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 Pang[2]

[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

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 non-convex 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

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:

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

where $f: R^n \to R^+$ 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):

$$\partial f(x) = \text{conv}\left\{ g \in R^n \mid \nabla f\left(x^k\right) \to g, \ x^k \to x, \ x^k \notin \Omega_f \right\}, \tag{2}$$

where Ω_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 ε-subdifferential $\partial_\varepsilon^G f(x)$ defined as

$$\partial_\varepsilon^G f(x) = \text{conv}\left\{ \partial f(y) \mid \|y - x\| \le \varepsilon \right\}, \tag{3}$$

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

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

$$f_x = f(x) - \eta_x,$$
$$g_x^G \in \partial_{\eta_x^g}^G f(x) = \text{conv}\left\{ \partial f(y) \mid \|y - x\| \le \eta_x^g \right\}, \tag{4}$$

where $\eta_x \in R$ and $\eta_x^g \ge 0$ are some unknown errors. Specifically, η_x is uniformly bounded; that is, there exists $\eta \ge 0$ such that

$$|\eta_x| \le \eta, \quad \forall x \in R^n. \tag{5}$$

Let $\hat{x}_k$ be the current stability center; the bundle I_k of available information is the set of elements:

$$\left(x_i, f_{x_i}, g_{x_i}^G, \alpha_i, a_i\right), \quad i \in I_k, \tag{6}$$

where α_i is the linearization error defined by

$$\alpha_i = f_{\hat{x}_k} - f_{x_i} - \left\langle g_{x_i}^G, \hat{x}_k - x_i \right\rangle \tag{7}$$

and $a_i = \|\hat{x}_k - x_i\|$. With this information, we create the linearization

$$\overline{f}_i(x) = f_{x_i} + \left\langle g_{x_i}^G, x - x_i \right\rangle. \tag{8}$$

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

$$\check{f}_k(x) = \max_{i \in I_k}\left\{ f_{x_i} + \left\langle g_{x_i}^G, x - x_i \right\rangle \right\}. \tag{9}$$

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

$$\min \quad \tilde{v}$$
$$\text{s.t.} \quad \tilde{v} \ge f_{x_i} + \left\langle g_{x_i}^G, x - x_i \right\rangle, \quad i \in I_k, \tag{10}$$

which is equivalent to solving

$$\min \quad v := \tilde{v} - 2\eta$$
$$\text{s.t.} \quad v := \tilde{v} - 2\eta \ge \left(g_{x_i}^G\right)^T d - \alpha_i - 2\eta, \quad i \in I_k, \tag{11}$$

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

$$I_k^+ = \{i \in I_k \mid \alpha_i \ge -2\eta\}, \qquad I_k^- = \{i \in I_k \mid \alpha_i < -2\eta\}. \tag{12}$$

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:

$$\Delta^+(d) = \max_{i \in I_k^+}\left\{ \left(g_{x_i}^G\right)^T d - \alpha_i \right\},$$
$$\Delta^-(d) = \min_{i \in I_k^-}\left\{ \left(g_{x_i}^G\right)^T d - \alpha_i \right\}. \tag{13}$$

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

$$S_f = \{d \in R^n \mid \Delta^+(d) \le \Delta^-(d)\}. \tag{14}$$

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

$$X_k = \left\{ x \in R^n \mid \check{f}_k(x) \le l_k \right\}. \tag{15}$$

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:

$$QP(u) \begin{cases} z_u = \min & uv + \frac{1}{2}\|d\|^2 \\ \text{s.t.} & v \ge \left(g_{x_i}^G\right)^T d - \alpha_i - 2\eta, \quad i \in I_k^+, \\ & v \le \left(g_{x_i}^G\right)^T d - \alpha_i - 2\eta, \quad i \in I_k^-, \\ & v \le l_k. \end{cases} \tag{16}$$

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

$$DP(u) \begin{cases} \min & \frac{1}{2}\|G_+\lambda - G_-\mu\|^2 + \alpha_+^T\lambda - \alpha_-^T\mu + \xi l_k \\ & + 2\eta\left(e^T\lambda - e^T\mu\right) \\ \text{s.t.} & \lambda, \mu, \xi \geq 0, \\ & e^T\lambda - e^T\mu - \xi = u, \end{cases} \tag{17}$$

where G_+ and G_- are matrices whose columns are, respectively, the vectors $g_{x_i}^G, i \in I_k^+$ and $g_{x_i}^G, i \in I_k^-$ and α_+ and α_- are vectors whose components are $\alpha_i, i \in I_k^+$ and $\alpha_i, i \in I_k^-$, respectively. λ and μ are the vectors with components $\lambda_i, i \in I_k^+, \mu_i, i \in I_k^-$, and $e = (1, 1, \ldots, 1)^T, \xi \in R^+$. The primal optimal solution (v_u, d_u) is related to the dual optimal solution $(\lambda_u, \mu_u, \xi_u)$ by the following formulae:

$$d_u = -G_+\lambda_u + G_-\mu_u, \tag{18}$$

$$v_u = -\frac{1}{u}\left[\|d_u\|^2 + (\alpha_+ + 2\eta e)^T\lambda_u - (\alpha_- + 2\eta e)^T\mu_u + \xi_u l_k\right]. \tag{19}$$

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) $z_{u_1} \leq z_{u_2}$;

(ii) $v_{u_1} \leq v_{u_2}$;

(iii) $\|d_{u_1}\| \geq \|d_{u_2}\|$.

Proof. We omit the proof since the conclusions can be obtained by imitating the proof of Lemma 2.1 [3]. □

Lemma 2. *For any* $u > 0$, *the following conclusions hold:*

(i) $\|d_u\| \leq 2u\|g_{\hat{x}_k}^G\| + 2u\eta$;

(ii) $z_u \geq -(1/2)u^2\|g_{\hat{x}_k}^G\|^2 - 2u\eta$;

(iii) $|v_u| \geq (1/2u)\|d_u\|^2$.

Proof. The conclusions can be obtained by imitating the proof of Lemma 2.2 [3]. □

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 $\delta > 0$, $\text{Tol}_\Delta \geq 0$, the proximity measure $\varepsilon \geq 0$, 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

$m_l \in (0, 1)$. Choose a starting point $x_0 \in R^n$; set $\hat{x}_0 = x_0$; the oracle provides $f_{x_0} \in R$ and $g_{x_0}^G \in R^n$ satisfying Assumption (4). The initial bundle is made up of just one element $(x_0, f_{x_0}, g_{x_0}^G, 0, 0)$ so that I_0^- is empty, while I_0^+ is a singleton. Since $f(x) > 0$, a lower bound f_0^{low} for $f^{\text{inf}} = \min\{f(x) \mid x \in R^n\}$ is available; set $v_0^l = (1 - m_l)(f_{\hat{x}_0} - f_0^{\text{low}})$. Set the iteration counter $k = 0$.

Step 1 (first stopping test). If $\|g_{\hat{x}_k}^G\| \leq \delta$, then terminate.

Step 2 (second stopping test). Set $\Delta_k = f_{\hat{x}_k} - f_k^{\text{low}}$. If $\Delta_k \leq \text{Tol}_\Delta$, then stop.

Step 3 (level feasibility checking). Set the level parameter $l_k = f_{\hat{x}_k} - v_k^l$. If the level set X_k defined by (15) is detected to be empty, then set $f_k^{\text{low}} = l_k, v_k^l = (1 - m_l)(f_{\hat{x}_k} - f_k^{\text{low}})$ and go back to Step 2. Otherwise set $u_{\min}^k = r\varepsilon/2\|g_{\hat{x}_k}^G\|, u_{\max}^k = Ru_{\min}^k$, and $\theta^{(k)} = r\delta u_{\min}^k$.

Step 4 (direction finding). Find the solution $(d_{\hat{u}}^k, v_{\hat{u}}^k)$ of $QP(\hat{u})$ for increasing value of u, such that

$$f_{\hat{x}_k + d_{\hat{u}}} > f_{\hat{x}_k} + mv_{\hat{u}}, \tag{20}$$

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

Step 5 (bundle updating). Set

$$I_k^+ = I_k^+ \setminus \{i \in I_k^+ \mid a_i > \varepsilon\}, \qquad I_k^- = I_k^- \setminus \{i \in I_k^- \mid a_i > \varepsilon\}. \tag{21}$$

Calculate

$$g^* = \arg\min\left\{\|g\| \mid g \in \text{conv}\left\{g_{x_i}^G \mid i \in I_k^+\right\}\right\}. \tag{22}$$

If $\|g^*\| \leq \delta$, then terminate.
Else set $u_{\max}^k = u_{\max}^k - r(u_{\max}^k - u_{\min}^k)$; go to Step 4.

Step 6 (trial point calculating). Set $x_{k+1} = \hat{x}_k + d_{\hat{u}}^k$, calculate $f_{x_{k+1}}, g_{x_{k+1}}^G \in \partial_{\eta_{x_{k+1}}^g}^G f(x_{x+1})$, and set $\alpha_{k+1} = f_{\hat{x}_k} - f_{x_{k+1}} + (g_{x_{k+1}}^G)^T d_{\hat{u}}^k$.

Step 7 (index insertion). (a) If $\alpha_{k+1} < -2\eta$ and $\|d_{\hat{u}}^k\| > \varepsilon$, then insert the element $(x_{k+1}, f_{x_{k+1}}, g_{x_{k+1}}^G, \alpha_{k+1}, \|d_{\hat{u}}^k\|)$ into the bundle for an appropriate value of $i \in I_k^-$ and set $\hat{u} = \hat{u} - r(\hat{u} - u_{\min}^k)$.

(b) Else, if $(g_{x_{k+1}}^G)^T d_{\hat{u}}^k \geq \rho v_{\hat{u}}^k - 2\eta$, then insert the element $(x_{k+1}, f_{x_{k+1}}, g_{x_{k+1}}^G, \max\{\alpha_{k+1}, -2\eta\}, \|d_{\hat{u}}^k\|)$ into the bundle for an appropriate value of $i \in I_k^+$.

(c) Else find a scalar $t \in (0, 1)$ such that $g(t) \in \partial_{\eta_{\hat{x}_k + td_{\hat{u}}^k}^g}^G f(\hat{x}_k + td_{\hat{u}}^k)$ satisfies the condition $g(t)^T d_{\hat{u}}^k \geq \rho v_{\hat{u}}^k - 2\eta$ and insert the element $(\hat{x}_k + td_{\hat{u}}^k, f_{\hat{x}_k + td_{\hat{u}}^k}, g(t), \max\{-2\eta, \alpha_t\}, t\|d_{\hat{u}}^k\|)$ into the bundle for an appropriate value of $i \in I_k^+$, where $\alpha_t = f_{\hat{x}_k} - f_{\hat{x}_k + td_{\hat{u}}^k} + tg(t)^T d_{\hat{u}}^k$.

Step 8 (descent test). If $\|d_{\hat{u}}^k\| \leq \theta$, go to Step 5. Choose $f_{k+1}^{\text{low}} \in [f_k^{\text{low}}, f^{\text{inf}}]$.

If

$$f_{x_{k+1}} \leq f_{\hat{x}_k} + m v_{\hat{u}}^k, \tag{23}$$

set the new stability center $\hat{x}_{k+1} = x_{k+1}$, $v_{k+1}^l = (1 - m_l)(f_{\hat{x}_{k+1}} - f_{k+1}^{\text{low}})$, set $k = k + 1$, and go to Step 1 (serious step).

Otherwise set $\hat{x}_{k+1} = \hat{x}_k$, $v_{k+1}^l = (1 - m_l)(f_{\hat{x}_{k+1}} - f_{k+1}^{\text{low}})$, 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 $(v_{\hat{u}}, d_{\hat{u}})$ and $(\lambda_{\hat{u}}, \mu_{\hat{u}}, \xi_{\hat{u}})$ 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 $l_k \leq \check{f}_k(x) \leq f(x) + \beta, \forall x \in R^n$, and since the update sets $f_k^{\text{low}} = l_k$, it holds that $f_k^{\text{low}} \leq f^{\text{inf}} + \beta$ for all k. If Algorithm 3 stops at Step 2, we have that

$$\text{Tol}_\Delta \geq f_{\hat{x}_k} - f_k^{\text{low}} \geq f_{\hat{x}_k} - f^{\text{inf}} - \beta = f(\hat{x}_k) - \eta_{\hat{x}_k} - f^{\text{inf}} - \beta$$

$$\geq f(\hat{x}_k) - \eta - f^{\text{inf}} - \beta, \tag{24}$$

which means that $\hat{x}_k$ is a $(\text{Tol}_\Delta + \eta + \beta)$-approximate solution to problem (1). If at Step 2 the rule $l_k = f_{\hat{x}_k} - v_k^l$ 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 solve efficiently sequences of related subproblems $QP(u)$ with varying u. The construction of $d_{\hat{u}}^k$ 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) $d_{\hat{u}}^k \leq \theta$ is never a consequence of the choice of too small $\hat{u}$. In fact, we note that if $\|g_{\hat{x}_k}^G\| > \delta$, it holds that

$$\left\|d_{u_{\min}}\right\| \leq 2u_{\min}\left\|g_{\hat{x}_k}^G\right\| + 2u_{\min}\eta = \frac{2\theta\left(\left\|g_{\hat{x}_k}^G\right\| + \eta\right)}{r\delta}. \tag{25}$$

The right-hand side of the above inequality is bigger than θ, so too small $\hat{u}$ cannot lead to $d_{\hat{u}}^k \leq \theta$; therefore we need to increase the value of u.

(e) We remark that the insertion of a bundle index into I_k^+ or I_k^- at Step 7 is not simply based on the sign of α_i.

(f) The level constraint $\check{f}_k(x) \leq l_k$ 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; (26)

infinitely many descent steps are generated; (27)

finitely many descent steps are generated. (28)

For the first case (26), if the stopping tolerance Tol_Δ 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 $\Delta_k \to 0$:*

$$\lim_{k \to \infty} f_{\hat{x}_k} \leq f^{\text{inf}} + \beta. \tag{29}$$

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

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

$$f_{\hat{x}_k} - l_k = f_{\hat{x}_k} - \left(f_{\hat{x}_k} - v_k^l\right) = v_k^l \leq \left(1 - m_l\right)\Delta_k, \tag{30}$$

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

$$\lim_{k \to \infty} f_{\hat{x}_k} - f^{\text{inf}} - \beta \leq \lim_{k \to \infty} f_{\hat{x}_k} - f_k^{\text{low}} = \lim_{k \to \infty} \Delta_k = 0, \tag{31}$$

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

$$f^{\text{inf}} + \beta \geq \lim_{j \to \infty} f_{\hat{x}_{k(j)}} = \lim_{j \to \infty} \left(f\left(\hat{x}_{k(j)}\right) - \eta_{\hat{x}_{k(j)}}\right) \geq f(\tilde{x}) - \eta, \tag{32}$$

which establishes the last conclusion. □

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 $S_0 = \{x \in R^n \mid f_x \leq f_{x_0}\}$ is compact;

(A2) for any $d \in R^n$, the directional directive $f'(x; d)$ of f at x exists and

$$f'(x; d) = \lim_{t \downarrow 0} g(t)^T d, \tag{33}$$

where $g(t) \in \partial^G_{\eta^g_{x+td}} f(x + td)$.

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 $\widehat{x}_k$ is the last stability center followed by a null step sequence $(d^j_{\widehat{u}}, v^j_{\widehat{u}})$ such that $\|d^j_{\widehat{u}}\| > \theta$ and*

$$f_{\widehat{x}_k + d^j_{\widehat{u}}} - f_{\widehat{x}_k} > m v^j_{\widehat{u}} \tag{34}$$

with the algorithm looping between Steps 6 and 9. Then the following conclusions hold.

   (i) *There exists an index $\bar{j}$ such that, for each $j \geq \bar{j}$, every new bundle index is inserted into I^+_k and $\widehat{u}$ remains unchanged.*

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

   (iii) *Whenever a new bundle index is inserted into I^+_k, the condition $(g^G_j)^T d^j_{\widehat{u}} \geq \rho v^j_{\widehat{u}} - 2\eta$ holds, where g^G_j 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 $\widehat{u}$ falls below the threshold $\varepsilon / (2\|g_{\widehat{x}_k}\| + 2\eta)$.

(ii) Since the sufficient decrease condition (23) is not satisfied, it follows that

$$\rho v^j_{\widehat{u}} < m v^j_{\widehat{u}} < f_{\widehat{x}_k + d^j_{\widehat{u}}} - f_{\widehat{x}_k}. \tag{35}$$

According to the inexact data in Assumption (4), $f_{\widehat{x}_k + d^j_{\widehat{u}}} - f_{\widehat{x}_k} \leq f(\widehat{x}_k + d^j_{\widehat{u}}) - f(\widehat{x}_k) + 2\eta$, by mean value theorem there exists a scalar $\bar{t}_j \in (0, 1)$ such that

$$\rho v^j_{\widehat{u}} < f'\left(\widehat{x}_k + \bar{t}^j d^j_{\widehat{u}}; d^j_{\widehat{u}_k}\right) + 2\eta. \tag{36}$$

Thus the conclusion follows from Assumption (A2).

(iii) Observe that the condition $(g^G_j)^T d^j_{\widehat{u}} \geq \rho v^j_{\widehat{u}} - 2\eta$ is ensured either by construction or by the fact that

$$\left(g^G_j\right)^T d^j_{\widehat{u}} + 2\eta \geq \left(g^G_j\right)^T d^j_{\widehat{u}} - \alpha^j_{\widehat{u}} = f_{\widehat{x}_k + d^j_{\widehat{u}}} - f_{\widehat{x}_k} > m v^j_{\widehat{u}} > \rho v^j_{\widehat{u}}, \tag{37}$$

whenever $\alpha^j_{\widehat{u}} \geq -2\eta$. □

Now 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 $\widehat{x}_{k+1} \neq \widehat{x}_k$.

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 $e^T \lambda - e^T \mu - \xi = \widehat{u}$ in the dual problem $DP(\widehat{u})$, there exists an index $\bar{j}$ such that, for all $j \geq \bar{j}$, $d^j_{\widehat{u}}$ can be expressed in the form (say $|I^+_k| = m$)

$$d^j_{\widehat{u}} = -G_+ \lambda^j = -\sum_{i \in I^+_k} \lambda_i g^G_{x_i}$$

$$= -\sum_{i \in I^+_k} \lambda_i \left(\frac{\lambda_1}{\sum_{i \in I^+_k} \lambda_i} g^G_{x_1} + \cdots + \frac{\lambda_m}{\sum_{i \in I^+_k} \lambda_i} g^G_{x_m} \right) \tag{38}$$

$$= -(\widehat{u} + \xi) g^{(j)},$$

where $g^{(j)} \in \text{conv}\{g^G_{x_i} \mid i \in I^+_k\}$. But since $\|d^j_{\widehat{u}}\| \leq \theta$ and $\|g^{*(j)}\| > \delta$, we have

$$\theta \geq \|d^j_{\widehat{u}}\| = |\widehat{u} + \xi| \|g^j\| \geq u_{\min} \|g^{*(j)}\| > \frac{\theta}{r\delta} \delta > \theta, \tag{39}$$

which leads to a contradiction.

Next we show that it is impossible to have $d^j_{\widehat{u}} > \theta$ 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 jth passage through such a loop, we observe that, by Lemma 6 (i), there exists an index $\bar{j}$ such that, for every $j \geq \bar{j}$, the sequence $\{z^j_{\widehat{u}}\}$ is nondecreasing, bounded, and hence convergent. Moreover, $\{d^j_{\widehat{u}}\}$ is bounded; it admits a convergent subsequence, say $\{d^j_{\widehat{u}}\}_{j \in J' \subset J}$. The above consideration implies also that $\{v^j_{\widehat{u}}\}_{j \in J' \subset J}$ 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 $\|d^j_{\widehat{u}}\| > \theta$. □

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 $\varepsilon > 0$ and $\delta > 0$, Algorithm 3 stops in a finite number of iterations at a point satisfying the approximate stationarity condition*

$$\|g^*\| \leq \delta \tag{40}$$

with $g^* \in \text{conv}\{\partial^G_{\eta^g_y} f(y) \mid \|y - \widehat{x}_k\| \leq \varepsilon\}$.

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 $\widehat{x}_k$ be the stability center at the kth passage; then $\|d_{\widehat{u}}^k\| > \theta$ and

$$f_{\widehat{x}_{k+1}} \leq f_{\widehat{x}_k} + m v_{\widehat{u}}^k,$$

$$f_{\widehat{x}_{k+1}} - f_{\widehat{x}_0} \leq m \sum_{i=0}^{k} v_{\widehat{u}}^i. \tag{41}$$

Since $u_{\min}^k = r\varepsilon/2\|g_{\widehat{x}_k}\|$ and $\theta^{(k)} = r\delta u_{\min}^k$, 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 $\|d_{\widehat{u}}^k\| > \theta^{(k)} \geq r^2 \varepsilon \delta / 2 L_0$ that $\|d_{\widehat{u}}^j\|$ is bounded away from zero. Then from Lemma 2(iii), $v_{\widehat{u}}^j$ is bounded away from zero as well. Therefore by passing to the limit we obtain

$$\lim_{k \to \infty} f_{\widehat{x}_{k+1}} - f_{\widehat{x}_0} \leq -\infty. \tag{42}$$

Since $f_{\widehat{x}_{k+1}} = f(\widehat{x}_{k+1}) - \eta_{\widehat{x}_{k+1}} \leq f(\widehat{x}_{k+1}) + \eta$, $f_{\widehat{x}_0} = f(\widehat{x}_0) - \eta_{\widehat{x}_0} \leq f(\widehat{x}_0) + \eta$, we have $\lim_{k \to \infty} f(\widehat{x}_{k+1}) = -\infty$, 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

References

[1] E. W. Cheney and A. A. Goldstein, "Newton's method for convex programming and Tchebycheff approximation," *Numerische Mathematik*, vol. 1, pp. 253–268, 1959.

[2] J. E. Kelley, Jr., "The cutting-plane method for solving convex programs," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, pp. 703–712, 1960.

[3] A. Fuduli, M. Gaudioso, and G. Giallombardo, "Minimizing nonconvex nonsmooth functions via cutting planes and proximity control," *SIAM Journal on Optimization*, vol. 14, no. 3, pp. 743–756, 2003.

[4] U. Brannlund, *On relaxation methods for nonsmooth convex optimization [Ph.D. thesis]*, Department of Mathematics, Royal Institute of Technology, Stockholm, Sweden, 1993.

[5] C. Lemaréchal, A. Nemirovskii, and Y. Nesterov, "New variants of bundle methods," *Mathematical Programming*, vol. 69, no. 1, pp. 111–147, 1995, Nondifferentiable and large-scale optimization (Geneva, 1992).

[6] W. de Oliveira and M. Solodov, "A doubly stabilized bundle method for nonsmooth convex optimization," Tech. Rep., 2013, http://www.optimization-online.org/DB_HTML/2013/04/3828.html.

[7] M. Hintermüller, "A proximal bundle method based on approximate subgradients," *Computational Optimization and Applications*, vol. 20, no. 3, pp. 245–266, 2001.

[8] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, NY, USA, 1983.

[9] A. A. Goldstein, "Optimization of Lipschitz continuous functions," *Mathematical Programming*, vol. 13, no. 1, pp. 14–22, 1977.

[10] K. C. Kiwiel, "A dual method for certain positive semidefinite quadratic programming problems," *SIAM Journal on Scientific and Statistical Computing*, vol. 10, no. 1, pp. 175–186, 1989.

[11] A. Fuduli and M. Gaudioso, "The proximal trajectory algorithm for convex minimization," Tech. Rep. 7/98, Laboratorio di Logistica, Dipartimento di Elettronica Informatica e Sistemistica, Universita della Calabria, Rende Cosenza, Italy, 1998.