A Proximal Bundle Variant with Optimal Iteration-Complexity for a Large Range of Prox Stepsizes

OP21 – MS56 Recent Developments in First-Order Methods for Composite Optimization

Jiaming Liang

School of Industrial and Systems Engineering
Georgia Institute of Technology

Joint work with Renato Monteiro

July 21, 2021

Introduction

Main problem:

$$\phi_* := \min \{\phi(x) := f(x) + h(x) : x \in \mathbb{R}^n\} \tag{1}$$

Main goal:
To show the iteration-complexity of the relaxed proximal bundle (RPB) method is optimal.
Main techniques:

- Inexact proximal point framework
- Bundle method

Convex nonsmooth problem

Consider (1), where

(A1) functions $f, h \in \overline{\mathrm{Conv}}\,(\mathbb{R}^n)$ are such that $\mathrm{dom}\, h \subset \mathrm{dom}\, f$ and function $f' : \mathrm{dom}\, h \to \mathbb{R}^n$ is such that $f'(x) \in \partial f(x)$ for all $x \in \mathrm{dom}\, h$;

(A2) the set of optimal solutions X^* of problem (1) is nonempty;

(A3) h is μ-convex and $\|f'(x)\| \leq M_f$ for all $x \in \mathrm{dom}\, h$;

(A4) h is M_h-Lipschitz continuous on $\mathrm{dom}\, h$, i.e.,

$$|h(u) - h(v)| \leq M_h \|u - v\| \quad \forall u, v \in \mathrm{dom}\, h.$$

1. Introduction
   - Assumptions
   - Literature review
2. The RPB method
   - Review of the composite subgradient method
   - Bundle method
   - RPB
   - CS as an instance of RPB
3. Main results
   - Upper complexity bounds
   - Complexity bounds for another proximal bundle variant
4. Optimal complexity
   - Lower complexity bounds
   - Optimal complexity bounds
5. Conclusion

Lower complexity bound results

- Convex, unconstrained, $\min f(x)$

$$\Omega\left(\frac{M_f^2 d_0^2}{\bar{\varepsilon}^2}\right)$$

where $d_0 := \inf\{\|x_0 - x_*\| : x_* \in X_*\}$ and $\bar{\varepsilon}$ is the tolerance.

- Strongly convex, unconstrained, $\min f(x)$

$$\Omega\left(\frac{M_f^2}{\mu\bar{\varepsilon}}\right)$$

where μ is the strong convexity of f.

Drawback: bounds are inconsistent when $\mu \to 0$

Upper bound complexity results

- Subgradient, Mirror descent and Bundle-level method are optimal.
- Bundle method
  - convex, Kiwiel 2000

$$\mathcal{O}_1\left(\frac{\tilde{M}^2\tilde{D}^4}{\lambda\bar{\varepsilon}^3}\right)$$

where $\tilde{D} = \tilde{D}[\tilde{f}] := \sup\{d(x_j, X^*) : j \geq 0\}$,
$\tilde{M} = \tilde{M}[\tilde{f}] := \sup\{\|\tilde{f}'(x_j)\| : j \geq 0\}$.

  - μ-strongly convex, Du and Ruszczyński 2017

$$\tilde{\mathcal{O}}_1\left(\frac{\tilde{M}^2\lambda}{\alpha^2\bar{\varepsilon}}\right)$$

where $\alpha := \min\{\lambda\mu, 1\}$.

Drawback: bounds are not optimal in general (i.e., for a large range of prox stepsizes λ)

Composite subgradient (CS) method

$$x_j = \underset{u\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ f(x_{j-1}) + \langle f'(x_{j-1}), u - x_{j-1}\rangle + h(u) + \frac{1}{2\lambda}\|u - x_{j-1}\|^2 \right\}$$

Theorem

For any given universal constant $C > 1$, CS with any stepsize λ such that $\bar{\varepsilon}/(CM_f^2) \le 4\lambda \le \bar{\varepsilon}/M_f^2$ has $\bar{\varepsilon}$-iteration complexity bound given by

$$\mathcal{O}_1\left(\min\left\{\frac{M_f^2 d_0^2}{\bar{\varepsilon}^2}, \left(\frac{M_f^2}{\mu\bar{\varepsilon}} + 1\right)\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\}\right) \tag{2}$$

where $d_0 = \inf\{\|x_0 - x^*\| : x^* \in X^*\} = \|x_0 - x_0^*\|$.

Bundle method

Solving the proximal problem

$$x^{+} \leftarrow \min_{u \in \mathbb{R}^n} \left\{ \phi(u) + \frac{1}{2\lambda} \|u - x\|^2 \right\} \tag{3}$$

can be as difficult as solving $\min\{\phi(u) : u \in \mathbb{R}^n\}$.

Bundle method approximately solves (3) and recursively builds up a model by using a standard cutting-plane approach.

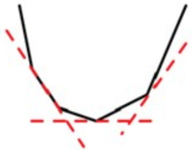
Bundle method

The **bundle method** solves a sequence of prox subproblems of the form

$$x_j = \underset{u\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ \Gamma_j^\lambda(u) := f_j(u) + h(u) + \frac{1}{2\lambda}\|u - x_{j-1}^c\|^2 \right\}, \qquad (4)$$

where x_{j-1}^c is the **prox-center**, f_j is the **cutting-plane** model defined as

$$f_j(u) = \max\{f(x) + \langle f'(x), u - x\rangle : x \in C_j\} \quad \forall u \in \mathbb{R}^n.$$



Bundle method

The **bundle method** solves a sequence of prox subproblems of the form

$$x_j = \underset{u\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ \Gamma_j^\lambda(u) := f_j(u) + h(u) + \frac{1}{2\lambda}\|u - x_{j-1}^c\|^2 \right\},$$

where x_{j-1}^c is the **prox-center**, f_j is the **cutting-plane** model defined as

$$f_j(u) = \max\{f(x) + \langle f'(x), u - x\rangle : x \in C_j\} \quad \forall u \in \mathbb{R}^n,$$

and decides to perform a **serious** or **null** iteration based on the **descent condition** $\phi(x_j) \le (1-\gamma)\phi\left(x_{j-1}^c\right) + \gamma\left(f_j + h\right)(x_j)$ for some $\gamma \in (0,1)$.

Relaxed proximal bundle (RPB) method

0. Let $x_0 \in \mathrm{dom}\, h$, $\lambda > 0$ and $\bar{\varepsilon} > 0$ be given, and set $x_0^c = x_0$, $C_1 = \{x_0\}$, and $j = 1$;

1. Compute

$$x_j = \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \Gamma_j^\lambda(u) := f_j(u) + h(u) + \frac{1}{2\lambda}\|u - x_{j-1}^c\|^2 \right\}, \; m_j = \Gamma_j^\lambda(x_j).$$

Moreover, consider the function

$$\phi_j^\lambda = \phi + \frac{1}{2\lambda}\|\cdot - x_{j-1}^c\|^2, \tag{5}$$

and let $\tilde{x}_j$ be such that

$$\tilde{x}_j \in \operatorname{Argmin} \left\{ \phi_j^\lambda(u) : u \in \{x_j, \tilde{x}_{j-1}\} \right\}; \tag{6}$$

2. **If**

$$t_j = \phi_j^\lambda(\tilde{x}_j) - m_j \le \frac{\bar{\varepsilon}}{2}, \tag{7}$$

2.a) **then** perform a serious iteration, i.e., set $x_j^c = x_j$, choose an arbitrary finite set C_{j+1} such that $\{x_j\} \subset C_{j+1}$;

2.b) **else** perform a null iteration, i.e., set $x_j^c = x_{j-1}^c$, choose C_{j+1} such that

$$A_j \cup \{x_j\} \subset C_{j+1} \subset C_j \cup \{x_j\} \tag{8}$$

where

$$A_j = \{x \in C_j : f(x) + \langle f'(x), x_j - x\rangle = f_j(x_j)\} \tag{9}$$

set $f_{j+1} = \max\{f(x) + \langle f'(x), \cdot - x\rangle : x \in C_{j+1}\}$;

3. Set $j \leftarrow j+1$ and go to step 1.

RPB vs. standard bundle method

- introduce an auxiliary iterate $\tilde{x}_j$, convergence in $\{\tilde{x}_j\}$
- null/serious iterate decision making based on t_j
- motivation for $\tilde{x}_j$ and t_j:
  define $m_j^* := \min\{\phi_j^\lambda(u) : u \in \mathbb{R}^n\}$, then we have

$$m_j \le m_j^* \le \phi_j^\lambda(\tilde{x}_j),$$

  and hence

$$\phi_j^\lambda(\tilde{x}_j) - m_j^* \le t_j \le \frac{\bar{\varepsilon}}{2}$$

  where $t_j = \phi_j^\lambda(\tilde{x}_j) - m_j$.

Exploration-exploitation trade-off

- RPB can be viewed as an inexact proximal point method that consists of a number of stages (exploration) and each stage aims to solve approximately a proximal subproblem by an iterative procedure (exploitation).
- inner complexity: $\mathcal{O}_1(\lambda M_f^2/\bar{\varepsilon})$, outer complexity: $\mathcal{O}_1(d_0^2/(\lambda\bar{\varepsilon}))$.
- smaller $\lambda \implies$ less work done inside stages, and more number of stages.
- CS only conducts exploration but no exploitation.
- If $\lambda = \bar{\varepsilon}/M_f^2$, then it can be shown that every iteration index of RPB is a serious one, and RPB reduces to CS.

A general complexity bound

Theorem

The total number of iterations performed until RPB obtains a $\bar{\varepsilon}$-solution is bounded by

$$\mathcal{O}\left(\left[\frac{M_f \min\{\lambda M, \lambda_\mu M_f + d_0\}}{\bar{\varepsilon}} + 1\right]\left[\min\left\{\frac{d_0^2}{\lambda\bar{\varepsilon}}, \frac{1}{\lambda_\mu \mu}\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\} + 1\right]\right)$$

where

$$M = M_f + M_h, \quad \lambda_\mu = \frac{\lambda}{1 + \lambda\mu}.$$

Reduction to the bound of CS in the strongly convex case

Theorem

Let universal constants $C, C' > 0$ be given and consider an instance $(x_0, (f, f'; h))$ of (1) which satisfies (A1)-(A4) with parameter triple (M_f, M_h, μ) such that

$$\frac{CM_f d_0}{\bar{\varepsilon}} \geq 1, \qquad M_h \in [0, +\infty], \qquad 0 \leq \mu \leq \frac{C' M_f}{d_0}. \tag{10}$$

Then, RPB with any λ lying in the (nonempty) interval

$$\frac{d_0}{M_f} \leq \lambda \leq \frac{C d_0^2}{\bar{\varepsilon}} \tag{11}$$

has $\bar{\varepsilon}$-iteration complexity bound given by (17).

Reduction to the bound of CS in the convex case

Theorem

Let universal constants $C, C' > 0$ be given and consider an instance $(x_0, (f, f'; h))$ of (1) which satisfies (A1)-(A4) with parameter triple (μ, M_f, M_h) such that

$$\frac{CM_f d_0}{\bar{\varepsilon}} \geq 1, \qquad M_h \leq C' M_f, \qquad \mu = 0. \tag{12}$$

Then, RPB with any λ lying in the (nonempty) interval

$$\frac{\bar{\varepsilon}}{CM_f^2} \leq \lambda \leq \frac{Cd_0^2}{\bar{\varepsilon}} \tag{13}$$

has $\bar{\varepsilon}$-iteration complexity bound $\mathcal{O}_1(M_f^2 d_0^2/\bar{\varepsilon}^2)$, and hence agrees with (17).

(Kiwiel 2000) and (Du and Ruszczyński 2017) study a proximal bundle variant (PBV) for solving the set constrained problem

$$\min\{\tilde{f}(x) : x \in X\} \tag{14}$$

where X is a nonempty closed convex set and $\tilde{f}$ is a μ-convex ($\mu \geq 0$) finite everywhere function.

- convex, Kiwiel 2000

$$\mathcal{O}_1\left(\frac{\tilde{M}^2\tilde{D}^4}{\lambda\bar{\varepsilon}^3}\right)$$

where

$$\tilde{D} = \tilde{D}[\tilde{f}] := \sup\{d(x_j, X^*) : j \geq 0\}, \quad \tilde{M} = \tilde{M}[\tilde{f}] := \sup\{\|\tilde{f}'(x_j)\| : j \geq 0\}.$$

- μ-strongly convex, Du and Ruszczyński 2017

$$\tilde{\mathcal{O}}_1\left(\frac{\tilde{M}^2\lambda}{\alpha^2\bar{\varepsilon}}\right)$$

where $\alpha := \min\{\lambda\mu, 1\}$.

Compare RPB with PBV

Consider $\tilde{f} = f + h$ with f satisfying (A1)-(A3), $h \equiv \mu\|\cdot - x_0\|^2/2$ and $X = \mathbb{R}^n$. When $\mu = 0$, PBV has the iteration complexity (Kiwiel)

$$\mathcal{O}_1\left(\frac{M_f^2(d_0 + \lambda M_f)^4}{\lambda \bar{\varepsilon}^3}\right), \tag{15}$$

when $\mu > 0$, PBV has the iteration complexity (Du and Ruszczyński)

$$\tilde{\mathcal{O}}_1\left(\frac{M_f^2}{\lambda \mu^2 \bar{\varepsilon}} + \frac{d_0^2}{\lambda \bar{\varepsilon}}\right). \tag{16}$$

In general, the above bounds are worse than that of PRB

$$\mathcal{O}_1\left(\min\left\{\frac{M_f^2 d_0^2}{\bar{\varepsilon}^2}, \left(\frac{M_f^2}{\mu \bar{\varepsilon}} + 1\right)\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\}\right). \tag{17}$$

Proof for the convex case

Note that the arithmetic-geometric mean inequality implies that

$$d_0 + \lambda M_f = \frac{d_0}{3} + \frac{d_0}{3} + \frac{d_0}{3} + \lambda M_f \geq 4\left(\frac{1}{27} d_0^3 \lambda M_f\right)^{1/4},$$

and hence that

$$\mathcal{O}_1\left(\frac{M_f^2(d_0 + \lambda M_f)^4}{\lambda \bar{\varepsilon}^3}\right)$$

is minorized by $\mathcal{O}_1(M_f^3 d_0^3/\bar{\varepsilon}^3)$, which in turn is minorized by $\mathcal{O}_1(M_f^2 d_0^2/\bar{\varepsilon}^2)$ in view of the assumption that $CM_f d_0/\bar{\varepsilon} \geq 1$.

Define a proper class of instances

Definition

Given $(M_f, \mu, R_0) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_{++}$, let $\mathcal{I}_\mu(M_f, R_0)$ denote the class consisting of all instances $(x_0, (f, f'; h))$ satisfying conditions (A1)-(A3) and the condition that $d_0 \le R_0$. Moreover, let $\mathcal{I}^u_\mu(M_f, R_0)$ denote the unconstrained class consisting of all instances $(x_0, (f, f'; h)) \in \mathcal{I}_\mu(M_f, R_0)$ such that $h \equiv \mu \|\cdot\|^2/2$.

$\bar{\varepsilon}$-lower complexity bound

Theorem

For any given quadruple $(M_f, \mu, R_0, \bar{\varepsilon}) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_{++} \times \mathbb{R}_{++}$, *there exists an instance* $(x_0, (f, f'; h))$ *such that:*

a) $(x_0, (f, f'; h)) \in \mathcal{I}_\mu^u(M_f, R_0)$;

b) *it has lower complexity bound with respect to* $\mathcal{A}(\mathcal{I}_\mu^u(M_f, R_0), \bar{\varepsilon})$ *given by*

$$\left\lfloor \min\left\{ \frac{M_f^2 R_0^2}{128\bar{\varepsilon}^2}, \frac{M_f^2}{8\mu\bar{\varepsilon}} \right\} \right\rfloor + 1. \tag{18}$$

As a consequence, (18) *is also a* $\bar{\varepsilon}$*-lower complexity bound for any instance class* $\mathcal{I} \supset \mathcal{I}_\mu^u(M_f, R_0)$.

Optimality in the strongly convex case

Using Theorem 3, we can show that RPB with λ satisfying

$$\frac{R_0}{M_f} \leq \lambda \leq \frac{CR_0^2}{\bar{\varepsilon}}$$

has a complexity bound

$$\mathcal{O}_1\left(\min\left\{\frac{M_f^2 R_0^2}{\bar{\varepsilon}^2}, \frac{M_f^2}{\mu\bar{\varepsilon}}\log\left(\frac{\mu R_0^2}{\bar{\varepsilon}}+1\right)\right\}\right),$$

and RPB is optimal for any instance class $\mathcal{I}$ and scalar $\mu \in [0, C' M_f/R_0]$ such that

$$\mathcal{I}_\mu^u(M_f, R_0) \subseteq \mathcal{I} \subseteq \mathcal{I}_\mu(M_f, R_0). \tag{19}$$

Optimality in the convex case

Using Theorem 4, we can show that RPB with λ satisfying

$$\frac{\bar{\varepsilon}}{CM_f^2} \le \lambda \le \frac{CR_0^2}{\bar{\varepsilon}}$$

has a complexity bound

$$\mathcal{O}_1\left(\frac{M_f^2 R_0^2}{\bar{\varepsilon}^2}\right),$$

and RPB is optimal for any instance class $\mathcal{I}$ such that

$$\mathcal{I}_0^u(M_f, R_0) \subseteq \mathcal{I} \subseteq \mathcal{I}_0(M_f, R_0; C) \tag{20}$$

where

$$\mathcal{I}_0(M_f, R_0; C)$$
$$:= \{(x_0, (f, f'; h)) \in \mathcal{I}_0(M_f, R_0) : \exists\ M_h \le CM_f \text{ such that } h \text{ satisfies (A4)}\}$$

Concluding remarks

- Iteration-complexity bound for RPB to find a $\bar{\varepsilon}$-solution.
- Optimal complexity bounds in both convex and strongly convex settings.
- RPB can be interpreted as an inexact proximal point method.
- RPB with sufficiently small constant prox stepsize becomes the composite subgradient method.

THE END

Thanks!