# Benders decomposition with scaled cuts for multistage stochastic mixed-integer programs

Ward Romeijnders* Niels van der Laan†

June 30, 2024

**Abstract**

We consider Benders decomposition algorithms for multistage stochastic mixed-integer programs (SMIPs) with general mixed-integer decision variables at every node in the scenario tree. We derive a hierarchy of convex polyhedral lower bounds for the value functions and expected cost to-go functions in multistage SMIPs using affine parametric cutting planes in extended spaces for the feasible regions in the problem. We improve this hierarchy of convex polyhedral lower bounds using so-called scaled cuts, and moreover we construct a scaled-cut decomposition algorithm that iteratively improves the convex polyhedral lower bounds of the expected cost to-go functions at every node of the scenario tree in such a way that the first-stage lower bound converges uniformly to the convex envelope of the first-stage expected cost to-go function. This is the best convex polyhedral lower bound possible. Our main convergence result depends on novel results for scaled cuts of expectations over lower semi-continuous value functions, and on the analysis of so-called $\delta$-exact scaled cuts.

## 1 Introduction

We consider multistage stochastic mixed-integer programs (SMIPs) with $T$ time stages, where the uncertainty in the problem can be modelled using a scenario tree $\mathcal{T}$. We use $n \in \mathcal{N}$ to denote the nodes of this scenario tree $\mathcal{T}$, and let $t_n$ denote the time stage in which node $n$ is located. Moreover, we let $n = 1$ denote the root node and $\mathcal{L} \subset \mathcal{N}$ the set of leaf nodes of the scenario tree, respectively, and we let $p_n \in [0, 1]$ denote the probability that node $n \in \mathcal{N}$ of the scenario tree is reached. Furthermore, for every $n \in \mathcal{N}$, we let $a(n)$ denote the unique ancestor node of $n$, where we define $a(1) = 0$ and $x_0 = 0$ for notational convenience, and we let $C(n)$ denote the set of children nodes of $n$. Using this notation, the large-scale deterministic formulation of our multistage SMIP is given by

$$\eta^* = \min_{x_n} \Big\{ \sum_{n\in\mathcal{N}} p_n c_n^\top x_n : x_n \in X_n,\ T_n x_{a(n)} + W_n x_n = h_n, \quad \forall\, n \in \mathcal{N} \Big\}. \tag{1}$$

Here, $c_n$ denotes the unit cost vector of the decisions $x_n$ in node $n \in \mathcal{N}$, and $x_n \in X_n$ and $T_n x_{a(n)} + W_n x_n = h_n$ represent the constraints in node $n \in \mathcal{N}$. We use the set $X_n$ to model simple bounds, including non-negativity constraints, and integrality constraints on the decisions $x_n$. Moreover, in what follows we will define $\mathcal{X}_n(x_{a(n)})$ as the set of feasible decisions in node $n$ for every $n \in \mathcal{N}$ and $x_{a(n)} \in X_{a(n)}$. That is, for all $n \in \mathcal{N}$, we define

$$\mathcal{X}_n(x_{a(n)}) = \Big\{ x_n \in X_n : T_n x_{a(n)} + W_n x_n = h_n \Big\}, \qquad x_{a(n)} \in X_{a(n)},$$

*Department of Operations, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherland (w.romeijnders@rug.nl).

†Department of Operations, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands (nielsvanderlaan@hotmail.com).

with the understanding that $X_0 = \{0\}$.

Equivalently, we can define our multistage SMIP in (1) using the following nested formulation, in which we define for every $n \in \mathcal{N}$ a value function $v_n$ and an expected cost to-go function $Q_n$ as

$$v_n(x_{a(n)}) = \min_{x_n \in \mathcal{X}_n(x_{a(n)})} c_n^\top x_n + Q_n(x_n), \qquad x_{a(n)} \in X_{a(n)}, \tag{2}$$

where the expected cost to-go function equals $Q_n \equiv 0$ if $n \in \mathcal{L}$, and

$$Q_n(x_n) = \sum_{m \in C(n)} q_{nm} v_m(x_n), \qquad x_n \in X_n,$$

if $n \in \mathcal{N} \setminus \mathcal{L}$, with $q_{nm}$ denoting the conditional probability of moving from node $n$ to node $m$ in the scenario tree $\mathcal{T}$. Observe that $\eta^* = v_1(x_0)$, that is, the optimal objective value $v_1(x_0)$ of the first-stage value function equals the optimal objective value $\eta^*$ of the large-scale deterministic equivalent formulation in (1).

If all decision variables $x_n$ in our multistage SMIP are continuous, then all value functions $v_n$ and expected cost to-go functions $Q_n$ are convex, see, e.g., [3, 18]. This property has been exploited to derive convex optimization based algorithms for such continuous multistage stochastic programs (SPs), including nested Benders decomposition [2], stochastic dual dynamic programming (SDDP) [10], progressive hedging [13], and variants thereof. The main challenge in solving multistage SMIPs, however, is that contrary to their continuous counterparts, the value functions $v_n$ and expected cost to-go functions $Q_n$ in SMIPs are typically non-convex due to the integer decision variables involved.

A successful approach to deal with these non-convex function in the special case of two-stage SMIPs, i.e., when $T = 2$, is to convexify the second-stage feasible region using cutting planes. In this way, the integrality restrictions on the decision variables $x_n$ can be relaxed, and we obtain a continuous multistage SP which is convex. See, e.g., Sen and Higle [16] for disjunctive cuts, Gade et al. [7] for Gomory cuts, Ntaimo [9] for Fenchel cuts, and [12, 14, 17, 19, 21] for other cutting plane approaches. The challenge, however, with applying cutting planes in a two-stage stochastic setting compared to deterministic MIPs, is that the second-stage feasible regions depend on the first-stage decision $x_1$, which means that the cutting planes need to be parametric in $x_1$ to be able to use them in any practical decomposition algorithm. Additionally, these parametric cutting planes are typically restricted to be *affine* in $x_1$, because this leads to affine optimality cuts in the master problem. As a result, the master problem remains computationally tractable. Affine parametric cutting planes have proven to be effective when the first-stage decision variables are pure binary. For general mixed-integer first-stage variables, however, they are generally not strong enough to obtain tight lower bounds.

Van der Laan and Romeijnders [20] show how to improve upon these affine parametric cutting planes when the first-stage variables are mixed-integer. They do so by constructing nonlinear lower bounds for the non-convex second-stage value functions, and by transforming the resulting nonlinear optimality cut in an affine so-called *scaled cut* for the master problem. Here, the nonlinear lower bounds for the second-stage value functions depend directly on $\hat{Q}_1$, the current convex polyhedral outer approximation, i.e., lower bound, of the first-stage expected cost to-go function $Q_1$. Van der Laan and Romeijnders [20] show that by iteratively improving the current outer approximation $\hat{Q}_1$ using scaled cuts, the outer approximation converges uniformly to the convex envelope of $Q_1$, which is the best possible convex polyhedral lower bound.

In this paper, inspired by their success in the two-stage setting, we consider affine parametric cutting planes and scaled cuts for multistage SMIPs. In particular, we investigate the strength of the convex polyhedral outer approximations of the value functions $v_n$ and expected cost to-go functions $Q_n$ that we may obtain using affine parametric cutting planes for the feasible regions in the model. A surprisingly subtle, but relevant difference between multistage SMIPs and their two-stage counterparts, is that in two-stage SMIPs the domain $\mathcal{X}_1(x_0)$ of the second-stage value functions $v_n(x_1)$, $n \in C(1)$, is fixed, whereas in later time stages of multistage SMIPs, the domain $\mathcal{X}_{a(n)}(x_{a^2(n)})$ of $v_n(x_{a(n)})$ depends on the previous-stage decision $x_{a^2(n)}$. Here, we define $a^2(n)$

as the ancestor node of $a(n)$, and in general for any $r = 1, \ldots, t_n - 1$, we let $a^r(n)$ denote the $r$-th ancestor node of $n$, going back $r$ generations. We will show that although $v_n$ itself does not directly depend on $x_{a^2(n)}$, the convex polyhedral outer approximation of $v_n$ may improve if we allow the affine parametric cutting planes for its feasible region to depend on $x_{a^2(n)}$. In fact, by constructing affine parametric cutting planes in extended spaces, i.e., depending on increasingly many previous-stage decisions $x_{a^2(n)}, \ldots, x_1$, we are able to derive a hierarchy of convex polyhedral outer approximations for the value functions and expected cost to-go functions in the model.

In general, however, using affine parametric cutting planes does not lead to tight outer approximations, similar as for two-stage SMIPs. That is why, we use scaled cuts to improve these outer approximations, similar as van der Laan and Romeijnders [20] for two-stage SMIPs, leading to a hierarchy of scaled cut based convex polyhedral lower bounds for the value functions and expected cost to-go functions in the model. The main additional challenge for multistage SMIPs compared to two-stage SMIPs, however, is that in two-stage SMIPS all value functions correspond to mixed-integer linear programs, whereas for multistage SMIPs, the value function at node $n$ of the scenario tree depends on the non-convex expected cost to-go function $Q_n$, that in turn is lower bounded by a convex polyhedral outer approximation. This implies that if the scaled cuts are applied with the expected cost to-go function $Q_n$ replaced by an outer approximation, then the scaled cuts may become inexact. Nevertheless, we are able to prove that by iteratively applying scaled cuts at different nodes $n \in \mathcal{N}$ of the scenario tree to convex polyhedral outer approximations of $Q_n$, where these scaled cuts depend on the outer approximations corresponding to all ancestor nodes of $n$, we obtain a sequence of outer approximations for which the first-stage outer approximation converges uniformly to the convex envelope of $Q_1$.

To derive our main convergence result, we first extend the scaled cuts results from [20] to expectations over lower semi-continuous (l.s.c.) value functions. Moreover, we introduce the concept of so-called $\delta$-exact scaled cuts, and analyze the performance of iteratively applying such $\delta$-exact scaled cuts. Finally, we derive a direct link between affine parametric cutting planes in extended spaces and scaled cuts.

Summarizing, our main contributions are as follows.

- We derive a hierarchy of convex polyhedral lower bounds of the value functions and expected cost functions in multistage SMIPs using affine parametric cutting planes in extended spaces for the feasible regions in the problem.

- We derive a direct link between affine parametric cutting planes in extended spaces and scaled cuts from [20].

- We extend the scaled cut results from [20] to expectations over l.s.c. value functions. Moreover, we introduce $\delta$-exact scaled cuts and analyze the performance of applying such cuts iteratively.

- We use scaled cuts to improve the hierarchy of affine parametric cutting plane based convex polyhedral lower bounds.

- We construct a scaled-cut decomposition algorithm that maintains a convex polyhedral outer approximation of $Q_n$ in an extended space at each node $n \in \mathcal{N}$ of the scenario tree. We prove that by iteratively applying scaled cuts at different nodes $n \in \mathcal{N}$, we obtain a sequence of convex polyhedral outer approximations for which the first-stage outer approximation converges uniformly to the convex envelope of $Q_1$, which is the best possible convex polyhedral lower bound.

Our scaled cut approach deviates from other work on multistage SMIPs in the following ways. First of all, compared to the stochastic dual dynamic integer programming (SDDiP) algorithm [22], we do not require to have binary state variables only, but allow for general mixed-integer state variables. Moreover, an advantage of our scaled cut approach compared to for example the stochastic Lipschitz dynamic programming algorithm from [1] and the non-convex Benders decomposition algorithm from [6] is that the outer approximations in our approach remain convex polyhedral. Alternative works on multistage SMIPs include the MIDAS algorithm from [11] which applies to multistage SMIPs with monotonic value functions, the dual decomposition algorithm from [4], and Lagrangian dual decision rules from [5]. See also Maggioni and Pflug [8] and Sandikçi and Özaltin [15] for bounds on the optimal objective value of multistage SMIPs.

The remainder of this paper is organized as follows. In Section 2 we discuss preliminary results in the context of two-stage SMIPs. Here, we discuss the potential of applying affine parametric cutting planes in extended spaces, and we derive a link between such cutting planes and scaled cuts from [20]. In Section 3, we extend scaled cut results from [20] to the expectation over l.s.c. value functions, and we introduce and analyze so-called $\delta$-exact scaled cuts. In Section 4, we use affine parametric cutting planes in extended spaces for the feasible region in the problem, to derive a hierarchy of convex polyhedral outer approximations for the value functions and expected cost to-go functions in multistage SMIPs, and in Section 5 we improve this hierarchy using scaled cuts, and we construct a converging scaled-cut decomposition algorithm for SMIPs. We end with a conclusion and discussion in Section 6. Finally, we note that throughout this paper all proofs are postponed to the Appendix.

### 1.1 Notation and assumptions

For any two nodes $n, m \in \mathcal{N}$ with $n = a^r(m)$ for some $r = 1, \ldots, t_n - 1$, we define $x_{[n:m]}$ as $x_{[n:m]} = (x_n, \ldots, x_{a(m)}, x_m)$, i.e., as a vector containing all decisions along the path in the scenario tree $\mathcal{T}$ from node $n$ to $m$. Similarly, we define $X_{[n:m]}$ as $X_{[n:m]} = X_n \times \cdots \times X_{a(m)} \times X_m$, and $S_{[n:m]}$ as

$$S_{[n:m]} = \Big\{ x_{[n:m]} \in X_{[n:m]} : \\ T_{a^{r+1}(m)} x_{a^{r+1}(m)} + W_{a^r(m)} x_{a^r(m)} = h_{a^r(m)}, \quad \forall r = 0, \ldots, t_m - t_n - 1 \Big\}.$$

Moreover, if $n$ equals the root node, i.e., if $n = 1$, then we write $x_{[m]}, X_{[m]}$, and $S_{[m]}$.

Throughout, we let $\text{conv}(X)$ denote the convex hull of a set $X$. Moreover, we let $\delta_X : \text{conv}(X) \to \mathbb{R} \cup \{+\infty\}$ denote the characteristic function of $X$, defined for all $x \in \text{conv}(X)$ as $\delta_X(x) = 0$, if $x \in X$, and $\delta_X(x) = +\infty$, if $x \notin X$. Furthermore, for a function $Q : X \to \mathbb{R}$, we define its closed convex envelope $\text{co}_X(Q) : \text{conv}(X) \to \mathbb{R}$ as the pointwise maximum of all affine lower bounds of $Q$. Instead of $\text{co}_X(Q)$, we may also write $\text{co}(Q)$ if the domain of $Q$ is clear from the context, or $\text{co}(Q + \delta_X)$. With slight abuse of notation, we also write $\text{co}(Q + \delta_\Theta)$ to denote the closed convex envelope of $\hat{Q} : \Theta \to \mathbb{R}$ with $\Theta \subset X \times Y$ and $\hat{Q}(x, y) = Q(x)$ for all $(x, y) \in \Theta$.

Throughout this paper we make the following assumptions.

(i) We assume relatively complete recourse. That is, $-\infty < v_n(x_{a(n)}) < +\infty$ for all $n \in \mathcal{N}$ and $x_{a(n)} \in X_{a(n)}$.

(ii) We assume that all data is rational. That is, $c_n, T_n, W_n$, and $h_n$ are rational for all $n \in \mathcal{N}$.

(iii) We assume that $X_n$ is bounded for all $n \in \mathcal{N}$.

(iv) As a result of (i)–(iii), for every $n \in \mathcal{N}$ there exists an upper bound $U_n \in \mathbb{R}$ such that $Q_n(x_n) \leq U_n$ for all $x_n \in X_n$.

## 2 Preliminaries

In this section we discuss how to obtain convex polyhedral lower bounds for the expectation of mixed-integer value functions using affine parametric cutting planes, extended spaces, and scaled cuts in Sections 2.1–2.3, respectively. We use these results extensively in Sections 4 and 5 to derive our hierarchy of convex polyhedral lower bounds for the expected cost to-go functions $Q_n$. The results in Sections 2.1 and 2.2 can also be found in, e.g., [19], whereas the results in Section 2.3 are new.

## 2.1 Affine parametric cutting planes

Throughout this section we restrict ourselves to a generic expected cost function $\mathcal{Q}(x)$, defined on a bounded mixed-integer polyhedral set $X$, with $\mathcal{Q}(x) := \mathbb{E}_\omega[v_\omega(x)]$, $x \in X$, where

$$v_\omega(x) := \min_{z \in Z_\omega(x)} q_\omega^\top z, \qquad x \in X, \omega \in \Omega, \tag{3}$$

represents a generic mixed-integer value function in which the feasible region $Z_\omega(x) = \{z \in Z : T_\omega x + W_\omega z \geq h_\omega\}$ is a mixed-integer polyhedral set parametrized by $x$.

Since $v_\omega(x)$ is the value function of a mixed-integer linear program for every $x \in X$, we can equivalently solve it by replacing the feasible region $Z_\omega(x)$ by its convex hull $\text{conv}(Z_\omega(x))$. That is,

$$v_\omega(x) := \min_{z \in \text{conv}(Z_\omega(x))} q_\omega^\top z, \qquad x \in X, \omega \in \Omega. \tag{4}$$

The problem with this approach, however, is that the cutting planes with which $Z_\omega(x)$ need to be strengthened to obtain $\text{conv}(Z_\omega(x))$ differ for every $x$, and thus cannot be reused. Hence, instead we may strengthen $Z_\omega(x)$ using *parametric cutting planes* in $x$. However, the form of these parametric cutting planes has significant impact on the computational tractability of the master problem, since non-linear parametric cutting planes for the feasible region of $v_\omega$ may translate to non-linear lower bounds for the expected cost functions $\mathcal{Q}$. A sufficient condition to obtain convex polyhedral lower bounds for $\mathcal{Q}$ is to use parametric cuts for $Z_\omega(x)$ that are *affine* in $x$.

**Definition 1.** *An affine parametric cutting plane $\pi^\top(x, z) \geq \pi_0$ is valid for the feasible region $Z_\omega(x)$ with respect to $X$ if and only if for every $\bar{x} \in X$ and $\bar{z} \in Z_\omega(\bar{x})$ it holds that $\pi^\top(\bar{x}, \bar{z}) \geq \pi_0$.*

An important observation is that any affine parametric cut in $x$ for $Z_\omega(x)$ is also valid for $P_\omega := \{(x, z) \in X \times Z : T_\omega x + W_\omega z \geq h_\omega\}$, and since it is an affine cut, also for $\text{conv}(P_\omega)$.

**Proposition 1.** *Let $\omega \in \Omega$ be given and consider the feasible region in (3) for some $x \in X$. Then. every affine parametric cutting plane for $Z_\omega(x)$ with respect to $X$ is valid for*

$$\text{conv}(P_\omega) = \mathit{conv}\Big\{(x, z) \in X \times Z : T_\omega x + W_\omega z \geq h_\omega\Big\}.$$

Proposition 1 implies that the best lower bound for $v_\omega$ that we may obtain using affine parametric cuts, is to use all of those required to construct $\text{conv}(P_\omega)$. In a practical algorithm, we never intend to add all those cuts. However, from a theoretical perspective we will use this observation to analyze the strength of the best possible lower bound that can be obtained using these affine parametric cuting planes.

**Definition 2.** *Consider the mixed-integer value function $v_\omega$ as defined in (3). Then, we define $\nu_\omega$ as the best possible convex polyhedral lower bound of $v_\omega$ that can be obtained using* affine parametric cutting planes, *given by*

$$\nu_\omega(x) := \min_z \Big\{q_\omega^\top z : (x, z) \in \mathit{conv}(P_\omega)\Big\}, \qquad x \in \text{conv}(X). \tag{5}$$

The function $\nu_\omega$, defined on $\text{conv}(X)$, is a convex polyhedral lower bound of $v_\omega$. However, contrary to the optimization problem in (4), the convexification in $\nu_\omega(x)$ does not necessarily yield the exact optimal objective value $v_\omega(x)$ unless $x$ is a vertex of $\text{conv}(X)$.

**Theorem 1.** *Consider the mixed-integer value function $v_\omega$ as defined in (3), and the convex poyhedral lower bound $\nu_\omega$ from Definition 2. Then,*

*(i) $\nu_\omega(x) \leq v_\omega(x)$ for all $x \in X$, and*

*(ii)* $\nu_\omega(x) = v_\omega(x)$ *if* $x$ *is an extreme point of* $\mathrm{conv}(X)$.

The next example shows that the inequality in Theorem 1 (i) may be strict.

**Example 1.** *Consider the one-dimensional mixed-integer value function*

$$v_\omega(x) = \min_{z \in \mathbb{Z}_+} \Big\{ z : z \geq \omega - 0.5x \Big\},$$

*for* $x \in X := \mathbb{Z}_+ \cap [0,2]$ *and* $\omega = 1$. *Figure 1 shows the set* $P_\omega = \{(x,z) \in \mathbb{Z}_+ \cap [0,2] \times \mathbb{Z}_+ : z \geq \omega - 0.5x\}$ *for* $\omega = 1$, *from which it is not hard to derive the feasible regions* $Z_\omega(x)$ *of* $v_\omega(x)$ *for different values of* $x$. *In particular, for* $x = 1$ *and* $\omega = 1$, *the feasible region* $Z_\omega(x)$ *of* $v_\omega(x)$

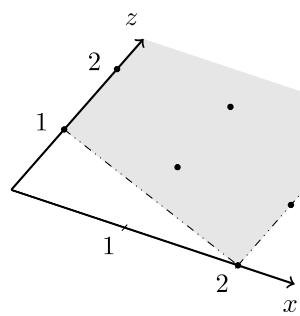


Figure 1: The regions $P_\omega = \{(x,z) \in \mathbb{Z}_+ \cap [0,2] \times \mathbb{Z}_+ : z \geq \omega - 0.5x\}$ and $\mathrm{conv}(P_\omega)$ corresponding to Example 1 for $\omega = 1$. The dots in the figure represent $P_\omega$ and the shaded region corresponds to $\mathrm{conv}(P_\omega)$.

*equals* $Z_\omega(x) = \{z \in \mathbb{Z}_+ : z \geq 0.5\}$. *Hence,* $z^* = 1$ *is the optimal solution in* $v_\omega(x)$, *and thus* $v_\omega(x) = 1$, *for* $x = 1$ *and* $\omega = 1$. *However, for* $\nu_\omega(x)$, *as defined in (5), it holds for* $\omega = 1$ *that* $\mathrm{conv}(P_\omega) = \{(x,z) \in [0,2] \times \mathbb{R}_+ : z \geq 1 - 0.5x\}$, *see Figure 1, and thus for* $x = 1$ *the optimal solution in*

$$\nu_\omega(x) = \min_z \{z : (x,z) \in \mathrm{conv}(P_\omega)\} = \min_{z \in \mathbb{R}_+} \{z : z \geq 1 - 0.5x\},$$

*equals* $z^* = 0.5$ *with corresponding objective value* $\nu_\omega(x) = 0.5$. *Hence, for* $\omega = 1$ *and* $x = 1$, *the convex polyhedral lower bound* $\nu_\omega(x)$ *is strictly smaller than* $v_\omega(x)$. *Note that indeed in this case*

$$\mathrm{conv}\Big(Z_\omega(x)\Big) = [1, \infty) \neq [0.5, \infty) = \Big\{ z \in \mathbb{R} : (x,z) \in \mathrm{conv}(P_\omega) \Big\}.$$

## 2.2 Affine parametric cutting planes in extended spaces

Example 1 shows that the lower bound $\nu_\omega$ is not necessarily tight for all $x \in X$, since affine parametric cutting planes are not strong enough to yield $\mathrm{conv}(Z_\omega(x))$ for all $x \in X$. A surprising way to improve this lower bound is to consider $v_\omega$ on an extended $(x,y)$-space, which we call $\Theta$. This may seem counterintuitive since $y$ itself does not impact the value of $v_\omega(x)$. However, Theorem 1 (ii) guarantees us that if $(\bar{x}, \bar{y})$ is a vertex of $\mathrm{conv}(\Theta)$, then we obtain a tight lower bound of $v_\omega(\bar{x})$ even when $\bar{x}$ is not a vertex of $\mathrm{conv}(X)$.

**Definition 3.** *Let* $X$ *be a bounded mixed-integer polyhedral set. Then, we call* $\Theta$ *an* extended space *of* $X$ *if* $\Theta$ *is of the form*

$$\Theta = \Big\{ (x,y) \in X \times Y : (x,y) \in \mathcal{F} \Big\},$$

*where* $\mathcal{F}$ *is a mixed-integer polyhedral set.*

**Definition 4.** *Consider the mixed-integer value function $v_\omega$ as defined in (3), and let $\Theta$ be an extended space of $X$. Then, we define the best possible convex polyhedral lower bound of $v_\omega$ that can be obtained using* affine parametric cutting planes in the extended space $\Theta$ *as*

$$\hat{\nu}_\omega(x, y) := \min_z \Big\{ q_\omega^\top z : (x, y, z) \in \mathrm{conv}(\mathcal{E}_\omega) \Big\}, \tag{6}$$

*where $\mathcal{E}_\omega := \{(x, y, z) \in \Theta \times Z : (x, z) \in P_\omega\}$.*

Note that by interpreting $v_\omega$ as a mixed-integer value function defined on the extended space $\Theta$ instead of $X$, it follows directly from Theorem 1 (i) that $\hat{\nu}_\omega$, defined on $\mathrm{conv}(\Theta)$, is a convex polyhedral lower bound for $v_\omega$. Moreover, by Theorem 1 (ii) we have that $\hat{\nu}_\omega(x, y) = v_\omega(x)$ if $(x, y)$ is a vertex of $\mathrm{conv}(\Theta)$. Furthermore, it turns out that the lower bound $\hat{\nu}_\omega(x, y)$ in the extended space $\Theta$ is always as least as good as the lower bound $\nu_\omega(x)$ in the original space $X$.

**Theorem 2.** *Consider the mixed-integer value function $v_\omega$ as defined in (3), and let $\Theta$ be an extended space of $X$. Consider the the convex poyhedral lower bounds $\nu_\omega(x)$ and $\hat{\nu}_\omega(x, y)$ from Definitions 2 and 4, respectively. Then,*

*(i)* $\nu_\omega(x) \leq \hat{\nu}_\omega(x, y)$ *for all* $(x, y) \in \mathrm{conv}(\Theta)$,

*(ii)* $\hat{\nu}_\omega(x, y) \leq v_\omega(x)$ *for all* $(x, y) \in \Theta$, *and*

*(iii)* $\hat{\nu}_\omega(x, y) = v_\omega(x)$ *if* $(x, y)$ *is an extreme point of* $\mathrm{conv}(\Theta)$.

Example 2 below shows that the inequality in Theorem 2 (i) may be strict.

**Example 2.** *Consider the mixed-integer value function $v_\omega(x)$ from Example 1, and consider the extended space*

$$\Theta = \Big\{ (x, y) \in X \times \mathbb{R} : y \geq |1 - x| \Big\}.$$

*Then, the convex polyhedral lower bound from Definition 4 is given by $\hat{\nu}_\omega(x, y) := \min_z \Big\{ z : (x, y, z) \in \mathrm{conv}(\mathcal{E}_\omega) \Big\}$, where $\mathcal{E}_\omega := \{(x, y, z) \in \Theta \times \mathbb{Z}_+ : z \geq 1 - 0.5x\}$. Figure 2 shows the sets $\mathcal{E}_\omega$ and $\mathrm{conv}(\mathcal{E}_\omega)$ for $\omega = 1$.*

*Based on the right figure in Figure 2, we conclude that for $(x, y) = (1, 0)$ and $\omega = 1$, we have that $\hat{\nu}_\omega(x, y) = 1$, and thus $\hat{\nu}_\omega(x, y)$ is a tight lower bound for $v_\omega(x)$. Observe that all $z$-solutions in $[0.5, 1)$ are cut away by affine parametric cutting planes in the extended $(x, y)$-space for $(x, y) = (1, 0)$, whereas this was not possible for $x = 1$ in the orginal $x$-space, see Example 1. Hence, in this case we have $\nu_\omega(x) < \hat{\nu}_\omega(x, y)$.*

### 2.2.1 Monotone extended spaces

In some cases, extended spaced exhibit a certain special structure, which we refer to as *monotone extended spaces.*

**Definition 5.** *We call an extended space $\Theta = \Big\{ (x, y) \in X \times Y : (x, y) \in \mathcal{F} \Big\}$* a monotone extended space *of $X$ if and only if*

*(i)* $y^1 \in Y$ *and* $y^2 \geq y^1 \Rightarrow y^2 \in Y$, *and*

*(ii) for every* $y^1, y^2 \in Y$ *with* $y^1 \leq y^2$, *it holds that*

$$\Big\{ x \in X : (x, y^1) \in \mathcal{F} \Big\} \subseteq \Big\{ x \in X : (x, y^2) \in \mathcal{F} \Big\},$$

*with the possibility that one of these two sets, or both, are empty.*

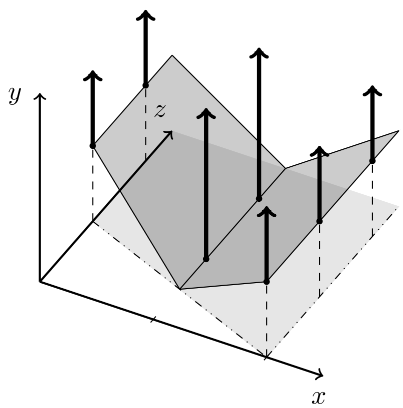


Figure 2: The right figure displays the regions $\mathcal{E}_\omega$ and $\text{conv}(\mathcal{E}_\omega)$ corresponding to Example 2 for $\omega = 1$. The black arrows represent $\mathcal{E}_\omega$ and the entire area above the shaded region represents $\text{conv}(\mathcal{E}_\omega)$. The left figure is added to show that an additional cutting plane $x + y + 2z \geq 3$ in the $(x, y, z)$-space is necessary to obtain $\text{conv}(\mathcal{E}_\omega)$ compared to the original constraints $y \geq |1 - x|$ and $z \geq 1 - 0.5x$. This additional cutting plane is left out in the left figure.

Obviously, the smaller $y$, the smaller the set $\{x \in X : (x, y) \in \mathcal{F}\}$, and the better the corresponding lower bound $\hat{\nu}_\omega(x, y)$.

**Theorem 3.** *Let $\Theta$ be a monotone extended space of $X$, and consider the convex polyhedral lower bound $\hat{v}_\omega$ from Definition 4 for some $\omega \in \Omega$. Then, for every $x \in \text{conv}(X)$ and $(x, y^1), (x, y^2) \in \text{conv}(\Theta)$ with $y^1 \leq y^2$, it holds that*

$$\hat{\nu}_\omega(x, y^1) \geq \hat{\nu}_\omega(x, y^2).$$

An example of a monotone extended space $\Theta$ is the extended space

$$\Theta = \Big\{(x, \theta) \in X \times \mathbb{R} : \theta \geq \varphi(x)\Big\},$$

where $\varphi : \text{conv}(X) \to \mathbb{R}$ is a convex polyhedral function. For this extended space it follows from Theorem 3 that $\hat{\nu}_\omega(x, \theta^1) \geq \hat{\nu}_\omega(x, \theta^2)$ if $\theta^1 \leq \theta^2$. In other words, the lower bound $\hat{\nu}_\omega(x, \theta)$ improves if $\theta$ decreases. Since $\hat{\nu}_\omega(x, \theta) = +\infty$ if $\theta < \varphi(x)$, we conclude that $\hat{\nu}_\omega(x, \varphi(x)) \geq \hat{\nu}_\omega(x, \theta)$ for all $(x, \theta) \in \text{conv}(\Theta)$. That is, $\hat{\nu}_\omega(x, \varphi(x))$ is the best lower bound that we may deduce from the lower bound $\hat{\nu}_\omega(x, \theta)$ based on the extended space $\Theta$, since this lower bound is best if $\theta$ is small. The larger $\theta$, however, the worse the lower bound.

**Lemma 1.** *Consider the monotone extended space*

$$\Theta = \Big\{(x, \theta) \in X \times \mathbb{R} : \theta \geq \varphi(x)\Big\}$$

*for some convex polyhedral function $\varphi : \text{conv}(X) \to \mathbb{R}$, and consider the corresponding convex polyhedral lower bound $\hat{\nu}_\omega$ from Definition 4 for some $\omega \in \Omega$. Then, there exists $\theta^* \in \mathbb{R}$ such that $\hat{\nu}_\omega(x, \theta) = \nu_\omega(x)$ for all $x \in \text{conv}(X)$ and $\theta \geq \theta^*$.*

Lemma 1 shows that the lower bound $\hat{\nu}_\omega(x, \theta)$ defined on the extended space $\text{conv}(\Theta)$ does not improve the lower bound $\nu_\omega(x)$ defined on the original space $\text{conv}(X)$ if $\theta$ is too large. Intuitively, this makes sense, since if $\theta \geq \sup_{x \in \text{conv}(X)} \varphi(x)$, then knowledge of $\theta$ does not restrict the set of feasible solutions $x$ in any way.

It turns out that there is a direct relation between the type of monotone extended spaces from Lemma 1 and nonlinear second-stage cuts that depend on $\varphi$, as Theorem 4 illustrates.

**Definition 6.** *Consider the mixed-integer value function $v_\omega$ from (3), and let $\varphi : \mathrm{conv}(X) \to \mathbb{R}$ be a convex polyhedral function. Then, we define $\Pi_\omega(\varphi)$ as the set of cut coefficients $(\alpha_\omega, \beta_\omega, \tau_\omega)$ with $\tau_\omega \geq 0$ such that the nonlinear function $\alpha_\omega - \beta_\omega^\top x - \tau_\omega \varphi(x)$ is a lower bound for $v_\omega$ on $X$. That is,*

$$\Pi_\omega(\varphi) = \Big\{(\alpha_\omega, \beta_\omega, \tau_\omega) : v_\omega(x) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega \varphi(x) \ \ \forall x \in X, \tau_\omega \geq 0\Big\}.$$

**Theorem 4.** *Consider the monotone extended space $\Theta = \{(x, \theta) \in X \times \mathbb{R} : \theta \geq \varphi(x)\}$ for some convex polyhedral function $\varphi : \mathrm{conv}(X) \to \mathbb{R}$. Then, for $(x, \theta) \in \mathrm{conv}(\Theta)$, the convex polyhedral lower bound from Definition 4 is given by*

$$\hat{\nu}_\omega(x, \theta) = \sup_{(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\varphi)} \alpha_\omega - \beta_\omega^\top x - \tau_\omega \theta.$$

Theorem 4 implies that for monotone extended spaces of the form $\Theta = \{(x, \theta) \in X \times \mathbb{R} : \theta \geq \varphi(x)\}$ for some convex polyhedral function $\varphi$, the convex polyhedral lower bound $\hat{\nu}_\omega(x, \theta)$ from Definition 4, obtained using affine parametric cuts in the extended space $\Theta$, can be equivalently expressed as the pointwise maximum of affine functions in $(x, \theta)$ depending on the cut coefficients $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\varphi)$.

## 2.3 Scaled cuts

We can use our analysis from the previous sections to derive a lower bound for the expected value function $\mathcal{Q}$, since for every extended space $\Theta$ of $X$, it holds that

$$\mathcal{Q}(x) \geq \mathbb{E}_\omega[\hat{\nu}_\omega(x, \theta)], \qquad \forall (x, \theta) \in \Theta.$$

This lower bound $\hat{\mathcal{Q}}$, defined as $\hat{\mathcal{Q}}(x, \theta) := \mathbb{E}_\omega[\hat{\nu}_\omega(x, \theta)]$ for $(x, \theta) \in \mathrm{conv}(\Theta)$, is convex in $(x, \theta)$ on $\mathrm{conv}(\Theta)$, but not necessarily in $x$ on $\mathrm{conv}(X)$ when we select $\theta := \theta(x)$ as a function of $x$. We will show in this section that for monotone extended spaces of the form

$$\Theta = \Big\{(x, \theta) \in X \times \mathbb{R} : \theta \geq \varphi(x)\Big\},$$

where $\varphi : \mathrm{conv}(X) \to \mathbb{R}$ is a convex polyhedral function *and a lower bound of $\mathcal{Q}$ with $\varphi(x) \leq \mathcal{Q}(x)$ for all $x \in X$*, it is possible to select $\theta(x)$ such that the resulting function $\hat{\mathcal{Q}}(x) = \mathbb{E}_\omega[\hat{\nu}_\omega(x, \theta(x))]$ is convex in $x$ and at least as good a lower bound of $Q$ as $\varphi$.

The intuition behind the selection of $\theta(x)$ is as follows. Theorem 3 implies that the best possible lower bound of $\mathcal{Q}(x)$ is obtained by selecting $\theta(x) = \varphi(x)$, $x \in \mathrm{conv}(X)$, yielding $\hat{\mathcal{Q}}(x) = \mathbb{E}_\omega[\hat{\nu}_\omega(x, \varphi(x))]$. It can be shown that $\mathbb{E}_\omega[\hat{\nu}_\omega(x, \varphi(x))] \geq \varphi(x)$ for all $x \in \mathrm{conv}(X)$. On the other hand, Lemma 1 shows that selecting $\theta(x)$ large enough, i.e., $\theta(x)$ constant and equal to $\sup_{\tilde{x} \in X} \mathcal{Q}(\tilde{x})$, yields $\hat{\mathcal{Q}}(x) = \mathbb{E}_\omega[\nu_\omega(x)], x \in \mathrm{conv}(X)$. The first lower bound is typically nonlinear and non-convex, whereas the second may not be strong enough, see Examples 1 and 2. However, these bounds do show that for a given $x \in \mathrm{conv}(X)$, we have $\theta < \mathbb{E}_\omega[\hat{\nu}_\omega(x, \theta)]$ when $\theta = \varphi(x)$, and $\theta > \mathbb{E}_\omega[\hat{\nu}_\omega(x, \theta)]$ for $\theta$ large enough. Since $\mathbb{E}_\omega[\hat{\nu}_\omega(x, \theta)]$ is non-increasing and continuous in $\theta$, it follows that there exists $\varphi(x) \leq \bar{\theta} \leq \mathcal{Q}(x)$ such that $\bar{\theta} = \mathbb{E}_\omega[\hat{\nu}_\omega(x, \bar{\theta})]$. Thus, selecting $\theta(x) \geq \varphi(x)$ such that

$$\theta(x) = \mathbb{E}_\omega\Big[\hat{\nu}_\omega(x, \theta(x))\Big], \quad x \in \mathrm{conv}(X), \tag{7}$$

guarantees that $\hat{\mathcal{Q}}(x) = \theta(x)$ is a lower bound of $\mathcal{Q}(x)$. Note that the solution $\theta(x)$ (7) can be determined without knowing $\mathcal{Q}(x)$. Surprisingly, this solution $\theta(x)$ turns out to be convex in $x$.

**Theorem 5.** *Consider the expected cost function $\mathcal{Q}(x) := \mathbb{E}_\omega[v_\omega(x)], x \in X$, with $X$ a bounded mixed-integer polyhedral set, and $v_\omega(x)$ as defined in (3). Let $\varphi : \mathrm{conv}(X) \to \mathbb{R}$ denote a convex polyhedral lower bound of $\mathcal{Q}$. Then, for $\theta(x)$, defined as*

$$\theta(x) := \sup_{(\alpha_\omega, \beta_\omega, \tau_\omega)} \left\{ \frac{\mathbb{E}_\omega[\alpha_\omega] - \mathbb{E}_\omega[\beta_\omega]^\top x}{1 + \mathbb{E}_\omega[\tau_\omega]} : (\alpha_w, \beta_\omega, \tau_\omega) \in \Pi_\omega(\varphi) \ \forall \omega \in \Omega \right\}, \tag{8}$$

*it holds that $\theta(x) = \mathbb{E}_\omega[\hat{\nu}_\omega(x, \theta(x))]$ for all $x \in \mathrm{conv}(X)$.*

Theorem 5 provides a direct link between affine parametric cutting planes in extended spaces and scaled cuts from [20], since $\theta(x)$ corresponds to the scaled cut closure of $\varphi(x)$ as defined in [20]. This scaled cut closure is the pointwise maximum of affine functions, and hence convex. Moreover, each such affine function, defined by a set of feasible cut coefficients $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\varphi)$ for all $\omega \in \Omega$, corresponds to a scaled cut of the form

$$\frac{\mathbb{E}_\omega[\alpha_\omega] - \mathbb{E}_\omega[\beta_\omega]^\top x}{1 + \mathbb{E}_\omega[\tau_\omega]}.$$

In Section 3 we analyze such scaled cuts in detail for the more general setting of the expectation of l.s.c. value functions. This is a generalization of the setting in this section since every mixed-integer value function is l.s.c.; this generalization is necessary for us to develop scaled cuts in a multistage context later.

# 3 Scaled cuts for the expectation of l.s.c. value functions

Let $Q$ denote the expected value function

$$Q(x) = \mathbb{E}_\omega\Big[v_\omega(x)\Big], \qquad x \in X, \tag{9}$$

where $v_\omega(\cdot)$ is an l.s.c. value function defined on a bounded mixed-integer polyhedral set $X$ for every $\omega \in \Omega$. We are interested in deriving convex polyhedral lower bounds $\varphi(x)$ of $Q(x)$ using scaled cuts. In fact, we can derive scaled cuts for $Q$ defined in (9) in a similar fashion as in Theorem 5 by using cut coefficients $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\varphi)$ for all $\omega \in \Omega$. For all such feasible cut coefficients, it holds that $v_\omega(x) + \tau_\omega \varphi(x) \geq \alpha_\omega - \beta_\omega^\top x$ for all $x \in X$ and $\omega \in \Omega$, and thus for all $x \in X$,

$$\Big(1 + \mathbb{E}_\omega[\tau_\omega]\Big) Q(x) \geq Q(x) + \mathbb{E}_\omega[\tau_\omega]\varphi(x) \geq \mathbb{E}_\omega[\alpha_\omega] - \mathbb{E}_\omega[\beta_\omega]^\top x. \tag{10}$$

Here, the first inequality holds since $\tau_\omega \geq 0$ for all $\omega \in \Omega$ and $\varphi$ is a lower bound for $Q$. By dividing the inequalities in (10) by $1 + \mathbb{E}_\omega[\tau_\omega]$, we directly derive a scaled cut as the right-hand side of the inequality

$$Q(x) \geq \frac{\mathbb{E}_\omega[\alpha_\omega] - \mathbb{E}_\omega[\beta_\omega]^\top x}{1 + \mathbb{E}_\omega[\tau_\omega]}, \qquad \forall\, x \in X. \tag{11}$$

Obviously, selecting different feasible cut coefficients $(\alpha_\omega, \beta_\omega, \tau_\omega)$ potentially leads to different scaled cuts in (11). Van der Laan and Romeijnders [20] define the *dominating scaled cut at $\bar{x} \in X$* as the scaled cut that is as large as possible at $\bar{x}$. Moreover, they define the *scaled cut closure* as the pointwise maximum over all possible scaled cuts.

**Definition 7.** *Let $\varphi : \mathrm{conv}(X) \to \mathbb{R}$ denote a convex polyhedral lower bound of $Q$. Then, we define the scaled cut closure $SCC(\varphi)$ of $Q$ with respect to $\varphi$ as*

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

*for all $x \in \mathrm{conv}(X)$.*

Van der Laan and Romeijnders [20] analyze this scaled cut closure operator in detail when $Q$ is the expected value function of a two-stage stochastic mixed-integer program. However, we observe that their proofs directly apply to the expectation of l.s.c. value functions. Hence, based on their results we directly conclude that the SCC operator from Definition 7 is monotone non-decreasing and Lipschitz continuous with Lipschitz constant 1.

**Proposition 2.** *Let $\varphi$ and $\varphi'$ denote convex polyhedral lower bounds of the expectation $Q$ of l.s.c. value functions, defined in (9), and consider the scaled cut closure operation defined in Definition 7. Then,*

*(i)* $SCC(\varphi) \geq \varphi$*,*

*(ii)* $SCC(\varphi) \geq SCC(\varphi')$ *if* $\varphi \geq \varphi'$*, and*

*(iii)* $\|SCC(\varphi) - SCC(\varphi')\|_\infty \leq \|\varphi - \varphi'\|_\infty$*.*

*Proof.* See [20]. □

Moreover, the main convergence theorem of [20] also holds, which means that iteratively applying the scaled cut closure operation to any initial convex polyhedral lower bound $\varphi^0$ of $Q$, converges uniformly to $\mathrm{co}_X(Q)$, the convex envelope of $Q$ with respect to $X$.

**Theorem 6.** *Let $Q : X \to \mathbb{R}$ be defined as the expectation $Q(x) = \mathbb{E}_\omega[v_\omega(x)]$ of* l.s.c. *value functions $v_\omega : X \to \mathbb{R}$, and let $\varphi^0 : \mathrm{conv}(X) \to \mathbb{R}$ be an initial convex polyhedral lower bound of $Q$. Then, the sequence $\varphi^k, k \in \mathbb{N}$, defined as*

$$\varphi^k := SCC(\varphi^{k-1}), \quad k \in \mathbb{N},$$

*converges uniformly to $\mathrm{co}_X(Q)$, i.e., the convex envelope of $Q$ with respect to $X$.*

*Proof.* See [20]. □

To be able to apply this convergence result to multistage stochastic mixed-integer programs, however, we need several additional results regarding the scaled cut closure. The first, i.e., Proposition 3 below, shows that instead of knowing the exact functions $v_\omega$ it suffices to have a sufficiently good lower bound for $v_\omega$ in an extended space. The remaining results show that even if we use a lower bound that is close but not sufficiently large, then we obtain inexact scaled cut closures that are close to the exact one.

**Proposition 3.** *Let $\varphi : \mathrm{conv}(X) \to \mathbb{R}$ denote a convex polyhedral lower bound of the expectation $Q$ of l.s.c. value functions, defined in (9), let $\Theta = \{(x, \theta) \in X \times \mathbb{R} : \theta \geq \varphi(x)\}$, and consider the set of feasible cut coefficients $\Pi_\omega(\varphi)$ as defined in Definition 6 for some $\omega \in \Omega$. Then, for all $\hat{v}_\omega : \Theta \to \mathbb{R}$ such that*

$$co(v_\omega + \delta_\Theta)(x, \theta) \leq \hat{v}_\omega(x, \theta) \leq v_\omega(x), \quad \forall (x, \theta) \in \Theta, \tag{12}$$

*it holds that $\Pi_\omega(\varphi)$ is equivalent to*

$$\Pi_\omega(\varphi) = \Big\{(\alpha_\omega, \beta_\omega, \tau_\omega) : \hat{v}_\omega(x, \theta) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega \theta,\ \forall (x, \theta) \in \Theta,\ \tau_\omega \geq 0\Big\}. \tag{13}$$

**Remark 1.** *Note that Proposition 3 also holds if inequality (12) is true for all $x \in X$ and $\varphi(x) \leq \theta \leq U$, with $U := \sup_{x \in X} \varphi(x) < +\infty$.*

Proposition 3 shows that the scaled cut closure $SCC(\varphi)$ of $Q$ may be computed using value functions $\hat{v}_\omega(x, \theta)$ in an extended space that are at least as large as the convex envelope of $v_\omega + \delta_\Theta$. If worse lower bounds $\hat{v}_\omega(x, \theta)$ of $v_\omega$ are used, then the corresponding scaled cut closure operation becomes inexact.

**Definition 8.** *Let $\varphi : \mathrm{conv}(X) \to \mathbb{R}$ denote a convex polyhedral lower bound of the expectation $Q$ of l.s.c. value functions, defined in (9), and let $\tilde{v}_\omega(x,\theta)$ be a lower bound of $v_\omega$ on the extended space $\Theta = \{(x,\theta) \in X \times \mathbb{R} : \theta \geq \varphi(x)\}$ for every $\omega \in \Omega$. Then, we define the corresponding* inexact scaled cut closure $\tilde{SCC}(\varphi)$ *as*

$$\tilde{SCC}(\varphi) = \sup \left\{ \frac{\mathbb{E}_\omega[\alpha_\omega] - \mathbb{E}_\omega[\beta_\omega]^\top x}{1 + \mathbb{E}_\omega[\tau_\omega]} : (\alpha_\omega, \beta_\omega, \tau_\omega) \in \tilde{\Pi}_\omega(\varphi), \quad \forall \omega \in \Omega \right\},$$

*with*

$$\tilde{\Pi}_\omega(\varphi) = \Big\{ (\alpha_\omega, \beta_\omega, \tau_\omega) : \tilde{v}_\omega(x,\theta) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega \theta, \ \forall (x,\theta) \in \Theta, \tau_\omega \geq 0 \Big\}. \tag{14}$$

*The scaled cut closure $\tilde{SCC}(\varphi)$ is called $\delta$-exact if $\|SCC(\varphi) - \tilde{SCC}(\varphi)\|_\infty \leq \delta$.*

**Remark 2.** *Note that the inexact scaled cut closure operation $\tilde{SCC}$ depends on the lower bounds $\tilde{v}_\omega$ on the extended space $\Theta$ of $X$. For notational convenience we will suppress this dependence in our notation.*

A sufficient condition for the inexact scaled cut closure operation $\tilde{SCC}$ to be $\delta$-exact is that every lower bound $\tilde{v}_\omega$, $\omega \in \Omega$, is close enough to $\mathrm{co}(v_\omega + \delta_\Theta)$.

**Theorem 7.** *Let $\varphi : \mathrm{conv}(X) \to \mathbb{R}$ denote a convex polyhedral lower bound of the expectation $Q$ of l.s.c. value functions, defined in (9), and consider lower bounds $\tilde{v}_\omega(x,\theta)$ of $v_\omega$ on the extended space $\Theta = \{(x,\theta) \in X \times \mathbb{R} : \theta \geq \varphi(x)\}$, such that for every $\omega \in \Omega$,*

$$\sup_{(x,\theta) \in \Theta} \Big\{ co(v_\omega + \delta_\Theta)(x,\theta) - \tilde{v}_\omega(x,\theta) \Big\} \leq \delta. \tag{15}$$

*Then, the corresponding inexact scaled cut operation $\tilde{SCC}$ as defined in Definition 8 is $\delta$-exact.*

In the final result of this section, we show how the error of an inexact scaled cut closure operation propagates when inexact scaled cut closures are computed iteratively.

**Corollary 1.** *Let $\varphi : \mathrm{conv}(X) \to \mathbb{R}$ denote a convex polyhedral lower bound of the expectation $Q$ of l.s.c. value functions, defined in (9), and consider sequences $\{\bar{\varphi}^k\}_{k \in \mathbb{N}}$ and $\{\hat{\varphi}^k\}_{k \in \mathbb{N}}$, obtained by exact and inexact SCC-operations, respectively, where $\bar{\varphi}^0 = \hat{\varphi}^0 = \varphi$, $\bar{\varphi}^k := SCC(\bar{\varphi}^{k-1}), k \in \mathbb{N}$, and $\hat{\varphi}^k := \tilde{SCC}(\hat{\varphi}^{k-1}), k \in \mathbb{N}$. If the $k$-th inexact SCC operation, with corresponding lower bounds $\tilde{v}^k_\omega(x,\theta)$, is $\delta_k$-exact for every $k \in \mathbb{N}$, then for every $K \in \mathbb{N}$, we have*

$$\|\bar{\varphi}^k - \hat{\varphi}^k\|_\infty \leq \sum_{k=1}^{K} \delta_k.$$

# 4 Affine parametric cutting planes for multistage SMIPs

In this section we discuss the strength of affine cutting planes for multistage SMIPs. That is, we construct convex polyhedral lower bounds for the value functions $v_n$, $n \in \mathcal{N}$, defined in (2), using affine parametric cutting planes for the feasible regions involved. By constructing these cutting planes in increasingly larger extended spaces, we derive a hierarchy of convex polyhedral lower bounds. We first derive this hierarchy for value functions corresponding to the leaf nodes in Section 4.1, and next for value functions in general in Section 4.2.

## 4.1 Leaf nodes

For leaf nodes $m \in \mathcal{L}$ of the scenario tree $\mathcal{N}$, the expected cost to-go function $Q_m$ equals zero, so that the value function $v_m(x_n)$ with $n = a(m)$ is given by

$$v_m(x_n) = \min_{x_m \in \mathcal{X}_m(x_n)} c_m^\top x_m,$$

for every $x_n \in X_n$. Thus, $v_m$ is the value function of a mixed-integer linear program, and of a similar form as the value functions $v_\omega$, $\omega \in \Omega$, considered in Section 2. Hence, we may strengthen the feasible region $\mathcal{X}_m(x_n)$ with affine parametric cutting planes in $x_n$. From Proposition 1 it follows that any affine parametric cutting plane in $x_n$ for $\mathcal{X}_m(x_n)$ is also valid for $\mathrm{conv}(S_{[n:m]})$, where $S_{[n:m]}$ is defined as

$$S_{[n:m]} := \Big\{(x_n, x_m) \in X_n \times X_m : x_m \in \mathcal{X}_m(x_n)\Big\}.$$

We conclude that the best lower bound for $v_m$ that we may obtain using affine parametric cutting planes in $X_n$ is to use all of those required to construct $\mathrm{conv}(S_{[n:m]})$. We do not intend to add all these cutting planes in a practical algorithm. However, from a theoretical perspective we will use this observation to analyze the strength of the best possible lower bound that can be obtained using affine parametric cutting planes. That is, we are interested in the strength of the lower bound

$$\nu_m(x_n) := \min_{x_m} \Big\{c_m^\top x_m : x_{[n:m]} \in \mathrm{conv}(S_{[n:m]})\Big\}, \qquad x_n \in \mathrm{conv}(X_n). \tag{16}$$

This lower bound is similar in nature to $\nu_\omega$, defined in Definition 2, however, there is a subtle, yet crucial difference: contrary to its two-stage counterpart $v_\omega$, the domain of $v_m$, i.e., the set of possible input values $x_n$, may depend on the decisions $x_{a(n)}$ made in an earlier time stage. Thus, interestingly, even though knowing $x_{a(n)}$ does not impact the value of $v_m(x_n)$, it may improve the quality of the lower bound $\nu_m(x_n)$ defined in (16), since the affine parametric cutting planes for $\mathcal{X}_m(x_n)$ only need to be valid for $X_n \cap \mathcal{X}_n(x_{a(n)})$, and not for $X_n$.

**Remark 3.** *It is even possible to improve the lower bound in (16) without knowledge of $x_{a(n)}$ if some $x_n \in X_n$ are never feasible for any $x_{a(n)} \in X_{a(n)}$, i.e., if $\cup_{x_{a(n)} \in X_{a(n)}} \mathcal{X}_n(x_{a(n)}) \subset X_n$. Moreover, taking more previous-stage constraints into account may reduce the set of feasible $x_n$, and thus improve the lower bound in (16). However, for ease of exposition we will assume that such improvements are not possible, i.e., we assume that*

$$X_n \subseteq \bigcup_{x_1 \in X_1} \cdots \bigcup_{x_{a(n)} \in \mathcal{X}_{a(n)}(x_{a^2(n)})} \mathcal{X}_n(x_{a(n)}).$$

The challenge, however, with using $x_{a(n)}$ to improve the lower bound $\nu_m$ in (16) is that we do not want to construct affine parametric cutting planes that cannot be reused for different values of $x_{a(n)}$. That is why instead we use affine parametric cutting planes in $(x_{a(n)}, x_n)$ that are valid for all $x_{a(n)} \in X_{a(n)}$. By constructing affine parametric cutting planes involving increasingly many previous-stage decisions, we derive a hierarchy of lower bounds.

**Definition 9.** *Consider a leaf node $m \in \mathcal{L}$ of the scenario tree $\mathcal{N}$. Then, for every $r = 1, \dots, T-1$, we let $n$ denote $n = a^r(m)$, and define $\bar{\nu}_m^r$ for all $x_{[n:a(m)]} \in \mathrm{conv}(S_{[n:a(m)]})$ as*

$$\bar{\nu}_m^r(x_{[n:a(m)]}) = \min_{x_m} \Big\{c_m^\top x_m : x_{[n:m]} \in conv(S_{[n:m]})\Big\}.$$

The convex polyhedral lower bounds of Definition 9 have an interpretation in terms of affine parametric cutting planes defined on extended spaces. In fact, for every $r = 1, \dots, T-1$, the lower bound $\bar{\nu}_m^r$ corresponds to the lower bound of Definition 4 using the extended space $S_{[a^r(m):a(m)]}$ of $X_{a(m)}$. Interestingly, if $r_1 > r_2$, then for $n_1 := a^{r_1}(m)$ and $n_2 := a^{r_2}(m)$, it holds that $S_{[n_1:a(m)]}$ is an extended space of $S_{[n_2:a(m)]}$. Hence, applying Theorem 2 repeatedly yields the following hierarchy of lower bounds.

**Theorem 8.** *Consider a leaf node $m \in \mathcal{L}$ of the scenario tree $\mathcal{N}$. Then, for every $r = 1, \ldots, T-1$ and for all $x_{[a(m)]} \in S_{[a(m)]}$, we have*

$$\bar{\nu}_m^r(x_{[a^r(m):a(m)]}) \leq v_m(x_{a(m)}).$$

*Moreover, for every $x_{[a(m)]} \in \operatorname{conv}(S_{[a(m)]})$, it holds that*

$$\bar{\nu}_m^1(x_{a(m)}) \leq \bar{\nu}_m^2(x_{[a^2(m):a(m)]}) \leq \ldots \leq \bar{\nu}_m^{T-1}(x_{[a(m)]}).$$

Theorem 8 shows that the strength of the lower bounds $\bar{\nu}_m^r$ improves when affine parametric cutting planes can be made in extended spaces involving more previous-stage decisions. However, even for the best lower bound in which we involve all decisions up to and including the root node of the scenario tree, the resulting lower bound $\bar{\nu}_m^{T-1}$ of $v_m$ is not necessarily tight.

Proposition 4 provides a sufficient condition for the lower bounds to be tight.

**Proposition 4.** *Consider a leaf node $m \in \mathcal{L}$ of the scenario tree, and let $\bar{\nu}_m^r$, $r = 1, \ldots, T-1$, denote the convex polyhedral lower bounds of $v_m$ from Definition 9. Then, for every $x_{[a(m)]} \in S_{[a(m)]}$, we have that*

$$\bar{\nu}_m^r(x_{[a^r(m):a(m)]}) = v_m(x_{a(m)}) \tag{17}$$

*if $x_{[a^\rho(m):a(m)]}$ is an exteme point of $\operatorname{conv}(S_{[a^\rho(m):a(m)]})$ for some $\rho = 1, \ldots, r$.*

## 4.2 Non-leaf nodes

For non-leaf nodes $n \in \mathcal{N} \backslash \mathcal{L}$ of the scenario tree, the expected cost to-go function $Q_n$ is typically non-zero, so that the epi-graph formulation of the value function $v_n(x_{a(n)})$ is given by

$$v_n(x_{a(n)}) = \min_{x_n \in \mathcal{X}_n(x_{a(n)}), \theta_n \in \mathbb{R}} \Big\{ c_n^\top x_n + \theta_n : \ \theta_n \geq Q_n(x_n) \Big\}. \tag{18}$$

However, in algorithms that only utilize affine parametric cutting planes for the feasible regions of the mixed-integer value functions involved, we do not know the exact expected cost to-go function $Q_n$, but only have a convex polyhedral lower bound of it based on convex polyhedral lower bounds for the value functions $v_m(x_n)$, $m \in C(n)$, derived using affine parametric cutting planes. In particular, for nodes $n \in \mathcal{N}$ with $t_n = T-1$, i.e., if all child nodes of $n$ are leaf nodes, then the expected cost to-go function $Q_n$ can be lower bounded by taking the expectation over convex polyhedral lower bounds $\bar{\nu}_m^r, m \in C(n)$, from Definition 9. That is, for every $r = 1, \ldots, T-1$,

$$\bar{Q}_n^r(x_{[a^{r-1}(n):n]}) := \sum_{m \in C(n)} q_{nm} \bar{\nu}_m^r(x_{[a^{r-1}(n):n]}),$$

is a lower bound of $Q_n(x_n)$ for every $x_{[a^{r-1}(n):n]} \in S_{[a^{r-1}(n):n]}$, with the understanding that $a^0(n) = n$. We define the corresponding lower bound $\hat{v}_n^r$ of $v_n$ as

$$\hat{v}_n^r(x_{[a^p(n):a(n)]}) = \min_{x_n \in \mathcal{X}_n(x_{a(n)}), \theta_n \in \mathbb{R}} \Big\{ c_n^\top x_n + \theta_n : \ \theta_n \geq \bar{Q}_n^r(x_{[a^{r-1}(n):n]}) \Big\}, \tag{19}$$

for every $x_{[a^p(n):a(n)]} \in S_{[a^p(n):a(n)]}$ with $p := \max\{1, r-1\}$. Note that since $\bar{Q}_n^r$ is convex polyhedral, we can interpret $\hat{v}_n^r$ as a mixed-integer value function similar as $v_\omega$ in Section 2, defined on the extended space $S_{[a^p(n):a(n)]}$ of $X_{a(n)}$. Thus, $\hat{v}_n^r$ is not necessarily convex since it involves the minimization over the mixed-integer polyhedral set $\mathcal{X}_n(x_{a(n)})$. However, similar as for leaf nodes $m \in \mathcal{L}$ in Section 4.1 we can use affine parametric cutting planes for the feasible region of $\hat{v}_n^r$, possibly in extended spaces of $S_{[a^p(n):a(n)]}$, to construct convex polyhedral lower bounds of $\hat{v}_n^r$. In fact, we can derive a similar hierarchy of lower bounds as in Theorem 8 by looking back $q = r-1, \ldots, T-2$ time periods. However, for ease of presentation we focus on affine parametric cutting planes in the extended space $S_{[a^r(n):a(n)]}$ of $S_{[a^p(n):a(n)]}$. The corresponding lower bound $\bar{\nu}_n^r$ is for all $x_{[a^r(n):a(n)]} \in \operatorname{conv}(S_{[a^r(n):a(n)]})$ given by

$$\bar{\nu}_n^r(x_{[a^r(n):a(n)]}) := \min_{x_n,\theta_n} \Big\{ c_n^\top x_n + \theta_n : \ (x_{[a^r(n):n]}, \theta_n) \in \operatorname{conv}(\Theta_n^r) \Big\},$$

where

$$\Theta_n^r := \Big\{ (x_{[a^r(n):n]}, \theta_n) \in S_{[a^r(n):n]} \times \mathbb{R} : \theta_n \geq \bar{Q}_n^r(x_{[a^{r-1}(n):n]}) \Big\}.$$

Since $\bar{\nu}_n^r$ is a tighter lower bound of $v_n$ for larger values of $r$, this again yields a hierarchy of convex polyhedral lower bounds. Moreover, by recursively defining $\bar{\nu}^r$ for all nodes $n \in \mathcal{N}$ of the scenario tree, we obtain the following result.

**Definition 10.** *Let $r = 1, \ldots, T-1$ be given, and consider the convex polyhedral lower bounds $\bar{\nu}_m^r$, $m \in \mathcal{L}$, from Definition 9. In backward recursion fashion we define for all $n \in \mathcal{N} \setminus \mathcal{L}$, the convex polyhedral lower bound $\bar{Q}_n^r$ of $Q_n$ for all $x_{[a^{r-1}(n):n]} \in \operatorname{conv}(S_{[a^{r-1}(n):n]})$ as*

$$\bar{Q}_n^r(x_{[a^{r-1}(n):n]}) := \sum_{m \in C(n)} q_{nm} \bar{\nu}_m^r(x_{[a^{r-1}(n):n]}),$$

*and the convex polyhedral lower bound $\bar{\nu}_n^r$ of $v_n$ for all $x_{[a^r(n):a(n)]} \in \operatorname{conv}(S_{[a^r(n):a(n)]})$ as*

$$\bar{\nu}_n^r(x_{[a^r(n):a(n)]}) := \min_{x_n,\theta_n} \Big\{ c_n^\top x_n + \theta_n : \ (x_{[a^r(n):n]}, \theta_n) \in \operatorname{conv}(\Theta_n^r) \Big\},$$

*where*

$$\Theta_n^r := \Big\{ (x_{[a^r(n):n]}, \theta_n) \in S_{[a^r(n):n]} \times \mathbb{R} : \bar{Q}_n^r(x_{[a^{r-1}(n):n]}) \leq \theta_n \leq U_n \Big\}.$$

*Note that for all $n \in \mathcal{N}$, we define $a^r(n)$ as the root node, i.e., $a^r(n) = 1$, if $r \geq t_n$.*

**Theorem 9.** *Consider the convex polyhedral lower bounds $\bar{\nu}_n^r$ from Definitions 9 and 10 for all $n \in \mathcal{N}$ and for all $r = 1, \ldots, T-1$. Then, for all $n \in \mathcal{N}$, $r = 1, \ldots, T-1$, and $x_{[n]} \in S_{[n]}$, it holds that*

$$\bar{Q}_n^r(x_{[a^{r-1}(n):n]}) \leq Q_n(x_n) \qquad \text{and} \qquad \bar{\nu}_n^r(x_{[a^r(n):a(n)]}) \leq v_n(x_{[a(n)]}). \tag{20}$$

*Moreover, for all $n \in \mathcal{N}$ and $x_{[n]} \in \operatorname{conv}(S_{[n]})$, it holds that*

$$\bar{Q}_n^1(x_n) \leq \ldots \leq \bar{Q}_n^{T-1}(x_{[n]}), \tag{21}$$

*and*

$$\bar{\nu}_n^1(x_{a(n)}) \leq \ldots \leq \bar{\nu}_n^{T-1}(x_{[a(n)]}). \tag{22}$$

The lower bounds in Theorem 9 for the value function $v_1(x_0)$ in the root node of the scenario tree, define lower bounds on the optimal objective value $\eta^*$ of the multistage SMIP, obtained by iteratively using affine parametric cutting planes involving previous-stage decisions looking back $r$ stages.

**Corollary 2.** *Consider the convex polyhedral lower bounds from Definition 10 for the root node $n = 1$, and define $\bar{\eta}^r := \bar{\nu}_1^r(x_0)$ for all $r = 1, \ldots, T-1$. Then,*

$$\bar{\eta}^1 \leq \ldots \leq \bar{\eta}^{T-1} \leq \eta^*.$$

*Proof.* Follows directly from Theorem 9. □

We remark that the inequality between $\bar{\eta}^{T-1}$ and $\eta^*$ may be strict. That is why, in the next section we will show that we can close this gap using scaled cuts. As a final result in this section, however, we provide sufficient conditions for the convex polyhedral lower bounds in Theorem 9 to be tight.

**Proposition 5.** *Consider the convex polyhedral lower bounds $\bar{\nu}_n^r$ from Definitions 9 and 10 for all $n \in \mathcal{N}$ and for all $r = 1, \ldots, T-1$. Then, for all $x_{[n]} \in S_{[n]}$, it holds that*

*(i)* $\bar{Q}_n^r(x_{[a^{r-1}(n):n]}) = Q_n(x_n)$ *if and only if* $\bar{\nu}_m^r(x_{[a^{r-1}(n):n]}) = v_m(x_n)$ *for all* $m \in C(n)$.

*(ii)* $\bar{\nu}_n^r(x_{[a^r(n):a(n)]}) = v_n(x_{a(n)})$ *if for some* $\rho = 1, \ldots, r$, *an optimal solution* $(x_n^*, \theta_n^*)$ *of the minimization problem in* $\hat{v}_n^\rho(x_{[a^\rho(n):a(n)]})$ *satisfies* $\theta^* = Q_n(x_n^*)$, *and* $(x_{[a^\rho(n):a(n)]}, \bar{Q}_n^\rho(x_{[a^{\rho-1}(n):n]}))$ *is an extreme point of* $\operatorname{conv}(\Theta_n^\rho)$.

# 5 Scaled cuts for multistage SMIPs

In this section, we discuss the intuition behind our scaled-cut decomposition algorithm in Section 5.1 and its proof of convergence in Section 5.2. In Section 5.3 we derive a hierarchy of scaled cut lower bounds that improves the hierarchy of Theorem 9.

## 5.1 The intuition behind the scaled-cut decomposition algorithm

Consider the first stage of a multistage stochastic mixed-integer programming problem, given by

$$\eta^* := \min_{x_1 \in X_1} c_1^\top x_1 + Q_1(x_1).$$

Here, $Q_1(x_1) := \sum_{n \in C(1)} q_{1n} v_n(x_1)$, $x_1 \in X_1$, is the expectation over l.s.c. value functions $v_n, n \in C(1)$. Hence, it follows directly from Theorem 6 that starting from an initial convex polyhedral lower bound $\varphi_1^0$ of $Q_1$, iteratively applying the scaled cut closure operation $\varphi_1^k := SCC_1(\varphi_1^{k-1})$, $k \in \mathbb{N}$, defined as

$$\mathrm{SCC}_1(\varphi_1)(x) := \sup \left\{ \frac{\displaystyle\sum_{n \in C(1)} q_{1n}\alpha_n - \sum_{n \in C(1)} q_{1n}\beta_n^\top x}{1 + \displaystyle\sum_{n \in C(1)} q_{1n}\tau_n} : (\alpha_n, \beta_n, \tau_n) \in \Pi_n(\varphi) \right\},$$

with for every $n \in C(1)$,

$$\Pi_n(\varphi_1) := \Big\{ (\alpha_n, \beta_n, \tau_n) \in \mathbb{R} \times \mathbb{R}^{n_1} \times \mathbb{R}_+ : v_n(x_1) \geq \alpha_n - \beta_n^\top x_1 - \tau_n \varphi_1(x_1) \quad, \forall x_1 \in X_1 \Big\},$$

yields a sequence $\{\varphi_1^k\}_{k=0}^\infty$ that converges uniformly to $\mathrm{co}_{X_1}(Q_1)$. That is, the objective value of

$$\min_{x_1 \in \mathrm{conv}(X_1)} c_1^\top x_1 + \varphi_1^k(x_1)$$

converges to the true optimal objective value $\eta^*$.

The challenge in applying the scaled cut closure operation $SCC_1$ is that the value functions $v_n, n \in C(1)$, defined as

$$v_n(x_1) = \min_{x_n \in \mathcal{X}_n(x_1)} c_n^\top x_n + Q_n(x_n), \qquad x_1 \in X_1,$$

depend on the expected cost to-go function $Q_n$. Similar as for $Q_1$, we do not know $Q_n$ exactly but intend to approximate $Q_n$ by iteratively improving convex polyhedral lower bounds $\varphi_n$ on extended spaces. If these lower bounds are not tight enough, then the resulting scaled cut closure operation becomes inexact.

Proposition 3 and Remark 1 show that to obtain the exact scaled cut closure $SCC_1(\varphi_1)$ it suffices to obtain lower bounds $\hat{v}_n(x_1, \theta_1)$, $n \in C(1)$, with

$$\text{co}(v_n + \delta_{\hat{\Theta}_1(\varphi_1)})(x_1, \theta_1) \leq \hat{v}_n(x_1, \theta_1) \leq v_n(x_1) \qquad \forall\,(x_1, \theta_1) \in \hat{\Theta}_1(\varphi_1), \tag{23}$$

where $\hat{\Theta}_1(\varphi_1) := \{(x_1, \theta_1) \in X_1 \times \mathbb{R} : \varphi_1(x_1) \leq \theta_1 \leq U_1\}$. We will show in Lemma 2 in Section 5.2 that such functions $\hat{v}_n$, $n \in C(1)$, can be obtained by replacing $Q_n$ in the minimization problem corresponding to $v_n$ by a convex polyhedral lower bound of $Q_n$ on the extended space $\Theta_{[n]}(\varphi_1) := \{(x_{[n]}, \theta_1) \in S_{[n]} \times \mathbb{R} : \varphi_1(x_1) \leq \theta_1 \leq U_1\}$ of $X_n$. That is, $\hat{v}_n$ defined for all $(x_1, \theta_1) \in \hat{\Theta}_1(\varphi_1)$, and $n \in C(1)$, as

$$\hat{v}_n(x_1, \theta_1) = \min_{x_n \in \mathcal{X}_n(x_1)} c_n^\top x_n + \varphi_n(x_{[n]}, \theta_1), \tag{24}$$

satisfies (23) if $\varphi_n(x_{[n]}, \theta_1) = \text{co}(Q_n + \delta_{\Theta_{[n]}(\varphi_1)})(x_{[n]}, \theta_1)$ for all $(x_{[n]}, \theta_1) \in \Theta_{[n]}(\varphi_1)$.

Thus, to be able to carry out an exact first-stage scaled cut closure operation it suffices to construct $\text{co}(Q_n + \delta_{\Theta_{[n]}(\varphi_1)})(x_{[n]}, \theta_1)$ for all $n \in C(1)$. Observe that for every $n \in C(1)$, $Q_n$ itself can be interpreted as a first-stage expected cost to-go function of a multistage SMIP consisting of $T-1$ stages with node $n$ as the root node of the scenario tree. Moreover, notice that in this case $Q_n$ is not defined on $X_n$, but on the extended space $\Theta_{[n]}(\varphi_1)$. Hence, we can apply our scaled cut closure procedure recursively to obtain $\text{co}(Q_n + \delta_{\Theta_{[n]}(\varphi_1)})$.

Moreover, it turns out that if we do not know $\text{co}(Q_n + \delta_{\Theta_{[n]}(\varphi_1)})$ exactly, but obtain a sufficiently accurate approximation, i.e., if for all $n \in C(1)$,

$$\|\varphi_n - \text{co}(Q_n + \delta_{\Theta_{[n]}(\varphi_1)})\|_\infty \leq \delta,$$

then the inexact scaled cut closure operation $S\tilde{C}C_1$, defined with $\Pi_n(\varphi_1)$ replaced by $\tilde{\Pi}_n(\varphi_1)$ defined as

$$\tilde{\Pi}_n(\varphi_1) = \Big\{(\alpha_n, \beta_n, \tau_n) : \hat{v}_n(x_1, \theta_1) \geq \alpha_n - \beta_n^\top x_1 - \tau_n \theta_1,\ \forall\,(x_1, \theta_1) \in \hat{\Theta}_1(\varphi_1)\Big\}$$

with $\hat{v}_n(x_1, \theta_1)$ given in (24), is $\delta$-exact; see Proposition 7 in Section 5.2.

## 5.2 A scaled-cut decomposition algorithms for multistage SMIPs

In our scaled-cut decomposition algorithm for multistage SMIPs, we assume that at each node $n \in \mathcal{N}$, we maintain a convex polyhedral outer approximation $\varphi_n$ of $Q_n$. For the first-stage expected cost to-go function $Q_1$, this outer approximation is defined on $X_1$. However, following the discussion in Section 5.1, for nodes $n \in \mathcal{N}$ in later stages, these outer approximations are recursively defined on extended spaces $\Theta_{[n]}(\varphi_{[a(n)]})$ of $X_n$, defined as

$$\Theta_{[n]}(\varphi_{[a(n)]}) := \Big\{(x_{[n]}, \theta_{[a(n)]}) : x_{[n]} \in S_{[n]}, \varphi_{[a(n)]}(x_{[a(n)]}, \theta_{[a^2(n)]}) \leq \theta_{[a(n)]} \leq U_{[a(n)]}\Big\},$$

where the latter constraint is shorthand notation for the set of constraints $\varphi_1(x_1) \leq \theta_1 \leq U_1, \ldots, \varphi_{a(n)}(x_{[a(n)]}, \theta_{[a^2(n)]}) \leq \theta_{a(n)} \leq U_{a(n)}$. Note that the extended spaces $\Theta_{[n]}(\varphi_{[a(n)]})$ depend on all outer approximations of all ancestor nodes $[a(n)]$ of $n$. Similarly, we define $\hat{\Theta}_{[n]}(\varphi_{[n]})$ as

$$\hat{\Theta}_{[n]}(\varphi_{[n]}) := \Big\{(x_{[n]}, \theta_{[n]}) : x_{[n]} \in S_{[n]}, \varphi_{[n]}(x_{[n]}, \theta_{[a(n)]}) \leq \theta_{[n]} \leq U_{[n]}\Big\}.$$

We make sure that at any point in time during the algorithm, it holds that

$$\varphi_n(x_{[n]}, \theta_{[a(n)]}) \leq Q_n(x_n) + \delta_{\Theta_{[n]}(\varphi_{[a(n)]})}(x_{[n]}, \theta_{[a(n)]}) \tag{25}$$

for all $(x_{[n]}, \theta_{[a(n)]}) \in \Theta_{[n]}(\varphi_{[a(n)]})$. Since $\varphi_n$ is convex polyhedral it follows that $\varphi_n$ is also a lower bound for the convex envelope of the right-hand side in (25). In our algorithm, we iteratively improve these outer approximations $\varphi_n$ using scaled cuts at different nodes $n \in \mathcal{N}$.

**Definition 11.** *Let $\varphi$ define a set of outer approximations $\{\varphi_n\}_{n\in\mathcal{N}}$. We call $\varphi$* feasible, *denoted $\varphi \in \Phi$, if and only if for all $n \in \mathcal{N}$, $\varphi_n$ is convex polyhedral on $\mathrm{conv}(\Theta_{[n]}(\varphi_{[a(n)]}))$, and moreover for all $(x_{[n]}, \theta_{[a(n)]}) \in \mathrm{conv}(\Theta_{[n]}(\varphi_{[a(n)]}))$,*

$$\varphi_n(x_{[n]}, \theta_{[a(n)]}) \leq \mathrm{co}\Big(Q_n(x_n) + \delta_{\Theta_{[n]}(\varphi_{[a(n)]})}\Big)(x_{[n]}, \theta_{[a(n)]}). \tag{26}$$

**Definition 12.** *Let $\varphi = \{\varphi_n\}_{n\in\mathcal{N}} \in \Phi$ denote a feasible set of outer approximations. Then, for every $n \in \mathcal{N}\backslash\mathcal{L}$, we define $\hat{\varphi} := SCC_n(\varphi)$ as the result of a scaled cut closure operation at node $n$, defined as $\hat{\varphi}_n(x_{[n]}, \theta_{[a(n)]})$ for every $(x_{[n]}, \theta_{[a(n)]}) \in \mathrm{conv}(\Theta_{[n]}(\varphi_{[a(n)]}))$ as*

$$\hat{\varphi}_n(x_{[n]}, \theta_{[a(n)]}) := \sup_{\substack{(\alpha_m,\beta_m,\tau_m)\in\Pi_m(\varphi)\\ \forall\, m\in C(n)}} \left\{ \frac{\displaystyle\sum_{m\in C(n)} q_{nm}\Big(\alpha_m - \beta_m^\top(x_{[n]}, \theta_{[a(n)]})\Big)}{1 + \displaystyle\sum_{m\in C(n)} q_{nm}\tau_m} \right\},$$

*where for every $m \in C(n)$, we have that $(\alpha_m, \beta_m, \tau_m) \in \Pi_m(\varphi)$ if and only if*

$$\hat{v}_m(x_{[n]}, \theta_{[n]}) \geq \alpha_m - \beta_m^\top(x_{[n]}, \theta_{[a(n)]}) - \tau_m\theta_n \qquad \forall\,(x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]}), \tag{27}$$

*with $\hat{v}_m(x_{[n]}, \theta_{[n]}) = \min_{x_m\in\mathcal{X}_m(x_n)} c_m^\top x_m + \varphi_m(x_{[m]}, \theta_{[n]})$ for all $(x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]})$, and defined for all $(x_{[m]}, \theta_{[a(m)]}) \in \mathrm{conv}(\Theta_m(\hat{\varphi}_{[a(m)]}))$ as*

$$\hat{\varphi}_m(x_{[m]}, \theta_{[a(m)]}) := \varphi_m(x_{[m]}, \theta_{[a(m)]})$$

*for $m \in \mathcal{N}$ with $m \neq n$, recursively.*

**Remark 4.** *Since $Q_n \equiv 0$ for all $n \in \mathcal{L}$, we define $\hat{\varphi}_n$ for $\hat{\varphi} = SCC_n(\varphi)$ with $n \in \mathcal{L}$ as $\hat{\varphi}_n(x_{[n]}, \theta_{[a(n)]}) = 0$ for all $(x_{[n]}, \theta_{[a(n)]}) \in \mathrm{conv}(\Theta_{[n]}(\varphi_{[a(n)]}))$.*

The scaled cut closure operations $SCC_n$ satisfy the following monotonicity properties: (i) applying any $SCC_n$-operator typically increases the lower bounds $\varphi = \{\varphi_n\}_{n\in\mathcal{N}}$, and at least does not reduce them, and (ii) applying any $SCC_n$ operator with a set of higher initial lower bounds typically leads to better lower bounds.

**Proposition 6.** *Let $\varphi, \varphi' \in \Phi$ denote sets of feasible convex polyhedral outer approximations. Then, for every $n \in \mathcal{N}$,*

*(i)* $SCC_n(\varphi) \in \Phi$,

*(ii)* $SCC_n(\varphi) \geq \varphi$, *and*

*(iii)* $SCC_n(\varphi) \geq SCC_n(\varphi')$ *if* $\varphi \geq \varphi'$.

**Remark 5.** *When comparing two sets of outer approximations $\varphi, \varphi' \in \Phi$ in Proposition 6, we say that $\varphi \geq \varphi'$ if and only if for all $n \in \mathcal{N}$, we have (i) $\Theta_{[n]}(\varphi_{[a(n)]}) \subseteq \Theta_{[n]}(\varphi'_{[a(n)]})$, and (ii) $\varphi_n(x_{[n]}, \theta_{[a(n)]}) \geq \varphi'(x_{[n]}, \theta_{[a(n)]})$ for all $(x_{[n]}, \theta_{[n]}) \in \mathrm{conv}(\Theta_{[n]}(\varphi_{[a(n)]}))$.*

For fixed $n \in \mathcal{N}$ and $\varphi_{[a(n)]}$, the best possible convex polyhedral lower bound $\varphi_n$ of $Q_n$ that we may obtain is $\varphi_n$ satisfying (26) with equality for all $(x_{[n]}, \theta_{[a(n)]}) \in \Theta_{n]}(\varphi_{[a(n)]})$. Indeed, if we interpret $Q_n$ as the expectation of l.s.c. value functions defined on the extended space $\Theta_{[n]}(\varphi_{[a(n)]})$ of $X_n$, then iteratively applying the exact scaled cut closure from Definition 7 to any initial convex polyhedral lower bound $\varphi_n^0$, converges uniformly to this best possible convex polyhedral lower bound. The main difference, however, between the SCC operation from Definition 7 and the $SCC_n$ operator defined in Definition 12 is that the $SCC$ operation in Definition 7 assumes that the expected cost to-go functions $Q_m$, $m \in C(n)$, are known, whereas the $SCC_n$ operator from Definition 12 uses convex polyhedral lower bounds $\varphi_m$ of $Q_m$. Theorem 10 shows that both scaled cut closure operators may coincide if for every $m \in C(n)$, the lower bound $\varphi_m$ of $Q_m$ is accurate enough. In this case, we call the $SCC_n$ operator exact.

**Definition 13.** *Let $\varphi = \{\varphi_n\}_{n\in\mathcal{N}} \in \Phi$ denote a feasible set of outer approximations and consider $n \in \mathcal{N}$. Interpret $Q_n$ as the expectation of l.s.c. value functions defined on the extended space $\Theta_{[n]}(\varphi_{[a(n)]})$ of $X_n$. Then, we call the $SCC_n$ operator from Definition 12* exact *if it coincides with the $SCC$ operator from Definition 7.*

**Lemma 2.** *Let $\varphi = \{\varphi_n\}_{n\in\mathcal{N}} \in \Phi$ denote a feasible set of outer approximations and consider $n \in \mathcal{N}\backslash\mathcal{L}$. Then, for all $m \in C(n)$, it holds that $\hat{v}_m$, defined for all $(x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]})$ as*

$$\hat{v}_m(x_{[n]}, \theta_{[n]}) := \min_{x_m \in \mathcal{X}_m(x_n)} c_m^\top x_m + co(Q_m + \delta_{\Theta_{[m]}(\varphi_{[n]})})(x_{[m]}, \theta_{[n]}),$$

*satisfies*

$$co(v_m + \delta_{\hat{\Theta}_{[n]}(\varphi_{[n]})})(x_{[n]}, \theta_{[n]}) \leq \hat{v}_m(x_{[n]}, \theta_{[n]}) \leq v_m(x_n)$$

*for all $(x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]})$.*

**Theorem 10.** *Let $\varphi = \{\varphi_n\}_{n\in\mathcal{N}} \in \Phi$ denote a feasible set of outer approximations and consider $n \in \mathcal{N}\backslash\mathcal{L}$. Then, the $SCC_n$ operator from Definition 12 is exact if for all $m \in C(n)$,*

$$\varphi_m(x_{[m]}, \theta_{[a(m)]}) = co(Q_m + \delta_{\Theta_{[m]}(\varphi_{[n]})})(x_{[m]}, \theta_{[n]}) \quad \forall (x_{[m]}, \theta_{[n]}) \in \Theta_{[m]}(\varphi_{[n]}). \tag{28}$$

Typically, however, the $SCC_n$ operations will be inexact. The next proposition provides a sufficient condition for the $SCC_n$ operations to be $\delta$-exact, conform Definition 8.

**Proposition 7.** *Let $\varphi = \{\varphi_n\}_{n\in\mathcal{N}} \in \Phi$ denote a feasible set of outer approximations and consider $n \in \mathcal{N}\backslash\mathcal{L}$. Then, the $SCC_n$ operation from Definition 12 is $\delta$-exact if for all $m \in C(n)$,*

$$\|\varphi_m - co(Q_m + \delta_{\Theta_{[m]}(\varphi_{[n]})})\|_\infty \leq \delta. \tag{29}$$

We are now ready to prove our main convergence result, which shows that iteratively applying scaled cut closure operations yields a sequence $\{\{\varphi_n^k\}_{n\in\mathcal{N}}\}_{k\in\mathbb{N}}$ for which $\varphi_1^k$ converges uniformly to $\text{co}_{X_1}(Q_1)$. To prove this, we only need mild conditions on the order of nodes $n_k \in \mathcal{N}$ at which scaled cut closure operations are carried out. Obviously, convergence does not necessarily hold if no, or only finitely many, $SCC_n$ operations are carried out at a particular node $n$.

**Definition 14.** *Let $\{n_k\}_{k\in\mathbb{N}} \subseteq \mathcal{N}$ denote a sequence of nodes. We call such a sequence* admissable *if and only if there exists a number $B \in \mathbb{N}$ such that for every $k \in \mathbb{N}$ and $n \in \mathcal{N}$ there exists $l \in \mathbb{N}$ with $k \leq l \leq k + B$ such that $n_l = n$.*

Intuitively, a sequence $\{n_k\}_{k\in\mathbb{N}}$ is admissable if every node $n \in \mathcal{N}$ occurs frequently enough, i.e., at least once after a fixed number $B$. In practice, this assumption is not very restrictive since we allow $B$ to be large. For example, an SDDP-like sequence based on moving through the network with forward and backward passes is admissable.

**Theorem 11.** *Consider a multistage SMIP as defined in (1), and let $\varphi^0 = \{\varphi_n^0\}_{n\in\mathcal{N}} \in \Phi$ be an initial feasible set of outer approximations. Then, for every admissable sequence $\{n_k\}_{k\in\mathbb{N}} \subseteq \mathcal{N}$, and corresponding sequence of outer approximations, defined as*

$$\varphi^k := SCC_{n_k}(\varphi^{k-1}), \qquad k \in \mathbb{N},$$

*it holds that $\{\varphi_1^k\}_{k\in\mathbb{N}}$ converges uniformly to $co_{X_1}(Q_1)$.*

## 5.3 A hierarchy of scaled cut lower bounds

To obtain a scaled-cut decomposition algorithm satisfying our main convergence result in Theorem 11, we have to keep track of convex polyhedral lower bounds $\varphi_n$ of $Q_n$, $n \in \mathcal{N}$, that depend on all lower bounds $\varphi_{[a(n)]}$ and decisions $x_{[a(n)]}$ of all ancestor nodes $[a(n)]$ of $n$. By letting these lower bounds $\varphi_n$ depend on fewer previous-stage decisions and lower bounds we may improve the computational tractability of the scaled cuts at the expense of their strength. We will show in this section that restricting the dependence of $\varphi_n$ at node $n \in \mathcal{N}$ on decisions $x_{[a^{r-1}(n):n]}$ only, for $r = 1, \dots, T-1$, improves the hierarchy of convex polyhedral lower bounds in Theorem 9, leading to a hierarchy of scaled cut lower bounds.

To understand why scaled cuts may improve the hierarchy of lower bounds in Theorem 9, consider the convex polyhedral lower bound $\bar{\nu}_n^r(x_1)$ from Definition 10 for some $n \in C(1)$, $r = 1, \dots, T-1$, and $x_1 \in X_1$, given by

$$\bar{\nu}_n^r(x_1) = \min_{x_n, \theta_n} \Big\{ c_n^\top x_n + \theta_n : x_{[a^{r-1}(n):n]} \in \operatorname{conv}(\Theta_n^r) \Big\},$$

where $\Theta_n^r := \{(x_{[a^{r-1}(n):n]}, \theta_n) \in S_{[a^{r-1}(n):n]} \times \mathbb{R} : \bar{Q}_n^r(x_{[a^{r-1}(n):n]}) \leq \theta_n \leq U_n\}$. Observe that $\bar{\nu}_n^r(x_1)$ is a convex polyhedral lower bound of

$$\hat{v}_n^r(x_1) := \min_{x_n \in \mathcal{X}_n(x_1)} \Big\{ c_n^\top x_n + \bar{Q}_n^r(x_{[a^{r-1}(n):n]}) \Big\},$$

where $\hat{v}_n^r$ is not necessarily convex and can be interpreted as $\bar{\nu}_n^r$ with $\operatorname{conv}(\Theta_n^r)$ replaced by $\Theta_n^r$. In Definition 10, the convex polyhedral lower bound $\bar{Q}_1^r$ of $Q_1$ is defined as $\bar{Q}_1^r(x_1) = \sum_{n \in C(1)} q_{1n} \bar{\nu}_n^r(x_1)$, $x_1 \in X_1$, which since $\bar{\nu}_n^r$ is convex can be interpreted as

$$\bar{Q}_1^r(x_1) = \sum_{n \in C(1)} q_{1n} \mathrm{co}_{X_1}(\bar{\nu}_n^r)(x_1), \qquad x_1 \in X_1.$$

Instead, we may iteratively apply scaled cuts to obtain the lower bound $\mathrm{co}_{X_1}(\hat{Q}_1^r)(x_1)$, with

$$\hat{Q}_1^r(x_1) := \sum_{n \in C(1)} q_{1n} \hat{v}_n^r(x_1), \qquad x_1 \in X_1.$$

Since for all $x_1 \in \operatorname{conv}(X_1)$, we have

$$\mathrm{co}_{X_1}(\hat{Q}_1^r)(x_1) \geq \sum_{n \in C(1)} q_{1n} \mathrm{co}_{X_1}(\hat{v}_n^r)(x_1) \geq \sum_{n \in C(1)} q_{1n} \mathrm{co}_{X_1}(\bar{\nu}_n^r)(x_1) = \bar{Q}_1^r(x_1),$$

where the first inequality is typically strict, this shows why scaled cuts may improve the hierarchy of lower bounds in Theorem 9.

Next, we define for every $r = 1, \dots, T-1$, similar scaled cut closure operators $SCC_n^r$ at each node $n \in \mathcal{N}$, and derive a hierarchy of scaled cut closure lower bounds.

**Definition 15.** *Let $r = 1, \dots, T-1$ be given, and let $\varphi$ define a set of outer approximations $\{\varphi_n\}_{n \in \mathcal{N}}$. We call $\varphi$* feasible, *denoted $\varphi \in \Phi^r$, if and only if for all $n \in \mathcal{N}$, $\varphi_n$ is convex polyhedral on $\operatorname{conv}(S_{[a^{r-1}(n):n]})$, and moreover for all $x_{[a^{r-1}(n):n]} \in \operatorname{conv}(S_{[a^{r-1}(n):n]})$,*

$$\varphi_n(x_{[a^{r-1}(n):n]}) \leq co\Big(Q_n(x_n) + \delta_{\operatorname{conv}(S_{[a^{r-1}(n):n]})}\Big)(x_{[a^{r-1}(n):n]}). \tag{30}$$

**Definition 16.** *Let $r = 1, \dots, T-1$, be given, and let $\varphi = \{\varphi_n\}_{n \in \mathcal{N}} \in \Phi^r$ denote a feasible set of outer approximations. Then, for every $n \in \mathcal{N} \setminus \mathcal{L}$, we define $\hat{\varphi} := SCC_n^r(\varphi)$ as the result of a scaled cut closure operation at node $n$, defined as $\hat{\varphi}_n(x_{[a^{r-1}(n):n]})$ for every $x_{[a^{r-1}(n):n]} \in \operatorname{conv}(S_{[a^{r-1}(n):n]})$ as*

$$\hat{\varphi}_n(x_{[a^{r-1}(n):n]}) := \sup_{\substack{(\alpha_m, \beta_m, \tau_m) \in \Pi_m^r(\varphi) \\ \forall\, m \in C(n)}} \left\{ \frac{\displaystyle\sum_{m \in C(n)} q_{nm} \Big(\alpha_m - \beta_m^\top x_{[a^{r-1}(n):n]}\Big)}{1 + \displaystyle\sum_{m \in C(n)} q_{nm} \tau_m} \right\},$$

*where for every $m \in C(n)$, we have that $(\alpha_m, \beta_m, \tau_m) \in \Pi_m^r(\varphi)$ if and only if for all $x_{[a^{r-1}(n):n]} \in S_{[a^{r-1}(n):n]}$ we have*

$$\hat{v}_m^r(x_{[a^{r-1}(n):n]}) \geq \alpha_m - \beta_m^\top x_{[a^{r-1}(n):n]} - \tau_m \varphi_n(x_{[a^{r-1}(n):n]}),$$

*with $\hat{v}_m^r(x_{[a^{r-1}(n):n]}) := \min_{x_m \in \mathcal{X}_m(x_n)} c_m^\top x_m + \varphi_m(x_{[a^{r-1}(m):m]})$ for all $(x_{[a^{r-1}(n):n]}) \in S_{[a^{r-1}(n):n]}$, and defined for all $x_{[a^{r-1}(m):m]} \in \operatorname{conv}(S_{[a^{r-1}(m):m]})$ as*

$$\hat{\varphi}_m(x_{[a^{r-1}(m):m]}) := \varphi_m(x_{[a^{r-1}(m):m]})$$

*for $m \in \mathcal{N}$ with $m \neq n$, recursively.*

**Remark 6.** *Similar as for $SCC_n$, we define for all $n \in \mathcal{L}$, $\hat{\varphi}_n$ for $\hat{\varphi} = SCC_n^r(\varphi)$ as $\hat{\varphi}_n(x_{[a^{r-1}(n):n]}) = 0$ for all $x_{[a^{r-1}(n):n]} \in \operatorname{conv}(S_{[a^{r-1}(n):n]})$.*

**Theorem 12.** *For any $r = 1, \ldots, T-1$, let $\bar{\varphi}^r := \lim_{k\to\infty} \varphi^{r,k}$, where $\varphi^{r,k} := SCC_{n_k}^r(\varphi^{r,k-1})$, $k \in \mathbb{N}$, for a feasible set of initial outer approximations $\varphi^{r,0} \in \Phi^r$ and any admissable sequence $\{n_k\}_{k\in\mathbb{N}} \subseteq \mathcal{N}$. Then, for every $n \in \mathcal{N}$, $r = 1, \ldots, T-1$, and $x_{[n]} \in S_{[n]}$, it holds that $\bar{\varphi}_n^r(x_{[a^{r-1}(n):n]}) \leq Q_n(x_n)$. Moreover, for all $n \in \mathcal{N}$ and $x_{[n]} \in \operatorname{conv}(S_{[n]})$, it holds that*

$$\bar{\varphi}_n^1(x_n) \leq \ldots \leq \bar{\varphi}_n^{T-1}(x_{[n]}),$$

*with $\bar{Q}_n^r$ from Definition 10 satisfying $\bar{Q}_n^r(x_{[a^{r-1}(n):n]}) \leq \bar{\varphi}_n^r(x_{[a^{r-1}(n):n]})$ for every $x_{[a^{r-1}(n):n]} \in$ conv$(S_{[a^{r-1}(n):n]})$ and $r = 1, \ldots, T-1$.*

# 6 Conclusion and discussion

In this paper, we consider Benders decomposition algorithms for multistage stochastic mixed-integer programs (SMIPs) with general mixed-integer decision variables. We analyze the strength of the convex polyhedral lower bounds for the expected cost to-go functions resulting from using affine parametric cutting planes for the feasible regions in the model. By constructing such affine parametric cutting planes in increasingly higher dimensional spaces, we derive a hierarchy of convex polyhedral lower bounds. Moreover, we improve this hierarchy using so-called scaled cuts, and we derive a scaled-cut decomposition algorithm for which the lower bound of the first-stage expected cost to-go function converges to its convex envelope.

We note that in this paper we do not discuss how to derive the affine parametric cutting planes in a practical algorithm, but merely discuss the strength of the lower bounds obtained when adding all possible affine parametric cuts. In a practical algorithm, however, we do not intend to add all of these cutting planes, but only those relevant for the direction in which we are optimizing. A future research direction is to investigate how to do this in a numerically efficient way.

Another numerical consideration for future research is the order in which to strengthen the feasible regions and convex polyhedral lower bounds at the various nodes of the scenario tree in the model. For example, at the leaf nodes first, or at the root node, or in SDDP-fashion. The same question applies to our scaled-cut decomposition algorithm. Theoretically, we have proven convergence of the algorithm under very mild conditions, however, the speed of convergence may significantly depend on the order in which the scaled cut closure operations are applied. Moreover, in practice it is possible to add only a single scaled cut in each iteration, instead of the entire scaled cut closure, similar as for two-stage SMIPs, see [20]. Finally, we remark that our affine parametric cutting planes and scaled cuts, possibly in extended spaces, may also be embedded in decomposition or B&B schemes, akin to branch-and-cut for deterministic mixed-integer linear programming.

# Appendix

**Proof of Proposition 1.** Let $\pi^\top(x,z) \geq \pi_0$ be an affine cutting plane in $x$ for $Z_\omega(x)$ with respect to $X$. Then, by definition, this affine cutting plane is valid for every $P_\omega$, which implies that for every $(\bar{x},\bar{z}) \in P_\omega$ we have $\pi^\top(\bar{x},\bar{z}) \geq \pi_0$. Moreover, observe that the cutting plane $\pi^\top(x,z) \geq \pi_0$ defines a half-space $H = \{(x,z) : \pi^\top(x,z) \geq \pi_0\}$, and thus $P_\omega \subseteq H$. Since every half-space is convex and $\mathrm{conv}(P_\omega)$ is the smallest convex set containing $P_w$, it now follows directly that $\mathrm{conv}(P_\omega) \subseteq H$, and thus the affine parametric cutting plane is valid for $\mathrm{conv}(P_\omega)$. $\square$

**Proof of Theorem 1.** Let $x \in X$ be given and observe that we can rewrite $v_\omega(x)$ as

$$v_\omega(x) = \min_z \{q_\omega^\top z : (x,z) \in P_\omega\}.$$

The feasible region corresponding to the minimization problem in $\nu_\omega(x)$ is at least as large as that of $v_\omega(x)$, since $\mathrm{conv}(P_\omega) \supseteq P_w$, and thus (i) holds.

To prove (ii), let $\bar{x} \in X$ be an extreme point of $\mathrm{conv}(X)$, and suppose for contradiction that $\nu_\omega(x) < v_\omega(x)$. This would mean that there exists $\bar{z} \in Z$ such that $(\bar{x},\bar{z}) \in \mathrm{conv}(P_\omega)$ and $q_\omega^\top \bar{z} = \nu_\omega(x) < v_\omega(\bar{x})$. Since $(\bar{x},\bar{z}) \in \mathrm{conv}(P_\omega)$, it follows from Caratheodory's theorem that there exist finitely many extreme points $(x^k,z^k) \in \mathrm{conv}(P_\omega)$ with $(x^k,z^k) \in P_\omega$ such that

$$(\bar{x},\bar{z}) = \sum_{k=1}^K \lambda^k (x^k,z^k),$$

for some $\lambda^k \in [0,1]$, $k = 1,\ldots,K$, with $\sum_{k=1}^K \lambda^k = 1$. However, by definition of $\mathrm{conv}(P_\omega)$ we need to have $x^k \in \mathrm{conv}(X)$ for all $k = 1,\ldots,K$, and since $\bar{x}$ is an extreme point of $\mathrm{conv}(X)$, it follows that $x^k = \bar{x}$ for all $k = 1,\ldots,K$. Hence, $(\bar{x},z^k) \in P_\omega$ for all $k = 1,\ldots,K$, and thus $z^k \in Z_\omega(\bar{x})$ for all $k = 1,\ldots,K$, from which it follows that $\bar{z} \in \mathrm{conv}(Z_\omega(\bar{x}))$. Hence, $\bar{z}$ is feasible in (4), implying that $v_\omega(\bar{x}) \leq q_\omega^\top \bar{z}$, contradicting that $\nu_\omega(\bar{x}) < v_\omega(\bar{x})$. We conlclude that $\nu_\omega(\bar{x}) = v_\omega(\bar{x})$, and thus (ii) holds. $\square$

**Proof of Theorem 2.** As already discussed, (ii) and (iii) follow directly from Theorem 1 applied to $\hat{v}_\omega(x,y) := v_\omega(x)$ defined on the extended space $\Theta$ of $X$. To prove the inequality in (i), let $(x,y) \in \mathrm{conv}(\Theta)$ be given, and consider $z$ such that $(x,y,z) \in \mathrm{conv}(\mathcal{E}_\omega)$. Then, by Caratheodory's theorem, there exist $\lambda_k \geq 0$, $k = 1,\ldots,K$, with $\sum_{k=1}^K \lambda_k = 1$, and $(x^k,y^k,z^k) \in \mathcal{E}_\omega$, $k = 1,\ldots,K$, such that $(x,y,z) = \sum_{k=1}^K \lambda_k (x^k,y^k,z^k)$. Moreover, since $(x^k,y^k,z^k) \in \mathcal{E}_\omega$, it holds that for all $k = 1,\ldots,K$, that $x^k \in X$, $z^k \in Z$, and $(x^k,z^k) \in P_\omega$. Hence,

$$(x,z) = \sum_{k=1}^K \lambda_k (x^k,z^k) \in \mathrm{conv}(P_\omega),$$

and thus $(x,y,z) \in \mathrm{conv}(\mathcal{E}_\omega)$ implies that $(x,z) \in \mathrm{conv}(P_\omega)$. We conclude that the feasible region of the minimization problem corresponding to the first problem is at least as large as that of the second problem. Hence, the desired result follows. $\square$

**Proof of Theorem 3.** Let $(x,y^1) \in \mathrm{conv}(\Theta)$ be given. Then, for every $z$ such that $(x,y^1,z) \in \mathrm{conv}(\mathcal{E}_\omega)$, it holds that there exists $\lambda_k \geq 0$, $k = 1,\ldots,K$, with $\sum_{k=1}^K \lambda^k = 1$, and $(x^k,y^k,z^k) \in \mathcal{E}_\omega$ such that $(x,y^1,z) = \sum_{k=1}^K \lambda_k (x^k,y^k,z^k)$. Observe that since $(x^k,y^k,z^k) \in \mathcal{E}_\omega$, it holds that $(x^k,y^k) \in \mathcal{F}$. Now define $\tilde{y}^k := y^k + (y^2 - y^1)$. By definition of a monotone extended space, it holds that $\tilde{y}^k \in Y$ and $(x^k,\tilde{y}^k) \in \mathcal{F}$ for all $k = 1,\ldots,K$, and thus, $(x^k,\tilde{y}^k,z^k) \in \mathcal{E}_\omega$. We conclude that

$$(x,y^2,z) = \sum_{k=1}^K \lambda_k (x^k,\tilde{y}^k,z^k) \in \mathrm{conv}(\mathcal{E}_\omega).$$

Hence, $(x,y^1,z) \in \mathrm{conv}(\mathcal{E}_\omega)$ implies that $(x,y^2,z) \in \mathrm{conv}(\mathcal{E}_\omega)$, and thus the second feasible region is larger. $\square$

**Proof of Lemma 1.** Since $X$ is compact, $\text{conv}(X)$ is compact, and since $\varphi$ is convex polyhedral it is continuous on $\text{conv}(X)$. Hence, by Weierstrass' theorem $\varphi$ has a finite supremum on $\text{conv}(X)$, which we denote by $\theta^* = \sup_{x\in\text{conv } X} \varphi(x)$. We will show that $\hat{\nu}_\omega(x,\theta) \leq \nu_\omega(x)$ for all $x \in \text{conv}(X)$ and $\theta \geq \theta^*$. The claim then follows since the reverse inequality holds by Theorem 2 (i).

Let $x \in \text{conv}(X)$ and $\theta \geq \theta^*$ be given. Then, for any feasible $z$ in the optimization problem of $\nu_\omega(x)$, i.e. for any $z \in Z$ such that $(x, z) \in \text{conv}(P_\omega)$, there exists $\lambda_k \geq 0$, $k = 1,\ldots,K$, with $\sum_{k=1}^K \lambda_k = 1$, and $(x^k, z^k) \in P$, $k = 1,\ldots,K$, such that $(x, z) = \sum_{k=1}^K \lambda_k(x^k, z^k)$. By definition of $\Theta$ and since $\theta \geq \theta^*$, if holds for all $k = 1,\ldots,K$, that $(x^k, \theta, z^k) \in \mathcal{E}_\omega$, so that $(x, \theta, z) = \sum_{k=1}^K \lambda_k(x^k, \theta, z^k) \in \text{conv}(\mathcal{E}_\omega)$. Thus, the feasible region in the optimization problem of $\hat{\nu}_\omega(x,\theta)$ is at least as large as that of $\nu_\omega(x)$, and thus $\nu_\omega(x) \geq \hat{\nu}_\omega(x,\theta)$. We conclude that $\nu_\omega(x) = \hat{\nu}_\omega(x,\theta)$ if $\theta \geq \theta^*$. □

**Proof of Theorem 4.** By Definition 4 it holds that for all $(x,\theta) \in \text{conv}(\Theta)$,

$$\hat{\nu}_\omega(x,\theta) := \min_z \Big\{ q_\omega^\top z : (x,\theta,z) \in \text{conv}(\mathcal{E}_\omega) \Big\}, \tag{31}$$

where $\mathcal{E}_\omega := \{(x,\theta,z) \in \Theta \times Z : (x,z) \in P_\omega\}$. Since $\hat{\nu}_\omega(x,\theta)$ is convex in $(x,\theta)$, it follows that $\hat{\nu}_\omega = \text{co}(\hat{\nu}_\omega)$, and thus $\hat{\nu}_\omega$ can be expressed as the supremum of all its affine lower bounds, i.e.,

$$\hat{\nu}_\omega(x,\theta) = \sup_{(\alpha_\omega,\beta_\omega,\tau_\omega)\in\tilde{\Pi}_\omega(\varphi)} \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta,$$

where $(\alpha_\omega, \beta_\omega, \tau_\omega)$ represents the coefficients of an affine function in $(x,\theta)$, and this function is a lower bound of $\hat{\nu}_\omega$ if $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \tilde{\Pi}_\omega(\varphi)$ with

$$\tilde{\Pi}_\omega(\varphi) := \Big\{ (\alpha_\omega, \beta_\omega, \tau_\omega) : \hat{\nu}_\omega(x,\theta) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta \ \ \forall (x,\theta) \in \Theta \Big\}.$$

To prove the desired result we will show that $\tilde{\Pi}_\omega(\varphi) = \Pi_\omega(\varphi)$. We have by Definition 6 that

$$\begin{aligned}
\Pi_\omega(\varphi) &= \Big\{ (\alpha_\omega, \beta_\omega, \tau_\omega) : v_\omega(x) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta \ \text{ for all } (x,\theta) \in \Theta \Big\} \\
&= \Big\{ (\alpha_\omega, \beta_\omega, \tau_\omega) : q_\omega^\top z \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta \ \text{ for all } (x,\theta) \in \Theta, (x,z) \in P_\omega \Big\} \\
&= \Big\{ (\alpha_\omega, \beta_\omega, \tau_\omega) : q_\omega^\top z \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta \ \text{ for all } (x,\theta,z) \in \mathcal{E}_\omega \Big\}.
\end{aligned}$$

Since the constraint $q_\omega^\top z \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta$ is affine in $(x,\theta,z)$, we may replace $\mathcal{E}_\omega$ in the last equation by $\text{conv}(\mathcal{E}_\omega)$. Moreover, it follows from (31) that $q_\omega^\top z \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta$ for all $(x,\theta,z) \in \text{conv}(\mathcal{E}_\omega)$ is equivalent to $\hat{\nu}_\omega(x,\theta) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta$ for all $(x,\theta) \in \Theta$. Hence,

$$\Pi_\omega(\varphi) = \Big\{ (\alpha_\omega, \beta_\omega, \tau_\omega) : \hat{\nu}_\omega(x,\theta) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta \ \text{ for all } (x,\theta) \in \Theta \Big\} = \tilde{\Pi}_\omega(\varphi).$$

□

**Proof of Theorem 5.** Let $x \in \text{conv}(X)$ be given and consider $\theta(x)$ as defined in (8). We define $(\alpha_\omega^*, \beta_w^*, \tau_\omega^*) \in \Pi_\omega(\varphi)$ for all $\omega \in \Omega$ as the maximizers of the supremum in (8), so that

$$\theta(x) = \frac{\mathbb{E}_\omega[\alpha_\omega^*] - \mathbb{E}_\omega[\beta_\omega^*]^\top x}{1 + \mathbb{E}_\omega[\tau_\omega^*]}.$$

Rewriting this expression yields

$$\theta(x) = \mathbb{E}_\omega[\alpha_\omega^*] - \mathbb{E}_\omega[\beta_\omega^*]^\top x - \mathbb{E}_\omega[\tau_\omega^*]\theta(x). \tag{32}$$

Recall that by Theorem 4,

$$\hat{\nu}_\omega(x,\theta(x)) = \sup_{(\alpha_\omega,\beta_\omega,\tau_\omega)\in\Pi_\omega(\varphi)} \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta(x).$$

Since $(\alpha_\omega^*, \beta_\omega^*, \tau_\omega^*)$ is feasible but not necessarily optimal for the supremization in $\hat{\nu}_\omega(x, \theta(x))$ for every $\omega \in \Omega$, it follows that

$$\begin{aligned}\mathbb{E}_\omega[\hat{v}_\omega(x, \theta(x))] &\geq \mathbb{E}_\omega[\alpha_\omega^*] - \mathbb{E}_\omega[\beta_\omega^*]^\top x - \mathbb{E}_\omega[\tau_\omega^*]\theta(x) \\ &= \theta(x),\end{aligned}$$

where we used (32) for the final equality. Similarly, it holds that

$$\theta(x) \geq \frac{\mathbb{E}_\omega[\tilde{\alpha}_\omega] - \mathbb{E}_\omega[\tilde{\beta}_\omega]^\top x}{1 + \mathbb{E}_\omega[\tilde{\tau}_\omega]}, \tag{33}$$

where $(\tilde{\alpha}_\omega, \tilde{\beta}_\omega, \tilde{\tau}_\omega) \in \Pi_\omega(\varphi)$ are optimal solutions to the supremization problems in $\hat{\nu}_\omega(x, \theta(x))$ for all $\omega \in \Omega$. Since $\mathbb{E}_\omega[\tilde{\tau}_\omega] \geq 0$, we can multiply both sides of the inequality in (33) by $1 + \mathbb{E}_\omega[\tilde{\tau}_\omega]$ to obtain $(1 + \mathbb{E}_\omega[\tilde{\tau}_\omega])\theta(x) \geq \mathbb{E}_\omega[\tilde{\alpha}_\omega] - \mathbb{E}_\omega[\tilde{\beta}_\omega]^\top x$, and thus

$$\theta(x) \geq \mathbb{E}_\omega[\tilde{\alpha}_\omega] - \mathbb{E}_\omega[\tilde{\beta}_\omega]^\top x - \mathbb{E}_\omega[\tilde{\tau}_\omega]\theta(x) = \mathbb{E}_\omega[\hat{\nu}_\omega(x, \theta(x))].$$

Hence, $\theta(x) = \mathbb{E}_\omega[\hat{\nu}_\omega(x, \theta(x))]$ for all $x \in \text{conv}(X)$. □

**Proof of Proposition 3.** We first rewrite $\Pi_\omega(\varphi)$ as given in Definition 6. Since $\tau_\omega \geq 0$, this set equals

$$\Pi_\omega(\varphi) = \Big\{(\alpha_\omega, \beta_\omega, \tau_\omega) : v_\omega(x) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta, \quad \forall\, (x, \theta) \in \Theta\Big\}.$$

By adding $\delta_\Theta(x, \theta)$ to the left-hand side of the inequality we make sure that the inequality holds for all $(x, \theta) \in \text{conv}(\Theta)$, that is,

$$\Pi_\omega(\varphi) = \Big\{(\alpha_\omega, \beta_\omega, \tau_\omega) : v_\omega(x) + \delta_\Theta(x, \theta) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta, \quad \forall (x, \theta) \in \text{conv}(\Theta)\Big\}.$$

Moreover, since the right-hand side of the inequality is affine in $(x, \theta)$, we may replace the left-hand side of the inequality by $\text{co}(v_\omega + \delta_\Theta)(x, \theta)$.

Now let $\hat{v}_\omega$ be given such that (12) holds. Then, for every $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\varphi)$, we have $\text{co}(v_\omega + \delta_\Theta)(x, \theta) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta$ for all $(x, \theta) \in \Theta$, and thus

$$(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Big\{(\alpha_\omega, \beta_\omega, \tau_\omega) : \hat{v}_\omega(x, \theta) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta, \quad \forall (x, \theta) \in \Theta\Big\},$$

since $\hat{v}_\omega(x, \theta) \geq \text{co}(v_\omega + \delta_\Theta)(x, \theta) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta$ for all $(x, \theta) \in \Theta$. On the other hand, for every $(\alpha_\omega, \beta_\omega, \tau_\omega)$ with $\tau_\omega \geq 0$ satisfying $\hat{v}_\omega(x, \theta) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta$ for all $(x, \theta) \in \Theta$, it holds that $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\varphi)$, since $v_\omega(x) \geq \hat{v}_\omega(x, \theta) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta$ for all $(x, \theta) \in \Theta$. We conclude that (13) holds. □

**Proof of Theorem 7.** Since the lower bounds $\tilde{v}_\omega$ of $v_\omega$ satisfy (15), it follows that for every $\omega \in \Omega$ there exists $\hat{v}_\omega(x, \theta)$ with $\text{co}(v_\omega + \delta_\Theta)(x, \theta) \leq \hat{v}_\omega(x, \theta) \leq v_\omega(x)$ for all $(x, \theta) \in \Theta$ such that $\|\hat{v}_\omega - \tilde{v}_\omega\|_\infty \leq \delta$. Comparing $\tilde{\Pi}_\omega(\varphi)$ as defined in (14) and $\Pi_\omega(\varphi)$, which by Proposition 3 equals

$$\Pi_\omega(\varphi) = \Big\{(\alpha_\omega, \beta_\omega, \tau_\omega) : \hat{v}_\omega(x, \theta) \geq \alpha_\omega - \beta_\omega^\top x - \tau_\omega\theta,\ \forall (x, \theta) \in \Theta, \tau_\omega \geq 0\Big\},$$

for every $\omega \in \Omega$, we observe that if $(\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\varphi)$, then $(\alpha_\omega - \delta, \beta_\omega, \tau_\omega) \in \tilde{\Pi}_\omega(\varphi)$, and vice versa, if $(\tilde{\alpha}_\omega, \tilde{\beta}_\omega, \tilde{\tau}_\omega) \in \tilde{\Pi}_\omega(\varphi)$, then $(\tilde{\alpha}_\omega - \delta, \tilde{\beta}_\omega, \tilde{\tau}_\omega) \in \Pi_\omega(\varphi)$. Hence, for every $x \in \text{conv}(X)$, we have

$$\begin{aligned}S\tilde{C}C(\varphi)(x) &= \sup\left\{\frac{\mathbb{E}_\omega[\tilde{\alpha}_\omega] - \mathbb{E}_\omega[\tilde{\beta}_\omega]^\top x}{1 + \mathbb{E}_\omega[\tilde{\tau}_\omega]} : (\tilde{\alpha}_\omega, \tilde{\beta}_\omega, \tilde{\tau}_\omega) \in \tilde{\Pi}_\omega(\varphi)\right\} \\ &\geq \sup\left\{\frac{\mathbb{E}_\omega[\alpha_\omega - \delta] - \mathbb{E}_\omega[\beta_\omega]^\top x}{1 + \mathbb{E}_\omega[\tau_\omega]} : (\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\varphi)\right\} \\ &\geq SCC(\varphi)(x) - \delta,\end{aligned}$$

where the second inequality holds since $\tau_\omega \geq 0$ for all $\omega \in \Omega$, and thus $-\delta/(1 + \mathbb{E}_\omega[\tau_\omega]) \geq -\delta$. Similarly, it holds that

$$\begin{aligned} SCC(\varphi)(x) = &\geq \sup\left\{\frac{\mathbb{E}_\omega[\alpha_\omega] - \mathbb{E}_\omega[\beta_\omega]^\top x}{1 + \mathbb{E}_\omega[\tau_\omega]} : (\alpha_\omega, \beta_\omega, \tau_\omega) \in \Pi_\omega(\varphi)\right\} \\ &\geq \sup\left\{\frac{\mathbb{E}_\omega[\tilde{\alpha}_\omega - \delta] - \mathbb{E}_\omega[\tilde{\beta}_\omega]^\top x}{1 + \mathbb{E}_\omega[\tilde{\tau}_\omega]} : (\tilde{\alpha}_\omega, \tilde{\beta}_\omega, \tilde{\tau}_\omega) \in \tilde{\Pi}_\omega(\varphi)\right\} \\ &\geq S\tilde{C}C(\varphi)(x) - \delta. \end{aligned}$$

We conclude that $\|SCC(\varphi) - S\tilde{C}C(\varphi)\|_\infty \leq \delta$. □

**Proof of Corollary 1.** We will prove the claim by mathematical induction. For $K = 1$, it holds by definition of a $\delta$-exact scaled cut closure operation. Next, we assume that the claim holds for some $K \in \mathbb{N}$, i.e.,

$$\|\bar{\varphi}^K - \hat{\varphi}^K\|_\infty \leq \sum_{k=1}^{K} \delta_k.$$

By Proposition 2 (ii), it holds that $\|SCC(\bar{\varphi}^K) - SCC(\hat{\varphi}^K)\|_\infty \leq \|\bar{\varphi}^K - \hat{\varphi}^K\|_\infty$, and thus using that $\bar{\varphi}^{K+1} = SCC(\bar{\varphi}^K)$, it holds that

$$\|\bar{\varphi}^{K+1} - SCC(\hat{\varphi}^K)\|_\infty \leq \sum_{k=1}^{K} \delta_k.$$

At the same time, since the $(K+1)$-th inexact $S\tilde{C}C$ is $\delta_{K+1}$-exact, we have that $\|SCC(\hat{\varphi}^K) - \hat{\varphi}^{K+1}\|_\infty \leq \delta_{K+1}$, and hence

$$\|\bar{\varphi}^{K+1} - \hat{\varphi}^{K+1}\|_\infty \leq \|\bar{\varphi}^{K+1} - SCC(\hat{\varphi}^K)\|_\infty + \|SCC(\hat{\varphi}^K) - \hat{\varphi}^{K+1}\|_\infty \leq \sum_{k=1}^{K+1} \delta_k.$$

□

**Proof of Theorem 8.** Let $m \in \mathcal{L}$ be a leaf node of the scenario tree $\mathcal{N}$ and consider $\bar{\nu}_m^r$ as defined in Definition 9 for $r = 1$. Observe that $v_m(x_{a(m)})$ can be interpreted as a mixed-integer value function $v_\omega(x)$ defined in (3) and that $\bar{\nu}_m^1(x_{a(m)})$ corresponds to the lower bound $\nu_\omega$ from Definition 2. Hence, by Theorem 1 (i) it holds that $\bar{\nu}_m^1(x_{a(m)}) \leq v_m(x_{a(m)})$ for all $x_{a(m)} \in X_{a(m)}$. Moreover, for any $r = 2, \ldots, T-1$, we have that $S_{[n:a(m)]}$ with $n = a^r(m)$ is an extended space of $X_{a(m)}$, see Definition 3, so that $\bar{\nu}_m^r$ corresponds to the lower bound $\bar{\nu}_\omega$ from Definition 4. By Theorem 2 (ii), it holds that

$$\bar{\nu}_m^r(x_{[n:a(m)]}) \leq v_m(x_{a(m)})$$

for all $x_{[a(m)]} \in S_{[a(m)]}$. Finally, let $r_1, r_2 \in \{1, \ldots, T-1\}$ with $r_1 > r_2$ be given and define $n_1 = a^{r_1}(m)$ and $n_2 = a^{r_2}(m)$. Note that it is possible to define $v_m$ on the extended space $S_{[n_2:a(m)]}$ as $v_m^{r_2}(x_{[n_2:a(m)]}) = v_m(x_{a(m)})$ for all $x_{[n_2:a(m)]} \in S_{[n_2:a(m)]}$. Hence, $v_m^{r_2}$ can be interpreted as a mixed-integer value function $v_\omega$ from (3) with $\bar{\nu}_m^{r_2}$ corresponding to the lower bound from Definition 2, and since $S_{[n_1:a(m)]}$ is an extended space of $S_{[n_2:a(m)]}$, with $\bar{\nu}_m^{r_1}$ corresponding to the lower bound from Definition 4. It follows directly from Theorem 2 (i) that for all $x_{[a(m)]} \in \operatorname{conv}(S_{[a(m)]})$,

$$\bar{\nu}_m^{r_2}(x_{[n_2:a(m)]}) \leq \bar{\nu}_m^{r_1}(x_{[n_1:a(m)]}).$$

This proves our hierarchy of lower bounds since $r_1 > r_2$ are arbitrarily given. □

**Proof of Proposition 4.** Let $x_{[a(m)]} \in S_{[a(m)]}$ be given. If $x_{[a^\rho(m):a(m)]}$ is an extreme point of $\text{conv}(S_{[a^\rho(m):a(m)]})$ for some $\rho = 1, \ldots, r$, then it follows from Theorem 2 (iii) that $\bar{\nu}^\rho_m(x_{[a^\rho(m):a(m)]}) = v_m(x_{a(m)})$, and thus by Theorem 8, it holds that

$$\bar{\nu}^\rho_m(x_{[a^\rho(m):a(m)]}) = \bar{\nu}^r_m(x_{[a^r(m):a(m)]}) = v_m(x_{a(m)}).$$

Hence, if there exists $\rho \in \{1, \ldots, r\}$ such that $x_{[a^\rho(m):a(m)]}$ is an extreme point of $\text{conv}(S_{[a^\rho(m):a(m)]})$, then (17) holds. $\square$

**Proof of Theorem 9.** We will prove the result by mathematical induction on the time stage $t = 1, \ldots, T$, starting at $t = T$ and moving back to $t = 1$. For $t = T$ and all nodes $n \in \mathcal{N}$, the inequalities in (20)–(22) hold by Theorem 8 and since $\bar{Q}^1_n(x_n) = \ldots = \bar{Q}^{T-1}_n(x_{[n]}) = Q_n(x_n) = 0$ for all $x_{[n]} \in S_{[n]}$. Next, assume that (20)–(22) hold for all $n \in \mathcal{N}$ with $t_n > \bar{t}$ for some $1 \leq \bar{t} \leq T-1$. Then, for $n \in \mathcal{N}$ with $t_n = \bar{t}$, we have $Q_n(x_n) = \sum_{m \in C(n)} q_{nm} v_m(x_n), x_n \in X_n$, and $\bar{Q}^r_n(x_{[a^{r-1}(n):n]}) = \sum_{m \in C(n)} q_{nm} \bar{\nu}^r_m(x_{[a^{r-1}(n):n]}), x_{[a^{r-1}(n):n]} \in S_{[a^{r-1}(n):n]}$, so that (21) holds for $n$ by (22) and the induction hypothesis. Similarly, the first inequality in (20) holds.

Next, it follows from (21) that $\hat{v}^r_n(x_{[a^p(n):a(n)]})$ as defined in (19) is a lower bound for $v_n(x_{a(n)})$ for every $r = 1, \ldots, T-1$. We can interpret $\hat{v}^r_n$ as a mixed-integer value function $v_\omega$ from Section 2 with $\bar{\nu}^r_n(x_{[a^r(n):a(n)]})$ corresponding to the lower bound $\bar{\nu}_\omega$ from Definition 4. By Theorem 2 (ii) it holds that $\bar{\nu}^r_n(x_{[a^r(n):a(n)]}) \leq \hat{v}^r_n(x_{[a^p(n):a(n)]})$, and thus $\bar{\nu}^r_n(x_{[a^r(n):a(n)]}) \leq v_n(x_{a(n)})$ for all $x_{[n]} \in S_{[n]}$.

Finally, let $r_1, r_2 \in \{1, \ldots, T-1\}$ with $r_1 > r_2$ be given. We define the auxiliary lower bound $\bar{\nu}^{r_1,r_2}_n(x_{[a^{r_1}(n):a(n)]})$ as

$$\bar{\nu}^{r_1,r_2}_n(x_{[a^{r_1}(n):a(n)]}) := \min_{x_n, \theta_n} \left\{ c_n^\top x_n + \theta_n : (x_{[a^{r_1}(n):a(m)]}, \theta_n) \in \text{conv}(\Theta^{r_1,r_2}_n) \right\},$$

where

$$\Theta^{r_1,r_2}_n := \left\{ (x_{[a^{r_1}(n):n]}, \theta_n) \in S_{[a^{r_1}(n):n]} \times \mathbb{R} : \theta_n \geq \bar{Q}^{r_2}_n(x_{[a^{r_2-1}(n):n]}) \right\}.$$

Since $\bar{Q}^{r_1}_n(x_{[a^{r_1-1}(n):n]}) \geq \bar{Q}^{r_2}_n(x_{[a^{r_2-1}(n):n]})$ by (21), it follows immediately that $\Theta^{r_1}_n \subseteq \Theta^{r_1,r_2}_n$, and thus $\bar{\nu}^{r_1}_n(x_{[a^{r_1}(n):a(n)]}) \geq \bar{\nu}^{r_1,r_2}_n(x_{[a^{r_1}(n):a(n)]})$. On the other hand, since $S_{[a^{r_1}(n):a(n)]}$ is an extended space of $S_{[a^{r_2}(n):a(n)]}$, we can consider

$$v^{r_2}_n(x_{[a^{r_2}(n):a(n)]}) := \min_{x_n, \theta_n} \left\{ c_n^\top x_n + \theta_n : (x_{[a^{r_2}(n):n]}, \theta_n) \in \Theta^{r_2}_n \right\},$$

as a mixed-integer value function $v_\omega$ from Section 2, where $\bar{\nu}^{r_2}_n$ corresponds to the convex lower bound $\nu_\omega$ from Definition 2, and $\bar{\nu}^{r_1,r_2}_n$ corresponds to the convex lower bound $\bar{\nu}_\omega$ from Definition 4. Hence, by Theorem 2 (i) it follows that $\bar{\nu}^{r_2}_n(x_{[a^{r_2}(n):a(n)]}) \leq \bar{\nu}^{r_1,r_2}_n(x_{[a^{r_1}(n):a(n)]})$, and thus it holds that $\bar{\nu}^{r_2}_n(x_{[a^{r_2}(n):a(n)]}) \leq \bar{\nu}^{r_1}_n(x_{[a^{r_1}(n):a(n)]})$ for all $x_{[n]} \in \text{conv}(S_{[n]})$. Since $r_1 > r_2$ are arbitrarily given, we conclude that (22) holds for all $n \in \mathcal{N}$ with $t_n = \bar{t}$. The proof now follows by mathematical induction. $\square$

**Proof of Proposition 5.** Since by definition $Q_n(x_n) = \sum_{m \in C(n)} q_{nm} v_m(x_n)$ and $\bar{Q}^r_n(x_{[a^{r-1}(n):n]}) = \sum_{m \in C(n)} q_{nm} \bar{\nu}^r_m(x_{[a^{r-1}(n):n]})$, and $\bar{\nu}^r_m$, $m \in C(n)$, are lower bounds for $v_m$, it follows directly that (i) holds.

To prove (ii), suppose that for some $\rho = 1, \ldots, T-1$, it holds that an optimal solution $(x^*_n, \theta^*_n)$ of the minimization problem in $\hat{v}^\rho_n(x_{[a^\rho(n):a(n)]})$ satisfies $\theta^* = Q_n(x^*_n)$. Then, $\hat{v}^\rho_n(x_{[a^\rho(n):a(n)]}) = v_n(x_{a(n)})$. Moreover, if $(x_{[a^\rho(n):a(n)]}, \bar{Q}^\rho_n(x_{[a^{\rho-1}(n):n]}))$ is an extreme point of $\text{conv}(\Theta^\rho_n)$, then it follows by Theorem 2 (iii) that $\bar{\nu}^\rho_n(x_{[a^\rho(n):a(n)]}) = \hat{v}^\rho_n(x_{[a^\rho(n):a(n)]})$. The result in (ii) follows by Theorem 9 since if the lower bound is tight for $\rho$, then it will be tight for all $r = 1, \ldots, T-1$ with $r \geq \rho$. $\square$

**Proof of Proposition 6.** We first prove (ii) before proving (i). To do so, it suffices to prove that for $\varphi \in \Phi$ and $n \in \mathcal{N}$, we have $\hat{\varphi}_n \geq \varphi_n$ for $\hat{\varphi} := SCC_n(\varphi)$. Indeed, then for $m \in \mathcal{N}$ with $m \neq m$, it holds that $\hat{\varphi}_m(x_{[m]}, \theta_{[a(m)]}) = \varphi_m(x_{[m]}, \theta_{[a(m)]})$ for all $(x_{[m]}, \theta_{[a(m)]}) \in \Theta_{[m]}(\hat{\varphi}_{[a(m)]}) \subseteq \Theta_{[m]}(\varphi_{[a(m)]})$.

To prove that $\hat{\varphi}_n \geq \varphi_n$, observe that $\Theta_{[n]}(\hat{\varphi}_{a(n)]}) = \Theta_{[n]}(\varphi_{[a(n)]})$. Moreover, if $n \in \mathcal{L}$, then$\hat{\varphi}_n \geq \varphi_n$ by definition. On the other hand, if $n \in \mathcal{N} \backslash \mathcal{L}$, then byDefinition 12 it holds for all $(x_{[n]}, \theta_{a(n)]}) \in \operatorname{conv}(\Theta_{[n]}(\varphi_{[a(n)]}))$ that

$$\hat{\varphi}_n(x_{[n]}, \theta_{[a(n)]}) \geq \sup_{\tau_m \geq 0} \sup_{\substack{(\alpha_m, \beta_m): \\ (\alpha_m, \beta_m, \tau_m) \in \Pi_m(\varphi_{[n]})}} \left\{ \frac{\sum_{m \in C(n)} q_{nm} \Big( \alpha_m - \beta_m^\top (x_{[n]}, \theta_{[a(n)]}) \Big)}{1 + \sum_{m \in C(n)} q_{nm} \tau_m} \right\}.$$

For all $m \in C(n)$, we have that $(\alpha_m, \beta_m, \tau_m) \in \Pi_m(\varphi_{[n]})$ if and only if

$$\hat{v}_m(x_{[n]}, \theta_{[a(n)]}) + \tau_m \theta_n \geq \alpha_m - \beta_m^\top (x_{[n]}, \theta_{[a(n)]}), \quad \forall (x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]}).$$

Since $\hat{\Theta}_{[n]}(\varphi_{[n]})$ is bounded, there exists $L \in \mathbb{R}$ such that for all $m \in C(n)$,

$$\begin{aligned} L &\leq \min_{(x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]})} \hat{v}_m(x_{[n]}, \theta_{[n]}) \\ &= \min_{(x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]})} \min_{x_m \in \mathcal{X}_m(x_n)} c_m^\top x_m + \varphi_m(x_{[m]}, \theta_{[n]}). \end{aligned}$$

Here, we use that $\varphi \in \Phi$, and thus $c_m^\top x_m + \varphi_m(x_{[m]}, \theta_{[n]})$ is a convex polyhedral function minimized over a compact set. Using this lower bound $L$, we conclude for all $m \in C(n)$ that if $(\alpha_m, \beta_m, \tau_m)$ satisfies

$$L + \tau_m \theta_n \geq \alpha_m - \beta_m^\top (x_{[n]}, \theta_{[a(n)]}), \quad \forall (x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]}),$$

then $(a_m, \beta_m, \tau_m) \in \Pi_m(\varphi_{[n]})$. Hence, for every $m \in C(n)$, $\tau_m > 0$, and $(x_{[n]}, \theta_{[a(n)]}) \in \operatorname{conv}(\Theta_{[n]}(\varphi_{[a(n)]}))$, we have

$$\sup_{\substack{(\alpha_m, \beta_m): \\ (\alpha_m, \beta_m, \tau_m) \in \Pi_m(\varphi_{[n]})}} \Big\{ \alpha_m - \beta_m^\top (x_{[n]}, \theta_{[a(n)]}) \Big\} \geq \operatorname{co}\Big( L + \tau_m \varphi_n + \delta_{\Theta_{[n]}(\varphi_{[a(n)]})} \Big)(x_{[n]}, \theta_{[a(n)]}).$$

Applying this inequality and letting $\tau_m \to +\infty$ for every $m \in C(n)$, yields for all $(x_{[n]}, \theta_{[a(n)]}) \in \operatorname{conv}(\Theta_{[n]}(\varphi_{[a(n)]}))$ that

$$\begin{aligned} \hat{\varphi}_n(x_{[n]}, \theta_{[a(n)]}) &\geq \sup_{\tau_m \geq 0} \left\{ \frac{\sum_{m \in C(n)} q_{nm} \operatorname{co}\Big( L + \tau_m \varphi_n + \delta_{\Theta_{[n]}(\varphi_{[a(n)]})} \Big)(x_{[n]}, \theta_{[a(n)]})}{1 + \sum_{m \in C(n)} q_{nm} \tau_m} \right\} \\ &\geq \lim_{\tau_m \to +\infty} \left\{ \frac{\sum_{m \in C(n)} q_{nm} \operatorname{co}\Big( L + \tau_m \varphi_n + \delta_{\Theta_{[n]}(\varphi_{[a(n)]})} \Big)(x_{[n]}, \theta_{[a(n)]})}{1 + \sum_{m \in C(n)} q_{nm} \tau_m} \right\} \\ &= \operatorname{co}\Big( \varphi_n + \delta_{\Theta_{[n]}(\varphi_{[a(n)]})} \Big)(x_{[n]}, \theta_{[a(n)]}) \\ &= \varphi_n(x_{[n]}, \theta_{[a(n)]}), \end{aligned}$$

where the last equality holds since $\varphi \in \Phi$, and thus $\varphi_n$ is convex. We conclude that (ii) holds.

To prove (i), observe that $\hat{\varphi}_m = \varphi_m$ for all $m \in \mathcal{N}$ with $m \neq n$, and thus $\hat{\varphi}_m$ is convex polyhedral, since $\varphi \in \Phi$. Moreover, by definition of the $SCC_n$ operator, $\hat{\varphi}_n$ is the pointwise maximum of affine functions, and hence convex. Similar to Proposition 1 in van der Laan and Romeijnders [20], we can show that $\hat{\varphi}_n$ is also polyhedral. Hence, it remains to show that (26) holds for $\hat{\varphi}$ for all $n \in \mathcal{N}$ and $(x_{[n]}, \theta_{[a(n)]}) \in \Theta_{[n]}(\varphi_{[a(n)]})$. Since $\hat{\varphi} \geq \varphi$ by (ii), and $\varphi \in \Phi$, we conclude that (26) holds for all $m \in \mathcal{N}$ with $m \neq n$. Indeed, since $\hat{\varphi}_m = \varphi_m$, only the right-hand side in (26) may increase by applying the scaled cut closure procedure at node $n$, and inaddition the inequality is require to hold for a potentially smaller set $\Theta_{[m]}(\hat{\varphi}_{[a(m)]})$. To show that (26) holds for $\hat{\varphi}_n$ and all $(x_{[n]}, \theta_{[a(n)]}) \in \Theta_{[n]}(\varphi_{[a(n)]})$, observe that $\Theta_{[n]}(\hat{\varphi}_{[a(n)]}) = \Theta_{[n]}(\varphi_{[a(n)]})$, and moreover since $\varphi \in \Phi$, the functions $\varphi_m(x_{[m]}, \theta_{[n]})$ used in the definition of $\hat{v}_m$ in Definition 12 to apply the $SCC_n$ operator satisfy (26), and thus the result of applying the inexact $SCC_n$ operator is upper bounded by applying the exact $SCC$ operator, cf. Definition 7, to $\varphi_n$, which is upper bounded by the right-hand side in (26). Hence, $\hat{\varphi} \in \Phi$.

Finally, to prove (iii), let $\varphi, \varphi' \in \Phi$ with $\varphi \geq \varphi'$ and $n \in \mathcal{N} \backslash \mathcal{L}$ be given. We will show that for $m \in C(n)$, we have $\Pi_m(\varphi_{[n]}) \supseteq \Pi_m(\varphi'_{[n]})$, implying that $SCC_n(\varphi) \geq SCC_n(\varphi')$. Hence, let $(\alpha_m, \beta_m \tau_m) \in \Pi_m(\varphi'_{[n]})$ be given for some $m \in C(n)$. Then, for all $(x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi'_{[n]})$ it holds that

$$\hat{v}'_m(x_{[n]}, \theta_{n]}) \geq \alpha_m - \beta_m^\top (x_{[n]}, \theta_{[a(n)]}) - \tau_m \theta_n, \tag{34}$$

where $\hat{v}'_m(x_{[n]}, \theta_n]) := \min_{x_m \in \mathcal{X}_m(x_n)} c_m^\top x_m + \varphi'_m(x_{[m]}, \theta_{[n]})$. Since $\varphi \geq \varphi'$, it holds that $\hat{\Theta}_{[n]}(\varphi_{[n]}) \subseteq \hat{\Theta}_{[n]}(\varphi'_{[n]})$ and $\varphi_m \geq \varphi'_m$, so that (34) implies that

$$\hat{v}_m(x_{[n]}, \theta_{n]}) \geq \alpha_m - \beta_m^\top (x_{[n]}, \theta_{[a(n)]}) - \tau_m \theta_n,$$

for all $(x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]})$, and thus $(\alpha_m, \beta_m, \tau_m) \in \Pi_m(\varphi_{[n]})$. We conclude that $\Pi_m(\varphi_{[n]}) \supseteq \Pi_m(\varphi'_{[n]})$ for all $m \in C(n)$, and thus (iii) holds. $\square$

**Proof of Lemma 2.** The second inequality holds since for every $(x_{[m]}, \theta_{[n]}) \in \Theta_{[m]}(\varphi_{[n]})$ we have that $\mathrm{co}(Q_m + \delta_{\Theta_{[m]}(\varphi_{[n]})})(x_{[m]}, \theta_{[n]}) \leq Q_m(x_m)$. To prove the first inequality we use that by definition of the closed convex envelope, we have

$$\mathrm{co}(v_m + \delta_{\Theta_{[n]}(\varphi_{[n]})})(x_{[n]}, \theta_{[n]}) = \sup_{(\alpha_n, \beta_n) \in \hat{\Pi}_n(\varphi_{[n]})} \left\{ \alpha_n - \beta_n^\top (x_{[n]}, \theta_{[n]}) \right\},$$

where $\hat{\Pi}_n(\varphi_{[n]}) := \{(\alpha_n, \beta_n) : v_m(x_n) \geq \alpha_n - \beta_n^\top (x_{[n]}, \theta_{[n]}) \ \forall (x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]})\}$. Since $v_m(x_n)$ is a minimization over $x_m$, it holds for all $(\alpha_n, \beta_n) \in \hat{\Pi}_n(\varphi_{[n]})$ that

$$c_m^\top x_m + Q_m(x_m) \geq \alpha_n - \beta_n^\top (x_{[n]}, \theta_{[n]}) \qquad \forall (x_{[m]}, \theta_{[n]}) \in \Theta_{[m]}(\varphi_{[n]}),$$

and since the right-hand side is affine in $(x_{[n]}, \theta_{[n]})$, it also holds that

$$c_m^\top x_m + \mathrm{co}(Q_m + \delta_{\Theta_{[m]}(\varphi_{[n]})})(x_{[m]}, \theta_{[n]}) \geq \alpha_n - \beta_n^\top (x_{[n]}, \theta_{[n]}) \quad \forall (x_{[m]}, \theta_{[n]}) \in \Theta_{[m]}(\varphi_{[n]}).$$

Since for all $(x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]})$ this inequality holds for all $x_m \in \mathcal{X}_m(x_n)$, we can minimize the left-hand side over $x_m \in \mathcal{X}_m(x_n)$, yielding for all $(x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]})$ that

$$\min_{x_m \in \mathcal{X}_m(x_n)} c_m^\top x_m + \mathrm{co}(Q_m + \delta_{\Theta_{[m]}(\varphi_{[n]})})(x_{[m]}, \theta_{[n]}) \geq \alpha_n - \beta_n^\top (x_{[n]}, \theta_{[n]}),$$

which is equivalent to

$$\hat{v}_m(x_{[n]}, \theta_{[n]}) \geq \alpha_n - \beta_n^\top (x_{[n]}, \theta_{[n]}) \qquad \forall (x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]}). \tag{35}$$

Since the inequality in (35) holds for all $(\alpha_n, \beta_n) \in \hat{\Pi}_n(\varphi_{[n]})$, we conclude that for all $(x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]})$, we have

$$\begin{aligned}\hat{v}_m(x_{[n]}, \theta_{[n]}) &\geq \sup_{(\alpha_n, \beta_n) \in \hat{\Pi}_n(\varphi_{[n]})} \left\{ \alpha_n - \beta_n^\top (x_{[n]}, \theta_{[n]}) \right\} \\ &= \mathrm{co}(v_m + \delta_{\Theta_{[n]}(\varphi_{[n]})})(x_{[n]}, \theta_{[n]}),\end{aligned}$$

which implies that also the first inequality holds. □

**Proof of Theorem 10.** Interpret $Q_n$ as the expectation of l.s.c. value functions defined on the extended space $\Theta_{[n]}(\varphi_{[a(n)]})$ of $X_n$. Then, it follows from Definition 13 that the $SCC_n$ operator from Definition 12 is exact if it coincides with the $SCC$ operator from Definition 7. This is true if (28) holds for all $m \in C(n)$, since then Lemma 2 implies that for every $m \in C(n)$, the function $\hat{v}_m$ from Definition 13 satisfies

$$\mathrm{co}(v_m + \delta_{\hat{\Theta}_{[n]}(\varphi_{[n]})})(x_{[n]}, \theta_{[n]}) \leq \hat{v}_m(x_{[n]}, \theta_{[n]}) \leq v_m(x_n)$$

for all $(x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]})$, which by Proposition 3 is a sufficient condition for the $SCC_n$ operator to be exact cf. Definition 7. □

**Proof of Proposition 7.** Observe that if (29) holds for all $m \in C(n)$, that then $\hat{v}_m$, defined for all $(x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]})$ as

$$\hat{v}_m(x_{[n]}, \theta_{[n]}) = \min_{x_m \in \mathcal{X}_m(x_n)} c_m^\top x_m + \varphi_m(x_{[m]}, \theta_{[n]}),$$

satisfies for all $(x_{[n]}, \theta_{[n]}) \in \hat{\Theta}_{[n]}(\varphi_{[n]})$ that

$$\begin{aligned}\hat{v}_m(x_{[n]}, \theta_{[n]}) &\geq \min_{x_m \in \mathcal{X}_m(x_n)} c_m^\top x_m + \mathrm{co}(Q_m + \delta_{\Theta_{[m]}(\varphi_{[n]})})(x_{[m]}, \theta_{[n]}) - \delta \\ &\geq \mathrm{co}(v_m + \delta_{\hat{\Theta}_{[n]}(\varphi_{[n]})})(x_{[n]}, \theta_{[n]}) - \delta,\end{aligned}$$

where the last inequality follows from Lemma 2. Hence, applying Theorem 7 yields that the $SCC_n$ operator is $\delta$-exact. □

**Proof of Theorem 11.** Consider any subsequence $\{\hat{n}_{k_l}\}_{l\in\mathbb{N}}$ of $\{n_k\}_{k\in\mathbb{N}}$, and let $\{\hat{\varphi}^l\}_{l\in\mathbb{N}}$ denote the corresponding sequence of outer approximations, defined as

$$\hat{\varphi}^l := SCC_{n_{k_l}}(\hat{\varphi}^{l-1}), \qquad l \in \mathbb{N},$$

with $\hat{\varphi}^0 := \varphi^0$. By monotonicity of the scaled cut closure operator, see Proposition 6, it holds that for every $l \in \mathbb{N}$,

$$\varphi^{k_l} \geq \hat{\varphi}^l.$$

Hence, $\varphi_1^k$ converges uniformly to $\mathrm{co}_{X_1}(Q_1)$ if $\hat{\varphi}_1^l$ does, since $\varphi_1^k \leq \mathrm{co}_{X_1}(Q_1)$ for all $k \in \mathbb{N}$. Thus, it suffices to prove that there exists a subsequence $\{\hat{n}_{k_l}\}_{l\in\mathbb{N}}$ of $\{n_k\}_{k\in\mathbb{N}}$ such that $\hat{\varphi}_1^l$ converges uniformly to $\mathrm{co}_{X_1}(Q_1)$.

Next, we prove the existence of such a subsequence $\{\hat{n}_{k_l}\}_{l\in\mathbb{N}}$. Observe that by definition of an admissable sequence, any sequence $\{\hat{n}_l\}_{l\in\mathbb{N}} \subseteq \mathcal{N}$ is a subsequence of $\{n_k\}_{k\in\mathbb{N}}$. To prove the result, we will prove the stronger claim that for any feasible initial set of outer approximations $\varphi^0 = \{\varphi_n^0\}_{n\in\mathcal{N}} \in \Phi$, for every $n \in \mathcal{N}$, and for every $\epsilon > 0$, there exists a finite sequence $\{n_k\}_{k=1}^N \subseteq \mathcal{C}(n)$ for some $N \in \mathbb{N}$ such that

$$\|\varphi_n^N - \mathrm{co}(Q_n + \delta_{\Theta_{[n]}(\varphi^0_{[a(n)]})})\|_\infty < \epsilon,$$

for $\varphi^N$ iteratively defined by $\varphi^k := SCC_{n_k}(\varphi^{k-1})$, $k = 1, \ldots, N$. The result for $n = 1$ then implies that $\varphi_1^k$ converges uniformly to $\mathrm{co}_{X_1}(Q_1)$ as desired. Here, we let $\mathcal{C}(n)$ denote te set of all descendant nodes of $n$ including $n$ itself, for all $n \in \mathcal{N}$.

We will prove the claim by mathematical induction on the time $t = 1, \ldots, T$, starting at $t = T$ and moving backward to $t = 1$. For all leaf nodes $n \in \mathcal{L}$, i.e., for all nodes $n$ in time stage $T$, it holds that $Q_n \equiv 0$, so that after a single $SCC_n$-operation it holds that

$$\|\varphi_n^1 - \mathrm{co}(Q_n + \delta_{\Theta_{[n]}(\varphi^0_{[a(n)]})})\|_\infty = 0.$$

Hence, the claim holds for all $n \in \mathcal{L}$. Next, assume that the claim holds for all nodes $n \in \mathcal{N}$ with $t_n \geq \bar{t}$ for some arbitrary $1 < \bar{t} \leq T$. We will prove that under this assumption the claim also holds for all nodes $n \in \mathcal{N}$ with $t_n = \bar{t} - 1$. To do so, let $n \in \mathcal{N}$ with $t_n = \bar{t} - 1$ and $\epsilon > 0$ be given. Define $\{\bar{\varphi}^\kappa\}_{\kappa \in \mathbb{N}}$ as the sequence of lower bounds obtained by iteratively applying exact scaled cut closure operators at node $n$. It follows from Theorem 6 that the sequence $\{\bar{\varphi}_n^\kappa\}_{\kappa \in \mathbb{N}}$ converges uniformly to $\mathrm{co}(Q_m + \delta_{\Theta_{[n]}(\varphi^0_{[a(n)]})})$. Hence, there exists $K \in \mathbb{N}$ such that for all $\kappa \in \mathbb{N}$ with $\kappa \geq K$,

$$\|\bar{\varphi}_n^\kappa - \mathrm{co}(Q_n + \delta_{\Theta_{[n]}(\varphi^0_{[a(n)]})})\|_\infty < \frac{\epsilon}{2}.$$

The exact scaled closure operations differ from the *inexact* scaled cut closure operations $SCC_n$ that we intend to apply at node $n$. However, Corollary 1 shows that if we apply $K$ times a $\delta$-exact scaled cut closure operations with $\delta := \epsilon/(2K)$, leading to the sequence $\{\hat{\varphi}^\kappa\}_{\kappa=0}^K$, with $\hat{\varphi}^0 = \varphi^0$, then $\|\bar{\varphi}^K - \hat{\varphi}^K\|_\infty \leq \epsilon/(2K) \times K = \epsilon/2$, and thus

$$\begin{aligned}\|\hat{\varphi}_n^K - \mathrm{co}(Q_n + \delta_{\Theta_{[n]}(\varphi^0_{a(n)]})})\|_\infty &\leq \|\hat{\varphi}_n^K - \bar{\varphi}_n^K\|_\infty + \|\bar{\varphi}_n^K - \mathrm{co}(Q_n + \delta_{\Theta_{[n]}(\varphi^0_{[a(n)]})})\|_\infty \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.\end{aligned}$$

Thus, to prove our claim it suffices to prove that before the $\kappa$-th $SCC_n$ operation, $\kappa = 1, \ldots, K$, there exists a finite sequence of $SCC_{n_l^\kappa}$ operations with $\{n_l^\kappa\}_{l=1}^{N_\kappa} \subseteq \mathcal{C}(n)$, such that the $\kappa$-th $SCC_n$ operation becomes $(\epsilon/(2N))$-exact when first these $SCC_{n_l^\kappa}$ operations are iteratively carried out.

By Proposition 7 a sufficient condition to achieve this is that for every $\kappa = 1, \ldots, K$, and $m \in C(n)$, there exists a finite sequence $\{n_l^\kappa\}_{l=1}^{N_{\kappa,m}} \subseteq \mathcal{C}(m)$ for some $N_{\kappa,m} \in \mathbb{N}$ such that

$$\|\hat{\hat{\varphi}}_m^{N_{\kappa,m}} - \mathrm{co}(Q_m + \delta_{\Theta_{[m]}(\hat{\varphi}^{\kappa-1}_{[n]})})\|_\infty \leq \frac{\epsilon}{2N}$$

for $\hat{\hat{\varphi}}^l$ iteratively defined by $\hat{\hat{\varphi}}^l := SCC_{n_l^\kappa}(\hat{\hat{\varphi}}^{l-1})$, $l = 1, \ldots, N_{\kappa,m}$, with $\hat{\hat{\varphi}}^0 := \hat{\varphi}^{\kappa-1}$. By the induction hypothesis, such sequences $\{n_l^\kappa\}_{l=1}^{N_{\kappa,m}}$ do indeed exist for every $\kappa = 1, \ldots, K$, and $m \in C(n)$. Hence, carrying out all scaled cut closure operations in the correct order, there exists a finite sequence $\{n_k\}_{k=1}^N \subseteq \mathcal{C}(n)$ of length $N = K + \sum_{\kappa=1}^K \sum_{m \in C(n)} N_{\kappa,m}$ such that $\|\varphi_n^N - \mathrm{co}(Q_n + \delta_{\Theta_{[n]}(\varphi^0_{a(n)]})})\|_\infty < \epsilon$ for $\varphi^N$ iteratively defined by $\varphi^k := SCC_{n_k}(\varphi^{k-1})$, $k = 1, \ldots, N$. $\square$

**Proof of Theorem 12.** To prove the theorem, we will first show by mathematical induction that the following claim holds.

*Claim:* Let $r = 1, \ldots, T-1$, be given. Then, for all $n \in \mathcal{N}$, and $x_{[a^{r-1}(n):n]} \in \mathrm{conv}(S_{[a^{r-1}:n]})$, it holds that

$$\bar{\varphi}_n^r(x_{[a^{r-1}(n):n]}) = \mathrm{co}\Big(\sum_{m \in C(n)} q_{nm} \hat{v}_m^r + \delta_{S_{[a^{r-1}(n):n]}}\Big)(x_{[a^{r-1}(n):n]}), \tag{36}$$

where $\hat{v}_m^r(x_{[a^{r-1}(n):n]}) = \min_{x_m \in \mathcal{X}_m(x_n)} c_m^\top x_m + \bar{\varphi}_m^r(x_{[a^{r-1}(m):m]})$ for all $m \in C(n)$ and $x_{[a^{r-1}(n):n]} \in S_{[a^{r-1}(n):n]}$. Moreover, $\bar{Q}_n^r(x_{[a^{r-1}(n):n]}) \leq \bar{\varphi}_n^r(x_{[a^{r-1}(n):n]})$ for all $x_{[a^{r-1}(n):n]} \in \mathrm{conv}(S_{[a^{r-1}(n):n]})$, and $\bar{\varphi}_n^r(x_{[a^{r-1}(n):n]}) \leq Q_n(x_n)$ for all $x_{[a^{r-1}(n):n]} \in S_{[a^{r-1}(n):n]}$.

For $n \in \mathcal{L}$, the claim holds since $Q_n \equiv 0$, and $\bar{\varphi}_n^r(x_{[a^{r-1}(n):n]}) = \bar{Q}_n^r(x_{[a^{r-1}(n):n]}) = 0$ for all $x_{[a^{r-1}(n):n]} \in \operatorname{conv}(S_{[a^{r-1}(n):n]})$. Next, assume that the claim holds for all $n \in \mathcal{N}$ with $t_n \geq \bar{t}$ for some $1 < \bar{t} \leq T$. Then, for $n \in \mathcal{N}$ with $t_n = \bar{t} - 1$, observe that the scaled cut closure operator $SCC_n^r$ does not depend on previous-stage lower bounds $\varphi_{[a(n)]}$. Hence, in the limit, $\bar{\varphi}_n^r$ will be equivalent to applying the exact scaled cut closure operator $SCC$ from Definition 7 iteratively with respect to the expected value function $Q_n^r(x_{[a^{r-1}(n):n]}) = \sum_{m \in C(n)} q_{nm} \hat{v}_m^r(x_{[a^{r-1}(n):n]})$, with $\hat{v}_m^r$ as in the claim. Hence, it follows directly from Theorem 6 that (36) holds. Moreover, by the induction hypothesis $\bar{\varphi}_m^r(x_{[a^{r-1}(m):m]}) \leq Q_m(x_m)$ for all $x_{[a^{r-1}(m):m]} \in S_{[a^{r-1}(m):m]}$, and thus $\hat{v}_m^r(x_{[a^{r-1}(n):n]}) \leq v_m(x_n)$ for all $x_{[a^{r-1}(n):n]} \in S_{[a^{r-1}(n):n]}$, which implies via (36) that $\bar{\varphi}_n^r(x_{[a^{r-1}(n):n]}) \leq Q_n(x_n)$ for all $x_{[a^{r-1}(n):n]} \in S_{[a^{r-1}(n):n]}$. Furthermore, since by the induction hypothesis $\bar{\varphi}_m^r(x_{[a^{r-1}(m):m]}) \geq \bar{Q}_m^r(x_{[a^{r-1}(m):m]})$ for all $x_{[a^{r-1}(m):m]} \in \operatorname{conv}(S_{[a^{r-1}(m):m]})$, $m \in C(n)$, it holds that for all $m \in C(n)$ and $x_{[a^{r-1}(n):n]} \in S_{[a^{r-1}(n):n]}$, that

$$\hat{v}_m^r(x_{[a^{r-1}(n):n]}) \min_{x_m \in \mathcal{X}_m(x_n)} c_m^\top x_m + \bar{Q}_m^r(x_{[a^{r-1}(m):m]}) \geq \bar{\nu}_m^r(x_{[a^{r-1}(n):n]}), \tag{37}$$

where the convex lower bound $\bar{\nu}_m^r(x_{[a^{r-1}(n):n]})$ is defined in Definition 10. Hence, for all $x_{[a^{r-1}(n):n]} \in \operatorname{conv}(S_{[a^{r-1}(n):n]})$, we have

$$\begin{aligned}
\bar{\varphi}_n^r(x_{[a^{r-1}(n):n]}) &= \operatorname{co}\Big(\sum_{m \in C(n)} q_{nm} \hat{v}_m^r + \delta_{S_{[a^{r-1}(n):n]}}\Big)(x_{[a^{r-1}(n):n]}) \\
&\geq \sum_{m \in C(n)} q_{nm} \operatorname{co}\Big(\hat{v}_m^r + \delta_{S_{[a^{r-1}(n):n]}}\Big)(x_{[a^{r-1}(n):n]}) \\
&\geq \sum_{m \in C(n)} q_{nm} \bar{\nu}_m^r(x_{[a^{r-1}(n):n]}) \\
&= \bar{Q}_n^r(x_{[a^{r-1}(n):n]}),
\end{aligned}$$

where the second inequality holds since $\bar{\nu}_m^r$ is a convex polyhedral lower bound of $\hat{v}_m^r$, see (37).

It remains to show that for all $n \in \mathcal{N}$ and $x_{[n]} \in \operatorname{conv}(S_{[n]})$, it holds that

$$\bar{\varphi}_n^1(x_n) \leq \ldots \leq \bar{\varphi}_n^{T-1}(x_{[n]}). \tag{38}$$

To do so, let $r_1, r_2 \in \{1, \ldots, T-1\}$ with $r_1 > r_2$ be given. We will prove that $\bar{\varphi}_n^{r_1}(x_{[a^{r_1-1}(n):n]}) \geq \bar{\varphi}_n^{r_2}(x_{[a^{r_2-1}(n):n]}$ for all $x_{[n]} \in \operatorname{conv}(S_{[n]})$ by mathematical induction, so that (38) holds. Observe that for $n \in \mathcal{L}$, we have $\bar{\varphi}_n^{r_1}(x_{[a^{r_1-1}(n):n]}) = \bar{\varphi}_n^{r_2}(x_{[a^{r_2-1}(n):n]}) = 0$ for all $x_{[n]} \in \operatorname{conv}(S_{[n]})$. Next, assume that $\bar{\varphi}_n^{r_1}(x_{[a^{r_1-1}(n):n]}) \geq \bar{\varphi}_n^{r_2}(x_{[a^{r_2-1}(n):n]})$ for all $x_{[n]} \in \operatorname{conv}(S_{[n]})$ for all $n \in \mathcal{N}$ with $t_n \geq \bar{t}$ for some $1 < \bar{t} \leq T$. Then, for $n \in \mathcal{N}$ with $t_n = \bar{t} - 1$, we have that

$$\bar{\varphi}_n^{r_1}(x_{[a^{r_1-1}(n):n]}) = \operatorname{co}\Big(\sum_{m \in C(n)} q_{nm} \hat{v}_m^{r_1} + \delta_{S_{[a^{r_1-1}(n):n]}}\Big)(x_{[a^{r_1-1}(n):n]}).$$

By the induction hypothesis, it holds that

$$\sum_{m \in C(n)} q_{nm} \hat{v}_m^{r_1}(x_{[a^{r_1-1}(n):n]}) \geq \sum_{m \in C(n)} q_{nm} \hat{v}_m^{r_2}(x_{[a^{r_2-1}(n):n]})$$

for all $x_{[n]} \in S_{[n]}$. Hence, $\bar{\varphi}_n^{r_2}$ is a convex polyhedral lower bound of the first function when defined on the extended space $\operatorname{conv}(S_{[a^{r_1-1}(n):n]})$. Since it is not necessarily the best convex polyhedral lower bound, we conclude that

$$\bar{\varphi}_n^{r_1}(x_{[a^{r_1-1}(n):n]}) \geq \bar{\varphi}_n^{r_2}(x_{[a^{r_2-1}(n):n]}) \qquad \forall\, x_{[n]} \in \operatorname{conv}(S_{[n]}).$$

$\square$

## Acknowledgments

The research of Ward Romeijnders has been supported by the Netherlands Organisation for Scientific Research [Grant VI.Vidi.211.066].

## References

[1] S. Ahmed, F. G. Cabral, and B. F. P. da Costa, *Stochastic Lipschitz dynamic programming*, Mathematical Programming, 191 (2022), pp. 755–793.

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

[3] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming*, Springer, New York, 1997.

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

[5] M. Daryalal, M. Bodur, and J. R. Luedtke, *Lagrangian dual decision rules for multistage stochastic mixed-integer programming*, Operations Research, 72 (2024), pp. 717–737.

[6] C. Füllner and S. Rebennack, *Non-convex nested Benders decomposition*, Mathematical Programming, 196 (2022), pp. 987–1024.

[7] D. Gade, S. Küçükyavuz, and S. Sen, *Decomposition algorithms with parametric Gomory cuts for two-stage stochastic integer programs*, Mathematical Programming, 144 (2014), pp. 39–64.

[8] F. Maggioni and G. C. Pflug, *Bounds and approximations for multistage stochastic programs*, SIAM Journal on Optimization, 26 (2016), pp. 831–855.

[9] L. Ntaimo, *Fenchel decomposition for stochastic mixed-integer programming*, Journal of Global Optimization, 55 (2013), pp. 141–163.

[10] M. Pereira and L. Pinto, *Multi-stage stochastic optimization applied to energy planning. mathematical programming*, Mathematical Programming, 52 (1991), pp. 359–375.

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

[12] Y. Qi and S. Sen, *The ancestral Benders' cutting plane algorithm with multi-term disjunctions for mixed-integer recourse decisions in stochastic programming*, Mathematical Programming, 161 (2017), pp. 193–235.

[13] R. T. Rockafellar and R. J.-B. Wets, *Scenarios and policy aggregation in optimization under uncertainty*, Mathematics of Operations Research, 16 (1991), pp. 119–147.

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

[15] B. Sandikçi and O. Y. Özaltin, *A scalable bounding method for multistage stochastic programs*, SIAM Journal on Optimization, 27 (2017), pp. 1772–1800.

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

[17] S. Sen and H. D. Sherali, *Decomposition with branch-and-cut approaches for two-stage stochastic mixed-integer programming*, Mathematical Programming, 106 (2006), pp. 203–223.

[18] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, 2009.

[19] H. D. Sherali and X. Zhu, *On solving discrete two-stage stochastic programs having mixed-integer first-and second-stage variables*, Mathematical Programming, 108 (2006), pp. 597–616.

[20] N. van der Laan and W. Romeijnders, *A converging Benders' decomposition algorithm for two-stage mixed-integer recourse models*, Operations Research, (2023). https://doi.org/10.1287/opre.2021.2223.

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

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