

Asymptotic convexity in multistage stochastic mixed-integer programs

Jinting Lin¹, Ward Romeijnders¹, E. Ruben van Beesten^{2,3}

¹Department of Operations, University of Groningen, Nettelbosje 2, 9747 AE, Groningen, The Netherlands.

²Econometric Institute, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA, Rotterdam, The Netherlands.

³Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology (NTNU), Alfred Getz' vei 3, 7043 Trondheim, Norway.

Contributing authors: jinting.lin@rug.nl; w.romeijnders@rug.nl; vanbeesten@ese.eur.nl;

Abstract

We consider multistage stochastic mixed-integer programs (M-SMIPs), and show that these problems are asymptotically convex. That is, we show that the expected cost-to-go functions in the problem converge to convex functions as the variability of the random parameters in the model increases. To derive this result, we first propose a novel adapted Gomory relaxation for deterministic mixed-integer value functions by selectively relaxing some of the non-negativity constraints of the basic variables. We show that the adapted Gomory relaxation is semi-periodic, and equal to the original mixed-integer value function on large parts of its domain. Then, we use these results to prove the desired asymptotic convexity of M-SMIPs. In contrast to earlier work on the two-stage case, our results also hold when only part of the right-hand side parameters are random, e.g., when some later-stage decisions are binary.

Keywords: Stochastic programming, Mixed-integer programming, Multistage recourse models, Asymptotic convexity

1 Introduction

Multistage stochastic mixed-integer programs (M-SMIPs) arise naturally in a wide range of decision-making problems under uncertainty, where mixed-integer decisions are made sequentially over multiple time stages, see, e.g., the textbooks [5, 25], and [12, 26, 35] for applications in capacity planning, supply chain optimization, and energy systems, respectively. However, the presence of integer decisions induces non-convexity in the expected cost-to-go functions across time stages. As a result, efficient decomposition approaches such as stochastic dual dynamic programming (SDDP) [17], which rely on convexity of the expected cost-to-go functions, cannot be applied directly or need to be adapted to the mixed-integer setting.

In this paper, we show that M-SMIPs are *asymptotically convex*: their expected cost-to-go functions converge to convex functions as the variability in the distribution of the random right-hand-side parameters in the model increases. The key observation is that randomness can have a convexifying effect that is absent in deterministic mixed-integer programs. While a deterministic mixed-integer value function is generally non-convex, the expected value function, i.e., the expected cost-to-go function, arising in M-SMIPs averages these non-convexities over different realizations of the random parameters. As the variability of the random parameters increases, this averaging effect becomes stronger, and the expected cost-to-go functions converge to convex functions. This convexifying effect may potentially be used to derive efficient approximate decomposition algorithms for M-SMIPs, analogous in spirit to SDDP for multistage stochastic linear programs.

In the literature, asymptotic convexity results have been established for two-stage SMIPs (2-SMIPs), initially for specific recourse structures, including simple integer recourse [13], mixed-integer recourse with a single recourse constraint [32], and totally unimodular integer recourse [21]. These studies show that as the variability of the random parameters increases, the expected mixed-integer value function asymptotically converges to a convex function. These results were subsequently extended to general risk-neutral [19, 28] and risk-averse [27, 29] 2-SMIPs.

These existing results apply only to 2-SMIPs in which *all* second-stage right-hand-side parameters are random. However, many practically relevant 2-SMIPs contain binary second-stage variables or constraints whose right-hand sides are independent of the random parameters. In addition, the value functions in general M-SMIPs are defined recursively through future expected cost-to-go functions, whose epigraph linearizations introduce additional constraints with deterministic right-hand side terms. In this paper, we derive asymptotic convexity of general M-SMIPs with right-hand side vectors that are not necessarily fully random.

We first study a generic mixed-integer value function, where some of the right-hand side components are fixed. We introduce a novel adaptation of the Gomory relaxation based on [11], in which we preserve the non-negativity constraints of some critical basic variables, and relax the remaining basic variables. This type of relaxation has been studied for pure integer programs [34], where it is shown that the relaxation is equal to the original integer value function on subsets of its domain. For our mixed-integer case, we prove that on large parts of its domain, the mixed-integer value function is equal to the adapted Gomory relaxation, and it is semi-periodic. That is, it can be expressed as

the sum of a linear and periodic function. This result may be of independent interest to the deterministic MIP community.

Using these properties of the generic mixed-integer value function, we construct a convex approximation of the expected mixed-integer value function, i.e., an α -approximation [30–32]. We derive a uniform error bound for this approximation that depends on the total variations of the probability density functions of the random parameters, and show that the expected approximation converges to the true expected value function as these total variations converge to zero. Thus, for any 2-SMIP whose value function satisfies the generic form considered in this paper, the expected value function converges to its α -approximation as the variability of the random parameters increases.

Subsequently, we use this two-stage result to prove the asymptotic convexity results for M-SMIPs. First, we consider the expected cost-to-go function at a given stage t . If the probability distributions of the later-stage random parameters are independent from those in stage t and can be represented by a finite scenario tree, then we can construct a convex approximation with total variation error bound that converges to zero if the total variations of the probability density functions of the random stage- t parameters converge to zero. Second, by recursively applying our α -approximation, we are able to prove that all expected cost-to-go functions in an M-SMIP can *simultaneously* converge to some convex functions if the variability of all the random parameters increases simultaneously.

Summarizing, our contributions are threefold.

- We define and study a novel adapted Gomory relaxation for a generic deterministic mixed-integer value function with some of its right-hand side parameters fixed. We show that, on large parts of the non-fixed domain, the value function is equal to the adapted Gomory relaxation, and it is semi-periodic, i.e., the sum of its LP-relaxation and a periodic function.
- For 2-SMIPs whose mixed-integer value function includes deterministic right-hand side parameters (e.g., in the case of binary second-stage decisions), we show that the expected value function converges to a convex function when the variability of the random parameters increases.
- For M-SMIPs, we prove two asymptotic convexity results.
 1. First, we show asymptotic periodicity of a single-stage expected cost-to-go function. If the uncertainty in stages $t+1, \dots, T$ can be represented by a finite scenario tree, then the stage- t expected cost-to-go function converges to a convex function as the variability of the stage- t random parameters increases.
 2. Second, we show simultaneous asymptotic convexity across stages: the expected cost-to-go functions in all stages converge to convex functions simultaneously as the variability of the random parameters in all stages increase simultaneously.

Our perspective differs from the algorithmic approaches in the literature, which tackle the M-SMIPs directly through decomposition methods. For example, [36] develop

stochastic dual dynamic integer programming (SDDiP), which extends SDDP to multistage stochastic integer programs with binary state variables through Lagrangian cuts. Then, [1] propose stochastic Lipschitz dynamic programming (SLDP), which constructs nonlinear Lipschitz cuts to approximate non-convex cost-to-go functions under a Lipschitz continuity assumption and complete continuous recourse. For a class of multistage non-convex mixed-integer nonlinear programs, [9] develop a non-convex nested Benders decomposition method based on dynamically refined mixed-integer linear outer approximations, regularization, binary-state approximations, and Lagrangian cuts. More recently, [20] construct a hierarchy of convex polyhedral lower bounds for the value functions and expected cost-to-go functions of M-SMIPs using scaled cuts. Finally, scenario decomposition approaches have also been proposed for multistage stochastic mixed-integer programs, including methods based on dual decomposition [6, 8] and progressive hedging [2, 4]. For a broader overview of decomposition methods for stochastic mixed-integer programming, we refer to [23].

These exact methods are useful for small to medium size problem instances, but tend to struggle computationally with larger instances. Our objective is not to develop another decomposition method that directly handles the non-convex cost-to-go functions. Instead, we study conditions under which these non-convex functions are close to convex approximations. This provides a complementary perspective: rather than tackling the non-convexity directly, we may identify conditions under which the non-convexity become negligible so that more efficient methods developed for convex multistage stochastic programs can be applied.

The remainder of this paper is organized as follows. In Section 2, we formally introduce M-SMIPs and state some general assumptions we make. In Section 3, we introduce our adapted Gomory relaxation and study its properties and relation to the original mixed-integer value function. In Section 4, we construct a convex approximation and derive a uniform error bound for the expected value function of 2-SMIPs. In Section 5, we derive the asymptotic convexity results for M-SMIPs. Finally, we end with a discussion in Section 6.

2 Formulation of M-SMIPs

We consider M-SMIPs of the form

$$\min_{x_1 \in X_1} \left\{ c_1^\top x_1 + Q_2(x_1) : W_1 x_1 = \omega_1 \right\}, \quad (1)$$

with the expected value function Q_{t+1} defined for stage $t = 1, \dots, T-1$, as

$$Q_{t+1}(x_t) := \mathbb{E}_{\omega_{t+1}} \left[v_{t+1}(\omega_{t+1}, x_t) \right]$$

with the underlying value function $v_{t+1}(\omega_{t+1}, x_t)$ defined for stage $t = 1, \dots, T-1$ as

$$v_{t+1}(\omega_{t+1}, x_t) = \min_{x_{t+1} \in X_{t+1}} \left\{ c_{t+1}^\top x_{t+1} + Q_{t+2}(x_{t+1}) : \right.$$

$$W_{t+1}x_{t+1} = \omega_{t+1} - T_{t+1}x_t \}, \quad (2)$$

with $Q_{T+1} \equiv 0$, ω_1 a fixed parameter, and X_t represents integrality and non-negativity constraints on x_t . Here, the random parameters ω_t are in the right-hand side vector of the constraints and they are stage-wise independent, i.e., independent from the random parameters of all other stages. Throughout this paper, we make the following assumptions for M-SMIPs.

- Assumption 1** (a) For every stage $t = 1, \dots, T$, we assume that the recourse matrix W_t consists of integer entries and is a complete recourse matrix. That is, for any right-hand side vector d , it holds that $\{y_t \in \mathbb{R}_+^{n_t} \times \mathbb{Z}_+^{\bar{n}_t} : W_t y_t = d\} \neq \emptyset$. Moreover, the technology matrix T_t and the cost vector c_t are rational.
- (b) For every stage $t = 2, \dots, T$, the value function v_t satisfies $v_t(\omega_t, x_{t-1}) > -\infty$.
- (c) The inequality $\mathbb{E}[\|\omega_t\|_\infty] < +\infty$ holds for every $t = 2, \dots, T$.

Here, Assumptions 1 (a)–(b) ensure that the value functions v_t are always finite, while Assumptions 1 (a)–(c) imply that the expected value functions Q_t are also finite. These assumptions are mild and are in line with the standard assumptions commonly imposed in the 2-SMIP literature, see, e.g., [19, 30].

Our goal is to prove that every expected value function Q_t is asymptotically convex. We will do so by constructing a convex approximation $\hat{Q}_t$ that is arbitrarily close to Q_t if the variability of the random parameters in the model is large enough.

In the literature, asymptotic convexity has been established for 2-SMIPs by exploiting the periodic behavior of the second-stage mixed-integer value function v_2 as a function of ω_2 [19]. This periodic behavior relies on the fact that in 2-SMIPs, v_2 is the optimal value of a mixed-integer linear program. In M-SMIPs, however, the objective function of the mixed-integer program defining v_t is not linear, due to its nested dependence on Q_{t+1} , meaning that existing results cannot be directly applied. If Q_{t+1} is convex polyhedral, which holds for example if later-stage decisions are continuous and all later-stage random parameters are finite discretely distributed, or Q_{t+1} represents a convex approximation of the true expected value function, then the objective function in v_t can be linearized via an epigraph formulation, i.e.,

$$v_t(\omega_t, x_{t-1}) = \min_{x_t \in X_t, \theta_{t+1}} \left\{ c_t^\top x_t + \theta_{t+1} : W_t x_t = \omega_t - T_t x_{t-1} \right. \\ \left. \theta_{t+1} - a_{t+1,i}^\top x_t \geq \beta_{t+1,i}, \forall i \in \mathcal{I}_{t+1} \right\}, \quad (3)$$

where Q_{t+1} is represented as $Q_{t+1}(x_t) = \max_{i \in \mathcal{I}_{t+1}} \{a_{t+1,i}^\top x_t + \beta_{t+1,i}\}$.

However, now part of the right-hand side parameters of the constraints in (3), i.e., $\beta_{t+1,i}$ for all $i \in \mathcal{I}_t$, are deterministic, violating the assumption in [19] that all right-hand side parameters are random. That is why in this paper we will first analyze the following generic mixed-integer value function v and its expectation Q , which represents any value function v_t of the form in (3) and any corresponding expected cost-to-go function Q_t , respectively.

Definition 1 We define the generic mixed-integer value function v as

$$v(\omega, x) := \min \left\{ q^\top y : Wy = (\omega - Tx, h), y \in \mathbb{Z}_+^n \times \mathbb{R}_+^{\bar{n}} \right\}, \quad (4)$$

where $x \in X \subseteq \mathbb{R}^p$, $\omega - Tx \in \mathbb{R}^{m_\omega}$, and $h \in \mathbb{R}^{m_h}$. Here, X denotes the feasible set for x , ω represents a random vector, and h is a vector with fixed elements. Moreover, we denote the expectation of $v(\omega, x)$ as $Q(x) := \mathbb{E}_\omega[v(\omega, x)]$.

We impose the following assumptions on the generic value function v , which are naturally equivalent to Assumption 1.

Assumption 2 For the generic mixed-integer value function $v(\omega, x)$ of (4), we assume that $W \in \mathbb{Z}^{(m_\omega + m_h) \times (n + \bar{n})}$, $-\infty < v(\omega, x) < \infty$, and $\mathbb{E}[\|\omega\|_\infty] < \infty$.

The generic mixed-integer value function v represents not only the value functions v_t of a broad class of M-SMIPs, but also the value functions of 2-SMIPs with deterministic second-stage constraints, for example when some second-stage decisions are restricted to be binary; see, e.g., [14–16].

In the next section, we begin by studying several properties of the deterministic generic value function v . Based on these properties, in Section 4 we establish asymptotic convexity of Q , and discuss the implications for 2-SMIPs. In Section 5, we show how this approximation can be extended to the multistage case.

3 Periodic behavior of the value function v

In this section, we show that, on large parts of its domain, the generic value function v is the sum of its LP-relaxation v_{LP} and a periodic function. To do so, we exploit the piecewise-linear structure of v_{LP} . In particular, let B_k , $k = 1, \dots, K$, denote its dual feasible basis matrices, so that

$$v_{\text{LP}}(\omega, x) = \max_{k=1, \dots, K} q_{B_k}^T B_k^{-1} (\omega - Tx, h). \quad (5)$$

For a fixed h , it follows that every dual feasible basis matrix B_k is optimal on a polyhedral subset $\Lambda_k(h)$ of the domain of $\omega - Tx$. Using the Basis Decomposition Theorem [33], we will select a collection $B_k, k \in \mathcal{K}(h)$, of such bases whose corresponding non-empty polyhedra $\Lambda_k(h)$, cover the entire domain $\mathbb{R}^{m_\omega}$. For each $k \in \mathcal{K}(h)$, we then identify a subset $\Pi_k(h) \subseteq \Lambda_k(h)$ on which

$$v(\omega, x) = v_{\text{LP}}(\omega, x) + \psi_k(\omega, x),$$

where ψ_k is periodic in ω . Thus, by considering the different basis matrices B_k , we obtain a semi-periodic representation of v on large parts of its domain.

To derive the semi-periodic representation, we first consider an arbitrary dual feasible basis matrix B_k . We can decompose W into $W \equiv (B_k \ N_k)$, and the decision variables into the basic variables y_{B_k} and non-basic variables y_{N_k} with corresponding

costs q_{B_k} and q_{N_k} . Using the equality $B_k y_{B_k} + N_k y_{N_k} = (\omega - Tx, h)$, we can write the basic variables y_{B_k} as a function of the non-basic variables y_{N_k}

$$y_{B_k}(\omega, x, y_{N_k}) := B_k^{-1} \left((\omega - Tx, h) - N_k y_{N_k} \right). \quad (6)$$

We further decompose the inverse matrix B_k^{-1} into $B_k^{-1} \equiv (B_{k,\omega}^{-1} \ B_{k,h}^{-1})$ with $B_{k,\omega}^{-1} \in \mathbb{R}^{(m_\omega+m_h) \times m_\omega}$ and $B_{k,h}^{-1} \in \mathbb{R}^{(m_\omega+m_h) \times m_h}$. Finally, we denote the reduced cost corresponding to the dual feasible basis matrix B_k as $\bar{q}_{N_k} \equiv q_{N_k}^\top - q_{B_k}^\top B_k^{-1} N_k \geq 0$. Then, the generic value function v of (4) can be written as

$$v(\omega, x) = q_{B_k}^\top B_{k,\omega}^{-1} (\omega - Tx) + q_{B_k}^\top B_{k,h}^{-1} h + \phi_k(\omega, x), \quad (7)$$

where $\phi_k(\omega, x)$ is given as

$$\begin{aligned} \phi_k(\omega, x) = & \min_{y_{N_k} \in \mathbb{Z}_+^{n_{N_k}} \times \mathbb{R}_+^{\bar{n}_{N_k}}} \bar{q}_{N_k}^\top y_{N_k} \\ \text{s.t.} \quad & y_{B_k}(\omega, x, y_{N_k}) \in \mathbb{Z}_+^{n_{B_k}} \times \mathbb{R}_+^{\bar{n}_{B_k}}. \end{aligned} \quad (8)$$

The first two terms in (7) are precisely the linear function induced by B_k in the LP-relaxation v_{LP} given in (5). Our aim is to show that the remainder ϕ_k exhibits periodic behavior on a sufficiently large subset of its domain. To establish this property, we construct an adapted Gomory relaxation v_{B_k} of v by relaxing selected non-negativity constraints on the basic variables. Analogously to (7), this relaxation consists of the linear function induced by B_k and a mixed-integer remainder, which we denote by ψ_k . We first show that ψ_k is periodic on its entire domain, and subsequently identify a subset on which the adapted Gomory relaxation v_{B_k} coincides with v , or equivalently, $\psi_k(\omega, x) = \phi_k(\omega, x)$. This establishes the desired periodic behavior of ϕ_k on this subset.

3.1 An adapted Gomory Relaxation for v

We now define our adapted Gomory relaxation v_{B_k} . For a given dual feasible basis matrix B_k , we take (7) as a starting point, and relax some of the non-negativity constraints in the minimization problem (8) defining ϕ_k . Our relaxation differs from the original Gomory relaxation $\bar{v}_{B_k}$ [11, 19] of v , in which all non-negativity constraints of the basic variables y_{B_k} are relaxed. Instead, we define an adapted Gomory relaxation v_{B_k} in which we only relax the non-negativity constraints of *non-critical* basic variables. This approach is similar in spirit to [34] for pure integer value functions. However, to our knowledge we are the first to apply it in the mixed-integer case.

The motivation for this adapted Gomory relaxation is that, in contrast to [19], we allow some right-hand side components h of v to be fixed. For this reason, as we will show in Example 1 below, the standard Gomory relaxation can be insufficient to prove asymptotic semi-periodicity of v . Instead, the adapted Gomory relaxation v_{B_k} is equal to v on a sufficiently large part of its domain and has the asymptotic semi-periodicity property.

Definition 2 Consider the generic mixed-integer value function v defined in (4), and suppose that its LP relaxation v_{LP} has K dual feasible basis matrices. For each $k = 1, \dots, K$, let $(B_{k,\omega}^{-1})_i$ denote the i -th row of the submatrix $B_{k,\omega}^{-1}$ of the inverse basis matrix B_k^{-1} . We define the set of critical indices $\bar{I}_k$ and non-critical indices I_k as

$$\bar{I}_k := \{i : (B_{k,\omega}^{-1})_i = 0\} \quad \text{and} \quad I_k := \{i : (B_{k,\omega}^{-1})_i \neq 0\},$$

respectively. Then, for each $k = 1, \dots, K$, our k -th *adapted Gomory relaxation* v_{B_k} is obtained by relaxing the non-negativity constraints associated with the non-critical indices in (7)-(8), i.e.,

$$v_{B_k}(\omega, x) = q_{B_k}^\top B_{k,\omega}^{-1}(\omega - Tx) + q_{B_k}^\top B_{k,h}^{-1}h + \psi_k(\omega, x),$$

where ψ_k is given as

$$\begin{aligned} \psi_k(\omega, x) = & \min_{y_{N_k} \in \mathbb{Z}_+^{n_{N_k}} \times \mathbb{R}_+^{\bar{n}_{N_k}}} \bar{q}_{N_k}^\top y_{N_k} \\ \text{s.t.} \quad & y_{B_k}(\omega, x, y_{N_k}) \in \mathbb{Z}^{n_{B_k}} \times \mathbb{R}^{\bar{n}_{B_k}}, \\ & [y_{B_k}(\omega, x, y_{N_k})]_i \geq 0, \quad \forall i \in \bar{I}_k. \end{aligned} \quad (9)$$

Example 1 Consider the mixed-integer value function

$$\begin{aligned} v(\omega, x) = & \min_{\substack{y_1, y_2, y_3 \in \mathbb{Z}_+ \\ y_4, y_5 \in \mathbb{R}_+}} \left\{ y_1 + y_2 + 2y_3 + 2y_4 + 2y_5 : 2y_1 - y_3 - y_4 = \omega - Tx, \right. \\ & \left. 2y_2 + 3y_3 - y_5 = h \right\}, \end{aligned} \quad (10)$$

where the right-hand side of the constraints consists of the random part $\omega - Tx$ and some deterministic part h . Observe that the recourse matrix W is given by $W = \begin{pmatrix} 2 & 0 & -1 & -1 & 0 \\ 0 & 2 & 3 & 0 & -1 \end{pmatrix}$ and we let y_1 and y_2 be the basic variables with corresponding basis matrix $B = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, and y_3, y_4 and y_5 be the non-basic variables with corresponding matrix $N = \begin{pmatrix} -1 & -1 & 0 \\ 3 & 0 & -1 \end{pmatrix}$. Using the fact that $B^{-1} \equiv (B_\omega^{-1} \ B_h^{-1}) = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$ and $\bar{q}_N^\top = (1, 2.5, 2.5)$, it follows from (6)-(8) that v is equivalently given by

$$\begin{aligned} v(\omega, x) = & \frac{1}{2}(\omega - Tx) + \frac{1}{2}h + \min_{\substack{y_3 \in \mathbb{Z}_+ \\ y_4, y_5 \in \mathbb{R}_+}} \left\{ y_3 + 2.5y_4 + 2.5y_5 : \right. \\ & \frac{1}{2}(\omega - Tx) + \frac{1}{2}y_3 + \frac{1}{2}y_4 \in \mathbb{Z}_+, \\ & \left. \frac{1}{2}h - \frac{3}{2}y_3 + \frac{1}{2}y_5 \in \mathbb{Z}_+ \right\}. \end{aligned} \quad (11)$$

Since the second row of $B_\omega^{-1} = \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix}$ is zero, the basic variable y_2 is a critical basic variable, whereas y_1 is non-critical.

We use this example to illustrate why it is important to preserve the non-negativity constraints of the critical basic variables. In particular, we show that the standard Gomory relaxation $\bar{v}_B$, obtained by relaxing the non-negativity constraints of both basic variables y_1 and y_2 , can differ from v for arbitrarily large values of $\omega - Tx$. We then show that

this discrepancy disappears for our adapted Gomory relaxation v_B , which preserves the non-negativity constraint of the critical basic variable y_2 .

We restrict attention to the non-negative orthant of $(\omega - Tx, h)$, and first consider the deterministic component h to be $h = 1$. For the original value function v , observe that the optimal solution always uses $y_3^* = 0$. Indeed, when $y_3 = 0$, the smallest feasible value for y_5 is 1. For any other $y_3 \geq 1$, feasibility requires y_5 to be at least $y_5 \geq 3y_3 - 1$; see (10). So, the additional cost from y_3 and y_5 is at least 3.5, whereas the possible saving in y_4 is at most 2.5; see (11). Hence, selecting any $y_3 \geq 1$ is more expensive than selecting $y_3 = 0$, and therefore $y_3^* = 0$. Consequently, it follows that y_4^* and y_5^* can be computed from (11), and the optimal solutions are $(y_1^*, y_2^*, y_3^*, y_4^*, y_5^*) = \left(\left\lceil \frac{\omega - Tx}{2} \right\rceil, 1, 0, 2 \left\lceil \frac{\omega - Tx}{2} \right\rceil - (\omega - Tx), 1 \right)$ for all $\omega - Tx \geq 0$.

For $\bar{v}_B$, it is sufficient to consider $y_3 \in \{0, 1\}$. To see this, observe that if a feasible solution satisfies $y_3 \geq 2$, replacing y_3 by $y_3 - 2$ preserves the integrality conditions and decreases the objective value. Hence, the optimal solution is either the same as for v or satisfies $y_3^* = 1$. In the latter case, we also select $y_2^* = -1$, which is possible because we have relaxed the non-negativity constraint on y_2 . The complete optimal solution is then $(y_1^*, y_2^*, y_3^*, y_4^*, y_5^*) = \left(\left\lceil \frac{\omega - Tx + 1}{2} \right\rceil, -1, 1, 2 \left\lceil \frac{\omega - Tx + 1}{2} \right\rceil - (\omega - Tx + 1), 0 \right)$; see panel (a) of Figure 1 for a geometric visualization. Furthermore, panel (b) of Figure 1 shows the value of v and $\bar{v}_B$ for $h = 1$ and $\omega - Tx \geq 0$. We observe that both value functions are semi-periodic but that $\bar{v}_B$ systematically and periodically underestimates v .

The reason why the original Gomory relaxation $\bar{v}_B$ may underestimate v is that it relaxes the non-negativity constraint of the critical basic variable y_2 . As can be observed in panel (c) of Figure 1 this is not a problem if h is large enough, but it is a problem for $h \in (0, 1]$. In contrast, our adapted Gomory relaxation v_B does not relax the non-negativity of y_2 , so that selecting $y_2 = -1$ becomes infeasible. Straightforward computations show that for all $h \geq 0$ and $\omega - Tx \geq 0$, it holds that $v_B(\omega, x) = v(\omega, x)$.

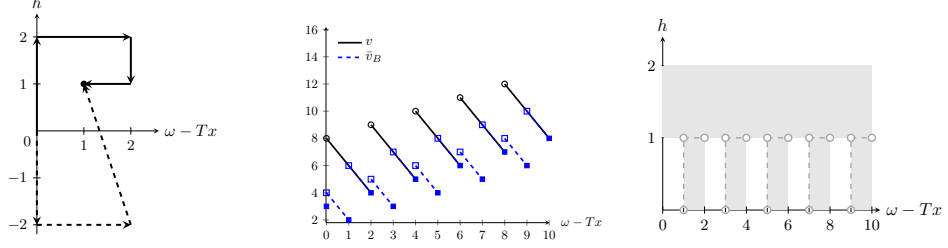


Fig. 1: We refer to the graphs from left to right as panel (a), (b), and (c). Panel (a) illustrates combinations of columns of W corresponding to the optimal solutions of v and v_B at $(\omega - Tx, h) = (1, 1)$, represented by the solid and dashed paths, respectively. In particular, the optimal solutions of v combine the columns $(2, 0), (0, 2), (-1, 0), (0, -1)$ scaled by $y_1^* = 1, y_2^* = 1, y_4^* = 1, y_5^* = 1$, respectively. In contrast, the optimal solutions of v_B combine $(2, 0), (0, 2), (-1, 3)$ scaled by $y_1^* = 1, y_2^* = -1, y_3^* = 1$. Panel (b) plots the value function v (black solid line) and the original Gomory relaxation $\bar{v}_B$ (blue dashed line) as functions of $\omega - Tx$ for $h = 1$. Panel (c) illustrates the domains in which v and $\bar{v}_B$ are equal (the shaded regions). In contrast, for $0 < h \leq 1$, the unshaded regions, shown either as blank areas or by dashed boundaries, correspond to the intervals $\omega - Tx \in (2k, 2k + 1]$, $k \in \mathbb{Z}_+$, on which $\bar{v}_B \neq v$.

3.2 Periodic function ψ_k

We now show that the function ψ_k from Definition 2 is periodic. We first define the concept of periodicity taken from [19], and then establish periodicity of ψ_k in Lemma 1.

Definition 3 A function $\psi(\omega, x) : \mathbb{R}^m \times \mathbb{R}^n \mapsto \mathbb{R}$ is B -periodic in ω for a fixed x with $B \in \mathbb{R}^{m \times m}$ if $\psi(\omega + Bl, x) = \psi(\omega, x)$ for any $l \in \mathbb{Z}^m$.

Lemma 1 For every $k = 1, \dots, K$, the function ψ_k associated with the k -th adapted Gomory relaxation v_{B_k} in Definition 2 is $|\det(B_k)|I_{m_\omega}$ -periodic in ω , where I_{m_ω} denotes the m_ω -dimensional identity matrix.

Proof To establish periodicity in ω , we need to show that given any $x \in X$ and $l \in \mathbb{Z}^{m_\omega}$, we have $\psi_k(\omega, x) = \psi_k(\omega + |\det(B_k)|I_{m_\omega}l, x)$. Denote $\omega' := \omega + |\det(B_k)|I_{m_\omega}l$ for some $l \in \mathbb{Z}^{m_\omega}$. Then, we have

$$[y_{B_k}(\omega', x, y_{N_k})]_i = (B_{k,\omega}^{-1})_i(\omega - Tx) + (B_{k,\omega}^{-1})_i|\det(B_k)|I_{m_\omega}l + (B_{k,h}^{-1})_ih - (B_k)_i^{-1}N^ky_{N_k}$$

for every $i \in I_k \cup \bar{I}_k$. Observe that for every $i \in \bar{I}_k$, it holds that $(B_{\omega,k}^{-1})_i = 0$ so that $[y_{B_k}(\omega', x, y_{N_k})]_i = [y_{B_k}(\omega, x, y_{N_k})]_i$, and thus the corresponding non-negativity constraints in $\psi_k(\omega, x)$ and $\psi_k(\omega', x)$ are identical. Moreover, for any $i \in I_k \cup \bar{I}_k$, it holds that

$$[y_{B_k}(\omega', x, y_{N_k})]_i - [y_{B_k}(\omega, x, y_{N_k})]_i = (B_{k,\omega}^{-1})_i|\det(B_k)|I_{m_\omega}l \in \mathbb{Z}^{m_\omega}.$$

This is true because the adjoint of B_k is given by $\text{adj}(B_k) = (B_{k,\omega}^{-1} \det(B_k) \quad B_{k,h}^{-1} \det(B_k))$, and the (i, j) -th entry of $\text{adj}(B_k)$ is given by $(-1)^{i+j} \det(B_k^{ij})$, where B_k^{ij} is the submatrix obtained by removing row i and column j of B_k , and where B_k is a submatrix of the integer matrix W .

Combining both observations, we conclude that the feasible regions with respect to $\psi_k(\omega, x)$ and $\psi_k(\omega', x)$ are identical, and thus $\psi_k(\omega, x) = \psi_k(\omega', x)$. $\square$

From Lemma 1, it follows that the adapted Gomory relaxation v_{B_k} is the sum of a linear function and a periodic function ψ_k , and thus semi-periodic.

3.3 Domains $\Pi_k(h)$ where v_{B_k} is equal to v

Next, we use a classical proximity result for mixed-integer programs (see Appendix A.1), to prove a uniform bound on the optimal solution $y_{N_k}^*(\omega, x)$ of the adapted Gomory relaxation v_{B_k} . We derive this bound for each adapted Gomory relaxation v_{B_k} with $k \in \bar{\mathcal{K}}(h)$, where the index set $\bar{\mathcal{K}}(h)$ is defined below.

Definition 4 Consider the generic mixed-integer value function v defined in (4), and let $\{1, \dots, K\}$ denote the index set of the dual feasible basis matrices of its LP-relaxation v_{LP} . Then, we define $\bar{\mathcal{K}}(h)$ as

$$\bar{\mathcal{K}}(h) := \left\{ k \in \{1, \dots, K\} : (B_{k,h}^{-1})_i h \geq 0, \forall i \in \bar{I}_k \right\}.$$

Remark 1 The set $\bar{\mathcal{K}}(h)$ corresponds to all indices k for which the optimal non-basic solution to the LP-relaxation of v_{B_k} satisfies $y_{N_k, \text{LP}}^* = 0$.

Lemma 2 For every $k \in \bar{\mathcal{K}}(h)$, $\omega \in \mathbb{R}^{m_\omega}$ and $x \in X$, there exists an optimal solution $y_{N_k}^*(\omega, x)$ of the adapted Gomory relaxation v_{B_k} that satisfies

$$0 \leq [y_{N_k}^*(\omega, x)]_\infty \leq (n + \bar{n})\Delta(W),$$

where $\Delta(W)$ denotes the maximum absolute value of the determinant of any square submatrix of W .

Proof We will apply Theorem 7 of Appendix A.1 for this proof. To use it, observe that since $v_{B_k}(\omega, x) := \min \left\{ q^\top y : Wy = (\omega - Tx, h), y \in \mathbb{Z}_+^{n_+} \times \mathbb{Z}^{n-n_+} \times \mathbb{R}_+^{\bar{n}_+} \times \mathbb{R}^{\bar{n}-\bar{n}_+} \right\}$, we can write the set of constraints of v_{B_k} as a MIP in the form that Theorem 7 requires. That is, we can represent the equality $Wy = (\omega - Tx, h)$ and the set of non-negativity constraints as $\hat{W}y \leq \hat{b}$,

where $\hat{b} := \begin{pmatrix} b \\ -b \\ 0 \end{pmatrix}$ with $b = \begin{pmatrix} \omega - Tx \\ h \end{pmatrix}$, and where $\hat{W} := \begin{pmatrix} W \\ -W \\ -\mathcal{I} \end{pmatrix}$ with $\mathcal{I}$ representing the

matrix whose i -th row is the i -th unit vector e_i if $i \in \bar{I}_k$, and the zero vector 0, otherwise.

Observe from (9) that if $k \in \bar{\mathcal{K}}(h)$, then $y_{N_k} = 0$ is optimal for v_{LP, B_k} since $\bar{q}_{N_k} \geq 0$. Thus, by Theorem 7, there exists an optimal solution $y_{N_k}^*(\omega, x)$ of v_{B_k} that satisfies $\|y_{N_k}^*(\omega, x)\|_\infty \leq (n + \bar{n})\Delta(\hat{W})$.

It remains to show that $\Delta(\hat{W}) = \Delta(W)$. Take any submatrix D of $\hat{W}$. If D contains a row of $-\mathcal{I}$, then from the Leibniz formula for determinants, it follows that the absolute value of $\det(D)$ is equal to the absolute value of the determinant of some submatrix of D . Inductively, this also holds when D contains multiple rows of $-\mathcal{I}$. Hence, without loss of generality, we can assume D only contains rows from W and $-W$. If D simultaneously contains the i -th row of W and $-W$, then $\det(D) = 0$ and thus the absolute value of the determinant of any other submatrix of $\hat{W}$ is at least as large or larger. Hence, we can assume that D is equal to a square submatrix of W with some of its rows multiplied by -1 . However, multiplying a row of a matrix by -1 does not change the absolute value of its determinant. Hence, we conclude that $\Delta(\hat{W}) = \Delta(W)$. $\square$

Corollary 1 *For every $k \in \bar{\mathcal{K}}(h)$, $\omega \in \mathbb{R}^{m_\omega}$, and $x \in X$, the function ψ_k associated with the adapted Gomory relaxation v_{B_k} as defined in Definition 2 is upper bounded by*

$$\psi_k(\omega, x) \leq \bar{q}_{N_k}^\top ((n + \bar{n})\Delta(W)e_{(n+\bar{n})}),$$

where $e_{(n+\bar{n})}$ denotes the all-one vector of dimension $(n + \bar{n})$.

Proof The corollary follows immediately from Lemma 2 and the definition of ψ_k in equation (9). $\square$

Using this boundedness property, for every $k \in \bar{\mathcal{K}}(h)$, we can now identify a domain $\Pi_k(h) \subseteq \mathbb{R}^{m_\omega}$ such that $v(\omega, x) = v_{B_k}(\omega, x)$ whenever $\omega - Tx \in \Pi_k(h)$.

Theorem 1 *Consider the generic mixed-integer value function v of Definition 1, and let $\bar{\mathcal{K}}(h)$ denote the index set defined in Definition 4. Then, there exists a set of indices $\mathcal{K}(h) \subseteq \bar{\mathcal{K}}(h)$ such that for the polyhedra $\Lambda_k(h) := \{s \in \mathbb{R}^{m_\omega} : (B_k)^{-1}(s, h) \geq 0\}$ with $k \in \mathcal{K}(h)$, the following properties hold*

$$(i) \quad \bigcup_{k \in \mathcal{K}(h)} \Lambda_k(h) = \mathbb{R}^{m_\omega},$$

$$(ii) \quad \text{int}(\Lambda_k(h)) \cap \text{int}(\Lambda_\ell(h)) = \emptyset \quad \text{for } k, \ell \in \mathcal{K}(h) \text{ with } k \neq \ell.$$

Moreover, for every $k \in \mathcal{K}(h)$, there exists constants $\delta_{k,i} \geq 0, \forall i \in I_k$, and a corresponding shifted polyhedron $\Pi_k(h) \subseteq \Lambda_k(h)$, defined by

$$\Pi_k(h) := \left\{ s \in \Lambda_k(h) : (B_{k,\omega}^{-1})_i s \geq \delta_{k,i} - (B_{k,h}^{-1})_i h, \forall i \in I_k \right\},$$

such that whenever $\omega - Tx \in \Pi_k(h)$, we have

$$\begin{aligned} v(\omega, x) &= v_{LP}(\omega, x) + \psi_k(\omega, x) \\ &= q_{B_k}^\top B_{k,\omega}^{-1}(\omega - Tx) + q_{B_k}^\top B_{k,h}^{-1}h + \psi_k(\omega, x). \end{aligned} \tag{12}$$

Proof We first prove properties (i) and (ii). Denoting the set Λ_k as $\Lambda_k := \{s' \in \mathbb{R}^{m_\omega + m_h} : (B_k)^{-1}s' \geq 0\}$, it follows from the Basis Decomposition theorem [33] that there exists a subset $\tilde{K}$ of $\{1, \dots, K\}$ such that

$$1: \bigcup_{k \in \hat{K}} \Lambda_k = \mathbb{R}^{m_\omega + m_h},$$

$$2: \text{int}(\Lambda_k) \cap \text{int}(\Lambda_l) = \emptyset \text{ for } k, l \in \hat{K} \text{ with } k \neq l.$$

From Lemma 10 in Appendix A.1, it now follows directly that there exists a subset of indices $\mathcal{K}(h) \subseteq \hat{K}$ such that properties (i) and (ii) hold. Moreover, it also follows that $\mathcal{K}(h) \subseteq \bar{\mathcal{K}}(h)$, since $\Lambda_k(h) = \emptyset$ for every $k \in \hat{K} \setminus \bar{\mathcal{K}}(h)$, so that excluding these indices does not affect either property.

Next, we prove the existence of $\delta_{k,i} \geq 0, \forall i \in I_k$, and the corresponding set $\Pi_k(h) \subseteq \Lambda_k(h)$ such that (12) holds for all $\omega - Tx \in \Pi_k(h)$ for every $k \in \mathcal{K}(h)$. Observe that the adapted Gomory relaxation v_{B_k} is equal to the original value function v if there exists an optimal solution $y_{N_k}^*(\omega, x)$ satisfying the relaxed non-negativity constraints of the non-critical basic variables, that is, if for all $i \in I_k$,

$$[y_{B_k}(\omega, x, y_{N_k})]_i = (B_{k,\omega}^{-1})_i(\omega - Tx) + (B_{k,h}^{-1})_i h - (B_k)_i^{-1} N_k y_{N_k}^*(\omega, x) \geq 0,$$

which is equivalent to

$$(B_{k,\omega}^{-1})_i(\omega - Tx) \geq (B_k)_i^{-1} N_k y_{N_k}^*(\omega, x) - (B_{k,h}^{-1})_i h. \quad (13)$$

It follows from Lemma 2 that for every $\omega \in \mathbb{R}^{m_\omega}$ and $x \in X$ there exists an optimal solution $0 \leq y_{N_k}^*(\omega, x) \leq (n + \bar{n})\Delta(W)e_{(n+\bar{n})}$, where $e_{(n+\bar{n})}$ is the vector of ones of dimension $(n + \bar{n})$. Hence, for all $i \in I_k$,

$$(B_k)_i^{-1} N_k y_{N_k}^*(\omega, x) \leq (n + \bar{n})\Delta(W)((B_k)_i^{-1} N_k)^+ e_{(n+\bar{n})},$$

where $((B_k)_i^{-1} N_k)^+$ is the row vector $(B_k)_i^{-1} N_k$ with each component replaced by its positive part. Hence, for every $i \in I_k$, we define the constant $\delta_{k,i}$ as

$$\delta_{k,i} := [(n + \bar{n})\Delta(W)((B_k)_i^{-1} N_k)^+ e_{(n+\bar{n})}]_i,$$

it follows that for all $\omega - Tx \in \Pi_k(h)$, it holds that $v(\omega, x) = v_{B_k}(\omega, x)$, and by definition of $v_{B_k}(\omega, x)$, (12) holds. $\square$

Theorem 1 proves that our generic mixed-integer value function v is asymptotically semi-periodic. This finding differs from the result in [19], where the entire right-hand side vector is treated as input, whereas we allow for fixing part of the right-hand side vector, namely, the part h . In fact, Theorem 1 extends the results in [19] by proving that any mixed-integer value function is asymptotically semi-periodic along any of the coordinate axes, or a combination thereof. This new observation turns out to be crucial to derive a convex approximation for v and Q in the next section.

4 Asymptotic convexity of the expected value function Q

In this section, we use the semi-periodicity of v , derived in Section 3, to show that the expected value function Q is asymptotically convex. To this end, we construct a convex approximation $\hat{v}_\alpha$ of v , called the α -approximation. By taking the expectation with respect to ω we obtain a convex approximation $\hat{Q}_\alpha$ of Q . We will show that these functions are asymptotically equal by deriving a uniform error bound on $\|Q - \hat{Q}_\alpha\|_\infty$ that decreases to zero when the variability of ω increases.

4.1 The α -approximation $\hat{v}_\alpha$

From Theorem 1 in Section 3, we observe that the mixed-integer value function v is the sum of its LP-relaxation v_{LP} and a periodic function ψ_k whenever $\omega - Tx \in \Pi_k(h)$. Thus, on $\Pi_k(h)$, the non-convexity of v is caused by the periodic function ψ_k . To construct an approximation of v that is convex in x , we replace $\psi_k(\omega, x)$ by $\psi_k(\omega, \alpha)$ for some constant vector α . Doing so for all $k \in \mathcal{K}(h)$ and maximizing over $k \in \mathcal{K}(h)$ then yields our α -approximation $\hat{v}_\alpha$. Intuitively, this approximation is not necessarily good for fixed values of ω, x , and α , but it may be good *on average* over ω since ψ_k is periodic in ω .

Definition 5 For every $\alpha \in \mathbb{R}^p$, we define the α -approximation $\hat{v}_\alpha$ of the generic value function v as

$$\hat{v}_\alpha(\omega, x) := \max_{k \in \mathcal{K}(h)} \left\{ q_{B_k}^\top B_{k,\omega}^{-1} (\omega - Tx) + q_{B_k}^\top B_{k,h}^{-1} h + \psi_k(\omega, \alpha) \right\}, \quad x \in X, \omega \in \Omega,$$

where $\mathcal{K}(h)$ is the index set from Theorem 1 and $\psi_k(\omega, \alpha)$ is the periodic function from Definition 2. Furthermore, we define the corresponding α -approximation $\hat{Q}_\alpha$ of Q as

$$\hat{Q}_\alpha(x) := \mathbb{E}_\omega[\hat{v}_\alpha(\omega, x)], \quad x \in X.$$

Compared with [30], our definition of the α -approximation differs in two ways. First, instead of using all dual feasible basis matrices $B_k, k = 1, \dots, K$, we use the subset $B_k, k \in \mathcal{K}(h)$, since those are sufficient to represent the periodic behavior of v on large parts of its domain as shown in Theorem 1. Second, we use the adapted Gomory relaxations instead of the standard ones; see Example 1 for the difference between the two.

4.2 Properties of the difference $v - \hat{v}_\alpha$

In this subsection, we derive properties of $\hat{v}_\alpha$ and of the difference $v - \hat{v}_\alpha$. In particular, we will derive a main theorem showing that for any $k \in \mathcal{K}(h)$ for which $\Pi_k(h)$ is full-dimensional, there exists an unbounded full-dimensional set $\mathcal{P}_k(h) \subseteq \Pi_k(h)$ such that $v(\omega, x) - \hat{v}_\alpha(\omega, x) = \psi_k(\omega, x) - \psi_k(\omega, \alpha)$ if $\omega - Tx \in \mathcal{P}_k(h)$. Hence, on large parts of its domain the difference $v - \hat{v}_\alpha$ is periodic. In addition, we show that the remainder of the domain can be covered by finitely many so-called hyperslices. These results will allow us to derive a uniform error bound on $\|Q - \hat{Q}_\alpha\|_\infty$ in Section 4.3, which establishes the asymptotic convexity of Q .

We first introduce the definition of a hyperslice and establish conditions under which a set can be covered by a hyperslice. We then use this concept to show that $v - \hat{v}_\alpha$ is periodic over large parts of its domain.

Definition 6 A *hyperslice* H is a set of the form $H := \{s \in \mathbb{R}^{m_\omega} : m \leq a^\top s \leq M\}$ for some constants $m, M \in \mathbb{R}$, and a non-zero vector $a \in \mathbb{R}^{m_\omega}$. We say that a set $\Pi \subseteq \mathbb{R}^m$ can be covered by a hyperslice if there exists a hyperslice H such that $\Pi \subseteq H$.

It is clear from Definition 6 that every bounded set can be covered by a hyperslice. The same applies to any set Π with $\dim(\Pi) < m$. For a general polyhedron $\Pi \in \mathbb{R}^m$, however, it depends on the dimension of the recession cone $\text{recc}(\Pi)$ of Π whether Π can be covered by a hyperslice.

Lemma 3 *Consider any non-empty polyhedron $\Pi \subseteq \mathbb{R}^m$, and let $\text{recc}(\Pi)$ denote the recession cone of Π . Then, Π can be covered by a hyperslice if and only if $\text{span}(\text{recc}(\Pi)) \neq \mathbb{R}^m$.*

Proof See Appendix A.2. □

Remark 2 Any empty polyhedron $\Pi \subseteq \mathbb{R}^m$ is naturally covered by any hyperslice.

Next, we present another sufficient condition under which a polyhedron Π can be covered by a hyperslice. In particular, if Π is lattice free, then it can be covered by a hyperslice.

Lemma 4 *Let $\Pi \subseteq \mathbb{R}^m$ be a lattice-free polyhedron. That is, $\Pi \cap \mathbb{Z}^m = \emptyset$. Then, Π can be covered by a hyperslice.*

Proof See Appendix A.2. □

In the following, we classify indices $k \in \mathcal{K}(h)$ according to whether $\Pi_k(h)$ can be covered by a hyperslice. The lemma below shows that, for every $k \in \mathcal{K}(h)$, this holds if and only if $\Lambda_k(h)$ can be covered by a hyperslice.

Definition 7 We define $\mathcal{K}_H(h) \subseteq \mathcal{K}(h)$ as the subset of indices $k \in \mathcal{K}(h)$ for which $\Pi_k(h)$ can be covered by a hyperslice. Moreover, we define $\mathcal{K}_{\mathcal{P}}(h) := \mathcal{K}(h) \setminus \mathcal{K}_H(h)$. That is, for every $k \in \mathcal{K}_{\mathcal{P}}(h)$, the associated set $\Pi_k(h)$ cannot be covered by any hyperslice.

Lemma 5 *For every $k \in \mathcal{K}(h)$, $\Pi_k(h)$ can be covered by a hyperslice if and only if $\Lambda_k(h)$ can be covered by a hyperslice.*

Proof See Appendix A.2. □

Next, we present two properties of the difference $v - \hat{v}_\alpha$. We first show that the absolute difference $|v(\omega, x) - \hat{v}_\alpha(\omega, x)|$ is uniformly bounded in the following lemma.

Lemma 6 Consider the value function v and its approximation $\hat{v}_\alpha$ from equation (4) and Definition 5. There exists a constant $\beta > 0$ such that for any $\omega \in \mathbb{R}^{m_\omega}$ and $x \in X$, we have

$$|v(\omega, x) - \hat{v}_\alpha(\omega, x)| \leq \beta. \quad (14)$$

Proof Observe that for any $\omega \in \mathbb{R}^{m_\omega}$ and $x \in X$, we have

$$|v(\omega, x) - \hat{v}_\alpha(\omega, x)| \leq |v(\omega, x) - v_{\text{LP}}(\omega, x)| + |\hat{v}_\alpha(\omega, x) - v_{\text{LP}}(\omega, x)|. \quad (15)$$

By Theorem 7, there exists a constant β_v so that the first term $|v(\omega, x) - v_{\text{LP}}(\omega, x)|$ on the right-hand side of (15) is bounded by $|v(\omega, x) - v_{\text{LP}}(\omega, x)| \leq \beta_v$. For the second term, observe from Theorem 1 that the sets $\Lambda_k(h), k \in \mathcal{K}(h)$, cover the entire domain $\mathbb{R}^{m_\omega}$. This implies that for any $\omega \in \mathbb{R}^{m_\omega}$ and $x \in X$, $v_{\text{LP}}(\omega, x)$ of (5) is attained by some index k of $\mathcal{K}(h)$. Hence, we have

$$|\hat{v}_\alpha(\omega, x) - v_{\text{LP}}(\omega, x)| \leq \max_{k \in \mathcal{K}(h)} |\psi_k(\omega, \alpha)|.$$

It follows from Corollary 1 that $|\psi_k(\omega, \alpha)|$ is upper bounded by $\beta_k := \bar{q}_{N_k}^\top ((n + \bar{n})\Delta(W)e_{(n+\bar{n})})$, so we have $\max_{k \in \mathcal{K}(h)} |\psi_k(\omega, \alpha)| \leq \max_{k \in \mathcal{K}(h)} \beta_k$. Therefore, we define the constant $\beta := \beta_v + \max_{k \in \mathcal{K}(h)} \beta_k$, and it follows that (14) indeed holds. $\square$

To derive the second property, we first present two auxiliary results in Lemma 7.

Lemma 7 Let $\mathcal{K}_{\mathcal{P}}(h)$ and $\mathcal{K}_H(h)$ be the set of indices of Definition 7. Let $k, l \in \mathcal{K}(h)$ be given with $k \in \mathcal{K}_{\mathcal{P}}(h)$ and assume that $q_{B_k}^\top B_{k,\omega}^{-1} = q_{B_l}^\top B_{l,\omega}^{-1}$. Then, we have

$$q_{B_k}^\top B_{k,h}^{-1} h + \psi_k(\omega, \alpha) \geq q_{B_l}^\top B_{l,h}^{-1} h + \psi_l(\omega, \alpha), \quad \forall \omega \in \mathbb{R}^{m_\omega}, \alpha \in \mathbb{R}^p, \quad (16)$$

and moreover if $l \in \mathcal{K}_{\mathcal{P}}(h)$, then

$$q_{B_k}^\top B_{k,h}^{-1} h + \psi_k(\omega, \alpha) = q_{B_l}^\top B_{l,h}^{-1} h + \psi_l(\omega, \alpha), \quad \forall \omega \in \mathbb{R}^{m_\omega}, \alpha \in \mathbb{R}^p. \quad (17)$$

Proof Recall from Theorem 1 that if $\omega - Tx \in \Pi_k(h)$, then

$$v(\omega, x) = v_{\text{LP}}(\omega, x) + \psi_k(\omega, x) = q_{B_k}^\top B_{k,\omega}^{-1} (\omega - Tx) + q_{B_k}^\top B_{k,h}^{-1} h + \psi_k(\omega, x).$$

Moreover, since every adapted Gomory relaxation v_{B_k} is a relaxation of v , it follows that for $\omega - Tx \in \Pi_k(h)$, we have

$$q_{B_k}^\top B_{k,\omega}^{-1} (\omega - Tx) + q_{B_k}^\top B_{k,h}^{-1} h + \psi_k(\omega, x) \geq q_{B_l}^\top B_{l,\omega}^{-1} (\omega - Tx) + q_{B_l}^\top B_{l,h}^{-1} h + \psi_l(\omega, x),$$

and since $q_{B_k}^\top B_{k,\omega}^{-1} = q_{B_l}^\top B_{l,\omega}^{-1}$, we have

$$q_{B_k}^\top B_{k,h}^{-1} h + \psi_k(\omega, x) \geq q_{B_l}^\top B_{l,h}^{-1} h + \psi_l(\omega, x). \quad (18)$$

Observe that since ψ_k and ψ_l are p_k and p_l -periodic in ω with $p_k := |\det(B_k)|I_{m_\omega}$ and $p_l := |\det(B_l)|I_{m_\omega}$, respectively; see Lemma 1. We will use this periodicity to show that (18) holds not only for $\omega - Tx \in \Pi_k(h)$, but for all $\omega - Tx \in \mathbb{R}^{m_\omega}$. This will be true if there exists some $u_k \in \mathbb{Z}^{m_\omega}$ such that $\omega + p_k p_l u_k - Tx \in \Pi_k(h)$, since we will then have

$$\begin{aligned}
\psi_k(\omega, x) + q_{B_k}^\top B_{k,h}^{-1} h &= \psi_k(\omega + p_k p_l u_k, x) + q_{B_k}^\top B_{k,h}^{-1} h \\
&\geq \psi_l(\omega + p_k p_l u_k, x) + q_{B_l}^\top B_{l,h}^{-1} h \\
&= \psi_l(\omega, x) + q_{B_l}^\top B_{l,h}^{-1} h.
\end{aligned}$$

where the equalities hold by the periodicity of ψ_k and ψ_l and the inequality follows from the fact that $\omega + p_k p_l u_k - Tx \in \Pi_k(h)$. Next, we will prove that such a u_k exists for any $\omega \in \mathbb{R}^{m_\omega}$ and $x \in X$. To do so, define the set $U \subseteq \mathbb{R}^{m_\omega}$ as

$$U := \{u \in \mathbb{R}^{m_\omega} : \omega + p_k p_l u_k - Tx \in \Pi_k(h)\}.$$

Observe that U is a shifted and scaled version of $\Pi_k(h)$, so that U is a non-empty polyhedron with $\text{recc}(U) = \text{recc}(\Pi_k(h))$. Indeed, since $k \in \mathcal{K}_{\mathcal{P}}(h)$, we have $\Pi_k(h) \neq \emptyset$, and thus $U \neq \emptyset$, and moreover, since $k \in \mathcal{K}_{\mathcal{P}}(h)$, the non-empty set $\Pi_k(h)$ cannot be covered by a hyperslice. Lemma 3 therefore gives $\text{span}(\text{recc}(U)) = \text{span}(\text{recc}(\Pi_k(h))) = \mathbb{R}^{m_\omega}$. Hence U cannot be covered by a hyperslice. So suppose for contradiction that U is lattice-free, i.e., $\text{int}(U) \cap \mathbb{Z}^{m_\omega} = \emptyset$ then it implies U can be covered by a hyperslice, contradicting the mentioned observation. Hence, there indeed exists some integer $u \in \mathbb{Z}^{m_\omega}$ such that $\omega + p_k p_l u_k - Tx \in \Pi_k(h)$ for every $\omega \in \mathbb{R}^{m_\omega}$ and $x \in X$, proving that (18) holds for all $\omega \in \mathbb{R}^{m_\omega}$ and $x \in X$, and thus (16) holds for all $\omega \in \mathbb{R}^{m_\omega}$, $\alpha \in \mathbb{R}^p$.

Finally, to prove (17), observe that if $l \in \mathcal{K}_{\mathcal{P}}(h)$, then by symmetry, (16) holds for the reversed inequality as well. $\square$

Now we are ready to present the main result of this subsection, showing the existence of subsets $\mathcal{P}_k(h) \subseteq \Pi_k(h)$ so that both v and v_{B_k} can be expressed in terms of the same periodic function ψ_k , i.e., $v(\omega, x) - \hat{v}_\alpha(\omega, x) = \psi_k(\omega, x) - \psi_k(\omega, \alpha)$, whenever $\omega - Tx \in \mathcal{P}_k(h)$.

Theorem 2 Consider the generic mixed-integer value function v and its α -approximation $\hat{v}_\alpha$, defined in (4) and Definition 5, respectively. For every $k \in \mathcal{K}_{\mathcal{P}}(h)$, with $\mathcal{K}_{\mathcal{P}}(h)$ given in Definition 7, there exists some non-negative constant $\hat{\delta}_{k,i}, \forall i \in I_k$, such that for any $\omega - Tx \in \mathcal{P}_k(h) \subseteq \Pi_k(h)$ with $\mathcal{P}_k(h)$ given as $\mathcal{P}_k(h) := \{s \in \Pi_k(h) : (B_{k,\omega}^{-1})_i s \geq \hat{\delta}_{k,i} - (B_{k,h}^{-1})_i h, \forall i \in I_k\}$, we have

$$\hat{v}_\alpha(\omega, x) = v_{LP}(\omega, x) + \psi_k(\omega, \alpha), \quad (19)$$

and

$$v(\omega, x) = v_{LP}(\omega, x) + \psi_k(\omega, x). \quad (20)$$

Proof Let $k \in \mathcal{K}_{\mathcal{P}}(h)$ be given. It follows from Theorem 1 that equality (20) holds for all $\omega - Tx \in \Pi_k(h)$, and thus it also holds for all $\omega - Tx \in \mathcal{P}_k(h)$ for any subset $\mathcal{P}_k(h)$ of $\Pi_k(h)$.

To prove equality (19), we will show that there exists a vector $\hat{\delta}_k \geq \delta$ with δ given in Theorem 1, such that for all $\omega - Tx \in \mathcal{P}_k(h)$ and every $l \in \mathcal{K}_{\mathcal{P}}(h)$, we have

$$\begin{aligned}
q_{B_k}^\top B_{k,\omega}^{-1}(\omega - Tx) + q_{B_k}^\top B_{k,h}^{-1} h + \psi_k(\omega, \alpha) \\
\geq q_{B_l}^\top B_{l,\omega}^{-1}(\omega - Tx) + q_{B_l}^\top B_{l,h}^{-1} h + \psi_l(\omega, \alpha),
\end{aligned} \quad (21)$$

implying that $\hat{v}_\alpha(\omega, x) = v_{LP}(\omega, x) + \psi_k(\omega, \alpha)$.

Observe that if $q_{B_k}^\top B_{k,\omega}^{-1} = q_{B_l}^\top B_{l,\omega}^{-1}$, then it follows directly from Lemma 7 that the inequality (21) holds for all $\omega \in \mathbb{R}^{m_\omega}$ and $\alpha \in \mathbb{R}^p$. Hence, from now on, let $l \in \mathcal{K}(h)$ be given such that $q_{B_k}^\top B_{k,\omega}^{-1} \neq q_{B_l}^\top B_{l,\omega}^{-1}$, and we equivalently write (21) as

$$(q_{B_k}^\top B_{k,\omega}^{-1} - q_{B_l}^\top B_{l,\omega}^{-1})(\omega - Tx) + \Delta_h^{k,l} \geq \bar{q}_{N_l}^\top y_{N_l}^*(\omega, \alpha) - \bar{q}_{N_k}^\top y_{N_k}^*(\omega, \alpha) \quad (22)$$

with $\Delta_h^{k,l} := q_{B_k}^\top B_{k,h}^{-1}h - q_{B_l}^\top B_{l,h}^{-1}h$, $y_{N_k}^*(\omega, \alpha)$ and $y_{N_l}^*(\omega, \alpha)$ being the optimal non-basic solutions of the optimization problems of $\psi_k(\omega, \alpha)$ and $\psi_l(\omega, \alpha)$, respectively. Moreover, it follows from Lemma 2 that both $y_{N_k}^*(\omega, \alpha)$ and $y_{N_l}^*(\omega, \alpha)$ are bounded, so that the right-hand side of (22) is uniformly bounded. That is, there exists a constant $C_{k,l}$ such that

$$\bar{q}_{N_l}^\top y_{N_l}^*(\omega, \alpha) - \bar{q}_{N_k}^\top y_{N_k}^*(\omega, \alpha) \leq C_{k,l}. \quad (23)$$

Next, we will define a $\hat{\delta}_k$ such that for all $\omega - Tx \in \mathcal{P}_k(h)$, it holds that the left-hand side of (22) exceeds $C_{k,l}$, i.e.,

$$(q_{B_k}^\top B_{k,\omega}^{-1} - q_{B_l}^\top B_{l,\omega}^{-1})(\omega - Tx) + \Delta_h^{k,l} \geq C_{k,l}. \quad (24)$$

To do so, we first analyze the LP-relaxation $v_{\text{LP}}(\omega, x)$ of $v(\omega, x)$. In general, it holds that

$$v_{\text{LP}}(\omega, x) = \max_{k \in \mathcal{K}(h)} \{q_{B_k}^\top B_{k,\omega}^{-1}(\omega - Tx) + q_{B_k}^\top B_{k,h}^{-1}h\}, \quad \forall \omega \in \mathbb{R}^{m_\omega}, x \in X \quad (25)$$

with $v_{\text{LP}}(\omega, x) = q_{B_k}^\top B_{k,\omega}^{-1}(\omega - Tx) + q_{B_k}^\top B_{k,h}^{-1}h$ if $\omega - Tx \in \Pi_k(h)$. Consequently, when $\omega - Tx \in \Pi_k(h)$, we have

$$v_{\text{LP}}(\omega, x) \geq q_{B_k}^\top B_{k,\omega}^{-1}(\omega - Tx) + q_{B_k}^\top B_{k,h}^{-1}h$$

for every $l \in \mathcal{K}(h)$ with $l \neq k$, and it follows directly that the left-hand side of (24) is non-negative if $\omega - Tx \in \Pi_k(h)$. Moreover, since $k \in \mathcal{K}_{\mathcal{P}}(h)$, Lemma 3 implies that $\text{span}(\text{recc}(\Pi_k(h))) = \mathbb{R}^{m_\omega}$. We may therefore select linearly independent vectors $r_1, \dots, r_{m_\omega} \in \text{recc}(\Pi_k(h))$. Let R_k be the matrix with these vectors as columns. Let $\omega - Tx \in \Pi_k(h)$ be given, then, for every selected recession direction r_i and $\eta \geq 0$, it holds that $\omega - Tx + \eta r_i \in \Pi_k(h)$, and thus the left-hand side elements in (22) is non-negative for $\omega - Tx + \eta r_i$, that is

$$(q_{B_k}^\top B_{k,\omega}^{-1} - q_{B_l}^\top B_{l,\omega}^{-1})(\omega - Tx + \eta r_i) + \Delta_h^{k,l} \geq 0. \quad (26)$$

Since $\Delta_h^{k,l}$ is fixed and (26) holds for all $\eta \geq 0$, we conclude that $(q_{B_k}^\top B_{k,\omega}^{-1} - q_{B_l}^\top B_{l,\omega}^{-1})r_i \geq 0$. In fact, this holds for all r_i , and thus

$$(q_{B_k}^\top B_{k,\omega}^{-1} - q_{B_l}^\top B_{l,\omega}^{-1})\mathcal{R}_k \geq 0.$$

Since the columns of $\mathcal{R}_k$ form a basis of $\mathbb{R}^{m_\omega}$, it follows that $\mathcal{R}_k$ has full rank and thus $(q_{B_k}^\top B_{k,\omega}^{-1} - q_{B_l}^\top B_{l,\omega}^{-1})\mathcal{R}_k = 0$ if and only if $q_{B_k}^\top B_{k,\omega}^{-1} = q_{B_l}^\top B_{l,\omega}^{-1}$. So, for $q_{B_k}^\top B_{k,\omega}^{-1} \neq q_{B_l}^\top B_{l,\omega}^{-1}$ there exists some $\bar{r}_i$ of $\mathcal{R}_k$ such that

$$(q_{B_k}^\top B_{k,\omega}^{-1} - q_{B_l}^\top B_{l,\omega}^{-1})\bar{r}_i > 0.$$

Next, define this $\bar{r}_i$ to be $r_{k,l}$, and define $\eta_{k,l}$ as a sufficiently large constant such that

$$(q_{B_k}^\top B_{k,\omega}^{-1} - q_{B_l}^\top B_{l,\omega}^{-1})\eta_{k,l}r_{k,l} \geq C_{k,l} \quad (27)$$

Now, for every $i \in I_k$, we define the constant $\hat{\delta}_{k,i}$ as

$$\hat{\delta}_{k,i} = \delta_{k,i} + \max_{\substack{l \in \mathcal{K}(h): \\ q_{B_k}^\top B_{k,\omega}^{-1} \neq q_{B_l}^\top B_{l,\omega}^{-1}}} (B_{k,\omega}^{-1})_i \eta_{k,l} r_{k,l}.$$

We will show in a moment that the constants $\hat{\delta}_{k,i}, i \in I_k$, are correctly constructed such that the corresponding set $\mathcal{P}_k(h)$ has the property that for every $\omega - Tx \in \mathcal{P}_k(h)$ and $l \in \mathcal{K}(h)$ with $q_{B_k}^\top B_{k,\omega}^{-1} \neq q_{B_l}^\top B_{l,\omega}^{-1}$, it holds that $\omega - Tx - \eta_{k,l} r_{k,l} \in \Pi_k(h)$. Consequently, for every $\omega - Tx \in \mathcal{P}_k(h) \subseteq \Pi_k(h)$, we have

$$\begin{aligned} (q_{B_k}^\top B_{k,\omega}^{-1} - q_{B_l}^\top B_{l,\omega}^{-1})(\omega - Tx) + \Delta_h^{k,l} &= (q_{B_k}^\top B_{k,\omega}^{-1} - q_{B_l}^\top B_{l,\omega}^{-1})(\omega - Tx - \eta_{k,l} r_{k,l}) \\ &\quad + \Delta_h^{k,l} + (q_{B_k}^\top B_{k,\omega}^{-1} - q_{B_l}^\top B_{l,\omega}^{-1})\eta_{k,l} r_{k,l} \\ &\geq 0 + C_{k,l}, \end{aligned}$$

where the inequality follows from (26) and (27), proving that (24), and consequently (21), holds.

We now prove that the constants $\hat{\delta}_{k,i}, i \in I_k$, satisfy the property mentioned above. Since $r_{k,l} \in \text{recc}(\Pi_k(h))$, with $\Pi_k(h)$ defined as

$$\Pi_k(h) := \{s \in \Lambda_k(h) : (B_{k,\omega}^{-1})_i s \geq (\delta_k)_i - (B_{k,h}^{-1})_i h, \forall i \in I_k\},$$

it holds that $(B_{k,\omega}^{-1})_i \eta_{k,l} r_{k,l} \geq 0$ for all $i \in I_k$ and $l \in \mathcal{K}(h)$ with $q_{B_k}^\top B_{k,\omega}^{-1} \neq q_{B_l}^\top B_{l,\omega}^{-1}$, and thus $\hat{\delta}_{k,i} \geq \delta_{k,i}$ for all $i \in I_k$. Let $\omega - Tx \in \mathcal{P}_k(h)$ be given, that is $(B_{k,\omega}^{-1})_i(\omega - Tx) \geq \hat{\delta}_{k,i} - (B_{k,h}^{-1})_i h, \forall i \in I_k$. Then, for every $i \in I_k$,

$$\begin{aligned} (B_{k,\omega}^{-1})_i(\omega - Tx - \eta_{k,l} r_{k,l}) &= (B_{k,\omega}^{-1})_i(\omega - Tx) - (B_{k,\omega}^{-1})_i(\eta_{k,l} r_{k,l}) \\ &\geq \hat{\delta}_{k,i} - (B_{k,h}^{-1})_i h - (B_{k,\omega}^{-1})_i(\eta_{k,l} r_{k,l}) \\ &= \delta_{k,i} - (B_{k,h}^{-1})_i h \\ &\quad + \max_{\substack{l \in \mathcal{K}(h): \\ q_{B_k}^\top B_{k,\omega}^{-1} \neq q_{B_l}^\top B_{l,\omega}^{-1}}} (B_{k,\omega}^{-1})_i \eta_{k,l} r_{k,l} - (B_{k,\omega}^{-1})_i \eta_{k,l} r_{k,l} \\ &\geq \delta_{k,i} - (B_{k,h}^{-1})_i h, \end{aligned}$$

and thus $\omega - Tx - \eta_{k,l} r_{k,l} \in \Pi_k(h)$. $\square$

From Theorem 2, it follows that whenever $\omega - Tx \in \mathcal{P}_k(h)$, we have

$$v(\omega, x) - \hat{v}_\alpha(\omega, x) = \psi_k(\omega, x) - \psi_k(\omega, \alpha), \quad (28)$$

and therefore $v - \hat{v}_\alpha$ is $|\det(B_k)|I_{m_\omega}$ -periodic in ω for any given $x \in X$.

4.3 Uniform error bound for the α -approximation $\hat{Q}_\alpha$

To derive a uniform error bound for the α -approximation $\hat{Q}_\alpha$, we first exploit the periodicity of $v - \hat{v}_\alpha$ over the sets $Tx + \mathcal{P}_k(h), k \in \mathcal{K}_\mathcal{P}(h)$. Indeed, for any given x and $k \in \mathcal{K}_\mathcal{P}(h)$, it holds that $v - \hat{v}_\alpha$ is periodic in ω whenever $\omega \in Tx + \mathcal{P}_k(h)$. We define the remaining space as $\mathcal{N}_x := \mathbb{R}^{m_\omega} \setminus \cup_{k \in \mathcal{K}_\mathcal{P}(h)} (Tx + \mathcal{P}_k(h))$. Then, the absolute error induced by $\hat{Q}_\alpha$ is given by

$$\left| Q(x) - \hat{Q}_\alpha(x) \right| = \left| \mathbb{E} \left(v(\omega, x) - \hat{v}_\alpha(\omega, x) \right) \right|$$

$$\begin{aligned}
&= \left| \int_{\mathbb{R}^{m_\omega}} \left(v(s, x) - \hat{v}_\alpha(s, x) \right) f(s) ds \right| \\
&\leq \left| \int_{\cup_{k \in \mathcal{K}_{\mathcal{P}}(h)} (Tx + \mathcal{P}_k(h))} \left(v(s, x) - \hat{v}_\alpha(s, x) \right) f(s) ds \right| \\
&\quad + \left| \int_{\mathcal{N}_x} \left(v(s, x) - \hat{v}_\alpha(s, x) \right) f(s) ds \right|, \quad (29)
\end{aligned}$$

where $f(s)$ denotes the joint density function of the random parameters ω . We will derive upper bounds for the two terms corresponding to the domains $\cup_{k \in \mathcal{K}_{\mathcal{P}}(h)} (Tx + \mathcal{P}_k(h))$ and $\mathcal{N}_x$, respectively. These bounds will depend on the total variations of the density functions of the random parameters. We first define the concept of total variation formally; see [22] for a detailed introduction to total variation error bounds.

Definition 8 ([19, Definition 4.1]) Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a real-valued function and let $I \subset \mathbb{R}$ be an interval. Denote $\Pi(I)$ as the set of all finite ordered sets

$$P = \{z_1, \dots, z_{N+1}\} \quad \text{with} \quad z_1 < \dots < z_{N+1} \in I$$

for some $N \in \mathbb{Z}_+$. Then, we denote the total variation of f on I as $|\Delta|f(I)$, defined as

$$|\Delta|f(I) := \sup_{P \in \Pi(I)} \sum_{i=1}^N |f(z_{i+1}) - f(z_i)|.$$

We write $|\Delta|f := |\Delta|f(\mathbb{R})$. Moreover, we say that f is of bounded variation if $|\Delta|f < +\infty$.

Definition 9 For any vector $s \in \mathbb{R}^{m_\omega}$, let $s_{-i} \in \mathbb{R}^{m_\omega-1}$ represent s without its i -th component. Then, we denote $f_i(s_i | s_{-i}) = \frac{f(s)}{f_{-i}(s_{-i})}$ as the i -th conditional density function, and define $\mathcal{H}^{m_\omega}$ as the set of joint density function $f(\cdot)$ whose conditional density functions $f_i(s_i | s_{-i}), i = 1, \dots, m_\omega$ are of bounded variation.

Intuitively, the total variation $|\Delta|f$ measures the total vertical distance traveled by the graph of the function f . As the graph of f flattens, the total variation $|\Delta|f$ converges to zero. If f_ω is a density function, then the total variation $|\Delta|f_\omega$ is inversely related to the variability of the random parameter ω . For example, for $\omega_1 \sim N(\mu, \sigma^2)$ and $\omega_2 \sim \text{Exp}(\lambda)$, it holds that $|\Delta|f_{\omega_1} = \frac{1}{\sigma} \sqrt{\frac{2}{\pi}}$ and $|\Delta|f_{\omega_2} = 2\lambda$, respectively.

4.3.1 Error bound for periodic functions

The following lemma presents a total variation error bound for the first term of the right-hand side of inequality (29).

Lemma 8 Consider the α -approximation $\hat{Q}_\alpha(x) = \mathbb{E}_\omega[\hat{v}_\alpha(\omega, x)]$ of Definition 5. Let $\mathcal{K}_{\mathcal{P}}(h)$ be the index set given in Definition 7 and the corresponding domain $\mathcal{P}_k(h)$ defined in Theorem 2, then, for any given $x \in X$ and every probability density function $f \in \mathcal{H}^{m_\omega}$, the expected approximation error over the set $\cup_{k \in \mathcal{K}_{\mathcal{P}}(h)} (Tx + \mathcal{P}_k(h))$ is bounded by

$$\left| \int_{\cup_{k \in \mathcal{K}_{\mathcal{P}}(h)} (Tx + \mathcal{P}_k(h))} \left(v(s, x) - \hat{v}_\alpha(s, x) \right) f(s) ds \right|$$

$$\leq \sum_{k \in \mathcal{K}_{\mathcal{P}}(h)} \gamma_k |\det(B_k)| \sum_{i=1}^{m_\omega} \mathbb{E}_{\omega-i} \left[|\Delta| f_i(\cdot | \omega-i) \right], \quad (30)$$

with $\gamma_k := \bar{q}_{N_k}^\top (n + \bar{n}) \Delta(W) e_{(n+\bar{n})} > 0$.

Proof Observe that

$$\begin{aligned} & \left| \int_{\cup_{k \in \mathcal{K}_{\mathcal{P}}(h)} (Tx + \mathcal{P}_k(h))} (v(s, x) - \hat{v}_\alpha(s, x)) f(s) ds \right| \\ &= \left| \sum_{k \in \mathcal{K}_{\mathcal{P}}(h)} \int_{(Tx + \mathcal{P}_k(h))} (v(s, x) - \hat{v}_\alpha(s, x)) f(s) ds \right| \\ &\leq \sum_{k \in \mathcal{K}_{\mathcal{P}}(h)} \left| \int_{(Tx + \mathcal{P}_k(h))} (v(s, x) - \hat{v}_\alpha(s, x)) f(s) ds \right|, \end{aligned} \quad (31)$$

where the first equality follows from the disjointness property of the sets $\Lambda_k(h)$ for $k \in \mathcal{K}_{\mathcal{P}}(h)$, established in Theorem 1, together with the fact that $\mathcal{P}_k(h) \subseteq \Lambda_k(h)$, and the first inequality follows from the triangle inequality.

Now, we will apply the total variation error bound of [19, Theorem 4.13] to bound the right-hand side of inequality (31) so that we can establish inequality (30). To apply this bound, it is required that $Tx + \mathcal{P}_k(h)$ is convex, and that $v - \hat{v}_\alpha$ is periodic in ω , with its absolute value uniformly bounded over $Tx + \mathcal{P}_k(h)$, and that its mean value is zero.

Indeed, from Theorem 2, we have shown that every $\mathcal{P}_k(h)$ with $k \in \mathcal{K}_{\mathcal{P}}(h)$ is convex polyhedral, and therefore $Tx + \mathcal{P}_k(h)$ is convex. In addition, we have shown that $v(\omega, x) - \hat{v}_\alpha(\omega, x)$ is equal to $\psi_k(\omega, x) - \psi_k(\omega, \alpha)$ when $\omega \in Tx + \mathcal{P}_k(h)$ and therefore it follows from Lemma 1 that it is $(\det(B_k)I_{m_\omega})$ -periodic in ω . Furthermore, from Corollary 1 it follows that

$$\begin{aligned} |v(\omega, x) - \hat{v}_\alpha(\omega, x)| &\leq |\psi_k(\omega, x)| + |\psi_k(\omega, \alpha)| \\ &\leq 2\bar{q}_{N_k}^\top ((n + \bar{n}) \Delta(W) e_{(n+\bar{n})}). \end{aligned}$$

It remains to show that the mean value of $v - \hat{v}_\alpha$ is zero. Indeed, observe that $v(\omega, x) - \hat{v}_\alpha(\omega, x) = \psi_k(\omega, x) - \psi_k(\omega, \alpha)$, and

$$\begin{aligned} & p_k^{-m_\omega} \int_0^{p_k} \dots \int_0^{p_k} \psi_k(s, x) ds_1 \dots ds_{m_\omega} \\ &= p_k^{-m_\omega} \int_0^{p_k} \dots \int_0^{p_k} \psi_k(s - T(x - \alpha), \alpha) ds_1 \dots ds_{m_\omega} \\ &= p_k^{-m_\omega} \int_{(-T(x-\alpha))_{m_\omega}}^{p_k - (-T(x-\alpha))_{m_\omega}} \dots \int_{(-T(x-\alpha))_{m_1}}^{p_k - (-T(x-\alpha))_{m_1}} \psi_k(u, \alpha) ds_1 \dots ds_{m_\omega} \\ &= p_k^{-m_\omega} \int_0^{p_k} \dots \int_0^{p_k} \psi_k(s, \alpha) ds_1 \dots ds_{m_\omega}. \end{aligned}$$

where the first equality follows from the definition of ψ_k , the second equality follows from a change of variables, and the final equality follows from the periodicity of ψ_k . Consequently, the mean value of $\psi_k(\omega, x)$ and $\psi_k(\omega, \alpha)$ are equal, and therefore the mean value of $\psi_k(s, x) - \psi_k(s, \alpha)$ is zero. Applying the error bound from [19, Theorem 4.13], this yields

$$\left| \int_{\cup_{k \in \mathcal{K}_{\mathcal{P}}(h)} (Tx + \mathcal{P}_k(h))} (v(s, x) - \hat{v}_\alpha(s, x)) f(s) ds \right|$$

$$\begin{aligned}
&\leq \sum_{k \in \mathcal{K}_{\mathcal{P}}(h)} \left| \int_{(Tx + \mathcal{P}_k(h))} (v(s, x) - \hat{v}_\alpha(s, x)) f(s) ds \right| \\
&\leq \sum_{k \in \mathcal{K}_{\mathcal{P}}(h)} \gamma_k |\det(B^k)| \sum_{i=1}^{m_\omega} \mathbb{E}_{\omega_{-i}} \left[|\Delta| f_i(\cdot | \omega_{-i}) \right], \tag{32}
\end{aligned}$$

where the constant γ_k is given by $\gamma_k := \bar{q}_{N_k}^\top ((n + \bar{n}) \Delta(W) e_{(n+\bar{n})})$. $\square$

4.3.2 Error bound for hyperslices

In the following lemma, we apply the probability bound from [19, Theorem 4.6] to bound the approximation error of the generic convex approximation over the domain $\mathcal{N}_x$.

Lemma 9 *Consider the α -approximation $\hat{Q}_\alpha(x) = \mathbb{E}_\omega[\hat{v}_\alpha(\omega, x)]$ from Definition 5 and the domain*

$$\mathcal{N}_x := \mathbb{R}^{m_\omega} \setminus \cup_{k \in \mathcal{K}_{\mathcal{P}}(h)} (Tx + \mathcal{P}_k(h)),$$

where the index set $\mathcal{K}_{\mathcal{P}}(h)$ is given in Definition 7 and the sets $\mathcal{P}_k(h)$ are defined in Theorem 2. Then, there exist some constants $\beta, D_{k,i}$ and D_k , $k \in \mathcal{K}_{\mathcal{P}}(h), i \in I_k$, such that for every probability density function $f \in \mathcal{H}^{m_\omega}$, the expected approximation error caused by $\hat{Q}_\alpha$ over $\mathcal{N}_x$ is bounded by

$$\begin{aligned}
&\left| \int_{\mathcal{N}_x} (v(s, x) - \hat{v}_\alpha(s, x)) f(s) ds \right| \\
&\leq \beta \left(\sum_{k \in \mathcal{K}_{\mathcal{P}}(h)} \sum_{i \in I_k} D_{k,i} + \sum_{k \in \mathcal{K}_H(h)} D_k \right) \sum_{j=1}^{m_\omega} \mathbb{E}_{\omega_{-j}} \left[|\Delta| f_j(\cdot | \omega_{-j}) \right], \tag{33}
\end{aligned}$$

with the index sets I_k and $\mathcal{K}_H(h)$ defined in Definitions 2 and 7, respectively.

Proof We first show that $\mathcal{N}_x$ can be covered by a collection of hyperslices so that the probability bound from [19, Theorem 4.6] can be applied. We have

$$\begin{aligned}
\mathcal{N}_x &= \mathbb{R}^{m_\omega} \setminus \cup_{k \in \mathcal{K}_{\mathcal{P}}(h)} (Tx + \mathcal{P}_k(h)) \\
&= \cup_{k \in \mathcal{K}(h)} (Tx + \Lambda_k(h)) \setminus \cup_{k \in \mathcal{K}_{\mathcal{P}}(h)} (Tx + \mathcal{P}_k(h)) \\
&= \left(\cup_{k \in \mathcal{K}_{\mathcal{P}}(h)} \left(Tx + (\Lambda_k(h) \setminus \mathcal{P}_k(h)) \right) \right) \\
&\quad \cup \left(\cup_{k \in \mathcal{K}_H(h)} (Tx + \Lambda_k(h)) \right), \tag{34}
\end{aligned}$$

where the second and third equality follow because of Property 1 and 2 of Theorem 1, respectively. In addition, for the first term of the right-hand side of (34), it follows that

$$\begin{aligned}
&\cup_{k \in \mathcal{K}_{\mathcal{P}}(h)} \left(Tx + (\Lambda_k(h) \setminus \mathcal{P}_k(h)) \right) \\
&\subseteq \cup_{k \in \mathcal{K}_{\mathcal{P}}(h)} \cup_{i \in I_k} \{s \in \mathbb{R}^{m_\omega} : (-B_{k,h}^{-1})_i h \leq (-B_{k,\omega}^{-1})_i (s - Tx) \\
&\quad \leq (-B_{k,h}^{-1})_i h + (\hat{\delta}_k)_i\} \\
&:= \cup_{k \in \mathcal{K}_{\mathcal{P}}(h)} \cup_{i \in I_k} H_{ki}(x),
\end{aligned}$$

where $H_{ki}(x)$ is a hyperslice for every $k \in \mathcal{K}_{\mathcal{P}}(h)$, $i \in I_k$ and $x \in X$. For the second term of (34), from Definition 7, we know that for every $k \in \mathcal{K}_H(h)$, the associated set $\Pi_k(h)$, and

consequently $Tx + \Pi_k(h)$, can be covered by some hyperslice. By Lemma 5, $\Lambda_k(h)$ can also be covered by a hyperslice. Shifting that hyperslice by Tx yields a hyperslice $H_k(x)$ covering $Tx + \Lambda_k(h)$. Therefore,

$$\bigcup_{k \in \mathcal{K}_H(h)} (Tx + \Lambda_k(h)) \subseteq \bigcup_{k \in \mathcal{K}_H(h)} H_k(x).$$

Hence, $\mathcal{N}_x$ is covered by a collection of hyperslices and it satisfies

$$\mathcal{N}_x \subseteq \left(\bigcup_{k \in \mathcal{K}_P(h)} \cup_{i \in I_k} H_{ki}(x) \right) \cup \left(\bigcup_{k \in \mathcal{K}_H(h)} H_k(x) \right).$$

To apply the probability bound from [19, Theorem 4.6] to the second term in (29), it remains to further show that there exists a constant $\beta > 0$ such that $|v(\omega, x) - \hat{v}_\alpha(\omega, x)| \leq \beta$ for all $\omega \in \mathcal{N}_x$ such that

$$\left| \int_{\mathcal{N}_x} (v(s, x) - \hat{v}_\alpha(s, x)) f(s) ds \right| \leq \beta \left| \int_{\mathcal{N}_x} f(s) ds \right|.$$

Indeed, it follows from Lemma 6 that such a β exists, and we have

$$\begin{aligned} & \left| \int_{\mathcal{N}_x} (v(s, x) - \hat{v}_\alpha(s, x)) f(s) ds \right| \\ & \leq \beta \left| \int_{\mathcal{N}_x} f(s) ds \right| \\ & \leq \beta \mathbb{P} \left(\omega \in \left(\bigcup_{k \in \mathcal{K}_P(h)} \cup_{i \in I_k} H_{ki}(x) \right) \cup \left(\bigcup_{k \in \mathcal{K}_H(h)} \Lambda_k(h) \right) \right) \\ & \leq \beta \left(\sum_{k \in \mathcal{K}_P(h)} \sum_{i \in I_k} \mathbb{P}(\omega \in H_{ki}(x)) + \sum_{k \in \mathcal{K}_H(h)} \mathbb{P}(\omega \in H_k(x)) \right). \end{aligned}$$

Applying the probability bound from [19, Theorem 4.6], we have

$$\begin{aligned} & \left| \int_{\mathcal{N}_x} (v(s, x) - \hat{v}_\alpha(s, x)) f(s) ds \right| \\ & \leq \beta \left(\sum_{k \in \mathcal{K}_P(h)} \sum_{i \in I_k} D_{k,i} + \sum_{k \in \mathcal{K}_H(h)} D_k \right) \sum_{j=1}^{m_\omega} \mathbb{E}_{\omega-j} \left[|\Delta| f_j(\cdot | \omega-j) \right], \end{aligned} \quad (35)$$

where $D_{k,i}$ and D_k denote the constants that correspond to the hyperslices $H_{ki}(x)$ and $H_k(x)$, respectively. $\square$

4.3.3 Uniform error bound for the generic $\hat{Q}_\alpha$

We are now set to present a uniform error bound for the α -approximation $\hat{Q}_\alpha$ of Q .

Theorem 3 *Consider the mixed-integer recourse function $Q(x) = \mathbb{E}_\omega[v(\omega, x)]$ and its α -approximation $\hat{Q}_\alpha(x) = \mathbb{E}_\omega[\hat{v}_\alpha(\omega, x)]$. Then, there exists some constant $C > 0$ such that for every $f \in \mathcal{H}^{m_\omega}$, we have*

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

Proof Let $x \in X$ be given, and recall from (29) that we have

$$\begin{aligned} & \left| Q(x) - \hat{Q}_\alpha(x) \right| \\ & \leq \left| \int_{\cup_{k \in \mathcal{K}_{\mathcal{P}}(h)} (Tx + \mathcal{P}_k(h))} (v(s, x) - \hat{v}_\alpha(s, x)) f(s) ds \right| \\ & \quad + \left| \int_{\mathcal{N}_x} (v(s, x) - \hat{v}_\alpha(s, x)) f(s) ds \right|. \end{aligned}$$

Applying the error bounds of Lemmas 8 and 9 to the first and second term of the right-hand side, respectively, we have

$$\|Q - \hat{Q}_\alpha\|_\infty \leq C \sum_{i=1}^{m_\omega} \mathbb{E}_{\omega_{-i}} \left[|\Delta| f_i(\cdot | \omega_{-i}) \right],$$

where the constant C is given as

$$C := \sum_{k \in \mathcal{K}(h)} \beta_k |\det(B_k)| + \beta \left(\sum_{k \in \mathcal{K}_{\mathcal{P}}(h)} \sum_{i \in I_k} D_{k,i} + \sum_{k \in \mathcal{K}_H(h)} D_k \right).$$

This completes the proof. $\square$

Theorem 3 shows that the approximation error of the α -approximation $\hat{Q}_\alpha$ depends only on the total variations of the density functions of the random parameters. Hence, if these total variations converge to zero, then the approximation error vanishes and Q approaches the convex function $\hat{Q}_\alpha$. Intuitively, this convergence occurs when the variability of the random parameters increases. This establishes the asymptotic convexity of the generic expected cost-to-go function Q .

This generic setting applies directly to two-stage models and extends the earlier two-stage results in [19]. In particular, it establishes asymptotic convexity for 2-SMIPs in which some constraints defining the value function are independent of the random parameters, as is the case, for example, with binary second-stage decisions.

5 Asymptotic convexity for M-SMIPs

In this section, we derive asymptotic convexity of M-SMIPs. First, in Section 5.1 we consider a single stage t and use the approximation $\hat{Q}_{t, \alpha_t}$ of Q_t and Theorem 3 to show that Q_t is asymptotically convex. Next, in Section 5.2, we show that all expected cost-to-go functions $Q_t, t = 2, \dots, T$, are *simultaneously* asymptotically convex if the total variations of the density functions of the random parameters in all stages are simultaneously small enough.

5.1 Asymptotic convexity for individual expected cost-to-go functions

Consider the expected cost-to-go function Q_t for a given stage $t = 2, \dots, T$, as defined in (1). We will show that Q_t converges to a convex function if the variability of ω_t is large enough, while keeping all remaining parameters and distributions the same. In particular, we will derive an α -approximation $\hat{Q}_{t, \alpha_t}$ of Q_t such that the maximal

approximation error $\|Q_t - \hat{Q}_{t,\alpha_t}\|_\infty$ converges to zero if the total variations of the probability density functions of ω_t converge to zero.

The stage-wise independence assumption can also be relaxed to allow dependence among the future-stage random vectors $\omega_{t+1}, \dots, \omega_T$, provided that their joint distribution is represented by a finite scenario tree. In this extension, ω_t remains continuously distributed, whereas the future-stage uncertainty is described by finitely many scenarios. Although this differs from the continuous distribution setting considered above, finite scenario trees can approximate many continuous distributions arbitrarily well [18].

Under this finite scenario tree assumption for $\omega_{t+1}, \dots, \omega_T$, the expected cost-to-go function Q_t is

$$Q_t(x_{t-1}) = \mathbb{E}_{\omega_t}[v_t(\omega_t, x_{t-1})],$$

with the value function $v_t(\omega_t, x_{t-1})$ given by

$$\begin{aligned} v_t(\omega_t, x_{t-1}) = \min_{\substack{x_t \in X_t \\ \tilde{x}_n \in X_{t(n)}, n \in \mathcal{S}}} & c_t^\top x_t + \sum_{n \in \mathcal{S}} p_n c_n^\top \tilde{x}_n \\ \text{s.t.} & W_t x_t + T_t x_{t-1} \geq \omega_t, \\ & W_{t(n)} \tilde{x}_n + T_{t(n)} x_t \geq \tilde{\omega}_n, \quad \forall n \in \mathcal{S}_{t+1}, \\ & W_{t(n)} \tilde{x}_n + T_{t(n)} \tilde{x}_{a(n)} \geq \tilde{\omega}_n, \quad \forall n \in \mathcal{S} \setminus \mathcal{S}_{t+1}, \end{aligned} \quad (36)$$

where $\mathcal{S}$ denotes the set of nodes in the finite scenario tree representing the uncertainty from stages $t+1, \dots, T$, $\mathcal{S}_{t+1} \subseteq \mathcal{S}$ denotes the set of nodes associated with stage $t+1$, $t(n)$ denotes the stage associated with node n , $a(n)$ denotes the ancestor node of n , and p_n denotes the probability of reaching node n . Moreover, $\tilde{\omega}_n$ denotes the realization of the random parameter associated with node n , $\tilde{x}_n$ denotes the decision variable associated with node n , and $W_{t(n)}$ and $T_{t(n)}$ denote the recourse and technology matrices, respectively, associated with stage $t(n)$.

Observe that v_t satisfies the generic form v given in Definition 1. Under this setting, we can define an α -approximation $\hat{Q}_{t,\alpha_t}$ from Definition 5 for Q_t and obtain a uniform error bound for $\|Q_t - \hat{Q}_{t,\alpha_t}\|_\infty$, similar as for the generic expected mixed-integer value function Q in Section 4.

Theorem 4 *Let $t = 2, \dots, T$ be given and consider the expected cost-to-go function Q_t of an M-SMIP model defined in (1), where the distributions of the random parameters $\omega_{t+1}, \dots, \omega_T$ can be represented by a finite scenario tree. Then, there exists a constant C_t such that for every probability density function $f \in \mathcal{H}^{m_{\omega_t}}$, the error between Q_t and a convex function $\hat{Q}_{t,\alpha_t}$, is uniformly bounded by*

$$\|Q_t - \hat{Q}_{t,\alpha_t}\|_\infty \leq C_t \sum_{i=1}^{m_{\omega_t}^t} \mathbb{E}_{\omega_{t,-i}} \left[|\Delta| f_i(\cdot | \omega_{t,-i}) \right].$$

Proof Under the deterministic equivalent formulation of the value function v_t in (36), we let

$$x_{\text{DE}} := \left(x_t^\top, \tilde{x}_{n_1}^\top, \dots, \tilde{x}_{n_S}^\top \right)^\top,$$

where the nodes in $\mathcal{S}$ are ordered such that each parent node precedes its child nodes. Moreover, we let W_{DE} denote the coefficient matrix of x_{DE} in the deterministic-equivalent MILP (36). In addition, the right-hand side components consist of the random elements $(\omega_t - T_{t+1}x_t)$ and the fixed elements $(\tilde{\omega}_{n_1}, \dots, \tilde{\omega}_{n_S})$.

Observe that since $W_1, \dots, W_{T-1}$ are complete recourse matrices, W_{DE} is a complete recourse matrix as well. Moreover, since the elements of every $T_{t(n_s)}$ are rational, the recourse matrix W_{DE} can be scaled to an integer matrix without changing the feasibility conditions. Therefore, we are set to apply the α -approximation $\hat{Q}_{t,\alpha_t}$ for Q_t as defined in Definition 5. Using Theorem 3, it follows that there exists a constant $C_t > 0$ such that the error of $\hat{Q}_{t,\alpha_t}$ is bounded by

$$\|Q_t - \hat{Q}_{t,\alpha_t}\|_\infty \leq C_t \sum_{i=1}^{m_\omega^t} \mathbb{E}_{\omega_{t,-i}} \left[|\Delta| f_i(\cdot | \omega_{t,-i}) \right].$$

This completes the proof. $\square$

Theorem 4 shows that the randomness in stage t has a *smoothing* effect on the expected cost-to-go function Q_t . If the variability of the stage- t parameters is large enough, then this smoothing effect is so strong that Q_t becomes arbitrarily close to a convex function. Interestingly, this means that the stage- t variability may also smooth any non-convexities induced by later time stages. Indeed, if ω_τ with $\tau > t$ follows a degenerate distribution, then the expected cost-to-go function Q_τ is possibly equal to a discontinuous, non-convex function. Thus, the variability of ω_t in stage t alone is sufficient to induce asymptotic convexity of Q_t .

If the total variation bound in Theorem 4 is sufficiently small, then we may replace Q_t by $\hat{Q}_{t,\alpha_t}$ in an approximation of the M-SMIP in (1). The convexity in $\hat{Q}_{t,\alpha_t}$ can then be exploited in decomposition algorithms, e.g., using optimality cuts, similarly as in [30]. This approach may work well if the large-scale deterministic MIP corresponding to the mixed-integer value function v_t in (36) is not too large. However, the number of variables and constraints in this MIP grows with the number of nodes in the future-stage scenario tree, which can rapidly increase with the number of remaining time stages. In this case, other results for M-SMIPs have to be used to efficiently derive (approximate) optimality cuts.

Alternatively, when the variability of the random parameters is sufficiently large at *every* stage, we can pursue a recursive approach based on α -approximations of the expected cost-to-go functions at all stages. Such an approach could avoid explicitly constructing the large-scale deterministic equivalent of the remaining stages and provide a theoretical basis for SDDP-like decomposition algorithms for approximately solving M-SMIPs with large scenario trees. We discuss this approach in the following section.

5.2 Simultaneous asymptotic convexity of every Q_t

Theorem 4 in the previous subsection shows that for a *fixed* stage t , the expected cost-to-go function Q_t is asymptotically convex as the variability of ω_t increases while

keeping the distributions of the random parameters in all future stages fixed. In this subsection, we prove that the expected cost-to-go functions in all stages $t = 2, \dots, T$ are also *simultaneously* asymptotically convex as the variability of $\omega_2, \dots, \omega_T$ increase simultaneously. We formalize this result in Theorem 6 below. To prove Theorem 6, we recursively derive a total variation error bound in the spirit of Theorem 3 to an α -approximation of each stage's expected cost-to-go function Q_t . It turns out that in order to apply the recursion, the α -approximations need to be polyhedral, i.e., piecewise affine. For this reason, we first present a result that shows that the α -approximations can be approximated by some polyhedral functions arbitrarily well.

Theorem 5 Consider any function $G : \mathbb{R}^m \mapsto \mathbb{R}$ that satisfies the form

$$G(x) := \mathbb{E}_\omega \left[\max_{k \in \{1, \dots, K\}} \{a_k^\top x + b_k(\omega)\} \right], \quad (37)$$

with $\mathbb{E}_\omega[|b_k(\omega)|] < \infty$ for every $k = 1, \dots, K$. Then for every $\epsilon > 0$, there exists a convex polyhedral function $\mathcal{G}$ of the form

$$\mathcal{G}(x) := \max_{j \in \{1, \dots, J\}} \{d_j^\top x + b_j\} \quad (38)$$

that satisfies

$$\|G - \mathcal{G}\|_\infty \leq \epsilon.$$

Moreover, if all vectors $a_k, k = 1, \dots, K$, are rational, then the vectors d_j and $b_j, j = 1, \dots, J$, can be chosen to be rational, too.

Proof See Appendix A.3. □

Remark 3 We call a convex polyhedral function *rational* if it is of the form in (38) with vectors d_j and b_j being rational for every $j = 1, \dots, J$.

Since the α -approximation $\hat{Q}_\alpha$ in Definition 5 is of the form (37), it follows from Theorem 5 that for every $\epsilon > 0$, there exists a convex polyhedral function $\hat{\mathcal{Q}}_\alpha$ such that $\|\hat{Q}_\alpha - \hat{\mathcal{Q}}_\alpha\|_\infty \leq \epsilon$. We use this observation to derive our simultaneous asymptotic convexity result.

Theorem 6 Consider the expected cost-to-go function $Q_t, t = 2, \dots, T$, of the M-SMIP model of (1). Under Assumption 1, for any $\epsilon > 0$, there exist density functions $f_{\omega_2}, \dots, f_{\omega_T}$ for the random parameters $\omega_2, \dots, \omega_T$ such that every expected cost-to-go function Q_t is ϵ -close to a convex polyhedral function $\hat{\mathcal{Q}}_t$, that is,

$$\|Q_t - \hat{\mathcal{Q}}_t\|_\infty \leq \epsilon, \quad \forall t = 2, \dots, T.$$

Proof Let any arbitrary $\epsilon > 0$ be given. We will prove by backward induction that for every $t = T + 1, \dots, 2$, there exist density functions $f_{\bar{\omega}_t}, \dots, f_{\bar{\omega}_T}$ and rational convex polyhedral functions $\hat{Q}_t, \dots, \hat{Q}_{T+1}$, so that

$$\|Q_\tau - \hat{Q}_\tau\|_\infty \leq \frac{T+1-\tau}{T-1}\epsilon, \quad \forall \tau = t, \dots, T+1.$$

The desired result then follows since $\frac{T+1-\tau}{T-1}\epsilon \leq \epsilon$ for all $\tau = 2, \dots, T+1$.

Observe that the claim directly holds for $t = T + 1$ since $Q_{T+1} \equiv 0$. Next, to prove the induction step, let $t = T, \dots, 2$, be given, and assume that the claim holds for $t + 1$. Then, there exists a rational convex polyhedral function $\hat{Q}_{t+1}$ such that $\|Q_{t+1} - \hat{Q}_{t+1}\|_\infty \leq \frac{T-t}{T-1}\epsilon$. Moreover, since $\hat{Q}_{t+1}$ is rational and convex polyhedral, it can be written as $\hat{Q}_{t+1}(x_t) := \max_{j \in \{1, \dots, J\}} \{d_j^\top x_t + b_j\}$ with d_j and $b_j, j = 1, \dots, J$, rational. Now, we define the approximation $\tilde{v}_t$ as the value function v_t with Q_{t+1} replaced by $\hat{Q}_{t+1}$, i.e.,

$$\begin{aligned} \tilde{v}_t(\omega_t, x_{t-1}) &= \min_{x_t \in X_t} \left\{ c_t^\top x_t + \hat{Q}_{t+1}(x_t) : T_t x_{t-1} + W_t x_t = \omega_t \right\} \\ &= \min_{\substack{x_t \in X_t, \theta_{t+1}, \\ s_1, \dots, s_J \geq 0}} \left\{ (c_t^\top - 1) \begin{pmatrix} x_t \\ \theta_{t+1} \end{pmatrix} : W_t x_t = \omega_t - T_t x_{t-1}, \right. \\ &\quad \left. \theta_{t+1} - d_j^\top x_t - s_j = b_j, \forall j \right\}, \end{aligned}$$

where we use an epigraph reformulation for $\hat{Q}_{t+1}$ with epigraph variable θ_{t+1} and constraints $\theta_{t+1} \geq d_j^\top x_t + b_j, j = 1, \dots, J$, and where we use slack variables $s_1, \dots, s_J \geq 0$ to transform these inequalities into equalities. Furthermore, we define the approximation $\tilde{Q}_t$ of the expected cost-to-go function Q_t as $\tilde{Q}_t(x_{t-1}) := \mathbb{E}_{\omega_t}[\tilde{v}_t(\omega_t, x_{t-1})]$.

Observe that $\tilde{v}_t$ is a mixed-integer value function of the form (4) with coefficient matrix

$$\tilde{W}_t := \begin{pmatrix} W_t & 0 & 0 \\ -D & e & -I \end{pmatrix},$$

where the rows of D are $d_j^\top, j = 1, \dots, J$, and where e represents the all-one vector, and with right-hand side $(\omega_t - T_t x_{t-1}, b)$ where ω_t is random and b is deterministic. Moreover, since the components of D are rational, we may scale the corresponding constraints and assume without loss of generality that $\tilde{W}_t$ is an integer matrix. In addition, it holds that $-\infty < \tilde{v}_t(\omega_t, x_{t-1}) < +\infty$, since

$$\|\tilde{v}_t - v_t\|_\infty \leq \|\hat{Q}_{t+1} - Q_{t+1}\|_\infty < \epsilon,$$

and v_t is finite by Assumption 1. Hence, $\tilde{v}_t$ satisfies all conditions in Assumption 2, which implies that we may construct an α -approximation $\hat{Q}_t$ for $\tilde{Q}_t$ as described in Definition 5, for which the error bound in Theorem 3 holds. Hence, there exists a density function $f_{\bar{\omega}_t}$ with sufficiently small total variation $|\Delta|f_{\bar{\omega}_t}|$, satisfying $\mathbb{E}_{\bar{\omega}_t}[\|\bar{\omega}_t\|_\infty] < +\infty$, so that

$$\|\hat{Q}_t - \tilde{Q}_t\|_\infty \leq \frac{1}{2(T-1)}\epsilon.$$

Moreover, observe that $\hat{Q}_t$ is of the form (37) with a_k rational, $k = 1, \dots, K$, since by Definition 5, for any generic α -approximation, the a_k correspond to $q_{B_k}^\top B_{k,\omega}^{-1}$, and in our case the cost vector is rational by Assumption 1, and $B_{\omega,k}^{-1}$ is rational since B_k is a submatrix of an integer matrix. Hence, by Theorem 5, there exists a rational convex polyhedral approximation $\hat{Q}_t$ of $\hat{Q}_t$ so that $\|\hat{Q}_t - \hat{Q}_t\|_\infty \leq \frac{1}{2(T-1)}\epsilon$. We conclude that by the triangle inequality

$$\begin{aligned} \|Q_t - \hat{Q}_t\|_\infty &\leq \|Q_t - \tilde{Q}_t\|_\infty + \|\tilde{Q}_t - \hat{Q}_t\|_\infty + \|\hat{Q}_t - \hat{Q}_t\|_\infty \\ &\leq \frac{T-t}{T-1}\epsilon + \frac{1}{2(T-1)}\epsilon + \frac{1}{2(T-1)}\epsilon \end{aligned}$$

$$= \frac{T+1-t}{T-1}\epsilon,$$

where the inequality $\|Q_t - \tilde{Q}_t\|_\infty \leq \frac{T-t}{T-1}\epsilon$ holds, because $\|Q_t - \tilde{Q}_t\|_\infty \leq \|Q_{t+1} - \tilde{Q}_{t+1}\|_\infty$ and $\|Q_{t+1} - \tilde{Q}_{t+1}\|_\infty \leq \frac{T-t}{T-1}\epsilon$ by the induction hypothesis. This proves the induction step and completes the backward induction proof. $\square$

Theorem 6 establishes asymptotic convexity of M-SMIPs at all stages simultaneously. In particular, we show that there exist density functions $f_{\omega_2}, \dots, f_{\omega_T}$ with sufficiently small total variations such that every expected cost-to-go function $Q_2, \dots, Q_T$ can be approximated arbitrarily well by some convex polyhedral approximation. This result provides a theoretical foundation for approximating M-SMIPs through α -approximations of their expected cost-to-go functions and suggests the potential for SDDP-style solution methods when the variability of the random parameters is sufficiently large.

6 Conclusion

The main insight of this paper is that randomness of the right-hand side vectors of M-SMIPs can have a convexifying effect on the expected cost-to-go functions of M-SMIPs. Although integer decisions make these expected cost-to-go functions non-convex, we show that these functions can become arbitrarily close to convex functions as the variability of the random parameters increases. Contrary to known results in the literature, this result applies to both 2-SMIPs and M-SMIPs, and also holds when some right-hand side parameters are fixed, e.g., when some of the later-stage decisions are binary.

An important tool for obtaining this result is the adapted Gomory relaxation introduced in this paper. In contrast to the original Gomory relaxation for mixed-integer value function, the adapted Gomory relaxation preserves the non-negativity constraints associated with critical basic variables. This allows us to derive the semi-periodicity for a generic mixed-integer value function with deterministic terms in the right-hand side, and in turn asymptotic convexity of M-SMIPs.

Our results provide a different computational perspective on M-SMIPs. Most existing solution approaches are designed to handle the non-convexity induced by mixed-integer decisions directly through specialized decomposition methods. Our results suggest that, when the variability of the random parameters is sufficiently large, it may instead be beneficial to replace the non-convex expected cost-to-go functions by convex approximations and solve the resulting approximate multistage problem using methods developed for convex stochastic programs.

Therefore, a direct potential direction for future research is to develop algorithms for the approximate M-SMIPs. The challenge in the algorithmic design of such algorithms is that the convex approximating expected cost-to-go functions are not given at the start but need to be iteratively constructed in SDDP-style using optimality cuts. Hence, the α -approximation of Definition 5 is applied in every iteration and depends on the current representation of the next-stage expected cost-to-go function. Future research is required to prove that our main convergence result also holds for

such an algorithm. Furthermore, since the total variation error bounds for these convex approximations require the random parameters to be continuously distributed, whereas practical implementations rely on finitely many scenarios in each iteration, the resulting approximation error may depend on the sampling method used. Therefore, additional analysis is needed to derive a performance guarantee for such algorithms. Another challenge is that an LP generally has exponentially many basis matrices, and that our approximation requires a specific subset of these basis matrices. It remains an open question how to identify these basis matrices efficiently, and whether an approximation in the spirit of [30] is possible.

Other future research directions include, e.g., constructing convex approximations for multistage (distributionally) robust mixed-integer programs, as well as for M-SMIPs involving alternative risk measures. In these settings, it is worthwhile to investigate whether similar asymptotic convexity properties can be derived when the expectation operator is replaced by a worst-case expectation, or a risk functional. Establishing accurate convex approximations in these settings could continue to broaden the approximation framework, and provide practical solutions to a broader class of multistage decision-making problems.

Appendix A.1. Auxiliary results for Section 3

In this appendix, we introduce several intermediate findings that are needed to derive some of the results in Section 3. We first present a classical proximity bound, originally introduced for integer programs and their LP relaxations [24] and later extended to the mixed-integer setting in [7]. We present the latter below, and we apply it to prove the boundedness of the elements of some optimal solution $y_{N_k}^*(\omega, x)$ of the adapted Gomory relaxation v_{B_k} .

Theorem 7 *Let $M \in \mathbb{Z}^{m \times n}$ be a matrix with integer entries, $c \in \mathbb{R}^n$ a cost vector, and $b \in \mathbb{R}^m$ a right-hand-side vector. Assume that the mixed-integer linear program*

$$\min_{z \in \mathbb{R}^n \times \mathbb{Z}^n} \{c^\top z : Mz \leq b\} \quad (\text{MIP})$$

attains an optimal solution. Let z^{LP} be any optimal basic feasible solution of the linear relaxation

$$\min_{z \in \mathbb{R}^{n+n}} \{c^\top z : Mz \leq b\}. \quad (\text{LP})$$

Then, there exists an optimal mixed-integer solution z^{MIP} to (MIP) such that

$$\|z^{\text{LP}} - z^{\text{MIP}}\|_\infty \leq n \Delta(M),$$

where $\Delta(M)$ denotes the maximum absolute value of the determinant of any square submatrix of M .

Next, Lemma 10 provides a key result used to derive Properties 1 and 2 of Theorem 1. This theorem is related to the Basis Decomposition Theorem of [33]. Here, we show that two relevant properties are preserved in a lower-dimensional setting.

Lemma 10 *Let $\Omega_k := \{s \in \mathbb{R}^m : A_k s \geq b_k\}, k = 1, \dots, K$, be polyhedra such that*

- (A) $\text{int}(\Omega_k) \neq \emptyset$ for all $k = 1, \dots, K$,
 (B) $\cup_{k=1}^K \Omega_k = \mathbb{R}^m$, and
 (C) $\text{int}(\Omega_k) \cap \text{int}(\Omega_l) = \emptyset$ for all $k, l = 1, \dots, K$ with $k \neq l$.

Moreover, let $h := (h_1, \dots, h_n) \in \mathbb{R}^n$ be given with $n < m$ and define the polyhedra $\Omega_k(h)$ for every $k = 1, \dots, K$ as

$$\Omega_k(h) := \{s \in \mathbb{R}^{m-n} : A_k(s, h) \geq b_k\}.$$

Then, there exists an index set $\mathcal{K}(h) \subseteq \{1, \dots, K\}$ such that

- (i) $\text{int}(\Omega_k(h)) \neq \emptyset$ for all $k \in \mathcal{K}(h)$,
 (ii) $\cup_{k \in \mathcal{K}(h)} \Omega_k(h) = \mathbb{R}^{m-n}$, and
 (iii) $\text{int}(\Omega_k(h)) \cap \text{int}(\Omega_l(h)) = \emptyset$ for all $k, l \in \mathcal{K}(h)$ with $k \neq l$.

Proof We will prove the result for $n = 1$. The claim then holds in general by applying mathematical induction on the dimension n . Hence, let $n = 1$ and polyhedra $\Omega_k, k = 1, \dots, K$, satisfying (A)-(C) be given. Denote the halfspaces $H_{>,h} \subseteq \mathbb{R}^m$ and $H_{\geq,h} \subseteq \mathbb{R}^m$, respectively, by

$$H_{>,h} := \{t \in \mathbb{R}^m : t_m > h\} \text{ and } H_{\geq,h} := \{t \in \mathbb{R}^m : t_m \geq h\}.$$

We will show that the index set $\mathcal{K}(h) \subseteq \{1, \dots, K\}$ defined as

$$\mathcal{K}(h) = \{k \in \{1, \dots, K\} : \Omega_k \cap H_{>,h} \neq \emptyset \text{ and } \text{int}(\Omega_k(h)) \neq \emptyset\}$$

satisfies (i)-(iii).

Observe that (i) holds immediately by definition of $\mathcal{K}(h)$. To prove (ii), we first consider $\hat{\mathcal{K}}(h) \supseteq \mathcal{K}(h)$ defined as $\hat{\mathcal{K}}(h) := \{k \in \{1, \dots, K\} : \Omega_k \cap H_{>,h} \neq \emptyset\}$. Since the $\Omega_k, k = 1, \dots, K$, cover $\mathbb{R}^m$ by assumption (B), it follows directly that

$$\cup_{k \in \hat{\mathcal{K}}(h)} (\Omega_k \cap H_{>,h}) = H_{>,h}. \quad (39)$$

Indeed, for every $t \in H_{>,h}$ there exists $k \in \hat{\mathcal{K}}(h)$ such that $t \in \Omega_k$, and moreover $\Omega_k \cap H_{>,h} \subseteq H_{>,h}$ for all $k \in \hat{\mathcal{K}}(h)$. Furthermore, observe that from (39) it follows that $\cup_{k \in \hat{\mathcal{K}}(h)} (\Omega_k \cap H_{\geq,h})$ is a closed superset of $H_{>,h}$ that is contained in $H_{\geq,h}$. That is, $H_{>,h} \subseteq \cup_{k \in \hat{\mathcal{K}}(h)} (\Omega_k \cap H_{\geq,h}) \subseteq H_{\geq,h}$. Since $H_{\geq,h}$ is the closure of $H_{>,h}$, i.e., the smallest closed set that contains $H_{>,h}$, it follows that $\cup_{k \in \hat{\mathcal{K}}(h)} (\Omega_k \cap H_{\geq,h}) = H_{\geq,h}$. Hence, for every $s \in \mathbb{R}^{m-1}$, there exists a $k \in \hat{\mathcal{K}}(h)$ such that $(s, h) \in \Omega_k$, and thus

$$\cup_{k \in \hat{\mathcal{K}}(h)} \Omega_k(h) = \mathbb{R}^{m-1}. \quad (40)$$

It remains to show that $\cup_{k \in \mathcal{K}(h)} \Omega_k(h) = \mathbb{R}^{m-1}$ as well. Given (40), it follows that one of those polyhedra $\Omega_k(h)$ satisfies $\text{int}(\Omega_k(h)) \neq \emptyset$ with $k \in \hat{\mathcal{K}}(h)$, and hence we have $\cup_{k \in \mathcal{K}(h)} \Omega_k(h) \neq \emptyset$. Moreover, $\cup_{k \in \mathcal{K}(h)} \Omega_k(h)$ is closed since it is the union of finitely many closed sets. Hence, its complement $\mathbb{R}^{m-1} \setminus (\cup_{k \in \mathcal{K}(h)} \Omega_k(h))$ is open. Moreover, observe that

$$\mathbb{R}^{m-1} \setminus (\cup_{k \in \mathcal{K}(h)} \Omega_k(h)) \subseteq (\cup_{k \in \hat{\mathcal{K}}(h) \setminus \mathcal{K}(h)} \Omega_k(h)), \quad (41)$$

since the polyhedra $\Omega_k(h), k \in \hat{\mathcal{K}}(h)$, cover $\mathbb{R}^{m-1}$; see (40). However, the right-hand side of (41) is the finite union of polyhedral sets with an empty interior. This implies that $\mathbb{R}^{m-1} \setminus (\cup_{k \in \hat{\mathcal{K}}(h)} \Omega_k(h))$ cannot be any other set than the empty set, and thus

$$\mathbb{R}^{m-1} \setminus (\cup_{k \in \hat{\mathcal{K}}(h)} \Omega_k(h)) = \emptyset,$$

which implies that $\cup_{k \in \hat{\mathcal{K}}(h)} \Omega_k(h) = \mathbb{R}^{m-1}$.

Finally, to prove (iii), let $k, l \in \mathcal{K}(h)$ with $k \neq l$ be given, and suppose for contradiction that $\text{int}(\Omega_k(h)) \cap \text{int}(\Omega_l(h)) \neq \emptyset$. That is, there exists some $s \in \mathbb{R}^{m-1}$ such that $s \in \text{int}(\Omega_k(h))$ and $s \in \text{int}(\Omega_l(h))$.

For this s , either $(s, h) \in \text{int}(\Omega_i)$ or $(s, h) \in \partial\Omega_i$ for both $i = k$ and $i = l$. However, because of assumption (C), it is not possible that $(s, h) \in \text{int}(\Omega_k)$ and $(s, h) \in \text{int}(\Omega_l)$. Moreover, if $(s, h) \in \text{int}(\Omega_k)$ and $(s, h) \in \partial(\Omega_l)$, then since Ω_l is a polyhedron with non-empty interior, it holds that for every $\epsilon > 0$, the open ball $\mathcal{B}((s, h), \epsilon)$ centered at (s, h) with radius ϵ , satisfies $\mathcal{B}((s, h), \epsilon) \cap \text{int}(\Omega_l) \neq \emptyset$. Moreover, since $(s, h) \in \text{int}(\Omega_k)$, there exists $\bar{\epsilon} > 0$ such that $\mathcal{B}((s, h), \bar{\epsilon}) \subseteq \text{int}(\Omega_k)$. Combining both observations, they imply that $\mathcal{B}((s, h), \bar{\epsilon})$ contains a point that is both in $\text{int}(\Omega_k)$ and $\text{int}(\Omega_l)$, contradicting assumption (C). By symmetry, the case $(s, h) \in \partial\Omega_k$ and $(s, h) \in \text{int}(\Omega_l)$ also yields a contradiction, hence the only case that remains to be investigated is $(s, h) \in \partial\Omega_k$ and $(s, h) \in \partial\Omega_l$.

Hence, suppose that $(s, h) \in \partial\Omega_k$ and $(s, h) \in \partial\Omega_l$. Since $k, l \in \mathcal{K}(h)$, there exist $t_i \in \Omega_i \cap H_{>,h}$ for $i = k, l$. Moreover, since $s \in \text{int}(\Omega_i(h))$, $i = k, l$, there exists $\bar{\epsilon} > 0$ such that $\mathcal{B}(s, \bar{\epsilon}) \subseteq \text{int}(\Omega_i(h))$ for $i = k, l$. Next, define for $i = k, l$,

$$C_i := \text{conv}\{\{t_i\} \cup \{(s', h) : s' \in \mathcal{B}(s, \bar{\epsilon})\}\}.$$

Since Ω_i is convex, $i = k, l$, it follows that $C_i \subseteq \Omega_i$, and hence $\text{int}(C_i) \subseteq \text{int}(\Omega_i)$ for $i = k, l$. Finally, we observe that by construction of C_i , $i = k, l$, there exists $\hat{\epsilon} > 0$ such that

$$\mathcal{B}((s, h), \hat{\epsilon}) \cap H_{>,h} \subseteq \text{int}(C_i), i = k, l,$$

implying that $\text{int}(C_k) \cap \text{int}(C_l) \neq \emptyset$, and then $\text{int}(\Omega_k) \cap \text{int}(\Omega_l) \neq \emptyset$, contradicting assumption (C). $\square$

Appendix A.2. Proofs of Lemmas 3, 4, and 5

In this section, we provide the proofs of Lemmas 3, 4, and 5.

Proof of Lemma 3. $\Leftarrow$: Suppose that $\text{span}(\text{recc}(\Pi)) \neq \mathbb{R}^m$. Let $V := \text{span}(\text{recc}(\Pi))$. Then, V is a proper subspace of $\mathbb{R}^m$. Denote its orthogonal complement by $V^\perp := \{w \in \mathbb{R}^m : w^\top v = 0 \text{ for all } v \in V\}$. Since $\dim(V) + \dim(V^\perp) = m$, there exists a nonzero $a \in V^\perp$ such that $a^\top d = 0$ for every $d \in \text{recc}(\Pi)$. By the Minkowski–Weyl theorem, there exist finitely many points $z_1, \dots, z_I \in \Pi$ and recession directions $r_1, \dots, r_J$ such that

$$\Pi = \left\{ s \in \mathbb{R}^m : s = \sum_{i=1}^I \mu_i z_i + \sum_{j=1}^J \lambda_j r_j : \sum_{i=1}^I \mu_i = 1, \mu_i, \lambda_j \geq 0, \forall i, j \right\}.$$

Since $a^\top r_j = 0$ for all $j = 1, \dots, J$, every $s \in \Pi$ satisfies $a^\top s = \sum_{i=1}^I \mu_i a^\top z_i$. Consequently, for every $s \in \Pi$, we have

$$\min_{1 \leq i \leq I} a^\top z_i \leq a^\top s \leq \max_{1 \leq i \leq I} a^\top z_i.$$

Since $a \neq 0$, these two finite bounds define a hyperslice that covers Π ; see Definition 6.

$\Rightarrow$: Suppose that $\text{span}(\text{recc}(\Pi)) = \mathbb{R}^m$. Then we can choose m linearly independent directions $r_1, \dots, r_m \in \text{recc}(\Pi)$. Let R be the matrix with these directions as columns.

Since R is invertible, $a^\top R \neq 0$ for every nonzero $a \in \mathbb{R}^m$. Thus, for every nonzero a , there is some $r_j \in \text{recc}(\Pi)$ such that $a^\top r_j \neq 0$. For any $s \in \Pi$, we have $s + \lambda r_j \in \Pi$ for every $\lambda \geq 0$. As $\lambda \rightarrow \infty$, we have $a^\top (s + \lambda r_j) \rightarrow +\infty$ if $a^\top r_j > 0$, and $a^\top (s + \lambda r_j) \rightarrow -\infty$ if $a^\top r_j < 0$. In both cases, we have $\sup_{s \in \Pi} a^\top s - \inf_{s \in \Pi} a^\top s = \infty$, and since this holds for every non-zero a , we conclude that Π cannot be covered by any hyperslice. $\square$

Proof of Lemma 4. If $\Pi = \emptyset$, the result immediately follows from Remark 2. If Π is non-empty and lattice free, then it follows from Lemma 4 in [3] that $\Pi + \text{span}(\text{recc}(\Pi))$ is also lattice free. Hence, $\text{span}(\text{recc}(\Pi)) \neq \mathbb{R}^m$. It then follows directly from Lemma 3 that Π can be covered by a hyperslice. $\square$

Proof of Lemma 5. Fix $k \in \mathcal{K}(h)$. Since $\Pi_k(h) \subseteq \Lambda_k(h)$, if $\Lambda_k(h)$ can be covered by a hyperslice, then $\Pi_k(h)$ can be covered by a hyperslice as well. For the converse, suppose that $\Lambda_k(h)$ cannot be covered by a hyperslice. Then, $\Lambda_k(h)$ is non-empty and the recession cone $\text{recc}(\Lambda_k(h))$ is given by

$$\text{recc}(\Lambda_k(h)) = \{d \in \mathbb{R}^{m_\omega} : (B_{k,\omega}^{-1})_i^\top d \geq 0 \text{ for all } i \in I_k\}.$$

Here, we use that $B_{k,\omega}^{-1} = 0$ for $i \in \bar{I}_k$. By Lemma 3, $\text{recc}(\Lambda_k(h))$ is full-dimensional in $\mathbb{R}^{m_\omega}$. We may therefore choose $d \in \text{int}(\text{recc}(\Lambda_k(h)))$, so that $(B_{k,\omega}^{-1})_i^\top d > 0$ for every $i \in I_k$. Then, choose $t \geq 0$ sufficiently large such that $t(B_{k,\omega}^{-1})_i^\top d \geq \delta_{k,i}$ for every $i \in I_k$ with $\delta_{k,i}$ given in Theorem 1. For any $s \in \Lambda_k(h)$ and $i \in I_k$, we then have

$$(B_{k,\omega}^{-1})_i^\top (s + td) + (B_{k,h}^{-1})_i h \geq t(B_{k,\omega}^{-1})_i^\top d \geq \delta_{k,i}.$$

Moreover, since $k \in \mathcal{K}(h)$, it holds that $(B_{k,h}^{-1})_i h \geq 0$ for all $i \in \bar{I}_k$, and consequently, $(B_{k,\omega}^{-1})_i^\top (s + td) + (B_{k,h}^{-1})_i h = (B_{k,h}^{-1})_i h \geq 0$, as required in $\Pi_k(h)$. Thus, we have

$$\Lambda_k(h) + td \subseteq \Pi_k(h).$$

Since $\Lambda_k(h)$ cannot be covered by a hyperslice, it follows directly that $\Pi_k(h)$ cannot be covered by any hyperslice either. $\square$

Appendix A.3. Proof of Theorem 5

In this part of the appendix, we prove Theorem 5, which states that for functions $G : \mathbb{R}^m \mapsto \mathbb{R}$ of the form (37), there exists an arbitrarily close convex polyhedral approximation $\mathcal{G}$. It turns out that the Fenchel conjugate G^* of G is crucial in the construction of $\mathcal{G}$.

Definition 10 The Fenchel conjugate $G^* : \mathbb{R}^m \mapsto \mathbb{R}$ of a function $G : \mathbb{R}^m \mapsto \mathbb{R}$ is given by

$$G^*(a) := \sup_x \{a^\top x - G(x)\}, \quad a \in \mathbb{R}^m.$$

Moreover, the effective domain of G^* , denoted as $\text{dom}(G^*)$, is given by $\text{dom}(G^*) := \{a \in \mathbb{R}^m : G^*(a) < +\infty\}$.

Given the form of G , we are able to derive a natural upper bound for G^* , and we are able to completely characterize $\text{dom}(G^*)$. These results are given in Lemmas 11 and 12 below, respectively.

Lemma 11 *For any given $\omega \in \mathbb{R}^m$, let $g_\omega : \mathbb{R}^m \mapsto \mathbb{R}$ be any function satisfying the form*

$$g_\omega(x) := \max_{k \in \{1, \dots, K\}} \{a_k^\top x + b_k(\omega)\},$$

and denote its expectation as $G(x) := \mathbb{E}_\omega[g_\omega(x)]$. Moreover, let $g_\omega^* : \mathbb{R}^m \mapsto \mathbb{R}$ and $G^* : \mathbb{R}^m \mapsto \mathbb{R}$ be the convex conjugate of g_ω and G , respectively. Then, the inequality

$$G^*(a) \leq \mathbb{E}_\omega[g_\omega^*(a)]$$

holds for all $a \in \mathbb{R}^m$.

Proof For any given $\omega \in \mathbb{R}^m$, define the function h_ω as $h_\omega(x) := a^\top x - g_\omega(x)$. Since the inequality $h_\omega(x) \leq \sup_x h_\omega(x)$ holds, taking expectations on both sides yields $\mathbb{E}_\omega[h_\omega(x)] \leq \mathbb{E}_\omega[\sup_x h_\omega(x)]$. Therefore, we have $\sup_x \mathbb{E}_\omega[h_\omega(x)] \leq \mathbb{E}_\omega[\sup_x h_\omega(x)]$, which implies that for all $a \in \mathbb{R}^m$, we have

$$\begin{aligned} G^*(a) &= \sup_x \{a^\top x - \mathbb{E}_\omega[g_\omega(x)]\} = \sup_x \{\mathbb{E}_\omega[h_\omega(x)]\} \\ &\leq \mathbb{E}_\omega[\sup_x \{h_\omega(x)\}] = \mathbb{E}_\omega[\sup_x \{a^\top x - g_\omega(x)\}] \\ &= \mathbb{E}_\omega[g_\omega^*(a)]. \end{aligned}$$

This completes the proof. $\square$

In the next lemma, we characterize the effective domain $\text{dom}(G^*)$.

Lemma 12 *For any given $\omega \in \mathbb{R}^m$, let $g_\omega : \mathbb{R}^m \mapsto \mathbb{R}$ be any function satisfying the form*

$$g_\omega(x) := \max_{k \in \{1, \dots, K\}} \{a_k^\top x + b_k(\omega)\},$$

with $b_k(\omega)$ satisfying $\mathbb{E}_\omega[|b_k(\omega)|] < +\infty$, and denote its expectation as $G(x) := \mathbb{E}_\omega[g_\omega(x)]$. Then, the effective domain $\text{dom}(G^*)$ of the Fenchel conjugate G^* of G is given by

$$\text{dom}(G^*) = \text{conv}(a_1, \dots, a_K).$$

Proof We first prove that $\text{conv}(a_1, \dots, a_K) \subseteq \text{dom}(G^*)$. To do so, it suffices to show that for any $a \in \text{conv}(a_1, \dots, a_K)$, we have $g_\omega^*(a)$ satisfying $g_\omega^*(a) < +\infty$, so that by Lemma 11 we have that $G^*(a) < +\infty$ also holds. Hence, let $a \in \text{conv}(a_1, \dots, a_K)$ be given, and observe that $g_\omega^*(a)$ can be written as

$$\begin{aligned} g_\omega^*(a) &= \sup_x \{a^\top x - \max_{k \in \{1, \dots, K\}} \{a_k^\top x + b_k(\omega)\}\} \\ &= \sup_{x, \theta} \{a^\top x - \theta : \theta \geq a_k^\top x + b_k(\omega), \forall k \in \{1, \dots, K\}\} \\ &= \sup_{x, \theta} \{a^\top x - \theta : \theta - a_k^\top x \geq b_k(\omega), \forall k \in \{1, \dots, K\}\}. \end{aligned}$$

Moreover, the dual of $g_\omega^*(a)$ is given by

$$g_{\text{dual},\omega}^*(a) = \min_{\lambda_k \geq 0} \left\{ - \sum_{k=1}^K \lambda_k b_k(\omega) : \sum_{k=1}^K \lambda_k a_k = a, \sum_{k=1}^K \lambda_k = 1 \right\}.$$

Observe that the dual $g_{\text{dual},\omega}^*(a)$ will be feasible, and consequently the primal $g_\omega^*(a)$ will be bounded from above, if $a \in \text{conv}(a_1, \dots, a_K)$. Therefore, for any $a \in \text{conv}(a_1, \dots, a_K)$, we have $g_\omega^*(a) < +\infty$. Using weak duality, denote $\bar{\lambda} := (\bar{\lambda}_1, \dots, \bar{\lambda}_K)$ as some feasible solution of $g_{\text{dual},\omega}^*(a)$, we have

$$\mathbb{E}_\omega[g_\omega^*(a)] \leq \mathbb{E}_\omega[g_{\text{dual},\omega}^*(a)] \leq \sum_{k=1}^K \bar{\lambda}_k \mathbb{E}_\omega[-b_k(\omega)] \leq \sum_{k=1}^K \mathbb{E}_\omega[|b_k(\omega)|] < +\infty.$$

Consequently, it follows from Lemma 11 that $G^*(a) \leq \mathbb{E}_\omega[g_\omega^*(a)] < +\infty$.

Next, we will prove that $\text{dom}(G^*) \subseteq \text{conv}(a_1, \dots, a_K)$. Hence, let $a \notin \text{conv}(a_1, \dots, a_K)$ be given. Then, using the hyperplane separation theorem, we know that there exists some non-zero vector $d \in \mathbb{R}^m$ such that $d^\top a > d^\top a_k$ for all $k \in \{1, \dots, K\}$, which implies that $d^\top a - \max_{k \in \{1, \dots, K\}} d^\top a_k > 0$.

Using the fact that

$$\begin{aligned} G(x) &= \mathbb{E}_\omega \left[\max_{k \in \{1, \dots, K\}} \{a_k^\top x + b_k(\omega)\} \right] \\ &\leq \max_k \{a_k^\top x\} + \mathbb{E}_\omega \left[\max_k \{b_k(\omega)\} \right] \\ &\leq \max_k \{a_k^\top x\} + \sum_{k=1}^K \mathbb{E}_\omega[|b_k(\omega)|], \end{aligned}$$

it holds that

$$\begin{aligned} G^*(a) &= \sup_{x \in \mathbb{R}^m} \{a^\top x - G(x)\} \\ &\geq \sup_{t > 0} \{a^\top (td) - G(td)\} \\ &\geq \sup_{t > 0} \left\{ t(a^\top d) - \max_{k \in \{1, \dots, K\}} \{a_k^\top (td)\} - \sum_{k=1}^K \mathbb{E}_\omega[|b_k(\omega)|] \right\} \\ &= \sup_{t > 0} \left\{ t(a^\top d - \max_{k \in \{1, \dots, K\}} \{a_k^\top d\}) - \sum_{k=1}^K \mathbb{E}_\omega[|b_k(\omega)|] \right\} \\ &= +\infty. \end{aligned}$$

Hence, it holds that $a \notin \text{dom}(G^*)$, and we conclude that $\text{dom}(G^*) = \text{conv}(a_1, \dots, a_K)$. $\square$

Next, for every $\epsilon > 0$, we construct a piecewise-constant approximation G_δ^* of G^* such that (i) $\|G^* - G_\delta^*\|_\infty < \epsilon$, (ii) the convex conjugate $\mathcal{G} := (G_\delta^*)^*$ of G_δ^* is convex polyhedral, and (iii) $\|G - \mathcal{G}\|_\infty < \epsilon$.

To do so, we partition $\text{dom}(G^*)$ into sufficiently small cells, each with diameter bounded by some $\delta > 0$, and make sure that the piecewise-constant approximation is constant on each cell.

We first define the piecewise-constant approximation G_δ^* in Definition 11 and visually illustrate the construction of G_δ^* in Figure 2. In Lemma 13, we show that the

absolute error caused by G_δ^* can be bounded by any arbitrarily small $\epsilon > 0$ if δ is made arbitrarily small.

Definition 11 Let $\delta > 0$ be given. For every $l \in \mathbb{Z}^m$, define the cell $\mathcal{D}_\delta^l \subset \mathbb{R}^m$ as $\mathcal{D}_\delta^l := \{a \in \mathbb{R}^m : \frac{\delta}{m}l_i \leq a_i \leq \frac{\delta}{m}(l_i + 1), i = 1, \dots, m\} \cap \text{dom}(G^*)$. Moreover, for every $l \in \mathbb{Z}^m$ for which $\mathcal{D}_\delta^l \neq \emptyset$, let $a_\delta^l \in \mathcal{D}_\delta^l$ be arbitrarily given. Then, the piecewise-constant approximation $G_\delta^*(a)$ is given by

$$G_\delta^*(a) := \inf_{l \in \mathbb{Z}^m : a \in \mathcal{D}_\delta^l} G^*(a_\delta^l).$$

Remark 4 Observe from Definition 11 that a point $a \in \mathbb{R}^m$ may belong to multiple cells $\mathcal{D}_\delta^l$. In this case, we select the minimizer a_δ^l associated with any one of these cells. In addition, observe that if $a \notin \text{dom}(G^*)$, by definition of $\mathcal{D}_\delta^l$, none of the candidate a_δ^l will be feasible and consequently we have $G_\delta^*(a) = +\infty$.

Lemma 13 For every $\epsilon > 0$, there exists a $\bar{\delta} > 0$ such that for all $0 < \delta \leq \bar{\delta}$, the piecewise-constant approximation G_δ^* of G^* satisfies

$$|G_\delta^*(a) - G^*(a)| \leq \epsilon$$

for all $a \in \text{dom}(G^*)$.

Proof Since G^* is the convex conjugate of G , it is lower semicontinuous on $\text{dom}(G^*)$. Moreover, since G^* is convex and $\text{dom}(G^*)$ is a polytope, it follows from [10] that G^* is also upper semicontinuous on $\text{dom}(G^*)$. Therefore, we know that G^* is continuous over the compact domain $\text{dom}(G^*)$, and hence it is also uniformly continuous on $\text{dom}(G^*)$. This implies that for every $\epsilon > 0$, there exists a $\bar{\delta} > 0$ such that for every $a, a' \in \text{dom}(G^*)$ with $\|a - a'\|_1 \leq \bar{\delta}$, it holds that $|G^*(a) - G^*(a')| \leq \epsilon$. Hence, for all $0 < \delta \leq \bar{\delta}$, it holds for all $a \in \text{dom}(G^*)$ that $|G^*(a) - G_\delta^*(a)| = |G^*(a) - G^*(a_\delta^l)|$ for some $l \in \mathbb{Z}^m$ with $a \in \mathcal{D}_\delta^l$; see Definition 11. Observe that $\|a - a_\delta^l\|_1 \leq \text{diam}(\mathcal{D}_\delta^l) := \sup_{a, a' \in \mathcal{D}_\delta^l} \|a - a'\|_1 = \delta$, we conclude that $|G^*(a) - G_\delta^*(a)| < \epsilon$ for all $a \in \text{dom}(G^*)$. $\square$

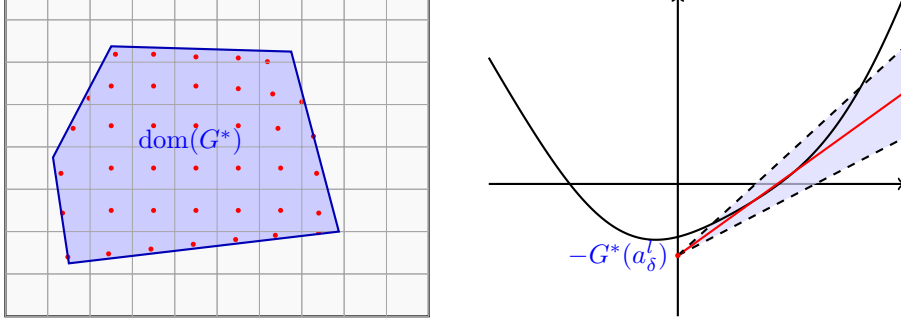


Fig. 2: The figure on the left illustrates several concepts of Definition 11: with the shaded polytope represents the effective domain $\text{dom}(G^*)$, and the non-empty cells $D_\delta^l, l \in \mathbb{Z}^m$ are represented by the intersection between the squares and $\text{dom}(G^*)$. Within each non-empty cell D_δ^l , the red point represents a_δ^l . In the figure on the right, the solid curve is the function G , and the red line is the supporting hyperplane $(a_\delta^l)^\top x - G^*(a_\delta^l)$. The shaded blue cone represents all approximate supporting hyperplanes $a^\top x - G^*(a_\delta^l)$ for $a \in D_\delta^l$ within a single cell. We note that these approximate supporting hyperplanes may exceed the function G . However, by construction they cannot exceed G by more than ϵ .

We now use G_δ^* to construct the approximation $\mathcal{G}$ of G in Theorem 5. Below, we present the proof for Theorem 5.

Proof of Theorem 5. Let $\epsilon > 0$ be given. By Lemma 13 there exists a $\delta > 0$ such that $|G^*(a) - G_\delta^*(a)| \leq \epsilon$ for all $a \in \text{dom}(G^*)$. Define $\mathcal{G} := (G_\delta^*)^*$. It holds for all $x \in \mathbb{R}^m$ that

$$\begin{aligned} \mathcal{G}(x) &= \sup_{a \in \text{dom}(G^*)} \{a^\top x - G_\delta^*(a)\} \\ &= \sup_{\substack{l \in \mathbb{Z}^m: \\ D_\delta^l \neq \emptyset}} \sup_{a \in D_\delta^l} \{a^\top x - G^*(a_\delta^l)\}. \end{aligned}$$

Observe that every non-empty cell D_δ^l is a polytope with finitely many vertices, denoted as $a_1^l, \dots, a_{K_l}^l$. Moreover, for every such cell D_δ^l , the supremum $\sup_{a \in D_\delta^l} \{a^\top x - G^*(a_\delta^l)\}$ is attained in one of these vertices, since the objective is linear in a . Hence,

$$\mathcal{G}(x) = \sup_{\substack{l \in \mathbb{Z}^m: \\ D_\delta^l \neq \emptyset}} \sup_{k=1, \dots, K_l} \{(a_k^l)^\top x - G^*(a_\delta^l)\}, \quad x \in \mathbb{R}^m,$$

and since there are finitely many cells D_δ^l with $D_\delta^l \neq \emptyset$, we conclude that $\mathcal{G}$ is convex polyhedral.

To prove that $\|G - \mathcal{G}\|_\infty \leq \epsilon$, we use the fact that $G = (G^*)^*$ since G is closed and convex. Then, for every $x \in \mathbb{R}^m$, we have that

$$\begin{aligned}
G(x) &= G^{**}(x) \\
&= \sup_{a \in \text{dom}(G^*)} \{a^\top x - G^*(a)\} \\
&= \sup_{a \in \text{dom}(G^*)} \{a^\top x - G_\delta^*(a) + G_\delta^*(a) - G^*(a)\} \\
&\leq \sup_{a \in \text{dom}(G^*)} \{a^\top x - G_\delta^*(a)\} + \sup_{a \in \text{dom}(G^*)} \{G_\delta^*(a) - G^*(a)\} \\
&\leq \mathcal{G}(x) + \epsilon,
\end{aligned}$$

and thus $G(x) - \mathcal{G}(x) \leq \epsilon$. By symmetry, we also have

$$\begin{aligned}
\mathcal{G}(x) &= \sup_{a \in \text{dom}(G^*)} \{a^\top x - G_\delta^*(a)\} \\
&= \sup_{a \in \text{dom}(G^*)} \{a^\top x - G_\delta^*(a) + G^*(a) - G^*(a)\} \\
&\leq \sup_{a \in \text{dom}(G^*)} \{a^\top x - G^*(a)\} + \sup_{a \in \text{dom}(G^*)} \{G^*(a) - G_\delta^*(a)\} \\
&\leq G(x) + \epsilon,
\end{aligned}$$

which leads to $\mathcal{G}(x) - G(x) \leq \epsilon$. Consequently, we conclude that $\|G - \mathcal{G}\|_\infty \leq \epsilon$.

Finally, it remains to show that the vectors d_j and b_j , $j = 1, \dots, J$, can be chosen to be rational if the vectors a_k , $k = 1, \dots, K$, are rational. To prove d_j , $j = 1, \dots, J$, being rational, observe that (i) by Lemma 13 we may choose $\delta > 0$ rational, and (ii) the vectors d_j , $j = 1, \dots, J$, correspond to vertices of the non-empty cells D_δ^l . Since $\text{dom}(G^*) = \text{conv}(a_1, \dots, a_K)$, and the vectors a_k , $k = 1, \dots, K$, are rational, it follows that the vertices of the non-empty cells D_δ^l are intersections of rational hyperplanes, and thus rational. For b_j , $j = 1, \dots, J$, observe from the definition of $\mathcal{G}$ that it is the intercept $G^*(a_\delta^l)$. In case the intercept $G^*(a_\delta^l)$ is irrational, since $\mathbb{Q}$ is dense in $\mathbb{R}$, we can also replace it by a rational approximation within an arbitrarily small error so that $\|G - \mathcal{G}\|_\infty \leq \epsilon$. $\square$

Acknowledgements. The research of Ward Romeijnders has been supported by the Netherlands Organisation for Scientific Research [Grant VI.Vidi.211.066]. We thank Niels van der Laan for his help with the initial analysis for some of the results in this paper.

Conflict of interest

The authors declare that they have no conflict of interest.

Data availability

This paper does not make use of any data.

References

- [1] Ahmed, S., Cabral, F.G., Freitas Paulo da Costa, B.: Stochastic Lipschitz dynamic programming. *Mathematical Programming* **191**(2), 755–793 (2022)
- [2] Atakan, S., Sen, S.: A progressive hedging based branch-and-bound algorithm for mixed-integer stochastic programs. *Computational Management Science* **15**(3), 501–540 (2018)
- [3] Averkov, G.: A proof of Lovász’s theorem on maximal lattice-free sets. *Beiträge zur Algebra und Geometrie/Contributions to Algebra and Geometry* **54**(1), 105–109 (2013)
- [4] Barnett, J., Watson, J.-P., Woodruff, D.L.: BBPH: Using progressive hedging within branch and bound to solve multi-stage stochastic mixed integer programs. *Operations Research Letters* **45**(1), 34–39 (2017)
- [5] Birge, J.R., Louveaux, F.: *Introduction to Stochastic Programming*. Springer, New York (1997)
- [6] Carøe, C.C., Schultz, R.: Dual decomposition in stochastic integer programming. *Operations Research Letters* **24**(1-2), 37–45 (1999)
- [7] Cook, W., Gerards, A.M., Schrijver, A., Tardos, É.: Sensitivity theorems in integer linear programming. *Mathematical Programming* **34**(3), 251–264 (1986)
- [8] Daryalal, M., Bodur, M., Luedtke, J.R.: Lagrangian dual decision rules for multi-stage stochastic mixed-integer programming. *Operations Research* **72**(2), 717–737 (2024)
- [9] Füllner, C., Rebennack, S.: Non-convex nested Benders decomposition. *Mathematical Programming* **196**(1), 987–1024 (2022)
- [10] Gale, D., Klee, V., Rockafellar, R.: Convex functions on convex polytopes. *Proceedings of the American Mathematical Society* **19**(4), 867–873 (1968)
- [11] Gomory, R.E.: Some polyhedra related to combinatorial problems. *Linear Algebra and its Applications* **2**(4), 451–558 (1969)
- [12] Kazemi Zanjani, M., Sanei Bajgiran, O., Noureldath, M.: A hybrid scenario cluster decomposition algorithm for supply chain tactical planning under uncertainty. *European Journal of Operational Research* **252**(2), 466–476 (2016)
- [13] Klein Haneveld, W.K., Stougie, L., van der Vlerk, M.H.: Simple integer recourse

- models: convexity and convex approximations. *Mathematical programming* **108**(2), 435–473 (2006)
- [14] Ljubić, I., Mutzel, P., Zey, B.: Stochastic survivable network design problems: Theory and practice. *European Journal of Operational Research* **256**(2), 333–348 (2017)
 - [15] Lozano, L., Smith, J.C.: A binary decision diagram based algorithm for solving a class of binary two-stage stochastic programs. *Mathematical Programming* **191**, 381–404 (2022)
 - [16] MacNeil, M., Bodur, M.: Leveraging decision diagrams to solve two-stage stochastic programs with binary recourse and logical linking constraints. *European Journal of Operational Research* **315**(1), 228–241 (2024)
 - [17] Pereira, M.V., Pinto, L.M.: Multi-stage stochastic optimization applied to energy planning. *Mathematical programming* **52**(1), 359–375 (1991)
 - [18] Pflug, G.C., Pichler, A.: From empirical observations to tree models for stochastic optimization: convergence properties. *SIAM Journal on Optimization* **26**(3), 1715–1740 (2016)
 - [19] Romeijnnders, W., Schultz, R., van der Vlerk, M.H., Klein Haneveld, W.K.: A convex approximation for two-stage mixed-integer recourse models with a uniform error bound. *SIAM Journal on Optimization* **26**(1), 426–447 (2016)
 - [20] Romeijnnders, W., van der Laan, N.: Benders decomposition with scaled cuts for multistage stochastic mixed-integer programs. *Optimization Online* (2024)
 - [21] Romeijnnders, W., van der Vlerk, M.H., Klein Haneveld, W.K.: Convex approximations for totally unimodular integer recourse models: a uniform error bound. *SIAM Journal on Optimization* **25**(1), 130–158 (2015)
 - [22] Romeijnnders, W., van der Vlerk, M.H., Klein Haneveld, W.K.: Total variation bounds on the expectation of periodic functions with applications to recourse approximations. *Mathematical Programming* **157**(1), 3–46 (2016)
 - [23] Romeijnnders, W., Zhang, Y., Sen, S.: Decomposition algorithms for computational stochastic mixed-integer programming: A survey. *European Journal of Operational Research* **335**(2), 335–353 (2026)
 - [24] Schrijver, A.: *Theory of Linear and Integer Programming*. John Wiley & Sons, Chichester, England (1986)
 - [25] Shapiro, A., Dentcheva, D., Ruszczyński, A.: *Lectures on Stochastic Programming: Modeling and Theory*, 3rd edn. Society for Industrial and Applied Mathematics, Philadelphia (2021)

- [26] Singh, K.J., Philpott, A.B., Wood, R.K.: Dantzig-Wolfe decomposition for solving multistage stochastic capacity-planning problems. *Operations Research* **57**(5), 1271–1286 (2009)
- [27] van Beesten, E.R., Romeijnders, W.: Convex approximations for two-stage mixed-integer mean-risk recourse models with conditional value-at-risk. *Mathematical Programming* **181**(2), 473–507 (2020)
- [28] van Beesten, E.R., Romeijnders, W.: Parametric error bounds for convex approximations of two-stage mixed-integer recourse models with a random second-stage cost vector. *Operations Research Letters* **50**(5), 541–547 (2022)
- [29] van Beesten, E.R., Romeijnders, W., Roodbergen, K.J.: Convex approximations of two-stage risk-averse mixed-integer recourse models. *Computational Optimization and Applications* **88**(1), 313–347 (2024)
- [30] van der Laan, N., Romeijnders, W.: A loose Benders decomposition algorithm for approximating two-stage mixed-integer recourse models. *Mathematical Programming* **190**(1), 761–794 (2021)
- [31] van der Vlerk, M.H.: Convex approximations for complete integer recourse models. *Mathematical Programming* **99**, 297–310 (2004)
- [32] van der Vlerk, M.H.: Convex approximations for a class of mixed-integer recourse models. *Annals of Operations Research* **177**(1), 139–150 (2010)
- [33] Walkup, D., Wets, R.: Lifting projections of convex polyhedra. *Pacific Journal of Mathematics* **28**(2), 465–475 (1969)
- [34] Wolsey, L.A.: The b-hull of an integer program. *Discrete Applied Mathematics* **3**(3), 193–201 (1981)
- [35] Zhong, Z., Fan, N., Wu, L.: Multistage stochastic optimization for mid-term integrated generation and maintenance scheduling of cascaded hydroelectric system with renewable energy uncertainty. *European Journal of Operational Research* **318**(1), 179–199 (2024)
- [36] Zou, J., Ahmed, S., Sun, X.A.: Stochastic dual dynamic integer programming. *Mathematical Programming* **175**, 461–502 (2019)