy 1 to avoid unboundedness, many cut coefficients are improved to smaller absolute values or even zero, making the master problem easier to solve compared to the original L-shaped cuts. Among all combinations, SB+R performs the best in most instances. Although the improved Benders cuts benefit from the LP-based cut-strengthening procedure, their performance remains less efficient than that of the strengthened Benders cuts. Moreover, the advantage of ReLU Lagrangian cuts becomes less significant as the number of stages increases.

This numerical example shows that for purely binary state variables, ReLU Lagrangian cuts, which are equivalent to Lagrangian cuts, can be alternatives to integer L-shaped cuts, which is quite different from existing literature (Zou et al. 2019). While previous studies have focused on improving Lagrangian cuts to achieve certain properties by solving Lagrangian duals with integrality constraints, the resulting Lagrangian cuts are still not as effective as strengthened Benders cuts. As a result, it may be more efficient to use the combination of strengthened Benders and integer L-shaped cuts. This is mainly due to two factors: obtaining the Lagrangian cuts requires solving multiple MIPs, whereas strengthened Benders only solves one. Second, Lagrangian cuts

may compromise the strength of the overall outer approximation to guarantee its tightness. As this effect accumulates across stages, it results in slower recovery of the epigraphs of cost-to-go functions compared to strengthened Benders cuts. Nevertheless, in our method, the combination of strengthened Benders cuts and ReLU Lagrangian cuts often outperforms that of strengthened Benders and integer L-shaped cuts, especially in providing better upper bounds.

6. Conclusion

This paper introduced a new family of nonlinear cuts, termed “ReLU Lagrangian cuts,” for solving stochastic integer programs. These cuts improve traditional methods by addressing nonanticipativity constraints through ReLU functions, enabling both local and expected recourse epigraphs to be tightly and efficiently recovered. The tightness of these cuts was established through strong duality. We also proved that ReLU Lagrangian cuts are a generalization of existing cut families, including integer L-shaped cuts, ordinary Lagrangian cuts, reversed norm cuts, and augmented Lagrangian cuts. Therefore, existing cuts can serve as a foundation for initiating the ReLU Lagrangian cuts. We also proposed efficient cut generation schemes that enhance cut coefficients while eliminating the need to solve multiple mixed integer programs at each iteration. Our numerical studies demonstrated the superior performance of ReLU Lagrangian cuts, particularly in reducing the number of iterations required for the cutting-plane method to converge, compared to existing approaches.

References

- Ahmed S, Cabral FG, Freitas Paulo da Costa B (2022) Stochastic Lipschitz dynamic programming. *Mathematical Programming* 191(2):755–793.
- Ahmed S, Garcia R (2003) Dynamic capacity acquisition and assignment under uncertainty. *Annals of Operations Research* 124:267–283.
- Angulo G, Ahmed S, Dey SS (2016) Improving the integer L-shaped method. *INFORMS Journal on Computing* 28(3):483–499.
- Bansal A, Küçükyavuz S (2024) A computational study of cutting-plane methods for multi-stage stochastic integer programs. *arXiv preprint arXiv:2405.02533* .
- Benders J (1962) Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik* 4:238–252.
- Carøe CC, Schultz R (1999) Dual decomposition in stochastic integer programming. *Operations Research Letters* 24(1-2):37–45.
- Chen R, Luedtke J (2022) On generating Lagrangian cuts for two-stage stochastic integer programs. *INFORMS Journal on Computing* 34(4):2332–2349.

- Feizollahi MJ, Ahmed S, Sun A (2017) Exact augmented Lagrangian duality for mixed integer linear programming. *Mathematical Programming* 161:365–387.
- Füllner C, Sun XA, Rebennack S (2024a) A new framework to generate Lagrangian cuts in multistage stochastic mixed-integer programming. Available at Optimization Online.
- Füllner C, Sun XA, Rebennack S (2024b) On Lipschitz regularization and Lagrangian cuts in multistage stochastic mixed-integer linear programming. Available at Optimization Online.
- Gade D, Küçükyavuz S, Sen S (2014) Decomposition algorithms with parametric Gomory cuts for two-stage stochastic integer programs. *Mathematical Programming* 144(1):39–64.
- Geoffrion A (1974) Lagrangean relaxation for integer programming. *Mathematical Programming Study* 82–114.
- Gupte A, Ahmed S, Cheon MS, Dey S (2013) Solving mixed integer bilinear problems using MILP formulations. *SIAM Journal on Optimization* 23(2):721–744.
- Keller B, Bayraksan G (2009) Scheduling jobs sharing multiple resources under uncertainty: A stochastic programming approach. *IIE Transactions* 42(1):16–30.
- Laporte G, Louveaux FV (1993) The integer L-shaped method for stochastic integer programs with complete recourse. *Operations research letters* 13(3):133–142.
- Laurent M, Sassano A (1992) A characterization of knapsacks with the max-flow—min-cut property. *Operations Research Letters* 11(2):105–110.
- Louveaux FV, Schultz R (2003) Stochastic integer programming. *Handbooks in operations research and management science* 10:213–266.
- Meyer RR (1975) Integer and mixed-integer programming models: general properties. *Journal of Optimization Theory and Applications* 16:191–206.
- Möller A, Römisch W, Weber K (2008) Airline network revenue management by multistage stochastic programming. *Computational Management Science* 5(4):355–377.
- Ntaimo L, Sen S (2005) The million-variable “march” for stochastic combinatorial optimization. *Journal of Global Optimization* 32:385–400.
- Rahmaniani R, Ahmed S, Crainic TG, Gendreau M, Rei W (2020) The Benders dual decomposition method. *Operations Research* 68(3):878–895.
- Romeijnders W, van der Laan N (2024) Benders decomposition with scaled cuts for multistage stochastic mixed-integer programs .
- Schrijver A (1998) *Theory of linear and integer programming* (John Wiley & Sons).
- 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* 104:1–20.

- Sen S, Sherali HD (2006) Decomposition with branch-and-cut approaches for two-stage stochastic mixed-integer programming. *Mathematical Programming* 106:203–223.
- van der Laan N, Romeijnders W (2023) A converging Benders’ decomposition algorithm for two-stage mixed-integer recourse models. *Operations Research* .
- 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.
- Zou J, Ahmed S, Sun XA (2019) Stochastic dual dynamic integer programming. *Mathematical Programming* 175:461–502.

Appendix A. Proofs

A.1. Proof of Corollary 2

COROLLARY 2. *A Lagrangian cut generated at $\hat{\mathbf{x}} \in \bar{X}$ is tight if and only if there exists $\boldsymbol{\alpha} \in \mathbb{R}^n$, $\boldsymbol{\alpha} \neq \mathbf{0}$, such that $\boldsymbol{\alpha}^\top \mathbf{x} + \theta \geq \boldsymbol{\alpha}^\top \hat{\mathbf{x}} + \mathcal{Q}_s(\hat{\mathbf{x}})$ for all $(\mathbf{x}, \theta) \in \text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s))$.*

Proof. It is equivalent to show that any cut coefficient $\boldsymbol{\pi} \in \mathcal{D}_{\hat{\mathbf{x}}}^s$ satisfies $\mathcal{L}_s(\boldsymbol{\pi}; \hat{\mathbf{x}}) = \mathcal{Q}_s(\hat{\mathbf{x}})$. Given $\boldsymbol{\pi} \in \mathcal{D}_{\hat{\mathbf{x}}}^s$, from the primal characterization we have $\mathcal{L}_s(\boldsymbol{\pi}; \hat{\mathbf{x}}) = \min_{\mathbf{x}} \{\theta : (\mathbf{x}, \theta) \in \text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s)), \mathbf{x} = \hat{\mathbf{x}}\} \geq \min_{\mathbf{x}} \{\theta : \theta \geq \boldsymbol{\alpha}^\top (\hat{\mathbf{x}} - \mathbf{x}) + \mathcal{Q}_s(\hat{\mathbf{x}}), \mathbf{x} = \hat{\mathbf{x}}\} = \mathcal{Q}_s(\hat{\mathbf{x}})$. Combined with (3a) completes the proof. \square

A.2. Proof of Theorem 1

THEOREM 1. *Under Assumptions 1, 2 and 3, $\underline{\mathcal{Q}}_s(\hat{\mathbf{x}}) = \mathcal{Q}_s(\hat{\mathbf{x}})$. Moreover, when*

$$\rho^* \geq \frac{\mathcal{Q}_s(\hat{\mathbf{x}}) - L}{d}, \quad (9)$$

$(-\mathbf{1}\rho^, -\mathbf{1}\rho^*)$ is optimal to (7), where $d = \min\{\|\mathbf{x}^k - \hat{\mathbf{x}}\|_1 : k \in K, \mathbf{x}^k \neq \hat{\mathbf{x}}\}$, and $\{(\mathbf{x}^k, \mathbf{y}^k, \omega_k)\}_{k \in K}$ are extreme points of the set $\text{conv}\{(\mathbf{x}, \mathbf{y}, \omega) : \mathbf{T}^s \mathbf{x} + \mathbf{W}^s \mathbf{y} \geq \mathbf{h}^s, \mathbf{x} \in \bar{X}, \mathbf{y} \in \mathbb{Z}^{m_1} \times \mathbb{R}^{m_2}, \|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \omega\}$.*

Proof. By weak duality, we have $\underline{\mathcal{Q}}_s(\hat{\mathbf{x}}) \leq \mathcal{Q}_s(\hat{\mathbf{x}})$. To show that $\underline{\mathcal{Q}}_s(\hat{\mathbf{x}}) \geq \mathcal{Q}_s(\hat{\mathbf{x}})$, we note

$$\underline{\mathcal{Q}}_s(\hat{\mathbf{x}}) = \sup_{\boldsymbol{\pi}^+, \boldsymbol{\pi}^- \in \mathbb{R}^n} \mathcal{L}_s(\boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}}) \geq \sup_{\rho > 0} \mathcal{L}_s(-\rho \mathbf{1}, -\rho \mathbf{1}; \hat{\mathbf{x}}),$$

where

$$\begin{aligned} \mathcal{L}_s(-\rho \mathbf{1}, -\rho \mathbf{1}; \hat{\mathbf{x}}) &= \inf_{\mathbf{x}} \{\mathcal{Q}_s(\mathbf{x}) + \rho \|\mathbf{x} - \hat{\mathbf{x}}\|_1 : \mathbf{x} \in \bar{X}\} = \inf_{\mathbf{x}, \mathbf{y}} \{(\mathbf{q}^s)^\top \mathbf{y} + \rho \|\mathbf{x} - \hat{\mathbf{x}}\|_1 : (\mathbf{x}, \mathbf{y}) \in \bar{Y}_s\} \\ &= \inf_{\mathbf{x}, \mathbf{y}, \omega} \{(\mathbf{q}^s)^\top \mathbf{y} + \rho \omega : (\mathbf{x}, \mathbf{y}) \in \bar{Y}_s, \|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \omega\}, \end{aligned} \quad (31)$$

and $\bar{Y}_s := \{(\mathbf{x}, \mathbf{y}) : \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{Z}^{m_1} \times \mathbb{R}^{m_2} : \mathbf{A} \mathbf{x} \geq \mathbf{b}, \mathbf{T}^s \mathbf{x} + \mathbf{W}^s \mathbf{y} \geq \mathbf{h}^s\}$. Since formulation (31) is MIP with rational data and $\inf_{\mathbf{x}, \mathbf{y}, \omega} \{(\mathbf{q}^s)^\top \mathbf{y} + \rho \omega : (\mathbf{x}, \mathbf{y}) \in \bar{Y}_s, \|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \omega\} \geq \inf_{\mathbf{x}, \mathbf{y}} \{(\mathbf{q}^s)^\top \mathbf{y} : (\mathbf{x}, \mathbf{y}) \in \bar{Y}_s\} \geq L$ by Assumption 2, the infimum of formulation (31) is attained at some $(\mathbf{x}_\rho, \mathbf{y}_\rho, \omega_\rho)$. From the weak duality, we have

$$\mathcal{L}_s(-\rho \mathbf{1}, -\rho \mathbf{1}; \hat{\mathbf{x}}) = (\mathbf{q}^s)^\top \mathbf{y}_\rho + \rho \omega_\rho = \mathcal{Q}_s(\mathbf{x}_\rho) + \rho \omega_\rho \leq \mathcal{Q}_s(\hat{\mathbf{x}}), \quad (32)$$

for all $\rho > 0$ and any optimal solution $(\mathbf{x}_\rho, \mathbf{y}_\rho, \omega_\rho)$ of problem (31). This implies that $\omega_\rho \leq \frac{\mathcal{Q}_s(\hat{\mathbf{x}}) - \mathcal{Q}_s(\mathbf{x}_\rho)}{\rho} \leq \frac{\mathcal{Q}_s(\hat{\mathbf{x}}) - L}{\rho}$. Letting $\rho \rightarrow \infty$, we have $\omega_\rho \rightarrow 0$. Hence, we must have $\lim_{\rho \rightarrow \infty} \mathbf{x}_\rho = \hat{\mathbf{x}}$, since $\|\mathbf{x}_\rho - \hat{\mathbf{x}}\|_1 \leq \omega_\rho$ for any $\rho > 0$. Notice that $\mathcal{Q}_s(\mathbf{x})$ is the value function of a rational MIP and is, therefore, lower semicontinuous with respect to \mathbf{x} (Meyer 1975). Thus, $\liminf_{\rho \rightarrow \infty} \mathcal{Q}_s(\mathbf{x}_\rho) \geq \mathcal{Q}_s(\hat{\mathbf{x}})$. For any $\rho > 0$, we have $\mathcal{L}_s(-\rho \mathbf{1}, -\rho \mathbf{1}; \hat{\mathbf{x}}) = \mathcal{Q}_s(\mathbf{x}_\rho) + \rho \omega_\rho \geq \mathcal{Q}_s(\mathbf{x}_\rho)$. Taking the liminf on both sides, we obtain $\liminf_{\rho \rightarrow \infty} \mathcal{L}_s(-\rho \mathbf{1}, -\rho \mathbf{1}; \hat{\mathbf{x}}) \geq \mathcal{Q}_s(\hat{\mathbf{x}})$. Meanwhile, taking the limsup on both sides

of (32) and using the fact that $\lim_{\rho \rightarrow \infty} \omega_\rho = \hat{\mathbf{x}}$, we have $\limsup_{\rho \rightarrow \infty} \mathcal{L}_s(-\rho \mathbf{1}, -\rho \mathbf{1}; \hat{\mathbf{x}}) \leq \mathcal{Q}_s(\hat{\mathbf{x}})$. Thus, $\lim_{\rho \rightarrow \infty} \mathcal{L}_s(-\rho \mathbf{1}, -\rho \mathbf{1}; \hat{\mathbf{x}}) = \mathcal{Q}_s(\hat{\mathbf{x}})$. Together with the weak duality, this shows that $\underline{\mathcal{Q}}_s(\hat{\mathbf{x}}) = \mathcal{Q}_s(\hat{\mathbf{x}})$.

Next, we prove the supremum can be attained at some finite ρ^* . Define a set

$$F := \{(\mathbf{x}, \mathbf{y}, \omega) : (\mathbf{x}, \mathbf{y}) \in \bar{Y}_s, \|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \omega\}. \quad (33)$$

When $\rho > 0$, we have

$$\begin{aligned} \mathcal{L}_s(-\rho \mathbf{1}, -\rho \mathbf{1}; \hat{\mathbf{x}}) &= \inf_{\mathbf{x}, \mathbf{y}, \omega} \{(\mathbf{q}^s)^\top \mathbf{y} + \rho \omega : (\mathbf{x}, \mathbf{y}, \omega) \in F\} = \inf_{\mathbf{x}, \mathbf{y}, \omega} \{(\mathbf{q}^s)^\top \mathbf{y} + \rho \omega : (\mathbf{x}, \mathbf{y}, \omega) \in \text{conv}(F)\} \\ &= \min_{k \in K} \{(\mathbf{q}^s)^\top \mathbf{y}^k + \rho \omega_k\} \geq \min_{k \in K} \{\mathcal{Q}_s(\mathbf{x}^k) + \rho \omega_k\}, \end{aligned}$$

where $\{(\mathbf{x}^k, \mathbf{y}^k, \omega_k)\}_{k \in K} \subseteq F$ are all extreme points of $\text{conv}(F)$. The third equality holds since $\rho > 0$ ensures that $\mathcal{L}_s(-\rho \mathbf{1}, -\rho \mathbf{1}; \hat{\mathbf{x}}) \geq L$, and the infimum is attained at some extreme point of $\text{conv}(F)$. The inequality follows because, for any feasible solution $(\mathbf{x}, \mathbf{y}) \in \bar{Y}_s$, we have $(\mathbf{q}^s)^\top \mathbf{y} \geq \mathcal{Q}_s(\mathbf{x})$ by the definition of the local recourse function. If $\mathbf{x}^k = \hat{\mathbf{x}}$ for all $k \in K$, then we have $\mathcal{L}_s(-\rho \mathbf{1}, -\rho \mathbf{1}; \hat{\mathbf{x}}) \geq \mathcal{Q}_s(\hat{\mathbf{x}})$ for any positive ρ^* . If not, let $d = \min\{\|\mathbf{x}^k - \hat{\mathbf{x}}\|_1 : k \in K, \mathbf{x}^k \neq \hat{\mathbf{x}}\}$ and let $\rho^* = \frac{\mathcal{Q}_s(\hat{\mathbf{x}}) - L}{d}$. Then, for any optimal ω_{ρ^*} , we have $\omega_{\rho^*} \leq \frac{\mathcal{Q}_s(\hat{\mathbf{x}}) - L}{\rho^*} \leq d$, according to (32). Therefore, we have

$$\mathcal{L}_s(-\rho^* \mathbf{1}, -\rho^* \mathbf{1}; \hat{\mathbf{x}}) = \min_{k \in K} \{(\mathbf{q}^s)^\top \mathbf{y}^k + \rho^* \omega_k : \omega_k \leq d\} = \min_{k \in K} \{(\mathbf{q}^s)^\top \mathbf{y}^k + \rho^* \omega_k : \omega_k \in \{0, d\}\} \geq \mathcal{Q}_s(\hat{\mathbf{x}}),$$

where the first equality follows since restricting $\omega_k \leq d$ preserves all optimal solutions, and the last inequality follows since when $\omega_k = 0$, we have $(\mathbf{q}^s)^\top \mathbf{y}^k + \rho^* \omega_k = \mathcal{Q}_s(\hat{\mathbf{x}})$, and when $\omega_k = d$, $(\mathbf{q}^s)^\top \mathbf{y}^k + \rho^* \omega_k \geq L + \frac{\mathcal{Q}_s(\hat{\mathbf{x}}) - L}{d} d = \mathcal{Q}_s(\hat{\mathbf{x}})$. Therefore, the optimal value of (7) is $\mathcal{Q}_s(\hat{\mathbf{x}})$ with an optimal solution $\boldsymbol{\pi}^{+*} = \boldsymbol{\pi}^{-*} = -\rho^* \mathbf{1}$. \square

A.3. Proof of Corollary 4

COROLLARY 4. *For any bounded set S such that $S \supseteq \bar{X}$, if $(\boldsymbol{\pi}^{+*}, \boldsymbol{\pi}^{-*})$ is optimal to $\sup_{\boldsymbol{\pi}^+, \boldsymbol{\pi}^- \in \mathbb{R}^n} \inf_{\mathbf{x} \in S} L_s(\mathbf{x}, \boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}})$, then it is also optimal to $\sup_{\boldsymbol{\pi}^+, \boldsymbol{\pi}^- \in \mathbb{R}^n} \inf_{\mathbf{x} \in \bar{X}} L_s(\mathbf{x}, \boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}})$, where*

$$L_s(\mathbf{x}, \boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}}) := \mathcal{Q}_s(\mathbf{x}) - \sum_{i \in [n]} \pi_i^+ (x_i - \hat{x}_i)^+ - \sum_{i \in [n]} \pi_i^- (x_i - \hat{x}_i)^-.$$

Proof. Since $(\boldsymbol{\pi}^{+*}, \boldsymbol{\pi}^{-*})$ is optimal to $\sup_{\boldsymbol{\pi}^+, \boldsymbol{\pi}^- \in \mathbb{R}^n} \inf_{\mathbf{x} \in S} L_s(\mathbf{x}, \boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}})$, we have $\inf_{\mathbf{x} \in S} L_s(\mathbf{x}, \boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}}) = \mathcal{Q}_s(\hat{\mathbf{x}})$. Taking the infimum over a subset $\bar{X} \subseteq S$, we have $\inf_{\mathbf{x} \in \bar{X}} L_s(\mathbf{x}, \boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}}) \geq \inf_{\mathbf{x} \in S} L_s(\mathbf{x}, \boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}}) = \mathcal{Q}_s(\hat{\mathbf{x}})$. At the same time, we have $L_s(\hat{\mathbf{x}}, \boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}}) = \mathcal{Q}_s(\hat{\mathbf{x}})$ and $\hat{\mathbf{x}} \in \bar{X}$. Thus, we can conclude that $\inf_{\mathbf{x} \in \bar{X}} L_s(\mathbf{x}, \boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}}) = \mathcal{Q}_s(\hat{\mathbf{x}})$. This implies that $(\boldsymbol{\pi}^{+*}, \boldsymbol{\pi}^{-*})$ is also optimal when restricting $\mathbf{x} \in \bar{X}$. \square

A.4. Proof of Corollary 5

COROLLARY 5. *For a given feasible first-stage decision $\hat{\mathbf{x}} \in \bar{X}$, $(\hat{\boldsymbol{\pi}}^+, \hat{\boldsymbol{\pi}}^-)$ is optimal to (7) if and only if*

$$\mathcal{Q}_s(\hat{\mathbf{x}}) \leq \mathcal{Q}_s(\mathbf{x}) - \sum_{i \in [n]} \hat{\pi}_i^+(x_i - \hat{x}_i)^+ - \sum_{i \in [n]} \hat{\pi}_i^-(x_i - \hat{x}_i)^-, \quad (10)$$

for all $x \in \bar{X}$, or equivalently,

$$\theta \geq \mathcal{Q}_s(\mathbf{x}) \geq \mathcal{Q}_s(\hat{\mathbf{x}}) + \sum_{i \in [n]} \hat{\pi}_i^+(x_i - \hat{x}_i)^+ + \sum_{i \in [n]} \hat{\pi}_i^-(x_i - \hat{x}_i)^-,$$

for all (\mathbf{x}, θ) in the epigraph of \mathcal{Q}_s .

Proof. Given $\hat{\mathbf{x}} \in \bar{X}$ and $(\hat{\boldsymbol{\pi}}^+, \hat{\boldsymbol{\pi}}^-) \in \mathbb{R}^{2n}$, if the condition (10) is satisfied, we have $\mathcal{L}_s(\hat{\boldsymbol{\pi}}^+, \hat{\boldsymbol{\pi}}^-; \hat{\mathbf{x}}) = \inf_{\mathbf{x}} \left\{ \mathcal{Q}_s(\mathbf{x}) - \sum_{i \in [n]} \hat{\pi}_i^+(x_i - \hat{x}_i)^+ - \sum_{i \in [n]} \hat{\pi}_i^-(x_i - \hat{x}_i)^- : \mathbf{x} \in \bar{X} \right\} \geq \mathcal{Q}_s(\hat{\mathbf{x}})$. Meanwhile, if $(\hat{\boldsymbol{\pi}}^+, \hat{\boldsymbol{\pi}}^-)$ is optimal to (7), we have $\mathcal{L}_s(\hat{\boldsymbol{\pi}}^+, \hat{\boldsymbol{\pi}}^-; \hat{\mathbf{x}}) = \mathcal{Q}_s(\hat{\mathbf{x}})$ according to the strong duality shown in Theorem 1.

Therefore, we have

$$\begin{aligned} \mathcal{Q}_s(\hat{\mathbf{x}}) &= \mathcal{L}_s(\hat{\boldsymbol{\pi}}^+, \hat{\boldsymbol{\pi}}^-; \hat{\mathbf{x}}) = \mathcal{L}_s(\boldsymbol{\pi}^+, \boldsymbol{\pi}^-; \hat{\mathbf{x}}) = \inf_{\mathbf{x}} \left\{ \mathcal{Q}_s(\mathbf{x}) - \sum_{i \in [n]} \hat{\pi}_i^+(x_i - \hat{x}_i)^+ - \sum_{i \in [n]} \hat{\pi}_i^-(x_i - \hat{x}_i)^- : \mathbf{x} \in \bar{X} \right\} \\ &\leq \mathcal{Q}_s(\mathbf{x}) - \sum_{i \in [n]} \hat{\pi}_i^+(x_i - \hat{x}_i)^+ - \sum_{i \in [n]} \hat{\pi}_i^-(x_i - \hat{x}_i)^- \end{aligned}$$

for all feasible $\mathbf{x} \in \bar{X}$. \square

A.5. Proof of Proposition 3

PROPOSITION 3. *Given a first-stage decision $\hat{\mathbf{x}} \in \bar{X}$, we have*

$$\text{conv}(S_1) = \{(\mathbf{x}, \boldsymbol{\theta}) \in \times_{i \in [n]} [0, B_i] \times \mathbb{R} : \exists (\boldsymbol{\omega}^+, \boldsymbol{\omega}^-, \mathbf{z}) \in \mathbb{R}^n \times \mathbb{R}^n \times [0, 1]^n, (13a), (13b)\}. \quad (14)$$

Proof. Let us denote the right-hand side of (14) by S_2 , which is a continuous relaxation of S_1 . Hence, we must have $S_1 \subseteq S_2$.

To prove $S_1 \supseteq S_2$, we first observe that $\text{conv}(S_1) = \text{conv}\{(\mathbf{x}, \boldsymbol{\theta}) : \exists (\boldsymbol{\omega}^+, \boldsymbol{\omega}^-, \mathbf{z}) \in \mathbb{R}^n \times \mathbb{R}^n \times [0, 1]^n, (13a), (13b), 0 \leq x_i \leq B_i, \forall i \in [n]\}$. For any $(\mathbf{x}, \boldsymbol{\theta}) \in S_2$, there exists a solution $(\boldsymbol{\omega}^+, \boldsymbol{\omega}^-, \mathbf{z})$ such that (14) holds. Without loss of generality, we assume $z_0 := 1 \geq z_1 \geq \dots \geq z_n$. To show $(\mathbf{x}, \boldsymbol{\theta}) \in \text{conv}(S_1)$, we construct $n+1$ points $(\mathbf{x}^k, \boldsymbol{\theta}_k)_{k \in [0, n]}$ as follows. For $k \in [0, n]$, let $\lambda_k = z_k - z_{k+1}$, and define $\mathbf{z}^k \in \{0, 1\}^n$ as

$$z_i^k = \begin{cases} 1, & i \leq k, \\ 0, & i > k, \end{cases}$$

for all $i \in [n]$. Define $(\boldsymbol{\omega}^+)^k, (\boldsymbol{\omega}^-)^k \in \mathbb{R}^n$ as follows: If $z_k = 1$, let

$$(\boldsymbol{\omega}^+)^k_i = \begin{cases} \omega_i^+, & i \leq k, \\ 0, & i > k, \end{cases} \quad \text{and } (\boldsymbol{\omega}^-)^k = \mathbf{0}.$$

If $0 < z_k < 1$, let

$$(\omega^+)_i^k = \begin{cases} \frac{\omega_i^+}{z_i}, & i \leq k, \\ 0, & i > k, \end{cases} \quad \text{and} \quad (\omega^-)_i^k = \begin{cases} 0, & i \leq k, \\ \frac{\omega_i^-}{1-z_i}, & i > k. \end{cases}$$

If $z_k = 0$, let

$$(\omega^+)^k = \mathbf{0} \quad \text{and} \quad (\omega^-)_i^k = \begin{cases} 0, & i \leq k, \\ \omega_i^-, & i > k. \end{cases}$$

With \mathbf{x}^k defined as $\mathbf{x}^k = \hat{\mathbf{x}} + (\omega^+)^k - (\omega^-)^k$, the above formulations ensure that constraints (13b) are satisfied. To satisfy condition (13a), we let $\theta^k = \theta - \sum_{i \in [n]} (\pi_i^+ \omega_i^+ + \pi_i^- \omega_i^-) + \sum_{i \in [n]} (\pi_i^+ (\omega^+)_i^k + \pi_i^- (\omega^-)_i^k)$. Thus, we have $\theta^k - \sum_{i \in [n]} (\pi_i^+ (\omega^+)_i^k + \pi_i^- (\omega^-)_i^k) = \theta - \sum_{i \in [n]} (\pi_i^+ \omega_i^+ + \pi_i^- \omega_i^-) \geq \mathcal{Q}_s(\hat{\mathbf{x}})$. It is easy to verify that $(\mathbf{x}, \theta, \omega^+, \omega^-, \mathbf{z}) = \sum_{k \in [0, n]} \lambda_k (\mathbf{x}^k, \theta^k, (\omega^+)^k, (\omega^-)^k, \mathbf{z}^k)$. Thus, $(\mathbf{x}, \theta) \in \text{conv}(S_1)$. Consequently, we have $\text{conv}(S_1) = S_2$. \square

A.6. Proof of Proposition 4

PROPOSITION 4. *Any tight Lagrangian cut is a ReLU Lagrangian cut.*

Proof. Given $\hat{\mathbf{x}} \in \bar{X}$ and $\boldsymbol{\pi} \in \mathbb{R}^n$ such that $\mathcal{L}(\boldsymbol{\pi}; \hat{\mathbf{x}}) = \mathcal{Q}_s(\hat{\mathbf{x}})$, we can derive a tight Lagrangian cut $\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) + \boldsymbol{\pi}^\top (\mathbf{x} - \hat{\mathbf{x}})$. It is equivalent to $\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) + \sum_{i \in [n]} \pi_i (x_i - \hat{x}_i)^+ - \sum_{i \in [n]} \pi_i (x_i - \hat{x}_i)^-$, which is a ReLU Lagrangian cut according to Corollary 5. \square

A.7. Proof of Proposition 8

PROPOSITION 8. *Given scenario $s \in [N]$, the set $\Pi_{\hat{\mathbf{x}}}^s$ is a polyhedron.*

Proof. According to Definition 4, from the optimality condition (10), we have

$$\begin{aligned} \Pi_{\hat{\mathbf{x}}}^s &= \left\{ (\boldsymbol{\pi}^+, \boldsymbol{\pi}^-) \in \mathbb{R}^{2n} : \min_{\mathbf{x}} \{ \mathcal{Q}_s(\mathbf{x}) - (\boldsymbol{\pi}^+)^\top (\mathbf{x} - \hat{\mathbf{x}})^+ - (\boldsymbol{\pi}^-)^\top (\mathbf{x} - \hat{\mathbf{x}})^- : \mathbf{x} \in \bar{X} \} \geq \mathcal{Q}_s(\hat{\mathbf{x}}) \right\} \\ &= \left\{ (\boldsymbol{\pi}^+, \boldsymbol{\pi}^-) \in \mathbb{R}^{2n} : \min_{\mathbf{x}, \omega^+, \omega^-, \mathbf{z}} \{ \mathcal{Q}_s(\mathbf{x}) - (\boldsymbol{\pi}^+)^\top \omega^+ - (\boldsymbol{\pi}^-)^\top \omega^- : \mathbf{x} \in \bar{X}, (13b), (13c) \} \geq \mathcal{Q}_s(\hat{\mathbf{x}}) \right\} \\ &= \left\{ (\boldsymbol{\pi}^+, \boldsymbol{\pi}^-) \in \mathbb{R}^{2n} : \min_{\mathbf{x}, \omega^+, \omega^-, \mathbf{z}, \mathbf{y}} \{ (\mathbf{q}^s)^\top \mathbf{y} - (\boldsymbol{\pi}^+)^\top \omega^+ - (\boldsymbol{\pi}^-)^\top \omega^- : (\mathbf{x}, \mathbf{y}) \in \bar{Y}_s, (13b), (13c) \} \geq \mathcal{Q}_s(\hat{\mathbf{x}}) \right\} \\ &= \left\{ (\boldsymbol{\pi}^+, \boldsymbol{\pi}^-) \in \mathbb{R}^{2n} : (\omega^{+k})^\top \boldsymbol{\pi}^+ + (\omega^{-k})^\top \boldsymbol{\pi}^- \leq (\mathbf{q}^s)^\top \mathbf{y}^k - \mathcal{Q}_s(\hat{\mathbf{x}}), \forall k \in K \right\}, \end{aligned}$$

where $\{(\mathbf{x}^k, \omega^{+k}, \omega^{-k}, \mathbf{z}^k, \mathbf{y}^k)\}_{k \in K}$ are all extreme points of the polyhedron $\text{conv}\{(\mathbf{x}, \omega^+, \omega^-, \mathbf{z}, \mathbf{y}) : (\mathbf{x}, \mathbf{y}) \in \bar{Y}_s, (13b), (13c)\}$. This implies that $\Pi_{\hat{\mathbf{x}}}^s$ is defined by finitely many linear inequalities and, therefore, is a polyhedron. \square

A.8. Proof of Lemma 1

LEMMA 1. *Given scenario $s \in [N]$, when the first-stage feasible decision is purely binary, the recession cone of $\Pi_{\hat{\mathbf{x}}}^s$ contains $\text{lin}\{\mathbf{e}_i, \mathbf{0}\}_{i \in I_{\hat{\mathbf{x}}}} + \text{lin}\{\mathbf{0}, \mathbf{e}_i\}_{i \notin I_{\hat{\mathbf{x}}}}$.*

Proof. Given $\hat{\mathbf{x}} \in X$, for any binary \mathbf{x} , we have $\omega_i^+ = (x_i - \hat{x}_i)^+ = 0$ for all $i \in I_{\hat{\mathbf{x}}}$ and $\omega_i^- = (x_i - \hat{x}_i)^- = 0$ for all $i \notin I_{\hat{\mathbf{x}}}$. Thus, $(\boldsymbol{\omega}^{+k})^\top \mathbf{e}_i + (\boldsymbol{\omega}^{-k})^\top \mathbf{0} = 0$ for all $i \in I_{\hat{\mathbf{x}}}$ and $(\boldsymbol{\omega}^{+k})^\top \mathbf{0} + (\boldsymbol{\omega}^{-k})^\top \mathbf{e}_i = 0$ for all $i \notin I_{\hat{\mathbf{x}}}$. \square

A.9. Proof of Theorem 2

THEOREM 2. *The following properties hold for the set $\mathcal{D}_{\hat{\mathbf{x}}}^s$ of dual optimal solutions and Lagrangian cuts generated at $\hat{\mathbf{x}} \in \bar{X}$: (i) The set $\mathcal{D}_{\hat{\mathbf{x}}}^s$ is a polyhedron; (ii) A Lagrangian cut $\theta \geq \text{co}(\mathcal{Q}_s)(\hat{\mathbf{x}}) + \boldsymbol{\pi}^\top(\mathbf{x} - \hat{\mathbf{x}})$ is facet-defining if and only if $\boldsymbol{\pi}$ is an extreme point of $\mathcal{D}_{\hat{\mathbf{x}}}^s$; (iii) The recession cone of $\mathcal{D}_{\hat{\mathbf{x}}}^s$ is the normal cone of \bar{X} at $\hat{\mathbf{x}}$; and (iv) The local convex hull can be described using the facet-defining Lagrangian cuts:*

$$\text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s)) = \{(\mathbf{x}, \theta) \in \text{conv}(\bar{X}) \times \mathbb{R} : \theta \geq \text{co}(\mathcal{Q}_s)(\hat{\mathbf{x}}) + \boldsymbol{\pi}^\top(\mathbf{x} - \hat{\mathbf{x}}), \forall \hat{\mathbf{x}} \in \bar{X}, \boldsymbol{\pi} \in \text{ext}(\mathcal{D}_{\hat{\mathbf{x}}}^s)\}.$$

Proof. For a given $\hat{\mathbf{x}} \in \bar{X}$, from the optimality condition (5), we have

$$\mathcal{D}_{\hat{\mathbf{x}}}^s = \left\{ \boldsymbol{\pi} \in \mathbb{R}^n : \min_{\mathbf{x}, \theta} \{ \theta + \boldsymbol{\pi}^\top(\hat{\mathbf{x}} - \mathbf{x}) : (\mathbf{x}, \theta) \in \text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s)) \} \geq \text{co}(\mathcal{Q}_s)(\hat{\mathbf{x}}) \right\},$$

where $\text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s)) = \text{conv}\{(\mathbf{x}^k, \mathcal{Q}(\mathbf{x}^k)) : k \in K\} + \text{cone}\{(\mathbf{0}, 1)\}$, and K is finite since $\text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s))$ is a polyhedron. We further obtain that

$$\mathcal{D}_{\hat{\mathbf{x}}}^s = \{ \boldsymbol{\pi} \in \mathbb{R}^n : \mathcal{Q}_s(\mathbf{x}^k) + \boldsymbol{\pi}^\top(\hat{\mathbf{x}} - \mathbf{x}^k) \geq \text{co}(\mathcal{Q}_s)(\hat{\mathbf{x}}), \forall k \in K, 1 - \boldsymbol{\pi}^\top \mathbf{0} \geq 0 \}. \quad (34)$$

which involves finitely many linear inequalities and, therefore, is a polyhedron.

Suppose that set $\text{conv}(\bar{X})$ is full-dimensional. Then $\text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s))$ has dimension $n + 1$. A cut $\theta \geq \text{co}(\mathcal{Q}_s)(\hat{\mathbf{x}}) + \boldsymbol{\pi}^\top(\mathbf{x} - \hat{\mathbf{x}})$ is facet-defining if and only if it is tight at $n + 1$ extreme points of $\text{conv}(\text{epi}_{\bar{X}}(\mathcal{Q}_s))$ that are affinely independent. That is, $\boldsymbol{\pi}$ satisfies n linearly independent inequalities with equality in (34), meaning that it is an extreme point of $\Pi_{\hat{\mathbf{x}}}$. When \bar{X} is not full dimensional, we can reduce it to a lower dimensional space where the same result holds.

From the optimality condition (5), we have

$$\mathcal{D}_{\hat{\mathbf{x}}}^s = \{ \boldsymbol{\pi} \in \mathbb{R}^n : \boldsymbol{\pi}^\top(\hat{\mathbf{x}} - \mathbf{x}) \geq \text{co}(\mathcal{Q}_s)(\hat{\mathbf{x}}) - \mathcal{Q}_s(\mathbf{x}), \forall \mathbf{x} \in \bar{X} \}.$$

Thus, $\text{rc}(\mathcal{D}_{\hat{\mathbf{x}}}^s) = \{ \boldsymbol{\pi} \in \mathbb{R}^n : \boldsymbol{\pi}^\top(\hat{\mathbf{x}} - \mathbf{x}) \geq 0, \forall \mathbf{x} \in \bar{X} \} = \{ \boldsymbol{\pi} \in \mathbb{R}^n : \boldsymbol{\pi}^\top(\hat{\mathbf{x}} - \mathbf{x}) \geq 0, \forall \mathbf{x} \in \text{conv}(\bar{X}) \} = N_{\text{conv}(\bar{X})}(\hat{\mathbf{x}})$. \square

A.10. Proof of Proposition 9

PROPOSITION 9. *The strengthening problem (Stren) is bounded if and only if $-\mathbf{a} \in \mathcal{T}_{\text{conv}(X)}(\hat{\mathbf{x}})$.*

Proof. Note that LP (Stren) is always feasible with $\boldsymbol{\eta} = \mathbf{0}$ being a feasible solution. Hence, the strengthening problem (Stren) is bounded if and only if its dual problem

$$\max \left\{ \sum_{k \in K} (\mathcal{Q}_s(\hat{\mathbf{x}}) - \mathcal{Q}_s(\mathbf{x}^k) - \hat{\boldsymbol{\pi}}^\top (\hat{\mathbf{x}} - \mathbf{x}^k)) y_k : \sum_{k \in K} (\hat{\mathbf{x}} - \mathbf{x}^k) y_k = \mathbf{a}, \mathbf{y} \geq \mathbf{0} \right\},$$

where $\{\mathbf{x}^k\}_{k \in K} = X$, is feasible. Therefore, by definition, the strengthening problem (Stren) is bounded if and only if $-\mathbf{a} \in \mathcal{T}_{\text{conv}(X)}(\hat{\mathbf{x}})$. \square

A.11. Proof of Proposition 10

PROPOSITION 10. *Any feasible solution of (24) is also feasible for (Stren).*

Proof. Since $\mathcal{Q}_s^{LP}(\mathbf{x})$ is obtained by solving the LP relaxation of the local recourse problem, we have $\mathcal{Q}_s^{LP}(\mathbf{x}) \leq \mathcal{Q}_s(\mathbf{x})$ for all $\mathbf{x} \in X$. Thus, for any $\boldsymbol{\eta}$ feasible for (24), we have $\min_{\mathbf{x}} \{\mathcal{Q}_s(\mathbf{x}) + (\hat{\boldsymbol{\pi}} + \boldsymbol{\eta})^\top (\hat{\mathbf{x}} - \mathbf{x}) : \mathbf{x} \in X\} \geq \min_{\mathbf{x}} \{\mathcal{Q}_s^{LP}(\mathbf{x}) + (\hat{\boldsymbol{\pi}} + \boldsymbol{\eta})^\top (\hat{\mathbf{x}} - \mathbf{x}) : \mathbf{x} \in X\} \geq \min_{\mathbf{x}} \{\mathcal{Q}_s^{LP}(\mathbf{x}) + (\hat{\boldsymbol{\pi}} + \boldsymbol{\eta})^\top (\hat{\mathbf{x}} - \mathbf{x}) : \mathbf{x} \in X^{LP}\} \geq \mathcal{Q}_s(\hat{\mathbf{x}})$. Therefore, $\boldsymbol{\eta}$ is also feasible for (Stren). \square

A.12. Proof of Proposition 11

PROPOSITION 11. *The set*

$$F_s^{LP} := \left\{ \boldsymbol{\eta} : \min_{\mathbf{x}} \left\{ \mathcal{Q}_s^{LP}(\mathbf{x}) + (\hat{\boldsymbol{\pi}} + \boldsymbol{\eta})^\top (\hat{\mathbf{x}} - \mathbf{x}) : \mathbf{x} \in \tilde{X} \right\} \geq \mathcal{Q}_s(\hat{\mathbf{x}}) \right\} \quad (25)$$

is always nonempty, where $\tilde{X} := \{\mathbf{x} \in X^{LP} : \boldsymbol{\chi}^\top (\hat{\mathbf{x}} - \mathbf{x}) \geq 1\}$.

Proof. According to (25), we have

$$\begin{aligned} F_s^{LP} &= \left\{ \boldsymbol{\eta} : \mathcal{Q}_s^{LP}(\mathbf{x}) + (\hat{\boldsymbol{\pi}} + \boldsymbol{\eta})^\top (\hat{\mathbf{x}} - \mathbf{x}) \geq \mathcal{Q}_s(\hat{\mathbf{x}}), \forall \mathbf{x} \in \tilde{X} \right\} \\ &= \left\{ \boldsymbol{\eta} : (\mathbf{x}^\top \boldsymbol{\theta}) \begin{pmatrix} -(\hat{\boldsymbol{\pi}} + \boldsymbol{\eta}) \\ 1 \end{pmatrix} \geq (\hat{\mathbf{x}}^\top \mathcal{Q}_s(\hat{\mathbf{x}})) \begin{pmatrix} -(\hat{\boldsymbol{\pi}} + \boldsymbol{\eta}) \\ 1 \end{pmatrix}, \forall (\mathbf{x}, \boldsymbol{\theta}) \in \text{epi}_{\tilde{X}} \mathcal{Q}_s^{LP} \right\}. \end{aligned}$$

Note that

$$\text{epi}_{\tilde{X}}(\mathcal{Q}_s^{LP}) = \{(\mathbf{x}, \boldsymbol{\theta}) \in X^{LP} \times \mathbb{R} : \exists \mathbf{y} \in \mathbb{R}^{m_1+m_2} : \boldsymbol{\theta} \geq (\mathbf{q}^s)^\top \mathbf{y}, \mathbf{T}^s \mathbf{x} + \mathbf{W}^s \mathbf{y} \geq \mathbf{h}^s, \boldsymbol{\chi}^\top (\hat{\mathbf{x}} - \mathbf{x}) \geq 1\}$$

is a polyhedron and, therefore, closed and convex. Since $(\hat{\mathbf{x}}^\top, \mathcal{Q}_s(\hat{\mathbf{x}})) \notin \text{epi}_{\tilde{X}} \mathcal{Q}_s^{LP}$, by the separation theorem, there exists $\boldsymbol{\alpha}, \beta, \gamma$ such that $\boldsymbol{\alpha}^\top \hat{\mathbf{x}} + \beta \mathcal{Q}_s(\hat{\mathbf{x}}) < \gamma$ and $\boldsymbol{\alpha}^\top \mathbf{x} + \beta \boldsymbol{\theta} \geq \gamma$ for all $(\mathbf{x}, \boldsymbol{\theta}) \in \text{epi}_{\tilde{X}}(\mathcal{Q}_s^{LP})$. If $\beta \neq 0$, since $(\mathbf{0}, 1)$ is an extreme ray of $\text{epi}_{\tilde{X}}(\mathcal{Q}_s^{LP})$, we have $\beta > 0$. Then, $-\boldsymbol{\alpha}/\beta - \hat{\boldsymbol{\pi}} \in F_s^{LP}$. If $\beta = 0$, let

$$\bar{\beta} = \frac{\gamma - \boldsymbol{\alpha}^\top \hat{\mathbf{x}}}{2 \max\{|\mathbf{L}|, \mathcal{Q}_s(\hat{\mathbf{x}}), 1\}} > 0.$$

Then we have a separation $\boldsymbol{\alpha}^\top \mathbf{x} + \bar{\beta} \boldsymbol{\theta} \geq \boldsymbol{\alpha}^\top \mathbf{x} + \bar{\beta} \mathbf{L} \geq \gamma - \frac{\gamma - \boldsymbol{\alpha}^\top \hat{\mathbf{x}}}{2} = \frac{\gamma + \boldsymbol{\alpha}^\top \hat{\mathbf{x}}}{2}$ and $\boldsymbol{\alpha}^\top \hat{\mathbf{x}} + \bar{\beta} \mathcal{Q}_s(\hat{\mathbf{x}}) \leq \boldsymbol{\alpha}^\top \hat{\mathbf{x}} + \frac{\gamma - \boldsymbol{\alpha}^\top \hat{\mathbf{x}}}{2} = \frac{\gamma + \boldsymbol{\alpha}^\top \hat{\mathbf{x}}}{2}$. Hence, $-\boldsymbol{\alpha}/\bar{\beta} - \hat{\boldsymbol{\pi}} \in F_s^{LP}$. \square

A.13. Proof of Proposition 12

PROPOSITION 12. *Given $\hat{\mathbf{x}} \in \bar{X}$, let us define $E_\Lambda^s = \{(\mathbf{x}, \theta) \in \mathcal{B} \times \mathbb{R} : (15)\}$ and*

$$E_L^s = \left\{ (\mathbf{x}, \theta) \in \mathcal{B} \times \mathbb{R} : \exists \boldsymbol{\delta} \in \{0, 1\}^{N_1+1} \times \dots \times \{0, 1\}^{N_n+1}, x_i = \sum_{j \in [0, N_i]} 2^j \delta_j^i, \forall i \in [n], (27) \right\},$$

where $\mathcal{B} := \times_{i \in [n]} [0, B_i]$. If $B_i \geq 3$ for all $i \in [n]$, then $\text{conv}(E_\Lambda^s) \supseteq \text{conv}(E_L^s)$.

Proof. According to Proposition 3, we have

$$\text{conv}(E_\Lambda^s) = \left\{ (\mathbf{x}, \theta) \in \times_{i \in [n]} [0, B_i] \times \mathbb{R} : \begin{cases} \exists (\boldsymbol{\omega}^+, \boldsymbol{\omega}^-, \mathbf{z}) \in \mathbb{R}^n \times \mathbb{R}^n \times [0, 1]^n, \\ \theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) + (L - \mathcal{Q}_s(\hat{\mathbf{x}})) \sum_{i \in [n]} (\omega_i^+ + \omega_i^-), \\ \omega_i^+ - \omega_i^- = x_i - \hat{x}_i, 0 \leq \omega_i^+ \leq (B_i - \hat{x}_i) z_i, \\ 0 \leq \omega_i^- \leq \hat{x}_i (1 - z_i), \forall i \in [n] \end{cases} \right\}.$$

it suffices to show that $\text{conv}(E_\Lambda^s) \supseteq E_L^s$. We observe that $(\mathbf{x}, \mathcal{Q}_s(\hat{\mathbf{x}}) + (L - \mathcal{Q}_s(\hat{\mathbf{x}})) \sum_{i \in [n]} \max\{B_i - \hat{x}_i, \hat{x}_i\}) \in \text{conv}(E_\Lambda^s)$ for all $\mathbf{x} \in \mathcal{B}$ by letting $z_i = 1$ if $B_i - \hat{x}_i > \hat{x}_i$ and 0, otherwise, for all $i \in [n]$. It remains to show that

CLAIM 1. $\|\boldsymbol{\delta}^i - \hat{\boldsymbol{\delta}}^i\| \leq \max\{B_i - \hat{x}_i, \hat{x}_i\}$ for all $\mathbf{x} \in \mathcal{B}$ and $i \in [n]$.

According to Claim 1, for any $(\mathbf{x}, \theta) \in E_L^s$, we have $\theta \geq \mathcal{Q}_s(\hat{\mathbf{x}}) + (L - \mathcal{Q}_s(\hat{\mathbf{x}})) \sum_{i \in [n]} \|\boldsymbol{\delta}^i - \hat{\boldsymbol{\delta}}^i\|_1 \geq \mathcal{Q}_s(\hat{\mathbf{x}}) + (L - \mathcal{Q}_s(\hat{\mathbf{x}})) \sum_{i \in [n]} \max\{B_i - \hat{x}_i, \hat{x}_i\}$. Hence, $(\mathbf{x}, \theta) \in \text{conv}(E_\Lambda^s)$ and $E_L^s \subseteq \text{conv}(E_\Lambda^s)$.

To prove Claim 1, notice that since B_i is an integer and $N_i = \lfloor \log_2 B_i \rfloor$, we have $\max\{B_i - \hat{x}_i, \hat{x}_i\} \geq \lfloor B_i/2 \rfloor \geq 2^{N_i-1}$. On the other hand, we observe that $\|\boldsymbol{\delta}^i - \hat{\boldsymbol{\delta}}^i\|_1 \leq N_i + 1$. There are three cases:

Case 1: $B_i \geq 5$ or $B_i = 3$. In this case, we have $\lfloor B_i/2 \rfloor \geq N_i + 1$.

Case 2: $B_i = 4$, if $\hat{x}_i = 0, 1, 3, 4$, we have $\|\boldsymbol{\delta}^i - \hat{\boldsymbol{\delta}}^i\|_1 \leq 3 \leq \max\{B_i - \hat{x}_i, \hat{x}_i\}$.

Case 3: $B_i = 4$ and $\hat{x}_i = 2$. In this case, we have $\max\{B_i - \hat{x}_i, \hat{x}_i\} = 2$, $\hat{\boldsymbol{\delta}}^i = (0, 1, 0)^\top$, and $\|\boldsymbol{\delta}^i - \hat{\boldsymbol{\delta}}^i\|_1 \leq 3$. Note that $\|\boldsymbol{\delta}^i - \hat{\boldsymbol{\delta}}^i\|_1 = 3$ if and only if $\boldsymbol{\delta}^i = (1, 0, 1)^\top$, i.e., $x_i = 5$, contradicting $x_i \leq B_i = 4$. Hence, we have $\|\boldsymbol{\delta}^i - \hat{\boldsymbol{\delta}}^i\|_1 \leq 2$.

Therefore, we must have $\max\{B_i - \hat{x}_i, \hat{x}_i\} \geq \|\boldsymbol{\delta}^i - \hat{\boldsymbol{\delta}}^i\|_1$ if $B_i \geq 3$ for any $i \in [n]$. \square

A.14. Proof of Proposition 13

PROPOSITION 13. *Suppose that integer $B_i = 2^{j_{1i}} + \dots + 2^{j_{\ell i}} + 2^{N_i}$, and set $J_i = \{j_{1i}, \dots, j_{\ell i}, N_i\}$ for each $i \in [n]$. Then,*

$$\text{conv}(E_L^s) = \left\{ (\mathbf{x}, \theta) \in \times_{i \in [n]} [0, B_i] \times \mathbb{R} : \begin{cases} \exists \boldsymbol{\delta} \in [0, 1]^{N_1+1} \times \dots \times [0, 1]^{N_n+1}, (27), \\ x_i = \sum_{j \in [0, N_i]} 2^j \delta_j^i \leq B_i, \forall i \in [n], \\ \delta_r^i + \sum_{\tau \in J_{ir}} \delta_\tau^i \leq |J_{ir}|, \forall r \in \{0, 1, \dots, N_i\} \setminus J_i, \forall i \in [n] \end{cases} \right\},$$

where we let $J_{ir} = \{\ell \in J_i : \ell > r\}$.

Proof. Let $D_i = \left\{ \boldsymbol{\delta}^i \in \{0, 1\}^{N_i+1} : \sum_{j \in [0, N_i]} 2^j \delta_j^i \leq B_i \right\}$ for each $i \in [n]$. Then, according to the decomposition structure of set E_L^s , we have

$$\text{conv}(E_L^s) = \left\{ \begin{array}{l} \exists \boldsymbol{\delta} \in [0, 1]^{N_1+1} \times \dots \times [0, 1]^{N_n+1}, (27), \\ (\boldsymbol{x}, \boldsymbol{\theta}) \in \times_{i \in [n]} [0, B_i] \times \mathbb{R} : x_i = \sum_{j \in [0, N_i]} 2^j \delta_j^i, \forall i \in [n], \\ \boldsymbol{\delta}^i \in \text{conv}(D_i), \forall i \in [n]. \end{array} \right\},$$

On the other hand, according to Laurent and Sassano (1992), Gupte et al. (2013), the convex hulls of the D_i 's can be described using their knapsack structure. Suppose that $B_i = 2^{j_{1i}} + \dots + 2^{j_{\ell i}} + 2^{N_i}$, and let $J_i = \{j_{1i}, \dots, j_{\ell i}, N_i\}$ for each $i \in [n]$. Then, we have

$$\text{conv}(D_i) = \left\{ \boldsymbol{\delta}^i \in [0, 1]^{N_i+1} : \delta_r^i + \sum_{\tau \in J_{ir}} \delta_\tau^i \leq |J_{ir}|, \forall r \in \{0, 1, \dots, N_i\} \setminus J_i \right\},$$

where $J_{ir} = \{\ell \in J_i : \ell > r\}$. This completes the proof. \square