On Lipschitz regularization and Lagrangian cuts in multistage stochastic mixed-integer linear programming

Christian Füllner[1*], X. Andy Sun[2], and Steffen Rebennack[1]

[1]Institute for Operations Research (IOR), Stochastic Optimization (SOP), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[*]Address correspondence to: christian.fuellner@kit.edu
[2]Sloan School of Management, Massachusetts Institute of Technology, USA

Abstract

We provide new theoretical insight on the generation of linear and non-convex cuts for value functions of multistage stochastic mixed-integer programs based on Lagrangian duality. First, we analyze in detail the impact that the introduction of copy constraints, and especially, the choice of the accompanying constraint set for the copy variable have on the properties of the Lagrangian dual and the obtained cuts. We show that the well-known tightness result for Lagrangian cuts in stochastic dual dynamic programming (SDDiP) crucially depends on this choice, and not on the introduction of copy constraints in itself. Afterwards, we generalize our results to the case where a Lipschitz regularization is applied to the value functions. In particular, we show a deep relation between norm-bounded Lagrangian dual problems and the closed convex envelope of the regularized value functions. For linear Lagrangian cuts, using an appropriate regularization, this result can be used to enhance the tightness result from SDDiP to the regularized case. For the generation of non-convex cuts, we pick up on the lift-and-project idea proposed by Füllner and Rebennack in their non-convex nested Benders decomposition (NC-NBD) method. We generalize this cut generation idea to the stochastic case. We then show that by careful choice of the norm used for regularization in the lifted space, Lipschitz continuity of the obtained non-convex cuts can be guaranteed. By that, we resolve an open theoretical question from the original NC-NBD paper. We highlight all our results by simple illustrative examples. Our work allows for a profound understanding of how and to which effect copy constraints and regularization may be used in decomposition methods in stochastic mixed-integer programming.

1 Introduction

1.1 Motivation and prior work

In many practical applications, sequential decisions have to be made over a finite number of stages, while some of the problem data of the following stages are subject to uncertainty. Such decision-making processes can be modeled as multistage stochastic programs, which under certain assumptions can be reformulated as large-scale deterministic problems. However, for a practically relevant number of scenarios, these problems get too large to be solved by off-the-shelf solvers. Therefore, they are usually approached by decomposition methods.

For multistage stochastic linear programs (MS-LP), these decomposition methods have a long tradition and are well-studied. Among the most prominent ones are nested Benders decomposition (NBD) [5] and stochastic dual dynamic programming (SDDP) [20]. One of their key ideas is to decompose the original multistage problem by stage and scenario into subproblems, which are linked by state variables and (expected) value functions. These functions are piecewise-linear and convex, and thus can be iteratively approximated by linear cutting-planes. Finitely many such cuts are sufficient to ensure (almost sure) convergence.

However, in many applications, some of the decisions have to be integer or binary, which yields a multistage stochastic mixed-integer linear program (MS-MILP). Problems of this class are very hard to solve, as they combine the challenges of dynamic and stochastic programming with the non-convexity of mixed-integer programs. In particular, the value functions become non-convex and discontinuous, which aggravates their approximation.

Various strategies have been proposed to solve MS-MILPs. A natural approach is to relax the integer constraints to obtain an MS-LP that can be solved by existing methods. However, in that case not the original MS-MILP is solved. The same issue occurs if the expected value functions are statically or dynamically convexified [8, 25, 24]. Another approach is to approximate the original value functions with linear Benders cuts [3, 28] or Lagrangian cuts [31] without convexifying the problem. In general, these cuts only yield a non-exact convex approximation of the expected value functions. Therefore, in such cases, convergence of decomposition methods is not guaranteed. As a relief, in two-stage stochastic programming linear cuts are often incorporated into branch-and-bound approaches, where convergence is guaranteed by additional branching [7, 9]. However, for multistage problems this is computationally intractable.

Still, two strategies have been proposed recently on how linear cuts can be used to solve stochastic MILPs to arbitrary precision. The first one is to use *scaled cuts* which are guaranteed to recover the convex envelope of the expected value functions [27]. This approach has only been applied to the two-stage case so far. The second one is to use stochastic dual dynamic integer programming (SDDiP) in a lifted space [31]. The SDDiP method uses special Lagrangian cuts to approximate the expected value functions of MS-MILPs. These cuts are tight, and thus ensuring convergence if all state variables are binary (or bounded integer). For general MS-MILPs it is therefore proposed to approximate the state variables using a static binary expansion [31]. Then, in the lifted binary state space, linear cuts can be used to approximate the expected value functions. In this case, however, not the original MS-MILP, but an approximation is solved. While it is possible to derive theoretical results on the approximation quality [31], it is not immediately clear how the static binary expansion should be chosen in practice.

The special Lagrangian cuts in SDDiP rely on the introduction of copy constraints, adding local copies of the state variables, which are then dualized in the Lagrangian relaxation. Even though this is barely discussed in [31], these copy constraints are accompanied by additional constraints for the new variables. Importantly, different choices of these constraints may lead to distinct cuts with different approximation quality. In this paper, we study this aspect in more theoretical detail and by that provide a new, generalized perspective on the generation of Lagrangian cuts.

Only recently, more focus has been put on deriving non-convex approximations (also called non-convex *cuts*) for the non-convex and discontinuous value functions. In [21], step functions are used instead of cutting-planes to approximate them, presuming their monotonicity. In *stochastic Lipschitz dynamic programming* (SLDP) [1] Lipschitz cuts are proposed under the assumption of Lipschitz continuity of the value functions, as well as knowledge of a Lipschitz constant. Moreover, non-convex cuts can be generated by solving augmented Lagrangian dual problems [1] instead of classical Lagrangian dual problems. However, the approach in [1] requires a strong recourse assumption, namely the complete continuous recourse.

In [29], the authors propose a new class of SDDP-type algorithms for solving multistage stochastic mixed-integer nonlinear programs with *non-Lipschitzian* value functions. In particular, the paper proposes a new cut generation framework using *generalized conjugacy* with *regularization*, which is guaranteed to obtain a global optimum without the assumption of complete recourse. This significantly generalizes SDDP, SDDiP, and SLDP. A complete oracle complexity analysis is also achieved in the paper. In [30], SDDP-type algorithms are extended to multistage distributionally robust convex optimization and a new type of SDDP algorithm that adaptively chooses the forward or backward direction at each node is proposed with complete oracle complexity analysis.

An alternative approach to obtain non-convex cuts is to use the binary approximation idea from [31] in a dynamic and temporary fashion, paired with *Lipschitz regularization*. This is one of the key ingredients of the non-convex nested Benders decomposition (NC-NBD) method proposed in [13], where MILP relaxations are solved iteratively in order to solve an MINLP. In the backward pass of this iteration, the state variables are temporarily lifted to a binary space, where tight Lagrangian cuts are generated as in SDDiP. These linear cuts are then projected back to the original state space, which yields a tight non-convex approximation of the value functions. Under some strong technical assumption, it is shown that these approximations are Lipschitz continuous. In order to improve the approximation quality, the binary approximation precision is iteratively refined if required. We generalize these results to the stochastic case in this paper and explore the regularization in more detail, which allows us to drop the technical assumption taken in [13].

Whereas regularization is particularly helpful for deriving non-convex, but Lipschitz continuous approximations of non-Lipschitzian value functions, it may also be useful when generating linear cuts. It naturally ensures feasibility of the subproblems, so that there is no need for an additional recourse assumption. Also, as shown in [12, 29], under mild conditions, regularization of an MS-MILP ensures exact solution of the original MS-MILP. Despite these amenities, there exists no study yet focusing on the effects and the theoretical backbone of using Lipschitz regularization in order to derive linear Lagrangian cuts. We close this scientific gap in this work.

1.2 Contribution

We investigate in detail the effects of copy constraints and Lipschitz regularization on the generation and properties of linear and non-convex cuts for the value functions of MS-MILPs, which are constructed based on Lagrangian duality. This allows for a more profound understanding of how these concepts may be utilized in multistage stochastic programming. Our work can be considered as complementary to recent work on Lagrangian cuts [31], on conjugacy cuts and regularization [30, 29], on augmented Lagrangian duality [12] and on Lagrangian-based non-convex cuts [13].

With respect to **linear Lagrangian cuts** we make the following key contributions.

1. In Sect. 3, we thoroughly explore the role of copy constraints when deriving Lagrangian cuts. More precisely, we show that accompanying the new copy vari-

ables with different types of constraints leads to cuts with different approximation characteristics. This way, we provide a new theoretical perspective on Lagrangian relaxation and cuts in general. In particular, we show that the well-known tightness result for Lagrangian cuts in SDDiP [31] actually relies on the accompanying constraints more than on the introduction of copy constraints itself.

2. In Sect. 4, we consider the case of Lipschitz regularization and generalize our previous results to this case. We prove a deep relation between norm-bounded Lagrangian dual problems and primal convexifications of the regularized subproblems in MS-MILPs, in the sense that they yield the same optimal value if dual norms are used in both cases. While such relation is known for non-regularized problems and unbounded Lagrangian duals, we are not aware of any literature covering this result for the regularized case.

3. We use this result to show that the obtained Lagrangian cuts are tight for the closed convex envelope of the regularized value functions. This also clarifies which kind of cuts are constructed if (artificial) multiplier bounds are introduced in Lagrangian dual problems in practice. Furthermore, for the 1-norm penalty function, tightness for the true regularized value functions can be achieved as long as all state variables are binary. This generalizes the tightness result from SDDiP [31] to the regularized case.

With respect to **non-convex cuts** we make the following key contributions.

4. We significantly extend the idea from the NC-NBD method by Füllner and Rebennack [13] to compute linear Lagrangian cuts in a lifted binary state space and to project them back to the original state space to obtain non-convex approximations of the value functions. First, we generalize it from the deterministic to the stochastic case. Additionally, we show that by using appropriate weighted norms in the Lipschitz regularization and in the Lagrangian dual, Lipschitz continuity of the obtained non-convex cuts is ensured. This is crucial to guarantee convergence of NC-NBD. In doing that, we show that the technical Assumption (A4) in [13] can be dropped, and thus close an open theoretical question from [13].

We underline all our results by providing illustrative examples.

1.3 Structure

This paper is structured as follows. In Sect. 2 we formally introduce the MS-MILP and state some basic concepts and assumptions. In Sect. 3 we discuss classical Lagrangian duality and cuts, but with special focus on the impact of copy constraints. In Sect. 4 we introduce Lipschitz regularization in a formal way, and then derive our main result on norm-bounded Lagrangian duality and the associated Lagrangian cuts. In Sect. 5 we enhance the cut generation idea from [13] to the stochastic case and show how Lipschitz continuity of the non-convex cuts can be ensured.

2 Problem formulation

We consider MS-MILPs with a finite number $T \in \mathbb{N}$ of stages, where some of the problem data are uncertain and evolve according to a known stochastic process $\xi := (\xi_1, \dots, \xi_T)$

with deterministic ξ_1. We assume that the random data vectors $\xi_t, t = 1, \ldots, T$, are discrete and finite, and thus the uncertainty can be modeled by a finite scenario tree.

Let $\mathcal{T} = (\mathcal{N}, \mathcal{E})$ denote such a tree with a set of nodes $\mathcal{N}$ and a set of edges $\mathcal{E}$. For each node $n \in \mathcal{N}$, the unique ancestor node is denoted by $a(n)$ and the set of child nodes is denoted by $\mathcal{C}(n)$. The probability for some node n is $p_n > 0$ and assumed to be known. The transition probabilities between adjacent nodes $n, m \in \mathcal{N}$ (*i.e.*, edges $(n, m) \in \mathcal{E}$) can then be determined as $p_{nm} := \frac{p_m}{p_n}$. For the root node r, we assume $a(r) = \emptyset$ and $p_r = 1$. We define $\overline{\mathcal{N}} := \mathcal{N} \setminus \{r\}$ to address the set of nodes without the root node and $\widetilde{\mathcal{N}}$ to address the set of nodes without the leaf nodes.

2.1 Dynamic programming equations

The MS-MILP can be expressed recursively by its dynamic programming equations. We obtain

$$\begin{aligned} v^* := \min_{x_r, y_r} \quad & f_r(x_r, y_r) + \mathcal{Q}_{\mathcal{C}(r)}(x_r) \\ \text{s.t.} \quad & (x_r, y_r) \in \mathcal{F}_r(x_{a(r)}) \end{aligned} \tag{1}$$

with $x_{a(r)} = 0$, and v^* is the optimal value of the MS-MILP. Let $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$. For all $n \in \widetilde{\mathcal{N}}$, the *expected value function* $\mathcal{Q}_{\mathcal{C}(n)}(\cdot) : \mathbb{R}^{d_{a(n)}} \to \overline{\mathbb{R}}$ is defined by

$$\mathcal{Q}_{\mathcal{C}(n)}(x_n) := \sum_{m \in \mathcal{C}(n)} p_{nm} Q_m(x_n), \tag{2}$$

with the *value function* $Q_n(\cdot) : \mathbb{R}^{d_{a(n)}} \to \overline{\mathbb{R}}$ defined by

$$\begin{aligned} Q_n(x_{a(n)}) := \min_{x_n, y_n} \quad & f_n(x_n, y_n) + \mathcal{Q}_{\mathcal{C}(n)}(x_n) \\ \text{s.t.} \quad & (x_n, y_n) \in \mathcal{F}_n(x_{a(n)}) \end{aligned} \tag{3}$$

for all $n \in \overline{\mathcal{N}}$. For the leaf nodes $n \in \mathcal{N} \setminus \widetilde{\mathcal{N}}$, we set $\mathcal{Q}_{\mathcal{C}(n)}(x_n) \equiv 0$. Moreover, we set $Q_n(x_{a(n)}) = +\infty$ if $\mathcal{F}_n(x_{a(n)}) = \emptyset$, and denote by

$$\operatorname{dom}(Q_n) := \left\{ x_{a(n)} \in \mathbb{R}^{d_{a(n)}} \ : \ Q_n(x_{a(n)}) < +\infty \right\}$$

the *effective domain* of $Q_n(\cdot)$. The same applies to $\mathcal{Q}_{\mathcal{C}(n)}(\cdot)$.

For each node $n \in \mathcal{N}$, we distinguish state variables $x_n \in \mathbb{R}^{d_n}$, which enter the child nodes' subproblems, and local variables $y_n \in \mathbb{R}^{\tilde{d}_n}$. Furthermore, $f_n(\cdot)$ denotes the linear objective function and $\mathcal{F}_n(x_{a(n)})$ denotes the feasible set which depends on the state $x_{a(n)}$ and is defined by

$$\mathcal{F}_n(x_{a(n)}) := \left\{ (x_n, y_n) \in \mathbb{R}^{d_n} \times \mathbb{R}^{\tilde{d}_n} \ : \ \begin{array}{l} x_n \in X_n, \ y_n \in Y_n, \\ A_n x_{a(n)} + B_n x_n + C_n y_n \geq b_n \end{array} \right\}. \tag{4}$$

Here, A_n, B_n, C_n, b_n denote appropriately defined data matrices and vectors. The sets X_n and Y_n are intersections of polyhedral sets $\bar{X}_n, \bar{Y}_n$ (*e.g.*, modeling non-negativity constraints) and possible integrality constraints. In the following, we also refer to X_n as the *state space*.

Remark 2.1. *We should emphasize that regarding $Q_n(\cdot)$ as a function on $\mathbb{R}^{d_{a(n)}}$ is not necessarily standard in stochastic programming. Often it is (implicitly) assumed to be* *defined only on the domain $X_{a(n)}$. However, from our view, allowing $Q_n(\cdot)$ to be defined on $\mathbb{R}^{d_{a(n)}}$ with extended real values proves beneficial when we discuss the impact of copy constraints later on. Moreover, as the co-domain of $Q_n(\cdot)$ is $\overline{\mathbb{R}}$, $Q_n(\cdot)$ should be more rigorously defined as the infimum of the objective values in problem* (3)*. However, below we take assumptions under which the minimization problem is bounded and finite infima are always attained. Therefore, we stick to the* min *operator in* (3) *with the additional definition of $Q_n(x_{a(n)}) = +\infty$ given that $\mathcal{F}_n(x_{a(n)}) = \emptyset$. This approach is also chosen for all other value functions later on.*

For the remainder of this article, we make some basic assumptions.

Assumption 1. *The following conditions are satisfied by* (1)-(4)*:*

(A1) For all $n \in \mathcal{N}$, the sets X_n and Y_n are compact.

(A2) For all $n \in \mathcal{N}$, all coefficients in A_n, B_n, C_n, b_n, f_n, $\bar{X}_n$ and $\bar{Y}_n$ are rational.

(A3) The MS-MILP has a feasible solution for each scenario, i.e., there exists some $(x_n, y_n)_{n\in\mathcal{N}}$ such that $(x_n, y_n) \in \mathcal{F}_n(x_{a(n)})$ for all $n \in \mathcal{N}$.

Note that the boundedness in (A1) immediately implies that $F_n(x_{a(n)})$ is bounded for all $x_{a(n)} \in \mathbb{R}^{d_{a(n)}}$ and $n \in \mathcal{N}$. By (A1) and [18, Theorem 2.1], it follows that the subproblems (1) and (3) are either infeasible or attain a finite infimum.

We obtain the following well-known properties for the value functions. For completeness, we provide a proof in Appendix A.

Lemma 2.2. *Under Assumption 1, for all $n \in \overline{\mathcal{N}}$, the functions $Q_n(\cdot)$, and for all $n \in \mathcal{N}$, the functions $\mathcal{Q}_{\mathcal{C}(n)}(\cdot)$ are proper, lsc (lower semicontinuous), and piecewise polyhedral with finitely many pieces. Moreover,* $\operatorname{dom}(Q_n)$ *is closed.*

Applying the same reasoning to the root node, we conclude that v^* is finite.

2.2 Closed convex envelopes

A key challenge in decomposition methods for MS-MILPs is that the (expected) value functions are not guaranteed to be continuous or convex. Therefore, approximations of the value functions based on linear cutting-planes may at best yield their closed convex envelopes. To deal with this concept, we denote by $\operatorname{conv}(S)$ the convex hull of some set $S \subseteq \mathbb{R}^d$. For a function $f : S \to \mathbb{R}$, its *closed convex envelope* $\overline{\operatorname{co}}(f) : \operatorname{conv}(S) \to \mathbb{R}$ (also called *convex closure*) is defined as the pointwise supremum of all affine functions majorized by f on S [4].

In our setting, for all $n \in \bar{N}$ the value functions $Q_n(\cdot)$ are defined on $\mathbb{R}^{d_{a(n)}}$. Hence, $\overline{\operatorname{co}}(Q_n)(\cdot)$ is the pointwise supremum of all affine functions defined on $\mathbb{R}^{d_{a(n)}}$ and majorized by $Q_n(\cdot)$ on $\mathbb{R}^{d_{a(n)}}$. With $Q_n(x_{a(n)}) = +\infty$ for all $x_{a(n)} \notin \operatorname{dom}(Q_n)$, the crucial part is $\overline{\operatorname{co}}(Q_n)(x_{a(n)}) \leq Q_n(x_{a(n)})$ for all $x_{a(n)} \in \operatorname{dom}(Q_n)$.

It is well-known that the closed convex envelope $\overline{\operatorname{co}}(Q_n)(\cdot)$ is equivalent to the biconjugate $(Q_n)^{**}(\cdot)$ of $Q_n(\cdot)$ in this setting. For a formal definition of biconjugate functions and a general introduction to the conjugacy theory we refer to [4].

Lemma 2.3. *Under Assumption 1, for all $n \in \bar{N}$ and all $x_{a(n)} \in \mathbb{R}^{d_{a(n)}}$*

$$\overline{\operatorname{co}}(Q_n)(x_{a(n)}) = (Q_n)^{**}(x_{a(n)}).$$

We provide a proof in Appendix B.

3 The role of copy constraints in Lagrangian cuts

We revisit Lagrangian duality and its usage to generate cuts. We focus specifically on the role that copy constraints and constraints accompanying them have on the obtained results. This yields a generalization of some known results from the literature.

3.1 Copy constraints and a family of value functions

We follow the SDDiP approach [31] and introduce local copies z_n together with copy constraints $x_{a(n)} = z_n$ to all subproblems (3). Crucially, in addition, we also impose the accompanying constraints $z_n \in Z_{a(n)}$ on z_n to restrict their potential values, given some set $Z_{a(n)} \subseteq \mathbb{R}^{d_{a(n)}}$. This yields a family of subproblems and value functions

$$
\begin{aligned}
Q_{n|Z}(x_{a(n)}) := Q_n(x_{a(n)}; Z_{a(n)}) := \quad \min_{x_n, y_n, z_n} \quad & f_n(x_n, y_n) + \mathcal{Q}_{\mathcal{C}(n)|Z}(x_n) \\
\text{s.t.} \quad & (x_n, x_n, z_n) \in \mathcal{F}_n \\
& z_n = x_{a(n)} \\
& z_n \in Z_{a(n)}
\end{aligned} \tag{5}
$$

for different choices of $Z_{a(n)}$. We use the short notation $Q_{n|Z}(x_{a(n)})$ instead of $Q_n(x_{a(n)}; Z_{a(n)})$ whenever the particular choice of $Z_{a(n)}$ is negligible. Additionally,

$$
\mathcal{F}_n := \Big\{ (x_n, y_n, z_n) \in \mathbb{R}^{d_n} \times \mathbb{R}^{\tilde{d}_n} \times \mathbb{R}^{d_{a(n)}} \ : \ (x_n, y_n) \in \mathcal{F}_n(z_n) \Big\}
$$

and $\mathcal{Q}_{\mathcal{C}(n)|Z}(x_n) := \sum_{m \in \mathcal{C}(n)} p_{nm} Q_{m|Z}(x_n)$.

We can interpret $Q_{n|Z}(\cdot)$ as $Q_n(\cdot)$ *restricted to* $Z_{a(n)}$ due to the copy constraint and $z_{a(n)} \in Z_{a(n)}$ immediately inducing infeasibility.

Lemma 3.1. *Let $X_{a(n)} \subseteq Z_{a(n)} \subseteq \mathbb{R}^{d_{a(n)}}$. Then, it follows*

$$
Q_{n|Z}(x_{a(n)}) = \begin{cases} Q_n(x_{a(n)}), & \text{for all } x_{a(n)} \in Z_{a(n)}, \\ +\infty & \text{for all } x_{a(n)} \notin Z_{a(n)}, \end{cases}
$$

and thus $Q_n(x_{a(n)}) \leq Q_{n|Z}(x_{a(n)})$ for all $x_{a(n)} \in \mathbb{R}^{d_{a(n)}}$.

This means that the additional constraints on z_n provide a natural way to restrict the effective domain of $Q_n(\cdot)$ such that $\operatorname{dom}(Q_{n|Z}) \subseteq Z_{a(n)}$.

In order to not exclude feasible points, we should choose $Z_{a(n)} \supseteq X_{a(n)}$. Given that, the actual choice of $Z_{a(n)}$ may seem of minor importance at first glance. However, it turns out that it has an important effect on the considered closed convex envelope, and by that on the quality of the obtained Lagrangian cuts (recall that $\overline{\text{co}}(Q_{n|Z})(\cdot)$ underestimates $Q_{n|Z}(\cdot)$ on $\operatorname{dom}(Q_{n|Z})$, so the convex envelope changes with $Z_{a(n)}$). As we shall see, choosing $Z_{a(n)}$ appropriately is also the main secret behind the tightness results for Lagrangian cuts in SDDiP [31] or Benders dual decomposition [22].

We discuss different choices for $Z_{a(n)}$:

- $\boldsymbol{Z_{a(n)} = X_{a(n)}}$. This is the most intuitive choice, as it yields $\operatorname{dom}(Q_{n|Z}) \subseteq X_{a(n)}$, *i.e.*, we restrict $Q_{n|Z}(\cdot)$ to the actual state space. This choice also yields the best possible polyhedral underestimators of $Q_n(\cdot)$ on $X_{a(n)}$. It is considered in [9, 22, 31] for instance.

- $\boldsymbol{Z_{a(n)} = \mathbf{conv}(X_{a(n)})}$. This choice may yield worse approximations of $Q_n(\cdot)$ on $X_{a(n)}$, as it leads to valid under-approximators on the larger set $\mathrm{conv}(X_{a(n)})$. However, this property may also be exploited on purpose [13], as we discuss in detail in Sect. 5. This choice is also considered in the original SDDiP work [31], but without further explanation.

- $\boldsymbol{Z_{a(n)} = \bar{X}_{a(n)}}$. In this case, $Z_{a(n)}$ is the LP relaxation of $X_{a(n)}$, so no additional integer variables appear in the reformulated subproblem (5).

- $\boldsymbol{Z_{a(n)} = \mathbb{R}^{d_{a(n)}}}$. This choice leads to the same Lagrangian cuts as if no copy constraints are introduced at all, but instead the original coupling constraints $A_n x_{a(n)} + B_n x_n + C_n y_n \geq b_n$ are dualized in the Lagrangian relaxation.

We take the following technical assumption.

Assumption 2. *The set $Z_{a(n)}$ is closed and either satisfies $Z_{a(n)} = \mathbb{R}^{d_{a(n)}}$ or is rational MILP-representable.*

In the remainder of this paper, we use two recurring examples to illustrate our results. We start with the first one to highlight the differences in the convex envelopes for different choices of $Z_{a(n)}$.

Example 3.2. *Consider the value function*

$$Q(x) = \min \left\{ y_1 + y_2 \ : \ 2y_1 + y_2 \geq 3x,\ 0 \leq y_1 \leq 2,\ 0 \leq y_2 \leq 3,\ y_1 \in \mathbb{Z} \right\} \tag{6}$$

with state space $X = \{0, 1\}$ [31, Example 2]. We introduce the local variable z, the copy constraint $z = x$, and constraint $z \in Z$. Depending on the choice of Z, the effective domain of $Q_{|Z}(\cdot)$ defined in (5) changes. If we set $Z = X$, then $\mathrm{dom}(Q_{|Z}) = \{0, 1\}$. If we set $Z = \mathrm{conv}(X)$, then $\mathrm{dom}(Q_{|Z}) = [0, 1]$. And if we set $Z = \mathbb{R}$ (or do not introduce copy constraints at all), then $\mathrm{dom}(Q_{|Z}) = [0, 2]$. Since $\overline{\mathrm{co}}(Q_{|Z})(\cdot)$ underestimates $Q_{|Z}(\cdot)$ on $\mathrm{dom}(Q_{|Z})$, the approximation quality on the actual state space X may vary. This is illustrated in Fig. 1, where the approximation at $x = 1$ is highlighted by dots. Clearly, $\overline{\mathrm{co}}(Q_{|Z})(\cdot)$ is tight for $Q(\cdot)$ at $x = 1$ in the first two cases, but not in the third one.

For the second example, the value function is not only non-convex, but also discontinuous.

Example 3.3. *Consider the value function*

$$\begin{aligned} Q_{|Z}(x) = \min_{y,z} \Big\{ y_1 - \frac{3}{4}y_2 + \frac{3}{4}y_3 + \frac{9}{4}y_4 \ : \ & \frac{5}{4}y_1 - y_2 + \frac{1}{2}y_3 + \frac{1}{3}y_4 = z, \\ & y_1, y_2, y_3, y_4 \geq 0,\ y_1, y_2 \in \mathbb{Z},\ z = x,\ z \in [0, 2] \Big\} \end{aligned} \tag{7}$$

with continuous state space $X = [0, 2]$, where we already introduced copy constraints with $Z = X$. The discontinuous value function and its closed convex envelope are illustrated in Fig. 2.

3.2 The approximate subproblem

In Benders-like decomposition methods such as SDDP or SDDiP, the functions $\mathcal{Q}_{\mathcal{C}(n)}(\cdot)$ are iteratively approximated by cutting-planes, so called *optimality cuts*. For this reason, in the recursion (5), $\mathcal{Q}_{\mathcal{C}(n)|Z}(\cdot)$ are replaced by polyhedral outer approximations

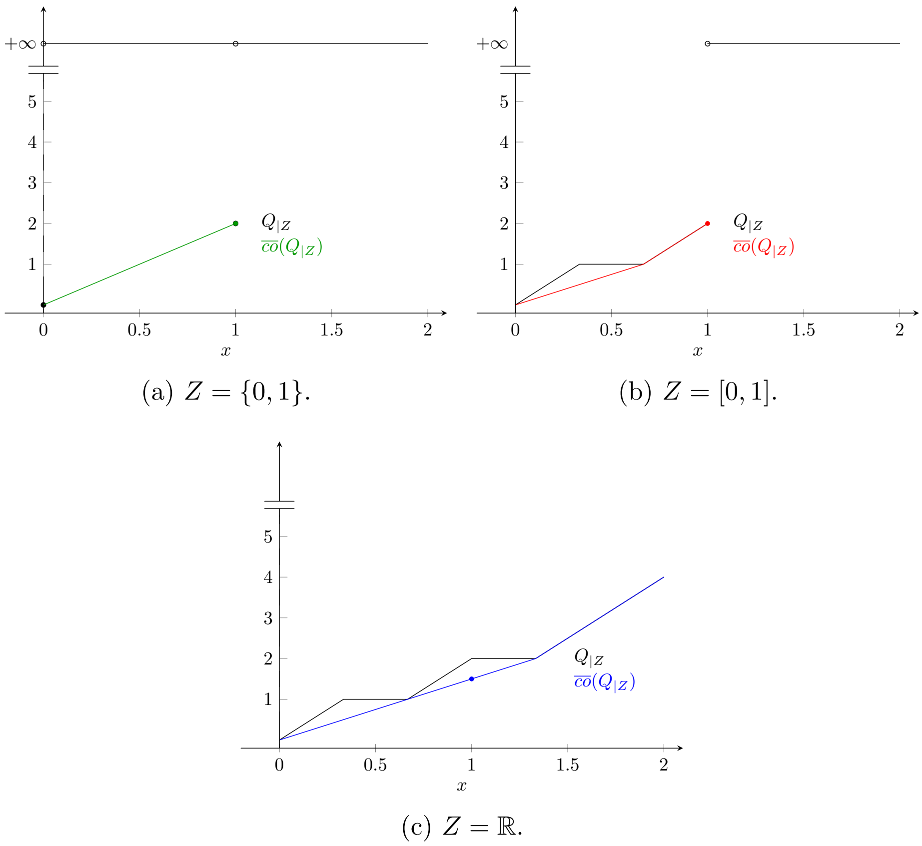


Figure 1: $Q_{|Z}(\cdot)$ on $[0, 2]$ and $\overline{\text{co}}(Q_{|Z})(\cdot)$ on $\text{dom}(Q_{|Z})$ for different choices of Z in Example 3.2.

$\mathfrak{Q}^i_{\mathcal{C}(n)|Z}(\cdot)$, which are then iteratively updated over the iterations i. Due to the nonconvex character of $\mathcal{Q}_{\mathcal{C}(n)|Z}(\cdot)$, for MS-MILPs these approximations will not be tight in general, and thus not convergence-guaranteeing. In this paper, we solely focus on the generation step for optimality cuts when solving MS-MILPs. For more details on SDDiP and its algorithmic procedure as a whole, we refer to [14, 31].

Whereas Assumption 1 guarantees the existence of a feasible solution for MS-MILP, for some $x_{a(n)}$, the subproblems may become infeasible. In such a case, also *feasibility cuts* are required, which iteratively approximate $\text{dom}(Q_{n|Z})$. Usually, this requirement is avoided by taking an appropriate recourse assumption, such as:

Assumption 3 (Relatively complete recourse)**.** *For all $n \in \mathcal{N}$, for all $x_{a(n)}$ feasible at node $a(n)$, there exist (z_n, x_n, y_n) satisfying the constraints in* (5)*.*

For the cut generation in some iteration i we consider some node $n \in \overline{\mathcal{N}}$, some incumbent $x^i_{a(n)}$ obtained from the ancestor node $a(n)$, and an (by traversing the nodes in backward direction) already updated approximation $\mathfrak{Q}^{i+1}_{\mathcal{C}(n)|Z}(\cdot)$ of $\mathcal{Q}_{\mathcal{C}(n)|Z}(\cdot)$. By a partial epigraph reformulation, we shift $\mathfrak{Q}^{i+1}_{\mathcal{C}(n)|Z}(\cdot)$ to the constraints. Then, the sub-

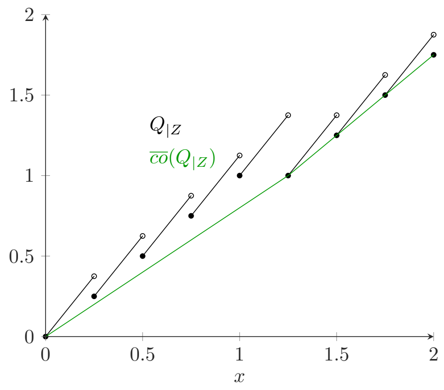


Figure 2: $Q_{|Z}(\cdot)$ and $\overline{\text{co}}(Q_{|Z})(\cdot)$ for Example 3.3.

problems can be expressed as

$$\begin{aligned} \underline{Q}_{n|Z}^{i+1}(x_{a(n)}^i) := \min_{x_n,y_n,z_n,\theta_{\mathcal{C}(n)}} \quad & f_n(x_n,y_n) + \theta_{\mathcal{C}(n)} \\ \text{s.t.} \quad & (x_n,y_n,z_n,\theta_{\mathcal{C}(n)}) \in \mathcal{M}_{n|Z}^{i+1} \\ & z_n = x_{a(n)}^i, \end{aligned} \tag{8}$$

where we define

$$\mathcal{M}_{n|Z}^{i+1} := \Big\{(x_n,y_n,z_n,\theta_{\mathcal{C}(n)}) \;:\; z_n \in Z_{a(n)},\ (x_n,y_n,z_n) \in \mathcal{F}_n,\ \theta_{\mathcal{C}(n)} \geq \mathfrak{Q}_{\mathcal{C}(n)}^{i+1}(x_n)\Big\}.$$

The polyhedral outer approximation $\mathfrak{Q}_{\mathcal{C}(n)|Z}^{i+1}(\cdot)$ is defined as the pointwise maximum of all linear cuts generated so far. To avoid unboundedness of subproblems (8), we initialize each $\mathfrak{Q}_{\mathcal{C}(n)|Z}^{0}(\cdot)$ with a valid lower bound $\underline{\theta}_{\mathcal{C}(n)} > -\infty$. We refer to $\underline{Q}_{n|Z}^{i+1}(\cdot)$ as the *approximate value function* given some set $Z_{a(n)}$.

In the same vein as Assumption 2, we impose another requirement, which is always satisfied in practice where cut coefficients are computed numerically.

Assumption 4. *For all $n \in \overline{\mathcal{N}}$ and all iterations i, all linear cuts defining the polyhedral set $\mathfrak{Q}_{\mathcal{C}(n)|Z}^{i+1}(x_n)$ are defined by rational coefficients.*

Our assumptions on rationality of coefficients and MILP-representability yield the following important result, which goes back to [18].

Lemma 3.4 (Theorem 11.13 in [10])**.** *Under Assumptions 1, 2, 4, the set* $\text{conv}(\mathcal{M}_{n|Z}^{i+1})$ *is a closed rational polyhedron, and the recession cones of* $\text{conv}(\mathcal{M}_{n|Z}^{i+1})$ *and* $\widehat{\mathcal{M}}_{n|Z}^{i+1}$ *coincide, where the latter set denotes the continuous relaxation of* $\mathcal{M}_{n|Z}^{i+1}$.

Importantly, however, neither $\mathcal{M}_{n|Z}^{i+1}$ nor $\text{conv}(\mathcal{M}_{n|Z}^{i+1})$ has to be bounded.

By additionally exploiting Assumptions 2 and 4 and that $\theta_{\mathcal{C}(n)}$ is bounded from below, we obtain similar properties to Lemma 2.2 for the functions $\underline{Q}_{n|Z}^{i+1}(\cdot)$:

Lemma 3.5. *Let $n \in \overline{\mathcal{N}}$. If Assumptions 1, 2, 4 are satisfied, then $\underline{Q}_{n|Z}^{i+1}(\cdot)$ is proper, lsc, and piecewise polyhedral with finitely many pieces. Moreover,* $\text{dom}(\underline{Q}_{n|Z}^{i+1}) = \text{dom}(Q_{n|Z})$, *and thus closed.*

3.3 The Lagrangian dual

In order to derive linear Lagrangian cuts, we consider a *Lagrangian relaxation* in which the copy constraints in subproblem (8) are relaxed. For a given vector of dual multipliers $\pi_n \in \mathbb{R}^{d_{a(n)}}$ for the copy constraints, this yields the problem

$$\begin{aligned} \mathcal{L}_{n|Z}^{i+1}(\pi_n) := \min_{x_n, y_n, z_n, \theta_{\mathcal{C}_n}} \quad & f_n(x_n, y_n) + \theta_{\mathcal{C}(n)} - \pi_n^\top z_n \\ \text{s.t.} \quad & (x_n, y_n, z_n, \theta_{\mathcal{C}(n)}) \in \mathcal{M}_{n|Z}^{i+1}, \end{aligned} \tag{9}$$

where we omit the constant $\pi_n^\top x_{a(n)}^i$ in the objective. Optimizing the associated function $\mathcal{L}_{n|Z}^{i+1}(\pi_n)$ over the dual multipliers π_n yields the *Lagrangian dual problem*

$$\underline{Q}_{n|Z}^{D,i+1}(x_{a(n)}^i) := \max_{\pi_n} \quad \mathcal{L}_{n|Z}^{i+1}(\pi_n) + \pi_n^\top x_{a(n)}^i. \tag{10}$$

As (9) is a relaxation of the primal subproblem (8), it yields a lower bound for $\underline{Q}_{n|Z}^{i+1}(\cdot)$ at $x_{a(n)}^i$. Solving the dual problem (10) can be interpreted as finding the tightest Lagrangian relaxation for (8), and thus the tightest such lower bound.

Remark 3.6. *In the light of Remark 2.1, note that unless $Z_{a(n)}$ is bounded, the dual function $\mathcal{L}_{n|Z}^{i+1}(\pi_n)$ may yield the trivial lower bound $-\infty$ for some π_n. However, based on Assumption 1 and $\theta_{\mathcal{C}(n)}$ being bounded from below, it is finite for $\pi_n = 0$. Therefore, $\underline{Q}_{n|Z}^{D,i+1}(\cdot)$ is proper. Moreover, we shall see that $\underline{Q}_{n|Z}^{D,i+1}(x_{a(n)})$ is guaranteed to be finite-valued for all $x_{a(n)} \in \operatorname{conv}(\operatorname{dom}(\underline{Q}_{n|Z}^{i+1}))$.*

A well-known result on Lagrangian relaxation for MILPs is that under some assumptions the optimal value of the dual (10) is the same as that of the following convexification of the primal subproblem (8)

$$\begin{aligned} \underline{Q}_{n|Z}^{C,i+1}(x_{a(n)}^i) := \min_{x_n, y_n, z_n, \theta_{\mathcal{C}_n}} \quad & f_n(x_n, y_n) + \theta_{\mathcal{C}(n)} \\ \text{s.t.} \quad & (x_n, y_n, z_n, \theta_{\mathcal{C}(n)}) \in \operatorname{conv}(\mathcal{M}_{n|Z}^{i+1}) \\ & z_n = x_{a(n)}^i. \end{aligned} \tag{11}$$

Here, the part of the constraints which is not relaxed in (9) is convexified, while the copy constraints keep their original form. We first derive an auxiliary result.

Lemma 3.7. *Under Assumptions 1, 2, 4, the function $\underline{Q}_{n|Z}^{C,i+1}(\cdot)$ is proper, lsc and convex with $\operatorname{dom}(\underline{Q}_{n|Z}^{C,i+1}) = \operatorname{conv}(\operatorname{dom}(\underline{Q}_{n|Z}^{i+1}))$. Moreover, on its effective domain it is piecewise linear.*

We provide a proof in Appendix C.

Based on this lemma, the equivalence between the primal convexification (11) and the dual (10) is given below.

Theorem 3.8 (Theorem 1 in [15])**.** *Under Assumptions 1, 2, 4, the Lagrangian dual* (10) *and the primal convexified problem* (11) *satisfy*

$$\underline{Q}_{n|Z}^{D,i+1}(x_{a(n)}^i) = \underline{Q}_{n|Z}^{C,i+1}(x_{a(n)}^i)$$

for all $x_{a(n)}^i \in \operatorname{conv}(\operatorname{dom}(\underline{Q}_{n|Z}^{i+1}))$.

The main idea behind this result is that problems (10) and (11) are LP duals of each other. Note that we even have $\underline{Q}_{n|Z}^{D,i+1}(x_{a(n)}^i) = \underline{Q}_{n|Z}^{C,i+1}(x_{a(n)}^i)$ for all $x_{a(n)}^i \in \mathbb{R}^{d_{a(n)}}$, since both functions are bounded from below and may only take the value $+\infty$ if non-finite.

Closely related is another interesting, and well-known, property of the Lagrangian dual, which relates to the closed convex envelope of the approximate value function. We provide a self-contained proof in Appendix D.

Theorem 3.9. *Under Assumption 1, the Lagrangian dual* (10) *satisfies*

$$\underline{Q}_{n|Z}^{D,i+1}(x_{a(n)}^i) = \overline{\text{co}}(\underline{Q}_{n|Z}^{i+1})(x_{a(n)}^i)$$

for all $x_{a(n)}^i \in \text{conv}(\text{dom}(\underline{Q}_{n|Z}^{i+1}))$.

3.4 Lagrangian cuts

We now focus on the generation of Lagrangian cuts at points $x_{a(n)}^i \in \text{conv}(\text{dom}(\underline{Q}_{n|Z}^{i+1}))$ using problem (10), as introduced in [31]. We first define these cuts formally.

Definition 3.10 (Lagrangian cut)**.** *For all* $n \in \overline{\mathcal{N}}$, *a Lagrangian cut is given by*

$$\theta_n \geq \mathcal{L}_{n|Z}^{i+1}(\pi_n^i) + (\pi_n^i)^\top x_{a(n)},$$

where π_n^i *denotes optimal dual multipliers in* (10) *for node* n *and some given* $x_{a(n)}^i \in \text{conv}(\text{dom}(\underline{Q}_{n|Z}^{i+1}))$.

Under relatively complete recourse, *i.e.*, Assumption 3, within the decomposition method it is ensured that for all iterations i and all nodes $n \in \mathcal{N}$, the condition $x_{a(n)}^i \in \text{dom}(\underline{Q}_{n|Z}^{i+1}) \subseteq \text{conv}(\text{dom}(\underline{Q}_{n|Z}^{i+1}))$ is satisfied, so there never occurs an $x_{a(n)}^i$ for which no cut can be computed due to infeasibility.

In general, the Lagrangian cuts have the following important properties:

Theorem 3.11. *Under Assumptions 1, the Lagrangian cuts defined in 3.10 are*

(a) valid lower approximations of $Q_{n|Z}(\cdot)$ *for all* $x_{a(n)} \in Z_{a(n)}$ *(and thus of* $Q_n(\cdot)$ *for all* $x_{a(n)} \in \mathbb{R}^{d_{a(n)}}$*),*

(b) tight for $\overline{\text{co}}(\underline{Q}_{n|Z}^{i+1})(\cdot)$ *at* $x_{a(n)}^i$*,*

(c) finite, i.e., only finitely many different cuts can be generated, if the dual multipliers π_n^i *are dual basic solutions.*

Proof. Properties (a) and (c) are proven in [31, Theorem 3], but without formalizing the dependence on $Z_{a(n)}$. The statement in brackets then follows from Lemma 3.1. Tightness of Lagrangian cuts is directly proven for $\underline{Q}_{n|Z}^{i+1}(\cdot)$ in [31]. Property (b) is about $\overline{\text{co}}(\underline{Q}_{n|Z}^{i+1})(\cdot)$ instead. But it directly follows from Theorem 3.9 and Definition 3.10. □

Property (a) implies that for $Z_{a(n)} = \text{conv}(X_{a(n)})$ Lagrangian cuts underestimate $Q_{n|Z}(\cdot)$ not only on $X_{a(n)}$, but also on the larger set $\text{conv}(X_{a(n)})$. Note that property (a) even holds if the Lagrangian dual (10) is not solved to optimality, *i.e.*, if suboptimal dual multipliers are used in Definition 3.10.

To approximate $\mathcal{Q}_{\mathcal{C}(n)}(\cdot)$ instead of each $Q_m(\cdot), m \in \mathcal{C}(n)$, separately, we construct an aggregated cut from those in Definition 3.10. We can then express $\mathfrak{Q}^{i+1}_{\mathcal{C}(n)|Z}(\cdot)$ by

$$\mathfrak{Q}^{i+1}_{\mathcal{C}(n)|Z}(x_n) := \min \Big\{ \theta_{\mathcal{C}(n)} \in \mathbb{R} \ : \ \theta_{\mathcal{C}(n)} \geq \sum_{m \in \mathcal{C}(n)} p_{nm} \big(\mathcal{L}^{i+1}_{m|Z}(\pi^r_m) + (\pi^r_m)^\top x_n \big) \ \forall r = 1, \dots, i+1 \Big\}.$$

Using Theorem 3.11, the validity of $\mathfrak{Q}^{i+1}_{\mathcal{C}(n)|Z}(\cdot)$ follows immediately.

Corollary 3.12. *Under Assumption 1, $\mathfrak{Q}^{i+1}_{\mathcal{C}(n)|Z}(\cdot)$ is a valid lower approximation of $\mathcal{Q}_{\mathcal{C}(n)|Z}(\cdot)$ for all $x_n \in Z_n$ (and thus of $\mathcal{Q}_{\mathcal{C}(n)}(\cdot)$ for all $x_n \in \mathbb{R}^{d_n}$).*

Importantly, we cannot directly generalize the tightness result from Theorem 3.11 to $\overline{\text{co}}(\mathbb{E}[Q_n])$, since in general $\mathbb{E}[\overline{\text{co}}(Q_n)] \neq \overline{\text{co}}(\mathbb{E}[Q_n])$ [27].

3.5 The case of tight Lagrangian cuts

In SDDiP [31], it is assumed that all state variables are binary, *i.e.*, $X_{a(n)} = \{0,1\}^{d_{a(n)}}$, and thus $\text{conv}(X_{a(n)}) = [0,1]^{d_{a(n)}}$. This assumption has two key effects, which ensure the almost sure finite convergence of SDDiP to an optimal policy of the considered MS-MILP. First, $X_{a(n)}$ is finite. Second, the Lagrangian cuts from Definition 3.10 are not only tight for $\overline{\text{co}}(\underline{Q}^{i+1}_{n|Z})(\cdot)$ at $x^i_{a(n)}$, but in fact for $\underline{Q}^{i+1}_{n|Z}(\cdot)$. This tightness is directly proven in [31, Theorem 3]. However, our previous analyses allow for a different perspective on this result, which is briefly mentioned in [31], but not used in the proof: It holds because $\overline{\text{co}}(\underline{Q}^{i+1}_{n|Z})(\cdot)$ and $\underline{Q}^{i+1}_{n|Z}(\cdot)$ coincide at $x^i_{a(n)}$. The main reason for this is that $X_{a(n)}$ is contained in the extreme points of $Z_{a(n)}$. Therefore, this tightness result crucially depends on the choice of $Z_{a(n)}$. This perspective also allows to extend the SDDiP tightness result to more general cases, as also touched upon in Remark 1 in [29].

Theorem 3.13. *Under Assumptions 1, 2, 4, for any iteration i and any node $n \in \overline{\mathcal{N}}$, let $Z_{a(n)}$ be bounded and let $X_{a(n)}$ be contained in the set of extreme points of $Z_{a(n)}$. Then, for all $x_{a(n)} \in X_{a(n)} \cap \text{dom}(\underline{Q}^{i+1}_{n|Z})$, we have*

$$\overline{\text{co}}(\underline{Q}^{i+1}_{n|Z})(x_{a(n)}) = \underline{Q}^{i+1}_{n|Z}(x_{a(n)}).$$

We present a proof in Appendix E.

Combined with the properties of the Lagrangian cuts from Theorem 3.11, Theorem 3.13 directly implies a tightness result for the Lagrangian cuts.

Corollary 3.14 (Theorem 3 in [31])**.** *Under Assumptions 1, 2, 4, for any iteration i and any node $n \in \overline{\mathcal{N}}$, let $Z_{a(n)}$ be bounded and let $X_{a(n)}$ be contained in the set of extreme points of $Z_{a(n)}$. Then the cuts defined in Definition 3.10 are* tight *for $\underline{Q}^{i+1}_{n|Z}(\cdot)$ at $x^i_{a(n)}$.*

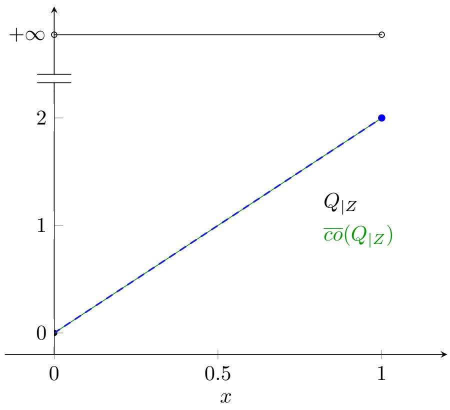
Remark 3.15. *Assume that $X_{a(n)} = \{0,1\}^{d_{a(n)}}$ and that $Z_{a(n)} = X_{a(n)}$ or $Z_{a(n)} = \text{conv}(X_{a(n)})$ as in SDDiP [31] and that we have relatively complete recourse (Assumption 3). Then, the conditions of Theorem 3.13 are satisfied, and for all feasible $x_{a(n)}$ the known tightness result of SDDiP follows.*

We highlight this result using the illustrative problems (6) and (7).

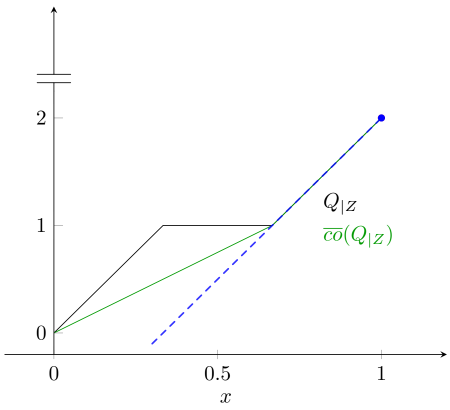
Example 3.16. *Consider the problem* (6) *with* $Z = X = \{0, 1\}$*. Solving the dual* (10) *yields the cut* $\theta \geq 2x$*. This cut underestimates* $Q_{|Z}(\cdot)$ *on* $\{0, 1\}$ *and is tight for* $\overline{\text{co}}(Q_{|Z})(\cdot)$ *at* $x = 1$*, see Fig. 3a. It is even tight for* $Q_{|Z}(\cdot)$ *at this point.*

If we choose $Z = \text{conv}(X) = [0, 1]$ *instead, we obtain the cut* $\theta \geq -1 + 3x$ *by solving* (10)*. This cut underestimates* $Q_{|Z}(\cdot)$ *on* $[0, 1]$ *and is tight for* $\overline{\text{co}}(Q_{|Z})(\cdot)$ *at* $x = 1$*, see Fig. 3b. Again, we observe tightness for* $Q_{|Z}(\cdot)$*.*

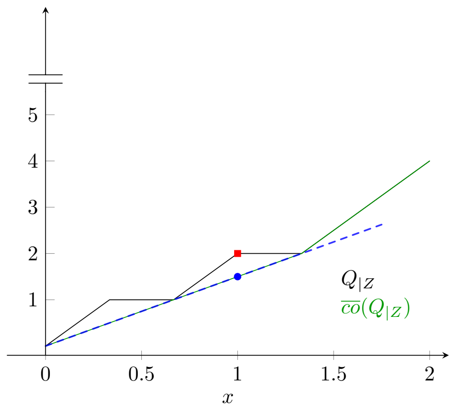
In contrast, as illustrated by Fig. 3c by the gap between the red square and the blue dot, Theorem 3.13 is not guaranteed to hold if we choose $Z = \mathbb{R}$ *or, equivalently, do not introduce copy constraint in the subproblems.*



(a) $Z = \{0, 1\}$.



(b) $Z = [0, 1]$.



(c) $Z = \mathbb{R}$.

Figure 3: Lagrangian cuts for $Q_{|Z}(\cdot)$ and different choices of Z in Example 3.16.

Example 3.17. *Consider the problem* (7) *and the incumbent* $x = \frac{6}{5}$*. Recall that* $X = Z = [0, 2]$*, so the extreme point condition in Corollary 3.14 is not satisfied. Solving the Lagrangian dual* (10) *yields the cut* $\theta \geq \frac{4}{5}x$*. As Fig. 4 shows, this cut is tight for* $\overline{\text{co}}(Q_{|Z})(\cdot)$ *at* $x = \frac{6}{5}$ *(blue dot), but not for* $Q_{|Z}(\cdot)$ *(red square).*

4 Lipschitz regularization and Lagrangian duality

In this section, we address the generation of *linear* Lagrangian cuts for $\mathcal{Q}_{\mathcal{C}(n)}(\cdot)$ when a Lipschitz regularization of the original MS-MILP is considered. We have shown in the

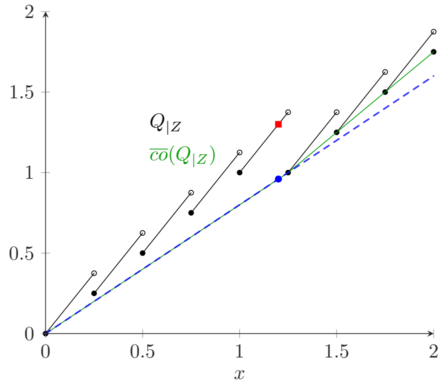


Figure 4: Lagrangian cut for $Q_{|Z}(\cdot)$ in Example 3.17.

previous section that such cuts can be generated even when no regularization is applied. In fact, regularization is particularly relevant for generating *non-convex* approximations of $\mathcal{Q}_{\mathcal{C}(n)}(\cdot)$, which we consider in Sect. 5. However, it may still be applied in cases where linear cuts are generated, for instance, to ensure feasibility of the subproblems. Moreover, the results for this case prove relevant to derive our results in Sect. 5. Therefore, we address the case of linear cuts first.

4.1 Applying a Lipschitz regularization

First, we formally introduce the considered Lipschitz regularization.

Definition 4.1 (Regularization). *For any $n \in \overline{\mathcal{N}}$, let $\sigma_n > 0$ and fix some norm $\|\cdot\|$. Then we call*

$$\begin{aligned} Q^R_{n|Z}(x_{a(n)}; \sigma_n \|\cdot\|) := \min_{x_n, y_n, z_n} \quad & f_n(x_n, y_n) + \sigma_n \|x_{a(n)} - z_n\| + \mathcal{Q}^R_{\mathcal{C}(n)|Z}(x_n; \sigma_{\mathcal{C}(n)}\|\cdot\|) \\ s.t. \quad & (z_n, x_n, y_n) \in \mathcal{F}_n \\ & z_n \in Z_{a(n)} \end{aligned} \tag{12}$$

the regularized subproblem *or the* regularized value function *for node n given some set $Z_{a(n)}$, respectively. The* regularized expected value function *is defined by*

$$\mathcal{Q}^R_{\mathcal{C}(n)|Z}(x_n; \sigma_{\mathcal{C}(n)}\|\cdot\|) := \sum_{m \in \mathcal{C}(n)} p_{nm} Q^R_{m|Z}(x_n; \sigma_m \|\cdot\|).$$

By writing $\sigma_{\mathcal{C}(n)}$, we indicate that $\mathcal{Q}^R_{\mathcal{C}(n)|Z}(\cdot; \sigma_{\mathcal{C}(n)}\|\cdot\|)$ depends on σ_m for all $m \in \mathcal{C}(n)$. For the root node, where no regularization is required, we obtain

$$\begin{aligned} v^R := \min_{x_r, y_r} \quad & f_r(x_r, y_r) + \mathcal{Q}^R_{\mathcal{C}(r)|Z}(x_r; \sigma_r \|\cdot\|) \\ s.t. \quad & (x_r, y_r) \in \mathcal{F}_r(x_{a(r)}). \end{aligned} \tag{13}$$

Remark 4.2. *The regularized problem defined by the recursion* (12)-(13) *can be interpreted as applying a special* inf*-convolution $f_n \square (\sigma_n \|\cdot\|)$, called* Lipschitz regularization *or* Pasch-Hausdorff envelope *[2], to the objective function of the original MS-MILP, see also [30, 29].*

This regularization comes with two main advantages. First, it naturally ensures feasibility of the considered subproblems, even if we take no recourse assumption for the original subproblems.

Lemma 4.3. *Under Assumption 1, problem* (12) *is feasible for all* $x_{a(n)} \in \mathbb{R}^{d_{a(n)}}$ *for any choice of* $Z_{a(n)}$, i.e., $\mathrm{dom}(Q^R_{n|Z}; \sigma_n\|\cdot\|) = \mathbb{R}^{d_{a(n)}}$.

In particular, we do not require the recourse assumption stated in Assumption 3.

Second, using a Lipschitz regularization ensures that the considered value functions are σ_n-Lipschitz continuous instead of only lsc. While related results have been shown before, see [29], our assumptions differ a bit, so we provide a self-contained proof in Appendix F.

Lemma 4.4. *Under Assumptions 1, 2, for all* $n \in \overline{\mathcal{N}}$, $Q^R_{n|Z}(\cdot; \sigma_n\|\cdot\|)$ *underestimates* $Q_{n|Z}(\cdot)$ *and is proper and* σ_n*-Lipschitz continuous on* $\mathbb{R}^{d_{a(n)}}$.

Considering regularized MS-MILPs comes at the price that we do not necessarily solve the original MS-MILP any longer. In general, $v^R \leq v^*$ [29, Proposition 2]. However, equality can be imposed by using a sufficiently large σ_n for all nodes $n \in \overline{\mathcal{N}}$, as shown in [12, Theorem 5]:

Lemma 4.5. *There exist finite* $\bar{\sigma}_n > 0$ *for all* $n \in \overline{\mathcal{N}}$ *such that given* $\sigma_n \geq \bar{\sigma}_n$*, for all* $n \in \mathcal{N}$*, the penalty reformulation in* (12) *is exact, i.e., any optimal solution* $(x_n, y_n, z_n)_{n \in \mathcal{N}}$ *of the regularized MS-MILP* (12)-(13) *satisfies* $z_n = x_{a(n)}$ *for all* $n \in \mathcal{N}$.

Hence, for sufficiently large, but finite $\sigma_n > 0$, we have $v^R = v^*$. This result also implies that for any optimal solution $(x^*_n, y^*_n)_{n \in \mathcal{N}}$ of the original MS-MILP (1)-(3) we have $Q^R_{n|Z}(x^*_{a(n)}; \sigma_n\|\cdot\|) = Q_{n|Z}(x^*_{a(n)})$ [29, Lemma 1].

Example 4.6. *Consider the problem* (7)*. We use the absolute value* $|\cdot|$ *as penalty function in* (12)*. The regularized value functions* $Q^R_{|Z}(\cdot; \sigma|\cdot|)$ *are depicted in Fig. 5 for different values of* σ*. It is visible that all of them underestimate* $Q_{|Z}(\cdot)$ *for all* $x \in [0, 2]$*, that all of them are Lipschitz continuous and that they are monotonically increasing in* $\sigma > 0$*. Assume that the optimal first-stage solution is* $x^* = 1$*. Then an exact penalization is achieved for any* $\sigma \geq 1$.

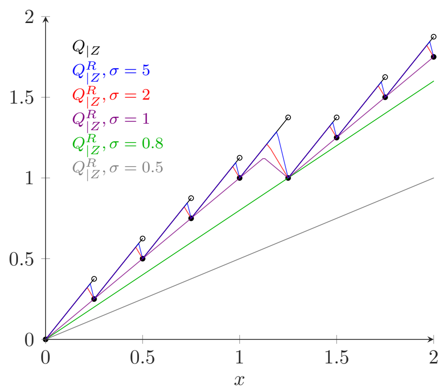


Figure 5: Functions $Q^R_{|Z}(\cdot; \sigma|\cdot|)$ in Example 4.6 for different σ and $Z = [0, 2]$.

Another effect is that using an arbitrary norm, the regularized subproblems (12) are no longer MILPs, but MINLPs. This is unfavorable from a computational perspective. However, at least for the (weighted) ℓ^1-norm or ℓ^∞-norm, an equivalent MILP reformulation can be achieved, so we do not leave the class of MILP subproblems [1].

Lemma 4.7. *If the norm* $\|\cdot\|$ *used in* (12) *is the (weighted)* ℓ^1*-norm or* ℓ^∞*-norm, the problem remains MILP-representable.*

4.2 Special regularized value functions

We introduce different variations of regularized subproblems and value functions, and some basic properties which we require in the next few subsections.

- Similarly to (8) in the non-regularized case, we define *approximate regularized value functions* for each $n \in \overline{\mathcal{N}}$ as

$$\begin{aligned} \underline{Q}_{n|Z}^{R;i+1}(x_{a(n)}^i;\sigma_n\|\cdot\|) := \min_{x_n,y_n,z_n,\theta_{\mathcal{C}(n)}} \quad & f_n(x_n,y_n) + \theta_{\mathcal{C}(n)} + \sigma_n\|x_{a(n)}^i - z_n\| \\ \text{s.t.} \quad & (x_n,y_n,z_n,\theta_{\mathcal{C}(n)}) \in \mathcal{M}_{n|Z}^{i+1}. \end{aligned} \tag{14}$$

  We also define $\underline{\mathcal{Q}}_{\mathcal{C}(n)}^{R;i+1}(x_n;\sigma_{\mathcal{C}(n)}\|\cdot\|) := \sum_{m\in\mathcal{C}(n)} p_{nm}\underline{Q}_m^{R;i+1}(x_n;\sigma_m\|\cdot\|)$ as the *expected approximate regularized value function* for all $x_n \in \mathbb{R}^{d_n}$.

- Similarly to (11) in the non-regularized case, for each $n \in \overline{\mathcal{N}}$ we consider the *convexified regularized value function*

$$\begin{aligned} \underline{Q}_{n|Z}^{CR;i+1}(x_{a(n)}^i;\sigma_n\|\cdot\|) := \min_{x_n,y_n,z_n,\theta_{\mathcal{C}(n)}} \quad & f_n(x_n,y_n) + \theta_{\mathcal{C}(n)} + \sigma_n\|x_{a(n)}^i - z_n\| \\ \text{s.t.} \quad & (x_n,y_n,z_n,\theta_{\mathcal{C}(n)}) \in \operatorname{conv}(\mathcal{M}_{n|Z}^{i+1}). \end{aligned} \tag{15}$$

- We denote the *closed convex envelope* of the approximate regularized value function by $\overline{\operatorname{co}}(\underline{Q}_{n|Z}^{R;i+1};\sigma_n\|\cdot\|)(\cdot)$. It underestimates $\underline{Q}_{n|Z}^{R;i+1}(\cdot;\sigma_n\|\cdot\|)$ on $\mathbb{R}^{d_{a(n)}}$.

The following properties are relevant in the next subsections.

Lemma 4.8. *Under Assumptions 1, 2, 4, given some arbitrary norm* $\|\cdot\|$ *and some* $\sigma_n > 0$*, for all* $n \in \overline{\mathcal{N}}$*, on* $\mathbb{R}^{d_{a(n)}}$ *the function*

(a) $\underline{Q}_{n|Z}^{R;i+1}(\cdot;\sigma_n\|\cdot\|)$ *is finite-valued and* σ_n*-Lipschitz continuous,*

(b) $\underline{Q}_{n|Z}^{CR;i+1}(\cdot;\sigma_n\|\cdot\|)$ *is finite-valued, convex and* σ_n*-Lipschitz continuous,*

(c) $\overline{\operatorname{co}}(\underline{Q}_{n|Z}^{R;i+1};\sigma_n\|\cdot\|)(\cdot)$ *is finite-valued and convex,*

(d) $\underline{Q}_{n|Z}^{CR;i+1}(\cdot;\sigma_n\|\cdot\|)$ *is equivalent to* $(\underline{Q}_{n|Z}^{CR;i+1};\sigma_n\|\cdot\|)^{**}(\cdot)$.

Moreover, we need the following auxiliary result.

Lemma 4.9. *For all* $x_{a(n)} \in \mathbb{R}^{d_{a(n)}}$ *we have*

$$(\underline{Q}_{n|Z}^{R;i+1};\sigma_n\|\cdot\|)^{**}(x_{a(n)}) = (\underline{Q}_{n|Z}^{CR;i+1};\sigma_n\|\cdot\|)^{**}(x_{a(n)}).$$

Lemma 4.8 and Lemma 4.9 are proven in Appendix G and Appendix H, respectively.

4.3 A primal convexification result

In the remainder of Sect. 4, we focus on the generation of linear Lagrangian cuts in the context of regularized subproblems and value functions (12).

Recall that in the non-regularized case (see Sect. 3.4), these cuts are generated based on a relaxation of the copy constraints. This approach cannot be applied in the regularized case, as the copy constraints are already relaxed and penalized in the regularized subproblems (12). However, we show that still cuts with similar properties can be obtained by considering specific *bounded* Lagrangian dual problems.

As a first key step, we introduce a primal convexification result for bounded Lagrangian dual problems. More precisely, we show that the convexified regularized problem (15) is closely related to the bounded Lagrangian dual problem

$$\underline{Q}_{n|Z}^{DR;i+1}(x_{a(n)}^i;\sigma_n\|\cdot\|) := \begin{array}{rl} \max\limits_{\pi_n} & \mathcal{L}_{n|Z}^{i+1}(\pi_n) + \pi_n^\top x_{a(n)}^i \\ \text{s.t.} & \|\pi_n\|_* \leq \sigma_n, \end{array} \tag{16}$$

where $\|\cdot\|_*$ denotes the dual norm to the norm $\|\cdot\|$ used in the regularized subproblem.

Theorem 4.10. *Under Assumptions 1, 2, and 4, given some arbitrary norm $\|\cdot\|$ and some $\sigma_n > 0$, the bounded Lagrangian dual* (16) *satisfies*

$$\underline{Q}_{n|Z}^{DR;i+1}(x_{a(n)}^i;\sigma_n\|\cdot\|) = \underline{Q}_{n|Z}^{CR;i+1}(x_{a(n)}^i;\sigma_n\|\cdot\|).$$

Proof. In this proof, we use sup and inf operators to be rigorous with regard to suprema and infima being attained. For notational simplicity, we set $\lambda_n := (x_n, y_n, \theta_{\mathcal{C}(n)})$ and then define $c_n^\top \lambda_n := f_n(x_n, y_n) + \theta_{\mathcal{C}(n)}$ with an appropriate coefficient vector c_n. Using this notation, we obtain

$$\begin{aligned} \underline{Q}_{n|Z}^{DR;i+1}(x_{a(n)}^i;\sigma_n\|\cdot\|) &= \sup_{\|\pi_n\|_*\leq\sigma_n} \mathcal{L}_{n|Z}^{i+1}(\pi_n) + \pi_n^\top x_{a(n)}^i \\ &= \sup_{\|\pi_n\|_*\leq\sigma_n} \inf_{(z_n,\lambda_n)\in\mathcal{M}_{n|Z}^{i+1}} \Big\{ c_n^\top \lambda_n + \pi_n^\top (x_{a(n)}^i - z_n) \Big\} \\ &= \sup_{\|\pi_n\|_*\leq\sigma_n} \inf_{(z_n,\lambda_n)\in\operatorname{conv}(\mathcal{M}_{n|Z}^{i+1})} \Big\{ c_n^\top \lambda_n + \pi_n^\top (x_{a(n)}^i - z_n) \Big\}. \end{aligned} \tag{17}$$

The last line follows since the objective of the inner problem is linear.

We now consider the dual problem where we swap the sup and inf operators. As we discuss below, strong duality holds.

$$\begin{aligned} & \inf_{(z_n,\lambda_n)\in\operatorname{conv}(\mathcal{M}_{n|Z}^{i+1})} \sup_{\|\pi_n\|_*\leq\sigma_n} \Big\{ c_n^\top \lambda_n + \pi_n^\top (x_{a(n)}^i - z_n) \Big\} \\ &= \inf_{(z_n,\lambda_n)\in\operatorname{conv}(\mathcal{M}_{n|Z}^{i+1})} \Big\{ c_n^\top \lambda_n + \sigma_n \sup_{\|\frac{\pi_n}{\sigma_n}\|_*\leq 1} \Big\{ \Big(\frac{\pi_n}{\sigma_n}\Big)^\top (x_{a(n)}^i - z_n) \Big\} \Big\} \\ &= \inf_{(z_n,\lambda_n)\in\operatorname{conv}(\mathcal{M}_{n|Z}^{i+1})} \Big\{ c_n^\top \lambda_n + \sigma_n \|x_{a(n)}^i - z_n\| \Big\}. \end{aligned} \tag{18}$$

Here, we used the definition of dual norms. As is shown in Appendix G, in problem (15) always a finite infimum is attained, so in the last line we may replace the infimum with a minimum. Substituting c_n and λ_n with their definitions, we obtain exactly the definition of the function $\underline{Q}_{n|Z}^{CR;i+1}(x_{a(n)}^i;\sigma_n\|\cdot\|)$.

It remains to be shown that we have strong duality between problems (17) and (18). First, according to Lemma 3.4, the set $\text{conv}(\mathcal{M}_{n|Z}^{i+1})$ is a closed polyhedron. By its relaxation property and Assumption 1, it is also non-empty. Therefore, both sets $\text{conv}(\mathcal{M}_{n|Z}^{i+1})$ and $\{\pi_n \in \mathbb{R}^{d_{a(n)}} \ : \ \|\pi_n\|_* \leq \sigma_n\}$ are closed convex and non-empty, with the latter also bounded. Moreover, the objective is linear in π_n for fixed (z_n, λ_n) and vice versa. Hence, we can apply the minimax theorem from [23, Corollary 37.3.2] to infer strong duality. □

Remark 4.11. *While the duality between multiplier bounds and a Lipschitz regularization (based on the duality of norms) is known in the literature on multistage stochastic programming, see for instance [17, Proposition 4.2], [30, Proposition 5], [29, Lemma 2], to our knowledge the result with respect to convexification in Theorem 4.10 has never been discussed and explicitly proven before. Also in the literature on convex analysis,* e.g. *[2, 4, 23], we are not aware of any mentioning of the above result, as the discussion is usually limited to results for* unbounded *Lagrangian dual problems, or exactness of general augmented Lagrangian dual problems with respect to the original primal problem, not its Lipschitz regularization or convexification. A related result to Theorem 4.10 is presented in [29, Proposition 4] for the* true *regularized value function $Q_{n|Z}^R(\cdot; \sigma_n\|\cdot\|)$ instead of its closed convex envelope, given that $Z_{a(n)} = X_{a(n)}$ is compact and that $Q_{n|Z}(\cdot)$ is convex. Similar results are proven in [30] in a distributionally robust setting. Another related, but different result is the primal characterization for general augmented Lagrangian dual problems in [12, Theorem 1].*

4.4 Convex envelopes from bounded Lagrangian duals

An important question is whether the primal convexification result from Theorem 4.10 can also be linked to the closed convex envelope $\overline{\text{co}}(\underline{Q}_{n|Z}^{R;i+1}; \sigma_n\|\cdot\|)(\cdot)$, as it was the case for the non-regularized case (see Theorems 3.8, 3.9). We prove this now.

Corollary 4.12. *Consider the regularized subproblem* (14) *and the corresponding bounded Lagrangian dual* (16) *given some arbitrary norm* $\|\cdot\|$ *and some* $\sigma_n > 0$. *Under Assumptions 1, 2, 4, for all* $x_{a(n)}^i \in \mathbb{R}^{d_{a(n)}}$ *we have*

$$\underline{Q}_{n|Z}^{DR;i+1}(x_{a(n)}^i; \sigma_n\|\cdot\|) = \overline{\text{co}}(\underline{Q}_{n|Z}^{R;i+1}; \sigma_n\|\cdot\|)(x_{a(n)}^i).$$

Proof. From Lemma 4.8 (d) and Lemma 4.9 we can conclude that for all $x_{a(n)} \in \mathbb{R}^{d_{a(n)}}$ we have

$$\underline{Q}_{n|Z}^{CR;i+1}(x_{a(n)}; \sigma_n\|\cdot\|) = (\underline{Q}_{n|Z}^{CR;i+1}; \sigma_n\|\cdot\|)^{**}(x_{a(n)}) = (\underline{Q}_{n|Z}^{R;i+1}; \sigma_n\|\cdot\|)^{**}(x_{a(n)}). \qquad (19)$$

Furthermore, from Lemma 4.8 (c) we know that $\overline{\text{co}}(\underline{Q}_{n|Z}^{R;i+1}; \sigma_n\|\cdot\|)(\cdot)$ is proper. From Proposition 1.6.1 (d) in [4] it then follows that $\overline{\text{co}}(\underline{Q}_{n|Z}^{R;i+1}; \sigma_n\|\cdot\|)(x_{a(n)}) = (\underline{Q}_{n|Z}^{R;i+1}; \sigma_n\|\cdot\|)^{**}(x_{a(n)})$ for all $x_{a(n)} \in \mathbb{R}^{d_{a(n)}}$. Hence, with (19) it follows that

$$\underline{Q}_{n|Z}^{CR;i+1}(x_{a(n)}^i; \sigma_n\|\cdot\|) = \overline{\text{co}}(\underline{Q}_{n|Z}^{R;i+1}; \sigma_n\|\cdot\|)(x_{a(n)}^i)$$

for all $x_{a(n)}^i \in \mathbb{R}^{d_{a(n)}}$. The primal convexification result in Theorem 4.10 yields

$$\underline{Q}_{n|Z}^{DR;i+1}(x_{a(n)}^i; \sigma_n\|\cdot\|) = \overline{\text{co}}(\underline{Q}_{n|Z}^{R;i+1}; \sigma_n\|\cdot\|)(x_{a(n)}^i).$$

□

Corollary 4.12 directly implies the following result for Lagrangian cuts that are computed using the bounded Lagrangian dual problem.

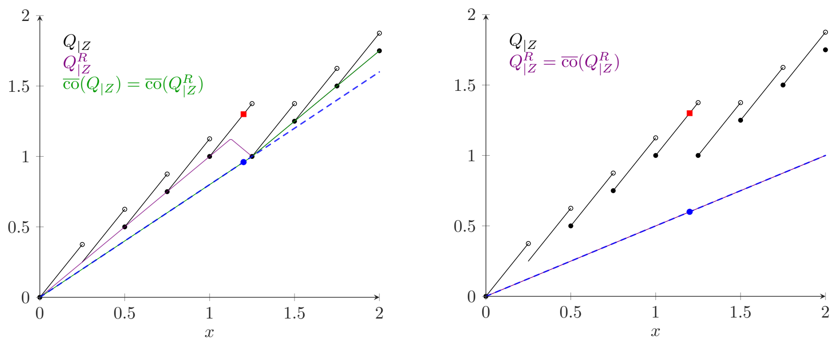
Corollary 4.13. *Consider Lagrangian cuts from Definition 3.10, but with optimal multipliers* π_n^i *for the bounded dual problem* (16)*. These cuts are*

(a) valid lower approximations of $Q_{n|Z}^R(\cdot;\sigma_n\|\cdot\|)$*, and thus also for* $Q_{n|Z}(\cdot)$ *for all* $x_{a(n)} \in Z_{a(n)}$*,*

(b) tight for $\overline{\text{co}}(\underline{Q}_{n|Z}^{R;i+1};\sigma_n\|\cdot\|)(\cdot)$ *at* $x_{a(n)}^i$*.*

We illustrate Corollary 4.13 with an example.

Example 4.14. *Consider the problem* (7) *with incumbent* $x = \frac{6}{5}$*. We use absolute value* $|\cdot|$ *as the penalty function in* (12) *and to bound the dual multipliers in* (16)*. Solving the Lagrangian dual problem for* $\sigma \geq \frac{4}{5}$*, we obtain the cut* $\theta \geq \frac{4}{5}x$*. For* $\sigma < \frac{4}{5}$*, in contrast, the resulting cut is* $\theta \geq \sigma x$*. Fig. 6 displays these cuts (blue broken lines) for* $\sigma = 1$ *and* $\sigma = \frac{1}{2}$*. As we can see, in both cases, the cut is tight for* $\overline{\text{co}}(Q_{|Z}^R;\sigma|\cdot|)(\cdot)$ *at* $x = \frac{6}{5}$ *(blue dots).*



(a) $\sigma = 1$. (b) $\sigma = \frac{1}{2}$.

Figure 6: (Regularized) value functions and Lagrangian cuts for Example 4.14.

4.5 The case of Lipschitz continuous value functions

Given that for all $n \in \overline{\mathcal{N}}$, the value functions $Q_{n|Z}(\cdot)$ are already Lipschitz continuous, *e.g.*, because a strong recourse assumption like complete continuous recourse is satisfied [1, 31], a close relation between the regularized and the non-regularized problem holds.

Lemma 4.15. *For all* $n \in \overline{\mathcal{N}}$*, let* $Q_{n|Z}(\cdot)$ *be Lipschitz continuous on* $\text{dom}(Q_{n|Z})$ *with respect to some norm* $\|\cdot\|$ *with Lipschitz constant* α_n*. Then for* $\sigma_n \geq \alpha_n$ *we have*

$$Q_{n|Z}^R(x_{a(n)};\sigma_n\|\cdot\|) = Q_{n|Z}(x_{a(n)})$$

for all $x_{a(n)} \in \text{dom}(Q_{n|Z})$*.*

For leaf nodes $n \in \mathcal{N}$, this result follows immediately from [2, Corollary 12.18], considering that the regularized value functions are the Pasch-Hausdorff envelopes of the non-regularized ones. For other nodes in $\mathcal{N}$ it can then be shown inductively.

As already noticed in [30], this result shows that computationally regularization may even prove beneficial if the original value functions are already guaranteed to be Lipschitz continuous. First, the Lagrangian dual problem can be bounded. Second, cuts with larger Lipschitz constant than the value function can be excluded from its approximation.

Additionally, Lemma 4.15 has a helpful implication that we use in the next section when discussing the SDDiP setting again. It can be used to show that for sufficiently large σ_n, the lower convex envelopes of the regularized and non-regularized approximate value functions do coincide.

4.6 The case of tight Lagrangian cuts

Similar to Sect. 3.5 for the non-regularized case, we consider cases where tightness for function $\underline{Q}_{n|Z}^{R;i+1}(\cdot;\sigma_n\|\cdot\|)$, and not only its closed convex envelope $\overline{\text{co}}(\underline{Q}_{n|Z}^{i+1})(\cdot)$, can be obtained using Lagrangian cuts. Importantly, the extreme point argument used in the proof of Theorem 3.13 is no longer valid in the regularized setting. The two functions $\overline{\text{co}}(\underline{Q}_{n|Z}^{R;i+1};\sigma_n\|\cdot\|)(\cdot)$ and $\underline{Q}_{n|Z}^{R;i+1}(\cdot;\sigma_n\|\cdot\|)$ share the effective domain $\mathbb{R}^{d_{a(n)}}$, and hence, it is not clear whether they coincide for all $x_{a(n)} \in X_{a(n)}$. Nonetheless, under some assumptions, the intended tightness result can be established.

Case 1: Using sufficiently large σ_n. Given the results from the previous subsections and the exact penalization result for Lipschitz regularization from Lemma 4.5, it can be shown that tightness of Lagrangian cuts for $\underline{Q}_{n|Z}^{R;i+1}(\cdot;\sigma_n\|\cdot\|)$ can also be guaranteed by choosing a sufficiently large, but finite regularization parameters $\sigma_n > 0$.

Case 2: Regularization with the ℓ^1-norm. Suppose the state space is binary and we use the (weighted) ℓ^1-norm for regularization. Then, we can derive the following auxiliary result, which is proven in [13, Lemma 3.8] in a slightly different form. The proof is given in Appendix I.

Lemma 4.16. *Let $X_{a(n)} = \{0,1\}^{d_{a(n)}}$ for all $n \in \mathcal{N}$ and $Z_{a(n)} = X_{a(n)}$ or $Z_{a(n)} = \text{conv}(X_{a(n)})$. Then, for any iteration i, any node $n \in \overline{\mathcal{N}}$ and any $\sigma_n > 0$, the regularized subproblem* (14) *and the bounded Lagrangian dual* (16) *for the ℓ^1-norm satisfy*

$$\underline{Q}_{n|Z}^{DR;i+1}(x_{a(n)}^i;\sigma_n\|\cdot\|_1) \geq \underline{Q}_{n|Z}^{R;i+1}(x_{a(n)}^i;\sigma_n\|\cdot\|_1). \tag{20}$$

Using this lemma, we obtain the intended tightness result.

Corollary 4.17. *The inequality in Lemma 4.16 is satisfied with equality.*

Proof. From Lemma 4.16 we have relation (20). On the other hand, from Corollary 4.12 and the definition of the closed convex envelope

$$\underline{Q}_{n|Z}^{DR;i+1}(x_{a(n)}^i;\sigma_n\|\cdot\|_1) = \overline{\text{co}}(\underline{Q}_{n|Z}^{R;i+1};\sigma_n\|\cdot\|_1)(x_{a(n)}^i) \leq \underline{Q}_{n|Z}^{R;i+1}(\cdot;\sigma_n\|\cdot\|_1).$$

□

As we already established a tightness result for the *true* value functions given a binary state space in Sect. 3.5, without the need of regularization, this result seems superfluous. However, it proves beneficial later in Sect. 5 when we deal with *non-convex* approximations of the value functions.

We illustrate Corollary 4.17 using problem (6) below.

Example 4.18. *Consider the problem* (6) *with incumbent* $x = 1$ *and* $Z = \mathrm{conv}(X) = [0, 1]$. *We use the absolute value* $|\cdot|$ *as penalty function in* (12) *and set* $\sigma = 2$. *Fig. 7 shows that the regularized value function* $Q^R_{|Z}(\cdot; \sigma|\cdot|)$ *is monotonically increasing outside of* Z, *and thus coincides with its convex envelope outside of* Z. *For this reason, both functions also coincide at extreme points of* Z, *such as* $x = 1$. *The obtained Lagrangian cut* $\theta \geq -\frac{1}{3} + 2x$ *is tight for* $Q^R_{|Z}(\cdot; \sigma|\cdot|)$ *at* $x = 1$, *in accordance with Corollary 4.17.*

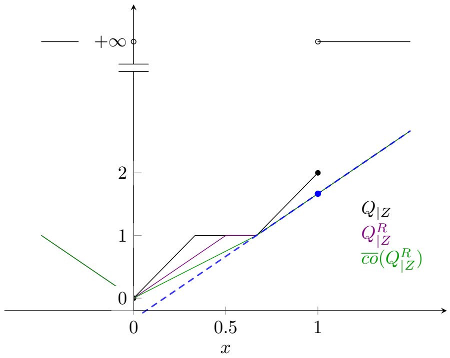


Figure 7: Lagrangian cut for the regularized value function with $\sigma = 2$ in problem (6).

In the multidimensional case for norms different than $\|\cdot\|_1$, *this is not necessarily true. We present an example for this in Example 5.15 (3) in the next section.*

5 Lipschitz regularization and non-convex cuts

As discussed before, the *linear* cuts analyzed in Sect. 3-4 are not guaranteed to be tight for the *true* value functions of MS-MILPs in general, and thus cannot guarantee convergence of decomposition methods for these problems.

In this section, we deal with the alternative approach to generate non-convex approximations $\mathfrak{Q}^{i+1}_{\mathcal{C}(n)}(\cdot)$ of $\mathcal{Q}_{\mathcal{C}(n)}(\cdot)$ in order to ensure convergence, again by exploiting Lipschitz regularization. We specifically focus on the alternative lift-and-project approach from [13], which is part of the NC-NBD method presented in the same paper. We first give a brief introduction into its main concepts and then present its theoretical backbone in a rigorous way. In particular, we close an open question on how to ensure Lipschitz continuity of the obtained non-convex approximations. This allows us to drop the technical Assumption 4 in [13]. While NC-NBD in [13] assumes a deterministic problem, we enhance its ideas to the stochastic setting here.

5.1 The lift-and-project idea

As a basis to describe the lift-and-project cut generation idea, we consider the approximate regularized value function from Sect. 4 for some node $n \in \overline{\mathcal{N}}$, some $\sigma_n > 0$, some norm $\|\cdot\|^\circ$ and a given incumbent $x^i_{a(n)}$:

$$\underline{Q}^{R;i+1}_{n|Z}(x^i_{a(n)}; \sigma_n\|\cdot\|^\circ) = \min_{x_n, y_n, z_n, \theta_{\mathcal{C}(n)}} \quad f_n(x_n, y_n) + \theta_{\mathcal{C}(n)} + \sigma_n \|x^i_{a(n)} - z_n\|^\circ \quad \text{s.t.} \quad (x_n, y_n, z_n, \theta_{\mathcal{C}(n)}) \in \mathcal{M}^{i+1}_{n|Z}. \tag{21}$$

The objective function $f_n(\cdot)$ and all but the cut constraints in (21) are still assumed linear. The only difference in (21) compared to Sect. 4 is that we now assume that the approximation $\mathfrak{Q}^{i+1}_{\mathcal{C}(n)|Z}(\cdot)$ of $\mathcal{Q}_{\mathcal{C}(n)|Z}(\cdot)$, which is contained in the set $\mathcal{M}^{i+1}_{n|Z}$, is non-convex. However, we assume that it can still be approximated by mixed-integer linear constraints, see Sect. 5.7 and [13], so $\mathcal{M}^{i+1}_{n|Z}$ has the same properties as in Sect. 4. Notation-wise, the symbol $\|\cdot\|^\circ$ is introduced to distinguish the norm used for regularization in the original state space from a second, possibly deviating norm $\|\cdot\|^\bullet$ that is used in a lifted space where cuts are generated later on.

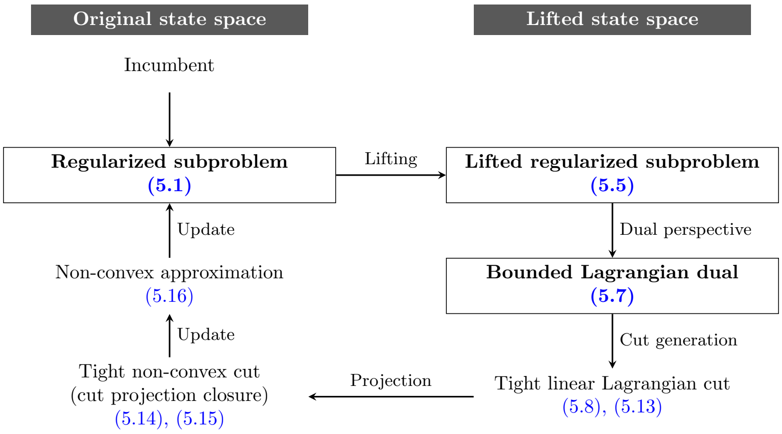


Figure 8: The lift-and-project approach used in [13].

In each iteration i, instead of directly generating Lagrangian cuts in the original state space, the subproblems (21) are first temporarily lifted to a binary state space. According to Corollary 4.17, by solving a bounded Lagrangian dual problem then *tight linear* Lagrangian cuts can be computed for the regularized value functions. However, these cuts are expressed in the lifted state space. In order to use them in the original state space, they are projected back to that space. The pointwise maximum of this projection, which we refer to as the *cut projection closure* (CPC), can be interpreted as a non-convex cut, and, as we show, it is tight for $\underline{Q}^{R;i+1}_{n|Z}(\cdot;\sigma_n\|\cdot\|^\circ)$.

Another key feature of this cut generation method is that we allow for the construction of cuts at points $x^i_{\mathcal{B},a(n)}$ differing from the current incumbents $x^i_{a(n)}$. These points are called *anchor points* in [13]. This also means that the non-convex cuts, *i.e.*, the CPC, is guaranteed to be tight for $\underline{Q}^{R;i+1}_{n|Z}(\cdot;\sigma_n\|\cdot\|^\circ)$ at $x^i_{\mathcal{B},a(n)}$ only. But as long as the distance between $x^i_{a(n)}$ and $x^i_{\mathcal{B},a(n)}$ can be controlled, we shall see that the approximation error of the non-convex cuts at $x^i_{a(n)}$ can be controlled as well.

5.2 Sufficient non-convex approximations

In order to ensure convergence when employing this lift-and-project cut generation approach in decomposition methods, the non-convex approximations $\mathfrak{Q}^{i+1}_{\mathcal{C}(n)}(\cdot)$ have to satisfy three main properties. We call an approximation satisfying them *sufficient*.

Definition 5.1. *Let $n \in \mathcal{N}$ and consider some arbitrary iteration $i \in \mathbb{N}$. Given some anchor point $x^i_{\mathcal{B},n}$, some norm $\|\cdot\|^\circ$ and some $\sigma_n > 0$ used in the regularized subproblem* (21)*, a non-convex approximation $\mathfrak{Q}^{i+1}_{\mathcal{C}(n)}(\cdot)$ is called* sufficient *if it*

(S1) is a valid under-approximation of $\mathcal{Q}_{\mathcal{C}(n)}(\cdot)$, i.e., *for all $x_n \in X_n$:*

$$\mathfrak{Q}^{i+1}_{\mathcal{C}(n)}(x_n) \leq \mathcal{Q}_{\mathcal{C}(n)}(x_n),$$

(S2) overestimates the expected approximate regularized value function at $x^i_{\mathcal{B},n}$:

$$\mathfrak{Q}^{i+1}_{\mathcal{C}(n)}(x^i_{\mathcal{B},n}) \geq \underline{\mathcal{Q}}^{R;i+1}_{\mathcal{C}(n)}(x^i_{\mathcal{B},n}; \sigma_{\mathcal{C}(n)} \|\cdot\|^\circ),$$

(S3) is Lipschitz continuous with respect to the norm $\|\cdot\|^\circ$ used in subproblem (21) with a finite Lipschitz constant (independent of i).

The reasoning behind these properties is the following: Using similar arguments as in Lemma 4.8, it can be shown that $\underline{\mathcal{Q}}^{R;i+1}_{\mathcal{C}(n)|Z}(\cdot; \sigma_{\mathcal{C}(n)} \|\cdot\|^\circ)$ is Lipschitz continuous. As $\mathfrak{Q}^{i+1}_{\mathcal{C}(n)|Z}(\cdot)$ is also Lipschitz continuous according to (S3), using property (S2), the approximation error at the incumbent x^i_n can be bounded by the Lipschitz constants and the distance between x^i_n and $x^i_{\mathcal{B},n}$ [13, Lemma 4.1]. In our case, the anchor points $x^i_{\mathcal{B},n}$ are determined using a binary approximation of x^i_n which goes along with the lifting to the binary state space, see Sect. 5.3. Thus, the distance between x^i_n and $x^i_{\mathcal{B},n}$ can be controlled by refining the associated approximation precision β_n if required. As a result, also the cut approximation error at x^i_n can be controlled, and reduced sufficiently [13, Lemma 4.2]. Together with the validity (S1), this ensures exactness and finite convergence of the decomposition method [13, Theorem 4.3]. For more details on NC-NBD and its convergence proof, we refer to [13].

In the remainder of this section, we focus on showing how sufficient non-convex approximations can be obtained in the lift-and-project framework. In [13], it is already shown how properties (S1) and (S2) can be achieved for the deterministic case, but we extend these results to the scenario tree setting. Ensuring property (S3), on the other hand, is more sophisticated. Whereas it can be easily shown that it holds for a fixed precision β_n of the binary approximation [13], we have to make sure that the Lipschitz constant does not diverge for $\beta_n \to 0$. Otherwise, the reduction in distance between x^i_n and $x^i_{\mathcal{B},n}$ may be redeemed by an increasing Lipschitz constant. In other words, we have to bound the Lipschitz constant in (S3) independently of β_n (and by that i). Instead of showing how this can be achieved, in [13] a technical assumption is taken (Assumption (A4) in [13]). We close this theoretical gap in this section. The key idea is to use a tailor-made norm $\|\cdot\|^\bullet$ for the regularization in the lifted space.

5.3 Lifting to the binary space

We lift the subproblems and value functions to a different space by temporarily applying a binary approximation of the state $x_{a(n)}$ [16]. For simplicity, we assume that all components of $x_{a(n)}$ satisfy bounded box constraints with a zero lower bound. Then, any component $x_{a(n),j} \in [0, U_{a(n),j}]$, $j = 1, \ldots, d_{a(n)}$, can be approximated by

$$x_{a(n),j} = \beta_{a(n),j} \sum_{\kappa=1}^{K_{a(n),j}} 2^{\kappa-1} \lambda_{a(n),\kappa j} + r_{a(n),j}, \tag{22}$$

with a discretization precision $\beta_{a(n),j} \in (0,1)$ if $x_{a(n),j}$ is continuous and $\beta_{a(n),j} = 1$ if it is integer. $r_{a(n),j} \in \left[-\frac{\beta_{a(n),j}}{2}, \frac{\beta_{a(n),j}}{2}\right]$ denotes the approximation error. For some vector $x_{a(n)}$, this requires $K_{a(n)} = \sum_{j=1}^{d_{a(n)}} K_{a(n),j}$ binary variables $\lambda_{a(n),\kappa j}$, with $K_{a(n),j} = \lfloor \log_2 \left(\frac{U_{a(n),j}}{\beta_{a(n),j}} \right) \rfloor + 1$.

We define a $(d_{a(n)} \times K_{a(n)})$-matrix $\mathcal{B}_{a(n)}$ containing all the coefficients of the binary approximation and collect all binary variables in one large vector $\lambda_{a(n)} \in \{0,1\}^{K_{a(n)}}$. Then, the binary expansion can be written compactly as

$$x_{a(n)} = \mathcal{B}_{a(n)}\lambda_{a(n)} + r_{a(n)}. \tag{23}$$

For some $k \in K_{a(n)}$, let $j(k)$ denote the component in the original space associated with k. Then, we define $\kappa(k) := k - \sum_{\ell=1}^{j(k)} K_{a(n),\ell}$ to access the correct κ in (22). In reverse, let $k(j,\kappa) := \sum_{\ell=1}^{j} K_{a(n),\ell} + \kappa$.

By applying (23) to a trial point $x^i_{a(n)}$ and omitting the error term, we define the anchor point as the approximation

$$x^i_{\mathcal{B},a(n)} := \mathcal{B}_{a(n)}\lambda^i_{a(n)}. \tag{24}$$

Example 5.2. *Again, we consider problem* (7) *with incumbent* $x^i = \frac{6}{5}$. *For different values of* β *or* K, *respectively, we obtain the anchor points*

$$\begin{aligned}
K = 2, \beta = \frac{2}{3}: &\qquad x^i_{\mathcal{B}} = \frac{2}{3}(2^0 \cdot 0 + 2^1 \cdot 1) = \frac{4}{3},\\
K = 3, \beta = \frac{2}{7}: &\qquad x^i_{\mathcal{B}} = \frac{2}{7}(2^0 \cdot 0 + 2^1 \cdot 0 + 2^2 \cdot 1) = \frac{8}{7},\\
K = 4, \beta = \frac{2}{15}: &\qquad x^i_{\mathcal{B}} = \frac{2}{15}(2^0 \cdot 1 + 2^1 \cdot 0 + 2^2 \cdot 0 + 2^3 \cdot 1) = \frac{6}{5}.
\end{aligned}$$

As we see, for $K = 4$ *and* $\beta = \frac{2}{15}$, *the approximation of the incumbent is exact.*

Instead of problem (21), we now consider $\underline{Q}^{R;i+1}_{n|Z}(x^i_{\mathcal{B},a(n)}; \sigma_n\|\cdot\|^{\circ})$, *i.e.*, we do not consider the subproblem at the incumbent $x^i_{a(n)}$, but at the anchor point $x^i_{\mathcal{B},a(n)}$.

Using relation (24) we can interpret $\lambda^i_{a(n),j}, j = 1,\dots,K_{a(n)}$, as the state variables in a lifted binary state space. We can thus express the function $\underline{Q}^{R;i+1}_{n|Z}(x^i_{\mathcal{B},a(n)}; \sigma_n\|\cdot\|^{\circ})$ in terms of these binary state variables. To this end, we also set

$$z_n = \mathcal{B}_{a(n)}\mathfrak{z}_n \tag{25}$$

with auxiliary variables $\mathfrak{z}_n \in [0,1]^{K_{a(n)}}$. Additionally, we define the norm $\|\lambda_n\|_{\mathcal{B}} := \|\mathcal{B}_{a(n)}\lambda_n\|$. This yields the reformulation

$$\begin{aligned}
\underline{Q}^{R;i+1}_{\mathcal{B};n|Z}(\lambda^i_{a(n)}; \sigma_n\|\cdot\|^{\circ}_{\mathcal{B}}) := \min_{x_n,y_n,z_n,\theta_{\mathcal{C}(n)},\mathfrak{z}_n} \quad & f_n(x_n,y_n) + \theta_{\mathcal{C}(n)} + \sigma_n\|(\lambda^i_{a(n)} - \mathfrak{z}_n)\|^{\circ}_{\mathcal{B}}\\
\text{s.t.} \quad & (x_n,y_n,z_n,\theta_{\mathcal{C}(n)}) \in \mathcal{M}^{i+1}_{n|Z}\\
& z_n = \mathcal{B}_{a(n)}\mathfrak{z}_n\\
& \mathfrak{z}_n \in [0,1]^{K_{a(n)}}.
\end{aligned} \tag{26}$$

of the regularized subproblem (21). It is exact in the sense that

$$\underline{Q}^{R;i+1}_{\mathcal{B};n|Z}(\lambda^i_{a(n)}; \sigma_n\|\cdot\|^{\circ}_{\mathcal{B}}) = \underline{Q}^{R;i+1}_{n|Z}(x^i_{\mathcal{B},a(n)}; \sigma_n\|\cdot\|^{\circ}). \tag{27}$$

In the same vein, we may define $Q_{\mathcal{B};n|Z}(\cdot)$ as the true value function $Q_{n|Z}(\cdot)$ expressed as a function in the lifted state space.

Remark 5.3. *In view of* (24) *and Sect. 3.1, the reformulation used in* (26) *can be interpreted as first introducing and then relaxing copy constraints for each binary variable*

$\lambda^i_{a(n),j}, j = 1, \dots, K_{a(n)}$, *separately, with an accompanying set* $\mathcal{Z}_{a(n)} = [0,1]^{K_{a(n)}}$.

Remark 5.4. *Recall our discussion on different choices of* $Z_{a(n)}$ *and their impact in Sect. 3.1. Whereas the choice of* $Z_{a(n)}$ *in the original state space is not particularly relevant in this lift-and-project setting here, the choice of bounding* $\mathfrak{z}_n$ *in the lifted space using the convex hull* $\mathcal{Z}_{a(n)} = [0,1]^{K_{a(n)}}$ *instead of* $\{0,1\}^{K_{a(n)}}$ *is crucial if some components of* $x_{a(n)}$ *are continuous. First, in contrast to (24) the reformulation (25) of* z_n *becomes exact, even for continuous variables. Second, it ensures that linear cuts generated in the lifted state space are valid on the whole set* $[0,1]^{K_{a(n)}}$. *This is inevitable in order to obtain non-convex cuts in the original state space which are valid underestimators of* $\mathcal{Q}_{\mathcal{C}(n)|Z}(\cdot)$ *even for points in* $X_{a(n)}$ *that cannot be exactly represented by the current binary approximation.*

5.4 Generating Lagrangian cuts in the lifted space

To generate Lagrangian cuts in the lifted binary state space, we follow Sect. 4 and consider a bounded Lagrangian dual problem

$$\underline{Q}^{DR;i+1}_{\mathcal{B};n|Z}(\lambda^i_{a(n)}; \sigma_n \|\cdot\|^\bullet) := \max_{\|\pi_n\|^\bullet_* \le \sigma_n} \quad \mathcal{L}^{i+1}_{\mathcal{B};n|Z}(\pi_n) + \pi_n^\top \lambda^i_{a(n)}. \tag{28}$$

The dual function $\mathcal{L}^{i+1}_{\mathcal{B};n|Z}(\cdot)$ is defined by

$$\begin{aligned} \mathcal{L}^{i+1}_{\mathcal{B};n|Z}(\pi_n) := \quad \min_{x_n, y_n, z_n, \theta_{\mathcal{C}(n)}, \mathfrak{z}_n} \quad & f_n(x_n, y_n) + \theta_{\mathcal{C}(n)} - \pi_n^\top \mathfrak{z}_n \\ \text{s.t.} \quad & (x_n, y_n, z_n, \theta_{\mathcal{C}(n)}) \in \mathcal{M}^{i+1}_{n|Z} \\ & z_n = \mathcal{B}_{a(n)} \mathfrak{z}_n \\ & \mathfrak{z}_n \in [0,1]^{K_{a(n)}}. \end{aligned}$$

Solving the dual problem (28), we obtain optimal multipliers π^i_n. We can then build the function

$$\phi_{\mathcal{B};n|Z}(\lambda_{a(n)}) := \mathcal{L}^{i+1}_{\mathcal{B};n|Z}(\pi^i_n) + (\pi^i_n)^\top \lambda_{a(n)}, \tag{29}$$

which defines a linear Lagrangian cut in the binary space $\{0,1\}^{K_{a(n)}}$.

The crucial part, and a new contribution compared to [13], is how we choose the norm $\|\cdot\|^\bullet_*$ in problem (28) to bound the dual multipliers. Let $\|\cdot\|$ be an arbitrary norm and $\|\cdot\|_*$ its dual norm. Furthermore, let W and $\widehat{W}$ be some diagonal weight matrices whose diagonal entries at row k and column k satisfy the relation $\widehat{w}_{kk} = w^{-1}_{kk}$ (for simplicity, we omit indices for W and $\widehat{W}$). Then, $\|x\|_w := \|Wx\|$ defines the *weighted* norm and $\|x\|_{w*} := \|\widehat{W}x\|_*$ defines the dual weighted norm for some vector x.

For each component $k = 1, \dots, K_{a(n)}$ in the binary state space, we now define weights

$$w_{kk} = 2^{\kappa(k)-1} \beta_{a(n),j(k)},$$

and choose $\|\cdot\|^\bullet = \|\cdot\|_w$ given some norm $\|\cdot\|$. The motivation behind this choice is to bound the dual multipliers in such a way that the effects of the binary approximation are compensated by the weights. Note that these bounds are tighter than the ones originally proposed in [13] where no weighted norms are used.

With this construction, we observe that the matrix W and the matrix $\mathcal{B}_{a(n)}$ are closely related. Both matrices contain the same non-negative entries, but W is a $(K_{a(n)} \times$

$K_{a(n)}$)-diagonal matrix, whereas $\mathcal{B}_{a(n)}$ is a $(d_{a(n)} \times K_{a(n)})$-matrix where non-negative entries corresponding to the same component j of the original state space occur in the same row. Hence, we can define a matrix G such that $\mathcal{B}_{a(n)} = GW$. This matrix contains only ones and zeros, with several ones in each row, but only a single one in each column.

For some matrix A, let $\|A\|$ be the matrix norm induced by $\|\cdot\|$. Then, the consistency of matrix norms and the inducing vector norm yields the relation

$$\begin{aligned}\|\mathcal{B}_{a(n)}(\lambda^i_{a(n)} - \mathfrak{z}_n)\| = \|GW(\lambda^i_{a(n)} - \mathfrak{z}_n)\| &\leq \|G\| \; \|W(\lambda^i_{a(n)} - \mathfrak{z}_n)\| \\ &= \|G\| \; \|\lambda^i_{a(n)} - \mathfrak{z}_n\|_w.\end{aligned} \tag{30}$$

5.5 Properties of Lagrangian cuts in the lifted space

Recall Remark 5.4. As shown in [13], by choosing $\mathfrak{z}_n \in [0,1]^{K_{a(n)}}$, the function $\phi_{\mathcal{B};n|Z}(\cdot)$ defined in (29) provides a valid underestimator for the true value function in the binary state space, but also everywhere in the original state space. This is crucial to prove property (S1) in the next section.

Lemma 5.5 (Lemma 3.7 in [13])**.** *The function $\phi_{\mathcal{B};n|Z}(\cdot)$ satisfies*

$$Q_{\mathcal{B};n|Z}(\lambda_{a(n)}) \geq \phi_{\mathcal{B};n|Z}(\lambda_{a(n)}),$$

for all $\lambda_{a(n)} \in [0,1]^{K_{a(n)}}$, and

$$Q_{n|Z}(x_{a(n)}) \geq \phi_{\mathcal{B};n|Z}(\lambda_{a(n)})$$

for all $x_{a(n)} \in \mathbb{R}^{d_{a(n)}}$ and any $\lambda_{a(n)} \in [0,1]^{K_{a(n)}}$, such that $x_{a(n)} = \mathcal{B}_{a(n)}\lambda_{a(n)}$.

Next, we use the results from Sect. 4.6 to obtain some overestimation results with respect to property (S2). Recall that one way to achieve tightness presented in Sect. 4.6 is to choose some sufficiently large but finite $\sigma_n > 0$. While this is sufficient to derive cuts that satisfy properties (S2) and (S3) for some fixed binary precision $\beta_{a(n),j}$, this is not the case if we consider binary refinements. In that case, after each refinement we may require a larger σ_n, such that the sequence of these values diverges. This is detrimental in ensuring property (S3). Therefore, we directly focus on the second case in Sect. 4.6 and choose the ℓ^1-norm. Importantly, Lemma 4.16 and Corollary 4.17 still hold if we use a weighted ℓ^1-norm.

Corollary 5.6. *Choosing $\|\cdot\|^\bullet = \|\cdot\|_{1,w}$ in* (28) *it follows*

$$\phi_{\mathcal{B};n|Z}(\lambda^i_{a(n)}) = \underline{Q}^{R;i+1}_{\mathcal{B};n|Z}(\lambda^i_{a(n)}; \sigma_n\|\cdot\|^\bullet). \tag{31}$$

Equation (31) is sufficient to achieve the overestimation property (S2) in Definition 28 if $\|\cdot\|^\circ$ is any ℓ_p-norm, as we show now.

Lemma 5.7. *Let $\|\cdot\|^\bullet = \|\cdot\|_{1,w}$ in problem* (28) *and let $\|\cdot\|^\circ$ in problem* (21) *be any ℓ_p-norm. Then,*

$$\phi_{\mathcal{B};n|Z}(\lambda^i_{a(n)}) \geq \underline{Q}^{R;i+1}_{n|Z}(x^i_{\mathcal{B},a(n)}; \sigma_n\|\cdot\|^\circ).$$

Proof. For the maximum absolute column sum norm we have $\|G\|_1 = 1$, since G contains at most a single one in each column. Hence, from (30) it follows

$$\|\mathcal{B}_{a(n)}(\lambda^i_{a(n)} - \mathfrak{z}_n)\|_1 \leq \|G\|_1 \; \|\lambda^i_{a(n)} - \mathfrak{z}_n\|_{1,w} = \|\lambda^i_{a(n)} - \mathfrak{z}_n\|_{1,w}. \tag{32}$$

Moreover, we have

$$\|\mathcal{B}_{a(n)}(\lambda^i_{a(n)} - \mathfrak{z}_n)\|_1 \geq \|\mathcal{B}_{a(n)}(\lambda^i_{a(n)} - \mathfrak{z}_n)\|_p \tag{33}$$

for any ℓ_p-norm. Combining some of our previous results and exploiting that $\|\cdot\|^\circ$ is an ℓ_p-norm, we obtain

$$\begin{aligned}\phi_{\mathcal{B};n|Z}(\lambda^i_{a(n)}) &\overset{(31)}{=} \underline{Q}^{R;i+1}_{\mathcal{B};n|Z}(\lambda^i_{a(n)}; \sigma_n\|\cdot\|_{1,w}) \overset{(32)}{\geq} \underline{Q}^{R;i+1}_{\mathcal{B};n|Z}(\lambda^i_{a(n)}; \sigma_n\|\cdot\|_{1,\mathcal{B}}) \\ &\overset{(33)}{\geq} \underline{Q}^{R;i+1}_{\mathcal{B};n|Z}(\lambda^i_{a(n)}; \sigma_n\|\cdot\|^\circ_{\mathcal{B}}) \overset{(27)}{=} \underline{Q}^{R;i+1}_{n|Z}(x^i_{\mathcal{B},a(n)}; \sigma_n\|\cdot\|^\circ).\end{aligned}$$

□

To derive a cut for $\mathcal{Q}_{\mathcal{C}(n)|Z}(\cdot)$, we aggregate the functions $\phi_{\mathcal{B};m|Z}(\cdot)$ for all $m \in \mathcal{C}(n)$:

$$\begin{aligned}\phi_{\mathcal{B};\mathcal{C}(n)|Z}(\lambda_n) &:= \sum_{m\in\mathcal{C}(n)} p_{nm}\big(\mathcal{L}^{i+1}_{\mathcal{B};m|Z}(\pi^i_m) + (\pi^i_m)^\top\lambda_n\big) \\ &= \underbrace{\sum_{m\in\mathcal{C}(n)} p_{nm}\mathcal{L}^{i+1}_{\mathcal{B};m|Z}(\pi^i_m)}_{=:\gamma^i_{\mathcal{C}(n)}} + \Big(\underbrace{\sum_{m\in\mathcal{C}(n)} p_{nm}\pi^i_m}_{=:\pi^i_{\mathcal{C}(n)}}\Big)^\top \lambda_n\end{aligned} \tag{34}$$

The validity and overestimation results naturally extend to this case.

Corollary 5.8. *The function $\phi_{\mathcal{B};\mathcal{C}(n)|Z}(\cdot)$ satisfies*

$$\mathcal{Q}_{\mathcal{C}(n)|Z}(x_n) \geq \phi_{\mathcal{B};\mathcal{C}(n)|Z}(\lambda_n)$$

for all $x_n \in \mathbb{R}^{d_n}$ and any $\lambda_n \in [0,1]^{K_n}$, such that $x_n = \mathcal{B}_n\lambda_n$.

Corollary 5.9. *Let $\|\cdot\|^\bullet = \|\cdot\|_{1,w}$ in* (28) *and let $\|\cdot\|^\circ$ in problem* (21) *be any ℓ_p-norm. Then,*

$$\phi_{\mathcal{B};\mathcal{C}(n)|Z}(\lambda^i_n) \geq \underline{Q}^{R;i+1}_{\mathcal{C}(n)|Z}(x^i_{\mathcal{B},n}; \sigma_n\|\cdot\|^\circ).$$

We now address the projection of these linear cuts back to the original state space.

5.6 The cut projection closure

As explained in Sect. 5.1, the lifting to the binary state space is carried out only temporarily to generate *tight* linear Lagrangian cuts. Importantly, according to Corollary 5.8, these cuts also allow us to obtain valid underapproximations of $\mathcal{Q}_{\mathcal{C}(n)|Z}(\cdot)$ for all points $x_n \in X_n$, even those which cannot be exactly represented by the current binary approximation. More precisely, for some given $x_n \in X_n$, each $\lambda_n \in [0,1]^{K_n}$ such that $x_n = \mathcal{B}_n\lambda_n$ provides a valid underestimator for $\mathcal{Q}_{\mathcal{C}(n)|Z}(x_n)$. We are interested in the pointwise supremum of all these underestimators, that is, the tightest underestimating function that can be gained from projecting the cut to x_n. We refer to this supremum as the cut projection closure.

Definition 5.10 (Cut projection closure)**.** *Let $\phi_{\mathcal{B};\mathcal{C}(n)|Z} : [0,1]^{K_n} \to \mathbb{R}$ be a cut-defining linear function given in* (34)*. Then, the cut projection closure (CPC) $\phi_{\mathcal{C}(n)|Z} : \mathbb{R}^{d_n} \to \mathbb{R}$*

is defined as

$$\phi_{\mathcal{C}(n)|Z}(x_n) := \max_{\lambda_n} \Big\{ \gamma_{\mathcal{C}(n)} + \pi_{\mathcal{C}(n)}^\top \lambda_n \;:\; \mathcal{B}_n \lambda_n = x_n, \lambda_n \leq e, \lambda_n \geq 0 \Big\}. \tag{35}$$

Here, e is a unit vector of dimension K_n.

By strong duality of linear programs, the CPC can be equivalently expressed as

$$\phi_{\mathcal{C}(n)|Z}(x_n) = \min_{\eta_n, \mu_n} \Big\{ \gamma_{\mathcal{C}(n)} + x_n^\top \eta_n + e^\top \mu_n \;:\; \mathcal{B}_n^\top \eta_n + \mu_n \geq \pi_{\mathcal{C}(n)}, \mu_n \geq 0 \Big\}. \tag{36}$$

The dual feasible region in (36) does not depend on x_n and has a finite number of extreme points for a given binary precision. Therefore, we can conclude:

Lemma 5.11. *The CPC $\phi_{\mathcal{C}(n)|Z}(\cdot)$ is piecewise linear and concave on $\mathbb{R}^{d_n}$.*

Moreover, the slope of each piece of the CPC is determined by the value of η_n in an extreme point of (36). Therefore, we analyze these extreme points in more detail. Based on our findings from the previous section, we choose $\|\cdot\|^\bullet$ as the weighted ℓ^1-norm again. Additionally, we define $\sigma_n^{\max} := \max_{m \in \mathcal{C}(n)} \sigma_m$. Then, as proven in Appendix J, we obtain:

Lemma 5.12. *Let $\|\cdot\|^\bullet = \|\cdot\|_{1,w}$ in problem (28). Then, for any binary precision $\beta_{n,j} \in (0,1), j = 1, \dots, d_n$, any extreme point of the feasible set in (36) satisfies $\|\eta_n\|_\infty \leq \sigma_n^{\max}$.*

The crucial idea here is that by a careful choice of the weighted norm to bound the Lagrangian dual (28), effects of the binary expansion are compensated, such that each component of η_n can be bounded independently of the current binary precision $\beta_{n,j} \in (0,1), j = 1, \dots, d_n$, and the number K_n of binary variables. Therefore, this bound remains valid for any refinement of the binary precision, and even with these refinements the CPC is prevented from becoming infinitely steep. This is stated in the following lemma, which is proven in Appendix K.

Lemma 5.13. *Let $\|\cdot\|^\bullet = \|\cdot\|_{1,w}$ in problem (28). Then, for any norm $\|\cdot\|^\circ$, the CPC is a $\tilde{\sigma}_{\mathcal{C}(n)}$-Lipschitz continuous function with $\tilde{\sigma}_{\mathcal{C}(n)} > 0$ independent of the binary precision $\beta_{n,j} \in (0,1), j = 1, \dots, d_n$.*

5.7 Main result

For any node $n \in \mathcal{N}$, using the CPC, we can now determine the non-convex outer approximation of $\mathcal{Q}_{\mathcal{C}(n)|Z}(\cdot)$ as

$$\mathfrak{Q}_{\mathcal{C}(n)|Z}^{i+1}(x_n) = \min \Big\{ \theta_{\mathcal{C}(n)} \in \mathbb{R} \;:\; \theta_{\mathcal{C}(n)} \geq \phi_{\mathcal{C}(n)|Z}^r(x_n) \;\forall r = 1, \dots, i+1 \Big\}. \tag{37}$$

Based on our previous findings we can now state conditions under which this non-convex approximation is sufficient in the sense of Definition 5.1, which is the main result of this section. We provide a proof in Appendix L.

Theorem 5.14. *Let $\|\cdot\|^\bullet = \|\cdot\|_{1,w}$ in problem (28), let $\|\cdot\|^\circ$ in problem (21) be any ℓ_p-norm and let $\sigma_n > 0$. Then, $\mathfrak{Q}_{\mathcal{C}(n)|Z}^{i+1}(\cdot)$ as defined in (37) is a sufficient non-convex approximation of $\mathcal{Q}_{\mathcal{C}(n)|Z}(\cdot)$.*

As pointed out in Sect. 5.2, a sufficient non-convex approximation of $\mathcal{Q}_{\mathcal{C}(n)|Z}(\cdot)$ (with appropriately chosen set $Z_{a(n)} \supseteq X_{a(n)}$) is sufficient to guarantee convergence of SDDiP-related decomposition methods, such as NC-NBD, without the requirement of the technical assumption (A4) in [13]. The condition of setting $\|\cdot\|^{\bullet} = \|\cdot\|_{1,w}$ in problem (28) in the lifted space to achieve such approximation is not really strict, since it still allows to choose any ℓ_p-norm for the regularization in problem (21) in the original state space.

Finally, let us emphasize that the CPC is a non-convex function outer approximating $\mathcal{Q}_{\mathcal{C}(n)|Z}(\cdot)$, and defined by the linear programs (35) and (36). Therefore, directly incorporating it into $\mathfrak{Q}^{i+1}_{\mathcal{C}(n)|Z}(\cdot)$ in subproblem (8) leads to a non-convex bilevel problem. In order to resolve this issue, in [13] it is proposed to first express the CPC by its KKT conditions. Using SOS-1 constraints or a Big-M formulation, this can be achieved without leaving the class of mixed-integer linear programs.

5.8 Illustrative example

We highlight the key take-aways from this section with an illustrative example.

Example 5.15. *Consider the problem* (7) *again, with* $Z = [0, 2]$.

(1) ***CPC for fixed*** σ ***and decreasing*** β***.*** *Let the incumbent be* $x = \frac{6}{5}$. *For the regularization, let* $\|\cdot\|^{\circ} = |\cdot|$ *and* $\sigma = 2$. *For the cut generation, we choose* $K = 3$ *(*$\beta = \frac{2}{7}$*) and* $\|\cdot\|^{\bullet} = \|\cdot\|_{1,w}$ *as proposed. According to Example 5.2, the anchor point is* $x_{\mathcal{B}} = \frac{8}{7}$. *By solving the Lagrangian dual problem* (28) *at that point, we obtain the CPC*

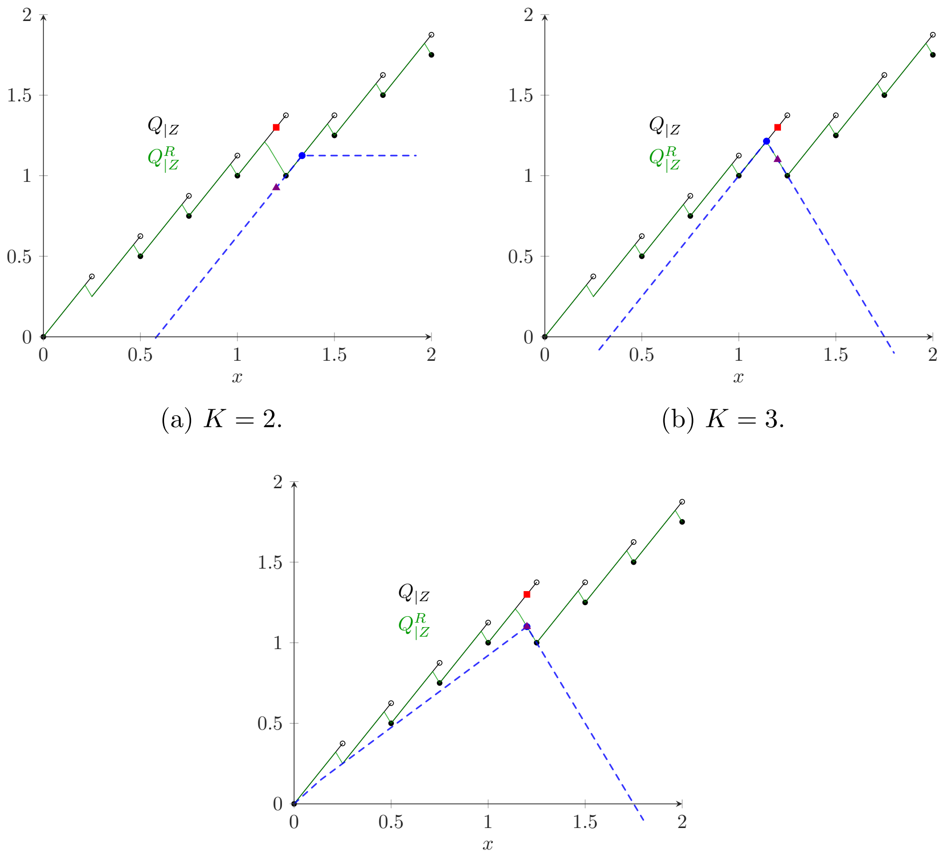
$$\phi_{|Z}(x) = \max_{\lambda} \Big\{ -\frac{1}{2} - \frac{4}{7}\lambda_1 - \frac{8}{7}\lambda_2 + \frac{12}{7}\lambda_3 \ : \ \frac{2}{7}(\lambda_1 + 2\lambda_2 + 4\lambda_3) = x, \ \lambda \in [0,1]^3 \Big\}.$$

We can apply the same procedure for $K = 2$ *and* $K = 4$. *For all three cases, the CPC is visualized in Fig. 9. Its value at* $x_{\mathcal{B}}$ *is highlighted by a blue dot, the value at* $x = \frac{6}{5}$ *by a violet triangle and the true value* $Q_{|Z}(\frac{6}{5})$ *by a red square. We see that in all three cases, the CPC is valid and the regularized value function* $Q^R_{|Z}(\cdot; \sigma|\cdot|)$ *is overestimated at* $x_{\mathcal{B}}$. *In fact, the overestimation is exact for the given example. Moreover, the anchor point* $x_{\mathcal{B}}$ *gets closer to* $x = \frac{6}{5}$ *with increasing the binary precision. For* $K = 4$ *both points coincide. This does not guarantee to monotonically improve the approximation at* $x = \frac{6}{5}$*, though, as Fig. 9b and Fig. 9c show.*

(2) ***CPC for increasing*** σ***.*** *Let the incumbent be* $x = \frac{6}{5}$ *again. As for case (1), we choose* $\|\cdot\|^{\circ} = |\cdot|$ *and* $\|\cdot\|^{\bullet} = \|\cdot\|_{1,w}$. *We fix the binary precision to* $K = 4$ *(*$\beta = \frac{2}{15}$*) and consider different values for* σ. *The obtained CPCs are visualized in Fig. 10. We observe that in each case, the CPC is valid and the corresponding regularized value function is (exactly) overestimated at* $x = \frac{6}{5}$. *Additionally, the slope of the CPC is bounded by* σ. *For increasing values of* σ, *the approximation of the true value function at* $x = \frac{6}{5}$ *is improved.*

(3) ***Using*** $\|\cdot\|^{\bullet} = \|\cdot\|_{\infty,w}$***.*** *Consider the setting from case (1), but with* $\|\cdot\|^{\bullet} = \|\cdot\|_{\infty,w}$ *instead of the weighted 1-norm. We obtain the CPC*

$$\phi_{|Z}(x) := \max_{\lambda} \Big\{ -\frac{2}{7}\lambda_1 - \frac{4}{7}\lambda_2 + 1.10714\lambda_3 \ : \ \frac{2}{7}(\lambda_1 + 2\lambda_2 + 4\lambda_3) = x, \ \lambda \in [0,1]^3 \Big\}.$$



(a) $K = 2$. (b) $K = 3$. (c) $K = 4$.

Figure 9: CPC at $x = \frac{6}{5}$ for $\sigma = 2$, $\|\cdot\|^\bullet = \|\cdot\|_{1,w}$ and different K in Example 5.15.

As Fig. 11a shows, in this case, the overestimation property is not satisfied.

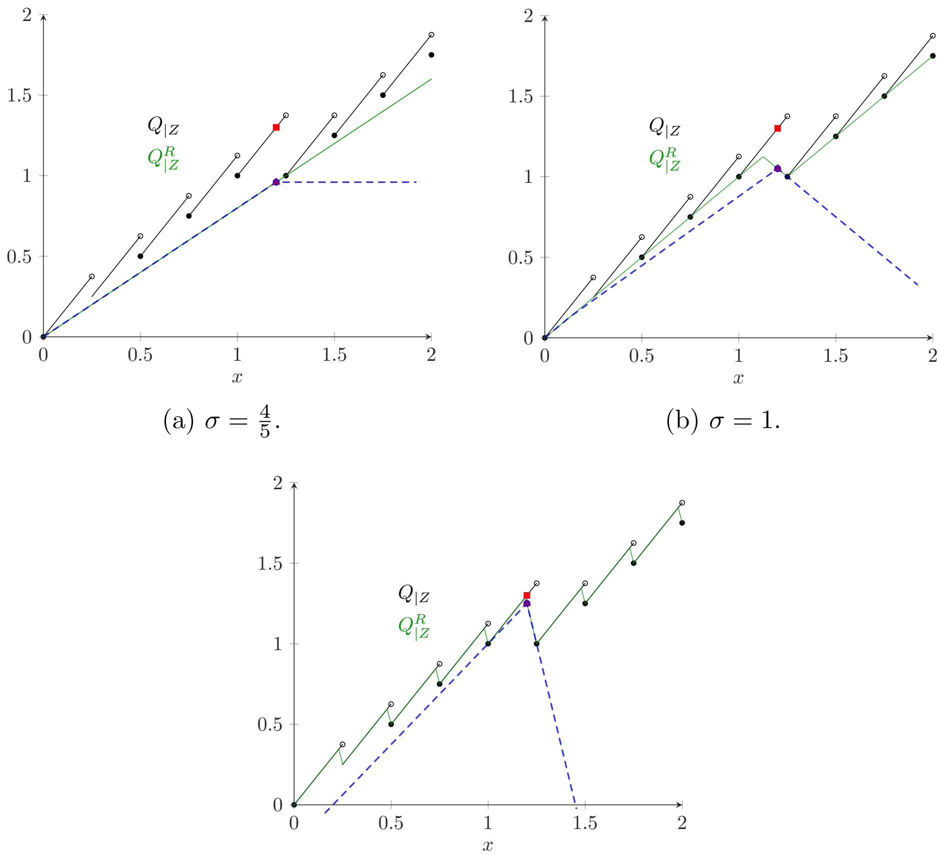
Setting $\sigma = 1$, we even observe a case in which $Q^R_{\mathcal{B}|Z}(\cdot; \sigma\|\cdot\|_{\infty,w})$ and its closed convex envelope $\overline{\text{co}}(Q^R_{\mathcal{B}|Z}; \sigma\|\cdot\|_{\infty,w})(\cdot)$ do not coincide at $\lambda = (0, 0, 1)$, which is the binary representation of the anchor point $x_\mathcal{B}$. This underlines the significance of choosing $\|\cdot\|^\bullet = \|\cdot\|_{1,w}$.

(4) ***Using $\mathfrak{z} \in \{0,1\}^K$.*** *Consider the same setting as for case (1), but with choosing $\mathfrak{z} \in \{0,1\}^K$ instead of $\mathfrak{z} \in [0,1]^K$ as the accompanying set in the lifted space. The CPC can be computed as*

$$\phi_{|Z}(x) := \max_{\lambda} \left\{ 0.914286\lambda_3 \ : \ \frac{2}{7}(\lambda_1 + 2\lambda_2 + 4\lambda_3) = x, \ \lambda \in [0,1]^3 \right\}.$$

As Fig. 11b shows, this is not a valid cut for the value function $Q_{|Z}(\cdot)$ in the original state space. The CPC is only guaranteed to be valid for points which can be exactly represented in the lifted state space.

(5) ***Bounding the slope of the CPC.*** *Let $x = 1.249$ now, i.e., very close to a point of discontinuity of $Q_{|Z}(\cdot)$. Let $\|\cdot\|^\circ = |\cdot|$, $\sigma = 5$, $K = 8$ for a sufficiently close approximation, and $\|\cdot\|^\bullet = \|\cdot\|_{1,w}$ as proposed. As Fig. 12a illustrates, in this case, we obtain a CPC that overestimates $Q^R_{|Z}(\cdot; \sigma|\cdot|)$, but is bounded in slope by $\sigma = 5$*



(a) $\sigma = \frac{4}{5}$.

(b) $\sigma = 1$.

(c) $\sigma = 5$.

Figure 10: CPC at $x = \frac{6}{5}$ for $K = 4$, $\|\cdot\|^\bullet = \|\cdot\|_{1,w}$ and different values of σ in Example 5.15.

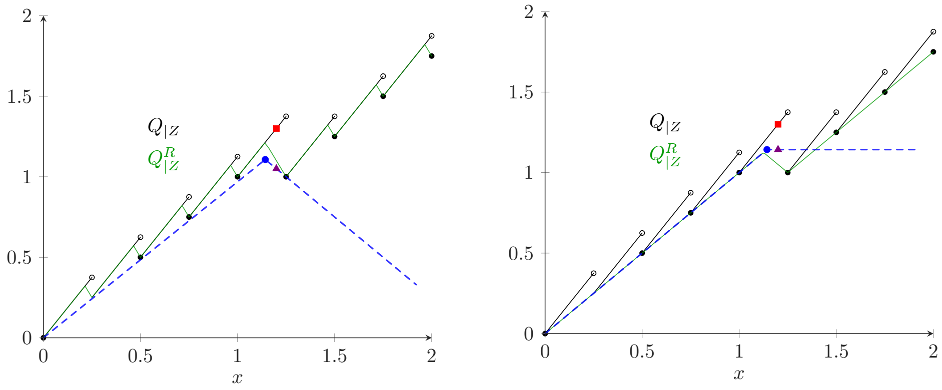
(blue dotted line).

In contrast, consider the case where we do not use the weighted norm $\|\cdot\|^\bullet = \|\cdot\|_{1,w}$*, but the unweighted one* $\|\cdot\|^\bullet = \|\cdot\|_1$ *instead. As proven in [13] we can then bound the dual multipliers in* (28) *by* $\sigma \max_j U_j$ *to achieve the intended overestimation. However, the CPC becomes extremely steep for* $K = 8$ *and is not bounded by* $\sigma = 5$*, see Fig. 12b. In general, its Lipschitz constant may diverge for* $\beta \to 0$*.*

For completeness, we should mention that Lagrangian dual problems are often degenerate with infinitely many optimal solutions. Therefore, given an incumbent $x^i_{a(n)}$, there may exist an infinite number of tight linear Lagrangian cuts (satisfying Corollary 3.14) or tight CPCs (satisfying Theorem 5.14) with varying approximation quality at $x_{a(n)} \neq x^i_{a(n)}$. For illustration, see the blue dashed, cyan dotted, and magenta dash-dotted CPCs for problem (7) shown in Fig. 12a.

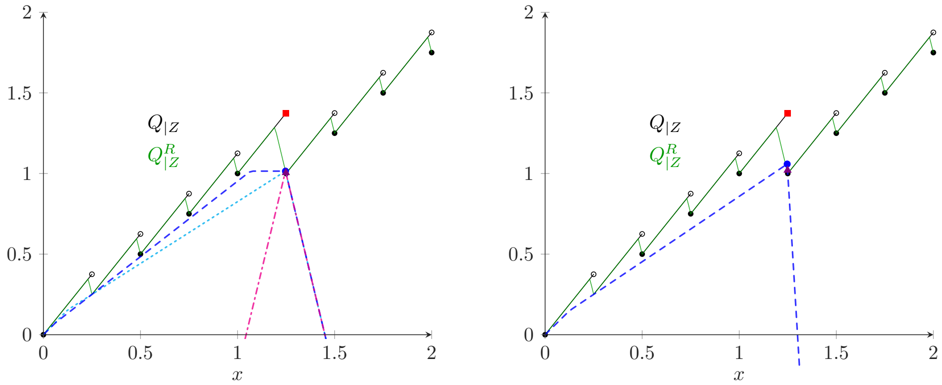
6 Conclusion

We provide new theoretical insight on the generation of linear and non-convex cuts for value functions of MS-MILPs based on Lagrangian duality, and the effects that copy constraints and a Lipschitz regularization of the subproblems have in this context. In



(a) $\sigma = 2$ and $\|\cdot\|^\bullet = \|\cdot\|_{\infty,w}$. (b) $\sigma = 1, \mathfrak{z} \in \{0,1\}^K$.

Figure 11: Non-sufficient CPCs for $K = 3$ at $x = \frac{6}{5}$ in Example 5.15.



(a) Using $\|\cdot\|^\bullet = \|\cdot\|_{1,w}$. (b) Using $\|\cdot\|^\bullet = \|\cdot\|_1$.

Figure 12: CPC at $x = 1.249$ for $\sigma = 5$ and $K = 8$ in Example 5.15.

particular, we point out the relation between bounded Lagrangian dual problems and the convex envelope of the regularized value functions. We further show that by a careful choice of the regularization, this relation can be exploited to generate non-convex cuts with favorable properties.

As future work directions, a computational comparison of generating linear Lagrangian cuts using non-regularized and regularized problems could be of interest. While the approximation quality is better in the first case, bounding the Lagrangian duals in the latter might accelerate the cut generation process. For non-convex approximations, the CPC from our lift-and-project approach could be compared in detail with the augmented Lagrangian cuts from [1]. Another challenge is to efficiently incorporate the non-convex CPC into the subproblems within SDDP-like methods.

Acknowledgments

Andy Sun's research is partially funded by the National Science Foundation CAREER award 2316675. Christian Füllner's research is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)–445857709. The authors thank Filipe Cabral and Shixuan Zhang for the fruitful discussions on this topic during Christian

Füllner's research visit at Georgia Institute of Technology. This visit was funded by the Karlsruhe House of Young Scientists (KHYS).

References

[1] S. Ahmed, F. G. Cabral, and B. Freitas Paulo da Costa. Stochastic Lipschitz dynamic programming. *Mathematical Programming*, 191:755–793, 2022.

[2] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2nd edition, 2017.

[3] J. F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4(1):238–252, 1962.

[4] D. P. Bertsekas. *Convex Optimization Theory*. Athena Scientific, 2009.

[5] J. R. Birge. Decomposition and partitioning methods for multistage stochastic linear programs. *Operations Research*, 33(5):989–1007, 1985.

[6] J. R. Birge and F. Louveaux. *Introduction to stochastic programming*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2nd edition, 2011.

[7] C.C. Carøe and R. Schultz. Dual decomposition in stochastic integer programming. *Operations Research Letters*, 24(1):37–45, 1999.

[8] S. Cerisola, J. M. Latorre, and A. Ramos. Stochastic dual dynamic programming applied to nonconvex hydrothermal models. *European Journal of Operational Research*, 218(3): 687–697, 2012.

[9] R. Chen and J. Luedtke. On generating Lagrangian cuts for two-stage stochastic integer programs. *INFORMS Journal on Computing*, 34(4):2332–2349, 2022.

[10] M. Conforti, G. Cornuéjols, and G. Zambelli. *Polyhedral approaches to mixed integer linear programming*, pages 343–385. 50 Years of Integer Programming 1958-2008: From the Early Years to the State-of-the-Art. Springer Berlin Heidelberg, 2010.

[11] J. E. Falk. Lagrange multipliers and nonconvex programs. *SIAM Journal on Control*, 7(4): 534–545, 1969.

[12] M.J. Feizollahi, S. Ahmed, and A. Sun. Exact augmented Lagrangian duality for mixed integer linear programming. *Mathematical Programming*, 161(1-2):365–387, 2017.

[13] C. Füllner and S. Rebennack. Non-convex nested Benders decomposition. *Mathematical Programming*, 196:987–1024, 2022.

[14] C. Füllner and S. Rebennack. Stochastic dual dynamic programming and its variants - a review. Preprint, available at `http://www.optimization-online.org/DB_FILE/2021/01/8217.pdf`, 2023.

[15] A. M. Geoffrion. Lagrangean relaxation for integer programming. In M.L. Balinski, editor, *Approaches to integer programming*, pages 82–114. Springer Berlin Heidelberg, Berlin, Heidelberg, 1974.

[16] F. Glover. Improved linear integer programming formulations of nonlinear integer problems. *Management Science*, 22(4):455–460, 1975.

[17] V. Leclère, P. Carpentier, J.-P. Chancelier, A. Lenoir, and F. Pacaud. Exact converging bounds for stochastic dual dynamic programming via Fenchel duality. *SIAM Journal on Optimization*, 30(2):1223–1250, 2020.

[18] R. R. Meyer. On the existence of optimal solutions to integer and mixed-integer programming problems. *Mathematical Programming*, 7:223–235, 1974.

[19] R. R. Meyer. Integer and mixed-integer programming models: general properties. *Journal of Optimization Theory and Applications*, 3(4), 1975.

[20] M. V. F. Pereira and L. M. V. G. Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52:359–375, 1991.

[21] A. B. Philpott, F. Wahid, and F. Bonnans. MIDAS: A mixed integer dynamic approximation scheme. *Mathematical Programming*, 181:19–50, 2020.

[22] R. Rahmaniani, S. Ahmed, T. G. Crainic, M. Gendreau, and W. Rei. The Benders dual decomposition method. *Operations Research*, 68(3):878–895, 2020.

[23] R. T. Rockafellar. *Convex analysis*. Princeton university press, 1970.

[24] G. Steeger and S. Rebennack. Dynamic convexification within nested Benders decomposition using Lagrangian relaxation: an application to the strategic bidding problem. *European Journal of Operational Research*, 257(2):669–686, 2017.

[25] G. Steeger, T. Lohmann, and S. Rebennack. Strategic bidding for a price-maker hydroelectric producer: stochastic dual dynamic programming and Lagrangian relaxation. *IISE Transactions*, 47:1–14, 2018.

[26] F. Tardella. On the existence of polyhedral convex envelopes. In C. A. Floudas and P. M. Pardalos, editors, *Frontiers in global optimization*, pages 563–573. Springer, 2004.

[27] N. van der Laan and W. Romeijnders. A converging Benders' decomposition algorithm for two-stage mixed-integer recourse models. *Operations Research*, 2023.

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

[29] S. Zhang and X. A. Sun. Stochastic dual dynamic programming for multistage stochastic mixed-integer nonlinear optimization. *Mathematical Programming*, 196:935–985, 2022.

[30] S. Zhang and X. A. Sun. On distributionally robust multistage convex optimization: new algorithms and complexity analysis. available at `https://arxiv.org/pdf/2010.06759.pdf`, 2022.

[31] J. Zou, S. Ahmed, and X. A. Sun. Stochastic dual dynamic integer programming. *Mathematical Programming*, 175:461–502, 2019.

A Proof of Lemma 2.2

Proof. Consider a leaf node n of $\mathcal{N}$. Since $f_n(\cdot)$ is linear and $F_n(x_{a(n)})$ is bounded for all $x_{a(n)}$ by Assumption 1 (A1), we conclude that $Q_n(\cdot)$ is bounded from below. Moreover, by feasibility assumption (A3), we have $\operatorname{dom}(Q_n) \neq \emptyset$. Hence, $Q_n(\cdot)$ is proper. For all other nodes in $\mathcal{N}$, a similar reasoning can be applied inductively.

The lsc of $Q_n(\cdot)$ and the closedness of $\operatorname{dom}(Q_n)$ follow from [19, Theorem 3.1] under Assumption 1 (A2), based on the observation that, apart from the integrality requirements, X_n and Y_n are representable by polyhedral constraints. The piecewise polyhedrality follows from the mixed-integer linear character of the subproblems, see also [19]. By taking expectations the assertion also holds for $\mathcal{Q}_{\mathcal{C}(n)}(\cdot)$. □

B Proof of Lemma 2.3

Proof. Since $Q_n(\cdot)$ is bounded from below (see Appendix A), also $\overline{\operatorname{co}}(Q_n)(\cdot)$ is bounded from below (there exists a constant convex function underestimating $Q_n(\cdot)$). By Assumption 1 (A3), $\operatorname{dom}(Q_n) \neq \emptyset$. As $\overline{\operatorname{co}}(Q_n)(\cdot)$ underestimates $Q_n(\cdot)$ on $\operatorname{dom}(Q_n)$, this implies that $\overline{\operatorname{co}}(Q_n)(\cdot)$ is proper. Then the assertion follows from [4, Proposition 1.6.1]. □

C Proof of Lemma 3.7

Proof. Recall from Lemma 3.5 that the objective function of subproblem (8) is bounded from below on $\mathcal{M}_{n|Z}^{i+1}$. The objective function is the same for (11), and by Lemma 3.4 also the recession cones of $\mathcal{M}_{n|Z}^{i+1}$ and $\operatorname{conv}(\mathcal{M}_{n|Z}^{i+1})$ do coincide. Therefore, the objective function is also bounded from below on $\operatorname{conv}(\mathcal{M}_{n|Z}^{i+1})$.

The result $\operatorname{dom}(\underline{Q}_{n|Z}^{C,i+1}) = \operatorname{conv}(\operatorname{dom}(\underline{Q}_{n|Z}^{i+1}))$ follows by standard convexity arguments considering the constraint set $\operatorname{conv}(\mathcal{M}_{n|Z}^{i+1})$ instead of $\mathcal{M}_{n|Z}^{i+1}$. By Assumption 1 (A3), we have $\operatorname{dom}(\underline{Q}_{n|Z}^{i+1}) \neq \emptyset$, and thus $\operatorname{dom}(\underline{Q}_{n|Z}^{C,i+1}) \neq \emptyset$. Together with the boundedness from below, the properness follows.

Due to Lemma 3.4, problem (11) can be rewritten as a linear program, and therefore the finite minimum is attained on $\operatorname{conv}(\operatorname{dom}(\underline{Q}_{n|Z}^{i+1}))$. Moreover, for linear programs, the lower semicontinuity and convexity on $\mathbb{R}^{d_{a(n)}}$ (with $\underline{Q}_{n|Z}^{C,i+1}(x_{a(n)}) = +\infty$ for all $x_{a(n)} \notin \operatorname{conv}(\operatorname{dom}(\underline{Q}_{n|Z}^{i+1}))$) and the piecewise linearity on $\operatorname{dom}(\underline{Q}_{n|Z}^{C,i+1})$ follow from standard results in stochastic programming, see [6, 19] for instance. □

D Proof of Theorem 3.9

Proof. Applying the arguments from Lemma 2.3 to $\underline{Q}_{n|Z}^{i+1}(\cdot)$, it follows that $(\underline{Q}_{n|Z}^{i+1})^{**}(x_{a(n)}) = \overline{\operatorname{co}}(\underline{Q}_{n|Z}^{i+1})(x_{a(n)})$ for all $x_{a(n)} \in \mathbb{R}^{d_{a(n)}}$, so it is sufficient for us to consider the biconjugate.

For any $\pi_n \in \mathbb{R}^{d_{a(n)}}$, the conjugate of $\underline{Q}_{n|Z}^{i+1}(\cdot)$ is defined as

$$(\underline{Q}_{n|Z}^{i+1})^*(\pi_n) = \max_{u_n} \big\{ \pi_n^\top u_n - \underline{Q}_{n|Z}^{i+1}(u_n) \big\} = - \min_{u_n} \big\{ -\pi_n^\top u_n + \underline{Q}_{n|Z}^{i+1}(u_n) \big\},$$

with $u_n, \pi_n \in \mathbb{R}^{d_{a(n)}}$ [4]. Then, the respective biconjugate is equal to $\underline{Q}^{D,i+1}_{n|Z}(x^i_{a(n)})$, since

$$(\underline{Q}^{i+1}_{n|Z})^{**}(x^i_{a(n)}) = \max_{\pi_n} \left\{ \pi_n^\top x^i_{a(n)} - (\underline{Q}^{i+1}_{n|Z})^*(\pi_n) \right\} = \max_{\pi_n} \min_{u_n} \left\{ \pi_n^\top (x^i_{a(n)} - u_n) + \underline{Q}^{i+1}_{n|Z}(u_n) \right\}.$$

Inserting the definition of $\underline{Q}^{i+1}_{n|Z}(\cdot)$ and utilizing the copy constraint to replace u_n with z_n, we obtain the Lagrangian dual (10). In particular, this holds true for the case where finite values are obtained, proving the assertion. □

E Proof of Theorem 3.13

Proof. First, we notice that $\text{dom}(\underline{Q}^{i+1}_{n|Z}) \subseteq Z_{a(n)}$ and bounded by assumption. By Lemma 3.5 it is also closed, thus compact. Under this condition, the closed convex envelope $\overline{\text{co}}(\underline{Q}^{i+1}_{n|Z})(\cdot)$ and the convex envelope $\text{co}(\underline{Q}^{i+1}_{n|Z})(\cdot)$ do coincide on $\text{dom}(\underline{Q}^{i+1}_{n|Z})$ [11, Theorem 2.2]. Therefore, it remains to be shown that

$$\text{co}(\underline{Q}^{i+1}_{n|Z})(x_{a(n)}) = \underline{Q}^{i+1}_{n|Z}(x_{a(n)})$$

for all $x_{a(n)} \in X_{a(n)} \cap \text{dom}(\underline{Q}^{i+1}_{n|Z})$.

Now choose $x_{a(n)}$ arbitrarily from this set. Since $x_{a(n)} \in X_{a(n)}$, by assumption $x_{a(n)}$ is an extreme point of $Z_{a(n)}$, meaning it cannot be expressed as a convex combination of points in $Z_{a(n)}$ different from itself [26]. However, we also have $x_{a(n)} \in \text{dom}(\underline{Q}^{i+1}_{n|Z}) \subseteq Z_{a(n)}$, so it cannot be expressed as a convex combination of points in $\text{dom}(\underline{Q}^{i+1}_{n|Z})$ different from itself either. Therefore, $x_{a(n)}$ is an extreme point of $\text{dom}(\underline{Q}^{i+1}_{n|Z})$. According to [26, Proposition 2.1], this implies that $\text{co}(\underline{Q}^{i+1}_{n|Z})(x_{a(n)}) = \underline{Q}^{i+1}_{n|Z}(x_{a(n)})$ for all $x_{a(n)} \in X_{a(n)}$. □

F Proof of Lemma 4.4

Proof. We prove all results for leaf nodes $n \in \mathcal{N}$. For other nodes, the same reasoning can be applied inductively, using that $\mathcal{Q}^R_{\mathcal{C}(n)|Z}(\cdot; \sigma_{\mathcal{C}(n)}\|\cdot\|)$ is Lipschitz and underestimating $\mathcal{Q}_{\mathcal{C}(n)|Z}(\cdot)$. First, we show that the minimum in subproblem (12) is well-defined. The feasible set is

$$\mathcal{M}^{R;i+1}_{n|Z} := \Big\{ (x_n, y_n, z_n) \ : \ z_n \in Z_{a(n)}, x_n \in X_n, y_n \in Y_n, A_n z_n + B_n x_n + C_n y_n \geq b_n \Big\}.$$

Under Assumptions 1 and 2, this set is closed as the intersection of closed sets.

By definition, $\mathcal{Q}^R_{\mathcal{C}(n)|Z}(\cdot; \sigma_{\mathcal{C}(n)}\|\cdot\|) \equiv 0$ is Lipschitz continuous. Therefore, the objective function $g_n(x_n, y_n, z_n) := f_n(x_n, y_n) + \sigma_n \|x_{a(n)} - z_n\| + \mathcal{Q}^R_{\mathcal{C}(n)|Z}(x_n; \sigma_{\mathcal{C}(n)}\|\cdot\|)$ is lsc. Together with the closedness of $\mathcal{M}^{R;i+1}_{n|Z}$, then for any $\alpha \in \mathbb{R}$, the level set

$$\text{lev}_\alpha(g_n) = \Big\{ (x_n, y_n, z_n) \in \mathcal{M}^{R;i+1}_{n|Z} \ : \ g_n(x_n, y_n, z_n) \leq \alpha \Big\}$$

is closed. Moreover, by Assumption 1, x_n and y_n are bounded, and by that $f_n(x_n, y_n) + \mathcal{Q}^R_{\mathcal{C}(n)|Z}(x_n; \sigma_{\mathcal{C}(n)}\|\cdot\|)$ is bounded from below by some finite constant $\tilde{\alpha}$. The remaining term $\sigma_n \|x_{a(n)} - z_n\|$ is bounded from below by 0 and bounded from above by $\alpha - \tilde{\alpha}$. Therefore, within $\text{lev}_\alpha(g_n)$, z_n is bounded as well. In total, $\text{lev}_\alpha(g_n)$ is compact. For α sufficiently large, it is also non-empty. Then, by extensions of the Theorem of

Weierstraß, a finite minimum is attained in subproblem (12). This immediately implies properness of $Q^R_{n|Z}(\cdot;\sigma_n\|\cdot\|)$.

Second, assuming that it exists, let (x^*_n, y^*_n, z^*_n) be an optimal solution to the original subproblem (5). Clearly, this point is feasible, but not necessarily optimal for subproblem (12). Due to the copy constraint in subproblem (5), the term $\sigma_n\|x_{a(n)} - z^*_n\|$ vanishes in the objective, and the underestimation property follows. In contrast, if subproblem (5) has no feasible solution given $x_{a(n)}$, then we have $Q_{n|Z}(x_{a(n)}) = +\infty$ and the assertion is trivial.

Finally, we notice that $Q^R_{n|Z}(\cdot;\sigma_n\|\cdot\|)$ can be interpreted as the Pasch-Hausdorff envelope (or Lipschitz regularization) of $Q_{n|Z}(\cdot)$ and $\sigma_n\|\cdot\|$:

$$Q^R_{n|Z}(x_{a(n)};\sigma_n\|\cdot\|) = Q_{n|Z}\Box(\sigma_n\|\cdot\|)(x_{a(n)}) = \min_{z_n\in Z_{a(n)}} Q_{n|Z}(z_n) + \sigma_n\|x_{a(n)} - z_n\|.$$

By a similar reasoning as above, a finite minimum is attained. Since $Q_{n|Z}(\cdot)$ is proper and lsc according to Lemma 2.2 and since $Q^R_{n|Z}(\cdot;\sigma_n\|\cdot\|)$ is proper as well, it follows that $Q^R_{n|Z}(\cdot;\sigma_n\|\cdot\|)$ is σ_n-Lipschitz by a general property of the Pasch-Hausdorff envelope, see [2, Proposition 12.17]. □

G Proof of Lemma 4.8

Proof. (a) Problem (14) is a relaxation of problem (12). Hence, by Lemma 4.3, it is feasible for all $x_{a(n)} \in \mathbb{R}^{d_{a(n)}}$. Additionally, x_n, y_n are bounded, whereas $\theta_{\mathcal{C}(n)}$ and $\|\cdot\|$ are at least bounded from below. Therefore, $\underline{Q}^{R;i+1}_{n|Z}(\cdot;\sigma_n\|\cdot\|)$ is finite-valued, and by that also proper. Using the same reasoning as in Appendix F, we can also show that finite infima are attained.

The Lipschitz continuity can be shown by exploiting that $\underline{Q}^{R;i+1}_{n|Z}(\cdot;\sigma_n\|\cdot\|)$ is the Pasch-Hausdorff envelope (or Lipschitz regularization) of $\underline{Q}^{i+1}_{n|Z}(\cdot)$ and $\sigma_n\|\cdot\|$ for all $n \in \mathcal{N}$. Since $\underline{Q}^{i+1}_{n|Z}(\cdot)$ is proper and lsc according to Lemma 3.5 and since $\underline{Q}^{R;i+1}_{n|Z}(\cdot;\sigma_n\|\cdot\|)$ is proper as well, it follows that $\underline{Q}^{R;i+1}_{n|Z}(\cdot;\sigma_n\|\cdot\|)$ is σ_n-Lipschitz continuous by a general property of the Pasch-Hausdorff envelope, see [2, Proposition 12.17].

(b) The convexity can be shown in a straightforward way given that the feasible set and the objective are convex. As problem (15) is a relaxation of problem (14), feasibility and boundedness of $\underline{Q}^{CR;i+1}_{n|Z}(\cdot;\sigma_n\|\cdot\|)$ from above follow from (a). Recall that x_n, y_n are bounded and $\theta_{\mathcal{C}(n)}$ is bounded from below in the set $\mathcal{M}^{i+1}_{n|Z}$. By convexity properties, it can be shown that these properties also must hold for elements in $\operatorname{conv}(\mathcal{M}^{i+1}_{n|Z})$. Therefore, the objective function is bounded from below. It follows that $\underline{Q}^{CR;i+1}_{n|Z}(\cdot;\sigma_n\|\cdot\|)$ is finite-valued, and by that also proper.

Using the same arguments as in (a) for $\underline{Q}^{CR;i+1}_{n|Z}(\cdot;\sigma_n\|\cdot\|)$ and $\underline{Q}^{C,i+1}_{n|Z}(\cdot)$ together with Lemma 3.7 we can conclude that $\underline{Q}^{CR;i+1}_{n|Z}(\cdot;\sigma_n\|\cdot\|)$ is σ_n-Lipschitz continuous on $\mathbb{R}^{d_{a(n)}}$, and thus also closed.

(c) The function $\overline{\text{co}}(\underline{Q}^{R;i+1}_{n|Z};\sigma_n\|\cdot\|)(\cdot)$ is convex by definition. As it underestimates $\underline{Q}^{R;i+1}_{n|Z}(\cdot;\sigma_n\|\cdot\|)$, which is finite-valued, it is finite-valued as well.

(d) As shown in (b), $\underline{Q}^{CR;i+1}_{n|Z}(\cdot;\sigma_n\|\cdot\|)$ is closed proper convex. Therefore, by [4, Proposition 1.6.1 (c)] it coincides with its biconjugate $(\underline{Q}^{CR;i+1}_{n|Z};\sigma_n\|\cdot\|)^{**}(\cdot)$ on $\mathbb{R}^{d_{a(n)}}$. □

H Proof of Lemma 4.9

Proof. We define c_n and λ_n as in the proof of Theorem 4.10. First, we consider problem (14), but introduce an additional variable and copy constraint to obtain

$$\underline{Q}_{n|Z}^{R;i+1}(x_{a(n)}^i;\sigma_n\|\cdot\|) = \min_{\lambda_n,z_n,u_n} \left\{c_n^\top \lambda_n + \sigma_n\|u_n\| \ : \ (\lambda_n,z_n)\in\mathcal{M}_{n|Z}^{i+1}, u_n = x_{a(n)}^i - z_n\right\}.$$

We relax the equality constraint, which yields the dual function

$$\begin{aligned}\Phi_{|Z}(\pi_n) &:= \min_{\lambda_n,z_n,u_n,w_n} \left\{c_n^\top \lambda_n + \sigma_n\|u_n\| + \pi_n^\top(u_n - (x_{a(n)}^i - z_n)) \ : \ (\lambda_n,z_n)\in\mathcal{M}_{n|Z}^{i+1}\right\}\\ &= \min_{(\lambda_n,z_n)\in\mathcal{M}_{n|Z}^{i+1}} c_n^\top\lambda_n - \pi_n^\top(x_{a(n)}^i - z_n) + \min_{u_n}\sigma_n\|u_n\| + \pi_n^\top u_n\\ &= \min_{(\lambda_n,z_n)\in\operatorname{conv}(\mathcal{M}_{n|Z}^{i+1})} c_n^\top\lambda_n - \pi_n^\top(x_{a(n)}^i - z_n) + \min_{u_n}\sigma_n\|u_n\| + \pi_n^\top u_n.\end{aligned}$$

The first equation follows from separability, while the second one follows from the linearity of the objective in the first minimization problem.

Using the same steps for the convexified problem (15), we obtain the dual function $\Phi_{|Z}^C(\pi_n)$, which satisfies $\Phi_{|Z}(\pi_n) = \Phi_{|Z}^C(\pi_n)$ for all $\pi_n\in\mathbb{R}^{d_{a(n)}}$. As they are defined by taking the supremum of $\Phi_{|Z}(\pi_n)$ or $\Phi_{|Z}^C(\pi_n)$ over all π_n respectively, also the biconjugates $(\underline{Q}_{n|Z}^{R;i+1};\sigma_n\|\cdot\|)^{**}(\cdot)$ and $(\underline{Q}_{n|Z}^{CR;i+1};\sigma_n\|\cdot\|)^{**}(\cdot)$ are equivalent. □

I Proof of Lemma 4.16

Proof. We construct a special feasible solution for the dual problem (16). Suppose $\widehat{\pi}_n\in\mathbb{R}^{d_{a(n)}}$ is a vector, for which each component j is defined by

$$\widehat{\pi}_{nj} := \begin{cases}\sigma_n & \text{if } x_{a(n),j}^i = 1\\ -\sigma_n & \text{if } x_{a(n),j}^i = 0.\end{cases} \tag{38}$$

Such a construction is always possible, since $x_{a(n)}^i\in\{0,1\}^{d_{a(n)}}$. The vector $\widehat{\pi}_n$ is feasible for (16) with ℓ^∞-norm. Therefore, we obtain

$$\begin{aligned}\underline{Q}_{n|Z}^{DR;i+1}(x_{a(n)}^i;\sigma_n\|\cdot\|_1) &= \max_{\|\pi_n\|_*\le\sigma_n}\min_{z_n\in Z_{a(n)}}\left\{\underline{Q}_{n|Z}^{i+1}(z_n) + \pi_n^\top(x_{a(n)}^i - z_n)\right\}\\ &\ge \min_{z_n\in Z_{a(n)}}\left\{\underline{Q}_{n|Z}^{i+1}(z_n) + \widehat{\pi}_n^\top(x_{a(n)}^i - z_n)\right\}\\ &= \min_{z_n\in Z_{a(n)}}\left\{\underline{Q}_{n|Z}^{i+1}(z_n) + \sum_{j=1}^{d_{a(n)}}\widehat{\pi}_{tj}(x_{a(n),j}^i - z_{nj})\right\}.\end{aligned} \tag{39}$$

Now we exploit the binary nature of $x_{a(n)}$. If $x_{a(n),j}^i = 1$, from (38) it follows

$$\widehat{\pi}_{tj}(x_{a(n),j}^i - z_{nj}) = \sigma_n(x_{a(n),j}^i - z_{nj}) = \sigma_n|x_{a(n),j}^i - z_{nj}|.$$

The last equality holds, since for $x_{a(n),j}^i = 1$ and $z_{nj}\in[0,1]$ (or $z_{nj}\in\{0,1\}$), the term $x_{a(n),j}^i - z_{nj}$ is always non-negative. Analogously, for $x_{a(n),j}^i = 0$, it follows

$$\widehat{\pi}_{tj}(x_{a(n),j}^i - z_{nj}) = -\sigma_n(x_{a(n),j}^i - z_{nj}) = \sigma_n|x_{a(n),j}^i - z_{nj}|.$$

Inserting this result in (39) yields

$$\begin{aligned} \underline{Q}_{n|Z}^{DR;i+1}(x_{a(n)}^i;\sigma_n\|\cdot\|_1) &\geq \min_{z_n\in Z_{a(n)}} \left\{\underline{Q}_{n|Z}^{i+1}(z_n) + \sum_{j=1}^{d_{a(n)}} \sigma_n |x_{a(n),j}^i - z_{nj}|\right\} \\ &= \underline{Q}_{n|Z}^{R;i+1}(x_{a(n)}^i;\sigma_n\|\cdot\|_1). \end{aligned} \tag{40}$$

□

J Proof of Lemma 5.12

Proof. First, we notice that the dual CPC problem (36) is separable in the dimensions $j = 1, \ldots, d_n$ of the original state space. Hence, we may analyze each case separately, which yields

$$\min_{\eta_{nj},\mu_{nj}} \left\{x_{nj}\eta_{nj} + e^\top \mu_{nj} \;:\; \eta_{nj}\mathfrak{b}_{nj} + \mu_{nj} \geq \pi_{\mathcal{C}(n)}^j, \mu_{nj} \geq 0\right\}.$$

$\mathfrak{b}_{nj} \in \mathbb{R}^{K_{nj}}$ is a vector which contains the non-zero entries from the j-th column of $\mathcal{B}_n^\top$. The variable η_{nj} is one-dimensional, and $\pi_{\mathcal{C}(n)}^j \in \mathbb{R}^{K_{nj}}$ contains all entries from $\pi_{\mathcal{C}(n)}$ referring to component j.

We introduce slack variables and split up η_{nj} to reformulate the constraints as:

$$\begin{aligned} \eta_{nj}^+\mathfrak{b}_{nj} - \eta_{nj}^-\mathfrak{b}_{nj} + \mu_{nj} - \nu_{nj} &= \pi_{\mathcal{C}(n)}^j \\ \eta_{nj}^+, \eta_{nj}^-, \mu_{nj}, \nu_{nj} &\geq 0. \end{aligned} \tag{41}$$

The set defined by (41) has $2+2K_{nj}$ variables. In a basic solution, $2+K_{nj}$ variables have to be zero and the K_{nj} columns associated with the remaining variables have to be linearly independent. We observe that for each row $\kappa = 1, \ldots, K_{nj}$, the variables μ_{njk} and ν_{njk} cannot be in the basis together, because otherwise the basic columns are not linearly independent. With the same reasoning, η_{nj}^+ and η_{nj}^- cannot be in the basis together. Moreover, for $K_{nj} > 1$, it is not sufficient to have only η_{nj}^+ or η_{nj}^- in the basis.

We now consider different cases of basic solutions.

Case 2. η_{nj}^+ in the basis. This implies $\eta_{nj}^- = 0$. The equation

$$\eta_{nj}^+ = \frac{\pi_{\mathcal{C}(n),k(j,\kappa)} - \mu_{nj\kappa} + \nu_{nj\kappa}}{2^{\kappa-1}\beta_{nj}}$$

has to be satisfied for all $\kappa = 1, \ldots, K_{nj}$ simultaneously. However, since η_{nj}^+ is in the basis, for some $\bar{\kappa}$, both $\mu_{nj\bar{\kappa}}$ and $\nu_{nj\bar{\kappa}}$ have to be zero. Therefore, this $\bar{\kappa}$ determines the value of η_{nj}^+. The largest possible value that $\eta_{nj}^+ \geq 0$ may take can be obtained by maximizing over k:

$$\eta_{nj}^+ \leq \max_{\kappa=1,\ldots,K_{nj}} \max\left\{\frac{\pi_{\mathcal{C}(n),k(j,\kappa)}}{2^{\kappa-1}\beta_{nj}}, 0\right\}. \tag{42}$$

If all $\pi_{\mathcal{C}(n),k(j,\kappa)} \leq 0$, then $\eta_{nj}^+ = 0 \leq \sigma_n^{\max}$. Otherwise, the second maximum in (42) is attained by the first term in the brackets.

We now exploit the bounds in the Lagrangian dual problem (28). We choose $\|\cdot\|^\bullet = \|\cdot\|_{1,w}$, hence for each $m \in \mathcal{C}(n)$, the dual multipliers are bounded by $\|\pi_m\|_{\infty,w} \leq$

σ_m. Recall that by our choice of the weight matrix W, this is equivalent to $|\pi_{mk}| \leq \sigma_m 2^{\kappa(k)-1}\beta_{n,j(k)}$ for all $k = 1, \dots, K_n$. Restricting to some component j, we have $|\pi_{\mathcal{C}(n),k(j,\kappa)}| \leq \sigma_m 2^{\kappa-1}\beta_{nj}$ for all $\kappa = 1, \dots, K_{nj}$, and thus $|\pi_{\mathcal{C}(n),k(j,\kappa)}| \leq \sigma_n^{\max} 2^{\kappa-1}\beta_{nj}$ for all $\kappa = 1, \dots, K_{nj}$. Consequently, in (42) it follows $\eta_{nj}^{+} \leq \sigma_n^{\max}$.

Case 3. η_{nj}^{-} **in the basis.** We can prove $\eta_{nj}^{-} \leq \sigma_n^{\max}$ by using the same reasoning as for Case 2.

Since $\eta_{nj} = \eta_{nj}^{+} - \eta_{nj}^{-}$, but only one of both variables can be non-zero, we conclude $\eta_{nj} \leq \sigma_n^{\max}$. Due to separability, the above reasoning can be applied for each $j = 1, \dots, d_n$ separately, so it follows $\eta_{nj} \leq \sigma_n^{\max}$ for all j. Hence, $\|\eta_n\|_\infty \leq \sigma_n^{\max}$. Note that the above reasoning is completely independent of the values of β_{nj} or K_{nj}. □

K Proof of Lemma 5.13

Proof. The CPC is defined as the minimum of finitely many linear functions, which we enumerate by $\ell = 1, \dots, L$. Each such function $\psi_{\mathcal{C}(n)}^{\ell}(\cdot)$ is determined by an extreme point $(\mu_n^{\ell}, \nu_n^{\ell}, \eta_n^{\ell})$ of (36). Consider two arbitrary points $x_n^1, x_n^2 \in \mathbb{R}_{d_n}$. Using the Hölder inequality and Lemma 5.12, we obtain

$$|\psi_{\mathcal{C}(n)}^{\ell}(x_n^2) - \psi_{\mathcal{C}(n)}^{\ell}(x_n^1)| = |(\eta_n^{\ell})^\top (x_n^2 - x_n^1)| \leq \|\eta_n^{\ell}\|_\infty \|x_n^2 - x_n^1\|_1 \leq \sigma_n^{\max} \|x_n^2 - x_n^1\|_1.$$

Hence, each $\psi_{\mathcal{C}(n)}^{\ell}(\cdot)$ is Lipschitz continuous w.r.t. $\|\cdot\|_1$ with Lipschitz constant $\sigma_n^{\max}$. Taking the maximum of all Lipschitz constants over $\ell = 1, \dots, L$, we obtain a Lipschitz constant for the CPC, which is again $\sigma_n^{\max}$. By equivalence of norms in $\mathbb{R}_{d_n}$, for any other norm than $\|\cdot\|_1$, we can obtain a Lipschitz constant $\tilde{\sigma}_{\mathcal{C}(n)} > 0$ by multiplying $\sigma_n^{\max}$ with an appropriate positive constant. □

L Proof of Theorem 5.14

Proof. We first prove validity property (S1). From Corollary 5.8 we obtain $\mathcal{Q}_{\mathcal{C}(n)|Z}(x_n) \geq \phi_{\mathcal{B};\mathcal{C}(n)|Z}(\lambda_n)$ for all $x_n \in \mathbb{R}^{d_n}$ and any $\lambda_n \in [0,1]^{K_n}$, such that $x_n = \mathcal{B}_n \lambda_n$. Hence

$$\mathcal{Q}_{\mathcal{C}(n)|Z}(x_n) \geq \max_{\lambda_n} \left\{ \phi_{\mathcal{B};\mathcal{C}(n)|Z}(\lambda_n) \ : \ \lambda_n \in [0,1]^{K_n}, \mathcal{B}_n \lambda_n = x_n \right\} = \phi_{\mathcal{C}(n)|Z}(x_n)$$

for all $x_n \in \mathbb{R}^{d_n}$, where the last equality applies the definition of the CPC in (35). Since this result is true for all $\phi_{\mathcal{C}(n)|Z}^{r}(\cdot), r = 1, \dots, i+1$, it also holds for their pointwise maximum $\mathfrak{Q}_{\mathcal{C}(n)|Z}^{i+1}(\cdot)$.

Next, we prove the overestimation property (S2). From the definitions of $x_{\mathcal{B},n}^{i}$ and the CPC, it follows $\phi_{\mathcal{B};\mathcal{C}(n)|Z}^{i}(\lambda_n^{i}) = \phi_{\mathcal{C}(n)|Z}^{i}(x_{\mathcal{B},n}^{i})$. Hence, under our assumptions, Corollary 5.9 yields

$$\phi_{\mathcal{C}(n)|Z}^{i}(x_{\mathcal{B},n}^{i}) \geq \underline{\mathcal{Q}}_{\mathcal{C}(n)|Z}^{R;i+1}(x_{\mathcal{B},n}^{i}; \sigma_{\mathcal{C}(n)} \|\cdot\|^\circ).$$

By definition of $\mathfrak{Q}_{\mathcal{C}(n)|Z}^{i+1}(\cdot)$ in (37) it directly follows

$$\mathfrak{Q}_{\mathcal{C}(n)|Z}^{i+1}(x_{\mathcal{B},n}^{i}) \geq \underline{\mathcal{Q}}_{\mathcal{C}(n)|Z}^{R;i+1}(x_{\mathcal{B},n}^{i}; \sigma_{\mathcal{C}(n)} \|\cdot\|^\circ).$$

Finally, we prove Lipschitz property (S3) using Lemma 5.13. Under our assumptions, for any $\|\cdot\|^\circ$, each $\phi_{\mathcal{C}(n)|Z}^{r}(\cdot), r = 1, \dots, i+1$, is $\tilde{\sigma}_{\mathcal{C}(n)}$-Lipschitz continuous with $\tilde{\sigma}_{\mathcal{C}(n)} >$

0, finite and independent of $\beta_{n,j} \in (0,1), j = 1, \dots, d_n$. The pointwise maximum of Lipschitz continuous functions is Lipschitz continuous with its Lipschitz constant the maximum of the individual constants. Therefore, $\mathfrak{Q}^{i+1}_{\mathcal{C}(n)|Z}(\cdot)$ is $\tilde{\sigma}_{\mathcal{C}(n)}$-Lipschitz continuous w.r.t. $\|\cdot\|^{\circ}$. $\square$