EC.1. Classical Lagrangian Cuts

For any $n \in \overline{\mathcal{N}}$, let $\mathfrak{Q}_n^{i+1}(\cdot)$ denote the current cut approximation for value function $Q_n(\cdot)$. Then, a Lagrangian cut can be generated by considering a special Lagrangian relaxation of the nodal subproblem (that is, subproblem (3) with $Q_m(\cdot), m \in \mathcal{C}(n)$, replaced by $\mathfrak{Q}_m^{i+1}(\cdot)$). More precisely, the copy constraints $z_n = x_{a(n)}^i$ are relaxed using a given vector of dual multipliers $\pi_n \in \mathbb{R}^{d_{a(n)}}$. This yields

$$\mathcal{L}_n^{i+1}(\pi_n) := \min_{x_n, y_n, z_n, (\theta_m)} \Big\{ f_n(x_n, y_n) + \sum_{m \in \mathcal{C}(n)} p_{nm}\theta_m - \pi_n^\top z_n \;:\; (z_n, x_n, y_n) \in \mathcal{F}_n, \\ z_n \in Z_{a(n)}, \theta_m \geq \mathfrak{Q}_m^{i+1}(\cdot)(x_n),\ m \in \mathcal{C}(n) \Big\}.$$

For varying π_n, this relaxation defines the *dual function* $\mathcal{L}_n^{i+1}(\cdot)$. The problem of optimizing the dual function over the dual multipliers π_n is the *Lagrangian dual problem*:

$$\max_{\pi_n} \Big\{ \mathcal{L}_n^{i+1}(\pi_n) + \pi_n^\top x_{a(n)}^i \Big\}. \tag{EC.1}$$

By solving problem (EC.1), a Lagrangian optimality cut for $Q_n(\cdot)$ can be derived as

$$\theta_n \geq \mathcal{L}_n^{i+1}(\pi_n^i) + (\pi_n^i)^\top x_{a(n)}, \tag{EC.2}$$

where π_n^i denotes feasible dual multipliers in (EC.1) for node n (Zou et al. 2019).

REMARK EC.1. Let $\pi_{n0}^i > 0$ in the new Lagrangian cut (10). As shown in Proposition 1 in Chen and Luedtke (2022), with division by π_{n0}^i, it follows $\theta_n \geq \mathcal{L}_n^{i+1}(\widehat{\pi}_n) - \widehat{\pi}_n^\top x_{a(n)}$ with $\widehat{\pi}_n := \frac{\pi_n^i}{\pi_{n0}^i}$ and $\mathcal{L}_n^{i+1}(\widehat{\pi}_n) := \mathscr{L}_n^{i+1}(\widehat{\pi}_n, 1)$. This is an equivalent representation of cut (10) in the form of a classical Lagrangian optimality cut (EC.2). Importantly, despite this scaling relation, our framework in general yields different cuts than the classical one as the choice of cut coefficients is based on a different dual problem. □

EC.2. Proofs

In this section, we present some auxiliary results and the proofs that are not displayed in the main text.

EC.2.1. Feasibility Problem 7

LEMMA EC.1. *Given a point $(x_{a(n)}^i, \theta_n^i)$, problem* (7) *is a feasibility problem for* $\mathrm{epi}(\underline{Q}_n^{i+1})$*, that is,*

$$v_n^{f,i+1}(x_{a(n)}^i, \theta_n^i) = \begin{cases} 0, & \textit{if } (x_{a(n)}^i, \theta_n^i) \in \mathrm{epi}(\underline{Q}_n^{i+1}) \\ +\infty, & \textit{else.} \end{cases}$$

Proof. Let $(x_{a(n)}^i, \theta_n^i) \in \mathrm{epi}(\underline{Q}_n^{i+1})$. Then according to (4) we have

$$\theta_n^i \geq \min_{\lambda_n, z_n} \Big\{ c_n^\top \lambda_n \;:\; (\lambda_n, z_n) \in \mathcal{W}_n^{i+1},\ z_n = x_{a(n)}^i \Big\}. \tag{EC.3}$$

This implies that there exists some (λ_n, z_n) such that for $(x_{a(n)}^i, \theta_n^i)$ all constraints of (7) are satisfied. Hence, $v_n^{f,i+1}(x_{a(n)}^i, \theta_n^i) = 0$. For the reverse direction, let $v_n^{f,i+1}(x_{a(n)}^i, \theta_n^i) = 0$. Then, there exist (λ_n, z_n) such that for $(x_{a(n)}^i, \theta_n^i)$ all constraints of (7) are satisfied. This implies (EC.3), and thus $(x_{a(n)}^i, \theta_n^i) \in \mathrm{epi}(\underline{Q}_n^{i+1})$. □

EC.2.2. Proof of Theorem 1

Proof. (i) is a standard result on Lagrangian relaxation (see Guignard 2003). Another well-known property of the Lagrangian dual is that it is equivalent to a primal convexification of the original subproblem (Geoffrion 1974). In our case, this convexification is given by

$$\min_{\lambda_n, z_n} \Big\{ 0 \, : \, (\lambda_n, z_n) \in \operatorname{conv}(\mathcal{W}_n^{i+1}), \; z_n = x_{a(n)}^i, \; \theta_n^i \geq c_n^\top \lambda_n \Big\}. \tag{EC.4}$$

The closed convex envelope $\overline{\operatorname{co}}(\underline{Q}_n^{i+1})(\cdot)$ can be expressed through the convex problem

$$\overline{\operatorname{co}}(\underline{Q}_n^{i+1})(x_{a(n)}^i) = \min_{\lambda_n, z_n} \Big\{ c_n^\top \lambda_n \, : \, (\lambda_n, z_n) \in \operatorname{conv}(\mathcal{W}_n^{i+1}), \; z_n = x_{a(n)}^i \Big\}, \tag{EC.5}$$

see for instance (Füllner et al. 2024, Theorems 3.8 and 3.9) for a formal proof. Therefore, by the same reasoning as in Lemma EC.1, problem (EC.4) is a feasibility problem for $\operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n^{i+1}))$. $\square$

EC.2.3. Proof of Lemma 2

Proof. This is proven in Chen and Luedtke (2022) for $\operatorname{epi}(Q_n)$, but we provide a customized proof here. Let $(x_{a(n)}, \theta_n) \in \operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n^{i+1}))$. Then,

$$\begin{aligned}
(\pi_n^i)^\top x_{a(n)} + \pi_{n0}^i \theta_n &\geq \min_{x_{a(n)}, \theta_n} \Big\{ (\pi_n^i)^\top x_{a(n)} + \pi_{n0}^i \theta_n \, : \, (x_{a(n)}, \theta_n) \in \operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n^{i+1})) \Big\} \\
&= \min_{\lambda_n, z_n} \Big\{ (\pi_n^i)^\top z_n + \pi_{n0}^i c_n^\top \lambda_n \, : \, (\lambda_n, z_n) \in \operatorname{conv}(\mathcal{W}_n^{i+1}) \Big\} \\
&= \min_{\lambda_n, z_n} \Big\{ (\pi_n^i)^\top z_n + \pi_{n0}^i c_n^\top \lambda_n \, : \, (\lambda_n, z_n) \in \mathcal{W}_n^{i+1} \Big\} \\
&= \mathscr{L}_n^{i+1}(\pi_n^i, \pi_{n0}^i).
\end{aligned}$$

The inequality follows by feasibility. The first equality uses the same relation that is also applied in (EC.4). The second equality exploits that the objective function is linear, and the last one follows from the definition of $\mathscr{L}_n^{i+1}(\cdot)$ in (8). The second part of the assertion follows with $\operatorname{epi}(Q_n) \subseteq \operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n^{i+1}))$. $\square$

EC.2.4. Proof of Lemma 3

Proof. According to Theorem 1, we have $\widehat{v}_n^{D,i+1}(x_{a(n)}^i, \theta_n^i) = 0$ for the non-normalized dual (9). By definition of (9) and its normalization (11), we conclude $\widehat{v}_n^{ND,i+1}(x_{a(n)}^i, \theta_n^i) \leq \widehat{v}_n^{D,i+1}(x_{a(n)}^i, \theta_n^i) = 0$.

Let $(\widehat{\pi}_n, \widehat{\pi}_{n0})$ be an optimal point for problem (9), *i.e.*, $\mathscr{L}_n^{i+1}(\widehat{\pi}_n, \widehat{\pi}_{n0}) - (\widehat{\pi}_n)^\top x_{a(n)}^i - \widehat{\pi}_{n0}\theta_n^i = 0$. If $\|\widehat{\pi}_n, \widehat{\pi}_{n0}\| \leq 1$, then it is also feasible for (11). As the objective of both problems is the same, $\widehat{v}_n^{ND,i+1}(x_{a(n)}^i, \theta_n^i) = 0$.

Otherwise, there exists $\mu > 0$ such that $\frac{1}{\mu}(\widehat{\pi}_n, \widehat{\pi}_{n0})$ is feasible for (11). By feasibility, it follows

$$\begin{aligned}
\widehat{v}_n^{ND,i+1}(x_{a(n)}^i, \theta_n^i) &\geq \mathscr{L}_n^{i+1}\left(\frac{1}{\mu}\widehat{\pi}_n, \frac{1}{\mu}\widehat{\pi}_{n0}\right) - \frac{1}{\mu}(\widehat{\pi}_n)^\top x_{a(n)}^i - \frac{1}{\mu}\widehat{\pi}_{n0}\theta_n^i \\
&= \frac{1}{\mu}\Big(\mathscr{L}_n^{i+1}(\widehat{\pi}_n, \widehat{\pi}_{n0}) - (\widehat{\pi}_n)^\top x_{a(n)}^i - \widehat{\pi}_{n0}\theta_n^i \Big) = 0,
\end{aligned} \tag{EC.6}$$

where we exploit that $\mathscr{L}_n^{i+1}(\cdot)$ is positive homogeneous. The reverse direction can be shown similarly. $\square$

EC.2.5. Proof of Lemma 4

Proof. According to Brandenberg and Stursberg (2021), $\mathcal{R}_n^{i+1}(x^i_{a(n)}, \theta^i_n)$ can be rewritten as

$$\mathcal{R}_n^{i+1}(x^i_{a(n)}, \theta^i_n) = \Big\{ (\gamma_n, \gamma_{n0}) \in \mathbb{R}^{d_{a(n)}} \times \mathbb{R} \,:\, \gamma_n^\top x^i_{a(n)} + \gamma_{n0}\theta^i_n - \operatorname{supp}_{\operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n^{i+1}))}(\gamma_n, \gamma_{n0}) \geq 1 \Big\}, \tag{EC.7}$$

where $\operatorname{supp}_{\operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n^{i+1}))}(\cdot)$ denotes the support function of $\operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n^{i+1}))$. It can be expressed as follows:

$$\begin{aligned}
&\operatorname{supp}_{\operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n^{i+1}))}(\gamma_n, \gamma_{n0}) \\
&= \max_{x_{a(n)}, \theta_n} \big\{ \gamma_n^\top x_{a(n)} + \gamma_{n0}\theta_n \,:\, (x_{a(n)}, \theta_n) \in \operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n^{i+1})) \big\} \\
&= \max_{x_{a(n)}, \theta_n, \lambda_n, z_n} \Big\{ \gamma_n^\top x_{a(n)} + \gamma_{n0}\theta_n \,:\, \tilde{A}_n \lambda_n + \tilde{B}_n z_n \geq \tilde{d}_n, z_n = x_{a(n)}, \theta_n - c_n^\top \lambda_n \geq 0 \Big\} \\
&= \min_{\mu_n, \pi_n, \pi_{n0}} \Big\{ \tilde{d}_n^\top \mu_n \,:\, \tilde{A}_n^\top \mu_n - c_n \pi_{n0} = 0, \tilde{B}_n^\top \mu_n - \pi_n = 0, \pi_n = \gamma_n, \pi_{n0} = \gamma_{n0}, \pi_{n0} \leq 0, \mu_n \leq 0 \Big\} \\
&= \min_{\mu_n} \Big\{ -\tilde{d}_n^\top \mu_n \,:\, -\tilde{A}_n^\top \mu_n - c_n \gamma_{n0} = 0, -\tilde{B}_n^\top \mu_n - \gamma_n = 0, \gamma_{n0} \leq 0, \mu_n \geq 0 \Big\}.
\end{aligned} \tag{EC.8}$$

The first equation applies the definition of support functions. The second one follows from Remark 4 and the third one exploits strong duality for LPs. We insert (EC.8) into (EC.7), and observe that the set remains unchanged if we replace the minimum operator using an existence quantor. □

EC.2.6. Proof of Lemma 8

Proof. By definition, the normalized Lagrangian dual problem is equivalent to

$$\begin{aligned}
&\max_{(\pi_n, \pi_{n0}) \in \Pi_n} \Big\{ \min_{(\lambda_n, z_n) \in \mathcal{W}_n^{i+1}} \Big\{ \pi_n^\top (z_n - x^i_{a(n)}) + \pi_{n0}(c_n^\top \lambda_n - \theta^i_n) \Big\} \Big\} \\
&= \max_{(\pi_n, \pi_{n0}) \in \Pi_n} \Big\{ \min_{(\lambda_n, z_n) \in \operatorname{conv}(\mathcal{W}_n^{i+1})} \Big\{ \pi_n^\top (z_n - x^i_{a(n)}) + \pi_{n0}(c_n^\top \lambda_n - \theta^i_n) \Big\} \Big\}
\end{aligned}$$

with $\Pi_n := \big\{ (\pi_n, \pi_{n0}) \in \mathbb{R}^{d_{a(n)}} \times \mathbb{R} \,:\, \pi_{n0} \geq 0, u_n^\top \pi_n + u_{n0}\pi_{n0} \leq 1 \big\}$. The equation follows from linearity. Using Remark 4 and then LP duality for the inner minimization problem, we obtain the equivalent problem

$$\begin{aligned}
\max_{\pi_n, \pi_{n0}, \mu_n} \Big\{ \tilde{d}_n^\top \mu_n - \pi_n^\top x^i_{a(n)} - \pi_{n0}\theta^i_n \,:\, &\mu_n \geq 0,\ \pi_{n0} \geq 0,\ u_n^\top \pi_n + u_{n0}\pi_{n0} \leq 1, \\
&\tilde{A}_n^\top \mu_n - \pi_{n0} c_n = 0,\ \tilde{B}_n^\top \mu_n - \pi_n = 0 \Big\}.
\end{aligned}$$

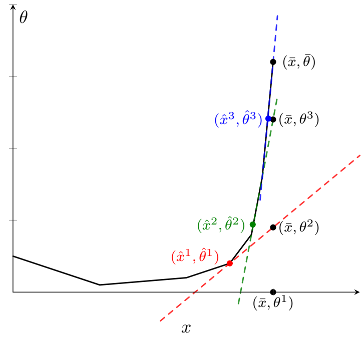
This is an LP. Using LP duality and Remark 4 again, the assertion follows. □

EC.2.7. Proof of Theorem 2

Proof. After reformulating the normalized Lagrangian dual problem (11) using Remark 4, and linking optimizing over the reverse polar set in (15) with optimizing over the so-called *relaxed alternative polyhedron* (see Theorem 2.1 and Corollary 2.2 in Brandenberg and Stursberg (2021)), we can conclude from Theorem 3.20 in Stursberg (2019) that the following result holds:

A point $(\widehat{\gamma}_n, \widehat{\gamma}_{n0})$ is optimal for problem (15) if and only if there exists some certificate $\widehat{\mu}_n \geq 0$ in $\mathcal{R}_n^{i+1}(x^i_{a(n)}, \theta^i_n)$ such that $(\mu_n^*, \pi_n^*, \pi_{n0}^*) = \alpha(-\widehat{\mu}_n, \widehat{\gamma}_n, \widehat{\gamma}_{n0})$ is optimal for problem (11), where $\alpha < 0$ is a

Figure EC.1 Illustration of deep cuts leading to tightness after finitely many steps.



Note. Sequence of deep cuts for fixed incumbent $\bar{x}$. The blue cut is tight at $(\bar{x}, \bar{\theta})$, thus equal to a classical Lagrangian cut.

negative scaling factor. Moreover, the optimal values of both problems multiply to -1. This result directly implies part (i) of the assertion.

We now prove (ii). Let $(\widehat{\gamma}_n, \widehat{\gamma}_{n0})$ be optimal for problem (15) with a valid certificate $\widehat{\mu}_n \geq 0$ in $\mathcal{R}_n^{i+1}(x_{a(n)}^i, \theta_n^i)$. Then, the valid cut $\tilde{d}_n^\top \widehat{\mu}_n + (\widehat{\gamma}_n)^\top x_{a(n)} + \widehat{\gamma}_{n0}\theta_n \leq 0$ is induced, separating $(x_{a(n)}^i, \theta_n^i)$ from $\operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n^{i+1}))$ (Cornuéjols and Lemaréchal 2006). According to the result that we used in part (i), this is equivalent to $-\frac{1}{\alpha}\tilde{d}_n^\top \mu_n^* + \frac{1}{\alpha}(\pi_n^*)^\top x_{a(n)} + \frac{1}{\alpha}\pi_{n0}^*\theta_n \leq 0$. However, given $\alpha < 0$, this is in turn equivalent to $\tilde{d}_n^\top \mu_n^* - (\pi_n^*)^\top x_{a(n)} - \pi_{n0}^*\theta_n \leq 0$ and by definition to the Lagrangian cut (10). The reverse direction follows in a similar way. $\square$

EC.2.8. Proof of Lemma 9

Proof. Suppose that condition (16) is satisfied. Then there exist some $(\tilde{x}_{a(n)}, \tilde{\theta}_n) \in \operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n^{i+1})) - (x_{a(n)}^i, \theta_n^i)$ and $\mu > 0$ such that $(u_n, u_{n0}) = \mu(\tilde{x}_{a(n)}, \tilde{\theta}_n)$. This implies $(\tilde{x}_{a(n)} + x_{a(n)}^i, \tilde{\theta}_n + \theta_n^i) \in \operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n^{i+1}))$. Therefore, the system $\Big\{(\lambda_n, z_n) \;:\; \tilde{\theta}_n + \theta_n^i \geq c_n^\top \lambda_n, (\lambda_n, z_n) \in \operatorname{conv}(\mathcal{W}_n^{i+1}), z_n = \tilde{x}_{a(n)} + x_{a(n)}^i\Big\}$ is non-empty. However, this set is equivalent to $\Big\{(\lambda_n, z_n) \;:\; \frac{1}{\mu}u_{n0} \geq c_n^\top \lambda_n - \theta_n^i, (\lambda_n, z_n) \in \operatorname{conv}(\mathcal{W}_n^{i+1}), z_n - x_{a(n)}^i = \frac{1}{\mu}u_n\Big\}$. With choosing $\eta_n = \frac{1}{\mu} > 0$ it immediately follows that problem (13) is feasible. By $\eta_n \geq 0$, its optimal value is also bounded from below, hence it is finite. By LP duality this implies that Assumption 3 is satisfied. $\square$

EC.2.9. Visualization of Convergence Idea

Fig. EC.1 visualizes the convergence idea for SDDiP using deep Lagrangian cuts from Sect. 4.

EC.2.10. Proof of Lemma 10

Proof. For any $\tilde{\nu} \in \mathbb{N}$, the incumbent $(\bar{x}_{a(n)}, \theta_n^{i_{\tilde{\nu}}})$ either satisfies $(\bar{x}_{a(n)}, \theta_n^{i_{\tilde{\nu}}}) \in \operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n^{i_{\tilde{\nu}}+1}))$, and by that $\theta_n^{i_{\tilde{\nu}}} = \bar{\theta}_n$, or it satisfies $\theta_n^{i_{\tilde{\nu}}} < \bar{\theta}_n$. In the first case, $(\bar{x}_{a(n)}, \theta_n^{i_{\tilde{\nu}}}) = (\bar{x}_{a(n)}, \bar{\theta}_n) = (\widehat{x}_{a(n)}^{i_{\tilde{\nu}}}, \widehat{\theta}_n^{i_{\tilde{\nu}}})$ and the assertion is trivially satisfied for all $\nu \geq \tilde{\nu}$. In the second case, the incumbent is separated from $\operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n^{i_{\tilde{\nu}}+1}))$ by a newly constructed Lagrangian cut. Therefore, the sequence $(\theta_n^{i_\nu})_{\nu \in \mathbb{N}}$ is monotonically increasing and bounded, thus converging. More precisely, by the above separation argument it converges to $\bar{\theta}_n$. By continuity of norms, it follows that the distance between the incumbent and $(\bar{x}_{a(n)}, \bar{\theta}_n)$ converges to zero: $\lim_{\nu\to\infty} \|\bar{x}_{a(n)} - x^{i_\nu}_{a(n)}, \bar{\theta}_n - \theta^{i_\nu}_n\|_* = 0$.

Since $(\bar{x}_{a(n)}, \bar{\theta}_n)$ is feasible for the projection problem (12), this first result also implies that any sequence $(\widehat{x}^{i_\nu}_{a(n)}, \widehat{\theta}^{i_\nu}_n)_{\nu\in\mathbb{N}}$ of (possibly non-unique) solutions of problem (12) satisfies the condition $\lim_{\nu\to\infty} \|\widehat{x}^{i_\nu}_{a(n)} - x^{i_\nu}_{a(n)}, \widehat{\theta}^{i_\nu}_n - \theta^{i_\nu}_n\|_* = 0$. By using the triangle inequality together with this and the first result we obtain $\lim_{\nu\to\infty} \|\widehat{x}^{i_\nu}_{a(n)} - \bar{x}_{a(n)}, \widehat{\theta}^{i_\nu}_n - \bar{\theta}_n\|_* = 0$. From the definition of norms, the assertion follows. □

EC.2.11. Proof of Lemma 12

Proof. We assume that ν is sufficiently large, that is, $\nu \geq \bar{\nu}$ from Lemma 11. If we have $(\widehat{x}^{i_\nu}_{a(n)}, \widehat{\theta}^{i_\nu}_n) = (\bar{x}_{a(n)}, \bar{\theta}_n)$ for all such ν, then by Corollary 1 the assertion follows immediately. Therefore, we assume that this is not true and distinguish three possible cases which are visualized in Fig. EC.2. Note that other cases cannot occur due to Lemma 10.

Case 1. Let $\overline{K} = \{\bar{k}\}$ for some $\bar{k} \in K$. Then, $(\bar{x}_{a(n)}, \bar{\theta}_n) \in \operatorname{int}(F_{\bar{k}})$. From non-emptiness and Lemma 11 it follows $\widehat{K}^\nu = \{\bar{k}\}$ for all ν, thus $(\widehat{x}^{i}_{a(n)}, \widehat{\theta}^{i_\nu}_n) \in \operatorname{int}(F_{\bar{k}})$ as well. According to Corollary 1, for each ν, the obtained cut supports $\operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n))$ at $(\widehat{x}^{i_\nu}_{a(n)}, \widehat{\theta}^{i_\nu}_n)$. The assertion then follows from Lemma 4.6 in Chapter D.4 of Füllner (2024), and the cut is facet-defining.

Case 2. Let $|\overline{K}| > 1$. This implies $(\bar{x}_{a(n)}, \bar{\theta}_n) \in \operatorname{bd}(F_k)$ for all $k \in \overline{K}$. For any ν there exist two possible sub-cases for the solution $(\widehat{x}^{i_\nu}_{a(n)}, \widehat{\theta}^{i_\nu}_n)$ to the projection problem (12).

(i) It satisfies $(\widehat{x}^{i_\nu}_{a(n)}, \widehat{\theta}^{i_\nu}_n) \in \operatorname{int}(F_{\bar{k}})$ for some $\bar{k} \in \overline{K}$. Then, the assertion follows again from Lemma 4.6 in Chapter D.4 of Füllner (2024).

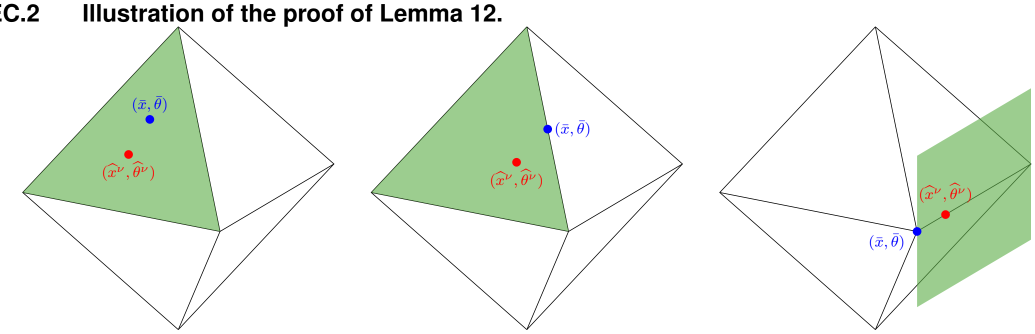
(ii) It satisfies $(\widehat{x}^{i_\nu}_{a(n)}, \widehat{\theta}^{i_\nu}_n) \in \operatorname{bd}(F_{k_\nu})$ for all $k^\nu \in \widehat{K}^\nu$, with $\widehat{K}^\nu \subseteq \overline{K}$ according to Lemma 11. Then, $(\widehat{x}^{i_\nu}_{a(n)}, \widehat{\theta}^{i_\nu}_n)$ and $(\bar{x}_{a(n)}, \bar{\theta}_n)$ are still located on a joint sub-facet (face) $\mathfrak{F}$ of $\operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n))$. Additionally, Lemma 11 states that $(\widehat{x}^{i_\nu}_{a(n)}, \widehat{\theta}^{i_\nu}_n)$ is not a vertex of $\operatorname{epi}(\overline{\operatorname{co}}(\underline{Q}_n))$. Therefore, we have $(\widehat{x}^{i_\nu}_{a(n)}, \widehat{\theta}^{i_\nu}_n) \in \operatorname{relint}(\mathfrak{F})$. Again, the assertion follows from Lemma 4.6 in Chapter D.4 of Füllner (2024). □

EC.3. Illustrative Example for Deep and LN Lagrangian Cuts

EXAMPLE EC.1. This example is inspired by illustrations in Brandenberg and Stursberg (2021) and Hosseini and Turner (2024). We consider a given epigraph $\operatorname{epi}(\overline{\operatorname{co}}(Q_n))$, an incumbent $(x^i_{a(n)}, \theta^i_n)$ and the associated reverse polar set $\mathcal{R}^{i+1}_n(x^i_{a(n)}, \theta^i_n)$.

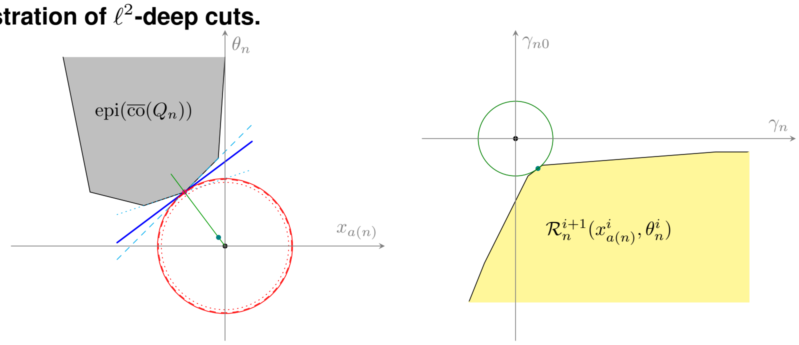
We first look at the obtained deep Lagrangian cuts for different norms (ℓ^2, ℓ^1, ℓ^∞ and a weighted ℓ^1-norm). The sets and cuts are illustrated in Fig. EC.3-EC.6. In each case, the illustration consists of two parts (a) and (b). In part (a), the incumbent (*black dot*) and the epigraph are depicted. Moreover, several norm balls are shown for the respective dual norms (*red lines*). We can see that the obtained deep cuts (*blue lines*) maximize the distance between the incumbent and the hyperplane in the dual norm. This is illustrated by

Figure EC.2 Illustration of the proof of Lemma 12.



Note. LEFT: Case 1. Both points contained in the interior of facet $F_{\bar{k}}$ (highlighted in green). The associated facet-defining cut is tight at $(\bar{x}, \bar{\theta})$. MIDDLE: Case 2 (i). Only $(\widehat{x}^{\nu}, \widehat{\theta}^{\nu})$ contained in the interior of facet $F_{\bar{k}}$ (highlighted in green), but that facet still contains $(\bar{x}, \bar{\theta})$, so the associated cut is tight. RIGHT: Case 2 (ii). Both points located on the boundary of facets but on a joint face. Any cut supporting $(\widehat{x}^{\nu}, \widehat{\theta}^{\nu})$ is also tight at $(\bar{x}, \bar{\theta})$ (with an example highlighted in green).

Figure EC.3 Illustration of ℓ^2-deep cuts.

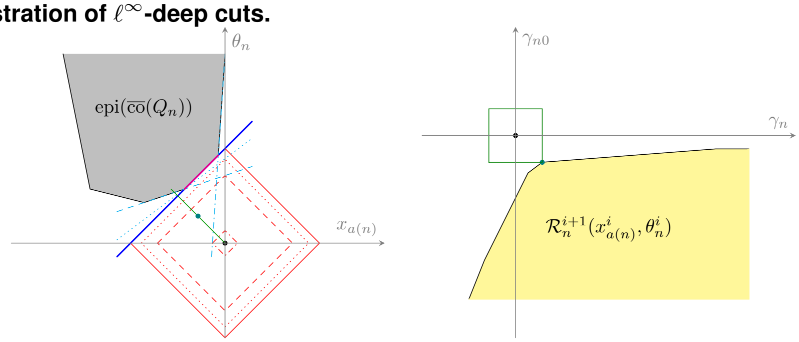


Note. LEFT: Epigraph and deep cut. RIGHT: Reverse polar set.

depicting different valid cuts (*dashed/dotted cyan lines*) with smaller distances. On the other hand, it is also shown that deep cuts minimize the distances between the incumbent and the epigraph in the dual norm and support the epigraphs at the corresponding projection to the epigraph (*violet line or point*). In part (b), the reverse polar set is depicted. Moreover, the optimal solutions (*teal line or point*) for optimizing the given norm (illustrated by a norm ball, *green line*) over this set are highlighted. These solutions (apart from sign changes) characterize the normal vectors of the obtained cuts (see Lemma 7), as is additionally illustrated in part (a). Note that for none of the considered cases, the deep cuts are tight at $\big(x_{a(n)}^i, \overline{\mathrm{co}}(Q_n)(x_{a(n)}^i)\big)$, contrary to classical Lagrangian cuts.

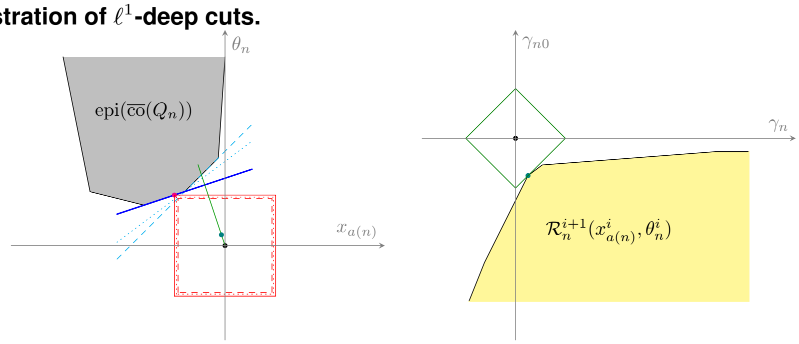
We next consider an exemplary LN cut. It is illustrated in Fig. EC.7. In part (a) the geometric idea of projection along a line segment is highlighted. The direction of this line segment is (u_n, u_{n0}) and obtained as the difference between a known core point (*yellow dot*) in $\mathrm{epi}(\overline{\mathrm{co}}(Q_n))$ and $(x_{a(n)}^i, \theta_n^i)$. In part (b) we can see that (apart from sign changes) the cut normal to the LN cut can be determined by maximizing the linear function $u_n^\top \gamma_n + u_{n0}\gamma_{n0}$ over $\mathcal{R}_n^{i+1}(x_{a(n)}^i, \theta_n^i)$. As the solution is an extreme point of $\mathcal{R}_n^{i+1}(x_{a(n)}^i, \theta_n^i)$ (*green dot*), the corresponding LN cut is facet-defining.

Figure EC.4 Illustration of ℓ^∞-deep cuts.



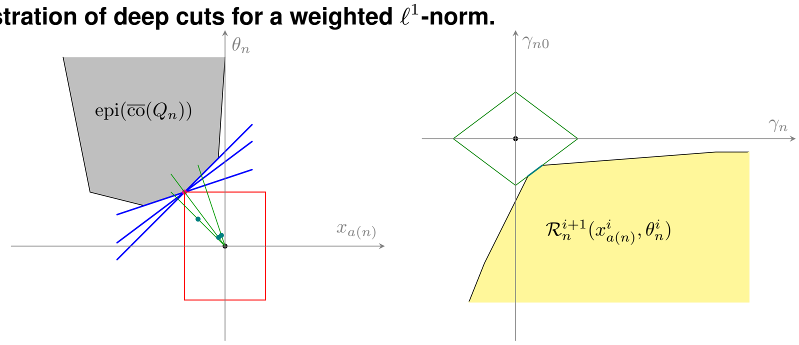
Note. LEFT: Epigraph and deep cut. RIGHT: Reverse polar set.

Figure EC.5 Illustration of ℓ^1-deep cuts.



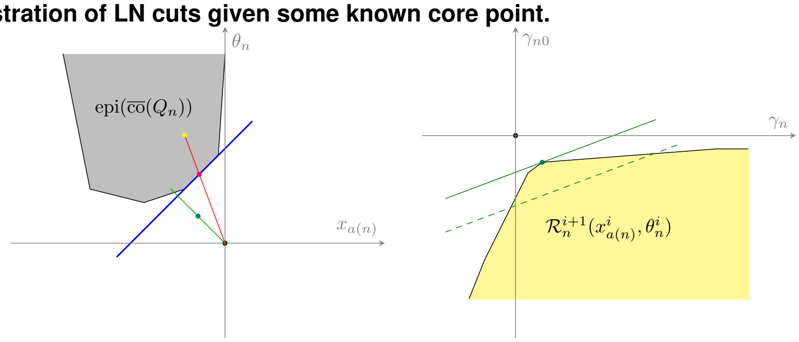
Note. LEFT: Epigraph and deep cut. RIGHT: Reverse polar set.

Figure EC.6 Illustration of deep cuts for a weighted ℓ^1-norm.



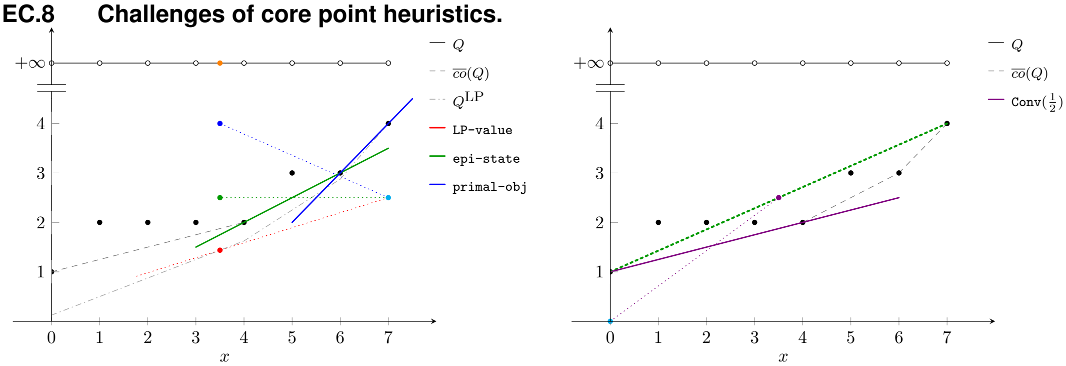
Note. LEFT: Epigraph and deep cut. RIGHT: Reverse polar set.

Figure EC.7 Illustration of LN cuts given some known core point.



Note. LEFT: Epigraph and LN cut. RIGHT: Reverse polar set.

Figure EC.8 Challenges of core point heuristics.



Note. LEFT: Example with integer state space $X = \{0, 1, 2, \dots, 7\}$ and incumbent $(x, \theta) = (7, \frac{5}{2})$. For `Mid`, due to infeasibility (see orange dot), no core point can be detected directly. However, fallback strategies `epi-state` and `primal-obj` yield a correct core point, whereas `LP-value` does not. RIGHT: Same example with incumbent $(x, \theta) = (0, 0)$. No fallback strategy yields a correct core point for `Mid`. However, `Conv` with $\alpha = \frac{1}{2}$ does.

EC.4. Challenges of Core Point Heuristics

Fig. EC.8 illustrates the challenges when identifying core points for MS-MILPs, as discussed in Sect. 3.6.

EC.5. Detecting Unboundedness for LN Cuts

Problem (13) provides a natural way to check for unboundedness, or the validity of direction (u_n, u_{n0}), respectively. However, it cannot be solved immediately, as we do not know $\operatorname{conv}(\mathcal{W}_n^{i+1})$ explicitly. We may instead solve the approximation

$$\min_{\lambda_n, z_n, \eta_n} \left\{ \eta_n \,:\, (\lambda_n, z_n) \in \mathcal{W}_n^{i+1},\ \eta_n \geq 0,\ u_{n0}\eta_n \geq c_n^\top \lambda_n - \theta_n^i,\ u_n \eta_n = z_n - x_{a(n)}^i \right\} \tag{EC.9}$$

using the known set $\mathcal{W}_n^{i+1}$ instead of $\operatorname{conv}(\mathcal{W}_n^{i+1})$. If the dual problem (11) is unbounded, then this problem is infeasible. Therefore, we can use infeasibility of (EC.9) as a proxy for unboundedness and to construct mitigation strategies between steps 7 and 8 of Algorithm 1. This is presented in Algorithm 2.

Unfortunately, in the presence of integrality constraints, infeasibility of (EC.9) may occur very often, even given an appropriate direction (u_n, u_{n0}), thus leading to a very conservative strategy of taking countermeasures. On the other hand, using the LP relaxation of (EC.9) is not sufficient to rule out all cases of unboundedness. For an effective implementation of LN cuts this remains a challenge.

EC.6. Kelley's Cutting-plane Method

Algorithm 3 states Kelley's cutting-plane method (Kelley 1960) for solving the dual problem (11).

EC.7. Additional Computational Results

We provide some additional results for our computational experiments.

- Table EC.1 complements Table 4 and provides results for `CLSP-Bin` and $T = 10$.
- Fig. EC.9 illustrates the iteration times for SDDiP, as well as the number of level Bundle method iterations required to solve the dual problems for each iteration of SDDiP.

Algorithm 2 Handling potential unboundedness for LN cuts

Require: Finished step 7 of Algorithm 1. Coefficients (u_m, u_{m0}) for node m and incumbent (x_n^i, θ_m^i).

Try to solve problem (EC.9) for m, (x_n^i, θ_m^i) and (u_m, u_{m0}).
if feasible **then**
    Problem (13), and thus the normalized dual (11) must be feasible as well. Proceed with step 8 of Algorithm 1.
else
    Potential infeasibility of problem (13), and thus unboundedness of the normalized dual (11) detected. Construct a (strengthened) Benders cut and proceed with step 11 of Algorithm 1.

Algorithm 3 Kelley's cutting-plane method to solve the dual problem (11)

Require: Iteration i, node n with incumbent $(x_{a(n)}^i, \theta_n^i)$. Initial solution $(\bar{\pi}_n^0, \bar{\pi}_{n0}^0)$.

Initialize $\text{LB} = -\infty, \text{UB} = +\infty$ and $k = 0$.
while $\text{UB} - \text{LB} \geq \varepsilon$ (and $\text{UB} > 0$) **do**
    Solve problem (8) for $(\bar{\pi}_n^k, \bar{\pi}_{n0}^k)$. Store the optimal solution $(\bar{\lambda}_n^k, \bar{z}_n^k)$.
    if $\text{LB} < (\bar{\pi}_n^k)^\top (\bar{z}_n^k - x_{a(n)}^i) + \bar{\pi}_{n0}^k (c_n^\top \bar{\lambda}_n^k - \theta_n^i)$ **then**
        Set $\text{LB} = (\bar{\pi}_n^k)^\top (\bar{z}_n^k - x_{a(n)}^i) + \bar{\pi}_{n0}^k (c_n^\top \bar{\lambda}_n^k - \theta_n^i)$ and $(\pi_n^*, \pi_{n0}^*) = (\bar{\pi}_n^k, \bar{\pi}_{n0}^k)$.
    Solve approximating problem

$$\max_{\pi_n, \pi_{n0}} \left\{ \nu_n : \nu_n \leq \pi_n^\top (\bar{z}_n^r - x_{a(n)}^i) + \pi_{n0} (c_n^\top \bar{\lambda}_n^r - \theta_n^i),\ r = 0, \ldots, k,\ g_n(\pi_n, \pi_{n0}) \leq 1,\ \pi_{n0} \geq 0 \right\}. \tag{EC.10}$$

    Store the solution $(\bar{\pi}_n^{k+1}, \bar{\pi}_{n0}^{k+1})$.
    Increase k by one.

Ensure: Optimal solution $(\pi_n^i, \pi_{n0}^i) = (\pi_n^*, \pi_{n0}^*)$ and $\mathscr{L}_n^{i+1}(\pi_n^i, \pi_{n0}^i)$.

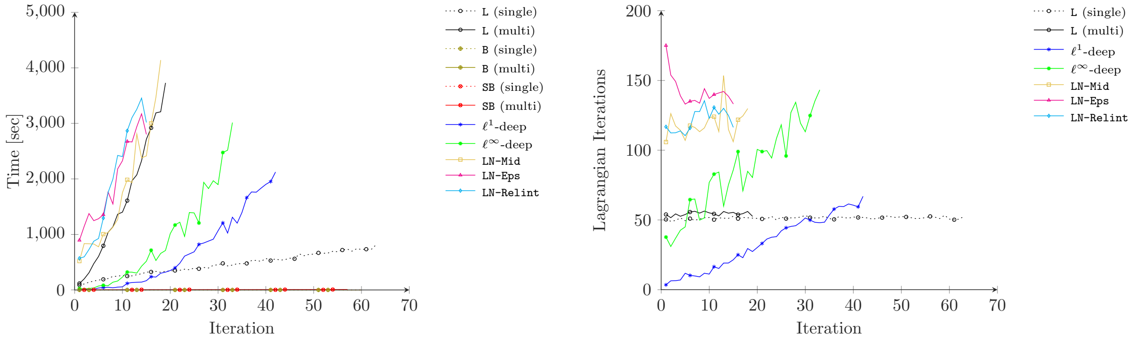
Table EC.1 SDDiP results for `CLSP-Bin`.

Stages	Method	Best LB	Stat. UB	Gap [%]	# Iter.	Time [s]	Time/It [s]	Avg. Lag-It	Time/Lag [s]	Cuts/Stage	Unb [%]
10	`B` (single)	2,165.3	5,120.8	58	136	26	0	-	-	73	-
	`B` (multi)	2,183.6	5,168.8	58	334	197	0	-	-	1,155	-
	`SB` (single)	3,377.1	4,949.8	32	89	122	1	-	-	22	-
	`SB` (multi)	3,472.9	**4,942.9**	30	206	1,066	5	-	-	746	-
	`L` (single)	616.4	5,466.6	89	69	lim	263	51	1.4	69	-
	`L` (multi)	682.4	5,443.1	88	20	lim	936	54	5.1	400	-
	ℓ^1-deep	3,782.5	-	-	38	lim	507	47	2.8	732	-
	ℓ^∞-deep	3,473.8	5,267.4	34	30	lim	648	112	3.6	581	-
	`LN-Mid`	**4,188.0**	5,163.4	**19**	18	lim	1,066	112	6.4	346	0
	`LN-Eps`	3,910.9	5,329.2	27	16	lim	1,150	139	6.4	309	0
	`LN-Relint`	4,121.1	5,185.6	21	16	lim	1,177	119	5.8	312	0
	`SB` + `L` (single)	3,475.5	**5,003.9**	31	132	lim	136	50	0.9	261	-
	`SB` + `L` (multi)	3,483.0	5,014.4	31	51	lim	360	48	3.0	1,889	-
	`SB` + ℓ^1-deep	3,804.1	5,161.6	26	35	lim	562	88	7.1	670	-
	`SB` + ℓ^∞-deep	3,856.5	5,290.3	27	25	lim	906	149	25.0	383	-
	`SB` + `LN-Mid`	**4,121.9**	5,121.8	**20**	31	lim	623	115	9.5	553	0
	`SB` + `LN-Eps`	4,091.1	5,092.3	**20**	31	lim	660	139	10.1	555	0
	`SB` + `LN-Relint`	4,041.7	5,157.5	22	31	lim	639	115	9.5	555	0

Time limits: $T = 10$: 5 hours (18,000 s). When the time limit is reached, this is indicated by 'lim'. 'Det. Equiv' refers to solving the deterministic equivalent of the MS-MILP.

- Table EC.2 shows average properties of all cuts obtained in the first 5 or 10 iterations of the experiments. For instance, it illustrates that for `CFLP`, ℓ^1-deep cuts are completely flat, whereas LN cuts obtained using `Eps` are extremely steep. It also illustrates that SB cuts are tight for `CFLP`.

Figure EC.9 Time requirements for experiments on `CLSP-Bin` and $T = 16$.



Note. LEFT: Time per iteration of SDDiP. RIGHT: Iterations required to solve Lagrangian dual over iterations of SDDiP.

Table EC.2 Cut properties for first iterations of computational experiments.

	`CLSP-Bin`		`CLSP`		`CFLP`	
Approach	Rel. Tightness	Cut Norm	Rel. Tightness	Cut Norm	Rel. Tightness	Cut Norm
`SB` (multi)	90.1	24,773	88.5	82	100.0	299
`L` (multi)	100.0	41,583	95.8	2,158	99.5	37
ℓ^1-deep	30.1	1,749	48.2	280	99.0	1
`Mid`	66.6	12,491	79.4	54	99.5	896
`Eps`	97.8	31,357	97.5	104	100.0	1,886
`Conv`(75)	-	-	-	-	100.0	1,459

Rel. Tightness: Relative tightness cutvalue$(x^i_{a(n)})/\underline{Q}^{i+1}_n(x^i_{a(n)})$ in %. *Cut Norm*: $\|\pi_n/\pi_{n0}\|_1$. A large cut norm indicates "steep" cuts, whereas a norm close to zero indicates very flat cuts. For `CLSP-Bin` and `CFLP` averages of all cuts generated in the first 5 iterations are presented, for `CLSP` averages over the first 10 iterations.

- Table EC.3 presents the values that *one specific* (*i.e.*, the first generated) cut in SDDiP takes at different points in the state space for problem `CFLP` and selected cut generation approaches. It illustrates that some cuts are tight at $x^i_{a(n)}$ (`SB`, `Eps`; note that `L` (multi) is only not tight because of the optimality tolerance with which the dual problem is solved), that some cuts are flat (ℓ^1`-deep`), that some cuts are steep, leading to bad approximations at parts of the state space (`Eps`), and that some provide a good approximation quality at all considered points of the state space (`Conv`(75)).

Table EC.3 Cuts generated for `CFLP` for first iteration, first scenario and last stage.

Approach	Cut value θ^i_n at $x_{a(n)} =$ $x^i_{a(n)}$	$\mathbf{0}$	$\mathbf{1}$
`SB` (multi)	**807.81**	807.81	507.81
`L` (multi)	801.17	801.17	636.27
ℓ^1-deep	670.75	670.75	**670.75**
`Mid`	793.74	1243.65	659.26
`Eps`	**807.81**	**1257.81**	25.77
`Conv`(75)	805.32	1253.88	653.40
$\underline{Q}^i_n(x_{a(n)})$	807.81	1257.81	670.75

$\mathbf{0}$ and $\mathbf{1}$ denote vectors of zeros and ones. All values scaled by 10^{-3}.