Improved Analysis of Restarted Accelerated Gradient and Augmented Lagrangian Methods via Inexact Proximal Point Frameworks

Jiaming Liang

Department of Computer Science
University of Rochester

March 5, 2026

Joint work with Matthew X. Burns (University of Rochester)

Applied Mathematics Seminar, Syracuse University

Overview

Outline

1. Introduction
2. LOrA
   - Dual: Inexact Augmented Lagrangian Method
3. FLOrA
   - Primal: Restarted ACG
   - Dual: Inexact Fast Augmented Lagrangian Method
4. Numerical Experiments

Problem I: CSCO

Problem: Convex Smooth Composite Optimization

$$\phi_* = \min_{x\in\mathbb{R}^n} \{\phi(x) := f(x) + h(x)\}$$

- f is closed proper convex and L_f-smooth on $\mathbb{R}^n$, i.e.,

$$\|\nabla f(x) - \nabla f(y)\| \le L_f \|x - y\|$$

  for all $x, y \in \mathbb{R}^n$.
- h is convex and simple, i.e.,

$$\mathrm{prox}_{\eta h}(x) := \operatorname*{argmin}_{y\in\mathbb{R}^n} \left\{ \eta h(y) + \frac{1}{2}\|y - x\|^2 \right\}$$

  is efficiently computable for all $\eta > 0$.

Goal ε-solution: find $x \in \mathrm{dom}\, h$ satisfying

$$\phi(x) - \phi_* \le \varepsilon.$$

Motivation

- Nesterov's accelerated composite gradient (ACG) method achieves optimal complexity $\mathcal{O}(\varepsilon^{-1/2})$ for solving CSCO.
- Restarted ACG methods are a widely used strategy to improve ACG empirical performance and suppress oscillations.
- Heuristic restart schemes: gradient (O'Donoghue and Candès [2015]), function value (O'Donoghue and Candès [2015]), speed (Su et al. [2016]). No rigorous analysis for global convergence.
- **Goal:** develop an efficient restarted ACG method with optimal complexity via the inexact proximal point (IPP) framework.

Problem II: LC-CSCO

Problem: Linearly Constrained Convex Smooth Composite Optimization

$$\hat{\phi}_* = \min_{x \in \mathbb{R}^n} \{\phi(x) := f(x) + h(x) : Ax = b\}$$

- $\mathrm{dom}\, f \supset \mathrm{dom}\, h$
- f is closed proper convex and L_f-smooth on $\mathbb{R}^n$.
- h is convex and simple.
- Slater's condition is satisfied.
- $\mathrm{dom}\, h$ is bounded with diameter $D \geq 1$.

Goal ε-primal-dual solution: find $(x, \lambda) \in \mathrm{dom}\, h \times \mathbb{R}^m$ satisfying

$$v \in \nabla f(x) + \partial h(x) + A^\top \lambda, \quad \|v\| \leq \varepsilon, \quad \|Ax - b\| \leq \varepsilon.$$

Alternative: ε-primal solution $x \in \mathrm{dom}\, h$

$$|\phi(x) - \hat{\phi}_*| \leq \varepsilon, \quad \|Ax - b\| \leq \varepsilon.$$

Motivation

A classical method for solving LC-CSCO problems is the augmented Lagrangian method (ALM), also known as the method of multipliers, which has the iteration

$$x_{k+1} = \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \mathcal{L}_\rho(u, \lambda_k),$$
$$\lambda_{k+1} = \lambda_k + \rho(Ax_{k+1} - b),$$

where the augmented Lagrangian with penalty coefficient $\rho > 0$ is

$$\mathcal{L}_\rho(x, \lambda) = \phi(x) + \langle \lambda, Ax - b \rangle + \frac{\rho}{2} \|Ax - b\|^2.$$

It is known from Rockafellar [1976] that ALM can be reformulated as a proximal point method in the dual,

$$\lambda_{k+1} = \underset{\lambda \in \mathbb{R}^m}{\operatorname{argmax}} \left\{ d(\lambda) - \frac{1}{2\rho} \|\lambda - \lambda_k\|^2 \right\},$$

where $d(\lambda) := \min_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda)$ and $\rho > 0$ is now the proximal stepsize.

Motivation

In practice, the ALM primal update is computed approximately (usually by ACG)

$$x_{k+1} \approx \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \mathcal{L}_\rho(u, \lambda_k).$$

Hence, referred to asI-ALM. It is thus natural to suppose that I-ALM is equivalent to an IPP method in the dual

$$\lambda_{k+1} \approx \underset{\lambda \in \mathbb{R}^m}{\operatorname{argmax}} \left\{ d(\lambda) - \frac{1}{2\rho} \|\lambda - \lambda_k\|^2 \right\},$$

for some suitable definition of "inexactness".

Current convergence analyses of I-ALM do not take either this perspective or the full advantage of powerful modern IPP frameworks.

Goal: Develop a straightforward analysis of I-ALM via IPP framework, further propose a "dual accelerated" I-ALM in spirit of Restarted ACG.

Preliminary: Gradient Mapping

For CSCO objective $\phi(x) := f(x) + h(x)$, define the **gradient mapping** as

$$\mathcal{G}_\phi^\lambda(x) = \frac{1}{\lambda}(x - \mathrm{prox}_{\lambda h}(x - \lambda \nabla f(x))) = \frac{x - x^+}{\lambda}.$$

Crucially: $\|\mathcal{G}_\phi^\lambda(x)\| \leq \varepsilon$ implies

- there exists a subgradient $v \in \partial\phi(x^+)$ satisfying $\|v\| \leq 2\varepsilon$
- if $\mathrm{dom}\, h$ has diameter $D < \infty$, then $\phi(x^+) - \phi_* \leq D\varepsilon$.

Preliminary: ACG I

Problem: Regularized CSCO

$$\min \{\psi(x) := g(x) + h(x) : x \in \mathbb{R}^n\}$$

where

- $g(\cdot)$ is $(L+\mu)$-smooth on $\mathbb{R}^n$ and μ-strongly convex with $\operatorname{dom} g \supset \operatorname{dom} h$
- $h(\cdot)$ is closed proper convex and simple

Goal: ε-small gradient mapping norm
Algorithm: Accelerated Composite Gradient (ACG)

Preliminary: ACG II

Algorithm Accelerated Composite Gradient

Require: given initial point $x_0 \in \operatorname{dom}\psi$, $L \geq 0$, and $\mu \geq 0$, set $A_0 = 0$, $\tau_0 = 1$, and $y_0 = x_0$.

for $j = 0, 1, \cdots$ **do**

1. Compute

$$a_j = \frac{\tau_j + \sqrt{\tau_j^2 + 8\tau_j A_j L}}{4L}, \quad A_{j+1} = A_j + a_j, \quad \tau_{j+1} = \tau_j + \mu a_j,$$

$$\tilde{x}_j = \frac{A_j}{A_{j+1}} y_j + \frac{a_j}{A_{j+1}} x_j.$$

2. Compute

$$\tilde{y}_{j+1} = \operatorname*{argmin}_{u \in \mathbb{R}^n} \left\{ \ell_g(u; \tilde{x}_j) + h(u) + \frac{2L + \mu}{2} \|u - \tilde{x}_j\|^2 \right\},$$

$$y_{j+1} = \operatorname{argmin} \{\psi(y_j), \psi(\tilde{y}_{j+1})\},$$

$$x_{j+1} = \frac{(2L + \mu) a_j \tilde{y}_{j+1} - \frac{2A_j a_j L}{A_{j+1}} y_j}{A_{j+1}\mu + 1}.$$

end for

ACG Convergence Rates

Standard Fact:

$$A_{j+1} \geq \frac{1}{2L} \max \left\{ \frac{(j+1)^2}{4}, \left(1 + \frac{1}{2}\sqrt{\frac{\mu}{2L}}\right)^{2j} \right\}.$$

Lemma

Define $R_0 = \min\{\|x - x_0\| : x \in X_*\}$*, where* X_* *is the set of optimal solutions to* $\min\{\psi(x) := g(x) + h(x) : x \in \mathbb{R}^n\}$*. Then, for all* $j \geq 1$*,*

$$\psi(y_j) - \psi(x_*) \leq \frac{R_0^2}{2A_j},$$

$$\|\mathcal{G}_\psi^{(2L+\mu)^{-1}}(\tilde{x}_{j-1})\| \leq \frac{(2L+\mu)R_0}{\sqrt{LA_j}}.$$

Outline

Lower Oracle Approximation (LOrA) Framework

Algorithm Lower Oracle Approximation (LOrA) Framework

Require: given initial point $x_0^{\mathrm{L}} \in \operatorname{dom}\Phi$, $\sigma_{\mathrm{L}} \in (0,1)$, $\lambda_{\mathrm{L}} > 0$, set $y_0^{\mathrm{L}} = x_0^{\mathrm{L}}$.

for $k = 0, 1, \ldots$ **do**

1. Choose $\delta_k^{\mathrm{L}} > 0$.

2. Find $(y_{k+1}^{\mathrm{L}}, \Gamma_k^{\mathrm{L}}) \in \operatorname{dom}\Phi \times \underset{1/\lambda_{\mathrm{L}}}{\overline{\mathrm{Conv}}}(\operatorname{dom}\Phi)$ such that

$$\Gamma_k^{\mathrm{L}}(\cdot) \le \Phi(\cdot) + \frac{1}{2\lambda_{\mathrm{L}}}\|\cdot - x_k^{\mathrm{L}}\|^2,$$

$$\|\lambda_{\mathrm{L}}\hat{u}_{k+1}^{\mathrm{L}}\|^2 + 2\lambda_{\mathrm{L}}\left[\Phi(y_{k+1}^{\mathrm{L}}) + \frac{1}{2\lambda_{\mathrm{L}}}\|y_{k+1}^{\mathrm{L}} - x_k^{\mathrm{L}}\|^2 - \Gamma_k^{\mathrm{L}}(x_{k+1}^{\mathrm{L}})\right] \le \sigma_{\mathrm{L}}\|y_{k+1}^{\mathrm{L}} - x_k^{\mathrm{L}}\|^2 + 2\lambda_{\mathrm{L}}\delta_k^{\mathrm{L}},$$

where for some $\mathcal{A}_k^{\mathrm{L}} \in (0,\infty]$,

$$x_{k+1}^{\mathrm{L}} = \underset{x\in\mathbb{R}^n}{\operatorname{argmin}}\left\{\Gamma_k^{\mathrm{L}}(x) + \frac{1}{2\mathcal{A}_k^{\mathrm{L}}}\|x - x_k^{\mathrm{L}}\|^2\right\}, \quad \hat{u}_{k+1}^{\mathrm{L}} = \frac{x_k^{\mathrm{L}} - x_{k+1}^{\mathrm{L}}}{\mathcal{A}_k^{\mathrm{L}}}.$$

end for

LOrA Conditions

1. Lower model
$$\Gamma_k^{\mathrm{L}}(\cdot) \le \Phi(\cdot) + \frac{1}{2\lambda_{\mathrm{L}}}\|\cdot - x_k^{\mathrm{L}}\|^2.$$
2. Error bound
$$\|\lambda_{\mathrm{L}}\hat{u}_{k+1}^{\mathrm{L}}\|^2 + 2\lambda_{\mathrm{L}}\left[\Phi(y_{k+1}^{\mathrm{L}}) + \frac{1}{2\lambda_{\mathrm{L}}}\|y_{k+1}^{\mathrm{L}} - x_k^{\mathrm{L}}\|^2 - \Gamma_k^{\mathrm{L}}(x_{k+1}^{\mathrm{L}})\right].$$
$$\le \sigma_{\mathrm{L}}\|y_{k+1}^{\mathrm{L}} - x_k^{\mathrm{L}}\|^2 + 2\lambda_{\mathrm{L}}\delta_k^{\mathrm{L}}$$

Proposition

Let $\hat{x}_*^{\mathrm{L}}$ be the minimizer of the proximal subproblem at iteration k,
$$\hat{x}_*^{\mathrm{L}} = \underset{z\in\mathbb{R}^n}{\operatorname{argmin}}\left\{\Phi(z) + \frac{1}{2\lambda_{\mathrm{L}}}\|z - x_k^{\mathrm{L}}\|^2\right\}.$$
Then, y_{k+1}^{L} obtained by the LOrA framework satisfies
$$\Phi(y_{k+1}^{\mathrm{L}}) + \frac{1}{2\lambda_{\mathrm{L}}}\|y_{k+1}^{\mathrm{L}} - x_k^{\mathrm{L}}\|^2 - \Phi(\hat{x}_*^{\mathrm{L}}) - \frac{1}{2\lambda_{\mathrm{L}}}\|\hat{x}_*^{\mathrm{L}} - x_k^{\mathrm{L}}\|^2 \le \frac{\sigma_{\mathrm{L}}}{2\lambda_{\mathrm{L}}}\|y_{k+1}^{\mathrm{L}} - x_k^{\mathrm{L}}\|^2 + \delta_k^{\mathrm{L}}.$$

LOrA Guarantees

Theorem (LOrA Suboptimality Guarantees)

Let X_* be the set of optimal solutions to $\min_{y\in\mathbb{R}^n}\Phi(y)$. Define $R_0^{\mathrm{L}} := \|x_0^{\mathrm{L}} - x_*\| = \min\{\|x_0^{\mathrm{L}} - x\| : x \in X_*\}$ and $\bar{\delta}_k^{\mathrm{L}} := k^{-1}\sum_{i=0}^{k-1}\delta_i^{\mathrm{L}}$. Suppose that $\mathcal{A}_k^{\mathrm{L}} = \infty$ at all iterations. Then, for every $k \geq 1$, we have

$$\min_{1\leq i\leq k}\|y_i^{\mathrm{L}} - x_{i-1}^{\mathrm{L}}\| \leq \frac{R_0^{\mathrm{L}}}{\sqrt{1-\sigma_{\mathrm{L}}}\sqrt{k}} + \sqrt{\frac{2\lambda_{\mathrm{L}}\bar{\delta}_k^{\mathrm{L}}}{1-\sigma_{\mathrm{L}}}}.$$

Moreover,

$$\min_{1\leq i\leq k}\Phi(y_i^{\mathrm{L}}) - \Phi(x_*) \leq \frac{\left(R_0^{\mathrm{L}}\right)^2}{2\lambda_{\mathrm{L}}k} + \bar{\delta}_k^{\mathrm{L}}.$$

The LHS $\min_{1\leq i\leq k}\Phi(y_i^{\mathrm{L}}) - \Phi(x_*)$ can be replaced with $\Phi(\hat{y}_k^{\mathrm{L}}) - \Phi(x_*)$ where

$$\hat{y}_k^{\mathrm{L}} = \frac{1}{k}\sum_{i=1}^{k} y_i^{\mathrm{L}}.$$

Warm-up: Proximal Gradient Method

Problem: CSCO
Goal: ε-solution
Algorithm: Proximal Gradient Method (PGM)

$$x_{k+1} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \langle \nabla f(x_k), x - x_k \rangle + h(x) + \frac{1}{2\eta} \|x - x_k\|^2 \right\}.$$

Folklore: $\mathcal{O}(\varepsilon^{-1})$ complexity

More examples: proximal subgradient method, mirror descent, extragradient method, primal-dual hybrid gradient method, proximal bundle method, Newton proximal extragradient method, ...

PGM as LOrA

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ \langle \nabla f(x_k), x - x_k \rangle + h(x) + \frac{1}{2\eta} \|x - x_k\|^2 \right\}.$$

$$\Phi(\cdot) = f(\cdot) + h(\cdot), \quad \Gamma_k^{\mathrm{L}}(\cdot) = \ell_f(\cdot; x_k) + h(\cdot) + \frac{1}{2\eta} \| \cdot - x_k \|^2, \quad \lambda_{\mathrm{L}} = \eta;$$
$$\sigma_{\mathrm{L}} = \eta L_f, \quad \mathcal{A}_k^{\mathrm{L}} = \infty, \quad \delta_k^{\mathrm{L}} = 0, \quad y_{k+1}^{\mathrm{L}} = x_{k+1}^{\mathrm{L}} = x_{k+1}, \quad \hat{u}_{k+1}^{\mathrm{L}} = 0. \tag{1}$$

1. Lower Model: $\ell_f(\cdot; x_k) + h(\cdot) + \frac{1}{2\eta} \| \cdot - x_k \|^2 \leq f(\cdot) + h(\cdot) + \frac{1}{2\eta} \| \cdot - x_k \|^2$
2. Error Bound: By L_f-smoothness and $\eta \leq 1/2L_f$:

$$2\lambda_{\mathrm{L}} \left[\Phi(y_{k+1}^{\mathrm{L}}) + \frac{1}{2\lambda_{\mathrm{L}}} \|y_{k+1}^{\mathrm{L}} - x_k^{\mathrm{L}}\|^2 - \Gamma_k^{\mathrm{L}}(x_{k+1}^{\mathrm{L}}) \right]$$
$$\overset{(1)}{=} 2\eta [f(x_{k+1}) - \ell_f(x_{k+1}; x_k)] \leq \eta L_f \|x_{k+1} - x_k\|^2 \overset{(1)}{=} \sigma_{\mathrm{L}} \|y_{k+1}^{\mathrm{L}} - x_k^{\mathrm{L}}\|^2.$$

$\mathcal{O}(\varepsilon^{-1})$ **complexity** ✓

Dual Application: Inexact Augmented Lagrangian Method

Problem: LC-CSCO
Goal: ε-primal-dual solution
Algorithm: I-ALM

$$x_{k+1} \approx \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ \mathcal{L}_\rho(x, \lambda_k) := \phi(x) + \langle \lambda_k, Ax - b \rangle + \frac{\rho}{2} \|Ax - b\|^2 \right\}$$

$$\lambda_{k+1} = \lambda_k + \rho(Ax_{k+1} - b)$$

Complexity: ?

Existing I-ALM Complexity Results (and Preview)

Paper	Alg.	Complexity	ρ	ϕ	Conv. Pt.
Lan and Monteiro [2016]	I-ALM	$\mathcal{O}(\varepsilon^{-7/4})$	Static	$f + \delta_Q$	Non-Erg.
Patrascu et al. [2017]	IFAL	$\mathcal{O}(\varepsilon^{-1})$	Static	$f + \delta_Q$	Erg.
Liu et al. [2019]	I-ALM	$\mathcal{O}(\varepsilon^{-2})$	Static	$f + h$	Non-Erg.
Xu [2021]	I-ALM	$\mathcal{O}(\varepsilon^{-1})$	Geo.	$f + h$	Non-Erg.
	I-ALM	$\mathcal{O}(\varepsilon^{-1})$	St./Geo.	$f + h$	Erg.
Lu and Zhou [2023]	aI-ALM	$\tilde{\mathcal{O}}(\varepsilon^{-1})$	Geo.	$f + h$	Non-Erg.
Li and Lin [2019]	LPALM	$\mathcal{O}(\varepsilon^{-1})$	Static	$f + h$	Non-Erg.
This work	**I-ALM**	$\tilde{\mathcal{O}}(\varepsilon^{-1})$	Static	$f + h$	Non-Erg.
This work	**I-FALM**	$\tilde{\mathcal{O}}(\varepsilon^{-1})$	Static	$f + h$	Non-Erg.

Existing I-ALM Complexity Results (and Preview)

Paper	Alg.	Complexity	ρ	ϕ	Conv. Pt.
Lan and Monteiro [2016]	I-ALM	$\mathcal{O}(\varepsilon^{-7/4})$	Static	$f + \delta_Q$	Non-Erg.
Patrascu et al. [2017]	IFAL	$\mathcal{O}(\varepsilon^{-1})$	Static	$f + \delta_Q$	Erg.
Liu et al. [2019]	I-ALM	$\mathcal{O}(\varepsilon^{-2})$	Static	$f + h$	Non-Erg.
Xu [2021]	I-ALM	$\mathcal{O}(\varepsilon^{-1})$	Geo.	$f + h$	Non-Erg.
	I-ALM	$\mathcal{O}(\varepsilon^{-1})$	St./Geo.	$f + h$	Erg.
Lu and Zhou [2023]	aI-ALM	$\tilde{\mathcal{O}}(\varepsilon^{-1})$	Geo.	$f + h$	Non-Erg.
Li and Lin [2019]	LPALM	$\mathcal{O}(\varepsilon^{-1})$	Static	$f + h$	Non-Erg.
This work	**I-ALM**	$\tilde{\mathcal{O}}(\varepsilon^{-1})$	Static	$f + h$	Non-Erg.
This work	**I-FALM**	$\tilde{\mathcal{O}}(\varepsilon^{-1})$	Static	$f + h$	Non-Erg.

Background: I-ALM as IPPM I

Exact ALM

Primal

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathbb{R}^n} \mathcal{L}_\rho(x, \lambda_k)$$

$$\lambda_{k+1} = \lambda_k + \rho(Ax_{k+1} - b)$$

Dual

$$\lambda_{k+1} = \operatorname*{argmax}_{\lambda \in \mathbb{R}^m} \left\{ d(\lambda) - \frac{1}{2\rho} \|\lambda - \lambda_k\|^2 \right\}$$

Inexact ALM

Primal

$$x_{k+1} \approx \operatorname*{argmin}_{x \in \mathbb{R}^n} \mathcal{L}_\rho(x, \lambda_k)$$

$$\lambda_{k+1} = \lambda_k + \rho(Ax_{k+1} - b)$$

Dual

$$\lambda_{k+1} \approx \operatorname*{argmax}_{\lambda \in \mathbb{R}^m} \left\{ d(\lambda) - \frac{1}{2\rho} \|\lambda - \lambda_k\|^2 \right\}$$

Background: I-ALM as IPPM II

Proposition (Theorem 4 of Rockafellar [1976])

Fix $\lambda_0 \in \mathbb{R}^m$, $x \in \operatorname{dom} h$ *and define*

$$\lambda_*^+ = \operatorname*{argmax}_{\lambda \in \mathbb{R}^m} \left\{ d(\lambda) - \frac{1}{2\rho} \|\lambda - \lambda_0\|^2 \right\},$$
$$\lambda^+ = \lambda_0 + \rho(Ax - b).$$

Then

$$\frac{1}{2\rho} \|\lambda^+ - \lambda_*^+\|^2 \le \mathcal{L}_\rho(x, \lambda_0) - \min_{y \in \mathbb{R}^n} \mathcal{L}_\rho(y, \lambda_0).$$

- Enables absolute error IPPM, requires $\rho_k \to \infty$ for $\mathcal{O}(\varepsilon^{-1})$ nonergodic complexity [Xu, 2021, Lu and Zhou, 2023].
- Can we obtain $\mathcal{O}(\varepsilon^{-1})$ nonergodic complexity with constant ρ?
- Can we use **relative error** IPPM?

I-ALM

Algorithm Inexact Augmented Lagrangian Method

Require: given initial point $x_0 \in \operatorname{dom} h$, $\rho > 0$, $\varepsilon_0 > 0$, $\alpha \in (0,1)$, $\varepsilon > 0$, set $\lambda_0 = 0$, and choose $\sigma \in (0,1)$ such that $2\sigma\rho \le D/\varepsilon$.

for $k = 0, 1, \cdots$ **do**

1. Define

$$\varepsilon_k = (\varepsilon_0 \alpha^k + \sigma\rho\varepsilon^2)/2$$

and use ACG to find a $\tilde{x}_k$ satisfying $\|\mathcal{G}^{(2L+\mu)^{-1}}_{\mathcal{L}_\rho(\cdot,\lambda_k)}(\tilde{x}_k)\| \le \varepsilon_k/(2D)$ and set

$$x_{k+1} = \tilde{x}_k - (2L+\mu)^{-1}\mathcal{G}^{(2L+\mu)^{-1}}_{\mathcal{L}_\rho(\cdot,\lambda_k)}(\tilde{x}_k).$$

2. Compute $\lambda_{k+1} = \lambda_k + \rho(Ax_{k+1} - b)$.

3. If $\|\mathcal{G}^{(2L+\mu)^{-1}}_{\mathcal{L}_\rho(\cdot,\lambda_k)}(\tilde{x}_k)\| \le \varepsilon$ and $\|Ax_{k+1} - b\| \le \varepsilon$, then **return** (x_{k+1}, λ_{k+1}).

end for

Inner/Outer Perspective

Inner Complexity (Proved by standard ACG arguments [Nesterov, 2013])

Proposition

Each call to ACG in Step 1 of I-ALM requires

$$\tilde{\mathcal{O}}\left(1 + \frac{D(\sqrt{L_f} + \sqrt{\rho}\|A\|)}{\sqrt{\sigma\rho\varepsilon}}\right)$$

ACG iterations.

Outer Complexity

Proposition

I-ALM terminates in

$$\mathcal{O}\left(1 + \frac{R_\Lambda^2 + \rho\varepsilon}{(1-\sigma)\rho^2\varepsilon^2}\right)$$

outer iterations, where $R_\Lambda = \min\{\|\lambda_*\| : d(\lambda_*) = \hat{\phi}_*\}$.

Proof by reduction to LOrA

I-ALM as LOrA

$$\Phi(\cdot) = -d(\cdot), \quad \Gamma_k^{\mathrm{L}}(\cdot) = -\mathcal{L}(x_{k+1}, \cdot) + \frac{1}{2\rho}\| \cdot - \lambda_k\|^2, \quad \mu_{\mathrm{L}} = 0, \quad \lambda_{\mathrm{L}} = \rho, \quad \sigma_{\mathrm{L}} = \sigma;$$
$$\mathcal{A}_k^{\mathrm{L}} = \infty, \quad \delta_k^{\mathrm{L}} = \varepsilon_0 \alpha^k, \quad y_k^{\mathrm{L}} = x_k^{\mathrm{L}} = \lambda_k, \quad \hat{u}_{k+1}^{\mathrm{L}} = 0, \quad \alpha_{\mathrm{L}} = \alpha.$$

1. **Lower Model:** By definition for all $\lambda \in \mathbb{R}^m$ we have

$$-\mathcal{L}(x_{k+1}, \lambda) \leq -\operatorname*{argmin}_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda) = -d(\lambda)$$

2. **Exact minimizer:** By definition $\lambda_{k+1} = \operatorname{argmin}_{\lambda \in \mathbb{R}^m}\{-\mathcal{L}(x_{k+1}, \lambda) + \frac{1}{2\rho}\|\lambda - \lambda_k\|^2\}$
3. **Error bound:** ?

I-ALM as LOrA

$$\Phi(\cdot) = -d(\cdot), \quad \Gamma_k^{\mathrm{L}}(\cdot) = -\mathcal{L}(x_{k+1}, \cdot) + \frac{1}{2\rho}\|\cdot - \lambda_k\|^2, \quad \mu_{\mathrm{L}} = 0, \quad \lambda_{\mathrm{L}} = \rho, \quad \sigma_{\mathrm{L}} = \sigma;$$
$$\mathcal{A}_k^{\mathrm{L}} = \infty, \quad \delta_k^{\mathrm{L}} = \varepsilon_0 \alpha^k, \quad y_k^{\mathrm{L}} = x_k^{\mathrm{L}} = \lambda_k, \quad \hat{u}_{k+1}^{\mathrm{L}} = 0, \quad \alpha_{\mathrm{L}} = \alpha.$$

1. **Lower Model:** By definition for all $\lambda \in \mathbb{R}^m$ we have

$$-\mathcal{L}(x_{k+1}, \lambda) \leq -\operatorname*{argmin}_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda) = -d(\lambda)$$

2. **Exact minimizer:** By definition $\lambda_{k+1} = \operatorname{argmin}_{\lambda \in \mathbb{R}^m}\{-\mathcal{L}(x_{k+1}, \lambda) + \frac{1}{2\rho}\|\lambda - \lambda_k\|^2\}$
3. **Error bound:** ?

I-ALM as LOrA

$$\Phi(\cdot) = -d(\cdot), \quad \Gamma_k^{\mathrm{L}}(\cdot) = -\mathcal{L}(x_{k+1}, \cdot) + \frac{1}{2\rho}\|\cdot - \lambda_k\|^2, \quad \mu_{\mathrm{L}} = 0, \quad \lambda_{\mathrm{L}} = \rho, \quad \sigma_{\mathrm{L}} = \sigma;$$
$$\mathcal{A}_k^{\mathrm{L}} = \infty, \quad \delta_k^{\mathrm{L}} = \varepsilon_0 \alpha^k, \quad y_k^{\mathrm{L}} = x_k^{\mathrm{L}} = \lambda_k, \quad \hat{u}_{k+1}^{\mathrm{L}} = 0, \quad \alpha_{\mathrm{L}} = \alpha.$$

1. **Lower Model:** By definition for all $\lambda \in \mathbb{R}^m$ we have
$$-\mathcal{L}(x_{k+1}, \lambda) \leq -\operatorname*{argmin}_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda) = -d(\lambda)$$
2. **Exact minimizer:** By definition
$\lambda_{k+1} = \operatorname{argmin}_{\lambda \in \mathbb{R}^m} \{-\mathcal{L}(x_{k+1}, \lambda) + \frac{1}{2\rho}\|\lambda - \lambda_k\|^2\}$
3. **Error bound: ?**

I-ALM as LOrA

$$\Phi(\cdot) = -d(\cdot), \quad \Gamma_k^{\mathrm{L}}(\cdot) = -\mathcal{L}(x_{k+1}, \cdot) + \frac{1}{2\rho}\|\cdot - \lambda_k\|^2, \quad \mu_{\mathrm{L}} = 0, \quad \lambda_{\mathrm{L}} = \rho, \quad \sigma_{\mathrm{L}} = \sigma;$$
$$\mathcal{A}_k^{\mathrm{L}} = \infty, \quad \delta_k^{\mathrm{L}} = \varepsilon_0 \alpha^k, \quad y_k^{\mathrm{L}} = x_k^{\mathrm{L}} = \lambda_k, \quad \hat{u}_{k+1}^{\mathrm{L}} = 0, \quad \alpha_{\mathrm{L}} = \alpha.$$

1. **Lower Model:** By definition for all $\lambda \in \mathbb{R}^m$ we have
$$-\mathcal{L}(x_{k+1}, \lambda) \leq -\operatorname*{argmin}_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda) = -d(\lambda)$$
2. **Exact minimizer:** By definition
$\lambda_{k+1} = \mathrm{argmin}_{\lambda \in \mathbb{R}^m} \{-\mathcal{L}(x_{k+1}, \lambda) + \frac{1}{2\rho}\|\lambda - \lambda_k\|^2\}$
3. **Error bound: ?**

I-ALM as LOrA: Error Bound

Condition to prove:

$$-d(\lambda_{k+1}) + \frac{1}{2\rho}\|\lambda_{k+1} - \lambda_k\|^2 - \Gamma_k^{\mathrm{L}}(\lambda_{k+1}) \leq \varepsilon_0 \alpha^k + \frac{\sigma}{2\rho}\|\lambda_{k+1} - \lambda_k\|^2, \qquad (2)$$

3 (Non-Exclusive) Case Proof

Case I) If $\|Ax_{k+1} - b\| \geq \varepsilon$, then (2) holds.

Case II) If $\|\mathcal{G}_{\mathcal{L}_\rho(\cdot,\lambda_k)}^{(2L+\mu)^{-1}}(\tilde{x}_k)\| \geq \varepsilon/2$, then (2) holds (using $2\sigma\rho \leq D/\varepsilon$).

Case III) If neither I) or II) hold, then (x_{k+1}, λ_{k+1}) is an ε-primal-dual solution, and I-ALM terminates.

I-ALM Complexity

Theorem

Given $\varepsilon > 0$, we choose $\varepsilon_0 = \varepsilon$, $\sigma = 1/2$, and $\rho = \varepsilon^{-1}$. Then, I-ALM finds an ε-primal-dual solution to the LC-CSCO problem in

$$\tilde{\mathcal{O}}\left((1+R_\Lambda^2)\left(1+D\left(\frac{\sqrt{L_f}}{\sqrt{\varepsilon}}+\frac{\|A\|}{\varepsilon}\right)\right)\right)$$

ACG iterations, where $R_\Lambda = \|\lambda_* - \lambda_0\| = \min\{\|\lambda - \lambda_0\| : \lambda \in \Lambda_*\}$, where Λ_* is the set of maximizers to the dual problem.

Corollary

Under the conditions and parameter choices of the above theorem, I-ALM finds an ε-primal solution to the LC-CSCO problem in

$$\tilde{\mathcal{O}}\left((1+R_\Lambda^2)\left(1+D\left(\frac{\sqrt{(R_\Lambda+D)L_f}}{\sqrt{\varepsilon}}+\frac{(R_\Lambda+D)\|A\|}{\varepsilon}\right)\right)\right)$$

ACG iterations.

Comparison with Previous Work

$$\tilde{\mathcal{O}}\left((1+R_\Lambda^2)\left(1+D\left(\frac{\sqrt{(R_\Lambda+D)L_f}}{\sqrt{\varepsilon}}+\frac{(R_\Lambda+D)\|A\|}{\varepsilon}\right)\right)\right)$$

Lower Bound: $\Omega\left(D\sqrt{\frac{L_f}{\varepsilon}}+\frac{\max\{R_\Lambda D,\|A\|\}}{\varepsilon}\right)$ [Ouyang and Xu, 2021]

- Near-optimal in ε, L_f, $\|A\|$, suboptimal in D, R_Λ
- Gradient mapping termination
  - Efficiently computable
  - Removes two-phase structure used by Lan and Monteiro [2016] by utilizing gradient mapping for inner termination

Outline

Combine LOrA with Acceleration

Algorithm FLOrA Framework

Require: given initial point $x_0^{\mathrm{F}} \in \mathrm{dom}\,\Phi$, $\mu_{\mathrm{F}} \geq 0$, $\sigma_{\mathrm{F}} \in (0, 1]$, $\lambda_{\mathrm{F}} > 0$, $\tau_0 = 1$, $\delta_0^{\mathrm{F}} \geq 0$, set $B_0 = 0$ and $y_0^{\mathrm{F}} = x_0^{\mathrm{F}}$ and choose an $\alpha_{\mathrm{F}} \geq 0$ satisfying $\alpha_{\mathrm{F}} < (1 + \sqrt{\lambda_{\mathrm{F}} \mu_{\mathrm{F}}})^{-2}$.

for $k = 0, 1, \ldots$ **do**

1. Set $\delta_k^{\mathrm{F}} = \delta_0^{\mathrm{F}} (\alpha_{\mathrm{F}})^k$ and compute

$$b_k = \frac{\lambda_{\mathrm{F}} \tau_k + \sqrt{\lambda_{\mathrm{F}}^2 \tau_k^2 + 4 \lambda_{\mathrm{F}} \tau_k B_k}}{2}, \quad B_{k+1} = B_k + b_k, \quad \tau_{k+1} = \tau_k + b_k \mu_{\mathrm{F}},$$

$$\tilde{x}_k^{\mathrm{F}} = \frac{B_k}{B_{k+1}} y_k^{\mathrm{F}} + \frac{b_k}{B_{k+1}} x_k^{\mathrm{F}}.$$

2. Find $(\tilde{y}_{k+1}^{\mathrm{F}}, z_{k+1}, \Gamma_k^{\mathrm{F}}) \in \mathrm{dom}\,\Phi \times \mathrm{dom}\,\Phi \times \overline{\mathrm{Conv}}_{\mu_{\mathrm{F}} + \lambda_{\mathrm{F}}^{-1}} (\mathrm{dom}\,\Phi)$ such that the FLOrA conditions are satisfied.

3. Choose y_{k+1}^{F} satisfying $\Phi(y_{k+1}^{\mathrm{F}}) \leq \Phi(\tilde{y}_{k+1}^{\mathrm{F}})$ and compute

$$u_{k+1}^{\mathrm{F}} = \hat{u}_{k+1}^{\mathrm{F}} + \frac{\tilde{x}_k^{\mathrm{F}} - z_{k+1}^{\mathrm{F}}}{\lambda_{\mathrm{F}}}, \quad x_{k+1}^{\mathrm{F}} = \frac{1}{\tau_{k+1}} \left(\tau_k x_k^{\mathrm{F}} + b_k \mu_{\mathrm{F}} z_{k+1}^{\mathrm{F}} - b_k u_{k+1}^{\mathrm{F}} \right).$$

end for

FLOrA Conditions

Find $(\tilde{y}_{k+1}^{\mathrm{F}}, z_{k+1}, \Gamma_k^{\mathrm{F}}) \in \mathrm{dom}\,\Phi \times \mathrm{dom}\,\Phi \times \overline{\mathsf{Conv}}_{\mu_{\mathrm{F}}+\lambda_{\mathrm{F}}^{-1}}(\mathrm{dom}\,\Phi)$ such that the following hold:

FLOrA Conditions

$$\Gamma_k^{\mathrm{F}}(\cdot) \le \Phi(\cdot) + \frac{1}{2\lambda_{\mathrm{F}}}\|\cdot - \tilde{x}_k^{\mathrm{F}}\|^2,$$

$$\|\lambda_{\mathrm{F}}\hat{u}_{k+1}^{\mathrm{F}}\|^2 + 2\lambda_{\mathrm{F}}\left[\Phi(\tilde{y}_{k+1}^{\mathrm{F}}) + \frac{1}{2\lambda_{\mathrm{F}}}\|\tilde{y}_{k+1}^{\mathrm{F}} - \tilde{x}_k^{\mathrm{F}}\|^2 - \Gamma_k^{\mathrm{F}}(z_{k+1}^{\mathrm{F}})\right] \le \sigma_{\mathrm{F}}\|\tilde{y}_{k+1}^{\mathrm{F}} - \tilde{x}_k^{\mathrm{F}}\|^2 + 2\lambda_{\mathrm{F}}\delta_k^{\mathrm{F}},$$

where for some $\mathcal{A}_k^{\mathrm{F}} \in (0, \infty]$,

$$z_{k+1}^{\mathrm{F}} = \underset{v\in\mathbb{R}^n}{\mathrm{argmin}}\left\{\Gamma_k^{\mathrm{F}}(v) + \frac{1}{2\mathcal{A}_k^{\mathrm{F}}}\|v - \tilde{x}_k^{\mathrm{F}}\|^2\right\}, \quad \hat{u}_{k+1}^{\mathrm{F}} = \frac{\tilde{x}_k^{\mathrm{F}} - z_{k+1}^{\mathrm{F}}}{\mathcal{A}_k^{\mathrm{F}}}.$$

FLOrA Complexity

Theorem

Let X_* *be the set of optimal solutions to* $\min_x \Phi(x)$. *Define* $R_0^{\mathrm{F}} := \|x_0^{\mathrm{F}} - x_*\| = \min\{\|x_0^{\mathrm{F}} - x\| : x \in X_*\}$. *Then, for every* $k \geq 0$,

$$\Phi(y_{k+1}^{\mathrm{F}}) - \Phi_* \leq \frac{\left(R_0^{\mathrm{F}}\right)^2}{2B_{k+1}} + \frac{\delta_0^{\mathrm{F}} C_{\mathrm{F}}}{B_{k+1}},$$

where $C_{\mathrm{F}} := \sum_{i=0}^{\infty} B_{i+1}(\alpha_{\mathrm{F}})^i < \infty$. *Furthermore, if* $\sigma_{\mathrm{F}} < 1$, *then for every* $k \geq 0$, *we have*

$$\|\tilde{y}_{k+1}^{\mathrm{F}} - \tilde{x}_k^{\mathrm{F}}\| \leq \frac{\sqrt{\lambda_{\mathrm{F}}} R_0^{\mathrm{F}} + \sqrt{2\lambda_{\mathrm{F}} \delta_0^{\mathrm{F}} C_{\mathrm{F}}}}{\sqrt{(1-\sigma_{\mathrm{F}}) B_{k+1}}},$$

$$\min_{0 \leq i \leq k} \|\tilde{y}_{i+1}^{\mathrm{F}} - \tilde{x}_i^{\mathrm{F}}\| \leq \frac{\sqrt{\lambda_{\mathrm{F}}} R_0^{\mathrm{F}} + \sqrt{2\lambda_{\mathrm{F}} \delta_0^{\mathrm{F}} C_{\mathrm{F}}}}{\sqrt{(1-\sigma_{\mathrm{F}}) \sum_{i=1}^{k+1} B_i}}.$$

$$B_k \geq \lambda_{\mathrm{F}} \max\left\{\frac{k^2}{4}, \left(1 + \frac{\sqrt{\lambda_{\mathrm{F}} \mu_{\mathrm{F}}}}{2}\right)^{2(k-1)}\right\}$$

Primal: Restarted ACG

Problem: CSCO. Additionally assume

- f is μ_f-strongly convex with $\mu_f \geq 0$
- $L_f \geq 2\mu_f$

Goal: ε-solution

Algorithm: Restarted ACG

Complexity (Preview):

$$\tilde{\mathcal{O}}\left(\min\left\{\sqrt{\frac{L_f}{\mu_f}}, \frac{\|x_0 - x_*\|\sqrt{L_f}}{\sqrt{\varepsilon}}\right\}\right)$$

Restarted ACG

Algorithm Restarted ACG

Require: given initial point $w_0 \in \operatorname{dom} h$, $\sigma \in (0,1)$, $L_f \geq 0$, $\mu_f \geq 0$, and $\lambda > 0$, set $B_0 = 0$, $\tau_0 = 1$, and $v_0 = w_0$.

for $k = 0, 1, \cdots$ **do**

1. Compute

$$b_k = \frac{\tau_k \lambda + \sqrt{\tau_k^2 \lambda^2 + 4\tau_k \lambda B_k}}{2}, \quad B_{k+1} = B_k + b_k, \quad \tau_{k+1} = \tau_k + b_k \mu_f,$$

$$\tilde{v}_k = \frac{B_k}{B_{k+1}} w_k + \frac{b_k}{B_{k+1}} v_k.$$

2. Call ACG with the proximal subproblem and perform j iterations until

$$\frac{\lambda}{A_j} \|y_j - \tilde{v}_k\|^2 + 2\lambda[\psi(y_j) - \Theta_j(x_j)] \leq \sigma \|y_j - \tilde{v}_k\|^2,$$

where y_j and x_j are the ACG iterates, respectively, and Θ_j and A_j^{-1} are the ACG estimating sequences.

3. Choose $w_{k+1} \in \operatorname{Argmin} \{\phi(u) : u \in \{w_k, y_j\}\}$ and compute

$$v_{k+1} = \frac{1}{\tau_{k+1}} \left(\tau_k v_k + b_k \mu_f x_j - b_k \frac{A_j + \lambda}{\lambda A_j} (\tilde{v}_k - y_j) \right).$$

end for

Inner Complexity

Proposition

Assume that $\lambda \geq 1/(L_f - \mu_f)$. *Then in each call to ACG in Step* ***2*** *of Restarted ACG, after at most*

$$1 + \left\lceil \min\left\{ 2\sqrt{10\sigma^{-1}\lambda(L_f - \mu_f)}, \left(\frac{1}{4} + \frac{1}{2}\sqrt{\frac{2\lambda(L_f - \mu_f)}{1 + \lambda\mu_f}}\right) \ln\left(10\sigma^{-1}\lambda(L_f - \mu_f)\right) \right\} \right\rceil .$$

ACG iterations, the termination condition is satisfied.

Restarted ACG as FLOrA

$$\Phi(\cdot) = \phi(\cdot), \quad \Gamma_k^{\mathrm{F}}(\cdot) = \Theta_j(\cdot), \quad \mathcal{A}_k^{\mathrm{F}} = A_j, \quad \delta_k^{\mathrm{F}} = \alpha_{\mathrm{F}} = 0, \quad \mu_{\mathrm{F}} = \mu_f;$$
$$\sigma_{\mathrm{F}} = \sigma, \ \lambda_{\mathrm{F}} = \lambda, y_k^{\mathrm{F}} = w_k, \ x_k^{\mathrm{F}} = v_k, \ \tilde{x}_k^{\mathrm{F}} = \tilde{v}_k;$$
$$\tilde{y}_{k+1}^{\mathrm{F}} = y_j, \ z_{k+1}^{\mathrm{F}} = x_j, \ u_{k+1}^{\mathrm{F}} = \frac{A_j + \lambda}{A_j \lambda}(\tilde{v}_k - y_j), \ \hat{u}_{k+1}^{\mathrm{F}} = \frac{\tilde{v}_k - y_j}{A_j}.$$

Proposition

For every $k \geq 1$, the function value gap $\phi(w_k) - \phi_*$ satisfies

$$\phi(w_k) - \phi_* \leq \min\left\{ \frac{2R_0^2}{\lambda k^2}, \quad \frac{R_0^2}{2\lambda}\left(1 + \frac{\sqrt{\lambda \mu_f}}{2}\right)^{-2(k-1)} \right\},$$

where R_0 denotes the distance from initial point w_0 to solution set X_*, i.e.,

$$R_0 = \|w_0 - x_*\| = \min\{\|w_0 - x\| : x \in X_*\}.$$

Restarted ACG Complexity

Theorem

For given $\varepsilon > 0$, the following statements hold:

a) *if $\mu_f = 0$ and $1/L_f \leq \lambda \leq R_0^2/\varepsilon$, then the total iteration-complexity of Restarted ACG to find an ε-solution is $\tilde{\mathcal{O}}\left(R_0\sqrt{L_f/\varepsilon}\right)$;*

b) *if $\mu_f > 0$ and $1/(L_f - \mu_f) \leq \lambda \leq \min\{1/\mu_f, R_0^2/\varepsilon\}$, then the total iteration-complexity of Restarted ACG to find an ε-solution is $\tilde{\mathcal{O}}(\min\{\sqrt{L_f/\mu_f}, R_0\sqrt{L_f/\varepsilon}\})$.*

- For $\lambda \leq 1/(L_f - \mu_f)$, the number of inner ACG iterations is $\mathcal{O}(1)$.
- If λ sufficiently small, Restarted ACG reduces to ACG.

Dual: Inexact Fast Augmented Lagrangian Method

Problem: LC-CSCO
Goal: ε-primal-dual solution
Algorithm: Inexact Fast Augmented Lagrangian Method (I-FALM)
Complexity (Preview):

$$\tilde{\mathcal{O}}\left(1 + \frac{\sqrt{D^2 + D\hat{R}_\Lambda}\|A\|}{\varepsilon} + \frac{\sqrt{D + \hat{R}_\Lambda}\|A\|}{\sqrt{L_f \varepsilon}} + \frac{\sqrt{DL_f}}{\sqrt{\varepsilon}}\right)$$

1. Add primal/dual perturbations

$$\tilde{\mathcal{L}}_\rho(x, \lambda) := \phi(x) + \frac{\gamma_p}{2}\|x - x_0\|^2 + \langle \lambda, Ax - b\rangle + \frac{\rho}{2}\|Ax - b\|^2 - \frac{\gamma_d}{2}\|\lambda - \lambda_0\|^2.$$

2. Apply FLOrA acceleration to outer loop

I-FALM

Algorithm Inexact Fast Augmented Lagrangian Method

Require: given initial $x_0 \in \operatorname{dom} h$, $\rho > 0$, $\gamma_d > 0$, $\varepsilon > 0$, and $\varepsilon_0 \geq \varepsilon$, set $B_0 = 0$, $\tau_0 = 1$, $\gamma_p = \varepsilon/(2D)$, and $\lambda_0 = \nu_0 = 0$, and choose $\sigma \in (0,1)$ such that $4\sigma\rho\varepsilon \leq 1$, and $\alpha \geq 0$ satisfying $\alpha < (1 + \sqrt{\gamma_d \rho})^{-2}$.

for $k = 0, 1, \cdots$ **do**

1. Set $\varepsilon_k = (7\varepsilon_0 \alpha^k + \sigma\rho\varepsilon^2)/8$ and compute

$$b_k = \frac{\rho\tau_k + \sqrt{\rho^2\tau_k^2 + 4\rho\tau_k B_k}}{2}, \quad B_{k+1} = B_k + b_k, \quad \tau_{k+1} = \tau_k + b_k\gamma_d;$$

$$\tilde{\nu}_k = \frac{B_k}{B_{k+1}}\lambda_k + \frac{b_k}{B_{k+1}}\nu_k.$$

2. Use ACG to find a $\tilde{x}_k$ satisfying $\|\mathcal{G}^{(2L+\mu)^{-1}}_{\mathcal{L}_\rho(\cdot,\tilde{\nu}_k)}(\tilde{x}_k)\| \leq \varepsilon_k/(2D)$ and set

$$x_{k+1} = \tilde{x}_k - (2L+\mu)^{-1}\mathcal{G}^{(2L+\mu)^{-1}}_{\mathcal{L}_\rho(\cdot,\tilde{\nu}_k)}(\tilde{x}_k).$$

3. Compute $\lambda_{k+1} = \tilde{\nu}_k + \rho(Ax_{k+1} - b)$.

4. If $\|v\| \leq \varepsilon/2$ and $\|Ax_{k+1} - b\| \leq \varepsilon$, then return (x_{k+1}, λ_{k+1}) and **terminate**; otherwise, compute

$$\nu_{k+1} = \frac{1}{\tau_{k+1}}\left(\tau_k\nu_k + b_k\gamma_d\frac{\lambda_{k+1}}{1+\gamma_d\rho} - \frac{b_k}{\rho}\left(\tilde{\nu}_k - \frac{\lambda_{k+1}}{1+\gamma_d\rho}\right)\right),$$

and continue.

end for

Inner/Outer Perspective

Proposition (Inner Complexity)

For $\gamma_p = \varepsilon/2D$, each call to ACG in Step 2 of I-FALM requires

$$\tilde{\mathcal{O}}\left(1 + \frac{\sqrt{D}(\sqrt{L_f} + \sqrt{\rho}\|A\|)}{\sqrt{\varepsilon}}\right).$$

ACG iterations[a].

[a] Reduced from $\tilde{\mathcal{O}}(\varepsilon^{-1})$ to $\tilde{\mathcal{O}}(\varepsilon^{-1/2})$ by primal γ_p-perturbation.

Proposition (Outer Complexity)

For $\gamma_d = \mathcal{O}(\varepsilon/(\hat{R}_\Lambda + D))$, I-FALM terminates in

$$\tilde{\mathcal{O}}\left(1 + \frac{\sqrt{\hat{R}_\Lambda + D}}{\sqrt{\rho\varepsilon}}\right)$$

outer iterations, where $\hat{R}_\Lambda = \max\{1, \min\{\|\lambda_*\| : d(\lambda_*) = \hat{\phi}_*\}\}$[a].

[a] Reduced from $\mathcal{O}(\rho^{-2}\varepsilon^{-2})$ to $\tilde{\mathcal{O}}(\rho^{-1/2}\varepsilon^{-1/2})$ by dual γ_d-perturbation.

I-FALM Complexity

Theorem (I-FALM Primal-Dual Complexity (Informal))

Let $\varepsilon > 0$ *satisfy* $\varepsilon \leq \|A\|^2/L_f$. *Choose* $\rho = L_f/\|A\|^2$, $\varepsilon_0 = \rho^{-1}$, $\sigma = 1/4$, $\gamma_p = \varepsilon/(2D)$, $\gamma_d = \mathcal{O}(\varepsilon/(\hat{R}_\Lambda + D))$. *Then, Algorithm 6 finds an* ε*-primal-dual solution to the LC-CSCO problem in*

$$\tilde{\mathcal{O}}\left(1 + \frac{\sqrt{D^2 + D\hat{R}_\Lambda}\|A\|}{\varepsilon} + \frac{\sqrt{D + \hat{R}_\Lambda}\|A\|}{\sqrt{L_f \varepsilon}} + \frac{\sqrt{DL_f}}{\sqrt{\varepsilon}}\right)$$

total ACG iterations, where $\hat{R}_\Lambda = \max\{1, \|\lambda_* - \lambda_0\|\}$, *with* $\lambda_* = \operatorname{argmin}\{\|\lambda - \lambda_0\| : \lambda \in \Lambda_*\}$ *and* Λ_* *is the set of maximizers to the dual problem.*

Corollary (I-FALM Primal Complexity (Informal))

Let $\varepsilon_g > 0$ *satisfy* $\varepsilon_g \leq 2\|A\|^2(D + \hat{R}_\Lambda)/L_f$. *Then, using the parameters above with* $\varepsilon = \varepsilon_g/(2(D + \hat{R}_\Lambda))$, *I-FALM finds an* ε_g*-primal solution to the LC-CSCO problem in*

$$\tilde{\mathcal{O}}\left(\sqrt{\hat{R}_\Lambda + D}\left(1 + \frac{\sqrt{D}(D + \hat{R}_\Lambda)\|A\|}{\varepsilon_g} + \frac{\sqrt{D + \hat{R}_\Lambda}\|A\|}{\sqrt{L_f \varepsilon_g}} + \frac{\sqrt{DL_f}}{\sqrt{\varepsilon_g}}\right)\right)$$

ACG iterations.

Remarks

$$\tilde{\mathcal{O}}\left(\sqrt{\hat{R}_\Lambda + D}\left(1 + \frac{\sqrt{D}(D + \hat{R}_\Lambda)\|A\|}{\varepsilon_g} + \frac{\sqrt{D + \hat{R}_\Lambda}\|A\|}{\sqrt{L_f \varepsilon_g}} + \frac{\sqrt{DL_f}}{\sqrt{\varepsilon_g}}\right)\right) \quad (3)$$

- Near-optimal complexity in terms of ε, L_f, $\|A\|$ if $\varepsilon_g \leq L_f$
    - $\varepsilon \leq \|A\|^2/L_f$: Complexity is $\tilde{\mathcal{O}}\left(\sqrt{L_f/\varepsilon_g} + \|A\|/\varepsilon_g\right)$
    - $\varepsilon \geq \|A\|^2/L_f$: By rescaling ϕ, ε_g, complexity is $\tilde{\mathcal{O}}\left(\sqrt{L_f/\varepsilon_g} + L_f/\|A\|\right)$
- Improved (but still suboptimal) dependence on D, R_Λ

Outline

Numerical Experiments: Restarted ACG

Problem: LASSO

$$\phi_* := \min_{x \in \mathbb{R}^n} \left\{ \phi(x) := \frac{1}{2}\|Ax - b\|_2^2 + \frac{1}{2}\|x\|_1 \right\}. \quad (4)$$

Randomly generated $A \in \mathbb{R}^{500 \times 1000}$ with 20% density, $b \in \mathbb{R}^{500}$, $x_0 = 0$.

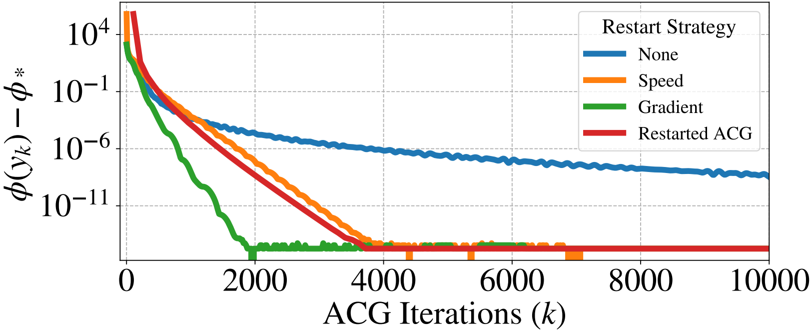

Figure: Comparison with baseline ACG, gradient restarting [O'Donoghue and Candès, 2015], and speed restarting [Su et al., 2016]

Numerical Experiments: I-(F)ALM

Problem: Linearly Constrained Quadratic Programming

$$\hat{\phi}_* := \min_{x \in \mathbb{R}^n} \left\{ \phi(x) := \frac{1}{2} x^\top M x + c^\top x + \delta_{\mathrm{Q}}(x) : Ax = b \right\},$$

where $Q = \{x : -10 \le x_i \le 10 \text{ for all } 1 \le i \le n\}$. Set $n = 200$, $m = 100$, $\|Q\| = 1$, rank$(Q) = 150$, A set to 20% density.

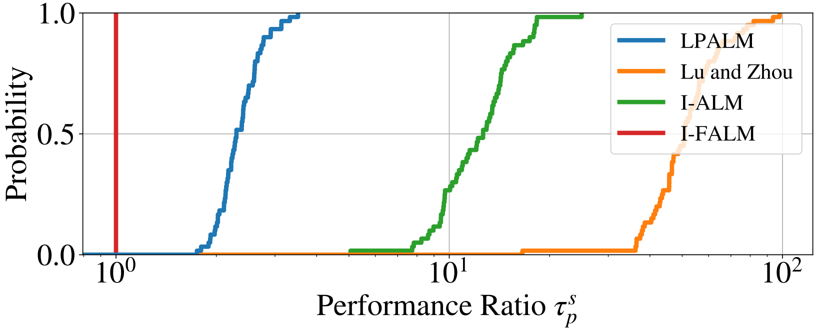

Figure: Performance profile of ALM variants: I-ALM, I-FALM, Lu and Zhou [2023], and LPALM Li and Lin [2019].

Thank you!

References I

G. Lan and R. D. C. Monteiro. Iteration-complexity of first-order augmented Lagrangian methods for convex programming. *Mathematical Programming*, 155(1-2):511–547, 2016. ISSN 0025-5610, 1436-4646. doi: 10.1007/s10107-015-0861-x.

H. Li and Z. Lin. Accelerated alternating direction method of multipliers: An optimal O(1 / K) nonergodic analysis. *Journal of Scientific Computing*, 79(2):671–699, 2019. ISSN 1573-7691. doi: 10.1007/s10915-018-0893-5.

Y. F. Liu, X. Liu, and S. Ma. On the nonergodic convergence rate of an inexact augmented Lagrangian framework for composite convex programming. *Mathematics of Operations Research*, 44(2):632–650, 2019. doi: 10.1287/moor.2018.0939.

Z. Lu and Z. Zhou. Iteration-complexity of first-order augmented Lagrangian methods for convex conic programming. *SIAM Journal on Optimization*, 33(2):1159–1190, 2023. ISSN 1052-6234. doi: 10.1137/21M1403837.

Y. Nesterov. Gradient methods for minimizing composite functions. *Mathematical Programming*, 140(1):125–161, 2013.

B. O'Donoghue and E. Candès. Adaptive restart for accelerated gradient schemes. *Foundations of Computational Mathematics*, 15(3):715–732, 2015. ISSN 1615-3383. doi: 10.1007/s10208-013-9150-3.

References II

Y. Ouyang and Y. Xu. Lower complexity bounds of first-order methods for convex-concave bilinear saddle-point problems. *Mathematical Programming*, 185 (1/2):1–35, 2021. ISSN 00255610. doi: 10.1007/s10107-019-01420-0.

A. Patrascu, I. Necoara, and Quoc Tran-Dinh. Adaptive inexact fast augmented Lagrangian methods for constrained convex optimization. *Optimization Letters*, 11 (3):609–626, 2017. ISSN 1862-4472, 1862-4480. doi: 10.1007/s11590-016-1024-6.

R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1(2): 97–116, 1976. ISSN 0364-765X.

W. Su, S. Boyd, and E. J. Candès. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, 17(153):1–43, 2016.

Y. Xu. Iteration complexity of inexact augmented Lagrangian methods for constrained convex programming. *Mathematical Programming*, 185(1):199–244, 2021. ISSN 1436-4646. doi: 10.1007/s10107-019-01425-9.