This article was downloaded by: [2610:148:2002:e000:3248:5755:4741:1cbc] On: 26 August 2023, At: 10:53
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA

Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A Unified Analysis of a Class of Proximal Bundle Methods for Solving Hybrid Convex Composite Optimization Problems

Jiaming Liang, Renato D. C. Monteiro

To cite this article:
Jiaming Liang, Renato D. C. Monteiro (2023) A Unified Analysis of a Class of Proximal Bundle Methods for Solving Hybrid Convex Composite Optimization Problems. Mathematics of Operations Research

Published online in Articles in Advance 05 Jun 2023

. https://doi.org/10.1287/moor.2023.1372

Full terms and conditions of use: https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2023, INFORMS

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

https://pubsonline.informs.org/journal/moor

MATHEMATICS OF OPERATIONS RESEARCH
Articles in Advance, pp. 1–24
ISSN 0364-765X (print), ISSN 1526-5471 (online)

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

A Unified Analysis of a Class of Proximal Bundle Methods for Solving Hybrid Convex Composite Optimization Problems

Jiaming Liang,[a,*] Renato D. C. Monteiro[b]

[a] Department of Computer Science, Yale University, New Haven, Connecticut 06511; [b] School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

*Corresponding author

Contact: jiaming.liang@yale.edu, https://orcid.org/0000-0001-7439-3845 (JL); renato.monteiro@isye.gatech.edu (RDCM)

Received: October 3, 2021
Revised: January 7, 2023
Accepted: March 26, 2023
Published Online in Articles in Advance: June 5, 2023

MSC2020 Subject Classifications: 49M37; 65K05; 68Q25; 90C25; 90C30; 90C60

https://doi.org/10.1287/moor.2023.1372

Copyright: © 2023 INFORMS

Abstract. This paper presents a proximal bundle (PB) framework based on a generic bundle update scheme for solving the hybrid convex composite optimization (HCCO) problem and establishes a common iteration-complexity bound for any variant belonging to it. As a consequence, iteration-complexity bounds for three PB variants based on different bundle update schemes are obtained in the HCCO context for the first time and in a unified manner. Although two of the PB variants are universal (i.e., their implementations do not require parameters associated with the HCCO instance), the other newly (as far as the authors are aware) proposed one is not, but has the advantage that it generates simple—namely, one-cut—bundle models. The paper also presents a universal adaptive PB variant (which is not necessarily an instance of the framework) based on one-cut models and shows that its iteration-complexity is the same as the two aforementioned universal PB variants.

Funding: Financial support from the Office of Naval Research [N00014-18-1-2077] and the Air Force Office of Scientific Research [Grant FA9550-22-1-0088] is gratefully acknowledged.

Keywords: hybrid convex composite optimization • iteration-complexity • proximal bundle method • universal method

1. Introduction

Let $f, h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper lower semicontinuous convex functions, such that $\operatorname{dom} h \subseteq \operatorname{dom} f$ and $h - \mu\|\cdot\|^2/2$ is convex for some $\mu \geq 0$, and consider the optimization problem

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

It is said that (1) is a hybrid convex composite optimization (HCCO) problem if there exist nonnegative scalars M_f and L_f and a first-order oracle $f' : \operatorname{dom} h \to \mathbb{R}^n$ (i.e., $f'(x) \in \partial f(x)$ for every $x \in \operatorname{dom} h$) satisfying the (M_f, L_f)-hybrid condition—namely $\|f'(u) - f'(v)\| \leq 2M_f + L_f\|u - v\|$ for every $u, v \in \operatorname{dom} h$. The main goal of this paper is to study the complexity of proximal bundle methods for solving the HCCO Problem (1) based on different bundle update schemes. Instead of focusing on a particular proximal bundle method, our unified approach considers a framework of generic proximal bundle methods (referred to as the GPB framework) based on a generic bundle update scheme and establishes a common iteration-complexity bound for all instances belonging to it.

Method Outline

Like all other proximal bundle methods, an iteration of a GPB variant solves the prox bundle subproblem

$$x = \arg\min_{u \in \mathbb{R}^n} \left\{ \Gamma(u) + \frac{1}{2\lambda}\|u - x^c\|^2 \right\}, \tag{2}$$

where λ is the prox stepsize, and x^c and Γ are the current prox-center and bundle function, respectively. Moreover, it also performs two types of iterations—that is, serious and null ones. In a serious iteration, the prox-center is updated to $x^c \leftarrow x$, and the updated bundle function Γ^+ is chosen so as to satisfy $\Gamma^+ \geq \ell_f(\cdot; x) + h$, where $\ell_f(\cdot; x) = f(x) + \langle f'(x), \cdot - x\rangle$. In a null iteration, the prox-center does not change, but Γ is updated according to a certain bundle update scheme (which is usually more restrictive than the ones in the serious iterations).

In order to illustrate the use of the generic bundle update scheme, this paper considers three specific well-known bundle update schemes and shows that they can all be viewed as special cases of the generic one. We now briefly describe the specific ones in the next three itemized paragraphs.

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

Example 1 (One-Cut Scheme). This scheme obtains Γ^+ as

$$\Gamma^+ = \Gamma^+_\tau := \tau\Gamma + (1-\tau)[\ell_f(\cdot;x) + h], \tag{3}$$

where x is as in (2), and $\tau \in (0,1)$ depends on (L_f, M_f, μ). Clearly, if Γ is the sum of h and an affine function underneath f, then so is Γ^+.

Example 2 (Two-Cuts Scheme). Assume that $\Gamma = \max\{A_f, \ell_f(\cdot;x^-)\} + h$, where A_f is an affine function satisfying $A_f \le f$, and x^- is the previous iterate. This scheme sets the next bundle function Γ^+ to one similar to Γ, but with (x^-, A_f) replaced by (x, A_f^+), where $A_f^+ = \theta A_f + (1-\theta)\ell_f(\cdot;x^-)$ for some $\theta \in [0,1]$ that does not depend on (L_f, M_f, μ).

Example 3 (Multiple-Cuts Scheme). The current bundle function Γ is of the form $\Gamma = \Gamma(\cdot;B)$, where $B \subset \mathbb{R}^n$ is a finite set (i.e., the current bundle set) and $\Gamma(\cdot;B)$ is defined as

$$\Gamma(\cdot;B) := \max\{\ell_f(\cdot;b) : b \in B\} + h. \tag{4}$$

This scheme obtains Γ^+ as $\Gamma^+ = \Gamma(\cdot;B^+)$, where B^+ is the updated bundle set obtained by possibly removing some points from B and then adding the most recent x to the resulting set.

Throughout out the paper, we refer to the GPB instances based on Example 1, Example 2, and Example 3 as 1C-PB, 2C-PB, and MC-PB, respectively.

Contribution

Regardless of the parameter triple (L_f, M_f, μ), it is shown that the iteration-complexity for any GPB variant to obtain a $\bar{\varepsilon}$-solution of the HCCO Problem (1) (i.e., a point $\bar{x} \in \operatorname{dom} h$ satisfying $\phi(\bar{x}) - \phi^* \le \bar{\varepsilon}$) is

$$\mathcal{O}\left(\min\left\{\frac{(M_f^2 + \bar{\varepsilon}L_f)d_0^2}{\bar{\varepsilon}^2}, \left(\frac{M_f^2 + \bar{\varepsilon}L_f}{\mu\bar{\varepsilon}} + 1\right)\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\} + 1\right), \tag{5}$$

for a large range of prox stepsizes λ, where d_0 denotes the distance of the initial point x_0 to the optimal solution set of (1). Because the 2C-PB and MC-PB methods do not rely on (L_f, M_f), a sharper iteration-complexity bound can be obtained for them by replacing (M_f, L_f) in (5) by $(\bar{M}_f, \bar{L}_f)$, respectively, where $(\bar{M}_f, \bar{L}_f)$ is the unique pair that minimizes $M_f^2 + \bar{\varepsilon}L_f$ over the set of pairs (M_f, L_f) satisfying the (M_f, L_f)-hybrid condition of f'. Moreover, even though this sharper complexity bound cannot be shown for 1C-PB, Section 5 presents an adaptive version of this variant, where τ in (3), instead of being chosen as a function of (L_f, M_f), is adaptively searched so as to satisfy a key inequality condition. Finally, Section 5 also shows that this adaptive variant has the same iteration-complexity as that of 2C-PB and MC-PB.

Related Literature

Proximal bundle methods are known to be efficient algorithms for solving nonsmooth convex composite optimization (NCCO) problems—that is, instances of (1) for which there exists $M_f \ge 0$ such that the hybrid condition holds with $L_f = 0$. Some preliminary ideas toward the development of the proximal bundle method were first presented in Lemaréchal [13] and Wolfe [25], and formal presentations of the method were given in Lemaréchal [14] and Mifflin [17]. Convergence analysis of the proximal bundle method for NCCO problems has been broadly discussed in the literature and can be found, for example, in the textbooks Ruszczyński [21] and Urruty and Lemaréchal [23]. Different bundle-management policies in the context of proximal bundle methods are discussed, for example, in de Oliveira et al. [5], Du and Ruszczyński [7], Frangioni [8], Kiwiel [10], Ruszczyński [21], and van Ackooij et al. [24].

Iteration-complexity bounds have been established for some proximal bundle methods in the context of the NCCO problem with $\mu = 0$ (see, for example, Astorino et al. [1], Díaz and Grimmer [6], Kiwiel [10], and Liang and Monteiro [16]). Astorino et al. [1] and Kiwiel [10] both consider the NCCO problem where h is the indicator function of a nonempty closed convex set, and Díaz and Grimmer [6] consider the NCCO problem where h is identically zero. Moreover, Kiwiel [10] obtains the first $\mathcal{O}(\bar{\varepsilon}^{-3})$ complexity bound, and Astorino et al. [1] and Díaz and Grimmer [6] subsequently also derive an $\mathcal{O}(\bar{\varepsilon}^{-3})$ bound. On the other hand, a previous authors' paper, Liang and Monteiro [16], proposes a proximal bundle variant using a novel condition to decide whether to perform a serious or null iteration, which does not necessarily yield a function value decrease. More importantly, Liang and Monteiro [16] establish the first $\mathcal{O}(\bar{\varepsilon}^{-2})$ complexity bound for a large range of prox stepsizes and show that the bound is indeed optimal.

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

More specialized iteration-complexity bounds for some proximal bundle methods in the context of the NCCO problem with $\mu > 0$ have also been established in Díaz and Grimmer [6], Du and Ruszczyński [7], and Liang and Monteiro [16]. More specifically, Du and Ruszczyński [7] derive a $\tilde{\mathcal{O}}(\bar{\varepsilon}^{-1})$ iteration-complexity bound for a proximal bundle method with prox stepsize set to $\lambda = 1/\mu$. Moreover, improving on the analysis of Du and Ruszczyński [7], Díaz and Grimmer [6] establish the optimal bound $\mathcal{O}(\bar{\varepsilon}^{-1})$ for the same method. Finally, Liang and Monteiro [16] also establish a $\tilde{\mathcal{O}}(\bar{\varepsilon}^{-1})$ iteration-complexity bound for its proximal bundle variant. In contrast to Díaz and Grimmer [6] and Du and Ruszczyński [7], the bound in Liang and Monteiro [16] is shown to be optimal (up to a logarithmic term) for a large range of prox stepsizes.

The current paper improves Liang and Monteiro [16] in the following aspects: (1) It deals with the more general HCCO problem; (2) in contrast to Liang and Monteiro [16], it nowhere assumes that h is Lipschitz continuous or imposes any condition on the parameter μ, and it shows that the Iteration-Complexity Bound (5) holds for prox stepsize ranges that are larger than the ones in Liang and Monteiro [16]; (3) whereas the proximal bundle variant of Liang and Monteiro [16] is based on the bundle update scheme Example 3, GPB is a framework based on a generic bundle update scheme that contains proximal bundle variants based on different update schemes (such as Examples 1–3); moreover, its unified analysis presented here applies to all these proximal bundle variants; and (4) as far as the authors are aware, it presents and analyzes for the first time a one-cut proximal bundle method for both NCCO and HCCO problems and also presents a universal variant of such a method.

Another method related to—and developed subsequently to—the proximal bundle method is the bundle-level method, which was first proposed in Lemaréchal et al. [15] and extended in many ways in Ben-Tal and Nemirovski [3], Kiwiel [9], and Lan [12]. These methods have been shown to have optimal iteration-complexity in the setting of the NCCO problem, with h being the indicator function of a compact convex set. Because their generated subproblems do not have a proximal term, and, hence, do not use a prox stepsize, they are different from the ones studied in this paper. Finally, de Oliveira and Solodov [4] present a doubly stabilized bundle method for solving NCCO problems whose prox subproblems combine elements from both proximal bundle and bundle-level methods and analyze its asymptotic convergence (but not its iteration-complexity).

Organization of the Paper

Subsection 1.5 presents basic definitions and notation used throughout the paper. Section 2 formally describes the assumptions on the HCCO Problem (1), reviews the constant stepsize composite subgradient (CS-CS) method, and discusses its iteration-complexity. Section 3 presents a generic bundle update scheme, describes the GPB framework, and states the main results of the paper—namely, the iteration-complexity of GPB. Section 4 contains three subsections, and they provide the analysis of bounds on the number of the serious, null, and total iterates, respectively. Section 5 presents the adaptive variant of 1C-PB and establishes the iteration-complexity of it. Section 6 presents some concluding remarks and possible extensions. Appendix A provides a few useful technical results. Appendix B presents two recursive formulas and their related results. Appendix C provides the proof of the iteration-complexity for the CS-CS method, describes an adaptive variant of CS-CS, and establishes its iteration-complexity. Finally, Appendix D provides the proofs of properties of bundle update schemes Example 2 and Example 3.

1.1. Basic Definitions and Notation

Let $\mathbb{R}$ denote the set of real numbers. Let $\mathbb{R}_+$ and $\mathbb{R}_{++}$ denote the set of nonnegative real numbers and the set of positive real numbers, respectively. Let $\mathbb{R}^n$ denote the standard n-dimensional Euclidean space equipped with inner product and norm denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, respectively. Let $\log(\cdot)$ denote the natural logarithm.

Let $\Psi : \mathbb{R}^n \to (-\infty, +\infty]$ be given. Let $\operatorname{dom} \Psi := \{x \in \mathbb{R}^n : \Psi(x) < \infty\}$ denote the effective domain of Ψ, and Ψ is proper if $\operatorname{dom} \Psi \neq \emptyset$. A proper function $\Psi : \mathbb{R}^n \to (-\infty, +\infty]$ is μ-convex for some $\mu \geq 0$ if

$$\Psi(\alpha z + (1-\alpha)z') \leq \alpha\Psi(z) + (1-\alpha)\Psi(z') - \frac{\alpha(1-\alpha)\mu}{2}\|z - z'\|^2,$$

for every $z, z' \in \operatorname{dom} \Psi$ and $\alpha \in [0,1]$. The set of all proper lower semicontinuous μ-convex functions is denoted by $\overline{\operatorname{Conv}}_\mu(\mathbb{R}^n)$. When $\mu = 0$, we simply denote $\overline{\operatorname{Conv}}_\mu(\mathbb{R}^n)$ by $\overline{\operatorname{Conv}}(\mathbb{R}^n)$. For $\varepsilon \geq 0$, the *ε-subdifferential* of Ψ at $z \in \operatorname{dom} \Psi$ is denoted by

$$\partial_\varepsilon \Psi(z) := \{s \in \mathbb{R}^n : \Psi(z') \geq \Psi(z) + \langle s, z' - z\rangle - \varepsilon,\ \forall z' \in \mathbb{R}^n\}.$$

The subdifferential of Ψ at $z \in \operatorname{dom} \Psi$, denoted by $\partial\Psi(z)$, is, by definition, the set $\partial_0\Psi(z)$.

Finally, even though $\mathcal{O}(\cdot)$ is a well-known concept in the study of complexity of algorithms, it is convenient for the purpose of our presentation to give a slightly stronger meaning to it—namely, if f and g are two positive functions defined in a certain set Ω, the notation $f(x) = \mathcal{O}(g(x))$ means that there exists constant $C > 0$, such that $f(x) \leq Cg(x)$ for all $x \in \Omega$.

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

2. Problem of Interest and a Review of the CS-CS Method

This section consists of two subsections. The first one describes the main problem and the assumptions imposed on it. The second one reviews the CS-CS method and an adaptive variant of it and describes their iteration-complexity bounds for obtaining a $\bar{\varepsilon}$-solution of the main problem.

2.1. Main Problem and Assumptions

The problem of interest in this paper is (1), which is assumed to satisfy the following conditions for some triple $(L_f, M_f, \mu) \in \mathbb{R}^3_+$:

Assumption 1. *$f \in \overline{\mathrm{Conv}}\,(\mathbb{R}^n)$ and $h \in \overline{\mathrm{Conv}}_\mu\,(\mathbb{R}^n)$ are such that $\mathrm{dom}\, h \subset \mathrm{dom}\, f$, and a subgradient oracle—that is, a function $f' : \mathrm{dom}\, h \to \mathbb{R}^n$ satisfying $f'(x) \in \partial f(x)$ for every $x \in \mathrm{dom}\, h$—is available.*

Assumption 2. *The set of optimal solutions X^* of Problem (1) is nonempty.*

Assumption 3. *For every $x, y \in \mathrm{dom}\, h$,*

$$\|f'(x) - f'(y)\| \le 2M_f + L_f\|x - y\|.$$

Throughout this paper, an instance of (1) means a triple $(f, f'; h)$ satisfying Assumptions 1–3 for some triple of parameters $(L_f, M_f, \mu) \in \mathbb{R}^3_+$.

We now add a few remarks about Assumptions 1–3. First, letting

$$\ell_f(\cdot; x) := f(x) + \langle f'(x), \cdot - x\rangle \quad \forall x \in \mathrm{dom}\, h, \tag{6}$$

then it is well-known that Assumption 3 implies that for every $x, y \in \mathrm{dom}\, h$,

$$f(x) - \ell_f(x; y) \le 2M_f\|x - y\| + \frac{L_f}{2}\|x - y\|^2. \tag{7}$$

Second, an obvious example of f satisfying Assumption 3 is the sum of an M_f-Lipschitz continuous function and a function whose gradient is L_f-Lipschitz continuous—for example, $f(x) = M_f\|x\| + L_f\|x\|^2/2$. Third, another way of obtaining functions f satisfying Assumption 3 is discussed in Proposition 1.

We now discuss other quantities, which, in addition to the parameters L_f, M_f, and μ, are also used in the complexity bounds obtained in this paper. For a given initial point $x_0 \in \mathrm{dom}\, h$, we denote its distance to X^* as

$$d_0 := \|x_0 - x_0^*\|, \text{ where } \quad x_0^* := \arg\min\{\|x_0 - x^*\| : x^* \in X^*\}. \tag{8}$$

Alternative quantities that are used in place of M_f and L_f are as follows. First, note that the set $\Omega \subset \mathbb{R}^2_+$ consisting of the pairs (M_f, L_f) satisfying Assumption 3 is easily seen to be a (nonempty) closed convex set. Moreover, for a given tolerance $\bar{\varepsilon} > 0$, it is easily seen that there exists a unique pair $(\bar{M}_f(\bar{\varepsilon}), \bar{L}_f(\bar{\varepsilon}))$, which minimizes $M_f^2 + \bar{\varepsilon}L_f$ over Ω, and, without any loss of clarity, we denote this pair simply by $(\bar{M}_f, \bar{L}_f)$ and define

$$T_{\bar{\varepsilon}} := (\bar{M}_f^2 + \bar{\varepsilon}\bar{L}_f)^{1/2}. \tag{9}$$

Moreover, if there exists a pair $(M_f, 0)$ satisfying Assumption 3, then the smallest M_f with this property is denoted by $\bar{M}_{f,0}$; otherwise, if no such pair exists, then we set $\bar{M}_{f,0} := \infty$. Finally, it is easily seen that $\bar{M}_{f,0} \ge T_{\bar{\varepsilon}} \ge \bar{M}_f$ and that any one of these two inequalities can hold strictly. For example, if $f = \|\cdot\| + \|\cdot\|^2/2$ and $h \equiv 0$, then we can easily see that $\bar{M}_{f,0} = \infty$, $\bar{M}_f = 1$, and $T_{\bar{\varepsilon}} \in (1, \infty)$ for any $\bar{\varepsilon} > 0$.

The following result, whose proof is postponed to Appendix A, gives conditions on (f, h), which guarantee that Assumption 3 holds.

Proposition 1. *Assume that Assumption 1 holds and that, for some $\nu \in (0, 1)$, the function f' in Assumption 1 satisfies*

$$\|f'(x) - f'(y)\| \le 2M_\nu + L_\nu\|x - y\|^\nu, \quad \forall x, y \in \mathrm{dom}\, h, \tag{10}$$

and, for any $\alpha > 0$, define

$$M_f(\alpha) := M_\nu + \frac{L_\nu\alpha}{2}, \quad L_f(\alpha) := L_\nu\nu\left(\frac{1-\nu}{\alpha}\right)^{\frac{1-\nu}{\nu}}. \tag{11}$$

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

Then, for any $\alpha > 0$, the pair $(M_f, L_f) = (M_f(\alpha), L_f(\alpha))$ satisfies Assumption 3 and

$$\inf_{\alpha>0}\left\{M_f(\alpha)^2 + \bar{\varepsilon} L_f(\alpha)\right\} \le 2\left(M_\nu^2 + \bar{\varepsilon}^{\frac{2\nu}{\nu+1}} L_\nu^{\frac{2}{\nu+1}}\right). \tag{12}$$

As a consequence,

$$T_{\bar{\varepsilon}} \le \sqrt{2}\left(M_\nu + \bar{\varepsilon}^{\frac{\nu}{\nu+1}} L_\nu^{\frac{1}{\nu+1}}\right). \tag{13}$$

We now make two remarks about (10). First, a trivial example of a pair (f, h) satisfying (10) is $f(\cdot) = M_\nu\|\cdot\| + L_\nu\|\cdot\|^{\nu+1}/(\nu+1)$ and $h \equiv 0$. More generally, the sum of an M_f-Lipschitz continuous function on dom h and a function whose gradient is ν-Hölder continuous on dom h satisfies (10). Second, if (10) holds with $M_\nu = 0$, it follows that f is differentiable on dom h, and its gradient is ν-Hölder continuous on dom h. Algorithms for solving instances of (1) satisfying (10) with $M_\nu = 0$ have been studied, for example, in Lan [12] and Nesterov [20].

Finally, for a given tolerance $\bar{\varepsilon} > 0$, it is said that an algorithm for solving (1) has $\bar{\varepsilon}$-iteration complexity $\mathcal{O}(T)$ if its total number of iterations until it obtains a $\bar{\varepsilon}$-solution is bounded by $C(T+1)$, where $C > 0$ is a universal constant.

2.2. Review of the CS-CS Method

We start by reviewing the CS-CS method. The CS-CS method with initial point $x_0 \in \operatorname{dom} h$ and constant prox stepsize $\lambda > 0$, denoted by CS-CS(x_0, λ), recursively computes its iteration sequence $\{x_j\}$, according to

$$x_{j+1} = \arg\min_{u\in\mathbb{R}^n}\left\{\ell_f(u; x_j) + h(u) + \frac{1}{2\lambda}\|u - x_j\|^2\right\} \qquad \forall j \ge 0. \tag{14}$$

For any given universal constant $C > 1$, pair (M_f, L_f) satisfying Assumption 3, and tolerance $\bar{\varepsilon} > 0$, it follows from Proposition C.1 that CS-CS(x_0, λ) with any stepsize λ such that $\bar{\varepsilon}/[4C(M_f^2 + \bar{\varepsilon} L_f)] \le \lambda \le \bar{\varepsilon}/[4(M_f^2 + \bar{\varepsilon} L_f)]$, has $\bar{\varepsilon}$-iteration complexity given by

$$\mathcal{O}\left(\min\left\{\frac{(M_f^2 + \bar{\varepsilon} L_f)d_0^2}{\bar{\varepsilon}^2}, \left(\frac{M_f^2 + \bar{\varepsilon} L_f}{\mu\bar{\varepsilon}} + 1\right)\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\} + 1\right), \tag{15}$$

(see our slightly modified definition of $\mathcal{O}(\cdot)$ in Subsection 1.5) with the convention that the second term is equal to the first one when $\mu = 0$. (It is worth noting that the second term converges to the first one as $\mu \downarrow 0$.)

In order to obtain the $\bar{\varepsilon}$-iteration Complexity (15), the CS-CS method requires the knowledge of (M_f, L_f) satisfying Assumption 3 to compute a suitable λ. Subsection C.2 presents an adaptive variant of the CS-CS method that does not require such knowledge. More precisely, this adaptive variant starts with any stepsize $\lambda_0 > 0$, employs a backtracking procedure to compute a nonincreasing sequence $\{\lambda_j\}$ such that each λ_j satisfies a key condition, and recursively performs iterations similar to (14). It is shown in Proposition C.2 that, without the prior knowledge of $T_{\bar{\varepsilon}}$, the adaptive variant of CS-CS has $\bar{\varepsilon}$-iteration complexity given by

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

It is worth noting that Bound (16) is better than the one for the CS-CS method (i.e., (15)) due to the fact that it is expressed in terms of the tighter quantity $T_{\bar{\varepsilon}}^2$ instead of the estimate $M_f^2 + \bar{\varepsilon} L_f$.

3. The GPB Framework

This section contains three subsections. Subsection 3.1 describes a generic bundle update scheme that is used to perform the null iterations of a method in the GPB framework. Subsection 3.2 presents the GPB framework, and Subsection 3.3 describes the main complexity results about it.

3.1. Bundle Update Schemes

Bundle methods discussed in the literature rely on different bundle update schemes—that is, schemes for updating the bundle function Γ in (2), which approximates the objective function of (1). Instead of focusing on a specific bundle update scheme, we describe in this subsection a generic scheme, which includes many of the ones

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

considered in the literature. This subsection also gives the details of the three concrete examples, Examples 1–3, of the generic bundle update scheme.

We start by describing the bundle update (BU) blackbox.

BU

Input: $(\lambda,\tau)\in\mathbb{R}_{++}\times(0,1)$ and $(x^c,x,\Gamma)\in\mathbb{R}^n\times\mathbb{R}^n\times\overline{\mathrm{Conv}}_\mu(\mathbb{R}^n)$ such that $\Gamma\le\phi$ and (2) holds.

- Find function Γ^+ such that

$$\Gamma^+\in\overline{\mathrm{Conv}}_\mu(\mathbb{R}^n),\qquad \tau\bar{\Gamma}(\cdot)+(1-\tau)[\ell_f(\cdot;x)+h(\cdot)]\le\Gamma^+(\cdot)\le\phi(\cdot), \tag{17}$$

where $\ell_f(\cdot;\cdot)$ is as in (6), and $\bar{\Gamma}(\cdot)$ is such that

$$\bar{\Gamma}\in\overline{\mathrm{Conv}}_\mu(\mathbb{R}^n),\quad \bar{\Gamma}(x)=\Gamma(x),\quad x=\arg\min_{u\in\mathbb{R}^n}\left\{\bar{\Gamma}(u)+\frac{1}{2\lambda}\|u-x^c\|^2\right\}. \tag{18}$$

Output: Γ^+.

Clearly, the above update scheme does not completely determine Γ^+, but, rather, gives minimal conditions on it which are suitable for the complexity analysis of this paper.

We now describe three concrete update schemes, Examples 1–3, which are special ways of implementing BU. Unless otherwise stated, it is assumed that their input is the same as in BU.

Example 1 (One-Cut Scheme). This scheme obtains Γ^+ as in (3). It is easy to see that if this update is used recursively, then Γ is always of the form

$$\Gamma(\cdot)=\sum_{x\in X}\alpha_x\ell_f(\cdot;x)+h(\cdot), \tag{19}$$

where X is a finite set in dom h and $\{\alpha_x : x\in X\}\subset\mathbb{R}_{++}$ are scalars, such that $\sum_{x\in X}\alpha_x=1$.

Example 2 (Two-Cuts Scheme). For this scheme, it is assumed that Γ has the form

$$\Gamma=\max\{A_f,\ell_f(\cdot;x^-)\}+h, \tag{20}$$

where $h\in\overline{\mathrm{Conv}}_\mu(\mathbb{R}^n)$, and A_f is an affine function satisfying $A_f\le f$. In view of (2), it can be shown that there exists $\theta\in[0,1]$, such that

$$\frac{1}{\lambda}(x-x^c)+\partial h(x)+\theta\nabla A_f+(1-\theta)f'(x^-)\ni 0, \tag{21}$$

$$\theta A_f(x)+(1-\theta)\ell_f(x;x^-)=\max\{A_f(x),\ell_f(x;x^-)\}. \tag{22}$$

The scheme then sets

$$A_f^+(\cdot):=\theta A_f(\cdot)+(1-\theta)\ell_f(\cdot;x^-), \tag{23}$$

and outputs the function Γ^+ defined as

$$\Gamma^+(\cdot):=\max\{A_f^+(\cdot),\ell_f(\cdot;x)\}+h(\cdot). \tag{24}$$

Example 3 (Multiple-Cuts Scheme). For this scheme, it is assumed that Γ of the form $\Gamma=\Gamma(\cdot;B)$, where $B\subset\mathbb{R}^n$ is a finite set (i.e., the current bundle set), and $\Gamma(\cdot;B)$ is defined as in (4). This scheme chooses the next bundle set B^+ so that

$$B(x)\cup\{x\}\subset B^+\subset B\cup\{x\}, \tag{25}$$

where

$$B(x):=\{b\in B:\ell_f(x;b)+h(x)=\Gamma(x)\}, \tag{26}$$

and then output $\Gamma^+=\Gamma(\cdot;B^+)$.

It is interesting to note that (24), (25), and the definition of Γ^+ in Example 3 imply that the updates Γ^+ output by schemes Example 2 and Example 3 have the property that $\Gamma^+(\cdot)$ is minorized by $\ell_f(\cdot;x)+h(\cdot)$, where x is as in (2). On the other hand, Γ^+ output by Example 1 does not necessarily have this property.

We now make some remarks to argue that all the update schemes above are special implementations of BU. It can be easily seen that the update Γ^+ in Example 1, together with $\bar{\Gamma}=\Gamma$, satisfies (17) and (18), and, hence, that this Γ^+ is a special way of implementing BU. On the other hand, the proofs that the updates Γ^+ of Example 2 and Example 3 are special implementations of BU are more involved and are given in Propositions D.1 and D.2, respectively.

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

3.2. The GPB Framework

This subsection states the GPB framework based on the BU blackbox presented in Subsection 3.1. It also gives several remarks about GPB and discusses how it relates to the classical proximal point method.

Before stating GPB, we first give a brief description for its j-th iteration. Given a prox-center x_j^c, it attempts to approximately solve the prox subproblem

$$m_j^* := \min\left\{\phi(u) + \frac{1}{2\lambda}\|u - x_j^c\|^2 : u \in \mathbb{R}^n\right\}, \tag{27}$$

(according to a certain termination criterion outlined below) by computing the exact solution x_j of the approximate prox subproblem of the Form (2), with $x^c = x_j^c$ and with bundle function $\Gamma = \Gamma_j$ obtained, for example, according to one of the update schemes Example 1, Example 2, or Example 3 described above. If it succeeds, then x_{j+1}^c is set to be x_j; otherwise, x_{j+1}^c is set to be x_j^c. Finally, j is updated to $j+1$, and the above iteration is repeated.

The method outlined above can be viewed as an inexact proximal point method. More specifically, consecutive iterations j such that x_j^c remains the same approximately solve the Prox Subproblem (27) (which does not depend on j). When that happens at an iteration j, the prox-center for the next iteration $j+1$ is then updated to a new one.

We now describe the aforementioned termination criterion. Given $\delta > 0$, it checks whether x_j and the iterate y_j defined as

$$y_j \in \operatorname{Arg\,min}\{\phi(x) : x \in \{x_0, x_1, \ldots, x_j\}\}, \tag{28}$$

satisfies

$$t_j := \phi(y_j) - \Gamma_j(x_j) - \frac{1}{2\lambda}\|x_j - x_j^c\|^2 \le \delta.$$

We are now ready to state GPB.

GPB

0. Let $x_0 \in \operatorname{dom} h$, $\lambda > 0$, $\bar{\varepsilon} > 0$ and $\tau \in (0,1)$ be given such that

$$\frac{\tau}{1-\tau} \ge \frac{8\lambda T_{\bar{\varepsilon}}^2}{(1+\lambda\mu)\bar{\varepsilon}}, \tag{29}$$

where $T_{\bar{\varepsilon}}$ is as in (9), and set $y_0 = x_0$, $t_0 = 0$ and $j = 1$;

1. if $t_{j-1} \le \bar{\varepsilon}/2$, then perform a **serious update**; that is, set $x_j^c = x_{j-1}$ and find Γ_j such that

$$\Gamma_j \in \overline{\operatorname{Conv}}_\mu(\mathbb{R}^n), \quad \ell_f(\cdot; x_{j-1}) + h \le \Gamma_j \le \phi; \tag{30}$$

else, perform a **null update**; that is, set $x_j^c = x_{j-1}^c$, and let Γ_j be the output of the BU blackbox with input (λ, τ) and $(x^c, x, \Gamma) = (x_{j-1}^c, x_{j-1}, \Gamma_{j-1})$;

2. Compute

$$x_j = \arg\min_{u \in \mathbb{R}^n}\left\{\Gamma_j(u) + \frac{1}{2\lambda}\|u - x_j^c\|^2\right\}, \tag{31}$$

choose y_j according to (28), and set

$$m_j = \Gamma_j(x_j) + \frac{1}{2\lambda}\|x_j - x_j^c\|^2, \qquad t_j = \phi(y_j) - m_j; \tag{32}$$

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

An iteration j such that $t_j \le \bar{\varepsilon}/2$ is called a serious iteration, in which case x_j (respectively (resp.), y_j) is called a serious iterate (resp., auxiliary serious iterate); otherwise, j is called a null iteration. Let $j_1 \le j_2 \le \ldots$ denote the sequence of all serious iterations and define the k-th cycle $\mathcal{C}_k$ to be the iterations j such that $j_{k-1} + 1 \le j \le j_k$; that is,

$$\mathcal{C}_k := \{j_{k-1} + 1, \ldots, j_k\}, \tag{33}$$

where $j_0 := 0$. Hence, only the last iteration of a cycle (which can be the first one if $\mathcal{C}_k$ contains only one iteration) is serious.

We make some basic remarks about GPB. First, we refer to it as a framework because it does not completely specify how some algorithmic quantities are generated. The framework, rather, gives minimal conditions on these quantities, which enables us to establish complexity bounds for all specific instances of it in a unified

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

manner. Second, in view of (30), or the fact that the output of BU satisfies (17), it follows that

$$\Gamma_j \le \phi, \quad \Gamma_j \in \overline{\mathrm{Conv}}_\mu\,(\mathbb{R}^n) \quad \forall j \ge 1. \tag{34}$$

Third, in view of the definition of $\mathcal{C}_k$ and the way the prox-center iterates are generated, it is easy to see that for every $k \ge 1$, we have

$$x_j^c = x_{j_{k-1}} \quad \forall j \in \mathcal{C}_k. \tag{35}$$

In words, all prox-centers in the k-th cycle are equal to the most recent serious iterate. Fourth, schemes Examples 1–3 in the previous subsection provide three possible concrete ways of implementing the BU blackbox in step 1. Fifth, although GPB does not specify a termination criterion for the sake of shortness, all iteration-complexity bounds established in this paper are relative to the effort of obtaining a $\bar{\varepsilon}$-solution of (1). Finally, although iteration-complexity bounds for GPB can also be established for other termination criteria (see, for example, Liang and Monteiro [16, section 6]), we have omitted the details of their derivation for the sake of shortness.

We now make some observations about possible simple ways of choosing the bundle function Γ_j in a serious update. Specifically, two simple ways are: (1) $\Gamma_j = \ell_f(\cdot; x_{j-1}) + h$, and (2) $\Gamma_j = \max\{\Gamma_{j-1}, \ell_f(\cdot; x_{j-1}) + h\}$. Moreover, under the assumption that every call to BU during a null update is carried out using Example 2 (resp., Example 3), another way to obtain Γ_j during a serious update is to also use update Example 2 (resp., Example 3). In view of the observation in the second-to-last paragraph in Subsection 3.1, it follows that the latter way yields a bundle function Γ_j, satisfying (30).

We now discuss the role played by the parameter τ of GPB. First, τ is only used in step 1 as input to the BU blackbox to obtain Γ_j. Second, even though the analysis of GPB depends on a scalar τ satisfying (29), the implementations of some specific instances of GPB do not require knowledge of such τ. For instance, because the updates Example 2 and Example 3 do not depend on τ, the GPB instances 2C-PB and MC-PB do not depend on τ either. (Recall the meaning of 1C-PB, 2C-PB, and MC-PB given in the sentence following Example 3 in Section 1.) Third, the GPB instance 1C-PB requires a scalar τ satisfying (29) because the update Example 2 depends on τ (see (3)). Finally, (29) implies that τ has to be sufficiently close to one, which, in the context of Example 1, means that the new bundle Γ_j is closer to Γ_{j-1} than the new cut $\ell_f(\cdot; x_{j-1}) + h(\cdot)$ in view of the nature of the one-cut scheme Example 1 (see Relation (3)).

Finally, we briefly discuss how accurately GPB solves the Prox Problem (27). Because $\Gamma_j \le \phi$, it follows from the definition of m_j in (32) that $m_j \le m_j^*$ and, hence, that

$$\begin{aligned} 0 &\le \phi(y_j) + \frac{1}{2\lambda}\|y_j - x_j^c\|^2 - m_j^* \\ &\le \phi(y_j) + \frac{1}{2\lambda}\|y_j - x_j^c\|^2 - m_j = t_j + \frac{1}{2\lambda}\|y_j - x_j^c\|^2. \end{aligned} \tag{36}$$

Thus, if j is a serious iteration, or, equivalently, $t_j \le \bar{\varepsilon}/2$, it follows that y_j is a $\bar{\varepsilon}_j$-solution of (27), where

$$\bar{\varepsilon}_j := \frac{\bar{\varepsilon}}{2} + \frac{1}{2\lambda}\|y_j - x_j^c\|^2.$$

The sequence of consecutive null iterations between two serious ones can be regarded as an iterative procedure to compute the aforementioned $\bar{\varepsilon}_j$-solution. More details of such an interpretation can be found in Liang and Monteiro [16, subsection 3.1].

Observe that even though the right-hand side of (36) contains two terms, our serious step condition used in GPB only checks the magnitude of the first one. It is possible to modify GPB to one whose serious step condition controls the magnitude of the right-hand side of (36). However, because the latter serious step condition is more restrictive, the resulting method will perform more null iterations, and, hence, its practical performance might not be as good as the one proposed in this paper.

We end this subsection by stating a general complexity bound that applies to any GPB variant. It assumes that the triple (M_f, L_f, μ) is known so that a parameter τ satisfying (29) can be computed.

Theorem 1. *Let universal constant $C > 0$, initial point $x_0 \in \mathrm{dom}\, h$, tolerance $\bar{\varepsilon} > 0$, and instance $(f, f'; h)$ of (1) satisfying Assumptions 1–3 for some parameter triple $(L_f, M_f, \mu) \in \mathbb{R}_+^3$ be given. Then, if λ satisfies*

$$\frac{\bar{\varepsilon}}{C(M_f^2 + \bar{\varepsilon} L_f)} \le \lambda \le \frac{C d_0^2}{\bar{\varepsilon}}, \tag{37}$$

and τ is given by

$$\tau = \left[1 + \frac{(1+\lambda\mu)\bar{\varepsilon}}{8\lambda(M_f^2 + \bar{\varepsilon} L_f)}\right]^{-1}, \tag{38}$$

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

then any variant of GPB *with input* $(x_0, \lambda, \bar{\varepsilon}, \tau)$ *obtains a* $\bar{\varepsilon}$*-solution of the above instance in a number of iterations bounded* (*up to a logarithmic term*) *by* (5).

3.3. Iteration-Complexity Results for τ-Free GPB Variants

This subsection considers the subclass of GPB methods, referred to as the τ-free GPB subclass, which do not depend on τ (and, hence, do not need τ as input), and derives improved iteration-complexity bounds for it that follow as immediate consequences of Theorem 1. Because 2C-PB and MC-PB do not depend on τ, the results below apply to both of them.

Corollary 1. *Let universal constant* $C > 0$, *initial point* $x_0 \in \operatorname{dom} h$, *and tolerance* $\bar{\varepsilon} > 0$ *be given, and consider an instance* $(f, f'; h)$ *of* (1) *satisfying Assumptions* 1–3. *Then, any variant of the* τ*-free* GPB *subclass with input* $(x_0, \lambda, \bar{\varepsilon})$ *satisfying*

$$\frac{\bar{\varepsilon}}{CT_{\bar{\varepsilon}}^2} \le \lambda \le \frac{Cd_0^2}{\bar{\varepsilon}}, \tag{39}$$

where $T_{\bar{\varepsilon}}$ *is as in* (9), *obtains a* $\bar{\varepsilon}$*-solution of the above instance in a number of iterations bounded* (*up to a logarithmic term*) *by* (16).

Proof. Observe that any variant of the τ-free GPB subclass can be viewed as an instance of GPB with input τ satisfying the equality in (29) because it does not depend on τ. Hence, it follows from (16) and Theorem 1 with (L_f, M_f) replaced by $(\bar{L}_f, \bar{M}_f)$ that the conclusion of the corollary holds. □

Recall that Proposition 1 shows that if f satisfies (10), then it satisfies Assumption 3 with $(M_f, L_f) = (M_f(\alpha), L_f(\alpha))$. The following result is a consequence of Corollary 1 when Condition (10) holds in place of Assumption 3. We omit its proof because it directly follows from Corollary 1 and (13).

Corollary 2. *Let universal constant* $C > 0$, *initial point* $x_0 \in \operatorname{dom} h$, *and tolerance* $\bar{\varepsilon} > 0$ *be given, and consider an instance* $(f, f'; h)$ *of* (1) *such that Assumption* 1, *Assumption* 2, *and Condition* (10) *hold for some quadruple* $(M_\nu, L_\nu, \mu, \nu) \in \mathbb{R}_+^3 \times (0, 1)$. *Then, any variant of the* τ*-free* GPB *subclass with input* $(x_0, \lambda, \bar{\varepsilon})$ *satisfying* (39) *obtains a* $\bar{\varepsilon}$*-solution of the above instance in a number of iterations bounded* (*up to a logarithmic term*) *by*

$$\mathcal{O}\left(\min\left\{\frac{M_\nu^2 d_0^2}{\bar{\varepsilon}^2} + \left(\frac{L_\nu}{\bar{\varepsilon}}\right)^{\frac{2}{\nu+1}} d_0^2, \left(\frac{M_\nu^2}{\mu\bar{\varepsilon}} + \frac{L_\nu^{\frac{2}{\nu+1}}}{\mu\bar{\varepsilon}^{\frac{1-\nu}{1+\nu}}} + 1\right)\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\} + 1\right). \tag{40}$$

We now make two remarks about Corollary 2. First, when $M_\nu = 0$, Bound (40) reduces to

$$\mathcal{O}\left(\min\left\{\left(\frac{L_\nu}{\bar{\varepsilon}}\right)^{\frac{2}{\nu+1}} d_0^2, \left(\frac{L_\nu^{\frac{2}{\nu+1}}}{\mu\bar{\varepsilon}^{\frac{1-\nu}{1+\nu}}} + 1\right)\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\} + 1\right).$$

Second, when $\mu = 0$, the above bound agrees with the one obtained for the primal universal method of Nesterov [20, (2.20) therein].

For the sake of comparing the results of this paper with the ones obtained in Liang and Monteiro [16], we now state another consequence of Theorem 1, in which an alternative $\bar{\varepsilon}$-iteration complexity for τ-free GPB instances applied to instances of (1) with $\bar{M}_{f,0}$ finite. (Recall the definition of $\bar{M}_{f,0}$ is in the line below (9).)

Corollary 3. *Let universal constant* $C > 0$, *initial point* $x_0 \in \operatorname{dom} h$, *and tolerance* $\bar{\varepsilon} > 0$ *be given, and consider an instance* $(f, f'; h)$ *of* (1) *such that Assumptions* 1–3 *hold and* $\bar{M}_{f,0}$ *is finite. Then, any variant of the* τ*-free* GPB *subclass with input* $(x_0, \lambda, \bar{\varepsilon})$ *satisfying*

$$\frac{\bar{\varepsilon}}{C(\bar{M}_{f,0})^2} \le \lambda \le \frac{Cd_0^2}{\bar{\varepsilon}}, \tag{41}$$

obtains a $\bar{\varepsilon}$*-solution of the above instance in a number of iterations bounded* (*up to a logarithmic term*) *by*

$$\mathcal{O}\left(\min\left\{\frac{(\bar{M}_{f,0})^2 d_0^2}{\bar{\varepsilon}^2}, \left(\frac{(\bar{M}_{f,0})^2}{\mu\bar{\varepsilon}} + 1\right)\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\} + 1\right). \tag{42}$$

Proof. Observe that any variant of the τ-free GPB subclass can be viewed as an instance of GPB with input $\tau = [1 + (1 + \lambda\mu)\bar{\varepsilon}/(8\lambda\bar{M}_{f,0}^2)]^{-1}$ because it does not depend on τ. Because the pair $(0, \bar{M}_{f,0})$ satisfies conditions Assumptions 1–3, it then follows from (16) with $T_{\bar{\varepsilon}}$ replaced by $\bar{M}_{f,0}$ and Theorem 1 with (L_f, M_f) replaced by $(0, \bar{M}_{f,0})$ that the conclusion of the corollary holds. □

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

Before comparing the RPB method of Liang and Monteiro [16] with the τ-free GPB instances of the paper, we first make two remarks about the first one in regard to the latter ones. First, the RPB method of Liang and Monteiro [16] with $\delta = \bar{\varepsilon}/2$ can be viewed as a special case of τ-free GPB because RPB uses the inequality $t_{j-1} \le \bar{\varepsilon}/2$ to decide whether to perform a serious or null update, and its serious and null updates, the latter of which are based on Example 3, fulfill the requirements of step 1 of GPB (see Lemma A.1). Second, whereas the RPB method of Liang and Monteiro [16] deals with instances of (1) such that $\bar{M}_{f,0}$ is finite (i.e., the nonsmooth setting), the analysis presented in this paper for τ-free GPB applies to the larger class of instances of (1) such that $T_{\bar{\varepsilon}}$ is finite (i.e., the hybrid or smooth/nonsmooth setting).

We now compare Corollary 3 of this paper with Liang and Monteiro [16, corollary 3.2]. Indeed, it follows from Liang and Monteiro [16, corollary 3.2] with $M_f = \bar{M}_{f,0}$ that RPB has $\bar{\varepsilon}$-iteration complexity given by (42) as long as $d_0/\bar{M}_{f,0} \le \lambda \le Cd_0^2/\bar{\varepsilon}$ and $\mu \le C\bar{M}_{f,0}/d_0$. On the other hand, Corollary 3 of this paper establishes Complexity Bound (42) for any λ lying in the larger range (41) without imposing any condition on μ.

We now compare Corollary 3 of this paper with Liang and Monteiro [16, corollary 3.3]. Indeed, it follows from Liang and Monteiro [16, corollary 3.3] with $M_f = \bar{M}_{f,0}$ that RPB has $\bar{\varepsilon}$-iteration complexity $\mathcal{O}((\bar{M}_{f,0})^2 d_0^2/\bar{\varepsilon}^2 + 1)$ as long as (41) holds and h is $(C\bar{M}_{f,0})$-Lipschitz continuous. On the other hand, Corollary 3 of this paper establishes the (possibly sharper) $\bar{\varepsilon}$-iteration Complexity (42) for any λ in the same range without imposing any condition Lipschitz continuity on h.

Even though 1C-PB depends on τ, it can be easily seen that its iteration-complexity is similar to the one of Corollary 1 if τ is close to the one satisfying the equality in (29). Section 5 describes an adaptive variant of 1C-PB, which adaptively chooses $\tau = \tau_j$ such that a key condition holds in every iteration j and which has the same $\bar{\varepsilon}$-iteration complexity as that of Corollary 1.

4. Complexity Analysis of GPB

This section consists of three subsections. The first one provides a bound on the number of serious iterates generated by the GPB framework. The second one derives a preliminary complexity bound on the number of possible consecutive null iterates. Finally, the last subsection combines the aforementioned bounds to obtain a complexity bound on the total number of iterations performed by any algorithm in the GPB framework with prox stepsize λ arbitrarily chosen. Moreover, it also provides the proof of Theorem 1 as a consequence of this general complexity result.

4.1. Bounding the Number of Serious Iterates

We start by introducing some notation and definitions. Recall from the paragraph following GPB that $j_1 < j_2 < \ldots$ denote the serious iterations of the GPB framework. Now, define $\hat{x}_0 := x_0$, and for every $k \ge 1$, let

$$\hat{x}_k := x_{j_k}, \quad \hat{y}_k := y_{j_k}, \quad \hat{\Gamma}_k := \Gamma_{j_k}, \quad \hat{m}_k := m_{j_k}. \tag{43}$$

The following result summarizes the basic properties of the above "hat" entities that follow as an immediate consequence of their definitions and the description of the GPB framework. It is worth noting that the complexity results developed in this subsection apply not only to the sequences defined in (43), but also to arbitrary sequences $\{\hat{x}_k\}$, $\{\hat{y}_k\}$ and $\{\hat{\Gamma}_k\}$ satisfying the basic properties stated below.

Lemma 1. *The following statements about GPB hold for every $k \ge 1$:*

a. $\hat{\Gamma}_k \in \overline{\mathrm{Conv}}_\mu(\mathbb{R}^n)$ *and* $\hat{\Gamma}_k \le \phi$;

b. $(\hat{x}_k, \hat{m}_k)$ *is the pair of optimal solution and optimal value of*

$$\min\left\{\hat{\Gamma}_k(u) + \frac{1}{2\lambda}\|u - \hat{x}_{k-1}\|^2 : u \in \mathbb{R}^n\right\};$$

c. *There holds* $\phi(\hat{y}_k) - \hat{m}_k \le \bar{\varepsilon}/2$.

Proof.

a. This statement follows from (34) and the definition of $\hat{\Gamma}_k$ in (43).

b. It follows from (31) with $j = j_k$, the first identity in (32) with $j = j_k$, and Relations (35) and (43), that (b) holds.

c. Because j_k is a serious iteration, we have that $t_{j_k} \le \bar{\varepsilon}/2$. Using this conclusion, (43), and the definition of t_j in (32), we conclude that (c) holds.

It is worth noting that (a), (b), and (c) can be viewed only as properties about the sequences $\{\hat{\Gamma}_k\}$ and $\{\hat{y}_k\}$, and the initial point $\hat{x}_0$, because $\{\hat{x}_k : k \ge 1\}$ is uniquely determined by $\{\hat{\Gamma}_k\}$. □

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

The next result provides an important recursive formula for the sequences in (43) and derives some important consequences that follow from it.

Lemma 2. *Let $u \in \operatorname{dom} h$ be given and define*

$$\lambda_\mu = \frac{\lambda}{1+\lambda\mu}. \tag{44}$$

Then, the following statements hold:

a. *For every $k \geq 1$, we have*

$$\phi(\hat{y}_k) - \phi(u) \leq \frac{1}{2\lambda}\|\hat{x}_{k-1} - u\|^2 - \frac{1}{2\lambda_\mu}\|\hat{x}_k - u\|^2 + \frac{\bar{\varepsilon}}{2}; \tag{45}$$

b. *We have $\min_{1\leq k\leq K}\{\phi(\hat{y}_k) - \phi(u)\} \leq \bar{\varepsilon}$ for every index K satisfying*

$$K \geq \min\left\{\frac{\|x_0 - u\|^2}{\lambda\bar{\varepsilon}}, \frac{1}{\mu\lambda_\mu}\log\left(\frac{\mu\|x_0 - u\|^2}{\bar{\varepsilon}} + 1\right)\right\};$$

c. *For every $k \geq 1$, we have $\|\hat{x}_k - u\|^2 \leq \|x_0 - u\|^2 + \lambda k\bar{\varepsilon}$.*

Proof.

a. It follows from Lemma 1(a) that $\hat{\Gamma}_k$ is μ-convex, and, hence, that the objective function in Lemma 1(b) is $(\mu + 1/\lambda)$-strongly convex. Using this observation, Lemma 1(b) and Beck [2, theorem 5.25(b)] with $f = \hat{\Gamma}_k + \|\cdot - \hat{x}_{k-1}\|^2/(2\lambda)$, $x^* = \hat{x}_k$ and $\sigma = \mu + 1/\lambda$, we have for the given $u \in \operatorname{dom} h$ and every $k \geq 1$,

$$\hat{m}_k + \frac{1}{2}\left(\mu + \frac{1}{\lambda}\right)\|u - \hat{x}_k\|^2 \leq \hat{\Gamma}_k(u) + \frac{1}{2\lambda}\|u - \hat{x}_{k-1}\|^2. \tag{46}$$

Using the above inequality and Lemma 1, (a) and (c), we conclude that

$$\begin{aligned}\phi(\hat{y}_k) - \phi(u) + \frac{1}{2}\left(\mu + \frac{1}{\lambda}\right)\|\hat{x}_k - u\|^2 &\leq \phi(\hat{y}_k) - \hat{\Gamma}_k(u) + \frac{1}{2}\left(\mu + \frac{1}{\lambda}\right)\|\hat{x}_k - u\|^2 \\ &\overset{(46)}{\leq} \phi(\hat{y}_k) - \hat{m}_k + \frac{1}{2\lambda}\|u - \hat{x}_{k-1}\|^2 \leq \frac{\bar{\varepsilon}}{2} + \frac{1}{2\lambda}\|u - \hat{x}_{k-1}\|^2\end{aligned},$$

and, hence, that (a) holds.

b and c. Because (45) is a special case of Inequality (B.2), in which

$$\eta_k = \phi(\hat{y}_k) - \phi(u), \quad \alpha_k = \frac{1}{2\lambda}\|\hat{x}_k - u\|^2, \quad \theta = 1 + \lambda\mu, \quad \delta = \frac{\bar{\varepsilon}}{2},$$

it follows from Corollary B.1 the fact that $\hat{x}_0 = x_0$ and the definition of λ_μ in (44) that (b) and (c) hold. □

We are now ready to present the main result of this subsection, which provides a bound on the number of serious iterates generated by GPB until it obtains a $\bar{\varepsilon}$-solution of (1).

Proposition 2. *The number of serious iterations K performed by* GPB *until it obtains for the first time an auxiliary serious iterate $\hat{y}_K$ such that $\phi(\hat{y}_K) - \phi^* \leq \bar{\varepsilon}$ is bounded by*

$$\min\left\{\frac{d_0^2}{\lambda\bar{\varepsilon}}, \frac{1}{\mu\lambda_\mu}\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\} + 1, \tag{47}$$

where λ_μ is as in (44). Moreover,

$$\|\hat{x}_k - x_0^*\| \leq \sqrt{2}d_0 \qquad \forall k \in \{0, 1, \ldots, K-1\}. \tag{48}$$

Proof. Lemma 2(b) with $u = x_0^*$ and the definition of d_0 in (8) imply the first conclusion of the proposition, and, hence, that $K - 1 \leq d_0^2/(\lambda\bar{\varepsilon})$. This conclusion, together with Lemma 2(c) with $u = x_0^*$, then implies (48). □

We note that Proposition 2 holds for any $\lambda > 0$.

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

4.2. Bounding the Number of Consecutive Null Iterates

Our goal in this subsection is to show that the set $\mathcal{C}_k$ is finite and also to provide a bound on its cardinality in terms of $\bar{M}_f$, $\bar{L}_f$, λ, $\bar{\varepsilon}$, d_0, and τ.

We start by noting that (35), the definition of m_j in (32), and the first identity in (43) imply that

$$m_j = \Gamma_j(x_j) + \frac{1}{2\lambda}\|x_j - \hat{x}_{k-1}\|^2 \quad \forall j \in \mathcal{C}_k. \tag{49}$$

The first result below describes some basic properties of a sequence of auxiliary bundle functions $\{\bar{\Gamma}_j\}$, whose existence is guaranteed by the nature of the BU blackbox.

Lemma 3. *For every $j \in \mathcal{C}_k \setminus \{j_k\}$, the following statements hold*:

a. *There exists function $\bar{\Gamma}_j(\cdot)$ such that*

$$\tau\bar{\Gamma}_j(\cdot) + (1-\tau)[\ell_f(\cdot; x_j) + h(\cdot)] \le \Gamma_{j+1}(\cdot) \le \phi(\cdot), \tag{50}$$

$$\bar{\Gamma}_j \in \overline{\mathrm{Conv}}_\mu(\mathbb{R}^n), \quad \bar{\Gamma}_j(x_j) = \Gamma_j(x_j), \quad x_j = \arg\min_{u\in\mathbb{R}^n}\left\{\bar{\Gamma}_j(u) + \frac{1}{2\lambda}\|u - \hat{x}_{k-1}\|^2\right\}; \tag{51}$$

b. *If λ_μ is as in (44), then for every $u \in \mathbb{R}^n$, we have*

$$\bar{\Gamma}_j(u) + \frac{1}{2\lambda}\|u - \hat{x}_{k-1}\|^2 \ge m_j + \frac{1}{2\lambda_\mu}\|u - x_j\|^2. \tag{52}$$

Proof.

a. This statement immediately follows from (17), (18), and the facts that Γ_{j+1} is the output of the BU blackbox with input (λ, τ) and $(x^c, x, \Gamma) = (x_j^c, x_j, \Gamma_j)$ (see the null update in step 1 of GPB) and $x_j^c = \hat{x}_{k-1}$.

b. It follows from $\bar{\Gamma}_j \in \overline{\mathrm{Conv}}_\mu(\mathbb{R}^n)$ and the definition of λ_μ in (44) that $\bar{\Gamma}_j + \|\cdot - \hat{x}_{k-1}\|^2/(2\lambda)$ is (λ_μ^{-1})-strongly convex. Using the second identity in (51) and Beck [2, theorem 5.25(b)] with $f = \bar{\Gamma}_j + \|\cdot - \hat{x}_{k-1}\|^2/(2\lambda)$, $x^* = x_j$ and $\sigma = \lambda_\mu^{-1}$, we have for every $u \in \mathrm{dom}\, h$,

$$\bar{\Gamma}_j(u) + \frac{1}{2\lambda}\|u - \hat{x}_{k-1}\|^2 \ge \bar{\Gamma}_j(x_j) + \frac{1}{2\lambda}\|x_j - \hat{x}_{k-1}\|^2 + \frac{1}{2\lambda_\mu}\|u - x_j\|^2.$$

The statement follows from the above inequality, the first identity in (51), and Relation (49). □

The following technical result provides an important recursive formula for $\{m_j\}$, which is used in Lemma 5 to give a recursive formula for $\{t_j\}$. It is worth observing that its proof uses for the first time the Condition (29).

Lemma 4. *Suppose (29) holds; then, for every $j \in \mathcal{C}_k \setminus \{j_k\}$, we have*

$$m_{j+1} \ge \tau m_j + (1-\tau)\left[\ell_f(x_{j+1}; x_j) + h(x_{j+1}) + \left(\frac{\bar{L}_f}{2} + \frac{4\bar{M}_f^2}{\bar{\varepsilon}}\right)\|x_{j+1} - x_j\|^2\right]. \tag{53}$$

Proof. First, it immediately follows from (29) and the definitions of $T_{\bar{\varepsilon}}$ and λ_μ in (9) and (44), respectively, that

$$\frac{\tau}{1-\tau} \ge \frac{8\lambda(\bar{M}_f^2 + \bar{\varepsilon}\bar{L}_f)}{(1+\lambda\mu)\bar{\varepsilon}} \ge \lambda_\mu\left(\bar{L}_f + \frac{8\bar{M}_f^2}{\bar{\varepsilon}}\right). \tag{54}$$

Using (49), (50), the fact that $\tau < 1$, and (52) with $u = x_{j+1}$, we have

$$\begin{aligned} m_{j+1} &\overset{(49)}{=} \Gamma_{j+1}(x_{j+1}) + \frac{1}{2\lambda}\|x_{j+1} - \hat{x}_{k-1}\|^2 \\ &\overset{(50)}{\ge} (1-\tau)[\ell_f(x_{j+1}; x_j) + h(x_{j+1})] + \tau\left(\bar{\Gamma}_j(x_{j+1}) + \frac{1}{2\lambda}\|x_{j+1} - \hat{x}_{k-1}\|^2\right), \\ &\overset{(52)}{\ge} (1-\tau)[\ell_f(x_{j+1}; x_j) + h(x_{j+1})] + \tau\left(m_j + \frac{1}{2\lambda_\mu}\|x_{j+1} - x_j\|^2\right) \end{aligned}$$

which, together with (54), implies (53). □

The next result, which plays an important role in the analysis of the null iterates, establishes a key recursive formula for the sequence $\{t_j\}$ defined in (32).

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

Lemma 5. *For every $j \in \mathcal{C}_k \setminus \{j_k\}$, we have*

$$t_{j+1} - \frac{\bar{\varepsilon}}{4} \le \tau\left(t_j - \frac{\bar{\varepsilon}}{4}\right). \tag{55}$$

Proof. Using (7) with $(M_f, L_f, x, y) = (\bar{M}_f, \bar{L}_f, x_{j+1}, x_j)$ and the fact that $\phi = f + h$, we have

$$\ell_f(x_{j+1}; x_j) + h(x_{j+1}) + \frac{\bar{L}_f}{2}\|x_{j+1} - x_j\|^2 \ge \phi(x_{j+1}) - 2\bar{M}_f\|x_{j+1} - x_j\|. \tag{56}$$

This inequality and (53) imply that

$$\begin{aligned} m_{j+1} - \tau m_j &\overset{(53)}{\ge} (1-\tau)\left[\ell_f(x_{j+1}; x_j) + h(x_{j+1}) + \left(\frac{\bar{L}_f}{2} + \frac{4\bar{M}_f^2}{\bar{\varepsilon}}\right)\|x_{j+1} - x_j\|^2\right] \\ &\overset{(56)}{\ge} (1-\tau)\phi(x_{j+1}) + \frac{1-\tau}{\bar{\varepsilon}}(4\bar{M}_f^2\|x_{j+1} - x_j\|^2 - 2\bar{M}_f\bar{\varepsilon}\|x_{j+1} - x_j\|) \\ &\ge (1-\tau)\phi(x_{j+1}) - \frac{(1-\tau)\bar{\varepsilon}}{4}, \end{aligned} \tag{57}$$

where the last inequality is due to the inequality $a^2 - 2ab \ge -b^2$ with $a = 2\bar{M}_f\|x_{j+1} - x_j\|$ and $b = \bar{\varepsilon}/2$. Using the above inequality and the definitions of y_{j+1} and t_{j+1} in (28) and (32), respectively, we conclude that

$$\begin{aligned} t_{j+1} &\overset{(32)}{=} \phi(y_{j+1}) - m_{j+1} \overset{(57)}{\le} \phi(y_{j+1}) - \tau m_j - (1-\tau)\phi(x_{j+1}) + \frac{(1-\tau)\bar{\varepsilon}}{4} \\ &\overset{(32)}{=} \phi(y_{j+1}) - \tau[\phi(y_j) - t_j] - (1-\tau)\phi(x_{j+1}) + \frac{(1-\tau)\bar{\varepsilon}}{4} \\ &\overset{(28)}{\le} \tau t_j + \frac{(1-\tau)\bar{\varepsilon}}{4}, \end{aligned}$$

and that the lemma holds. □

The next lemma gives a uniform bound on t_{j_k+1}, which is used in Proposition 3 to derive a uniform bound on the maximum number of consecutive null iterates generated by GPB. Its proof uses Lemma A.3 in Appendix A, where a crucial bound on $\|x_{j_k+1} - \hat{x}_k\| = \|x_{j_k+1} - x_{j_k}\|$ is obtained.

Lemma 6. *For every $k \ge 0$, we have $t_{j_k+1} \le \bar{t}$, where*

$$\bar{t} := \bar{M}_f^2 + 4(\bar{L}_f + 2)(\max\{1, 2\lambda\bar{L}_f\}d_0 + \lambda\bar{M}_f)^2. \tag{58}$$

Proof. Using both (28) and (32) with $j = j_k + 1$, Relation (49), and the facts that $\phi = f + h$ and $\Gamma_{j_k+1} \ge \ell_f(\cdot; x_{j_k}) + h$ (see the serious update in step 1 of GPB), we have

$$\begin{aligned} t_{j_k+1} &\overset{(32)}{=} \phi(y_{j_k+1}) - m_{j_k+1} \overset{(28),(49)}{\le} \phi(x_{j_k+1}) - \Gamma_{j_k+1}(x_{j_k+1}) \le f(x_{j_k+1}) - \ell_f(x_{j_k+1}; x_{j_k}) \\ &\overset{(7)}{\le} 2\bar{M}_f\|x_{j_k+1} - x_{j_k}\| + \frac{\bar{L}_f}{2}\|x_{j_k+1} - x_{j_k}\|^2 \le \bar{M}_f^2 + \left(\frac{\bar{L}_f}{2} + 1\right)\|x_{j_k+1} - x_{j_k}\|^2 \end{aligned},$$

where the third inequality is due to (7) with $(M_f, L_f, x, y) = (\bar{M}_f, \bar{L}_f, x_{j_k+1}, x_{j_k})$, and the last inequality is due to the fact that $2ab \le a^2 + b^2$ for every $a, b \in \mathbb{R}$. The conclusion of the lemma now follows from the above inequality and Lemma A.3 in Appendix A. □

We are now ready to present the main result of this subsection, where a bound on $|B(\ell_0)|$ is obtained in terms of τ, $\bar{t}$, and $\bar{\varepsilon}$.

Proposition 3. *The set $\mathcal{C}_k$ is finite and*

$$|\mathcal{C}_k| \le \frac{1}{1-\tau}\log\left(\frac{4\bar{t}}{\bar{\varepsilon}}\right) + 1, \tag{59}$$

where $\bar{t}$ is as in (58) and τ is as in step 0 of GPB. In particular, if τ is as in (38), then

$$|\mathcal{C}_k| \le \left(1 + \frac{8\lambda_\mu(M_f^2 + \bar{\varepsilon}L_f)}{\bar{\varepsilon}}\right)\log\left(\frac{4\bar{t}}{\bar{\varepsilon}}\right) + 1. \tag{60}$$

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

Proof. Using the inequality $\tau \le e^{\tau-1}$ and Lemmas 5 and 6, we then conclude that for every $j \in \mathcal{C}_k$,

$$t_j - \frac{\bar{\varepsilon}}{4} \le \tau^{j-j_{k-1}-1}\left(t_{j_{k-1}+1} - \frac{\bar{\varepsilon}}{4}\right) \le \tau^{j-j_{k-1}-1} t_{j_{k-1}+1} \le e^{(\tau-1)(j-j_{k-1}-1)}\bar{t}.$$

Using this observation, and noting that step 1 of GPB and the definition of $\mathcal{C}_k$ imply that $t_j > \bar{\varepsilon}/2$ for every $j \in \mathcal{C}_k \setminus \{j_k\}$, it is now easy to see that (59) follows. Because τ as in (38) satisfies (29), it immediately follows that (60) holds in view of (38) and (59). □

4.3. The Total Iteration-Complexity of GPB

This subsection establishes the total iteration-complexity of GPB.

We start by providing a more general version of Theorem 1 that does not impose any condition on λ.

Proposition 4. *Let $(x_0, \lambda, \bar{\varepsilon}) \in \operatorname{dom} h \times \mathbb{R}_{++} \times \mathbb{R}_{++}$ and τ as in (38) be given. Then, any variant of* GPB *with input $(x_0, \lambda, \bar{\varepsilon}, \tau)$ obtains a $\bar{\varepsilon}$-solution of (1) in a number of iterations bounded by*

$$\left[\left(1 + \frac{8\lambda_\mu(M_f^2 + \bar{\varepsilon}L_f)}{\bar{\varepsilon}}\right)\log\left(\frac{4\bar{t}}{\bar{\varepsilon}}\right) + 1\right]\left[\min\left\{\frac{d_0^2}{\lambda\bar{\varepsilon}}, \frac{1}{\mu\lambda_\mu}\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\} + 1\right], \tag{61}$$

where $\bar{t}$ is as in (6).

Proof. This proposition is a direct consequence of Propositions 2 and 3. □

Because τ-free GPB instances do not depend on τ, we can choose τ as in (38) with (M_f, L_f) replaced by $(\bar{M}_f, \bar{L}_f)$. Hence, the $\bar{\varepsilon}$-iteration complexity for τ-free GPB instances is (61), with $M_f^2 + \bar{\varepsilon}L_f$ replaced by $T_{\bar{\varepsilon}}$.

Proposition 4 allows us to make one additional remark about Theorem 1—namely, in the unusual case where the range of λ (39) is empty—that is, $C^2(M_f^2 + \bar{\varepsilon}L_f)d_0^2/\bar{\varepsilon}^2 < 1$—it can be easily seen that (61), up to a logarithmic term, reduces to $\mathcal{O}([\kappa+1][C^{-2}\kappa^{-1}+1])$, where $\kappa := \lambda(M_f^2 + \bar{\varepsilon}L_f)/\bar{\varepsilon}$. Hence, the $\bar{\varepsilon}$-iteration complexity of GPB with $\lambda = \bar{\varepsilon}/[C(M_f^2 + \bar{\varepsilon}L_f)]$ becomes $\mathcal{O}((1+C^{-1})^2)$, which shows that the instances of (1) for which (39) does not hold can be trivially solved by GPB with a proper choice of the prox stepsize.

We are now ready to prove Theorem 1.

Proof of Theorem 1. Defining

$$a = \frac{\lambda_\mu(M_f^2 + \bar{\varepsilon}L_f)}{\bar{\varepsilon}}, \quad b = \min\left\{\frac{d_0^2}{\lambda\bar{\varepsilon}}, \frac{1}{\mu\lambda_\mu}\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\}, \tag{62}$$

and using (61), we conclude that $\mathcal{O}((a+1)(b+1))$ is a $\bar{\varepsilon}$-iteration complexity bound for GPB up to a logarithmic term. We break the proof into two cases: (1) $\mu \le C(M_f^2 + \bar{\varepsilon}L_f)/\bar{\varepsilon}^2$; and (2) $\mu \ge C(M_f^2 + \bar{\varepsilon}L_f)/\bar{\varepsilon}^2$.

First, assume that case 1 holds. Using the definition of λ_μ in (44), the fact that $\mu \le C(M_f^2 + \bar{\varepsilon}L_f)/\bar{\varepsilon}^2$, and the first inequality in (37), we have

$$\frac{1}{\lambda_\mu} = \frac{1}{\lambda} + \mu \le \frac{2C(M_f^2 + \bar{\varepsilon}L_f)}{\bar{\varepsilon}}, \tag{63}$$

and, hence, $a \ge 1/(2C)$. Moreover, it follows from the definition of b in (62) and the second inequality in (37) that

$$b \ge \min\left\{\frac{1}{C}, \frac{1}{\mu\lambda_\mu}\log\left(\frac{\lambda\mu}{C} + 1\right)\right\}. \tag{64}$$

Using the fact that $\log(1+t) \ge t/(1+t)$ for every $t > 0$, we easily see that $\log(1+t) \ge t/2$ if $t \le 1$ and $\log(1+t) \ge \log 2 > 0$ if $t \ge 1$. This observation with $t = \lambda\mu/C$ and the definition of λ_μ in (44) then imply that

$$\frac{1}{\mu\lambda_\mu}\log\left(\frac{\lambda\mu}{C} + 1\right) \ge \min\left\{\frac{\lambda}{2\lambda_\mu C}, \left(1 + \frac{1}{\lambda\mu}\right)\log 2\right\} \ge \min\left\{\frac{1}{2C}, \log 2\right\},$$

and, hence, that $b \ge \min\{1/(2C), \log 2\}$. This inequality and the fact that $a \ge 1/(2C)$ imply that $\mathcal{O}((a+1)(b+1))$ is equal to $\mathcal{O}(ab+1)$. Using this observation, the definitions of a and b in (62), and the fact that $\lambda_\mu \le \lambda$, we then conclude that the bound $\mathcal{O}((a+1)(b+1))$ reduces to (5), and, hence, that the theorem holds for case 1.

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

Assume now that case 2 holds. Then, it follows from the definition of λ_μ in (44) and the first inequality in (37) that

$$\frac{1}{\lambda_\mu} = \mu + \frac{1}{\lambda} \geq \mu \geq \frac{C(M_f^2 + \bar{\varepsilon} L_f)}{\bar{\varepsilon}}, \quad \lambda\mu \geq 1. \tag{65}$$

The first inequality then implies that $a \leq 1/C$ in view of the first identity in (62), and, hence, that $\mathcal{O}((a+1)(b+1))$ is $\mathcal{O}(b+1)$. We will now derive a bound on b. Indeed, using the definitions of b and λ_μ in (62) and (44), respectively, we have

$$b = \min\left\{\frac{d_0^2}{\lambda\bar{\varepsilon}}, \left(1 + \frac{1}{\lambda\mu}\right)\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\} \leq \min\left\{\frac{C(M_f^2 + \bar{\varepsilon} L_f)d_0^2}{\bar{\varepsilon}^2}, 2\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\}, \tag{66}$$

where the inequality is due to the second inequality in (65) and the first inequality in (37). Hence, the bound $\mathcal{O}(b+1)$ becomes

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

Finally, it is easy to see that Bound (5) becomes the above bound when $\mu \geq C(M_f^2 + \bar{\varepsilon} L_f)/\bar{\varepsilon}^2$, and, hence, that the theorem holds for case 2. □

It is worth pointing out how Condition (37) on the prox stepsize is used in the proof of Theorem 1. Indeed, the first inequality in (37) is used to obtain the inequality in (63), the last inequality in (65), and the inequality in (66), whereas the second inequality in (37) is used to obtain (64).

5. A One-Cut Adaptive Proximal Bundle Method

This section presents an adaptive version of the 1C-PB method, referred to as the 1C-APB method, which, in contrast to 1C-PB, does not require the availability of a triple (L_f, M_f, μ) satisfying Assumptions 1 and 3 and which has the same $\bar{\varepsilon}$-iteration complexity as described in Corollary 1 for an arbitrary τ-free GPB variant.

We start by stating the 1C-APB method.

1C-APB

0. Let $x_0 \in \operatorname{dom} h$, $\lambda > 0$, $\beta \geq 1$ and $\bar{\varepsilon} > 0$ be given, and set $y_0 = x_0$, $t_0 = 0$, $\tau_0 = 0$, and $j = 1$;
1. Set $\tau = \tau_{j-1}/\beta$;
2. If $t_{j-1} \leq \bar{\varepsilon}/2$, then perform a **serious update**—that is, set $x_j^c = x_{j-1}$ and $\Gamma_j = \ell_f(\cdot; x_{j-1}) + h$; else, perform a **null update**—that is, set $x_j^c = x_{j-1}^c$ and $\Gamma_j = \tau\Gamma_{j-1} + (1-\tau)[\ell_f(\cdot; x_{j-1}) + h]$;
3. Compute x_j, y_j, m_j, and t_j as in step 2 of GPB;
4. If $t_{j-1} > \bar{\varepsilon}/2$ and $t_j > \tau t_{j-1} + (1-\tau)\bar{\varepsilon}/4$, then set $\tau = (1+\tau)/2$ and go to step 2; else, set $\tau_j = \tau$ and $j \leftarrow j+1$, and go to step 1.

We use the same terminology (e.g., serious iteration) as defined in the paragraph following GPB. For ease of discussion in this subsection, we define $\bar{\tau}$ as follows:

$$\bar{\tau} := \left[1 + \frac{(1+\lambda\mu)\bar{\varepsilon}}{8\lambda T_{\bar{\varepsilon}}^2}\right]^{-1}, \tag{67}$$

where $T_{\bar{\varepsilon}}$ is as in (9). We note that $\bar{\tau}$ is the smallest $\tau \in (0,1)$, satisfying (29).

We now make some remarks about the 1C-APB method. First, in contrast to the GPB framework, which does not specify how some quantities are generated, 1C-APB is a well-determined method because it specifies Γ_j in both the serious and null updates, the latter of which computes Γ_j based on the one-cut bundle update scheme Example 1. Second, the iteration count j is only increased in step 4, and when that happens, the key inequality

$$t_j - \frac{\bar{\varepsilon}}{4} \leq \tau_j\left(t_{j-1} - \frac{\bar{\varepsilon}}{4}\right), \tag{68}$$

is satisfied. Before that happens, 1C-APB can loop a few times between steps 2 and 4 and, in the process, computes intermediate quantities, which depend on τ and (with some abuse of notation) are all denoted by Γ_j, x_j, y_j, m_j, and t_j. Third, because $\tau_0 = 0 < \bar{\tau}$, it may happen that many τ_j's will also be less than $\bar{\tau}$. Hence, 1C-APB cannot be viewed as a special case of GPB because the latter one requires its constant τ to be at least $\bar{\tau}$. Finally, $\{\tau_j\}$ is a nondecreasing sequence if $\beta = 1$, but it can decrease if $\beta > 1$.

The following lemma summarizes some basic properties of 1C-APB.

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

Lemma 7. *The following statements about the* 1C-*APB method hold*:

a. $0 \le \tau_j \le (1+\bar{\tau})/2$ *for every* $j \ge 0$;

b. *For every serious iteration* j_k, $t_{j_k} \le \bar{\varepsilon}/2$, *and* (68) *holds for every* $j \in \mathcal{C}_k \setminus \{j_{k-1}+1\}$.

Proof.

a. It follows from Lemma 5 that if $\tau_j \ge \bar{\tau}$, then $\tau_\ell = \tau_j$ for every $\ell > j$. This statement now immediately follows from this observation, the fact that $\tau_0 = 0$, and the way the sequence $\{\tau_j\}$ is generated.

b. This statement follows immediately from steps 2 and 4 of 1C-APB. □

The following result is similar to Proposition 3 and establishes a bound on the maximum number of consecutive null iterates generated by 1C-APB.

Proposition 5. *The following statements about* 1C-*APB hold*:

a. *In each iteration, the number of times* τ *is updated in step* 4 *is at most*

$$1 + \left\lceil \log\left(1 + \frac{8\lambda_\mu T_{\bar{\varepsilon}}^2}{\bar{\varepsilon}}\right)\right\rceil; \tag{69}$$

b. *If* j_{k-1} *is a serious iteration of the* 1C-*APB method, then the next serious iteration* j_k *happens and satisfies*

$$j_k - j_{k-1} \le 2\left(1 + \frac{8\lambda_\mu T_{\bar{\varepsilon}}^2}{\bar{\varepsilon}}\right)\log\left(\frac{4\bar{t}}{\bar{\varepsilon}}\right) + 1,$$

where $T_{\bar{\varepsilon}}$, λ_μ, *and* $\bar{t}$ *are as in* (9), (44), *and* (58), *respectively*.

Proof.

a. It follows from the way τ is updated in step 4 that $1 - \tau^+ = (1-\tau)/2$, where τ^+ is the updated τ. Using this observation and Lemma 7(a), we then easily conclude that the number of times τ changes is bounded by $1 + \lceil \log(1/(1-\bar{\tau}))\rceil$. The conclusion in (a) now follows from the last conclusion and the definition of $\bar{\tau}$ in (67).

b. It follows from Lemma 7, (a) and (b) that for every $j \in \mathcal{C}_k \setminus \{j_{k-1}+1\}$,

$$t_j - \frac{\bar{\varepsilon}}{4} \le \frac{1+\bar{\tau}}{2}\left(t_{j-1} - \frac{\bar{\varepsilon}}{4}\right).$$

Using the inequality above, the fact that $t_{j_k} \le \bar{\varepsilon}/2$ (see Lemma 7(b)) and Proposition 3, we conclude that

$$j_k - j_{k-1} \le \frac{2}{1-\bar{\tau}}\log\left(\frac{4\bar{t}}{\bar{\varepsilon}}\right) + 1.$$

The above inequality, (44), and the definition of $\bar{\tau}$ in (67) immediately imply (b). □

We now discuss the $\bar{\varepsilon}$-iteration complexity of 1C-APB.

Theorem 2. *Let initial point* $x_0 \in \operatorname{dom} h$, *tolerance* $\bar{\varepsilon} > 0$, *and prox stepsize* $\lambda > 0$ *be given, and consider an instance* $(f, f'; h)$ *of* (1) *satisfying conditions Assumptions* 1–3. *Then, the* $\bar{\varepsilon}$-*iteration complexity for* 1C-*APB is*

$$\left[2\left(1 + \frac{8\lambda_\mu T_{\bar{\varepsilon}}^2}{\bar{\varepsilon}}\right)\log\left(\frac{4\bar{t}}{\bar{\varepsilon}}\right) + 1\right]\left[\min\left\{\frac{d_0^2}{\lambda\bar{\varepsilon}}, \frac{1}{\mu\lambda_\mu}\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\} + 1\right]. \tag{70}$$

As a consequence, if, in addition, the instance $(f, f'; h)$ *and the input triple* $(x_0, \lambda, \bar{\varepsilon})$ *satisfy* (39), *then the* $\bar{\varepsilon}$-*iteration complexity for* 1C-*APB is* (*up to a logarithmic term*) *given by* (16).

Proof. First, the same analysis as in Subsection 4.1 shows that the number of serious iterations of 1C-APB is bounded by (47). Hence, this conclusion and Proposition 5(b) imply that the $\bar{\varepsilon}$-iteration complexity for 1C-APB is given by (70). Letting $a = \lambda_\mu T_{\bar{\varepsilon}}^2/\bar{\varepsilon}$ and b be as in (62) and using (70), we have $\mathcal{O}((a+1)(b+1))$ is the $\bar{\varepsilon}$-iteration complexity for 1C-APB up to a logarithmic term. Using the Assumption (39) and following a similar argument as in the proof of Theorem 1, we conclude that the $\bar{\varepsilon}$-iteration complexity for 1C-APB is (up to a logarithmic term) given by (16). □

It is worth noting that a result similar to Corollary 2 dealing with instances $(f, f'; h)$ of (1) satisfying Assumption 1, Assumption 2, and (10) can also be established for 1C-APB.

We end this section by discussing the complexity of 1C-APB in terms of the total number of resolvent evaluations of ∂h—that is, an evaluation of the point-to-point operator $(I + \alpha\partial h)^{-1}(\cdot)$ for some $\alpha > 0$. Observe first that the computation of x_j in step 3 of 1C-APB requires one resolvent evaluation of ∂h due to (31) and the fact that Γ_j

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

has the form (19). Hence, the total number of resolvent evaluations of ∂h is bounded by the number that step 3 is performed. Thus, it follows from Theorem 2 and Proposition 5(a) that the total number of resolvent evaluations of ∂h is bounded by the product of (69) and (70) if $\beta > 1$ or the sum of (69) and (70) if $\beta = 1$.

6. Concluding Remarks

This paper presents a generic proximal bundle framework—namely, GPB—for solving the HCCO Problem (1). Instead of focusing on a specific bundle update scheme, GPB is based on a generic one—that is, the BU blackbox, which includes three schemes—namely, multiple-cuts Example 3, two-cuts Example 2, and a novel one-cut scheme Example 1. Moreover, this paper considers the hybrid case where Assumption 3 holds and presents a unified and simple analysis for GPB. It establishes two $\bar{\varepsilon}$-iteration complexity for GPB instances—namely, (5) for 1C-PB and (16) for the τ-free GPB instances (i.e., 2C-PB and MC-PB). Finally, this paper presents the 1C-APB method, which is an adaptive version of 1C-PB, and shows that 1C-APB has the same $\bar{\varepsilon}$-iteration complexity as the τ-free GPB instances.

We briefly discuss the relationship between GPB instances and other methods. First, the CS-CS method can be viewed as a special instance of any GPB variant with a relatively small prox stepsize. Second, it is worth noting that 1C-PB has slight similarity with the dual averaging (DA) method of Nesterov [19] because both methods explore the idea of aggregating cuts into a single one. However, there are essential differences between the two methods: (1) DA uses variable prox stepsizes, whereas 1C-PB uses a constant one; and (2) most importantly, 1C-PB updates the prox-center immediately after every serious iteration, whereas DA uses a fixed prox-center throughout the process.

Finally, we discuss some possible extensions of our analysis in this paper.

First, under the assumption that the diameter D of dom h is finite, it follows from Lan [11, subsection 3.1, last inequality] that the $\bar{\varepsilon}$-iteration complexity of an accelerated composite subgradient method proposed in Lan [11] is

$$\mathcal{O}\left(\frac{\sqrt{L_f}D}{\sqrt{\bar{\varepsilon}}} + \frac{M_f^2 D^2}{\bar{\varepsilon}^2}\right).$$

Moreover, it follows from Lan [11, introduction, paragraph containing equation (6)] that the above bound is optimal for the HCCO problem class determined by L_f, M_f, and D. In this regard, the $\bar{\varepsilon}$-iteration complexity of GPB is optimal when $L_f = 0$ (i.e., in the pure nonsmooth case), but it is not optimal when $L_f > 0$. It would be interesting to design an accelerated variant of GPB that is optimal for the aforementioned HCCO problem class.

Second, proximal bundle methods have not been studied in the context of stochastic subgradient oracles with continuous distribution, and, hence, it is interesting to investigate such methods by using the techniques developed in this paper.

Third, a drawback of GPB is that its cycle-termination criterion—namely, $t_j \le \bar{\varepsilon}/2$—depends on the tolerance $\bar{\varepsilon}$ specified for it. An interesting question is whether it is possible to develop a variant of GPB with a cycle-termination criterion that does not depend on the tolerance $\bar{\varepsilon}$.

Finally, we address issues related to the strongly convex case (i.e., $\mu > 0$). Our analysis assumes that f is convex and h is μ-convex and shows that (see Theorem 1), even though GPB does not require $\mu > 0$, the dependence of its Iteration-Complexity Bound (5) on μ and $\bar{\varepsilon}$ is (up to a logarithmic term) the same as that for the CS-CS method (see Proposition C.1). An interesting question is whether GPB or a related variant that does not require μ either directly applied to the HCCO Problem (1) also has the above iteration-complexity bound under the assumption that f is μ_f-convex, h is μ_h-convex, and $\mu = \mu_f + \mu_h$.

We now mention some papers and observations related to the topic of the previous paragraph. Under the assumption that $L_f = 0$ and $\mu_h = 0$ (and, hence, $\mu = \mu_f$), the proximal bundle method of Du and Ruszczyński [7] is shown to have an $\tilde{\mathcal{O}}(M_f^2/(\mu\bar{\varepsilon}))$ iteration-complexity bound, but requires μ_f as input because its prox stepsize is chosen as $\lambda = 1/\mu_f$. Moreover, for the same method and under the same assumptions, Díaz and Grimmer [6] improves the latter bound to $\mathcal{O}(M_f^2/(\mu\bar{\varepsilon}))$ by removing a logarithmic term. Finally, if μ_f is known and the new composite structure $(\tilde{f}, \tilde{h})$ defined as $\tilde{f} = f - \mu_f\|\cdot\|^2/2$ and $\tilde{h} = h + \mu_f\|\cdot\|^2/2$ is considered in place of (f, h), then GPB with this composite structure has iteration-complexity equal to (5), where $\mu = \mu_f + \mu_h$.

Acknowledgments

The authors are grateful to the anonymous referee and the associate editor for providing helpful comments on earlier versions of this manuscript.

Appendix A. Technical Results

The main result of this section is Lemma A.3, which was used in the proof of Lemma 6. It also presents the proof of Proposition 1.

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

Before stating and proving Lemma A.3, we first present two technical results.

Lemma A.1. *Let $x \in \mathbb{R}^n$, $0 < \tilde{\lambda} < \lambda$ and $\Gamma \in \overline{\mathrm{Conv}}(\mathbb{R}^n)$ be given, and define*

$$x^+ = \arg\min_{u\in\mathbb{R}^n}\left\{\Gamma(u) + \frac{1}{2\lambda}\|u-x\|^2\right\}, \quad \tilde{x}^+ = \arg\min_{u\in\mathbb{R}^n}\left\{\Gamma(u) + \frac{1}{2\tilde{\lambda}}\|u-x\|^2\right\}.$$

Then, we have $\|x^+ - x\| \le (\lambda/\tilde{\lambda})\|\tilde{x}^+ - x\|$.

Proof. Denote $\partial\Gamma$ by A, and define

$$y_A(\lambda;x) := (I+\lambda A)^{-1}(x), \quad \varphi_A(\lambda;x) := \lambda\|y_A(\lambda;x) - x\|.$$

It is easy to see that

$$\|x^+ - x\| = \|y_A(\lambda;x) - x\| = \frac{1}{\lambda}\varphi_A(\lambda;x), \|\tilde{x}^+ - x\| = \|y_A(\tilde{\lambda};x) - x\| = \frac{1}{\tilde{\lambda}}\varphi_A(\tilde{\lambda};x).$$

The conclusion of the lemma now follows from the above observation and Monteiro and Svaiter [18, the second inequality in (39)], which claims that

$$\varphi_A(\lambda;x) \le \frac{\lambda^2}{\tilde{\lambda}^2}\varphi_A(\tilde{\lambda};x). \quad \square$$

Lemma A.2. *Let $(\Gamma, z_0, \lambda) \in \overline{\mathrm{Conv}}_\mu(\mathbb{R}^n) \times \mathbb{R}^n \times (0, 1/\bar{L}_f)$ be a triple such that*

$$\ell_f(\cdot;z_0) + h \le \Gamma \le \phi, \tag{A.1}$$

and define

$$z := \arg\min_{u\in\mathbb{R}^n}\left\{\Gamma(u) + \frac{1}{2\lambda}\|u-z_0\|^2\right\}. \tag{A.2}$$

Then, for every $u \in \mathrm{dom}\, h$, we have

$$\frac{1}{2}\left(\mu + \frac{1}{\lambda}\right)\|u-z\|^2 + \phi(z) - \phi(u) \le \frac{1}{2\lambda}\|u-z_0\|^2 + \frac{2\lambda\bar{M}_f^2}{1-\lambda\bar{L}_f}. \tag{A.3}$$

Proof. It follows from the assumption that $\Gamma \in \overline{\mathrm{Conv}}_\mu(\mathbb{R}^n)$ that the function $\Gamma + \|\cdot - z_0\|^2/(2\lambda)$ is $(\mu+\lambda^{-1})$-strongly convex. This conclusion, (A.1), (A.2), and Beck [2, theorem 5.25(b)] with $f = \Gamma + \|\cdot - z_0\|^2/(2\lambda)$, $x^* = z$ and $\sigma = \mu + \lambda^{-1}$ then imply that for every $u \in \mathrm{dom}\, h$,

$$\begin{aligned}
\phi(u) + \frac{1}{2\lambda}\|u-z_0\|^2 &\overset{(A.1)}{\ge} \Gamma(u) + \frac{1}{2\lambda}\|u-z_0\|^2 \\
&\overset{(A.2)}{\ge} \Gamma(z) + \frac{1}{2\lambda}\|z-z_0\|^2 + \frac{1}{2}\left(\mu+\frac{1}{\lambda}\right)\|u-z\|^2 \\
&\overset{(A.1)}{\ge} \ell_f(z;z_0) + h(z) + \frac{1}{2\lambda}\|z-z_0\|^2 + \frac{1}{2}\left(\mu+\frac{1}{\lambda}\right)\|u-z\|^2.
\end{aligned}$$

The above inequality, the fact that $\phi = f + h$ and (7) with $(M_f, L_f, x, y) = (\bar{M}_f, \bar{L}_f, z, z_0)$ then imply that

$$\begin{aligned}
\frac{1}{2}\left(\mu+\frac{1}{\lambda}\right)\|u-z\|^2 + \phi(z) - \phi(u) &\le \frac{1}{2\lambda}\|u-z_0\|^2 + \phi(z) - \ell_f(z;z_0) - h(z) - \frac{1}{2\lambda}\|z-z_0\|^2 \\
&\overset{(7)}{\le} \frac{1}{2\lambda}\|u-z_0\|^2 + 2\bar{M}_f\|z-z_0\| - \frac{1-\lambda\bar{L}_f}{2\lambda}\|z-z_0\|^2.
\end{aligned}$$

The lemma now follows from the above inequality, the fact that $\lambda\bar{L}_f < 1$, and the inequality $2ab - a^2 \le b^2$ with $a^2 = (1-\lambda\bar{L}_f)\|z-z_0\|^2/(2\lambda)$ and $b^2 = 2\lambda\bar{M}_f^2/(1-\lambda\bar{L}_f)$. $\square$

We are now ready to prove the main technical result of this section, which provides a bound on the distance between a serious iterate generated by GPB and its consecutive (possibly null or serious) iterate. It is worth noting that this result is quite general and makes no use of the generic bundle update scheme of Subsection 3.1 because the step from x_{ℓ_0} to x_{ℓ_0+1} does not use this update.

Lemma A.3. *If ℓ_0 is a serious iteration, then*

$$\|x_{\ell_0} - x_{\ell_0+1}\| \le 2\sqrt{2}(\max\{1, 2\lambda\bar{L}_f\}d_0 + \lambda\bar{M}_f). \tag{A.4}$$

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

Proof. For the sake of this proof only, we define the auxiliary stepsize $\tilde{\lambda} := \min\{\lambda, 1/(2\bar{L}_f)\}$ and auxiliary point

$$w_{\ell_0} := \arg\min_{u\in\mathbb{R}^n}\left\{\Gamma_{\ell_0+1}(u) + \frac{1}{2\tilde{\lambda}}\|u - x_{\ell_0}\|^2\right\}.$$

Because $j = \ell_0$ is a serious index, it follows from step 1 of GPB that $\Gamma_{\ell_0+1} \geq \ell_f(\cdot; x_{\ell_0}) + h$, and, hence, that $(\Gamma, z_0, \lambda) = (\Gamma_{\ell_0+1}, x_{\ell_0}, \tilde{\lambda})$ and $z = w_{\ell_0}$ satisfy the assumptions of Lemma A.2. The conclusion of Lemma A.2 with $(u, z, z_0, \lambda) = (x_0^*, w_{\ell_0}, x_{\ell_0}, \tilde{\lambda})$ and the fact that $\tilde{\lambda} \leq 1/(2\bar{L}_f)$ then imply that

$$\frac{1}{2}\left(\mu + \frac{1}{\tilde{\lambda}}\right)\|x_0^* - w_{\ell_0}\|^2 + \phi(w_{\ell_0}) - \phi(x_0^*) \leq \frac{1}{2\tilde{\lambda}}\|x_0^* - x_{\ell_0}\|^2 + 4\tilde{\lambda}\bar{M}_f^2,$$

which, in turn, in view of the facts that $\phi(w_{\ell_0}) \geq \phi^* = \phi(x_0^*)$ and $\mu \geq 0$, and the inequality $(a+b)^{1/2} \leq a^{1/2} + b^{1/2}$ for any $a, b \geq 0$, yields

$$\|x_0^* - w_{\ell_0}\| \leq \|x_{\ell_0} - x_0^*\| + 2\sqrt{2}\tilde{\lambda}\bar{M}_f.$$

This inequality and the triangle inequality then imply that

$$\|x_{\ell_0} - w_{\ell_0}\| \leq \|x_{\ell_0} - x_0^*\| + \|x_0^* - w_{\ell_0}\| \leq 2\|x_{\ell_0} - x_0^*\| + 2\sqrt{2}\tilde{\lambda}\bar{M}_f \leq 2\sqrt{2}(d_0 + \tilde{\lambda}\bar{M}_f), \tag{A.5}$$

where the last inequality is due to (48) and the fact that x_{ℓ_0} is equal to one of serious iterates $\hat{x}_k$ preceding the last one generated by GPB. On the other hand, because $0 < \tilde{\lambda} < \lambda$ and $\Gamma_{\ell_0+1} \in \overline{\mathrm{Conv}}(\mathbb{R}^n)$, it follows from Lemma A.1 with $(\Gamma, x) = (\Gamma_{\ell_0+1}, x_{\ell_0})$ that

$$\|x_{\ell_0+1} - x_{\ell_0}\| \leq \frac{\lambda}{\tilde{\lambda}}\|w_{\ell_0} - x_{\ell_0}\|.$$

This inequality, together with (A.5) and the fact that $\lambda/\tilde{\lambda} = \max\{1, 2\lambda\bar{L}_f\}$, clearly implies (A.4). □

We end this section by providing the proof of Proposition 1.

Proof of Proposition 1. Using Young's inequality

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q},$$

with

$$a = \|x - y\|^\nu\left(\frac{1-\nu}{\alpha}\right)^{1-\nu}, \quad b = \left(\frac{\alpha}{1-\nu}\right)^{1-\nu}, \quad p = \frac{1}{\nu}, \quad q = \frac{1}{1-\nu},$$

where $\alpha > 0$ is arbitrary, we have

$$\|x - y\|^\nu \leq \nu\left(\frac{1-\nu}{\alpha}\right)^{\frac{1-\nu}{\nu}}\|x - y\| + \alpha.$$

It follows from (10) and the above inequality that

$$\|f'(x) - f'(y)\| \leq 2M_\nu + L_\nu\alpha + L_\nu\nu\left(\frac{1-\nu}{\alpha}\right)^{\frac{1-\nu}{\nu}}\|x - y\|,$$

and, hence, that Assumption 3 holds with $(M_f, L_f) = (M_f(\alpha), L_f(\alpha))$ in view of (11). Moreover, using (11) and the fact that $(a+b)^2 \leq 2a^2 + 2b^2$ for every $a, b \in \mathbb{R}$, we have

$$\begin{aligned}\inf_{\alpha>0}\{M_f(\alpha)^2 + \bar{\varepsilon}L_f(\alpha)\} &\leq \min_{\alpha>0}\left\{2M_\nu^2 + \frac{L_\nu^2\alpha^2}{2} + \bar{\varepsilon}L_\nu\nu\left(\frac{1-\nu}{\alpha}\right)^{\frac{1-\nu}{\nu}}\right\}\\ &= 2M_\nu^2 + \bar{\varepsilon}^{\frac{2\nu}{\nu+1}}L_\nu^{\frac{2}{\nu+1}}\left[\frac{1}{2}(1-\nu)^{\frac{2}{\nu+1}} + \nu(1-\nu)^{\frac{1-\nu}{\nu+1}}\right]\\ &\leq 2M_\nu^2 + 2\bar{\varepsilon}^{\frac{2\nu}{\nu+1}}L_\nu^{\frac{2}{\nu+1}},\end{aligned}$$

where the minimization problem is minimized at

$$\alpha = \left(\frac{\bar{\varepsilon}}{L_\nu}\right)^{\frac{\nu}{\nu+1}}(1-\nu)^{\frac{1}{\nu+1}},$$

and the second inequality is due to the fact that $\nu \in (0, 1)$. Hence, (12) holds. Finally, (13) immediately follows from the definition of $T_{\bar{\varepsilon}}$ in (9), (12), and the fact that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for every $a, b \in \mathbb{R}_{++}$. □

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

Appendix B. Useful Recursive Formulas

The following two technical results play important roles in the complexity analysis of both GPB and CS-CS. We start by stating the following simple result for general sequences of nonnegative scalars.

Lemma B.1. *Assume that sequences of nonnegative scalars $\{\theta_j\}$, $\{\delta_j\}$, $\{\eta_j\}$ and $\{\alpha_j\}$ satisfy for every $j \geq 1$, $\theta_j \geq 1$, $\delta_j > 0$ and*

$$\eta_j \leq \alpha_{j-1} - \theta_j \alpha_j + \delta_j. \tag{B.1}$$

Let $\Theta_0 := 1$ and $\Theta_j := \Pi_{i=1}^{j} \theta_i$ for every $j \geq 1$; then, we have for every $k \geq 1$,

$$\sum_{j=1}^{k} \Theta_{j-1}\eta_j \leq \alpha_0 - \Theta_k \alpha_k + \sum_{j=1}^{k} \Theta_{j-1}\delta_j.$$

Proof. Multiplying (B.1) by Θ_{j-1} and summing the resulting inequality from $j = 1$ to k, we have

$$\sum_{j=1}^{k} \Theta_{j-1}\eta_j \leq \sum_{j=1}^{k} \Theta_{j-1}(\alpha_{j-1} - \theta_j\alpha_j + \delta_j) = \alpha_0 - \Theta_k\alpha_k + \sum_{j=1}^{k} \Theta_{j-1}\delta_j.$$

Hence, the lemma holds. □

The next result discusses a special case of the previous lemma in which $\theta_j = \theta$ and $\delta_j = \delta$ for every $j \geq 1$.

Corollary B.1. *Assume that scalars $\theta \geq 1$ and $\delta > 0$ and sequences of nonnegative scalars $\{\eta_j\}$ and $\{\alpha_j\}$ satisfy*

$$\eta_j \leq \alpha_{j-1} - \theta\alpha_j + \delta \quad \forall j \geq 1. \tag{B.2}$$

Then, the following statements hold:

a. $\min_{1\leq j\leq k}\eta_j \leq 2\delta$ *for every $k \geq 1$ such that*

$$k \geq \min\left\{\frac{\alpha_0}{\delta}, \frac{\theta}{\theta-1}\log\left(\frac{\alpha_0(\theta-1)}{\delta}+1\right)\right\},$$

with the convention that the second term is equal to the first term when $\theta = 1$ (note that the second term converges to the first term as $\theta \downarrow 1$);

b. $\alpha_k \leq \alpha_0 + k\delta$ *for every $k \geq 1$.*

Proof.

a. It follows from Lemma B.1 with $\theta_j = \theta$ and $\delta_j = \delta$ for every $j \geq 1$ that

$$\sum_{j=1}^{k} \theta^{j-1}\left[\min_{1\leq j\leq k}\eta_j\right] \leq \sum_{j=1}^{k} \theta^{j-1}\eta_j \leq \alpha_0 - \theta^k\alpha_k + \sum_{j=1}^{k}\theta^{j-1}\delta. \tag{B.3}$$

Using the fact that $\theta \geq e^{(\theta-1)/\theta}$ for every $\theta \geq 1$, we have

$$\sum_{j=1}^{k}\theta^{j-1} = \max\left\{k, \frac{\theta^k-1}{\theta-1}\right\} \geq \max\left\{k, \frac{e^{(\theta-1)k/\theta}-1}{\theta-1}\right\}.$$

This inequality, (B.3), and the fact that $\alpha_k \geq 0$ imply that for every $k \geq 1$,

$$\min_{1\leq j\leq k}\eta_j \leq \alpha_0 \min\left\{\frac{1}{k}, \frac{\theta-1}{e^{(\theta-1)k/\theta}-1}\right\} + \delta,$$

which can be easily seen to imply (a).

b. This statement follows from (B.3), the fact that $\eta_j \geq 0$, and the assumption that $\theta \geq 1$. □

Appendix C. The Composite Subgradient Method

This section contains two subsections. The first one provides the analysis of the CS-CS method, which is used to derive the $\bar{\varepsilon}$-iteration complexity of CS-CS in Subsection 2.2. The second one presents an adaptive variant of CS-CS and establishes the $\bar{\varepsilon}$-iteration complexity of it.

C.1. Analysis of CS-CS

Proposition C.1. *Let an initial point $x_0 \in \operatorname{dom} h$, $(L_f, M_f) \in \mathbb{R}^2_+$ and instance $(f, f'; h)$ satisfying conditions Assumptions 1–3 be given. Then, the number of iterations performed by CS-CS(x_0, λ) with $\lambda \leq \bar{\varepsilon}/[4(M_f^2 + \bar{\varepsilon}L_f)]$ until it finds a $\bar{\varepsilon}$-solution is bounded by*

$$\left\lfloor \min\left\{\frac{d_0^2}{\lambda\bar{\varepsilon}}, \frac{1+\lambda\mu}{\lambda\mu}\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}}+1\right)\right\}\right\rfloor + 1.$$

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

Proof. Recall that an iteration of CS-CS(x_0, λ) is as in (14). Noting that (14) satisfies (A.2) with $(z_0, z, \Gamma) = (x_j, x_{j+1}, \ell_f(\cdot; x_j) + h)$, and using the facts that $\lambda \le \bar{\varepsilon}/[4(M_f^2 + \bar{\varepsilon} L_f)] \le \bar{\varepsilon}/(4T_{\bar{\varepsilon}}^2) < 1/\bar{L}_f$, $\ell_f(\cdot; x_j) + h \in \overline{\mathrm{Conv}}_\mu(\mathbb{R}^n)$, and $\ell_f(\cdot; x_j) + h \le \phi$, we conclude that the assumptions of Lemma A.2 is satisfied. Hence, it follows from (A.3) with $(u, z, z_0) = (x_0^*, x_{j+1}, x_j)$ that

$$\phi(x_{j+1}) - \phi^* - \frac{1}{2\lambda}\|x_0^* - x_j\|^2 + \frac{1+\lambda\mu}{2\lambda}\|x_0^* - x_{j+1}\|^2 \le \frac{2\lambda \bar{M}_f^2}{1 - \lambda \bar{L}_f} \le \frac{\bar{\varepsilon}}{2},$$

where the last inequality is due to the facts that $2\lambda\bar{M}_f^2/(1-\lambda\bar{L}_f)$ is an increasing function in λ and $\lambda \le \bar{\varepsilon}/(4T_{\bar{\varepsilon}}^2)$. Because the above inequality with $j = j-1$ satisfies (B.2) with

$$\eta_j = \phi(x_j) - \phi^*, \quad \alpha_j = \frac{1}{2\lambda}\|x_j - x_0^*\|^2, \quad \theta = 1 + \lambda\mu, \quad \delta = \frac{\bar{\varepsilon}}{2},$$

it follows from Corollary B.1(a) and the fact that $\alpha_0 = d_0^2/(2\lambda)$ that $\min_{1\le j\le k}\phi(x_j) - \phi^* \le \bar{\varepsilon}$ for every index $k \ge 1$ such that

$$k \ge \min\left\{\frac{d_0^2}{\lambda\bar{\varepsilon}}, \frac{1+\lambda\mu}{\lambda\mu}\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\},$$

and, hence, that the lemma holds. □

C.2. An Adaptive CS Method

This subsection presents an adaptive variant of the CS-CS method—namely, the A-CS method—and establishes $\bar{\varepsilon}$-iteration complexity of the adaptive method. The proposed method is a universal method for solving the HCCO Problem (1) because it does not rely on any problem parameters.

A-CS

0. Let $x_0 \in \mathrm{dom}\, h$, $\lambda_0 > 0$ and $\bar{\varepsilon} > 0$ be given, and set $\lambda = \lambda_0$ and $j = 0$;
1. Compute
$$x = \arg\min_{u\in\mathbb{R}^n}\left\{\ell_f(u; x_j) + h(u) + \frac{1}{2\lambda}\|u - x_j\|^2\right\};$$
2. **If** $f(x) - \ell_f(x; x_j) - \|x - x_j\|^2/(2\lambda) > \bar{\varepsilon}/2$, **then** set $\lambda = \lambda/2$ and go to step 1; **else**, go to step 3;
3. Set $\lambda_{j+1} = \lambda$, $x_{j+1} = x$ and $j \leftarrow j+1$, and go to step 1.

Lemma C.1. *The following statements hold for A-CS$(\lambda_0, \bar{\varepsilon})$:*

a. *For every $j \ge 0$, we have*
$$x_{j+1} = \arg\min_{u\in\mathbb{R}^n}\left\{\ell_f(u; x_j) + h(u) + \frac{1}{2\lambda_{j+1}}\|u - x_j\|^2\right\}, \tag{C.1}$$
$$f(x_{j+1}) - \ell_f(x_{j+1}; x_j) - \frac{1}{2\lambda_{j+1}}\|x_{j+1} - x_j\|^2 \le \frac{\bar{\varepsilon}}{2}; \tag{C.2}$$
b. *If $\lambda_j \le \bar{\varepsilon}/(4T_{\bar{\varepsilon}}^2)$, where $T_{\bar{\varepsilon}}$ is as in (9), then (C.2) holds with $\lambda_{j+1} = \lambda_j$;*
c. *$\{\lambda_j\}$ is a nonincreasing sequence;*
d. *For every $j \ge 0$,*
$$\lambda_j \ge \underline{\lambda} := \min\left\{\frac{\bar{\varepsilon}}{8T_{\bar{\varepsilon}}^2}, \lambda_0\right\}. \tag{C.3}$$

Proof.

a. This statement directly follows from the description of A-CS.
b. Using (7) with $(M_f, L_f, x, y) = (\bar{M}_f, \bar{L}_f, x_{j+1}, x_j)$ and the inequality that $a^2 + b^2 \ge 2ab$ for $a, b \in \mathbb{R}$, we have
$$\begin{aligned} f(x_{j+1}) - \ell_f(x_{j+1}; x_j) - \frac{1}{2\lambda_j}\|x_{j+1} - x_j\|^2 &\le 2\bar{M}_f\|x_{j+1} - x_j\| - \frac{1-\lambda_j\bar{L}_f}{2\lambda_j}\|x_{j+1} - x_j\|^2 \\ &\le \frac{2\lambda_j\bar{M}_f^2}{1-\lambda_j\bar{L}_f} \le \frac{\bar{\varepsilon}}{2} \end{aligned},$$
where the last inequality is due to the assumption that $\lambda_j \le \bar{\varepsilon}/(4T_{\bar{\varepsilon}}^2)$. Hence, (C.2) holds with $\lambda_{j+1} = \lambda_j$.
c. This statement clearly follows from steps 2 and 3 of A-CS.
d. This statement follows trivially from (b) and (c) and the way λ is updated in step 2. □

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

Proposition C.2. *Let an initial point x_0 and a universal constant $C > 0$ be given, and consider an instance $(f, f'; h)$ of (1) satisfying conditions Assumptions 1–3. Moreover, assume $(\lambda_0, \bar{\varepsilon}) \in \mathbb{R}^2_{++}$ is such that $\lambda_0 \geq \bar{\varepsilon}/(CT^2_{\bar{\varepsilon}})$, where $T_{\bar{\varepsilon}}$ is as in (9). Then, the following statements hold*:

a. *A-CS$(\lambda_0, \bar{\varepsilon})$ has $\bar{\varepsilon}$-iteration complexity given by (16)*;

b. *The total number of times λ is halved in step 2 is bounded by*

$$\left\lceil \log\left(\max\left\{\frac{8\lambda_0 T^2_{\bar{\varepsilon}}}{\bar{\varepsilon}}, 1\right\}\right)\right\rceil.$$

Proof.

a. It follows from the fact that h is μ-convex that the objective function in (C.1) is $(\mu + \lambda_{j+1}^{-1})$-strongly convex. Using this conclusion, (C.1), and Beck [2, theorem 5.25(b)], we have for every $u \in \operatorname{dom} h$,

$$\begin{aligned}\ell_f(x_{j+1}; x_j) + h(x_{j+1}) + \frac{1}{2\lambda_{j+1}}\|x_{j+1} - x_j\|^2 + \frac{1}{2}\left(\mu + \frac{1}{\lambda_{j+1}}\right)\|u - x_{j+1}\|^2 \\ \leq \ell_f(u; x_j) + h(u) + \frac{1}{2\lambda_{j+1}}\|u - x_j\|^2 \leq \phi(u) + \frac{1}{2\lambda_{j+1}}\|u - x_j\|^2.\end{aligned}$$

It follows from the above inequality with $u = x_0^*$ and (C.1) that

$$\begin{aligned}(1 + \lambda_{j+1}\mu)\|x_0^* - x_{j+1}\|^2 + 2\lambda_{j+1}[\phi(x_{j+1}) - \phi^*] - \|x_0^* - x_j\|^2 \\ \leq 2\lambda_{j+1}\left[f(x_{j+1}) - \ell_f(x_{j+1}; x_j) - \frac{1}{2\lambda_{j+1}}\|x_{j+1} - x_j\|^2\right] \leq \bar{\varepsilon}\lambda_{j+1}.\end{aligned}$$

Because the above inequality with $j = j - 1$ satisfies (B.1) with

$$\eta_j = 2\lambda_j[\phi(x_j) - \phi^*], \quad \alpha_j = \|x_j - x_0^*\|^2, \quad \theta_j = 1 + \lambda_j\mu, \quad \delta_j = \bar{\varepsilon}\lambda_j,$$

it follows from Lemma B.1 and the fact that $\alpha_0 = d_0^2$ that

$$\left(\sum_{j=1}^{k} 2\lambda_j\Theta_{j-1}\right)\min_{1\leq j\leq k}[\phi(x_j) - \phi^*] \leq \sum_{j=1}^{k} 2\lambda_j\Theta_{j-1}[\phi(x_j) - \phi^*] \leq d_0^2 + \left(\sum_{j=1}^{k} 2\lambda_j\Theta_{j-1}\right)\frac{\bar{\varepsilon}}{2}, \tag{C.4}$$

where $\Theta_j = \Pi_{i=1}^{j}(1 + \lambda_j\mu)$ for every $j \geq 1$. Note that it follows from Lemma C.1(d) that $\Theta_j \geq (1 + \underline{\lambda}\mu)^j$ for every $j \geq 1$. Using this observation, (C.4), and Lemma C.1(d), and following the argument in the proof of Corollary B.1(a), we conclude that $\min_{1\leq j\leq k}\phi(x_j) - \phi^* \leq \bar{\varepsilon}$ for k, satisfying

$$k \geq \min\left\{\frac{d_0^2}{\underline{\lambda}\bar{\varepsilon}}, \frac{1 + \underline{\lambda}\mu}{\underline{\lambda}\mu}\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\},$$

and, hence, that the statement holds in view of (C.3) and the assumption that $\lambda_0 \geq \bar{\varepsilon}/(CT^2_{\bar{\varepsilon}})$.

b. This statement immediately follows from the update rule in λ_j and Lemma C.1(d). □

It is worth noting that a result similar to Corollary 2 dealing with instances $(f, f'; h)$ of (1) satisfying Assumption 1, Assumption 2, and (10) can also be established for A-CS.

Appendix D. Properties of Bundle Update Schemes Example 2 and Example 3

This section shows that the update schemes Example 2 and Example 3 of Subsection 3.1 are special implementations of BU.

Proposition D.1. *Consider the update Γ^+ of Example 2 and set $\bar{\Gamma} = A_f^+ + h$, where A_f^+ is as in (23). Then, $(\Gamma^+, \bar{\Gamma})$ satisfies (17) and (18). As a consequence, Γ^+ is a special implementation of BU.*

Proof. First, using the facts that $h \in \overline{\operatorname{Conv}}_\mu(\mathbb{R}^n)$ and $\ell_f(\cdot; z) \leq f$ for any $z \in \mathbb{R}^n$, the fact that $\bar{\Gamma} = A_f^+ + h$, and the definition of Γ^+ in (24), we have

$$\Gamma^+, \bar{\Gamma} \in \overline{\operatorname{Conv}}_\mu(\mathbb{R}^n), \quad \Gamma^+ \leq \phi.$$

We have thus shown the inclusion and the second inequality in (17) and the inclusion in (18). It follows from the definitions of Γ^+ and $\bar{\Gamma}$ that

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

and, hence, that Γ^+ satisfies the first inequality in (17) for any $\tau \in (0, 1)$. Moreover, using the fact that $\bar{\Gamma} = A_f^+ + h$, Relation (22), and the definitions of Γ and A_f^+ in (20) and (23), respectively, we have

$$\bar{\Gamma}(x) = A_f^+(x) + h(x) \overset{(23)}{=} \theta A_f(x) + (1 - \theta)\ell_f(x; x^-) + h(x) \overset{(22)}{=} \max\{A_f(x), \ell_f(x; x^-)\} + h(x) \overset{(20)}{=} \Gamma(x),$$

and, hence, the first identity in (18) holds. Finally, we prove $\bar{\Gamma}$ satisfies the second identity in (18).

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

It follows from the definition of $\ell_f(\cdot;\cdot)$ in (6) and Relations (21) and (23) that

$$\frac{1}{\lambda}(x-x^c)+\partial h(x)+\nabla A_f^+ \overset{(6),(23)}{=} \frac{1}{\lambda}(x-x^c)+\partial h(x)+\theta\nabla A_f+(1-\theta)f'(x^-) \overset{(21)}{\ni} 0. \quad \square$$

In conclusion, Γ^+ as in Example 2 is a special way of implementing BU.

Proposition D.2. *Consider the update Γ^+ of Example 3 and set $\bar{\Gamma}=\Gamma(\cdot;B(x))$, where $B(x)$ is as in (26) and $\Gamma(\cdot;B(x))$ is as in (4). Then, $(\Gamma^+,\bar{\Gamma})$ satisfies (17) and (18). As a consequence, Γ^+ is a special implementation of* BU.

Proof. First, using the facts that $h\in\overline{\mathrm{Conv}}_\mu(\mathbb{R}^n)$ and $\ell_f(\cdot;b)\le f$ for any $b\in\mathbb{R}^n$, and the definition of $\Gamma(\cdot;B)$ in (4), it is easy to see that for any $B\subset\mathbb{R}^n$, we have

$$\Gamma(\cdot;B)\in\overline{\mathrm{Conv}}_\mu(\mathbb{R}^n),\quad \Gamma(\cdot;B)\le\phi. \tag{D.1}$$

Recall that

$$\Gamma^+=\Gamma(\cdot;B^+),\quad \bar{\Gamma}=\Gamma(\cdot;B(x)), \tag{D.2}$$

hence, it follows from (D.1) that

$$\Gamma^+,\bar{\Gamma}\in\overline{\mathrm{Conv}}_\mu(\mathbb{R}^n),\quad \Gamma^+\le\phi.$$

We have thus shown the inclusion and the second inequality in (17) and the inclusion in (18). Also, it is easy to see from (D.2), the first inclusion in (25), and the definition of $\Gamma(\cdot;B)$ in (4) that

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

and, hence, that Γ^+ satisfies the first inequality in (17) for any $\tau\in(0,1)$. Moreover, it follows from the fact that $\Gamma=\Gamma(\cdot;B)$, (4), (D.2), and the definition of $B(x)$ in (26) that the first identity in (18) holds. Finally, we prove $\bar{\Gamma}$ satisfies the second identity in (18). Using the definitions of $\Gamma(\cdot;B)$ and $B(x)$ in (4) and (26), respectively, and a well-known formula for the subdifferential of the pointwise maximum of finitely many convex functions (e.g., see Urruty and Lemaréchal [22, corollary 4.3.2]), we conclude that

$$\partial\Gamma(x)=\overline{\mathrm{co}}\left(\cup\{f'(b):b\in B(x)\}\right)+\partial h(x).$$

Using the same reasoning, but with Γ replaced by $\bar{\Gamma}$, we conclude that the above set is also $\partial\bar{\Gamma}(x)$, and, hence, that

$$\frac{1}{\lambda}(x_0-x)\in\partial\Gamma(x)=\partial\bar{\Gamma}(x),$$

where the inclusion is due to (2). Now, the second identity in (18) immediately follows. In conclusion, Γ^+ as in Example 3 is a special way of implementing BU. $\square$

References

[1] Astorino A, Frangioni A, Fuduli A, Gorgone E (2013) A nonmonotone proximal bundle method with (potentially) continuous step decisions. *SIAM J. Optim.* 23(3):1784–1809.
[2] Beck A (2017) *First-Order Methods in Optimization*, *MOS-SIAM Series on Optimization*, vol. 25 (Society for Industrial and Applied Mathematics, Philadelphia).
[3] Ben-Tal A, Nemirovski A (2005) Non-Euclidean restricted memory level method for large-scale convex optimization. *Math. Program.* 102(3):407–456.
[4] de Oliveira W, Solodov M (2016) A doubly stabilized bundle method for nonsmooth convex optimization. *Math. Program.* 156(1–2):125–159.
[5] de Oliveira W, Sagastizábal C, Lemaréchal C (2014) Convex proximal bundle methods in depth: A unified analysis for inexact oracles. *Math. Program.* 148(1–2):241–277.
[6] Díaz M, Grimmer B (2021) Optimal convergence rates for the proximal bundle method. Preprint, submitted May 17, https://arxiv.org/abs/2105.07874.
[7] Du Y, Ruszczyński A (2017) Rate of convergence of the bundle method. *J. Optim. Theory Appl.* 173(3):908–922.
[8] Frangioni A (2002) Generalized bundle methods. *SIAM J. Optim.* 13(1):117–156.
[9] Kiwiel KC (1995) Proximal level bundle methods for convex nondifferentiable optimization, saddle-point problems and variational inequalities. *Math. Program.* 69(1–3):89–109.
[10] Kiwiel KC (2000) Efficiency of proximal bundle methods. *J. Optim. Theory Appl.* 104(3):589–603.
[11] Lan G (2012) An optimal method for stochastic composite optimization. *Math. Program.* 133(1):365–397.
[12] Lan G (2015) Bundle-level type methods uniformly optimal for smooth and nonsmooth convex optimization. *Math. Program.* 149(1–2):1–45.
[13] Lemaréchal C (1975) An extension of davidon methods to non differentiable problems. Balinski ML, Wolfe P, eds. *Nondifferentiable Optimization*, *Mathematical Programming Studies*, vol. 3 (Springer, Berlin, Heidelberg), 95–109.
[14] Lemaréchal C (1978) Nonsmooth optimization and descent methods. Research Report 78-4, IIASA, Laxenburg, Austria.
[15] Lemaréchal C, Nemirovski A, Nesterov Y (1995) New variants of bundle methods. *Math. Program.* 69(1–3):111–147.

Downloaded from informs.org by [2610:148:2002:e000:3248:5755:4741:1cbc] on 26 August 2023, at 10:53 . For personal use only, all rights reserved.

[16] Liang J, Monteiro RDC (2021) A proximal bundle variant with optimal iteration-complexity for a large range of prox stepsizes. *SIAM J. Optim.* 31(4):2955–2986.

[17] Mifflin R (1982) A modification and an extension of Lemaréchal's algorithm for nonsmooth minimization. Sorensen DC, Wets RJ-B, eds. *Nondifferential and Variational Techniques in Optimization*, *Mathematical Programming Studies*, vol. 17 (Springer Verlag, Berlin, Heidelberg), 77–90.

[18] Monteiro RDC, Svaiter BF (2012) Iteration-complexity of a Newton proximal extragradient method for monotone variational inequalities and inclusion problems. *SIAM J. Optim.* 22(3):914–935.

[19] Nesterov Y (2009) Primal-dual subgradient methods for convex problems. *Math. Program.* 120(1):221–259.

[20 Nesterov Y (2015) Universal gradient methods for convex optimization problems. *Math. Program.* 152(1):381–404.

[21] Ruszczyński A (2011) *Nonlinear Optimization* (Princeton University Press, Princeton, NJ).

[22] Urruty J-BH, Lemaréchal C (1993) *Convex Analysis and Minimization Algorithms I* (Springer-Verlag, Berlin, Heidelberg).

[23] Urruty J-BH, Lemaréchal C (1993) *Convex Analysis and Minimization Algorithms II.* (Springer-Verlag, Berlin, Heidelberg).

[24] van Ackooij W, Berge V, de Oliveira W, Sagastizábal C (2017) Probabilistic optimization via approximate p-efficient points and bundle methods. *Comput. Oper. Res.* 77:177–193.

[25] Wolfe P (1975) A method of conjugate subgradients for minimizing nondifferentiable functions. Balinski ML, Wolfe P, eds. *Nondifferentiable Optimization*, *Mathematical Programming Studies*, vol. 3 (Springer, Berlin, Heidelberg), 145–173.