A Bundle-based Augmented Lagrangian Framework: Algorithm, Convergence, and Primal-dual Principles

Feng-Yi Liao

Joint work with Yang Zheng

ECE department, UC San Diego
Scalable Optimization and Control (SOC) Lab

ICCOPT 2025, Los angeles, USA
July 22, 2025

Outline

- Consider the following constrained convex optimization problem:

$$\begin{aligned} p^\star := \min_{x \in \mathbb{R}^n} \quad & \langle c, x \rangle \\ \text{subject to} \quad & \mathcal{A}x = b, \\ & x \in \Omega, \end{aligned} \tag{P}$$

where $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, and $\Omega \subseteq \mathbb{R}^n$ is a compact convex set.

Applications

- Semidefinite programs

$$\min_{X \in \mathbb{S}^n} \{ \langle C, X \rangle \mid \mathcal{A}X = b, X \in \Omega \},$$

where $\Omega = \{X \in \mathbb{S}^n_+ : \mathrm{tr}(X) \leq \gamma\}$, where γ is large enough.

- Max-Cut

$$\min_{X \in \mathbb{S}^n} \{ \langle L, X \rangle \mid X_{ii} = 1, i = 1, \ldots, n, X \in \mathbb{S}^n_+ \}.$$

Augmented Lagrangian methods

- A conceptually simple framework for constrained optimization (Rockafellar 1976; Hestenes 1969; Powell 1969)
- Define the augmented Lagrangian function $\mathcal{L}(x, y)$ with $\rho > 0$

$$\mathcal{L}_\rho(x, y) = \langle c, x \rangle + \langle y, b - \mathcal{A}x \rangle + \frac{\rho}{2}\|b - \mathcal{A}x\|^2.$$

- For iterations $k = 1, 2, \ldots,$ the augmented Lagrangian method (ALM) repeats the two steps

$$x_{k+1} \in \operatorname*{argmin}_{x \in \Omega} \; \mathcal{L}_\rho(x, y_k), \tag{1a}$$

$$y_{k+1} = y_k + \rho(b - \mathcal{A}x_{k+1}). \tag{1b}$$

- The minimization in (1a) is difficult. Often consider the inexact ALM (Rockafellar 1976):

$$x_{k+1} \approx \operatorname*{argmin}_{x \in \Omega} \; \mathcal{L}_\rho(x, y_k),$$

$$y_{k+1} = y_k + \rho(b - \mathcal{A}x_{k+1}).$$

Proximal point methods

- The ALM is the same as the proximal point method applied to the dual (Rockafellar 1976)
- **Proximal point method (PPM):** Consider $\min_{y \in \mathbb{R}^n} f(y)$, where f is a convex function. For iterations $k = 1, 2, \ldots$, the PPM performs the proximal update

$$y_{k+1} = \mathrm{prox}_{\alpha f}(y_k), \quad k = 1, 2, \ldots,$$

where $\mathrm{prox}_{\alpha f}$ is the proximal mapping defined as

$$\mathsf{prox}_{\alpha f}(y_k) := \underset{y \in \mathbb{R}^n}{\mathsf{argmin}} \; f(y) + \frac{1}{2\alpha} \|y - y_k\|^2 .$$

- However, the proximal update is difficult to evaluate. Often consider inexact PPM

$$y_{k+1} \approx \mathrm{prox}_{\alpha f}(y_k).$$

- The **proximal bundle method** can be viewed as an efficient realization of the inexact PPM (Liang and Monteiro 2021)

Proximal bundle methods

- The difficulty of the proximal operator $\mathrm{prox}_{\alpha f}(y_k)$ comes from the function f
- Approximate f by a simplier lower approximation f_k (i.e., $f_k \leq f$) (Lemarechal and Zowe 1994; Liang and Monteiro 2021; Kiwiel 2000)
- Acquire a candidate point z_{k+1} by

$$z_{k+1} = \mathrm{prox}_{\alpha f_k}(y_k)$$

- Test if z_{k+1} provides sufficient descent by the test

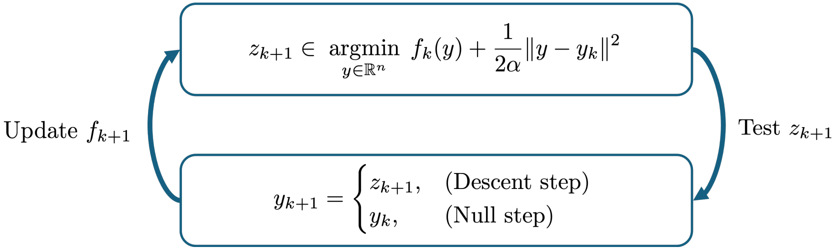
$$f(z_{k+1}) \leq f(y_k) - \beta \underbrace{(f(y_k) - f_k(z_{k+1}))}_{\text{Approximated drop}}, \quad \text{where } \beta \in (0,1) \text{ is fixed.} \tag{2}$$

- Update the iterate

$$y_{k+1} = \begin{cases} z_{k+1}, & \text{if (2) holds} \quad \text{(Descent step)} \\ y_k, & \text{otherwise} \quad \text{(Null step).} \end{cases}$$

Proximal bundle methods

- Update the approximation f_{k+1} satisfying (Díaz and Grimmer 2023)
  - **Lower approximation:** $f_{k+1} \leq f$.
  - **Subgradient:** There exists $v_{k+1} \in \partial f(z_{k+1})$ s.t. $f_{k+1}(\cdot) \geq f(z_{k+1}) + \langle v_{k+1}, \cdot - z_{k+1} \rangle$.
  - **Aggregation:** For a null step, we require $f_{k+1}(\cdot) \geq f_k(z_{k+1}) + \langle s_{k+1}, \cdot - z_{k+1} \rangle$, where $s_{k+1} = (y_k - z_{k+1})/\alpha \in \partial f_k(z_{k+1})$.



Big Picture: Primal and Dual Perspective

- Let $g(y) = -\min_{x\in\Omega} \mathcal{L}(x,y)$ be the dual function and $\mathcal{L}(x,y) = \langle c, x\rangle + \langle y, b - \mathcal{A}x\rangle$.

	Primal		**Dual**	
	$\min\limits_{x\in\mathbb{R}^n} \{\langle c, x\rangle \mid \mathcal{A}x = b, x \in \Omega\}.$		$\min\limits_{y\in\mathbb{R}^m} \quad g(y).$	
ALM	$x_{k+1} \in \operatorname{argmin}\limits_{x\in\Omega} \; \mathcal{L}_\rho(x, y_k),$ $y_{k+1} = y_k + \rho(b - \mathcal{A}x_{k+1}).$	$\Leftrightarrow$	$y_{k+1} = \mathrm{prox}_{\rho g}(y_k)$	**PPM**
	$\Uparrow$		$\Uparrow$	
Inexact ALM	$x_{k+1} \approx \operatorname{argmin}\limits_{x\in\Omega} \; \mathcal{L}_\rho(x, y_k),$ $y_{k+1} = y_k + \rho(b - \mathcal{A}x_{k+1}).$		$y_{k+1} \approx \mathrm{prox}_{\rho g}(y_k)$	**Inexact PPM**
	$\Uparrow$		$\Uparrow$	
	?	$\Leftrightarrow$	$z_{k+1} = \mathrm{prox}_{\rho g_k}(y_k)$ $y_{k+1} = z_{k+1}$ or y_k	**Proximal bundle method**

Big Picture: Primal and Dual Perspective

- Let $g(y) = -\min_{x\in\Omega} \mathcal{L}(x,y)$ be the dual function and $\mathcal{L}(x,y) = \langle c, x\rangle + \langle y, b - \mathcal{A}x\rangle$.

	Primal		**Dual**	
	$\min_{x\in\mathbb{R}^n} \{\langle c, x\rangle \mid \mathcal{A}x = b, x \in \Omega\}.$		$\min_{y\in\mathbb{R}^m} \quad g(y).$	
ALM	$x_{k+1} \in \operatorname{argmin}_{x\in\Omega} \; \mathcal{L}_\rho(x, y_k),$ $y_{k+1} = y_k + \rho(b - \mathcal{A}x_{k+1}).$	$\Leftrightarrow$	$y_{k+1} = \mathrm{prox}_{\rho g}(y_k)$	**PPM**
	$\Uparrow$		$\Uparrow$	
Inexact ALM	$x_{k+1} \approx \operatorname{argmin}_{x\in\Omega} \; \mathcal{L}_\rho(x, y_k),$ $y_{k+1} = y_k + \rho(b - \mathcal{A}x_{k+1}).$		$y_{k+1} \approx \mathrm{prox}_{\rho g}(y_k)$	**Inexact PPM**
	$\Uparrow$		$\Uparrow$	
	This work	$\Leftrightarrow$	$z_{k+1} = \mathrm{prox}_{\rho g_k}(y_k)$ $y_{k+1} = z_{k+1}$ or y_k	**Proximal bundle method**

This work: A Bundle-based Augmented Lagrangian Framework

Contribution 1: A new Bundle-based Augmented Lagrangian Algorithm (BALA):

$$w_{k+1} \in \operatorname*{argmin}_{x \in \Omega_k} \mathcal{L}_\rho(x, y_k),$$

$$z_{k+1} = y_k + \rho(b - \mathcal{A} w_{k+1})$$

Update Ω_{k+1} Test (w_{k+1}, z_{k+1})

$$(x_{k+1}, y_{k+1}) = \begin{cases} (w_{k+1}, z_{k+1}), & \text{(Descent step)}, \\ (x_k, y_k), & \text{(Null step)}. \end{cases}$$

Contribution 2: Sublinear convergence: Under the proper choice of the parameters, for any $\epsilon > 0$, BALA finds a pair of primal and dual solutions (x_k, y_k) satisfying

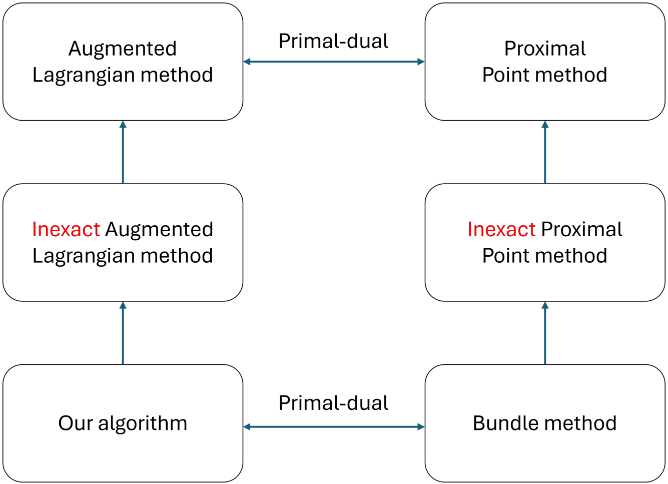
$$\max\{|\langle c, x_k \rangle - p^\star|, \|\mathcal{A} x_k - b\|, g(y_k) - g^\star\} \le \epsilon$$

in at most $\mathcal{O}(\epsilon^{-2})$ iterations.

Contribution 3: Linear convergence: Under mild assumptions, the complexity becomes $\mathcal{O}(\log(\frac{1}{\epsilon}))$.

This work: A Bundle-based Augmented Lagrangian Framework

Contribution 4: Primal and dual interplay



Outline

Bundle-based Augmented Lagrangian Framework

Under the inexact ALM framework

$$x_{k+1} \approx \operatorname*{argmin}_{x \in \Omega} \; \mathcal{L}_\rho(x, y_k),$$
$$y_{k+1} = y_k + \rho(b - \mathcal{A}x_{k+1}).$$

How to solve the minimization efficiently?

- The fact that the minimization is difficult is due to the constraint $x \in \Omega$
- What if we approximate the set Ω by a **simple** inner approximation Ω_k, i.e.,

  $$\Omega_k \subseteq \Omega,$$

  and solve the simpler subproblem **exactly**

  $$w_{k+1} \in \operatorname*{argmin}_{x \in \Omega_k} \; \mathcal{L}_\rho(x, y_k)$$

- For example, $\Omega_k = \mathrm{conv}(v_k, w_k)$ where $v_k, w_k \in \Omega$. An analytical solution exists.

Bundle-based Augmented Lagrangian Framework

Challenges of the proposal $w_{k+1} \in \operatorname{argmin}_{x\in\Omega_k} \mathcal{L}_\rho(x, y_k)$

- How to decide w_{k+1} is a good solution?
- In general, it is difficult to relate w_{k+1} with the true solution $\operatorname{argmin}_{x\in\Omega} \mathcal{L}_\rho(x, y_k)$.

Potential solutions

- A very good approximation $\Omega_k \subseteq \Omega$, i.e., $\Omega_k \approx \Omega$
- However, a good approximation Ω_k leads to a **harder** subproblem

We aim to design **a sequence of inner approximation** $\{\Omega_k\}$ such that the sequence

$$\{w_{k+1} \in \operatorname*{argmin}_{x\in\Omega_k} \mathcal{L}_\rho(x, y_k)\}$$

- finds an ϵ_k-solution to $\operatorname{argmin}_{x\in\Omega} \mathcal{L}_\rho(x, y_k)$;
- the sequence of sets do not need to **be nested** $\Omega_k \not\subseteq \Omega_{k+1}$

Bundle-based Augmented Lagrangian Framework

Decide if w_{k+1} is a good candidate:

Idea: Look at the descent generated by $z_{k+1} = y_k + \rho(b - \mathcal{A}w_{k+1})$ in the dual.

- An inner approximation $\Omega_k \subseteq \Omega$ naturally defines an **approximated** dual function g_k :

$$g_k(y) := -\min_{x \in \Omega_k} \mathcal{L}(x, y) \leq g(y), \quad \forall y \in \mathbb{R}^m,$$

  where g is the dual function.
- Test the descent progress

$$g(z_{k+1}) \leq g(y_k) - \beta \underbrace{\left(g(y_k) - g_k(z_{k+1})\right)}_{\text{Approximated drop}}, \quad \text{where } \beta \in (0,1) \text{ is fixed.} \tag{3}$$

- Update the iterate

$$(x_{k+1}, y_{k+1}) = \begin{cases} (w_{k+1}, z_{k+1}), & \text{if (3) holds (descent step)}, \\ (x_k, y_k), & \text{otherwise (null step)}. \end{cases}$$

Bundle-based Augmented Lagrangian Framework

Update the inner approximation Ω_{k+1}:

- **Inner approximation:** we have $\Omega_{k+1} \subseteq \Omega$ closed and convex;
- **Dual information:** we require $v_{k+1} \in \Omega_{k+1}$, where $v_{k+1} \in \Omega$ satisfies

$$g(z_{k+1}) = -\mathcal{L}(v_{k+1}, y_k) \quad \text{or equivalently} \quad v_{k+1} \in \operatorname*{argmin}_{x \in \Omega} \mathcal{L}(x, z_{k+1});$$

- **Primal information:** if the step k is a null step, then we require $w_{k+1} \in \Omega_{k+1}$.

$$w_{k+1} \in \operatorname*{argmin}_{x \in \Omega_k} \mathcal{L}_\rho(x, y_k),$$

$$z_{k+1} = y_k + \rho(b - \mathcal{A}w_{k+1})$$

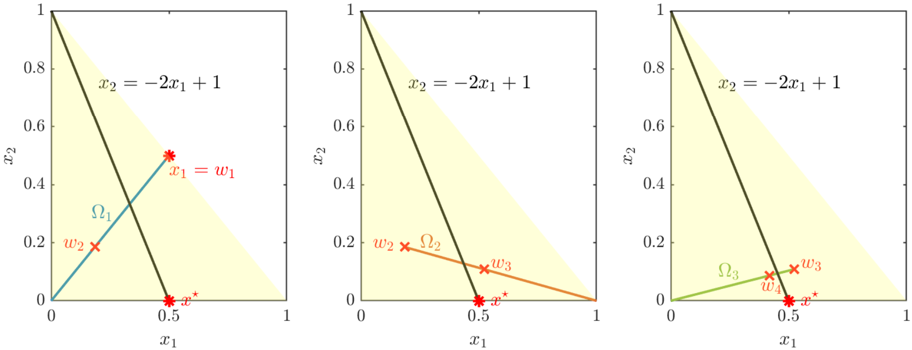
Test (w_{k+1}, z_{k+1})

$$(x_{k+1}, y_{k+1}) = \begin{cases} (w_{k+1}, z_{k+1}), & \text{(Descent step)}, \\ (x_k, y_k), & \text{(Null step)}. \end{cases}$$

Update Ω_{k+1}

Illustration in 2 dimensions

$$\begin{aligned} \min_{x\in\mathbb{R}^2} \quad & x_1 + x_2 \\ \text{subject to} \quad & 2x_1 + x_2 = 1, \\ & x \in \mathbb{R}^2_+, |x|_1 \leq 1. \end{aligned}$$



Subproblem in `BALA`

$$w_{k+1} \in \operatorname*{argmin}_{x \in \Omega_k} \; \mathcal{L}_\rho(x, y_k), \quad z_{k+1} = y_k + \rho(b - \mathcal{A}x_{k+1}).$$

(Lemma) The dual update z_{k+1} is the same as a proximal step on the approximated dual function, i.e.,

$$z_{k+1} = \min_{y \in \mathbb{R}^m} g_k(y) + \frac{1}{2\rho}\|y - y_k\|^2,$$

where $g_k = -\min_{x \in \Omega_k} \mathcal{L}(x, \cdot)$ an approximated dual function of g.

(Lemma) The construction of Ω_{k+1} implies that the function $g_{k+1} = -\min_{x \in \Omega_{k+1}} \mathcal{L}(x, \cdot)$ satisfies the assumptions for the proximal bundle method.

`BALA` **can be viewed as a proximal bundle method applied to the dual.**

Outline

Theorem 1 (Sublinear convergences)

For any $\epsilon > 0$, $\mathtt{BALA}$ *with parameters* $\beta \in (0,1)$ *and* $\rho > 0$ *finds a dual iterate* y_k *satisfying*

$$g(y_k) - g^\star \leq \epsilon$$

in at most $\mathcal{O}\left(\epsilon^{-3}\right)$ *number of iterations, and a primal iterate* x_k *satisfying*

$$|\langle c, x_k\rangle - p^\star| \leq \epsilon \text{ and } \|\mathcal{A}x_k - b\| \leq \epsilon$$

in at most $\mathcal{O}\left(\epsilon^{-6}\right)$ *number of iterations. Moreover, if we choose* $\rho = 1/\epsilon$, *then the iteration complexities are improved to* $\mathcal{O}(\epsilon^{-2})$ *for both the primal and dual iterates, respectively.*

Theorem 2 ((Local) Linear convergences)

Suppose that the dual function g satisfies quadratic growth

$$g(y) - g^\star \geq \frac{\alpha}{2} \cdot \mathrm{dist}^2(y, \Omega_\mathrm{D}), \quad \forall y \in \mathbb{R}^m \tag{4}$$

and the approximation function g_k satisfies

$$g_k(y) \leq g(y) \leq g_k(y) + \frac{\gamma}{2}\|y - y_k\|^2, \ \forall y \in \mathbb{R}^m, \tag{5}$$

for all $k \geq T$ with $\gamma > 0$. Under a proper choice of parameters, for all iterations $k \geq T$, there exists two constants $\mu_1 \in (0,1), \mu_2 > 0$ such that

$$\mathrm{dist}(y_{k+1}, \Omega_\mathrm{D}) \leq \mu_1 \cdot \mathrm{dist}(y_k, \Omega_\mathrm{D})$$

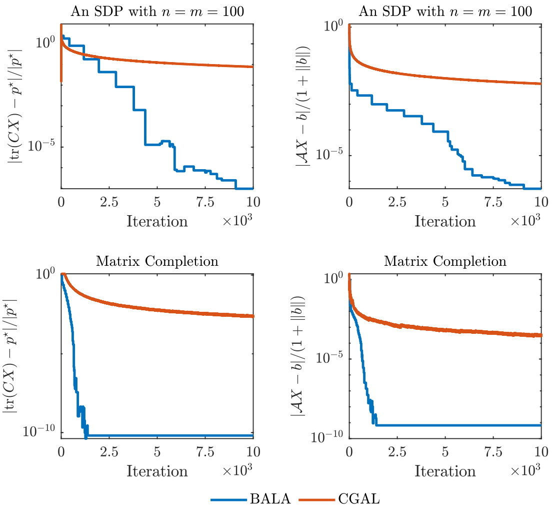
and

$$\max\{g(y_k) - g^\star, \|\mathcal{A}x_k - b\|^2, |\langle c, x_k\rangle - p^\star|^2\} \leq \mu_2 \cdot \mathrm{dist}(y_k, \Omega_\mathrm{D}).$$

Outline

Numerical results

Comparison with the algorithm CGAL in (Yurtsever, Fercoq, and Cevher 2019)



Comparison with the SCS (O'Donoghue 2021) and SDPNAL+ (Sun et al. 2020)

Instance	Metric	SCS	SDPNAL+	BALA	Instance	Metric	SCS	SDPNAL+	BALA
	ϵ_{p}	9.6e−3	3.6e−4	2.5e−4		ϵ_{p}	2.4e−2	2.9e−4	3.6e−4
	ϵ_{d}	1.1e−4	3.1e−4	0		ϵ_{d}	3.4e−4	4.5e−4	0
qpp100	ϵ_{g}	2.8e−7	3.2e−4	9.1e−6	**qpp250-1**	ϵ_{g}	2e−5	5.4e−3	1.4e−5
	Cost	4.5e1	4.5e1	4.5e1		Cost	1.6e1	1.6e1	1.5e1
	Time(s)	1.1e1	4.9e−1	1.6		Time(s)	1.2e2	1.1	5.1
	ϵ_{p}	1.2e−2	2.6e−4	4.5e−4		ϵ_{p}	N/A	2.1e−4	2.3e−4
	ϵ_{d}	1.1e−4	3.5e−4	0		ϵ_{d}	N/A	1.6e−5	0
qpp500-1	ϵ_{g}	7.7e−5	5.4e−3	1.1e−4	**qpG51**	ϵ_{g}	N/A	4.8e−5	2.1e−5
	Cost	4.5e1	4.5e1	4.5e1		Cost	N/A	−1.2e4	1.2e4
	Time(s)	1.0e2	7.5	1.1e1		Time(s)	36e1	6.1e2	1.2e2

$$\epsilon_{\mathrm{p}} = \frac{\|\mathcal{A}x - b\|}{1+\|b\|}, \quad \epsilon_{\mathrm{d}} = \frac{\|C - \mathcal{A}^* y - Z\|}{1+\|C\|}, \quad \text{and} \quad \epsilon_{\mathrm{g}} = \frac{|\langle C, X\rangle - \langle b, y\rangle|}{1+|\langle C, X\rangle| + |\langle b, y\rangle|}.$$

- Solved until $\max\{\epsilon_{\mathrm{p}}, \epsilon_{\mathrm{d}}, \epsilon_{\mathrm{g}}\} \le 5 \times 10^{-4}$

Outline

Conclusion

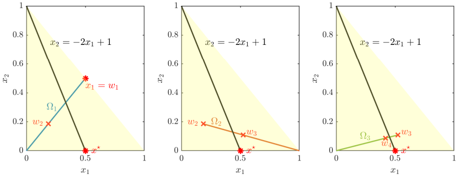
- A new <u>B</u>undle-based <u>A</u>ugmented <u>L</u>agrangian <u>A</u>lgorithm (`BALA`):

$$w_{k+1} \in \underset{x \in \Omega_k}{\operatorname{argmin}} \ \mathcal{L}_\rho(x, y_k),$$
$$z_{k+1} = y_k + \rho(b - \mathcal{A}w_{k+1})$$

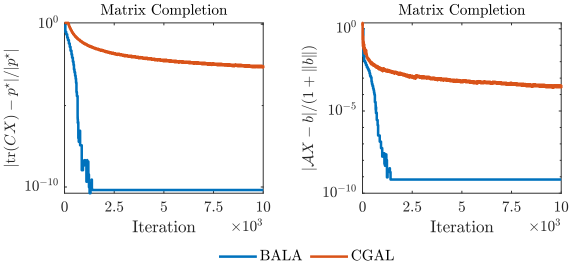
Test (w_{k+1}, z_{k+1})

$$(x_{k+1}, y_{k+1}) = \begin{cases} (w_{k+1}, z_{k+1}), & \text{(Descent step)}, \\ (x_k, y_k), & \text{(Null step)}. \end{cases}$$

Update Ω_{k+1}

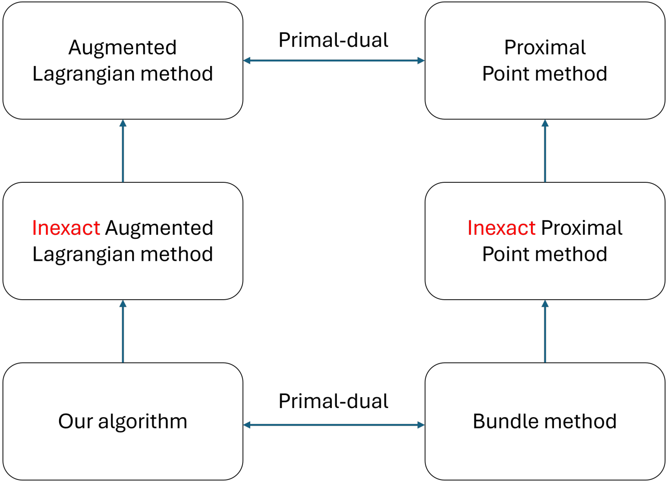


- `BALA` has sublinear and linear convergence guarantees



Conclusion

- Primal and dual interplay



Thank you for your attention!

Q & A

- Feng-Yi Liao and Yang Zheng (2025). "A Bundle-based Augmented Lagrangian Framework: Algorithm, Convergence, and Primal-dual Principles". In: *arXiv preprint arXiv:2502.08835*

Supported by NSF ECCS 2154650, NSF CMMI 2320697, and NSF CAREER 2340713.

References I

- Díaz, Mateo and Benjamin Grimmer (2023). "Optimal convergence rates for the proximal bundle method". In: *SIAM Journal on Optimization* 33.2, pp. 424–454.
- Hestenes, Magnus R (1969). "Multiplier and gradient methods". In: *Journal of optimization theory and applications* 4.5, pp. 303–320.
- Kiwiel, Krzysztof C (2000). "Efficiency of proximal bundle methods". In: *Journal of Optimization Theory and Applications* 104, pp. 589–603.
- Lemarechal, Claude and Jochem Zowe (1994). "A condensed introduction to bundle methods in nonsmooth optimization". In: *Algorithms for continuous optimization: the state of the art*. Springer, pp. 357–382.
- Liang, Jiaming and Renato DC Monteiro (2021). "A proximal bundle variant with optimal iteration-complexity for a large range of prox stepsizes". In: *SIAM Journal on Optimization* 31.4, pp. 2955–2986.
- Liao, Feng-Yi and Yang Zheng (2025). "A Bundle-based Augmented Lagrangian Framework: Algorithm, Convergence, and Primal-dual Principles". In: *arXiv preprint arXiv:2502.08835*.
- O'Donoghue, Brendan (Aug. 2021). "Operator Splitting for a Homogeneous Embedding of the Linear Complementarity Problem". In: *SIAM Journal on Optimization* 31 (3), pp. 1999–2023.
- Powell, Michael JD (1969). "A method for nonlinear constraints in minimization problems". In: *Optimization*, pp. 283–298.

References II

Rockafellar, R Tyrrell (1976). "Augmented Lagrangians and applications of the proximal point algorithm in convex programming". In: *Mathematics of operations research* 1.2, pp. 97–116.

Sun, Defeng et al. (2020). "SDPNAL+: A Matlab software for semidefinite programming with bound constraints (version 1.0)". In: *Optimization Methods and Software* 35.1, pp. 87–115.

Yurtsever, Alp, Olivier Fercoq, and Volkan Cevher (2019). "A conditional-gradient-based augmented Lagrangian framework". In: *International Conference on Machine Learning*. PMLR, pp. 7272–7281.