Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH

Open Access

# An alternating linearization bundle method for a class of nonconvex nonsmooth optimization problems

Chunming Tang[1], Jinman Lv[1] and Jinbao Jian[2*]

*Correspondence:
jianjb@gxu.edu.cn
[2]College of Science, Guangxi University of Nationalities, Nanning, P.R. China
Full list of author information is available at the end of the article

**Abstract**

In this paper, we propose an alternating linearization bundle method for minimizing the sum of a nonconvex function and a convex function, both of which are not necessarily differentiable. The nonconvex function is first locally "convexified" by imposing a quadratic term, and then a cutting-planes model of the local convexification function is generated. The convex function is assumed to be "simple" in the sense that finding its proximal-like point is relatively easy. At each iteration, the method solves two subproblems in which the functions are alternately represented by the linearizations of the cutting-planes model and the convex objective function. It is proved that the sequence of iteration points converges to a stationary point. Numerical results show the good performance of the method.

**Keywords:** Bundle method; Alternating linearization; Local convexification; Global convergence

## 1 Introduction

In this paper, we consider the structured nonconvex minimization problem

$$\min_{x\in\mathbb{R}^n}\bigl\{F(x):=f(x)+h(x)\bigr\},\tag{1}$$

where $f:\mathbb{R}^n\to\mathbb{R}$ is possibly a nonconvex nonsmooth function and $h:\mathbb{R}^n\to(-\infty,\infty]$ is a closed proper convex function.

Problems of the form (1) often appear in practice, such as signal processing, image reconstruction, engineering, optimal control, and so on. Three typical examples are given below.

*Example* 1 (Unconstrained transformation of a constrained problem) Consider the constrained problem

$$\min\bigl\{f(x):x\in C\bigr\},\tag{2}$$

where $f$ is possibly a nonsmooth nonconvex function and $C$ is a convex subset of $\mathbb{R}^n$. Problem (2) can be written equivalently as

$$\min_{x\in\mathbb{R}^n} f(x) + \iota_C(x), \tag{3}$$

where $\iota_C$ is the indicator function of $C$, i.e., $\iota_C(x)$ equals 0 on $C$ and infinity elsewhere. Clearly, problem (3) is a special case of problem (1) with $h(x) = \iota_C(x)$. We note that the proximal point of $\iota_C$ can easily be calculated or even has a closed-form solution if $C$ has some special structure.

*Example* 2 (Nonconvex regularization of a convex function) Consider the $l_q$ $(0 < q < 1)$ regularization problem

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_q, \tag{4}$$

which has many practical applications in compressed sensing and imaging science (see e.g., [1]), where $\|x\|_q = (\sum_{i=1}^n |x_i|^q)^{1/q}$. The objective function of problem (4) is also the sum of a convex function and a nonconvex function.

*Example* 3 (Convex regularization of a nonconvex function) Hare et al. [2] studied the function of the form

$$F(x) = \sum_{i=1}^n \bigl|f_i(x)\bigr| + \frac{1}{2}\|x\|^2, \tag{5}$$

where $f_i(x) : \mathbb{R}^n \to \mathbb{R}, i = 1, \dots, n$ are Ferrier polynomials defined as

$$f_i(x) = \bigl(ix_i^2 - 2x_i\bigr) + \sum_{j=1}^n x_j.$$

It is well known that $f(x) = \sum_{i=1}^n |f_i(x)|$ is a nonconvex nonsmooth function, and $h(x) = \frac{1}{2}\|x\|^2$ is a simple convex function.

The methods for minimizing the sum of two functions have been well studied during the past several decades. Different methods are developed based on these two types of functions; see e.g., [3–12]. In particular, Kiwiel [9] proposed an alternating linearization bundle method for the sum of two convex functions and one of them is "simple" (i.e., minimizing this function plus a separable convex quadratic function is "easy"). Goldfarb et al. [7] proposed a fast alternating linearization methods for the sum of two convex functions both of which are "simple". Li et al. [10] presented a proximal alternating linearization method for the sum of two nonconvex functions based on the assumption that the proximal point of the two functions at a given point can easily be calculated. Attouch et al. [3] and Bolte et al. [4] considered a broad class of nonconvex and nonsmooth minimization problems that include as a special case minimizing the sum of two nonconvex functions, in which the proximal alternating minimization technique is used and the Kurdyka–Łojasiewicz property is assumed.

In this paper, we consider to minimize the sum of a nonconvex function and a convex function with the form of (1). In particular, we assume that $f$ is lower-$C^2$ and $h$ is "simple" in the sense that minimizing $h$ plus a quadratic term is relatively easy. The method presented in this paper can be viewed as a generalized version of the methods given in [9] and [13]. On one hand, we generalize the method of [9] from minimizing the sum of two convex functions to the sum of a nonconvex function and a convex function. On the other hand, we generalize the method of [13] from minimizing a single nonconvex nonsmooth function to the sum of two functions.

Our method will produce three sequences of points: $\{z^\ell\}$, $\{y^\ell\}$ and $\{x^{k(\ell)}\}$, where $\{z^\ell\}$ is the sequence of proximal points, $\{y^\ell\}$ is the sequence of trial points, and $\{x^{k(\ell)}\}$ is the sequence of stability centers (i.e., $x^{k(\ell)} \in \{y^\ell\}$ is the "best" point obtained so far for iteration $\ell$, which will be abbreviated as $x^k$ if there is no confusion). More precisely, our method will alternately solve the following two subproblems:

$$z^{\ell+1} := \arg\min\left\{\check{\varphi}_\ell(\cdot) + \bar{h}_{\ell-1}(\cdot) + \frac{1}{2}\mu_\ell\left\|\cdot - x^k\right\|^2\right\}, \tag{6}$$

$$y^{\ell+1} := \arg\min\left\{\bar{\varphi}_\ell(\cdot) + h(\cdot) + \frac{1}{2}\mu_\ell\left\|\cdot - x^k\right\|^2\right\}, \tag{7}$$

where $\check{\varphi}_\ell$ is a cutting-planes model [14, 15] of the local convexification function of $f$ at iteration $\ell$, which is based on the idea of the redistributed proximal bundle method in [13] and will be made more precise later; $\bar{h}_{\ell-1}$ is a linearization of $h$ at iteration $\ell - 1$; $\bar{\varphi}_\ell$ is a linearization of $\check{\varphi}_\ell$; $\mu_\ell$ is the proximal parameter. Our convergence analysis shows that, under suitable assumptions, any accumulation point of the sequence $\{x^k\}$ is a stationary point of $F$ if there is an infinite number of serious steps; otherwise, the last stability center is a stationary point of $F$.

This paper is organized as follows. In Sect. 2, we review some basic definitions and results required for this work. In Sect. 3, we present the alternating linearization bundle method for problem (1). Section 4 examines the convergence properties of the algorithm. Some preliminary numerical results are given in Sect. 5. The Euclidean inner product in $\mathbb{R}^n$ is denoted by $\langle x, y\rangle = x^T y$, and the associated norm by $\|\cdot\|$.

## 2 Preliminaries

In this section, we recall some basic definitions and results that are closely relevant to our method, which can be found in [13, 16, 17].

– The *limiting subdifferential* of $f$ at $\bar{x}$ is defined by

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

where $\hat{\partial} f(\bar{x})$ is the *regular subdifferential* defined by

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

An element $g \in \partial f(\bar{x})$ is called a *subgradient* of $f$ at $\bar{x}$.

– The function $f$ is *prox-bounded* if there exists $R \geq 0$ such that the function $f(\cdot) + \frac{1}{2}R\|\cdot\|^2$ is bounded below. The corresponding threshold is the smallest $r_{pb} \geq 0$ such that $f(\cdot) + \frac{1}{2}R\|\cdot\|^2$ is bounded below for all $R > r_{pb}$.
– The function $f$ is *lower-$C^2$* on an open set $V$ if for each $\bar{x} \in V$ there is a neighborhood $V'$ of $\bar{x}$ upon which a representation $f(x) = \max_{t\in T} f_t(x)$ holds, where $T$ is a compact set and the functions $f_t$ are of class $C^2$ on $V$ such that $f_t$, $\triangledown f_t$ and $\triangledown^2 f_t$ depend continuously on $(t,x) \in T \times V$.
– The *proximal point mapping* of the function $f$ at the point $x \in \mathbb{R}^n$ is defined by

$$p_R f(x) := \arg\min_{\omega\in\mathbb{R}^n}\left\{f(\omega) + \frac{1}{2}R\|\omega - x\|^2\right\}.$$

**Lemma 1** ([16]) *Suppose that the function $f$ is lower-$C^2$ on $V$ and $\bar{x} \in V$. Then there exist $\varepsilon > 0, K > 0$, and $\rho > 0$ such that*
(i) *for any point $x^0$ and parameter $R \geq \rho$ the function $f + \frac{1}{2}R\|\cdot - x^0\|^2$ is convex and finite valued on the closed ball $\bar{B}_\varepsilon(\bar{x})$, and*
(ii) *the function $f$ is Lipschitz continuous with constant $K$ on $\bar{B}_\varepsilon(\bar{x})$.*

**Theorem 1** ([16]) *Suppose that the lower semicontinuous function $f$ is prox-bounded with threshold $r_{pb}$ and lower-$C^2$ on $V$. Let $\bar{x} \in V$ and let $\varepsilon > 0, K > 0$ and $\rho > 0$ be given by Lemma* 1. *Then $\bar{x}$ is a stationary point of $f$ if and only if $\bar{x} = p_R f(\bar{x})$ for any $R > R_{\bar{x}} := \max\{4K/\varepsilon, \rho, r_{pb}\}$.*

**Assumption 1** ([13]) Given $x^0 \in \mathbb{R}^n$ and $M_0 \geq 0$, there exist an open bounded set $\mathcal{O}$ and a function $H$ such that $\mathcal{L}_0 := \{x \in \mathbb{R}^n : f(x) \leq f(x^0) + M_0\} \subset \mathcal{O}$, and $H$ is lower-$C^2$ on $\mathcal{O}$ with $H \equiv f$ on $\mathcal{L}_0$.

**Theorem 2** ([13]) *For a function $f$ satisfying Assumption* 1, *the following results hold*:
(i) *The level set $\mathcal{L}_0$ is nonempty and compact.*
(ii) *There exists $\rho^{\mathrm{id}} > 0$ such that, for any $\rho \geq \rho^{\mathrm{id}}$ and any given $y \in \mathcal{L}_0$, the function $f + \frac{1}{2}\rho\|\cdot - y\|^2$ is convex on $\mathcal{L}_0$.*
(iii) *The function $f$ is Lipschitz continuous on $\mathcal{L}_0$.*

## 3 The alternating linearization bundle method

### 3.1 Motivation and framework

The classic proximal point algorithm (see e.g. [18]) for solving problem (1) generates the new iterate by

$$y^{\ell+1} = \arg\min\left\{f(\cdot) + h(\cdot) + \frac{1}{2}R_\ell\left\|\cdot - y^\ell\right\|^2\right\}, \tag{8}$$

where $R_\ell > 0$ is the proximal parameter.

However, since $f$ is a nonconvex function, solving problem (8) may not be easy and is usually as difficult as the original problem (1). Therefore, we will tackle the difficulty via the following three steps.

1. *Generate the local convexification function of the nonconvex function $f$*. Following the redistribution idea of [13], we split the prox-parameter $R_\ell$ into two dynamic parameters

$\eta_\ell$ and $\mu_\ell$ which are nonnegative and satisfy $R_\ell = \eta_\ell + \mu_\ell$. Then the problem (8) (after replacing $y^\ell$ by $x^k$) can be written as

$$y^{\ell+1} = \arg\min\left\{\varphi_\ell(\cdot) + h(\cdot) + \frac{1}{2}\mu_\ell\left\|\cdot - x^k\right\|^2\right\}, \tag{9}$$

where

$$\varphi_\ell(\cdot) = f(\cdot) + \frac{1}{2}\eta_\ell\left\|\cdot - x^k\right\|^2 \tag{10}$$

is called the local convexification function of $f$, since it is convex whenever $\eta_\ell$ is large enough (see Theorem 2).

2. *Generate the cutting-planes model of* $\varphi_\ell$. Let $\ell$ be the current iteration index, $y^i$, $i \in J_\ell \subseteq \{0, 1, \ldots, \ell\}$ be trial points generated in the previous iterations, and $g_f^i \in \partial f(y^i)$. Define the cutting-planes model of $\varphi_\ell$ by

$$\check{\varphi}_\ell(\cdot) = \max_{i\in J_\ell}\left\{f\left(y^i\right) + \frac{1}{2}\eta_\ell\left\|y^i - x^k\right\|^2 + \left\langle g_f^i + \eta_\ell\left(y^i - x^k\right), \cdot - y^i\right\rangle\right\}, \tag{11}$$

where $g_f^i = g_f(y^i) \in \partial f(y^i)$. Therefore, we obtain an approximate version of problem (9) as follows:

$$y^{\ell+1} := \arg\min\left\{\check{\varphi}_\ell(\cdot) + h(\cdot) + \frac{1}{2}\mu_\ell\left\|\cdot - x^k\right\|^2\right\}. \tag{12}$$

3. *Apply the alternating linearization bundle strategy to solve problem* (12). Since problem (12) may still be difficult, motivating by the idea of the alternating linearization bundle [9], we consider to alternately solve the following two subproblems:

$$z^{\ell+1} := \arg\min\left\{\check{\varphi}_\ell(\cdot) + \bar{h}_{\ell-1}(\cdot) + \frac{1}{2}\mu_\ell\left\|\cdot - x^k\right\|^2\right\}, \tag{13}$$

$$y^{\ell+1} := \arg\min\left\{\bar{\varphi}_\ell(\cdot) + h(\cdot) + \frac{1}{2}\mu_\ell\left\|\cdot - x^k\right\|^2\right\}. \tag{14}$$

The above two subproblems are much easier to solve, whose objective functions are alternately represented by linear models of $h(\cdot)$ and $\check{\varphi}_\ell(\cdot)$, respectively.

### 3.2 Further description via bundle terminologies

Bundle methods [19–21] are among the most robust and reliable methods to solve general nonsmooth optimization problems, which can be considered stabilized variants of cutting-planes method [14, 15]. In general, for a convex function $h$, bundle methods store the trial points $y^i$, $i \in J_\ell$ with their function values and subgradients in a bundle of information:

$$\bigcup_{i\in J_\ell}\left\{\left(y^i, h\left(y^i\right), g_h^i \in \partial h\left(y^i\right)\right)\right\}, \tag{15}$$

and a point $x^k := x^{k(\ell)}$ (called *stability center*) which is the "best" point obtained so far. A storage-saving form of (15) (refer to the current stability center $x^k$) is given by

$$\bigcup_{i\in J_\ell}\bigl\{\bigl(e_h^{i,k}, g_h^i \in \partial_{e_h^{i,k}} h\bigl(x^k\bigr)\bigr)\bigr\},$$

where $\partial_e h$ is the $e$-subdifferential of $h$ in convex analysis, and $e_h^{i,k}$ are the linearization errors of $h$ defined by

$$e_h^{i,k} = h\bigl(x^k\bigr) - \bigl(h\bigl(y^i\bigr) + \bigl\langle g_h^i, x^k - y^i\bigr\rangle\bigr). \tag{16}$$

Following the notations above, the bundle information of the function $\varphi_\ell(\cdot)$ can be written as (see also [13]):

$$\bigcup_{i\in J_\ell}\bigl\{\bigl(e_f^{i,k}, d_i^k, \Delta_i^k, g_f^i\bigr)\bigr\} \quad \text{with} \quad \begin{cases} e_f^{i,k} = f(x^k) - (f(y^i) + \langle g_f^i, x^k - y^i\rangle), \\ d_i^k = \frac{1}{2}\|y^i - x^k\|^2, \\ g_f^i \in \partial f(y^i), \\ \Delta_i^k = y^i - x^k, \end{cases} \tag{17}$$

where $e_f^{i,k}$ and $d_i^k$ are the linearization errors of $f$ and $\frac{1}{2}\|\cdot - x^k\|^2$, respectively, $g_f^i$ is a subgradient of $f$ at $y^i$, and $\Delta_i^k$ is the gradient of $\frac{1}{2}\|\cdot - x^k\|^2$ at $y^i$. From (17), we know that $e_f^{i,k}$, $d_i^k$ and $\Delta_i^k$ depend on the point $x^k$, so they should be updated whenever a new stability center is generated (details are given below).

By the optimality conditions of subproblem (13), there exists a multiplier vector $(\alpha_i^\ell, i \in J_\ell) \in S^\ell$ such that

$$z^{\ell+1} = x^k - \frac{1}{\mu_\ell}\Biggl(\sum_{i\in J_\ell}\alpha_i^\ell\bigl(g_f^i + \eta_\ell\Delta_i^k\bigr) + g_h^{\ell-1}\Biggr), \tag{18}$$

where $S^\ell$ denotes the unit simplex in $\mathbb{R}^{|J_\ell|}$ and $g_h^{\ell-1} = \nabla\bar{h}_{\ell-1}(z^{\ell+1})$.

As iterations go along, the number of elements in the bundle may increase infinitely, which could lead to serious problems with storage and computation. The subgradient aggregation strategy [22] is the most popular and efficient way to overcome such a difficulty. We use the notation $g_{\eta_\ell}^{-\ell}$ to denote the aggregate subgradient, i.e.,

$$g_{\eta_\ell}^{-\ell} := \sum_{i\in J_\ell}\alpha_i^\ell\bigl(g_f^i + \eta_\ell\Delta_i^k\bigr) \in \partial\check{\varphi}_\ell\bigl(z^{\ell+1}\bigr). \tag{19}$$

Define the strongly active set of subgradients by

$$J_\ell^{\text{act}} := \bigl\{i \in J_\ell : \alpha_i^\ell > 0\bigr\}.$$

Then the corresponding aggregate bundle elements are given by

$$\bigl(e_f^{-\ell}, d_{-\ell}^k, \Delta_{-\ell}^k, g_f^{-\ell}\bigr) := \sum_{i\in J_\ell}\alpha_i^\ell\bigl(e_f^{i,k}, d_i^k, \Delta_i^k, g_f^i\bigr) = \sum_{j\in J_\ell^{\text{act}}}\alpha_j^\ell\bigl(e_f^{j,k}, d_j^k, \Delta_j^k, g_f^j\bigr). \tag{20}$$

Therefore

$$g_{\eta_\ell}^{-\ell} = \sum_{i\in J_\ell} \alpha_i^\ell \big(g_f^i + \eta_\ell \Delta_i^k\big) = g_f^{-\ell} + \eta_\ell \Delta_{-\ell}^k = \mu_\ell \big(x^k - z^{\ell+1}\big) - g_h^{\ell-1}.$$

Here, as in [13, 23], we use negative index $-\ell$ to express the aggregate bundle elements, hence $J_\ell \subseteq \{-\ell, -\ell+1, \ldots, 0, 1, \ldots, \ell-1, \ell\}$ in general.

By making use of the notations above, the cutting-planes model $\check{\varphi}_\ell$ in (11) can be rewritten as:

$$\check{\varphi}_\ell(\cdot) = f\big(x^k\big) + \max_{i\in J_\ell}\big\{-\big(e_f^{i,k} + \eta_\ell d_i^k\big) + \big\langle g_f^i + \eta_\ell \Delta_i^k, \cdot - x^k\big\rangle\big\}. \tag{21}$$

Note that, for all $j \in J_\ell^{\mathrm{act}}$ we have

$$\check{\varphi}_\ell\big(z^{\ell+1}\big) = f\big(x^k\big) - e_f^{j,k} - \eta_\ell d_j^k + \big\langle g_f^j + \eta_\ell \Delta_j^k, z^{\ell+1} - x^k\big\rangle, \tag{22}$$

and the aggregate model of $\check{\varphi}_\ell$ in $J_\ell$ is

$$\tilde{\varphi}_{-\ell}(\cdot) = f\big(x^k\big) - e_f^{-\ell,k} - \eta_\ell d_{-\ell}^k + \big\langle g_f^{-\ell} + \eta_\ell \Delta_{-\ell}^k, \cdot - x^k\big\rangle.$$

For a new stability center $x^{k+1}$, the bundle elements can be updated by (see [13])

$$\begin{aligned}
e_f^{i,k+1} &= e_f^{i,k} + f\big(x^{k+1}\big) - \big(f\big(x^k\big) + \big\langle g_f^i, x^{k+1} - x^k\big\rangle\big),\\
d_i^{k+1} &= d_i^k + \frac{1}{2}\big\|x^{k+1} - x^k\big\|^2 + \big\langle \Delta_i^k, x^{k+1} - x^k\big\rangle,\\
\Delta_i^{k+1} &= \Delta_i^k + x^k - x^{k+1}.
\end{aligned} \tag{23}$$

On the other hand, since $f$ is possibly nonconvex, the linearization errors $e_f^{i,k} + \eta_\ell d_i^k, i \in J_\ell$ may be negative, and therefore the model $\check{\varphi}_\ell$ is not necessarily a lower approximation to $\varphi_\ell$. In the case of $e_f^{i,k} + \eta_\ell d_i^k \geq 0$, one has

$$g_f^i + \eta_\ell \Delta_i^k \in \partial_{e_f^{i,k} + \eta_\ell d_i^k} \check{\varphi}_\ell\big(x^k\big). \tag{24}$$

In order to ensure that the linearization errors are all nonnegative, the convexification parameter $\eta_\ell$ should be adjusted to asymptotically estimate the ideal convexity threshold $\rho^{\mathrm{id}}$ in Theorem 2. Hare et al. [2] suggested a lower bound for $\eta_\ell$ as follows:

$$\eta_\ell^{\min} := \max_{i\in J_\ell, d_i^k > 0} -\frac{e_f^{i,k}}{d_i^k}, \tag{25}$$

which guarantees that $e_f^{i,k} + \eta d_i^k \geq 0$ for all $i \in J_\ell$ whenever $\eta \geq \eta_\ell^{\min}$.

Finally, in our algorithm, we define the predicted descent $\delta_\ell$ and the linearization error $\varepsilon_\ell$ as follows:

$$\delta_\ell := f\big(x^k\big) + \frac{1}{2}\eta_\ell \big\|y^{\ell+1} - x^k\big\|^2 + h\big(x^k\big) - \big[\bar{\varphi}_\ell\big(y^{\ell+1}\big) + h\big(y^{\ell+1}\big)\big], \tag{26}$$

$$\varepsilon_\ell := F(x^k) - \left[\bar{\varphi}_\ell(x^k) + \bar{h}_\ell(x^k)\right]. \tag{27}$$

For a fixed parameter $\kappa \in (0,1)$, a *descent step* is taken if

$$F(y^{\ell+1}) \leq F(x^k) - \kappa\delta_\ell, \tag{28}$$

holds, and then update the stability center $x^{k+1} = y^{\ell+1}$. Otherwise, a *null step* occurs, and then the aggregate linearization and the new linearization are used to produce a better model $\check{\varphi}_{\ell+1}$.

### 3.3 The algorithm

**Algorithm 1**

*Step* 0. (*Initialization*). Select a starting point $y^0$ and set $x^0 = y^0$. Set parameters $M > 0$, $R_0 > 0$, $\kappa \in (0,1)$, $\epsilon \geq 0$, and $\Gamma \geq 1$. Initialize the iteration counter $\ell = 0$, the descent step counter $k := k(\ell) = 0$ with $i_0 = 0$. Set $(\mu_0, \eta_0) = (R_0, 0)$ and $J_0 := \{0\}$. Compute $f(x^0), g_f^0 \in \partial f(x^0)$ and the bundle information $(e_f^{0,0}, d_0^0, \triangle_0^0) := (0,0,0)$. Set $s_h^{-1} = g_h^0 \in \partial h(x^0)$.

*Step* 1. Find $z^{\ell+1}$ by solving subproblem (13), and set

$$\bar{\varphi}_\ell(\cdot) = \check{\varphi}_\ell(z^{\ell+1}) + \langle s_\varphi^\ell, \cdot - z^{\ell+1}\rangle \quad \textit{with } s_\varphi^\ell = \mu_\ell(x^k - z^{\ell+1}) - s_h^{\ell-1}. \tag{29}$$

*Step* 2. Find $y^{\ell+1}$ by solving subproblem (14), and set

$$\bar{h}_\ell(\cdot) = h(y^{\ell+1}) + \langle s_h^\ell, \cdot - y^{\ell+1}\rangle \quad \textit{with } s_h^\ell = \mu_\ell(x^k - y^{\ell+1}) - s_\varphi^\ell. \tag{30}$$

*Step* 3. (*Stopping criterion*). Compute $f(y^{\ell+1})$, $h(y^{\ell+1})$, $g_f^{\ell+1} \in \partial f(y^{\ell+1})$ and $g_h^{\ell+1} \in \partial h(y^{\ell+1})$. If $\delta_\ell \leq \epsilon$, then STOP. Otherwise, compute the new bundle elements by

$$\Delta_{\ell+1}^k := y^{\ell+1} - x^k, \qquad d_{\ell+1}^k := \left\|\Delta_{\ell+1}^k\right\|^2/2,$$
$$e_f^{\ell+1,k} := f(x^k) - \left(f(y^{\ell+1}) + \langle g_f^{\ell+1}, \Delta_{\ell+1}^k\rangle\right).$$

Select a new index set $J_{\ell+1}$ satisfying

$$J_{\ell+1} \supseteq \{\ell+1, i_k\} \quad \textit{and} \quad \begin{cases} \textit{either } J_{\ell+1} \supseteq J_\ell^{\text{act}}, \\ \textit{or } J_{\ell+1} \supseteq \{-\ell\}. \end{cases} \tag{31}$$

*Step* 4. (*Descent test*). If (28) holds, declare a descent step, set $k(\ell+1) = k+1, i_{k+1} = \ell+1, x^{k+1} = y^{\ell+1}$, and update the bundle elements by (23). Otherwise, declare a null step, and set $k(\ell+1) = k(\ell)$.

*Step* 5. (*Update* $\eta$). Update the convexification parameter by

$$\begin{cases} \eta_{\ell+1} := \eta_\ell \quad \textit{if } \eta_{\ell+1}^{\min} \leq \eta_\ell, \\ \eta_{\ell+1} := \Gamma\eta_{\ell+1}^{\min} \quad \textit{and} \quad R_{\ell+1} := \mu_\ell + \eta_{\ell+1} \quad \textit{otherwise}. \end{cases} \tag{32}$$

*Step* 6. (*Update* $\mu$). If $F(y^{\ell+1}) > F(x^k) + M$, then the objective increase is unacceptable, let $\mu_{\ell+1} := \Gamma\mu_\ell$ and loop to Step 1; otherwise, set $\mu_{\ell+1} := \mu_\ell$.

*Step* 7. (*Loop*). Increase $\ell$ by 1 and go to Step 1.

*Remark* 1 (1) The predicted descent $\delta_\ell$ and the linearization error $\varepsilon_\ell$ are nonnegative (the details are given below); (2) in Step 6, if $F(y^{\ell+1}) > F(x^k) + M$ holds, then the current model is considered as "bad", so we should become more "conservative", and therefore increase the proximal parameter $\mu$ by setting $\mu_{\ell+1} = \Gamma\mu_\ell$. In the next section, we will prove that the number of increasing $\mu$ is finite; (3) the parameters $\mu_\ell, \eta_\ell$ in the algorithm will be stable eventually.

**Lemma 2** *The predicted descent $\delta_\ell$ and the linearization error $\varepsilon_\ell$ are nonnegative, and satisfy*

$$\varepsilon_\ell = \delta_\ell - \frac{R_\ell + \mu_\ell}{2\mu_\ell^2}\|s^\ell\|^2, \quad \textit{with } s^\ell = s_\varphi^\ell + s_h^\ell. \tag{33}$$

*Proof* In Step 2, from (30) and (18) we know $g_h^{\ell-1} = s_h^{\ell-1}$, hence

$$g_{\eta_\ell}^{-\ell} = \mu_\ell\left(x^k - z^{\ell+1}\right) - g_h^{\ell-1} = \mu_\ell\left(x^k - z^{\ell+1}\right) - s_h^{\ell-1} = s_\varphi^\ell. \tag{34}$$

Next, we prove $\delta_\ell \geq 0$ and $\varepsilon_\ell \geq 0$. From (26) and (29) we have

$$\begin{aligned}\delta_\ell &= f\left(x^k\right) + \frac{1}{2}\eta_\ell\left\|y^{\ell+1} - x^k\right\|^2 + h\left(x^k\right) - \left[\bar{\varphi}_\ell\left(y^{\ell+1}\right) + \bar{h}_\ell\left(y^{\ell+1}\right)\right]\\ &= f\left(x^k\right) + \frac{1}{2}\eta_\ell\left\|y^{\ell+1} - x^k\right\|^2 + h\left(x^k\right) - \check{\varphi}_\ell\left(z^{\ell+1}\right) - \left\langle s_\varphi^\ell, y^{\ell+1} - z^{\ell+1}\right\rangle - h\left(y^{\ell+1}\right)\\ &= f\left(x^k\right) - \check{\varphi}_\ell\left(z^{\ell+1}\right) - \left\langle s_\varphi^\ell, y^{\ell+1} - z^{\ell+1}\right\rangle + \frac{1}{2}\eta_\ell\left\|y^{\ell+1} - x^k\right\|^2 + h\left(x^k\right) - h\left(y^{\ell+1}\right).\end{aligned} \tag{35}$$

Let $j = -\ell$ in (22) and from (34) we can obtain

$$\begin{aligned}&f\left(x^k\right) - \check{\varphi}_\ell\left(z^{\ell+1}\right) - \left\langle s_\varphi^\ell, y^{\ell+1} - z^{\ell+1}\right\rangle\\ &\quad= f\left(x^k\right) - \left[f\left(x^k\right) - e_f^{-\ell} - \eta_\ell d_{-\ell}^k + \left\langle g_f^{-\ell} + \eta_\ell\triangle_{-\ell}^k, z^{\ell+1} - x^k\right\rangle\right] - \left\langle s_\varphi^\ell, y^{\ell+1} - z^{\ell+1}\right\rangle\\ &\quad= e_f^{-\ell} + \eta_\ell d_{-\ell}^k - \left\langle s_\varphi^\ell, z^{\ell+1} - x^k\right\rangle - \left\langle s_\varphi^\ell, y^{\ell+1} - z^{\ell+1}\right\rangle\\ &\quad= e_f^{-\ell} + \eta_\ell d_{-\ell}^k + \left\langle s_\varphi^\ell, x^k - y^{\ell+1}\right\rangle.\end{aligned} \tag{36}$$

On the other hand, from (30), we have

$$\begin{aligned}-h\left(y^{\ell+1}\right) &= -h\left(y^{\ell+1}\right) - \left\langle s_h^\ell, x^k - y^{\ell+1}\right\rangle + \left\langle s_h^\ell, x^k - y^{\ell+1}\right\rangle\\ &= -\bar{h}_\ell\left(x^k\right) + \left\langle s_h^\ell, x^k - y^{\ell+1}\right\rangle.\end{aligned} \tag{37}$$

Hence, by combining (36), (37) and (30), (35) can be written as

$$\begin{aligned}\delta_\ell &= e_f^{-\ell} + \eta_\ell d_{-\ell}^k + \left\langle s_\varphi^\ell, x^k - y^{\ell+1}\right\rangle + \frac{1}{2}\eta_\ell\left\|y^{\ell+1} - x^k\right\|^2 + h\left(x^k\right) - \bar{h}_\ell\left(x^k\right)\\ &\quad + \left\langle s_h^\ell, x^k - y^{\ell+1}\right\rangle \qquad (38)\\ &= e_f^{-\ell} + \eta_\ell d_{-\ell}^k + \left\langle \mu_\ell\left(x^k - y^{\ell+1}\right), x^k - y^{\ell+1}\right\rangle + \frac{1}{2}\eta_\ell\left\|y^{\ell+1} - x^k\right\|^2 + h\left(x^k\right) - \bar{h}_\ell\left(x^k\right)\\ &= e_f^{-\ell} + \eta_\ell d_{-\ell}^k + \frac{R_\ell + \mu_\ell}{2}\left\|y^{\ell+1} - x^k\right\|^2 + h\left(x^k\right) - \bar{h}_\ell\left(x^k\right). \qquad (39)\end{aligned}$$

In Step 5, the update for $\eta_\ell$ is done to ensure $\eta_\ell \geq \eta_\ell^{\min}$ for all iterations, so that $e_f^{-\ell} + \eta_\ell d_{-\ell}^k \geq 0$. Therefore, the predicted descent $\delta_\ell \geq 0$ since $h$ is convex.

For $\varepsilon_\ell$, from (27), one has

$$\begin{aligned}\varepsilon_\ell &= F(x^k) - [\bar{\varphi}_\ell(x^k) + \bar{h}_\ell(x^k)] \\ &= f(x^k) + h(x^k) - \check{\varphi}_\ell(z^{\ell+1}) - \langle s_\varphi^\ell, x^k - z^{\ell+1}\rangle - \bar{h}_\ell(x^k) \\ &= f(x^k) - \check{\varphi}_\ell(z^{\ell+1}) - \langle s_\varphi^\ell, x^k - z^{\ell+1}\rangle + h(x^k) - \bar{h}_\ell(x^k). \end{aligned} \tag{40}$$

Similar to (36), we have

$$\begin{aligned}&f(x^k) - \check{\varphi}_\ell(z^{\ell+1}) - \langle s_\varphi^\ell, x^k - z^{\ell+1}\rangle \\ &\quad = f(x^k) - [f(x^k) - e_f^{-\ell} - \eta_\ell d_{-\ell}^k + \langle g_f^{-\ell} + \eta_\ell \Delta_{-\ell}^k, z^{\ell+1} - x^k\rangle] \\ &\qquad - \langle s_\varphi^\ell, x^k - z^{\ell+1}\rangle \\ &\quad = e_f^{-\ell} + \eta_\ell d_{-\ell}^k - \langle s_\varphi^\ell, z^{\ell+1} - x^k\rangle - \langle s_\varphi^\ell, x^k - z^{\ell+1}\rangle \\ &\quad = e_f^{-\ell} + \eta_\ell d_{-\ell}^k. \end{aligned} \tag{41}$$

Thus, we have

$$\varepsilon_\ell = e_f^{-\ell} + \eta_\ell d_{-\ell}^k + h(x^k) - \bar{h}_\ell(x^k) \geq 0. \tag{42}$$

Equation (33) follows immediately from (39) and (42). □

From (33), We know that $\delta_\ell \geq \varepsilon_\ell$. Therefore, if $\delta_\ell \leq \epsilon$, then $\varepsilon_\ell \leq \epsilon$. So, we only use $\delta_\ell \leq \epsilon$ as the termination criterion in Step 5.

**Lemma 3** *The vectors $s_\varphi^\ell$ and $s_h^\ell$ of* (30) *and* (29) *are in fact subgradients, i.e.,*

$$s_\varphi^\ell \in \partial\check{\varphi}_\ell(z^{\ell+1}) \quad \textit{and} \quad s_h^\ell \in \partial h_\ell(y^{\ell+1}). \tag{43}$$

*Furthermore, we have*

$$\bar{\varphi}_\ell \leq \check{\varphi}_\ell \quad \textit{and} \quad \bar{h}_\ell \leq h. \tag{44}$$

*Proof* Let $\phi_f^\ell$ and $\phi_h^\ell$ denote the objectives of (13) and (14), respectively, i.e.,

$$\phi_f^\ell(\cdot) := \check{\varphi}_\ell(\cdot) + \bar{h}_{\ell-1}(\cdot) + \frac{1}{2}\mu_\ell \|\cdot - x^k\|^2, \tag{45}$$

$$\phi_h^\ell(\cdot) := \bar{\varphi}_\ell(\cdot) + h(\cdot) + \frac{1}{2}\mu_\ell \|\cdot - x^k\|^2. \tag{46}$$

By (13), (29) and the optimality condition of (45), we have

$$0 \in \partial\check{\varphi}_\ell(z^{\ell+1}) + s_h^{\ell-1} + \mu_\ell(z^{\ell+1} - x^k) = \partial\check{\varphi}_\ell(z^{\ell+1}) - s_\varphi^\ell,$$

which implies $s_\varphi^\ell \in \partial\check{\varphi}_\ell(z^{\ell+1})$. Similarly, by (14) and the optimality condition of (46), we obtain

$$0 \in \partial h\big(y^{\ell+1}\big) + s_\varphi^\ell + \mu_\ell\big(x^k - y^{\ell+1}\big) = \partial h\big(y^{\ell+1}\big) - s_h^\ell,$$

which implies $s_h^\ell \in \partial h_\ell(y^{\ell+1})$. So (43) holds.

Equation (44) follows immediately from (43). □

## 4 Convergence

In this section, we will study the convergence properties of Algorithm 1. Firstly, based on the objective function of problem (1), we need to slightly modify Assumption 1 as follows.

**Assumption 2** Given $x^0 \in \mathbb{R}^n$ and $M_0 \geq 0$, there exist an open bounded set $\mathcal{O}$ and a function $H$ such that $\mathcal{L}_0 := \{x \in \mathbb{R}^n : F(x) \leq F(x^0) + M_0\} \subset \mathcal{O}$, and $H$ is lower-$C^2$ on $\mathcal{O}$ with $H \equiv f$ on $\mathcal{L}_0$.

For convenience, we assume that Assumption 2 holds throughout the rest of convergence analysis.

In addition, from [24] we know that, if $f$ is a locally Lipschitz continuous function, then the subgradients of $f$ are locally bounded, i.e.,

$$\big\{g_f^\ell\big\} \text{ is bounded if } \big\{y^\ell\big\} \text{ is bounded.} \tag{47}$$

Further, as in [9], it follows that the model subgradients $s_\varphi^\ell$ in (43) satisfy

$$\big\{s_\varphi^\ell\big\} \text{ is bounded if } \big\{y^\ell\big\} \text{ is bounded.} \tag{48}$$

*Remark* 2 Note that (47) implies that $\{g_\varphi^\ell := g_f^\ell + \eta_\ell \triangle_i^\ell\}$ ($g_\varphi^\ell \in \partial\check{\varphi}_\ell$) is bounded if $\{y^\ell\}$ is bounded, since $\{\triangle_i^\ell\}$ in (17) is bounded if $\{y^\ell\}$ is bounded. Since $s_\varphi^\ell \in \partial\check{\varphi}_\ell$, then $s_\varphi^\ell \in \operatorname{conv}\{g_\varphi^j\}_{j\in J_\ell}$, thus we have $\|s_\varphi^\ell\| \leq \max_{j=1}^{\ell} \|g_\varphi^j\|$, and the model $\check{\varphi}_\ell$ satisfies condition (48) automatically when (47) holds.

The following lemma shows the properties of the model function $\check{\varphi}_\ell$, whose proof can be found in [13, 16].

**Lemma 4** *For the model function $\check{\varphi}_\ell$ and convexification parameter $\eta_\ell$, we have*

(i) *$\check{\varphi}_\ell$ is a convex function.*

(ii) *If $\eta_\ell \geq \eta_\ell^{\min}$, then*

$$\check{\varphi}_\ell\big(x^k\big) \leq f\big(x^k\big). \tag{49}$$

(iii) *If $\eta_{\ell+1} = \eta_\ell$, and either $J_{\ell+1} \supseteq J_\ell^{\mathrm{act}}$ or $J_{\ell+1} \supseteq \{-\ell\}$, then*

$$\check{\varphi}_{\ell+1}(\cdot) \geq \check{\varphi}_\ell\big(z^{\ell+1}\big) + \big\langle s_\varphi^\ell, \cdot - z^{\ell+1}\big\rangle$$

*if $y^{\ell+1}$ is a null step.*

(iv) *If* $J_\ell \supseteq \{\ell\}$, *then*

$$\check{\varphi}_\ell(\cdot) \geq f(y^\ell) + \frac{1}{2}\eta_\ell \|y^\ell - x^k\|^2 + \langle g_f^\ell + \eta_\ell(y^\ell - x^k), \cdot - y^\ell \rangle,$$

*for some* $g_f^\ell \in \partial f(y^\ell)$.

(v) *If* $\eta_\ell \geq \rho^{\mathrm{id}}$, *then*

$$\check{\varphi}_\ell(\omega) \leq f(\omega) + \frac{1}{2}\eta_\ell \|\omega - x^k\|^2 \quad \text{for all } \omega \in \mathcal{L}_0. \tag{50}$$

From the updating rule in Step 5 of Algorithm 1, the convexification parameter $\eta_\ell$ is either unchanged or increasing. The following lemma shows that $\eta_\ell$ can be fixed in a finite number of iterations, whose proof can be found in [13].

**Lemma 5** *There exist an index* $\ell_1$ *and a positive constant* $\overline{\eta} > 0$ *such that*

$$\eta_\ell \equiv \overline{\eta}, \quad \text{for all } \ell \geq \ell_1.$$

**Lemma 6** *Suppose that there exists an integer* $K$ *such that, for all* $\ell \geq K$, *only null steps occur without increasing* $\mu$. *Then the following results hold*:

(i) *The sequences*

$$\left\{\phi_f^\ell(z^{\ell+1}) = \check{\varphi}_\ell(z^{\ell+1}) + \bar{h}_{\ell-1}(z^{\ell+1}) + \frac{1}{2}\mu_\ell \|z^{\ell+1} - x^k\|^2\right\}_{\ell \geq K},$$
$$\left\{\phi_h^\ell(y^{\ell+1}) = \bar{\varphi}_\ell(y^{\ell+1}) + h(y^{\ell+1}) + \frac{1}{2}\mu_\ell \|y^{\ell+1} - x^k\|^2\right\}_{\ell \geq K}$$

*are nondecreasing and convergent.*

(ii) *The sequences* $\{y^{\ell+1}\}$ *and* $\{z^{\ell+1}\}$ *are bounded,* $\|z^{\ell+1} - y^{\ell+1}\| \to 0$ *and* $\|z^{\ell+2} - y^{\ell+1}\| \to 0$ *as* $\ell \to \infty$.

*Proof* First, using partial linearizations of the subproblems to show (i) is hold. Fixed $\ell \geq K$. By the definitions in (13) and (29), we have $\bar{\varphi}_\ell(z^{\ell+1}) = \check{\varphi}_\ell(z^{\ell+1})$ and

$$z^{\ell+1} = \arg\min\left\{\bar{\phi}_f^\ell(\cdot) := \bar{\varphi}_\ell(\cdot) + \bar{h}_{\ell-1}(\cdot) + \frac{1}{2}\mu_\ell \|\cdot - x^k\|^2\right\}, \tag{51}$$

from $\nabla\bar{\phi}_f^\ell(z^{\ell+1}) = 0$. Since $\bar{\phi}_f^\ell$ is quadratic and $\bar{\phi}_f^\ell(z^{\ell+1}) = \phi_f^\ell(z^{\ell+1})$, by Taylor's expansion

$$\begin{aligned}\bar{\phi}_f^\ell(\cdot) &= \bar{\phi}_f^\ell(z^{\ell+1}) + \nabla\bar{\phi}_f^\ell(z^{\ell+1})(\cdot - z^{\ell+1}) + \frac{1}{2}\mu_\ell \|\cdot - z^{\ell+1}\|^2 \\ &= \phi_f^\ell(z^{\ell+1}) + \frac{1}{2}\mu_\ell \|\cdot - z^{\ell+1}\|^2.\end{aligned} \tag{52}$$

Similarly, by the definitions in (14) and (30), we have $\bar{h}_\ell(y^{\ell+1}) = h(y^{\ell+1})$, and

$$y^{\ell+1} = \arg\min\left\{\bar{\phi}_h^\ell(\cdot) := \bar{\varphi}_\ell(\cdot) + \bar{h}_\ell(\cdot) + \frac{1}{2}\mu_\ell \|\cdot - x^k\|^2\right\}, \tag{53}$$

$$\bar{\phi}_h^\ell(\cdot) = \phi_h^\ell\big(y^{\ell+1}\big) + \frac{1}{2}\mu_\ell\big\|\cdot - y^{\ell+1}\big\|^2. \tag{54}$$

Next, to bound the objective values of the linearized subproblem (51) and (53) from above, we use $\bar{\varphi}_\ell \le \check{\varphi}_\ell$ and $\bar{h}_{\ell-1} \le h$, $\bar{h}_\ell \le h$ of (44) and $\check{\varphi}_\ell(x^k) \le f(x^k)$ in (ii) of Lemma 4

$$\phi_f^\ell\big(z^{\ell+1}\big) + \frac{1}{2}\mu_\ell\big\|x^k - z^{\ell+1}\big\|^2 = \bar{\phi}_f^\ell\big(x^k\big) \le \check{\varphi}_\ell\big(x^k\big) + h\big(x^k\big) \le F\big(x^k\big), \tag{55}$$

$$\phi_h^\ell\big(y^{\ell+1}\big) + \frac{1}{2}\mu_\ell\big\|x^k - y^{\ell+1}\big\|^2 = \bar{\phi}_h^\ell\big(x^k\big) \le \check{\varphi}_\ell\big(x^k\big) + h\big(x^k\big) \le F\big(x^k\big). \tag{56}$$

From (14) and (51), we have $\bar{\phi}_f^\ell \le \phi_h^\ell$. On the other hand, since only null step occurred, so $x^{k+1} = x^k$, the algorithm ensures that $\mu_\ell = \mu_{\ell+1}$, and $\bar{\varphi}_\ell \le \check{\varphi}_{\ell+1}$ by (iii) of Lemma 4, we can obtain $\bar{\phi}_h^\ell \le \phi_f^{\ell+1}$. By (52) and (54), we see that

$$\phi_f^\ell\big(z^{\ell+1}\big) + \frac{1}{2}\mu_\ell\big\|y^{\ell+1} - z^{\ell+1}\big\|^2 = \bar{\phi}_f^\ell\big(y^{\ell+1}\big) \le \phi_h^\ell\big(y^{\ell+1}\big), \tag{57}$$

$$\phi_h^\ell\big(y^{\ell+1}\big) + \frac{1}{2}\mu_\ell\big\|z^{\ell+2} - y^{\ell+1}\big\|^2 = \bar{\phi}_h^\ell\big(z^{\ell+2}\big) \le \phi_f^{\ell+1}\big(z^{\ell+2}\big). \tag{58}$$

In particular, from (57) and (58), we have the relation

$$\phi_f^\ell\big(z^{\ell+1}\big) \le \phi_h^\ell\big(y^{\ell+1}\big) \le \phi_f^{\ell+1}\big(z^{\ell+2}\big)$$

which implies that $\{\phi_f^\ell(z^{\ell+1})\}_{\ell \ge K}$ and $\{\phi_h^\ell(y^{\ell+1})\}_{\ell \ge K}$ are nondecreasing sequences. Together with the bound of $F(x^k)$ from (55) and (56), the convergence is established.

For (ii), we have proved the convergence of $\{\phi_f^\ell(z^{\ell+1})\}$ and $\{\phi_h^\ell(y^{\ell+1})\}$ in (i) when $\ell \ge K$, so there must have a common limit, say $\phi_\infty \le F(x^k)$, such that

$$\phi_f^\ell\big(z^{\ell+1}\big) \to \phi_\infty, \qquad \phi_h^\ell\big(y^{\ell+1}\big) \to \phi_\infty \tag{59}$$

and we have $\|z^{\ell+1} - y^{\ell+1}\| \to 0$ and $\|z^{\ell+2} - y^{\ell+1}\| \to 0$ from (57) and (58), $\{y^{\ell+1}\}$ and $\{z^{\ell+1}\}$ are bounded from (55) and (56). Then the sequences $\{g_f^\ell\}$ and $\{s_\varphi^\ell\}$ are bounded by (47) and (48). □

The following lemma shows that the number of times of increasing $\mu$ is finite.

**Lemma 7** *Suppose that $i_k \in J_\ell$, and let $N_\ell$ be the number of times of increasing $\mu$. Then there exists a positive constant $L$ such that*

$$N_\ell \le \left\lceil \frac{\ln \frac{L^2}{M\mu_0}}{\ln \Gamma} \right\rceil, \tag{60}$$

*where $\lceil a \rceil$ is the smallest integer greater than or equal to $a$. As a result, there exists an index $\ell_2$ such that*

$$\mu_\ell = \bar{\mu}, \quad \textit{for all } \ell \ge \ell_2.$$

*Proof* Let $\ell_r$ be the index corresponding to the $r$th time that $\mu$ increases, then when $\ell_r + 1 \leq \ell < \ell_{r+1}$, we have

$$\mu_\ell = \Gamma^r \mu_0. \tag{61}$$

Since $i_k \in J_\ell$, from (24), we obtain $g^{i_k} \in \partial\check{\varphi}(x^k)$ by writing $i = i_k$, and it also holds that $g_f^{i_k} \in \partial f(x^k)$ from (21), so $\|g_f^{i_k}\|$ is bounded. Hence

$$\begin{aligned}
&p_{\mu_\ell}(\bar{\varphi}_\ell + h)(x^k) \\
&\quad = \arg\min\left\{\bar{\varphi}_\ell(y) + h(y) + \frac{1}{2}\mu_\ell\|y - x^k\|^2\right\} \\
&\quad \in \left\{y \middle| \bar{\varphi}_\ell(y) + h(y) + \frac{1}{2}\mu_\ell\|y - x^k\|^2 \leq \bar{\varphi}_\ell(x^k) + h(x^k)\right\} \\
&\quad \subseteq \Bigl\{y \Big| \bar{\varphi}_\ell(x^k) + \langle g_f^{i_k}, y - x^k\rangle + h(x^k) + \langle g_h^k, y - x^k\rangle \\
&\qquad + \frac{1}{2}\mu_\ell\|y - x^k\|^2 \leq \bar{\varphi}_\ell(x^k) + h(x^k)\Bigr\} \\
&\quad = \left\{y \middle| \langle g_f^{i_k}, y - x^k\rangle + \langle g_h^k, y - x^k\rangle + \frac{1}{2}\mu_\ell\|y - x^k\|^2 \leq 0\right\} \\
&\quad = \left\{y \middle| \frac{1}{2}\mu_\ell\|y - x^k\|^2 \leq -\langle g_f^{i_k} + g_h^k, y - x^k\rangle\right\} \\
&\quad \subseteq \left\{y \middle| \frac{1}{2}\mu_\ell\|y - x^k\|^2 \leq \|g_f^{i_k} + g_h^k\|\|y - x^k\|\right\} \\
&\quad = \left\{y \middle| \|y - x^k\| \leq \frac{2\|g_f^{i_k}\| + 2\|g_h^k\|}{\mu_\ell}\right\}.
\end{aligned}$$

When $\ell_r + 1 \leq \ell < \ell_{r+1}$, if $y^{\ell+1}$ is a null step, we know $g_f^\ell$ is bounded from Lemma 6, else $y^{\ell+1}$ is a descent step, and the corresponding subgradient $g^{i_{k(\ell+1)}} \in \partial f(x^{k+1})$ is also bounded. Therefore, there exists a constant $L > 0$ such that $\max\{\|g_f^{i_k}\|, \|g_f^{\ell+1}\|, \|g_h^k\|, \|g_h^{\ell+1}\|\} \leq \frac{L}{2}$. Thus we have

$$y^{\ell+1} = p_{\mu_\ell}(\bar{\varphi}_\ell + h)(x^k) \in \left\{y \middle| \|y - x^k\| \leq \frac{2L}{\mu_\ell}\right\}.$$

This together with (61) shows that

$$\begin{aligned}
&\left|F(y^{\ell+1}) - F(x^k)\right| \\
&\quad = \left|f(x^k) + \langle g_f^{\ell+1}, y^{\ell+1} - x^k\rangle - f(x^k) + h(x^k) + \langle g_h^{\ell+1}, y^{\ell+1} - x^k\rangle - h(x^k)\right| \\
&\quad \leq \|g_f^{\ell+1} + g_h^{\ell+1}\|\|y^{\ell+1} - x^k\| \\
&\quad \leq 2 \cdot \frac{L}{2} \cdot \frac{2L}{\Gamma^r\mu_0} \leq M.
\end{aligned}$$

Thus, if

$$r \geq \frac{\ln\frac{L^2}{M\mu_0}}{\ln\Gamma},$$

then

$$f\left(y^{\ell+1}\right) \leq f\left(x^k\right) + M, \quad \forall \ell_r + 1 \leq \ell \leq \ell_{r+1}.$$

This means that the number of times $N_\ell$ of increasing $\mu$ satisfies (60). The latter part of the lemma follows immediately from the above result. □

**Theorem 3** *If $\delta_\ell = 0$ and $\eta_\ell \geq \rho^{\mathrm{id}}$, then $x^k$ is a stationary point of $F$.*

*Proof* From (33), we have the relation

$$\delta_\ell = \varepsilon_\ell + \frac{R_\ell + \mu_\ell}{2\mu_\ell^2} \left\| s^\ell \right\|^2.$$

If $\delta_\ell = 0$, then $\varepsilon_\ell = 0$ and $s^\ell = 0$, and therefore

$$x^k = y^{\ell+1} = p_{\mu_\ell}(\bar{\varphi}_\ell + h)\left(x^k\right). \tag{62}$$

From the last result of Theorem 1, we know that $x^k$ is a stationary point of $\bar{\varphi}_\ell + h$. In addition, from $\varepsilon_\ell = 0$ in (27), we have

$$f\left(x^k\right) + h\left(x^k\right) = \bar{\varphi}_\ell\left(x^k\right) + \bar{h}_\ell\left(x^k\right).$$

This together with $\bar{h}_\ell(x^k) \leq h(x^k)$ shows that

$$f\left(x^k\right) \leq \bar{\varphi}_\ell\left(x^k\right). \tag{63}$$

On one hand, for $\omega \in \mathcal{L}_0$, if $\eta_\ell \geq \rho^{\mathrm{id}}$, we obtain by (62) and (63)

$$F\left(x^k\right) = f\left(x^k\right) + h\left(x^k\right) \leq \bar{\varphi}_\ell\left(x^k\right) + h\left(x^k\right) \leq \bar{\varphi}_\ell(\omega) + h(\omega) + \frac{1}{2}\mu_\ell \left\| \omega - x^k \right\|^2. \tag{64}$$

From the convexity of $\check{\varphi}_\ell$ and (50), we have

$$\bar{\varphi}_\ell(\omega) \leq \check{\varphi}_\ell(\omega) \leq f(\omega) + \frac{1}{2}\eta_\ell \left\| \omega - x^k \right\|^2.$$

So, (64) can be written as

$$F\left(x^k\right) \leq f(\omega) + \frac{1}{2}\eta_\ell \left\| \omega - x^k \right\|^2 + h(\omega) + \frac{1}{2}\mu_\ell \left\| \omega - x^k \right\|^2. \tag{65}$$

On the other hand, for $\omega \notin \mathcal{L}_0$, from (28) we can obtain

$$F\left(x^k\right) \leq F\left(x^0\right) \leq F\left(x^0\right) + M \leq F(\omega) = F(\omega) + \frac{1}{2}R_\ell \left\| \omega - x^k \right\|^2. \tag{66}$$

Combining (64) and (66), we have

$$F\left(x^k\right) \leq F(\omega) + \frac{1}{2}R_\ell \left\| \omega - x^k \right\|^2 \quad \text{for all } \omega \in \mathbb{R}^n.$$

Hence

$$x^k = p_{R_\ell} F(x^k),$$

which together with Theorem 1 shows that $x^k$ is a stationary point of $F$. □

We are now in a position to present the main convergence result of our algorithm. As usual in bundle methods, two cases are considered: the algorithm generates finite number of descent steps; and the algorithm generates infinite number of descent steps. We set the stopping parameter $\epsilon = 0$.

**Theorem 4** *Let $\bar{\eta}$ be stabilized value for the convexification parameter sequence and assume $\bar{\eta} \geq \rho^{\mathrm{id}}$. Then the following mutually exclusive situations hold*:

(i) *Algorithm* 1 *generates finite number of descent steps followed by infinitely many null steps. Let $\bar{x}$ be the last stability center. Then $y^{\ell+1} \to \bar{x}$, and $\bar{x}$ is a stationary point of $F$.*

(ii) *Algorithm* 1 *generates an infinite sequence $\{x^k\}$ of stability centers. Then any accumulation point of $\{x^k\}$ is a stationary point of $F$.*

*Proof* For (i), without loss of generality, we may assume $\eta_\ell = \overline{\eta}$, $\mu_\ell = \overline{\mu}$, and $R_\ell = \overline{R}$ throughout. As in Lemma 6, for the bounded sequences $\{y^\ell\}$ and $\{z^\ell\}$ we showed that $\|y^\ell - z^\ell\| \to 0$ and $\|z^{\ell+1} - y^\ell\| \to 0$ as $\ell \to \infty$. Therefore $y^{\ell_i} \to p$ as $i \to \infty$ implies $z^{\ell_i} \to p$ and $z^{\ell_i+1} \to p$ as $i \to \infty$. For $\omega \in \mathcal{L}_0$ near $p$, by (50) and the convexity of $h$, we have

$$\begin{aligned}
F(\omega) &= f(\omega) + h(\omega) \\
&\geq \check{\varphi}_{\ell_i}(\omega) - \frac{1}{2}\bar{\eta}\|\omega - \bar{x}\|^2 + h(\omega) \\
&\geq \bar{\varphi}_{\ell_i}(\omega) - \frac{1}{2}\bar{\eta}\|\omega - \bar{x}\|^2 + \bar{h}_{\ell_i-1}(\omega) \\
&= \check{\varphi}_{\ell_i}(z^{\ell_i+1}) + \langle s_\varphi^{\ell_i}, \omega - z^{\ell_i+1}\rangle - \frac{1}{2}\bar{\eta}\|\omega - \bar{x}\|^2 + h(y^{\ell_i}) + \langle s_h^{\ell_i-1}, \omega - y^{\ell_i}\rangle.
\end{aligned} \tag{67}$$

Let $x^k = \bar{x}$, $\mu_\ell = \bar{\mu}$, from (29) and the boundedness of $s_\varphi^\ell + s_h^{\ell-1} = \overline{\mu}(\bar{x} - z^{\ell+1})$, we know

$$\begin{aligned}
&\langle s_\varphi^{\ell_i}, \omega - z^{\ell_i+1}\rangle + \langle s_h^{\ell_i-1}, \omega - y^{\ell_i}\rangle \\
&\quad = \langle s_\varphi^{\ell_i}, \omega - z^{\ell_i+1}\rangle + \langle s_h^{\ell_i-1}, \omega - z^{\ell_i+1} + z^{\ell_i+1} - y^{\ell_i}\rangle \\
&\quad = \bar{\mu}\langle \bar{x} - z^{\ell_i+1}, \omega - z^{\ell_i+1}\rangle + \langle s_h^{\ell_i-1}, z^{\ell_i+1} - y^{\ell_i}\rangle.
\end{aligned} \tag{68}$$

Note that

$$-\frac{1}{2}\bar{\eta}\|\omega - \bar{x}\|^2 \tag{69}$$

$$\begin{aligned}
&= -\frac{1}{2}\bar{\eta}\left\|\omega - z^{\ell_i+1} + z^{\ell_i+1} - \bar{x}\right\|^2 \\
&= -\frac{1}{2}\bar{\eta}\left\|\omega - z^{\ell_i+1}\right\|^2 - \frac{1}{2}\bar{\eta}\left\|z^{\ell_i+1} - \bar{x}\right\|^2 + \bar{\eta}\langle \bar{x} - z^{\ell_i+1}, \omega - z^{\ell_i+1}\rangle.
\end{aligned} \tag{70}$$

Combining (68) and (70), (67) can be written as

$$\begin{aligned}F(\omega) &= \check{\varphi}_{\ell_i}\bigl(z^{\ell_i+1}\bigr) - \frac{1}{2}\bar{\eta}\bigl\|z^{\ell_i+1}-\bar{x}\bigr\|^2 + h\bigl(y^{\ell_i}\bigr) + (\bar{\eta}+\bar{\mu})\bigl\langle \bar{x}-z^{\ell_i+1}, \omega - z^{\ell_i+1}\bigr\rangle \\ &\quad + \bigl\langle s_h^{\ell_i-1}, z^{\ell_i+1}-y^{\ell_i}\bigr\rangle - \frac{1}{2}\bar{\eta}\bigl\|\omega - z^{\ell_i+1}\bigr\|^2.\end{aligned}$$

By Claim (iv) of Lemma 4 written with $\ell = \ell_i$ for $\omega = z^{\ell_i+1}$, we have the following inequality:

$$\begin{aligned}\check{\varphi}_{\ell_i}\bigl(z^{\ell_i+1}\bigr) - \frac{1}{2}\bar{\eta}\bigl\|z^{\ell_i+1}-\bar{x}\bigr\|^2 &\geq f\bigl(y^{\ell_i}\bigr) + \frac{1}{2}\bar{\eta}\bigl\|y^{\ell_i}-\bar{x}\bigr\|^2 - \frac{1}{2}\bar{\eta}\bigl\|z^{\ell_i+1}-\bar{x}\bigr\|^2 \\ &\quad + \bigl\langle g_f^{\ell_i} + \overline{\eta}\bigl(y^{\ell_i}-\bar{x}\bigr), z^{\ell_i+1}-y^{\ell_i}\bigr\rangle.\end{aligned}$$

Then

$$\begin{aligned}F(\omega) &= f(\omega) + h(\omega) \\ &\geq \check{\varphi}_{\ell_i}\bigl(z^{\ell_i+1}\bigr) - \frac{1}{2}\bar{\eta}\bigl\|z^{\ell_i+1}-\bar{x}\bigr\|^2 + h\bigl(y^{\ell_i}\bigr) - \frac{1}{2}\bar{\eta}\bigl\|\omega - z^{\ell_i+1}\bigr\|^2 \\ &\quad + (\bar{\eta}+\bar{\mu})\bigl\langle \bar{x}-z^{\ell_i+1}, \omega - z^{\ell_i+1}\bigr\rangle + \bigl\langle s_h^{\ell_i-1}, z^{\ell_i+1}-y^{\ell_i}\bigr\rangle \\ &\geq f\bigl(y^{\ell_i}\bigr) + h\bigl(y^{\ell_i}\bigr) + \frac{1}{2}\bar{\eta}\bigl\|y^{\ell_i}-\bar{x}\bigr\|^2 - \frac{1}{2}\bar{\eta}\bigl\|z^{\ell_i+1}-\bar{x}\bigr\|^2 - \frac{1}{2}\bar{\eta}\bigl\|\omega - z^{\ell_i+1}\bigr\|^2 \\ &\quad + \bigl\langle g_f^{\ell_i} + \bar{\eta}\bigl(y^{\ell_i}-\bar{x}\bigr), z^{\ell_i+1}-y^{\ell_i}\bigr\rangle + (\bar{\eta}+\bar{\mu})\bigl\langle \bar{x}-z^{\ell_i+1}, \omega - z^{\ell_i+1}\bigr\rangle \\ &\quad + \bigl\langle s_h^{\ell_i-1}, z^{\ell_i+1}-y^{\ell_i}\bigr\rangle.\end{aligned}$$

From $s_h^{\ell-1} = \overline{\mu}(\bar{x}-z^{\ell+1}) - s_\varphi^\ell$, the bounded sequence $\{\overline{\mu}(\bar{x}-z^{\ell+1})\}$, $\{s_\varphi^\ell\}$, $\{g_f^\ell\}$ and $\{y^\ell\}$, we know that $\{s_h^{\ell-1}\}$ and $\{g_f^{\ell_i} + \overline{\eta}(y^{\ell_i}-\bar{x})\}$ are bounded. Taking the limit as $i \to \infty$, and using the fact that $f$ is continuous at $p$, we obtain

$$\begin{aligned}F(\omega) &= f(\omega) + h(\omega) \\ &\geq \lim_{i\to\infty} \check{\varphi}_{\ell_i}\bigl(z^{\ell_i+1}\bigr) - \frac{1}{2}\bar{\eta}\|p-\bar{x}\|^2 + h(p) - \frac{1}{2}\bar{\eta}\|\omega - p\|^2 + \bar{R}\langle \bar{x}-p, \omega-p\rangle \\ &\geq f(p) + h(p) - \frac{1}{2}\bar{\eta}\|\omega-p\|^2 + \bar{R}\langle \bar{x}-p, \omega-p\rangle \\ &= F(p) - \frac{1}{2}\bar{\eta}\|\omega-p\|^2 + \bar{R}\langle \bar{x}-p, \omega-p\rangle,\end{aligned} \tag{71}$$

for all $\omega \in \mathcal{L}_0$ near $p$. Since $\frac{1}{2}\overline{\eta}\|\omega-p\|^2 = o(\|\omega-p\|)$, the last inequality means that $\overline{R}(\bar{x}-p) \in \partial F(p)$ by Definition 8.3 in [17], which implies $p = p_{\bar{R}}F(\bar{x})$ by Theorem 1. Since $f$ is continuous and condition (50) holds at all accumulation points of $\{y^\ell\}$, then the entire sequence $\{y^\ell\}$ converges to the proximal point $p_{\bar{R}}F(\bar{x})$. Furthermore, evaluating the relations at $\omega = p$ shows that the following equation holds for the entire sequence by Theorem 2 in [16]:

$$\lim_{i\to\infty} \check{\varphi}_{\ell_i}\bigl(z^{\ell_i+1}\bigr) = f(p) + \frac{1}{2}\overline{\eta}\|p-\bar{x}\|^2. \tag{72}$$

So as $\ell \to \infty$, the whole sequence

$$\{y^\ell\} \to p = p_{\bar{R}}F(\bar{x}) \quad \text{with } \check{\varphi}_\ell(z^{\ell+1}) \to f(p) + \frac{1}{2}\bar{\eta}\|p - \bar{x}\|^2.$$

Thus, from (26) we have

$$\begin{aligned}
\delta_\ell &= f(\bar{x}) + \frac{1}{2}\bar{\eta}\|y^{\ell+1} - \bar{x}\|^2 + h(\bar{x}) - [\bar{\varphi}_\ell(y^{\ell+1}) + h(y^{\ell+1})] \\
&= f(\bar{x}) + \frac{1}{2}\bar{\eta}\|y^{\ell+1} - \bar{x}\|^2 + h(\bar{x}) - \check{\varphi}_\ell(z^{\ell+1}) - \langle s_\varphi^\ell, y^{\ell+1} - z^{\ell+1}\rangle - h(y^{\ell+1}) \\
&\to f(\bar{x}) + \frac{1}{2}\bar{\eta}\|p - \bar{x}\|^2 + h(\bar{x}) - f(p) - \frac{1}{2}\bar{\eta}\|p - \bar{x}\|^2 - h(p) \\
&= f(\bar{x}) + h(\bar{x}) - f(p) - h(p) \\
&= F(\bar{x}) - F(p).
\end{aligned}$$

Since null step does not satisfy the descent test in Step 4 of the algorithm, we have $F(y^{\ell+1}) > F(\bar{x}) - \kappa\delta_\ell$. Taking the limit as $\ell \to \infty$ gives the relation $F(p) \geq F(\bar{x}) - \kappa(F(\bar{x}) - F(p))$, so $F(\bar{x}) \leq F(p)$ because $\kappa \in (0,1)$. But $p = p_{\bar{R}}F(\bar{x})$ implies

$$F(p) + \bar{R}\|p - \bar{x}\|^2 \leq F(\bar{x}),$$

which shows that $\bar{x} = p$. That is, $\bar{x} = p_{\bar{R}}F(\bar{x})$, so $\bar{x}$ is a stationary point of $F$ from Theorem 1.

For (ii), $\mathcal{L}_0$ is a compact set, and the sequence $\{x^k\} \subset \mathcal{L}_0$, so it has an accumulation point, i.e., there exists some infinite set $K$ such that $x^k \to \hat{x} \in \mathcal{L}_0$ as $K \ni k \to \infty$. Since $x^{k+1} = y^{i_{k+1}}$, let $j_k = i_{k+1} - 1$ so that $x^{k+1} = p_{\bar{\mu}}(\bar{\varphi}_{j_k} + h)(x^k)$. The descent test

$$F(x^{k+1}) \leq F(x^k) - \kappa\delta_{j_k}$$

implies that, as $k \to \infty$, either $F(x^k) \searrow -\infty$, or $\delta_{j_k} \to 0$. By Assumption 2, $F(x^k)$ is bounded below, therefore, $\delta_{j_k} \to 0$. From (39), this means that

$$\|y^{j_k+1} - x^k\|, \qquad e_f^{-j_k} + \bar{\eta}d_{-j_k}^k, \qquad h(x^k) - \bar{h}_{j_k}(x^k)$$

must converge to 0. By

$$\|z^{j_k+1} - x^k\| \leq \|z^{j_k+1} - y^{j_k+1}\| + \|y^{j_k+1} - x^k\|$$

and

$$\|z^{j_k+1} - y^{j_k+1}\| \to 0$$

in Lemma 6, we have

$$\|z^{j_k+1} - x^k\| \to 0.$$

By (21), $\check{\varphi}_{j_k}(z^{j_k+1}) - f(x^k) \to 0$ as $k \to \infty$, from (29) we know that

$$\bar{\varphi}_{j_k}\left(y^{j_k+1}\right) = \check{\varphi}_{j_k}\left(z^{j_k+1}\right) + \left\langle s_{\varphi}^{j_k}, y^{j_k+1} - z^{j_k+1}\right\rangle.$$

Therefore,

$$\bar{\varphi}_{j_k}\left(y^{j_k+1}\right) - f\left(x^k\right) \to 0$$

as $k \to \infty$. Consider now $k \in K$. Since $\|x^{k+1} - x^k\| = \|y^{j_k+1} - x^k\| \to 0$, both $x^{k+1}$ and $x^k$ converge to $x^{\inf}$ as $K \ni k \to \infty$ with

$$\bar{\varphi}_{j_k}\left(x^{k+1}\right) \to f(\hat{x}).$$

And from $x^{k+1} = p_{\bar{\mu}}(\bar{\varphi}_{j_k} + h)(x^k)$, $\bar{\eta} \geq \rho^{\mathrm{id}}$ and (50), for all $\omega \in \mathcal{L}_0$,

$$\begin{aligned}
&\bar{\varphi}_{j_k}\left(x^{k+1}\right) + h\left(x^{k+1}\right) + \frac{1}{2}\bar{\mu}\left\|x^{k+1} - x^k\right\|^2 \\
&\quad \leq \bar{\varphi}_{j_k}(\omega) + h(\omega) + \frac{1}{2}\bar{\mu}\left\|\omega - x^k\right\|^2 \\
&\quad \leq \check{\varphi}_{j_k}(\omega) + h(\omega) + \frac{1}{2}\bar{\mu}\left\|\omega - x^k\right\|^2 \\
&\quad \leq f(\omega) + h(\omega) + \frac{1}{2}\bar{R}\left\|\omega - x^k\right\|^2.
\end{aligned}$$

Therefore, taking the limit $k \in K$, we have

$$f(\hat{x}) + h(\hat{x}) \leq f(\omega) + h(\omega) + \frac{1}{2}\bar{R}\|\omega - \hat{x}\|^2, \quad \text{for all } \omega \in \mathcal{L}_0.$$

On the other hand, $x^{\inf} \in \mathcal{L}_0$ and for any $\omega \notin \mathcal{L}_0$, it follows

$$F(\hat{x}) \leq F\left(x^0\right) \leq F\left(x^0\right) + M < F(\omega) < F(\omega) + \frac{1}{2}\bar{R}\|\omega - \hat{x}\|^2.$$

Hence,

$$F(\hat{x}) \leq F(\omega) + \frac{1}{2}\bar{R}\|\omega - \hat{x}\|^2, \quad \text{for all } \omega \in \mathbb{R}^n.$$

Therefore, $\hat{x} = p_{\bar{R}}F(\hat{x})$ with $\bar{R} \geq \rho^{\mathrm{id}}$, hence $\hat{x}$ is a stationary point of $F$ from Theorem 1. □

## 5 Numerical results

This section aims to test the practical effectiveness of Algorithm 1. We tested a set of nine problems. The first set of seven problems are generalized from the unconstrained versions in [25] by imposing suitable constraints, the second set of two nonconvex unconstrained problems are taken from [13, 26] which are the sum of a nonconvex function and a convex function.

All numerical experiments were implemented by using MATLAB R2014a, and on a ThinkPad laptop with Windows 7 operating system. The first seven problems have the form of (2) with $C = \{x : \|x - a\| \le b\}$, where $a \in \mathbb{R}^n$ and $0 < b \in \mathbb{R}$ are given below. These problems are transformed to the form of (3) by using the indicator function. The detailed data for the seven problems are listed below. For simplicity, we use the MATLAB notations: `ones(p,q)` and `zeros(p,q)` denote `p`-by-`q` matrices of ones and zeros, respectively.

`CB2`: $f(x) = \max\{x_1^2 + x_2^4, (2 - x_1)^2 + (2 - x_2)^2, 2e^{x_2 - x_1}\}$, $y^0 = (3,3)^T$, $a = (0,0)^T$, $b = 1$.

`CB3`: $f(x) = \max\{x_1^4 + x_2^2, (2 - x_1)^2 + (2 - x_2)^2, 2e^{x_2 - x_1}\}$, $y^0 = (3,3)^T$, $a = (3,3)^T$, $b = 1$.

`LQ`: $f(x) = \max\{-x_1 - x_2, -x_1 - x_2 + x_1^2 + x_2^2 - 1\}$, $y^0 = (1,1)^T$, $a = (1,-1)^T$, $b = 1$.

`Mifflin1`: $f(x) = -x_1 + 20\max\{x_1^2 + x_2^2 - 1, 0\}$, $y^0 = (1.5, 0.5)^T$, $a = (-2,2)^T$, $b = 1$.

`Rosen-Suzuki`: $f(x) = \max_{1\le i\le 4} f_i(x)$, $y^0 = (1, 2.1, -3, -0.9)^T$, $a = (1,2,3,4)^T$, $b = 2$, with

$$
\begin{aligned}
f_1(x) &= x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4, \\
f_2(x) &= f_1(x) + 10\big(x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8\big), \\
f_3(x) &= f_1(x) + 10\big(x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10\big), \\
f_4(x) &= f_1(x) + 10\big(2x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5\big).
\end{aligned}
$$

`Shor`: $f(x) = \max_{1\le i\le 10}\{d_i \sum_{j=1}^{5} (x_j - c_{ij})^2)\}$, $y^0 =$ `zeros(10,1)`, $a =$ `zeros(10,1)`, $b = 3$, $d = (1, 5, 10, 2, 4, 3, 1.7, 2.5, 6, 3.5)^T$,

$$
C = \begin{pmatrix}
0 & 2 & 1 & 1 & 3 & 0 & 1 & 1 & 0 & 1 \\
0 & 1 & 2 & 4 & 2 & 2 & 1 & 0 & 0 & 1 \\
0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 \\
0 & 1 & 1 & 2 & 0 & 0 & 1 & 2 & 1 & 0 \\
0 & 3 & 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0
\end{pmatrix}.
$$

`MAXL`: $f(x) = \max_{1\le i\le 20} |x_i|$, $a = (-$`ones(1,10)`$,$`ones(1,10)`$)^T$, $b = 4$, $y^0 = (1, 1.1, 3, 1.1, 5, 1.1, 7, 1.1, 9, 1.1, -11, 0.1, -13, 0.1, -15, 0.1, -17, 0.1, -19, 0.1)^T$.

The second set of two problems are:

`Regular`: $F(x) = \sum_{i=1}^{n} |f_i(x)| + \frac{1}{2}\|x\|^2$, where

$$
f_i(x) := \big(ix_i^2 - 2x_i\big) + \sum_{j=1}^{n} x_j, \quad i = 1, 2, \ldots, n
$$

are the Ferrier polynomials.

`L-Mifflin`: $F(x) = 2(x_1^2 + x_2^2 - 1) + 1.75|x_1^2 + x_2^2 - 1|$.

The nonconvexity of the above two problems can be seen from Fig. 1.

In the test, the parameters are selected as $M = 5$, $R_0 = 10$, $\kappa = 0.3$, $\epsilon = 10^{-5}$, $\Gamma = 2$. The numerical results are reported in Tables 1, 2 and 3. The notations are: the dimension of problem $n$; the number of iterations NI; the number of descent steps ND; the number of function evaluations NF; the approximately optimal solution $x^*$; the approximately optimal objective value $F^*$. The comparisons between Algorithm 1 and PPBM (the algorithm in [27]) for the first seven problems are listed in Table 1. From Table 1, we see that Algorithm 1 performs better than PPBM. In Table 2, we compare our algorithm with RedistProx in [13] for problem `Regular` with various $n$. From Table 2, under the same conditions (terminates if NF is more than 300), we see that the approximately optimal values

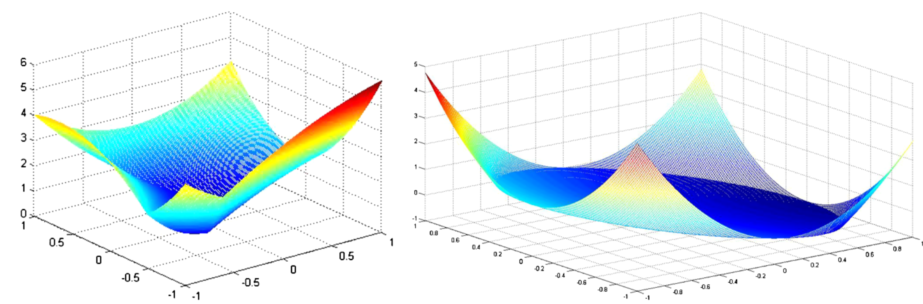

**Figure 1** The 3D images of Regular and L-Mifflin

**Table 1** Numerical results for the first set of seven problems

| Problem | $n$ | Algorithm | NI | ND | NF | $F^*$ |
|---|---|---|---|---|---|---|
| CB2 | 2 | Algorithm 1 | 2 | 1 | 2 | 3.343146 |
| | | PPBM | 2 | 1 | 2 | 3.343146 |
| CB3 | 2 | Algorithm 1 | 9 | 4 | 9 | 24.479795 |
| | | PPBM | 11 | 8 | 11 | 24.479795 |
| LQ | 2 | Algorithm 1 | 15 | 12 | 15 | −0.999989 |
| | | PPBM | 18 | 17 | 18 | −0.999996 |
| Mifflin1 | 2 | Algorithm 1 | 3 | 2 | 3 | 48.153612 |
| | | PPBM | 4 | 2 | 4 | 48.153612 |
| Rosen-Suzuki | 4 | Algorithm 1 | 7 | 6 | 7 | 39.715617 |
| | | PPBM | 9 | 8 | 9 | 39.715617 |
| Shor | 5 | Algorithm 1 | 5 | 1 | 5 | 50.250278 |
| | | PPBM | 4 | 1 | 4 | 50.250278 |
| MAXL | 20 | Algorithm 1 | 19 | 6 | 19 | 0.552786 |
| | | PPBM | 22 | 2 | 22 | 0.552786 |

**Table 2** Numerical results for "Regular" compared with RedistProx

| | ALBM | | | RedistProx | | |
|---|---|---|---|---|---|---|
| $n$ | $F^*$ | $\delta^*$ | NF | $F^*$ | $\delta^*$ | NF |
| 1 | 0.000000 | 0.000001 | 52 | 0.500000 | 0.000010 | 5 |
| 2 | 0.012188 | 0.038685 | 35 | 0.000188 | 0.002720 | 301 |
| 3 | 0.000001 | 0.000001 | 88 | 0.000006 | 0.002360 | 90 |
| 4 | 0.000000 | 0.000001 | 105 | 0.000005 | 0.000682 | 301 |
| 5 | 0.000000 | 0.000001 | 134 | 7.948708 | 1.179115 | 3 |
| 6 | 0.000001 | 0.000001 | 220 | 0.000002 | 0.000001 | 289 |
| 7 | 0.000004 | 0.000007 | 301 | 0.008876 | 0.037044 | 103 |
| 8 | 0.064289 | 0.004921 | 301 | 0.000043 | 0.000240 | 301 |
| 9 | 0.000920 | 0.003238 | 301 | 0.281491 | 0.020101 | 301 |
| 10 | 0.004182 | 0.014245 | 301 | 0.426285 | 0.056585 | 301 |

**Table 3** Numerical results for "Regular" and "L-Mifflin"

| | Regular | | L-Mifflin | |
|---|---|---|---|---|
| $x^0$ | $x^*$ | $F^*$ | $x^*$ | $F^*$ |
| [1, 1] | [0.0267, 0.0283] | 0.003103 | [−0.1362e−009, −0.1362e−009] | −0.250000 |
| [−1, −1] | [0.0208, 0.0204] | 0.001707 | [0.2353e−009, 0.2353e−009] | −0.250000 |
| [10, 10] | [0.0779, 0.1313] | 0.089998 | [0.4776e−008, 0.4776e−008] | −0.250000 |
| [−10, −10] | [0.0626, 0.0635] | 0.015955 | [−0.4817e−008, −0.4817e−008] | −0.250000 |

and accuracies of Algorithm 1 are better than RedistProx. Finally, in Table 3, we report the approximately optimal solutions and values for problem L-Mifflin with different starting points.

**Acknowledgements**
Project supported by the National Natural Science Foundation (11761013, 11771383) and Guangxi Natural Science Foundation (2013GXNSFAA019013, 2014GXNSFFA118001, 2016GXNSFDA380019) of China.

**Competing interests**
The authors declare that they have no competing interests.

**Authors' contributions**
All authors read and approved the final manuscript. CT mainly contributed to the algorithm design and convergence analysis; JL mainly contributed to the convergence analysis and numerical results; and JJ mainly contributed to the idea of the method and algorithm design.

**Author details**
[1]College of Mathematics and Information Science, Guangxi University, Nanning, P.R. China. [2]College of Science, Guangxi University of Nationalities, Nanning, P.R. China.

## Publisher's Note


Received: 2 March 2018 Accepted: 28 March 2018 Published online: 27 April 2018

**References**
1. Chartrand, R.: Exact reconstruction of sparse signals via nonconvex minimization. IEEE Signal Process. Lett. **14**(10), 707–710 (2007)
2. Hare, W., Sagastizábal, C., Solodov, M.V.: A proximal bundle method for nonsmooth nonconvex functions with inexact information. Comput. Optim. Appl. **63**(1), 1–28 (2016)
3. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka–Łojasiewicz inequality. Math. Oper. Res. **35**(2), 438–457 (2010)
4. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Program. **146**, 459–494 (2014)
5. Dinh, Q.T., Diehl, M.: Proximal methods for minimizing the sum of a convex function and a composite function (2011). arXiv:1105.0276
6. Eckstein, J., Svaiter, B.F.: A family of projective splitting methods for the sum of two maximal monotone operators. Math. Program. **111**(1), 173–199 (2007)
7. Goldfarb, D., Ma, S., Scheinberg, K.: Fast alternating linearization methods for minimizing the sum of two convex functions. Math. Program. **141**, 349–382 (2013)
8. Kiwiel, K.C.: A method for minimizing the sum of a convex function and a continuously differentiable function. J. Optim. Theory Appl. **48**(3), 437–449 (1986)
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. Li, D., Pang, L., Chen, S.: A proximal alternating linearization method for nonconvex optimization problems. Optim. Methods Softw. **29**(4), 771–785 (2014)
11. Mine, H., Fukushima, M.: A minimization method for the sum of a convex function and a continuously differentiable function. J. Optim. Theory Appl. **33**(1), 9–23 (1981)
12. Tuy, H., Tam, B.T., Dan, N.D.: Minimizing the sum of a convex function and a specially structured nonconvex function. Optimization **28**, 237–248 (1994)
13. Hare, W.L., Sagastizábal, C.: A redistributed proximal bundle method for nonconvex optimization. SIAM J. Optim. **20**(5), 2442–2473 (2010)
14. Cheney, E.W., Goldstein, A.A.: Newton's method for convex programming and Tchebycheff approximations. Numer. Math. **1**, 253–268 (1959)
15. Kelley, J.E.: The cutting-plane method for solving convex programs. J. Soc. Ind. Appl. Math. **8**, 703–712 (1960)
16. Hare, W., Sagastizábal, C.A.: Computing proximal points of nonconvex functions. Math. Program. **116**(1), 221–258 (2009)
17. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer, Berlin (1998)
18. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**(5), 877–898 (1976)
19. Lemaréchal, C.: An extension of Davidon methods to nondifferentiable problems. Math. Program. Stud. **3**, 95–109 (1975)
20. Wolfe, P.: A method of conjugate subgradients for minimizing nondifferentiable functions. Math. Program. Stud. **3**, 145–173 (1975)
21. Bonnans, J.F., Gilbert, J.C., Lemaréchal, C., Sagastizábal, C.: Numerical Optimization: Theoretical and Practical Aspects, 2nd edn. Springer, Berlin (2006)
22. Kiwiel, K.C.: Methods of Descent for Nondifferentiable Optimization. Lecture Notes in Mathematics, vol. 1133. Springer, Berlin (1985)
23. Kiwiel, K.C.: A method of centers with approximate subgradient linearizations for nonsmooth convex optimization. SIAM J. Optim. **18**(4), 1467–1489 (2008)
24. Kiwiel, K.C.: An algorithm for linearly constrained convex nondifferentiable minimization problems. J. Math. Anal. Appl. **105**(2), 452–465 (1985)
25. Lukšan, L., Vlček, J.: Test problems for nonsmooth unconstrained and linearly constrained optimization. Technical Report No. 798, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague (2000)

26. Bagirov, A.M., Karmitsa, N., Mäkelä, M.: Introduction to Nonsmooth Optimization: Theory, Practice and Software. Springer, Cham (2014)
27. Kiwiel, K.C.: A proximal-projection bundle method for Lagrangian relaxation, including semidefinite programming. SIAM J. Optim. **17**(4), 1015–1034 (2006)