Iteration Complexity of First-order Methods for Linear Constrained Composite Nonconvex Problems

Wei Liu

Chinses academy of sciences

Joint with Yangyang Xu (RPI) and Qihang Lin (U. Iowa)

Beihang University, Nov. 2024

中国科学院数学与系统科学研究院

Academy of Mathematics and Systems Science

Chinese Academy of Sciences

Paper

- Wei Liu, Qihang Lin, Yangyang Xu, *First-order Methods for Affinely Constrained Composite Non-convex Non-smooth Problems: Lower Complexity Bound and Near-optimal Methods.* Arxiv

- Wei Liu, Qihang Lin, Yangyang Xu, *Lower Complexity Bounds of First-order Methods for Affinely Constrained Composite Non-convex Problems.* Submitted to Mathematics of Operations Research (second round)

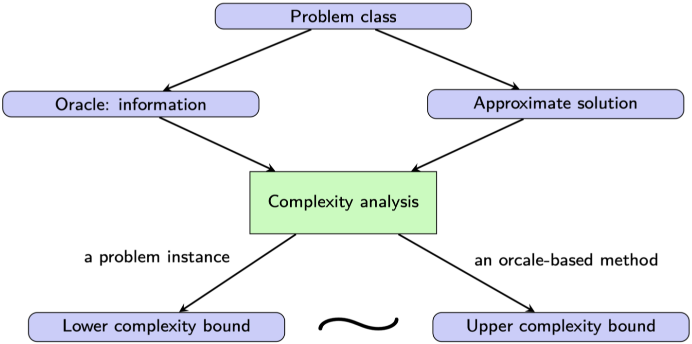
Problem class

$$\min_{\mathbf{x}\in\mathbb{R}^d} F_0(\mathbf{x}) := f_0(\mathbf{x}) + g(\mathbf{x}), \quad \text{s.t. } \mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{0} \tag{P}$$

- ∇f_0 is L-Lipschitz continuous
- $\inf_{\mathbf{x}} F_0(\mathbf{x}) > -\infty$
- $g(\mathbf{x}) = \bar{g}(\bar{\mathbf{A}}\mathbf{x} + \bar{\mathbf{b}})$, where $\bar{\mathbf{A}} \in \mathbb{R}^{\bar{n}\times d}$, $\bar{\mathbf{b}} \in \mathbb{R}^{\bar{n}}$, and $\bar{g} : \mathbb{R}^{\bar{n}} \to \mathbb{R} \cup \{+\infty\}$ is a proper closed convex function but potentially non-smooth
- relint(dom$(\bar{g})$) $\neq \emptyset$

Why lower complexity bounds

- provide understanding of the fundamental limit of a class of methods and the difficulty of a class of problems
- tell if existing methods could be improved
- guide to design "optimal" methods
- Opposite: Worst-case complexity/Upper complexity bounds


Problem class
Oracle: information
Approximate solution
Complexity analysis
a problem instance
an orcale-based method
Lower complexity bound
∼
Upper complexity bound

Two parts

Lower Complexity Bounds

Upper Complexity Bounds

A few existing lower bounds and motivating questions

Lower/upper bounds for unconstrained convex problems

Consider problem

$$f^* = \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

where f is convex and L-smooth, i.e., $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \le L\|\mathbf{x} - \mathbf{y}\|, \forall\, \mathbf{x}, \mathbf{y}$

- lower bound [Nesterov'04]: given $\varepsilon > 0$, there is a convex L-smooth function f such that $\Omega\left(\frac{\sqrt{L}\|\mathbf{x}^0 - \mathbf{x}^*\|}{\sqrt{\varepsilon}}\right)$ gradients needed to produce an ε-optimal solution $\bar{\mathbf{x}}$, i.e., $f(\bar{\mathbf{x}}) - f^* \le \varepsilon$, if

$$\mathbf{x}^k \in \mathbf{x}^0 + \mathsf{span}\{\nabla f(\mathbf{x}^0), \nabla f(\mathbf{x}^1), \ldots, \nabla f(\mathbf{x}^{k-1})\}$$

- upper bound: $O\left(\frac{\sqrt{L}\|\mathbf{x}^0 - \mathbf{x}^*\|}{\sqrt{\varepsilon}}\right)$ gradients enough

- [Nesterov'13]: the same upper bound holds for $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) + g(\mathbf{x})$ with oracle ∇f and $\mathbf{prox}_{\eta g}$

Observation: a nonsmooth proximable term g does not make a harder problem

Lower/upper bounds for unconstrained nonconvex problems

Consider problem

$$f^* = \min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$$

where f is L-smooth but may be nonconvex, and f^* is finite

- lower bound [Carmon et. al'20]: given $\varepsilon > 0$, there is an L-smooth function f such that $\Omega(\frac{L(f(\mathbf{x}^0)-f^*)}{\varepsilon^2})$ gradients needed to produce an ε-stationary point $\bar{\mathbf{x}}$, i.e., $\|\nabla f(\bar{\mathbf{x}})\| \le \varepsilon$
- upper bound: $O(\frac{L(f(\mathbf{x}^0)-f^*)}{\varepsilon^2})$

- the same upper bound holds for $\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) + g(\mathbf{x})$ with oracle ∇f and $\mathbf{prox}_{\eta g}$ to produce $\bar{\mathbf{x}}$ such that $\mathrm{dist}(\mathbf{0}, \nabla f(\bar{\mathbf{x}}) + \partial g(\bar{\mathbf{x}})) \le \varepsilon$

Observation: a nonsmooth proximable term g does not make a harder problem

Lower/upper bounds for linear-constrained nonconvex problems

Consider problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}), \text{ s.t. } \mathbf{A}\mathbf{x}+\mathbf{b}=\mathbf{0}$$

where f is L-smooth but may be nonconvex

- lower bound [Sun-Hong'19]: given $\varepsilon>0$, there is an L-smooth function f such that $\Omega\left(\frac{L\kappa(\mathbf{A})(f(\mathbf{x}^0)-\inf_{\mathbf{x}} f(\mathbf{x}))}{\varepsilon^2}\right)$ oracle calls to $\nabla f(\cdot)$, $\mathbf{A}(\cdot)$, and $\mathbf{A}^\top(\cdot)$ to produce an ε-KKT point
- upper bound: $\tilde{O}\left(\frac{L\kappa(\mathbf{A})(f(\mathbf{x}^0)-\inf_{\mathbf{x}} f(\mathbf{x}))}{\varepsilon^2}\right)$

Questions: for the regularized problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} F(\mathbf{x}) := f(\mathbf{x})+g(\mathbf{x}), \text{ s.t. } \mathbf{A}\mathbf{x}+\mathbf{b}=\mathbf{0}$$

where g is closed convex and $\mathbf{prox}_{\eta g}$ can be accessed

Q1. Will g make the problem harder?

Q2. What is a "tight" lower complexity bound?

Upper bounds for linear-constrained regularized nonconvex problems

- [Zhang-Luo'22]: when g is the indicator function of $\{\mathbf{x} : \bar{\mathbf{A}}\mathbf{x} \leq \bar{\mathbf{b}}\}$

$$O\left(\frac{\hat{\kappa}^2 L^3 (F(\mathbf{x}^0) - \inf_{\mathbf{x}} F(\mathbf{x}))}{\varepsilon^2}\right)$$

  where $\hat{\kappa}$ the condition number of $[\mathbf{A}; \bar{\mathbf{A}}]$

- [Lin-Ma-X.'22; Kong-Melo-Monteiro'23]: when Slater's condition holds, i.e., $\exists\, \mathbf{x} \in \mathrm{relint}(\mathrm{dom}(g)) : \mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{0}$, then $\tilde{O}(\varepsilon^{-2.5})$ oracles are enough

Observation: higher than the lower bound for linear-constrained nonregularized nonconvex problems; thus we guess g may make a harder problem

This talk: (nearly) close the gap between lower and upper bounds for a class of linear-constrained regularized nonconvex problems

Problem class

$$\min_{\mathbf{x}\in\mathbb{R}^d} F_0(\mathbf{x}) := f_0(\mathbf{x}) + g(\mathbf{x}), \quad \text{s.t. } \mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{0} \tag{P}$$

- ∇f_0 is L-Lipschitz continuous
- $\inf_{\mathbf{x}} F_0(\mathbf{x}) > -\infty$
- $g(\mathbf{x}) = \bar{g}(\bar{\mathbf{A}}\mathbf{x} + \bar{\mathbf{b}})$, where $\bar{\mathbf{A}} \in \mathbb{R}^{\bar{n}\times d}$, $\bar{\mathbf{b}} \in \mathbb{R}^{\bar{n}}$, and $\bar{g} : \mathbb{R}^{\bar{n}} \to \mathbb{R} \cup \{+\infty\}$ is a proper closed convex function but potentially non-smooth
- relint(dom$(\bar{g})$) $\neq \emptyset$

Algorithm Class I by a strong oracle

- **a strong oracle:** $\left(\nabla f_0(\cdot), \mathbf{A}(\cdot), \mathbf{A}^\top(\cdot)\right)$ and $\mathbf{prox}_{\eta g}(\cdot)$ for any $\eta > 0$
- iterates $\left\{\mathbf{x}^{(t)}\right\}_{t=0}^{\infty}$ satisfies: $\mathbf{x}^{(t)} \in \mathbf{span}\left(\left\{\boldsymbol{\xi}^{(t)}, \boldsymbol{\zeta}^{(t)}\right\}\right) \,\forall\, t \geq 1$, where

$$\boldsymbol{\xi}^{(t)} \in \mathbf{span}\left(\left\{\mathbf{A}^\top \mathbf{b}\right\} \bigcup \cup_{s=0}^{t-1} \left\{\mathbf{x}^{(s)}, \nabla f_0(\mathbf{x}^{(s)}), \mathbf{A}^\top \mathbf{A}\mathbf{x}^{(s)}\right\}\right)$$

$$\boldsymbol{\zeta}^{(t)} \in \left\{\mathbf{prox}_{\eta g}(\boldsymbol{\xi}^{(t)}) \mid \eta > 0\right\}$$

  - **Examples:** quadratic penalty or augmented Lagrangian based first-order methods [Kong-Melo-Monteiro'19 '20]

Lower complexity bound

Let $\epsilon > 0$ and $L > 0$ be given. Then for any $\omega \in [0, \frac{150\pi\epsilon}{L})$, there exists an instance of problem (P) such that at least

$$\left\lceil \frac{\kappa([\bar{\mathbf{A}};\mathbf{A}])L\Big(F_0(\mathbf{x}^{(0)}) - \inf_{\mathbf{x}} F_0(\mathbf{x})\Big)}{36000\pi^2}\epsilon^{-2} \right\rceil$$

oracles are needed by any algorithm in the considered class to obtain a point that is ω-close to an ϵ-KKT point of the instance

- It can hold $\kappa([\bar{\mathbf{A}};\mathbf{A}]) \gg \kappa(\mathbf{A})$, $[\bar{\mathbf{A}};\mathbf{A}] := \begin{bmatrix} \bar{\mathbf{A}} \\ \mathbf{A} \end{bmatrix}$
- Hence, existence of g makes a harder problem even with $\mathbf{prox}_{\eta g}$
  This answers Q1

Algorithm Class II by a more practical oracle

- $g(\mathbf{x}) = \bar{g}(\bar{\mathbf{A}}\mathbf{x} + \bar{\mathbf{b}})$
- **a practical oracle:** $\left(\nabla f_0(\cdot), \mathbf{A}(\cdot), \mathbf{A}^\top(\cdot), \bar{\mathbf{A}}(\cdot), \bar{\mathbf{A}}^\top(\cdot)\right)$, and $\mathbf{prox}_{\eta\bar{g}}(\cdot)$
- iterate sequence $\left\{(\mathbf{x}^{(t)}, \mathbf{y}^{(t)})\right\}_{t=0}^{\infty}$ satisfies: for all $t \geq 1$

$$\mathbf{x}^{(t)} \in \mathbf{span}\Big(\left\{\mathbf{A}^\top\mathbf{b}, \bar{\mathbf{A}}^\top\bar{\mathbf{b}}\right\} \bigcup \cup_{s=0}^{t-1}\left\{\mathbf{x}^{(s)}, \nabla f_0(\mathbf{x}^{(s)}), \mathbf{A}^\top\mathbf{A}\mathbf{x}^{(s)}, \bar{\mathbf{A}}^\top\bar{\mathbf{A}}\mathbf{x}^{(s)}, \bar{\mathbf{A}}^\top\mathbf{y}^{(s)}\right\}\Big)$$

$$\mathbf{y}^{(t)} \in \mathbf{span}\left(\left\{\boldsymbol{\xi}^{(t)}, \boldsymbol{\zeta}^{(t)}\right\}\right), \text{ where}$$

$$\boldsymbol{\xi}^{(t)} \in \mathbf{span}\left(\left\{\bar{\mathbf{b}}\right\} \bigcup \cup_{s=0}^{t-1}\left\{\mathbf{y}^{(s)}, \bar{\mathbf{A}}\bar{\mathbf{A}}^\top\mathbf{y}^{(s)}, \bar{\mathbf{A}}\mathbf{x}^{(s)}\right\}\right)$$

$$\boldsymbol{\zeta}^{(t)} \in \left\{\mathbf{prox}_{\eta\bar{g}}(\boldsymbol{\xi}^{(t)}) \mid \eta > 0\right\}$$

  - **Examples:** quadratic penalty or augmented Lagrangian based first-order methods (such as linearized ADMM) for

$$\min_{\mathbf{x},\mathbf{y}} F(\mathbf{x},\mathbf{y}) := f_0(\mathbf{x}) + \bar{g}(\mathbf{y}), \quad \text{s.t. } \mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{0}, \bar{\mathbf{A}}\mathbf{x} + \bar{\mathbf{b}} = \mathbf{y}$$

  - **Examples:** existing FOMs on solving minimax-structured optimization [Thekumparampil et al.'19, Lin et al.'20, Xu et al.'23]

Lower complexity bound under the practical oracle

Let $\epsilon > 0$ and $L > 0$ be given. Then for any $\omega \in [0, \frac{150\pi\epsilon}{L})$, there exists an instance of problem (P) such that at least

$$\left\lceil \frac{\kappa([\bar{\mathbf{A}}; \mathbf{A}]) L \Big(F_0(\mathbf{x}^{(0)}) - \inf_{\mathbf{x}} F_0(\mathbf{x}) \Big)}{18000\pi^2} \epsilon^{-2} \right\rceil$$

oracles are needed by any algorithm in the considered class to obtain a point that is ω-close to an ϵ-KKT point of the instance

- The lower complexity bound almost the same as the previous one

Worst-case instance inspired by [Sun-Hong'19]

- objective: let $m = 3m_1m_2$ be even for two integers $m_1 \geq 2, m_2 \geq 1$. Set $\mathbf{x} = (\mathbf{x}_1; \ldots; \mathbf{x}_m)$ and

$$f_0(\mathbf{x}) := \sum_{i=1}^{m} f_i(\mathbf{x}_i), \text{ with } f_i(\mathbf{z}) := \frac{300\pi\epsilon^2}{mL} h_i\left(\frac{\sqrt{m}L\mathbf{z}}{150\pi\epsilon}\right)$$

  where $\{h_i\}$ are designed to satisfy certain properties

- linear constraint: let $\mathcal{M} := \{im_1\}_{i=1}^{3m_2-1}$ and $\mathcal{M}^C := \{1, 2, \ldots, m-1\} \backslash \mathcal{M}$

$$\bar{\mathbf{A}} := mL \cdot \mathbf{J}_{\mathcal{M}} \otimes \mathbf{I}_{\bar{d}}, \ \mathbf{A} := mL \cdot \mathbf{J}_{\mathcal{M}^C} \otimes \mathbf{I}_{\bar{d}}, \ \bar{\mathbf{b}} = \mathbf{0}, \ \mathbf{b} = \mathbf{0}$$

  where $\mathbf{J}_{\mathcal{M}}$ and $\mathbf{J}_{\mathcal{M}^C}$ are the rows of $\mathbf{J}_m$ indexed by $\mathcal{M}$ and $\mathcal{M}^C$ and

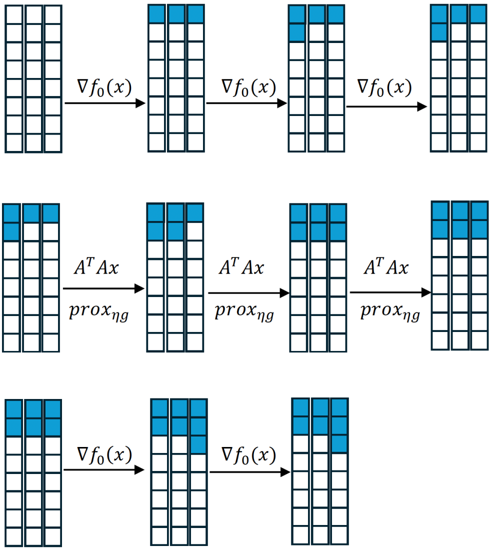
$$\mathbf{J}_m := \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(m-1)\times m}$$

- regularizer: $\bar{g}(\mathbf{y}) := \frac{\beta}{mL}\|\mathbf{y}\|_1$ and $g(\mathbf{x}) := \bar{g}(\bar{\mathbf{A}}\mathbf{x})$

Properties

- $\kappa(\mathbf{A}) = O(1), \kappa(\bar{\mathbf{A}}) = O(1)$ but $\kappa([\bar{\mathbf{A}}; \mathbf{A}]) = \Theta(m)$
- a KKT point $\mathbf{x}$ satisfies $\mathbf{x}_1 = \mathbf{x}_2 = \cdots = \mathbf{x}_m$
- KKT violation$(\mathbf{x}) \geq \varepsilon$, if some $|[\bar{\mathbf{x}}]_j| < \frac{150\pi\epsilon}{\sqrt{m}L}$, where $\bar{\mathbf{x}} := \frac{1}{m}\sum_{i=1}^m \mathbf{x}_i$
- From $\mathbf{x}^{(0)} = \mathbf{0}$, it takes $\Theta(m)$ oracles to kill one zero in the block average

Zero-respecting sequences



Recap

- With a regularizer, lower complexity bounds become higher
- Hence, a regularizer makes the problem harder
- But still gap between the new lower bounds and existing upper bounds

Are the lower complexity bounds tight?

Answer: The lower complexity bounds under Algorithm Class II is tight, see the second part of this talk

Conclusions and Open Questions

- established two lower complexity bounds of first-order methods for linear-constrained regularized nonconvex problems
  - under two different oracles
  - lower bounds higher than that for linear-constrained non-regularied nonconvex problems
  - **Take-away message:** regularizer makes linear-constrained nonconvex problems harder even if its proximal mapping is used

Open question:

- What is the "tight" lower bound if g is a general convex function?
- If high-order derivative is available, what does the lower bound look like?

Extension

If high-order derivative is available

- when f_0 has L_f-Lipschitz continuous p-th order derivatives, the lower bound would be
$$\Omega\left(\frac{\kappa([\bar{\mathbf{A}};\mathbf{A}])^{\frac{p+1}{2p}} L_f^{\frac{1}{p}} \Delta_{F_0}}{\epsilon^{\frac{p+1}{p}}}\right)$$
- compared to problem (P) without the regularization term, the lower bound would be [Carmon et. al'20]
$$\Omega\left(\frac{L_f^{\frac{1}{p}} \Delta_{F_0}}{\epsilon^{\frac{p+1}{p}}}\right)$$
- Answer: existence of g makes a more challenging problem in some cases
  - the latter lower bound is tight
  - The upper bound for solving problem (P) without the regularization term is **lower than the former lower bound**

Relax for a while

New upper complexity bounds by a new algorithm

Problem formulation

$$\min_{\mathbf{x}} F(\mathbf{x}) := f(\mathbf{x}) + g(\bar{\mathbf{A}}\mathbf{x} + \bar{\mathbf{b}}),$$
$$\text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \; \mathbf{h}(\mathbf{x}) := [h_1(\mathbf{x}), \ldots, h_m(\mathbf{x})] \leq \mathbf{0} \tag{QCP}$$

- f is smooth and ρ-weakly convex, i.e., $f(\cdot) + \frac{\rho}{2}\|\cdot\|^2$ is convex for some $\rho > 0$
- $h_i(\mathbf{x}) := \frac{1}{2}\mathbf{x}^\top \mathbf{Q}_i \mathbf{x} + \mathbf{r}_i^\top \mathbf{x} + c_i$ is convex quadratic
- $\mathbf{A}, \bar{\mathbf{A}}$ and $\mathbf{b}, \bar{\mathbf{b}}$ are given matrix and vector; g is convex

Two cases of f to be explored

1. $m = 0$ (reduce to Problem (P))
2. $m \geq 1$

Remark: these cases will be handled differently

Many applications

- all linear (equality or inequality) constrained and/or quadratically-constrained problems
- linear constrained nonlinear least squares [Orban-Siqueira'20]
- constrained machine learning: Neyman-Pearson classification [Scott'07], logical neural network [Riegel al et.'20]
- robust phase retrieval [Davis-Drusvyatskiy'19] and linear-constrained variant

Key assumptions

- Continuity of F, $\inf F(\mathbf{x}) > \infty$
- Lipschitz continuity of g
- Slater-type condition
- It holds that either (a) $\widetilde{\mathbf{A}} = \left[\overline{\mathbf{A}}^\top, \mathbf{A}^\top\right]^\top$ has a full-row rank; or (b) $g(\mathbf{y}) = \max\left\{\mathbf{u}^\top \mathbf{y} \,:\, \mathbf{C}\mathbf{u} \leq \mathbf{d}, \mathbf{u} \in \mathbb{R}^{\bar{n}}\right\}$ for some $\mathbf{C}$ and $\mathbf{d}$

Instances

- the assumption on g includes ℓ_1 norm
- the worst-case instance given in page 16 meets all the assumptions

Framework of inexact proximal method when $m = 0$

$$(\mathbf{x}^{(k+1)}, \mathbf{y}^{(k+1)}) \approx \arg\min_{\mathbf{x},\mathbf{y}} \ \left\langle \nabla f(\mathbf{x}^{(k)}), \mathbf{x} - \mathbf{x}^{(k)} \right\rangle + \frac{\tau}{2} \left\| \mathbf{x} - \mathbf{x}^{(k)} \right\|^2 + g(\mathbf{y})$$

$$\text{s.t.} \ \ \mathbf{y} = \bar{\mathbf{A}}\mathbf{x} + \bar{\mathbf{b}}, \ \mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{0}$$

where $\tau > L$ to have a strongly convex problem

- Let $\mathbf{z}^{k+1}$ be an approximate solution of the Lagrangian dual problem

$$\min_{\mathbf{z}} \frac{1}{2\tau} \|\bar{\mathbf{A}}^\top \mathbf{z}_1 + \mathbf{A}^\top \mathbf{z}_2 + \nabla f(\mathbf{x}^{(k)}) - \tau \mathbf{x}^{(k)}\|^2 + g^\star(\mathbf{z}_1) - \mathbf{z}_1^\top \bar{\mathbf{b}} - \mathbf{z}_2^\top \mathbf{b}$$

- Set $(\mathbf{x}^{(k+1)}, \mathbf{y}^{(k+1)})$ to

$$\begin{aligned} \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} - \frac{1}{\tau} \left(\bar{\mathbf{A}}^\top \mathbf{z}_1^{(k+1)} + \mathbf{A}^\top \mathbf{z}_2^{(k+1)} + \nabla f(\mathbf{x}^{(k)}) \right) \\ \mathbf{y}^{(k+1)} &= \mathbf{prox}_{\sigma^{-1}\bar{g}} \left(\sigma^{-1} \mathbf{z}_1^{(k+1)} + \bar{\mathbf{A}}\mathbf{x}^{(k+1)} + \bar{\mathbf{b}} \right) \end{aligned}$$

Worst-case complexity

- Assumption: Lipschitz continuity of ∇f
- A (restarted) APG has linear convergence
- Total oracle complexity to find a point that is ω-close to an ϵ-KKT point of problem (P) is

  $$O\left(\kappa([\bar{\mathbf{A}};\mathbf{A}])\log\left(\tfrac{\Delta_F}{\epsilon}\right)\frac{L\Delta_F}{\epsilon^2}\right)$$

  where $\Delta_F = F(\mathbf{x}^{(0)}) - \inf_{\mathbf{x}} F(\mathbf{x})$, and $\omega > 0$ arbitrarily close to 0
- Upper bound almost matches the lower bound
  This answers Q2

APG: linear convergence

- an initial point $\boldsymbol{\gamma}^{\text{ini}} = ((\mathbf{y}^{(0)})^\top, (\mathbf{A}\mathbf{x}^{(0)})^\top)^\top$
- $\boldsymbol{\gamma}^{(0,k)} \leftarrow \boldsymbol{\gamma}^{\text{ini}}$, $\widehat{\boldsymbol{\gamma}}^{(0,k)} \leftarrow \boldsymbol{\gamma}^{(0,k)}$, $\alpha_0 \leftarrow 1$
- **For** $i = 0, \ldots, i_k - 1$
  - **For** $j = 0, \ldots, j_k - 1$
    - $\widehat{\mathcal{G}}_1^j \leftarrow \frac{1}{\tau}\bar{\mathbf{A}}\left(\bar{\mathbf{A}}^\top\widehat{\boldsymbol{\gamma}}_1^{(j,k)} + \mathbf{A}^\top\widehat{\boldsymbol{\gamma}}_2^{(j,k)} + \boldsymbol{\xi}_f^{(k)} - \tau\mathbf{x}^{(k)}\right) - \bar{\mathbf{b}}$
    - $\widehat{\mathcal{G}}_2^j \leftarrow \frac{1}{\tau}\mathbf{A}\left(\bar{\mathbf{A}}^\top\widehat{\boldsymbol{\gamma}}_1^{(j,k)} + \mathbf{A}^\top\widehat{\boldsymbol{\gamma}}_2^{(j,k)} + \boldsymbol{\xi}_f^{(k)} - \tau\mathbf{x}^{(k)}\right) - \mathbf{b}$
    - $\boldsymbol{\gamma}_1^{(j+1,k)} \leftarrow \mathbf{prox}_{L_{\mathcal{D}}^{-1}g^*}\left(\widehat{\boldsymbol{\gamma}}_1^{(j,k)} - \frac{1}{L_{\mathcal{D}}}\widehat{\mathcal{G}}_1^j\right)$, $\boldsymbol{\gamma}_2^{(j+1,k)} \leftarrow \widehat{\boldsymbol{\gamma}}_2^{(j,k)} - \frac{1}{L_{\mathcal{D}}}\widehat{\mathcal{G}}_2^j$
    - $\alpha_{j+1} \leftarrow \frac{1+\sqrt{1+4\alpha_j^2}}{2}$
    - $\widehat{\boldsymbol{\gamma}}^{(j+1,k)} \leftarrow \boldsymbol{\gamma}^{(j+1,k)} + \left(\frac{\alpha_j - 1}{\alpha_{j+1}}\right)\left(\boldsymbol{\gamma}^{(j+1,k)} - \boldsymbol{\gamma}^{(j,k)}\right)$
  - $\boldsymbol{\gamma}^{(0,k)} \leftarrow \boldsymbol{\gamma}^{(j_k,k)}$, $\widehat{\boldsymbol{\gamma}}^{(0,k)} \leftarrow \boldsymbol{\gamma}^{(0,k)}$, $\alpha_0 \leftarrow 1$
- **Output:** $\boldsymbol{\gamma}^{(k+1)} = \boldsymbol{\gamma}^{(0,k)}$

Framework of inexact proximal method when $m = O(1)$

$$
(\mathbf{x}^{(k+1)}, \mathbf{y}^{(k+1)}) \approx \arg\min_{\mathbf{x},\mathbf{y}} \; \left\langle \nabla f(\mathbf{x}^{(k)}), \mathbf{x} - \mathbf{x}^{(k)} \right\rangle + \frac{\tau}{2} \left\| \mathbf{x} - \mathbf{x}^{(k)} \right\|^2 + g(\mathbf{y})
$$

$$
\text{s.t.} \;\; \mathbf{y} = \bar{\mathbf{A}}\mathbf{x} + \bar{\mathbf{b}}, \;\; \mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{0}, h_i(\mathbf{x}) \le 0
$$

where $\tau > L$ to have a strongly convex problem

Lagrangian function:

$$
\mathcal{L}_k(\mathbf{x}, \mathbf{y}; \boldsymbol{\gamma}, \boldsymbol{\lambda}) = \bar{f}^k(\mathbf{x}) + g(\mathbf{y}) - \boldsymbol{\gamma}_1^\top \left(\mathbf{y} - \left(\bar{\mathbf{A}}\mathbf{x} + \bar{\mathbf{b}} \right) \right) + \boldsymbol{\gamma}_2^\top \left(\mathbf{A}\mathbf{x} + \mathbf{b} \right) + \sum_{i=1}^{m} \lambda_i h_i(\mathbf{x})
$$

Let $\mathcal{D}_k(\boldsymbol{\gamma}, \boldsymbol{\lambda}) := -\min_{\mathbf{x},\mathbf{y}} \mathcal{L}_k(\mathbf{x}, \mathbf{y}; \boldsymbol{\gamma}, \boldsymbol{\lambda})$

$$
\Omega^{(k+1)} := \operatorname*{Arg\,min}_{\boldsymbol{\gamma}, \boldsymbol{\lambda} \ge 0} \mathcal{D}_k(\boldsymbol{\gamma}, \boldsymbol{\lambda}) \quad \text{and} \quad \mathcal{D}_k^* := \min_{\boldsymbol{\gamma}, \boldsymbol{\lambda} \ge 0} \mathcal{D}_k(\boldsymbol{\gamma}, \boldsymbol{\lambda})
$$

Framework of inexact proximal method when $m = O(1)$ (Cont'd)

A primal-dual step

Find a near-optimal point $(\boldsymbol{\gamma}^{(k+1)}, \boldsymbol{\lambda}^{(k+1)})$ by solving the convex problem $\min_{\boldsymbol{\gamma}, \boldsymbol{\lambda} \geq 0} \mathcal{D}_k(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ inexactly

Let $\sigma > 0$ be a given scalar, $\widetilde{\mathbf{Q}}_i = (\mathbf{Q}_i + {\mathbf{Q}_i}^\top)/2$

Obtain $(\mathbf{x}^{(k+1)}, \mathbf{y}^{(k+1)})$ by solving

$$\left(\tau \mathbf{I}_d + \sum_{i=1}^m \lambda_i^{(k+1)} \widetilde{\mathbf{Q}}_i\right) \mathbf{x}^{(k+1)} = \tau \mathbf{x}^{(k)} - \left(\widetilde{\mathbf{A}}^\top \boldsymbol{\gamma}^{(k+1)} + \nabla f(\mathbf{x}^{(k)}) + \sum_{i=1}^m \lambda_i^{(k+1)} \mathbf{r}_i\right)$$

and letting

$$\mathbf{y}^{(k+1)} = \mathbf{prox}_{\sigma^{-1} g}\left(\sigma^{-1} \gamma_1^{(k+1)} + \bar{\mathbf{A}} \mathbf{x}^{(k+1)} + \bar{\mathbf{b}}\right)$$

The x-solution is

$$\mathbf{x}^{k+1}(\boldsymbol{\gamma}, \boldsymbol{\lambda}) := \left(\tau \mathbf{I}_d + \boldsymbol{\lambda} \widetilde{\mathbf{Q}}\right)^{-1} \left(\tau \mathbf{x}^{(k)} - \boldsymbol{\lambda} \mathbf{r} - \widetilde{\mathbf{A}}^\top \boldsymbol{\gamma} - \nabla f(\mathbf{x}^{(k)})\right)$$

Worst-case complexity

For outer loop

- Assumption: Lipschitz continuity of ∇f
- Outer oracle complexity to find a point that is ω-close to an ϵ-KKT point of problem (P) is

$$O\left(\frac{L\Delta_F}{\epsilon^2}\right)$$

  where $\Delta_F = F(\mathbf{x}^{(0)}) - \inf_{\mathbf{x}} F(\mathbf{x})$, and $\omega > 0$ arbitrarily close to 0

For inner loop: solve the convex problem $\min_{\boldsymbol{\gamma},\boldsymbol{\lambda}\geq 0} \mathcal{D}_k(\boldsymbol{\gamma},\boldsymbol{\lambda})$ inexactly

- use a bisection method on $\boldsymbol{\lambda}$
- $O(m)$ iterations

Comparison to existing methods

- For composite nonconvex optimization with affinely constraints and smooth f, $O\left(\kappa([\bar{\mathbf{A}};\mathbf{A}])\log\left(\frac{\Delta_F}{\epsilon}\right)\frac{L\Delta_F}{\epsilon^2}\right) \Rightarrow$ Near Optimal
- For composite nonconvex optimization with convex quadratic constraints and smooth f, $\tilde{O}(m/\varepsilon^2)$ V.S. $\tilde{O}(1/\varepsilon^{2.5})$

Conclusion and Future Work

Conclusion

- The lower complexity bound under Algorithm Class II is tight
- The upper complexity bound for solving problem (OCP) is improved

Future Work

- For Algorithm Class 1 on solving Problem (P), can the lower bound be achieved, i.e., is it tight?
- How to formulate a lower complexity bound of Algorithm Class 1 on solving problem (P) without $g = \bar{g}(\bar{\mathbf{A}}\mathbf{x} + \bar{\mathbf{b}})$?
- What will the lower bound look like if there are convex nonlinear inequality constraints?

Thank you!!!