SIAM J. OPTIM.
Vol. 31, No. 4, pp. 2955–2986

A PROXIMAL BUNDLE VARIANT WITH OPTIMAL ITERATION-COMPLEXITY FOR A LARGE RANGE OF PROX STEPSIZES*

JIAMING LIANG† AND RENATO D. C. MONTEIRO†

Abstract. This paper presents a proximal bundle variant, namely, the relaxed proximal bundle (RPB) method, for solving convex nonsmooth composite optimization problems. Like other proximal bundle variants, RPB solves a sequence of prox bundle subproblems whose objective functions are regularized composite cutting-plane models. Moreover, RPB uses a novel condition to decide whether to perform a serious or null iteration which does not necessarily yield a function value decrease. Optimal iteration-complexity bounds for RPB are established for a large range of prox stepsizes, in both convex and strongly convex settings. To the best of our knowledge, this is the first time that a proximal bundle variant is shown to be optimal for a large range of prox stepsizes. Finally, iteration-complexity results for RPB to obtain iterates satisfying practical termination criteria, rather than near optimal solutions, are also derived.

Key words. nonsmooth composite optimization, iteration-complexity, proximal bundle method, optimal complexity bound

AMS subject classifications. 49M37, 65K05, 68Q25, 90C25, 90C30, 90C60

DOI. 10.1137/20M1327513

1. Introduction.

The main goal of this paper is to present a proximal bundle variant, namely, the relaxed proximal bundle (RPB) method, whose iteration-complexity is optimal (possibly up to a logarithmic term), for a large range of prox stepsizes, in the context of convex nonsmooth composite optimization (CNCO) problems.

RPB is presented in the context of the CNCO problem

$$\phi^* := \min \left\{ \phi(x) := f(x) + h(x) : x \in \mathbb{R}^n \right\}, \tag{1.1}$$

where (i) $f, h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ are proper closed convex functions such that $\operatorname{dom} h \subseteq \operatorname{dom} f$; (ii) h is M_h-Lipschitz continuous and μ-convex on $\operatorname{dom} h$ for some $M_h \in [0, \infty]$ and $\mu \geq 0$; and (iii) a zeroth-order (resp., first-order) oracle, which for each $x \in \operatorname{dom} h$ returns $(f(x), h(x))$ (resp., $f'(x) \in \partial f(x)$ such that $\|f'(x)\| \leq M_f$), is available. Like other proximal bundle variants, the jth iteration of RPB considers the cutting-plane model

$$f_j(\cdot) = \max \left\{ f(x) + \langle f'(x), \cdot - x \rangle : \, x \in C_j \right\}, \tag{1.2}$$

where C_j is a suitable subset of the iterates $\{x_0, x_1, \ldots, x_{j-1}\}$ generated so far. RPB then solves the prox bundle subproblem

$$x_j := \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \underline{\phi}_j^\lambda(u) := f_j(u) + h(u) + \frac{1}{2\lambda} \|u - x_{j-1}^c\|^2 \right\} \tag{1.3}$$

*Received by the editors March 30, 2020; accepted for publication (in revised form) August 10, 2021; published electronically November 18, 2021.
https://doi.org/10.1137/20M1327513
Funding: The authors were partially supported by ONR grant N00014-18-1-2077. The first author was also partially supported by NSF grant CCF-1740776 through an ARC-TRIAD fellowship and an IDEaS-TRIAD research scholarship from Georgia Tech.
†School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205 USA (jliang76@gatech.edu, monteiro@isye.gatech.edu).

for x_j where λ is the prox stepsize (which for simplicity is assumed constant throughout the execution of RPB) and x^c_{j-1} is the prox-center. It is also assumed that a solver oracle that can exactly solve (1.3) is available. Complexity bounds described in this paper are relative to the number of RPB iterations performed, each consisting of two zeroth-order oracle calls (f and h), a subgradient call for f, and the resolution of the prox bundle subproblem (1.3).

Like many other proximal bundle methods, RPB performs two types of iterations, namely (i) serious ones during which the prox-centers are changed and (ii) null ones where the prox-centers are left unchanged. Moreover, RPB uses a novel condition to decide whether to perform a serious or null iteration which does not necessarily yield a function value decrease. A nice feature of our complexity analysis of RPB is that it considers a flexible bundle management policy (i.e., the way C_j is updated) which allows for some cuts to be removed but not aggregated (i.e., combined as a convex combination).

Contributions. This paper establishes an iteration-complexity bound for RPB with an arbitrary prox stepsize $\lambda > 0$ to obtain an $\bar{\varepsilon}$-solution of (1.1) (i.e., a point $\bar{x} \in \operatorname{dom} h$ satisfying $\phi(\bar{x}) - \phi^* \le \bar{\varepsilon}$). As a consequence, letting d_0 denote the distance of the initial point x_0 to the set of optimal solutions of (1.1), it is shown that the iteration-complexity of RPB is similar to that of the constant stepsize composite subgradient (CS-CS) method under either one of the following two cases:

(1) $\lambda \in [d_0/M_f, Cd_0^2/\bar{\varepsilon}]$ and $\mu \in [0, C'M_f/d_0]$;

(2) $\lambda \in [\bar{\varepsilon}/(CM_f^2), Cd_0^2/\bar{\varepsilon}]$, $M_h \le C'M_f$ and $\mu = 0$,

where C, C' are positive universal constants. It is worth noting that (a) case (1) allows μ to be zero and M_h to be arbitrary, but its λ-range is smaller than the one in case (2), and (b) case (2) covers all instances of (1.1) for which h is the indicator function of a closed convex set. Using these results, it is then argued that RPB has optimal iteration-complexity with respect to some important instance classes of (1.1).

Iteration-complexity results are also established for RPB to obtain iterates satisfying practical termination criteria rather than an $\bar{\varepsilon}$-solution. Another interesting conclusion of our analysis is that the CS-CS method can be viewed as a special instance of RPB as long as its prox stepsize λ is sufficiently small.

Related works. Some preliminary ideas toward the development of the proximal bundle method were first presented in [13, 31] and formal presentations of the method were given in [14, 18]. Convergence analysis of the proximal bundle method for CNCO problems has been broadly discussed in the literature and can be found, for example, in the textbooks [24, 27]. Different bundle management policies in the context of proximal bundle methods are discussed, for example, in [5, 6, 11, 22, 24, 28].

Previous iteration-complexity analysis of some proximal bundle variants can be found in [1, 5, 11]. More specifically, papers [1, 11] consider proximal bundle variants for the special case of the CNCO problem where h is the indicator function of a nonempty closed convex set (and hence $\mu = 0$). Paper [5] analyzes the complexity of the proximal bundle method considered in [11] under the condition that $h = 0$ and f is strongly convex. A detailed discussion of how the complexity bounds obtained in these papers compare to the ones obtained in this work is given in subsection 3.3 and the conclusion is that the bounds in [1, 5, 11] are generally much worse than the ones obtained in this work for most (in some cases, all) values of the prox stepsize λ.

Another method related, and developed subsequently, to the proximal bundle method is the bundle-level method, which was first proposed in [15] and extended in many ways in [3, 10, 12]. These methods have been shown to have optimal iteration-complexity in the setting of the CNCO problem with h being the indicator function of a compact convex set. Since their generated subproblems do not have a prox term, and hence do not use a prox stepsize, they are different from the ones studied in this paper.

Organization of the paper. Subsection 1.1 presents basic definitions and complexity theory notation used throughout the paper. Section 2 formally describes the assumptions on the CNCO problem (1.1), reviews the CS-CS method, and discusses its iteration-complexity. Subsections 3.1 and 3.2 present the RPB method and state the main results of the paper, namely, the general iteration-complexity for RPB and its implications in convex and strongly convex settings. Subsection 3.3 discusses, in the unconstrained CNCO context, results established for other proximal bundle variants in light of the ones obtained for RPB in this paper. Section 4 establishes a bound on the number of null iterations between two consecutive serious iterations and discusses the relationship between CS-CS and RPB. Section 5 provides the proof of the general iteration-complexity for RPB stated in section 3. Section 6 describes two alternative notions of approximate solutions for (1.1) and presents iteration-complexity results with respect to them. Section 7 reviews basic concepts from complexity theory, presents the lower complexity bound, and shows both CS-CS and RPB are optimal with respect to some instance classes of (1.1) introduced in this section. Section 8 presents some concluding remarks and possible extensions. Finally, Appendix A provides the proof of the iteration-complexity for the CS-CS method and Appendix B provides the proof of optimal complexity of the RPB method.

1.1. Basic definitions and notation.

The set of real numbers is denoted by $\mathbb{R}$. The set of nonnegative real numbers and the set of positive real numbers are denoted by $\mathbb{R}_+$ and $\mathbb{R}_{++}$, 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. Given a set $S \subset \mathbb{R}^n$, its linear (resp., convex) hull is denoted by $\mathrm{Lin}\, S$ (resp., $\mathrm{conv}\, S$). Let $\log(\cdot)$ denote the natural logarithm and define $\log_1^+(\cdot) := \max\{\log(\cdot), 1\}$.

Let $\psi : \mathbb{R}^n \to (-\infty, +\infty]$ be given. The effective domain of ψ is denoted by $\mathrm{dom}\, \psi := \{x \in \mathbb{R}^n : \psi(x) < \infty\}$ and ψ is proper if $\mathrm{dom}\, \psi \neq \emptyset$. Moreover, a proper function $\psi : \mathbb{R}^n \to (-\infty, +\infty]$ is μ-convex for some $\mu \geq 0$ if

$$\psi(\alpha z + (1-\alpha)u) \leq \alpha \psi(z) + (1-\alpha)\psi(u) - \frac{\alpha(1-\alpha)\mu}{2}\|z-u\|^2$$

for every $z, u \in \mathrm{dom}\, \psi$ and $\alpha \in [0,1]$. The set of all proper lower semicontinuous convex functions $\psi : \mathbb{R}^n \to (-\infty, +\infty]$ is denoted by $\overline{\mathrm{Conv}}\,(\mathbb{R}^n)$. For $\varepsilon \geq 0$, the *ε-subdifferential* of ψ at $z \in \mathrm{dom}\, \psi$ is denoted by

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

The subdifferential of ψ at $z \in \mathrm{dom}\, \psi$, denoted by $\partial \psi(z)$, is by definition the set $\partial_0 \psi(z)$.

Let constant $\bar{c} \in (1, \infty)$ and functions $p, q : \mathcal{Y} \to \mathbb{R}_+$ defined in an arbitrary set $\mathcal{Y}$ be given. We write $p(\cdot) = \mathcal{O}(q(\cdot))$ (with underlying constant $\bar{c}$) if $p(y) \leq \bar{c} q(y)$ for every $y \in \mathcal{Y}$. Finally, we write $p(\cdot) = \mathcal{O}_1(q(\cdot))$ if $p(\cdot) = \mathcal{O}(q(\cdot) + 1)$. It is worth emphasizing that the above $\mathcal{O}(\cdot)$ concept depends on the prespecified constant $\bar{c}$. Clearly, it follows from the above definition that if $p_i(y) = \mathcal{O}(q_i(y))$ with underlying constant $\bar{c}_i \in (1, \infty)$ for $i = 1, 2$, then $p_1(y)p_2(y) = \mathcal{O}(q_1(y)q_2(y))$ with underlying constant $\bar{c}_1 \bar{c}_2$.

2. Assumptions and the CS-CS method.

This section contains two subsections. The first one formally describes the assumptions made on the CNCO problem

(1.1). The second one presents the CS-CS method and the iteration-complexity of it for solving (1.1).

2.1. Assumptions. For some triple $(M_f, M_h, \mu) \in \mathbb{R}_+ \times [0,\infty] \times \mathbb{R}_+$, the following conditions on (1.1) are assumed to hold:

(A1) functions $f, h \in \overline{\text{Conv}}\,(\mathbb{R}^n)$ are such that $\text{dom}\, h \subset \text{dom}\, f$ and function f' : $\text{dom}\, h \to \mathbb{R}^n$ is such that $f'(x) \in \partial f(x)$ for all $x \in \text{dom}\, h$;
(A2) the set of optimal solutions X^* of problem (1.1) is nonempty;
(A3) h is μ-convex and $\|f'(x)\| \le M_f$ for all $x \in \text{dom}\, h$;
(A4) h is M_h-Lipschitz continuous on $\text{dom}\, h$, i.e.,

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

As already mentioned in section 1, in addition to the above assumptions, it is assumed that a zeroth-order oracle, which for each $x \in \text{dom}\, h$ returns $(f(x), h(x))$, and a solver oracle that can exactly solve (1.3), are also available. Complexity bounds developed throughout this paper are in terms of RPB iterations. Since each RPB iteration involves two zeroth-order oracle calls, one first-order oracle call, and one solver oracle call, they are also complexity bounds for the number of oracle calls.

We now make remarks about assumptions (A1)–(A4). First, function $f'(\cdot)$ should be viewed as an oracle which, for given $u \in \text{dom}\, h$, returns a subgradient of f at u whose magnitude is bounded by M_f. Second, it follows as a consequence of (A3) that

$$|f(u) - f(v)| \le M_f \|u - v\| \quad \forall u, v \in \text{dom}\, h. \tag{2.1}$$

Third, if $\mu > 0$ and $\text{dom}\, h$ is unbounded, then M_h cannot be finite. Fourth, if $u \in \text{dom}\, h$, (A4) does not imply that $\partial h(u)$ is bounded, even when M_h is finite. For example, an indicator function of a closed convex set satisfies (A4) but its subdifferential at a point in its relative boundary is unbounded.

For a given tolerance $\bar\varepsilon > 0$, a point x is called a $\bar\varepsilon$-solution of (1.1) if

$$\phi(x) - \phi^* \le \bar\varepsilon, \tag{2.2}$$

where ϕ^* is as in (1.1). Note that while (2.2) is theoretically appealing from a complexity point of view, it can rarely be used as a stopping criterion since ϕ^* is generally not known. Other more practical stopping criteria are discussed in section 6.

Throughout this paper, an instance of (1.1) means a quadruple $(x_0, (f, f'; h))$ where $x_0 \in \text{dom}\, h$ is the initial point and the triple $(f, f'; h)$ satisfies conditions (A1)–(A4) for some triple of parameters $(M_f, M_h, \mu) \in \mathbb{R}_+ \times [0, \infty] \times \mathbb{R}_+$. Moreover, for a given tolerance $\bar\varepsilon > 0$, it is said that an algorithm for solving an instance $(x_0, (f, f'; h))$ of (1.1) has $\bar\varepsilon$-iteration complexity $\mathcal{O}(T)$ if the total number of iterations it performs until it obtains an $\bar\varepsilon$-solution is bounded by $C'T$ where $C' > 0$ is a universal constant. In addition to $\bar\varepsilon$ and (M_f, M_h, μ), two other important quantities that are used to express complexity bounds associated with $(x_0, (f, f'; h))$ are

$$x_0^* := \text{argmin}\, \{\|x_0 - x^*\| : x^* \in X^*\}, \quad d_0 := \|x_0 - x_0^*\|. \tag{2.3}$$

2.2. Review of the CS-CS method. For a given scalar $\lambda > 0$ and instance $(x_0, (f, f'; h))$ of (1.1), the CS-CS method with initial point x_0 and constant prox stepsize λ, denoted by CS-CS(x_0, λ), recursively computes its iterates according to

$$x_j = \underset{u \in \mathbb{R}^n}{\text{argmin}} \left\{ f(x_{j-1}) + \langle f'(x_{j-1}), u - x_{j-1} \rangle + h(u) + \frac{1}{2\lambda} \|u - x_{j-1}\|^2 \right\}. \tag{2.4}$$

For any given universal constant $C > 1$, it follows from Proposition A.2 that CS-CS(x_0, λ) with any stepsize λ such that $\bar{\varepsilon}/(CM_f^2) \le 4\lambda \le \bar{\varepsilon}/M_f^2$ has $\bar{\varepsilon}$-iteration complexity bound given by

$$\mathcal{O}_1\left(\min\left\{\frac{M_f^2 d_0^2}{\bar{\varepsilon}^2}, \left(\frac{M_f^2}{\mu\bar{\varepsilon}}+1\right)\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}}+1\right)\right\}\right) \tag{2.5}$$

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 $\mu \downarrow 0$.)

We now make some remarks about bound (2.5). First, for the case in which $\mu = 0$ and under the extra assumption that $x_0 \in \mathrm{Argmin}\,\{h(x) : x \in \mathbb{R}^n\}$, it is well-known that the $\bar{\varepsilon}$-iteration complexity bound for CS-CS(x_0, λ) with $\bar{\varepsilon}/(CM_f^2) \le \lambda \le \bar{\varepsilon}/M_f^2$ for a universal constant $C > 1$ is as in (2.5) with $\mu = 0$ (e.g., see Theorem 9.26 of [2]). Hence, the result described in the previous paragraph generalizes the one in the previous sentence in that it removes the extra assumption above but requires changing the range on λ to $\bar{\varepsilon}/(CM_f^2) \le 4\lambda \le \bar{\varepsilon}/M_f^2$ for a universal constant $C > 1$. Second, it follows as a special case of the analysis in Chapter 3.2.3 of [21] that a certain variable stepsize projected subgradient method for the case in which $\mu = 0$ has $\bar{\varepsilon}$-iteration complexity bound for instances $(x_0, (f, f'; h))$ satisfying (A1)–(A3) and such that h is the indicator function of a closed convex set. In this regard, the result in the previous paragraph with $\mu = 0$ extends the result just mentioned to all instances satisfying (A1)–(A3) but replaces the variable stepsize projected subgradient method with CS-CS(x_0, λ) with $\bar{\varepsilon}/(CM_f^2) \le 4\lambda \le \bar{\varepsilon}/M_f^2$ for a universal constant $C > 1$.

3. The RPB method and main results.

This section contains three subsections. The first one describes the RPB method and discusses serious/null decision policies, storage requirements of RPB, and bundle management policies. The second one states a general $\bar{\varepsilon}$-iteration complexity bound for RPB and two consequences of the general bound in the convex and strongly convex settings. The third one derives $\bar{\varepsilon}$-iteration complexity bounds for another proximal bundle variant with respect to unconstrained CNCO instances and compares them with the ones obtained for RPB in subsection 3.2.

3.1. The RPB method.

We start by formally stating the RPB method. Its description below uses the cutting-plane model f_j defined in (1.2) and the availability of the subgradient oracle function $f'(\cdot)$ as in (A3). Note that the model f_j is used in the construction of subproblem (1.3) and is defined in terms of a finite set $C_j \subset \{x_0, x_1, \ldots, x_{j-1}\}$ which is updated according to step 2 below. Moreover, RPB is stated without a specific termination criterion with the intent of making it as flexible as possible. Subsection 3.2 (resp., section 6) then describes iteration-complexity bounds for it to obtain an $\bar{\varepsilon}$-solution (resp., other types of approximate solutions).

RPB

---

0. Let $x_0 \in \mathrm{dom}\, h$, $\lambda > 0$, and $\delta > 0$ be given, invoke the oracle $f'(\cdot)$ to obtain $f'(x_0) \in \partial f(x_0)$, and set $x_0^c = x_0$, $\tilde{x}_0 = x_0$, $\hat{z}_0 = x_0$, $C_1 = \{x_0\}$, $j = 1$, and $k = 1$;
1. Compute x_j according to (1.3), the function values $f(x_j)$, $h(x_j)$, and $f_j(x_j)$, and the optimal value $m_j := \underline{\phi}_j^\lambda(x_j)$ of subproblem (1.3), and invoke the oracle $f'(\cdot)$ to obtain $f'(x_j) \in \partial f(x_j)$. Moreover, consider the function ϕ_j^λ defined as

$$\phi_j^\lambda := \phi + \frac{1}{2\lambda}\|\cdot - x_{j-1}^c\|^2 \tag{3.1}$$

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

$$\tilde{x}_j \in \text{Argmin}\left\{\phi_j^\lambda(u) : u \in \{x_j, \tilde{x}_{j-1}\}\right\}; \tag{3.2}$$

2. **If**

$$t_j := \phi_j^\lambda(\tilde{x}_j) - m_j \le \delta, \tag{3.3}$$

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

$$\hat{z}_k \in \text{Argmin}\left\{\phi(u) : u \in \{\hat{z}_{k-1}, \tilde{x}_j\}\right\}; \tag{3.4}$$

if $\hat{z}_k$ satisfies the termination criterion, then **stop** and return $\hat{z}_k$; else, set $k \leftarrow k+1$, and go to step 3;

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

$$A_j \cup \{x_j\} \subset C_{j+1} \subset C_j \cup \{x_j\} \tag{3.5}$$

where

$$A_j := \{x \in C_j : f(x) + \langle f'(x), x_j - x\rangle = f_j(x_j)\} \tag{3.6}$$

and f_j is defined in (1.2); go to step 3;

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

We sometimes refer to RPB as RPB(x_0, λ, δ) whenever it is necessary to make its input (x_0, λ, δ) explicit. An iteration index j for which (3.3) is satisfied is called a serious one in which case x_j (resp., $\tilde{x}_j$) is called a serious iterate (resp., auxiliary serious iterate); otherwise, j is called a null iteration index. Moreover, we assume throughout our presentation that $j = 0$ is also a serious iteration index.

We now make some basic observations about RPB. First, the index j denotes the total iteration count and $k = k(j)$ equals the number of serious iteration indices (including 0) less than j. Second, for any $j \ge 1$, if ℓ_0 denotes the largest serious iteration index less than or equal to j, then $\tilde{x}_j$ is the best point (in terms of ϕ_j^λ) among the set $\{\tilde{x}_{\ell_0}, x_{\ell_0+1}, \ldots, x_j\}$. Third, the iterate $\hat{z}_k$ can be easily seen to be the best auxiliary serious iterate $\tilde{x}_j$ (in terms of ϕ) found up to and including the kth serious iteration. Fourth, the complexity results established in Theorems 3.1 and 6.4, and Corollary 6.5 below, are with respect to $\hat{z}_k$. This is in contrast to the iteration-complexity analysis of [5, 11], which establish complexity bounds with respect to the best (in terms of ϕ) serious iterate x_j (instead of $\tilde{x}_j$ as above) found so far. Fifth, the bundle set C_j consists of the set of points that are used to construct the cutting-plane model f_j which minorizes f. Sixth, A_j consists of the subset of points from C_j which are active at the most recent point x_j, i.e., the set of points which attains the maximum in (1.2).

We now provide some insights on how RPB can be viewed as an inexact proximal point method (see, for example, [7, 9, 17, 23, 25]) for solving (1.1). First, recall that each RPB iteration performs either a serious iteration (step 2.a) or a null one (step 2.b). Letting $(z_{k-1}, \tilde{z}_{k-1})$ denote the $(k-1)$th serious pair generated after x_0, it follows that the sequence of consecutive null pairs $\{(x_j, \tilde{x}_j) : j \in J_k\}$ obtained immediately after z_{k-1} together with the next serious pair $(z_k, \tilde{z}_k)$ can be viewed as an iterative procedure to compute a δ-solution of the proximal subproblem $\min\{\phi(u) + \|u - z_{k-1}\|^2/(2\lambda) : u \in \mathbb{R}^n\}$. Indeed, first note that this subproblem is equivalent to the problem $\min\{\phi_j^\lambda(u) : u \in \mathbb{R}^n\}$, where ϕ_j^λ is as in (3.1), due to the fact that $x_{j-1}^c = z_{k-1}$ for every index $j \in J_k$. Second, using the definition of t_j in (3.3) and the fact that $m_j \le m_j^* \le \phi_j^\lambda(\tilde{x}_j)$ where $m_j^* := \min\{\phi_j^\lambda(u) : u \in \mathbb{R}^n\}$ (see Lemma 4.1), we conclude that $\phi_j^\lambda(\tilde{x}_j) - m_j^* \le t_j$. This observation together with the role played by (3.3) in step 2 implies that $\tilde{z}_k$ is a δ-solution of the above proximal subproblem. Third, once such an approximate solution pair $(z_k, \tilde{z}_k)$ is obtained, the prox-center z_{k-1} of the above proximal subproblem is updated to z_k (see step 2.a) and this essentially corresponds to performing an inexact proximal step to problem (1.1). Section 5 and subsection 6.2 develop complexity bounds on the total number of proximal steps that can be performed as above, and hence on the number of serious iterations performed by RPB, until a prespecified termination criterion is satisfied.

We now discuss some serious/null decision policies that were used in other proximal bundle methods. First, the ones in references [4, 5, 11, 22, 24, 27, 30] all rely on the unified condition

$$\phi(x_{j-1}^c) - \phi(x_j) \ge \frac{\gamma}{1-\gamma}\left[f(x_j) - f_j(x_j) - \frac{\alpha_j}{2\lambda}\|x_j - x_{j-1}^c\|^2\right], \tag{3.7}$$

where $\alpha_j \in [0, 2]$ and $\gamma \in (0, 1)$. Under the assumption that $\alpha_j = 0$, the above condition together with the fact that $f \ge f_j$ (see Lemma 4.1) implies that $\phi(x_{j-1}^c) \ge \phi(x_j)$ and hence that $\phi(x_{j-1}^c) \ge \phi(x_j^c)$ in view of the way x_j^c is defined in step 2 of RPB. In view of the latter inequality, condition (3.7) with $\alpha_j = 0$ is referred to as the descent condition, and proximal bundle variants based on it have been studied in [5, 11, 24]. Moreover, the one with $\alpha_j \in (0, 2]$ can be viewed as a relaxation of the descent condition which does not necessarily imply monotonicity of $\{\phi(x_j^c)\}$ but guarantees the pointwise convergence of $\{x_j^c\}$ and $\{x_j\}$. Proximal bundle variants based on this relaxed condition have been studied in [4, 27] for $\alpha_j = 1$ and in [22, 30] for the more general case where $\alpha_j \in [0, 2]$. Second, paper [29] proposes a proximal bundle variant where the serious/null decision policies are not necessarily mutually exclusive. Third, as opposed to RPB and the algorithms in the above references which rely on serious/null decision policies, paper [1] proposes a proximal bundle variant which allows the next prox-center x_j^c to be a specific point in the line segment $[x_{j-1}^c, x_j]$. Finally, the selection rule (3.3) involving $\tilde{x}_j$ differs from the other aforementioned selection rules for x_j^c since they do not rely on $\tilde{x}_j$ (which generally differs from x_j and does not necessarily lie in $[x_{j-1}^c, x_j]$).

We now add a few remarks about the RPB storage requirement. First, at the beginning of each iteration of RPB, it is assumed that the following information is available: (1) the data $\{(x, f(x), f'(x)) : x \in C_j\}$ of the model (1.2) in order to solve (1.3) for x_j in step 1, and (2) the triple $(x_{j-1}^c, \tilde{x}_{j-1}, \hat{z}_{k-1})$ where $k = k(j)$ (see the first remark in the second paragraph following RPB for the definition of $k(j)$). Second, $\tilde{x}_j$ is updated in every iteration of RPB according to (3.2). Third, x_j^c and $\hat{z}_k$ change only during a serious iteration, and are updated as described in step 2.a. Hence, the size of the storage requirement of RPB is directly proportional to the cardinality of the bundle set C_j.

We end this subsection by briefly discussing bundle management policies in the context of proximal bundle methods that have been investigated in many works (see, for example, in [5, 6, 11, 22, 24, 28]). In the context of RPB, this means the way the bundle set C_j is updated in both steps 2.a and 2.b. The following two paragraphs specifically comment on these updates.

Consider first the (flexible) rule imposed on the next bundle set C_{j+1} relative to the current bundle set C_j in the execution of step 2.b when a null iteration happens. First, this rule has already been considered in [5, 6, 11, 24]. Second, it can be seen from (3.5) that $x_j \in C_{j+1}$ holds for every $j \geq 0$, and it is shown in Lemma 4.2(d) below that $x_j \notin C_j$ for every null iteration index j. This remark together with (3.5) then implies that, in every null iteration, one new point $x_j \notin C_j$ is added to C_{j+1}, while some of the points in $C_j \setminus A_j$ are possibly removed from it.

Consider next the rule imposed on the next bundle set C_{j+1} in the execution of step 2.a when a serious iteration happens. First, this rule has already been considered in [22]. Second, this rule, which requires C_{j+1} to satisfy $C_{j+1} \supset \{x_j\}$, allows for the possibility of completely refreshing the bundle set by setting it to $C_{j+1} = \{x_j\}$. Third, if C_{j+1} is chosen as $\{x_j\}$ at every serious iteration, then it follows from Theorem 3.1(b) below that the size of any bundle set C_j is always bounded by (3.9). Finally, since the prox bundle subproblem (1.3) generally becomes harder to solve as the size of the bundle set C_j grows, it might be convenient to choose C_{j+1} as lean as possible, i.e., $C_{j+1} = \{x_j\}$ if j is a serious iteration index and $C_{j+1} = A_j \cup \{x_j\}$ if j is a null iteration index.

3.2. A general $\bar{\varepsilon}$-iteration complexity bound for RPB. The following result, whose proof will be given at the end of section 5, presents, among other facts, an $\bar{\varepsilon}$-iteration complexity bound for RPB, which is a bound on the total number of (both serious and null) iterations performed by RPB until an $\bar{\varepsilon}$-solution is obtained. The iterate used to obtain such a solution is $\hat{z}_k$ which, as already mentioned in the second paragraph following RPB, is the best (in terms of ϕ) auxiliary serious iterate $\tilde{x}_j$ generated up to and including the kth serious iteration. The use of this iterate as a candidate to obtain an $\bar{\varepsilon}$-solution plays a fundamental role in the complexity analysis of RPB and clearly differs from the iteration-complexity analyses of [5, 11] which are based on the best (in terms of ϕ) serious iterate x_j (instead of $\tilde{x}_j$ as above) generated up to and including the kth serious iteration.

THEOREM 3.1. *Assume that* $(f, f'; h)$ *satisfies* (A1)–(A4) *for some* $(M_f, M_h, \mu) \in \mathbb{R}_+ \times [0, \infty] \times \mathbb{R}_+$. *Then, for any given* $(x_0, \lambda, \bar{\varepsilon}) \in \operatorname{dom} h \times \mathbb{R}_{++} \times \mathbb{R}_{++}$, *the following statements about* $RPB(x_0, \lambda, \delta)$ *with* $\delta = \bar{\varepsilon}/2$ *hold:*

(a) *the number of serious iterations performed until it obtains a best auxiliary serious iterate* $\hat{z}_k$ *such that* $\phi(\hat{z}_k) - \phi^* \leq \bar{\varepsilon}$ *is bounded by*

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

where

$$\tilde{\lambda} := \frac{\lambda}{1 + \lambda\mu}; \tag{3.8}$$

(b) *if* ℓ_0 *denotes an arbitrary serious iteration index and all the auxiliary serious iterates* $\hat{z}_k$ *generated up to and including the* ℓ_0*th iteration satisfy* $\phi(\hat{z}_k) - \phi^* > \bar{\varepsilon}$, *then the next serious iteration index* $\ell_1 > \ell_0$ *occurs and satisfies*

$$\ell_1 - \ell_0 \leq \frac{2(16)^{4/3} M_f \min\left\{\lambda M, 4\tilde{\lambda} M_f + \sqrt{2} d_0\right\}}{\bar{\varepsilon}} + 1; \tag{3.9}$$

(c) *the total number of iterations performed until it obtains an auxiliary serious iterate $\hat{z}_k$ such that $\phi(\hat{z}_k) - \phi^* \le \bar{\varepsilon}$ is bounded by*

$$\left(\frac{2(16)^{4/3} M_f \min\{\lambda M, 4\tilde{\lambda} M_f + \sqrt{2} d_0\}}{\bar{\varepsilon}} + 1\right)\left[\min\left\{\frac{d_0^2}{\lambda \bar{\varepsilon}}, \frac{1}{\tilde{\lambda}\mu} \log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\} + 1\right], \tag{3.10}$$

where ϕ^* and d_0 are as in (1.1) *and* (2.3)*, respectively, and $M = M_f + M_h$.*

We now make some comments about Theorem 3.1. First, the behavior of RPB clearly depends on the choice of the prox stepsize λ in its step 0. More specifically, as λ decreases, Theorem 3.1(a) implies that the total number of serious iteration indices increases, while Theorem 3.1(b) implies that bound (3.9) on the number of null iterations between any two consecutive serious iterations decreases. Second, in the unusual case where $CM_f d_0/\bar{\varepsilon} \le 1$ for a given universal constant $C > 0$, it can be easily seen that (3.10) reduces to $\mathcal{O}([\kappa + C^{-1} + 1][C^{-2}\kappa^{-1} + 1])$ where $\kappa := \lambda M_f^2/\bar{\varepsilon}$. Hence, the $\bar{\varepsilon}$-iteration complexity bound of RPB reduces to $\mathcal{O}((1 + C^{-1})^2)$ when λ is chosen as $\lambda = \bar{\varepsilon}/(CM_f^2)$.

We now make a few remarks about some of the input required by the CS-CS and RPB methods as well as the assumptions made to obtain iteration-complexity bounds for them. First, neither of the two methods requires the availability of a Lipschitz constant M_h as in (A4). Second, while the CS-CS method uses M_f as input, RPB has the advantage of not needing it. Third, iteration-complexity bounds for both of them have been established for any choice of initial point $x_0 \in \operatorname{dom} h$ and regardless of whether M_h is finite or not (see Theorem 3.1(c) and Proposition A.2). Fourth, complexity bound (2.5) for the CS-CS method and the one for RPB implied by (3.10), where $\min\{\lambda M, \tilde{\lambda} M_f + d_0\}$ is replaced by $\tilde{\lambda} M_f + d_0$, do not depend on M_h.

Under some reasonable conditions on the triple (μ, M_f, M_h), the next two results describe ranges on the prox stepsize λ which guarantee that the $\bar{\varepsilon}$-iteration complexity (3.10) of RPB reduces to that of the CS-CS method, namely (2.5). The first result covers the strongly convex case where μ is not too large and allows M_h to be arbitrary.

COROLLARY 3.2. *Let universal constants $C, C' > 0$ be given and consider an instance $(x_0, (f, f'; h))$ of* (1.1) *which satisfies* (A1)–(A4) *with parameter triple (M_f, M_h, μ) such that*

$$\frac{CM_f d_0}{\bar{\varepsilon}} \ge 1, \qquad M_h \in [0, +\infty], \qquad 0 \le \mu \le \frac{C' M_f}{d_0}. \tag{3.11}$$

Then, RPB$(x_0, \lambda, \bar{\varepsilon}/2)$ with any λ lying in the (nonempty) interval

$$\frac{d_0}{M_f} \le \lambda \le \frac{C d_0^2}{\bar{\varepsilon}} \tag{3.12}$$

has $\bar{\varepsilon}$-iteration complexity bound given by (2.5).

Proof. First, the conclusion that the set of λ satisfying (3.12) is nonempty follows directly from the first inequality in (3.11). Now, the assumption that $C' M_f/d_0 \ge \mu$, the first inequality in (3.12), and the definition of $\tilde{\lambda}$ in (3.8) imply that

$$(C' + 1)\frac{M_f}{d_0} \ge \mu + \frac{1}{\lambda} = \frac{1}{\tilde{\lambda}}$$

and hence that

$$(C' + 1)\tilde{\lambda} M_f \ge d_0. \tag{3.13}$$

Defining

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

and using Theorem 3.1(c) and (3.13), we conclude that $\mathcal{O}((a+1)(b+1))$ is an $\bar{\varepsilon}$-iteration complexity bound for RPB$(x_0, \lambda, \bar{\varepsilon}/2)$. Now, (3.13) and the first inequality in (3.11) can be easily seen to imply that $a \geq 1/[C(C'+1)]$. Moreover, it follows from the definition of b and the second inequality in (3.12) that

$$b \geq \min\left\{\frac{1}{C}, \frac{1}{\tilde{\lambda}\mu}\log\left(\frac{\lambda\mu}{C}+1\right)\right\}. \tag{3.14}$$

Using the fact that $\log(1+t) \geq t/(1+t)$ for every $t > 0$, we easily see that $\log(1+t) \geq t/2$ if $t \leq 1$ and $\log(1+t) \geq \log 2 > 0$ if $t \geq 1$. This observation with $t = \lambda\mu/C$ and the definition of $\tilde{\lambda}$ in (3.8) then imply that

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

and hence that $b \geq \min\{1/(2C), \log 2\}$. This inequality and the fact that $a \geq 1/[C(C'+1)]$ imply that $\mathcal{O}((a+1)(b+1))$ is equal to $\mathcal{O}(ab+1)$. Using this observation, the definitions of a and b, and the fact that $\tilde{\lambda} \leq \lambda$, we then conclude that the bound $\mathcal{O}((a+1)(b+1))$ reduces to (2.5) and hence that the lemma holds. ◻

We now make a few remarks about an instance $(x_0, (f, f'; h))$ which satisfies the assumptions of Corollary 3.2. First, if $M_h = +\infty$, then it is possible to have $\operatorname{dom} h = \mathbb{R}^n$. Actually, if $\operatorname{dom} h$ is unbounded and $\mu > 0$, then M_h must be $+\infty$.

The second result covers the convex case (i.e., $\mu = 0$) under the condition that M_h/M_f is bounded and shows that (3.10) reduces to (2.5) for a larger range of λ's.

Corollary 3.3. *Let universal constants $C, C' > 0$ be given and consider an instance $(x_0, (f, f'; h))$ of* (1.1) *which satisfies* (A1)-(A4) *with parameter triple* (μ, M_f, M_h) *such that*

$$\frac{CM_f d_0}{\bar{\varepsilon}} \geq 1, \qquad M_h \leq C' M_f, \qquad \mu = 0. \tag{3.15}$$

Then, RPB$(x_0, \lambda, \bar{\varepsilon}/2)$ with any λ lying in the (nonempty) interval

$$\frac{\bar{\varepsilon}}{CM_f^2} \leq \lambda \leq \frac{Cd_0^2}{\bar{\varepsilon}} \tag{3.16}$$

has $\bar{\varepsilon}$-iteration complexity bound $\mathcal{O}_1(M_f^2 d_0^2/\bar{\varepsilon}^2)$ and hence agrees with (2.5).

Proof. First, the conclusion that the set of λ satisfying (3.16) is nonempty follows immediately from the first inequality in (3.15). Moreover, it follows from the second inequality in (3.15) and Theorem 3.1(c) with $\mu = 0$ that the $\bar{\varepsilon}$-iteration complexity bound for RPB$(x_0, \lambda, \bar{\varepsilon}/2)$ is $\mathcal{O}([1 + \lambda M_f^2/\bar{\varepsilon}][1 + d_0^2/(\lambda\bar{\varepsilon})])$. Since (3.16) implies that $\max\{\lambda M_f^2/\bar{\varepsilon}, d_0^2/(\lambda\bar{\varepsilon})\} = \mathcal{O}(M_f^2 d_0^2/\bar{\varepsilon}^2)$, we then conclude that the previous bound reduces to $\mathcal{O}_1(M_f^2 d_0^2/\bar{\varepsilon}^2)$. ◻

We now make two remarks about an instance $(x_0, (f, f'; h))$ which satisfies the assumptions of Corollary 3.3. First, if h is an indicator function, then $M_h = 0$ and hence the second inequality in (3.15) is trivially satisfied. Second, if h is μ-convex with $\mu > 0$, then μ cannot be large. Indeed, it can be easily seen that $\mu \le 4M_h/D_h$, where D_h denotes the diameter of $\operatorname{dom} h$, and hence that D_h is finite.

We now make some remarks about Corollary 3.3 in light of Corollary 3.2. First, range (3.16) is larger than range (3.12) since they both have the same right endpoints and the left endpoint of the first one is the geometric mean of the endpoints of the latter one. Second, Corollary 3.2 also holds when $\mu = 0$ but, since it does not assume any condition on M_h, its conclusion is only guaranteed for a smaller range on λ.

We end this subsection by arguing that the first inequality in (3.11) or (3.15) is a mild assumption. Indeed, for those instances which violate this inequality, i.e., it satisfies $CM_f d_0/\bar\varepsilon \le 1$, the second remark in the paragraph following Theorem 3.1 implies that RPB with $\lambda = \bar\varepsilon/(CM_f^2)$ finds an $\bar\varepsilon$-solution of (1.1) in $\mathcal{O}((1 + C^{-1})^2)$ iterations. Hence, instances that do not satisfy this inequality are trivial.

3.3. Complexity bounds for other proximal bundle variants. Papers [5, 11] study a proximal bundle variant for solving the set constrained problem

$$\min\{\tilde f(x) : x \in X\}, \tag{3.17}$$

where X is a nonempty closed convex set[1] and $\tilde f$ is a μ-convex ($\mu \ge 0$) finite everywhere function such that a first-order oracle $\tilde f' : \mathbb{R}^n \to \mathbb{R}$ satisfying $\tilde f'(x) \in \partial \tilde f(x)$ for every $x \in \mathbb{R}^n$ is available. The method of [5, 11] starts from some $x_0 \in X$ and also uses a constant prox stepsize λ, and hence is referred to as $\text{PBV}(x_0, \lambda)$ below. If $\{x_j\}$ denotes the sequence of iterates generated by $\text{PBV}(x_0, \lambda)$ and

$$\tilde D = \tilde D[\tilde f] := \sup\{d(x_j, X^*) : j \ge 0\}, \quad \tilde M = \tilde M[\tilde f] := \sup\{\|\tilde f'(x_j)\| : j \ge 0\}, \tag{3.18}$$

then, under the assumption that the set X^* of optimal solutions of the above problem is nonempty, [11] shows that $\text{PBV}(x_0, \lambda)$ has $\bar\varepsilon$-iteration complexity bound

$$\mathcal{O}_1\left(\frac{\tilde M^2 \tilde D^4}{\lambda \bar\varepsilon^3}\right) \tag{3.19}$$

for the case in which $\mu = 0$, and [5] shows that $\text{PBV}(x_0, \lambda)$ has $\bar\varepsilon$-iteration complexity bound[2] given by

$$\mathcal{O}_1\left(\left[\frac{\tilde M^2 \lambda}{\alpha^2 \bar\varepsilon} \log_1^+\left(\frac{1}{\alpha^2}\right) + \frac{1}{\alpha}\right] \log_1^+\left(\frac{\tilde f(x_0) - \tilde f^*}{\alpha \bar\varepsilon}\right) + \frac{\tilde M^2 \lambda}{\alpha \bar\varepsilon} \log_1^+\left(\frac{\tilde M^2 \lambda}{\alpha \bar\varepsilon}\right)\right) \tag{3.20}$$

for the case $\mu > 0$, where $\alpha := \min\{\lambda\mu, 1\}$ and $\log_1^+(\cdot)$ is defined in subsection 1.1.

For the purpose of comparing the implication of the above bounds with the $\bar\varepsilon$-iteration complexity bounds established in Corollaries 3.2 and 3.3, we restrict our attention to the unconstrained CNCO problem (1.1), where f satisfies (A1)–(A3), $h \equiv \mu\|\cdot - x_0\|^2/2$, and x_0 is the initial point.

Clearly, such an unconstrained CNCO problem can be solved by applying $\text{PBV}(x_0, \lambda)$ to (3.17) with $\tilde f = f + h$ and $X = \mathbb{R}^n$ and with first-order oracle $\tilde f' := f' + \mu(\cdot - x_0)$. As a consequence, the $\bar\varepsilon$-iteration complexity bound of $\text{PBV}(x_0, \lambda)$ for solving the aforementioned unconstrained CNCO problem in the above manner is given by (3.19) with $\tilde f = f + h$ if $\mu = 0$ and (3.20) if $\mu > 0$.

[1] Actually, [5] only considers the case where $X = \mathbb{R}^n$.

[2] Actually, bound (3.20) has been formally derived in [8], which corrects a small error in the one derived in [5].

We will now derive $\bar\varepsilon$-iteration complexity bounds for PBV(x_0,λ) in terms of M_f, d_0, λ, and $\bar\varepsilon$ for any $\mu \ge 0$. We first claim that, for some constant $C'' > \sqrt{2}$ determined by the input of PBV(x_0,λ), we have

(a) $\tilde D \le \sup_{j\ge 0}\{\|x_j - x_0^*\|\} \le C''(d_0+\lambda\tilde M)$ where x_0^* is as in the line below (2.3);

(b) if $2C''\lambda\mu \le 1$, then $\tilde M \le 2[M_f + \mu(1+C'')d_0]$.

Indeed, (a) is proved in Lemma 4.1 of [8]. To prove (b), first note that the definition of $\tilde f'$, (2.3), the assumption that f satisfies (A3), and the triangle inequality imply that for every $j \ge 0$,

$$\|\tilde f'(x_j)\| \le \|f'(x_j)\| + \mu\|x_j - x_0\| \le M_f + \mu\|x_j - x_0\| \le M_f + \mu(d_0 + \|x_j - x_0^*\|),$$

and hence that $\tilde M \le M_f + \mu(d_0 + \sup_{j\ge 0}\|x_j - x_0^*\|)$, due to the definition of $\tilde M$ in (3.18). This inequality together with the second inequality in (a) implies that $\tilde M \le M_f + \mu d_0 + \mu C''(d_0 + \lambda \tilde M)$ and hence that (b) holds.

Now, using (3.19), (3.20), and statements (a) and (b) above, we conclude that PBV(x_0,λ) has $\bar\varepsilon$-iteration complexity bound

$$\mathcal{O}_1\left(\frac{M_f^2(d_0+\lambda M_f)^4}{\lambda\bar\varepsilon^3}\right) \tag{3.21}$$

if $\mu = 0$ and

$$\mathcal{O}_1\left(\left[\frac{M_f^2}{\lambda\mu^2\bar\varepsilon} + \frac{d_0^2}{\lambda\bar\varepsilon}\right]\log_1^+\left(\frac{1}{\lambda\mu}\right)\log_1^+\left(\frac{\tilde f(x_0)-\tilde f^*}{\lambda\mu\bar\varepsilon}\right)\right) \tag{3.22}$$

if $\mu > 0$ and $2C''\lambda\mu \le 1$. In summary, we have argued that bound (3.19) (resp., (3.20)) obtained in [11] (resp., [5]) yields the $\bar\varepsilon$-iteration complexity bound (3.21) (resp., (3.22)) if $\mu = 0$ (resp., if $\mu > 0$).

In the remaining part of this subsection, we compare the $\bar\varepsilon$-iteration complexity bounds (3.21) and (3.22) established for PBV(x_0,λ) and those for RPB$(x_0,\lambda,\bar\varepsilon/2)$ presented in Corollaries 3.3 and 3.2. We first discuss the case of bound (3.21) under the same assumption made in Corollary 3.3, i.e., the inequality $CM_fd_0/\bar\varepsilon \ge 1$ holds. Note that the arithmetic-geometric mean inequality implies that

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

and hence that (3.21) is minorized by $\mathcal{O}_1(M_f^3d_0^3/\bar\varepsilon^3)$, which in turn is minorized by $\mathcal{O}_1(M_f^2d_0^2/\bar\varepsilon^2)$ in view of the above assumption. Moreover, if $M_fd_0/\bar\varepsilon$ is significantly larger than 1, then it also follows from the above reasoning that, for any $\lambda > 0$, bound (3.21) is much worse than the $\bar\varepsilon$-iteration complexity bound $\mathcal{O}_1(M_f^2d_0^2/\bar\varepsilon^2)$ established in Corollary 3.3.

We now discuss the case of bound (3.22) under the assumptions of Corollary 3.2, i.e., the two inequalities $CM_fd_0/\bar\varepsilon \ge 1$ and $C'M_f/d_0 \ge \mu$ hold. Since (3.22) was proved under the condition that $2C''\lambda\mu \le 1$, we also assume that this condition holds in this paragraph. The assumption that $C'M_f/d_0 \ge \mu$ implies that (3.22) is equivalent to $\mathcal{O}_1(M_f^2/(\lambda\mu^2\bar\varepsilon))$. In view of (2.5) and the fact that $2C''\lambda\mu \le 1$, the latter bound is and can only be as good as the bound in Corollary 3.2 (i.e., (2.5)) when $\lambda\mu$ is bounded away from zero and μ is not too small. On the other hand, it follows from Corollary 3.2 that the $\bar\varepsilon$-iteration complexity of RPB$(x_0,\lambda,\bar\varepsilon/2)$ is given by (2.5) regardless of the sizes of the quantities $(\lambda\mu)^{-1}$ and μ^{-1} and this happens for a reasonably large λ-range which is independent of μ. Moreover, Corollary 3.2 does not assume the restrictive condition that $2C''\lambda\mu \le 1$.

Finally, [1] establishes an $\mathcal{O}_1(M_f^3 D^3/\bar\varepsilon^3)$ $\bar\varepsilon$-iteration complexity bound for an alternative proximal bundle variant, where D is the diameter of X. Clearly, this bound is much worse than the bound established in Corollary 3.3.

4. Analysis of null iterations. This section contains two subsections. The first one establishes a preliminary upper bound on the number of null iterations between two consecutive serious iterations. The second one discusses the relationship between CS-CS and RPB and presents a result showing that the former one can be viewed as a special instance of the latter one.

4.1. An upper bound on the number of consecutive null iterations. We assume throughout this subsection that ℓ_0 denotes an arbitrary serious iteration index (and hence it can be equal to zero) and $B(\ell_0)$ denotes the set consisting of the next serious iteration index ℓ_1 (if any) and all null iteration indices between ℓ_0 and ℓ_1, i.e., $B\left(\ell_0\right) = \{\ell_0+1,\ldots,\ell_1\}$.

We start by making some simple observations that immediately follow from the description of RPB. For any $j \in B(\ell_0)$, it follows from the definition of x_j^c in step 2 of RPB that $x_{j-1}^c = x_{\ell_0}$ and hence that

$$\phi_j^\lambda = \phi + \frac{1}{2\lambda}\|\cdot - x_{\ell_0}\|^2, \tag{4.1}$$

$$\underline{\phi}_j^\lambda = f_j + h + \frac{1}{2\lambda}\|\cdot - x_{\ell_0}\|^2, \tag{4.2}$$

in view of the definitions of $\underline{\phi}_j^\lambda$ and ϕ_j^λ in (1.3) and (3.1), respectively. Hence, it follows from the last identity and (1.3) that

$$x_j = \underset{u\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ f_j(u) + h(u) + \frac{1}{2\lambda}\|u - x_{\ell_0}\|^2 \right\} \quad \forall j \in B(\ell_0). \tag{4.3}$$

We now make a few immediate observations that will be used in the analysis of this subsection. First, it follows from the above equation that

$$\frac{1}{\lambda}(x_{\ell_0} - x_j) \in \partial\left(f_j + h\right)(x_j). \tag{4.4}$$

Second, since (4.1) implies that the function ϕ_j^λ remains the same whenever $j \in B(\ell_0)$ and ℓ_0 remains fixed throughout the analysis of this section, we will simply denote the function ϕ_j^λ for $j \in B(\ell_0)$ by ϕ^λ, i.e.,

$$\phi^\lambda = \phi_j^\lambda \quad \forall j \in B(\ell_0). \tag{4.5}$$

Third, in view of the definition of $\tilde{x}_j$ in (3.2) (see also the second remark in the second paragraph following RPB) and the above relation, it then follows that

$$\tilde{x}_j \in \operatorname{Argmin}\left\{\phi^\lambda(x) : x \in \{\tilde{x}_{\ell_0}, x_{\ell_0+1},\ldots,x_j\}\right\}. \tag{4.6}$$

Fourth, it directly follows from (4.3) and (4.6) that $\{x_j, \tilde{x}_j\} \subset \operatorname{dom} h$. Fifth, ℓ_1 is characterized as the first index $j > \ell_0$ satisfying condition (3.3). Sixth, it will be shown below that the sequence $\{t_j : j \in B(\ell_0)\}$, where t_j is defined in (3.3), is nonincreasing (see Lemma 4.5(b)) and converges to zero with an $\mathcal{O}(1/j)$ convergence rate (see Proposition 4.7).

The following result describes some basic facts about the prox subproblem (4.7) and the prox bundle subproblem (1.3).

LEMMA 4.1. *For every* $j \in B(\ell_0)$*, define*

$$m_j^* := \min\left\{\phi^\lambda(u) : u \in \mathbb{R}^n\right\}, \tag{4.7}$$

where ϕ^λ *is as in* (4.5)*. Then, for every* $j \in B(\ell_0)$ *and* $u \in \operatorname{dom} h$*, we have*

$$f(u) \geq f_j(u), \quad \phi^\lambda(u) \geq \underline{\phi}_j^\lambda(u), \quad \phi^\lambda(u) \geq m_j^* \geq m_j. \tag{4.8}$$

As a consequence, $t_j \geq \phi^\lambda(\tilde{x}_j) - m_j^* \geq 0$*, where* t_j *is as in* (3.3)*.*

Proof. It follows from the definition of f_j in (1.2) and (A1) that the first inequality in (4.8) holds. This inequality, and relations (4.1), (4.2), and (4.5), imply that the second inequality in (4.8) holds. It follows from the definition of m_j^* in (4.7) that $\phi^\lambda(u) \geq m_j^*$ for every $u \in \operatorname{dom} h$. Using the second inequality in (4.8), and the definitions of m_j and m_j^* in step 1 in RPB and (4.7), respectively, we have $m_j^* \geq m_j$. Moreover, it follows from the fact that $\{\tilde{x}_j\} \subset \operatorname{dom} h$ (see the fourth remark below (4.6)), the last two inequalities in (4.8) with $u = \tilde{x}_j$, and the definition of t_j in (3.3) that $t_j \geq \phi^\lambda(\tilde{x}_j) - m_j^* \geq 0$. ∎

The following technical result provides basic properties of RPB that are used in our analysis.

LEMMA 4.2. *The following statements about RPB hold for every* $j \in B(\ell_0)$*:*

(a) *for every* $x \in C_j$*, we have* $f(x) = f_j(x)$*;*
(b) *for every* $i \in B(\ell_0)$ *such that* $i < j$*, we have* $\phi^\lambda(\tilde{x}_j) \leq \phi^\lambda(\tilde{x}_i)$*;*
(c) $t_j \leq f(x_j) - f_j(x_j) \leq 2M_f\|x_j - x_{j-1}\|$*;*
(d) *if* $x_j \in C_j$*, then* $t_j = 0$ *and* j *coincides with* ℓ_1 *(i.e., the only serious iteration index in* $B(\ell_0)$*);*
(e) f_j *is* M_f*-Lipschitz continuous on* $\operatorname{dom} h$*.*

Proof. (a) Let $x \in C_j$ be given. Using the first inequality in (4.8), the assumption that $x \in C_j$, and the definition of f_j in (1.2), we conclude $f \geq f_j \geq f(x) + \langle f'(x), \cdot - x\rangle$ and hence that $f(x) \geq f_j(x) \geq f(x) + \langle f'(x), x - x\rangle = f(x)$. Thus, (a) follows.

(b) This statement follows immediately from (4.6).

(c) Using the definitions of t_j and m_j in (3.3) and step 1 of RPB, respectively, relations (4.2), (4.5), and (4.6), and the fact that $\phi = f + h$, we have

$$t_j = \phi^\lambda(\tilde{x}_j) - m_j \leq \phi^\lambda(x_j) - \underline{\phi}_j^\lambda(x_j) = f(x_j) - f_j(x_j),$$

and hence the first inequality in the statement holds. Next we show the second inequality in the statement. It follows from (3.5) with $j = j - 1$ that $x_{j-1} \in C_j$. This inclusion and the definition of f_j in (1.2) imply that

$$f_j(\cdot) \geq f(x_{j-1}) + \langle f'(x_{j-1}), \cdot - x_{j-1}\rangle$$

and hence that

$$\begin{aligned} f(x_j) - f_j(x_j) &\leq f(x_j) - [f(x_{j-1}) + \langle f'(x_{j-1}), x_j - x_{j-1}\rangle] \\ &\leq |f(x_j) - f(x_{j-1})| + \|f'(x_{j-1})\|\|x_j - x_{j-1}\|, \end{aligned} \tag{4.9}$$

where the second inequality is due to the triangle and Cauchy–Schwarz inequalities. The second inequality in the statement now follows from (A3), (2.1), the fact that $\{x_j\} \subset \operatorname{dom} h$ (see the fourth remark below (4.6)), and inequality (4.9).

(d) Assume that $x_j \in C_j$. It then follows from statement (a) with $x = x_j$ and the first inequality in statement (c) that $t_j \le 0$. In view of Lemma 4.1, we then conclude that $t_j = 0$. In view of step 2 of RPB, this implies that j is a serious iteration index. Thus, since ℓ_1 is the only serious iteration index in $B(\ell_0)$, we must have $j = \ell_1$.

(e) It follows from (A3), the definition of f_j in (1.2), and a well-known formula for the subdifferential of the pointwise maximum of finitely many affine functions (e.g., see Example 3.4 of [26]) that f_j is M_f-Lipschitz continuous on $\operatorname{dom} h$. ◻

The following result gives a few useful properties about the relationship between the active sets $\{A_j : j \in B(\ell_0)\}$ and the iterates $\{x_j : j \in B(\ell_0)\}$.

Lemma 4.3. *Define*

$$f_{A_j}(\cdot) := \max\{f(x) + \langle f'(x), \cdot - x\rangle : x \in A_j\} \quad \forall j \in B(\ell_0), \tag{4.10}$$

where A_j is as in (3.6). *Then, the following statements hold for every $j \in B(\ell_0)$:*

(a) $(f_{A_j} + h)(x_j) = (f_j + h)(x_j)$ *and* $\partial(f_{A_j} + h)(x_j) = \partial(f_j + h)(x_j)$;

(b) $f_{A_j} \le \min\{f_j, f_{j+1}\}$;

(c) *we have*

$$\begin{aligned} x_j &= \operatorname*{argmin}_{u\in\mathbb{R}^n} \left\{ (f_{A_j} + h)(u) + \frac{1}{2\lambda}\|u - x_{\ell_0}\|^2 \right\}, \\ m_j &= \min_{u\in\mathbb{R}^n} \left\{ (f_{A_j} + h)(u) + \frac{1}{2\lambda}\|u - x_{\ell_0}\|^2 \right\}, \end{aligned} \tag{4.11}$$

where m_j is as in step 1 *of RPB;*

(d) *for every $u \in \mathbb{R}^n$, we have*

$$(f_{A_j} + h)(u) + \frac{1}{2\lambda}\|u - x_{\ell_0}\|^2 \ge m_j + \frac{1}{2\tilde\lambda}\|u - x_j\|^2,$$

where $\tilde\lambda$ is as in (3.8).

Proof. (a) The first conclusion immediately follows from the definitions of A_j and f_{A_j} in (3.6) and (4.10), respectively. Using the definition of A_j in (3.6), the definition of f_j in (1.2), and a well-known formula for the subdifferential of the pointwise maximum of finitely many convex functions (e.g., see Corollary 4.3.2 of [26]), we conclude that $\partial f_j(x_j)$ is the convex hull of $\cup\{f'(x) : x \in A_j\}$. Using the same reasoning but with (1.2) replaced by (4.10), we conclude that the latter set is also the subdifferential of f_{A_j} at x_j. Hence, statement (a) follows.

(b) Note that $A_j \subset C_j$ due to (3.6). Also, it follows from rule (3.5) regarding the choice of C_{j+1} that $A_j \subset C_{j+1}$. Hence, the definitions of f_j and f_{A_j} in (1.2) and (4.10), respectively, imply that $f_{j+1} \ge f_{A_j}$ and $f_j \ge f_{A_j}$. Thus, (b) holds.

(c) It follows from (4.4) and the second identity in (a) that

$$\frac{1}{\lambda}(x_{\ell_0} - x_j) \in \partial(f_j + h)(x_j) = \partial(f_{A_j} + h)(x_j).$$

Using the definition of m_j in step 1 of RPB, (4.2), the first identity in (a), and the fact that the above inclusion implies that x_j satisfies the optimality condition of (4.11), we conclude that (c) holds.

(d) This statement follows immediately from (c), the definition of $\tilde\lambda$ in (3.8), the fact that the objective function of (4.11) is $(\mu + 1/\lambda)$-strongly convex, and Theorem 5.25(b) of [2] with $f = f_{A_j} + h + \|\cdot - x_{\ell_0}\|^2/(2\lambda)$, $x^* = x_j$, and $\sigma = \mu + 1/\lambda$. ◻

The following lemma provides a bound on $\|x_j - x_{\ell_0}\|$ for $j \in B(\ell_0)$.

Lemma 4.4. *Let $M = M_f + M_h$ and define $d_{\ell_0} := \|x_{\ell_0} - x_0^*\|$, where x_0^* is as in the line below* (2.3)*. Then, the following statements hold:*

(a) $\|x_j - x_{\ell_0}\| \le 2\lambda M$ *for every* $j \in B(\ell_0)$*;*

(b) *if* $j \in B(\ell_0)$ *is such that the bundle set* C_j *contains* x_{ℓ_0}*, then* $\|x_j - x_{\ell_0}\| \le 2d_{\ell_0} + 8\tilde\lambda M_f$*;*

(c) $\|x_{\ell_0+1} - x_{\ell_0}\| \le 2\min\{\lambda M, d_{\ell_0} + 4\tilde\lambda M_f\}$.

Proof. (a) Using Lemma 4.3(b) and (d), and the definitions of m_j and in step 1 of RPB and (3.8), respectively, we conclude that for every $u \in \operatorname{dom} h$,

$$\frac{1}{2\tilde\lambda}\|u - x_j\|^2 + (f_j + h)(x_j) + \frac{1}{2\lambda}\|x_j - x_{\ell_0}\|^2 \le (f_j + h)(u) + \frac{1}{2\lambda}\|u - x_{\ell_0}\|^2, \tag{4.12}$$

which upon setting $u = x_{\ell_0}$ yields

$$\frac{1}{2\lambda}\|x_j - x_{\ell_0}\|^2 \le (f_j + h)(x_{\ell_0}) - (f_j + h)(x_j) \le M\|x_{\ell_0} - x_j\|,$$

where the last inequality is due to Lemma 4.2(e) and (A4). Hence, (a) follows.

(b) It follows from (4.12) with $u = x_0^*$, the fact that $(f_j + h)(x_0^*) \le \phi(x_0^*) = \phi^* \le \phi(x_j)$, and the definition of d_{ℓ_0} that

$$\frac{1}{2\tilde\lambda}\|x_0^* - x_j\|^2 \le \frac{1}{2\tilde\lambda}\|x_0^* - x_j\|^2 + \phi(x_j) - \phi^* \le \frac{d_{\ell_0}^2}{2\lambda} + f(x_j) - f_j(x_j) - \frac{1}{2\lambda}\|x_j - x_{\ell_0}\|^2.$$

Using the assumption that $x_{\ell_0} \in C_j$ and an argument similar to one in the proof of Lemma 4.2(c), we can see that $f(x_j) - f_j(x_j) \le 2M_f\|x_j - x_{\ell_0}\|$. This conclusion, the above inequality, and the fact that $\tilde\lambda \le \lambda$ (see (3.8)) imply that

$$\|x_0^* - x_j\|^2 \le d_{\ell_0}^2 + 4\tilde\lambda M_f\|x_j - x_{\ell_0}\|. \tag{4.13}$$

It follows from the above inequality, the definition of d_{ℓ_0}, and the Cauchy–Schwarz inequality that

$$\|x_j - x_{\ell_0}\|^2 \le 2\|x_j - x_0^*\|^2 + 2\|x_0^* - x_{\ell_0}\|^2 \le 4d_{\ell_0}^2 + 8\tilde\lambda M_f\|x_j - x_{\ell_0}\|.$$

Solving the above quadratic inequality in terms of $\|x_j - x_{\ell_0}\|$, we conclude (b) holds.

(c) It is easy to see that (3.5) with $j = \ell_0$ implies that $x_{\ell_0} \in C_{\ell_0+1}$. The conclusion of (c) now follows from (a) and (b) with $j = \ell_0 + 1$. ∎

We now make some remarks about Lemma 4.4. First, while the bound in (a) is meaningless when $M_h = \infty$, the one in (b) is finite but it requires the mild condition that C_j contain x_{ℓ_0}. Second, the results in (a) and (b) can be used in conjunction with (5.5) to show that the whole RPB sequence $\{x_j : j \ge 0\}$ is bounded. Third, the complexity analysis of RPB does not make use of the last observation but only of the fact stated in Lemma 4.4(c).

The following lemma presents a few technical results about the set of scalars $\{t_j : j \in B(\ell_0)\}$ and plays an important role in the estimation of the cardinality of the set $B(\ell_0)$.

Lemma 4.5. *Consider the sequence $\{t_j\}$ as in* (3.3) *and the sequences $\{m_j\}$ and $\{x_j\}$ as in step* 1 *of RPB. Then, the following statements hold:*

(a) *for every* $i, j \in B(\ell_0)$ *such that* $i < j$, *we have*

$$t_i \geq m_j - m_i \geq \frac{1}{2\tilde{\lambda}} \sum_{l=i+1}^{j} \|x_l - x_{l-1}\|^2 \,; \tag{4.14}$$

(b) $\{t_j : j \in B(\ell_0)\}$ *is nonincreasing;*

(c) $t_j \leq 4M_f \min\{\lambda M, d_{\ell_0} + 4\tilde{\lambda}M_f\}$ *for every* $j \in B(\ell_0)$.

Proof. (a) It follows from the last two inequalities in (4.8) with $u = \tilde{x}_j$ and Lemma 4.2(b) that

$$m_j \leq \phi^\lambda(\tilde{x}_j) \leq \phi^\lambda(\tilde{x}_i)$$

and hence that the first inequality in (4.14) holds in view of the definition of t_i in (3.3). Using the definition of m_{j+1} in step 1 of RPB, (4.2), and statements (b) and (d) with $u = x_{j+1}$ of Lemma 4.3, we conclude that

$$\begin{aligned} m_{j+1} &= (f_{j+1} + h)(x_{j+1}) + \frac{1}{2\lambda}\|x_{j+1} - x_{\ell_0}\|^2 \\ &\geq (f_{A_j} + h)(x_{j+1}) + \frac{1}{2\lambda}\|x_{j+1} - x_{\ell_0}\|^2 \geq m_j + \frac{1}{2\tilde{\lambda}}\|x_{j+1} - x_j\|^2. \end{aligned}$$

The second inequality in (4.14) now follows by adding the above inequality from $j = i$ to $j = j - 1$ and simplifying the resulting inequality.

(b) It immediately follows from (4.14) that $\{m_j\}$ is nondecreasing, which together with Lemma 4.2(b) and the definition of t_j in (3.3) implies that $\{t_j\}$ is nonincreasing.

(c) It follows from Lemma 4.2(c) with $j = \ell_0 + 1$ and Lemma 4.4(c) that

$$t_{\ell_0+1} \leq 4M_f \min\{\lambda M, d_{\ell_0} + 4\tilde{\lambda}M_f\}.$$

The statement now follows from (b). ∎

The following technical result relates t_j and the minimum distance Δ_j between two consecutive iterates among $\{x_{\ell_0}, \ldots, x_j\}$, a quantity that plays an important role in the complexity analysis of the null iterations.

LEMMA 4.6. *Let*

$$\Delta_j := \min\{\|x_i - x_{i-1}\| : i \in B(\ell_0),\, i \leq j\} \quad \forall j \in B(\ell_0). \tag{4.15}$$

Then, the following statements hold:

(a) *for every* $j \in B(\ell_0)$, *we have* $t_j \leq 2M_f\Delta_j$*;*

(b) *for every* $j \in B(\ell_0)$ *such* $j \geq \ell_0 + 4$, *we have*

$$\Delta_j^2 \leq \frac{32\tilde{\lambda}M_f}{(j-\ell_0)^2}\sqrt{2\tilde{\lambda}\lceil (j-\ell_0)/2\rceil t_{\ell_0+\lfloor (j-\ell_0)/2\rfloor -1}},$$

where $\tilde{\lambda}$ *is as in* (3.8).

Proof. (a) Let $j \in B(\ell_0)$ and an arbitrary $i \in B(\ell_0)$ such that $i \leq j$ be given. Using Lemmas 4.5(b) and 4.2(c) with $j = i$, we conclude that

$$t_j \leq t_i \leq 2M_f\|x_i - x_{i-1}\|.$$

The statement now follows from the definition of Δ_j in (4.15) and the fact that the above inequality holds for every $i \in B(\ell_0)$ such that $i \leq j$.

(b) Let $j \in B(\ell_0)$ such that $j \geq \ell_0 + 4$ be given. For any $i \in B(\ell_0)$ such that $i < j$, it follows from Lemmas 4.5(a), Lemma 4.2(c) with $j = i$, and the definition of Δ_j in (4.15) that

$$\frac{1}{2\tilde{\lambda}}(j-i)\Delta_j^2 \leq \frac{1}{2\tilde{\lambda}}\sum_{l=i+1}^{j}\|x_l - x_{l-1}\|^2 \leq t_i \leq 2M_f\|x_i - x_{i-1}\|.$$

Since the set of indices $\mathcal{I} := \{\ell_0 + \lfloor (j-\ell_0)/2\rfloor, \ldots, j-1\}$ is clearly in $\{i \in B(\ell_0) : i < j\}$ and $|\mathcal{I}| = \lceil (j-\ell_0)/2\rceil$, we conclude by adding the above inequality as i varies in $\mathcal{I}$ that

$$\frac{(j-\ell_0)^2}{16\tilde{\lambda}}\Delta_j^2 \leq \frac{\lceil (j-\ell_0)/2\rceil(\lceil (j-\ell_0)/2\rceil + 1)}{4\tilde{\lambda}}\Delta_j^2 \leq 2M_f\sum_{i\in\mathcal{I}}\|x_i - x_{i-1}\|. \tag{4.16}$$

On the other hand, using the fact that $j \geq \ell_0 + 4$ implies that $\ell_0 + \lfloor (j-\ell_0)/2\rfloor - 1 \geq \ell_0 + 1$, the Cauchy–Schwarz inequality, and Lemma 4.5(a) with $(i,j) = (\ell_0 + \lfloor (j - \ell_0)/2\rfloor - 1, j-1)$, we conclude that

$$\begin{aligned}\sum_{i\in\mathcal{I}}\|x_i - x_{i-1}\| &\leq \left\lceil \frac{j-\ell_0}{2}\right\rceil^{1/2}\left(\sum_{i\in\mathcal{I}}\|x_i - x_{i-1}\|^2\right)^{1/2}\\ &\leq \sqrt{2\tilde{\lambda}\left\lceil \frac{j-\ell_0}{2}\right\rceil t_{\ell_0+\lfloor (j-\ell_0)/2\rfloor - 1}}.\end{aligned}$$

Statement (b) now follows by plugging the above inequality into (4.16) and rearranging the resulting inequality. ∎

The following proposition shows that the sequence $\{t_j : j \in B(\ell_0)\}$ converges to zero with an $\mathcal{O}(1/j)$ convergence rate.

PROPOSITION 4.7. *For every $j \in B(\ell_0)$, we have*

$$t_j \leq \frac{(16)^{4/3}M_f \min\left\{\lambda M, 4\tilde{\lambda}M_f + d_{\ell_0}\right\}}{j-\ell_0}. \tag{4.17}$$

Proof. The proof of the proposition is by induction on $j \in B(\ell_0)$. First note that (4.17) holds for every $j \in B(\ell_0)$ such that $j \leq \ell_0 + 5$ in view of Lemma 4.5(c). Now, let $j \in B(\ell_0)$ be such that $j \geq \ell_0 + 6$ and assume for the induction argument that (4.17) holds for the indices $\ell_0 + 1, \ldots, j-1$. Also, define $a := (16)^{4/3}M_f \min\{\lambda M, 4\tilde{\lambda}M_f + d_{\ell_0}\}$. Since $\ell_0 + 1 \leq \ell_0 + \lfloor (j-\ell_0)/2\rfloor - 1 \leq j-1$ when $j \geq \ell_0 + 6$, we conclude that

$$\lceil (j-\ell_0)/2\rceil\, t_{\ell_0+\lfloor (j-\ell_0)/2\rfloor - 1} \leq \frac{\lceil (j-\ell_0)/2\rceil}{\lfloor (j-\ell_0)/2\rfloor - 1}a \leq 2a,$$

where the last inequality is due to the assumption that $j \geq \ell_0 + 6$ and the definition of a. The last conclusion together with Lemma 4.6(b) then implies that

$$\Delta_j^2 \leq \frac{64\tilde{\lambda}M_f\sqrt{\tilde{\lambda}a}}{(j-\ell_0)^2}.$$

Now, using Lemma 4.6(a) and the last inequality, we then conclude that

$$t_j \leq 2M_f\Delta_j \leq \frac{16M_f^{3/2}\tilde{\lambda}^{3/4}a^{1/4}}{j-\ell_0} \leq \frac{a}{j-\ell_0},$$

where the last inequality follows from the definition of a and the fact that $\lambda \ge \tilde{\lambda}$ (see (3.8)). We have thus shown that the conclusion of the proposition holds. ∎

We are now ready to state the main result of this subsection.

PROPOSITION 4.8. *Let $(x_0, \lambda, \delta) \in \operatorname{dom} h \times \mathbb{R}_{++} \times \mathbb{R}_{++}$ be given and assume that* (A1)–(A4) *hold and $j = \ell_0$ is a serious iteration index of RPB(x_0, λ, δ). Then, the next serious iteration index $j = \ell_1 > \ell_0$ exists and satisfies*

$$\ell_1 - \ell_0 \le \frac{(16)^{4/3} M_f \min\left\{\lambda M, 4\tilde{\lambda} M_f + d_{\ell_0}\right\}}{\delta} + 1,$$

where M_f is as in (A3)*, and M and d_{ℓ_0} are as in Lemma* 4.4.

Proof. If $\ell_1 = \ell_0 + 1$, then (3.9) is obviously true. Assume then $\ell_1 > \ell_0 + 1$. This clearly implies that $\ell_1 - 1 \in B(\ell_0)$ and hence is a null iteration index. Using this observation and the fact that an iteration index j is null if and only if (3.3) does not hold, we conclude that $t_{\ell_1 - 1} > \delta$. This conclusion, the fact that $\ell_1 - 1 \in B(\ell_0)$, and Proposition 4.7 with $j = \ell_1 - 1$ then imply that

$$\delta < t_{\ell_1 - 1} \le \frac{(16)^{4/3} M_f \min\left\{\lambda M, 4\tilde{\lambda} M_f + d_{\ell_0}\right\}}{\ell_1 - 1 - \ell_0},$$

from which the conclusion of the proposition immediately follows. ∎

4.2. Relationship between RPB and CS-CS. We start by making a few trivial remarks about the relationship between CS-CS and RPB. First, if they use the same stepsize λ, then they both generate the same first iterate x_1. Second, if $d_0 \le \bar{\varepsilon}/(4M_f)$, then it follows from Proposition A.2 that the CS-CS method, and hence RPB, with $\lambda = \bar{\varepsilon}/(4M_f^2)$ finds a $\bar{\varepsilon}$-solution of (1.1) in one iteration.

The following result describes a less trivial relationship between RPB and the CS-CS method. More specifically, it shows that the first remark in the previous paragraph can be extended to the other iterates as well as long as λ is sufficiently small.

PROPOSITION 4.9. *Let $(x_0, \lambda, \delta) \in \operatorname{dom} h \times \mathbb{R}_{++} \times \mathbb{R}_{++}$ satisfying $\lambda \le \delta/(2MM_f)$ (and hence $M_h < \infty$) be given, where M is as in Lemma* 4.4*. Then, every iteration index of RPB(x_0, λ, δ) is a serious one. As a consequence, if the set C_{j+1}, which necessarily contains x_j, is always set to be $\{x_j\}$ in step* 2.a *of RPB, then RPB(x_0, λ, δ) reduces to CS-CS(x_0, λ).*

Proof. Using Lemma 4.5(c) and the assumption that $\lambda \le \delta/(2MM_f)$, we have $t_j \le 2\lambda M M_f \le \delta$ for every $j \in B(\ell_0)$. Hence, we have $t_{\ell_0+1} \le \delta$, and in view of (3.3), we conclude that every iteration index j is serious. We now show that under the assumption of the proposition, RPB(x_0, λ, δ) reduces to CS-CS(x_0, λ). Since every iteration index is a serious one, using step 2.a of RPB, the definition of f_j in (1.2), and the assumption of this proposition that $C_{j+1} = \{x_j\}$, we conclude that $x_j^c = x_j$ and $f_j(\cdot) = f(x_{j-1}) + \langle f'(x_{j-1}), \cdot - x_{j-1}\rangle$ for every $j \ge 1$. In view of this observation and (2.4), it is now easy to see that RPB(x_0, λ, δ) reduces to CS-CS(x_0, λ). ∎

5. Proof of Theorem 3.1. This section provides the proof of Theorem 3.1, which describes a general iteration-complexity for RPB to find an $\bar{\varepsilon}$-solution of (1.1).

We start by introducing some notation and definitions. Consider the sequences $\{f_j\}$, $\{x_j\}$, and $\{\tilde{x}_j\}$ as in (1.2), (1.3), and (3.2), respectively, and let $\{j_k : k \ge 0\}$ denote the sequence of serious iteration indices generated by RPB (and hence $j_0 = 0$). Moreover, define $z_0 := x_0$, $\tilde{z}_0 := x_0$, and, for every $k \ge 1$,

$$z_k := x_{j_k}, \quad \tilde{z}_k := \tilde{x}_{j_k}, \quad \tilde{f}_k := f_{j_k}. \tag{5.1}$$

Using the definitions of $\hat{z}_k$ and $\tilde{z}_k$ in (3.4) and (5.1), respectively, we have

$$\hat{z}_k \in \text{Argmin}\,\{\phi(z) : z \in \{\tilde{z}_0, \tilde{z}_1, \ldots, \tilde{z}_k\}\} \quad \forall k \geq 1. \tag{5.2}$$

The following lemma provides some technical results that will be used in the proof of Theorem 3.1.

LEMMA 5.1. *The following statements about RPB(x_0, λ, δ) hold for every $k \geq 1$:*
(a) $z_{k-1} = x_j^c$ *for every* $j = j_{k-1}, \ldots, j_k - 1$;
(b) $z_k = \text{argmin}\,\{(\tilde{f}_k + h)(u) + \|u - z_{k-1}\|^2/(2\lambda) : u \in \mathbb{R}^n\}$;
(c) $\delta_k + \|\tilde{z}_k - z_{k-1}\|^2/(2\lambda) \leq \delta$, *where* $\delta_k := \phi(\tilde{z}_k) - (\tilde{f}_k + h)(z_k) - \|z_k - z_{k-1}\|^2/(2\lambda)$;
(d) $\phi(\tilde{z}_k) - \phi(z) + (1 + \lambda\mu)\|z_k - z\|^2/(2\lambda) \leq \delta_k + \|z_{k-1} - z\|^2/(2\lambda)$ *for every* $z \in \text{dom}\, h$;
(e) $\|\tilde{z}_k - z_k\|^2 \leq 2\lambda\delta$.

Proof. (a) This statement follows from the definition of z_k in (5.1) and the prox-center update policy in step 2 of RPB.

(b) This statement follows from (1.3) with $j = j_k$, (5.1), and (a).

(c) Using the fact that $m_j = \underline{\phi}_j^\lambda(x_j)$ (see step 1) with $j = j_k$ and (a), we have

$$m_{j_k} = (\tilde{f}_k + h)(z_k) + \frac{1}{2\lambda}\|z_k - z_{k-1}\|^2.$$

Relation (3.1) with $j = j_k$, (5.1), and (a) imply that

$$\phi_{j_k}^\lambda(\tilde{x}_{j_k}) = \phi(\tilde{z}_k) + \frac{1}{2\lambda}\|\tilde{z}_k - z_{k-1}\|^2.$$

Since j_k is a serious iteration index, (3.3) holds with $j = j_k$. Using this conclusion, the above two identities, and the definition of δ_k in the statement, we conclude that

$$\delta_k + \frac{1}{2\lambda}\|\tilde{z}_k - z_{k-1}\|^2 = \phi_{j_k}^\lambda(\tilde{x}_{j_k}) - m_{j_k} \leq \delta.$$

(d) Noting that the objective function in (b) is $(\mu + 1/\lambda)$-strongly convex, and using (b) and Theorem 5.25(b) of [2] with $f = \tilde{f}_k + h + \|\cdot - z_{k-1}\|^2/(2\lambda)$, $x^* = z_k$, and $\sigma = \mu + 1/\lambda$, we have for every $k \geq 1$ and $z \in \mathbb{R}^n$,

$$(\tilde{f}_k + h)(z_k) + \frac{1}{2\lambda}\|z_k - z_{k-1}\|^2 \leq (\tilde{f}_k + h)(z) + \frac{1}{2\lambda}\|z - z_{k-1}\|^2 - \frac{1}{2}\left(\mu + \frac{1}{\lambda}\right)\|z - z_k\|^2.$$

The above inequality, the fact that $\phi \geq \tilde{f}_k + h$, and the definition of δ_k in (c) imply

$$\begin{aligned}\phi(\tilde{z}_k) - \phi(z) + \frac{1}{2}\left(\mu + \frac{1}{\lambda}\right)\|z_k - z\|^2 &\leq \phi(\tilde{z}_k) - (\tilde{f}_k + h)(z) + \frac{1}{2}\left(\mu + \frac{1}{\lambda}\right)\|z_k - z\|^2 \\ &\leq \phi(\tilde{z}_k) - (\tilde{f}_k + h)(z_k) - \frac{1}{2\lambda}\|z_k - z_{k-1}\|^2 + \frac{1}{2\lambda}\|z_{k-1} - z\|^2 \\ &= \delta_k + \frac{1}{2\lambda}\|z_{k-1} - z\|^2.\end{aligned}$$

(e) This statement follows from (d) with $z = \tilde{z}_k$ and (c). ∎

We are now ready to prove Theorem 3.1.

Proof of Theorem 3.1. Recall that Theorem 3.1 deals with RPB(x_0, λ, δ) with $\delta = \bar{\varepsilon}/2$. Using Lemma 5.1(d) with $z = x_0^*$, and the fact that $\delta_k \le \delta$ (see Lemma 5.1(c)), we have

$$\phi(\tilde{z}_k) - \phi^* \le \frac{1}{2\lambda}\|z_{k-1} - x_0^*\|^2 - \frac{1+\lambda\mu}{2\lambda}\|z_k - x_0^*\|^2 + \delta \quad \forall k \ge 1. \tag{5.3}$$

Since (5.3) satisfies (A.1) with $\eta_k = \phi(\tilde{z}_k) - \phi^*$, $\alpha_k = \|z_k - x_0^*\|^2/(2\lambda)$, $\theta = 1 + \lambda\mu$, and $\delta = \bar{\varepsilon}/2$, it follows from Lemma A.1, the fact that $\alpha_0 = d_0^2/(2\lambda)$, and relation (5.2) that every $k \ge 1$ such that $\phi(\hat{z}_k) - \phi^* > \bar{\varepsilon}$ satisfies

$$k < \min\left\{\frac{d_0^2}{\lambda\bar{\varepsilon}}, \frac{1+\lambda\mu}{\lambda\mu}\log\left(\frac{\mu d_0^2}{\bar{\varepsilon}} + 1\right)\right\} \tag{5.4}$$

and

$$\|z_k - x_0^*\| \le \sqrt{d_0^2 + 2\lambda k\delta} = \sqrt{d_0^2 + \lambda k\bar{\varepsilon}} \le \sqrt{2}d_0, \tag{5.5}$$

where the identity is due to the fact that $\delta = \bar{\varepsilon}/2$ and the last inequality is due to (5.4). Clearly, the first conclusion above (i.e., (5.4)) and the definition of $\tilde{\lambda}$ in (3.8) imply (a). Moreover, the second one (i.e., (5.5)) together with the first identity in (5.1) and Proposition 4.8 with $\delta = \bar{\varepsilon}/2$ implies that (b) holds. Finally, (c) follows immediately from (a) and (b). ∎

6. Complexity results for other termination criteria. This section contains two subsections. The first one describes two alternative notions of approximate solutions for problem (1.1). The second one states iteration-complexity results with respect to these approximate solutions. For simplicity, we assume in this section that $\mu = 0$ and M_h is finite.

6.1. Other termination criteria. Usually, algorithms for solving (1.1) naturally generate pairs (x, η) satisfying the inclusion $0 \in \partial_\eta \phi(x)$ or, equivalently, the inequality $\phi(x) - \phi^* \le \eta$, in all of their iterations (see the discussion in the second and third paragraphs following Definition 6.1 below). For the purpose of our discussion in this section, we refer to such a pair (x, η) as a ϕ-compatible pair. Moreover, a ϕ-compatible pair (x, η) is called an $\bar{\varepsilon}$-solution pair of (1.1) if its residual η satisfies $\eta \le \bar{\varepsilon}$. We now make a few remarks about a given ϕ-compatible pair (x, η). First, if $\eta \le \bar{\varepsilon}$, then x is an $\bar{\varepsilon}$-solution. Second, checking whether $\eta \le \bar{\varepsilon}$ is satisfied is much simpler than checking whether (2.2) holds. Third, it is possible for (x, η) to satisfy the inequalities (2.2) and $\eta > \bar{\varepsilon}$, which means that x is already a desired $\bar{\varepsilon}$-solution but the certificate (or residual) η is not suitable to detect this fact.

More generally, the following definition of an approximate solution triple of (1.1) will be useful.

Definition 6.1. *A triple (x, v, η) is called ϕ-compatible if it satisfies the inclusion $v \in \partial_\eta \phi(x)$. For a given tolerance pair $(\hat{\rho}, \hat{\varepsilon})$, a ϕ-compatible triple (x, v, η) is called a $(\hat{\rho}, \hat{\varepsilon})$-solution triple of* (1.1) *if it satisfies $\|v\| \le \hat{\rho}$ and $\eta \le \hat{\varepsilon}$.*

At this point, it is interesting to illustrate the notion of a ϕ-compatible triple in the specific setting of (1.1) where $h(\cdot) = I_K(\cdot)$ and K is a nonempty closed convex cone. In such a setting, (x, v, η) is ϕ-compatible if and only if there exists $s \in \partial f(x)$ such that $s - v \in K^*$ and $\langle x, s - v\rangle \le \eta$ (see Lemma 3.3 in [19]). Clearly, when

$v = 0$ and $\eta = 0$, the latter condition implies that x is an optimal solution of (1.1). In general, v is a perturbation made on s to obtain a dual feasible point $s - v \in K^*$ and η is an upper bound on the complementarity gap of the primal-dual feasible pair $(x, s - v)$ (see Proposition 3.4 in [19]). This specific setting shows that the two residuals v and η have their own natural meanings. This same phenomenon can also be observed in the context of other constrained convex optimization problems and monotone variational inequalities (see, for example, [19, 20]).

We now make some comments about the use of the above definition as a natural algorithmic stopping criterion. Many algorithms, including the one considered in this paper, are able to naturally generate a sequence of ϕ-compatible triples $\{(\hat z_k, \hat v_k, \hat\varepsilon_k)\}$ for which the residual pair $(\hat v_k, \hat\varepsilon_k)$ can be made arbitrarily small (see, for example, Proposition 6.3 below). As a consequence, some $(\hat z_k, \hat v_k, \hat\varepsilon_k)$ will eventually become a $(\hat\rho, \hat\varepsilon)$-solution triple of (1.1) and verifying this simply amounts to checking whether the two inequalities $\|\hat v_k\| \le \hat\rho$ and $\hat\varepsilon_k \le \hat\varepsilon$ hold.

It is natural to wonder whether these same algorithms can also produce a sequence as above but with $\hat v_k = 0$ for every $k \ge 0$. It turns out that when $\operatorname{dom} h$ is unbounded, such a sequence is generally difficult or impossible to obtain. However, when $\operatorname{dom} h$ is bounded, we can construct such a sequence using the one as in the previous paragraph. Indeed, let S be a compact convex set containing $\operatorname{dom} h$ and, for every k, define

$$\hat\eta_k := \hat\varepsilon_k + \sup\{\langle \hat v_k, \hat z_k - x\rangle : x \in S\}. \tag{6.1}$$

Then, using the assumption that $(\hat z_k, \hat v_k, \hat\varepsilon_k)$ is a ϕ-compatible triple, the definition of ε-subdifferential in subsection 1.1, and the above definition of $\hat\eta_k$, we conclude that

$$\phi(x) \ge \phi(\hat z_k) + \langle \hat v_k, x - \hat z_k\rangle - \hat\varepsilon_k \ge \phi(\hat z_k) - \hat\eta_k \quad \forall x \in \operatorname{dom} h$$

or, equivalently, $0 \in \partial\phi_{\hat\eta_k}(\hat z_k)$. Hence, $\{(\hat z_k, 0, \hat\eta_k)\}$ is a sequence of ϕ-compatible triples with $\hat v_k = 0$ for every k or, equivalently, $\{(\hat z_k, \hat\eta_k)\}$ is a sequence of ϕ-compatible pairs. Moreover, using (6.1) and the assumptions that S is bounded and $(\hat v_k, \hat\varepsilon_k)$ can be made arbitrarily small, we easily see that $\hat\eta_k$ can also be made arbitrarily small. Observe that this implies that, for any given tolerance $\bar\varepsilon > 0$, an index k will eventually be generated such that $(\hat z_k, \hat\eta_k)$ is an $\bar\varepsilon$-solution pair, and detecting the latter property simply amounts to checking whether the inequality $\hat\eta_k \le \bar\varepsilon$ holds.

6.2. Iteration-complexity results. The following lemma states some bounds on the magnitude of the sequences $\{z_k\}$ and $\{\hat z_k\}$ which are used in establishing the iteration-complexity for RPB to obtain a $(\hat\rho, \hat\varepsilon)$-solution triple.

LEMMA 6.2. *For every $k \ge 1$, we have*

$$\|z_k - z_0\| \le \sqrt{2k\lambda\delta} + 2d_0, \tag{6.2}$$

$$\|\hat z_k - z_0\|^2 \le 2\lambda\delta + 5\sqrt{k}\lambda\delta + 3k\lambda\delta + \frac{15 d_0^2}{2}, \tag{6.3}$$

where $\hat z_k$, d_0, and z_k are as in (3.4)*,* (2.3)*, and* (5.1)*, respectively, and δ is as in step 0 of RPB.*

Proof. Using the first inequality in (5.5), the triangle inequality, and the facts that $d_0 = \|z_0 - x_0^*\|$ and $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ for every $a, b \in \mathbb{R}_+$, we have

$$\|z_k - z_0\| \le \|z_k - x_0^*\| + \|z_0 - x_0^*\| \le \sqrt{2k\lambda\delta} + 2d_0,$$

and hence (6.2) holds. Using the fact that $(\sum_{i=1}^n a_i)^2 \le (\sum_{i=1}^n s_i)(\sum_{i=1}^n a_i^2/s_i)$ for every $(a_1,\ldots,a_n) \in \mathbb{R}^n$ and $(s_1,\ldots,s_n) \in \mathbb{R}^n_{++}$, the triangle inequality, the first inequality in (5.5), and Lemma 5.1(e), we conclude that for every $k \ge 1$,

$$\begin{aligned}
\|\tilde z_k - z_0\|^2 &\le (\|\tilde z_k - z_k\| + \|z_k - x_0^*\| + \|x_0^* - z_0\|)^2 \\
&\le \left(\frac{1}{\sqrt{k}} + 1 + \frac{1}{2}\right)\left(\sqrt{k}\|\tilde z_k - z_k\|^2 + \|z_k - x_0^*\|^2 + 2\|x_0^* - z_0\|^2\right) \\
&\le \left(\frac{1}{\sqrt{k}} + \frac{3}{2}\right)\left(2\sqrt{k}\lambda\delta + 2k\lambda\delta + 3d_0^2\right) \\
&= 2\lambda\delta + 5\sqrt{k}\lambda\delta + 3k\lambda\delta + \frac{3d_0^2}{\sqrt{k}} + \frac{9d_0^2}{2}.
\end{aligned}$$

Since (5.2) implies that there exists $i \in \{0,1,\ldots,k\}$ such that $\hat z_k = \tilde z_i$, the above inequality with $k = i$ then implies that

$$\|\hat z_k - z_0\|^2 = \|\tilde z_i - z_0\|^2 \le 2\lambda\delta + 5\sqrt{i}\lambda\delta + 3i\lambda\delta + \frac{15d_0^2}{2},$$

from which (6.3) immediately follows due to the fact that $i \le k$. ◻

We now make a remark about the above result. Bound (6.3) and its proof can be significantly simplified at the expense of obtaining a bound whose constant multiplying the term $k\lambda\delta$ is not as tight as its current value, namely 3. The current value is the best we could obtain and, as we will see from the second inequality for $\hat\varepsilon_k$ in (6.5), the smaller this constant is, the closer δ can be chosen to the tolerance $\hat\varepsilon$.

The following two results establish the iteration-complexity for RPB to find a $(\hat\rho,\hat\varepsilon)$-solution triple (see Definition 6.1). The first of these two results describes the convergence rate of a certain sequence of triples $\{(\hat z_k, \hat v_k, \hat\varepsilon_k)\}$ generated by RPB.

PROPOSITION 6.3. *Define*

$$\hat v_k := \frac{z_0 - z_k}{\lambda k}, \quad \hat\varepsilon_k := \frac{1}{k}\sum_{i=1}^k \delta_i + \frac{\|\hat z_k - z_0\|^2 - \|\hat z_k - z_k\|^2}{2\lambda k} \quad \forall k \ge 1, \tag{6.4}$$

where λ is as in step 0 of RPB. Then, the following statements hold for every $k \ge 1$:

(a) $\hat v_k \in \partial_{\hat\varepsilon_k}\phi(\hat z_k)$;

(b) *the residual pair $(\hat v_k, \hat\varepsilon_k)$ is bounded by*

$$\|\hat v_k\| \le \frac{2d_0}{\lambda k} + \frac{\sqrt{2\delta}}{\sqrt{\lambda k}}, \qquad 0 \le \hat\varepsilon_k \le \frac{5\delta}{2}\left(1 + \frac{1}{\sqrt{k}} + \frac{2}{5k}\right) + \frac{15d_0^2}{4\lambda k}, \tag{6.5}$$

where d_0 is as in (2.3) and δ is as in step 0 of RPB.

Proof. (a) It follows from Lemma 5.1(d) that

$$\phi(\tilde z_k) - \phi(z) \le \delta_k + \frac{1}{2\lambda}\left(\|z_{k-1} - z\|^2 - \|z_k - z\|^2\right).$$

Summing the above inequality from $k = 1$ to $k = k$ and using (5.2), we have

$$\phi(\hat z_k) - \phi(z) \le \frac{1}{k}\sum_{i=1}^k \delta_i + \frac{1}{2\lambda k}(\|z_0 - z\|^2 - \|z_k - z\|^2).$$

This inequality, the obvious identity

$$\|z-z_0\|^2 - \|z-z_k\|^2 = \|\hat{z}_k - z_0\|^2 - \|\hat{z}_k - z_k\|^2 + 2\langle z_0 - z_k, \hat{z}_k - z\rangle \quad \forall z \in \mathbb{R}^n,$$

and the definitions of $\hat{v}_k$ and $\hat{\varepsilon}_k$ in (6.4) imply that for every $z \in \operatorname{dom} h$,

$$\begin{aligned}\phi(\hat{z}_k) - \phi(z) &\le \frac{1}{k}\sum_{i=i}^{k}\delta_i + \frac{1}{2\lambda k}\left(\|\hat{z}_k - z_0\|^2 - \|\hat{z}_k - z_k\|^2 + 2\langle z_0 - z_k, \hat{z}_k - z\rangle\right)\\ &= \hat{\varepsilon}_k + \langle \hat{v}_k, \hat{z}_k - z\rangle, \end{aligned} \tag{6.6}$$

from which we conclude that (a) holds due to the definition of ε-subdifferential.

(b) The first inequality in (6.5) follows by plugging (6.2) into the definition of $\hat{v}_k$ in (6.4). The first inequality for $\hat{\varepsilon}_k$, i.e., $\hat{\varepsilon}_k \ge 0$, follows from (6.6) with $z = \hat{z}_k$. Using the fact that $\delta_k \le \delta$ (see Lemma 5.1(c)), the definition of $\hat{\varepsilon}_k$ in (6.4), and relation (6.3), we have

$$\hat{\varepsilon}_k \le \frac{1}{k}\sum_{i=1}^{k}\delta_i + \frac{\|\hat{z}_k - z_0\|^2}{2\lambda k} \le \delta + \frac{1}{2\lambda k}\left(2\lambda\delta + 5\sqrt{k}\lambda\delta + 3k\lambda\delta + \frac{15 d_0^2}{2}\right),$$

from which the second inequality for $\hat{\varepsilon}_k$ immediately follows. ∎

We now make some remarks about the above result. First, Proposition 6.3(a) shows that RPB naturally generates a sequence $\{(\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)\}$ of ϕ-compatible triples. Second, Proposition 6.3(b) implies that the sequence $\{\hat{\varepsilon}_k\}$ can be made arbitrarily small, say $\hat{\varepsilon}_k \le \hat{\varepsilon}$, for sufficiently large k, as long as δ is chosen in $(0, 2\hat{\varepsilon}/5)$. Third, the two previous remarks ensure that RPB is able to generate a $(\hat{\rho}, \hat{\varepsilon})$-solution triple $(\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)$. Fourth, the three previous remarks in turn show that RPB is able to generate a sequence $\{(\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)\}$ satisfying the properties outlined in the second paragraph following Definition 6.1.

We are now ready to describe the iteration-complexity for RPB to find a $(\hat{\rho}, \hat{\varepsilon})$-solution triple of (1.1).

THEOREM 6.4. *For a given tolerance pair $(\hat{\rho}, \hat{\varepsilon}) \in \mathbb{R}^2_{++}$, the following statements about the RPB method hold with $\delta = \hat{\varepsilon}/3$:*

(a) *the number of serious iterations performed until it obtains a $(\hat{\rho}, \hat{\varepsilon})$-solution triple $(\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)$ is bounded by*

$$\mathcal{O}_1\left(\max\left\{\frac{\hat{\varepsilon}}{\lambda\hat{\rho}^2}, \frac{d_0^2}{\lambda\hat{\varepsilon}}\right\}\right);$$

(b) *the total number of iterations performed until it obtains a $(\hat{\rho}, \hat{\varepsilon})$-solution triple $(\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)$ is bounded by*

$$\mathcal{O}_1\left(\max\left\{\frac{MM_f}{\hat{\rho}^2}, \frac{MM_f d_0^2}{\hat{\varepsilon}^2}\right\} + \max\left\{\frac{\hat{\varepsilon}}{\lambda\hat{\rho}^2}, \frac{d_0^2}{\lambda\hat{\varepsilon}}\right\} + \frac{\lambda M M_f}{\hat{\varepsilon}}\right), \tag{6.7}$$

where λ and δ are two of the inputs to RPB (see step 0), d_0 is as in (2.3), $M = M_f + M_h$, and M_f and M_h are as in (A3) and (A4), respectively.

Proof. (a) It follows from Proposition 6.3(a) and Definition 6.1 that $(\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)$ is a ϕ-compatible triple for every $k \ge 1$. Moreover, the first inequality in (6.5) and the fact that $\delta = \hat{\varepsilon}/3$ imply that for every $k \ge \max\{4d_0/(\lambda\hat{\rho}), 8\hat{\varepsilon}/(3\lambda\hat{\rho}^2)\}$,

$$\|\hat{v}_k\| \le \frac{2d_0}{\lambda k} + \frac{\sqrt{2\delta}}{\sqrt{\lambda k}} \le \frac{\hat{\rho}}{2} + \frac{\hat{\rho}}{2} = \hat{\rho},$$

and the second inequality for $\hat{\varepsilon}_k$ in (6.5) and the fact that $\delta = \hat{\varepsilon}/3$ imply that for every $k \geq \max\{405d_0^2/(2\lambda\hat{\varepsilon}), 36\}$,

$$\hat{\varepsilon}_k \leq \frac{5\delta}{2}\left(1 + \frac{1}{\sqrt{k}} + \frac{2}{5k}\right) + \frac{15d_0^2}{4\lambda k} \leq \frac{5\hat{\varepsilon}}{6}\left(1 + \frac{1}{6} + \frac{1}{90}\right) + \frac{\hat{\varepsilon}}{54} = \hat{\varepsilon}.$$

The above two observations then imply that $(\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)$ must satisfy the two inequalities in Definition 6.1 with $(v, \eta) = (\hat{v}_k, \hat{\varepsilon}_k)$, and hence that $(\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)$ is a $(\hat{\rho}, \hat{\varepsilon})$-solution triple (see Definition 6.1), for every index k satisfying

$$k \geq \max\left\{\frac{4d_0}{\lambda\hat{\rho}}, \frac{8\hat{\varepsilon}}{3\lambda\hat{\rho}^2}, \frac{405d_0^2}{2\lambda\hat{\varepsilon}}, 36\right\}.$$

The complexity bound in (a) now follows from the last conclusion and the inequality $2\sqrt{ab} \leq a + b$ with $a = \hat{\varepsilon}/(\lambda\hat{\rho}^2)$ and $b = d_0^2/(\lambda\hat{\varepsilon})$.

(b) This statement immediately follows from (a), Proposition 4.8 with $\delta = \hat{\varepsilon}/3$, and the assumption that M_h is finite in the beginning of this section. ∎

The following result describes the iteration-complexity for RPB to find an $\bar{\varepsilon}$-solution pair $(x, \eta) = (\hat{z}_k, \hat{\eta}_k)$ for the case in which $\operatorname{dom} h$ is bounded. (Recall the definition of an $\bar{\varepsilon}$-solution pair is given in the first paragraph of subsection 6.1.) Observe that the major difference between the result below and Theorem 3.1 is that the one below provides a certificate $\eta = \hat{\eta}_k$ of the $\bar{\varepsilon}$-optimality of $x = \hat{z}_k$ while Theorem 3.1 does not. Although it is possible to derive a complexity bound for any value of λ with little extra effort, the result below assumes for simplicity that λ lies in a certain range and obtains a simpler iteration-complexity bound under this assumption.

COROLLARY 6.5. *Assume that $S \subset \mathbb{R}^n$ is a compact convex set containing* $\operatorname{dom} h$ *and let $\bar{\varepsilon} > 0$ be a given tolerance. Consider RPB with inner tolerance $\delta = \bar{\varepsilon}/6$ and prox stepsize λ satisfying $\bar{\varepsilon}/(CMM_f) \leq \lambda \leq CD_S^2/\bar{\varepsilon}$, where $C > 0$ is a universal constant, $M = M_f + M_h$, M_f, and M_h are as in* (A3) *and* (A4)*, respectively, and $D_S := \sup\{\|u - u'\| : u, u' \in S\}$, and let $\{(\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)\}$ and $\{\hat{\eta}_k\}$ denote the sequences obtained according to* (3.4)*,* (6.4)*, and* (6.1)*. Then, the overall iteration-complexity of RPB until it finds an $\bar{\varepsilon}$-solution pair $(\hat{z}_k, \hat{\eta}_k)$ is $\mathcal{O}_1(MM_fD_S^2/\bar{\varepsilon}^2)$.*

Proof. The assumption on S and the fact that $x_0 \in \operatorname{dom} h$ clearly imply that $D_S \geq d_0$. Using this fact, the assumption on λ, and Theorem 6.4(b) with the tolerance pair $(\hat{\rho}, \hat{\varepsilon}) = (\bar{\varepsilon}/(2D_S), \bar{\varepsilon}/2)$, we conclude that the overall iteration-complexity for RPB with $\delta = \bar{\varepsilon}/6$ to find a $(\hat{\rho}, \hat{\varepsilon})$-solution triple $(\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)$ is bounded by $\mathcal{O}_1(MM_fD_S^2/\bar{\varepsilon}^2)$. In view of the definition of a $(\hat{\rho}, \hat{\varepsilon})$-solution triple in Definition 6.1, we have

$$\hat{v}_k \in \partial_{\hat{\varepsilon}_k}\phi(\hat{z}_k), \quad \|\hat{v}_k\| \leq \hat{\rho} = \bar{\varepsilon}/(2D_S), \quad \hat{\varepsilon}_k \leq \hat{\varepsilon} = \bar{\varepsilon}/2. \tag{6.8}$$

The above inclusion and the remarks following (6.1) then imply that the pair $(\hat{z}_k, \hat{\eta}_k)$ satisfies the inclusion $0 \in \partial_{\hat{\eta}_k}\phi(\hat{z}_k)$. Moreover, the definition of $\hat{\eta}_k$ in (6.1), the Cauchy–Schwarz inequality, and the two inequalities in (6.8) imply that

$$\hat{\eta}_k \leq \hat{\varepsilon}_k + \|\hat{v}_k\|D_S \leq \frac{\bar{\varepsilon}}{2} + \frac{\bar{\varepsilon}}{2} = \bar{\varepsilon}$$

and hence that $(\hat{z}_k, \hat{\eta}_k)$ is a $\bar{\varepsilon}$-solution pair. We have thus shown the corollary. ∎

7. Optimal complexity results for RPB. This section contains two subsections. The first one presents a lower complexity result. The second one shows the optimality of CS-CS and RPB with respect to some important instance classes introduced in the first subsection.

7.1. A lower complexity bound. Before stating a lower complexity result, we first introduce some complexity concepts and define some important instance classes.

Given a tolerance $\bar\varepsilon$ and an arbitrary class $\mathcal{I}$ of instances $(x_0,(f,f';h))$ such that $x_0 \in \operatorname{dom} h$ and $(f,f';h)$ satisfies (A1)–(A2), let $\mathcal{A}(\mathcal{I},\bar\varepsilon)$ denote the class of algorithms $\mathcal{A}$ which, for some given $(x_0,(f,f';h)) \in \mathcal{I}$, start from x_0 and generate a finite sequence $\{x_{j-1}\}_{j=1}^J$, $J \ge 1$, satisfying the following two properties: (a) within $\{x_0,\ldots,x_{J-1}\}$, the iterate x_{J-1} is the only one which is a $\bar\varepsilon$-solution of (1.1); and (b) if h is a quadratic function and $\nabla^2 h$ is a multiple of the identity matrix I, then for every $j \in \{1,\ldots,J-1\}$, there holds

$$x_j \in x_0 + \operatorname{Lin}\{f'(x_0),\ldots,f'(x_{j-1}),\nabla h(x_0),\ldots,\nabla h(x_{j-1})\}, \tag{7.1}$$

where $\operatorname{Lin}\{\cdot\}$ is defined in subsection 1.1. Clearly, the index $J = J_{x_0}^{\bar\varepsilon}((f,f';h);\mathcal{A})$ above is uniquely determined by the tolerance $\bar\varepsilon$, instance $(x_0,(f,f';h))$, and algorithm $\mathcal{A}$. The function $J_{x_0}^{\bar\varepsilon}(\cdot;\mathcal{A})$, defined on $\mathcal{I}$, is referred to as the $\bar\varepsilon$-iteration complexity bound of $\mathcal{A}$ (with respect to $\mathcal{I}$).

For any given $\bar\varepsilon > 0$ and $\mathcal{A} \in \mathcal{A}(\mathcal{I},\bar\varepsilon)$, an $\bar\varepsilon$-upper complexity bound for $\mathcal{A}$ with respect to $\mathcal{I}$ is defined to be an upper bound on the supremum of $J_{x_0}^{\bar\varepsilon}((f,f';h);\mathcal{A})$ as $(x_0,(f,f';h))$ varies in $\mathcal{I}$. Moreover, an $\bar\varepsilon$-upper complexity bound for some algorithm $\mathcal{A} \in \mathcal{A}(\mathcal{I},\bar\varepsilon)$ with respect to $\mathcal{I}$ is said to be an $\bar\varepsilon$-upper complexity bound for the class $\mathcal{I}$. For a given instance $(x_0,(f,f';h)) \in \mathcal{I}$, a lower bound on the infimum of $J_{x_0}^{\bar\varepsilon}((f,f';h),\mathcal{A})$ as $\mathcal{A}$ varies in $\mathcal{A}(\mathcal{I},\bar\varepsilon)$ is called a lower complexity bound of $(x_0,(f,f';h))$ with respect to $\mathcal{A}(\mathcal{I},\bar\varepsilon)$. Moreover, a lower complexity bound for some instance in $\mathcal{I}$ with respect to $\mathcal{A}(\mathcal{I},\bar\varepsilon)$ is called an $\bar\varepsilon$-lower complexity bound for the class $\mathcal{I}$. Clearly, if M_1 and M_2 are $\bar\varepsilon$-lower and $\bar\varepsilon$-upper complexity bounds for the class $\mathcal{I}$, respectively, then $M_1 \le M_2$. Moreover, if $M_2 = \mathcal{O}(M_1)$, then either M_1 or M_2 is said to be an $\bar\varepsilon$-optimal complexity bound for $\mathcal{I}$ and any algorithm $\mathcal{A} \in \mathcal{A}(\mathcal{I},\bar\varepsilon)$ which has an $\bar\varepsilon$-upper complexity bound equal to $\mathcal{O}(M_2)$ is said to be $\bar\varepsilon$-optimal for $\mathcal{I}$.

We now define some important instance classes for (1.1).

DEFINITION 7.1. *Given $(M_f,\mu,R_0) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_{++}$, let $\mathcal{I}_\mu(M_f,R_0)$ denote the class consisting of all instances $(x_0,(f,f';h))$ satisfying conditions* (A1)–(A3) *and the condition that $d_0 \le R_0$, where d_0 is as in* (2.3). *Moreover, let $\mathcal{I}_\mu^u(M_f,R_0)$ denote the unconstrained class consisting of all instances $(x_0,(f,f';h)) \in \mathcal{I}_\mu(M_f,R_0)$ such that $h \equiv \mu\|\cdot\|^2/2$.*

The following result describes an $\bar\varepsilon$-lower complexity bound for any instance class $\mathcal{I} \supset \mathcal{I}_\mu^u(M_f,R_0)$. Its proof is presented in an extended version of this paper, i.e., [16].

THEOREM 7.2. *For any given quadruple $(M_f,\mu,R_0,\bar\varepsilon) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_{++} \times \mathbb{R}_{++}$, there exists an instance $(x_0,(f,f';h)) \in \mathcal{I}_\mu^u(M_f,R_0)$ which has a lower complexity bound with respect to $\mathcal{A}(\mathcal{I}_\mu^u(M_f,R_0),\bar\varepsilon)$ given by*

$$\left\lfloor \min\left\{\frac{M_f^2R_0^2}{128\bar\varepsilon^2},\frac{M_f^2}{8\mu\bar\varepsilon}\right\}\right\rfloor + 1. \tag{7.2}$$

As a consequence, (7.2) *is also an $\bar\varepsilon$-lower complexity bound for any instance class $\mathcal{I} \supset \mathcal{I}_\mu^u(M_f,R_0)$.*

It is worth mentioning that the second minimand in (7.2) is smaller than the first one if and only if $\mu \ge 16\bar\varepsilon/R_0^2$ and converges to ∞ as μ approaches zero.

We now make a few remarks regarding the relationship of Theorem 7.2 with the ones derived in Theorems 3.2.1 and 3.2.5 of [21]. First, the class of algorithms considered in these three results is the same and hence based on the linear hull condition (7.1). Second, the above three results show the existence of bad instances $(x_0, (f', f; h))$ such that $h = \mu\|\cdot\|^2/2$ (and hence $h \equiv 0$ when $\mu = 0$) but the functions f of the ones of Theorems 3.2.1 and 3.2.5 of [21] are M_f-Lipschitz on the ball $\bar{B}(x^*; R_0)$ for some $x^* \in X^*$ while the f for the one of Theorem 7.2 is M_f-Lipschitz on the whole $\mathbb{R}^n$. In contrast to the bad instances of [21], this additional property of the bad instance of Theorem 7.2 allows us to show that (7.2) is an $\bar{\varepsilon}$-lower complexity bound for a smaller instance class, namely $\mathcal{I}^u_\mu(M_f, R_0)$, than the one considered in [21]. Third, Theorem 3.2.5 (resp., Theorem 3.2.1) in [21] obtains the $\bar{\varepsilon}$-lower complexity bound $M_f^2/(2\mu\bar{\varepsilon})$ (resp., $M_f^2R_0^2/(4\bar{\varepsilon}^2)$) only for $\mu \geq 2\bar{\varepsilon}/R_0^2$ (resp, $\mu = 0$), and hence (7.2) is a valid $\bar{\varepsilon}$-lower complexity bound for any $\mu \in \{0\} \cup [2\bar{\varepsilon}/R_0^2, \infty)$. This contrasts with Theorem 7.2, which establishes the $\bar{\varepsilon}$-lower complexity bound (7.2) for any $\mu \geq 0$.

7.2. Optimal complexity results for the CS-CS and RPB methods. This subsection establishes the $\bar{\varepsilon}$-optimality of the CS-CS and RPB with respect to some of the instance classes introduced in Definition 7.1 as well as in this section.

We first tackle the $\bar{\varepsilon}$-optimalilty of the CS-CS method. Let $(x_0, \lambda) \in \operatorname{dom} h \times \mathbb{R}_{++}$ and $(M_f, \mu, R_0) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_{++}$ be given. It is easy to see that CS-CS(x_0, λ) satisfies property (b) in the paragraph containing (7.1). Hence, for any given universal constant $C > 1$, it follows from Proposition A.2 and the definition of $\mathcal{I}_\mu(M_f, R_0)$ in Definition 7.1 that CS-CS(x_0, λ) with $\bar{\varepsilon}/(CM_f^2) \leq 4\lambda \leq \bar{\varepsilon}/M_f^2$ is in $\mathcal{A}(\mathcal{I}_\mu(M_f, R_0), \bar{\varepsilon})$ and has an $\bar{\varepsilon}$-upper complexity bound for $\mathcal{I}_\mu(M_f, R_0)$ given by

$$\mathcal{O}_1\left(\min\left\{\frac{M_f^2R_0^2}{\bar{\varepsilon}^2}, \left(\frac{M_f^2}{\mu\bar{\varepsilon}} + 1\right)\log\left(\frac{\mu R_0^2}{\bar{\varepsilon}} + 1\right)\right\}\right). \tag{7.3}$$

This observation together with the $\bar{\varepsilon}$-lower complexity bound in Theorem 7.2 implies that (7.3) is an $\bar{\varepsilon}$-optimal complexity bound (up to a logarithmic term) for any instance class $\mathcal{I}$ satisfying $\mathcal{I}^u_\mu(M_f, R_0) \subseteq \mathcal{I} \subseteq \mathcal{I}_\mu(M_f, R_0)$ and that CS-CS(x_0, λ) with $\bar{\varepsilon}/(CM_f^2) \leq 4\lambda \leq \bar{\varepsilon}/M_f^2$ for a universal constant $C > 1$ is $\bar{\varepsilon}$-optimal (up to a logarithmic term) for $\mathcal{I}$.

We next tackle the $\bar{\varepsilon}$-optimalilty of the RPB method. The following result describes conditions on $\bar{\varepsilon}$ and (M_f, μ, R_0) that guarantee the $\bar{\varepsilon}$-optimality of RPB with respect to some suitable instance classes. Its statement makes use of the two instance classes $\mathcal{I}^u_\mu(M_f, R_0)$ and $\mathcal{I}_\mu(M_f, R_0)$ introduced in Definition 7.1, as well as the instance class $\mathcal{I}_0(M_f, R_0; C)$ defined as

$$\begin{aligned}&\mathcal{I}_0(M_f, R_0; C)\\ &:= \{(x_0, (f, f'; h)) \in \mathcal{I}_0(M_f, R_0) : \exists\, M_h \leq CM_f \text{ such that } h \text{ satisfies (A4)}\},\end{aligned} \tag{7.4}$$

where C is a universal constant.

THEOREM 7.3. *Let a universal constant $C > 0$, tolerance $\bar{\varepsilon} > 0$, and pair $(M_f, R_0) \in \mathbb{R}_+ \times \mathbb{R}_{++}$ be given such that $CM_fR_0/\bar{\varepsilon} \geq 1$. Then, the following statements hold:*

(a) *for any universal constant $C' > 0$, RPB$(x_0, \lambda, \bar{\varepsilon}/2)$ with λ satisfying* (3.12) *with d_0 replaced by R_0 is (up to a logarithmic term) $\bar{\varepsilon}$-optimal for any instance class $\mathcal{I}$ and scalar $\mu \in [0, C'M_f/R_0]$ such that*

$$\mathcal{I}^u_\mu(M_f, R_0) \subseteq \mathcal{I} \subseteq \mathcal{I}_\mu(M_f, R_0); \tag{7.5}$$

(b) *RPB*$(x_0, \lambda, \bar\varepsilon/2)$ *with* λ *satisfying* (3.16) *with* d_0 *replaced by* R_0 *is* $\bar\varepsilon$*-optimal for any instance class* $\mathcal{I}$ *such that*

$$\mathcal{I}_0^u(M_f, R_0) \subseteq \mathcal{I} \subseteq \mathcal{I}_0(M_f, R_0; C). \tag{7.6}$$

We now make some remarks about Theorem 7.3.

The inclusion $\mathcal{I}_0^u(M_f, R_0) \subseteq \mathcal{I}_0(M_f, R_0; C)$ always holds in view of (7.4) and the fact that the composite component h of any instance in $\mathcal{I}_0^u(M_f, R_0)$ is identically zero. Hence, in view of the last conclusion of Theorem 7.2, (7.2) is also an $\bar\varepsilon$-lower complexity bound for $\mathcal{I}_0(M_f, R_0; C)$.

Theorem 7.3(a) shows that RPB with $R_0/M_f \le \lambda \le CR_0^2/\bar\varepsilon$, similar to the CS-CS method with $\bar\varepsilon/(CM_f^2) \le 4\lambda \le \bar\varepsilon/M_f^2$ (see subsection 2.2), is $\bar\varepsilon$-optimal (up to a logarithmic term) for the instance class $\mathcal{I}_\mu(M_f, R_0)$ for any $\mu \ge 0$. Note that the two ranges of λ above do not overlap when $C \le 1$ due to the assumption that $CM_f R_0/\bar\varepsilon \ge 1$ in Theorem 7.3.

On the other hand, Theorem 7.3(b) asserts that RPB with λ within the much wider range $\bar\varepsilon/(CM_f^2) \le \lambda \le CR_0^2/\bar\varepsilon$ is $\bar\varepsilon$-optimal for the smaller instance class $\mathcal{I}_0(M_f, R_0; C)$, which includes the instance subclass where h is the indicator function of a closed convex set.

8. Concluding remarks. This paper presents a proximal bundle variant, i.e., the RPB method, for solving CNCO problems. Like many other proximal bundle variants, (i) RPB solves a sequence of prox bundle subproblems whose objective functions are obtained by a usual regularized composite cutting-plane strategy, and (ii) RPB performs either serious iterations during which the prox-centers are changed or null iterations where the prox-centers are left unchanged. However, RPB uses the novel condition (3.3) involving $\tilde x_j$ to decide whether to perform a serious or null iteration. Our analysis shows that the consideration of the sequence $\{\tilde x_j\}$ plays an important role in the derivation of optimal complexity bounds for RPB over a large range of prox stepsizes λ in the context of CNCO problems.

As far as the authors are aware of, this is the first time that such results are obtained in the context of a proximal bundle variant. A nice feature of our analysis is that it is carried out in the context of CNCO problems and takes into account a flexible bundle management policy which allows cut removal but no cut aggregation. Moreover, it places the CS-CS method under the umbrella of RPB in that the former can be viewed as an instance of the latter with a relatively small prox stepsize. This paper also establishes iteration-complexity results for RPB to obtain iterates satisfying practical termination criteria.

We now discuss some possible extensions of our analysis in this paper. First, recall that we have assumed throughout this paper that the prox stepsize λ is constant. We believe a slightly modified version of our analysis can be used to study the case in which λ is allowed to change (possibly within a positive closed bounded interval) at every iteration j for which j is a serious iteration index. Second, if f is μ_f-strongly convex and h is μ_h-strongly convex, then the CNCO problem (1.1) is clearly equivalent to another CNCO problem (1.1) in which f is convex, h is μ-strongly convex, and $\mu = \mu_f + \mu_h$. Hence, if μ_f is known, then there is no loss of generality in assuming that only h is strongly convex. Third, a natural question is whether, under the weaker assumption that ϕ is μ-strongly convex, the results are still valid for RPB directly applied to the CNCO problem (1.1) without using the above transformation. The advantage of the latter approach, if doable, is that it does not require the knowledge of μ_f (nor μ_h). Fourth, it would be interesting to investigate a variant of RPB under the assumption that f shares properties of both a smooth and a nonsmooth function, i.e., for some nonnegative scalars M_f and L_f, there holds $\|f'(x)-f'(x')\| \le 2M_f + L_f\|x-x'\|$ for every $x, x' \in \operatorname{dom} h$. Fifth, it would be interesting to consider an RPB variant which, instead of using the cutting-plane model f_j in (1.2), uses the cut aggregation model considered, for example, in Chapter 7.4.4 of [24] (see also [5, 22]). A clear advantage of the latter model is that the cardinality of the bundle is no more than two and, as a consequence, subproblem (1.3) becomes easier to solve. Sixth, it would be interesting to extend the conclusion of Corollary 3.2 to the one where (3.12) is replaced by the wider range (3.16). Note that such a version of Corollary 3.2, if correct, would imply Corollary 3.3 as a special case.

Appendix A. Proof of the iteration-complexity of the CS-CS method. The goal of this section is to establish a complexity bound for CS-CS(x_0, λ) with λ satisfying $\bar{\varepsilon}/(CM_f^2) = 4\lambda \le \bar{\varepsilon}/M_f^2$ for a universal constant $C > 1$ without assuming any condition on the initial point x_0 other than just being in $\operatorname{dom} h$. Before presenting the complexity bound result, we first state a useful technical lemma.

Lemma A.1. *Assume that scalars $\theta \ge 1$ and $\delta > 0$ and sequences of nonnegative scalars $\{\eta_j\}$ and $\{\alpha_j\}$ satisfy*

$$\eta_j \le \alpha_{j-1} - \theta\alpha_j + \delta \quad \forall j \ge 1. \tag{A.1}$$

Then, the following statements hold:

(a) $\min_{1\le j\le k} \eta_j \le 2\delta$ *for every* $k \ge 1$ *such that*

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

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) $\alpha_k \le \alpha_0 + k\delta$ *for every* $k \ge 1$.

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

$$\sum_{j=1}^{k}\theta^{j-1}\left[\min_{1\le j\le k}\eta_j\right] \le \sum_{j=1}^{k}\theta^{j-1}\eta_j \le \sum_{j=1}^{k}\theta^{j-1}\left(\alpha_{j-1}-\theta\alpha_j+\delta\right)$$

$$= \alpha_0 - \theta^k\alpha_k + \sum_{j=1}^{k}\theta^{j-1}\delta. \tag{A.2}$$

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

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

This inequality, (A.2), and the fact that $\alpha_k \ge 0$ imply that for every $k \ge 1$,

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

which can be easily seen to imply (a).

(b) This statement follows from (A.2) and the facts that $\eta_j \ge 0$ and $\theta \ge 1$. ∎

Now we are ready to present the main result of the section.

PROPOSITION A.2. *Let* $(M_f, \mu) \in \mathbb{R}_+ \times \mathbb{R}_+$ *and instance* $(x_0, (f, f'; h))$ *satisfying conditions* (A1)–(A3) *be given. Then, the number of iterations performed by CS-CS*(x_0, λ) *with* $\lambda \le \bar\varepsilon/(4M_f^2)$ *until it finds a* $\bar\varepsilon$*-solution is bounded by*

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

Proof. Recall that an iteration of CS-CS(x_0, λ) is as in (2.4). Using the fact that the objective function in (2.4) is $(\mu + 1/\lambda)$-strongly convex and Theorem 5.25(b) of [2], we conclude that for every $j \ge 1$ and $u \in \operatorname{dom} h$,

$$\begin{aligned} &\ell_f(x_j; x_{j-1}) + h(x_j) + \frac{1}{2\lambda}\|x_j - x_{j-1}\|^2 + \frac{1}{2}\left(\mu + \frac{1}{\lambda}\right)\|u - x_j\|^2 \\ &\qquad \le \ell_f(u; x_{j-1}) + h(u) + \frac{1}{2\lambda}\|u - x_{j-1}\|^2, \end{aligned} \tag{A.3}$$

where $\ell_f(u; v) := f(v) + \langle f'(v), u - v\rangle$ for every $u, v \in \operatorname{dom} h$. Noting that (A4), (2.1), the definition of ℓ_f, the triangle inequality, and the Cauchy–Schwarz inequality imply

$$f(x_j) - \ell_f(x_j; x_{j-1}) \le |f(x_j) - f(x_{j-1})| + \|f'(x_{j-1})\|\|x_j - x_{j-1}\| \le 2M_f\|x_j - x_{j-1}\|,$$

and using the definition of ϕ in (1.1), and the fact that $\ell_f(\cdot; v) \le f(\cdot)$ for every $v \in \operatorname{dom} h$, we then conclude from (A.3) with $u = x_0^*$ that

$$\begin{aligned} &\frac{1}{2}\left(\mu + \frac{1}{\lambda}\right)\|x_0^* - x_j\|^2 + \phi(x_j) - \phi^* \\ &\quad \le \frac{1}{2\lambda}\|x_0^* - x_{j-1}\|^2 + f(x_j) - \ell_f(x_j; x_{j-1}) - \frac{1}{2\lambda}\|x_j - x_{j-1}\|^2 \\ &\quad \le \frac{1}{2\lambda}\|x_0^* - x_{j-1}\|^2 + 2M_f\|x_j - x_{j-1}\| - \frac{1}{2\lambda}\|x_j - x_{j-1}\|^2 \\ &\quad \le \frac{1}{2\lambda}\|x_0^* - x_{j-1}\|^2 + 2\lambda M_f^2, \end{aligned}$$

where the last inequality is due to the fact that $a^2 + b^2 \ge 2ab$ for every $a, b \in R$. Since the above inequality satisfies (A.1) with $\eta_j = \phi(x_j) - \phi^*$, $\alpha_j = \|x_j - x_0^*\|^2/(2\lambda)$, $\theta = 1 + \lambda\mu$, and $\delta = \bar\varepsilon/2$ in view of $\lambda \le \bar\varepsilon/(4M_f^2)$, it follows from Lemma A.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 $k \ge 1$ 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 conclusion of the lemma holds. ∎

Appendix B. Proof of Theorem 7.3.

LEMMA B.1. *For any* $\bar\varepsilon, \lambda > 0$ *and* $(M_f, \mu, R_0) \in \mathbb{R}^3_+$, *RPB*$(x_0, \lambda, \bar\varepsilon/2)$ *is in* $\mathcal{A}(\mathcal{I}^u_\mu(M_f, R_0), \bar\varepsilon)$.

Proof. To simplify notation within this proof, denote $\mathcal{I}^u_\mu(M_f, R_0)$ simply by $\mathcal{I}^u_\mu$. Our goal is to show that RPB satisfies properties (a) and (b) in the definition (see the paragraph containing (7.1)) of $\mathcal{A}(\mathcal{I}^u_\mu, \bar\varepsilon)$. Indeed, (a) follows from Theorem 3.1(c). In order to show property (b), assume that there exists $\alpha \ge \mu$ such that $\nabla^2 h(x) = \alpha I$ for every $x \in \mathbb{R}^n$. Note first that the optimality condition of (1.3), the above assumption on h, and the facts that $x^c_{j-1} = x_{\ell_0}$ and $\partial f_j(x_j) = \mathrm{conv}\{f'(x) : x \in A_j\}$ (see Corollary 4.3.2 of [26]) imply that for any two consecutive serious iteration indices ℓ_0 and ℓ_1 and any index j such that $\ell_0 < j \le \ell_1$,

$$0 \in \partial f_j(x_j) + \nabla h(x_j) + \frac{1}{\lambda}(x_j - x_{\ell_0}) = \mathrm{conv}\{f'(x) : x \in A_j\} + \nabla h(x_j) + \frac{1}{\lambda}(x_j - x_{\ell_0})$$
$$= \mathrm{conv}\{f'(x) : x \in A_j\} + \nabla h(x_{\ell_0}) + \left(\frac{1}{\lambda} + \alpha\right)(x_j - x_{\ell_0})$$

and hence that (7.1) holds with x_0 replaced by x_{ℓ_0}. Using this inclusion and a simple induction argument, it is easy to see that (7.1) holds for every $j \ge 1$ and hence that property (b) holds. ∎

We are now ready to present the proof of Theorem 7.3.

Proof of Theorem 7.3. For brevity RPB$(x_0, \lambda, \bar{\varepsilon}/2)$ is referred to below to as RPB.

(a) In view of Theorem 7.2, (a) will follow from the claim that (7.3) is an $\bar{\varepsilon}$-upper complexity bound for RPB with respect to the instance class $\mathcal{I}_\mu(M_f, R_0)$. To show the latter claim, first note that RPB is in $\mathcal{A}(\mathcal{I}_\mu(M_f, R_0), \bar{\varepsilon})$ in view of Lemma B.1. It follows from Corollary 3.2 and the fact $d_0 \le R_0$ that we conclude that (7.3) is an $\bar{\varepsilon}$-upper complexity bound for RPB with respect to $\mathcal{I}_\mu(M_f, R_0)$ and hence that the aforementioned claim holds.

(b) In view of Theorem 7.2, (b) will follow from the claim that (7.3) with $\mu = 0$ is an $\bar{\varepsilon}$-upper complexity bound for RPB with respect to the instance class $\mathcal{I}_0(M_f, R_0; C)$ (and hence to any instance class $\mathcal{I}$ satisfying (7.6)). To show the latter claim, first note that RPB is in $\mathcal{A}(\mathcal{I}_0(M_f, R_0; C), \bar{\varepsilon})$ in view of Lemma B.1 with $\mu = 0$. Moreover, since the second inclusion of (7.6) and the definition of $\mathcal{I}_0(M_f, R_0; C)$ in (7.4) imply that $M_h \le C M_f$ and $d_0 \le R_0$, it follows from Corollary 3.3 that $\mathcal{O}_1(M_f^2 R_0^2/\bar{\varepsilon}^2)$ is an $\bar{\varepsilon}$-upper complexity bound for RPB with respect to $\mathcal{I}_0(M_f, R_0; C)$. Clearly, the previous bound is equal to (7.3) with $\mu = 0$. ∎

Acknowledgments. We are grateful to the two anonymous referees and the associate editor, Marc Teboulle, for providing helpful comments on earlier versions of this manuscript.

REFERENCES

[1] A. Astorino, A. Frangioni, A. Fuduli, and E. Gorgone, *A nonmonotone proximal bundle method with (potentially) continuous step decisions*, SIAM J. Optim., 23 (2013), pp. 1784–1809.

[2] A. Beck, *First-Order Methods in Optimization*, MOS-SIAM Ser. Optim. 25, SIAM, Philadelphia, 2017.

[3] A. Ben-Tal and A. Nemirovski, *Non-euclidean restricted memory level method for large-scale convex optimization*, Math. Program., 102 (2005), pp. 407–456.

[4] D. P. Bertsekas, *Convex Optimization Algorithms*, Athena Scientific, Belmont, MA, 2015.

[5] Y. Du and A. Ruszczyński, *Rate of convergence of the bundle method*, J. Optim. Theory Appl., 173 (2017), pp. 908–922.

[6] A. Frangioni, *Generalized bundle methods*, SIAM J. Optim., 13 (2002), pp. 117–156.

[7] M. Fuentes, J. Malick, and C. Lemaréchal, *Descentwise inexact proximal algorithms for smooth optimization*, Comput. Optim. Appl., 53 (2012), pp. 755–769.

[8] B. Grimmer, *General Convergence Rates Follow from Specialized Rates Assuming Growth Bounds*, https://arxiv.org/abs/1905.06275, 2019.

[9] O. Güler, *New proximal point algorithms for convex minimization*, SIAM J. Optim., 2 (1992), pp. 649–664.

[10] K. C. Kiwiel, *Proximal level bundle methods for convex nondifferentiable optimization, saddle-point problems and variational inequalities*, Math. Program., 69 (1995), pp. 89–109.

[11] K. C. Kiwiel, *Efficiency of proximal bundle methods*, J. Optim. Theory Appl., 104 (2000), pp. 589–603.
[12] G. Lan, *Bundle-level type methods uniformly optimal for smooth and nonsmooth convex optimization*, Math. Program., 149 (2015), pp. 1–45.
[13] C. Lemaréchal, *An extension of davidon methods to non differentiable problems*, in Nondifferentiable Optimization, Springer, New York, 1975, pp. 95–109.
[14] C. Lemaréchal, *Nonsmooth Optimization and Descent Methods*, Research report, IIASA, Laxenburg, Austria, 1978.
[15] C. Lemaréchal, A. Nemirovski, and Y. Nesterov, *New variants of bundle methods*, Math. Program., 69 (1995), pp. 111–147.
[16] J. Liang and R. D. C. Monteiro, *A Proximal Bundle Variant with Optimal Iteration-Complexity for a Large Range of Prox Stepsizes*, https://arxiv.org/abs/2003.11457, 2020.
[17] H. Lin, J. Mairal, and Z. Harchaoui, *Catalyst acceleration for first-order convex optimization: From theory to practice*, J. Mach. Learn. Res., 18 (2017), pp. 7854–7907.
[18] R. Mifflin, *A modification and an extension of Lemaréchal's algorithm for nonsmooth minimization*, in Nondifferential and Variational Techniques in Optimization, Springer, New York, 1982, pp. 77–90.
[19] R. D. C. Monteiro and B. F. Svaiter, *On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean*, SIAM J. Optim., 20 (2010), pp. 2755–2787.
[20] R. D. C. Monteiro and B. F. Svaiter, *Complexity of variants of Tseng's modified F-B splitting and Korpelevich's methods for hemivariational inequalities with applications to saddle-point and convex optimization problems*, SIAM J. Optim., 21 (2011), pp. 1688–1720.
[21] Y. Nesterov, *Lectures on Convex Optimization*, Springer Optim. Appl. 137, Springer, New York, 2018.
[22] W. de. Oliveira, C. Sagastizábal, and C. Lemaréchal, *Convex proximal bundle methods in depth: A unified analysis for inexact oracles*, Math. Program., 148 (2014), pp. 241–277.
[23] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.
[24] A. Ruszczyński, *Nonlinear Optimization*, Princeton University Press, Princeton, NJ, 2011.
[25] M. V. Solodov and B. F. Svaiter, *Error bounds for proximal point subproblems and associated inexact proximal point algorithms*, Math. Program., 88 (2000), pp. 371–389.
[26] J.-B. H. Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms* I, Springer, New York, 1993.
[27] J.-B. H. Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms* II, Springer, New York, 1993.
[28] W. van Ackooij, V. Berge, W. de Oliveira, and C. Sagastizábal, *Probabilistic optimization via approximate p-efficient points and bundle methods*, Comput. Oper. Res., 77 (2017), pp. 177–193.
[29] W. van Ackooij and A. Frangioni, *Incremental bundle methods using upper models*, SIAM J. Optim., 28 (2018), pp. 379–410.
[30] W. van Ackooij and C. Sagastizábal, *Constrained bundle methods for upper inexact oracles with application to joint chance constrained energy problems*, SIAM J. Optim., 24 (2014), pp. 733–765.
[31] P. Wolfe, *A method of conjugate subgradients for minimizing nondifferentiable functions*, in Nondifferentiable Optimization, Springer, New York, 1975, pp. 145–173.