Lower complexity bounds of first-order methods for convex-concave bilinear saddle-point problems

Yuyuan Ouyang
Mathematical and Statistical Sciences, Clemson University

Yangyang Xu
Mathematical Sciences, Rensselaer Polytechnic Institute

August 06, 2019

What and why first-order methods

- First-order method only inquires gradient and/or function value information of a problem, and possibly other simple operations
- Generally, lower per-update complexity, lower memory requirement, and better scalability compared to second or higher order methods
- Difficult to achieve very high accuracy
- Favorable for very "big" problems that do not require high accuracy

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

First-order methods for smooth convex problems [Nesterov'04]

Consider problem

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\, f(\mathbf{x})$$

- 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 complexity bound: $f(\mathbf{x}^k) - f(\mathbf{x}^*) \ge \frac{3L\|\mathbf{x}^0-\mathbf{x}^*\|^2}{32(k+1)^2}$ if $k \le \frac{n-1}{2}$ and

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

- upper complexity bound: $f(\mathbf{x}^k) - f(\mathbf{x}^*) \le \frac{4L\|\mathbf{x}^0-\mathbf{x}^*\|^2}{(k+1)^2}$

First-order methods for nonsmooth convex problems [Nesterov'04]

Consider problem

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\, f(\mathbf{x})$$

- f is convex and M-Lipschitz continuous on $X = \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}^0\| \le R\}$, i.e.,

$$|f(\mathbf{x}) - f(\mathbf{y})| \le M\|\mathbf{x} - \mathbf{y}\|, \,\forall\, \mathbf{x}, \mathbf{y} \in X$$

- lower complexity bound: $f(\mathbf{x}^k) - f(\mathbf{x}^*) \ge \frac{MR}{2(1+\sqrt{k+1})}$ if $k \le n-1$, and

$$\mathbf{x}^k \in \mathbf{x}^0 + \text{span}\{\mathbf{g}^0, \mathbf{g}^1, \ldots, \mathbf{g}^{k-1}\}$$

  where $\mathbf{g}^t \in \partial f(\mathbf{x}^t)$

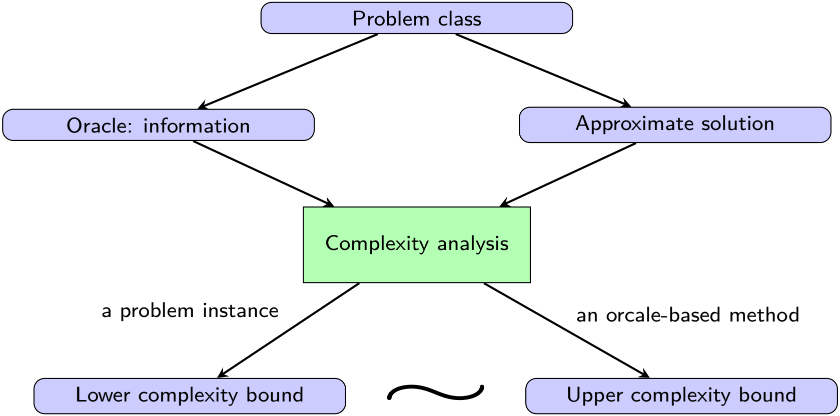
- upper complexity bound: $f(\mathbf{x}^k) - f(\mathbf{x}^*) \le \frac{2MR}{\sqrt{k+1}}$

Remark: non-smooth problems are harder than smooth ones.

More examples

- first-order methods for stochastic convex problems [Agarwal et. al'12]
- first-order methods for finite-sum convex problems [Woodworth-Srebro'16]
- first-order and higher-order methods for nonconvex problems [Carmon et. al'17a, Carmon et. al'17b]
- $\cdots\cdots$

Diagram: iteration complexity analysis



This talk: convex-concave bilinear saddle-point problem

Problem setup:

$$\min_{\mathbf{x}\in X}\Big\{f(\mathbf{x})+\max_{\mathbf{y}\in Y}\langle \mathbf{A}\mathbf{x}-\mathbf{b},\mathbf{y}\rangle - g(\mathbf{y})\Big\}$$

where $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ are simple closed convex sets, and f and g are closed convex functions.

Assumptions:

- ∇f is L_f-Lipschitz: $\|\nabla f(\mathbf{u}) - \nabla f(\mathbf{x})\| \leq L_f\|\mathbf{u}-\mathbf{x}\|, \forall \mathbf{u},\mathbf{x}\in X$
- g is a proximable function, i.e., easy proximal mapping $\mathbf{prox}_{\eta g}$
- large-scale: information of f through $f(\mathbf{x}), \nabla f(\mathbf{x})$, information of $\mathbf{A}$ through $\mathbf{A}\mathbf{x}$ and $\mathbf{A}^\top\mathbf{y}$ for inquiry $(\mathbf{x},\mathbf{y})$

Special cases:

- If $X = \mathbb{R}^n, Y = \mathbb{R}^m$, $g \equiv 0$: linearly constrained smooth convex optimization

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

- If $Y = \mathrm{dom}(g)$:

$$\min_{\mathbf{x} \in X} \phi(\mathbf{x}) := f(\mathbf{x}) + g^*(\mathbf{A}\mathbf{x} - \mathbf{b})$$

  convex composite optimization with two components; ϕ is usually nonsmooth due to g^*

Existing rate by smoothing [Nesterov'05]

Assume X and Y are both compact. Let

$$\phi(\mathbf{x}) = f(\mathbf{x}) + \max_{\mathbf{y}\in Y}\langle \mathbf{A}\mathbf{x} - \mathbf{b}, \mathbf{y}\rangle - g(\mathbf{y}),$$

$$\psi(\mathbf{y}) = \min_{\mathbf{x}\in X} f(\mathbf{x}) + \langle \mathbf{A}\mathbf{x} - \mathbf{b}, \mathbf{y}\rangle - g(\mathbf{y}).$$

Then there is a first-order method such that

$$0 \le \phi(\mathbf{x}^k) - \psi(\mathbf{y}^k) \le \frac{4L_f D_X^2}{(k+1)^2} + \frac{4D_X D_Y \|\mathbf{A}\|_2}{k+1}$$

- better than $O(1/\sqrt{k})$ that is a lower bound for nonsmooth problems
- but worse than $O(1/k^2)$ that is an upper bound for smooth problems
- unknown before if the rate is optimal

Rate for nonsmooth linearly constrained problems [X.'17]

For the problem

$$\underset{\mathbf{x}}{\text{minimize}}\, f(\mathbf{x}) + g(\mathbf{x}), \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{b},$$

assume exact solvability to subproblem in the form of

$$\min_{\mathbf{x}} \langle \nabla f(\hat{\mathbf{x}}) + \mathbf{z}, \mathbf{x} \rangle + g(\mathbf{x}) + \frac{\beta}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \frac{\eta}{2}\|\mathbf{x}\|^2.$$

- Can have $\{\mathbf{x}^k\}$ such that

$$|f(\mathbf{x}^k) - f(\mathbf{x}^*)| + \|\mathbf{A}\mathbf{x}^k - \mathbf{b}\| = O\left(\frac{L_f\|\mathbf{x}^0 - \mathbf{x}^*\|^2 + \frac{\|\mathbf{y}^0 - \mathbf{y}^*\|^2}{\gamma}}{k^2}\right)$$

  where $\gamma > 0$ is a constant, and k is the number of solved subproblems
- $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$ can be further prox-linearized to achieve $O(1/k^2)$ if strong convexity assumed
- unknown before if $O(1/k^2)$ can be achieved with just convexity

roadmap

lower iteration complexity bounds

1. affinely constrained problems with iterate in spanned space of gradient
2. affinely constrained problems by any method based on first-order oracle
3. bilinear saddle-point problems by any method based on first-order oracle

comparison to existing upper complexity bounds

Case I: linearly constrained problems by linear span

linearly constrained problems by linear span

First consider

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

- problem class I: ∇f is L_f-smooth, and $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $\|\mathbf{A}\|_2 = L_A > 0$
- problem class II: f is μ-strongly convex, and $\mathbf{A}$ with $\|\mathbf{A}\|_2 = L_A > 0$
- algorithm class:

$$\mathbf{x}^t \in \mathbf{x}^0 + \mathrm{Span}\{\nabla f(\mathbf{x}^0), \mathbf{A}^\top \mathbf{r}^0, \ldots, \nabla f(\mathbf{x}^{t-1}), \mathbf{A}^\top \mathbf{r}^{t-1}\} \qquad \text{(Span)}$$

  where $\mathbf{r}^t = \mathbf{A}\mathbf{x}^t - \mathbf{b}$
  - Without loss of generality, assume that $\mathbf{x}^0 = \mathbf{0}$
- error measure: $|f(\mathbf{x}^t) - f^*|$ and $\|\mathbf{A}\mathbf{x}^t - \mathbf{b}\|$, or $\|\mathbf{x}^t - \mathbf{x}^*\|^2$

lower complexity bound for convex case [Ouyang-X.'18]

Setting of problem class:

- given positive integers $m \le n$, and $t < \frac{m}{2}$
- given positive numbers L_A and L_f

Conclusion: there exists an instance of smooth linearly constrained problem such that

- ∇f is L_f-Lipschitz continuous, $\|\mathbf{A}\|_2 = L_A$
- it has a unique primal-dual solution $(\mathbf{x}^*, \mathbf{y}^*)$
- in addition, for (Span), it holds

$$|f(\mathbf{x}^t) - f(\mathbf{x}^*)| \ge \frac{3L_f\|\mathbf{x}^*\|^2}{64(t+1)^2} + \frac{\sqrt{3}L_A\|\mathbf{x}^*\| \cdot \|\mathbf{y}^*\|}{16(t+1)},$$

$$\|\mathbf{A}\mathbf{x}^t - \mathbf{b}\| \ge \frac{\sqrt{3}L_A\|\mathbf{x}^*\|}{4\sqrt{2}(t+1)}.$$

Worst-case instance

$$\underset{\mathbf{x}}{\text{minimize}} \frac{1}{2}\mathbf{x}^\top \mathbf{H}\mathbf{x} - \mathbf{h}^\top \mathbf{x}, \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}. \qquad \text{(QP-Inst)}$$

Here,

$$\mathbf{H} = \frac{L_f}{4}\begin{bmatrix} \mathbf{B}^\top \mathbf{B} & \\ & \mathbf{I}_{n-2k} \end{bmatrix} \in \mathbb{R}^{n\times n}, \mathbf{h} = \frac{L_f}{2}\mathbf{e}_{2k,n}, \mathbf{A} = \frac{L_A}{2}\mathbf{\Lambda}, \mathbf{b} = \frac{L_A}{2}\mathbf{c},$$

and

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{B} & \mathbf{O} \\ \mathbf{O} & \mathbf{G} \end{bmatrix} \in \mathbb{R}^{m\times n},\ \mathbf{c} = \begin{bmatrix} \mathbf{1}_{2k} \\ \mathbf{0} \end{bmatrix},\ \mathbf{B} := \begin{bmatrix} & & & -1 & 1 \\ & & \iddots & \iddots & \\ -1 & 1 & & & \\ 1 & & & & \end{bmatrix} \in \mathbb{R}^{2k\times 2k}$$

with $\mathbf{G} \in \mathbb{R}^{(m-2k)\times(n-2k)}$ is any matrix of full row rank such that $\|\mathbf{G}\| = 2$.

Remark: condition number of $\mathbf{B}$ proportional to k

Sketch of proof

1. primal-dual solution $(\mathbf{x}^*, \mathbf{y}^*)$:

$$x_i^* = \begin{cases} i, & \text{if } 1 \le i \le 2k, \\ 0, & \text{if } i \ge 2k+1, \end{cases} \quad y_i^* = \begin{cases} -\frac{L_f}{2L_A} & \text{if } 1 \le i \le 2k \\ 0 & \text{if } i \ge 2k+1. \end{cases}$$

2. optimal objective: $f^* = -\frac{3L_f}{4}k$
3. property of iterate: if $\mathbf{x}^0 = \mathbf{0}$, then $\mathbf{x}^t \in \mathcal{K}_{k-1}$ for any $t \le k$, where

$$\mathcal{J}_i := \mathrm{Span}\{\mathbf{c}, (\mathbf{\Lambda}\mathbf{\Lambda}^\top)\mathbf{c}, (\mathbf{\Lambda}\mathbf{\Lambda}^\top)^2\mathbf{c}, \ldots, (\mathbf{\Lambda}\mathbf{\Lambda}^\top)^i\mathbf{c}\}$$
$$\mathcal{K}_i := \mathbf{\Lambda}^\top \mathcal{J}_i = \mathrm{Span}\{\mathbf{e}_{2k-i,n}, \mathbf{e}_{2k-i+1,n}, \ldots, \mathbf{e}_{2k,n}\}$$

4. objective value and feasibility at points in $\mathcal{K}_{k-1}$:

$$\min_{\mathbf{x} \in \mathcal{K}_{k-1}} f(\mathbf{x}) - f^* \ge \frac{3L_f \|\mathbf{x}^*\|^2}{64(k+1)^2} + \frac{\sqrt{3}L_A \|\mathbf{x}^*\| \cdot \|\mathbf{y}^*\|}{16(k+1)},$$

$$\min_{\mathbf{x} \in \mathcal{K}_{k-1}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\| \ge \frac{\sqrt{3}L_A \|\mathbf{x}^*\|}{4\sqrt{2}(k+1)}.$$

lower complexity bound for strongly convex case [Ouyang-X.'18]

Setting of problem class:

- given positive integers $m \leq n$, and $t < \frac{m}{2}$
- given positive numbers L_A and μ

Conclusion: there exists an instance of smooth linearly constrained problem such that

- ∇f is μ-strongly convex, $\|\mathbf{A}\|_2 = L_A$
- it has a unique primal-dual solution $(\mathbf{x}^*, \mathbf{y}^*)$
- in addition, for (Span), it holds

$$\|\mathbf{x}^t - \mathbf{x}^*\|^2 \geq \frac{5L_A^2\|\mathbf{y}^*\|^2}{256\mu^2(t+1)^2}.$$

Worst-case instance

$$\underset{\mathbf{x}}{\text{minimize}} \, \frac{\mu}{2}\mathbf{x}^\top \mathbf{x}, \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}$$

where $\mathbf{A}$ and $\mathbf{b}$ are the same as in (QP-Inst).

Sketch of proof:

1. primal-dual solution $(\mathbf{x}^*, \mathbf{y}^*)$:

$$x_i^* = \begin{cases} i, & \text{if } 1 \le i \le 2k, \\ 0, & \text{if } i \ge 2k+1, \end{cases} \quad y_i^* = \begin{cases} \frac{\mu}{L_A} i(4k - i + 1), & \text{if } 1 \le i \le 2k, \\ 0, & \text{if } i \ge 2k+1. \end{cases}$$

2. property of iterate: if $\mathbf{x}^0 = \mathbf{0}$, then $\mathbf{x}^t \in \mathcal{K}_{k-1}$ for any $t \le k$
3. distance of iterate to optimal solution: $\|\mathbf{x}^k - \mathbf{x}^*\|^2 \ge \sum_{i=1}^k i^2$

Case II: linearly constrained problems by general first-order methods

linearly constrained problems by general first-order methods

Still consider

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

- problem class I: ∇f is L_f-smooth, and $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $\|\mathbf{A}\|_2 = L_A > 0$
- problem class II: f is μ-strongly convex, and $\mathbf{A}$ with $\|\mathbf{A}\|_2 = L_A > 0$
- algorithm class ($\{\mathcal{I}_t\}$ is a sequence of **fixed** rules):

$$\left(\mathbf{x}^{t+1}, \mathbf{y}^{t+1}\right) = \mathcal{I}_t\left(\boldsymbol{\theta}; \mathcal{O}(\mathbf{x}^0, \mathbf{y}^0), \ldots, \mathcal{O}(\mathbf{x}^t, \mathbf{y}^t)\right), \ \forall\, t \geq 0, \qquad \text{(FOM)}$$

  where $\mathcal{O}(\mathbf{x}, \mathbf{y}) := \left(\nabla f(\mathbf{x}), \mathbf{A}\mathbf{x}, \mathbf{A}^\top \mathbf{y}\right)$.
- error measure: $|f(\mathbf{x}^t) - f^*|$ and $\|\mathbf{A}\mathbf{x}^t - \mathbf{b}\|$, or $\|\mathbf{x}^t - \mathbf{x}^*\|^2$

lower complexity bound for convex case [Ouyang-X.'18]

Setting of problem class:

- given positive integers $m \leq n$, and $t < \frac{m}{4} - 1$
- given positive numbers L_A and L_f

Conclusion: for (FOM), there exists an instance of smooth linearly constrained problem such that

- ∇f is L_f-Lipschitz continuous, $\|\mathbf{A}\| = L_A$
- it has a unique primal-dual solution $(\mathbf{x}^*, \mathbf{y}^*)$
- in addition, it holds

$$f(\mathbf{x}^t) - f^* \geq \frac{3L_f\|\mathbf{x}^*\|^2}{64(2t+5)^2} + \frac{\sqrt{3}L_A\|\mathbf{x}^*\| \cdot \|\mathbf{y}^*\|}{16(2t+5)},$$

$$\|\mathbf{A}\mathbf{x}^t - \mathbf{b}\| \geq \frac{\sqrt{3}L_A\|\mathbf{x}^*\|}{4\sqrt{2}(2t+5)}.$$

Key proposition

Proposition 3.1 [Ouyang-X.'18]: let $\min_{\mathbf{x}}\{f(\mathbf{x}),\ \text{s.t. } \mathbf{A}\mathbf{x}=\mathbf{b}\}$ be one instance. There is a rotated instance $\min_{\mathbf{x}}\{\tilde{f}(\mathbf{x}),\ \text{s.t. } \tilde{\mathbf{A}}\mathbf{x}=\mathbf{b}\}$ with

$$\tilde{f}(\mathbf{x}) = f(\mathbf{U}\mathbf{x}),\ \tilde{\mathbf{A}} = \mathbf{V}^\top \mathbf{A}\mathbf{U},\ \mathbf{V}\mathbf{b} = \mathbf{b},$$

where $\mathbf{V}$ and $\mathbf{U}$ are orthogonal. In addition,

- $(\mathbf{x}^*, \mathbf{y}^*)$ is a primal-dual solution to the *original instance* if and only if $(\hat{\mathbf{x}}, \hat{\mathbf{y}}) := (\mathbf{U}^\top \mathbf{x}^*, \mathbf{V}^\top \mathbf{y}^*)$ is a primal-dual solution to the *rotated instance*.
- When (FOM) applied to the rotated instance, for any $1 \le t \le \frac{k}{2} - 1$:

$$\left|\tilde{f}(\mathbf{x}^t) - \tilde{f}^*\right| \ge \min_{\mathbf{x}\in\mathcal{K}_{k-1}} \left|f(\mathbf{x}) - f^*\right|,$$

$$\|\tilde{\mathbf{A}}\mathbf{x}^t - \mathbf{b}\| \ge \min_{\mathbf{x}\in\mathcal{K}_{k-1}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|,$$

$$\|\mathbf{x}^t - \hat{\mathbf{x}}\|^2 \ge \min_{\mathbf{x}\in\mathcal{K}_{k-1}} \|\mathbf{x} - \mathbf{x}^*\|^2,$$

Key idea of proving proposition: rotation

- Without linear span assumption, the property $\mathbf{x}^t \in \mathcal{K}_{t-1}$ does not hold any more.
- **Fact:** $\mathcal{X} \subsetneq \bar{\mathcal{X}} \subseteq \mathbb{R}^p$ be two subspaces. For any $\bar{\mathbf{x}} \in \mathbb{R}^p$, there is an orthogonal $\mathbf{V}$ such that

$$\mathbf{V}\mathbf{x} = \mathbf{x}, \forall \mathbf{x} \in \mathcal{X}, \ \mathbf{V}\bar{\mathbf{x}} \in \bar{\mathbf{X}}$$

- If $\mathbf{x}^t \notin \mathcal{K}_{t-1}$, rotate it by an orthogonal matrix $\mathbf{U}_t$ such that $\mathbf{U}_t\mathbf{x}^t \in \mathcal{K}_{t-1}$ and thus $\mathbf{x}^t \in \mathbf{U}_t^\top \mathcal{K}_{t-1}$
- similarly, apply rotation to dual iterate
- repeatedly use the **Fact** and rotation:

$$\mathbf{x}^i \in \mathbf{U}^\top \mathcal{K}_{2t+1}, \ \mathbf{y}^i \in \mathbf{V}^\top \mathcal{J}_{2t+1}, \forall i \leq t$$

lower complexity bound for strongly convex case [Ouyang-X.'18]

Setting of problem class:

- given positive integers $m \le n$, and $t < \frac{m}{4} - 1$
- given positive numbers L_A and μ

Conclusion: for (FOM), there exists an instance of smooth linearly constrained problem such that

- ∇f is μ-strongly convex, $\|\mathbf{A}\| = L_A$
- it has a unique primal-dual solution $(\mathbf{x}^*, \mathbf{y}^*)$
- in addition, it holds

$$\|\mathbf{x}^t - \mathbf{x}^*\|^2 \ge \frac{5L_A^2\|\mathbf{y}^*\|^2}{256\mu^2(2t+5)^2}$$

Case III: saddle-point problems by general first-order methods

saddle-point problems by general first-order methods

Finally consider

$$\min_{\mathbf{x} \in X} \phi(\mathbf{x}) := \Big\{ f(\mathbf{x}) + \max_{\mathbf{y} \in Y} \langle \mathbf{A}\mathbf{x} - \mathbf{b}, \mathbf{y} \rangle - g(\mathbf{y}) \Big\}$$

where g is closed convex and proximable.

- problem class I: ∇f is L_f-smooth, and $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $\|\mathbf{A}\|_2 = L_A > 0$
- problem class II: f is μ-strongly convex, and $\mathbf{A}$ with $\|\mathbf{A}\|_2 = L_A > 0$
- algorithm class: general oracle-based first-order method (FOM)
- error measure: primal-dual gap $\phi(\mathbf{x}^t) - \psi(\mathbf{y}^t)$

lower complexity for convex saddle-point problems [Ouyang-X.'18]

Setting of problem class:

- given positive integers $m \leq n$, and $t < \frac{m}{4} - 1$
- given positive numbers L_A and L_f

Conclusion: for (FOM), there exists an instance of convex-concave bilinear saddle-point problem such that

- ∇f is L_f-Lipschitz continuous, $\|\mathbf{A}\| = L_A$
- X and Y are Euclidean balls with radii R_X and R_Y
- it has a unique primal-dual solution $(\mathbf{x}^*, \mathbf{y}^*)$
- in addition, it holds

$$\phi(\mathbf{x}^{(t)}) - \psi(\mathbf{y}^{(t)}) \geq \frac{L_f R_X^2}{4(4t+5)^2} + \frac{L_A R_X R_Y}{4(4t+5)},$$

where ϕ and ψ are the associated primal and dual objective functions.

Idea of proof

- Note that R_X and R_Y are not fixed.
- Set $g(\mathbf{y}) = \lambda\|\mathbf{y}\|$ and choose R_X, R_Y such that the primal-dual solution $(\mathbf{x}^*, \mathbf{y}^*)$ of previous worst-case of linearly constrained problems is in $X \times Y$.
- Hence, $(\mathbf{x}^*, \mathbf{y}^*)$ is also the solution of the saddle-point problem

$$\min_{\mathbf{x}\in X} \max_{\mathbf{y}\in Y} \left\{ f(\mathbf{x}) + \langle \mathbf{A}\mathbf{x} - \mathbf{b}, \mathbf{y} \rangle \right\}$$

- Also, note $\phi^* \leq f(\mathbf{x}^*) + \langle \mathbf{A}\mathbf{x}^* - \mathbf{b}, \mathbf{y}^* \rangle = f(\mathbf{x}^*)$
- Choose $\lambda > 0$: $\phi(\mathbf{x}) = f(\mathbf{x}) + R_Y(\|\mathbf{A}\mathbf{x} - \mathbf{b}\| - \lambda)$ for all $\mathbf{x} \in \mathcal{K}_{k-1}$
- Then estimate the bound $\min_{\mathbf{x}\in\mathcal{K}_{k-1}} \phi(\mathbf{x}) - \phi^*$

lower complexity for strongly convex SP problems [Ouyang-X.'18]

Setting of problem class:

- given positive integers $m \le n$, and $t < \frac{m}{4} - 1$
- given positive numbers L_A and μ

Conclusion: for (FOM), there exists an instance of convex-concave bilinear saddle-point problem such that

- ∇f is μ-strongly convex, $\|\mathbf{A}\| = L_A$
- X and Y are Euclidean balls with radii R_X and R_Y
- it has a unique primal-dual solution $(\mathbf{x}^*, \mathbf{y}^*)$
- in addition, it holds

$$\phi(\mathbf{x}^t) - \psi(\mathbf{y}^t) \ge \frac{5L_A^2 R_Y^2}{512\mu(4t+5)^2},$$

existing upper complexity bounds

Near tightness of established bounds

1. For linearly constrained problems, the rate of an accelerated linearized ADMM [Ouyang et. al'15] is

$$f(\mathbf{x}^t) - f^* \leq \frac{2L_f D_X^2}{t(t+1)} + \frac{2D_X D_Y \|\mathbf{A}\|_2}{t+1}$$

2. For smooth and strongly convex equality and inequality constrained problems, to have an ε-solution, i.e., $|f(\bar{\mathbf{x}}) - f^*| \leq \varepsilon$ and $\|\mathbf{A}\bar{\mathbf{x}} - \mathbf{b}\| + \|[\mathbf{f}(\bar{\mathbf{x}})]_+\| \leq \varepsilon$, the iteration complexity by inexact ALM [X.'17] is

$$t \leq 2\left(\sqrt{\frac{L_f}{\mu}} + \frac{2L_A\|\mathbf{y}^*\|}{\mu\sqrt{\varepsilon}}\right)\left(O(1) + \log\frac{1}{\varepsilon}\right)$$

3. For bilinear saddle-point problems, the rate of a first-order method by smoothing [Nesterov'05] is

$$\phi(\mathbf{x}^k) - \psi(\mathbf{y}^k) \leq \frac{4L_f D_X^2}{(k+1)^2} + \frac{4D_X D_Y \|\mathbf{A}\|_2}{k+1}$$

Open questions

1. How should feasibility lower bound depend on constraint and also objective?
2. If the number of gradient evaluation is constrained, what is the lower bound of matrix-vector multiplication $\mathbf{A}\mathbf{x}, \mathbf{A}^\top \mathbf{y}$?
3. How should the lower bound depend on condition number for strongly convex case?

Conclusions

1. reviewed lower complexity bounds of first-order methods for a few classes of problems
2. established lower complexity bounds of
   - gradient type first-order method for linearly constrained problems
   - general first-order method for linearly constrained problems
   - general first-order method for bilinear saddle-point problems
3. showed near tightness of the established lower bounds

References

Y. Ouyang and Y. Xu. Lower complexity bound of first-order methods for bilinear saddle-point problems, Accepted in *MPA*, arXiv:1808.02901, 2018.

Thank you!!!