J. of Inequal. & Appl., 1998, Vol. 2, pp. 157–179
Reprints available directly from the publisher
Photocopying permitted by license only

A New Method for Nonsmooth Convex Optimization*

Z. WEI [a], L. QI [a,†] and J.R. BIRGE [b]

[a] *School of Mathematics, University of New South Wales, Sydney NSW 2052, Australia;* [b] *Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, USA*

(Received 3 March 1997)

A new method for minimizing a proper closed convex function f is proposed and its convergence properties are studied. The convergence rate depends on both the growth speed of f at minimizers and the choice of proximal parameters. An application of the method extends the corresponding results given by Kort and Bertsekas for proximal minimization algorithms to the case in which the iteration points are calculated approximately. In particular, it relaxes the convergence conditions of Rockafellar's results for the proximal point algorithm.

Keywords: Proper closed convex function; Proximal minimization; Global and superlinear convergence

AMS Subject Classification: Primary 90C47, Secondary 90C05

1 INTRODUCTION

Consider the proximal minimization algorithm for solving the following convex minimization problem

$$\min\{f(x):\ x \in R^n\} \tag{1}$$

* This material is based on work supported by the Australian Research Council and the National Science Foundation under Award Number DDM-9215921.

† Corresponding author.

defined by

$$x_{k+1} = \arg\min\left\{ f(x) + \frac{1}{(\tau+1)\lambda_k} \|x - x_k\|^{\tau+1} :\ x \in R^n \right\},$$

where f is a proper closed convex function defined in R^n, the proximal parameters λ_k and τ are positive constants, $\|\cdot\|$ denotes the standard Euclidean norm on R^n, (see Martinet [15] which is the case $\tau = 1$ based on [16]). The algorithm is globally convergent, and its local speed of convergence depends on: (i) The growth speed of f at the minimizers; (ii) The choice of λ_k; and (iii) The choice of τ (see [1,2,11]). For a survey of results of the proximal type minimization we refer the reader to [3–6,8–10,14,20–24], especially to [3,5,9,20].

Rockafellar [20] made a nice extension for the proximal point algorithm. In his algorithm, the iteration points are calculated approximately. A question is: do corresponding results of Kort and Bertsekas [11] hold for the algorithm of Rockafellar [20] when it is applied to a proper closed convex function?

In the next section, a new method for solving nonsmooth convex optimization problems is presented. This algorithm is globally convergent, and its local speed of convergence depends on both the growth speed at the minimizers and the choice of proximal parameters. These results are similar to those in [1,2,11]. Furthermore, a general proximal point algorithm introduced in [20] can be regarded as a special case of this method. In Section 3, an application of this algorithm relaxes a key condition of Rockafellar [20] for convergence of the proximal point algorithm. The condition,

$$\sum_{k=1}^{\infty} \delta_k < +\infty,$$

becomes

$$\limsup\{\delta_k\} < 1,$$

where δ_k is a parameter in the algorithm. In the same section, some special cases of the algorithm are also discussed. Some concluding remarks are in Section 4.

2 METHOD AND ITS CONVERGENCE PROPERTIES

2.1 The Algorithm and its Global Convergence

Algorithm 1 Let $q > 1$ be a given positive number. Let $x_1 \in R^n$ be arbitrary. In general, given $x_k \in R^n$, generate a proximal parameter $t_k > 0$ and $x_{k+1} \in R^n$ such that there exists $g_{k+1} \in \partial f(x_{k+1})$ satisfying the following inequality:

$$g_{k+1}^T(x_k - x_{k+1}) \geq t_k \|g_{k+1}\|^q. \tag{2}$$

THEOREM 2.1 *Let $\{x_k\}_{k=1}^{\infty}$ be any sequence generated by Algorithm* 1.

(i) *If $\sum_{k=1}^{\infty} t_k = +\infty$, then either $f(x_k) \to -\infty$ or $\liminf_{k\to\infty} \|g_k\| = 0$. In the latter case, if $\{x_k\}_{k=1}^{\infty}$ is a bounded set, then every accumulation point of $\{x_k\}_{k=1}^{\infty}$ is an optimal point of* (1).

(ii) *If $\inf\{t_k\} > 0$, then either $f(x_k) \to -\infty$ or $g_k \to 0$. In the latter case, every accumulation point of $\{x_k\}_{k=1}^{\infty}$ (if there is any) is an optimal point of* (1).

Furthermore, Let X^* be the set of minimizers of f, $f^* = \inf\{f(x): x \in R^n\}$. Then

$$f(x_{k+1}) - f^* \leq \inf\{\|x_{k+1} - x^*\|:\ x^* \in X^*\}[t_k^{-1}\|x_{k+1} - x_k\|]^{(q-1)^{-1}}.$$

Proof From the subgradient inequality,

$$f(x_k) \geq f(x_{k+1}) + g_{k+1}^T(x_k - x_{k+1}).$$

By (2), we have

$$f(x_k) \geq f(x_{k+1}) + t_k\|g_{k+1}\|^q. \tag{3}$$

This implies that $f(x_{k+1}) \leq f(x_k)$. If the decreasing sequence $\{f(x_k)\}_{k=1}^{\infty}$ tends to $-\infty$, we are done. Otherwise, this sequence is bounded from below.

By (3),

$$\sum_{k=1}^{\infty} t_k\|g_{k+1}\|^q < +\infty.$$

If $\|g_k\|$ is bounded below by a positive number, then the above inequality implies that $\sum_{k=1}^{\infty} t_k < +\infty$, which contradicts our assumptions in (i). Thus, $\liminf_{k\to\infty} \|g_k\| = 0$. If $\{x_k\}_{k=1}^{\infty}$ is a bounded set, then there exists an infinite index set K such that $\lim_{k\in K} g_k = 0$. Without loss of generality, we may assume that $\lim_{k\in K} x_k = x^*$.

Applying

$$f(x) \geq f(x_k) + g_k^T(x - x_k)$$

and letting $k \to \infty$, $k \in K$, we have that for all $x \in R^n$, $f(x) \geq f(x^*)$, which implies that x^* is a minimum point of f. By $f(x_{k+1}) \leq f(x_k)$, we deduce that every accumulation point of $\{x_k\}_{k=1}^{\infty}$ has the same optimal value, $f(x^*)$, and is, hence, also a minimum point of (1).

We now prove the second conclusion. From (3), we have that either $f(x_k) \to -\infty$ or $g_k \to 0$ by using $\inf\{t_k\} > 0$. Let $\{x_k$: $k \in K\}$ be any convergence subsequence of $\{x_k\}_{k=1}^{\infty}$ and $\lim_{k\in K} x_k = x^*$. Then,

$$\lim_{k\in K} \partial f(x_k) \subseteq \partial f(x^*)$$

yields that $0 \in \partial f(x^*)$. So x^* is an optimal point of (1), since f is a convex function.

For each $x^* \in X^*$ and any k, since

$$f^* = f(x^*) \geq f(x_{k+1}) + g_{k+1}^T(x^* - x_{k+1}),$$

by (2), we have

$$\begin{aligned} f(x_{k+1}) - f^* &\leq \|g_{k+1}\| \|x_{k+1} - x^*\| \\ &\leq \|x_{k+1} - x^*\| [t_k^{-1} \|x_{k+1} - x_k\|]^{(q-1)^{-1}}. \end{aligned}$$

Hence

$$f(x_{k+1}) - f^* \leq \inf\{\|x_{k+1} - x^*\|\colon x^* \in X^*\} [t_k^{-1} \|x_{k+1} - x_k\|]^{(q-1)^{-1}}.$$

The second conclusion follows.

COROLLARY 2.1 *Suppose that X^* is nonempty, compact and $\{t_k\}_{k=1}^{\infty}$ is bounded away from zero. Let $\{x_k\}_{k=1}^{\infty}$ be any sequence generated by*

Algorithm 1. *Then* $\{x_k\}_{k=1}^{\infty}$ *is bounded*, $f(x_k) \to f^*$ *and*

$$\lim_{k \to \infty} \rho(x_k; X^*) = 0, \tag{4}$$

where $\rho(x; X^*)$ *is the distance from* x *to* X^* *given by*

$$\rho(x; X^*) = \min\{\|x - x^*\|: \ x^* \in X^*\}.$$

Proof By Theorems 8.4 and 8.7 of [18], the level sets of f are bounded. On the other hand, for all k, $f(x_{k+1}) \geq f(x_k)$. Hence $\{x_k\}_{k=1}^{\infty}$ is contained in some level sets of f. Therefore, $\{x_k\}_{k=1}^{\infty}$ is a bounded set. By (ii) of Theorem 2.1 and the compactness of X^*, we have $f(x_k) \to f^*$ and (4).

2.2 Local Convergence

We from now on assume that $\inf\{t_k\} > 0$. Like [3,11], we need the following key assumption which was used to analyze the convergence rate of the proximal minimization algorithm for quadratic as well as certain types of nonquadratic proximal terms (see [3,11]).

Assumption 1 X^* is nonempty and compact. Furthermore, there exist scalars $\alpha > 0$, $p \geq 1$ and $\delta > 0$ such that for any x with $\rho(x; X^*) \leq \delta$,

$$f(x) - f^* \geq \alpha(\rho(x; X^*))^p. \tag{5}$$

LEMMA 2.1 *Suppose that* (5) *holds. Then for each* x *with* $\rho(x; X^*) \leq \delta$ *and* $g \in \partial f(x)$,

$$\|g\| \geq \alpha^{1/p}[f(x) - f^*]^{(p-1)/p}. \tag{6}$$

Proof Since for any $g \in \partial f(x)$ and any $x^* \in X$,

$$f^* = f(x^*) \geq f(x) + g^T(x^* - x).$$

So we have

$$\begin{aligned} f^* - f(x) &\geq \max\{g^T(x^* - x): \ x^* \in X^*\} \\ &\geq \max\{-\|g\|\|x^* - x\|: \ x^* \in X^*\} \\ &= -\|g\|\rho(x; X^*). \end{aligned}$$

Hence,

$$f(x) - f^* \leq \|g\|\rho(x; X^*) \leq \|g\| \left[\frac{f(x) - f^*}{\alpha} \right]^{1/p}$$

by using (5). This implies (6).

LEMMA 2.2 *Suppose that* (5) *holds,* $\{x_k\}_{k=1}^{\infty}$ *is generated by Algorithm* 1. *Then for all large* k,

$$\frac{f(x_{k+1}) - f^*}{f(x_k) - f^*} \leq \frac{1}{1 + t_k \alpha^{q/p} [f(x_{k+1}) - f^*]^{((q-1)(p-1)-1)/p}}. \tag{7}$$

Proof Since $\rho(x_k; X^*) \to 0$ and $f(x_k) > f(x_{k+1})$, we have $k_1 > 1$ such that for all $k \geq k_1$, $\rho(x_k; X^*) \leq \delta$. Using $g_{k+1} \in \partial f(x_{k+1})$, we have

$$\begin{aligned} f(x_k) &\geq f(x_{k+1}) + g_{k+1}^T (x_k - x_{k+1}) \\ &\geq f(x_{k+1}) + t_k \|g_{k+1}\|^q \\ &\geq f(x_{k+1}) + t_k \alpha^{q/p} [f(x_{k+1}) - f^*]^{q(p-1)/p} \end{aligned}$$

by Lemma 2.1. Therefore,

$$f(x_k) - f^* \geq f(x_{k+1}) - f^* + t_k \alpha^{q/p} [f(x_{k+1}) - f^*]^{q(p-1)/p}.$$

This implies the conclusion by using $f(x_{k+1}) > f^*$.

THEOREM 2.2 *Suppose that* (5) *holds,* $\{x_k\}_{k=1}^{\infty}$ *is generated by Algorithm* 1. *Then* $f(x_k)$ *tends to* f^* *superlinearly under any one of the following conditions* ((a), (b) *or* (c)):

(a) $(q-1)(p-1) < 1$.
(b) $(q-1)(p-1) = 1$ *and* $t_k \to +\infty$. (*If* $t_k \to t_* < +\infty$, *then* $f(x_k)$ *tends to* f^* *linearly.*)
(c) $(q-1)(p-1) > 1$ *and*

$$\lim_{k \to \infty} t_k [f(x_k) - f^*]^{((q-1)p-q)/p} = +\infty. \tag{8}$$

If $p = 1$, *then there exists* $k_2 > k_1$ *such that* $x_{k_2} \in X^*$, *i.e., the minimizer of* f *can be determined within a finite number of steps.*

Furthermore, if there exists a scalar $\beta < +\infty$ *such that*

$$f(x) - f^* \leq \beta(\rho(x; X^*))^p \quad \textit{for all } x \textit{ with } \rho(x; X^*) \leq \delta, \tag{9}$$

then x_k *tends to* X^* *superlinearly under any of the above conditions* ((a), (b) *or* (c)).

Proof The fact, $f(x_{k+1}) \to f^*$, implies that

$$\lim_{k \to \infty} t_k \alpha^{q/p} [f(x_{k+1}) - f^*]^{((q-1)(p-1)-1)/p} = +\infty, \tag{10}$$

if one of the conditions (a)–(b) holds. Then

$$\lim_{k \to \infty} \frac{f(x_{k+1}) - f^*}{f(x_k) - f^*} = 0 \tag{11}$$

by Lemma 2.2.

In case (b) with $t_k \to t_* < +\infty$, from (7), we have

$$\lim_{k \to \infty} \frac{f(x_{k+1}) - f^*}{f(x_k) - f^*} \leq \frac{1}{1 + t_* \alpha^{q/p}} < 1.$$

Hence $f(x_k)$ tends to f^* linearly.

For (c), using

$$\frac{f(x_{k+1}) - f^*}{f(x_k) - f^*} \leq 1 \Big/ \Big\{ 1 + \alpha^{q/p} [f(x_{k+1}) - f^* / f(x_k) - f^*]^{((q-1)(p-1)-1)/p} \\ t_k [f(x_k) - f^*]^{((q-1)(p-1)-1)/p} \Big\},$$

(11) follows.

Recalling (5) and $\rho(x_k; X^*) \leq \delta$ for all $k > k_1$, which has been used in the proof of Lemma 2.2, we have

$$f(x_k) - f^* \geq \alpha(\rho(x_k; X^*))^p. \tag{12}$$

But from the proof of the Lemma 2.1,

$$f(x_k) - f^* \leq \|g_k\| \rho(x_k; X^*). \tag{13}$$

Thus,

$$\alpha(\rho(x_k; X^*))^p \leq \|g_k\|\rho(x_k; X^*).$$

If $p=1$, since $\lim_{k\to\infty} g_k = 0$ and X^* is a compact set, we have $k_2 > k_1$ such that $x_{k_2} \in X^*$ (otherwise $\alpha \leq 0$).

We now prove the second conclusion. Let $\|x_k - X^*\|$ denote $\rho(x_k; X^*)$. By (5) and (9) we have $k_3 > k_1$ such that for all $k > k_3$,

$$\alpha\|x_k - X^*\|^p \leq f(x_k) - f^* \leq \beta\|x_k - X^*\|^p. \tag{14}$$

This yields that $f(x_k)$ tends to f^* superlinearly if and only if x_k tends to X^* superlinearly. The second conclusion follows from (11).

Corollary 2.2 *Suppose that X^* is singleton, i.e., $X^* = \{x^*\}$, $\{x_k\}_{k=1}^{\infty}$ is generated by Algorithm* 1.

(i) *Assume that* (5) *and* (9) *hold. If there exist $\tau_0 > \max\{0, (q-1)p-q\}$ and $\tau_1 > 0$, such that for all k,*

$$t_k\|x_k - x_{k-1}\|^{\tau} \geq \tau_1,$$

then

$$\liminf_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0;$$

if, for all k,

$$t_k\|x_{k+1} - x_k\|^{\tau} \geq \tau_1,$$

then

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0.$$

(ii) *Assume that* (5) *holds. If there exists $\tau_{1'} > 0$ such that for all k,*

$$t_k[f(x_k) - f(x_{k-1})]^{q-1} \geq \tau_{1'},$$

then

$$\liminf_{k\to\infty} \frac{f(x_{k+1}) - f^*}{f(x_k) - f^*} = 0;$$

if, for all k,

$$t_k[f(x_{k+1}) - f(x_k)]^{q-1} \geq \tau_{1'},$$

then

$$\lim_{k\to\infty} \frac{f(x_{k+1}) - f^*}{f(x_k) - f^*} = 0.$$

Furthermore, if (9) *is true, then*

$$\liminf_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0$$

or

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0$$

holds respectively.

Proof (i) It is very clear that $t_k \to +\infty$. Hence, we only need to consider case (c) of Theorem 2.2. If the conclusion is false, then we have a positive number $\tau_2 < +\infty$, such that for all k,

$$\frac{\|x_k - x^*\|}{\|x_{k-1} - x^*\|} \geq \frac{1}{\tau_2}.$$

Thus, we have

$$\begin{aligned}\frac{\|x_k - x^*\|}{\|x_k - x_{k-1}\|} &\geq \frac{\|x_k - x^*\|}{\|x_k - x^*\| + \|x_{k-1} - x^*\|}\\ &= \frac{1}{1 + \|x_{k-1} - x^*\|/\|x_k - x^*\|}\\ &\geq \frac{1}{1 + \tau_2}.\end{aligned}$$

This inequality and (5) yield that

$$\begin{aligned} & t_k[f(x_k) - f^*]^{((q-1)p-q)/p} \\ & = t_k\|x_k - x_{k-1}\|^{\tau}\left[\frac{\|x_k - x^*\|}{\|x_k - x_{k-1}\|}\right]^{[(q-1)p-q]}\left[\frac{f(x_k) - f^*}{\|x_k - x^*\|^p}\right]^{((q-1)p-q)/p} \\ & \quad \times \frac{1}{\|x_k - x_{k-1}\|^{\tau-[(q-1)p-q]}} \\ & \geq \tau_1\alpha^{((q-1)p-q)/p}\left[\frac{1}{1+\tau_2}\right]^{[(q-1)p-q]}\frac{1}{\|x_k - x_{k-1}\|^{\tau-[(q-1)p-q]}}. \end{aligned}$$

This last result implies that (8) holds. Hence x_k tends to x^* superlinearly by Theorem 2.2 and

$$\frac{1}{\tau_2} \leq \frac{\|x_k - x^*\|}{\|x_{k-1} - x^*\|} \to 0.$$

This is a contradiction. So the first conclusion of (i) follows.

We now prove the second conclusion of (i). If the conclusion is false, then we have a positive number $\tau_2 < +\infty$ and K such that for all $k \in K$,

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \geq \frac{1}{\tau_2}.$$

Using this inequality, similarly to the proof of the first conclusion, we have for all $k \in K$,

$$\begin{aligned} & t_k[f(x_{k+1}) - f^*]^{((q-1)p-q)/p} \\ & \quad \geq \tau_1\alpha^{((q-1)p-q)/p}\left[\frac{1}{1+\tau_2}\right]^{[(q-1)p-q]}\frac{1}{\|x_{k+1} - x_k\|^{\tau-[(q-1)p-q]}}. \end{aligned}$$

This and Lemma 2.2 imply that

$$\lim_{k\in K}\frac{f(x_{k+1}) - f^*}{f(x_k) - f^*} = 0.$$

By (5) and (9), we have

$$\lim_{k\in K}\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|}\le\left[\frac{\alpha}{\beta}\right]^{1/p}\lim_{k\in K}\left[\frac{f(x_{k+1})-f^*}{f(x_k)-f^*}\right]^{1/p}=0.$$

So $\tau_2=+\infty$. Again, this is a contradiction. This proves the second conclusion of (i).

(ii) Assume that (5) holds. We prove the first conclusion.
Using

$$\frac{f(x_k)-f^*}{f(x_k)-f(x_{k-1})}\ge\frac{1}{1+(f(x_{k-1})-f^*)/(f(x_k)-f^*)}$$

and

$$\begin{aligned}&t_k[f(x_k)-f^*]^{((q-1)p-q)/p}\\&=t_k\left[f(x_k)-f(x_{k-1})\right]^{q-1}\left[\frac{f(x_k)-f^*}{f(x_k)-f(x_{k-1})}\right]^{((q-1)p-q)/p}\\&\quad\times\left[f(x_k)-f(x_{k-1})\right]^{-q/p},\end{aligned}$$

similarly to the proof of (i), if the conclusion is false, we have

$$\lim_{k\to\infty}t_k\left[f(x_k)-f^*\right]^{((q-1)p-q)/p}=+\infty.$$

Using conclusion (c) of Theorem 2.2, we have $f(x_k)$ tends to f^* superlinearly which contradicts our assumption that the conclusion is false. Hence, the conclusion follows.

Similarly to the proof of the first conclusion of (ii) and the proof of the second conclusion of (i), we can prove the second conclusion of (ii).

If (9) holds, by

$$\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|}\le\left[\frac{\beta}{\alpha}\frac{f(x_{k+1})-f^*}{f(x_k)-f^*}\right]^{1/p}$$

and the conclusions which we have just proved, the last conclusion of the corollary follows.

Using the results of this subsection, we may give some special cases of Algorithm 1 according to the choice of t_k. One such special case follows.

Algorithm 1.1 In Algorithm 1, generate $x_{k+1} \in \partial f(x_{k+1})$ satisfying the following inequality:

$$g_{k+1}^T(x_k - x_{k+1}) \geq \|x_k - x_{k-1}\|^{-\tau} \|g_{k+1}\|^q$$

or

$$g_{k+1}^T(x_k - x_{k+1}) \geq \left[f(x_k) - f(x_{k-1})\right]^{1-q} \|g_{k+1}\|^q.$$

In general, we can choose $q = 2$.

It is worth to note that, if we choose t_k as following:

$$t_k = r \frac{g_k^T(x_{k-1} - x_k)}{\|g_k\|^q},$$

where $r > 1$ is a constant, then from (2), we can deduce that for all k, $t_k \geq r^k$. Hence $t_k \to +\infty$. Hence if $p = q = 2$, then, by Theorem 2.2, $f(x_k)$ tends to f^* superlinearly under the condition (5).

Kort and Bertsekas in their novel paper [11] presented a combined primal–dual and penalty method for solving constrained convex programming. It essentially is a proximal point algorithm (see [2,19]). Hence, from the results of [1,2,11], under (5), the convergence rate of the ordinary proximal minimization algorithm is shown to depend on the growth speed of f at the minimizers as well as the proximal parameters λ_k and τ. More precisely, if $\tau = 1$, the ordinary minimization algorithm has finite, superlinear and linear convergence rate depending on whether $p = 1$, $1 < p < 2$ and $p = 2$, respectively. The convergence rate is also superlinear if $p = 2$ and $\lambda_k \to \infty$. In the case where the proximal term has a growth rate $\tau > 1$ other than quadratic ($\tau = 1$), the convergence rate is superlinear if $1 < p < \tau + 1$ even in the case where $p \geq 2$ (see [1,2,11]). From Theorem 2.2 (in the case of $q = 1 + 1/\tau$), we obtain corresponding results for our algorithm.

3 APPLICATION TO THE PROXIMAL POINT METHOD

In [20], Rockafellar presented two general proximal point algorithms for solving (1). One of them is as follows.

Algorithm A For x_k, generate $x_{k+1} \in R^n$ satisfying

$$\rho(0; S_k(x_{k+1})) \leq \frac{\delta_k}{c_k} \|x_{k+1} - x_k\|, \tag{15}$$

where $c_k > 0$,

$$\sum_{k=1}^{\infty} \delta_k < \infty, \tag{16}$$

and

$$S_k(x) = \partial f(x) + \frac{1}{c_k}(x - x_k). \tag{17}$$

In the following, we present a generalization of Algorithm A. In the new algorithm, we relax (16) to the following condition:

$$\lim \sup\{\delta_k\} < 1. \tag{18}$$

The new algorithm possesses a convergence rate similar to the rate for Algorithm A. The relaxation may be important in practice.

Algorithm B For any given $\tau > 0$, let $x_1 \in R^n$. In general, for $x_k \in R^n$, generate $x_{k+1} \in R^n$ satisfying

$$\rho(0; S_k^\tau(x_{k+1})) \leq \frac{\delta_k}{c_k} \|x_{k+1} - x_k\|^\tau, \tag{19}$$

$$S_k^\tau(x) = \partial f(x) + \frac{1}{c_k} \|x - x_k\|^{\tau-1}(x - x_k). \tag{20}$$

The following theorem proves that Algorithm B is a special case of Algorithm 1, hence the conclusions of Theorem 2.2 hold for Algorithm B. It is very clear that the ordinary proximal minimization described in the introduction of this paper is a special case of Algorithm B. Therefore, we extended the corresponding results [1,2,11], for ordinary proximal minimization algorithms to the case when the iteration points are calculated approximately.

THEOREM 3.1 *Suppose that x_{k+1} is generated by Algorithm B, and g_{k+1} is a vector in $\partial f(x_{k+1})$ such that the norm of $g_{k+1} + 1/c_k\|x_{k+1} - x_k\|^{\tau-1}(x_{k+1} - x_k)$ is not bigger than the right-hand side of* (19). *Then for all large k, (x_{k+1}, g_{k+1}) satisfies* (2) *with $q = 1 + 1/\tau$ and*

$$t_k = \frac{1 - \delta_k}{(1 + \delta_k)^{1+1/\tau}} c_k^{1/\tau}.$$

Proof Without loss of generality, we may assume that

$$\sup\{\delta_k\} < 1. \tag{21}$$

By (19), (21) and (20), we have

$$\left\| g_{k+1} + \frac{1}{c_k}\|x_{k+1} - x_k\|^{\tau-1}(x_{k+1} - x_k) \right\| \le \frac{\delta_k}{c_k}\|x_{k+1} - x_k\|^{\tau}. \tag{22}$$

The inequality,

$$\begin{aligned} &\left[g_{k+1} + \frac{1}{c_k}\|x_{k+1} - x_k\|^{\tau-1}(x_{k+1} - x_k) \right]^T (x_{k+1} - x_k) \\ &\quad \le \left\| g_{k+1} + \frac{1}{c_k}\|x_{k+1} - x_k\|^{\tau-1}(x_{k+1} - x_k) \right\| \|x_{k+1} - x_k\|, \end{aligned}$$

and (22) imply that

$$g_{k+1}^T(x_{k+1} - x_k) \le -\frac{1 - \delta_k}{c_k}\|x_{k+1} - x_k\|^{\tau+1}.$$

Therefore,

$$g_{k+1}^T(x_k - x_{k+1}) \ge \frac{1 - \delta_k}{c_k}\|x_{k+1} - x_k\|^{\tau+1}. \tag{23}$$

On the other hand, using (22), we have

$$\|g_{k+1}\| \leq \frac{1+\delta_k}{c_k}\|x_{k+1}-x_k\|^{\tau}. \tag{24}$$

Now, the conclusion of the theorem follows (23) and (24).

By the above results, we may state global convergence and convergence rate results for Algorithm B. Noticing that $t_k \rightarrow +\infty$ if and only if $c_k \rightarrow +\infty$, we only need to change q to $1+1/\tau$ in Theorem 2.2 and Corollary 2.2. We do not go into this direction in details. We are more interested in the special case in which $\tau = 1$. In the following analysis, we always assume $\tau = 1$.

THEOREM 3.2 *Let $\{x_k\}_{k=1}^{\infty}$ be any sequence generated by Algorithm B with $\{c_k\}_{k=1}^{\infty}$ bounded away from zero. Suppose that $\{x_k\}_{k=1}^{\infty}$ is bounded. Then every accumulation point of $\{x_k\}_{k=1}^{\infty}$ is an optimal solution of* (1). *Furthermore, if x^* is an optimal solution of* (1), *then*

$$\begin{aligned} f(x_{k+1}) - f(x^*) &\leq \|x_{k+1}-x^*\| \\ &\quad \times \left[\|g_{k+1}+c_k^{-1}(x_{k+1}-x_k)\| + c_k^{-1}\|x_{k+1}-x_k\|\right] \\ &\leq \frac{1+\delta_k}{c_k}\|x_{k+1}-x^*\|\|x_{k+1}-x_k\|. \end{aligned}$$

Proof By Theorem 2.1, we have the first conclusion. The proof of the first inequality in the second conclusion is similar to the proof of Theorem 4 in [18]. The second inequality follows (22).

For rates of convergence in optimization methods, a key condition is called the quadratic growth at x^* of f, i.e., in the neighbourhood of x^*, $O(x^*; r_1) = \{x: \|x-x^*\| < r_1\}$,

$$f(x) \geq f(x^*) + \alpha\|x-x^*\|^2 \quad \text{for all } x \in O(x^*; r_1), \tag{25}$$

where $\alpha > 0$. Under (25) and some other additional assumptions ($c_k \rightarrow +\infty$ and $\sum_{k=1}^{+\infty}\delta_k < +\infty$), Rockafellar proved the superlinear convergence of Algorithm A. The conditions in (16) and $c_k \rightarrow +\infty$ are, however, quite strict. Very large c_k and very small δ_k, in fact, imply that the calculation of x_{k+1} can be almost as difficult as the original minimization problem ($0 \in \partial f(x_{k+1})$). Moreover, the condition (25) is false for a large class convex functions (for example, $f(x)=\|x\|^p, p>2$). So it is necessary to discuss the possibility of relaxation of the requirements for c_k, δ_k and (25).

For given $x\in R^n$ and $S\subseteq R^+=(0,+\infty)$, a collection of proper closed convex functions (PCCF) is defined by

$$CC(x;S)=\{f:\ f\in \text{PCCF and there exists } p\in S, 0<\alpha_p(f;x) \le \beta_p(f;x)<\infty\},$$

where

$$\alpha_p(f;x)=\liminf_{y\to x}\left\{\frac{f(y)-f(x)}{\|y-x\|^p}:\ f(y)>f(x)\right\} \tag{26}$$

and

$$\beta_p(f;x)=\limsup_{y\to x}\left\{\frac{f(y)-f(x)}{\|y-x\|^p}:\ f(y)>f(x)\right\}. \tag{27}$$

We now have the following theorem by Theorems 2.2 and 3.2.

THEOREM 3.3 *Suppose that the conditions of Theorem* 3.2 *hold,* $\{x_k\}_{k=1}^{\infty}$ *is generated by Algorithm B and* $x_k\to x^*$. *Then* $f(x_k)$ *tends to* $f(x^*)$ *superlinearly if one of the following conditions* ((a$_1$), (b$_1$) *or* (c$_1$)) *holds*:

(a$_1$) $p\in(1,2)$, $\alpha_p(f;x^*)>0$;
(b$_1$) $p=2$, $\alpha_2(f;x^*)>0$ *and* $c_k\to+\infty$;
(c$_1$) $p\in(2,+\infty)$, $\alpha_p(f;x^*)>0$, *and* $\lim_{k\to\infty} c_k[f(x_k)-f(x^*)]^{(p-2)/p}=+\infty$.

If $\alpha_1(f;x^*)>0$, *then the minimizer of* f *can be determined after a finite number of steps.*

Furthermore, x_k *tends to* x^* *superlinearly if one of the following conditions* ((a$_2$) *or* (b$_2$) *or* (c$_2$)) *holds*:

(a$_2$) $f\in CC(x^*;(1,2))$;
(b$_2$) $f\in CC(x^*;\{2\})$ *and* $c_k\to+\infty$;
(c$_2$) $f\in CC(x^*;(2,+\infty))$ *and* $\lim_{k\to\infty} c_k\left[f(x_k)-f(x^*)\right]^{(p-2)/p}=+\infty$.

COROLLARY 3.1 *Assume that f has a unique minimum, say x^*. Let $\{x_k\}_{k=1}^{\infty}$ be a sequence generated by Algorithm B.*

(i) *If* $f \in CC(x^*; \{p\})$ *and there exist* $\tau_3 > \max\{0, p-2\}$, *and* $\tau_4 > 0$, *such that for all k*

$$c_k \|x_k - x_{k-1}\|^{\tau_3} \geq \tau_4,$$

then x_k tends to the unique minima of f, i.e., $x_k \to x^*$, and

$$\liminf_{k \to \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0;$$

if for all k,

$$c_k \|x_{k+1} - x_k\|^{\tau_3} \geq \tau_4,$$

then

$$\lim_{k \to \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0.$$

(ii) *Assume* $\alpha_p(f; x^*) > 0$. *If there exists* $\tau_{1'} > 0$ *such that for all k,*

$$c_k [f(x_k) - f(x_{k-1})] \geq \tau_{4'},$$

then

$$\liminf_{k \to \infty} \frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} = 0;$$

if, for all k,

$$c_k [f(x_{k+1}) - f(x_k)] \geq \tau_{4'},$$

then

$$\lim_{k \to \infty} \frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} = 0.$$

Furthermore, if $f \in CC(x^*; p)$, *then*

$$\liminf_{k \to \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0$$

or

$$\lim_{k \to \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0$$

holds respectively.

Proof Using the assumption of this theorem, we can deduce that $\{x_k\}_{k=1}^{\infty}$ is bounded (see the proof of Corollary 2.1). By the boundedness of $\{x_k\}_{k=1}^{\infty}$,

$$c_k \geq \frac{\tau_4}{\|x_k - x_{k-1}\|^{\tau_3}} \geq \frac{\tau_4}{2^{\tau_3} \sup\{\|x_k\|^{\tau_3}\}} > 0,$$

and Theorem 3.2, we have that x_k tends to x^*, the unique minimum of f, i.e., $x_k \to x^*$. The conclusions follow Corollary 2.2 now.

It should be noted that the condition $\alpha_2(f; x^*) > 0$ is weaker than (25). In fact, $\alpha_2(f; x^*) > 0$ does not imply the uniqueness of the minimizers of f, but the following proposition holds.

PROPOSITION 3.1 *If* (25) *holds, then* $X^* = \{x^*\}$.

Let Y^* be the set of all accumulation points of $\{x_k\}_{k=1}^{\infty}$ generated by Algorithm B. Then we have the following result.

LEMMA 3.1 *Suppose that* $Y^* \neq \emptyset$ *has an isolated point and*

$$\lim_{k \to \infty} \|x_{k+1} - x_k\| = 0. \tag{28}$$

Then $\{x_k\}_{k=1}^{\infty}$ *is a convergent sequence.*

Proof By the assumption on Y^*, there is a point $y^* \in Y^*$ such that it is an isolation point in Y^*. Then there exist $r_2 > 0$ and $O(y^*; r_2) = \{x\colon \|x - y^*\| < r_2\}$ such that $Y^* \cap O(y^*; r_2) \backslash \{y^*\} = \emptyset$. Let $K(y^*) = \{k\colon x_k \in O(y^*; r_2)\}$. Then $K(y^*)$ is an infinite set and

$$\lim_{k \in K(y^*)} x_k = y^*.$$

Furthermore, let $K(y^*) = \{k_j\colon j_1 < j_2 \text{ implies } k_{j_1} \le k_{j_2}\}$ and N be the set of natural numbers. If $N\backslash K(y^*)$ is an infinite set, then there exists an infinite set $K_1(y^*) \subseteq K(y^*)$ such that for all $k_i \in K_1(y^*)$, $k_i + 1 \notin K(y^*)$. Let l_1 be the smallest number in $N\backslash K(y^*)$.

In general, suppose that we have $\{l_1, l_2, \ldots, l_j\}$. Then let l_{j+1} be the smallest number in $N\backslash K(y^*)\backslash\{l_1, l_1+1, \ldots, l_j, l_j+1\}$. By induction, we have an infinite set $\{l_i\colon i \in N\}$. It is clear that $K_1(y^*) = \{k_i = l_i - 1\colon i \in N\}$ is the desired set. Since $\lim_{k_i \to \infty} x_{k_i} = y^*$, we have $\lim_{k_i \to \infty} x_{k_i+1} = y^*$ by (28). This is impossible from the definition of $K(y^*)$ and $k_i + 1 \notin K(y^*)$. So $N\backslash K(y^*)$ is a finite set. This implies that $\lim_{k \to \infty} x_k = y^*$ by the definition of $K(y^*)$.

Indeed, if both $L(f; x_1) = \{x\colon f(x) \le f(x_1)\}$ and $\{c_k\}$ are bounded, then Theorem 3.2 implies that $\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0$. In addition, if the condition (a_2) in Theorem 3.3 holds and Y^* is not a dense subset of R^n, then x_k tends to one of the minimizers of f superlinearly. In this sense, (25) relaxed by $\alpha_2(f; x^*) > 0$ is a useful consideration for those problems with nonunique minima.

Example 3.1 Let $f(x) \in LEC$, and $f(x^*) = 0$. For any given number $l < -(p-2)/p$, choose

$$c_k = \begin{cases} c > 0 & \text{if } 1 \le p < 2, \\ [f(x_k)]^l & \text{if } p \ge 2 \end{cases} \tag{29}$$

in Algorithm B, then $f(x_k)$ tends to $f(x^*)$ superlinearly if $\alpha_p(f; x^*) > 0$. Furthermore, x_k tends to x^* superlinearly if $f \in CC(x^*; \{p\})$.

In Examples 3.2–3.4, we assume that c_k satisfies (29) in Algorithm B and that x_k tends to x^*.

Example 3.2 Suppose that we wish to find a zero point of a function $h(x) \in LEC$. Then we can choose $f(x) = \max\{0, h(x)\}$ and start Algorithm B from the point x_1, $h(x_1) > 0$. Since $f(x_k) > f(x_{k+1}) \to f(x^*)$, also $h(x_k) > h(x_{k+1}) \to h(x^*)$. Thus, $h(x^*) = 0$. Furthermore, we have the following proposition.

PROPOSITION 3.2 *If $\alpha_p(h; x^*) > 0$, then $h(x_k)$ tends to 0 superlinearly. If $h \in CC(x^*; \{p\})$, then x_k tends to x^*superlinearly.*

Proof Since $\alpha_p(f; x^*) = \alpha_p(h; x^*)$ and $\beta_p(f; x^*) = \beta_p(h; x^*)$, the conclusions follow from Theorem 3.3.

Example 3.3 An important special case of Example 3.1 is called the inequality feasibility problem which is the most fundamental in convex mathematical programming.

Let f_i ($i \in I$ a finite index set) be convex functions. Find $x \in R^n$ such that

$$f_i(x) \leq 0 \quad \text{for all } i \in I.$$

Let $f(x) = \max\{0, f_i(x){:}\ i \in I\}$. Then f is a convex function, $f(x) \geq 0$. Furthermore

$$f(x) = 0 \text{ if and only if } f_i(x) \leq 0 \quad \text{for all } \ i \in I.$$

Hence, for the inequality feasibility problem, we have a method to solve it quickly if there exists $p \in (0, +\infty)$ such that for all $i \in I(x^*) = \{i{:} f_i(x^*) = 0\}, f_i \in CC(x^*; \{p\})$. More precisely, we have the following proposition.

PROPOSITION 3.3 *If for all* $i \in I(x^*)$, $\alpha_p(f_i; x^*) > 0$, *then* $f(x_k)$ *tends to* 0 *superlinearly; if for all* $i \in I(x^*)$, $f_i \in CC(x^*; \{p\})$, *then* x_k *tends to* x^* *superlinearly.*

Proof First of all, we note that $I(x^*) \neq \emptyset$ from $f(x_k) > 0$ and $f(x_k) \to f(x^*) = 0$. On the other hand, for $i \notin I(x^*)$, since $f_i(x^*) < 0$, so there exists $r_2 > 0$, such that for all $x \in \mathrm{O}(x^*; r_2)$, $f_i(x) < 0$. This implies that for $i \notin I(x^*)$,

$$f_i(x) \neq f(x) \quad \text{for all } x \in \mathrm{O}(x^*; r_2).$$

Hence, if for all $i \in I(x^*)$, $\alpha_p(f_i; x^*) > 0$, then

$$\begin{aligned}
\alpha_p(f; x^*) &= \lim_{x \to x^*} \inf \left\{ \frac{f(x) - f(x^*)}{\|x - x^*\|^p} : \ f(x) > f(x^*) \right\} \\
&\geq \lim_{x \to x^*} \inf \left\{ \frac{f_i(x) - f_i(x^*)}{\|x - x^*\|^p} : \ i \in I(x^*), f_i(x) = f(x) > 0 \right\} \\
&\geq \min\{\alpha_p(f_i; x^*) : \ i \in I(x^*)\} > 0.
\end{aligned}$$

Therefore, the first conclusion follows Theorem 3.3.

By the same approach, if for all $i \in I(x^*), f_i \in CC(x^*; \{p\})$, we have

$$\beta_p(f; x^*) < +\infty.$$

The second conclusion follows Theorem 3.3.

Example 3.4 Another case of Example 3.1 is finding the zero points of a given function $F(x) : R^n \to R^m$. More precisely, find $x^* \in R^n$, such that

$$F(x^*) = 0$$

if $f(x) = \frac{1}{2}F(x)^T F(x)$ is a convex function (example: $F(x) = Ax + b$, $A \in R^{n \times m}$). For this problem, we have the following result.

PROPOSITION 3.4 *Suppose that f is a convex function from $R^n \to R$,*

$$\lim_{x \to x^*} \inf \left\{ \frac{\|F(x)\|^2}{\|x - x^*\|^p} : F(x) \neq 0 \right\} > 0.$$

Then $\|F(x_k)\|$ tends to 0 superlinearly. Furthermore, if the following condition

$$\lim_{x \to x^*} \sup \left\{ \frac{\|F(x)\|^2}{\|x - x^*\|^p} : F(x) \neq 0 \right\} < +\infty$$

is added, then x_k tends to x^* superlinearly.

4 CONCLUDING REMARKS

The original motivation for the Moreau–Yosida regularization was to solve ill-conditioned systems of linear equations. It is noticed that Algorithm B always converges superlinearly when $p < 2$, $c_k = constant$, $\delta_k = \delta < 1$ and, for a large class of functions, a smaller p (< 2) implies greater ill-conditioning. Comparing with other methods (for example, Newton or quasi-Newton methods), the proximal point method takes advantage of solving ill-conditioned problems (also see [13,17,7] for more details on this topic).

The benefit of Algorithm B can be explained by using the univariate function $f(x) = \frac{1}{3}|x|^3$. For this function, the algorithm given in [12] may not converge superlinearly when starting from $x_1 > 0$ (see Proposition 15 of [12]). If we choose $c_k = [f(x_k)]^l$ ($l < -\frac{1}{3}$ is a constant) and apply Algorithm B, then $\{x_k\}_{k=1}^{\infty}$ converges to the solution 0 superlinearly by the results of Example 3.1. Also, for this function, since (25) does not hold, the superlinear convergence results of [20] cannot be applied.

The results of Corollaries 2.2 and 3.1 are important for constructing a *good* algorithm in the spirit of the *variable metric* methods (see [12]). In [12], the authors gave a way to construct c_k such that $c_k \to +\infty$. They proved the algorithm possesses superlinear convergence under (25) when f is an LC^1 function (in this case, $f \in CC(x^*; \{2\})$). Since Algorithm 13 of [12] is a special case of Algorithm B (if for all k, $\delta_k = 0$), x_k tends to x^* superlinearly from Theorem 3.3. On the other hand, if we let $c_k = \|x_k - x_{k-1}\|^{-\tau_3}$ in Algorithm B, then we can expect that x_k tends to x^* superlinearly even if f is not a smooth function using Corollary 3.1. Hence, the results of Corollary 3.1 provide a rule to construct the variable matrices based on the ideas of [12] (for penalty-type bundle algorithms) and quasi-Newton formulae.

It is not very hard to give a partial proximal minimization algorithm (see [3]) by a way similar to Algorithm 1. In fact, in Algorithm 1, if we let g_{k+1} be the vector with components zero for $i \in I$ (where I is a subset of the index set $\{1, \ldots, n\}$, see [3]) then one of the special cases of the PPM algorithm given by [3] is contained in the Algorithm 1. The reader can go into this direction in details by following [3].

The essential content of this paper is a theoretical investigation of algorithms for nonsmooth convex optimization. We extended convergence results from those previously given but one question remains.

How can one solve (2) effectively if f is a nonsmooth function?

Acknowledgement

The authors are thankful to the referee for his or her helpful comments.

References

[1] D.P. Bertsekas, Necessary and sufficient conditions for a penalty method to be exact, *Mathematical Programming*, **9** (1975), 87–99.
[2] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[3] D.P. Bertsekas and P. Tseng, Partial proximal minimization algorithms for convex programming, *SIAM Journal on Optimization*, **4** (1994), 551–572.

[4] J.R. Birge, L. Qi and Z. Wei, Convergence analysis of some methods for minimizing a nonsmooth convex function, *Journal of Optimization Theory and Applications*, to appear.

[5] G. Chen and M. Teboulle, A proximal-based decomposition method for convex minimization problems, *Mathematical Programming*, **64** (1994), 81–101.

[6] J. Eckstein and D.P. Bertsekas, On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, **55** (1992), 293–318.

[7] M. Fukushima and L. Qi, A globally and superlinearly convergent algorithm for nonsmooth convex minimization, *SIAM Journal on Optimization*, **6** (1996), 1106–1120.

[8] O. Güler, On the convergence of the proximal point algorithm for convex minimization, *SIAM Journal on Control and Optimization*, **29** (1991), 403–419.

[9] O. Güler, New Proximal point algorithms for convex minimization, *SIAM Journal on Optimization*, **4** (1992), 649–664.

[10] C.D. Ha, A generalization of the proximal point algorithm, *SIAM Journal on Control and Optimization*, **28** (1990), 503–512.

[11] B.W. Kort and D.P. Bertsekas, Combined primal–dual and penalty methods for convex programming, *SIAM Journal on Control and Optimization*, **2** (1976), 268–294.

[12] C. Lemaréchal and C. Sagastizábal, An approach to variable metric methods, in: *Systems Modelling and Optimization*, J. Henry and J.-P. Yvon, Eds., no. 197 in Lecture Notes in Control and Information Sciences, Springer-Verlag, 1994.

[13] C. Lemaréchal and C. Sagastizábal, Practical aspects of the Moreau–Yosida regularization I: Theoretical properties, *SIAM Journal on Optimization*, to appear.

[14] F.J. Luque, Asymptotic convergence analysis of the proximal point algorithm, *SIAM Journal on Control and Optimization*, **22** (1984), 277–293.

[15] B. Martinet, Perturbation des methods d'optimisation: Application, *RAIRO Anal. France*, **93**(12) (1978), 153–171.

[16] J.J. Moreau, Proximite et dualite dans un espace hilbertien, *Bulletin de la Societe Mathematique de France*, **93** (1965), 273–299.

[17] L. Qi, Second-order analysis of the Moreau–Yosida regularization of a convex function, AMR 94/20, Applied Mathematics Report, University of New South Wales, Revised, 1995.

[18] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[19] R.T. Rockafellar, A dual approach to solving nonlinear programming problems by unconstrained minimization, *Mathematical Programming*, **5** (1973), 354–373.

[20] R.T. Rockafellar, Monotone operators and proximal point algorithm, *SIAM Journal on Control and Optimization*, **14** (1976), 877–898.

[21] R.T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, *Mathematics of Operations Research*, **1** (1976), 97–116.

[22] C. Zhu, Modified proximal point algorithm for extended linear–quadratic programming, *Computational Optimization and Applications*, **1** (1992), 185–206.

[23] C. Zhu, Asymptotic convergence analysis of the forward–backward splitting algorithm, *Mathematics of Operations Research*, **20** (1995), 449–464.

[24] C. Zhu, Asymptotic convergence analysis of some inexact proximal point algorithms for minimization, *SIAM Journal on Optimization*, to appear.