Journal of Global Optimization (2023) 86:589–620
https://doi.org/10.1007/s10898-023-01279-8

A proximal bundle method for a class of nonconvex nonsmooth composite optimization problems

Liping Pang[1] · Xiaoliang Wang[2] · Fanyun Meng[3]

Received: 2 May 2020 / Accepted: 27 February 2023 / Published online: 10 April 2023

Abstract
In this paper, a proximal bundle method is proposed for a class of nonconvex nonsmooth composite optimization problems. The composite problem considered here is the sum of two functions: one is convex and the other is nonconvex. Local convexification strategy is adopted for the nonconvex function and the corresponding convexification parameter varies along iterations. Then the sum of the convex function and the extended function is dynamically constructed to approximate the primal problem. To choose a suitable cutting plane model for the approximation function, here we consider the sum of two cutting planes, which are designed respectively for the convex function and the extended function. By choosing appropriate descent condition, our method can keep track of the relationship between primal problem and approximate models. Under mild conditions, the convergence is proved and the accumulation point of iterations is a stationary point of the primal problem. Two polynomial problems and twelve DC (difference of convex) problems are referred in numerical experiments. The preliminary numerical results show that the proposed method is effective for solving these testing problems.

Keywords Proximal bundle method · Nonconvex and nonsmooth · Redistributed strategy · DC problems

Mathematics Subject Classification 65K05 · 90C26

This work is supported by Science Foundation of Zhejiang Sci-Tech Uni-versity (ZSTU) under Grant No. 21062347-Y and also supported by Zhejiang Provincial Natural Science Foundation of China under Grant No. LQ23A010020.

✉ Xiaoliang Wang
xliangwang@126.com

1 School of Mathematical Sciences, Dalian University of Technology, Key Laboratory for Computational Mathematics and Data Intelligence of Liaoning Province, Dalian 116024, China

2 School of Science, Department of Mathematics Science, Zhejiang Sci-Tech University, Hangzhou 310018, China

3 School of Information and Control Engineering, Qingdao University of Technology, Qingdao 266520, China

1 Introduction

In this article, we focus on a class of composite problem, which has the following form:

$$\min_{x \in \mathbf{R}^N} \psi(x) := f(x) + h(x), \tag{1.1}$$

where $f : \mathbf{R}^N \to \mathbf{R}$ is a convex function and $h : \mathbf{R}^N \to \mathbf{R}$ is a nonconvex function. Note that the functions f and h are all not necessarily differentiable. In fact, the problem (1.1) may be seen in many practical applications: image reconstruction, engineering, finance, optimal control and statistics.

Example 1 (Regularized minimization mapping) Many signal and image reconstruction problems minimize a function having the following form ($0 < q \leq 1$):

$$\frac{1}{2}\|Ax - b\|^2 + \tau\|x\|_q,$$

for $x \in \mathbf{R}^N$, $b \in \mathbf{R}^M$, A is a matrix with $M \times N$ dimensions, τ is a positive parameter and $\|x\|_q = (\sum_{i=1}^N |x_i|^q)^{\frac{1}{q}}$. The above function can be seen as the sum of a convex function and a lower-c^2 function.

Example 2 (Transformation of constraint problems) Consider the following constraint problem:

$$\min f(x), \quad \text{s.t.} \quad c_i(x) \leq 0, \quad i = 1, \cdots, m,$$

where f is finite convex and c_i, $i = 1, \cdots, m$ is nonconvex function. It is clear that the above constraint problem can be converted into a problem which has only one constraint $c(x)$, where $c(x) := \max\{c_i(x), i = 1, \cdots, m\}$. By the definition of $c(x)$, the function c is not necessary convex and smooth. By the penalty strategy, the new generated problem can be converted into the following unconstrained problem:

$$\min_{x \in \mathbf{R}^N} f(x) + \lambda c_+(x),$$

where $c_+(x) = \max\{c(x), 0\}$ and $\lambda \geq 0$. Setting $h(x) = \lambda c_+(x)$, the above problem is in the form of (1.1).

Example 3 (DC problem) The minimization of difference of convex functions (DC) is defined as

$$\min_{x \in \mathbf{R}^N} f_1(x) - f_2(x),$$

where f_1, f_2 are finite convex and all not necessarily differentiable. If we set $h(x) = -f_2(x)$, then the above problem has the form of problem (1.1).

Many researchers [4–18] are devoted to developing the implementable algorithms for solving the problems with the form of sum of two functions. These algorithms include the well known operator splitting methods [5, 6, 13], alternating direction methods [7–10, 14, 15], primal-dual extragradient methods [11] and bundle methods [12, 24] and so on. Concretely, the sum of two smooth and convex functions can be seen in [4, 15]; the sum of a nonsmooth convex function and a smooth convex function can be found in [4, 15]; the sum of two separable convex functions with diffrent variables can be seen in [12, 16, 17]; the sum of a nonsmooth convex function and a nonconvex smooth function can be found in [18]; the sum of two nonconvex functions can be seen in [8] and so on. By viewing these papers, their central ideas are exploiting the separable structure of the primal objective functions. As for nonconvex cases, the prox-regularity and prox-boundedness techniques are considered as the essential theoretical basis to design practical implementable algorithms. In this paper, we consider the convexification strategy and utilize the above theoretical basis to design a class of implementable proximal bundle algorithms. Bundle methods [12, 19, 22, 23] are recognized as highly effective methods for solving nonsmooth optimization problems. Proximal bundle method is one class of bundle methods, which has been successfully used to solve unconstrained convex optimization problems with discontinuous first derivatives [19–23]. In bundle methods, the nonnegative linearization error plays an important role in ensuring the convergence of the algorithms. In convex cases, the linearization error is always nonnegative. However, in nonconvex cases, the linearization error may be negative. In order to circumvent this drawback, some methods [24–30] about refining the subgradient locality measure are proposed. In this paper, we adopt the convexification technique for handling the nonconvex function h to ensure the corresponding linearization error nonnegative.

Motivated by the redistributed strategy in [24] and the idea of building the cutting plane models in [12], we propose a nonconvex proximal bundle method for solving the composite problem (1.1) and this new method also works well for some DC problems. Although our algorithm is similar to that in [12, 24], they are much different. In [24], the authors focused on a single nonconvex function (lower-c^2). The authors in [12] considered the sum of two convex functions. In this article, we present a proximal bundle method for solving the primal optimization problem (1.1). The advantages of the proposed method are as follows: (i) the proximal parameter is dynamically divided into two parameters: one is used to convexificate function $h(x)$ to ensure its linearization errors nonnegative and the other is regarded as the proximal parameter; (ii) The cutting planes for the convex function and convexification function are designed respectively. Since the sum of maxima is bigger than the maximum of the sum, then we sum up the two cutting models in order to obtain a better approximate model for the modified optimization problem; (iii) The proposed method can keep track of the relationship between the original problem (1.1) and the generated piecewise linear model. This maintains a powerful connection between the model functions and the primal function.

The paper is organized as follows. In Sect. 2, we review some definitions in variational analysis and some preliminaries in proximal bundle methods and give some notations. Section 3 introduces our proximal bundle method. In Sect. 4, the global convergence is analyzed under some mild conditions. The numerical results about polynomial functions and some DC problems are presented in Sect. 5. Some conclusions are given in Sect. 6.

2 Preliminaries

In this section, we firstly recall some concepts and results in variational analysis, which will be used in our paper. Then the redistributed strategy is discussed. Meanwhile, some notations for the bundles of functions f and h are considered.

2.1 Concepts and properties

Throughout the paper, we assume the functions f and h are all proper and regular (see [1, Definition 7.25]). Notably, we make use of the limiting subdifferential for the nonconvex function h, denoted by $\partial h(x)$ in [1]. The limiting subdifferential of function h at point $\hat{x}$ is defined by

$$\partial h(\hat{x}) := \limsup_{x \to \hat{x}\ h(x) \to h(\hat{x})} \left\{ v : \liminf_{x \to \hat{x}\ x \neq \hat{x}} \frac{h(x) - h(\hat{x}) - \langle v, x - \hat{x} \rangle}{\|x - \hat{x}\|} \geq 0 \right\}.$$

For the convex function f, the common subdifferential in convex analysis is adopted, which is defined as

$$\partial f(\hat{x}) := \{v \mid f(x) \geq f(\hat{x}) + \langle v, x - \hat{x} \rangle\},$$

for all $x \in dom(f)$. In convex optimization, $\epsilon-$subdifferential is defined by

$$\partial_\epsilon f(x) := \{g \in \mathbf{R}^N : f(y) \geq f(x) + \langle g, y - x \rangle - \epsilon\}, \tag{2.1}$$

where $\epsilon \geq 0, x, y \in dom(f)$. Specially, if $\epsilon = 0$, by the above definition, we have $\partial_0 f(x) = \partial f(x)$.

In this paper, we consider a class of special nonconvex functions, i.e., lower-c^2 function which is presented in the following definition.

Definition 2.1 [1] A function $\psi : \mathcal{O} \to \mathbf{R}$, where $\mathcal{O}$ is an open subset of $\mathbf{R}^N$, is said to be lower-c^2 on $\mathcal{O}$, if on some neighborhood V of each $\hat{x} \in \mathcal{O}$, there is a representation:

$$\psi(x) = \max_{t \in T} \psi_t(x),$$

in which the function ψ_t are of class C^2 on V and the index set T is a compact space such that $\psi_t(x)$ and $\nabla_x \psi_t(x)$ depend continuously not just on x but on $(t, x) \in T \times V$.

In the following, we state a theorem which presents a much important property of lower-c^2 functions which can also be found in [1, Theorem 10.33]. For completeness, we state it but omit its proof.

Theorem 2.1 [1, Theorem 10.33] *Any finite convex function is lower-c^2. In fact, a function F is lower-c^2 on an open set $\mathcal{O} \in \mathbf{R}^N$, if and only if, relative to some neighborhood of each point of $\mathcal{O}$, there is an expression $F(x) = G(x) - H(x)$, in which G is a finite convex function and function H is C^2; indeed, H can be taken to be a convex, quadratic function, even of form $\frac{\rho}{2}\|\cdot\|^2$.*

Another statement of Theorem 2.1 is that the function F is lower-c^2 on an open set $\mathcal{O} \in \mathbf{R}^N$ if F is finite on $\mathcal{O}$ and for any $x \in \mathcal{O}$, there exists a threshold $\bar{\rho} \geq 0$ such that $F + \frac{\rho}{2}\|\cdot\|^2$ is convex on an open neighborhood $\mathcal{O}'$ of x for all $\rho \geq \bar{\rho}$, where $\|\cdot\|$ stands for standard Euclidean norm, unless specified otherwise. Specifically, if F is finite and convex, then F is a lower-c^2 function and the threshold $\bar{\rho} \equiv 0$. In this paper, it is reasonable to deal with the primal problem (1.1) by adding quadric terms to convexificate function h.

2.2 Nonconvex setting and compression technique

In the convex case, at each iteration i, the classical bundle methods keep track of the bundle of information in each iteration, which includes the past function values and subgradient values. For example, for a convex function f, the bundle has the form of $\mathcal{B}_f^n = \{(x^i, f_i = f(x^i), g^i = g(x^i) \in \partial f(x^i)), i \in I_n\}$, where $I_n \subseteq \{0, 1, \cdots, n\}$ is an index set. Along iterations, a sequence of iterative points are generated. These iterative points are divided into two types: one type improves the approximate models' accuracy; the other type not only improves the models' accuracy but also decreases significantly the objective function. The corresponding iterations of the algorithmic scheme are respectively called null steps and serious steps. Usually, the serious steps are referred to as proximal center or stability center in the literature, which are denoted by $\hat{x}^{k(n)}$. When it is clear in the context, we drop the explicit dependence of k on the current iteration index n to alleviate notation. It is obvious that a cutting plane model is a lower approximate to the primal problem in convex cases and $\widetilde{f}(y) \approx f(x^i)$ holds for any y sufficiently approximates to x^i. In convex case, the linearization error $e^k_{f,i} := f(\hat{x}^k) - f(x^i) - \langle g^i_f, \hat{x}^k - x^i\rangle$ is always nonnegative which plays an important role to ensure the convergence of bundle algorithms. However, in nonconvex cases, the linearization error may be negative and it becomes much difficult for the proof of convergence and the algorithms for convex optimization can not be directly applied.

For the composite function (1.1), we consider a special nonconvex function h, lower-c^2 function. It is clear that the linearization errors $e^k_{h,i}$ may be negative. By Theorem 2.1, we know that there exists a threshold $\bar{\rho}$ such that the augmented function $\varphi_n(\cdot)$

$$\varphi_n(x) := h(x) + \frac{\eta_n}{2}\|x - \hat{x}^k\|^2, \tag{2.2}$$

is convex in the set $\mathcal{V}$ whenever $\eta_n \geq \bar{\rho}$ holds. But the value $\bar{\rho}$ is not known in prior, then if $\eta_n < \bar{\rho}$, the augmented function φ_n may not be a convex function in the set $\mathcal{V}$. However by the definition of the linearization error, we have that if we choose the convexification parameter η_n suitably, the linearization error of the augmented function $\varphi_n(\cdot)$ may be nonnegative. Concretely for any $i \in I_n$, the linearization errors for functions h and φ_n are as follows:

$$e^k_{h,i} = h(\hat{x}^k) - h(x^i) - \langle g^i_h, \hat{x}^k - x^i\rangle, \quad \text{for } g^i_h \in \partial h(x^i).$$

and

$$e^k_{\varphi,i} = \varphi_n(\hat{x}^k) - \varphi_n(x^i) - \langle g^i_\varphi, \hat{x}^k - x^i\rangle, \quad \text{for } g^i_\varphi \in \partial \varphi_n(x^i).$$

By the definition of function φ_n, we have $\varphi_n(\hat{x}^k) = h(\hat{x}^k)$. With the calculation of the subdifferential and the relationship (2.2), we have

$$g^i_\varphi = g^i_h + \eta_n(x^i - \hat{x}^k).$$

and

$$e^k_{\varphi,i} = h(\hat{x}^k) - h(x^i) - \frac{\eta_n}{2}\|x^i - \hat{x}^k\|^2 - \langle g^i_h + \eta_n(x^i - \hat{x}^k), \hat{x}^k - x^i\rangle = e^k_{h,i} + \frac{\eta_n}{2}\|x^i - \hat{x}^k\|^2. \tag{2.3}$$

Then by choosing suitable parameter η_n, we may make sure that the linearization errors of function φ_n in index set I_n to be nonnegative. Next we give the rule for parameter η_n,

$$\eta_n \geq \overline{\eta}_n := \max\left\{\max\left\{-\frac{2e^k_{h,i}}{\|x^i - \hat{x}^k\|^2}, \right\}, 0\right\} + \gamma \quad \text{with } i \in I_n \text{ and } \|x^i - \hat{x}^k\| \neq 0, \tag{2.4}$$

where γ is a positive constant. By the rule, we have that for all $i \in I_n$, $e^k_{\varphi,i} \geq \frac{\gamma}{2}\|x^i - \hat{x}^k\|^2 \geq 0$ holds. Similar rules for the parameter η_n can also be seen in [19, 24] and reference therein. Note also that the parameter η_n may vary along iterations.

In the following, our attention will be paid to the approximation function $\phi_n(x)$ of $\psi(x)$, which is defined as follows:

$$\phi_n(x) := f(x) + \varphi_n(x). \tag{2.5}$$

In fact, function ϕ_n can also be regarded as a local "convexification" of the problem (1.1). For convex function f, its linearization error is defined by

$$e^k_{f,i} = f(\hat{x}^k) - f(x^i) - \langle g^i_f, \hat{x}^k - x^i \rangle \geq 0, \quad \text{for } g^i_f \in \partial f(x^i), \quad \text{and } \forall i \in I_n.$$

Note that since function f is finite convex, by (2.1), we have $g^i_f \in \partial_{e^k_{f,i}} f(\hat{x}^k)$. Meanwhile, if $\eta_n \geq \bar{\rho}$, it also holds $g^i_\varphi \in \partial_{e^k_{\varphi,i}} \varphi_n(\hat{x}^k)$.

Based on the linearization errors, the bundles of f and h can respectively be

$$\mathscr{B}^n_f = \left\{ \left(e^k_{f,i} \in \mathbf{R}_+, g^i_f \in \partial_{e^k_{f,i}} f(\hat{x}^k) \right), i \in I_n \right\}, \quad \text{and} \quad \left(\hat{x}^k, f(\hat{x}^k), g^k_f \in \partial f(\hat{x}^k) \right),$$
$$\mathscr{B}^n_h = \left\{ \left(x^i, h_i = h(x^i), g^i_h \in \partial h(x^i), e^k_{h,i} \right), i \in I_n \right\}.$$

Then the cutting planes for f and φ_n at the current serious step $\hat{x}^k$ are stated as follows:

$$\widetilde{f}_n(x) = \max_{i \in I_n} \left\{ f_i + \langle g^i_f, x - x^i \rangle \right\} = f(\hat{x}^k) + \max_{i \in I_n} \left\{ -e^k_{f,i} + \langle g^i_f, x - \hat{x}^k \rangle \right\}, \tag{2.6a}$$

and

$$\begin{aligned} \widetilde{\varphi}_n(x) =& \max_{i \in I_n} \left\{ \varphi_n(x^i) + \langle g^i_h + \eta_n(x^i - \hat{x}^k), x - x^i \rangle \right\} \\ =& h(\hat{x}^k) + \max_{i \in I_n} \left\{ - \left(e^k_{h,i} + \frac{\eta_n}{2} \|x^i - \hat{x}^k\|^2 \right) + \langle g^i_h + \eta_n(x^i - \hat{x}^k), x - \hat{x}^k \rangle \right\}. \end{aligned} \tag{2.6b}$$

To ensure the sequence of approximate points is well defined and the corresponding algorithm is convergent, the approximate models must be carefully chosen. In this paper, we regard $\widetilde{f}_n(x) + \widetilde{\varphi}_n(x)$ as the cutting plane model for the approximate function ϕ_n. For simplify, we note

$$\widetilde{\phi}_n(x) := \widetilde{f}_n(x) + \widetilde{\varphi}_n(x). \tag{2.6c}$$

Then the next point x^{n+1} may be solved by the following quadratic programming (QP) subproblems:

$$x^{n+1} := \arg \min_{x \in \mathbf{R}^N} \left\{ \widetilde{\phi}_n(x) + \frac{\mu_n}{2} \|x - \hat{x}^k\|^2 \right\}, \tag{2.7}$$

where $\mu_n > 0$ is called proximal parameter. By (2.6a) and (2.6b), we have that function $\widetilde{\phi}_n$ is convex. Since $\mu_n > 0$, it holds that x^{n+1} is the unique solution of subproblem (2.7) with $x^{n+1} = p_{\mu_n} \widetilde{\phi}_n(\hat{x}^k)$.

The following lemma shows the relationship between the new point x^{n+1} and the current stability center $\hat{x}^k$. Similar conclusion can also be found in [33, Lemma 10.6]. Let $S_n = \{z \in [0, 1]^{|I_n|} : \sum_{i \in I_n} z^i = 1\}$ denote a unit simplex in $\mathbf{R}^{|I_n|}$.

Lemma 2.1 *Let x^{n+1} be the unique solution to the subproblem (2.7) and the proximal parameter $\mu_n > 0$. Then we have*

$$x^{n+1} = \hat{x}^k - \frac{1}{\mu_n} \hat{G}^k, \tag{2.8}$$

where

$$\hat{G}^k := \sum_{i \in I_n} \alpha^i_1 g^i_f + \sum_{i \in I_n} \alpha^i_2 g^i_\varphi, \tag{2.9}$$

and $\alpha_1 = (\alpha_1^1, \cdots, \alpha_1^{|I_n|}) \in S_n$, $\alpha_2 = (\alpha_2^1, \cdots, \alpha_2^{|I_n|}) \in S_n$ *is a solution to*

$$\min_{\alpha_1,\ \alpha_2 \in \mathbf{R}_+^{|I_n|}} \left\{ \frac{1}{2\mu_n} \left\| \sum_{i \in I_n} \alpha_1^i g_f^i + \sum_{i \in I_n} \alpha_2^i g_\varphi^i \right\|^2 + \sum_{i \in I_n} \alpha_1^i e_f^{i,k} + \sum_{i \in I_n} \alpha_2^i e_\varphi^{i,k} \right\}. \tag{2.10}$$

In addition, the following relations hold:

(i) $\hat{G}^k \in \partial \left(\widetilde{f}_n + \widetilde{\varphi}_n \right)(x^{n+1})$*;*
(ii) $f(\hat{x}^k) + h(\hat{x}^k) - \widetilde{f}_n(x^{n+1}) - \widetilde{\varphi}_n(x^{n+1}) = \mu_n \|x^{n+1} - \hat{x}^k\|^2 + \varepsilon_{n+1}$, *where* $\varepsilon_{n+1} = \sum_{i \in I_n} \alpha_1^i e_f^{i,k} + \sum_{i \in I_n} \alpha_2^i e_\varphi^{i,k}$.

Proof Write the objective function of (2.7) as a QP subproblem with two extra scalar variables r_1 and r_2 as follows

$$\begin{cases} \min\limits_{(x,\ r_1, r_2) \in \mathbf{R}^N \times \mathbf{R} \times \mathbf{R}} & r_1 + r_2 + \frac{\mu_n}{2} \|x - \hat{x}^k\|^2 \\ \text{s. t.} & r_1 \geq f(\hat{x}^k) - e_f^{i,k} + \langle g_f^i, x - \hat{x}^k \rangle, \quad \forall i \in I_n, \\ & r_2 \geq h(\hat{x}^k) - e_\varphi^{i,k} + \langle g_\varphi^i, x - \hat{x}^k \rangle, \quad \forall i \in I_n. \end{cases}$$

The corresponding Lagrange function is, for $\alpha_1,\ \alpha_2 \in \mathbf{R}_+^{|I_n|}$,

$$\begin{aligned} L(x, r_1, r_2, \alpha_1, \alpha_2) =& r_1 + r_2 + \frac{\mu_n}{2} \|x - \hat{x}^k\|^2 \\ &+ \sum_{i \in I_n} \alpha_1^i \left(f(\hat{x}^k) - e_f^{i,k} + \langle g_f^i, x - \hat{x}^k \rangle - r_1 \right) \\ &+ \sum_{i \in I_n} \alpha_2^i \left(h(\hat{x}^k) - e_\varphi^{i,k} + \langle g_\varphi^i, x - \hat{x}^k \rangle - r_2 \right) \\ =& \left(1 - \sum_{i \in I_n} \alpha_1^i \right) r_1 + \left(1 - \sum_{i \in I_n} \alpha_2^i \right) r_2 \\ &+ \frac{\mu_n}{2} \|x - \hat{x}^k\|^2 + \sum_{i \in I_n} \alpha_1^i \left(f(\hat{x}^k) - e_f^{i,k} + \langle g_f^i, x - \hat{x}^k \rangle \right) \\ &+ \sum_{i \in I_n} \alpha_2^i \left(h(\hat{x}^k) - e_\varphi^{i,k} + \langle g_\varphi^i, x - \hat{x}^k \rangle \right). \end{aligned} \tag{2.11}$$

In view of strong convexity, (2.7) has the unique solution x^{n+1}. Furthermore, the equivalent problem (2.11) has affine constraints, and there exists optimal multiplier (α_1, α_2) associate with x^{n+1}. Since there is no duality gap, the optimal solution $(x^{n+1}, \alpha_1, \alpha_2)$ can also be obtained either by solving the primal problem or solving its dual, i.e.,

$$(2.11) \equiv \min_{(x, r_1, r_2)} \max_{(\alpha_1, \alpha_2)} L(x, r_1, r_2, \alpha_1, \alpha_2) \equiv \max_{(\alpha_1, \alpha_2)} \min_{(x, r_1, r_2)} L(x, r_1, r_2, \alpha_1, \alpha_2).$$

All the problems above have the same finite optimal value. However, the dual problem involves the unconstrained minimization of the function L with respect to r_1 and r_2. When the dual value is finite, the terms multiplying r_1 and r_2 in L have to vanish, i.e., α_1 and α_2 must lie in the unit simplex S_n. As a result, $(x^{n+1}, \alpha_1, \alpha_2)$ solves the primal and dual problems

$$\min_{x \in \mathbf{R}^N} \max_{(\alpha_1, \alpha_2) \in \mathbf{R}_+^{|I_n|} \times \mathbf{R}_+^{|I_n|}} L(x, \alpha_1, \alpha_2) \equiv \max_{(\alpha_1, \alpha_2) \in \mathbf{R}_+^{|I_n|} \times \mathbf{R}_+^{|I_n|}} \min_{x \in \mathbf{R}^N} L(x, \alpha_1, \alpha_2).$$

where

$$L(x,\alpha_1,\alpha_2) = f(\hat{x}^k) + h(\hat{x}^k) + \frac{\mu_n}{2}\|x-\hat{x}^k\|^2 + \sum_{i\in I_n}\alpha_1^i\left(-e_f^{i,k} + \langle g_f^i, x-\hat{x}^k\rangle\right)$$
$$+\sum_{i\in I_n}\alpha_2^i\left(-e_\varphi^{i,k} + \langle g_\varphi^i, x-\hat{x}^k\rangle\right).$$

Consider the last dual problem, for each $(\alpha_1,\alpha_2)\in S_n\times S_n$ fixed, defining $x(\alpha_1,\alpha_2) = \arg\min_x \ L(x,\alpha_1,\alpha_2)$, the optimality condition is $0 = \nabla_x L(x,\alpha_1,\alpha_2)$, i.e.,

$$0 = \mu_n(x-\hat{x}^k) + \sum_{i\in I_n}\alpha_1^i g_f^i + \sum_{i\in I_n}\alpha_2^i g_\varphi^i, \tag{2.12}$$

then the relation (2.8) holds.

To prove (α_1,α_2) solves (2.10),we can multiply (2.12) by $x-\hat{x}^k$ and by $\frac{1}{\mu_n}\left(\sum_{i\in I_n}\alpha_1^i g_f^i + \sum_{i\in I_n}\alpha_2^i g_\varphi^i\right)$ respectively, then it holds

$$\begin{aligned}0 =&\mu_n\left\|x-\hat{x}^k\right\|^2 + \sum_{i\in I_n}\alpha_1^i\langle g_f^i, x-\hat{x}^k\rangle + \sum_{i\in I_n}\alpha_2^i\langle g_\varphi^i, x-\hat{x}^k\rangle\\ =&\sum_{i\in I_n}\alpha_1^i\langle g_f^i, x-\hat{x}^k\rangle + \sum_{i\in I_n}\alpha_2^i\langle g_\varphi^i, x-\hat{x}^k\rangle\\ &+\frac{1}{\mu_n}\left\|\sum_{i\in I_n}\alpha_1^i g_f^i + \sum_{i\in I_n}\alpha_2^i g_\varphi^i\right\|^2,\end{aligned}$$

which implies that

$$\mu_n\left\|x-\hat{x}^k\right\|^2 = \frac{1}{\mu_n}\left\|\sum_{i\in I_n}\alpha_1^i g_f^i + \sum_{i\in I_n}\alpha_2^i g_\varphi^i\right\|^2,$$

and

$$L(x,\alpha_1,\alpha_2) = f(\hat{x}^k) + h(\hat{x}^k) - \frac{1}{2\mu_n}\left\|\sum_{i\in I_n}\alpha_1^i g_f^i + \sum_{i\in I_n}\alpha_2^i g_\varphi^i\right\|^2 - \sum_{i\in I_n}\alpha_1^i e_f^{i,k}$$
$$-\sum_{i\in I_n}\alpha_2^i e_\varphi^{i,k}.$$

Based on above discussions, (α_1,α_2) solves

$$\begin{aligned}&\max_{(\alpha_1,\alpha_2)\in S_n\times S_n} L(x,\alpha_1,\alpha_2) = f(\hat{x}^k) + h(\hat{x}^k)\\ &-\min_{(\alpha_1,\alpha_2)\in S_n\times S_n}\left\{\frac{1}{2\mu_n}\left\|\sum_{i\in I_n}\alpha_1^i g_f^i + \sum_{i\in I_n}\alpha_2^i g_\varphi^i\right\|^2 + \sum_{i\in I_n}\alpha_1^i e_f^{i,k} + \sum_{i\in I_n}\alpha_2^i e_\varphi^{i,k}\right\},\end{aligned} \tag{2.13}$$

so the relation (2.10) holds. The assertion (i) follows from the optimality condition of (2.7) and (2.9). Next we show the assertion (ii). Since there is no duality gap, the primal optimal value in (2.7) is equal to the dual optimal value in (2.13), i.e.,

$$\begin{aligned}&\widetilde{f}_n(x^{n+1})+\widetilde{\varphi}_n(x^{n+1})+\frac{\mu_n}{2}\|x^{n+1}-\hat{x}^k\|^2\\&\quad=f(\hat{x}^k)+h(\hat{x}^k)\\&\quad-\frac{1}{2\mu_n}\left\|\sum_{i\in I_n}\alpha_1^i g_f^i+\sum_{i\in I_n}\alpha_2^i g_\varphi^i\right\|^2-\left(\sum_{i\in I_n}\alpha_1^i e_f^{i,k}+\sum_{i\in I_n}\alpha_2^i e_\varphi^{i,k}\right).\end{aligned}$$

By (2.8), the above inequality can be rewritten as

$$\begin{aligned}&f(\hat{x}^k)+h(\hat{x}^k)-\widetilde{f}_n(x^{n+1})-\widetilde{\varphi}_n(x^{n+1})\\&\quad=\frac{\mu_n}{2}\|x^{n+1}-\hat{x}^k\|^2+\frac{1}{2\mu_n}\left\|\sum_{i\in I_n}\alpha_1^i g_f^i+\sum_{i\in I_n}\alpha_2^i g_\varphi^i\right\|^2+\varepsilon_{n+1}\\&\quad=\frac{1}{2\mu_n}\left\|\hat{G}^k\right\|^2+\frac{1}{2\mu_n}\left\|\hat{G}^k\right\|^2+\varepsilon_{n+1}\\&\quad=\frac{1}{\mu_n}\left\|\hat{G}^k\right\|^2+\varepsilon_{n+1}=\mu_n\|x^{n+1}-\hat{x}^k\|^2+\varepsilon_{n+1},\end{aligned}$$

where $\varepsilon_{n+1}=\sum_{i\in I_n}\alpha_1^i e_f^{i,k}+\sum_{i\in I_n}\alpha_2^i e_\varphi^{i,k}$. This completes the proof. □

By the convexity of $\widetilde{f}_n$ and $\widetilde{\varphi}_n$ and the calculus rule of subdifferentials in convex analysis, it holds

$$\partial(\widetilde{f}_n+\widetilde{\varphi}_n)(x^{n+1})=\partial\widetilde{f}_n(x^{n+1})+\partial\widetilde{\varphi}_n(x^{n+1}).$$

As the iteration proceeds, the data in the bundle become too much, which will affect the algorithm's efficiency. For the sake of efficiency, the cardinality for I_n should not grow beyond control as n increases. To overcome this shortcoming, the reformulated bundle opens the ways of mechanism. One is known as the bundle compression, which allows to keep bounded the cardinality of I_n as $n\to\infty$, without impairing the convergence of the method. Another way is to keep only active data, and this technique is called bundle aggregate. The aggregate information of the bundle includes as least as two: one is the aggregation information and the other is the new generated information. In this paper, we use both techniques, and also append the current serious step information to the bundle. In the following, we regard $J_{1,n}^{act}$ and $J_{2,n}^{act}$ as the active index set, i.e.,

$$J_{1,n}^{act}:=\{i\in I_n\mid\alpha_1^i>0\},\quad\alpha_1\in S_n\text{ satisfying (2.10)}.$$

and

$$J_{2,n}^{act}:=\{i\in I_n\mid\alpha_2^i>0\},\quad\alpha_2\in S_n\text{ satisfying (2.10)}.$$

At the same time, we introduce the negative indices which can also be seen in [24, 32] for the aggregate bundle, that is, $I_n\subseteq\{-n,-n+1,\cdots,0,1,\cdots,n\}$. Meanwhile, we give the following definitions

$$e_{f,-n}^k=\sum_{i\in I_n}\alpha_1^i e_{f,i}^k\qquad e_{\varphi,-n}^k=\sum_{i\in I_n}\alpha_2^i e_{\varphi,i}^k=\sum_{i\in I_n}\left(e_{h,i}^k+\frac{\eta_n}{2}\|x^i-\hat{x}^k\|^2\right)$$

and

$$g_f^{-n} = \sum_{i\in I_n} \alpha_1^i g_f^i \qquad g_\varphi^{-n} = \sum_{i\in I_n} \alpha_2^i g_\varphi^i = \sum_{i\in I_n} \alpha_2^i \left(g_h^i + \eta_n(x^i - \hat{x}^k)\right).$$

Then, for the aggregative error ε^{n+1}, by the above definitions, we have

$$\begin{aligned}\varepsilon^{n+1} &= \sum_{i\in I_n} \alpha_1^i e_{f,i}^k + \sum_{i\in I_n} \alpha_2^i e_{\varphi,i}^k = e_{f,-n}^k + e_{\varphi,-n}^k \\ &= \sum_{i\in J_{1,n}^{act}} \alpha_1^i e_{f,i}^k + \sum_{i\in J_{2,n}^{act}} \alpha_2^i \left(e_{h,i}^k + \frac{\eta_n}{2}\|x^i - \hat{x}^k\|^2\right).\end{aligned}$$

For the aggregate subgradient $\hat{G}^k$, it holds

$$\begin{aligned}\hat{G}^k &= \sum_{i\in I_n} \alpha_1^i g_f^i + \sum_{i\in I_n} \alpha_2^i \left(g_h^i + \eta_n(x^i - \hat{x}^k)\right) = g_f^{-n} + g_\varphi^{-n} \\ &= \sum_{i\in J_{1,n}^{act}} \alpha_1^i g_f^i + \sum_{i\in J_{2,n}^{act}} \alpha_2^i \left(g_h^i + \eta_n(x^i - \hat{x}^k)\right).\end{aligned}$$

Note that for all $\iota \in J_{1,n}^{act} \bigcap J_{2,n}^{act}$ and $\iota = -n$, the following two equalities hold

$$\widetilde{f}_n(x^{n+1}) = f(\hat{x}^k) - e_{f,\iota}^k + \langle g_f^\iota, x^{n+1} - \hat{x}^k\rangle, \tag{2.14a}$$

$$\widetilde{\varphi}_n(x^{n+1}) = \varphi_n(\hat{x}^k) - \left(e_{h,\iota}^k + \frac{\eta_n}{2}\|x^\iota - \hat{x}^k\|^2\right) + \langle g_h^\iota + \eta_n(x^\iota - \hat{x}^k), x^{n+1} - \hat{x}^k\rangle. \tag{2.14b}$$

3 The proximal bundle algorithm

In this section, we will present our algorithm for the problem (1.1). Before that, we firstly give some necessary notations for the algorithm. The predicted decrease at x^{n+1} for functions f and φ_n are respectively defined by:

$$\delta_{n+1}^f := f(\hat{x}^k) - \widetilde{f}_n(x^{n+1}),$$

and

$$\delta_{n+1}^\varphi := \varphi_n(\hat{x}^k) + \frac{\eta_n}{2}\|x^{n+1} - \hat{x}^k\|^2 - \widetilde{\varphi}_n(x^{n+1}) = h(\hat{x}^k) + \frac{\eta_n}{2}\|x^{n+1} - \hat{x}^k\|^2 - \widetilde{\varphi}_n(x^{n+1}).$$

By the convexity of function f, we have $\delta_{n+1}^f \geq 0$. By (2.2), (2.6b) and the choice of the parameter η_n, we have $\delta_{n+1}^\varphi \geq 0$. For the approximate function ϕ_n, we definite its predict decrease as follows

$$\delta_{n+1} := \delta_{n+1}^f + \delta_{n+1}^\varphi. \tag{3.1}$$

By the above discussions, it holds that $\delta_{n+1} \geq 0$. In fact, according to Lemma 2.1 and (3.1), if we set $R_n = \eta_n + \mu_n$, then it holds that

$$\delta_{n+1} = \mu_n\|x^{n+1} - \hat{x}^k\|^2 + \frac{\eta_n}{2}\|x^{n+1} - \hat{x}^k\|^2 + \varepsilon^{n+1} = \frac{R_n + \mu_n}{2}\|x^{n+1} - \hat{x}^k\|^2 + \varepsilon^{n+1}.$$

Now we present our proximal bundle algorithm for problem (1.1).

Algorithm 1 *(Nonconvex Nonsmooth Proximal Bundle Method for a class of composite optimization)*

Step 0 (Input and Initialization):

Select an initial starting point x^0, an unacceptable increase parameter $M_0 > 0$, an initial parameter $R_0 > 0$, a stopping tolerance $TOL_{stop} \geq 0$, a convexification grow parameter $\Gamma_\eta > 1$, a proximal grow parameter $\Gamma_\mu > 1$ and an Armijo-like parameter $m \in (0, 1)$. Set the initial iteration counter $n = 0$, the serious step counter $k = k(n) = 0$, the bundle index set $I_0 = \{0\}$ and $\hat{x}^0 = x^0$. Call the black box to compute f^0, g_f^0, h^0, g_h^0 and set $e_{f,0}^0 = 0$, $e_{h,0}^0 = 0$. Choose the starting parameter distribution $(\mu_0, \eta_0) = (R_0, 0)$;

Step 1 (Model generation and QP subproblem):

Having the current proximal center $\hat{x}^k$, the current bundle $\mathcal{B}_f^n$ and $\mathcal{B}_h^n$ with index set I_n, and the current proximal parameter distribution (μ_n, η_n) with $\eta_n \leq R_n$. Compute $\widetilde{f}_n(x)$ from (2.6a) and $\widetilde{\varphi}_n(x)$ from (2.6b). Solve (2.7) and (3.1) to get x^{n+1} and δ_{n+1}, respectively. Meanwhile, we get the optimal simplicial multipliers $(\alpha_1, \alpha_2) \in S_n \times S_n$ for (2.8). We give two choices for a new index set, which satisfies

$$\text{Either } J_n^{act} \subseteq I_{n+1} \text{ and } \{n+1, i_k\} \subseteq I_{n+1} \quad \text{Or } \{-n\} \subseteq I_{n+1} \text{ and } \{n+1, i_k\} \subseteq I_{n+1}$$

where i_k is the iterative index giving the current serious points, i.e., $\hat{x}^k = x^{i_k}$ and $J_n^{act} := J_{1,n}^{act} \bigcup J_{2,n}^{act}$. Then, go to Step 2;

Step 2 (Stopping criterion):

If $\delta_{n+1} \leq TOL_{stop}$, then stop. Otherwise, go to Step 3;

Step 3 (Serious step testing):

If the descent condition

$$\psi(x^{n+1}) \leq \psi(\hat{x}^k) - m\delta_{n+1}, \tag{3.2}$$

holds, then declare a serious step, and set $k(n+1) = k+1$, $i_{k+1} = n+1$, and $\hat{x}^{k+1} = x^{n+1}$. Update the bundles, go to Step 4; otherwise, declare a null step, and set $k(n+1) = k(n)$, go to Step 4;

Step 4 (Update convexification parameter η):

Set

$$\begin{cases} \eta_{n+1} = \eta_n, & if \ \ \overline{\eta}_{n+1} \leq \eta_n; \\ \eta_{n+1} = \Gamma \overline{\eta}_{n+1} & if \ \ \overline{\eta}_{n+1} > \eta_n; \end{cases} \tag{3.3}$$

where $\overline{\eta}_{n+1}$ is computed by (2.4) with n replaced by $n+1$;

Step 5 (Update model proximal parameter μ):

If the following testing condition

$$\psi(x^{n+1}) > \psi(\hat{x}^k) + M_0,$$

holds, then the objective increase is unacceptable, restart the algorithm by setting:

$$\eta_0 := \eta_n, \ \ \mu_0 := \Gamma \mu_n, \ \ R_0 := \eta_0 + \mu_0, \ x^0 := \hat{x}^k, \ k(0) := 0, \ i_0 := 0, \ I_0 := \{0\}, \ n := 0,$$

then go to Step 1; otherwise, increase k by 1 in the case of the serious step. In all cases, increase n by 1, and loop to Step 1. □

Remark 3.1 Note that the updating strategies for serious steps and null steps are much different in bundle methods. If a serious step occurs, the linearization errors and the stability center are updated, that is, the new generated point is regarded as a new stability center and the corresponding linearization errors update along the new center. If a null step happens, the new generated information should be added into the bundle to improve approximate accuracy and the linearization errors keep unchanged.

When the descent condition (3.2) is satisfied, the new generated iterative point x^{n+1} is regarded as a new serious point, i.e., $\hat{x}^{k+1} = x^{n+1}$, the linearization errors for functions f, h and φ_n can be respectively updated as follows:

$$e_{f,i}^{k+1} = e_{f,i}^{k} + f(\hat{x}^{k+1}) - f(\hat{x}^{k}) - \langle g_f^i, \hat{x}^{k+1} - \hat{x}^{k}\rangle, \ \ \forall\, i \in I_{n+1}, \tag{3.4a}$$

$$e_{h,i}^{k+1} = e_{h,i}^{k} + h(\hat{x}^{k+1}) - h(\hat{x}^{k}) - \langle g_h^i, \hat{x}^{k+1} - \hat{x}^{k}\rangle, \ \ \forall\, i \in I_{n+1}, \tag{3.4b}$$

and

$$e_{\varphi,i}^{k+1} = e_{h,i}^{k+1} + \frac{\eta_n}{2}\|x^i - \hat{x}^{k+1}\|^2, \quad \forall\, i \in I_{n+1}. \tag{3.4c}$$

In the following, we prove that the stopping test is reasonable and the algorithm is well defined. Before that, we first present our assumption which is always used in bundle methods and can also be seen in [11, 22, 24] and reference therein.

Assumption 3.1 The level set

$$\mathcal{T} := \{x \in \mathbf{R}^N :\ \psi(x) \le \psi(\hat{x}^0) + M_0\ \}$$

is compact.

For $\eta_n > 0$, the augmented function $h(x) + \frac{\eta_n}{2}\|x - \hat{x}^k\|^2$ is somehow "more convex" than the function $h(x)$. By Assumption 3.1 and Theorem 2.1, there exists a threshold $\bar{\rho}$ such that for any $\eta \ge \bar{\rho}$, it holds that

$$\text{The function } h(\cdot) + \frac{\eta}{2}\|\cdot - \hat{x}^k\|^2 \text{ is convex on } \mathcal{T}.$$

where $\hat{x}^k$ is current stability center. At this moment, for functions f and φ_n, the cutting planes model functions satisfy $\widetilde{f}_n(\cdot) \le f(\cdot)$ and $\widetilde{\varphi}_n(\cdot) \le \varphi_n(\cdot)$, thus $\mu_n(\hat{x}^k - x^{n+1}) \in \partial(\widetilde{f}_n + \widetilde{\varphi}_n)(x^{n+1})$ implies that for all $x \in \mathcal{T}$, we have

$$\begin{aligned}\phi_n(x) \ge \widetilde{f}_n(x) + \widetilde{\varphi}_n(x) &\ge \phi_n(x^{n+1}) + \mu_n\left\langle \hat{x}^k - x^{n+1}, x - x^{n+1}\right\rangle \\ &- \left[\phi_n(x^{n+1}) - \widetilde{f}_n(x^{n+1}) - \widetilde{\varphi}_n(x^{n+1})\right],\end{aligned}$$

where $\phi_n(x)$ is defined in (2.5). In particular, for $x = p_{\mu_n}\phi_n(\hat{x}^k)$, we obtain

$$\begin{aligned}\phi_n\left(p_{\mu_n}\phi_n(\hat{x}^k)\right) &\ge \widetilde{f}_n\left(p_{\mu_n}\phi_n(\hat{x}^k)\right) + \widetilde{\varphi}_n\left(p_{\mu_n}\phi_n(\hat{x}^k)\right) \\ &\ge \phi_n(x^{n+1}) + \mu_n\left\langle \hat{x}^k - x^{n+1}, p_{\mu_n}\phi_n(\hat{x}^k) - x^{n+1}\right\rangle \\ &\quad - \left[\phi_n(x^{n+1}) - \widetilde{f}_n(x^{n+1}) - \widetilde{\varphi}_n(x^{n+1})\right].\end{aligned}$$

Similarly, since $p_{\mu_n}\phi_n(\hat{x}^k)$ is characterized by $\mu_n(\hat{x}^k - p_{\mu_n}\phi_n(\hat{x}^k)) \in \partial\phi_n(\hat{x}^k)$, the associate subgradient inequality becomes

$$\phi_n(x^{n+1}) \ge \phi_n\left(p_{\mu_n}\phi_n(\hat{x}^k)\right) + \mu_n\left\langle \hat{x}^k - p_{\mu_n}\phi_n(\hat{x}^k), x^{n+1} - p_{\mu_n}\phi_n(\hat{x}^k)\right\rangle.$$

Combining above two inequalities and it yields that

$$\phi_n(x^{n+1}) - \widetilde{f}_n(x^{n+1}) - \widetilde{\varphi}_n(x^{n+1}) \ge \mu_n\|x^{n+1} - p_{\mu_n}\phi_n(\hat{x}^k)\|^2.$$

By the descent condition (3.2), we get that the stopping criterion in Algorithm 1 is reasonable.

In the following, we will show that the restart steps in Algorithm 1 is finite. Before that, we should note that all lower-c^2 functions are locally Lipschitz continuous (see [1, Theorem 10.31]), then f and h are all locally Lipschitz continuous. The compactness of $\mathcal{T}$ allows us to find the Lipschitz constants for functions f and h, denoted by L_f and L_h respectively.

Lemma 3.1 *Consider $\{x^n\}$ generated by Algorithm 1 and $i_k \in I_n$, then there can be only a finite number of restarts in Step 5 in Algorithm 1. Moreover, the sequence $\{\mu_n\}$ finally becomes a constant sequence.*

Proof Firstly, the new iterative point x^{n+1} is well defined by the convexity of $\widetilde{\phi}_n(\cdot)$. As functions f and h are Lipschitz continuous, ψ is also Lipschitz continuous in $\mathcal{T}$ and denoting its Lipschitz constant is $L := L_f + L_h$. By the Lipschitz continuity of ψ, there exists $\epsilon > 0$ such that, for any $\widetilde{x} \in \{x : \psi(x) \leq \psi(\hat{x}^0)\}$, the open set $B_\epsilon(\widetilde{x})$ is contained in $\mathcal{T}$. Note that

$$
\begin{aligned}
p_{\mu_n}(\widetilde{f}_n + \widetilde{\varphi}_n)(\hat{x}^k) &= \arg\min_x \left\{ \widetilde{f}_n(x) + \widetilde{\varphi}_n(x) + \mu_n \|x - \hat{x}^k\|^2/2 \right\} \\
&\in \left\{ x : \widetilde{f}_n(x) + \widetilde{\varphi}_n(x) + \mu_n \|x - \hat{x}^k\|^2/2 \leq \widetilde{f}_n(\hat{x}^k) + \widetilde{\varphi}_n(\hat{x}^k) \right\} \\
&\in \Big\{ x : \widetilde{f}_n(\hat{x}^k) + \widetilde{\varphi}_n(\hat{x}^k) + \langle g^{i_k}, x - \hat{x}^k \rangle \\
&\quad + \mu_n \|x - \hat{x}^k\|^2/2 \leq \widetilde{f}_n(\hat{x}^k) + \widetilde{\varphi}_n(\hat{x}^k) \Big\} \\
&\in \left\{ x : -\|g^{i_k}\| \, \|x - \hat{x}^k\| + \mu_n \|x - \hat{x}^k\|^2/2 \leq 0 \right\} \\
&\in \left\{ x : \|x - \hat{x}^k\| \leq 2L/\mu_n \right\}.
\end{aligned}
$$

where $i_k \in I_n$, and $g^{i_k} \in \partial(\widetilde{f}_n + \widetilde{\varphi}_n)(\hat{x}^k)$. It holds $g^{i_k} \in \partial \phi_n(\hat{x}^k)$ and $\|g^{i_k}\| \leq L$. In Step 5, μ_n increases when the restart steps happen, or keep unchanged when the restart steps do not happen, eventually μ_n becomes large enough so that the relationship $2L/\mu_n < \epsilon$ holds. Noting that $\psi(\hat{x}^k) \leq \psi(\hat{x}^0)$ for any new generated point $\hat{x}^k$ in Algorithm 1 completes the proof. □

4 The convergence theory

In this section, we give the convergence of Algorithm 1. There are different techniques in dealing with $\widetilde{\phi}_n(x)$. In [19], the authors use the bundle compression to handle the information in bundles, but they don't allow to erase the active bundle information. In [24], the authors also utilize the bundle compression technique and they allow to erase active bundle elements in order to use aggregate technique. In our algorithm, we use the latter technique.

Lemma 4.1 *Consider the approximate model functions* (2.6a) *and* (2.6b) *and $\widetilde{\phi}_n(x)$. then we have:*

(i) Condition

$$\widetilde{\phi}_n(x) \text{ is a convex function,}$$

is always satisfied.

(ii) If for any index n, $\eta_n \geq \bar{\eta}_n$ with $\bar{\eta}_n$ defined in (2.4)*, then*

$$\widetilde{\phi}_n(\hat{x}^k) \leq \psi(\hat{x}^k),$$

(iii) If the condition $\eta_{n+1} = \eta_n \geq \bar{\rho}$ holds, x^{n+1} is a null point and either $J_n^{act} \subseteq I_{n+1}$ or $-n \in I_{n+1}$, then

$$
\begin{aligned}
\widetilde{\phi}_{n+1}(x) \geq \widetilde{\phi}_n(x^{n+1}) + \mu_n \langle \hat{x}^k - x^{n+1}, x - x^{n+1} \rangle,& \\
\text{for all } x \in \mathbf{R}^N, \text{ and if } x^{n+1} \text{ is a null step,}& \qquad (4.1)
\end{aligned}
$$

(iv) If index $n \in I_n$, then for some $g_f^n \in \partial f(x^n)$, $g_h^n \in \partial h(x^n)$ and all $x \in \mathbf{R}^N$, we have

$$\widetilde{\phi}_n(x) \geq \psi(x^n) + \frac{\eta_n}{2}\|x^n - \hat{x}^k\|^2 + \langle g_f^n + g_h^n + \eta_n(x^n - \hat{x}^k), x - \hat{x}^k \rangle.$$

Proof (i) By (2.6a), (2.6b) and (2.6c), $\widetilde{\phi}_n(x)$ is defined as a sum of the point-wise maximum of a finite collection of affine functions, therefore, $\widetilde{\phi}_n(x)$ is convex, and item (i) is always satisfied.

(ii) By the definition of $\widetilde{\phi}_n(x)$, i.e., (2.6c), it holds that

$$\widetilde{\phi}_n(\hat{x}^k) = \widetilde{f}_n(\hat{x}^k) + \widetilde{\varphi}_n(\hat{x}^k) = f(\hat{x}^k) + \max_{i \in I_n}\left\{-e_{f,i}^k\right\} + h(\hat{x}^k) + \max_{i \in I_n}\left\{-\left(e_{h,i}^k + \frac{\eta_n}{2}\|x^i - \hat{x}^k\|^2\right)\right\}.$$

By the convexity of f and the choice for the parameter η_n, it holds that for any $i \in I_n$

$$e_{f,i}^k \geq 0, \quad \text{and} \quad e_{h,i}^k + \frac{\eta_n}{2}\|x^i - \hat{x}^k\|^2 \geq 0.$$

Combining the above two inequalities, we have

$$\widetilde{\phi}_n(\hat{x}^k) = \widetilde{f}_n(\hat{x}^k) + \widetilde{\varphi}_n(\hat{x}^k) \leq f(\hat{x}^k) + h(\hat{x}^k) = \psi(\hat{x}^k),$$

then (ii) holds.

(iii) Suppose that $\eta_{n+1} = \eta_n \geq \bar{\rho}$ and x^{n+1} is a null step, then we have $k(n+1) = k(n) = k$. By (2.6a), (2.6b), (2.6c) and $\eta_{n+1} = \eta_n$, for all $x \in \mathbf{R}^N$ and all $l \in I_{n+1}$, we have

$$\begin{aligned} \phi_{n+1}(x) &\geq \widetilde{\phi}_{n+1}(x) = \widetilde{f}_{n+1}(x) + \widetilde{\varphi}_{n+1}(x) \\ &\geq f(\hat{x}^k) - e_{f,l}^k + \langle g_f^l, x - \hat{x}^k \rangle + h(\hat{x}^k) - e_{h,l}^k \\ &\quad - \tfrac{\eta_n}{2}\|x^l - \hat{x}^k\|^2 + \langle g_h^l + \eta_n(x^l - \hat{x}^k), x - \hat{x}^k \rangle, \end{aligned} \tag{4.2}$$

where the inequality holds by (2.6a), (2.6b) and $\eta_{n+1} = \eta_n$. In particular, (4.2) also holds for all $l \in J_n^{act}$ or $l = -n$ in I_{n+1} by the assumption. For such indices and $\eta_{n+1} = \eta_n$, (2.14b) and (2.14a) imply that

$$f(\hat{x}^k) - e_{f,l}^k = \widetilde{f}_n(x^{n+1}) - \langle g_f^l, x^{n+1} - \hat{x}^k \rangle,$$

and

$$\varphi_n(\hat{x}^k) - e_{f,l}^k - \frac{\eta_n}{2}\|x^l - \hat{x}^k\|^2 = \widetilde{\varphi}_n(x^{n+1}) - \langle g_h^l + \eta_n(x^l - \hat{x}^k), x^{n+1} - \hat{x}^k \rangle.$$

With (4.2), for all $l \in J_n^{act}$ or $l = -n$, we have

$$\begin{aligned} \widetilde{\phi}_{n+1}(x) &\geq \widetilde{f}_n(x^{n+1}) + \widetilde{\varphi}_n(x^{n+1}) - \langle g_f^l + g_h^l + \eta_n(x^l - \hat{x}^k), x - x^{n+1} \rangle \\ &= \widetilde{\phi}_n(x^{n+1}) - \langle g_f^l + g_h^l + \eta_n(x^l - \hat{x}^k), x - x^{n+1} \rangle. \end{aligned}$$

For the case $\{-n\} \in I_{n+1}$, by Lemma 2.1, we have that item (iii) holds. Next, we show the case of $J_n^{act} \subseteq I_{n+1}$. Since QP subproblem (2.7) is strictly convex, then the solution x^{n+1} is unique. Meanwhile, there exists optimal nonnegative multiplier $(\overline{\alpha}_1, \overline{\alpha}_2) \in S_n \times S_n$ and they satisfy

$$\begin{cases} \overline{\alpha}_1^i > 0, \quad \text{for all } i \in J_{1;n}^{act}, \quad \overline{\alpha}_2^i > 0, \\ \text{for all } i \in J_{2;n}^{act}, \quad \sum_{i \in J_{1;n}^{act}} \overline{\alpha}_1^i = 1, \quad \text{and} \sum_{i \in J_{2;n}^{act}} \overline{\alpha}_2^i = 1, \\ \widetilde{\phi}_n(x^{n+1}) = \sum_{i \in J_{1,n}^{act}} \overline{\alpha}_i \left(f_i + \langle g_f^i, x^{n+1} - x^i \rangle \right) \\ + \sum_{i \in J_{2,n}^{act}} \overline{\alpha}_i \left[h(x^i) + \frac{\eta_n}{2} \|x^i - \hat{x}^k\|^2 + \langle g_\varphi^i, x^{n+1} - x^i \rangle \right], \\ \hat{G}^k = \mu_n(\hat{x}^k - x^{n+1}) = \sum_{i \in J_{1,n}^{act}} \overline{\alpha}_1^i g_f^i \\ + \sum_{i \in J_{2,n}^{act}} \overline{\alpha}_2^i g_\varphi^i \in \partial(f + \varphi_n)(x^{n+1}) = \partial f(x^{n+1}) + \partial \varphi_n(x^{n+1}). \end{cases}$$

It follows that

$$\widetilde{\phi}_n(x^{n+1}) = \sum_{i \in J_{1,n}^{act}} \overline{\alpha}_1^i f_i + \sum_{i \in J_{2,n}^{act}} \overline{\alpha}_2^i \left(h(x^i) + \frac{\eta_n}{2} \|x^i - \hat{x}^k\|^2 \right) + \left\langle \hat{G}^k, x^{n+1} - x^i \right\rangle.$$

Note that for each $x \in \mathbf{R}^N$ and $i \in I_{n+1}$, it holds that

$$\widetilde{f}_{n+1}(x) \geq f_i + \langle g_f^i, x - x^i \rangle,$$

and

$$\widetilde{\varphi}_{n+1}(x) \geq h(x^i) + \frac{\eta_{n+1}}{2} \|x^i - \hat{x}^k\|^2 + \langle g_h^i + \eta_{n+1}(x^i - \hat{x}^k), x - x^i \rangle.$$

As a result, writing the above inequalities for the convex sum of indices $i \in J_{1;n}^{act}$, $i \in J_{2;n}^{act}$ and using that $\eta_{n+1} = \eta_n$, we have

$$\begin{aligned} \widetilde{\phi}_{n+1}(x) &= \sum_{i \in J_{1;n}^{act}} \overline{\alpha}_1^i \widetilde{f}_{n+1}(x) + \sum_{i \in J_{2;n}^{act}} \overline{\alpha}_2^i \widetilde{\varphi}_{n+1}(x) \\ &\geq \sum_{i \in J_{1,n}^{act}} \overline{\alpha}_1^i f_i + \sum_{i \in J_{2,n}^{act}} \overline{\alpha}_2^i \left(h(x^i) + \frac{\eta_n}{2} \|x^i - \hat{x}^k\|^2 \right) \\ &\quad + \left\langle \sum_{i \in J_{1;n}^{act}} \overline{\alpha}_1^i g_f^i + \sum_{i \in J_{2;n}^{act}} \overline{\alpha}_2^i (g_h^i + \eta_n(x^i - \hat{x}^k)), x - x^i \right\rangle \\ &= \widetilde{\phi}_n(x^{n+1}) + \mu_n \langle \hat{x}^k - x^{n+1}, x - x^{n+1} \rangle. \end{aligned}$$

Then by the definition of J_n^{act}, we have that item (iii) holds for the case of $J_n^{act} \subseteq I_{n+1}$.
By (2.6a), (2.6b), (2.6c) and setting $i = n$, it is straightforward to item (iv). □

Lemma 4.2 *Consider the approximate model functions* (2.6c) *and Algorithm 1. Then there exists index $\tilde{n}$, such that for all $n \geq \tilde{n}$, the parameters η_n, μ_n and R_n stabilize:*

$$\tilde{\eta} \equiv \eta_n, \quad \tilde{\mu} \equiv \mu_n, \quad \textit{and} \quad \tilde{R} \equiv R_n = \tilde{\eta} + \tilde{\mu}.$$

Proof By Lemma 3.1, there is a finite restart steps in Algorithm 1. Once there are no more restart steps, according Algorithm 1, μ_n will be kept unchanged. Then there exists an index n_μ and from the index n_μ on, $\mu_n = \tilde{\mu}$ holds. In the algorithm, η_n is nondecreasing: either $\eta_{n+1} = \eta_n$ or $\eta_{n+1} = \Gamma_\eta \overline{\eta}_{n+1}$ where $\Gamma_\eta > 1$ and $\overline{\eta}_{n+1} > \eta_n$. Suppose that $\{\eta_n\}$ is not to stabilize, then there must be an infinite subsequence of iterations at which η_{n+1} is increased at least in a factor of Γ_η. But this leads to a contradiction. Since function h is lower-c^2 in the compact set $\mathcal{T}$, there exists a threshold $\bar{\rho}$ such that for any $\eta \geq \bar{\rho}$, the augment function $\varphi_n(\cdot) = h(\cdot) + \frac{\eta}{2}\|\cdot - \hat{x}^k\|$ is convex in the set $\mathcal{T}$, then the approximate function $\phi_n(x)$ is also convex in the set $\mathcal{T}$. Since the sequence $\{\eta_n\}$ is increased, then there exists an index n_η such that from that index on, η_n keeps unchanged,i.e, $\eta_{n_\eta + j} \equiv \eta_{n_\eta}$ for all $j \geq 0$. Let $\tilde{\eta}$ be $\eta_{\tilde{n}}$ and $\tilde{n} = \max\{n_\mu, n_\eta\}$, then the conclusion is followed. □

By Lemma 4.1 and Lemma 4.2, we have that our approximate model (2.6c) is reasonable, which satisfies the criteria in [19]. In addition, if $\tilde{\eta} \geq \bar{\rho}$, the approximate model (2.6c) is a lower approximation to function $\phi_n(x)$.

In order to verify the convergent properties, we set $TOL_{stop} = 0$. Note that if Algorithm 1 stops at some index n, this means that $\delta_{n+1} = 0$, i.e.,

$$f(\hat{x}^k) - \widetilde{f}_n(x^{n+1}) = 0, \quad \text{and} \quad h(\hat{x}^k) + \frac{\eta_n}{2}\|x^{n+1} - \hat{x}^k\|^2 - \widetilde{\varphi}_n(x^{n+1}) = 0. \tag{4.3}$$

By the above equalities, we have

$$\psi(\hat{x}^k) + \frac{\eta_n}{2}\|x^{n+1} - \hat{x}^k\|^2 = \widetilde{\phi}_n(x^{n+1}). \tag{4.4}$$

Since $x^{n+1} = p_{\mu_n}\widetilde{\phi}_n(\hat{x}^k)$, then it holds that

$$\widetilde{\phi}_n(x^{n+1}) + \frac{\mu_n}{2}\|x^{n+1} - \hat{x}^k\|^2 \leq \widetilde{\phi}_n(\hat{x}^k) + \frac{\mu_n}{2}\|\hat{x}^k - \hat{x}^k\|^2 = \widetilde{\phi}_n(\hat{x}^k).$$

Based on (4.4) and the above inequality, we have

$$\psi(\hat{x}^k) + \frac{\eta_n + \mu_n}{2}\|x^{n+1} - \hat{x}^k\|^2 \leq \widetilde{\phi}_n(\hat{x}^k).$$

By Lemma 4.1, it holds $\widetilde{\phi}_n(\hat{x}^k) \leq \psi(\hat{x}^k)$. By the above discussions, we have

$$\psi(\hat{x}^k) + \frac{\eta_n + \mu_n}{2}\|x^{n+1} - \hat{x}^k\|^2 \leq \psi(\hat{x}^k),$$

then it holds that $\hat{x}^k = x^{n+1} = p_{\mu_n}\widetilde{\phi}_n(\hat{x}^k)$. However, suppose that η_n largely enough to guarantee $h(\cdot) + \eta_n\|\cdot - \hat{x}^k\|^2/2$ is convex in $\mathcal{T}$. This implies that $\psi(\hat{x}^k) = \widetilde{\phi}_n(\hat{x}^k) = \widetilde{f}_n(\hat{x}^k) + \widetilde{\varphi}_n(\hat{x}^k)$ holds. By $\hat{x}^k = x^{n+1} = p_{\mu_n}\widetilde{\phi}_n(\hat{x}^k)$ and (2.6c), it always holds that

$$\begin{aligned}
\psi(\hat{x}^k) &= \widetilde{f}_n(\hat{x}^k) + \widetilde{\varphi}_n(\hat{x}^k) \leq \widetilde{f}_n(x) + \widetilde{\varphi}_n(x) + \frac{\mu_n}{2}\|x - \hat{x}^k\|^2, \quad \text{with } \forall x \in \mathbf{R}^N, \\
&\leq f(x) + h(x) + \frac{\eta_n}{2}\|x - \hat{x}^k\|^2 \\
&+ \frac{\mu_n}{2}\|x - \hat{x}^k\|^2, \quad \text{with } \forall x \in \mathcal{T}, \\
&\leq \psi(x) + \frac{\eta_n + \mu_n}{2}\|x - \hat{x}^k\|^2, \quad \text{with } \forall x \in \mathbf{R}^N.
\end{aligned} \tag{4.5}$$

Note that in the second inequality, we can also return to $x \in \mathbf{R}^N$ by the definition of $\mathcal{T}$, i.e., for any $x \notin \mathcal{T}$

$$\psi(\hat{x}^k) \leq \psi(\hat{x}^0) < \psi(\hat{x}^0) + M_0 < \psi(x) < \psi(x) + \frac{\eta_n + \mu_n}{2}\|x - \hat{x}^k\|^2.$$

Therefore, we have

$$\hat{x}^k = p_{\mu_n+\eta_n}\psi(\hat{x}^k).$$

This means that if Algorithm 1 stops in finite iterations, the current stability center $\hat{x}^k$ is a solution to the composite problem (1.1). In the following, we assume Algorithm 1 never stops.

Lemma 4.3 *Consider the sequence $\{x^n\}$ generated by Algorithm 1, then $\{\widetilde{\phi}_n(x^{n+1}) + \frac{\mu_n}{2}\|x^{n+1} - \hat{x}^k\|^2\}$ is eventually strictly increasing and convergent.*

Proof Taking index n large enough so that η_n and μ_n keep unchanged, i.e, $\tilde{\eta} = \eta_n$ and $\tilde{\mu} = \mu_n$. Consider the following function:

$$L_n(y) := \widetilde{\phi}_n(x^{n+1}) + \frac{\tilde{\mu}}{2}\|x^{n+1} - \hat{x}^k\|^2 + \frac{\tilde{\mu}}{2}\|y - x^{n+1}\|^2, \tag{4.6}$$

and set $s^{n+1} := \tilde{\mu}(\hat{x}^k - x^{n+1})$. By the optimal condition of QP subproblem (2.7), it holds that $s^{n+1} \in \partial\widetilde{\phi}_n(x^{n+1})$. Since x^{n+1} is the unique solution of QP problem (2.7), then we have

$$L_n(x^{n+1}) \le \widetilde{\phi}_n(\hat{x}^k) + \frac{\tilde{\mu}}{2}\|\hat{x}^k - \hat{x}^k\|^2 = \widetilde{\phi}_n(\hat{x}^k).$$

By stem (ii) in Lemma 4.1 and the above inequality, it holds that

$$L_n(x^{n+1}) \le \psi(\hat{x}^k), \quad \forall\ n = 0, 1, \cdots. \tag{4.7}$$

That means the function L_n is upper bounded. Considering (4.1) with $x = x^{n+2}$, we have

$$\widetilde{\phi}_{n+1}(x^{n+2}) \ge \widetilde{\phi}_n(x^{n+1}) + \tilde{\mu}\langle \hat{x}^k - x^{n+1},\ x^{n+2} - x^{n+1}\rangle.$$

By (4.6) with index $n+1$ and taking $y = x^{n+2}$, we have

$$\begin{aligned} L_{n+1}(x^{n+2}) &= \widetilde{\phi}_{n+1}(x^{n+2}) + \frac{\tilde{\mu}}{2}\|x^{n+2} - \hat{x}^k\|^2 + \frac{\tilde{\mu}}{2}\|x^{n+2} - x^{n+2}\|^2 \\ &= \widetilde{\phi}_{n+1}(x^{n+2}) + \frac{\tilde{\mu}}{2}\|x^{n+2} - \hat{x}^k\|^2. \end{aligned}$$

Hence, by the above discussions, it holds that

$$L_{n+1}(x^{n+2}) \ge L_n(x^{n+1}) + \langle s^{n+1},\ x^{n+2} - x^{n+1}\rangle + \frac{\tilde{\mu}}{2}\|x^{n+2} - \hat{x}^k\|^2 - \frac{\tilde{\mu}}{2}\|x^{n+1} - \hat{x}^k\|^2.$$

Note that the two rightmost terms in the above inequality satisfy the following relationship:

$$\begin{aligned} &\frac{\tilde{\mu}}{2}\|x^{n+2} - \hat{x}^k\|^2 - \frac{\tilde{\mu}}{2}\|x^{n+1} - \hat{x}^k\|^2 \\ &= \tilde{\mu}\langle \hat{x}^k - x^{n+1},\ x^{n+1} - x^{n+2}\rangle + \frac{\tilde{\mu}}{2}\|x^{n+2} - x^{n+1}\|^2 \\ &= \langle -s^{n+1}, x^{n+2} - x^{n+1}\rangle + \frac{\tilde{\mu}}{2}\|x^{n+2} - x^{n+1}\|^2. \end{aligned}$$

Therefore, it holds that

$$L_{n+1}(x^{n+2}) \ge L_n(x^{n+1}) + \frac{\tilde{\mu}}{2}\|x^{n+2} - x^{n+1}\|^2.$$

By the definition of L_n, it holds that

$$L_{n+1}(x^{n+2}) = \widetilde{\phi}_{n+1}(x^{n+1}) + \frac{\tilde{\mu}}{2}\|x^{n+2} - \hat{x}^k\|^2, \text{ and}$$

$$L_n(x^{n+1}) = \widetilde{\phi}_n(x^{n+1}) + \frac{\tilde{\mu}}{2}\|x^{n+1} - \hat{x}^k\|^2.$$

Hence, the sequence $\left\{\widetilde{\phi}_n(x^{n+1}) + \frac{\mu_n}{2}\|x^{n+1} - \hat{x}^k\|^2\right\}$ is strictly increasing for n large enough. By (4.7), it is also bounded above, so it converges. □

The convergent analysis in bundle methods is usually divided into two different cases depending on whether a finite or an infinite number of serious steps is generated. Now we give the convergent analyses of Algorithm 1.

Theorem 4.1 *Consider Algorithm 1 with $TOL_{stop} = 0$ and suppose there is no termination and $\tilde{\eta} > \bar{\rho}$ holds, then the following conclusions hold:*

(i) If only finite number of serious steps are generated in Algorithm 1 and the last serious step is denoted by $\hat{x}$, followed by an infinite sequence of null steps, then $x^{n+1} \to \hat{x}$, as $n \to +\infty$ and $\hat{x}$ is a stationary point of the composite problem (1.1).
(ii) If there is an infinite number of serious steps generated by Algorithm 1, then any accumulation point of the sequence $\{\hat{x}^k\}$ is a stationary point of the composite problem (1.1).

Proof (i) Consider the iterations n, which is generated after the last serious point $\hat{x}$. There are only null steps. By Lemma 4.3, we have that, as $n \to \infty$, the algorithm generates an infinite sequence $\{x^n\}$ converging to $\bar{p} := p_{\tilde{\eta}+\tilde{\mu}}\psi(\hat{x})$ with

$$\lim_{n\to\infty}\left(\widetilde{\phi}_n(x^{n+1}) + \frac{\mu_n}{2}\|x^{n+1} - \hat{x}\|^2\right) = \psi(\bar{p}) + \frac{\tilde{\eta}+\tilde{\mu}}{2}\|\bar{p} - \hat{x}\|^2.$$

By the definition of the predict descent and (3.1), it is followed as

$$\begin{aligned}
\delta_{n+1} &= \delta^f_{n+1} + \delta^\varphi_{n+1} = f(\hat{x}^k) - \widetilde{f}_n(x^{n+1}) + h(\hat{x}) \\
&\quad + \frac{\eta_n}{2}\|x^{n+1} - \hat{x}\|^2 - \widetilde{\varphi}_n(x^{n+1}) \\
&\to f(\hat{x}) + h(\hat{x}) + \frac{\tilde{\eta}}{2}\|\bar{p} - \hat{x}\|^2 - f(\bar{p}) - h(\bar{p}) \\
&\quad - \frac{\tilde{\eta}}{2}\|\bar{p} - \hat{x}\|^2, \quad \text{when } I_n \ni n \to \infty \\
&= \psi(\hat{x}) - \psi(\bar{p}).
\end{aligned}$$

Since x^{n+1} is a null step, then the serious step test is not satisfied, i.e.

$$\psi(x^{n+1}) > \psi(\hat{x}) - m\delta_{n+1},$$

Taking $n \to \infty$, it holds

$$\psi(\bar{p}) \geq \psi(\hat{x}) - m \lim_{n\to\infty} \delta_{n+1} = \psi(\hat{x}) - m(\psi(\hat{x}) - \psi(\bar{p})).$$

By the above inequality and $m \in (0, 1)$, we have

$$\psi(\bar{p}) \geq \psi(\hat{x}).$$

Meanwhile, $\bar{p} = p_{\tilde{R}}\psi(\hat{x})$ means that

$$\psi(\bar{p}) + \frac{\tilde{\eta}+\tilde{\mu}}{2}\|\bar{p} - \hat{x}\|^2 \leq \psi(\hat{x}) + \frac{\tilde{\eta}+\tilde{\mu}}{2}\|\hat{x} - \hat{x}\|^2 = \psi(\hat{x}).$$

So we have $\psi(\bar{p}) = \psi(\hat{x})$ and $\bar{p} = \hat{x}$. That is, $\hat{x} := p_{\tilde{\eta}+\tilde{\mu}}\psi(\hat{x})$, so $\hat{x}$ is a stationary point of $\psi(x)$.

(ii) By Lemma 3.1, the sequence $\{x^k\} \in \mathcal{T}$, where $\mathcal{T}$ is compact, has an accumulation point, that is, for some infinite set Λ, $\hat{x}^k \to \widetilde{x} \in \mathcal{T}$ as $\Lambda \ni k \to \infty$. Since the iteration $k+1$ is a serious step, then we have $\hat{x}^{k+1} = x^{i_{k+1}}$. To simply the notation, we set $j_k = i_{k+1} - 1$, then $\hat{x}^{k+1} = p_{\tilde{\mu}}\ \widetilde{\phi}_{j_k}(\hat{x}^k)$. The telescopic sum of descent test for the subsequence of serious steps implies that as $k \to \infty$, either $\psi(\hat{x}^k) \downarrow -\infty$ or $\delta_{i_{k+1}} \to 0$. Since ψ is the sum of two finite functions, then we have $\delta_{i_{k+1}} \to 0$.
Consider $k \in \Lambda$, since $\|\hat{x}^{k+1} - \hat{x}^k\|^2 \to 0$, $\hat{x}^{k+1}$ and $\hat{x}^k$ converge to $\widetilde{x}$ as $\Lambda \ni k \to \infty$, then $\widetilde{\phi}_{j_k}(\hat{x}^{k+1}) \to \psi(\widetilde{x})$ holds. But the conditions $\hat{x}^{k+1} = p_{\tilde{\mu}}\ \widetilde{\phi}_n \hat{x}^k)$ and $\tilde{\eta} > \bar{\rho}$ imply that for all $x \in \mathcal{L}$, we have

$$\widetilde{\phi}_{j_k}(\hat{x}^{k+1}) + \frac{\tilde{\mu}}{2}\|\hat{x}^{k+1} - \hat{x}^k\|^2 \leq \psi(x) + \frac{\tilde{\eta} + \tilde{\mu}}{2}\|x - \hat{x}^k\|^2.$$

Therefore, for $k \to +\infty$, we have

$$\psi(\widetilde{x}) \leq \psi(x) + \frac{\tilde{\eta} + \tilde{\mu}}{2}\|x - \widetilde{x}\|^2, \quad \text{for all } x \in \mathcal{T}.$$

As $\widetilde{x} \in \mathcal{T}$, and for any $x \notin \mathcal{T}$, it holds that

$$\psi(\widetilde{x}) \leq \psi(\hat{x}^0) + M_0 \leq \psi(x) \leq \psi(x) + \frac{\tilde{\eta} + \tilde{\mu}}{2}\|x - \widetilde{x}\|^2.$$

Hence, by the above discussions,we have

$$\psi(\widetilde{x}) \leq \psi(x) + \frac{\tilde{\eta} + \tilde{\mu}}{2}\|x - \widetilde{x}\|^2, \text{for all } x \in \mathbf{R}^N.$$

In other words, that means $\widetilde{x} = p_{\tilde{\eta}+\tilde{\mu}}\psi(x)$. Since $\tilde{\eta} > \bar{\rho}$, we have $0 \in \partial\psi(\widetilde{x})$. The conclusion holds. □

5 Numerical results

In this section, we focus on the numerical performance of Algorithm 1. The numerical experiments will be divided into two parts. Two polynomial functions which be seen in [24, 34, 35] are provided in the first subsection. In the second subsection, we devote to deal with some DC problems, which are taken from [36–38]. Meanwhile, we code Algorithm 1 in MATLAB R2016 and run it on a PC with 2.10 GHz CPU. The Quadratic programming solver for Algorithm 1 is Quadprog.m, which is available in the Optimization Toolbox in MATLAB.

5.1 Polynomial functions optimization

In this subsection, we firstly introduce two nonconvex polynomial problems developed in [24, 34, 35]. Concretely, for each $i = 1, \cdots, N$, let the function $g_i : \mathbf{R}^N \to \mathbf{R}$ be defined by

$$g_i(x) = (ix_i^2 - 2x_i - K) + \sum_{j=1}^{N} x_j, \quad \text{where } x \in \mathbf{R}^N, \text{ and } K \text{ is a constant.}$$

Fig. 1 The 3D image of $h(x)$

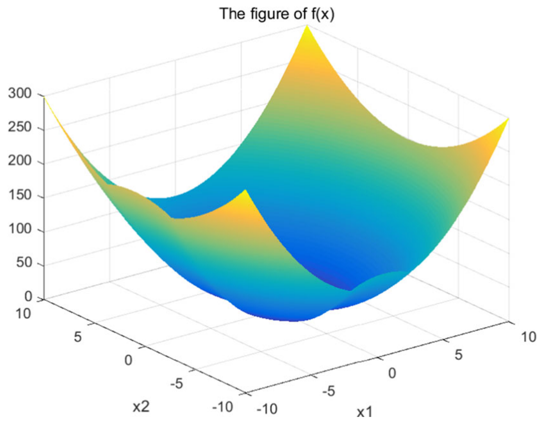
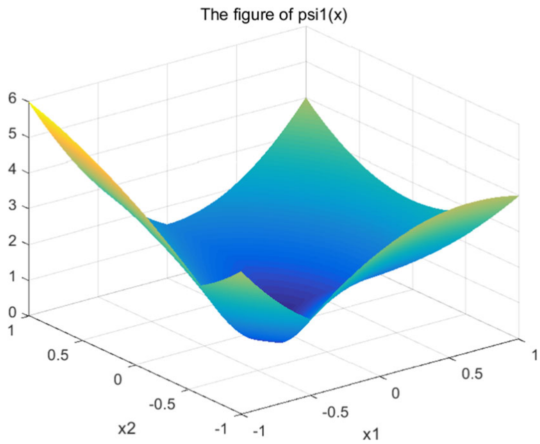


Fig. 2 The 3D image of $\psi_1(x)$



The two polynomial functions have the following forms:

$$\psi_1(x) := \sum_{i=1}^{N} |g_i(x)| + \frac{\|x\|^2}{2}, \quad \text{and} \quad \psi_2(x) := \sum_{i=1}^{N} |g_i(x)| + \frac{\|x\|}{2}.$$

The above two polynomial functions have some important properties (see in [24]): they are nonconvex and nonsmooth; they are globally lower-c^2, bounded below and level coercive; If $K = 0$, they have the same optimal value 0 and the same global minimization solution $\mathbf{0}$. If we denote $h(x) := \sum_{i=1}^{N} |g_i(x)|$ and $f(x) = \frac{1}{2}\|x\|^2$ or $f(x) = \frac{1}{2}\|x\|$, then the above functions satisfy the conditions in the problem (1.1). Figures 1, 2, 3 show that functions h, ψ_1 and ψ_2 are nonconvex with $K = 0$ and $N = 2$.

In the following, we first check the numerical behaviour of Algorithm 1 for $\psi_1(x)$ by comparing it with the the RedistProx method in [24]. Similar to that in [24], here we consider the cases of $K = 0$ and $N \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. The parameters referred in Algorithm 1 are that

$$x_0 = [1, 1, \cdots, 1]^T, \quad m = 0.01, \quad \Gamma_\eta = 2, \quad \Gamma_\mu = 2, \quad M_0 = 10,$$
$$R_0 = 10 \quad I_n = \{0, 1, \cdots, n\}.$$

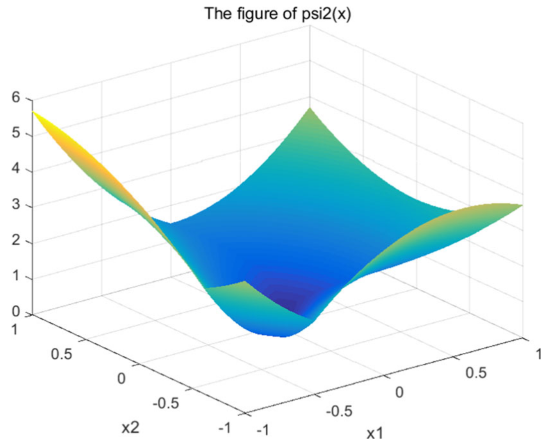


Fig. 3 The 3D image of $\psi_2(x)$

Table 1 The numerical results of Algorithm 1 and RedistProx algorithm for $\psi_1(x)$

Dim	Algorithm 1					RedistProx		
	Nu	Ns	Nf	fk	δ_k	Nf	fk	δ_k
1	2	25	28	3.0702e−09	9.4028e−08	5	0.5000	1.0000e−4
2	30	46	77	9.2727e−09	7.4453e−07	10	3.6623e−4	8.0019e−4
3	5	12	18	5.1480e−08	5.3128e−07	12	5.2922e−4	2.52121e−1
4	7	11	19	7.0451e−06	4.1736e−07	18	2.5722e−2	1.23599e−1
5	2	15	18	5.4569e−2	8.9098e−07	26	5.1166e−2	3.04311e−1
6	–	–	–	–	–	60	0.000000	2.0000e−5
7	43	29	73	1.2567e−07	8.2491e−07	34	2.39281e−1	2.57493e−1
8	127	40	168	1.0969e−07	8.2662e−07	56	6.9823e−2	1.54684e−1
9	22	31	54	3.7028e−07	9.0585e−07	150	0.000000	1.4000e−5
10	28	41	70	1.4399e−07	5.934e−07	61	2.17352e−1	1.71962e−1

The parameters for the RedistProx method in [24] are as default. Meanwhile, we also stop the method when the number of function evaluations is greater than 300, which can also be found in [24]. To compare the numerical performance of Algorithm 1, we adopt the computational results of the RedistProx algorithm in [24]. The numerical results are reported in Table 1 for $\psi_1(x)$. The columns of Tables 1 and 2 have the following meanings:

Dim: the the tested problem dimension. Nu: the number of null steps.
Ns: the number of serious steps. Nf: the number of oracle function evaluations used.
fk: the minimal function value found. δ_k: the value of δ at the final iteration.

From Table 1, we see that Algorithm 1 can not solve successfully $\psi_1(x)$ with $N = 6$. However, it can be solved by selecting $m = 0.05$ with the results $Nu = 202$, $Ns = 75$, $Nf = 278$, $fk = 1.69e-07$ and $\delta_k = 5.7577e-07$. Even under $m = 0.1$, the number of iteration decreases: $Nu = 77$, $Ns = 60$, $Nf = 138$, $fk = 1.8373e-07$ and $\delta_k = 6.0042e-07$. This denotes there exists different parameter values for variable dimensions. Although the number of iteration in Algorithm 1 is relatively a little bigger than that in the RedistProx method, the minimal function values in Algorithm 1 are more precise than that

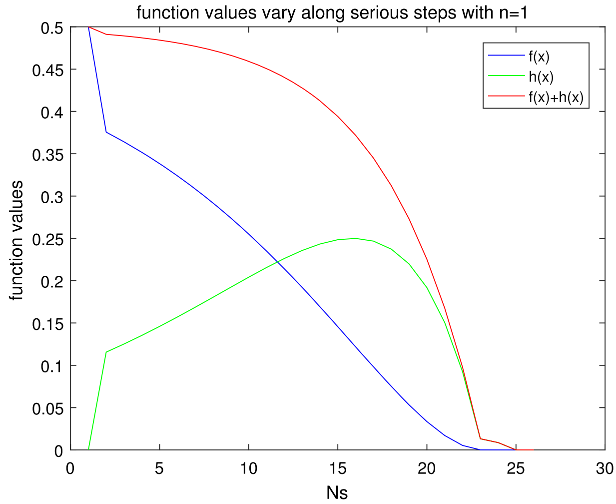


Fig. 4 The values of $\psi_1(x)$ and its components with $N = 1$

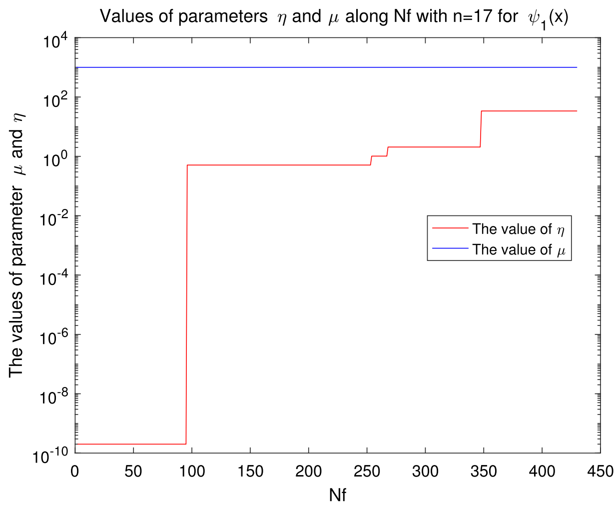


Fig. 5 The values of components of proximal parameter in function $\psi_1(x)$ with $N = 17$ along Nf iterations

in RedistProx method. The more precise minimal function values indicate that our model is more accurate. The descent condition (3.2) means the objective function values decrease along all serious steps. In fact, it is a weaker condition than that of its components decreasing along serious steps. Figure 4 illustrates that the objective function value decreases along all serious steps comparing one of its components increasing along some serious steps.

Next, we focus on the numerical performance of $\psi_2(x)$ and the results and the comparisons can be found in Table 2. The parameters are the same as above except R_0, here we select $R_0 = 100$. In Table 2, although function $\psi_2(x)$ with $N = 1$ does not be solved effectively by Algorithm 1, Algorithm 1 obtains more precise minimal function values except the case of $N = 1, 2, 3$. Hence, Algorithm 1 is more attractive.

Table 2 The numerical results of Algorithm 1 and RedistProx algorithm for $\psi_2(x)$

Dim	Algorithm 1					RedistProx		
	Nu	Ns	Nf	fk	δ_k	Nf	fk	δ_k
1	2	0	3	0.5000	2.8854e−09	13	0.0000	7.0000e−6
2	10	88	99	3.8463e−07	4.4659e−07	16	0.0000	0.0000
3	1	47	49	6.7233e−07	9.6252e−07	17	0.0000	2.0000e−6
4	6	38	45	5.4771e−07	7.5772e−07	23	1.9105e−2	1.28998e−1
5	7	38	46	1.4583e−07	9.266e−07	31	3.51332e−1	3.74674e−1
6	13	35	49	1.6685e−07	3.8116e−07	35	1.13835e−1	2.17732e−1
7	27	44	72	2.4414e−07	9.8858e−07	41	1.20009e−1	1.82087e−1
8	15	27	43	0.017576	7.6601e−07	41	1.084600	8.6662e−2
9	59	44	104	4.6022e−07	9.6337e−07	42	0.782526	0.238298
10	52	56	109	4.0989e−07	9.0711e−07	66	3.6327e−2	8.7404e−2

In the following, we consider higher dimension to Algorithm 1. The parameters for the cases of $N \in \{11, 12, 13, 14, 15, 16, 17, 18, 19\}$ are that: $m = 0.01$, $\Gamma_\eta = 2$, $\Gamma_\mu = 2$, $M_0 = 10$ and $R_0 = 1000$. For the stopping rules, TOL_{stop} here is 10^{-5} and the maximal function evaluations is 800. For $N \in \{20, 50, 80\}$, we take $TOL_{stop} = 10^{-4}$ and other parameters keep unchanged. For the case of $N = 100$, we take $TOL_{stop} = 10^{-4}$ and $M_0 = 100$ and the other parameters keep unchanged. The corresponding results are reported in Table 3. From Table 3, Algorithm 1 may solve all the problems successfully which indicate Algorithm 1 is effective. Figure 5 shows that the convexification parameter η and proximal parameter μ ultimately keep unchanged, which illustrates the conclusions of Lemma 3.1 and Lemma 4.2.

5.2 The results for some DC problems

In this subsection, twelve nonsmooth DC problems are firstly tested for the effectiveness of Algorithm 1. These functions can be expressed as $\psi(x) = f(x) - g(x)$ where functions f and g are convex. By taking $h(x) = -g(x)$, it can be converted into the form of the objective function in this paper. The testing problems are presented in Table 4 where Pr denotes the index of problems and ψ^* means the optimal value of the objective function.

The parameters in Algorithm 1 for the above DC problems are as follows: $m = 0.1$, $\Gamma_\eta = 2$, $\Gamma_\mu = 2$, $M_0 = 10$ and $R_0 = 10$. We also stop Algorithm 1 when the number of function evaluations is bigger than 1000. For parameter TOL_{stop}, we consider it for different choices. For $N \leq 10$, we take $TOL_{stop} = 10^{-5}$, otherwise, Meanwhile, we take $TOL_{stop} = 10^{-4}$. Tables 5 and 6 show the numerical results of Algorithm 1 to these problems. The columns Pr, N, δ_k and ψ_k denote the index of problems, the dimension of variable x, the value of δ_n at the finial iteration and the minimum function value found, respectively.

From Tables 5, 6, these problems can be successfully solved by Algorithm 1 that indicates Algorithm 1 is effective and reliable. Meanwhile, Fig. 6 shows that parameters η and μ ultimately keep unchanged, which also illustrates the conclusions of Lemma 3.1 and Lemma 4.2. Next, we compare the numerical results of Algorithm 1 with that of the PALNCO method in [8] for Problems 1,2 and 9. he PALNCO method was designed by the authors

Table 3 The numerical results of Algorithm 1 with higher dimensions for $\psi_1(x)$ and $\psi_2(x)$

Dim	Algorithm 1 for $\psi_1(x)$					Algorithm 1 for $\psi_2(x)$				
	Nu	Ns	Nf	fk	δ_k	Nu	Ns	Nf	fk	δ_k
11	28	175	204	1.3299e−06	9.9096e−06	6	170	177	4.0278e−06	7.1249e−06
12	11	173	185	4.0939e−06	9.4234e−06	14	168	183	3.3838e−06	9.2796e−06
13	38	166	205	8.2124e−07	7.3391e−06	34	161	196	2.7104e−06	7.9650e−06
14	22	167	190	1.9585e−06	3.3183e−06	16	160	177	1.9583e−06	6.9973e−06
15	53	177	231	1.2213e−06	9.5084e−06	6	167	174	3.8891e−02	7.7433e−06
16	43	166	210	2.0755e−06	5.0300e−06	44	167	212	3.1144e−06	6.5387e−06
17	37	178	216	2.8304e−06	8.5472e−06	14	170	185	2.9795e−06	7.4924e−06
18	70	188	259	2.6160e−06	9.8221e−06	44	150	195	3.6197e−06	5.0552e−06
19	201	187	389	2.6364e−06	6.7693e−06	17	173	191	2.0003e−02	9.0980e−06
20	14	168	183	4.4680e−02	7.7570e−06	6	142	149	9.3093e−02	9.8616e−05
50	136	315	452	3.4743e−03	9.4607e−05	69	266	336	3.4251e−05	9.7936e−05
80	100	290	391	8.1293e−02	8.5187e−05	199	349	549	1.5508e−01	8.8977e−05
100	254	291	546	9.9077e−02	8.7952e−05	263	281	545	2.8990e−01	8.2718e−05

Table 4 The testing problems

Pr	Problem	ψ^*	Pr	Problem	ψ^*	Pr	Problem	ψ^*
1	Problem 2 [36]	0	5	Problem 1 [36]	2	9	Problem 5 [36]	0
2	Problem 3 [36]	0	6	Problem 9 [36]	11/6	10	Problem 4 [37]	0
3	Problem 6 [36]	−2.5	7	Problem 10 [36]	−0.5	11	Problem 7 [37]	0
4	Problem 4 [38]	0	8	Problem 7 [36]	0.5	12	Problem 8 [37]	0

Table 5 The numerical results of Algorithm 1 for Problems 1-11

Pr	N	Nu	Ns	Nf	δ_k	ψ_k	ψ^*
1	2	3	200	204	7.4516e−07	7.4219e−10	0.0000
2	4	7	18	26	2.5316e−08	6.0247e−13	0.0000
3	2	0	235	236	7.8417e−06	−2.4998e+00	−2.5000
4	2	0	5	6	3.8330e−11	2.0000e−11	0.0000
4	5	1	20	22	1.4221e−07	1.4221e−07	0.0000
4	10	2	46	49	6.6541e−07	5.0980e−08	0.0000
4	20	9	96	106	6.2524e−06	5.7540e−06	0.0000
4	50	25	246	272	7.6744e−06	1.2201e−06	0.0000
4	80	38	396	435	9.1123e−06	6.8365e−06	0.0000
4	100	53	496	550	8.3967e−05	4.8405e−05	0.0000
5	2	1	21	23	9.5406e−06	2.0000e+00	2.0000
6	4	0	15	16	8.5727e−06	1.8333e+00	11/6
7	2	0	24	25	3.9179e−06	−4.9999e−01	−0.5000
8	2	8	30	39	8.0415e−06	5.0000e−01	0.5000
9	2	1	2	4	2.0943e−07	1.1275e−12	0.0000
9	5	3	7	11	2.3907e−06	2.3262e−06	0.0000
9	10	9	5	15	4.8611e−06	3.8969e−06	0.0000
9	50	8	16	25	9.8931e−05	1.6987e−03	0.0000
9	100	10	10	21	6.6163e−05	1.4956e−04	0.0000
9	150	11	29	41	7.8237e−05	8.1479e−04	0.0000
9	200	9	19	29	7.4887e−05	6.4618e−04	0.0000
9	300	9	20	30	7.8354e−05	1.0954e−03	0.0000

for minimizing the sum of two nonsmooth functions and proximal alternating linearization technique was referred. The parameters for Algorithm 1 keep unchanged except $R_0 = 20$ and the parameters for the PALNCO method are as default. The results are reported in Table 7 which shows that our algorithm is comparable to the PALNCO method.

In the following, we compare the numerical performance of Algorithm 1 with the PBDC method [36], the MPBNGC method [39], the nonsmooth DCA method [40, 41], the TCM method [38], the NCVX method [30] and the penalty NCVX method [31] in small dimension ($N < 10$). Here we take the data in [36]. For the nonsmooth DCA method [40, 41], we take the maximum number of the number of evaluation f and h as the number of function evaluation. The numerical results are presented in Tables 8 and 9. Meanwhile, the column with * denotes that the obtained objective function value is not optimal. From Tables 8 and 9, It is clear

Table 6 The numerical results of Algorithm 1 for Problem 2

Pr	N	Nu	Ns	Nf	δ_k	ψ_k	ψ^*
9	400	11	8	20	5.2230e−05	7.3640e−05	0.0000
9	500	12	14	27	4.4978e−05	6.8267e−05	0.0000
9	1000	13	10	24	7.7988e−05	2.0104e−04	0.0000
9	1500	13	9	23	6.3465e−05	1.6718e−04	0.0000
9	2000	14	8	23	9.1341e−05	2.1689e−04	0.0000
9	3000	14	8	23	7.8129e−05	1.2992e−04	0.0000
10	2	48	140	189	9.6668e−06	6.3182e−05	0.0000
10	4	2	74	77	7.8904e−06	5.5639e−05	0.0000
10	5	120	118	239	9.5363e−06	4.6750e−05	0.0000
10	10	4	139	144	4.2608e−05	7.8598e−04	0.0000
10	20	10	107	118	5.0623e−05	4.2504e−03	0.0000
10	35	3	244	248	5.3753e−05	6.6062e−03	0.0000
10	50	21	423	445	9.8859e−05	1.0707e−02	0.0000
11	2	4	2	7	2.1964e−13	1.7957e−14	0.0000
11	5	7	9	17	4.9187e−06	3.3561e−06	0.0000
11	10	10	13	24	8.1033e−06	4.2077e−06	0.0000
11	50	43	35	79	6.0951e−05	4.5213e−04	0.0000
11	100	22	51	74	8.7378e−05	3.4136e−04	0.0000
11	500	19	61	81	8.6205e−05	3.2987e−03	0.0000
11	1000	20	41	62	8.9359e−05	1.4876e−03	0.0000
11	3000	20	19	40	9.9680e−05	4.4435e−04	0.0000
12	2	4	3	8	9.9660e−09	2.8599e−12	0.0000
12	5	9	7	17	9.8805e−07	7.6711e−07	0.0000
12	10	10	20	31	8.5379e−06	2.1073e−06	0.0000
12	50	13	8	22	6.7411e−05	7.4966e−05	0.0000
12	100	16	12	29	9.6775e−05	1.5114e−04	0.0000
12	500	16	78	95	9.8179e−05	3.4981e−04	0.0000
12	1000	17	50	68	9.9081e−05	2.3764e−03	0.0000
12	3000	19	26	46	9.5126e−05	1.5896e−03	0.0000

Table 7 The numerical results of Algorithm 1 and PALNCO method

Pr	N	Algorithm 1					PALNCO			
		Nu	Ns	Nf	δ_k	fk	Nu	Ns	Nf	fk
1	2	2	200	203	7.4516e−07	7.4318e−10	6	130	137	2.609026e−07
4	10	2	90	93	8.2443e−06	8.2329e−06	80	157	238	4.869462e−06
9	10	3	139	143	4.2608e−05	7.8598e−04	84	129	214	1.638234e−04

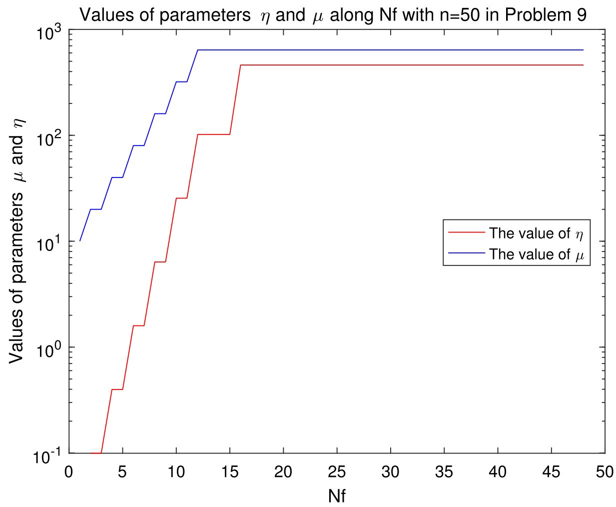


Fig. 6 The values of η and μ along Nf for Problem 9 with N=50

that Algorithm 1 can successfully solve all the above DC problems. Although Algorithm 1 is comparable with the PBDC method and the MPBNGC method, it is competitive to the nonsmooth DCA method, the NCVX method and the penalty NCVX method for the testing DC problems.

Based on continuous approximations to the Demyanov-Rubinov quasidifferential, the authors [37] proposed an effective method (called the DCQF method) for the unconstrained DC problems. In the following, we focus on the numerical performance of Algorithm 1 and compare it with the DCQF method. Similar to that in [37], we take $TOL_{stop} = 10^{-4}$ and keep other parameters unchanged. The numerical results can be found in Table 10. It is easy to see that our algorithm is much effective for problems 9,11-12 even in a higher dimension and the number of function values evaluated in our work is smaller that that in [37]. In order to compare the number of iteration intuitively, we adopt the method in [42] and present the performance profiles of iterations in Fig. 7 which indicates Algorithm 1 is more effective and reliable.

6 Conclusions

In this paper, we focus on a class of composite problem which is the sum of a finite convex function and a special nonconvex function. For the special nonconvex function (lower-c^2), we adopt the convexification technique. The corresponding parameter η_n is computed dynamically along iterations. We utilize the sum of the convex function and the augment function to approximate the primal problem. Meanwhile, the sum of the cutting plan models for the convex function and the augment function is regarded as the cutting plan model for the approximate function. A class of proximal bundle methods are designed. Under mild conditions, the convergence is obtained and the accumulation point of iterative points is a stationary point of primal problem. Two polynomial functions and twelve DC problems are referred in

Table 8 The numerical results of Algorithm 1, PBDC, MPBNGC and DCA methods

Pr	N	Algorithm 1		PBDC method		MPBNGC method		Nonsmooth DCA		ψ^*
		Nf	fk	Nf	fk	Nf	fk	Nf	fk	
1	2	204	7.4219e−10	21	1.1098e−12	39	2.3093e−14	139	1.0000*	0
2	4	26	6.0247e−13	25	2.2689e−12	112	2.2209e−11	439	3.8708e−09	0
3	2	236	−2.4998	22	−2.5000	53	−2.5000	78	−2.5000	−2.5
4	2	6	2.0001e−11	6	8.4377e−15	7	4.4409e−16	100	2.4337e−10	0
4	5	22	1.4221e−07	13	0.0000	30	3.5527e−15	325	2.8532e−10	0
5	2	23	2.0000	22	2.0000	17	2.0000	116	3.1638*	2
6	4	16	1.8333	85	1.8333	4	9.2000*	149	9.2000*	11/6
7	2	25	−0.5000	19	−0.5000	18	−0.5000	120	−0.5000	−0.5
8	2	39	0.5000	72	0.5000	27	1.0000*	6671	1.0000*	0.5
9	2	4	1.1275e−12	10	0.0000	5	8.8818e−16	70	1.6753e−09	0
9	5	11	2.3262e−06	15	6.1284e−14	25	3.6082e−15	943	1.5305e−10	0

Table 9 The numerical results of Algorithm 1, TCM, NCVX and penalty NCVX methods

Pr	N	Algorithm 1		TCM method		NCVX method		penalty NCVX		ψ^*
		Nf	fk	Nf	fk	Nf	fk	Nf	fk	
1	2	204	7.4219e−10	526	8.4453e−09	20	1.0000*	48	1.0000*	0
2	4	26	6.0247e−13	517	3.7104e−09	59	6.8183e−06	59	5.2088e−07	0
3	2	236	−2.4998	164	−2.5000	9	−2.5000	8	−2.5000	−2.5
4	2	6	2.0001e−11	125	4.0552e−11	19	1.2801e−08	4	1.2500e−07	0
4	5	22	1.4221e−07	399	3.6698e−10	25	2.5204e−05	87	7.1731e−08	0
5	2	23	2.0000	333	2.0000	20	2.0000	20	2.0000	2.0
6	4	16	1.8333	307	9.2000*	5	9.2000*	34	1.8333	11/6
7	2	25	−0.5000	270	−0.5000	4	−0.5000	3	−0.5000	−0.5
8	2	39	0.5000	482	0.5000	29	1.0000*	41	0.5000	0.5
9	2	4	1.1275e−12	149	1.9922e−10	19	1.1842e−08	24	7.5223e−09	0
9	5	11	2.3262e−06	1174	7.1262e−10	13	9.6573e−09	12	1.1546e−10	0

Table 10 The numerical results of Algorithm 1 and DCQF method

Pr	N	Algorithm 1		DCQF algorithm		Pr	N	Algorithm 1		DCQF algorithm	
		Iter	Nf	Iter	Nf			Iter	Nf	Iter	Nf
1	2	204	207	12	57	10	20	118	121	59	436
2	4	26	30	41	278	10	35	248	251	190	1667
4	5	22	23	26	128	10	50	445	449	236	2105
4	10	48	49	46	429	10	70	885	889	539	5799
4	20	106	107	141	1926	11	5	14	19	172	1381
4	35	189	190	293	5163	11	10	19	25	294	2495
4	50	272	273	434	7960	11	20	43	50	538	5312
5	2	18	20	11	82	11	35	37	44	843	8245
9	5	8	11	30	158	11	50	79	86	868	8414
9	10	14	18	44	305	11	70	68	76	1445	13, 117
9	20	32	37	41	399	12	5	16	23	36	180
9	35	19	25	28	263	12	10	26	34	62	399
9	50	25	32	36	350	12	20	29	38	76	602
9	70	24	31	49	805	12	35	60	70	131	618
10	5	96	98	21	122	12	50	22	32	109	667
10	10	144	147	20	123	12	70	58	69	76	402

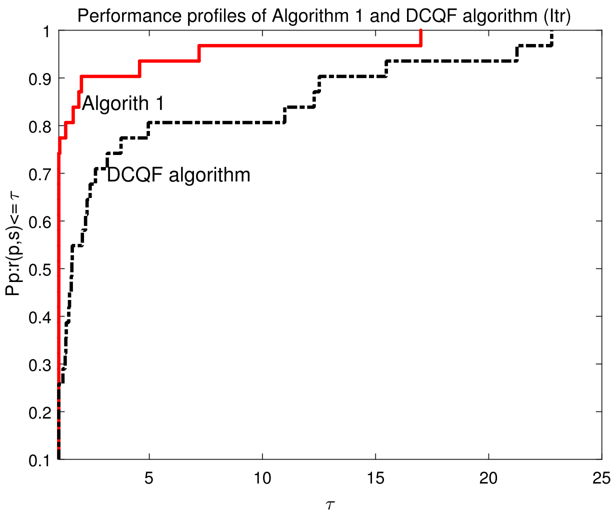


Fig. 7 Performance profiles of Algorithm 1 and DCQF algorithm in iterations

the numerical experiment. The numerical results show that our algorithm is interesting and attractive.

Data availability All data included in this study are available upon reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer-Verlag, Berlin (1998)
2. Horst, R., Thoai, N.V.: DC programming: overview. J. Optim. Theory Appl. **103**(1), 1–43 (1999)
3. Rockafellar, R.T.: Convex Analysis. Princeton University Press, New Jersey (1970)
4. Goldfarb, D., Ma, S., Scheinberg, K.: Fast alternating linearization methods for minimizing the sum of two convex functions. Math. Program. **141**, 349–382 (2013)
5. Jonathan, E., Dimitri, P. Bertsekas.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**(1–3), 293–318 (1992)
6. Tseng, P.: Further applications of a splitting algorithm to decomposition in variational inequalities and convex programming. Math. Program. **48**(1–3), 249–263 (1990)
7. Wen, Z., Goldfarb, D., Yin, W.: Alternating direction augmented Lagrangian methods for semidefinite programming. Math. Program. Comput. **2**(3–4), 203–230 (2010)
8. Li, D., Pang, L.P., Chen, S.: A proximal alternating linearization method for nonconvex optimization problems. Optim. Methods Softw. **29**(4), 771–785 (2014)
9. Kiwiel, K.C.: An alternating linearization bundle method for convex optimization and nonlinear multicommodity flow problems. Math. Program. **130**(1), 59–84 (2011)
10. Tang, C., Lv, J., Jian, J.: An alternating linearization bundle method for a class of nonconvex nonsmooth optimization problems. J. Inequal. Appl. **2018**, 101 (2018)
11. Clason, C., Valkonen, T.: Primal-dual extragradient methods for nonlinear nonsmooth PDE constrained optimization. SIAM J. Optim. **27**(3), 1314–1339 (2017)
12. Emiel, G., Sagastizábal, C.: Incremental-like bundle methods with application to energy planning. Comput. Optim. Appl. **46**(2), 305–332 (2010)
13. Cruz, J.Y.B.: On proximal subgradient splitting method for minimizing the sum of two nonsmooth convex functions. Set-Valued Var. Anal. **25**(2), 245–263 (2017)
14. Suzuki, T.: Dual averaging and proximal gradient descent for online alternating direction multiplier method. In: *International Conference on Machine Learning*, pp. 392–400 (2013)
15. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imag. Sci. **2**(1), 183–202 (2009)
16. Fukushima, M.: Application of the alternating direction method of multipliers to separable convex programming problems. Comput. Optim. Appl. **1**(1), 93–111 (1992)
17. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**(1), 1–122 (2011)
18. Tuy, H., Tam, B.T., Dan, N.D.: Minimizing the sum of a convex function and a specially structured nonconvex function. Optimization **28**(3–4), 237–248 (1994)
19. Hare, W., Sagastizábal, C.: Computing proximal points of nonconvex functions. Math. Program. **116**(1–2), 221–258 (2009)
20. Kiwiel, K.C.: A proximal bundle method with approximate subgradient linearizations. SIAM J. Optim. **16**(4), 1007–1023 (2006)
21. Hare, W.L., Poliquin, R.A.: Prox-regularity and stability of the proximal mapping. J. Convex Anal. **14**(3), 589–606 (2007)
22. Sagastizábal, C.: Composite proximal bundle method. Math. Program. **140**(1), 189–233 (2013)
23. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer Science and Business Media, Berlin (2013)
24. Sagastizábal, C., Hare, W.: A redistributed proximal bundle method for nonconvex optimization. SIAM J. Optim. **20**(5), 2442–2473 (2010)
25. Kiwiel, K.C.: Restricted step and Levenberg-Marquardt techniques in proximal bundle methods for nonconvex nondifferentiable optimization. SIAM J. Optim. **6**(1), 227–249 (1996)
26. Haarala, N., Miettinen, K., Mäkelä, M.M.: Globally convergent limited memory bundle method for large-scale nonsmooth optimization. Math. Program. **109**(1), 181–205 (2007)
27. Schramm, H., Zowe, J.: A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results. SIAM J. Optim. **2**(1), 121–152 (1992)
28. Lukšan, L., Vlček, J.: A bundle-Newton method for nonsmooth unconstrained minimization. Math. Program. **83**(1–3), 373–391 (1998)

29. Fuduli, A., Gaudioso, M., Giallombardo, G.: A DC piecewise affine model and a bundling technique in nonconvex nonsmooth minimization. Optim. Methods Softw. **19**(1), 89–102 (2004)
30. Fuduli, A., Gaudioso, M., Giallombardo, G.: Minimizing nonconvex nonsmooth functions via cutting planes and proximity control. SIAM J. Optim. **14**(3), 743–756 (2004)
31. Fuduli, A., Gaudioso, M., Nurminski, E.A.: A splitting bundle approach for non-smooth non-convex minimization. Optimization **64**(5), 1131–1151 (2015)
32. Kiwiel, K.C.: A method of centers with approximate subgradient linearizations for nonsmooth convex optimization. SIAM J. Optim. **18**(4), 1467–1489 (2008)
33. Bonnans, J.F., Gilbert, J.C., Lemaréchal, C., Sagastizábal, C.A.: Numerical Optimization: Theoretical and Practical Aspects. Springer-Verlag, Berlin (2006)
34. Ferrier, C.: Computation of the distance to semi-algebraic sets. Esaim Control Optim. Calc. Var. **5**(5), 139–156 (2000)
35. Lv, J., Pang, L.P., Meng, F.Y.: A proximal bundle method for constrained nonsmooth nonconvex optimization with inexact information. J. Global Optim. **70**, 517–549 (2018)
36. Joki, K., Bagirov, A.M., Karmitsa, N., Mäkelä, M.M.: A proximal bundle method for nonsmooth DC optimization utilizing nonconvex cutting planes. J. Global Optim. **68**(3), 501–535 (2017)
37. Bagirov, A.M.: A method for minimization of quasidifferentiable functions. Optim. Methods Softw. **17**(1), 31–60 (2002)
38. Bagirov, A.M., Ugon, J.: Codifferential method for minimizing nonsmooth DC functions. J. Global Optim. **50**(1), 3–22 (2011)
39. Mäkelä, M. M.: Multiobjective proximal bundle method for nonconvex nonsmooth optimization: Fortran subroutine MPBNGC 2.0. Rep. Dep. Math. Inf. Technol. Ser. B Sci. Comput. B **13**, 203 (2003)
40. Le, T.H.A., Dinh, T.P., Van Ngai, H.: Exact penalty and error bounds in DC programming. J. Glob. Optim. **52**(3), 509–535 (2012)
41. Tao, P.D.: The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. Ann. Oper. Res. **133**(1–4), 23–46 (2005)
42. Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Math. Program. **91**(2), 201–213 (2002)