Journal of Industrial and Management Optimization
Vol. 19, No. 12, December 2023, pp. 8691-8708
doi:10.3934/jimo.2023057

A REDISTRIBUTED PROXIMAL BUNDLE METHOD FOR NONSMOOTH NONCONVEX FUNCTIONS WITH INEXACT INFORMATION

Ming Huang*[1,2], Hui-Min Niu[1],
Si-Da Lin[1], Zi-Ran Yin[1] and Jin-Long Yuan*[1]

[1]School of Science, Dalian Maritime University, Dalian 116026, China

[2]Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

(Communicated by Changzhi Wu)

Abstract. In this paper, we propose a redistributed proximal bundle method for a class of nonconvex nonsmooth optimization problems with inexact information, i.e., we consider the problem of computing the approximate critical points when only the inexact information about the function values and subgradients are available and show that reasonable convergence properties are obtained. We assume that the errors in the computation of functions and subgradients are only bounded and in principle do not have to vanish within the limits. For the nonconvex functions, we design the convexification technique, which ensures that the linearization error of its augmentation function is nonnegative. Meanwhile, for the inexact information, we utilize noise management strategies and update approximate parameters to reduce the impact of inexact information. Based on this method, we can obtain the approximate solution.

1. **Introduction.** We consider the nonsmooth unconstrained optimization problem

$$\min f(x), \quad x \in R^n, \tag{1}$$

where $f(x)$ is a nonsmooth and nonconvex function.

Proximal bundle methods are currently one of the most effective optimization methods for unconstrained convex problems with discontinuous first derivatives [26]. The bundle method is based on the subdifferential estimation, which asymptotically ensures that the first-order optimality conditions are satisfied [3, 24, 25, 33, 34]. Of these dual methods, little attention is paid to how to model the objective function using the tangent hyperplane. The original form of the convex bundle methods are sometimes referred to as the stabilized cutting-plane or proximal bundle method, were mostly developed in the 1990s. The minimization of nonsmooth convex functions given precise information has been successfully achieved in several ways. The most popular of these is the level bundle method and the proximal bundle method [11]. Recently this method is applied to space-decomposition scheme for solving the eigenvalue problems [18]-[12] and convex problems [30, 31]. In fact, such methods

2020 *Mathematics Subject Classification.* Primary: 49J52, 90C26; Secondary: 49M37, 90C30.
Key words and phrases. Nonconvex optimization, nonsmooth optimization, inexact information, bundle method, $lower - C^2$.
*Corresponding author: Ming Huang, Jin-Long Yuan.

are currently considered to be the most efficient optimization methods for non-smooth problems; see [27, 29, 43, 42] for more detailed reviews.

The nonsmooth and nonconvex function can be found in many optimization problems such as large-scale lagrangian or semidefinite relaxations and stochastic simulations [6, 26, 38, 40, 42]. There are errors in the calculation of the function and the subgradient, making it impossible to determine the exact values. Particularly, in some circumstances, the function being considered is so complex that it is challenging to compute or exact information is not accessible, and the problem is tackled by getting inexact information in order to speed up the computation.

Example 1(The eigenvalue optimization or semidefinite programming) Considering the maximum eigenvalue function,

$$f(x) = \lambda_{max}(A(x)),$$

where the matrix $A : R^n \to S^m$, $m \times m$ is the space of symmetric matrices. Each element in $A(x)$ is a smooth function of x.

Obviously, for fixed $x \in R^n$, the maximum eigenvalue function $h(C) = \lambda_{max}(C)$ is a positive homogeneous function, where $C = A(x)$. Since the maximum eigenvalue function is non-smooth, the composite function $f(x)$ is also non-smooth, and it is also a nonconvex function [2], unless every element of $A(x)$ is a affine function of x.

Example 2(Sum of Euclidean parametrization) Given a series of smooth vector-valued functions $\{\varphi_1, \cdots, \varphi_I\}$, where $\varphi_i : R^N \to R^{m_i}$, $i = 1, \cdots, I$, $\Sigma_{i=1}^I |\varphi_i(x)|$ is a composite function with $m = \Sigma_{i=1}^I m_i$ components, where the inner layer mapping is

$$c(x) = (\varphi_1, \cdots, \varphi_I).$$

The outer layer function is

$$h(C_1, \cdots, C_m) = \Sigma_{i=1}^I |(C = \Sigma_{k=1}^{i-1} m_k + 1, \cdots, \Sigma_{k=1}^{i} m_k)|,$$

where $C_i \in R^{m_i}$. When $I = m_1 = 1$, let $\varphi_1 = a_2 x^2 + a_1 x + a_0$, $a_2 \neq 0$, then the function $h(c(x))$ is convex, if and only if $a_1^2 \leq 4a_0 a_2$. This function is nonconvex and nonsmooth in other cases.

Example 3 (Minimax optimization problem) The form to the minimization of a function is

$$f(x) = \max_{i \in I} f_i(x),$$

the functions $f_i(\cdot)$ that are often smooth but occasionally may be nonconvex. When the set I is bounded, as it is in [6, 42], it is simple to achieve controllable accuracy even though it may be impossible to assess f accurately in some situations.

Example 4 When a Monte-Carlo simulation-estimated expected value is used to represent the objective function [40]. By utilizing the central limit theorem and running the simulation numerous times, errors can be managed and minimized. Similar to the stochastic simulation providing the objective function, probability distribution functions can be used to understand the errors in the function and subgradient values.

Bundle method can be viewed from a "primitive" perspective as substituting the actual objective function with a model made by information bundles of data from previous evaluation points and their corresponding function values f and the subgradients g. Particularly, the proximal bundle method [11] computes the model function's proximal point in order to provide bundle elements and better minimizer estimations. The goal of this work is to modify one such approach to accommodate both nonconvex objective functions and inexact information. The direction of this work is to adopt this approach to deal with non-convex objective functions and inexact information. After the first study of bundle methods for convex problems, the problem of (locally) minimizing non-smooth and non-convex functions using exact information was considered in [19, 34] and more recently [7, 8, 21, 32]. They prioritize satisfying first-order optimality requirements by directing some convex combinations of subgradients in that direction [25, 28, 32]. All of these strategies, deal with nonconvexity by reducing the so-called linearization errors if they are negative.

In the convex case, for various algorithms, studies are carried out in [1, 22, 35]. In the nonconvex setting, inexact evaluations of subgradient algorithms have been researched in [45]. In this paper, only inexact information is available, based on previous studies on determining the proximal points of nonconvex functions, in order to develop the proximal bundle method applicable for nonsmooth and nonconvex functions with inexact information. The problem can only call the inexact oracle. Given a point x^j, we can not obtain exact function values and subgradients, just some estimates of the function value x^j and the subgradient at x^j instead. Therefore, the obtained information is $f^j \approx f(x^j)$ and $g^j \approx g(x^j) \in \partial f(x^j)$. Handling the inexact information is a natural challenge in many modern applications. We assume that inexact information is provided in such a way that the errors in the function values and subgradients are bounded by a universal constant. Although the algorithm and convergence analysis do not require knowledge of these constants, they do need to be present over the entire compact set of constraints. Unlike past research on inexact subgradient method, both [35] and [45] allow for nonvanishing noise. That is to say, the evaluation of subgradients being asymptotically tightened is required. The inexact estimation of function and subgradient values in the convex bundle method can be traced back to [20]. However, the noise in [20] disappears asymptotically. The first work considering nonvanishing perturbations in bundle methods seems to be [10]; but only function evaluations were required to be exact, whereas subgradient values could only be approximated. Nonvanishing inexactness was first introduced in [44] and extensively researched in [23] for both functions and subgradient values (still in the convex case). A unified theory on convex inexact bundle methods is elaborated in [37]. In our work, the behavior of the redistributed proximal bundle method given by exact information [8] for non-convex functions is considered. To our knowledge, the other work dealing with inexact information in bundle methods for nonconvex functions is [9, 36]. If the linearization errors are negative, the approach of [36] uses the “downshift” process to modify them. Our technique tilts its slopes in addition to downshifting the cutting-planes, and there are of course additional algorithmic changes. In this study, we apply the redistribued proximal bundle method to solve the nonsmooth and nonconvex problem with inexact information. Other significant variations from [8] include: we use inexact data in our study, only exact data are used in that one, it’s going to be more complicated to deal with and more difficult to solve.

In this paper, for a class of inexact nonconvex optimization problems, we propose a redistribued proximal bundle method to solve problem (1). Specifically, we design “convexifying” techniques for nonconvex functions to ensure that the corresponding linearization errors are non-negative, and then we employ a redistributed proximal bundle method to solve. The rest of this article is organized as follows. In Section 2 we review some variational analysis definitions and results required for this work. Section 2 also includes the main assumption for the functions considered in this work, and some basic results arising from the assumptions. In Section 3 we review how bundle methods work in a convex setting and discuss how requirements change in a nonconvex setting with inexact information. We provide the details of the algorithm developed in this paper. Section 4 examines the convergence properties of the algorithm.

2. **Preliminaries.** In this section, we review variational analysis concepts and findings that will be relevant to this paper. It is worth noting that we make use of the limiting subdifferential, denoted by $\partial f(\overline{x})$. Specifically, if f has a regular subdifferential at $\bar{x}$,

$$\hat{\partial} f(\bar{x}) := \left\{ g \in R^n : \lim_{x\to\bar{x}} \inf_{x\neq\bar{x}} \frac{f(x) - f(\bar{x}) - \langle g, x-\bar{x}\rangle}{|x-\bar{x}|} \geq 0 \right\}, \tag{2}$$

the limiting subdifferential is defined by

$$\partial f(\bar{x}) := \lim_{x\to\bar{x}} \sup_{f(x)\to f(\bar{x})} \hat{\partial} f(x). \tag{3}$$

We call these elements of subdifferential *subgradients*, which are usually denoted by the element g.

In this article, the objective function $f(x)$ is a nonconvex nonsmooth $lower-C^2$ function. The definition, properties and some assumptions of the $lower-C^2$ function are described below [[41], Definition 1.23, p.20].

(1): A function f is a $lower-C^2$ function on an open set $\mathcal{O}$, if f is finite on the open set $\mathcal{O}$ and for any point x, there exists a threshold r^0 such that for all $r \geq r^0$, $f + \frac{r}{2}|\cdot|^2$ is convex on an open neighborhood $\mathcal{O}'$ of x.

(2): A function f is prox-bounded when there exists $R \geq 0$, such that the function $f + \frac{R}{2}|\cdot|^2$ is bounded below and the corresponding threshold is the smallest $r_{pb} \geq 0$, $f + \frac{R}{2}|\cdot|^2$ is bounded below for all $R \geq r_{pb}$.

Now for problem (1.1), the basic assumptions are given that depend on the given parameters x^0 and M_0.

Assumption 2.1. *Given x^0 and M_0, there exists an open bounded sets M_0, such that*

$$\mathcal{L}_0 := \left\{ x \in R^n : f(x) \leq f\left(x^0\right) + M_0 \right\} \subset \mathcal{O}.$$

The following proposition gives an important property about the $lower-C^2$ function, basically related to the existence of uniform bounds for the various thresholds involved.

Proposition 2.2. *([41] Prop. 10.54)For a function f satisfying Assumption 2.1, the following holds:*

(1): *The level set $\mathcal{L}_0$ is nonempty and compact.*

(2): *The function f is bounded below and prox-bounded with threshold $r_{pb} = 0$.*

(3): *There exists $\rho^{id} > 0$ such that, for any $\rho \geq \rho^{id}$ and given any $y \in \mathcal{L}_0$, the function $f + \frac{\rho}{2}|\cdot - y|^2$ is convex on $\mathcal{L}_0$.*

(4): *The function f is Lipschitz continuous on $\mathcal{L}_0$.*

Another result of the Assumption 2.1 is that the proximal point mapping p_R, defined as:

$$p_R f(x) := \operatorname*{argmin}_y \left\{ f(y) + R\frac{1}{2}|x-y|^2 \right\}, \tag{4}$$

is single-valued and Lipschitz continuous on $\mathcal{L}_0$ [41], provided the prox-parameter R sufficiently large. By [[41], Thm 2.26] it is clear that, in this case, R sufficiently large means that $R \geq \rho^{id}$, where ρ^{id} is the value in item (3) in Proposition 2.2.

It is relatively simple to define the concept of inexact information for function values. Given a point x and a certain amount of error tolerance $\kappa \geq 0$, $F \in R$ approximates the value $f(x)$ within κ, meaning $|F - f(x)| \leq \kappa$. For the subgradient values, of course, the notion of inexact information allows for more interpretation. The closed ball in R^n centered at x and radius $\rho \geq 0$ is denoted by $B_\rho(x)$. We will consider the following estimates, which make good sense especially in the nonconvex case. At a point x, an element $g \in R^n$ approximates within tolerance $\theta \geq 0$ some subgradient of f at point x if $g \in \partial f(x) + B_\theta(0)$.

The algorithm deals with inexact information. We denote the current iteration index j with iterate point x^j, and the stability center $\hat{x}^k$. We present inexact function values and subgradients as follows:

$$\begin{aligned} f^j &= f(x^j) - \kappa^j, && \text{where} \quad \kappa^j \text{ is an unknown error,} \\ \hat{f}^k &= f(\hat{x}^k) - \hat{\kappa}^j, && \text{where} \quad \hat{\kappa}^j \text{ is an unknown error,} \\ g^j &\in \partial f(x^j) + B_{\theta^j}(0), && \text{where} \quad \theta^j \text{ is an unknown error.} \end{aligned}$$

It should be noted that the sign of errors κ^j is not specified, the true function value can be either overestimated or underestimated. The error terms κ^j and θ^j are assumed bounded:

$$\left|\kappa^j\right| \leq \kappa, \qquad 0 \leq \theta^j \leq \theta, \qquad \text{for all } j,$$

their bounds κ, θ are usually unknown.

If the function f is convex and exact in data, the linearization error e_j^k:

$$e_j^k = f\left(\hat{x}^k\right) - f\left(x^j\right) - \left\langle g(x^j),\ \hat{x}^k - x^j \right\rangle \geq 0, \tag{5}$$

where $\hat{x}^k$ is a stability center, $g(x^j)$ is an exact subgradient of $\partial f(x^j)$.

The accumulate linearization for f is:

$$f_j^L(x) = f\left(x^j\right) + \left\langle g(x^j),\ x - x^j \right\rangle. \tag{6}$$

The classical cutting - plane model for f is:

$$\begin{aligned} &\max\left\{f_j^L(x) : j \in J^k = \{1, \cdots k\}\right\} \\ &= \max_{j \in J^k}\left\{f\left(x^j\right) + \left\langle g(x^j), x - x^j \right\rangle\right\} \\ &= f\left(\hat{x}^k\right) + \max_{j \in J^k}\left\{-e_j^k + \left\langle g(x^j), x^j - \hat{x}^k \right\rangle\right\}. \end{aligned} \tag{7}$$

We use inexact information in our setting. We also need to take into account the possible nonconvexity of $f(x)$. In accordance with the redistributed proximal bundle method of [8], we have the following strategies.

The linearization error of the objective function $f(x)$ is as follows,

$$e_{f,j}^k = \hat{f}^k - f^j - \left\langle g^j, \hat{x}^k - x^j \right\rangle, \tag{8}$$

and the argument function is:

$$\varphi_k\left(x^j\right) = f^j + \frac{\eta_k}{2}|x^j - \hat{x}^k|^2. \tag{9}$$

According to the properties of $lower - C^2$ functions, η is a threshold, when $\eta_k > \eta$, the argument function is convex. Specifically, $\varphi_k\left(\hat{x}^k\right) = \hat{f}^k$.

The linearization error of $e^k_{\varphi,j}$ is expressed as:

$$\begin{aligned} e^k_{\varphi,j} &= \varphi_k\left(\hat{x}^k\right) - \varphi_k\left(x^j\right) - \left\langle g^j_\varphi, \hat{x}^k - x^j\right\rangle \\ &= e^k_{f,j} + \frac{\eta_k}{2}|x^j - \hat{x}^k|^2 \\ &\geq 0, \end{aligned} \tag{10}$$

there exists $g^j \in \partial f\left(x^i\right) + B_{\theta^j}(0)$, $g^j_\varphi = g^j + \eta_k\left(x^j - \hat{x}^k\right) \in \partial\varphi_k\left(x^i\right) + B_{\theta^j}(0)$.

A convex piecewise - linear model is:

$$\begin{aligned} \widetilde{\varphi}_k(x) &= \max_{j\in J^k}\left\{\varphi_k(x^j) + \left\langle g^j_\varphi, x - x^j\right\rangle\right\} \\ &= \hat{f}^k + \max_{j\in J^k}\left\{-e^k_{\varphi,j} + \left\langle g^j_\varphi, x - \hat{x}^k\right\rangle\right\}. \end{aligned} \tag{11}$$

The new iterative point x^{k+1} is given by the following QP subproblem,

$$x^{k+1} = p_{\mu_k}\widetilde{\varphi}_k\left(\hat{x}^k\right) = \text{argmin}_x\left\{\widetilde{\varphi}_k(x) + \frac{\mu_k}{2}|x - \hat{x}^k|^2\right\}, \tag{12}$$

where μ_k is the pro-parameter. Notice that $\eta_k > \eta$, φ_k is convex, so the quadratic subproblem is a strongly convex problem. The relationship between the new generated point and the current stationary center is demonstrated by the next lemma.

Lemma 2.3. *$\widetilde{\varphi}_k$ is piecewise-linear, this means that there exists a simplicial multiplier α^k_j, $\alpha^k_j \in R^{|J^k|}$, $\alpha^k_j \geq 0$, $\sum_{j=1}^{|k|}\alpha^k_j = 1$,*

$$\begin{cases} \hat{g}^k = \sum\limits_{j\in J^k}\alpha^k_j g^j_\varphi = \sum\limits_{j\in J^k}\alpha^k_j\left(g^j + \eta_k\left\langle x^j - \hat{x}^k\right\rangle\right) \\ x^{k+1} = \hat{x}^k - \dfrac{1}{\mu_k}\hat{g}^k. \end{cases}.$$

According to the optimality condition of the above subproblem, the following relation are drawn,

$$0 \in \widetilde{\varphi}_k\left(\hat{x}^k\right) + \frac{\mu_k}{2}|x^{k+1} - \hat{x}^k|^2, \tag{13}$$

and $\alpha_j = \left(\alpha^1_j, \alpha^2_j, \cdots\alpha^k_j\right)$ is the solution to

$$\begin{array}{ll} \min\limits_{\alpha^k_j\in R^{|J^k|}} & \frac{1}{2}|\hat{g}^k|^2 + \mu_k\sum\limits_{j\in J^k}\alpha^k_j e^k_{\varphi,j} \\ \text{s.t.} & \alpha^k_j \geq 0, \quad \sum_{j=1}^{|k|}\alpha^k_j = 1. \end{array}$$

Proof. The QP subproblem (12) is

$$\min_{x\in R^n}\widetilde{\varphi}_k\left(\hat{x}^k\right) + \frac{\mu_k}{2}|x - \hat{x}^k|^2.$$

It is equivalent to the following problem,

$$\begin{array}{ll} \min\limits_{t,x\in R^n} & t + \frac{\mu_k}{2}|x - \hat{x}^k|^2 \\ \text{s.t.} & -e^k_{\varphi,j} + \left\langle g^j_\varphi, x - \hat{x}^k\right\rangle \leq t. \end{array}$$

Introducing $\alpha_j \in R_+^{|J_k|}$, the corresponding Lagrangian function is:

$$\mathcal{L}(x,t,\alpha) = t + \frac{\mu_k}{2}|x-\hat{x}^k|^2 + \sum_{j\in J^k} \alpha_j \left(-e^k_{\varphi,j} + \left\langle g^j_\varphi, x-\hat{x}^k \right\rangle - t\right)$$
$$= \frac{\mu_k}{2}|x-\hat{x}^k|^2 + \sum_{j\in J^k} \alpha_j \left\langle g^j_\varphi, x-\hat{x}^k \right\rangle + \left(1 - \sum_{j\in J^k} \alpha_j\right) t - \sum_{j\in J^k} \alpha_j e^k_{\varphi,j}.$$

Because the dual gap is 0, we can get it by solving the original problem or the dual problem:

$$\min_{(x,t)\in R^n\times R} \max_{\alpha\in R^{|J^k|}} \mathcal{L}(x,t,\alpha) = \max_{\alpha\in R^{|J^k|}} \min_{(x,t)\in R^n\times R} \mathcal{L}(x,t,\alpha). \tag{14}$$

Because both the original problem and the dual problem have finite optimal values, and $t \in R$, $\sum\limits_{j\in J^k} \alpha^k_j = 1$, then (14) can be transformed into

$$\min_{x\in R^n} \max_{\alpha\in R^{|J^k|}} \mathcal{L}(x,\alpha) = \max_{\alpha\in R^{|J^k|}} \min_{x\in R^n} \mathcal{L}(x,\alpha), \tag{15}$$

where

$$\mathcal{L}(x,\alpha) = f(\hat{x}^k) + \frac{\mu_k}{2}|x-\hat{x}^k|^2 + \sum_{j\in J^k} \alpha_j \left\langle g^j_\varphi, x-\hat{x}^k \right\rangle - \sum_{j\in J^k} \alpha_j e^k_{\varphi_j}.$$

Consider the above dual problem, optimal conditions of the minimization problem for a fixed α is $x(\alpha) = \operatorname{argmin}_x \mathcal{L}(x,\alpha)$, if and only if $0 = \nabla_x \mathcal{L}(x,\alpha)$, that is,

$$0 = \mu_k(x(\alpha) - \hat{x}^k) + \sum_{j\in J^k} \alpha_j g^j_\varphi.$$

Specifically, when $\alpha = \alpha_j$, $x(\alpha_j) = x^{k+1}$. Next we prove $\hat{g}^k \in \partial\widetilde{\varphi}_k(x^{k+1})$,

$$\begin{aligned} 0 &\in \partial\varphi_k\left(x^{k+1}\right) + \mu_k\left(x^{k+1} - \dot{x}^k\right) \\ &= \partial\varphi_k\left(x^{k+1}\right) - \sum_{j\in j^k} \alpha_j g^j_\varphi \\ &= \partial\varphi_k\left(x^{k+1}\right) - \hat{g}^k. \end{aligned}$$

Multiply each side of the equation by $x(\alpha) - \hat{x}^k$ and $\frac{1}{\mu_k}\sum_{j\in J^k} \alpha_j g^j_\varphi$,

$$\begin{aligned} 0 &= \mu_k \left\|x(\alpha) - \hat{x}^k\right\|^2 + \sum_{j\in J^k} \alpha_j \left\langle g^j_\varphi, x(\alpha) - x^k \right\rangle \\ &= \sum_{j\in J^k} \alpha_j \left\langle g^j_\varphi, x(\alpha) - \hat{x}^k \right\rangle - \frac{1}{\mu_k}|\sum_{j\in j^k} \dot{\alpha}_j g^j_\varphi|^2. \end{aligned}$$

This means that

$$\mu_k|x(\alpha) - \hat{x}^k|^2 = \frac{1}{\mu_k}|\sum_{j\in J^k} \alpha_j g^j_\varphi|^2,$$

and

$$\mathcal{L}(x(\alpha),\alpha) = \frac{1}{2\mu_k}|\sum_{j\in J^k} \alpha_j g^j_\varphi|^2 - \sum_{j\in J^k} \alpha_j e^k_{\varphi,j}.$$

In summary, α_j is the solution of

$$\max_{\alpha\in J^k} \mathcal{L}(x(\alpha),\alpha) = -\min_{\alpha\in J^k} \{\frac{1}{2\mu_k}|\sum_{j\in J^k} \alpha_j g^j_\varphi|^2 - \sum_{j\in J^k} \alpha_j e^k_{\varphi,j}\}.$$

We finish the proof. □

In addition, the following relations hold:

$$\hat{g}^k \in \partial\widetilde{\varphi}_k\left(x^{k+1}\right),$$

$$\begin{aligned}\varphi_k\left(\hat{x}^k\right)+\widetilde{\varphi}_k\left(x^{k+1}\right) &= \hat{f}^k-\varphi_k\left(\hat{x}^k\right)+\widetilde{\varphi}_k\left(x^{k+1}\right)\\ &= \mu_k|x^{k+1}-\hat{x}^k|^2+\sum_{j\in J^k}\alpha_j^k e_{\varphi,j}^k.\end{aligned}$$

The following conclusions can be drawn from Lemma 2.3,

$$d^k = x^{k+1}-\hat{x}^k = -\frac{1}{\mu^k}\hat{g}^k, \quad \text{where} \quad \hat{g}^k=\sum_{j\in J^k}\alpha_j^k g_\varphi^j. \tag{16}$$

Once the new iterate is known, we define the *aggregation linearization*,

$$\varphi_{-k}(x)=\widetilde{\varphi}_k\left(x^{k+1}\right)+\left\langle\hat{g}^k, x-x^{k+1}\right\rangle. \tag{17}$$

Thus we have,

$$\begin{aligned}\varphi_{-k}\left(x^{k+1}\right)=\varphi_k\left(x^{k+1}\right),\ \hat{g}^k\in\partial\widetilde{\varphi}_k\left(x^{k+1}\right),\ \text{and}\\ \hat{g}^k=\nabla\varphi_{-k}\left(x\right), \qquad \text{for all } x\in R^n.\end{aligned} \tag{18}$$

By the subgradient inequality, we can obtain:

$$\varphi_{-k}\left(x\right)\le\widetilde{\varphi}_k\left(x\right),\ \text{for all } x\in R^n. \tag{19}$$

The *aggregation errors* are defined by

$$\begin{aligned}e_{\tilde{\varphi}} &= \widetilde{\varphi}_k(\hat{x}^k)-\widetilde{\varphi}_k(x^{k+1})+\left\langle\hat{g}^k, x^{k+1}-\hat{x}^k\right\rangle\\ &= \hat{f}^k-\widetilde{\varphi}_k(x^{k+1})+\left\langle\hat{g}^k, x^{k+1}-\hat{x}^k\right\rangle\\ &\ge 0.\end{aligned} \tag{20}$$

Using $\varphi_k\left(\hat{x}^k\right)=\hat{f}^k$ and the optimal multipliers from (16), this gives the following alternative *aggregate error* expressions:

$$e_{\tilde{\varphi}}=\sum_{j\in J^k}\alpha_j^k e_{\varphi,j}^k. \tag{21}$$

Similarly, we get the *aggregate linearization* by (17),

$$\varphi_{-k}(x)=\hat{f}^k+\left\langle\hat{g}^k, x-\hat{x}^k\right\rangle-e_{\tilde{\varphi}}.$$

Any choice for convexification parameter that keeps $e_{\varphi,j}^k$ in (10) is acceptable. We employ a redistributed proximal bundle method in our analysis,

$$\eta_k^{min}\ge\max\left\{\max_{j\in J^k, x^j\ne\hat{x}^k}\frac{-2e_{f,j}^k}{\left|x^j-\hat{x}^k\right|^2}, 0\right\}+\gamma, \tag{22}$$

for the (small) positive parameter γ in the above equation, the term $\max_{j\in J^k, x^j\ne\hat{x}^k}\frac{-2e_{f,j}^k}{|x^j-\hat{x}^k|^2}$ represents the smallest value of η to suggest that linearization errors of the "locally convexified" function remain nonnegative for all $j\in J^k$: $e_{f,j}^k+\frac{\eta}{2}|x^j-\hat{x}^k|^2\ge 0$. The nonnegativity of η_k^{min} is obtained by taking the maximum of this term with 0, adding the"safeguarding" small positive parameter makes η_k^{min} strictly greater than the minimal value.

3. **Algorithm statement.** As described in Section 2, previous extensions of the bundle method to nonconvex settings usually depend on redefining the linearization error so that it is non-negative. After computing the new iteration, we first check whether it provides sufficient decrease of the objective function to the previous stabilization center (both are, of course, inexact values in our setting).

Since x^{k+1} is the solution to the QP issue in (12), the predicted descent is defined,

$$\begin{aligned}\delta_{k+1} &= \hat{f}^k + \frac{\eta_k}{2}|x^{k+1} - \hat{x}^k|^2 - \widetilde{\varphi}_k\left(x^{k+1}\right) \\ &= \mu_k|x^{k+1} - \hat{x}^k|^2 + \sum_{j\in J^k} \alpha_j^k e_{\varphi,j}^k + \frac{\eta_k}{2}|x^{k+1} - \hat{x}^k|^2. \end{aligned} \tag{23}$$

For convenience, we have temporarily removed the iteration index from the notation. To define the QP problem, the current *prox-parameter* R is divided into two non-negative terms η and μ. Using η as the *convexification parameter* and μ as the *model prox-parameter*, satisfying $R = \eta + \mu$,

$$p_R\hat{f}^k = p_\mu\varphi_k\left(\hat{x}^k\right). \tag{24}$$

Along the iterative process, R, μ and η have to be suitably modified.

So equation (23) can be transformed into

$$\delta_{k+1} = \frac{R_k + \mu_k}{2}|x^{k+1} - \hat{x}^k|^2 + e_{\widetilde{\varphi}}. \tag{25}$$

Note that since $e_{\widetilde{\varphi}} \geq 0$ by (20), it follows from (25) that

$$\delta_{k+1} \geq 0. \tag{26}$$

Hence,

$$\delta_{k+1} > \frac{\mu_k}{2}|x^{k+1} - \hat{x}^k|^2 + \frac{\eta_k}{2}|x^{k+1} - \hat{x}^k|^2 = \frac{R_k}{2}|x^{k+1} - \hat{x}^k|^2. \tag{27}$$

Algorithm 3.1 (Noconvex Nonsmooth Redistributed Proximal Bundle Method with Inexact Information) Given a procedure that provides an approximation f of $f(x)$ for each x, and the corresponding approximate subgradient value g.

Step 1 (Input and Initialization) Choose an initial starting point $x^0 \in R^n$, $\tau \in [\frac{1}{2}, 1)$ and unacceptable increase parameters $M_0 > 0$, and $R_0 > 0$. Select parameter $\gamma > 0$ and a stopping tolerance $TOL_{stop} \geq 0$, $m \in (0,1)$, and a convexification growth parameter $\Gamma > 1$, place the initial iteration counter $k = 0$, the bundle index set $J^0 = 0$, and the first candidate point $x^0 = \hat{x}^0$.

Compute the inexact oracle values $\hat{f}^0 = f^0$, and the initial prox-center $\hat{x}^0 = x^0$, $g^0 \in \partial f(x^0) + B_{\theta^0}(0)$. Setting the starting prox-parameter distribution $(\mu_0, \eta_0) = (R_0, 0)$.

Step 2 (Model Generation and QP Subproblem) According to (12),

$$x^{k+1} = P_{\mu_k}\widetilde{\varphi}_k\left(\hat{x}^k\right) = \operatorname{argmin}\left\{\widetilde{\varphi}_k(\hat{x}^k) + \frac{\mu_k}{2}|x - \hat{x}^k|^2\right\}. \tag{28}$$

Define the predicted decrease from (25),

$$\delta_{k+1} = \frac{R_k + \mu_k}{2}|x^{k+1} - \hat{x}^k|^2 + e_{\widetilde{\varphi}}. \tag{29}$$

Step 3 (Trial Point Finding and Stopping Test) Given the model $\widetilde{\varphi}_k(x)$ defined by (11), compute the direction $x^{k+1} - \hat{x}^k$. Define the associated $\hat{g}^k$ by (16), $e_{\widetilde{\varphi}}$ by (20), and δ_{k+1} by (23). If $\delta_{k+1} \leq TOL_{stop}$, stop.

Step 4 (Noise Management) If relation (27) does not hold, set $\mu_{k+1} = \tau\mu_k$, $k = k + 1$, go to Step 1.

Step 5 (Serious Step Test) Call the inexact oracle to obtain f^{k+1}, and $g^{k+1} \in \partial f(x^{k+1}) + B_{\theta^{k+1}}(0)$,

$$f^{k+1} \leq \hat{f}^k - m\delta_{k+1}, \tag{30}$$

and loop to Step 2.

If this condition is hold, declare a serious step, set $\hat{x}^{k+1} = x^{k+1}$, $\hat{f}^{k+1} = f^{k+1}$. Otherwise, declare a null step, set $k+1 = k$, $\hat{x}^{k+1} = x^k$, $\hat{f}^{k+1} = \hat{f}^k$.

Step 6 (Update η)

$$\begin{cases} \eta_{k+1} := \eta_k & \text{if } \eta_{k+1}^{\min} \leq \eta_k; \\ \eta_{k+1} := \Gamma\eta_{k+1}^{\min} \quad \text{and } R_k := \mu_k + \eta_{k+1} & \text{if } \eta_{k+1}^{\min} > \eta_k, \end{cases}$$

where $\eta_{k+1}^{\min}$ is gived by (22).

Step 7 (Update μ)

If $f^{k+1} > \hat{f}^k + M_0$, then target growth is unacceptable.

Restart algorithm via settings,

$$\eta_0 := \eta_k, \mu_0 := \Gamma\mu_k, R_0 := \eta_0 + \mu_0$$
$$x^0 := \hat{x}^k, k(0) := 0, J_0 := \{0\}.$$

and loop to Step 2.

Otherwise, in the case of serious steps, increase k by 1 and go to Step 2.

Remark 3.1. Note that the update of elements in the bundle is not explicitly stated in Algorithm 3.1. The update strategy is different for null and serious steps. When a serious step occurs, the newly generated point is considered as a new proximal center and the corresponding linearization error in the bundle is updated. When there is a null step, the proximal center remains unchanged and only the newly generated information is added to the bundle to improve the accuracy of the model. As the iterations proceed, the elements in the bundle may be too large, thus reducing the efficiency of the algorithm. So, we can adopt the compression strategy. For the compression strategy, the number of elements in the bundles can be at least two, the aggregate information and the new generated information. It should be noted that although the compression strategy does not impair the convergence of the algorithm, it may affect the effectiveness of the model, if the number of elements in the bundles is too small.

In order to prove the convergence of the algorithm, we will make the following assumptions.

Assumption 3.2. The cardinality of the set $\{j \in J^k \mid \alpha_j^k > 0\}$ is uniformly bounded in k.

The above assumption can also be seen in [47]. The reason is that most (if not all) active set QP solvers choose linearly independent bases. In the expression of $\hat{g}^k$ in (16), this means that the QP solver gives solutions with no more than $k+1$ positive simplex multipliers (such a solution always exists by the Carathéodory). Similar assumptions of QP solvers have been used in different QP-based methods, especially in the bundling process in [5].

Assumption 3.3. The model convexification parameter sequences $\{\eta_k\}$ is bounded.

When the function information is exact, Boundedness of $\{\eta_k\}$ for the *lower* $-C^2$ case has been shown in [8]. However, it is theoretically possible in our setting that inexactness leads to an unbounded $\{\eta_k\}$, even if the objective function is convex.

4. **Asymptotic analysis.** We now analyze the different cases that may occur when Algorithm 3.1 loops forever. In the convergence analysis of the bundle method, we usually consider the following two possible cases.

- Either there are an infinite number of serious steps;
- or there are a finite number of serious steps, and then an infinite number of null steps.

Lemma 4.1. *Suppose the above Assumption 3.2 holds.*
(i) If $e_{\widetilde{\varphi}} \to 0$ *as* $k \to \infty$, $\sum_{j \in J^k} \alpha_j^k \left|x^j - \hat{x}^k\right| \to 0$ *as* $k \to \infty$.
(ii) If, in addition, Assumption 3.3 holds on the subset $K \subset \{1, 2, \ldots\}$ *and* $\hat{x}^k \to \bar{x}$, $\hat{g}^k \to \bar{g}$ *as* $K \ni k \to \infty$, *then we also have*

$$\bar{g} \in \partial f(\bar{x}) + B_\theta(0). \tag{31}$$

(iii) If, in addition, $\hat{g}^k \to 0$ *as* $K \ni k \to \infty$, *then* $\bar{x}$ *satisfies the following approximate stationarity condition:*

$$0 \in \partial f(\bar{x}) + B_\theta(0). \tag{32}$$

(iv) In addition, for each $\varepsilon > 0$ *there exists* $\rho > 0$, *such that* $f(y) \geq f(\bar{x}) - (\theta + \varepsilon)|y - \bar{x}| - 2\kappa$, *for all* $y \in B_\rho(\bar{x})$.

Proof. For the interpretation of the convexification parameters, in (22),

$$\max_{j \in J^k, x^j \neq \hat{x}^k} \frac{-2e_{f,j}^k}{\left|x^j - \hat{x}^k\right|^2},$$

represents the minimal value of η to imply that for all $j \in J^k$ the linearization errors of the "locally convexified" function remain nonnegative:

$$e_{f,j}^k + \frac{\eta}{2}\left|x^j - \hat{x}^k\right|^2 \geq 0$$

for all $j \in J^k$. It is then easily seen that, for such η_k and for η and $\eta_k > \eta + \gamma$, we have that

$$e_{f,j}^k + \frac{\eta_k}{2}\left|x^j - \hat{x}^k\right|^2 \geq \frac{\gamma}{2}\left|x^j - \hat{x}^k\right|^2.$$

Since α_j^k and $e_{f,j}^k + \frac{\eta_k}{2}\left|x^j - \hat{x}^k\right|^2$ are nonnegative, and $\alpha_j^k \leq 1$, if $e_{\widetilde{\varphi}} \to 0$ then it follows from (21) that $\alpha_j^k(e_{f,j}^k + \frac{\eta_k}{2}\left|x^j - \hat{x}^k\right|^2) \to 0$ for all $j \in J^k$. Hence,

$$\alpha_j^k(e_{f,j}^k + \frac{\eta_k}{2}\left|x^j - \hat{x}^k\right|^2) \geq (\alpha_j^k)^2(e_{f,j}^k + \frac{\eta_k}{2}\left|x^j - \hat{x}^k\right|^2) \geq \frac{\gamma}{2}(\alpha_j^k\left|x^j - \hat{x}^k\right|)^2 \to 0.$$

Thus, $\alpha_j^k\left|x^j - \hat{x}^k\right| \to 0$ for all $j \in J^k$. Hence, by the Assumption 3.3, the sum in the item (i) is over a finite set of indices and each element in the sum tends to zero, this demonstrates the claim (i).

For each j, let h^j be the orthogonal projection of g^j on the convex and closed set $\partial f(x^j)$, it holds that $|g^j - h^j| \leq \theta^j \leq \theta$. By (16), we can get,

$$\begin{aligned}\hat{g}^k &= \sum_{j \in J^k} \alpha_j^k g^j + \eta^k \sum_{j \in J^k} \alpha_j^k \left(x^j - \hat{x}^k\right) \\ &= \sum_{j \in J^k} \alpha_j^k h^j + \sum_{j \in J^k} \alpha_j^k \left(g^j - h^j\right) + \eta^k \sum_{j \in J^k} \alpha_j^k \left(x^j - \hat{x}^k\right).\end{aligned} \tag{33}$$

Let J be the set of all $j \in J^k$ such that $\liminf \alpha_j^k > 0$. Then item (i) implies that $|x^j - \hat{x}^k| \to 0$, thus,

$$|x^j - \bar{x}| \leq |x^j - \hat{x}^k| + |\hat{x}^k - \bar{x}| \to 0.$$

As $h^j \in \partial f(x^j) + B_{\theta^j}(0)$ and $x^j \to \bar{x}$ for $j \in J$, and $\{\alpha_j^k\} \to 0$ for $j \notin J$, passing onto a further subsequence in the set K(if necessary), the outer semicontinuity of the Clarke subdifferential ([41], Thm 6.6) implies that

$$\lim_{k\to\infty} \sum_{j\in J^k} \alpha_j^k h^j \in \partial f(\bar{x}).$$

Note that the second term in (33) is clearly in $B_{\theta^j}(0)$, while the last term tends to zero by item (i), this shows the assertion (ii).

The term (iii) follows from noting that $\hat{g}^k \to 0$ as $K \ni k \to \infty$. Adding the inclusion and result (ii) gives (32).

We finally prove item (iv). Fix any $\varepsilon > 0$, let $\rho > 0$ be such that (ii) applies to $\bar{x}$. Let $y \in B_\rho(\bar{x})$ be arbitrary but fixed. We can consider that J^k is a fixed index set, let J be the set of $j \in J^k$ for which $|x^j - \hat{x}^k| \to 0$. In particular, it then holds that $x^j \in B_\rho(\bar{x})$. By (i), we get that $\{\alpha_j^k\} \to 0$ for $j \notin J$. Since the objective function f is a $lower - C^2$ function, given an open set Ω, combining ([4] Thm.2, Cor.3) with ([46], Prop.2.4), the following statements are equivalent.

$$\left\{ \begin{array}{l} \forall \bar{x} \in \Omega, \forall \varepsilon > 0, \exists \rho > 0 : \\ \forall x \in B_\rho(\bar{x}) \text{ and } g \in \partial f(x). \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} f(x+a) \geq f(x) + \langle g, a\rangle - \varepsilon |a| \\ \text{ whenever } |a| \leq \rho \text{ and } x + a \in B_\rho(\bar{x}). \end{array} \right\}$$

Using the above information, for $j \in J$, we can obtain that

$$\begin{aligned} f(y) &\geq f^j + \langle g^j, y - x^j\rangle + \kappa^j + \langle h^j - g^j, y - x^j\rangle - \varepsilon \left|y - x^j\right| \\ &\geq f^j + \langle g^j, y - x^j\rangle + \kappa^j - \left(\theta^j + \varepsilon\right)\left|y - x^j\right|. \end{aligned}$$

According to (8) and the definition of linearization error, we have

$$f^j + \langle g^j, -x^j\rangle = \hat{f}^k - \langle g^j, \hat{x}^k\rangle - e_{f,j}^k.$$

Therefore, it is obvious that

$$f(y) \geq \hat{f}^k - e_{f,j}^k + \langle g^j, y - \hat{x}^k\rangle + \kappa^j - \left(\theta^j + \varepsilon\right)\left|y - x^j\right|.$$

Since $\frac{\eta_k}{2}|x^j - \hat{x}^k|^2 \geq 0$ and $g_\varphi^j = g^j + \eta_k\left(x^j - \hat{x}^k\right)$, we obtain that

$$f(y) \geq f\left(\hat{x}^k\right) - e_{\varphi,j}^k + \langle g_\varphi^j, y - \hat{x}^k\rangle - \eta_k \langle x^j - \hat{x}^k, y - \hat{x}^k\rangle + \kappa^j + \hat{\kappa} - \left(\theta^j + \varepsilon\right)\left|y - x^j\right|.$$

Taking the convex combination in the latter relation using the simplicial multipliers in (21) and using (16), this gives

$$\begin{aligned} f(y)\sum_{j\in J}\alpha_j^k \geq& \sum_{j\in J}\alpha_j^k\left(f\left(\hat{x}^k\right) - e_{\varphi,j}^k + \langle g_\varphi^j, y - \hat{x}^k\rangle\right) - \eta_k\left\langle \sum_{j\in J}\alpha_j^k\left(x^j - \hat{x}^k\right), y - \hat{x}^k\right\rangle \\ &+ \sum_{j\in J}\alpha_j^k\left(\kappa^j + \hat{\kappa}\right) - \left(\theta^j + \varepsilon\right)\sum_{j\in J}\alpha_j^k\left|y - x^j\right| \\ \geq& f\left(\hat{x}^k\right)\sum_{j\in J}\alpha_j^k - e_{\tilde{\varphi}} + \langle \hat{g}^k, y - \hat{x}^k\rangle - \sum_{j\notin J}\alpha_j^k\langle g_\varphi^j, y - \hat{x}^k\rangle \\ &- \eta^k\left\langle \sum_{j\in J}\alpha_j^k\left(x^j - \hat{x}^k\right), y - \hat{x}^k\right\rangle \\ &- 2\kappa - (\theta + \varepsilon)\sum_{j\in J}\alpha_j^k\left(\left|y - \hat{x}^k\right| + \left|x^j - \hat{x}^k\right|\right), \end{aligned}$$

using item (i) and also that $\{\alpha_j^k\} \to 0$ for $j \notin J$, we obtain that,

$$f(y) \geq f(\bar{x}) - (\theta + \varepsilon)|y - \bar{x}| - 2\kappa.$$

The proof is done. □

4.1. **Infinite number of serious steps.**

Theorem 4.2. *Let the algorithm generate an infinite number of serious steps. Then $\delta_{k+1} \to 0$, as $k \to \infty$. In addition, if the Assumption 3.3 hold, we have,*
(i) If $\sum_{k=1}^{\infty} \frac{1}{\mu_k} = +\infty$, then as $k \to \infty$, we have $e_{\tilde{\varphi}} \to 0$. There exist $K \subset \{1, 2, \ldots\}$ and $\bar{x}$, $\bar{g}$ such that $\{\hat{x}^k\} \to \bar{x}$, $\hat{g}^k \to \bar{g}$, and $\hat{g}^k \to 0$ as $K \ni k \to \infty$. In particular, if Assumption 3.2 holds, then Lemmas 4.1 holds;
(ii) If $\liminf_{k\to\infty} \frac{1}{\mu_k} > 0$, then these assertions hold for all accumulation points $\bar{x}$ of the serious step sequence $\{\hat{x}^k\}$.

Proof. At each serious step k, the relation (30) holds. We have that

$$f^{k+1} \le \hat{f}^k - m\delta_{k+1}, \tag{34}$$

where $\delta_{k+1} \ge 0$. Therefore, the sequence $\{\hat{f}^k\}$ is nonincreasing. Since the sequence $\{\hat{x}^k\}$ is bounded, by our assumptions on f and $\hat{\kappa}$, the sequence $\{f(\hat{x}^k) + \hat{\kappa}\}$ is bounded below, i.e., $\{\hat{f}^k\}$ is bounded below. Since $\{\hat{f}^k\}$ is also nonincreasing, we conclude that it converges.

Using (4.1), we have that,

$$0 \le m \sum_{k=1}^{p} \delta_{k+1} \le \sum_{k=1}^{p-1} \left(\hat{f}^k - \hat{f}^{k+1}\right), \tag{35}$$

so that, letting $p \to \infty$,

$$0 \le m \sum_{k=1}^{\infty} \delta_{k+1} \le \hat{f}^1 - \lim_{k\to\infty} \hat{f}^k. \tag{36}$$

As a result,

$$\sum_{k=1}^{\infty} \delta_{k+1} = \sum_{k=1}^{\infty} \left(e_{\tilde{\varphi}} + \frac{R_k + \mu_k}{2} |x^{k+1} - \hat{x}^k|^2 \right) < +\infty. \tag{37}$$

Hence, $\delta_{k+1} \to 0$, as $k \to \infty$. Since all the quantities above are nonnegative, it also holds that

$$e_{\tilde{\varphi}} \to 0, \quad \frac{R_k + \mu_k}{2} |x^{k+1} - \hat{x}^k|^2 \to 0, \quad \text{as} \quad k \to \infty. \tag{38}$$

If $\sum_{k=1}^{\infty} \frac{1}{\mu_k} = +\infty$, but for some $\beta > 0$ and all k, $\left|\hat{g}^k\right| \ge \beta$, then (37) with the relation (16) results in a contradiction. The fact shows that no such β exists, by the relationship between $\hat{g}^k$ and $x^{k+1} - \hat{x}^k$ in (16), means precisely that there exists an index set $K \subset \{1, 2 \cdots\}$ such that

$$\hat{g}^k \to 0, \qquad K \ni k \to \infty. \tag{39}$$

Passing onto a further subsequence, if necessary, we can assume that $\{\hat{x}^k\} \to \bar{x}$, and $\hat{g}^k \to \bar{g}$ as $K \ni k \to \infty$. Item (i) is now proved.

If $\liminf_{k\to\infty} \frac{1}{\mu_k} > 0$, the second relation in (38) readily implied (39) for $K \subset \{1, 2 \cdots\}$. Thus, the same assertions can be seen to hold for all accumulation points of $\{\hat{x}^k\}$. This completes the proof. □

In the following, we consider the case where the number of serious step is finite. That means after a finite number of iterations, the stability center does not change, i.e., there exists an index $\bar{k}$ and from then on, the stability center is $\hat{x}^k = \hat{x}$ for all $k > \bar{k}$, and only null steps follow. The following formula is referred in the proof process which are easy to be verified by the corresponding definitions. Specifically, by (8), we see that,

$$\begin{aligned} & - e^{k+1}_{\varphi,k+1} + \left\langle g^{k+1}_{\varphi,k+1}, x^{k+1} - \hat{x}^k \right\rangle \qquad (40) \\ &= -e^{k+1}_{k+1} - \{\frac{\mu_k}{2}|x^{n+1} - \hat{x}^k|^2\}^{k+1}_{k+1} + \left\langle g^{k+1} + \mu_{k+1}\left(x^{k+1} - \hat{x}^k\right), x^{k+1} - \hat{x}^k \right\rangle \\ &= -\left(\hat{f}^k - f^{k+1} - \left\langle g^{k+1}, \hat{x}^k - x^{k+1}\right\rangle\right) - \frac{\mu_{k+1}}{2}\left|x^{k+1} - \hat{x}^k\right|^2 \\ &\quad + \left\langle g^{k+1}, x^{k+1} - \hat{x}^k \right\rangle + \mu_{k+1}\left|x^{k+1} - \hat{x}^k\right|^2 \\ &= f^{k+1} - \hat{f}^k + \frac{\mu_{k+1}}{2}\left|x^{k+1} - \hat{x}^k\right|^2. \end{aligned}$$

Hence, whenever x^{k+1} is declared as a null step, the descent condition (30) does not hold. Meanwhile, by (40), we have

$$-e^{k+1}_{\varphi,k+1} + \left\langle g^{k+1}_{\varphi,k+1}, x^{k+1} - \hat{x}^k \right\rangle \geq -m\delta_{k+1}. \qquad (41)$$

4.2. **Finite serious steps followed by infinitely many null steps.**

Theorem 4.3. *Let a finite number of serious iterates be followed by infinite null steps. Let the Assumption 3.3 hold and* $\liminf_{k\to\infty} \frac{1}{\mu_k} > 0$. *Then* $\{\hat{x}^k\} \to \hat{x}$, $\delta_{k+1} \to 0$, $e_{\tilde{\varphi}} \to 0$, *and there exist* $K \subset \{1, 2, \ldots\}$ *and* $\bar{g}$ *such that* $\hat{g}^k \to \bar{g}$ *as* $K \ni k \to \infty$. *In particular, the conclusions of Lemma 4.1 hold for* $\hat{x} = \bar{x}$ *if Assumption 3.2 holds.*

Proof. Let k large enough, so that $k \geq \bar{k}$ and $\hat{x}^k = \hat{x}$, $\hat{f}^k = \hat{f}$ are fixed.

Define the optimal value of the subproblem (12) by,

$$\Psi^k = \widetilde{\varphi}_k(x^{k+1}) + \frac{\mu_k}{2}|x^{k+1} - \hat{x}^k|^2. \qquad (42)$$

We first show that the sequence $\{\Psi^k\}$ is bounded above. Recall that, by (17),

$$\varphi_{-k}(\hat{x}^k) = \widetilde{\varphi}_k\left(x^{k+1}\right) - \left\langle \hat{g}^k, x^{k+1} - \hat{x}^k \right\rangle. \qquad (43)$$

We then obtain that

$$\begin{aligned} \Psi^k + \frac{\mu_k}{2}|x^{k+1} - \hat{x}^k|^2 &= \varphi_{-k}(\hat{x}^k) + \left\langle g^k, x^{k+1} - \hat{x}^k \right\rangle + \mu_k|x^{k+1} - \hat{x}^k|^2 \\ &= \varphi_{-k}(\hat{x}^k) \leq \widetilde{\varphi}_k(\hat{x}^k) = \hat{f}^k, \end{aligned}$$

since $\Psi^k \leq \hat{f}^k$, it can be shown that $\{\Psi^k\}$ is bounded above.

We next show that $\{\Psi^n\}$ is increasing. To that end, we obtain that

$$\begin{aligned} \Psi^{k+1} &= \widetilde{\varphi}_{k+1}(x^{k+2}) + \frac{\mu_{k+1}}{2}|x^{k+2} - \hat{x}^{k+1}|^2 \\ &\geq \varphi_{-k}(x^{k+2}) + \frac{\mu_k}{2}|x^{k+2} - \hat{x}^k|^2 \\ &= \widetilde{\varphi}_k(x^{k+1}) + \left\langle \hat{g}^k, x^{k+2} - x^{k+1} \right\rangle + \frac{\mu_k}{2}|x^{k+2} - \hat{x}^{k+1}|^2 \\ &= \Psi^k - \frac{\mu_k}{2}|x^{k+1} - \hat{x}^k|^2 - \mu_k\left\langle d^k, d^{k+1} - d^k \right\rangle + \frac{\mu_k}{2}|x^{k+2} - \hat{x}^{k+1}|^2 \\ &\geq \Psi^k + \frac{\mu_k}{2}|d^{k+1} - d^k|^2, \end{aligned}$$

where d^k is defined in (16).

It converges when the sequence $\{\Psi^k\}$ is bounded above and increasing. Consequently, taking also into account that $\mu_k \geq \mu_{\bar{k}}$, it follows that,

$$|d^{k+1} - d^k| \to 0, \quad k \to \infty. \tag{44}$$

Next by the definition of δ_{k+1} in (23) and the characterization of $e_{\tilde{\varphi}}$ in (20),

$$\begin{aligned}\hat{f} &= \delta_{k+1} - \frac{\eta_k}{2}\left|x^{k+1} - \hat{x}^k\right|^2 + \widetilde{\varphi}_k\left(x^{k+1}\right) \\ &\geq \delta_{k+1} + \widetilde{\varphi}_k\left(x^{k+1}\right),\end{aligned}$$

therefore,

$$\delta_{k+1} \leq \hat{f}^k - \widetilde{\varphi}_k\left(x^{k+1}\right). \tag{45}$$

According to the model assumptions, written for $d = d^{k+1}$,

$$-\hat{f}^{k+1} + e^{k+1}_{\varphi,k+1} - \left\langle g^{k+1}_{\varphi,k+1}, d^{k+1}\right\rangle \geq -\widetilde{\varphi}_{k+1}\left(\hat{x} + d^{k+1}\right).$$

As $\hat{f}^{k+1} = \hat{f}$, adding condition (41) to the inequality above, we obtain that,

$$m\delta_{k+1} + \left\langle g^{k+1}_{\varphi,k+1}, d^{k+1} - d^k\right\rangle \geq \hat{f} - \widetilde{\varphi}_{k+1}\left(\hat{x} + d^{k+1}\right).$$

Combining this relation with (45) yields,

$$0 \leq \delta_{k+2} \leq m\delta_{k+1} + \left\langle g^{k+1}_{\varphi,k+1}, d^{k+1} - d^k\right\rangle.$$

Since $m \in (0,1)$, by (44) and $\{\eta_k\}$ is bounded, $\left\langle g^{k+1}_{\varphi,k+1}, x^{k+1} - \hat{x}^k\right\rangle \to 0$, as $k \to \infty$, using [[39] Lemma 3], it follows that,

$$\lim_{k\to\infty} \delta_{k+1} = 0.$$

Since $\delta_{k+1} = \frac{R_k+\mu_k}{2}|x^{k+1}-x^k|^2+e_{\widetilde{\varphi}}$ and $\liminf_{k\to\infty} \frac{1}{\mu_k} > 0$, we have $\lim_{k\to\infty} e_{\widetilde{\varphi}} = 0$ and $\lim_{k\to\infty} |\hat{g}^k| = 0$. Also $\lim_{k\to\infty} d^k = 0$, so that $\lim_{k\to\infty} x^k = \hat{x}$. Passing onto a subsequence if necessary, we may also conclude that $\hat{g}^k$ converges to some $\bar{g}$. Finally, as $\hat{x}^k = \bar{x}$ for all k, we clearly have all of conditions in Lemma 4.1 fulfilled. This completes the proof. □

5. **Conclusion.** In this paper, we provide an implementable algorithm for solving nonconvex nonsmooth optimization problems with inexact information. Our method just requires that the objective function is a $lower-C^2$ function, and only gets inexact function values and subgradients, this assumption covers a rich family of interesting problems.

By redefining the linearization error, we can deal with the inexact information. Based on the convexification techniques, a locally convex model is created by the construction of the $lower-C^2$ function. Then we can solve the minimization problem by creating a quadratic subproblem. Finally, we present redistributed proximal bundle method. Any nonnegative initial values can be used for the convexification parameter and prox-parameter in the algorithm, and they are changed according to specific rules while the algorithm is executed. By specifying the inexact information of the convex function, our approach can be used to solve the convex programming problem if the convexification parameter equals to zero or any other positive constant.

Acknowledgments. The authors would like to thank the associate editor and the anonymous referees for their valuable comments and suggestions which helped a lot to improve this article. This paper is supported by the National Key Research and Development Program of China (No.2022YFB3304602), the International Postdoctoral Exchange Fellowship Program 2020 by the Office of China Postdoctoral Council (No.2020003), the National Natural Science Foundation of China (No.11701063, 11901075, 12101101, 12071342), the Project funded by China Postdoctoral Science Foundation (No.2019M651091, 2019M661073 and 2022M720631), the Fundamental Research Funds for the Central Universities of DMU(No.3132023204, 3132023205, 3132023206 and 3132022204), the Fundamental Research Funds for the Central Universities(DUT22LAB305), Dalian Youth Science and Technology Star (No.2020RQ047), the Natural Science Foundation of Liaoning Province in China (No.2020-BS-074), and the Fundamental Scientific Research Projects of Higher Education Institutions of Liaoning Provincial Department of Education (Youth Project) under Grant LJKQZ2021011, the Natural Science Foundation of Hebei Province, No.A2020202030.

REFERENCES

[1] G. d'Antonio and A. Frangioni, Convergence analysis of deflected conditional approximate subgradient methods, *J. Optim. Theory Appl.*, **20** (2009), 357-386.

[2] P. Apkarian, D. Noll and O. Prot, A proximity control algorithm to minimize nonsmooth and nonconvex semi-infinite maximum eigenvalue functions, *J. Convex Anal.*, **16** (2009), 641-666.

[3] J. R. Birge, L. Qi and Z. Wei, Convergence analysis of some methods for minimizing a nonsmooth convex function, *J. Optim. Theory Appl.*, **97** (1998), 357-383.

[4] A. Daniilidis and P. Georgiev, Approximate convexity and submonotonicity, *J. Math. Anal. Appl.*, **291** (2004), 292-301.

[5] A. Daniilidis, C. Sagastizábal and M. Solodov, Identifying structure of nonsmooth convex functions by the bundle technique, *J. Optim. Theory Appl.*, **20** (2009), 820-840.

[6] G. Emiel and C. Sagastizábal, Incremental-like bundle methods with application to energy planning, *Comp. Optim. Appl.*, **46** (2010), 305-332.

[7] W. Hare and C. Sagastizábal, Computing proximal points of nonconvex functions, *Math. Program.*, **116** (2009), 221-258.

[8] W. Hare and C. Sagastizábal, A redistributed proximal bundle method for nonconvex optimization, *J. Optim. Theory Appl.*, **20** (2010), 2442-2473.

[9] W. Hare, C. Sagastizábal and M. Solodov, A proximal bundle method for nonsmooth nonconvex functions with inexact information, *Comp. Optim. Appl.*, **63** (2016), 1-28.

[10] M. Hintermüller, A proximal bundle method based on approximate subgradients, *Comp. Optim. Appl.*, **20** (2001), 245-266.

[11] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, II Volume 306 of Grundlehren der MathematischenWissenschaften [Fundamental Principles of Mathematical Sciences], Springer, Berlin, 1993.

[12] M. Huang, C. Cheng, Y. Li and Z. Xia, The space decomposition method for the sum of nonlinear convex maximum eigenvalues and its applications, *J. Indus. Man. Optim.*, **16** (2020), 1885-1905.

[13] M. Huang, X.-J. Liang, L.-P. Pang and Y. Lu, The bundle scheme for solving arbitrary eigenvalue optimizations, *J. Indus. Man. Optim.*, **13** (2017), 659-680.

[14] M. Huang, S. Lin, J. Yuan, Z. Gong and X. Liang, A VU-decomposition technique for a class of eigenvalue optimizations, *Pac. J. Optim.*, **16** (2020), 129-142.

[15] M. Huang, Y. Lu, L. Pang and Z. Xia, A space decomposition scheme for maximum eigenvalue functions and its applications, *Math. Meth. Oper. Res.*, **85** (2017), 453-490.

[16] M. Huang, L.-P. Pang, X.-J. Liang and F.-Y. Meng, A second-order bundle method based on UV-decomposition for a special class of eigenvalue optimizations, *Numer. Func. Anal. Optim.*, **37** (2016), 554-582.

[17] M. Huang, L.-P. Pang, Y. Lu and Z.-Q. Xia, A fast space-decomposition scheme for nonconvex eigenvalue optimization, *Set-Val. Var. Anal.*, **25** (2017), 43-67.

[18] M. Huang, L.-P. Pang and Z.-Q. Xia, The space decomposition theory for a class of eigenvalue optimizations, *Comp. Optim. Appl.*, **58** (2014), 423-454.

[19] K. C. Kiwiel, A linearization algorithm for nonsmooth minimization, *Math. Oper. Res.*, **10** (1985), 185-194.

[20] K. C. Kiwiel, Approximations in proximal bundle methods and decomposition of convex programs, *J. Optim. Theory Appl.*, **84** (1995), 529-548.

[21] K. C. Kiwiel, Restricted step and Levenberg-Marquardt techniques in proximal bundle methods for nonconvex nondifferentiable optimization, *SIAM J. Optim.*, **6** (1996), 227-249.

[22] K. C. Kiwiel, Convergence of approximate and incremental subgradient methods for convex optimization, *J. Optim. Theory Appl.*, **14** (2004), 807-840.

[23] K. C. Kiwiel, A proximal bundle method with approximate subgradient linearizations, *J. Optim. Theory Appl.*, **16** (2006), 1007-1023.

[24] C. Lemaréchal, An extension of Davidon methods to nondifferentiable problems, *Math. Program. Study.*, **3** (1975), 95-109.

[25] C. Lemaréchal, Bundle methods in nonsmooth optimization, *Nonsmooth Optimization (Proceedings of IIASA Workshop, Laxenburg, 1977), IIASA Proceeding Series. Pergamon, Oxford.*, **3** (1978), 79-102.

[26] C. Lemaréchal, Lagrangian relaxation, *Computational Combinatorial Optimization, Lecture Notes in Computer Science*, Springer, Berlin, **2241** (2001), 112-156.

[27] C. Lemaréchal, A. Renaud and C. Sagastizábal, Bundle methods in stochastic optimal power management: A disaggregated approach using preconditioners, *Comput. Optim. Appl.* , **20** (2001), 227-244.

[28] C. Lemaréchal, J.-J. Strodiot and A. Bihain, On a bundle algorithm for nonsmooth optimization, *Nonlinear Programming*, **4** (1981), 245-282.

[29] S. Lin, M. Huang, Z. Xia and D. Li, Quasidifferentiabilities of the expectation functions of random quasidifferentiable functions, *Optim.*, **71** (2022), 1509-1524.

[30] Y. Lu, L. Pang, F. Guo and Z. Xia, A superlinear space decomposition algorithm for constrained nonsmooth convex program, *J. Comput. Appl. Math.*, **234** (2010), 224-232.

[31] Y. Lu, L.-P. Pang, X.-J. Liang and Z.-Q. Xia, An approximate decomposition algorithm for convex minimization, *J. Comput. Appl. Math.*, **234** (2010), 658-666.

[32] L. Lukšan and J. Vlček, A bundle-Newton method for nonsmooth unconstrained minimization, *Math. Program.*, **83** (1998), 373-391.

[33] R. Mifflin, Semismooth and semiconvex functions in constrained optimization, *SIAM J. Control Optim.*, **15** (1977), 959-972.

[34] R. Mifflin, A modification and extension of Lemaréchal's algorithm for nonsmooth minimization, *Math. Program. Study.*, **17** (1982), 77-90.

[35] A. Nedić and D. P. Bertsekas, The effect of deterministic noise in subgradient methods, *Math. Program.*, **125** (2010), 75-99.

[36] D. Noll, Bundle method for nonconvex minimization with inexact subgradients and function values, *Computational and Analytical Mathematics*, **50** (2013), 555-592.

[37] W. de Oliveira and C. Sagastizábal, Convex proximal bundle methods in depth: A unified analysis for inexact oracles, *Math. Program.*, **148** (2014), 241-277.

[38] M. Pinar and M. Teboulle, On semidefinite bounds for maximization of a non-convex quadratic objective over the L1-unit ball, *RAIRO Oper. Res.*, **40** (2006), 253-265.

[39] B. T. Polyak, *Introduction to Optimization*, Optimization Software, Inc., Publications Division, New York, 2004.

[40] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2^{nd} edition, Springer Texts in Statistics, Springer, New York, 2004.

[41] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, Berlin, 1998.

[42] C. Sagastizábal, Divide to conquer: Decomposition methods for energy optimization, *Math. Program. B.*, **134** (2012), 187-222.

[43] C. Sagastizábal and M. Solodov, An infeasible bundle method for nonsmooth convex constrained optimization without a penalty function or a filter, *SIAM J. Optim.*, **16** (2005), 146-169.

[44] M. V. Solodov, On approximations with finite precision in bundle methods for nonsmooth optimization, *J. Optim. Theory Appl.*, **119** (2003), 151-165.

[45] M. V. Solodov and S. K. Zavriev, Error stabilty properties of generalized gradient-type algorithms, *J. Optim. Theory Appl.*, **98** (1998), 663-680.
[46] J. E. Spingarn, Submonotone subdifferentials of Lipschitz functions, *Trans. Am. Math. Soc.*, **264** (1981), 77-89.
[47] Y. Yang, L. Pang, X. Ma and J. Shen, Constrained nonconvex nonsmooth optimization via proximal bundle method, *J. Optim. Theory Appl.*, **163** (2014), 900-925.

Received November 2022; revised March 2023; early access May 2023.