Research Article

A Cutting Plane and Level Stabilization Bundle Method with Inexact Data for Minimizing Nonsmooth Nonconvex Functions

Jie Shen,1 Dan Li,2 and Li-Ping Pang2

1 School of Mathematics, Liaoning Normal University, Dalian 116029, China
2 School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China

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

Received 12 January 2014; Accepted 20 March 2014; Published 14 April 2014

Academic Editor: Gaohang Yu

Copyright © 2014 Jie Shen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Under the condition that the values of the objective function and its subgradient are computed approximately, we introduce a cutting plane and level bundle method for minimizing nonsmooth nonconvex functions by combining cutting plane method with the ideas of proximity control and level constraint. The proposed algorithm is based on the construction of both a lower and an upper polyhedral approximation model to the objective function and calculates new iteration points by solving a subproblem in which the model is employed not only in the objective function but also in the constraints. Compared with other proximal bundle methods, the new variant updates the lower bound of the optimal value, providing an additional useful stopping test based on the optimality gap. Another merit is that our algorithm makes a distinction between affine pieces that exhibit a convex or a concave behavior relative to the current iterate. Convergence to some kind of stationarity point is proved under some looser conditions.

1. Introduction

Bundle methods family is based on the cutting plane method, first described in [1, 2], where the convexity of the objective function is the fundamental assumption. The extension of bundle methods to the nonconvex case is not straightforward; however, it is apparent that a number of ideas valid in the convex framework are also valuable in the treatment of the nonconvex case. In [3], by combining cutting plane method with proximity control, for minimizing nonsmooth nonconvex function, the authors propose an iterative algorithm which makes a distinction between affine pieces that exhibit a convex or a concave behavior relative to the current point in the iterative procedure, but the exact information of the objective function and its subgradient is needed.

Level bundle method is another method which is easier to implement and has shown encouraging numerical results [4, 5]. Some evidence shows that constrained level bundle methods are preferable under certain conditions. Also, strategies for updating the level parameter $l_k$ are readily available [6].

In various real-world applications, the objective function and/or its subgradient can be costly (sometimes impossible) to compute. This is particularly true when $f$ is given by some optimization problem; for example, $f(x) = \max_{u \in U} \phi(x, u)$. In such situations, approximate values must be used. Various inexact bundle methods that use approximate functions values and subgradient evaluations have been studied [7].

In this work, we combine the level bundle method with cutting plane method involving the idea of proximal control by employing inexact objective function information for minimizing a nonconvex function. The proposed algorithm makes a distinction between affine pieces that exhibit a convex or a concave behavior relative to the current iteration point such that the downward shifting of the affine pieces is not arbitrary. The rest of this paper is organized as follows. Section 2 gives some preliminary results and introduces the basic idea of the proposed algorithm. Section 3 describes formally the cutting plane and level stabilization bundle method with inexact data for minimizing nonsmooth nonconvex functions. Section 4 is devoted to convergence analysis of the
2 Abstract and Applied Analysis

Algorithm with looser conditions than [3]. Section 5 contains some conclusions of the proposed method.

2. The Construction of the Model

In this paper, we consider the following unconstrained minimization problem:

$$\min \quad f(x)$$
$$\text{s.t.} \quad x \in \mathbb{R}^n,$$

where $f: \mathbb{R}^n \to \mathbb{R}$ is not necessarily differentiable. We assume that $f$ is locally Lipschitz; then $f$ is differentiable almost everywhere. It is well known that, in [8], under the above hypotheses, there is defined at each point $x$ the generalized gradient (Clarke's gradient):

$$\partial f(x) = \{ y \in \mathbb{R}^n \mid \forall f(y) \to g, x^k \to x, x^k \notin \Omega \},$$

where $\Omega_f$ is the set, where $f$ is not differentiable, and $\partial f(x)$ is locally bounded. An extension of the generalized gradient is the Goldstein $\varepsilon$-subdifferential $\partial^\varepsilon f(x)$ defined as

$$\partial^\varepsilon f(x) = \{ y \in \mathbb{R}^n \mid f(y) - f(x) \leq \varepsilon \},$$

which is first described by Goldstein in [9]. The Goldstein subdifferential is both outer and inner semicontinuous as a multifunction of $x$ and $\varepsilon$.

Throughout the paper, we assume that, given $x \in \mathbb{R}^n$, the oracle provides some approximate values $f_x \in \mathbb{R}$ and $g^G_x \in \mathbb{R}^n$ of the objective function and its subgradient, respectively, such that

$$f_x = f(x) - \eta_x,$$
$$g^G_x \in \partial^\varepsilon f(x) = \{ y \in \mathbb{R}^n \mid \forall f(y) \to g, x^k \to x, x^k \notin \Omega \}.$$  \hspace{1cm} (4)

where $\eta_x \in \mathbb{R}$ and $\eta_x^G \geq 0$ are some unknown errors. Specifically, $\eta_x$ is uniformly bounded; that is, there exists $\eta \geq 0$ such that

$$|\eta_x| \leq \eta, \quad \forall x \in \mathbb{R}^n.$$  \hspace{1cm} (5)

Let $\bar{s}_x$ be the current stability center; the bundle $I_k$ of available information is the set of elements:

$$(x_i, f_{x_i}, g^G_{x_i}, x_i, \alpha_i, \tilde{a}_i), \quad i \in I_k,$$  \hspace{1cm} (6)

where $\alpha_i$ is the linearization error defined by

$$\alpha_i = f_{x_i} - f_{\bar{s}_x} - \langle g^G_{x_i}, \bar{s}_x - x_i \rangle.$$  \hspace{1cm} (7)

and $\tilde{a}_i = \|\tilde{s}_x - x_i\|$. With this information, we create the linearization

$$\bar{f}_k(x) = f_{x_k} + \langle g^G_{x_k}, x - x_k \rangle.$$  \hspace{1cm} (8)

At iteration $k$, a polyhedral cutting plane model of $f$ is available:

$$\bar{f}_k(x) = \max_{i \in I_k} \{ f_{x_i} + \langle g^G_{x_i}, x - x_i \rangle \}.$$  \hspace{1cm} (9)

Suppose there exists $\beta_k > 0$ such that $\bar{f}_k(x) \leq f(x) + \beta_k$, $\forall x \in \mathbb{R}^n$, and $\beta_k$ is uniformly bounded; that is, there exists $\beta > 0$ such that $\beta_k \leq \beta$, $\forall k$. We recall that the classical cutting plane methods [1, 2] minimize $f_k(x)$ at each iteration, and the minimization of $\bar{f}_k(x)$ can be written in linear programming form:

$$\min \quad \bar{v}$$
$$\text{s.t.} \quad \bar{v} \geq f_{x_i} + \langle g^G_{x_i}, x - x_i \rangle, \quad i \in I_k,$$

which is equivalent to solving

$$\min \quad v := \bar{v} - 2\eta$$
$$\text{s.t.} \quad v \geq f_{x_k} + \langle g^G_{x_k}, x - x_k \rangle - 2\eta, \quad i \in I_k,$$

where $d = x - \bar{s}_k$. We divide the set $I_k$ into two sets $I_k^+$ and $I_k^-$ defined as follows:

$$I_k^+ = \{ i \in I_k | \alpha_i \geq -2\eta \}, \quad I_k^- = \{ i \in I_k | \alpha_i < -2\eta \}.$$  \hspace{1cm} (12)

We observe that $I_k^+$ is never empty since $k \in I_k^+$. By using $I_k^+$ and $I_k^-$, we define two piecewise affine functions:

$$\Delta^+(d) = \max_{i \in I_k^+} \{ \langle g^G_{x_i}, d - \alpha_i \rangle \},$$
$$\Delta^-(d) = \min_{i \in I_k^-} \{ \langle g^G_{x_i}, d - \alpha_i \rangle \}.$$  \hspace{1cm} (13)

In fact, $\Delta^-(d)$ can be regarded as an approximation of the difference function $h(d) = f_{x_k + d} - f_{x_k}$, because $\Delta^+(0) \leq 2\eta$ and $\Delta^-(0) \geq 2\eta$, thus $\Delta^+(0) < \Delta^-(0)$; it is reasonable to consider the approximation $\Delta^+(d)$ significant as far as $\Delta^+(d) \leq \Delta^-(d)$. Therefore we introduce a kind of trust model:

$$S_f = \{ d \in \mathbb{R}^n | \Delta^+(d) \leq \Delta^-(d) \}.$$  \hspace{1cm} (14)

Let $\nu_k^f$ be a nonnegative scalar representing how much we aim to reduce the value $f_{x_k}$ at the current iteration. Define the corresponding level parameter $l_k = f_{x_k} - \nu_k^f$. Since $f(x) > 0$, $l_k > 0$. Then the level set associated with $\bar{f}_k(c)$ and $l_k$ is given by

$$X_k = \{ x \in \mathbb{R}^n | \bar{f}_k(x) \leq l_k \}.$$  \hspace{1cm} (15)

We obtain the search direction by solving the following convex quadratic subproblem, parameterized in the nonnegative scalar $u$, where the first two constraints ensure $d \in S_f$ and the last constraint represents the idea of level bundle method:

$$\begin{align*}
Q(P_u) = \min & \quad z_u = \min \langle uv, \frac{1}{2} \|d\|^2 \rangle \\
\text{s.t.} & \quad v \geq \langle g^G_{x_k}, d - \alpha_i - 2\eta \rangle, \quad i \in I_k^+,
& \quad v \leq \langle g^G_{x_k}, d - \alpha_i - 2\eta \rangle, \quad i \in I_k^-,
& \quad v \leq l_k.
\end{align*}$$  \hspace{1cm} (16)
Observe that \( z_u \leq 0 \) since \( (v, d) = (0, 0) \) is feasible. We have consequently that the optimal value \( v \) cannot be positive. The dual problem of \( QP(u) \) can be written in the form

\[
\begin{align*}
\text{DP} (u) & \quad \min \frac{1}{2} \| G_r \lambda - G_d \mu \|^2 + \alpha^T \lambda - \alpha^T \mu + \xi_k \\
& \quad \text{s.t.} \quad \lambda, \mu, \xi \geq 0, \quad e^T \lambda - e^T \mu - \xi = u,
\end{align*}
\]

where \( G_r \) and \( G_d \) are matrices whose columns are, respectively, the vectors \( g_{x_i}^r, i \in I_k^r \) and \( g_{x_i}^d, i \in I_k^d \) and \( \alpha_r \) and \( \alpha_d \) are vectors whose components are \( \alpha_i, i \in I_k^r \) and \( \alpha_i, i \in I_k^d \), respectively. \( \lambda \) and \( \mu \) are the vectors with components \( \lambda_i, i \in I_k^r \), \( \mu_i, i \in I_k^d \), and \( e = (1, 1, \ldots, 1)^T, \xi \in R^7 \). The primal optimal solution \( (\nu_u, d_u, \xi_u) \) is related to the dual optimal solution \( (\lambda, \mu, \xi) \) by the following formulas:

\[
d_u = -G_r \lambda + G_d \mu, \\
\nu_u = -\frac{1}{u} \| d_u \|^2 + (\alpha + 2 \eta e)^T \lambda - (\alpha - 2 \eta e)^T \mu + \xi dk.
\]

Before giving a description of the algorithm, we state some simple properties of \( QP(u) \).

**Lemma 1.** Let \( 0 < u_2 < u_1 \); then the following conclusions hold:

(i) \( z_u \leq z_{u_2} \);
(ii) \( v_u \leq v_{u_2} \);
(iii) \( \| d_u \| \geq \| d_{u_2} \| \).

**Proof.** We omit the proof since the conclusions can be obtained by imitating the proof of Lemma 2.1 [3]. \( \square \)

**Lemma 2.** For any \( u > 0 \), the following conclusions hold:

(i) \( \| d_u \| \leq 2u \| g_{x_0}^r \| + 2u \eta ;
(ii) \( z_u \geq -(1/2)u \| g_{x_0}^d \|^2 - 2u \eta ;
(iii) \| v_u \| \geq (1/2u) \| d_u \|^2 .

**Proof.** The conclusions can be obtained by imitating the proof of Lemma 2.2 [3]. \( \square \)

### 3. Algorithm

In this section, we state the algorithm in full detail and give some comments on it.

**Algorithm 3.** We have the following.

**Step 0** (initialization). The following global parameters are to be set: the stopping tolerances \( \delta > 0, \delta_0 \geq 0, \) the proximity measure \( \epsilon \geq 0, \) the descent parameter \( m \in (0, 1), \) the cut parameter \( \rho \in (m, 1), \) the reduction parameter \( r \in (0, 1), \) the increase parameter \( R > 1, \) and the level measure parameter \( m_i \in (0, 1) \). Choose a starting point \( x_0 \in R^n; \) set \( \bar{x}_0 = x_0; \) the oracle provides \( f_{x_0} \in R \) and \( g_{x_0} \in R^n \) satisfying Assumption (4). The initial bundle is made up of just one element \( (x_0, f_{x_0}, g_{x_0}, 0, 0) \) so that \( I^l_0 \) is empty, while \( I^l_0 \) is a singleton. Since \( f(x) > 0, \) a lower bound \( f_{x_0}^\text{low} \) for \( f(x) \in \max \{ f(x) \mid x \in R^n \} \) is available; set \( v_0 = (1 - m)(f_{x_0} - f_{x_0}^\text{low} \) . Set the iteration counter \( k = 0 \).

**Step 1** (first stopping test). If \( \| g_{x_k}^r \| \leq \delta, \) then terminate.

**Step 2** (second stopping test). Set \( \Delta_k = f_{x_k} - f_{x_k}^\text{low} \). If \( \Delta_k \leq T \delta \), then stop.

**Step 3** (level feasibility checking). Set the level parameter \( l_k = f_{x_k} - v_k \). If the level set \( X_k \) defined by (15) is detected to be empty, then set \( f_{x_k}^\text{low} = l_k, v_k = (1 - m)(f_{x_k} - f_{x_k}^\text{low} \) and go back to Step 2. Otherwise set \( u_k = e/2 \frac{1}{\| g_{x_k}^r \|^2}, \) \( v_k = R \| u \|^2, \) and \( g_k = r \delta u_k \).

**Step 4** (direction finding). Find the solution \( (d_k^r, v_k^d) \) of \( QP(\bar{u}) \) for increasing value of \( u_k \), such that

\[
f_{x_k} + d_k^r > f_{x_k} + mv_k^d,
\]

where \( \bar{u} \) equals the minimum value of \( u \in [u_{\text{min}}^k, u_{\text{max}}^k] \) if such \( u \) does exist; otherwise set \( \bar{u} = u_{\text{max}}^k \).

**Step 5** (bundle updating). Set

\[
I^l_k = I^l_k \setminus \{ i \in I^l_k | a_i > \epsilon \}, \quad I^l_k = I^l_k \setminus \{ i \in I^l_k | a_i > \epsilon \}.
\]

**Calculate**

\[
g^* = \text{arg min} \{ \| g \| \mid g \in \text{conv} \{ g_{x_i}^r | i \in I^l_k \} \}.
\]

If \( \| g^* \| \leq \delta, \) then terminate.

Else set \( u_{\text{max}}^k = u_{\text{max}}^k - r(u_{\text{max}}^k - u_{\text{min}}^k) \) and go to Step 4.

**Step 6** (trial point calculating). Set \( x_{k+1} = x_k + d_k^r \), calculate \( f_{x_{k+1}}^r, g_{x_{k+1}}^r \in g_{x_k}^r f_{x_{k+1}}^r + (\alpha_{x_{k+1}}^r)^T d_k^r \), and set \( \alpha_{x_{k+1}} = f_{x_k} - f_{x_{k+1}} + (\alpha_{x_{k+1}}^r)^T d_k^r \).

**Step 7** (index insertion). (a) If \( \alpha_{x_{k+1}} < -2 \eta \) and \( \| d_k^r \| > \epsilon, \) then insert the element \( (x_{k+1}, f_{x_{k+1}}, g_{x_{k+1}}, \alpha_{x_{k+1}}^r, \| d_k^r \| ) \) into the bundle for an appropriate value of \( i \in I^l_k \) and set \( \bar{u} = \bar{u} - r(\bar{u} - u_{\text{min}}^k) \).

(b) Else, if \( (\alpha_{x_{k+1}}^r)^T d_k^r \geq \rho v_k^d - 2 \eta, \) then insert the element \( (x_{k+1}, f_{x_{k+1}}, g_{x_{k+1}}, \max(\alpha_{x_{k+1}}^r - 2 \eta), \| d_k^r \| ) \) into the bundle for an appropriate value of \( i \in I^l_k \).

(c) Else find a scalar \( t \in (0, 1) \) such that \( g(t) = \alpha_{x_{k+1}}^r f_{x_{k+1}} + t d_k^r \) satisfies the condition \( g(t)^T d_k^r \geq \rho v_k^d - 2 \eta \) and insert the element \( (\bar{x}_k + t d_k^r, f_{\bar{x}_k + t d_k^r}, g(t), \max(-2 \eta, \alpha_i), t \| d_k^r \| ) \) into the bundle for an appropriate value of \( i \in I^l_k \), where \( \alpha_i = f_{\bar{x}_k} - f_{\bar{x}_k + t d_k^r} + t g(t)^T d_k^r \).
Step 8 (descent test). If \( \|d_k^p\| \leq \theta \), go to Step 5. Choose \( f_{k+1}^{low} \in [f_k^{low}, f_k^{inf}] \).

If

\[
f_{\hat{x}_{k+1}} \leq f_{\hat{x}} + m \eta^k,
\]

set the new stability center \( \hat{x}_{k+1} = x_{k+1}, v_k^{\prime} = (1-m)(f_{\hat{x}_{k+1}} - f_k^{low}), \) set \( k = k + 1 \), and go to Step 1 (serious step).

Otherwise set \( \hat{x}_{k+1} = \hat{x}_k, v_k^{\prime} = (1-m)(f_{\hat{x}_{k+1}} - f_k^{low}), \) set \( k = k + 1 \), and go to Step 9 (null step).

Step 9 (resolving QU(u) or DP(u)). Solve QP(u) or, equivalently, DP(u), to obtain the primal and dual optimal solutions \( (v_k, d_k) \) and \( (\lambda_k, \mu_k, \xi_k) \) and go to Step 6.

That is the end of the algorithm.

A few comments on the algorithm are in the following order.

(a) Step 1 is justified by the optimality estimate.

(b) Step 2 is another stopping criterion. \( X_k = \emptyset \) means that \( l_k \leq f_k(x) \leq f(x) + \beta, \forall x \in R^n \), and since the update sets \( f_k = l_k \), it holds that \( f_k^{low} \leq f_k^{inf} + \beta \) for all \( k \). If Algorithm 3 stops at Step 2, we have that

\[
Tol_k \geq f_{\hat{x}} - f_{\hat{x}}^{low} \geq f_{\hat{x}}^{inf} - f_{\hat{x}}^{inf} - \beta = f(\hat{x}_k) - \eta \xi_k - f_{\hat{x}}^{inf} - \beta
\]

which means that \( \hat{x}_k \) is a \((Tol_k + \eta + \beta)\)-approximate solution to problem (1). If at Step 2 the rule \( l_k = f_{\hat{x}} - v_k^{\prime} \) is replaced by \( l_k = +\infty \) for all \( k \), Algorithm 3 becomes an approximate proximal bundle algorithm.

(c) Step 4 may use the dual quadratic programming method of [10], which can efficiently solve sequences of related subproblems QP(u) with varying \( u \). The construction of \( d_k^G \) at Step 4 may be discretized by repeatedly solving QP(u) for increasing values of \( u \) or by adopting techniques of the type described in [11].

(d) \( d_k^u \leq \theta \) is never a consequence of the choice of too small \( \bar{u} \). In fact, we note that if \( \|\bar{d}_{\bar{s}}\| > \delta \), it holds that

\[
\|\bar{d}_{\bar{s}}\| \leq 2\mu_{\min} \|\bar{d}_{\bar{s}}^G\| + 2\mu_{\min} \eta \frac{2\theta (\|\bar{d}_{\bar{s}}^G\| + \eta)}{r\delta},
\]

The right-hand side of the above inequality is bigger than \( \theta \), so too small \( \bar{u} \) cannot lead to \( d_k^u \leq \theta \); therefore we need to increase the value of \( u \).

(e) We remark that the insertion of a bundle index into \( I_k^r \) or \( I_k^l \) at Step 7 is not simply based on the sign of \( \xi_k \).

(f) The level constraint \( f_k(x) \leq l_k \) provides an additional useful stopping test based on a certain optimality gap, something not present in cutting plane methods. It should be noted that this additional stopping test is useful, as usual proximal bundle methods sometimes “take time” to accumulate enough information to recognize that an acceptable approximate solution has been already computed.

(g) To keep the size of problem (16) manageable, the number of elements in bundle should be kept bounded, without impairing convergence. For this, the usual aggregation techniques of proximal bundle methods can be employed here. Specifically, the model can be composed of as few as only two cutting planes, corresponding to the new linearization and the aggregate linearization.

4. Convergence Analysis

Convergence analysis of Algorithm 3 has to account for all the following cases:

- the level sets \( X_k \) are empty for infinitely many times;
- infinitely many descent steps are generated;
- finitely many descent steps are generated.

For the first case (26), if the stopping tolerance \( Tol_k \) is positive, then the method terminates with an approximate solution in a finite number of iterations.

**Lemma 4.** Suppose the level sets \( X_k \) are empty for infinitely many times; then \( \Delta_k \to 0 \):

\[
\lim_{k \to \infty} f_{\hat{x}} \leq f^{inf} + \beta.
\]

And every cluster point of the sequence \( \{\hat{x}_k\} \) (if any exists) is a \((\eta + \beta)\)-approximate solution to problem (1).

**Proof.** Since \( f_k^{low} > -\infty, \Delta_k < +\infty \). Also, by Steps 3 and 8, \( v_k^{\prime} \leq (1-m)\Delta_k \); thus

\[
f_{\hat{x}} - l_k = f_{\hat{x}} - (f_{\hat{x}} - v_k^{\prime}) = v_k^{\prime} \leq (1-m)\Delta_k,
\]

which means that if \( X_k \) is empty at iteration \( k \), then the update decreases the optimality gap \( \Delta_k \) by a factor of at least \( (1-m) \). Hence if this happens infinitely many times, we have \( \Delta_k \to 0 \) as \( k \to \infty \). Moreover, we have \( f_{\hat{x}} - f_{\hat{x}}^{low} \geq f_{\hat{x}}^{inf} - \beta \), which means that \( f_{\hat{x}} \) is decreasing and bounded below \((f_{\hat{x}} = f(\hat{x}_k) - \eta \xi_k > f(\hat{x}) - \eta > f^{inf} - \eta > 0)\), we conclude that

\[
\lim_{k \to \infty} f_{\hat{x}} - f^{inf} - \beta \leq \lim_{k \to \infty} f_{\hat{x}} - f^{low} = \lim_{k \to \infty} \Delta_k = 0,
\]

which gives (29). Let \( \bar{x} \) be any cluster point of \( \{\hat{x}_k\} \), and let \( \{\hat{x}_{k(j)}\} \) be a subsequence converging to \( \bar{x} \) as \( j \to \infty \). Then

\[
f^{inf} + \beta \geq \lim_{j \to \infty} f_{\hat{x}_{k(j)}} = \lim_{j \to \infty} \left( f(\hat{x}_{k(j)}) - \eta \xi_{k(j)} \right) \geq f(\bar{x}) - \eta,
\]

which establishes the last conclusion. \( \square \)

From now on, we consider the case when \( X_k \neq \emptyset \) for \( k \) large enough. Without loss of generality, we can simply assume \( X_k \neq \emptyset \) for all \( k \). For cases (27) and (28) we make the following assumptions:

(A1) the set \( S_0 = \{x \in R^n | f_x \leq f_{x_0}\} \) is compact;
Abstract and Applied Analysis

(A2) for any \( d \in \mathbb{R}^n \), the directional directive \( f'(x; d) \) of \( f \) at \( x \) exists and
\[
f'(x; d) = \lim_{t \to 0} \frac{g(t)d}{t}
\]
where \( g(t) \in d_{\text{grad}} f(x + td) \).

Remark 5. Assumption (A2) is a looser condition than the one emerging in [3], where the condition that \( f \) is weakly semismooth is required.

Next we introduce Lemma 6 to see what will happen if no really new stability center is generated; that is, there exists the last stability center and from then on only null steps are generated.

Lemma 6. Suppose \( \bar{x}_k \) is the last stability center followed by a null step sequence \( (d_n^i, v_n^i) \) such that \( \|d_n^i\| > \theta \) and
\[
f_{\bar{x}_k+d_n^i} - f_{\bar{x}_k} > m_n^i
\]
with the algorithm looping between Steps 6 and 9. Then the following conclusions hold.

(i) There exists an index \( j \) such that, for each \( j \geq \bar{j} \), every new bundle index is inserted into \( I_k \) and \( \bar{u} \) remains unchanged.

(ii) Step 7(c) is well defined.

(iii) Whenever a new bundle index is inserted into \( I_k^* \), the condition \( (g_j^i)^T d_n^i \geq \rho v_n^i - 2\eta \) holds, where \( d_n^i \) is the subgradient corresponding to the new bundle element.

Proof. (i) The conclusion can be obtained by noting that no bundle index can be inserted into \( I_k \) as soon as \( \bar{u} \) falls below the threshold \( \varepsilon/(2\|g_{x_k}\| + 2\eta) \).

(ii) Since the sufficient decrease condition (23) is not satisfied, it follows that
\[
\rho v_n^i < m_n^i < f_{\bar{x}_k+d_n^i} - f_{\bar{x}_k}.
\]

According to the inexact data in Assumption (4), \( f_{\bar{x}_k+d_n^i} - f_{\bar{x}_k} \leq f(\bar{x}_k + d_n^i) - f(\bar{x}_k) + 2\eta \), by mean value theorem there exists a scalar \( \bar{r}_j^i \in (0, 1) \) such that
\[
\rho v_n^i < f'\left(\bar{x}_k + \bar{r}_j^i d_n^i; d_n^i\right) + 2\eta.
\]

Thus the conclusion follows from Assumption (A2).

(iii) Observe that the condition \((g_j^i)^T d_n^i \geq \rho v_n^i - 2\eta\) is ensured either by construction or by the fact that
\[
(g_j^i)^T d_n^i + 2\eta \geq (g_j^i)^T d_n^i - \alpha_n^i = f_{\bar{x}_k+d_n^i} - f_{\bar{x}_k} > m_n^i > \rho v_n^i,
\]
whenever \( \alpha_n^i \geq -2\eta \).

Finally we consider the case of infinitely many descent steps; we prove that either an approximate solution is achieved at Step 5 or we obtain a really new stability center \( \bar{x}_{k+1} \neq \bar{x}_k \).

Lemma 7. For Algorithm 3, if infinitely many descent steps are generated one either obtains an approximate solution at Step 5 or one executes descent step at Step 8.

Proof. Firstly we prove that the algorithm cannot pass through Step 5 for infinitely many times. Just like the result of Lemma 4.2 in [3], we can obtain that the indices of the new bundle elements are inserted into \( I_k^* \) and are never removed. Moreover, when a passage at Step 5 occurs all the elements with index \( i \in I_k^* \) are removed. Taking into account (18) and the constraint \( e_j^i \lambda - e_j^i \mu - \xi = \bar{u} \) in the dual problem \( DP(\bar{u}) \), there exists an index \( j \) such that, for all \( j \geq j^*, d_n^i \) can be expressed in the form (say \( |I_k^*| = m \))
\[
d_n^i = -G_i \lambda^i = -\sum_{i \in I_k^*} \lambda_i g_i^x_i
\]
\[
= -\sum_{i \in I_k^*} \lambda_i \left( \frac{1}{\sum_{i \in I_k^*} \lambda_i} g_i^x_1 + \cdots + \frac{1}{\sum_{i \in I_k^*} \lambda_i} g_i^x_m \right)
\]
\[
= -(\bar{u} + \xi) g_j^i,
\]
where \( g_j^i \in \text{conv}\{g_i^x_1 | i \in I_k^*\} \). But since \( \|d_n^i\| \leq \theta \) and \( \|g_j^*\| \geq 0 \), we have
\[
\theta \geq \|d_n^i\| = |\bar{u} + \xi| \|g_j^i\| \geq u_{\min} \|g_j^*\| > \frac{\theta}{r_0} > \theta,
\]
which leads to a contradiction.

Next we show that it is impossible to have \( d_n^i = \theta \) for infinitely many times and the descent condition (23) is not satisfied with the algorithm looping between Steps 6 and 9. Indexing by \( j \in J \) the \( j \)th passage through such a loop, we observe that, by Lemma 6 (i), there exists an index \( j^* \) such that, for every \( j \geq j^* \), the sequence \( \{z_n^j\} \) is nondecreasing, bounded, and hence convergent. Moreover, \( |d_n^i| \) is bounded; it admits a convergent subsequence, say \( \{d_n^i\}_{j \in J^*} \). The above consideration implies also that \( \{v_n^j\}_{j \in J^*} \) is convergent to a nonpositive limit, say \( \bar{v} \). Next, we can imitate the proof of Lemma 4.2 in [3] to show that \( \bar{v} = 0 \), which, by Lemma 2 (iii), contradicts the fact that \( \|d_n^i\| > \theta \).

Finally we show after a finite number of descent steps that Algorithm 3 stops at a point satisfying the condition of some kind of approximate solution.

Theorem 8. For any \( e \geq 0 \) and \( \delta > 0 \), Algorithm 3 stops in a finite number of iterations at a point satisfying the approximate stationarity condition
\[
\|g^*\| \leq \delta
\]
with \( g^* \in \text{conv}\{\tilde{g}_{n_i}^G f(y) | \|y - \bar{x}_k\| \leq e\} \).

Proof. Suppose that the conclusion does not hold. It follows from Lemma 7 that for infinitely many times the descent
condition (23) is satisfied. Let $\tilde{x}_k$ be the stability center at the $k$th passage; then $\|d_k\| > \theta$ and

$$\begin{align*}
f_{\tilde{x}_{k+1}} & \leq f_{\tilde{x}_k} + m v_k^k, \\
f_{\tilde{x}_{k+1}} - f_{x_0} & \leq m \sum_{i=0}^{k} v_i.
\end{align*} \tag{41}$$

Since $u^k_{\min} = r \frac{\|\tilde{x}_k\|}{2}$ and $\theta^{(k)} = r \theta u^k_{\min}$, so $\theta^{(k)} \geq r^2 \varepsilon \delta / 2 L_0$, where $L_0$ is the Lipschitz constant of $f$ on $S_0$. It follows from $\|d_k\| > \theta^{(k)} \geq r^2 \varepsilon \delta / 2 L_0$ that $\|d_k\|$ is bounded away from zero. Then from Lemma 2(iii), $v^k_0$ is bounded away from zero as well. Therefore by passing to the limit we obtain

$$\lim_{k \to \infty} f_{\tilde{x}_{k+1}} - f_{\tilde{x}_k} \leq -\infty. \tag{42}$$

Since $f_{\tilde{x}_{k+1}} = f(\tilde{x}_{k+1}) - \eta_{\tilde{x}_{k+1}} \leq f(\tilde{x}_{k+1}) + \eta$, $f_{\tilde{x}_k} = f(\tilde{x}_0) - \eta_{\tilde{x}_0} \leq f(\tilde{x}_0) + \eta$, we have $\lim_{k \to \infty} f(\tilde{x}_{k+1}) = -\infty$, which contradicts the fact that the value of $f$ at $x \in R^n$ is positive.

5. Conclusions

In this paper, we propose a new algorithm for nonsmooth nonconvex minimization by employing the approximate values of the objective function and its subgradient. It combines the cutting plane method, level bundle method, and the idea of proximal control. The aim is to take advantage of good properties of all above-mentioned methods, thus speeding up the optimization process. In addition, the algorithm provides a useful stopping test based on the optimality gap, something not present in the proximal bundle methods. Compared with bundle methods for nonsmooth nonconvex functions, the amount of the shifting of affine pieces appears somehow arbitrary, but in our paper, the use of downward shifting is restricted to some particular cases.

Conflict of Interests

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

Acknowledgments

The authors would like to acknowledge the valuable suggestions and helpful comments from the referees. This research was supported by the National Natural Science Foundation of China (Grants 11301246, 11171049, and 11171138).

References


Submit your manuscripts at http://www.hindawi.com