Decomposition Methods for Global Solutions of MILPs

Kaizhao Sun [1] Mou Sun [2] Dr. Wotao Yin [2]

[1]School of ISyE, Georgia Tech [2]Decision Intelligence Lab, DAMO Academy, Alibaba Group

Georgia Tech

DAMO ALIBABA DAMO ACADEMY

An Overview

- **Problem**: two-block mixed-integer linear program (MILP)

$$p^* := \min\{c^\top x + g^\top z \mid Ax + Bz = 0,\ x \in X, z \in Z\};$$

  - X and Z are compact and mixed-integer representable
- **Goal**: decomposition algorithms with global optimality guarantees.
  - ϵ-solution: $c^\top x + g^\top z \le p^* + \epsilon$, $\|Ax + Bz\|_1 \le \epsilon$.
- **Two Algorithms**:
  - a framework based on the augmented Lagrangian method (ALM);
  - a variant of the alternating direction method of multipliers (ADMM).
- **Features**:
  - both algorithms converge to globally optimal solutions;
  - iteration complexity upper bounds to ϵ-solutions are derived;
  - when A is block-angular, subproblems can be updated in parallel;
  - applicable to multi-block settings (with reformulation):

$$\min\left\{\sum_{i=1}^{p} c_i^\top x_i \mid \sum_{i=1}^{p} A_i x_i \le b, x_i \in X_i\ \forall i \in [p]\right\}.$$

Augmented Lagrangian Duality for MILPs

- **ℓ_1-Augmented Lagrangian Function**:

$$L(x, z, \lambda, \rho) := c^\top x + g^\top z + \langle \lambda, Ax + Bz\rangle + \rho\|Ax + Bz\|_1.$$

- **Augmented Lagrangian Relaxation**:

$$d(\lambda, \rho) := \min\{L(x, z, \lambda, \rho) \mid x \in X, z \in Z\}.$$

- **Augmented Lagrangian Dual Problem**:

$$\max_{\lambda,\rho} d(\lambda, \rho) = \max_{\lambda,\rho} \min_{x\in X, z\in Z} L(x, z, \lambda, \rho) \le p^*. \quad \text{[Weak Duality]}$$

- **Exact Penalization for MILP (Feizollahi-Ahmed-Sun 17)**:
  - given **any** λ^*, there exists a **finite** $\rho^* > 0$ such that

$$\begin{aligned}&\mathrm{Argmin}\{L(x, z, \lambda^*, \rho^*) \mid x \in X\ z \in Z\}\\ =&\mathrm{Argmin}\{c^\top x + g^\top z \mid Ax + Bz = 0,\ x \in X, z \in Z\};\end{aligned}$$

  - consequently,

$$\max_{\lambda,\rho} d(\lambda, \rho) = d(\lambda^*, \rho^*) = p^*. \quad \text{[Strong Duality]}$$

Challenges in Algorithm Design

- Efficient evaluation of the AL dual function $d(\lambda, \rho)$ is missing:
  - existing works assume such an (inexact) oracle is available;
  - block coordinate descent (BCD) does not converge with ℓ_1 coupling.
- It is attempting to use ADMM, but:
  - ADMM may converge to local solutions or diverge;
  - only *stationary point* can be guaranteed on structured problems.

AUSAL for AL Dual Function Evaluation

- **Goal**: given (λ, ρ) and $\epsilon \ge 0$, find (x, z) such that

$$L(x, z, \lambda, \rho) \le d(\lambda, \rho) + \epsilon.$$

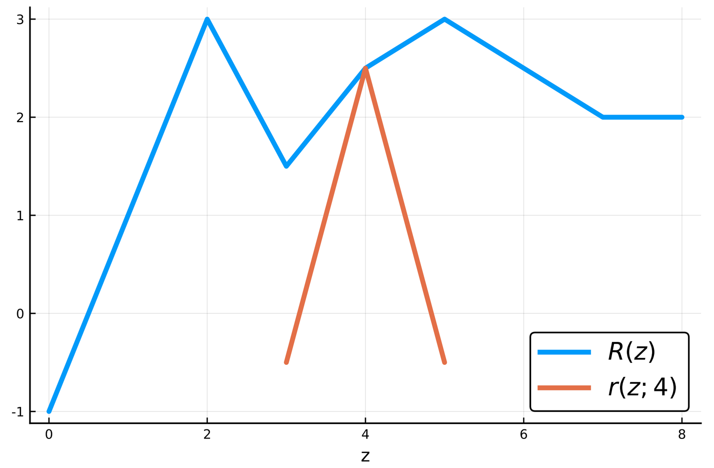
- **Our Approach**: decompose the minimization into two stages:

$$\min_{x\in X, z\in Z} L(x, z, \lambda, \rho) = \min_{z\in Z}\langle g + B^\top\lambda, z\rangle + \underbrace{\min_{x\in X}\langle c + A^\top\lambda, x\rangle + \rho\|Ax + Bz\|_1}_{=:R(z)}.$$

- **Lemma:** $R(z)$ is piece-wise linear and K-Lipschitz continuous where $K = \rho\|B\|_1$.
- **Reverse Norm Cut:**

$$r(z; \bar{z}) = R(\bar{z}) - K\|z - \bar{z}\|_1;$$

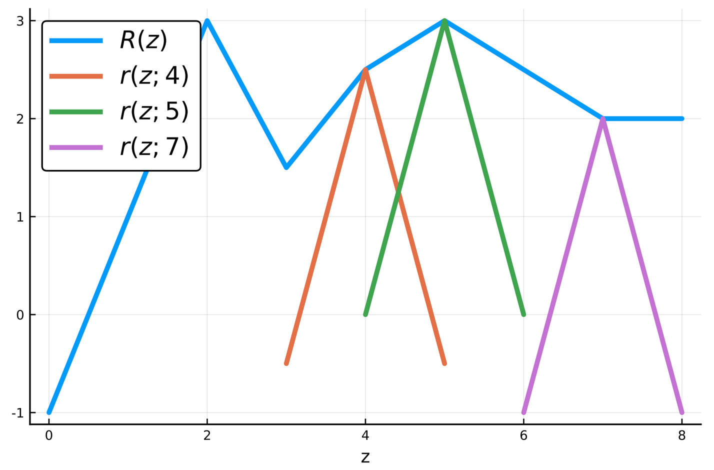
  - lower approximation of R: $r(z; \bar{z}) \le R(z)\ \forall z \in Z$;
  - locally tight at $\bar{z}$: $r(\bar{z}; \bar{z}) = R(\bar{z})$.



- **Point-wise Maximum of a Collection of Reverse Norm Cuts:**

$$\max\{r(z; z^i) \mid i = 1, \cdots, k\}$$

  - a better approximation as the number of trial points increases.



- **Alternating Update Scheme for the Sharp Augmented Lagrangian (AUSAL)**:
  - decide the next trial point and update `LowerBound`

$$z^{k+1} \in \mathop{\mathrm{Argmin}}_{z\in Z}\ \langle g + B^\top\lambda, z\rangle + \max\{r(z; z^i) \mid i = 1, \cdots, k\};$$

  - evaluate $R(z^{k+1})$ and update `UpperBound`;
  - stop if `UpperBound` − `LowerBound` $\le \epsilon$.
- **Convergence of AUSAL**:
  - AUSAL subsequentially converges to an optimal solution of the AL relaxation.
  - Given $\epsilon > 0$, AUSAL terminates with an ϵ-optimal solution of the AL relaxation in $\mathcal{O}(\rho^d/\epsilon^d)$ iterations, where $Z \subseteq \mathbb{R}^d$.

First Algorithm: ALM Empowered by AUSAL

- **Penalty Method**: a single call of AUSAL with proper (λ, ρ).
  - If (λ, ρ) supports exact penalization, then AUSAL returns an ϵ-solution of the MILP in $\mathcal{O}(\rho^d/\epsilon^d)$ iterations.
  - Fix any λ and choose $\rho = \Theta(1/\epsilon) + \|\lambda\|_\infty$, then AUSAL returns an ϵ-solution of the MILP in $\mathcal{O}(1/\epsilon^{2d})$ iterations.
- **ALM**: inexact subgradient updates on (λ, ρ).
  - $d(\lambda, \rho)$ is concave and upper-semicontinuous.
  - If $(x, z) = \text{AUSAL}(\lambda, \rho, \epsilon)$, then

$$-\begin{bmatrix} Ax + Bz \\ \|Ax + Bz\|_1 \end{bmatrix} \in \partial_\epsilon(-d)(\lambda, \rho).$$

  - Inexact subgradient ascent: with proper choices of α_k,

$$\begin{aligned}(x^k, z^k) =&\text{AUSAL}(\lambda^k, \rho^k, \epsilon);\\ \lambda^{k+1} =&\lambda^k + \alpha_k(Ax^k + Bz^k);\\ \rho^{k+1} =&\rho^k + \alpha_k\|Ax^k + Bz^k\|_1.\end{aligned}$$

Second Algorithm: An ADMM Variant

- **Motivation for ADMM**:
  - ALM is double-looped;
  - historical reverse norm cuts cannot be reused efficiently.
- **Solve the AL Relaxation in Variable x**:
  - define

$$\begin{aligned}R(\bar{z}) :=& \min_{x\in X} c^\top x + \rho\|Ax + B\bar{z}\|_1.\\ P(\bar{z}, \bar{\mu}, \bar{\beta}) :=& \min_{x} c^\top x + \langle\bar{\mu}, Ax + B\bar{z}\rangle + \bar{\beta}\|Ax + B\bar{z}\|_1.\end{aligned}$$

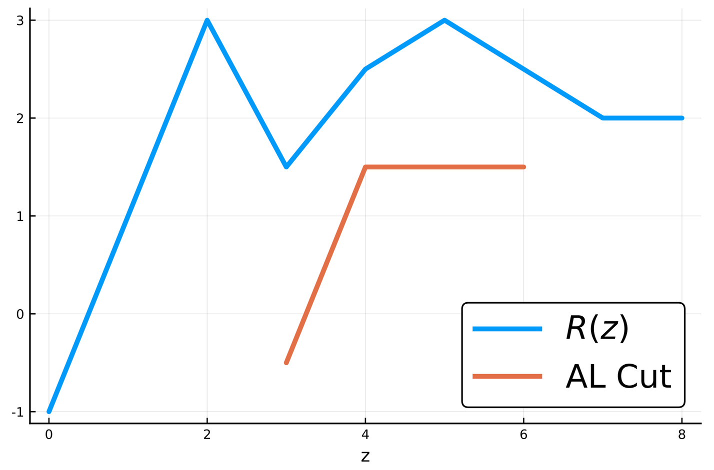
  - a strong duality in x:

$$\max_{(\bar{\mu},\bar{\beta})\in\Lambda(\rho)} P(\bar{z}, \bar{\mu}, \bar{\beta}) = R(\bar{z}),$$

where $\Lambda(\rho) := \{(\mu, \beta) : \beta \ge 0, \|\mu\|_\infty + \beta \le \rho\}$;
- **Augmented Lagrangian Cuts (AL Cuts)**: pick $(\bar{\mu}, \bar{\beta}) \in \Lambda(\rho)$, define

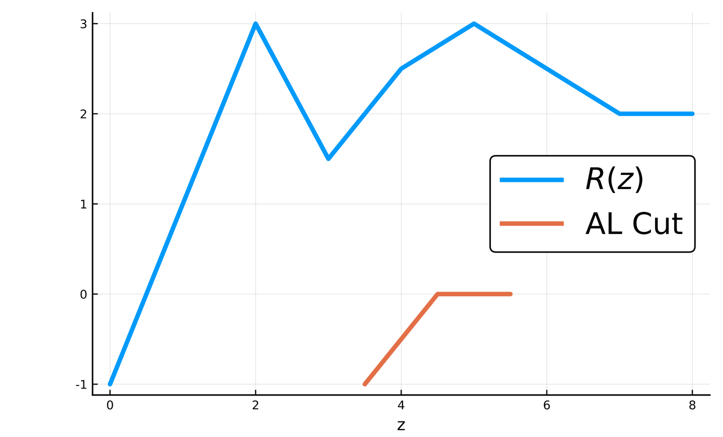
$$r(z; \bar{z}, \bar{\mu}, \bar{\beta}) := P(\bar{z}, \bar{\mu}, \bar{z}) + \langle\bar{\mu}, Bz - B\bar{z}\rangle - \bar{\beta}\|Bz - B\bar{z}\|_1;$$

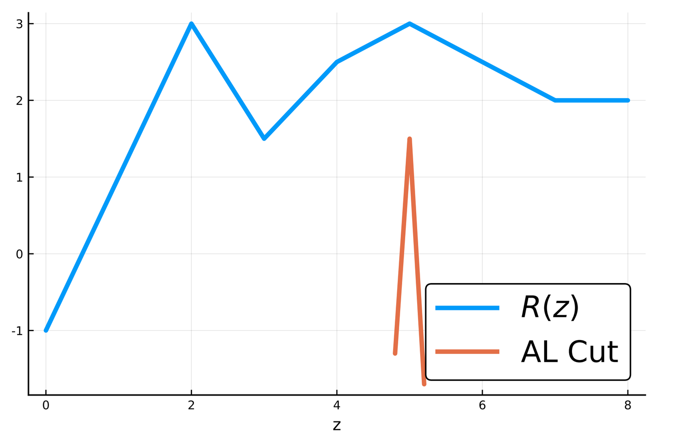
  - lower approximation of $R(z)$: $r(z; \bar{z}, \bar{\mu}, \bar{\beta}) \le R(z)\ \forall z \in Z$;
  - not necessarily tight;
  - allows rotation and a smaller Lipschitz constant (fatter in shape).



Second Algorithm: An ADMM Variant (Cont.)

- **An ADMM Variant**: use AL cuts to approximate the dependency on z.
  - $x^k \in \mathrm{Argmin}_{x\in X}\ c^\top x + \langle\mu^k, Ax + Bz^{k-1}\rangle + \beta^k\|Ax + Bz^{k-1}\|_1$;
  - $z^k \in \mathrm{Argmin}_{z\in Z}\ g^\top z + \max_{j\in[k]}\{r(z; z^{j-1}, \mu^j, \beta^j)\}$;
  - flexible update on (μ^{k+1}, β^{k+1}).
- **Assumptions**: Let $\underline{\rho} > 0$ be a finite penalty that supports exact penalization. Suppose the sequence $\{(\mu^k, \beta^k)\}_{k\in\mathbb{N}}$ are chosen such that
  - $\beta^k - \|\mu^k\|_\infty \ge \underline{\rho}$ for large enough $k \in \mathbb{N}$;
  - $\beta^k + \|\mu^k\|_\infty \le \bar{\rho}$ for all $k \in \mathbb{N}$.
- **Geometric Intuition**: prevents loose and slim AL cuts.





- **Convergence of the ADMM Variant**:
  - The ADMM variant subsequentially converges to an optimal solution of the two-block MILP problem.
  - The ADMM variant finds an ϵ-solution of the two-block MILP in $\mathcal{O}((\bar{\rho} + \underline{\rho})^d/\epsilon^d)$ iterations, where $Z \subseteq \mathbb{R}^d$.

Comparison with Primal/Dual Decomposition

- Multi-block problem:
  - 10 linear constraints couple 100 blocks;
  - each block has local constraints: $X_i = \{x \in \{0, \cdots, 60\} \times [-60, 60] | E_i x \le f_i\}$
- Compare to primal (Camisa et al. 18) and dual (Vujanic et al. 16) Decomposition.
  - ALM and ADMM find optimal solutions and are faster in most cases;
  - may fail to find feasible solution due to subproblem slow-down;
  - need better strategy to manage nonconvex cuts.

	ALM		ADMM		Primal		Dual	
b	Gap(%)	Time(s)	Gap(%)	Time(s)	Gap(%)	Time(s)	Gap(%)	Time(s)
	0.00	2.12	0.00	2.04	151.11	356.16	14.96	2.50
	0.00	2.16	0.00	1.79	148.19	328.91	9.95	2.62
1200$\mathbf{e}$	*5.97	MAX	0.01	51.35	134.21	1195.83	10.79	2.56
	0.00	2.08	0.00	2.22	115.17	762.54	4.63	2.47
	0.00	2.45	0.00	4.16	135.81	982.31	16.46	2.49
	0.00	2.14	0.00	2.53	162.58	619.04	12.97	2.90
	*6.57	MAX	0.02	160.13	144.29	420.62	24.02	2.53
1000$\mathbf{e}$	0.00	2.08	0.00	2.60	122.71	320.78	251.43	0.99
	0.00	2.31	0.00	1.91	110.11	900.52	2.70	2.24
	0.00	2.55	0.00	2.19	109.39	47.56	13.37	2.67
	*4.82	MAX	0.00	233.12	136.37	948.73	36.94	2.95
	*5.95	MAX	0.00	10.76	126.77	778.19	10.28	2.12
800$\mathbf{e}$	0.00	2.14	0.00	2.62	144.48	1046.39	8.54	2.94
	0.00	2.30	0.00	1.90	161.41	1173.69	63.09	2.71
	0.00	2.14	0.00	2.44	135.91	689.21	4.12	2.44

arXiv preprint: https://arxiv.org/abs/2102.11980

IPCO 2021

ksun46@gatech.edu