First-order Methods for Affinely Constrained Composite Non-convex Non-smooth Problems: Lower Complexity Bound and Near-optimal Methods

Wei Liu
liuw16@rpi.edu
Rensselaer Polytechnic Institute

Joint work with **Qihang Lin** (iowa), and **Yangyang Xu** (RPI)

Talk in MOPTA
August, 2023

References and Photos

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.* Submitted to Mathematical Programming.

(a) Qihang Lin

(b) Yangyang Xu

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 “tight/near-optimal” methods to find an ϵ stationary point

Here, if there are only constants and log terms difference between lower bound and upper bound, we call them tight/near-optimal

First-order Methods for Nonsmooth Nonconvex Problems

Consider Problem

$$\min_{\mathbf{x}} f_0(\mathbf{x}) + g(\mathbf{x})$$

- g is nonsmooth, the proximal operator of g is allowed
- f_0 is nonconvex and L_f-smooth, i.e.,

$$\|\nabla f_0(\mathbf{x}_1) - \nabla f_0(\mathbf{x}_2)\| \leq L_f \|\mathbf{x}_1 - \mathbf{x}_2\|, \forall \mathbf{x}_1, \mathbf{x}_2$$

- upper complexity bound $O(L_f \epsilon^{-2})$, provided by a proximal gradient method [Nesterov 2012]:

$$\mathbf{x}^{k+1} = \mathbf{prox}_{\eta g}(\mathbf{x}^k - \eta \nabla f_0(\mathbf{x}^k))$$

- lower complexity bound: $O(L_f \epsilon^{-2})$ [Carmon et al. 2020], by accessing the gradient of f_0 and the proximal operator of g $\Rightarrow$ Tight!

Consider Problem

$$\min_{\mathbf{x}} \ f_0(\mathbf{x}),$$
$$\text{s.t. } \mathbf{A}\mathbf{x} + \mathbf{b} = 0$$

- f_0 is non-convex and L_f-smooth
- $\mathbf{A}$ has a full-row rank
- upper complexity bound $O(\kappa(\mathbf{A})L_f\epsilon^{-2})$, provided by a projection gradient method [Nesterov 2013]:

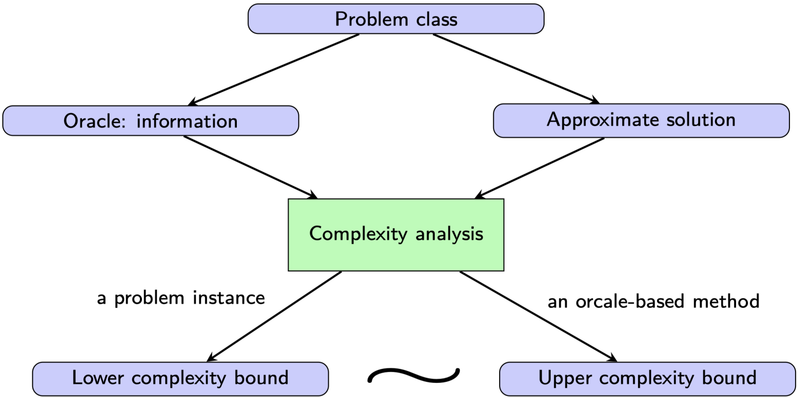
$$\mathbf{x}^{k+1} = \textbf{proj}_{\mathbf{x}:\mathbf{A}\mathbf{x}+\mathbf{b}=0}(\mathbf{x}^k - \eta\nabla f_0(\mathbf{x}^k))$$

- lower complexity bound: $O(\kappa(\mathbf{A})L_f\epsilon^{-2})$ [Sun-Hong 2019], by accessing the information of ∇f_0 and $\mathbf{A}\cdot \Rightarrow$ Tight!

More Examples

- first-order methods for convex problems [Nesterov 2004]
- first-order methods for stochastic convex problems [Agarwal et al. 2012]
- first-order methods for finite-sum convex problems [Woodworth-Srebro 2016]
- first-order and higher-order methods for nonconvex problems [Carmon et al. 2017]

Diagram: iteration complexity analysis



This talk: Affinely Constrained Nonsmooth Nonconvex Problems

Problem Class I

$$\begin{aligned} \min_{\mathbf{x}} \ & F_0(\mathbf{x}) := f_0(\mathbf{x}) + g(\mathbf{x}) \\ \text{s.t. } & \mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{0} \end{aligned} \tag{P}$$

$\mathbf{A} \in \mathbb{R}^{n \times d}$, $\mathbf{b} \in \mathbb{R}^n$, f_0 is smooth nonconvex, and g is convex non-smooth

Assumptions

- ∇f_0 is L_f-Lipschitz continuous
- $\inf_{\mathbf{x}} F_0(\mathbf{x}) > -\infty$. $\Delta := F_0(\mathbf{x}^{(0)}) - \inf_{\mathbf{x}} F_0(\mathbf{x})$
- $g(\mathbf{x}) = \bar{g}(\bar{\mathbf{A}}\mathbf{x} + \bar{\mathbf{b}})$, where $\bar{\mathbf{A}} \in \mathbb{R}^{\bar{n} \times d}$, $\bar{\mathbf{b}} \in \mathbb{R}^{\bar{n}}$ and $\bar{g}$ is convex non-smooth.
- g is a proximable function, i.e., easy proximal mapping $\mathbf{prox}_{\eta g}$

Problem Class II: Splitting case

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

- $\bar{g}$ is a proximable function (weaker than that of g)
- $\inf_{\mathbf{x},\mathbf{y}} F(\mathbf{x},\mathbf{y}) > -\infty$. $\Delta := F(\mathbf{x}^{(0)},\mathbf{y}^{(0)}) - \inf_{\mathbf{x},\mathbf{y}} F(\mathbf{x},\mathbf{y})$

Algorithm Setup (first-order "Oracle: Information")

Linear Combinations of

- information of f_0 through $f_0(\mathbf{x}), \nabla f_0(\mathbf{x})$ (for (P) and (SP))
- information of $\mathbf{A}$ through $\mathbf{A}\mathbf{x}$ and $\mathbf{A}^\top\mathbf{y}$ for inquiry $(\mathbf{x},\mathbf{y})$ (for (P) and (SP))
- information of $\bar{\mathbf{A}}$ through $\bar{\mathbf{A}}\mathbf{x}$ and $\bar{\mathbf{A}}^\top\mathbf{y}$ for inquiry $(\mathbf{x},\mathbf{y})$ (for (P) and (SP))
- information of $\bar{\mathbf{b}}$ and $\mathbf{b}$ (for (P) and (SP))

and information of g (for (P)) or $\bar{g}$ (for (SP)) through the proximal operator

Problem Class II: Splitting case

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

- $\bar{g}$ is a proximable function (weaker than that of g)
- $\inf_{\mathbf{x},\mathbf{y}} F(\mathbf{x},\mathbf{y}) > -\infty$. $\Delta := F(\mathbf{x}^{(0)}, \mathbf{y}^{(0)}) - \inf_{\mathbf{x},\mathbf{y}} F(\mathbf{x},\mathbf{y})$

Algorithm Setup (first-order "Oracle: Information")

Linear Combinations of

- information of f_0 through $f_0(\mathbf{x}), \nabla f_0(\mathbf{x})$ (for (P) and (SP))
- information of $\mathbf{A}$ through $\mathbf{A}\mathbf{x}$ and $\mathbf{A}^\top\mathbf{y}$ for inquiry $(\mathbf{x},\mathbf{y})$ (for (P) and (SP))
- information of $\bar{\mathbf{A}}$ through $\bar{\mathbf{A}}\mathbf{x}$ and $\bar{\mathbf{A}}^\top\mathbf{y}$ for inquiry $(\mathbf{x},\mathbf{y})$ (for (P) and (SP))
- information of $\bar{\mathbf{b}}$ and $\mathbf{b}$ (for (P) and (SP))

and information of g (for (P)) or $\bar{g}$ (for (SP)) through the proximal operator

Our Desired "Approximate Solution"

To obtain an ϵ-stationary point of (P) and (SP) (due to nonconvexity)

- a point $\mathbf{x}^*$ is called an ϵ-stationary point of (P) if for some $\mathbf{z} \in \mathbb{R}^n$,

$$\max\left\{\mathrm{dist}\left(\mathbf{0}, \nabla f_0(\mathbf{x}^*) + \mathbf{A}^\top \mathbf{z} + \partial g(\mathbf{x}^*)\right), \left\|\mathbf{A}\mathbf{x}^* + \mathbf{b}\right\|\right\} \le \epsilon$$

  $\bar{\mathbf{x}}$ is a near ϵ-stationary point of (P) if it is ω-close to an ϵ-stationary point $\mathbf{x}^*$ of (P) with $\omega = O(\epsilon)$

- a point $(\mathbf{x}^*, \mathbf{y}^*)$ is called an ϵ-stationary point of (SP) if for some $\mathbf{z}_1 \in \mathbb{R}^{\bar{n}}$ and $\mathbf{z}_2 \in \mathbb{R}^n$

$$\max\Big\{\mathrm{dist}(\mathbf{0}, \partial \bar{g}(\mathbf{y}^*) - \mathbf{z}_1), \|\nabla f_0(\mathbf{x}^*) + \bar{\mathbf{A}}^\top \mathbf{z}_1 + \mathbf{A}^\top \mathbf{z}_2\|,$$
$$\|\mathbf{y}^* - \bar{\mathbf{A}}\mathbf{x}^*\|, \|\mathbf{A}\mathbf{x}^*\|\Big\} \le \epsilon$$

Existing Upper Bound Complexity Results

For (P)

- inexact proximal point method: $O(\epsilon^{-2.5})$ [Lin et al. 200]
- inexact proximal accelerated augmented Lagrangian (ALM): $O(\epsilon^{-2.5})$ [Melo et al. 2020]
- smoothed proximal ALM with g being indicator function: $O(\epsilon^{-2})$ [Zhang-Luo 2022]
- alternating direction method of multipliers (ADMM) with $g = 0$ or $\bar{\mathbf{A}} = I_d$: $O(\epsilon^{-2})$ [Jiang et al. 2019]

For (SP)

- ADMM with Kurdyka-Łojasiewicz coefficient: $O(\epsilon^{-2})$ [Yashtini 2022]
- ADMM: $O(\epsilon^{-2})$ [Goncalves et al. 2017]

Questions

For problems (P) and (SP)

- are existing lower bound results tight?

  (there is no lower bound results for our problem, except for some special instances)

- are existing upper bound results tight?

  (examine the hidden constant)

Are These Upper Bound Complexity Tight?

Our lower bound complexity results: $O(\kappa([\mathbf{A};\bar{\mathbf{A}}])L_f\Delta\epsilon^{-2})$ for (P) and (SP)

For (P)

- smoothed proximal ALM: $O(\hat{\kappa}^2 L_f^3 \Delta\epsilon^{-2})$ [Zhang-Luo 2022]
- ADMM with $g = 0$ or $\bar{\mathbf{A}} = I_d$: $O(\kappa([\mathbf{A};\bar{\mathbf{A}}])^2 L_f^2 \Delta\epsilon^{-2})$ [Jiang et al. 2019]

For (SP)

- ADMM with Kurdyka-Łojasiewicz coefficient: not comparable [Yashtini 2022]
- ADMM: $O(\kappa([\mathbf{A};\bar{\mathbf{A}}])^2 L_f^2 \Delta\epsilon^{-2})$ [Goncalves et al. 2017]

the existing upper bound is not tight!

Roadmap

lower iteration complexity bounds

- $O(\kappa([\mathbf{A}; \bar{\mathbf{A}}])L_f \Delta \epsilon^{-2})$ to find a near ϵ stationary point of (P)
- $O(\kappa([\mathbf{A}; \bar{\mathbf{A}}])L_f \Delta \epsilon^{-2})$ to find an ϵ stationary point of (SP)
- κ: condition number

upper iteration complexity bounds

- design a new inexact proximal gradient (IPG) method
- find an ϵ stationary point $(\mathbf{x}^*, \mathbf{y}^*)$ of (SP) within $\tilde{O}(\kappa([\mathbf{A}; \bar{\mathbf{A}}])L_f \Delta \epsilon^{-2})$
  $\Rightarrow$tight/near-optimal!
- $\mathbf{x}^*$ is arbitrary close to an ϵ stationary point of (P)
  $\Rightarrow$ find a near ϵ stationary point $\mathbf{x}^*$ of (P) within $\tilde{O}(\kappa([\mathbf{A}; \bar{\mathbf{A}}])L_f \Delta \epsilon^{-2})$
  $\Rightarrow$tight/near-optimal!

Roadmap

lower iteration complexity bounds

- $O(\kappa([\mathbf{A}; \bar{\mathbf{A}}])L_f \Delta\epsilon^{-2})$ to find a near ϵ stationary point of (P)
- $O(\kappa([\mathbf{A}; \bar{\mathbf{A}}])L_f \Delta\epsilon^{-2})$ to find an ϵ stationary point of (SP)
- κ: condition number

upper iteration complexity bounds

- design a new inexact proximal gradient (IPG) method
- find an ϵ stationary point $(\mathbf{x}^*, \mathbf{y}^*)$ of (SP) within $\tilde{O}(\kappa([\mathbf{A}; \bar{\mathbf{A}}])L_f \Delta\epsilon^{-2})$
  $\Rightarrow$tight/near-optimal!
- $\mathbf{x}^*$ is arbitrary close to an ϵ stationary point of (P)
  $\Rightarrow$ find a near ϵ stationary point $\mathbf{x}^*$ of (P) within $\tilde{O}(\kappa([\mathbf{A}; \bar{\mathbf{A}}])L_f \Delta\epsilon^{-2})$
  $\Rightarrow$tight/near-optimal!

1. Lower bound complexity results

Variable $\mathbf{x}$, Matrices $\mathbf{A}$ and $\bar{\mathbf{A}}$, Vectors $\mathbf{b}$ and $\bar{\mathbf{b}}$, Function g

- $m = 3m_1m_2$ with $2 \mid m_1$, $n = (m - 3m_2)\bar{d}$, $\bar{n} = (3m_2 - 1)\bar{d}$
- $\mathbf{x} = \left(\mathbf{x}_1^\top, \ldots, \mathbf{x}_m^\top\right)^\top \in \mathbb{R}^d$, with $\mathbf{x}_i \in \mathbb{R}^{\bar{d}}$, $2 \nmid \bar{d}$, $\bar{d} \geq 5$, $d = m\bar{d}$
- $$\mathbf{J}_p := \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(p-1)\times p}.$$
- $\mathbf{H} := mL_f \cdot \mathbf{J}_m \otimes \mathbf{I}_{\bar{d}}$
- $\mathcal{M} := \{im_1 | i = 1, 2, \ldots, 3m_2 - 1\}$, $\mathcal{M}^C := \{1, 2, \ldots, m-1\} \backslash \mathcal{M}$
- $\bar{\mathbf{A}} := mL_f \cdot \mathbf{J}_{\mathcal{M}} \otimes \mathbf{I}_{\bar{d}}$, $\mathbf{A} := mL_f \cdot \mathbf{J}_{\mathcal{M}^C} \otimes \mathbf{I}_{\bar{d}}$, $\bar{\mathbf{b}} = \mathbf{0} \in \mathbb{R}^{\bar{n}}$, $\mathbf{b} = \mathbf{0} \in \mathbb{R}^n$
- $g(\mathbf{x}) := \beta \sum_{i \in \mathcal{M}} \|\mathbf{x}_i - \mathbf{x}_{i+1}\|_1$
- $\beta > (50\pi + 1 + \|\mathbf{A}\|)\sqrt{m}\epsilon$

Variable $\mathbf{x}$, Matrices $\mathbf{A}$ and $\bar{\mathbf{A}}$, Vectors $\mathbf{b}$ and $\bar{\mathbf{b}}$, Function g

- $m = 3m_1m_2$ with $2 \mid m_1$, $n = (m - 3m_2)\bar{d}$, $\bar{n} = (3m_2 - 1)\bar{d}$
- $\mathbf{x} = \left(\mathbf{x}_1^\top, \ldots, \mathbf{x}_m^\top\right)^\top \in \mathbb{R}^d$, with $\mathbf{x}_i \in \mathbb{R}^{\bar{d}}$, $2 \nmid \bar{d}$, $\bar{d} \geq 5$, $d = m\bar{d}$
- $$\mathbf{J}_p := \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(p-1)\times p}.$$
- $\mathbf{H} := mL_f \cdot \mathbf{J}_m \otimes \mathbf{I}_{\bar{d}}$
- $\mathcal{M} := \{im_1 | i = 1, 2, \ldots, 3m_2 - 1\}$, $\mathcal{M}^C := \{1, 2, \ldots, m-1\} \backslash \mathcal{M}$
- $\bar{\mathbf{A}} := mL_f \cdot \mathbf{J}_{\mathcal{M}} \otimes \mathbf{I}_{\bar{d}}$, $\mathbf{A} := mL_f \cdot \mathbf{J}_{\mathcal{M}^C} \otimes \mathbf{I}_{\bar{d}}$, $\bar{\mathbf{b}} = \mathbf{0} \in \mathbb{R}^{\bar{n}}$, $\mathbf{b} = \mathbf{0} \in \mathbb{R}^n$
- $g(\mathbf{x}) := \beta \sum_{i \in \mathcal{M}} \|\mathbf{x}_i - \mathbf{x}_{i+1}\|_1$
- $\beta > (50\pi + 1 + \|\mathbf{A}\|)\sqrt{m}\epsilon$

Variable $\mathbf{x}$, Matrices $\mathbf{A}$ and $\bar{\mathbf{A}}$, Vectors $\mathbf{b}$ and $\bar{\mathbf{b}}$, Function g

- $m = 3m_1m_2$ with $2 \mid m_1$, $n = (m - 3m_2)\bar{d}$, $\bar{n} = (3m_2 - 1)\bar{d}$
- $\mathbf{x} = \left(\mathbf{x}_1^\top, \ldots, \mathbf{x}_m^\top\right)^\top \in \mathbb{R}^d$, with $\mathbf{x}_i \in \mathbb{R}^{\bar{d}}$, $2 \nmid \bar{d}$, $\bar{d} \geq 5$, $d = m\bar{d}$
- $$\mathbf{J}_p := \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(p-1)\times p}.$$
- $\mathbf{H} := mL_f \cdot \mathbf{J}_m \otimes \mathbf{I}_{\bar{d}}$
- $\mathcal{M} := \{im_1 | i = 1, 2, \ldots, 3m_2 - 1\}$, $\mathcal{M}^C := \{1, 2, \ldots, m-1\} \backslash \mathcal{M}$
- $\bar{\mathbf{A}} := mL_f \cdot \mathbf{J}_{\mathcal{M}} \otimes \mathbf{I}_{\bar{d}}$, $\quad \mathbf{A} := mL_f \cdot \mathbf{J}_{\mathcal{M}^C} \otimes \mathbf{I}_{\bar{d}}$, $\quad \bar{\mathbf{b}} = \mathbf{0} \in \mathbb{R}^{\bar{n}}$, $\quad \mathbf{b} = \mathbf{0} \in \mathbb{R}^n$
- $g(\mathbf{x}) := \beta \sum_{i\in\mathcal{M}} \|\mathbf{x}_i - \mathbf{x}_{i+1}\|_1$
- $\beta > (50\pi + 1 + \|\mathbf{A}\|)\sqrt{m}\epsilon$

Function f_0

Inspired by [Sun-Hong 2019]

- $$f_0(\mathbf{x}) := \sum_{i=1}^{m} f_i(\mathbf{x}_i) := \sum_{i=1}^{m} \frac{300\pi\epsilon^2}{mL_f} h_i\left(\frac{\sqrt{m}L_f\mathbf{x}_i}{150\pi\epsilon}\right), \forall\, \mathbf{x} \in \mathbb{R}^d$$

- $$h_i(\mathbf{z}) := \begin{cases} l(\mathbf{z},1) + 3\sum_{j=1}^{\lfloor \bar{d}/2 \rfloor} l(\mathbf{z},2j) & \text{if } i \in \left[1, \frac{m}{3}\right] \\ l(\mathbf{z},1) & \text{if } i \in \left[\frac{m}{3}+1, \frac{2m}{3}\right] \\ l(\mathbf{z},1) + 3\sum_{j=1}^{\lfloor \bar{d}/2 \rfloor} l(\mathbf{z},2j+1) & \text{if } i \in \left[\frac{2m}{3}+1, m\right] \end{cases}$$

- $$l(\mathbf{z},j) := \begin{cases} -\Psi(1)\Phi(z_1) & \text{if } j = 1 \\ \Psi(-z_{j-1})\Phi(-z_j) - \Psi(z_{j-1})\Phi(z_j) & \text{if } j = 2,\ldots,\bar{d} \end{cases}$$

- $$\Psi(w) := \begin{cases} 0, & \text{if } w \le 0 \\ 1 - e^{-w^2}, & \text{if } w > 0 \end{cases} \quad \text{and} \quad \Phi(w) := 4\arctan w + 2\pi$$

Function f_0

Inspired by [Sun-Hong 2019]

- $$f_0(\mathbf{x}) := \sum_{i=1}^{m} f_i(\mathbf{x}_i) := \sum_{i=1}^{m} \frac{300\pi\epsilon^2}{mL_f} h_i\left(\frac{\sqrt{m}L_f\mathbf{x}_i}{150\pi\epsilon}\right), \forall\, \mathbf{x} \in \mathbb{R}^d$$

- $$h_i(\mathbf{z}) := \begin{cases} l(\mathbf{z},1) + 3\sum_{j=1}^{\lfloor \bar{d}/2 \rfloor} l(\mathbf{z},2j) & \text{if } i \in \left[1, \frac{m}{3}\right] \\ l(\mathbf{z},1) & \text{if } i \in \left[\frac{m}{3}+1, \frac{2m}{3}\right] \\ l(\mathbf{z},1) + 3\sum_{j=1}^{\lfloor \bar{d}/2 \rfloor} l(\mathbf{z},2j+1) & \text{if } i \in \left[\frac{2m}{3}+1, m\right] \end{cases}$$

- $$l(\mathbf{z},j) := \begin{cases} -\Psi(1)\Phi(z_1) & \text{if } j = 1 \\ \Psi(-z_{j-1})\Phi(-z_j) - \Psi(z_{j-1})\Phi(z_j) & \text{if } j = 2,\ldots,\bar{d} \end{cases}$$

- $$\Psi(w) := \begin{cases} 0, & \text{if } w \le 0 \\ 1 - e^{-w^2}, & \text{if } w > 0 \end{cases} \quad \text{and} \quad \Phi(w) := 4\arctan w + 2\pi$$

Worst-case Instance

Definition 1 (instance $\mathcal{P}$)

Given $\epsilon \in (0, 1)$ and $L_f > 0$, let m_1, m_2 and $\bar{d}$ be integers such that $m_1 \geq 2$ is even and $\bar{d} \geq 5$ is odd. We call as *instance* $\mathcal{P}$ the instance of problem (P) where f_0, g, $(\mathbf{A}, \mathbf{b})$, $(\bar{\mathbf{A}}, \bar{\mathbf{b}})$ are defined aforementioned.

Remark:

- condition number of $[\mathbf{A}; \bar{\mathbf{A}}] = \mathbf{H}$ proportional to m
- the instance $\mathcal{P}$ is difficult because of the coexistence of the regularization term and the affine constraints
- f_0 is nonconvex, and do not have a bounded level set

Lower bound complexity Result for (P)

Noting

$$\bar{d} \geq \frac{L_f\left(F_0\left(\mathbf{x}^{(0)}\right) - \inf_{\mathbf{x}} F_0(\mathbf{x})\right)}{3000\pi^2}\epsilon^{-2} = \frac{L_f \Delta_{F_0}}{3000\pi^2}\epsilon^{-2}$$

Theorem 2

Let $\epsilon > 0$ and $L_f > 0$ be given. Suppose an first-order algorithm is applied to problem (P) *and generates a sequence $\{\mathbf{x}^{(t)}\}_{t\geq 0}$. There exists an instance of problem* (P)*, i.e., instance $\mathcal{P}$, such that the algorithm requires at least*

$$t \geq \frac{\kappa([\bar{\mathbf{A}}; \mathbf{A}])L_f \Delta}{36000\pi^2}\epsilon^{-2}$$

iterations to obtain a near ϵ-stationary point of instance $\mathcal{P}$.

Hard Instance

Let $\bar{g}(\mathbf{y}) := \frac{\beta}{mL_f}\|\mathbf{y}\|_1$

Definition 3 (instance $\mathcal{P}$)

Given $\epsilon \in (0, 1)$ and $L_f > 0$, let m_1, m_2 and $\bar{d}$ be integers such that $m_1 \geq 2$ is even and $\bar{d} \geq 5$ is odd. We call as *instance* $\mathcal{P}$ the instance of problem (SP) where f_0, $\bar{g}$, $(\mathbf{A}, \mathbf{b})$, $(\bar{\mathbf{A}}, \bar{\mathbf{b}})$ are defined aforementioned.

Lower bound complexity Result for (SP)

Theorem 4

Let $\epsilon > 0$ and $L_f > 0$ be given. Suppose an first-order algorithm is applied to problem (SP) and generates a sequence $\{\mathbf{x}^{(t)}, \mathbf{y}^{(t)}\}_{t\geq 0}$. There exists an instance of problem (SP), i.e., instance $\mathcal{P}$, such that the algorithm requires at least

$$t \geq \frac{\kappa([\bar{\mathbf{A}}; \mathbf{A}])L_f \Delta}{72000\pi^2}\epsilon^{-2}$$

iterations to obtain an ϵ-stationary point of instance $\mathcal{P}$.

3. Upper bound complexity for Problem (SP)

An Inexact Proximal Gradient Method

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

A new inexact proximal gradient (IPG) method for problem (SP)

1. **Input:** A feasible solution $(\mathbf{x}^{(0)}, \mathbf{y}^{(0)})$ of problem (SP), $\sigma > 0$, and a proximal parameter $\tau > L_f$, set $k = 0$.
2. Inexactly solve subproblem

$$(\mathbf{x}^{(k+1)}, \mathbf{y}^{(k+1)}) \approx \arg\min_{\mathbf{x},\mathbf{y}} \left\langle \nabla f_0(\mathbf{x}^{(k)}), \mathbf{x} - \mathbf{x}^{(k)} \right\rangle + \frac{\tau}{2} \left\| \mathbf{x} - \mathbf{x}^{(k)} \right\|^2 + \bar{g}(\mathbf{y}) \qquad \text{s.t.} \;\; \mathbf{y} = \bar{\mathbf{A}}\mathbf{x} + \bar{\mathbf{b}}, \;\; \mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{0}. \tag{1}$$

3. set $k := k + 1$, return to step 2.

How to Solve Problem (1)

- Dual problem: convex and smooth

$$\min_{\mathbf{z}=(\mathbf{z}_1^\top,\mathbf{z}_2^\top)^\top} \frac{1}{2\tau}\|\bar{\mathbf{A}}^\top\mathbf{z}_1 + \mathbf{A}^\top\mathbf{z}_2 + \nabla f_0(\mathbf{x}^{(k)}) - \tau\mathbf{x}^{(k)}\|^2 + \bar{g}^\star(\mathbf{z}_1) - \mathbf{z}_1^\top\bar{\mathbf{b}} - \mathbf{z}_2^\top\mathbf{b}, \tag{2}$$

- if $[\bar{\mathbf{A}};\mathbf{A}]$ has a full-row rank
  $\Rightarrow$ the objective function in problem (2) is strongly convex
  $\Rightarrow$ using accelerated proximal gradient methods [Nesterov 2013]

- if $\bar{g}(\mathbf{y}) = \max\{\mathbf{u}^\top\mathbf{y} : \mathbf{C}\mathbf{u} \le \mathbf{d}, \mathbf{u} \in \mathbb{R}^{\bar{n}}\}$ for some $\mathbf{C}$ and $\mathbf{d}$
  $\Rightarrow$ the objective function in problem (2) has a quadratic growth
  $\Rightarrow$ restarted accelerated proximal gradient [Necoara et al. 2019]

Recover a Primer Solution from the Inexact Dual Solution

Let $\mathbf{z}^{(k+1)} = (\mathbf{z}_1^{(k+1)}, \mathbf{z}_2^{(k+1)})$ be a nearly optimal point of problem (2)

- $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \frac{1}{\tau}\left(\bar{\mathbf{A}}^\top \mathbf{z}_1^{(k+1)} + \mathbf{A}^\top \mathbf{z}_2^{(k+1)} + \nabla f_0(\mathbf{x}^{(k)})\right)$
- $\mathbf{y}^{(k+1)} = \mathbf{prox}_{\sigma^{-1}\bar{g}}\left(\sigma^{-1}\mathbf{z}_1^{(k+1)} + \bar{\mathbf{A}}\mathbf{x}^{(k+1)} + \bar{\mathbf{b}}\right).$

Upper Bound Complexity Results for (SP)

Theorem 5

Let $\tau = 2L_f$. Then, an ϵ-stationary point $(\mathbf{x}^K, \mathbf{y}^K)$ of problem (SP) *can be found within K main iterations of Algorithm IPG, where*

$$K := \left\lceil 12 L_f \Delta \epsilon^{-2} \right\rceil = O(L_f \Delta \epsilon^{-2})$$

- lower bound: $\frac{\kappa([\bar{\mathbf{A}};\mathbf{A}]) L_f \Delta}{72000\pi^2} \epsilon^{-2}$
- if $[\bar{\mathbf{A}}; \mathbf{A}]$ has a full-row rank
  $\Rightarrow$ accelerated proximal gradient methods [Nesterov 2013]
  $\Rightarrow$ Total complexity: $\tilde{O}(\kappa([\mathbf{A}; \bar{\mathbf{A}}]) L_f \Delta \epsilon^{-2}) \Rightarrow$ Near optimal!
- if $\bar{g}(\mathbf{y}) = \max\{\mathbf{u}^\top \mathbf{y} : \mathbf{C}\mathbf{u} \le \mathbf{d}, \mathbf{u} \in \mathbb{R}^{\bar{n}}\}$ for some $\mathbf{C}$ and $\mathbf{d}$, $\mathbf{b} = 0$, $\bar{\mathbf{b}} = 0$
  $\Rightarrow$ restarted accelerated proximal gradient [Necoara et al. 2019]
  $\Rightarrow$ Total complexity: $\tilde{O}(\kappa([\mathbf{A}; \bar{\mathbf{A}}]) L_f \Delta \epsilon^{-2}) \Rightarrow$ Near optimal!

Upper Bound Complexity Results for (P)

Theorem 6

Let $\tau = 2L_f$. Then, a near ϵ-stationary point $\mathbf{x}^K$ of problem (P) *can be found within K main iterations of Algorithm IPG, where*

$$K := \left\lceil 192 L_f \Delta \epsilon^{-2} \right\rceil = O(L_f \Delta \epsilon^{-2})$$

- lower bound: $\frac{\kappa([\bar{\mathbf{A}};\mathbf{A}])L_f\Delta}{36000\pi^2}\epsilon^{-2}$
- if $[\bar{\mathbf{A}};\mathbf{A}]$ has a full-row rank
  $\Rightarrow$ using accelerated proximal gradient methods [Nesterov 2013]
  $\Rightarrow$ Total complexity: $\tilde{O}(\kappa([\mathbf{A};\bar{\mathbf{A}}])L_f\Delta\epsilon^{-2}) \Rightarrow$ Near optimal!
- if $\bar{g}(\mathbf{y}) = \max\{\mathbf{u}^\top\mathbf{y} : \mathbf{C}\mathbf{u} \le \mathbf{d}, \mathbf{u} \in \mathbb{R}^{\bar{n}}\}$ for some $\mathbf{C}$ and $\mathbf{d}$, $\mathbf{b} = 0$, $\bar{\mathbf{b}} = 0$
  $\Rightarrow$ restarted accelerated proximal gradient [Necoara et al. 2019]
  $\Rightarrow$ Total complexity: $\tilde{O}(\kappa([\mathbf{A};\bar{\mathbf{A}}])L_f\Delta\epsilon^{-2}) \Rightarrow$ Near optimal!

Upper Bound Complexity Results for (P)

Theorem 6

Let $\tau = 2L_f$. *Then, a near* ϵ*-stationary point* $\mathbf{x}^K$ *of problem* (P) *can be found within* K *main iterations of Algorithm IPG, where*

$$K := \left\lceil 192 L_f \Delta \epsilon^{-2} \right\rceil = O(L_f \Delta \epsilon^{-2})$$

- lower bound: $\frac{\kappa([\bar{\mathbf{A}};\mathbf{A}])L_f\Delta}{36000\pi^2}\epsilon^{-2}$
- if $[\bar{\mathbf{A}};\mathbf{A}]$ has a full-row rank
  $\Rightarrow$ using accelerated proximal gradient methods [Nesterov 2013]
  $\Rightarrow$ Total complexity: $\tilde{O}(\kappa([\mathbf{A};\bar{\mathbf{A}}])L_f\Delta\epsilon^{-2}) \Rightarrow$ Near optimal!
- if $\bar{g}(\mathbf{y}) = \max\{\mathbf{u}^\top\mathbf{y} : \mathbf{C}\mathbf{u} \le \mathbf{d}, \mathbf{u} \in \mathbb{R}^{\bar{n}}\}$ for some $\mathbf{C}$ and $\mathbf{d}$, $\mathbf{b} = 0$, $\bar{\mathbf{b}} = 0$
  $\Rightarrow$ restarted accelerated proximal gradient [Necoara et al. 2019]
  $\Rightarrow$ Total complexity: $\tilde{O}(\kappa([\mathbf{A};\bar{\mathbf{A}}])L_f\Delta\epsilon^{-2}) \Rightarrow$ Near optimal!

Conclusions and Future Work

Conclusions

- established lower complexity bounds of affinely constrained nonconvex nonsmooth problems
- prove the tightness of the lower complexity results

Future Work

- determine whether an ϵ stationary point of problem (P) can be achieved within $O(\epsilon^{-2})$ iterations
- determine whether the lower bound complexity for problem (SP) can be achieved by an algorithm with **weaker assumptions**

Conclusions and Future Work

Conclusions

- established lower complexity bounds of affinely constrained nonconvex nonsmooth problems
- prove the tightness of the lower complexity results

Future Work

- determine whether an ϵ stationary point of problem (P) can be achieved within $O(\epsilon^{-2})$ iterations
- determine whether the lower bound complexity for problem (SP) can be achieved by an algorithm with **weaker assumptions**

Thanks for your listening!

Email: liuw16@rpi.edu