Normalization of the ReLU Dual for Cut Generation in Stochastic Mixed-Integer Programs

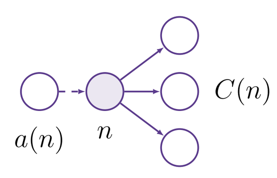
Akul Bansal   Simge Küçükyavuz

Department of Industrial Engineering & Management Sciences, Northwestern University

Introduction

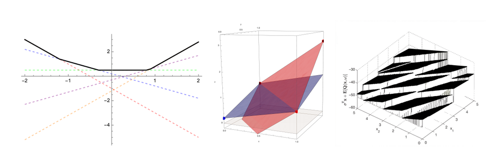
- A **multistage stochastic integer program** (MSIP) optimizes decisions over a stochastic process represented by a scenario tree. The subproblem at node n is

$$Q_n(x_{a(n)}) = \min_{x_n, y_n} f_n(x_n, y_n) + \sum_{m \in C(n)} q_{nm}\, Q_m(x_n),$$
$$(x_n, y_n) \in H_n\left(x_{a(n)}\right) \cap (X_n \times Y_n),$$



where **mixed-integer** restrictions on X_n, Y_n make Q_n **non-convex and discontinuous**.

- To **decompose** the problem, we replace each child cost $Q_m(x_n)$ with an approximation θ_m and iteratively add **cuts** of the form $\theta_m \geq g(x_n)$ to refine the approximation.
- **Existing methods** restrict state variables to be *purely binary* or *purely integer*, or rely on classical *Benders* and *Lagrangian* cuts that do not ensure convergence with mixed-integer state variables.



- Deng and Xie 2024 introduce **Rectified Linear Unit (ReLU) cuts**, which guarantee convergence with mixed-integer states.
- ReLU cuts arise from the **ReLU dual**, which admits **multiple optimal solutions**: every solution yields a cut violating the incumbent, but these cuts give different approximations of the future cost.
- **Goal**: generate **strong** ReLU cuts to solve MSIPs with mixed-integer state variables.

Contributions

- **Normalize the ReLU dual** (extending Fischetti et al. 2010; Füllner et al. 2024) to obtain strong cuts with asymptotic convergence for general MSIPs.
- Prove that with the right choice of normalization constraint, the resulting cuts are strong, specifically: (i) **tight** at the incumbent solution, and (ii) **Pareto-optimal** in the original state space.
- Show **regularization ⊆ normalization**: every regularization-based cut (Magnanti and Wong 1981; Deng and Xie 2024; Yang and Yang 2025) is recoverable from the normalized dual, but not conversely.

ReLU dual and ReLU cuts

- Replace the **copy constraint** $z_n = \hat{x}_{a(n)}$ in the subproblem with ReLU constraints:

$$Q_n(\hat{x}_{a(n)}) = \min_{z_n \in X_{a(n)}} Q_n(z_n),$$
$$(z_{nk} - \hat{x}_{a(n),k})^+ = 0,\ (z_{nk} - \hat{x}_{a(n),k})^- = 0,\ \forall k \in [d_{a(n)}],$$

where $(a)^+ := \max\{a, 0\}$ and $(a)^- := \max\{-a, 0\}$ are the positive and negative parts.

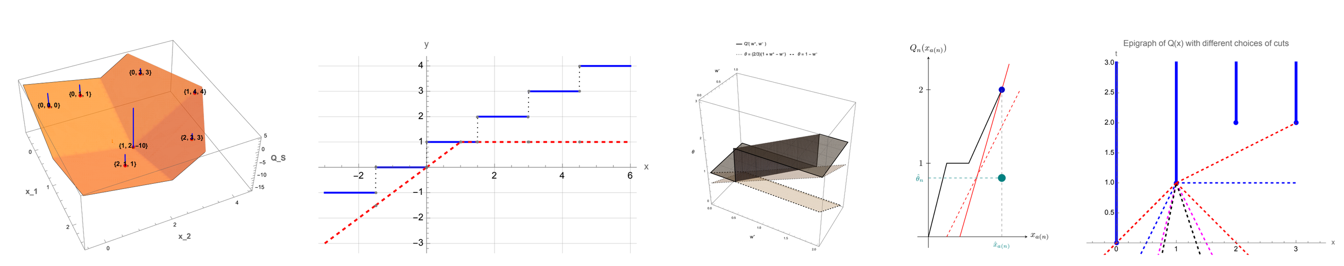
- Relaxing them with multipliers $\pi_n^+, \pi_n^- \in \mathbb{R}^{d_{a(n)}}$ and $Z_{a(n)} \subseteq X_{a(n)}$ gives the **ReLU dual**

$$\max_{\pi_n^+, \pi_n^-} \mathcal{L}_n^R(\pi_n^+, \pi_n^-; \hat{x}_{a(n)}) := \max_{\pi_n^+, \pi_n^-} \min_{z_n \in Z_{a(n)}} Q_n(z_n) + \sum_k \pi_{nk}^+ (z_{nk} - \hat{x}_{a(n),k})^+ + \sum_k \pi_{nk}^- (z_{nk} - \hat{x}_{a(n),k})^-.$$

- A valid **ReLU cut**:

$$\theta_n \geq \mathcal{L}_n^R(\pi_n^+, \pi_n^-; \hat{x}_{a(n)}) - \sum_k \pi_{nk}^+ (z_{nk} - \hat{x}_{a(n),k})^+ - \sum_k \pi_{nk}^- (z_{nk} - \hat{x}_{a(n),k})^-.$$

- A ReLU cut is a **Lagrangian cut in the lifted space** $Z_{\hat{x}}^L$ with variables $w_n^+, w_n^- \in \mathbb{R}_+^{d_{a(n)}}$ encoding the positive and negative parts $(z_n - \hat{x}_{a(n)})^{\pm}$, lifted cost $Q_n'(w_n^+, w_n^-; \hat{x}_{a(n)}) := Q_n(\hat{x}_{a(n)} + w_n^+ - w_n^-)$, and incumbent $(\mathbf{0}, \mathbf{0})$.
- **Strong duality** holds; an optimal ReLU dual solution gives a cut tight at $\hat{x}_{a(n)}$ even with mixed-integer states.
- The ReLU dual is **degenerate**: many (π_n^+, π_n^-) are optimal, each yielding a ReLU cut tight at $\hat{x}_{a(n)}$ but with very different overall approximation of the cost function.



Examples of cost-to-go function $Q_n(x_{a(n)})$ and ReLU cuts.

Normalized ReLU dual

- For incumbent $(\hat{x}_{a(n)}, \hat{\theta}_n)$ with $\hat{\theta}_n < Q_n(\hat{x}_{a(n)})$, consider the **feasibility version** of the subproblem:

$$v_n(\hat{x}_{a(n)}, \hat{\theta}_n) := \min_{z_n \in Z_{a(n)}} 0,$$
$$Q_n(z_n) \leq \hat{\theta}_n,$$
$$(z_{nk} - \hat{x}_{a(n),k})^+ = 0,\ (z_{nk} - \hat{x}_{a(n),k})^- = 0,\ \forall k \in [d_{a(n)}].$$

- Dualizing the epigraph constraint with $\pi_{n0} \geq 0$ and the copy constraints with $\pi_n^+, \pi_n^- \in \mathbb{R}^{d_{a(n)}}$ (and adding a **normalization** on dual variables with **coefficients u_n**) gives the **normalized dual**

$$v_n^{ND} := \max_{\pi_n^+, \pi_n^-, \pi_{n0} \geq 0} \mathcal{L}_n(\pi_n^+, \pi_n^-, \pi_{n0}; \hat{x}_{a(n)}) - \pi_{n0}\, \hat{\theta}_n,$$
$$u_n^+ \pi_n^+ + u_n^- \pi_n^- + u_{n0}\, \pi_{n0} \leq 1.$$

with Lagrangian relaxation given by:

$$\mathcal{L}_n(\pi_n^+, \pi_n^-, \pi_{n0}; \hat{x}_{a(n)}) = \min_{z_n \in Z_{a(n)}} \left\{ \pi_{n0}\, Q_n(z_n) + \sum_k \pi_{nk}^+ (z_{nk} - \hat{x}_{a(n),k})^+ + \sum_k \pi_{nk}^- (z_{nk} - \hat{x}_{a(n),k})^- \right\}.$$

- The choice of coefficients u_n^+, u_n^-, u_{n0} in the normalization constraint is important for preventing unboundedness of the dual and selecting a ray that yields a valid normalized ReLU cut:

$$\pi_{n0}\, \theta_n \geq \mathcal{L}_n(\pi_n^+, \pi_n^-, \pi_{n0}; \hat{x}_{a(n)}) - \sum_k \pi_{nk}^+ (z_{nk} - \hat{x}_{a(n),k})^+ - \sum_k \pi_{nk}^- (z_{nk} - \hat{x}_{a(n),k})^-.$$

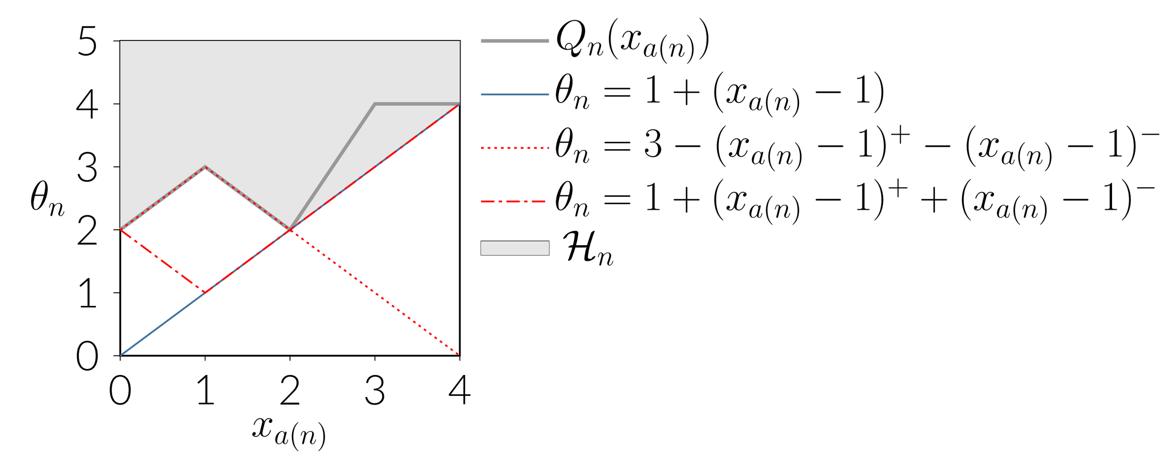
- Normalized ReLU cuts preserve the separation property needed for (asymptotic) convergence.

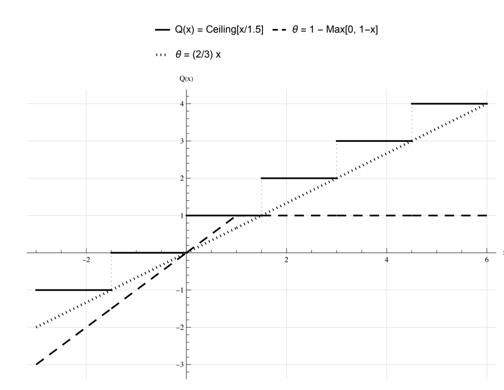
Proposition 1 (Cut existence). If $\hat{\theta}_n \geq Q_n(\hat{x}_{a(n)})$, then $v_n^{ND} = 0$. Otherwise, for any choice of normalization coefficients, there exists a feasible solution to the normalized ReLU dual, yielding a cut which violates the current incumbent $(\hat{x}_{a(n)}, \hat{\theta}_n)$.

Cut strength

- A cut is **Pareto-optimal** on a reference set $\mathcal{H}_n$ if no other valid cut dominates it on $\mathcal{H}_n$. We take $\mathcal{H}_n$ to be the **epigraph of the closed convex envelope** of the lifted cost-to-go Q_n' over $\mathrm{conv}(Z_{\hat{x}}^L)$, the tightest convex relaxation of the value function in the lifted space.

Proposition 2 (Pareto-optimal cut). For any **core point** $(u_n^+, u_n^-, u_{n0}) \in \mathrm{relint}\big(\mathrm{epi}(\overline{\mathrm{co}}(Q_n')) - (\mathbf{0}, \mathbf{0}, \hat{\theta}_n)\big)$, any optimal normalized dual solution with $\pi_{n0} > 0$ defines a **Pareto-optimal** cut on the reference set $\mathcal{H}_n$.





Proposition 3 (Tight cuts). There exists an ϵ-ball $B_\epsilon(\mathbf{0}, \mathbf{0})$ in $\mathrm{conv}(Z_{\hat{x}}^L)$ such that whenever $(u_n^+, u_n^-) \in B_\epsilon(\mathbf{0}, \mathbf{0}) \cap \mathrm{relint}(\mathrm{conv}(Z_{\hat{x}}^L))$ and $u_{n0} \geq Q_n'(u_n^+, u_n^-; \hat{x}_{a(n)}) - \hat{\theta}_n$, the optimal cut is **tight** at $\hat{x}_{a(n)}$ *and* **Pareto-optimal** on $\mathcal{H}_n$.

Normalization vs. regularization

- Let $\Pi_n(\hat{x}_{a(n)})$ be the set of all optimal solutions of the ReLU dual. Given a core point $(\tilde{u}_n^+, \tilde{u}_n^-) \in \mathrm{relint}(\mathrm{conv}(Z_{\hat{x}}^L))$, the **regularized dual** of Yang and Yang (2025)

$$\max_{\pi_n^+, \pi_n^- \in \Pi_n(\hat{x}_{a(n)})} \mathcal{L}_n^R(\pi_n^+, \pi_n^-; \hat{x}_{a(n)}) - (\pi_n^+, \pi_n^-) \begin{pmatrix} \tilde{u}_n^+ \\ \tilde{u}_n^- \end{pmatrix}$$

always yields a tight *and* Pareto-optimal cut.

- The optimal set $\Pi_n(\hat{x}_{a(n)})$ is approximated by the constraint

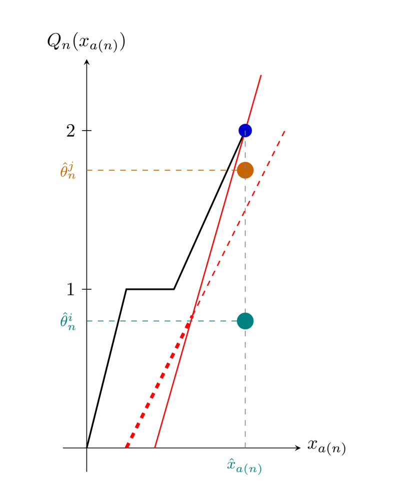
$$\mathcal{L}_n^R(\pi_n^+, \pi_n^-; \hat{x}_{a(n)}) \ \geq\ Q_n(\hat{x}_{a(n)}) - \epsilon, \qquad \epsilon > 0 \text{ small.}$$

- Normalization gives **flexibility**: cuts need not be tight at the incumbent, and Pareto-optimal cuts that are *not* tight at the incumbent **cannot be obtained via regularization**.

Proposition 4 (Regularization ⊆ Normalization). Every cut produced by regularization (up to scaling) is also optimal for the normalized dual, but the converse is false.

Alternating cut criterion

- Introduced by Angulo et al. 2016 to alternate between Benders and integer L-shaped cuts for two-stage stochastic programs.
- **Alternate** between cheap Benders' cuts (from the LP relaxation) and ReLU cuts: only solve the ReLU dual when the Benders' cut fails to cut off the incumbent.
- Benders' cuts provide computationally cheap and **strong approximations** of the cost-to-go across the domain, mitigating the weakness of Lagrangian-type cuts away from the incumbent.



Computational experiments

Two-stage **capacitated lot-sizing problem (CLSP)**; P products, N scenarios. **Regularization** (Yang and Yang 2025) vs. **Normalization** (this work); 1 h time limit (T), stopping gap 0.1%. Criterion `Def' stands for default (only ReLU cuts), `Alt' stands for alternating with Benders' cuts.

			Regularization					Normalization				
P	N	Crit.	Iter	Time	Gap(%)	D-Iter	Prop.	Iter	Time	Gap(%)	D-Iter	Prop.
5	10	Def	35	79	**0.085**	35	--	29	43	0.095	21	--
		Alt	46	77	**0.052**	29	0.90	35	57	0.080	18	0.90
5	100	Def	33	667	0.090	36	--	20	164	**0.081**	17	--
		Alt	38	593	0.093	27	0.88	26	192	**0.071**	14	0.87
10	10	Def	80	T	3.677	94	--	65	T	**1.802**	57	--
		Alt	105	T	3.750	73	0.86	71	T	**1.723**	50	0.90
20	10	Def	56	T	8.930	99	--	50	T	**2.419**	98	--
		Alt	116	T	8.111	47	0.58	68	T	**1.908**	72	0.77
20	100	Def	27	T	11.916	92	--	20	T	**6.351**	80	--
		Alt	75	T	12.448	23	0.46	48	T	**3.805**	42	0.63

- **Normalization needs fewer iterations and obtains tighter gaps** than regularization, in both settings.
- Normalization also converges in **fewer level-bundle iterations** (D-Iter) per dual solve.
- Alternating with Benders' cuts further **narrows the gap** on the hardest instances.
- The preprint (Bansal and Küçükyavuz 2026) also reports results on the **dynamic capacity allocation problem (DCAP)**; extension to **multistage** problems is ongoing work in revision.

References

G. Angulo, S. Ahmed, and S. S. Dey. Improving the integer L-shaped method. *INFORMS Journal on Computing*, 28(3):483--499, 2016.

A. Bansal and S. Küçükyavuz. Normalization of ReLU dual for cut generation in stochastic mixed-integer programs. *arXiv preprint arXiv:2602.05974*, 2026.

H. Deng and W. Xie. On the ReLU lagrangian cuts for stochastic mixed integer programming. *Optimization Online*, 2024.

M. Fischetti, D. Salvagnin, and A. Zanette. A note on the selection of Benders' cuts. *Mathematical Programming*, 124(1):175--182, 2010.

C. Füllner, X. A. Sun, and S. Rebennack. A new framework to generate Lagrangian cuts in multistage stochastic mixed-integer programming. *Optimization Online*, 2024.

T. L. Magnanti and R. T. Wong. Accelerating Benders decomposition: Algorithmic enhancement and model selection criteria. *Operations Research*, 29(3):464--484, 1981.

H. Yang and H. Yang. Globally converging algorithm for multistage stochastic mixed-integer programs via enhanced Lagrangian cuts. *Optimization Online*, 2025.