A unified analysis of a class of proximal bundle methods for smooth-nonsmooth convex composite optimization

Jiaming Liang

School of Industrial and Systems Engineering
Georgia Institute of Technology

Joint work with Renato Monteiro

October 27, 2021

This talk is based on the following papers:

- J. Liang and R. D. C. Monteiro. A unified analysis of a class of proximal bundle methods for smooth-nonsmooth convex composite optimization. Technical report, 2021.
- J. Liang and R. D. C. Monteiro. A proximal bundle variant with optimal iteration-complexity for a large range of prox stepsizes. To appear in SIAM Journal on Optimization, available on arXiv:2003.11457, 2020.

Introduction

Main problem:

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

Main goal:

To present a framework consisting of most proximal bundle methods for convex smooth-nonsmooth composite optimization.

Convex hybrid (smooth-nonsmooth) composite problem

Consider (1), where

(A1) $f, h \in \overline{\mathrm{Conv}}\,(\mathbb{R}^n)$ are such that $\mathrm{dom}\, h \subset \mathrm{dom}\, f$ and a subgradient oracle $f' : \mathrm{dom}\, h \to \mathbb{R}^n$ satisfying $f'(x) \in \partial f(x)$ for every $x \in \mathrm{dom}\, h$ is available;

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

(A3) $\|f'(u) - f'(v)\| \leq 2M_f + L_f\|u - v\|$ for every $u, v \in \mathrm{dom}\, h$;

(A4) h is μ-convex.

In a previous paper [1], we proposed a relaxed proximal bundle (RPB) method that is optimal for convex nonsmooth optimization.

In this work, we generalize and improve RPB in the following aspects:

1. hybrid cases;
2. a general framework including 3 bundle update schemes;
3. a unified and much simpler analysis;
4. stronger complexity results;
5. an adaptive variant.

[1] J. Liang and R. D. C. Monteiro. A proximal bundle variant with optimal iteration-complexity for a large range of prox stepsizes. To appear in SIAM Journal on Optimization, available on arXiv:2003.11457, 2020.

Bundle method

Solving the proximal problem

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

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

Bundle method approximately solves (2) and recursively builds up a model by using a standard cutting-plane approach.

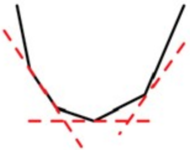
Bundle method

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

$$x_j = \underset{u\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ \Gamma_j^\lambda(u) := f_j(u) + h(u) + \frac{1}{2\lambda}\|u - x_{j-1}^c\|^2 \right\}, \qquad (3)$$

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

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

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** $\phi(x_j) \le (1-\gamma)\phi\left(x_{j-1}^c\right) + \gamma\,(f_j + h)\,(x_j)$ for some $\gamma \in (0,1)$.

A generic bundle update scheme

Definition

Let $\mathcal{C}_\mu(\phi)$ denote a class of convex functions Γ satisfying $\Gamma \le \phi$ and Γ is μ-convex.

For a given quadruple $(\Gamma, x_0, \lambda, \tau) \in \mathcal{C}_\mu(\phi) \times \mathbb{R}^n \times \mathbb{R}_{++} \times (0,1)$, the generic bundle update scheme returns $\Gamma^+ \in \mathcal{C}_\mu(\phi)$ satisfying

$$\tau\bar{\Gamma} + (1-\tau)[\ell_f(\cdot;x) + h] \le \Gamma^+ \tag{4}$$

where

$$x = \operatorname*{argmin}_{u\in\mathbb{R}^n} \left\{ \Gamma(u) + \frac{1}{2\lambda}\|u - x_0\|^2 \right\}$$

and $\bar{\Gamma} \in \mathcal{C}_\mu(\phi)$ is such that

$$\bar{\Gamma}(x) = \Gamma(x), \quad x = \operatorname*{argmin}_{u\in\mathbb{R}^n} \left\{ \bar{\Gamma}(u) + \frac{1}{2\lambda}\|u - x_0\|^2 \right\}. \tag{5}$$

Examples

(E1) **convex combination update:**
$\Gamma^+ = \Gamma^+_\tau := \tau\Gamma + (1-\tau)[\ell_f(\cdot;x) + h]$ with $\bar{\Gamma} = \Gamma$.

(E2) **multiple cuts update:** assume $\Gamma = \Gamma(\cdot;C)$ where $C \subset \mathbb{R}^n$ is the current bundle set and $\Gamma(\cdot;C) := \max\{\ell_f(\cdot;c) : c \in C\} + h$, choose the next bundle set C^+ satisfying

$$C(x)\cup\{x\} \subset C^+ \subset C\cup\{x\}, \quad C(x) := \{c \in C : \ell_f(x;c)+h(x) = \Gamma(x)\},$$

and then set $\Gamma^+ = \Gamma(\cdot;C^+)$ and $\bar{\Gamma} = \Gamma(\cdot;C(x))$.

Examples

(E3) **cut aggregation update:** assume $\Gamma = \max\{A_f, \ell_f(\cdot; x^-)\} + h$ where A_f is an affine function satisfying $A_f \le f$, set

$$\Gamma^+ = \max\{A_f^+, \ell_f(\cdot; x)\} + h$$

where $A_f^+ = \theta A_f + (1-\theta)\ell_f(\cdot; x^-)$ and

$$\theta \begin{cases} = 1, & \text{if } A_f(x) > \ell_f(x; x^-), \\ = 0, & \text{if } A_f(x) < \ell_f(x; x^-), \\ \in [0,1], & \text{if } A_f(x) = \ell_f(x; x^-). \end{cases}$$

Also set $\bar{\Gamma} = A_f^+ + h$.

Hybrid composite proximal bundle (HCPB) framework

0. Let $x_0 \in \mathrm{dom}\, h$, $\lambda > 0$, $\bar{\varepsilon} > 0$ and $\tau \in [\bar{\tau}, 1)$ be given where

$$\bar{\tau} = \left[1 + \frac{(1+\lambda\bar{\mu})\bar{\varepsilon}}{8\lambda T_{\bar{\varepsilon}}}\right]^{-1}, \tag{6}$$

and set $y_0 = x_0$, $t_0 = 0$ and $j = 0$;

1. if $t_j \le \bar{\varepsilon}/2$, then perform a **serious update**, i.e., set $x^c_{j+1} = x_j$ and find $\Gamma_{j+1} \in \mathcal{C}_\mu(\phi)$ such that $\Gamma_{j+1} \ge \ell_f(\cdot; x_j) + h$; else, perform a **null update**, i.e., set $x^c_{j+1} = x^c_j$ and find $\Gamma_{j+1} \in \mathcal{C}_\phi(\Gamma_j, x^c_j, \lambda, \tau)$;

2. compute

$$x_{j+1} = \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \Gamma^{\lambda}_{j+1}(u) := \Gamma_{j+1}(u) + \frac{1}{2\lambda} \|u - x^c_{j+1}\|^2 \right\}, \qquad (7)$$

choose $y_{j+1} \in \{x_{j+1}, y_j\}$ such that

$$\phi^{\lambda}_{j+1}(y_{j+1}) = \min \left\{ \phi^{\lambda}_{j+1}(x_{j+1}), \phi^{\lambda}_{j+1}(y_j) \right\} \qquad (8)$$

where ϕ^{λ}_j is defined as

$$\phi^{\lambda}_j := \phi + \frac{1}{2\lambda} \| \cdot - x^c_j \|^2, \qquad (9)$$

and set

$$m_{j+1} = \Gamma^{\lambda}_{j+1}(x_{j+1}), \quad t_{j+1} = \phi^{\lambda}_{j+1}(y_{j+1}) - m_{j+1}; \qquad (10)$$

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

HCPB vs. standard bundle method

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

$$m_j \le m_j^* \le \phi_j^\lambda(y_j),$$

  and hence

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

  where $t_j = \phi_j^\lambda(y_j) - m_j$.

RPB can be viewed as HCPB with bundle update scheme (E2).

While RPB only deals with the nonsmooth case ($L_f = 0$), HCPB extends the analysis to the hybrid case ($L_f \geq 0$).

Complexity for HCPB variants

Theorem

Let $x_0 \in \operatorname{dom} h$, $\bar{\varepsilon} > 0$ and $C > 0$ be given, and assume

$$\frac{C^2(M_f^2 + \bar{\varepsilon} L_f) d_0^2}{\bar{\varepsilon}^2} \geq 1. \tag{11}$$

Then, any variant of HCPB with input $(x_0, \lambda, \bar{\varepsilon}, \tau)$ satisfying

$$\tau = \left[1 + \frac{(1+\lambda\mu)\bar{\varepsilon}}{8\lambda(M_f^2 + \bar{\varepsilon} L_f)}\right]^{-1}, \quad \frac{\bar{\varepsilon}}{C(M_f^2 + \bar{\varepsilon} L_f)} \leq \lambda \leq \frac{C d_0^2}{\bar{\varepsilon}}, \tag{12}$$

has $\bar{\varepsilon}$-iteration complexity given (up to a logarithmic term) by

$$\mathcal{O}_1\left(\min\left\{\frac{(M_f^2 + \bar{\varepsilon} L_f) d_0^2}{\bar{\varepsilon}^2}, \left(\frac{M_f^2 + \bar{\varepsilon} L_f}{\mu \bar{\varepsilon}} + 1\right) \log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\}\right). \tag{13}$$

Complexity for τ-free HCPB variants in the nonsmooth case

Theorem

Let $x_0 \in \mathrm{dom}\, h$, $\bar{\varepsilon} > 0$ *and* $C > 0$ *be given, and assume*

$$\frac{CM_f d_0}{\bar{\varepsilon}} \geq 1. \tag{14}$$

Then, any variant of the τ*-free HCPB subclass with input* $(x_0, \lambda, \bar{\varepsilon})$ *satisfying*

$$\frac{\bar{\varepsilon}}{CM_f^2} \leq \lambda \leq \frac{Cd_0^2}{\bar{\varepsilon}}, \tag{15}$$

has $\bar{\varepsilon}$*-iteration complexity given (up to a logarithmic term) by*

$$\mathcal{O}_1\left(\min\left\{\frac{M_f^2 d_0^2}{\bar{\varepsilon}^2}, \left(\frac{M_f^2}{\mu\bar{\varepsilon}} + 1\right)\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\}\right). \tag{16}$$

Complexity of RPB in the strongly convex case

Theorem

Let $x_0 \in \mathrm{dom}\, h$, $\bar\varepsilon > 0$ *and* $C > 0$ *be given, and assume*

$$\frac{CM_f d_0}{\bar\varepsilon} \ge 1, \qquad 0 \le \mu \le \frac{CM_f}{d_0}. \tag{17}$$

Then, RPB with input $(x_0, \lambda, \bar\varepsilon)$ *satisfying*

$$\frac{d_0}{M_f} \le \lambda \le \frac{Cd_0^2}{\bar\varepsilon} \tag{18}$$

has $\bar\varepsilon$*-iteration complexity given by*

$$\mathcal{O}_1\left(\min\left\{\frac{M_f^2 d_0^2}{\bar\varepsilon^2}, \left(\frac{M_f^2}{\mu\bar\varepsilon}+1\right)\log\left(\frac{\mu d_0^2}{\bar\varepsilon}+1\right)\right\}\right). \tag{19}$$

Complexity of RPB in the convex case

Theorem

Let $x_0 \in \mathrm{dom}\, h$, $\bar{\varepsilon} > 0$ *and* $C > 0$ *be given, and assume*

$$\frac{CM_f d_0}{\bar{\varepsilon}} \geq 1, \qquad M_h \leq CM_f, \qquad \mu = 0. \tag{20}$$

Then, RPB with input $(x_0, \lambda, \bar{\varepsilon})$ *satisfying*

$$\frac{\bar{\varepsilon}}{CM_f^2} \leq \lambda \leq \frac{Cd_0^2}{\bar{\varepsilon}} \tag{21}$$

has $\bar{\varepsilon}$*-iteration complexity given by* $\mathcal{O}_1(M_f^2 d_0^2 / \bar{\varepsilon}^2)$.

Adaptive HCPB (A-HCPB)

0. Let $x_0 \in \mathrm{dom}\, h$, $\lambda > 0$, $\tau_0 = 0$ and $\bar{\varepsilon} > 0$ be given, and set $y_0 = x_0$, $t_0 = 0$ and $j = 0$;

1. set $\tau = \tau_j$;

2. if $t_j \le \bar{\varepsilon}/2$, then perform a **serious update**, i.e., set $x^c_{j+1} = x_j$ and $\Gamma_{j+1} = \ell_f(\cdot; x_j) + h$; else, perform a **null update**, i.e., set $x^c_{j+1} = x^c_j$ and $\Gamma_{j+1} = \tau\Gamma_j + (1-\tau)[\ell_f(\cdot; x_j) + h]$;

3. compute x_{j+1}, y_{j+1}, m_{j+1} and t_{j+1} as in step 2 of HCPB;

4. if $t_j > \bar{\varepsilon}/2$ and $t_{j+1} > \tau t_j + (1-\tau)\bar{\varepsilon}/4$, then set $\tau = (1+\tau)/2$ and go to step 2; else, set $\tau_{j+1} = \tau$ and $j \leftarrow j+1$, and go to step 1.

Complexity of A-HCPB

The general $\bar{\varepsilon}$-iteration complexity for A-HCPB is

$$\left[2\left(1+\frac{8\lambda_\mu(M_f^2+\bar{\varepsilon}L_f)}{\bar{\varepsilon}}\right)\log\left(\frac{4\bar{t}}{\bar{\varepsilon}}\right)+1\right]\left[\min\left\{\frac{d_0^2}{\lambda\bar{\varepsilon}},\frac{1}{\mu\lambda_\mu}\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}}+1\right)\right\}+1\right]$$

Under same assumptions as in previous theorems of HCPB, A-HCPB has the same iteration complexity as HCPB.

The total number of times τ is updated in step 4 is at most

$$\left\lceil\log\left(1+\frac{8\lambda_\mu(M_f^2+\bar{\varepsilon}L_f)}{\bar{\varepsilon}}\right)\right\rceil.$$

Concluding remarks

- A generic HCPB framework for convex hybrid composite optimization
- Including most proximal bundle variants such as RPB, and a new convex combination bundle update scheme
- A unified and simple analysis
- Stronger complexity results
- An adaptive variant that requires no prior knowledge of problem parameters

THE END

Thanks!