A proximal bundle variant with optimal iteration-complexity for a large range of prox stepsizes

Jiaming Liang *  Renato D.C. Monteiro *

March 26, 2020

Abstract

This paper presents a proximal bundle variant, namely, the RPB method, for solving convex nonsmooth composite optimization problems. Like other proximal bundle variants, the RPB method solves a sequence of prox subproblems whose objective functions are regularized cutting-plane models. In contrast to other variants, instead of deciding whether to perform a serious or null iteration based on a descent condition, the RPB method uses a relaxed non-descent condition, which does not necessarily yield a function value decrease. Iteration-complexity bounds for the RPB method in the convex and strongly convex settings are derived which are optimal (possibly up to a logarithmic term) for a large range of prox stepsizes. 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 the RPB method 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

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, both in the context of convex and strongly convex nonsmooth composite optimization (NCO) problems.

RPB is presented in the context of the convex NCO problem of the form

\[ \phi_* := \min \{ \phi(x) := f(z) + h(x) : x \in \mathbb{R}^n \} \]

where \( f, h : \mathbb{R}^n \to \bar{\mathbb{R}} \) are proper closed convex functions such that \( \text{dom} \ h \subseteq \{ x \in \mathbb{R}^n : \partial f(x) \neq \emptyset \} \). Moreover, it is assumed that there exists a scalar \( M_f \geq 0 \) (resp., \( M_h \geq 0 \)) such that \( \partial f(X) \subset \{ s \in \mathbb{R}^n : \| s \| \leq M_f \} \) (resp., \( h \) is \( M_h \)-Lipschitz continuous on \( X \)) where \( X = \text{dom} \ h \). Like other proximal bundle variants, the RPB method solves a sequence of prox subproblems of the form

\[ x_j := \arg\min_{u \in \text{dom} \ h} \left\{ \phi^\lambda_j(u) := f_j(u) + h(u) + \frac{1}{2\lambda} \| u - x_{j-1} \|^2 \right\}, \]

\*School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205. (email: jiaming.liang@gatech.edu and renato.monteiro@isye.gatech.edu). This work was partially supported by ONR Grant N00014-18-1-2077.
where \( x_{j-1}^c \) is the prox-center, \( f_j \) is the cutting-plane model defined as

\[
f_j(u) = \max \{ f(x_i) + \langle g_i, u - x_i \rangle : g_i \in \partial f(x_i), x_i \in C_j \} \quad \forall u \in \mathbb{R}^n,
\]

and \( C_j \) is a suitable subset of \( \{ x_0, x_1, \ldots, x_{j-1} \} \). RPB also performs two types of iterations, namely: i) serious ones during which the prox-center is changed; or ii) null ones where the prox-center is left unchanged. As opposed to other proximal bundle variants that decide whether to perform a serious or null iteration based on a descent condition, namely, \( f(x_j) \leq (1 - \gamma)f(x_{j-1}^c) + \gamma f_j(x_j) \), for some \( \gamma \in (0, 1) \), RPB uses a relaxed non-descent condition which, may not necessarily result in a function value decrease but, plays an important role in establishing that its iteration-complexity is optimal for a large range of prox stepsizes \( \lambda \). Moreover, the RPB method provides more flexibility in the construction of the cutting-plane model \( f_j \) as follows. When a serious iteration occurs, the new bundle \( C_{j+1} \) can be chosen to be an arbitrary set satisfying \( C_{j+1} \supset \{ x_j \} \), and hence can be simply reduced to the singleton \( \{ x_j \} \). Moreover, as the number of null iterations between two consecutive serious iterations is shown to be \( O(\lambda) \), the size of the bundle \( C_j \) is at most \( O(\lambda) \) for every \( j \geq 1 \), and hence can be kept under control if \( \lambda \) is chosen not too large. This is in contrast to other proximal bundle variants (see [2, 3, 6, 14, 16]) where the size of \( C_j \) might grow indefinitely.

It is shown in this paper that the iteration-complexity (see (22)) of the RPB method to find a \( \bar{\varepsilon} \)-optimal solution (i.e., a point \( \bar{x} \in \text{dom} h \) satisfying \( \phi(\bar{x}) - \phi_* \leq \bar{\varepsilon} \)) is optimal in the context of the convex NCO problem for a large range of prox stepsizes \( \lambda \), namely, \( \Omega(\bar{\varepsilon}/M^2) = \lambda = O(d_0^2/\bar{\varepsilon}) \) where \( M = M_f + M_h \) and \( d_0 \) is the distance of the initial point to the set of optimal solutions. It is also shown under the assumptions that \( h \) is \( \mu \)-strongly convex and \( \text{dom} h \) is bounded that the iteration-complexity of the RPB method (see (65)) is nearly optimal (up to a logarithmic term) for a large range of prox stepsizes \( \lambda \), namely, \( \Omega(\bar{\varepsilon}/M^2) = \lambda = O(1/\mu) \). In addition, iteration-complexity results are also established for the RPB method to obtain iterates satisfying practical termination criteria rather than a \( \bar{\varepsilon} \)-optimal solution. Finally, another interesting conclusion of our analysis is that the composite subgradient method can be viewed as a special instance of the RPB method with prox stepsize \( \lambda \) satisfying \( \lambda = \Theta(\bar{\varepsilon}/M^2) \).

**Related works.** Some preliminary ideas towards the development of the proximal bundle method were first presented in [8, 17] and formal presentations of the method were given in [9, 11]. Convergence analysis of the proximal bundle method for convex NCO problems has been broadly discussed in the literature and can be found for example in the textbooks [14, 16].

We now discuss iteration-complexity bounds, that have been previously derived in [3, 6], for two proximal bundle variants with the aim of obtaining a \( \bar{\varepsilon} \)-optimal solution of (1). Paper [6] establishes the first bound in the context of the convex NCO problem with \( h \) being the indicator function of a nonempty closed convex set, namely, \( O(M^2 d_0^3/(\lambda \bar{\varepsilon}^3)) \), which is optimal only when \( \lambda = \Theta(d_0^2/\bar{\varepsilon}) \). Paper [3] considers the unconstrained version of the NCO problem (1) in which \( h \) is identically zero and \( f \) is \( \mu \)-strongly convex everywhere, and establishes the iteration-complexity bound (68) below which, only when \( \lambda = \Theta(1/\mu) \), reduces to the optimal complexity bound (up to the product of two logarithmic terms). In conclusion, as opposed to the results established in this paper, the iteration-complexity bounds obtained in these two papers are optimal only for specific values of \( \lambda \).

Another method related, and developed subsequently, to the proximal bundle method is the bundle-level method, which was first proposed in [10] and extended in many ways in [1, 5, 7]. These methods have been shown to have optimal iteration-complexity in the setting of the convex NCO 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 notations used throughout the paper. Section 2 contains two subsections. The first one describes the problem of interest, its corresponding assumptions, and some notions of approximate solutions for it. The second subsection presents the RPB method for solving the problem of interest and its corresponding iteration-complexity bounds with respect to the different notions of approximate solutions introduced in the first subsection. Section 3 provides the proof of one of the results stated in Section 2, which establishes a bound on the number of null iterates between two consecutive serious iterates. It also presents a result showing that the composite subgradient method is a special instance of the RPB method. Section 4 provides the proofs of two results stated in Section 2, which establish bounds on the total number of serious iterates generated by RPB as well as the overall iteration-complexity for RPB. Section 5 describes an iteration-complexity result for the RPB method in the context of the strongly convex NCO problem, and discusses its implication to the convex setting. Moreover, this section also compares the complexity bounds obtained by RPB with those obtained by other proximal bundle variants. Section 6 presents some concluding remarks. Finally, the appendix states a technical result which is useful in our analysis.

1.1 Basic definitions and notation

This subsection provides some basic definitions and notation used throughout this paper.

The set of real numbers is denoted by \( \mathbb{R} \). Let \( \mathbb{R} \) denote the set \( \mathbb{R} \cup \{ \pm \infty \} \). 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. For \( t > 0 \), define \( \log^+ (t) := \max \{ \log t, 1 \} \).

Let \( \psi : \mathbb{R}^n \to (-\infty, +\infty] \) be given. The effective domain of \( \psi \) is denoted by \( \text{dom} \psi := \{ x \in \mathbb{R}^n : \psi(x) < \infty \} \) and \( \psi \) is proper if \( \text{dom} \psi \neq \emptyset \). Moreover, a proper function \( \psi : \mathbb{R}^n \to (-\infty, +\infty] \) is \( \mu \)-strongly 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 \( z, u \in \text{dom} \psi \) and \( \alpha \in [0, 1] \). The set of all proper lower semicontinuous convex functions \( \psi : \mathbb{R}^n \to (-\infty, +\infty] \) is denoted by \( \text{Conv} (\mathbb{R}^n) \). For \( \varepsilon \geq 0 \), its \( \varepsilon \)-subdifferential at \( z \in \text{dom} \psi \) is denoted by
\[
\partial \varepsilon \psi(z) := \{ v \in \mathbb{R}^n : \psi(u) \geq \psi(z) + \langle v, u - z \rangle - \varepsilon, \forall u \in \mathbb{R}^n \}.
\] (4)
The subdifferential of \( \psi \) at \( z \in \text{dom} \psi \), denoted by \( \partial \psi(z) \), corresponds to \( \partial 0 \psi(z) \).

2 Main problem, assumptions and the RPB method

This section contains two subsections. The first one describes the main problem, its corresponding assumptions, and some notions of approximate solutions for it. The second one presents the RPB method for solving the main problem and states iteration-complexity results for the different notions of approximate solutions mentioned above. The proof of most of these results are postponed to Sections 3 and 4.

2.1 Main problem and assumptions

This subsection describes the main problem and the assumptions made on it in detail. It also discusses some notions of approximate solutions that will be used as stopping criteria of the RPB method for solving the main problem.
The problem of interest in this paper is (1), where the following conditions are assumed to hold:

(A1) \( h \in \overline{\text{Conv}}(\mathbb{R}^n) \) is \( M_h \)-Lipschitz continuous on \( \text{dom} \ h \), i.e.,
\[
|h(u) - h(v)| \leq M_h \|u - v\| \quad \forall u, v \in \text{dom} \ h;
\]

(A2) \( f \in \overline{\text{Conv}}(\mathbb{R}^n) \) is such that
\[
\text{dom} \ h \subset \{u \in \mathbb{R}^n : \partial f(u) \neq \emptyset \}
\]
and there exists a scalar \( M_f \geq 0 \) satisfying
\[
\|g_f\| \leq M_f, \quad \forall g_f \in \partial f(u), \forall u \in \text{dom} \ h.
\]

We make some remarks about the above assumptions. First, the iterates generated by the RPB method of Subsection 2.2 all lie in \( \text{dom} \ h \) and (5) implies that there always exists a subgradient of \( f \) at any one of these iterates. Second, in view of Theorem 23.4 in [13], a sufficient condition for (5) to hold is that \( \text{dom} \ h \subset \text{ri}(\text{dom} \ f) \). Third, it follows as a consequence of (6) that
\[
|f(u) - f(v)| \leq M_f \|u - v\| \quad \forall u, v \in \text{dom} \ h.
\]

Fourth, any indicator function of a closed convex set satisfies (A1). Fifth, an alternative and weaker condition that can be used in place of (A2) is described in the third paragraph of Section 6.

We now define some notions of approximate solutions considered in our analysis. For a given tolerance \( \bar{\varepsilon} > 0 \), a point \( x \) is called a \( \bar{\varepsilon} \)-optimal solution of (1) if
\[
\phi(x) - \phi_* \leq \bar{\varepsilon}.
\]

We now make some trivial comments about the above notion of approximate solution. First, while (8) is theoretically appealing from a complexity point of view, it can rarely be used as a stopping criterion since \( \phi_* \) is generally not known. Second, (8) is obviously equivalent to the condition that \( 0 \in \partial \varepsilon \phi(x) \). These two remarks naturally lead us to consider the following stronger notion of approximate solution: a pair \((x, \varepsilon)\) is called a \( \varepsilon \)-optimal pair of (1) if
\[
0 \in \partial \varepsilon \phi(x), \quad \varepsilon \leq \bar{\varepsilon}.
\]

The advantage of the latter notion compared to the first one is that a residual \( \varepsilon \) satisfying the above inclusion is given in order to certify that \( x \) is an \( \varepsilon \)-optimal solution. Clearly, if the residual \( \varepsilon \) also satisfies the inequality in (9), then it follows that \( x \) is a \( \varepsilon \)-optimal solution. More generally, for a given tolerance pair \((\hat{\rho}, \hat{\varepsilon})\), a triple \((x, v, \varepsilon)\) is called a \((\hat{\rho}, \hat{\varepsilon})\)-approximate solution of (1) if it satisfies
\[
v \in \partial \varepsilon \phi(x), \quad \|v\| \leq \hat{\rho}, \quad \varepsilon \leq \hat{\varepsilon}.
\]

Clearly, (10) with \( \hat{\varepsilon} \geq \bar{\varepsilon} \) is a relaxation of (9) and reduces to it when \( \hat{\varepsilon} = \bar{\varepsilon} \) and \( \hat{\rho} = 0 \).

We now make some comments about (9) and (10) for the case in which \( \text{dom} \ h \) is unbounded. Many algorithms, including the one considered in this paper, are able to generate a sequence (or subsequence) \( \{(\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)\} \) for which: i) \((x, v, \varepsilon) = (\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)\) satisfies the inclusion in (10) for every \( k \); and ii) \((\|\hat{v}_k\|, \hat{\varepsilon}_k)\) can be made arbitrarily small (possibly by choosing some input for the algorithm also small). As a consequence, one of the iterates of the above sequence will eventually satisfy (10) and verification of this fact simply amounts to checking that \((\hat{v}_k, \hat{\varepsilon}_k)\) satisfies the two inequalities in (10).
It is natural then to wonder if these same algorithms can also produce a sequence with the same
properties as above but with \( \hat{v}_k = 0 \) for every \( k \geq 0 \). It turns out that, when \( \text{dom} \ h \) is unbounded,
such sequence is generally difficult or impossible to obtain. However, when \( \text{dom} \ h \) is bounded, we
can easily obtain such a sequence from a sequence \( \{ (\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k) \} \) satisfying properties i) and ii) in
the last paragraph as follows. Let \( \Omega \) be a compact convex set containing \( \text{dom} \ h \) and, for every \( k \), define
\[
\hat{\eta}_k := \hat{\varepsilon}_k + \sup \{ \langle \hat{v}_k, \hat{z}_k - x \rangle : x \in \Omega \}. \tag{11}
\]
Then, it follows from property i) and the above definition of \( \hat{\eta}_k \) that
\[
\phi(x) \geq \phi(\hat{z}_k) + \langle \hat{v}_k, x - \hat{z}_k \rangle - \hat{\varepsilon}_k \geq \phi(\hat{z}_k) - \hat{\eta}_k \quad \forall x \in \text{dom} \ h,
\]
or equivalently, \( 0 \in \partial \phi(\hat{z}_k) \). Clearly, the sequence \( \{ (\hat{z}_k, 0, \hat{\eta}_k) \} \) satisfies properties i) and ii), and
has residual \( \hat{v}_k = 0 \) for every \( k \). As a consequence, it eventually yields a pair \( (x, \varepsilon) = (\hat{z}_k, \hat{\eta}_k) \) such
that (9) holds.

### 2.2 The RPB method and main results

This subsection describes the RPB method and states iteration-complexity results with respect to
the different notions of approximate solutions introduced in (8), (9) and (10).

We start by stating the RPB method which is based on the cutting-plane models \( f_j \) defined in
(3). Note that these models are defined in terms of a finite set \( C_j \subset \{ x_0, x_1, \ldots, x_{j-1} \} \) and are
used to construct the subproblems (2) which are generated by the RPB method according to its
description given below.

#### RPB method

0. Let \( x_0 \in \text{dom} \ h, \lambda > 0 \) and \( \delta > 0 \) be given, select \( g_0 \in \partial f(x_0) \), and set \( x_0^c = x_0 \), \( X_0 = \{ x_0 \} \),
\( C_1 = \{ x_0 \} \), and \( j = 1 \);

1. Compute \( x_j \) according to (2) and let \( m_j := \phi_j^\lambda(x_j) \) denote the optimal value of (2). Moreover,
let \( \check{x}_j \) be a point from the finite set \( X_{j-1} \cup \{ x_j \} \) with the lowest \( \phi_j^\lambda(\cdot) \) value, where
\[
\phi_j^\lambda = \phi + \frac{1}{2 \lambda} \| x_{j-1}^c \|^2; \tag{12}
\]

2. If
\[
t_j := \phi_j^\lambda(\check{x}_j) - m_j \leq \delta, \tag{13}
\]

2.a) **then** perform a serious iteration, i.e.,: set \( x_j^c = x_j \) and \( X_j = \{ x_j \} \), choose \( C_{j+1} \) such
that \( \{ x_j \} \subset C_{j+1} \), and go to step 3;

2.b) **else** perform a null iteration, i.e.,: set \( x_j^c = x_{j-1}^c \) and \( X_j = X_{j-1} \cup \{ x_j \} \), choose \( C_{j+1} \) such
that
\[
A_j \cup \{ x_j \} \subset C_{j+1} \subset C_j \cup \{ x_j \} \tag{14}
\]
where
\[
A_j := \{ x_i \in C_j : f(x_i) + \langle g_i, x_j - x_i \rangle = f_j(x_j), g_i \in \partial f(x_i) \}, \tag{15}
\]
and go to step 3;
3. Select \( g_j \in \partial f(x_j) \) and set \( j \leftarrow j + 1 \), and go to step 1.

An iteration index \( j \) for which (13) 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 remarks about the RPB method. First, the existence of \( g_j \in \partial f(x_j) \) in step 3 follows from (5) and the fact that (2) implies that \( x_j \in \text{dom } h \). Second, \( C_j \) consists of the set of points that are used to construct the cutting-plane model \( f_j \) which minorizes \( f \). Third, \( A_j \) consists of the subset of points from \( C_j \) which are active at the most recent point \( x_j \). Fourth, the prox-center \( x^c_{j-1} \) remains the same when \( j \) is a null iteration index and it is updated to the most recent \( x_j \) only when \( j \) is a serious iteration index. Fifth, no termination criterion is added to RPB in order to later consider different ways of stopping it, namely, (8)-(10).

We now discuss the above RPB method in light of other well-known proximal bundle variants discussed in the literature (see for example [2, 3, 6, 14, 16]). First, while the RPB method uses the criterion (13) to decide whether to perform a serious or null iteration, the ones in references [3, 6, 14] use the condition \( \phi(x_j) \leq (1 - \gamma) \phi(x^c_{j-1}) + \gamma f_j(x_j) \), or equivalently,

\[
\phi(x^c_{j-1}) - \phi(x_j) \geq \frac{\gamma}{1 - \gamma} [f(x_j) - f_j(x_j)],
\]

which implies that \( \phi(x^c_{j-1}) \geq \phi(x_j) \) since \( f \geq f_j \). Moreover, the ones in [2, 16] use the condition \( \phi(x_j) \leq (1 - \gamma) \phi(x^c_{j-1}) + \gamma m_j \) where \( m_j \) is as in (2), or equivalently,

\[
\phi(x^c_{j-1}) - \phi(x_j) + \frac{\gamma}{2(1 - \gamma)} \|x_j - x^c_{j-1}\|^2 \geq \frac{\gamma}{1 - \gamma} [f(x_j) - f_j(x_j)].
\]

Note that proximal bundle variants based on (16), which is sometimes referred to as the descent condition, have the property that \( \{\phi(x^c_j)\} \) is a non-increasing sequence. It is worth noting that condition (13) is equivalent to

\[
\phi(\tilde{x}_j) - (f_j + h)(x_j) - \frac{1}{2\lambda} \|x_j - x^c_{j-1}\|^2 \leq -\frac{1}{2\lambda} \|\tilde{x}_j - x^c_{j-1}\|^2 + \delta,
\]

which, due to the definition of \( x_j \) in (2), implies that

\[
\phi(x^c_{j-1}) - \phi(\tilde{x}_j) \geq f(x^c_{j-1}) - f_j(x^c_{j-1}) + \frac{1}{2\lambda} \|\tilde{x}_j - x^c_{j-1}\|^2 - \delta,
\]

and hence that \( \phi(x^c_{j-1}) - \phi(x_j) \geq 0 \) when \( \tilde{x}_j = x_j \) and \( \delta = 0 \). However, we believe that the RPB method generally does not have the property that \( \{\phi(x^c_j)\} \) is a non-increasing sequence even when \( \delta = 0 \) due to the fact that the two points \( \tilde{x}_j \) and \( x_j \) (the latter of which is set to be \( x^c_j \) when (13) holds) may differ. We refer to (13) as a relaxed non-descent condition, which in some sense justifies the terminology we have chosen for the proximal bundle variant studied in this paper.

Second, at the end of a serious iteration index \( j \) of the RPB method, the set \( C_{j+1} \) is only required to satisfy \( C_{j+1} \supseteq \{x_j\} \), which allows for the possibility of completely refreshing it, i.e., setting it as \( C_{j+1} = \{x_j\} \). As far as we know, other proximal bundle variants studied in the literature only allow for the possibility of choosing a set \( C_{j+1} \) satisfying (14) with \( A_j \) chosen as either \( C_j \) (e.g., see [2, 16]), as in (15) (e.g., see [3]), or as a possibly smaller subset of (15) obtained by taking into consideration the positive dual variables of a reformulation of (2) (e.g., see [6, 14]).
Third, the complexity result established below is about the best (in terms of $\phi$) auxiliary serious iterate $\tilde{x}_j$ found so far. Moreover, auxiliary serious iterate $\tilde{x}_j$ in turn is the best (in terms of $\phi^\lambda_{j}$) among $x_j$, its previous serious iterate $x^c_{j-1}$ and all the null iterates generated between them. This is in contrast to the iteration-complexity analysis of [3, 6], which establish complexity bounds with respect to the best (in terms of $\phi$) serious iterate found so far.

The following result whose proof is given in Section 3 gives a bound on the number of null iterates between two consecutive serious iterates.

**Proposition 2.1.** Assume that $j = \ell_0$ is a serious iteration index of the RPB method. Then, the next serious iteration index $j = \ell_1$ must occur within

\[
O\left(\frac{M^2 \lambda}{\delta} + 1\right)
\]  

iterations of the RPB method (i.e., $\ell_1 - \ell_0$ is on the above order), where

\[
M := M_f + M_h.
\]  

We now state two main results, namely Theorems 2.2 (whose proof is given in Section 4) and 2.4, about the iteration-complexity of the RPB method. More specifically, the first (resp., second) result gives the complexity of RPB to find a $\bar{\varepsilon}$-optimal (resp., $(\hat{\rho}, \hat{\varepsilon})$-approximate) solution of (1). The bounds below are all expressed in terms of the distance

\[
d_0 := \inf\{\|x_0 - x_*\| : x_* \in X_*\}
\]

of the initial iterate $x_0$ to the set of optimal solutions $X_*$.  

**Theorem 2.2.** Let a prox stepsize $\lambda > 0$ and a tolerance $\bar{\varepsilon} > 0$ be given. Then, the following statements about the RPB method hold:

a) the number of serious iteration indices generated by the RPB method with $\delta = \bar{\varepsilon}/2$ until an auxiliary serious iterate $\tilde{x}_j$ satisfying (8) occurs is bounded by

\[
O\left(\frac{\bar{\varepsilon}^2}{\lambda \bar{\varepsilon}} + 1\right);
\]

b) if $\lambda$ is such that

\[
\Omega\left(\frac{\bar{\varepsilon}}{M^2}\right) = \lambda = O\left(\frac{\bar{\varepsilon}^2}{\bar{\varepsilon}}\right),
\]

then the total number of iterations performed by the RPB method with $\delta = \bar{\varepsilon}/2$ until it finds an auxiliary serious iterate $\tilde{x}_j$ satisfying (8) is bounded by

\[
O\left(\frac{M^2 \hat{\varepsilon}^2}{\bar{\varepsilon}^2}\right).
\]
iteration-complexity of the RPB method does not depend on \( \lambda \) as long as \( \lambda \) is chosen within the range specified in its statement.

The complexity bound (22) is in regards to the termination criterion (8). The following two results on the other hand establish the iteration-complexity for the RPB method to find a \((\hat{\rho}, \hat{\varepsilon})\)-approximate solution, i.e., one satisfying (10). The first result whose proof is given in Section 4 describes the convergence rate of a certain sequence of serious triple \((\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)\) generated by the RPB method.

**Proposition 2.3.** Let \( \{j_k : k \geq 0\} \) denote the sequence of serious iteration indices generated by the RPB method (and hence \( j_0 = 0 \)), and define \( z_0 = x_0 \) and for every \( k \geq 1 \), where

\[
\delta_k := \phi(\hat{z}_k) - (\tilde{f}_k + h)(z_k) - \frac{1}{2\lambda} \|z_k - z_{k-1}\|^2.
\]

Then, for every \( k \geq 1 \), we have

\[
\hat{v}_k \in \partial_{\hat{\varepsilon}_k} \phi(\hat{z}_k),
\]

and the residual pair \((\hat{v}_k, \hat{\varepsilon}_k)\) is bounded by

\[
\|\hat{v}_k\| \leq \frac{2d_0}{\lambda k} + \frac{\sqrt{2\delta}}{\sqrt{\lambda k}}, \quad 0 \leq \hat{\varepsilon}_k \leq \frac{5\delta}{2} \left( 1 + \frac{1}{\sqrt{k}} + \frac{2}{5k} \right) + \frac{3d_0^2}{2\lambda k^{3/2}} \left( 1 + \frac{3\sqrt{k}}{2} \right)
\]

where \( d_0 \) is as in (19) and \( \delta \) is as in step 0 of the RPB method.

We now make some remarks about the above result. First, the sequence \( \{\hat{\varepsilon}_k\} \) can be made arbitrarily small, say \( \hat{\varepsilon}_k \leq \hat{\varepsilon} \), for sufficiently large \( k \), as long as \( \delta \) is chosen in \((0, 2\hat{\varepsilon}/5)\). Second, Proposition 2.3 and the previous remark ensure that RPB is able to generate a sequence \( \{\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k\} \) satisfying the two properties in the paragraph below (10). As a consequence, (10) can be used as a suitable criterion to terminate RPB (see the discussion in the last paragraph of Subsection 2.1).

We are now ready to describe the iteration-complexity for the RPB method to find a \((\hat{\rho}, \hat{\varepsilon})\)-approximate solution of (1).

**Theorem 2.4.** Let a prox stepsize \( \lambda > 0 \) and a tolerance pair \((\hat{\rho}, \hat{\varepsilon}) \in \mathbb{R}^2_{++} \) be given. Then, the following statements about the RPB method hold:

a) the number of serious iteration indices generated by the RPB method with \( \delta = \hat{\varepsilon}/3 \) until a serious triple \((\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)\) satisfying (10) occurs is bounded by

\[
\mathcal{O} \left( \max \left\{ \frac{\hat{\varepsilon}}{\lambda \hat{\rho}^2}, \frac{d_0^2}{\lambda \hat{\varepsilon}} \right\} + 1 \right);
\]
b) if \( \lambda \) is such that \( \lambda = \Omega \left( \varepsilon / M^2 \right) \) and \( \lambda = \mathcal{O} \left( \max \left\{ \varepsilon / \rho^2, d_0^2 / \varepsilon \right\} \right) \), then the total number of iterations performed by the RPB method with \( \delta = \hat{\varepsilon} / 3 \) until it finds a serious triple \((\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)\) satisfying (10) is bounded by

\[
\mathcal{O} \left( \max \left\{ M^2 / \rho^2, M^2 d_0^2 / \varepsilon^2 \right\} \right).
\]

**Proof:** a) It follows from (27) that every serious triple \((\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)\) satisfies the inclusion in (10). Moreover, using the assumption that \( \delta = \hat{\varepsilon} / 3 \) and (28), we easily see that every serious triple \((\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)\) with index \( k \) such that

\[
k \geq \max \left\{ \frac{4d_0}{\lambda \rho}, \frac{8\varepsilon}{3\lambda \rho^2}, \frac{243d_0^2}{\lambda \varepsilon}, \left( \frac{162d_0^2}{\lambda \varepsilon} \right)^{2/3}, 36 \right\}
\]

must satisfy the two inequalities in (10). Hence, statement a) follows.

b) This statement immediately follows from Proposition 2.1, statement a) and the assumptions on \( \lambda \).

The following result gives the iteration-complexity for the RPB method to find a pair \((x, \varepsilon) = (\hat{z}_k, \hat{\eta}_k)\) satisfying (9) for the case in which \( \text{dom} \ h \) is bounded. Observe that the major difference between the result below and Theorem 2.2 is that the one below provides a certificate \( \varepsilon \) of the near optimality of \( x \) (see the discussion following (9)) while Theorem 2.2 does not.

**Corollary 2.5.** Assume that \( \Omega \subset \mathbb{R}^n \) is a compact convex set containing \( \text{dom} \ h \) and, for some given tolerance \( \varepsilon > 0 \), consider the RPB method with inner tolerance \( \delta = \hat{\varepsilon} / 6 \) and with prox stepsize \( \lambda \) satisfying \( \lambda = \Omega \left( \varepsilon / M^2 \right) \) and \( \lambda = \mathcal{O} \left( D_\Omega^2 / \varepsilon \right) \) where \( D_\Omega := \sup \{ \| u - u' \| : u, u' \in \Omega \} \). Moreover, consider the sequence of triples \((\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)\) obtained according to (24) and (25) and, for every \( k \geq 1 \), define \( \hat{\eta}_k \) as in (11). Then, the overall iteration-complexity of the RPB method until it finds a pair \((\hat{z}_k, \hat{\eta}_k)\) satisfying (9) is bounded by

\[
\mathcal{O} \left( \frac{M^2 D_\Omega^2}{\varepsilon^2} \right).
\]

**Proof:** Using Theorem 2.4(b) with the tolerance pair \((\hat{\rho}, \hat{\varepsilon}) = (\varepsilon/(2D_\Omega), \varepsilon/2)\), we conclude that the overall iteration-complexity of RPB with \( \delta = \hat{\varepsilon} / 6 \) until it finds a triple \((\hat{z}_k, \hat{v}_k, \hat{\varepsilon}_k)\) satisfying (10) is bounded by (29). The inclusion in (10) and the remarks following (11) then imply that the pair \((\hat{z}_k, \hat{\eta}_k)\) satisfies the inclusion in (9). Moreover, it follows from the definition of \( \hat{\eta}_k \) in (11), the Cauchy-Schwarz inequality, and the facts that \( \| \hat{v}_k \| \leq \hat{\rho} = \varepsilon / (2D_\Omega) \) and \( \hat{\varepsilon}_k \leq \hat{\varepsilon} = \varepsilon / 2 \) (see (10)), that \( \hat{\eta}_k \) satisfies the inequality in (9). We have thus shown the corollary.

**3 Analysis of null iterates**

This section provides the proof of Proposition 2.1, which gives a bound on the null iterates between two consecutive serious iterates. It also presents a result showing that the composite subgradient method can be viewed as a special instance of the RPB method.

We assume throughout this section that \( \ell_0 \) 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 \( \ell_1 \) (if any) and all null iteration indices between \( \ell_0 \) and \( \ell_1 \). In terms of this notation, we easily see that Proposition 2.1 is equivalent to show that \( \ell_1 \) is such that \( \ell_1 - \ell_0 \) is bounded by (17), and hence that \( \ell_1 \) will indeed occur.
We start by making some simple observations that immediately follow from the description of the RPB method. Using the definition of \( x_j^c \) in step 2 of the RPB method, we immediately see that for every \( j \in B(\ell_0) \),

\[
x_{j-1}^c = x_{\ell_0},
\]

and hence in view of the definitions of \( \phi_j^\lambda \) and \( x_j \) in (12) and (2), respectively, we have

\[
\phi_j^\lambda = \phi + \frac{1}{2\lambda} \| -x_{\ell_0} \|^2, \tag{30}
\]

\[
(x_j, m_j) = \arg \min_{u \in \mathbb{R}^n} \left\{ \phi_j^\lambda(u) := f_j(u) + h(u) + \frac{1}{2\lambda} \| u - x_{\ell_0} \|^2 \right\}, \tag{31}
\]

with the understanding that \( x_j \) (resp., \( m_j \)) is the optimal solution (resp., value) of (31).

We now make a few observations in connection to the above notation. First, it follows from (31) that

\[
m_j = \phi_j^\lambda(x_j), \quad \frac{1}{\lambda}(x_{\ell_0} - x_j) \in \partial(f_j + h)(x_j). \tag{32}
\]

Second, (30) implies that the function \( \phi_j^\lambda \) remains the same whenever \( j \in B(\ell_0) \). Since \( \ell_0 \) remains fixed for the analysis in this section, we simply denote the function \( \phi_j^\lambda \) for \( j \in B(\ell_0) \) by \( \phi^\lambda \), i.e.,

\[
\phi^\lambda = \phi_j^\lambda \quad \forall j \in B(\ell_0). \tag{33}
\]

Third, in view of the definition of \( \tilde{x}_j \) in step 1 of the RPB method and the above relation, it then follows that

\[
\tilde{x}_j = \arg \min \left\{ \phi^\lambda(x) : x \in \{x_{\ell_0}\} \cup \{x_i : i \in B(\ell_0), i \leq j\} \right\}. \tag{34}
\]

Fourth, \( \ell_1 \) is characterized as the first index \( j > \ell_0 \) satisfying the null iteration termination criterion (13). It will be shown below that the sequence \( \{t_j : j \in B(\ell_0)\} \) is non-increasing (see Lemma 3.4(b)) and converges to zero with an \( O(1/j) \) convergence rate (see Proposition 3.6).

For the sake of generality, we state all the results below in terms of \( \ell_0 \). However, we assume in their proofs that \( \ell_0 = 0 \) in order to keep their notation and formulae simple.

**Lemma 3.1.** The following statements hold for the RPB method:

a) for every \( j \in B(\ell_0) \), we have

\[
f \geq f_j, \quad \phi^\lambda \geq \overline{\phi}_j^\lambda; \tag{35}
\]

b) for every \( i, j \in B(\ell_0) \) such that \( j > i \), we have

\[
\overline{\phi}_j^\lambda(\tilde{x}_j) \leq \overline{\phi}_i^\lambda(\tilde{x}_i); \tag{36}
\]

c) for every \( j \in B(\ell_0) \), we have

\[
\phi^\lambda(x_{j-1}) = \overline{\phi}_j^\lambda(x_{j-1}). \tag{37}
\]

**Proof:** a) The first inequality in (35) simply follows from (3) and the convexity of \( f \), and it together with the definitions of \( \phi^\lambda \) and \( \overline{\phi}_j^\lambda \) in (30) and (31), respectively, implies the second inequality in (35).

b) The inequality immediately follows from relation (34).

c) In view of (14), it is obvious that \( x_{j-1} \in C_j \), and by the definition of \( f_j \) in (3), we conclude \( f_j(x_{j-1}) = f(x_{j-1}) \). Therefore, in view of (30) and (31), the conclusion (37) follows.

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 3.2. Define
\[ f_{A_j} := \max \{ f(x_i) + \langle g_i, x_i - x \rangle : x_i \in A_j, g_i \in \partial f(x_i) \} \quad \forall j \in B(\ell_0) \] (38)
where \( A_j \) is as defined in (15). 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) \(f_{A_j} \leq \min\{f_j, f_{j+1}\}\);

c) we have
\[ (x_j, m_j) = \arg\min_{u \in \text{dom} h} \left\{ (f_{A_j} + h)(u) + \frac{1}{2\lambda} \| u - x_{\ell_0} \|^2 \right\}; \] (39)

d) for every \( u \in \text{dom} h \), we have
\[ (f_{A_j} + h)(u) + \frac{1}{2\lambda} \| u - x_{\ell_0} \|^2 \geq m_j + \frac{1}{2\lambda} \| u - x_j \|^2. \]

Proof: Recall that even though the lemma is stated for a general serious iteration index \( \ell_0 \), we assume in its proof that \( \ell_0 = 0 \) in order to keep the notation simple.

a) The first conclusion immediately follows from the definitions of \( A_j, f_j \) and \( f_{A_j} \) in (15), (3) and (38), respectively. It follows from the definition of \( A_j \), the definition of \( f_j \) in (3), and a well-known formula for the subdifferential of the pointwise maximum of finitely many affine functions (e.g., see Example 3.4 of [15]), that \( \partial f_j(x_j) \) is the convex hull of \( \cup \{ g_i : x_i \in A_j \} \). Using the same reasoning but with (3) replaced by (38), we conclude that the latter set is also the subdifferential of \( f_{A_j} \) at \( x_j \). Hence, statement a) follows.

b) In view of (14), it is clear that \( A_j \subset C_j \) and \( A_j \subset C_{j+1} \), and hence that this statement follows from the definitions of \( f_j \) and \( f_{A_j} \) in (3) and (38), respectively.

c) It follows from the inclusion in (32) and the second identity in a) that
\[ \frac{1}{\lambda}(x_0 - x_j) \in \partial(f_j + h)(x_j) = \partial(f_{A_j} + h)(x_j). \]
Using the identity in (32), the first identity in a), and the fact that the above inclusion implies that \( x_j \) satisfies the optimality condition of (39), we conclude that c) holds.

d) This statement follows immediately from c) and the fact that the objective function of (39) is \((1/\lambda)\)-strongly convex.

The following lemma provides a few key facts involving the Lipschitz continuity of the objective function.

Lemma 3.3. Consider the RPB method and let \( j \in B(\ell_0) \) be given. Then, the following statements hold:

a) \( f_j + h \) is \( M \)-Lipschitz continuous on \( \text{dom} h \) where \( M \) is as in (18);

b) \( \| x_{\ell_0} - x_j \| \leq \lambda M \);

c) for every \( u \in \text{dom} h \), we have
\[ \phi_j^\lambda(u) - m_j \leq 2M \| u - x_j \| + \frac{1}{2\lambda} \| u - x_j \|^2. \]
Proof: Recall that even though the lemma is stated for a general serious iteration index \( \ell_0 \), we assume in its proof that \( \ell_0 = 0 \) in order to keep the notation simple.

a) It follows from (6), the definition of \( f_j \) in (3), and a well-known formula for the subdifferential of the pointwise maximum of finitely many affine functions (e.g., see Example 3.4 of [15]) that \( f_j \) is \( M \)-Lipschitz continuous on \( \text{dom } h \), which together with (A1) and the definition of \( M \) in (18) proves this statement.

b) Using the inclusion in (32), statement a) and the definition of the subdifferential in the line below (4), we have

\[
\frac{1}{\lambda} \| x_0 - x_j \|^2 \leq (f_j + h)(x_0) - (f_j + h)(x_j) \leq M \| x_0 - x_j \|,
\]

and hence this statement holds.

c) Using (31), the equality in (32), statement a), and Cauchy-Schwarz inequality, we have

\[
\phi_j^\lambda(u) - m_j = (f_j + h)(u) - (f_j + h)(x_j) + \frac{1}{2\lambda} \| u - x_0 \|^2 - \frac{1}{2\lambda} \| x_j - x_0 \|^2 \\
\leq M \| u - x_j \| + \frac{1}{2\lambda} \| u - x_j \|^2 + \frac{1}{\lambda} \langle x_j - x_0, u - x_j \rangle \\
\leq M \| u - x_j \| + \frac{1}{2\lambda} \| u - x_j \|^2 + \frac{1}{\lambda} \| x_j - x_0 \| \| u - x_j \|.
\]

Statement c) now follows from the above inequality and b).

The following lemma presents a few technical results about \( t_j \) that will play important roles in the estimation of the cardinality of the set \( B(\ell_0) \).

Lemma 3.4. The following statements hold for the RPB method:

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

\[
t_i \geq m_j - m_i \geq \frac{1}{2\lambda} \sum_{l=i+1}^{j} \| x_l - x_{l-1} \|^2;
\]

(40)

b) \( \{ t_j : j \in B(\ell_0) \} \) is non-increasing and \( t_j \leq 2\lambda M^2 \) for every \( j \in B(\ell_0) \).

Proof: Recall that even though the lemma is stated for a general serious iteration index \( \ell_0 \), we assume in its proof that \( \ell_0 = 0 \) in order to keep the notation simple.

a) Using the definition of \( m_{j+1} \) in (31), and statements b) and d) with \( u = x_{j+1} \) of Lemma 3.2, we conclude that

\[
m_{j+1} = (f_{j+1} + h)(x_{j+1}) + \frac{1}{2\lambda} \| x_{j+1} - x_0 \|^2 \\
\geq (f_{A_j} + h)(x_{j+1}) + \frac{1}{2\lambda} \| x_{j+1} - x_0 \|^2 \geq m_j + \frac{1}{2\lambda} \| x_{j+1} - x_j \|^2.
\]

The second inequality in (40) now follows by summing the above inequality from \( j = i \) to \( j = j - 1 \).

Using (31), (35), (36), and the assumption that \( i, j \in B(\ell_0) \) and \( j > i \), we conclude that

\[
m_j = \min_{u \in \text{dom } h} \phi_j^\lambda(u) \leq \min_{u \in \text{dom } h} \phi_j^\lambda(u) \leq \phi_j^\lambda(\check{x}_j) \leq \phi_j^\lambda(\check{x}_i),
\]

and hence that the first inequality in (40) holds in view of (33) and the definition of \( t_j \) in (13).
b) It immediately follows from (40) that $m_j \geq m_i$, which together with (36) and the definition of $t_j$ in (13) implies that $\{t_j\}$ is non-increasing. In order to show the second conclusion in b), it then suffices to show that $t_1 \leq 2\lambda M^2$. Indeed, it follows from the definitions of $\tilde{x}_j$, $\phi^\lambda$ and $m_j$ in (34), (33) and (31), respectively, that

$$
\phi^\lambda(\tilde{x}_1) \leq (f + h)(x_1) + \frac{1}{2\lambda} \|x_1 - x_0\|^2,
$$

$$
m_1 = (f_1 + h)(x_1) + \frac{1}{2\lambda} \|x_1 - x_0\|^2 \geq f(x_0) + \langle g_0, x_1 - x_0 \rangle + h(x_1) + \frac{1}{2\lambda} \|x_1 - x_0\|^2
$$

where $g_0$ is as in step 3 of the RPB method. Using the above two relations and the definition of $t_j$ in (13), we have

$$
t_1 \leq f(x_1) - f(x_0) - \langle g_0, x_1 - x_0 \rangle
$$

$$
\leq |f(x_1) - f(x_0)| + \|g_0\| \|x_1 - x_0\| \leq 2Mf \|x_1 - x_0\|
$$

where the last inequality is due to (6) and (7). The above claim now follows by combining the above inequality and Lemma 3.3(b).

The following technical result connects $\tilde{t}_j$ to the minimum of distance between two consecutive null iterates up to the $j$-th iteration, i.e., $\Delta_j$ defined below, and it will in turn give a recursive formula in terms of $\tilde{t}_j$ and $t_{j/2}$, which will be useful in the proof of the next proposition.

**Lemma 3.5.** Let

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

Then, the following statements hold:

a) for every $j \in B(\ell_0)$, we have

$$
t_j \leq 2M \Delta_j + \frac{1}{2\lambda} \Delta_j^2;
$$

b) for every $j \in B(\ell_0)$, we have

$$
\frac{j^2 - 2j}{16\lambda} \Delta_j^2 \leq 2M \sqrt{2[j/2] \lambda t_{\lfloor j/2 \rfloor} + t_{\lfloor j/2 \rfloor}}.
$$

**Proof:** Recall that even though the lemma is stated for a general serious iteration index $\ell_0$, we assume in its proof that $\ell_0 = 0$ in order to keep the notation simple.

a) Let $j \in B(\ell_0)$ be given. It directly follows from Lemma 3.3(c) with $u = x_{j-1}$ and Lemma 3.1(c)

$$
\phi^\lambda(x_{j-1}) - m_j \leq 2M \|x_j - x_{j-1}\| + \frac{1}{2\lambda} \|x_j - x_{j-1}\|^2.
$$

Using the first conclusion in Lemma 3.4(b), the definitions of $t_j$ and $\tilde{x}_j$ in (13) and (34), respectively, and the inequality above with $j = i$, we conclude for every $i \in B(\ell_0)$ such that $i \leq j$ that

$$
t_j \leq t_i \leq \phi^\lambda(x_{i-1}) - m_i \leq 2M \|x_i - x_{i-1}\| + \frac{1}{2\lambda} \|x_i - x_{i-1}\|^2.
$$

The inequality (42) now follows from the above inequality and the definition of $\Delta_j$ in (41).
b) Using Lemma 3.4(a), relation (43) with $j = i$, and the definition of $\Delta_j$ in (41), we conclude that
\[
\frac{1}{2\lambda}(j-i)\Delta_j^2 \leq \frac{1}{2\lambda} \sum_{t=|j/2|+1}^{j} \|x_t - x_{t-1}\|^2 \leq t_i \leq 2M \|x_i - x_{i-1}\| + \frac{1}{2\lambda} \|x_i - x_{i-1}\|^2.
\]
Summing the above inequality from $i = |j/2| + 1$ to $i = j$, we obtain
\[
\frac{1}{4\lambda}(j/2)(j/2 - 1)\Delta_j^2 \leq 2M \sum_{i=|j/2|+1}^{j} \|x_i - x_{i-1}\| + \frac{1}{2\lambda} \sum_{i=|j/2|+1}^{j} \|x_i - x_{i-1}\|^2.
\]
On the other hand, it follows from Lemma 3.4(a) and Cauchy-Schwarz inequality that
\[
\sum_{i=|j/2|+1}^{j} \|x_i - x_{i-1}\| \leq \left( \left\lfloor \frac{j}{2} \right\rfloor \sum_{t=|j/2|+1}^{j} \|x_t - x_{t-1}\|^2 \right)^{1/2} \leq \sqrt{2|j/2|\lambda |j/2|}.
\]
Plugging (40) with $i = |j/2|$ and the above inequality into (44), we conclude that b) holds.

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

**Proposition 3.6.** Define $a := 128(3)^{2/3}M^2\lambda$. Then, for every $j \in B(\ell_0)$, we have
\[
t_j \leq \frac{a}{j - \ell_0}.
\]

**Proof:** Recall that even though the lemma is stated for a general serious iteration index $\ell_0$, we assume in its proof that $\ell_0 = 0$ in order to keep the notation simple.

The proof of the lemma is by induction. We first show that (45) holds for every $j \in B(\ell_0)$ such that $j \leq 11$. Indeed, it follows from the second conclusion in Lemma 3.4(b) and the definition of $a$ that $t_j \leq a/j$ for $j \in \{1, \ldots, 11\}$. Now, let $j \in B(\ell_0)$ be such that $j \geq 12$ and assume for simplicity and without any loss of generality that $j$ is an even number. For the induction argument, assume also that (45) holds for every $i \leq j - 1$. Using Lemma 3.5(b) and the induction hypothesis for $j/2$, i.e., $t_{j/2} \leq 2a/j$, we have for $j \geq 12$,
\[
\frac{j^2}{24\lambda} \Delta_j^2 \leq 2M\sqrt{2\lambda a} + \frac{2a}{j}.
\]
Plugging the above inequality into (42), we have for $j \geq 12$,
\[
t_j \leq \frac{8\sqrt{3}M^{3/2}\lambda^{3/4}(2a)^{1/4}}{j} + \frac{8M\sqrt{3}\lambda}{j^{3/2}} + \frac{24M\sqrt{2\lambda a}}{j^2} + \frac{24a}{j^3}.
\]
Using the definition of $a$ in the statement of the proposition and the fact that $j \geq 12$, we can easily see that
\[
\frac{8\sqrt{3}M^{3/2}\lambda^{3/4}(2a)^{1/4}}{j} \leq \frac{a}{4j}, \quad \frac{8M\sqrt{3}\lambda}{j^{3/2}} \leq \frac{a}{4j}, \quad \frac{24M\sqrt{2\lambda a}}{j^2} \leq \frac{a}{4j}, \quad \frac{24a}{j^3} \leq \frac{a}{4j}.
\]
Therefore, we conclude from (46) and (47) that $t_j \leq a/j$. We have thus shown that the conclusion of the proposition holds.

We are now ready to give the proof of Proposition 2.1.
Proof of Proposition 2.1 As in the preceding proofs in this section, without loss of generality, we assume $\ell_0 = 0$, and hence we have $t_j \leq a/j$ for every $j \in B(\ell_0)$ in view of Proposition 3.6. Now the conclusion of the proposition immediately follows from the last observation and the fact that an iteration index is serious if the stopping criterion (13) holds.

Before ending this section, it is worth pointing out the relationship between the RPB method and the composite subgradient method with constant prox stepsize. In view of Proposition 2.1, it is natural to conjecture whether the RPB method with small prox stepsize reduces to the composite subgradient method with constant prox stepsize. The following result shows that this is indeed the case.

**Proposition 3.7.** If $\lambda = \delta/(2M^2)$, then every iteration index 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), then the RPB method reduces to the composite subgradient method with constant prox stepsize $\lambda$.

**Proof:** As in the preceding proofs in this section, without loss of generality, we assume $\ell_0 = 0$. It follows from the assumption that $\lambda = \delta/(2M^2)$ and Lemma 3.4(b) that $t_j \leq \delta$ for every $j \in B(\ell_0)$. Hence, we have $t_1 \leq \delta$, and in view of (13), we conclude that every iteration index $j$ is serious. We now show that, under the assumptions of the proposition, the RPB method reduces to the composite subgradient method with constant prox stepsize $\lambda$. Since every iteration index is a serious one, it follows from step 2.a) of the RPB method, the definition of $f_j$ in (3), and the assumption that $C_{j+1} = \{x_j\}$, that $x^c_j = x_j$ and $f_j(\cdot) = f(x_{j-1}) + \langle g_{j-1}, \cdot - x_{j-1} \rangle$ for every $j \geq 1$. Using this observation, it is now easy to see that RPB method reduces to the composite subgradient method with constant prox stepsize $\lambda = \delta/(2M^2)$.

4 Analysis of serious iterates

This section provides the proofs of Theorem 2.2 and Proposition 2.3 that deal with the overall iteration-complexity of the RPB method for finding the two types of approximate solutions introduced in (8) and (10). Since the analysis of the number of null iterates between any two consecutive serious iterates has already been given in the previous section, this section focuses on the derivation of bounds on the overall number of serious iterates generated by the RPB method.

Recall from Proposition 2.3 that $\{j_k : k \geq 0\}$ denotes the sequence of serious iteration indices generated by the RPB method. From the description of the RPB method, namely, relations (2) and (13), and the definitions of $z_k$, $\tilde{z}_k$ and $\tilde{f}_k$ in (23) and of $\delta_k$ in (26), we have that for every $k \geq 1$,

$$x_j^c = z_{k-1}, \quad \forall j = j_{k-1}, \ldots, j_k - 1,$$

$$z_k = \arg\min_{u \in \mathbb{R}^n} \left( \tilde{f}_k + h(u) + \frac{1}{2\lambda} ||u - z_{k-1}||^2 \right),$$

$$\delta_k \leq -\frac{1}{2\lambda} ||\tilde{z}_k - z_{k-1}||^2 + \delta.$$

The following lemma gives a basic recursive formula that is the starting point for the serious iteration complexity analysis.

**Lemma 4.1.** For $k \geq 1$ and every $z \in \text{dom } f$, we have

$$\phi(\tilde{z}_k) - (\tilde{f}_k + h)(z) \leq \delta_k + \frac{1}{2\lambda} (||z - z_{k-1}||^2 - ||z - z_k||^2)$$

(50)
where \( z_k, \tilde{z}_k \) and \( \tilde{f}_k \) are as defined in (23), and \( \delta_k \) is as defined in (26). As a consequence, we have
\[
\|\tilde{z}_k - z_k\|^2 \leq 2\lambda\delta.
\]  

(51)

**Proof:** Noting that the objective function of (48) is \( 1/\lambda \)-strongly convex, it follows from (48) that for every \( k \geq 1 \) and \( z \in \mathbb{R}^n \),
\[
(f_k + h)(z) + \frac{1}{2\lambda}\|z - z_{k-1}\|^2 \leq (f_k + h)(z_k) + \frac{1}{2\lambda}\|z - z_{k-1}\|^2 - \frac{1}{2\lambda}\|z - z_k\|^2
\]
and hence that
\[
\phi(\tilde{z}_k) - (\tilde{f}_k + h)(z) \leq \phi(\tilde{z}_k) - (\tilde{f}_k + h)(z_k) - \frac{1}{2\lambda}\|z - z_{k-1}\|^2 + \frac{1}{2\lambda}\|z - z_{k-1}\|^2 - \frac{1}{2\lambda}\|z - z_k\|^2,
\]
from which (50) follows due to the definition of \( \delta_k \) in (26). Moreover, the inequality (51) follows from (50) with \( z = \tilde{z}_k \), (49) and the fact that \( f \geq \tilde{f}_k \).

The following result gives a bound that follows by aggregating the first \( k \) bounds in Lemma 4.1.

**Lemma 4.2.** For every \( k \geq k_0 \geq 1 \), consider \( \tilde{f}_k \) as in (23) and define
\[
\Gamma_{k_0,k} := \frac{1}{k - k_0} \sum_{i=k_0+1}^{k} (\tilde{f}_i + h).
\]

(52)

Then, for every \( k \geq k_0 \geq 0 \) and \( z \in \text{dom } h \), we have \( \Gamma_{k_0,k} \leq \phi \) and
\[
\phi(\tilde{z}_k) - \Gamma_{k_0,k}(z) \leq \frac{1}{k - k_0} \sum_{i=k_0+1}^{k} \delta_i + \frac{1}{2\lambda(k - k_0)} (\|z - z_{k_0}\|^2 - \|z - z_k\|^2)
\]

(53)

where \( z_k, \tilde{z}_k \) and \( \delta_k \) are as defined in (23), (24) and (26), respectively.

**Proof:** First, using the fact that \( f \geq \tilde{f}_i \) for every \( i \geq 1 \) and the definitions of \( \phi \) and \( \Gamma_{k_0,k} \) in (1) and (52), respectively, we have \( \Gamma_{k_0,k} \leq \phi \). Summing (50) from \( k = k_0 + 1 \) to \( k = k \) and dividing the resulting inequality by \( k - k_0 \), we have
\[
\frac{1}{k - k_0} \sum_{i=k_0+1}^{k} (\phi(\tilde{z}_i) - (\tilde{f}_i + h)(z)) \leq \frac{1}{k - k_0} \sum_{i=k_0+1}^{k} \delta_i + \frac{1}{2\lambda(k - k_0)} (\|z - z_{k_0}\|^2 - \|z - z_k\|^2),
\]
which, together with the definitions of \( \tilde{z}_k \) and \( \Gamma_{k_0,k} \) in (24) and (52), respectively, immediately implies (53).

The following result provides a bound on the primal gap \( \phi(\tilde{z}_k) - \phi_* \) from which the bound on the number of serious iteration indices stated in Theorem 2.2(a) will immediately follow.

**Lemma 4.3.** Let \( d_0 \) be defined as in (19) and let \( z_* \) be the unique point in \( X_* \) such that \( d_0 = \|z_0 - z_*\| \). Then, for every \( k \geq 1 \), we have
\[
\phi(\tilde{z}_k) - \phi_* + \frac{1}{2\lambda k} \sum_{i=1}^{k} \|\tilde{z}_i - z_{i-1}\|^2 \leq \delta + \frac{1}{2\lambda k} (d_0^2 - \|z_k - z_*\|^2)
\]

(54)

where \( z_k \) and \( \tilde{z}_k \) are as defined in (23), \( \tilde{z}_k \) is as defined in (24) and \( \delta \) is as in step 0 of the RPB method. As a consequence, we have
\[
\phi(\tilde{z}_k) - \phi_* \leq \delta + \frac{d_0^2}{2\lambda k},
\]

(55)

\[
\|z_k - z_*\|^2 \leq 2k\lambda\delta + d_0^2.
\]

(56)
Proof: Using (53) with $z = z_*$ and $k_0 = 0$, relation (49), and the fact that $\Gamma_{k_0,k} \leq \phi$, we conclude that (54) holds. Inequalities (55) and (56) immediately follow from (54) and the fact that $\phi(\hat{z}_k) \geq \phi_*$.  

We are now ready to prove Theorem 2.2.

Proof of Theorem 2.2 a) Using the assumption that $\delta = \bar{\varepsilon}/2$ and (55), we easily see that every serious auxiliary iterate $\hat{z}_k$ with an index $k$ such that $k \geq d_0^2/(\lambda \bar{\varepsilon})$ must satisfy (8). Hence, statement a) follows in view of the definitions of $\hat{z}_k$ and $\hat{z}_k$ in (23) and (24), respectively.

b) This statement immediately follows from Proposition 2.1, statement a) and the assumptions on $\lambda$ in (21).

The following lemma states some bounds on the magnitude of the sequences $\{z_k\}$ and $\{\hat{z}_k\}$ which are used in the proof of Proposition 2.3 to obtain convergence rate bounds on the sequence of residual pairs $(\hat{v}_k, \hat{z}_k)$. Recall that the latter bounds have been used in the proof of Theorem 2.4 to establish the second iteration-complexity for RPB to obtain a $(\bar{\rho}, \bar{\varepsilon})$-approximate solution.

Lemma 4.4. For every $k \geq 1$, we have

$$\|z_k - z_0\| \leq \sqrt{2k\lambda\delta} + 2d_0, \tag{57}$$

$$\|\hat{z}_k - z_0\|^2 \leq 2\lambda\delta + 5\sqrt{k}\lambda\bar{\varepsilon} + 3k\lambda\delta + \frac{3d_0^2}{\sqrt{k}} + \frac{9d_0^2}{2}, \tag{58}$$

where $d_0$, $z_k$ and $\hat{z}_k$ are as defined in (19), (23) and (24), respectively, and $\delta$ is as in step 0 of the RPB method.

Proof: Let $z_*$ be the unique point in $X_*$ such that $d_0 = \|z_0 - z_*\|$. It then follows from (56) and this choice of $z_*$ that

$$\|z_k - z_0\| \leq \|z_k - z_*\| + \|z_0 - z_*\| \leq \sqrt{2k\lambda\delta} + 2d_0,$$

and hence that (57) holds. Using Cauchy-Schwarz inequality, relations (51) and (56), and the fact that $d_0 = \|z_0 - z_*\|$, we conclude that for every $k \geq 1$,

$$\|\hat{z}_k - z_0\|^2 \leq \left( \frac{1}{\sqrt{k}} + 1 + \frac{1}{2} \right) \left( \sqrt{k}\|\hat{z}_k - z_k\|^2 + \|z_k - z_*\|^2 + 2\|z_* - z_0\|^2 \right)$$

$$\leq \left( \frac{1}{\sqrt{k}} + \frac{3}{2} \right) \left( 2\sqrt{k}\lambda\delta + 2k\lambda\delta + 3d_0^2 \right)$$

$$= 2\lambda\delta + 5\sqrt{k}\lambda\bar{\varepsilon} + 3k\lambda\delta + \frac{3d_0^2}{\sqrt{k}} + \frac{9d_0^2}{2}.$$

Inequality (58) immediately follows from the above inequality and the fact that $\hat{z}_k$ is one of the points $\hat{z}_i$ for some $i \leq k$ in view of (24).

We now make a remark about the above result. Bound (58) 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 proof of Proposition 2.3, the smaller this constant is, the closer $\delta$ can be chosen to the tolerance $\bar{\varepsilon}$.

We are now ready to prove the Proposition 2.3.

Proof of Proposition 2.3 First, we show (27) holds for every $k \geq 1$. Using the obvious identity

$$\|z - z_0\|^2 - \|z - z_k\|^2 = \|\hat{z}_k - z_0\|^2 - \|\hat{z}_k - z_k\|^2 + 2\langle z_0 - z_k, \hat{z}_k - z \rangle \quad \forall z \in \mathbb{R}^n,$$
inequality (53) with $k_0 = 0$, the fact that $\phi \geq \Gamma_{k_0,k}$, and the definitions of $\hat{v}_k$ and $\hat{z}_k$ in (24) and (25), respectively, we conclude that for every $z \in \text{dom } f$, 

$$
\phi(\hat{z}_k) - \phi(z) \leq \frac{1}{k} \sum_{i=z}^{k} \delta_i + \frac{1}{2\lambda k} (\|\hat{z}_k - z_0\|^2 - \|\hat{z}_k - z_k\|^2 + 2\langle z_0 - z_k, \hat{z}_k - z \rangle) = \hat{\varepsilon}_k + \langle \hat{v}_k, \hat{z}_k - z \rangle,
$$

and hence that (27) holds. Next, we show (28) holds for every $k \geq 1$. The first inequality in (28) follows from plugging (57) into the definition of $\hat{v}_k$ in (24). The first inequality for $\hat{\varepsilon}_k$, i.e. $\hat{\varepsilon}_k \geq 0$, follows from taking $z = \hat{z}_k$ in (59). Using the definition of $\hat{\varepsilon}_k$ in (25), and relations (49) and (58), we obtain the second inequality for $\hat{\varepsilon}_k$.

\section{Analysis of RPB in the strongly convex setting}

This section derives an iteration-complexity bound for the RPB method to find an iterate satisfying (8) under the assumptions that $h$ is strongly convex and $\text{dom } h$ is bounded. It also provides a discussion of how the aforementioned bound can be alternatively used to derive the corresponding one in the convex setting, namely, bound (22). Finally, it discusses how the complexity bounds derived in this paper relate to those derived in [3] and [6].

In this section, in addition to assumptions (A1) and (A2) stated in Subsection 2.1, the following conditions are assumed to hold:

(B1) $h$ is $\mu$-strongly convex for some $\mu > 0$;

(B2) $\text{dom } h$ is bounded.

The following result is a stronger version of Lemma 4.3, which follows as a consequence of the stronger additional conditions (B1) and (B2) stated above.

\textbf{Lemma 5.1.} Let $d_0$ be defined as in (19) and $z_\ast$ be the unique point in $X_\ast$ such that $d_0 = \|z_0 - z_\ast\|$. Then, for every $k \geq 1$, we have

$$
\phi(\hat{z}_k) - \phi_\ast \leq \delta + \frac{2\delta}{\lambda \mu k} + \frac{d_0^2}{\lambda k(1 + \lambda \mu)^{[k/2]}},
$$

where $\hat{z}_k$ is as defined in (24) and $\delta$ is as in step 2 of the RPB method.

\textbf{Proof:} Noting that the objective function of (48) is $(\mu + 1/\lambda)$-strongly convex, it follows from (48) that for every $k \geq 1$ and $z \in \mathbb{R}^n$,

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

Taking $z = z_\ast$ in the above inequality, rearranging the terms, and using the fact that $\phi \geq \hat{f}_k + h$, the definition of $\delta_k$ in (26) and relation (49), we have

$$
\phi(\hat{z}_k) - \phi_\ast + \frac{1}{2} \left( \mu + \frac{1}{\lambda} \right) \|z_k - z_\ast\|^2 \leq \phi(\hat{z}_k) - (\hat{f}_k + h)(z_\ast) + \frac{1}{2} \left( \mu + \frac{1}{\lambda} \right) \|z_k - z_\ast\|^2
$$

$$
\leq \phi(\hat{z}_k) - (\hat{f}_k + h)(z_k) - \frac{1}{2\lambda} \|z_k - z_{k-1}\|^2 + \frac{1}{2\lambda} \|z_{k-1} - z_\ast\|^2
$$

$$
= \delta_k + \frac{1}{2\lambda} \|z_{k-1} - z_\ast\|^2 \leq \delta + \frac{1}{2\lambda} \|z_{k-1} - z_\ast\|^2
$$


which, in view of the fact that $\phi(\tilde{z}_k) \geq \phi_*$, then yields
\begin{equation}
\|z_k - z_*\|^2 \leq \frac{2\lambda}{1 + \lambda\mu} \delta + \frac{1}{1 + \lambda\mu} \|z_{k-1} - z_*\|^2.
\end{equation}

Using the fact that $d_0 = \|z_0 - z_*\|$, it is easy to see that the above inequality further implies that
\begin{equation}
\|z_k - z_*\|^2 \leq \frac{2\mu}{\lambda} \delta + \frac{d_0^2}{(1 + \lambda\mu)^{k/2}}.
\end{equation}

Using (53) with $z = z_*$ and $k_0 = \lfloor k/2 \rfloor$, relation (49), and the facts that $\Gamma_k \leq \phi$ and $\lfloor k/2 \rfloor \leq k/2$, we have
\begin{equation}
\phi(\hat{z}_k) - \phi_* \leq \frac{1}{\lambda\mu} \|z_{\lfloor k/2 \rfloor} - z_*\|^2,
\end{equation}

which, together with relation (61) with $k = \lfloor k/2 \rfloor$, immediately implies (60).

We are ready to present the main result of this section, which derives an iteration-complexity bound for the RPB method to find an iterate satisfying (8) under the additional assumptions (B1) and (B2).

**Theorem 5.2.** Let a prox stepsize $\lambda > 0$ and a tolerance $\bar{\varepsilon} > 0$ be given. Then, the following statements about the RPB method hold:

a) the number of serious iteration indices generated by the RPB method with $\delta = \bar{\varepsilon}/4$ until an auxiliary serious iterate $\hat{z}_k$ satisfying (8) occurs is bounded by
\begin{equation}
O \left( \frac{1}{\lambda\mu} \log_1^+ \left( \frac{\mu d_0^2}{\bar{\varepsilon}} \right) + 1 \right)
\end{equation}

where $\log_1^+ (\cdot)$ is as defined in Subsection 1.1;

b) the total number of iterations performed by the RPB method with $\delta = \bar{\varepsilon}/4$ until it finds an auxiliary serious iterate $\hat{z}_k$ satisfying (8) is bounded by
\begin{equation}
O \left( \left( \frac{M^2 \lambda}{\bar{\varepsilon}} + 1 \right) \left[ \frac{1}{\lambda\mu} \log_1^+ \left( \frac{\mu d_0^2}{\bar{\varepsilon}} \right) + 1 \right] \right).
\end{equation}

**Proof:** a) In view of the definition of auxiliary serious iterate in the paragraph below the description of the RPB method, we know by the definition of $\tilde{z}_k$ in (24) that $\hat{z}_k$ is an auxiliary serious iterate. Using the assumption that $\delta = \bar{\varepsilon}/4$ and Lemma 5.1, we easily see that
\begin{equation}
\phi(\hat{z}_k) - \phi_* \leq 2\delta + \frac{\mu d_0^2}{2(1 + \lambda\mu)^{k/2-1}} = \bar{\varepsilon}/2 + \frac{\mu d_0^2}{2(1 + \lambda\mu)^{k/2-1}} \quad \forall k \geq \frac{2}{\lambda\mu}.
\end{equation}

Using the above conclusion and the fact that $1 + x > e^{x/2}$ for every $0 < x \leq 1$, we now easily see that there exists $\hat{k}$ that is bounded by (62) and has the property that any $\hat{z}_k$ with $k \geq \hat{k}$ satisfies (8).

b) This statement immediately follows from statement a) and Proposition 2.1 with $\delta = \bar{\varepsilon}/4$. 

We now make a few remarks about the iteration-complexity obtained in the above result under the mild (and usual) condition that $\bar{\varepsilon} = O(M^2/\mu)$, which clearly implies the existence of prox stepsizes $\lambda > 0$ satisfying
\begin{equation}
\Omega(\bar{\varepsilon}/M^2) = \lambda = O(1/\mu).
\end{equation}
First, the iteration-complexity in (63) with any such \( \lambda \) reduces to
\[
\mathcal{O} \left( 1 + \frac{M^2}{\mu \bar{\varepsilon}} \log_1^+ \left( \frac{\mu d_0^2}{\bar{\varepsilon}} \right) \right),
\]
which does not depend on \( \lambda \). Second, up to a logarithmic term, this bound is optimal up since it agrees with the lower bound described in Theorem 3.2.5 of [12]. Third, it can also be shown that the above bound (65) still holds for the less interesting case in which \( \bar{\varepsilon} \) agrees with the lower bound described in Theorem 3.2.5 of [12]. Third, it can also be shown that which does not depend on \( \lambda \) of [3] with \( \mu = \bar{\varepsilon}/d_0^2 \).

It is worth noting that, under assumption (64), it is possible to derive the complexity bound of Theorem 2.2(b) by applying the RPB method to a strongly convex perturbation of (1), namely,
\[
\min \left\{ \phi^\mu := \phi(u) + \frac{\mu}{2} \|u - x_0\|^2 \right\}
\]
where \( \mu = \bar{\varepsilon}/d_0^2 \). Indeed, it follows from Theorem 5.2(b) and Lemma A.1(a) that the total number of iterations performed by the RPB method with \( \delta = \bar{\varepsilon}/8 \) until it finds an auxiliary serious iterate \( \hat{z}_k \) satisfying
\[
\phi^\mu (\hat{z}_k) - \phi^\mu_* \leq \frac{\bar{\varepsilon}}{2}
\]
is bounded by (65), and hence by (22) in view of the fact that \( \mu = \bar{\varepsilon}/d_0^2 \). Now, using Lemma A.1(b) with \( \varepsilon = \hat{z}_k \), inequality (67), and the fact that \( \mu = \bar{\varepsilon}/d_0^2 \), we conclude that Theorem 5.2(b) holds.

Comparison with other proximal bundle variants. A related result to Theorem 5.2 is also described in [3] for a different proximal bundle variant which, similar to RPB, consists of solving a sequence of subproblems of the form (2). More specifically, it is shown that the latter method can find a \( \bar{\varepsilon} \)-optimal solution of the non-composite problem \( \min \{ f(x) : x \in \mathbb{R}^n \} \), where \( f \) is a \( \mu \)-strongly convex nonsmooth function, in
\[
\mathcal{O} \left( \left[ \frac{M^2 \lambda}{\alpha^2 \bar{\varepsilon}} \log_1^+ \left( \frac{1}{\alpha^2} \right) + 1 \right] \log_1^+ \left( \frac{f(x_0) - f_*}{\alpha \bar{\varepsilon}} \right) + \frac{M^2 \lambda}{\alpha \bar{\varepsilon}} \log_1^+ \left( \frac{M^2 \lambda}{\alpha \bar{\varepsilon}} \right) + 1 \right)
\]
iterations\(^1\) where \( \alpha := \min \{ \lambda, \mu, 1 \} \). Note that the above complexity reduces to the optimal complexity (65) only when \( \lambda = \Theta(1/\mu) \). This contrasts with the first remark following Theorem 5.2 where it is observed that the iteration-complexity of RPB reduces to the optimal bound (65) for a considerably larger range of prox stepsizes \( \lambda \), i.e., the ones satisfying (64).

It is also worth commenting on the implications of bound (68) to the convex setting of Sections 2-4 where only conditions (A1) and (A2) are assumed. Indeed, it follows from the discussion in previous two paragraphs that the number of iterations performed by the proximal bundle variant of [3] with \( \mu = \bar{\varepsilon}/d_0^2 \) applied to (66) until it finds a \( \bar{\varepsilon} \)-optimal solution of (1) is bounded by (68) with \( \mu = \bar{\varepsilon}/d_0^2 \). It is easy to see that, up to logarithmic terms, the latter bound reduces to the optimal bound (22) only when \( \lambda = \Theta(d_0^2/\bar{\varepsilon}) \), and that its general iteration-complexity is
\[
\mathcal{O} \left( \left[ \frac{M^2 d_0^4}{\lambda \bar{\varepsilon}^3} \log_1^+ \left( \frac{d_0^2}{\lambda \bar{\varepsilon}} \right) + \frac{d_0^2}{\lambda \bar{\varepsilon}} \log_1^+ \left( \frac{f(x_0) - f_* d_0^2}{\lambda \bar{\varepsilon}^2} \right) + 1 \right) \right)
\]
which, in terms of \( \bar{\varepsilon} \) alone, is \( \mathcal{O} \left( \| \log_1^+(\bar{\varepsilon}^{-1}) \|^2/\bar{\varepsilon}^3 \right) \). As already noted in Corollary 4.3 of [4], the complexity bound (69) in turn agrees, up to logarithmic terms, with the one obtained in [6] for a different proximal bundle variant, namely, \( \mathcal{O} \left( M^2 d_0^4/(\lambda \bar{\varepsilon}^3) \right) \). In contrast, Theorem 2.2 shows that

\(^1\) Actually, bound (68) has been formally derived in [4], which corrects a small error in the one derived in [3].
the iteration-complexity for the RPB method is always optimal (and hence is $O\left(1/\varepsilon^2\right)$) for a large range of prox stepsizes $\lambda$, i.e., the ones satisfying (21). Finally, since the size of the cutting-plane model can be bounded by $O(\lambda)$, which decreases with $\lambda$, it might be desirable to choose $\lambda$ relatively small in order to keep its size under control. It follows from the previous observation that such a choice of $\lambda$ still yields the optimal bound (22).

6 Concluding remarks

This paper presents a proximal bundle variant, namely, the RPB method, for solving convex NCO problems. Like other proximal bundle variants, the RPB method solves a sequence of prox subproblems whose objective functions are obtained by a usual regularized cutting-plane strategy. Moreover, it also performs either serious iterations during which the prox-center is changed or null iterations where the prox-center is left unchanged. As opposed to other proximal bundle variants: 1) RPB uses condition (13) instead of the descent condition (16) to decide when to perform a serious iteration; 2) RPB uses the auxiliary serious iterates to obtain optimal complexity bounds for a large range of prox stepsizes $\lambda$; and 3) RPB offers more flexibility in the construction of the cutting-plane model in that it allows to shrink it to a single cut whenever a serious iteration occurs. Finally, the size of the cutting-plane model can be chosen to be $O(\lambda)$, and hence can be nicely controlled by choosing $\lambda$ not too large.

This paper shows that the iteration-complexity of RPB is optimal for a large range of prox stepsizes $\lambda$ both in the context of convex and strongly convex NCO problems. As far as the authors are aware of, this is the first time that such result is obtained for a proximal bundle variant. A nice feature of our analysis is that it is carried out in the context of convex NCO problems. Moreover, it places the composite subgradient method under the umbrella of the RPB method in that the former can be viewed as an instance of the latter with a relatively small prox stepsize, namely, $\lambda = \Theta(\varepsilon/M^2)$ (which lies in the aforementioned range where RPB has optimal complexity). This paper also establishes iteration-complexity results for the RPB method to obtain iterates satisfying practical termination criteria such as (10) (see Theorem 2.4) or (9) (see Corollary 2.5) rather than the theoretical termination criterion (8).

An alternative and more general condition which can be used in place of (A2) is: $f \in \text{Conv}(\mathbb{R}^n)$ and there exists a scalar $M_f \geq 0$ such that

$$\partial f(u) \cap \{g \in \mathbb{R}^n : \|g\| \leq M_f\} \neq \emptyset \quad \forall u \in \text{dom} \ h.$$ 

If the subgradient of $f$ at $x_j$ in step 3 of the RPB method is chosen from the above set with $u = x_j$, then it can be easily verified that the whole analysis of this paper still holds.

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 $\lambda$ is constant. We believe that a slightly modified version of our analysis can be used to study the case in which $\lambda$ is allowed to change (possibly within a positive closed bounded interval) at every iteration $j$ for which $j$ is a serious iteration index. Second, we have assumed in Section 5 that $\text{dom} \ h$ is bounded but we also believe that a slight, although more complicated, modification of our analysis can handle the case in which the latter assumption is removed. Third, if $f$ is $\mu_f$-strongly convex and $h$ is $\mu_h$-strongly convex, then NCO (1) is clearly equivalent to another NCO (1) in which $f$ is convex, $h$ is $\mu$-strongly convex and $\mu = \mu_f + \mu_h$. Hence, there is no loss of generality in assuming in Section 5 that only $h$ is strongly convex although the aforementioned transformation requires knowledge of $\mu_f$. Fourth, a natural question is whether, under the weaker assumption that $\phi$ is $\mu$-strongly convex, the results
of Section 5 are still valid for the RPB method directly applied to the NCO problem (1) without using the above transformation. The advantage of the latter approach, if doable, is that it does not require the knowledge of $\mu_f$ (nor $\mu_h$).

References


**A Appendix**

**Lemma A.1.** The following statements about the perturbed function $\phi^\mu$ defined in (66) hold:

a) $d^\mu_0 := \|z_0 - z^\mu_*\| \leq d_0$ where $z^\mu_* := \arg\min_{u \in \mathbb{R}^n} \phi^\mu(u)$ and $d_0$ is as defined in (19);

b) for every $z \in \text{dom } h$,

$$\phi(z) - \phi_* \leq \phi^\mu(z) - \phi^\mu_* + \frac{\mu d^2_0}{2}$$

where $\phi^\mu_* := \phi^\mu(z^\mu_*)$.

**Proof:** a) Let $z_*$ be the unique point in $X_*$ such that $d_0 = \|z_0 - z_*\|$. Using the definitions of $z_*$ and $z^\mu_*$, we have

$$\phi_* + \frac{\mu}{2} \|z^\mu_* - z_0\|^2 \leq \phi(z^\mu_*) + \frac{\mu}{2} \|z^\mu_* - z_0\|^2 \leq \phi_* + \frac{\mu}{2} \|z_* - z_0\|^2.$$

Statement a) now follows from the above inequality, and the facts that $d^\mu_0 = \|z_0 - z^\mu_*\|$ and $d_0 = \|z_0 - z_*\|$.

b) It follows from the definitions of $\phi^\mu$ and $\phi^\mu_*$, and the fact that $d_0 = \|z_0 - z_*\|$ that for every $z \in \text{dom } h$,

$$\phi^\mu(z) - \phi^\mu_* \geq \phi(z) - \phi^\mu(z_*) = \phi(z) - \phi_* - \frac{\mu d^2_0}{2},$$

and hence statement b) follows. \(\blacksquare\)