

Research Article

An Approximate Redistributed Proximal Bundle Method with Inexact Data for Minimizing Nonsmooth Nonconvex Functions

Jie Shen,¹ Xiao-Qian Liu,¹ Fang-Fang Guo,² and Shu-Xin Wang¹

¹*School of Mathematics, Liaoning Normal University, Dalian 116029, China*

²*School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China*

Correspondence should be addressed to Jie Shen; tt010725@163.com

Received 6 January 2015; Revised 28 April 2015; Accepted 28 April 2015

Academic Editor: Anna Pandolfi

Copyright © 2015 Jie Shen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We describe an extension of the redistributed technique from classical proximal bundle method to the inexact situation for minimizing nonsmooth nonconvex functions. The cutting-planes model we construct is not the approximation to the whole nonconvex function, but to the local convexification of the approximate objective function, and this kind of local convexification is modified dynamically in order to always yield nonnegative linearization errors. Since we only employ the approximate function values and approximate subgradients, theoretical convergence analysis shows that an approximate stationary point or some double approximate stationary point can be obtained under some mild conditions.

1. Introduction and Motivation

Consider the following unconstrained nonsmooth nonconvex optimization problem:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in R^m, \end{aligned} \quad (1)$$

where f is a function from $R^m \rightarrow R$. Nonsmooth optimization problems (NSO) arise from many fields of applications. There exist several approaches to solve these kinds of problems; see [1–4]. Bundle methods [5] are based on the cutting-planes methods, first described in [6, 7], where the convexity of the objective function is the fundamental assumption. If the objective function f is convex, tangent lines are cutting planes supporting the epigraph of f and the linearization errors are always nonnegative, and the model functions usually defined by the maximum of tangent lines are lower approximations to the objective function. However, in the nonconvex case the linearization errors can be negative, and the corresponding model function does not stay below f and may even cut off a region containing a minimizer. Very little systematic research has been performed on extending convex

bundle methods to nonconvex cases. The bundle methods for nonconvex functions [8–12] are of proximal type, which were developed in the 90's; see [13]. They still use subgradients locality measures to redefine negative linearization errors, and primal information, corresponding to function values, is again ignored.

Note that in some cases computing the exact function value is not easy. The assumptions for using approximate subgradients and approximate values of the function are realistic, for instance, the Lagrangian relaxation problem: if f is a max-type function of the form $f(y) = \sup\{F_z(y) \mid z \in Z\}$, where each $F_z(y)$ is convex and Z is an infinite set, then it may be impossible to calculate $f(y)$ since f itself is defined by a minimization problem involving another function F . However, we may still consider two cases. In the first case, for each positive $\varepsilon > 0$ one can find an element $z_y \in Z$ satisfying $F_{z_y}(y) \geq f(y) - \varepsilon$; in the second case, this may be possible only for some fixed (any possibly unknown) $\varepsilon < \infty$. In both cases we may set $\bar{f}_y = F_{z_y}(y) \geq f(y) - \varepsilon$. Besides that, the study of approximate subgradients of convex functions is deserved since in some cases a subgradient $\xi(x) \in \partial f(x)$ is expensive to compute. But if we know an already computed subgradient

$\xi(\bar{x}) \in \partial f(\bar{x})$, where \bar{x} is near to x , then we have $\xi(\bar{x}) \in \partial_\varepsilon f(x)$ because

$$\begin{aligned} f(x) + \xi(\bar{x})^T(z - x) &= f(\bar{x}) + \xi(\bar{x})^T(z - \bar{x}) + \varepsilon \\ &\leq f(z) + \varepsilon, \quad \forall z \in R^m, \end{aligned} \quad (2)$$

where $\varepsilon = f(x) - f(\bar{x}) - \xi(\bar{x})^T(x - \bar{x}) \geq 0$. For more details and papers involving the approximate function values and subgradients, we refer to [14–16] and the references therein.

In our work, we attempt to explore the possibility of using the approximate function values and approximate subgradients of f instead of the exact ones for solving problem (1). From the point of view of primal problem, by separating the traditional prox-parameter into two parts, local convexification parameter and new model prox-parameter, and employing the inexact information of the objective function, we construct a cutting-planes model which is the approximation to the local convexification function, and the iterate points are obtained approximately by computing its double approximate proximal points. The cutting-planes model is special in the sense that it no longer approximates the objective function, but rather certain local convexification, centered at the proximal center.

This paper is organized as follows: in Section 2, some preliminary results and assumptions required in our paper are provided. In Section 3 we pay more attention to the primal pattern of (1) instead of the dual insight and the cutting-planes model of local approximation to the objective function is constructed in this part. In Section 4, the concrete approximate redistributed algorithm for solving (1) is presented. Convergence results are examined and discussed in Section 5, and the iterate points generated by the proposed algorithm converge to an approximate (double approximate) stationary point of the objective function. In the last section, some conclusions are given.

2. Preliminaries and Assumptions

In this part, we first present some basic definitions and results [17].

- (i) The **regular subdifferential** of f at \bar{x} is defined by

$$\begin{aligned} \widehat{\partial}f(\bar{x}) &= \left\{ g \right. \\ &\left. \in R^m \mid \liminf_{x \rightarrow \bar{x}, x \neq \bar{x}} \frac{f(x) - f(\bar{x}) - \langle g, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0 \right\}. \end{aligned} \quad (3)$$

- (ii) The **limiting subdifferential** of f at \bar{x} is defined by

$$\partial f(\bar{x}) = \lim_{x \rightarrow \bar{x}} \sup_{f(x) \rightarrow f(\bar{x})} \widehat{\partial}f(x). \quad (4)$$

If f is finite at \bar{x} , $\widehat{\partial}f(\bar{x})$ and $\partial f(\bar{x})$ are closed and $\widehat{\partial}f(\bar{x})$ is convex.

- (iii) For given $\varepsilon > 0$, **ε -limiting subdifferential** of f at \bar{x} is defined by

$$\partial_\varepsilon f(\bar{x}) = \text{conv} \{ \partial f(y) \mid \|y - \bar{x}\| \leq \varepsilon \}. \quad (5)$$

We call elements of this approximate subdifferential approximate subgradients. And if f is finite at \bar{x} , $\partial_\varepsilon f(\bar{x})$ is convex and closed.

- (iv) We say f is prox-bounded if there exists $R \geq 0$ such that $f + (R/2)\|\cdot\|^2$ is bounded below. The corresponding threshold is the smallest $r_{pb} \geq 0$ such that $f + (R/2)\|\cdot\|^2$ is bounded below for all $R > r_{pb}$.
- (v) The function f is lower- C^2 on an open set O if f is finite on O and for any x in O there exists a threshold $r_{lc2} > 0$ such that $f + (r/2)\|\cdot\|^2$ is convex on an open neighbourhood O' of x for all $r > r_{lc2}$.

Next we give some assumptions needed in our paper.

Assumption 1. For fixed accuracy tolerances $\varepsilon_f \geq 0$, $\varepsilon_g \geq 0$, for each $x \in R^m$, the oracle can provide an approximate function value $\tilde{f}_x \in [f(x) - \varepsilon_f, f(x) + \varepsilon_g]$ and an approximate subgradient g_x^ε of f at x , where $\varepsilon = \varepsilon_f + \varepsilon_g$.

Assumption 2. Given $x^0 \in R^m$ and $M_0 > 0$, there exist an open bounded set O and a function H such that $L_0 = \{x \in R^m \mid \tilde{f}_x \leq \tilde{f}_{x^0} + M_0\} \subset O$, and H is lower- C^2 on O satisfying $H \equiv \tilde{f}$ on L_0 .

Lemma 3. For a function f satisfying Assumptions 1 and 2, the following conclusions hold:

- (a) The approximate level set L_0 is nonempty and compact.
- (b) The approximate function \tilde{f} is bounded below and prox-bounded with threshold $r_{pb} = 0$.
- (c) There exists $\rho^{id} > 0$ such that, for any $\rho \geq \rho^{id}$ and given $y \in L_0$, the function $f + (\rho/2)\|\cdot - y\|^2$ is convex on L_0 .
- (d) The approximate function \tilde{f} is Lipschitz continuous on L_0 .

Proof. Since H is the lower- C^2 function defined in by Assumption 2 which agrees with \tilde{f} on L_0 , H is continuous and finite valued on the open set O , which contains L_0 . Thus L_0 is closed. By the fact that L_0 is bounded from the boundedness of O , (a) is proved. For item (b), without loss of generality, we assume that \tilde{f} and H are the same function. Since \tilde{f} is finite and continuous on the approximate compact level set L_0 , this implies \tilde{f} is bounded below. Any function which is bounded below is prox-bounded with threshold $r_{pd} = 0$. Item (c) can be obtained by [17]. The last item (d) is obvious by noting that all lower- C^2 functions are locally Lipschitz continuous. \square

Define the approximate proximal point mapping

$$\tilde{p}_R \tilde{f}(x) = \arg \min_{y \in R^m} \left\{ \tilde{f}_y + \frac{R}{2} \|y - x\|^2 \right\}, \quad x \in L_0, \quad (6)$$

which is single-valued and Lipschitz continuous on L_0 , provided R is sufficiently large. By imitating the conclusion in [18] and the optimality condition, we define a new kind of approximate stationary point of the objective function.

Definition 4. One calls $x^* \in L_0$ an approximate stationary point of f if $x^* = \tilde{p}_R \tilde{f}(x^*)$, where $\tilde{p}_R \tilde{f}(x^*)$ is defined by (6), written with x replaced by x^* .

Note that R is sufficiently large meaning that $R > \rho^{id}$, where ρ^{id} is the value in item (c) in Lemma 3 and

$$\tilde{p}_R \tilde{f}(x) = \tilde{p}_{R-\rho^{id}} \left(\tilde{f} + \frac{\rho^{id}}{2} \|\cdot - x\|^2 \right) (x). \quad (7)$$

The above relation with the local convexification property plays a fundamental role in the development of our algorithm if we have already known the ideal proximal threshold ρ^{id} .

3. Construction of the Model

For a convex function f , the exact linearization error of f at \hat{x}^k is defined by

$$e_i^k = f(\hat{x}^k) - (f(x^i) + \langle g^i, \hat{x}^k - x^i \rangle), \quad (8)$$

where $g^i \in \partial f(x^i)$ and \hat{x}^k is the current stability center. Obviously we have $e_i^k \geq 0$, and the reformulated bundle data consists of $\{\hat{x}^k, f(\hat{x}^k), g^k \in \partial f(\hat{x}^k)\}$ and the approximate subgradients $\cup_{i \in I_n} \{(e_i^k, g^i \in \partial_{e_i^k} f(\hat{x}^k))\}$. In our method, at any iteration, bundle method keeps memory of the iterative process in a bundle of inexact information:

$$\bigcup_{i \in I_n} \{(x^i, \tilde{f}_i, g_\varepsilon^i \in \partial_\varepsilon f(x^i))\}, \quad (9)$$

where $I_n \subseteq \{0, 1, 2, \dots, n\}$, $\tilde{f}_i \in [f(x^i) - \varepsilon_f, f(x^i) + \varepsilon_g]$, and $\tilde{f}_{\hat{x}^{k(n)}}$ is the best approximate value obtained until iteration n , evaluated at the serious step k , corresponding to some past iterate i_k . For nonconvex function f , we work with augmented functions

$$\tilde{f}_{\eta_n}^{\hat{x}^k} := \tilde{f} + \frac{\eta_n}{2} \|\cdot - \hat{x}^k\|^2. \quad (10)$$

We consider an augmented bundle of inexact information: $\cup_{i \in I_n} \{(\tilde{e}_i^k, d_i^k, \Delta_i^k, g_\varepsilon^i)\}$, where

$$\begin{aligned} \tilde{e}_i^k &= \tilde{f}_{\hat{x}^k} - (\tilde{f}_i + \langle g_\varepsilon^i, \hat{x}^k - x^i \rangle), \\ d_i^k &= \frac{\|x^i - \hat{x}^k\|^2}{2}, \\ \Delta_i^k &= x^i - \hat{x}^k, \\ g_\varepsilon^i &\in \partial_\varepsilon f(x^i). \end{aligned} \quad (11)$$

Note that the following relation holds:

$$\tilde{p}_R \tilde{f}(\hat{x}) = \tilde{p}_\mu (\tilde{f}_{\eta}^{\hat{x}})(\hat{x}), \quad (12)$$

where $\tilde{f}_{\eta}^{\hat{x}} = \tilde{f} + (\eta/2) \|\cdot - \hat{x}\|^2$, $R = \mu + \eta$, and η and μ are called the convexification parameter and model

prox-parameter, respectively. We use the past information in bundle to construct a cutting-planes model of the function $\tilde{f}_{\eta}^{\hat{x}}$:

$$\begin{aligned} \tilde{\phi}(y) &:= \max_i \left\{ \left(\tilde{f}_i + \frac{\eta}{2} \|x^i - \hat{x}\|^2 \right) \right. \\ &\quad \left. + \langle g_\varepsilon^i + \eta(x^i - \hat{x}), y - x^i \rangle - 2\varepsilon \right\}. \end{aligned} \quad (13)$$

An equivalent expression, written with all the iteration indices, is the following:

$$\begin{aligned} \tilde{\phi}_n(y) &= \tilde{f}_{\hat{x}^k} + \max_{i \in I_n} \left\{ -(\tilde{e}_i^k + \eta_n d_i^k + 2\varepsilon) \right. \\ &\quad \left. + \langle (g_\varepsilon^i + \eta_n \Delta_i^k), y - \hat{x}^k \rangle \right\}. \end{aligned} \quad (14)$$

The next candidate point is chosen as $x^{n+1} = \tilde{p}_{\mu_n} \tilde{\phi}_n(\hat{x}^k)$. The corresponding optimality condition is that there exists $\alpha^n \in S_n$ such that

$$\begin{aligned} x^{n+1} &= \hat{x}^k - \frac{1}{\mu_n} g_{\eta_n, \varepsilon}^{-n}, \\ g_{\eta_n, \varepsilon}^{-n} &= \sum_{i \in I_n} \alpha_i^n (g_\varepsilon^i + \eta_n \Delta_i^k) \in \partial \tilde{\phi}_n(x^{n+1}), \end{aligned} \quad (15)$$

where S_n denotes the unit simplex in R^{I_n} . Denote $J_n^{\text{act}} := \{i \in I_n \mid \alpha_i^n > 0 \text{ in (15)}\}$. $g_{\eta_n, \varepsilon}^{-n}$ is called the aggregate approximate subgradient, and the corresponding aggregate bundle element is the quadruplet:

$$\begin{aligned} (\tilde{e}_{-n}^k, d_{-n}^k, \Delta_{-n}^k, g_\varepsilon^{-n}) &:= \sum_{i \in I_n} (\tilde{e}_i^k, d_i^k, \Delta_i^k, g_\varepsilon^i) \\ &= \sum_{i \in J_n^{\text{act}}} (\tilde{e}_i^k, d_i^k, \Delta_i^k, g_\varepsilon^i). \end{aligned} \quad (16)$$

Note that for $l = -n$ and all $l \in J_n^{\text{act}}$ we have

$$\begin{aligned} \tilde{\phi}_n(x^{n+1}) &= \tilde{f}_{\hat{x}^k} - (\tilde{e}_l^k + \eta_n d_l^k + 2\varepsilon) \\ &\quad + \langle (g_\varepsilon^l + \eta_n \Delta_l^k), x^{n+1} - \hat{x}^k \rangle, \\ g_{\eta_n, \varepsilon}^{-n} &= g_\varepsilon^{-n} + \eta_n \Delta_{-n}^k = \mu_n (\hat{x}^k - x^{n+1}). \end{aligned} \quad (17)$$

For all $i \in I_n$, \tilde{e}_i^{k+1} , d_i^{k+1} , and Δ_i^{k+1} can be updated according to the following formula:

$$\begin{aligned} \tilde{e}_i^{k+1} &= \tilde{e}_i^k + \tilde{f}_{\hat{x}^{k+1}} - \tilde{f}_{\hat{x}^k} - \langle g_\varepsilon^i, \hat{x}^{k+1} - \hat{x}^k \rangle, \\ d_i^{k+1} &= d_i^k + \frac{\|\hat{x}^{k+1} - \hat{x}^k\|^2}{2} - \langle \Delta_i^k, \hat{x}^{k+1} - \hat{x}^k \rangle, \\ \Delta_i^{k+1} &= \Delta_i^k + \hat{x}^k - \hat{x}^{k+1}. \end{aligned} \quad (19)$$

In classical bundle method with inexact information for convex functions, the bundle consists of the pair $(\tilde{e}_i^k, g_\varepsilon^i)$ for which $g_\varepsilon^i \in \partial_{2\varepsilon + \tilde{e}_i^k} f(\hat{x}^k)$ holds whenever $2\varepsilon + \tilde{e}_i^k \geq 0$, which

can be seen by noting that $g_\varepsilon^i \in \partial f(x^i)$, the definition of \tilde{e}_k^i , and

$$\begin{aligned} f(y) &\geq f(x^i) + \langle g_\varepsilon^i, y - x^i \rangle - \varepsilon \\ &= f(\hat{x}^k) + \langle g_\varepsilon^i, y - \hat{x}^k \rangle - f(\hat{x}^k) + f(x^i) \\ &\quad + \langle g_\varepsilon^i, \hat{x}^k - x^i \rangle - \varepsilon \\ &\geq f(\hat{x}^k) + \langle g_\varepsilon^i, y - \hat{x}^k \rangle - \tilde{f}_{\hat{x}^k} - \varepsilon_f + \tilde{f}_{x^i} - \varepsilon_g \\ &\quad + \langle g_\varepsilon^i, \hat{x}^k - x^i \rangle - \varepsilon \\ &= f(\hat{x}^k) + \langle g_\varepsilon^i, y - \hat{x}^k \rangle - \tilde{e}_i^k - 2\varepsilon. \quad \forall y \in R^m. \end{aligned} \quad (20)$$

For our method, this pair is replaced by a quadruplet $(\tilde{e}_i^k, d_i^k, \Delta_i^k, g_\varepsilon^i)$ for which the relation

$$g_\varepsilon^i + \eta_n \Delta_i^k \in \partial_{\tilde{e}_i^k + \eta_n d_i^k + 2\varepsilon} \tilde{\phi}_n(\hat{x}^k), \quad (21)$$

whenever $\tilde{e}_i^k + \eta_n d_i^k + 2\varepsilon \geq 0$, $\forall i \in I_n$,

holds, since for all $i \in I_n$, whenever $\tilde{e}_i^k + \eta_n d_i^k + 2\varepsilon \geq 0$, we have

$$\begin{aligned} \tilde{\phi}_n(y) &\geq \tilde{f}_{\hat{x}^k} - (\tilde{e}_i^k + \eta_n d_i^k + 2\varepsilon) \\ &\quad + \langle g_\varepsilon^i + \eta_n \Delta_i^k, y - \hat{x}^k \rangle \\ &\geq \tilde{\phi}_n(\hat{x}^k) - (\tilde{e}_i^k + \eta_n d_i^k + 2\varepsilon) \\ &\quad + \langle g_\varepsilon^i + \eta_n \Delta_i^k, y - \hat{x}^k \rangle. \end{aligned} \quad (22)$$

We want to make the parameter η_n asymptotically estimate the ideal convexity threshold ρ^{id} , and when η_n is sufficiently large, $\tilde{f} + (\eta_n/2)\|\cdot - \hat{x}^k\|^2$ is a convex function. As a result, the model function $\tilde{\phi}_n$ becomes eventually a lower approximation to a locally convex function $\tilde{f}_{\rho^{id}}^{\hat{x}^k}$. Set

$$\bar{\eta}_n := \max_{i \in I_n, d_i^{k(n)} > 0} - \frac{\tilde{e}_i^{k(n)} + 2\varepsilon}{d_i^{k(n)}}, \quad (23)$$

and clearly $\tilde{e}_i^k + \eta_n d_i^k + 2\varepsilon \geq 0$ for all $i \in I_n$ whenever $\eta \geq \bar{\eta}_n$.

Remark 5. By comparing with $\bar{\eta}_n$ [19] and η_n^{\min} [18], we find that $\bar{\eta}_n \geq \eta_n^{\min} \geq \bar{\eta}_n$, which means that the domain of η that ensures that the nonnegativeness of the linearization errors is enlarged.

4. Algorithm

Algorithm 6 (approximate redistributed proximal bundle method). Consider the following steps.

Step 1 (initialization). Choose an initial point x^0 , two parameters $M_0 > 0$, $R_0 > 0$, one stopping tolerance $\text{TOL}_{\text{stop}} \geq 0$, an Armijo-like parameter $\bar{m} \in (0, 1)$, and a convexification

growth parameter $\tau > 1$. Set $n = 0$, $k = k(n) = 0$, $i_0 = 1$, $I_0 = \{0\}$, and $x^0 = \hat{x}^0$. Compute $\tilde{f}_{x^0} \in [f(x^0) - \varepsilon_f, f(x^0) + \varepsilon_g]$, $g_\varepsilon^0 \in \partial_\varepsilon f(x^0)$, and the additional bundle information $(\tilde{e}_0^0, d_0^0, \Delta_0^0) := (0, 0, 0)$. Choose $(\mu_0, \eta_0) = (R_0, 0)$.

Step 2 (computation of trail point). Having the current serious step \hat{x}^k , the bundle $\{(\tilde{e}_i^k, d_i^k, \Delta_i^k, g_\varepsilon^i)\}_{i \in I_n}$, and the prox-parameter distribution (μ_n, η_n) with $\eta_n \leq R_n$, $\mu_n = R_n - \eta_n$, define the convex piecewise linear approximate model function $\tilde{\phi}_n$ from (14). Compute

$$x^{n+1} = \tilde{p}_{\mu_n} \tilde{\phi}_n(\hat{x}^k). \quad (24)$$

The optimal simplicial multipliers α^n from (15) and the aggregate quadruplet from (16) are available. Define the predicted decrease

$$\delta_{n+1} = \left(\tilde{f}_{\hat{x}^k} + \frac{\eta_n}{2} \|x^{n+1} - \hat{x}^k\|^2 + 2\varepsilon \right) - \tilde{\phi}_n(x^{n+1}). \quad (25)$$

Step 3 (bundle management). Compute $\tilde{f}_{x^{n+1}}$ and $g_\varepsilon^{n+1} \in \partial_\varepsilon f(x^{n+1})$. Set

$$\begin{aligned} \Delta_{n+1}^k &:= x^{n+1} - \hat{x}^k, \\ d_{n+1}^k &:= \frac{\|x^{n+1} - \hat{x}^k\|^2}{2}, \\ \tilde{e}_{n+1}^k &:= \tilde{f}_{\hat{x}^k} - (\tilde{f}_{x^{n+1}} + \langle g_\varepsilon^{n+1}, x^{n+1} - \hat{x}^k \rangle). \end{aligned} \quad (26)$$

Choose a new index set I_{n+1} satisfying

$$\begin{aligned} I_{n+1} &\supseteq \{n+1, i_k\} \\ \text{and either } I_{n+1} &\supseteq J_n^{\text{act}} := \{i \in I_n \mid \alpha_i^n > 0\}, \\ \text{or } I_{n+1} &\supseteq \{-n\}. \end{aligned} \quad (27)$$

Step 4 (descent test). If

$$\tilde{f}_{x^{n+1}} \leq \tilde{f}_{\hat{x}^k} - \bar{m}(\delta_{n+1} - 2\varepsilon), \quad (28)$$

declare a serious step: set $k(n+1) = k+1$, $i_{k+1} = n+1$, and $\hat{x}^{k+1} = x^{n+1}$, and update bundle elements according to (19).

Otherwise, declare a null step: set $k(n+1) = k(n)$.

Step 5 (update of local convexification parameter). Set

$$\begin{aligned} \eta_{n+1} &:= \eta_n, \quad \text{if } \bar{\eta}_{n+1} \leq \eta_n; \\ \eta_{n+1} &:= \tau \bar{\eta}_{n+1}, \quad \text{if } \bar{\eta}_{n+1} > \eta_n, \\ R_n &= \mu_n + \eta_{n+1}, \quad \text{if } \bar{\eta}_{n+1} > \eta_n, \end{aligned} \quad (29)$$

where $\bar{\eta}_{n+1}$ is given in (23), written with n replaced by $n+1$.

Step 6 (update of model prox-parameter). If $\tilde{f}_{x^{n+1}} > \tilde{f}_{\hat{x}^k} + M_0$, restart the algorithm by setting

$$\begin{aligned} \eta_0 &:= \eta_n, \\ \mu_0 &:= \tau\mu_n, \\ R_0 &:= \eta_0 + \mu_0, \\ x^0 &:= \hat{x}^k, \\ (\tilde{e}_0^0, d_0^0, \Delta_0^0) &:= (0, 0, 0), \\ k(0) &:= 0, \\ i_0 &:= 0, \\ I_0 &:= \{0\}, \\ n &:= 0, \end{aligned} \quad (30)$$

and loop to Step 2.

Otherwise go to Step 7.

Step 7 (stopping criterion). If $\delta_{n+1} - 2\varepsilon \leq \text{TOL}_{\text{stop}}$, then stop with the message “Algorithm successfully terminated at x^{n+1} .”

Otherwise, in case of serious step increase k by 1. In all cases increase n by 1, and loop to Step 2.

From the definition of δ_{n+1} , we obtain its equivalent expression

$$\delta_{n+1} = \tilde{e}_{-n}^k + \eta_n d_{-n}^k + 4\varepsilon + \frac{R_n + \mu_n}{2} \|x^{n+1} - \hat{x}^k\|^2 \geq 0. \quad (31)$$

Lemma 7. Consider the sequence $\{x^n\}$ generated by Algorithm 6. If f satisfies Assumptions 1 and 2, and $I_n \supseteq \{i_k\}$, there can be only a finite number of restarts in Step 6. Eventually the sequence $\{x^n\}$ lies in L_0 and the sequence $\{\mu_n\}$ becomes constant.

Proof. Because the functions $\tilde{\phi}_n$ are convex, iterates x^{n+1} are always well defined. By Lemma 3(d), the function \tilde{f} is Lipschitz continuous on L_0 and let L be the Lipschitz constant. Take $\bar{\varepsilon} = M_0/L$, and then for any $\bar{x} \in \{x \in R^m \mid \tilde{f}_x \leq \tilde{f}_{x^0}\}$, the open ball $B_{\bar{\varepsilon}}(\bar{x}) \subseteq L_0$. Note that

$$\begin{aligned} \tilde{p}_{\mu_n} \tilde{\phi}_n(\hat{x}^k) &= \arg \min_y \left\{ \tilde{\phi}_n(y) + \frac{\mu_n}{2} \|y - \hat{x}^k\|^2 \right\} \\ &\in \left\{ y \mid \tilde{\phi}_n(y) + \frac{\mu_n}{2} \|y - \hat{x}^k\|^2 \leq \tilde{\phi}_n(\hat{x}^k) \right. \\ &\quad \left. + \frac{\mu_n}{2} \|\hat{x}^k - \hat{x}^k\|^2 \right\}. \end{aligned} \quad (32)$$

Since $i_k \in I_n$, we have $g_\varepsilon^{i_k} \in \partial_{2\varepsilon} \tilde{\phi}_n(\hat{x}^k)$. From the convexity of $\tilde{\phi}_n$, we know there exists $\bar{L} > 0$ such that $\|g_\varepsilon^{i_k}\| \leq \bar{L}$, and hence

$$\begin{aligned} \tilde{p}_{\mu_n} \tilde{\phi}_n(\hat{x}^k) &\in \left\{ y \mid -2\varepsilon + \tilde{\phi}_n(\hat{x}^k) + \langle g_\varepsilon^{i_k}, y - \hat{x}^k \rangle \right. \\ &\quad \left. + \frac{\mu_n}{2} \|y - \hat{x}^k\|^2 \leq \tilde{\phi}_n(\hat{x}^k) \right\} \in \left\{ y \mid -\|g_\varepsilon^{i_k}\| \right. \\ &\quad \left. \cdot \|y - \hat{x}^k\| + \frac{\mu_n}{2} \|y - \hat{x}^k\|^2 \leq 2\varepsilon \right\} \\ &\in \left\{ y \mid \frac{\mu_n}{2} \|y - \hat{x}^k\|^2 \leq \bar{L} \|y - \hat{x}^k\| + 2\varepsilon \right\} \\ &= \left\{ y \mid \|y - \hat{x}^k\| \leq \frac{\bar{L}}{\mu_n} + \sqrt{\frac{\bar{L}^2}{\mu_n^2} + \frac{4\varepsilon}{\mu_n}} \right\}. \end{aligned} \quad (33)$$

During each Step 6, μ_n increases, eventually we will have $\bar{L}/\mu_n + \sqrt{\bar{L}^2/\mu_n^2 + 4\varepsilon/\mu_n} \leq \bar{\varepsilon}$, and it means that $x^{n+1} = \tilde{p}_{\mu_n} \tilde{\phi}_n(\hat{x}^k) \in B_{\bar{\varepsilon}}(\hat{x}^k)$. So $\tilde{f}_{x^{n+1}} - \tilde{f}_{\hat{x}^k} \leq L\|x^{n+1} - \hat{x}^k\| \leq L\bar{\varepsilon} = M_0$; that is, $\tilde{f}_{x^{n+1}} \leq \tilde{f}_{\hat{x}^k} + M_0$. \square

5. Convergence Results

Lemma 8. For the function $\tilde{\phi}_n$, the following conclusions hold:

(a) One always has

$$\tilde{\phi}_n \text{ is a convex function.} \quad (34)$$

(b) If $\eta_n \leq \bar{\eta}_n$, which is defined in (23), then

$$\tilde{\phi}_n(\hat{x}^k) \leq \tilde{f}_{\hat{x}^k}. \quad (35)$$

(c) If x^{n+1} is a null step, $\eta_{n+1} = \eta_n$, and either $I_{n+1} \supseteq I_n^{\text{act}}$ or $I_{n+1} \supseteq \{-n\}$, then for all $w \in R^m$

$$\tilde{\phi}_{n+1}(w) \geq \tilde{\phi}_n(x^{n+1}) + \mu_n \langle \hat{x}^k - x^{n+1}, w - x^{n+1} \rangle. \quad (36)$$

(d) If $I_n \supseteq \{n\}$, then, for some $g_\varepsilon^n \in \partial_\varepsilon f(x^n)$,

$$\tilde{\phi}_{n+1}(w) \geq \tilde{f}_{x^n} + \eta_n d_n^k + \langle g_\varepsilon^n + \eta_n \Delta_n^k, w - x^n \rangle - 2\varepsilon. \quad (37)$$

Proof. (a) is clear since $\tilde{\phi}_n$ is the maximum of a collection of affine functions.

(b) follows from $\tilde{\phi}_n(\hat{x}^k) = \tilde{f}_{\hat{x}^k} + \max\{-\langle \tilde{e}_i^k + \eta_n d_i^k, \hat{x}^k - \hat{x}^k \rangle + 2\varepsilon\}$ and (23) ensures $\tilde{e}_i^k + \eta_n d_i^k + 2\varepsilon \geq 0$ for all $i \in I_n$.

For (c), suppose that x^{n+1} is a null step and $\eta_{n+1} = \eta_n$; we know

$$\begin{aligned} \tilde{\phi}_{n+1}(w) &\geq \tilde{f}_{\hat{x}^k} - \tilde{e}_l^k - \eta_n d_l^k - 2\varepsilon \\ &\quad + \langle g_\varepsilon^l + \eta_n \Delta_l^k, w - \hat{x}^k \rangle, \end{aligned} \quad (38)$$

for all $l \in I_{n+1}$ and $w \in R^m$. Equation (17) implies that

$$\begin{aligned} \tilde{f}_{\hat{x}^k} - (\tilde{e}_l^k + \eta_n d_l^k + 2\varepsilon) \\ = \tilde{\phi}_n(x^{n+1}) - \langle g_\varepsilon^l + \eta_n \Delta_l^k, x^{n+1} - \hat{x}^k \rangle. \end{aligned} \quad (39)$$

Then we obtain in (38) for all $l \in J_n^{\text{act}}$ and $l = -n$

$$\begin{aligned}\tilde{\phi}_{n+1}(w) &\geq \tilde{\phi}_n(x^{n+1}) \\ &\quad + \langle g_\varepsilon^l + \eta_n \Delta_l^k, w - \hat{x}^k + \hat{x}^k - x^{n+1} \rangle \\ &= \tilde{\phi}_n(x^{n+1}) + \langle g_\varepsilon^l + \eta_n \Delta_l^k, w - x^{n+1} \rangle.\end{aligned}\quad (40)$$

If $I_{n+1} \supseteq \{-n\}$, since $g_\varepsilon^{-n} + \eta_n \Delta_{-n}^k = \mu_n(\hat{x}^k - x^{n+1})$, the relation above for $l = -n$ is just item (c). For the case $I_{n+1} \supseteq J_n^{\text{act}}$, we sum the above inequality by using the convex multipliers α_l^n and obtain the desired result.

For the last conclusion (d), since $I_n \supseteq \{n\}$, from the definition of $\tilde{\phi}_n$ and $g_\varepsilon^n \in \partial_\varepsilon f(x^n)$, for all $w \in R^m$,

$$\begin{aligned}\tilde{\phi}_{n+1}(w) &\geq \tilde{f}_{\hat{x}^k} - (e_n^k + \eta_n d_n^k + 2\varepsilon) \\ &\quad + \langle g_\varepsilon^n + \eta_n \Delta_n^k, w - \hat{x}^k \rangle \\ &= \tilde{f}_{\hat{x}^k} - \tilde{f}_{\hat{x}^k} + \tilde{f}_{x^n} + \langle g_\varepsilon^n, \hat{x}^k - x^n \rangle - \eta_n d_n^k \\ &\quad - 2\varepsilon + \langle g_\varepsilon^n + \eta_n \Delta_n^k, w - \hat{x}^k \rangle \\ &= \tilde{f}_{x^n} + \langle g_\varepsilon^n, w - x^n \rangle + \eta_n \langle \Delta_n^k, w - x^n \rangle \\ &\quad + \eta_n \langle x^n - \hat{x}^k, x^n - \hat{x}^k \rangle - \eta_n d_n^k - 2\varepsilon \\ &= \tilde{f}_{x^n} + \langle g_\varepsilon^n + \eta_n \Delta_n^k, w - x^n \rangle + \eta_n d_n^k - 2\varepsilon.\end{aligned}\quad (41)$$

□

Lemma 9. Consider the functions $\tilde{\phi}_n(y)$ given by (14). If f satisfies Assumptions 1 and 2, there exists an iteration $\bar{n} > 0$ such that all the parameters sequences stabilize; that is,

$$\begin{aligned}\eta_n &= \bar{\eta}, \\ \mu_n &= \bar{\mu}, \\ R_n &= \bar{\eta} + \bar{\mu}, \\ \forall n &\geq \bar{n}.\end{aligned}\quad (42)$$

Therefore, condition (3) in [19] is eventually satisfied. If, in addition, $\bar{\eta} \geq \rho^{\text{id}}$, then

$$\tilde{\phi}_n(w) \leq \tilde{f}_w + \frac{\bar{\eta}}{2} \|w - \hat{x}^{k(n)}\|^2, \quad \forall w \in L_0, \quad n \geq \bar{n}; \quad (43)$$

that is, condition (6) in [19] holds.

Proof. By Lemma 7 there are a finite number of restarts in Step 6. Once there are no more restarts, $\mu_n = \bar{\mu}$ and the update of η in Step 5 is nondecreasing. If the sequence $\{\eta_n\}$ is not to be stabilized at some value $\bar{\eta}$, there must be an infinite subsequence of iterations at which η is increased by a factor of at least τ . In this case for some iteration η_{n_c} , the function $\tilde{f} + (\eta_{n_c}/2) \|\cdot - \hat{x}^{k(n_c)}\|^2$ is convex on L_0 . And one will have $\tilde{e}_i^k + \eta_{n_c} d_i^k + 2\varepsilon \geq 0$ for all $i \in I_{n_c}$. Hence,

$$\eta_{n_c} \geq \max_{i \in I_{n_c}} \left\{ -\frac{\tilde{e}_i^k + 2\varepsilon}{d_i^k} \right\} = \bar{\eta}_{n_c+1}, \quad (44)$$

and therefore, from that iteration on, the convexification parameter remains unchanged; that is, $\eta_{n_c} + j = \eta_{n_c}$ for all $j \geq 0$, and this leads to a contradiction.

When $\bar{\eta} \geq \rho^{\text{id}}$, the augmented function $\tilde{f}_{\bar{\eta}}^{\hat{x}^{k(n)}}$ is convex on L_0 , and hence the model function $\tilde{\phi}_n$ remains below the augmented function. □

If Algorithm 6 stops at some iteration n with $\delta_{n+1} = 2\varepsilon$ and $\text{TOI}_{\text{stop}} = 0$, by the definition of δ_{n+1} , $\tilde{\phi}_n(x^{n+1}) = \tilde{f}_{\hat{x}^k} + (\eta_n/2) \|x^{n+1} - \hat{x}^k\|^2$. Note that $\tilde{f}_{\hat{x}^k} \geq \tilde{\phi}_n(\hat{x}^k)$ and $x^{n+1} = \tilde{p}_{\mu_n} \tilde{\phi}_n(\hat{x}^k)$, and we see

$$\begin{aligned}\tilde{\phi}_n(x^{n+1}) + \frac{\mu_n}{2} \|x^{n+1} - \hat{x}^k\|^2 &\leq \tilde{\phi}_n(\hat{x}^k), \\ \tilde{f}_{\hat{x}^k} + \frac{\eta_n + \mu_n}{2} \|x^{n+1} - \hat{x}^k\|^2 &\leq \tilde{\phi}_n(\hat{x}^k), \\ \tilde{f}_{\hat{x}^k} + \frac{R_n}{2} \|x^{n+1} - \hat{x}^k\|^2 &\leq \tilde{f}_{\hat{x}^k},\end{aligned}\quad (45)$$

which shows $x^{n+1} = \hat{x}^k$; this implies that $\hat{x}^k = \tilde{p}_{\mu_n} \tilde{\phi}_n(\hat{x}^k)$. Suppose that η_n is sufficiently large for $\tilde{f} + (\eta_n/2) \|\cdot - \hat{x}^k\|^2$ to be convex on L_0 , and this would imply that

$$\begin{aligned}\tilde{f}_{\hat{x}^k} &\leq \tilde{\phi}_n(\hat{x}^k) \leq \tilde{\phi}_n(w) + \frac{\mu_n}{2} \|w - \hat{x}^k\|^2, \quad \forall w \in R^m \\ &\leq \tilde{f}_w + \frac{\eta_n}{2} \|w - \hat{x}^k\|^2 + \frac{\mu_n}{2} \|w - \hat{x}^k\|^2, \quad \forall w \in L_0 \\ &= \tilde{f}_w + \frac{R_n}{2} \|w - \hat{x}^k\|^2, \quad \forall w \in R^m.\end{aligned}\quad (46)$$

In other words, $\hat{x}^k = \tilde{p}_{R_n} \tilde{f}(\hat{x}^k)$; \hat{x}^k is an approximate stationary point of f .

Definition 10. x^* is called ε -double approximate stationary point of f if

$$\tilde{f}_{x^*} \leq \tilde{f}_w + \frac{\bar{R}}{2} \|w - x^*\|^2 + \varepsilon, \quad \forall w \in R^m, \quad (47)$$

where $\bar{R} = \bar{\mu} + \bar{\eta}$, $\bar{\eta}$, and $\bar{\mu}$ are the stabilized convexification parameter and model prox-parameter.

Theorem 11. Suppose that $\text{TOL}_{\text{stop}} = 0$ and there is no termination. Let $\bar{\eta}$ be the stabilized value for the local convexification parameter sequence, as in Lemma 9, and $\bar{\eta} > \rho^{\text{id}}$. The following two mutually exclusive conclusions hold:

- There is a last serious step \hat{x} , followed by infinitely many null steps. Then $x^{n+1} \rightarrow \hat{x}$ and \hat{x} is an approximate stationary point of f .
- There is an infinite number of serious steps. Then any accumulation point of the sequence $\{\hat{x}^k\}$ is ε -double approximate stationary point of f .

Proof. For item (a), consider that iteration n after the last serious step \hat{x} was generated, so there are only null steps. First, we prove $x^{n+1} \rightarrow p := \tilde{p}_{\bar{R}} \tilde{f}(\hat{x})$. Consider the function:

$L_n(w) := \tilde{\phi}_n(x^{n+1}) + (\bar{\mu}/2)\|x^{n+1} - \hat{x}\|^2 + (\bar{\mu}/2)\|w - x^{n+1}\|^2$. And let $S^{n+1} := \bar{\mu}(\hat{x} - x^{n+1}) \in \partial\tilde{\phi}_n(x^{n+1})$. Since $x^{n+1} = \tilde{p}_n\tilde{\phi}_n(\hat{x})$, $L_n(x^{n+1}) \leq \tilde{\phi}_n(\hat{x})$, and (35) implies that

$$L_n(x^{n+1}) \leq \tilde{f}_{\hat{x}}, \quad \forall n \geq \bar{n}. \quad (48)$$

Condition (36) gives the inequality

$$\tilde{\phi}_{n+1}(x^{n+2}) \geq \tilde{\phi}_n(x^{n+1}) + \langle S^{n+1}, x^{n+2} - x^{n+1} \rangle. \quad (49)$$

By the definition of L_n , $L_{n+1}(x^{n+2}) = \tilde{\phi}_{n+1}(x^{n+2}) + (\bar{\mu}/2)\|x^{n+2} - \hat{x}\|^2$, and hence

$$\begin{aligned} L_{n+1}(x^{n+2}) &\geq \tilde{\phi}_n(x^{n+1}) + \langle S^{n+1}, x^{n+2} - x^{n+1} \rangle \\ &\quad + \frac{\bar{\mu}}{2}\|x^{n+1} - \hat{x}\|^2. \end{aligned} \quad (50)$$

But $\tilde{\phi}_n(x^{n+1}) = L_n(x^{n+1}) - (\bar{\mu}/2)\|x^{n+1} - \hat{x}\|^2$, so

$$\begin{aligned} L_{n+1}(x^{n+2}) &\geq L_n(x^{n+1}) + \langle S^{n+1}, x^{n+2} - x^{n+1} \rangle \\ &\quad + \frac{\bar{\mu}}{2}\|x^{n+2} - \hat{x}\|^2 - \frac{\bar{\mu}}{2}\|x^{n+1} - \hat{x}\|^2. \end{aligned} \quad (51)$$

After expanding squares, the two rightmost terms above satisfy the relation

$$\begin{aligned} &\frac{\bar{\mu}}{2}\|x^{n+2} - \hat{x}\|^2 - \frac{\bar{\mu}}{2}\|x^{n+1} - \hat{x}\|^2 \\ &= -\bar{\mu}\langle \hat{x} - x^{n+1}, x^{n+2} - x^{n+1} \rangle + \frac{\bar{\mu}}{2}\|x^{n+2} - x^{n+1}\|^2 \\ &= -\langle S^{n+1}, x^{n+2} - x^{n+1} \rangle + \frac{\bar{\mu}}{2}\|x^{n+2} - x^{n+1}\|^2. \end{aligned} \quad (52)$$

So we obtain $L_{n+1}(x^{n+2}) \geq L_n(x^{n+1}) + (\bar{\mu}/2)\|x^{n+2} - x^{n+1}\|^2$ and the sequence $\{L_n(x^{n+1})\}$ is eventually increasing and convergent since it is bounded above by $\tilde{f}_{\hat{x}}$. By conditions (35) and (36)

$$L_n(\hat{x}) = \tilde{\phi}_n(x^{n+1}) + \langle S^{n+1}, \hat{x} - x^{n+1} \rangle \leq \tilde{\phi}_n(\hat{x}) \leq \tilde{f}_{\hat{x}}. \quad (53)$$

Since $L_n(x^{n+1}) + (\bar{\mu}/2)\|\hat{x} - x^{n+1}\|^2 = L_n(\hat{x})$, we obtain that $L_n(x^{n+1}) + (\bar{\mu}/2)\|\hat{x} - x^{n+1}\|^2 \leq \tilde{f}_{\hat{x}}$. By the convergence of $\{L_n(x^{n+1})\}$, the sequence $\{x^n\}$ is bounded. Note that

$$0 \leq \frac{\bar{\mu}}{2}\|x^{n+1} - x^n\|^2 \leq L_n(x^{n+1}) - L_{n-1}(x^n) \rightarrow 0, \quad (54)$$

so $\|x^{n+1} - x^n\| \rightarrow 0$ as $n \rightarrow \infty$. Since $\{x^n\} \subseteq L_0$ (when n is sufficiently large), $\{x^n\}$ is bounded. Therefore, the sequence of approximate proximal point $\{x^n\}$ converges. By using the similar technique of Theorem 3(ii) in [19], we can prove that $x^{n+1} \rightarrow \tilde{p}_R\tilde{f}(\hat{x})$. For the sequence $\{\tilde{\phi}_n(x^{n+1})\}$, we show that there exists $N_1 \subseteq N$ such that

$$\left| \lim_{n \in N_1, n \rightarrow \infty} \tilde{\phi}_n(x^{n+1}) - \left(\tilde{f}_p + \frac{\bar{\eta}}{2}\|\hat{x} - p\|^2 \right) \right| \leq 2\varepsilon. \quad (55)$$

By Lemma 8(d), when n is sufficiently large,

$$\tilde{\phi}_n(x^{n+1}) \geq \tilde{f}_{x^n} + \bar{\eta}d_n^k + \langle g_\varepsilon^n + \bar{\eta}\Delta_n^k, x^{n+1} - x^n \rangle - 2\varepsilon. \quad (56)$$

And at the same time, according to Lemma 9 $\tilde{\phi}_n(x^{n+1}) \leq \tilde{f}_{x^{n+1}} + (\bar{\eta}/2)\|x^{n+1} - \hat{x}\|^2$. As $x^{n+1} \rightarrow p$, we have $\tilde{\phi}_n(x^{n+1})$ is bounded, so it has an accumulation point; say for some iterate set $N_1 \subseteq N$, $\lim_{n \in N_1, n \rightarrow \infty} \tilde{\phi}_n(x^{n+1}) = \tilde{\phi}^*$. Obviously,

$$\left| \tilde{\phi}^* - \left(\tilde{f}_p + \frac{\bar{\eta}\|\hat{x} - p\|^2}{2} \right) \right| \leq 2\varepsilon. \quad (57)$$

Since $\tilde{\phi}_n(x^{n+1}) = \tilde{f}_{\hat{x}} + (\bar{\eta}/2)\|\hat{x} - p\|^2 + 2\varepsilon - \delta_{n+1}$, so $\delta^* = \lim_{n \in N_1, n \rightarrow \infty} \delta_{n+1} = \tilde{f}_{\hat{x}} + (\bar{\eta}/2)\|p - \hat{x}\|^2 + 2\varepsilon - \tilde{\phi}^*$, and

$$\begin{aligned} &\left| \tilde{\phi}^* - \left(\tilde{f}_p + \frac{\bar{\eta}\|\hat{x} - p\|^2}{2} \right) \right| \\ &= \left| \tilde{f}_{\hat{x}} + \frac{\bar{\eta}}{2}\|p - \hat{x}\|^2 - \left(\tilde{f}_p + \frac{\bar{\eta}\|\hat{x} - p\|^2}{2} \right) + 2\varepsilon - \delta^* \right| \\ &= |\delta^* - (\tilde{f}_{\hat{x}} - \tilde{f}_p) - 2\varepsilon| \leq 2\varepsilon. \end{aligned} \quad (58)$$

Because the descent test in Step 4 is not satisfied, taking the limit as $n \rightarrow \infty$ we have

$$\begin{aligned} \tilde{f}_p &\geq \tilde{f}_{\hat{x}} - \bar{m}(\delta^* - 2\varepsilon) \geq \tilde{f}_{\hat{x}} + \bar{m}(-\tilde{f}_{\hat{x}} + \tilde{f}_p - 2\varepsilon) \\ &= \tilde{f}_{\hat{x}} - \bar{m}\tilde{f}_{\hat{x}} + \bar{m}\tilde{f}_p - 2\bar{m}\varepsilon. \end{aligned} \quad (59)$$

Therefore, $\tilde{f}_p - \tilde{f}_{\hat{x}} \geq \bar{m}(\tilde{f}_p - \tilde{f}_{\hat{x}}) - 2\bar{m}\varepsilon$, and from $\bar{m} \in (0, 1)$ we obtain $\tilde{f}_p - \tilde{f}_{\hat{x}} \geq -2\bar{m}\varepsilon/(1 - \bar{m})$. But $p = \tilde{p}_R\tilde{f}(\hat{x})$ implies that $\tilde{f}_p + (\bar{R}/2)\|p - \hat{x}\|^2 \leq \tilde{f}_{\hat{x}}$, which shows that $\hat{x} \in B_{2\bar{m}\varepsilon/(1 - \bar{m})\bar{R}}(p)$. That is, \hat{x} equals approximately $\tilde{p}_R\tilde{f}(\hat{x})$; \hat{x} is an approximate stationary point of f .

To see item (b), $\{\hat{x}^k\}$ has an accumulation point since $\{\hat{x}^k\} \subseteq L_0$; say for some iterate set K , $\hat{x}^k \rightarrow x^{\text{inf}} \in L_0$ as $k \in K$ and $k \rightarrow \infty$. Since $\hat{x}^{k+1} = x^{i_{k+1}}$, we set $j_k = i_{k+1} - 1$, so that $\hat{x}^{k+1} = \tilde{p}_R\tilde{\phi}_{j_k}(\hat{x}^k)$. The descent test

$$\tilde{f}_{\hat{x}^{k+1}} \leq \tilde{f}_{\hat{x}^k} - \bar{m}(\delta_{i_{k+1}} - 2\varepsilon) \quad (60)$$

implies that as $k \rightarrow \infty$, either $\tilde{f}_{\hat{x}^k} \searrow \infty$ or $\delta_{i_{k+1}} \rightarrow 2\varepsilon$. By Lemma 3(b), \tilde{f} is bounded below; therefore $\delta_{i_{k+1}} \rightarrow 2\varepsilon$. From (31) both $\|\hat{x}^{k+1} - \hat{x}^k\|^2$ and $\bar{e}_{-j_k}^k + \bar{\eta}d_{-j_k}^k$ must converge to zero. Therefore by (14), $\tilde{\phi}_{j_k}(\hat{x}^{k+1}) - \tilde{f}_{\hat{x}^k} \rightarrow -2\varepsilon$ as $k \rightarrow \infty$. Consider now $k \in K$; since $\|\hat{x}^{k+1} - \hat{x}^k\|^2 \rightarrow 0$, both \hat{x}^{k+1}

and \hat{x}^k converge to x^{inf} , with $\tilde{\phi}_{j_k}(\hat{x}^{k+1}) \rightarrow \tilde{f}_{x^{\text{inf}}} - 2\varepsilon$. But $\hat{x}^{k+1} = \tilde{p}_{\bar{\mu}}\tilde{\phi}_{j_k}(\hat{x}^k)$ and $\bar{\eta} > \rho^{\text{id}}$ implies that, for all $w \in L_0$,

$$\begin{aligned} & \tilde{\phi}_{j_k}(\hat{x}^{k+1}) + \frac{\bar{\mu}}{2} \|\hat{x}^{k+1} - \hat{x}^k\|^2 \\ & \leq \tilde{\phi}_{j_k}(w) + \frac{\bar{\mu}}{2} \|w - \hat{x}^k\|^2 \\ & \leq \tilde{f}_w + \frac{\bar{\eta}}{2} \|w - \hat{x}^k\|^2 + \frac{\bar{\mu}}{2} \|w - \hat{x}^k\|^2 \\ & = \tilde{f}_w + \frac{\bar{R}}{2} \|w - \hat{x}^k\|^2. \end{aligned} \quad (61)$$

Therefore, we have

$$\tilde{f}_{x^{\text{inf}}} - \varepsilon \leq \tilde{f}_w + \frac{\bar{R}}{2} \|w - x^{\text{inf}}\|^2, \quad \forall w \in L_0. \quad (62)$$

As $x^{\text{inf}} \in L_0$, we also have that, for any $w \notin L_0$,

$$\begin{aligned} \tilde{f}_{x^{\text{inf}}} - \varepsilon & \leq \tilde{f}_{x^{\text{inf}}} \leq \tilde{f}_{x^0} + M_0 < \tilde{f}_w \\ & \leq \tilde{f}_w + \frac{\bar{R}}{2} \|w - x^{\text{inf}}\|^2. \end{aligned} \quad (63)$$

Hence, $\tilde{f}_{x^{\text{inf}}} \leq \tilde{f}_w + (\bar{R}/2)\|w - x^{\text{inf}}\|^2 + \varepsilon$ for all $w \in R^m$. That is, x^{inf} is ε -double approximate stationary point of f . \square

Remark 12. Note that if $\varepsilon = 0$, the results obtained in Theorem 11 are exactly the ones in [18], which means that our work is really the generalization of previous work.

6. Conclusion

We propose an approximate redistributed proximal bundle method for nonsmooth nonconvex optimization by employing the inexact information from the objective function. With the inexact data we prove that the cutting-planes model constructed in this paper is eventually the local lower approximation to the approximate objective function. Analysis of the convergence proceeds by first showing that the convexification parameter eventually stabilized and once stabilized, the convergence to the approximate or ε -double approximate stationary point of the objective function is obtained under the condition that the stabilized convexification parameter is greater than the ideal proximal threshold ρ^{id} . The local convexification approach opens a new way for future study on nonsmooth nonconvex optimization and can shed a new insight on the first order models from [20].

In [21], the authors present a framework of general bundle methods which are capable of handling inexact oracles, and the framework generalizes in various ways a number of algorithms proposed in the literatures. Next we discuss the relationship between our algorithm and [21]. In [21], the objective function f is a finite-valued convex function, and

the authors make the following assumptions. For each given $u \in R^n$, the oracle delivers the inexact information:

$$\begin{aligned} f_u &= f(u) - \eta_u, \\ g_u &\in R^n \text{ such that } f(\cdot) \geq f_u + \langle g_u, \cdot - u \rangle - \eta_u^g \end{aligned} \quad (64)$$

$$\text{with } \eta_u \leq \eta, \quad \eta_u^g \leq \eta,$$

where the error bound η is possibly unknown. By (64) we have $f(\cdot) \geq f(u) + \langle g_u, \cdot - u \rangle - (\eta_u^g + \eta_u)$. As a result,

$$g_u \in \partial_{\eta_u^g + \eta_u} f(u) \text{ with } \eta_u^g + \eta_u \geq 0, \quad \forall u. \quad (65)$$

Even when the value of the upper error bound η is unknown, inequalities (64) and (65) imply that $\eta \geq \eta_u \geq -\eta_u^g \geq -\eta$, which means that

$$\begin{aligned} f_u &\in [f(u) - \eta, f(u) + \eta], \\ g_u &\in \partial_{2\eta} f(u). \end{aligned} \quad (66)$$

In our paper, if \tilde{f} is the function which is locally convex by adding a quadratic term with convexification parameter, we suppose that, for each given $x \in R^m$, an oracle delivers the inexact information:

$$\begin{aligned} \tilde{f}_x &\in [f(x) - \varepsilon_f, f(x) + \varepsilon_g], \\ g_\varepsilon^x &\in R^m \text{ such that } f(\cdot) \geq \tilde{f}_x + \langle g_\varepsilon^x, \cdot - x \rangle - \varepsilon \end{aligned} \quad (67)$$

with $\varepsilon = \varepsilon_f + \varepsilon_g$.

Assumption (67) is more general than (64) since if we choose $\varepsilon_f = \varepsilon_g = \eta$, (67) will become $\tilde{f}_x \in [f(x) - \eta, f(x) + \eta]$, $g_\varepsilon^x \in \partial_{2\eta} f(x)$, which is exactly (66). Therefore, our assumption is, in a sense, some kind of generalization of assumptions in [21].

In [21], the authors use the linearizations and the cutting-plane model

$$\begin{aligned} f_j^L(u) &= f_{u_j} + \langle g_{u_j}, u - u_j \rangle, \quad j = 1, 2, \dots, k, \\ \tilde{f}_k(\cdot) &= \max \{f_j^L(\cdot) \mid j \in J_k = \{1, 2, \dots, k\}\}, \end{aligned} \quad (68)$$

and the next trial point is obtained by minimizing the stabilized model function

$$\min_{u \in R^n} \tilde{f}_k(\cdot) + \frac{1}{2t_k} \|u - \hat{u}_k\|^2. \quad (69)$$

In our paper, we choose the cutting-plane model function to be

$$\begin{aligned} \tilde{\phi}_n(y) &= \max_{i \in I_n} \left\{ \left(\tilde{f}_i + \frac{\eta_n}{2} \|x^i - \hat{x}^k\|^2 \right) \right. \\ &\quad \left. + \langle g_\varepsilon^i + \eta_n(x^i - \hat{x}^k), y - x^i \rangle - 2\varepsilon \right\}, \end{aligned} \quad (70)$$

which becomes a lower approximation to the locally convex objective function. It is similar to (68) except for the appearance of 2ε . At the same time, the next trail point is chosen

as $x^{n+1} = \tilde{p}_{\mu_n} \tilde{\phi}_n(\hat{x}^k)$, the approximate proximal point of \tilde{f} defined by (6), which is quite different from the general traditional techniques introduced in [21].

In [21], the authors mention that, for inexact oracles, the progress made by the algorithm can be measured relative to the model, some nominal reference value, or the approximate objective function, and they are called and denoted by

model decrease:

$$\delta_k^M = l_k - \tilde{f}_k(u_{k+1}), \text{ where } l_k \text{ is chosen so that } l_k \geq f_{\hat{u}_k},$$

nominal decrease:

$$\delta_k^N = \delta_k^M - \frac{\alpha_k}{t_k} \|u_{k+1} - \hat{u}_k\|^2, \quad \alpha \in [0, 1],$$

effective decrease:

$$\delta_k^E = f_{\hat{u}_k} - f_{u_{k+1}}.$$

In our paper, we define a predicted decrease $\delta_{n+1} = (\tilde{f}_{\hat{x}^k} + (\eta_n/2)\|x^{n+1} - \hat{x}^k\|^2 + 2\varepsilon) - \tilde{\phi}_n(x^{n+1})$, which is similar to δ_k^M , but a bit different from δ_k^M since l_k is only associated with current stability center \hat{u}^k , while the first term in δ_{n+1} is not only related to current stability center \hat{x}^k but also involved with the new trial point x_{n+1} . And this predicted decrease is nonnegative as long as η_n is sufficiently large, and this requirement is exactly coincident with the condition which guarantees that $\tilde{\phi}_n(\cdot)$ is a lower approximation to $\tilde{f}(\cdot)$. Therefore, we can further employ it in the descent test to decide between making a descent step or a null step.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to acknowledge the valuable suggestions and helpful comments from referees. This project is supported by the National Natural Science Foundation of China (nos. 11301246, 11171049, and 11471151).

References

- [1] A. Fuduli, M. Gaudioso, and G. Giallombardo, "Minimizing nonconvex nonsmooth functions via cutting planes and proximity control," *SIAM Journal on Optimization*, vol. 14, no. 3, pp. 743–756, 2003.
- [2] K. C. Kiwiel, "Exact penalty functions in proximal bundle methods for constrained convex nondifferentiable minimization," *Mathematical Programming*, vol. 52, no. 2, pp. 285–302, 1991.
- [3] L.-P. Pang, Z.-Q. Xia, and L.-W. Zhang, "On a second order parallel variable transformation approach," *Journal of Applied Mathematics & Computing*, vol. 11, no. 1-2, pp. 201–213, 2003.

- [4] L.-P. Pang and Z.-Q. Xia, "A PVT-type algorithm for minimizing a nonsmooth convex function," *Serdica. Mathematical Journal*, vol. 29, no. 1, pp. 11–32, 2003.
- [5] C. Lemaréchal, "Lagrangian relaxation," in *Computational Combinatorial Optimization*, vol. 2241 of *Lecture Notes in Computer Science*, pp. 112–156, Springer, Berlin, Germany, 2001.
- [6] E. W. Cheney and A. A. Goldstein, "Newton's method for convex programming and Tchebycheff approximation," *Numerische Mathematik*, vol. 1, pp. 253–268, 1959.
- [7] J. E. Kelley Jr., "The cutting-plane method for solving convex programs," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 4, pp. 703–712, 1960.
- [8] N. Haara, K. Miettinen, and M. Makela, "Globally convergent limited memory bundle method for large-scale nonsmooth optimization," *Mathematical Programming Series A*, vol. 109, no. 1, pp. 181–205, 2007.
- [9] A. Fuduli, M. Gaudioso, and G. Giallombardo, "A DC piecewise affine model and a bundling technique in nonconvex nonsmooth minimization," *Optimization Methods & Software*, vol. 19, no. 1, pp. 89–102, 2004.
- [10] K. C. Kiwiel, "A linearization algorithm for nonsmooth minimization," *Mathematics of Operations Research*, vol. 10, no. 2, pp. 185–194, 1985.
- [11] K. C. Kiwiel, "Restricted step and Levenberg-Marquardt techniques in proximal bundle methods for nonconvex nondifferentiable optimization," *SIAM Journal on Optimization*, vol. 6, no. 1, pp. 227–249, 1996.
- [12] L. Luksan and J. Vlcek, "A bundle-Newton method for nonsmooth unconstrained minimization," *Mathematical Programming*, vol. 83, no. 1-3, pp. 373–391, 1998.
- [13] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*, vol. 306 of *Grundlehren der Mathematischen Wissenschaften*, Springer, Berlin, Germany, 1993.
- [14] J. Shen, Z.-Q. Xia, and L.-P. Pang, "A proximal bundle method with inexact data for convex nondifferentiable minimization," *Nonlinear Analysis. Theory, Methods & Applications*, vol. 66, no. 9, pp. 2016–2027, 2007.
- [15] J. Shen and L.-P. Pang, "A bundle-type auxiliary problem method for solving generalized variational-like inequalities," *Computers and Mathematics with Applications*, vol. 55, no. 12, pp. 2993–2998, 2008.
- [16] J. Shen and L.-P. Pang, "An approximate bundle-type auxiliary problem method for solving generalized variational inequalities," *Mathematical and Computer Modelling*, vol. 48, no. 5-6, pp. 769–775, 2008.
- [17] R. T. Rockafellar and J. J.-B. Wets, *Variational Analysis*, vol. 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, Springer, Berlin, Germany, 1998.
- [18] W. Hare and C. Sagastizabal, "A redistributed proximal bundle method for nonconvex optimization," *SIAM Journal on Optimization*, vol. 20, no. 5, pp. 2442–2473, 2010.
- [19] W. Hare and C. Sagastizabal, "Computing proximal points of nonconvex functions," *Mathematical Programming*, vol. 116, no. 1-2, pp. 221–258, 2009.
- [20] D. Noll, O. Prot, and A. Rondepierre, "A proximity control algorithm to minimize nonsmooth and nonconvex functions," *Pacific Journal of Optimization*, vol. 4, no. 3, pp. 569–602, 2008.
- [21] W. de Oliveira, C. Sagastizabal, and C. Lemaréchal, "Convex proximal bundle methods in depth: a unified analysis for inexact oracles," *Mathematical Programming Series B*, vol. 148, no. 1-2, pp. 241–277, 2014.