A UNIFIED APPROACH TO CONVERGENCE PROPERTIES OF SOME METHODS FOR NONSMOOTH CONVEX OPTIMIZATION

John R. Birge
Department of Industrial & Operations Engineering
The University of Michigan
Ann Arbor, Michigan 48109-2117

Liqun Qi and Zengxin Wei
School of Mathematics
University of New South Wales
Sydney, NSW 2052
Australia

Technical Report 95-28

November 1995

A Unified Approach to Convergence Properties of Some Methods for Nonsmooth Convex Optimization [1]

JOHN R. BIRGE
Department of Industrial and Operations Engineering
University of Michigan, Ann Arbor, MI 48109, USA

LIQUN QI and ZENGXIN WEI
School of Mathematics,
University of New South Wales, Sydney NSW 2052, Australia

November 10, 1995

Abstract. Based on the notion of the ϵ-subgradient, we present a unified technique to establish convergence properties of several methods for nonsmooth convex minimization problems. Starting from the technical results, we obtain the global convergence of: (i) the variable metric proximal methods presented by Bonnans, Gilbert, Lemaréchal and Sagastizábal, (ii) some algorithms proposed by Correa and Lemaréchal and (iii) proximal point algorithm given by Rockafellar. In particular, we prove that the Rockafellar-Todd phenomenon does not occur for each of the above mentioned methods. Moreover, we explore the convergence rate of $\{\|x_k\|\}$ *and* $\{f(x_k)\}$ *when* $\{x_k\}$ *is unbounded and* $\{f(x_k)\}$ *is bounded for the nonsmooth* minimization methods (i), (ii) and (iii).

Key Words. Nonsmooth convex minimization, global convergence, convergence rate.

AMS Subject classifications. 90C25, 90C30, 90C33.

1. Introduction

To establish convergence of algorithms for convex minimization, a usual assumption is, at least, the existence of a minimum. This assumption has been removed for some methods [11, 9, 8, 10, 17, 6, 15, 16]. The study of the minimizing sequence was pioneered by Auslender, Crouzeix and their colleagues [3, 2, 1], the relation between minimizing and stationary sequences of unconstrained and constrained optimization problems has appeared recently, (see [5]). Similar results for complementarity problems and variational inequalities appeared in [7]. These papers were motivated by examples that presented in

[1]The research is supported by the National Science Foundation under Grant DDM-9215921 and the Australian Research Council.

Rockafellar [12] and Todd [14], which show that in general a stationary sequence is not necessarily a minimizing sequence.

Todd's example has the following properties:

- $h : R^n \to R$ is convex and continuously differentiable.
- The sequence $\{h(x_k)\}$ is monotonically decreasing and $\lim_{k\to\infty} \nabla h(x_k) = 0$.
- $\lim_{k\to\infty} h(x_k) > \inf_{x\in R^n} h(x)$.

We call the above phenomenon the Rockafellar-Todd (RT) phenomenon. Since most optimization algorithms produce a sequence $\{x_k\}$ that is only stationary, i.e., $\lim_{k\to\infty} \nabla h(x_k) = 0$, it is therefore important to know what kind of algorithms generate such sequences that are minimizing, i.e., $\lim_{k\to\infty} h(x_k) =$ the infimal value of h.

The purpose of this paper is to propose a general model algorithm for minimizing a proper lower-semicontinuous extended-valued convex function $f : R^n \to R\bigcup\{\infty\}$ and to establish the convergence properties without any additional assumption on f. We shall focus on two aspects: the RT phenomenon and the convergence rates of $\{\|x_k\|\}$ and $\{f(x_k)\}$ when $\{x_k\}$ is unbounded and $\{f(x_k)\}$ is bounded from below. These two issues have not been discussed in the literature.

Let $\|x\|^2$ denote the Euclidean norm of the vector $x \in R^n$. The subdifferential of f at x is a nonempty convex compact set

$$\partial f(x) = \{g : g \in R^n, f(y) \geq f(x) + < g, y - x >, \quad \text{for all } y \in R^n\}. \tag{1.1}$$

For any $\epsilon \geq 0$, let

$$\partial_\epsilon f(x) = \{g : g \in R^n, f(y) \geq f(x) + < g, y - x > -\epsilon, \text{ for all } y \in R^n\}. \tag{1.2}$$

We are interested in estimating the ϵ-minimum value of f, say f^*_ϵ, and also in identifying a ϵ-minimum point, for example, x^*, where

$$f(x) \geq f^*_\epsilon - \epsilon, \text{ for all } x \in R^n, \tag{1.3}$$

and

$$f(x) \geq f(x^*) - \epsilon, \text{ for all } x \in R^n. \tag{1.4}$$

Let $f^* = f^*_0$, the infimal value of f, $R^+ = \{\alpha \in R : \alpha > 0\}$ and $R^+_0 = R^+ \bigcup\{0\}$. With the above notation, we may now state the method in detail:

Algorithm 1:

Let $x_1 \in R^n$ be given. At the kth iteration, given $x_k \in R^n$, generate $(\epsilon_{1,k}, \epsilon_{2,k}, t_k, x_{k+1}, g_{k+1}) \in R^+_0 \times R^+_0 \times R^+ \times R^n \times R^n$ and $g_{k+1} \in \partial_{\epsilon_{2,k}} f(x_{k+1})$ satisfying the following inequality:

$$f(x_{k+1}) \leq f(x_k) - t_k < g_{k+1}, g_{k+1} > +\epsilon_{1,k}. \tag{1.5}$$

The remainder of the paper is organized as follows. In Section 2, we give some basic global convergence results for Algorithm 1 without any additional assumption on f. We in particular give a sufficient condition for avoiding the RT phenomenon. In Section 3, we discuss the convergence rate of Algorithm 1. In Section 4, we demonstrate that a number of methods for convex optimization problems are special cases of Algorithm 1.

In addition to results on the convergence rates of $\{\|x_k\|\}$ and $\{f(x_k)\}$, a class of descent algorithms for minimizing a continuously differentiable function is studied in [16].

2. Global Convergence

Theorem 2.1 *Let* $(\epsilon_{1,k}, \epsilon_{2,k}, t_k, x_{k+1}, g_{k+1})$ *be any sequence generated by Algorithm 1,* $\sum_{k=1}^{\infty} \epsilon_{1,k} < +\infty$ *and* $\sum_{k=1}^{\infty} t_k = +\infty$.

(i) Either $\lim_{k\to\infty} f(x_k) = -\infty$ *or* $\liminf_{k\to\infty} \|g_k\| = 0$. *In particular, if* $\{x_k\}$ *is bounded, then every accumulation point,* $x^* \in R^n$ *is an* ϵ^**-minimum point of* f, $f^*_{\epsilon^*} = f(x^*)$ *and* $\lim_{k\to\infty} f(x_k) = f^*_{\epsilon^*}$, *where*

$$\epsilon^* = \sup\{ \limsup_{k\in K_1, k\to\infty} \epsilon_{2,k-1} : K_1 \ \textit{ is an index set such that } \lim_{k\in K_1, k\to\infty} \|g_k\| = 0\}. \tag{2.1}$$

(ii) If $\inf\{t_k\} > 0$, *then either* $\lim_{k\to\infty} f(x_k) = -\infty$ *or* $\|g_k\| \to 0$. *In this case, every accumulation point of* $\{x_k\}$ *(if one exists) is an* ϵ^**-minimum point of* f.

(iii) If $\epsilon_{2,k} \to 0$ *and there exists a sequence* $\{m_k\}$ *with* $m_k \ge m > 0$ *such that for all large* k,

$$m_k\|x_{k+1} - x_k\| \le t_k\|g_{k+1}\|, \tag{2.2}$$

then $f(x_k) \to f_0^*$. *Furthermore, if for all large* k,

$$x_{k+1} - x_k = t_k g_{k+1}, \tag{2.3}$$

then $\{x_k\}$ *converges to a minimum point of* f *if one exists.*

Proof: It is clear that either $\lim_{k\to\infty} f(x_k) = -\infty$ or $\{f(x_k)\}$ is a convergent sequence.

(i) Suppose that $f(x_k)$ is bounded from below. From (1.5), we obtain

$$\sum_{k=1}^{\infty}[f(x_{k+1}) - f(x_k)] > -\infty$$

which implies that

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{2.4}$$

Thus, there exists an infinite index set K_1, such that $\lim_{k\in K_1, k\to\infty} \|g_k\| = 0$. If $\{x_k\}$ is a bounded set, then $\{x_k\}$ has at least one accumulation point. Since $\{x_k : k \in K_1\}$ is also a bounded set, without loss of generality, we may assume that $\lim_{k\in K_1,k\to\infty}\|x_k - x^*\| = 0$. Applying the ϵ-subgradient inequality,

$$f(x) \geq f(x_k) + < g_k, x - x_k > -\epsilon_{2,k-1}, \tag{2.5}$$

we have that, for all $x \in R^n$,

$$f(x) \geq f(x^*) - \epsilon^*.$$

This implies that $f(x^*) = f^*_{\epsilon^*}$. Conclusion (i) follows from $f(x_k) \to f(x^*)$.

(ii) Suppose that $f(x_k)$ is bounded from below. Using (1.5), we have

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{2.6}$$

Let $\{x_k : k \in K\}$ be any convergent sequence of $\{x_k\}$, $\lim_{k\in K,k\to\infty}\|x_k - x^*\| = 0$. For these $\{x_k : k \in K\}$, by (2.5) and (2.6), we have that x^* is an ϵ^*-minimum point of f.

(iii) From (i) of this theorem, it sufficies to consider the case that $\{f(x_k)\}$ is bounded and $\{x_k\}$ is unbounded. Suppose that there exist $\bar{x} \in R^n, \tau > 0$ and k_0 such that for all $k \geq k_0$,

$$< g_k, \bar{x} - x_k > < -\tau.$$

This inequality and (1.5) imply that

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq -t_k\|g_{k+1}\|^2 + \epsilon_{1,k} \\ &\leq t_k\|g_{k+1}\|\tfrac{<g_{k+1},\bar{x}-x_{k+1}>}{\|\bar{x}-x_{k+1}\|} + \epsilon_{1,k}. \end{aligned}$$

Therefore, we have that for all k,

$$f(x_{k+1}) - f(x_k) \leq -\tau t_k \frac{\|g_{k+1}\|}{\|\bar{x} - x_{k+1}\|} + \epsilon_{1,k}. \tag{2.7}$$

¿From (2.2),

$$\|x_{k+1} - x_1\| \leq \sum_{i=1}^{k} \|x_{i+1} - x_i\| \leq m^{-1}\sum_{i=1}^{k} t_i\|g_{i+1}\|,$$

which implies that

$$\sum_{k=1}^{\infty} t_k\|g_{k+1}\| = +\infty \tag{2.8}$$

by using the unboundedness of $\{\|x_k\|\}$. Therefore, there exists k_1, such that for all $k > k_1$,

$$\|x_1 - \bar{x}\| \leq m^{-1}\sum_{i=1}^{k} t_i\|g_{i+1}\|.$$

Hence for all $k > k_1$,

$$\|x_{k+1} - \bar{x}\| \leq \|x_1 - \bar{x}\| + \sum_{i=1}^{k} \|x_{i+1} - x_i\| \leq 2m^{-1}\sum_{i=1}^{k} t_i\|g_{i+1}\|. \tag{2.9}$$

¿From (2.7) and (2.9), we have

$$f(x_{k+1}) - f(x_k) \le -\frac{m\tau}{2} \frac{t_k \|g_{k+1}\|}{\sum_{i=1}^{k} t_i \|g_{i+1}\|} + \epsilon_{1,k}. \tag{2.10}$$

This inequality, (2.8) and Lemma 3.1 given in [17] yield that

$$\sum_{k=1}^{\infty} [f(x_{k+1}) - f(x_k)] = -\infty,$$

which contradicts that $\{f(x_k)\}$ is bounded from below. Therefore, for all $x \in R^n$,

$$\limsup_{k \to \infty} < g_k, x - x_k > \ge 0. \tag{2.11}$$

The ϵ-subgradient inequality,

$$f(x) - f(x_k) \ge < g_k, x - x_k > -\epsilon_{2,k-1}$$

and (2.11) yield that for all $x \in R^n$,

$$f(x) \ge \limsup_{k \to \infty} f(x_k) \ge f_0^*.$$

This implies that $f(x_k) \to f_0^*$, which completes the proof of the first conclusion of (iii).

We now prove the second conclusion of (iii). It is easy to verify that for all $x \in R^n$ and all k:

$$\|x_k - x\|^2 = \|x_{k+1} - x\|^2 + \|x_k - x_{k+1}\|^2 + 2 < x_k - x_{k+1}, x_{k+1} - x > . \tag{2.12}$$

¿From (2.3) and the definition of g_{k+1}, for all large k,

$$< x_k - x_{k+1}, x_{k+1} - x > \ge a t_k [f(x_{k+1}) - f(x) - \epsilon_{2,k}].$$

Combining it with (2.12), we have,

$$\|x_k - x\|^2 \ge \|x_{k+1} - x\|^2 + a^2 t_k^2 \|g_{k+1}\|^2 + 2 a t_k [f(x_{k+1}) - f(x) - \epsilon_{2,k}]. \tag{2.13}$$

Using any given minimum point x^* in (2.13), we have

$$\|x_k - x^*\|^2 \ge \|x_{k+1} - x^*\|^2 - 2 a t_k \epsilon_{2,k}.$$

This implies that $\{x_k\}$ is bounded and the existence of

$$\lim_{k \to \infty} \|x_k - x^*\|^2 = l < +\infty.$$

Suppose there are two accumulation points x_1^* and x_2^* of $\{x_k\}$. From the first part of this theorem, x_1^* and x_2^* are minimum points of f. As above, we have the existences of

$$\lim_{k \to \infty} \|x_k - x_i^*\|^2 = l_i < +\infty, \text{ for } i = 1, 2.$$

This implies that $l_i = 0$ for $i = 1, 2$ and $x_1^* = x_2^*$. Therefore $\{x_k\}$ has a unique accumulation point. □

3. Local Convergence

In this section, we discuss the convergence rate of Algorithm 1 in the following two cases:
Case 1: x^* minimizes f and $\lim_{k\to\infty} x_k = x^*$.
Case 2: A global minimizer of f does not exist but $\inf_{x\in R^n} f(x) > -\infty$. In this case, $\{\|x_k\|\}$ is unbounded and $f_0^* > -\infty$.

Let
$$a_k = \frac{\epsilon_{1,k}}{\|g_{k+1}\|^2}, b_k = \frac{\epsilon_{2,k}}{\|g_{k+1}\|},$$
and
$$c_k = (t_k - a_k)\frac{f(x_k) - f(x^*)}{(b_k + \|x_{k+1} - x^*\|)^2}.$$

Theorem 3.1 *Suppose that* $\inf\{t_k\} > 0, x^*$ *minimizes* f *and* $\lim_{k\to\infty} x_k = x^*$. *If, for all* k, $t_k \geq a_k$, *then*
$$\frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} \leq \sqrt{1 + (\frac{1}{2}c_k)^2} - \frac{1}{2}c_k. \tag{3.1}$$
Consequently,

(i) if for all $k, c_k \geq c^* \in (0, +\infty)$, *then*
$$\overline{\lim_{k\to\infty}} \frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} \leq \sqrt{1 + (\frac{1}{2}c^*)^2} - \frac{1}{2}c^* < 1.$$

(ii) if $\lim_{k\in K, k\to\infty} c_k = +\infty$, *then*
$$\lim_{k\in K, k\to\infty} \frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} = 0.$$

Proof: Using $g_{k+1} \in \partial_{\epsilon_{2,k}} f(x_{k+1})$, we have
$$f(x^*) \geq f(x_{k+1}) + g_{k+1}^T(x^* - x_{k+1}) - \epsilon_{2,k} \geq f(x_{k+1}) - \|g_{k+1}\|\|x^* - x_{k+1}\| - b_k\|g_{k+1}\|.$$
This implies
$$\|g_{k+1}\| \geq \frac{f(x_{k+1}) - f(x^*)}{b_k + \|x_{k+1} - x^*\|}$$
which combined with (1.5) yields
$$\begin{aligned} f(x_k) &\geq f(x_{k+1}) + t_k\|g_{k+1}\|^2 - \epsilon_{1,k} \\ &= f(x_{k+1}) + t_k\|g_{k+1}\|^2 - a_k\|g_{k+1}\|^2 \\ &\geq f(x_{k+1}) + \tfrac{t_k - a_k}{(b_k + \|x_{k+1} - x^*\|)^2}[f(x_{k+1}) - f(x^*)]^2. \end{aligned}$$
Hence
$$f(x_k) - f(x^*) \geq [f(x_{k+1}) - f(x^*)][1 + \frac{(t_k - a_k)(f(x_{k+1}) - f(x^*))}{(b_k + \|x_{k+1} - x^*\|)^2}].$$

Therefore

$$\frac{f(x_{k+1})-f(x^*)}{f(x_k)-f(x^*)} \leq \frac{1}{1+\frac{f(x_{k+1})-f(x^*)}{f(x_k)-f(x^*)}\frac{(t_k-a_k)(f(x_k)-f(x^*))}{(b_k+\|x_{k+1}-x^*\|)^2}}.$$

Thus, (3.1) follows.

Since $\sqrt{1+(\frac{1}{2}t)^2}-\frac{1}{2}t$ ($t \geq 0$) is a decreasing function, by (3.1), we have conclusions (i) and (ii). □

The following theorem extends the related result of [16] for smooth optimization to the case where f is only a proper lower-semicontinuous extended-valued function.

Theorem 3.2 *Suppose that $\{x_k\}$ is generated by Algorithm 1 with $\epsilon_{2,k} \to 0$ and $\sum_{k=1}^{\infty}\sqrt{\epsilon_{1,k}} < +\infty$. Suppose that (2.2) holds and $\{t_k\}$ is a bounded set. If $\{x_k\}$ is unbounded and $\{f(x_k)\}$ is bounded, then the rate of $\sqrt{f(x_k)-f_0^*}$ converging to zero is less than geometric. Furthermore, $\{\frac{\|x_k\|^2}{k}\}$ is bounded.*

Proof: From (2.2), we have for all k

$$\begin{aligned} \|x_{k+1}-x_1\| &\leq \textstyle\sum_{i=1}^{k}\|x_{i+1}-x_i\| \\ &\leq m^{-1}\textstyle\sum_{i=1}^{k}\|t_i g_{i+1}\|, \end{aligned}$$

which implies that

$$\sum_{k=1}^{\infty}\|t_k g_{k+1}\| = +\infty, \tag{3.2}$$

using that $\{x_k\}$ is unbounded.

Since for all $k, f(x_k) \geq f_0^*$ and $\{t_k\}$ is bounded, we obtain from (1.5) that

$$\begin{aligned} f_0^*-f(x_k) &\leq f(x_{k+1})-f(x_k) \\ &\leq -t_k\|g_{k+1}\|^2+\epsilon_{1,k} \\ &\leq -\tfrac{1}{t_k}(t_k\|g_{k+1}\|)^2+\epsilon_{1,k}. \end{aligned}$$

Hence

$$\sqrt{f(x_k)-f_0^*} \geq \max\{0, \sqrt{\frac{1}{t_k}}(t_k\|g_{k+1}\|)-\sqrt{\epsilon_{1,k}}\}.$$

This inequality and our assumptions on t_k and $\epsilon_{1,k}$ yield

$$\sum_{k=1}^{\infty}\sqrt{f(x_k)-f_0^*} = +\infty,$$

which implies that $\sqrt{f(x_k) - f_0^*}$ cannot converge to 0 with geometric rate.

We now prove the second part of the theorem. From (1.5) and (2.2),

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq -t_k\|g_{k+1}\|^2 + \epsilon_{1,k} \\ &\leq -\tfrac{m^2}{t_k}\|x_{k+1} - x_k\|^2 + \epsilon_{1,k}, \end{aligned}$$

which implies

$$f(x_{k+1}) - f(x_1) \leq -\inf\{\frac{m^2}{t_i} : i = 1, ..., k,\}\sum_{i=1}^{k}\|x_{i+1} - x_i\|^2 + \sum_{i=1}^{k}\epsilon_{1,i}.$$

On the other hand,

$$\begin{aligned} \|x_{k+1} - x_1\|^2 &\leq (\textstyle\sum_{i=1}^k \|x_{i+1} - x_i\|)^2 \\ &\leq k\textstyle\sum_{i=1}^k \|x_{i+1} - x_i\|^2. \end{aligned}$$

Hence we obtain the following inequality by combining the above two inequalities

$$f(x_{k+1}) - f(x_1) \leq -\inf\{\frac{m^2}{t_i} : i = 1, ..., k,\}\frac{\|x_{k+1} - x_1\|^2}{k} + \sum_{i=1}^{k}\epsilon_{1,i},$$

which implies that $\{\frac{\|x_{k+1}-x_1\|^2}{k}\}$ is bounded since $\{f(x_k)\}$ is bounded from below. Therefore, $\{\frac{\|x_k\|^2}{k}\}$ is bounded. □

4. Applications

In this section, we demonstrate that a number of methods for convex optimization problems are special cases of Algorithm 1. These contain

- *A family of variable metric proximal methods proposed in [4].*
- *The methods for convex minimization given in [6].*
- *Proximal point algorithms introduced in [13].*

Example 4.1 A family of variable metric proximal methods [4].

In [4], the authors proposed a family of variable metric proximal algorithms based on the Moreau-Yosida regularization and quasi-Newton approximations. Given $x \in R^n$ and a symmetric positive definite $n \times n$ matrix B, let

$$\varphi_B(z) := f(z) + \frac{1}{2} < B(z - x), z - x >, \tag{4.1}$$

$$x^p = p_B(x) := \text{argmin}\{\varphi_B(z) : z \in R^N\}, \tag{4.2}$$

$$\delta_k := f(x_k) - f(x_k^p) - \frac{1}{2} < g_k^p, W_k g_k^p >, \tag{4.3}$$

where $W_k = B_k^{-1}$, $g_k^p \in \partial f(x_k^p)$.

With the notation in (4.1)-(4.3), we can state the algorithm of [4] as follows:

Algorithm 4.1: (GAP of [4])
Step 0: Start with some initial point x_1 and matrix B_1; choose some parameter $m_0 \in (0,1)$; set k=1.
Step 1: With δ_k given by (4.3), compute x_{k+1} satisfying

$$f(x_{k+1}) \leq f(x_k) - m_0\delta_k. \tag{4.4}$$

Step 2: Update B_k, increase k by 1 and loop to Step 1.

Lemma 4.1 *Suppose that (x_{k+1}, g_k^p) is generated by Algorithm 4.2. Let*

$$\epsilon_{2,k} = max\{0, [f(x_{k+1}) - f(x_k) - < g_k^p, x_{k+1} - x_k >] + [f(x_k) - f(x_k^p) - < g_k^p, W_k g_k^p >]\}, \tag{4.5}$$

then

$$g_k^p \in \partial_{\epsilon_{2,k}} f(x_{k+1}) \tag{4.6}$$

and

$$f(x_{k+1}) \leq f(x_k) - \frac{m_0}{2} < g_k^p, W_k g_k^p > . \tag{4.7}$$

Thus, (1.5) holds with $t_k = \frac{m_0}{2}\lambda_{\min}(W_k)$, where $\lambda_{\min}(W)$ denotes the smallest eigenvalue of a symmetric matrix W.

Proof: Using $g_k^p \in \partial f(x_k^p)$, we have for all $x \in R^N$:

$$\begin{aligned} f(x) &\geq f(x_k^p) + < g_k^p, x - x_k^p > \\ &= f(x_{k+1}) + < g_k^p, x - x_{k+1} > - [f(x_{k+1}) - f(x_k) - < g_k^p, x_{k+1} - x_k >] \\ &\quad - [f(x_k) - f(x_k^p) - < g_k^p, x_k - x_k^p >]. \end{aligned} \tag{4.8}$$

On the other hand, from the definitions of x_k^p and g_k^p, we have

$$x_k^p = x_k - W_k g_k^p. \tag{4.9}$$

Substituting (4.9) into (4.8), (4.6) follows.

Since for all $x \in R^N$,

$$f(x_k^p) + \frac{1}{2} < B_k(x_k^p - x_k), x_k^p - x_k > \leq f(x) + \frac{1}{2} < B_k(x - x_k), x - x_k > .$$

Setting $x = x_k$, we have

$$f(x_k^p) \leq f(x_k) - < g_k, W_k g_k^p > \tag{4.10}$$

by (4.9). Relations (4.10), (4.3) and (4.4) imply (4.7). □

Conclusion (a) in the Theorem 4.1 is the global convergence result of [5], while conclusion (b) is a new one for this algorithm. Indeed, the assumptions in (b) can be satisfied by viewing (qN-AP3) (or BFGS-AP3) and the Remark 4.1 of [4].

Theorem 4.1 *(a) (Theorem 2.3 of [5]). Assume that f has a nonempty bounded set of minima, and let $\{x_k\}$ be a sequence generated by (GAP). Then $\{x_k\}$ is bounded and, if*

$$\sum_{k=1}^{\infty} \lambda_{\min}(W_k) = +\infty, \tag{4.11}$$

any accumulation point of $\{x_k\}$ minimizes f. The same properties hold for the sequence of proximal points $\{x_k^p\}$. It also holds that $\liminf_{k\to\infty} \|g_k^p\| = 0$.

(b) Suppose there exists $\bar{t}_k$ such that for all large k,

$$x_{k+1} = x_k + \bar{t}_k(x_k^p - x_k), \tag{4.12}$$

where $\bar{t}_k \leq \bar{t} < +\infty$. If $\{\|W_k\|\}$ is bounded and (4.11) holds, then $f(x_k) \to f_0^*$.

Proof: (a) Since f has a nonempty bounded set of minima, the level sets of f are bounded. Hence, $\{x_k\}$ and $\{x_k^p\}$ are bounded by (4.7) and (4.9). (In fact, this conclusion follows due to [5, Theorem 2.3].) Using (4.4), we have

$$\delta_k \to 0. \tag{4.13}$$

¿From (4.7), we obtain

$$< g_k^p, W_k g_k^p > \to 0. \tag{4.14}$$

Hence,

$$f(x_k) - f(x_k^p) - < g_k, W_k g_k^p > = \delta_k - \frac{1}{2} < g_k^p, W_k g_k^p > \to 0. \tag{4.15}$$

Results (4.13), (4.14) and (4.15) imply that if $\lim_{k\in K} \|g_k^p\| = 0$, then $\lim_{k\in K} \epsilon_{2,k} = 0$. ¿From the definition of ϵ^*, we have $\epsilon^* = 0$. Let $t_k = \frac{m_0}{2}\lambda_{\min}(W_k)$, then (4.11) and (i) of Theorem 2.1 yield the first conclusion.

¿From (4.13) and (4.14), we have

$$\lim_{k\to\infty} f(x_k) = \lim_{k\to\infty} f(x_k^p).$$

This implies that every accumulation point of $\{x_k^p\}$ minimizes f by the first conclusion. Using (4.7) and (4.11), we have $\liminf_{k\to\infty} \|g_k^p\| = 0$.

(b) From (4.12) and (4.9), we have, for all large k, that

$$\frac{1}{t_k\|W_k\|}\|x_{k+1} - x_k\| \leq \|g_k^p\|,$$

which implies that $f(x_k) \to f_0^*$ by using the assumptions in (b) and result (iii) in Theorem 2.1. In fact, we only need to note that we can let $m_k = \frac{1}{t_k\|W_k\|}$ and let $t_k = 1$ for all large k. □

Theorem 4.2 *Suppose that the assumptions of (b) in Theorem 4.1 hold. If $\{x_k\}$ is unbounded and $f_0^* > -\infty$, then the rate of $\sqrt{f(x_k) - f_0^*}$ converging to zero is less than geometric and $\{\frac{\|x_k\|^2}{k}\}$ is bounded.*

Proof: The conclusions follow Lemma 4.1, Theorem 3.2 and (b) of Theorem 4.1. □

Example 4.2: Algorithms given in [6].

In [6], Correa and Lemaréchal presented a simple and unified technique to establish convergence of a number of minimization methods. These contain (i) the exact prox-iteration, (ii) its implementable approximations, which include in particular (iii) bundle methods, and finally (iv) the classical subgradient optimization scheme. Their methods can be summarized as follows:

Algorithm 4.2: From an arbitrary point $x_1 \in R^n$, the sequence $\{x_k\}$ constructed with the following formulas:

$$x_{k+1} = x_k - \tau_k \gamma_k, \tag{4.16}$$

$$\gamma_k \in \partial_{\epsilon_{3,k}} f(x_k), \tag{4.17}$$

$$f(x_{k+1}) \leq f(x_k) - m\tau_k\|\gamma_k\|^2, \tag{4.18}$$

where $\epsilon_{3,k}$ is nonnegative, $\tau_k > 0$ is the stepsize and m is a positive constant.

The following lemma shows that Algorithm 4.2 is a special case of Algorithm 1.

Lemma 4.2 *Suppose that $(\epsilon_{3,k}, \tau_k, x_{k+1}, \gamma_k)$ is generated by Algorithm 4.2. Let*

$$t_k = m\tau_k, \tag{4.19}$$

$$\epsilon_{2,k} = \max\{0, \epsilon_{3,k} + (1-m)\tau_k\|\gamma_k\|^2\}. \tag{4.20}$$

Then

$$\gamma_k \in \partial_{\epsilon_{2,k}} f(x_{k+1}) \tag{4.21}$$

and (1.5) holds for any $\epsilon_{1,k} \geq 0$.

Proof: It sufficies to prove that (4.21) holds. By (4.17), we have for all $x \in R^n$,

$$\begin{aligned} f(x) &\geq f(x_k) + < \gamma_k, x - x_k > -\epsilon_{3,k} \\ &= f(x_{k+1}) + < \gamma_k, x - x_{k+1} > + f(x_k) - f(x_{k+1}) + < \gamma_k, x_{k+1} - x_k > -\epsilon_{3,k}. \end{aligned}$$

This inequality, (4.16) and (4.18) imply that

$$f(x) \geq f(x_{k+1}) + < \gamma_k, x - x_{k+1} > -[\epsilon_{3,k} + (1-m)\tau_k \|\gamma_k\|^2].$$

So (4.21) follows. □

The following is a main result of [6].

Theorem 4.3 *(Proposition 2.2 of [6].) Suppose that* $(\epsilon_{3,k}, \tau_k, x_{k+1}, \gamma_k)$ *is generated by Algorithm 4.2.*

(i) Assume that

$$\sum_{k=1}^{\infty} \tau_k = +\infty, \tag{4.22}$$

$$\epsilon_{3,k} \to 0, \tag{4.23}$$

then $f(x_k) \to f_0^*$.

(ii) If $\{\tau_k\}$ *is bounded and*

$$\sum_{k=1}^{\infty} \epsilon_{3,k} < +\infty, \tag{4.24}$$

then $\{x_k\}$ *converges to a minimum point of* f *if there is such a minimum point.*

Proof: (i) If the decreasing $\{f(x_k)\}$ tends to $-\infty$, then the conclusion follows. Otherwise, from (4.18), we have

$$\sum_{k=1}^{\infty} \tau_k \|\gamma_k\|^2 < +\infty. \tag{4.25}$$

This and (4.20) imply that $\{\epsilon_{3,k}\}$ tends to 0 if and only if $\{\epsilon_{2,k}\}$ tends to 0. So the conclusion follows by Lemma 4.2 and (iii) of Theorem 2.1.

(ii) Suppose f has a minimum point. Then $\{f(x_k)\}$ is bounded from below. Thus, (4.25) holds. Inequalities (4.25) and (4.24) imply that

$$\sum_{k=1}^{\infty} t_k \epsilon_{2,k} < +\infty$$

by the boundedness of $\{\tau_k\}$ and (4.19). The results of (iii) in Theorem 2.1 imply that $\{x_k\}$ converges to a minimum point of f. $\square$

Note that from (4.16) and Theorem 3.2, we obtain the following new convergence rate for Algorithm 4.2.

Theorem 4.4 *Let $(\epsilon_{3,k}, \tau_k, x_{k+1}, \gamma_k)$ be generated by Algorithm 4.1 with (4.22) and (4.23). If $\{\tau_k\}$ is bounded, $\{x_k\}$ is unbounded and $f_0^* > -\infty$, then the rate of $\sqrt{f(x_k) - f_0^*}$ converging to zero is less than geometric and $\{\frac{\|x_k\|^2}{k}\}$ is bounded.*

Example 4.3: A proximal point algorithm introduced in [13].

In [13], Rockafellar introduced two general criteria for finding the zero of an arbitrary maximal monotone operator when the iteration points are given approximately. As an application, he applied the results to a lower semicontinuous proper convex function f. In this case, one of the algorithms follows:

Algorithm 4.3: For x_k, generate $(\sigma_k, \lambda_k, x_{k+1}, g_{k+1}) \in R^{+0} \times R^{+} \times R^n \times R^n$ ($g_{k+1} \in \partial f(x_{k+1})$) satisfying

$$\mathrm{dist}(0, S_k(x_{k+1})) \le \frac{\sigma_k}{\lambda_k}\|x_{k+1} - x_k\|, \tag{4.26}$$

where

$$\sum_{k=1}^{\infty} \sigma_k < \infty, \tag{4.27}$$

and

$$S_k(x) = \partial f(x) + \frac{1}{\lambda_k}(x - x_k). \tag{4.28}$$

In the following discussion, we only assume that for all k,

$$\sigma_k \in [0, \frac{1}{2}]. \tag{4.29}$$

Lemma 4.3 *Suppose that $(\sigma_k, \lambda_k, x_{k+1}, g_{k+1}) \in R^{+0} \times R^{+0} \times R^{+} \times R^n \times R^n$ is generated by Algorithm 4.3. Let*

$$t_k = \frac{1 - \sigma_k}{(1 + \sigma_k)^2}\lambda_k, \tag{4.30}$$

Then (1.5) holds for $\epsilon_{1,k} = \epsilon_{2,k} = 0$. Furthermore,

$$\frac{(1 - \sigma_k)^2}{(1 + \sigma_k)^2}\|x_{k+1} - x_k\| \le t_k\|g_{k+1}\|. \tag{4.31}$$

Proof: By (4.26), (4.29) and (4.28), we have

$$\|g_{k+1} + \frac{1}{\lambda_k}(x_{k+1} - x_k)\| \le \frac{\sigma_k}{\lambda_k}\|x_{k+1} - x_k\|. \tag{4.32}$$

The inequality,

$$< g_{k+1} + \frac{1}{\lambda_k}(x_{k+1} - x_k), x_{k+1} - x_k >\le \|g_{k+1} + \frac{1}{\lambda_k}(x_{k+1} - x_k)\|\|x_{k+1} - x_k\|,$$

and (4.32) imply that

$$< g_{k+1}, x_{k+1} - x_k >\le -\frac{1-\sigma_k}{\lambda_k}\|x_{k+1} - x_k\|^2.$$

Therefore,

$$< g_{k+1}, x_k - x_{k+1} >\ge \frac{1-\sigma_k}{\lambda_k}\|x_{k+1} - x_k\|^2. \tag{4.33}$$

On the other hand, by (4.32), we have

$$\|g_{k+1}\| \le \frac{1+\sigma_k}{\lambda_k}\|x_{k+1} - x_k\|. \tag{4.34}$$

Inequalities(4.33) and (4.34) yield

$$< g_{k+1}, x_k - x_{k+1} > \ge t_k < g_{k+1}, g_{k+1} > . \tag{4.35}$$

Applying the subgradient inequality for convex functions,

$$f(x_k) \ge f(x_{k+1}) + t_k\|g_{k+1}\|^2.$$

Hence, (1.5) follows.

¿From (4.32), we have

$$\frac{1-\sigma_k}{\lambda_k}\|x_{k+1} - x_k\| \le \|g_{k+1}\|,$$

which implies that (4.31) holds. □

The following theorem indicates that the RT phenomenon does not occur for the well-known Algorithm 4.3. In view of Theorem 2.1 and Lemma 4.3, it does not require proof.

Theorem 4.5 *Suppose that* $\{(\sigma_k, \lambda_k, x_{k+1}, g_{k+1})\}$ *is generated by Algorithm 4.3 with* $\sum_{k=1}^{\infty} \lambda_k = +\infty$.

(i) Either $\lim_{k\to\infty} f(x_k) = -\infty$ *or* $\liminf_{k\to\infty} \|g_k\| = 0$. *In particular, if* $\{x_k\}$ *is bounded, then* $f(x_k) \to f_0^*$ *and every accumulation point of* $\{x_k\}$ *is a minimum point of* f.

(ii) If $\inf\{\lambda_k\} > 0$*, then either* $\liminf_{k\to\infty} f(x_k) = -\infty$ *or* $\|g_k\| \to 0$*. In this case, every accumulation point of* $\{x_k\}$ *(if one exists) is a minimum point of* f*.*

(iii) $f(x_k) \to f_0^*$*.*

(iv) If for all $k, \sigma_k = 0$*, then* $\{x_k\}$ *converges to a minimum point of* f *if such one exists.*

For Algorithm 4.3, we obtain the following two basic convergence rate results from Theorem 3.1 and Theorem 3.2.

Theorem 4.6 *(a) Suppose that* x^* *minimizes* f*,* $x_k \to x^*$*, and there exist two scalars* $r > 0$ *and* $M > 0$ *such that for any* x *satisfying* $\|x - x^*\| \le r$*,*

$$f(x) - f(x^*) \ge M\|x - x^*\|^2. \tag{4.36}$$

Then

(a1) If

$$\lim_{k\to\infty} \lambda_k = \lambda^* \in (0, +\infty),$$

then $f(x_k)$ *tends to* $f(x^*)$ *linearly.*

(a2) If

$$\lim_{k\to\infty} \lambda_k = +\infty,$$

then $f(x_k)$ *tends to* $f(x^*)$ *superlinearly.*

(b) Suppose that $\sum_{k=1}^{\infty} \lambda_k = +\infty$*,* $\{\lambda_k\}$ *is bounded,* $\{x_k\}$ *is unbounded and* $f_0^* > -\infty$*. Then the rate of* $\sqrt{f(x_k) - f_0^*}$ *converging to zero is less than geometric and* $\{\frac{\|x_k\|^2}{k}\}$ *is bounded.*

Proof: We first prove (a). Since for all k,

$$c_k = \frac{1 - \sigma_k}{(1 + \sigma_k)^2} \lambda_k \frac{f(x_k) - f(x^*)}{\|x_{k+1} - x^*\|^2}$$

and

$$\frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} \le 1.$$

Hence, we have, from (4.29), that

$$c_k = \frac{1 - \sigma_k}{(1 + \sigma_k)^2} \lambda_k \frac{f(x_k) - f(x^*)}{f(x_{k+1}) - f(x^*)} \frac{f(x_{k+1}) - f(x^*)}{\|x_{k+1} - x^*\|^2} \ge \frac{2}{9} M \lambda_k.$$

which implies that (a1) and (a2) hold by Theorem 3.1.

(b) From the definition of m_k in (2.2), we have, for Algorithm 4.3, that

$$m_k = \frac{(1-\sigma_k)^2}{(1+\sigma_k)^2} \geq \frac{1}{9}$$

by (4.29). This implies that the results of (b) hold by using Theorem 3.2. $\square$

It is worth noting that the conclusions (iii) in Theorem 4.5 and (b) in Theorem 4.6 are not contained in the convergence results given in [13]. Since Algorithm 4.3 is different from those using linear search to produce the next iteration $x_{k+1} = x_k + t_k d_k$, where d_k is a linear search direction at the kth iteration, it is surprising that we can easily obtain the same convergence properties for these two types of methods. We believe that the tool in this paper is useful in the convergence analysis for optimization problems under a unified framework.

References

[1] A. Auslender, Convergence of stationary sequences for variational inequalities with maximal monotone operators, *Applied Mathematics and Optimization* 28 (1993) 161-172.

[2] A. Auslender, R. Cominetti and J.-P. Crouzeix, Convex functions with unbounded level sets and applications to duality theory, *SIAM Journal on Optimization* 3 (1993) 669-687.

[3] A. Auslender and J.-P. Crouzeix, Well behaved asymptotical convex functions, *Analyse Non-linémare* (1989) 101-122.

[4] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal and C. A. Sagastizábal, A family of variable metric proximal methods, *Mathematical Programming* 68 (1995) 15-47.

[5] C. C. Chou, K. F. Ng and J. S. Pang, Minimizing and stationary sequences for optimization problems, manuscript, Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, Maryland, August 1995.

[6] R. Correa and C. Lemaréchal, Convergence of some algorithms for convex minimization, *Mathematical Programming* 62 (1993) 261-275.

[7] M. Fukushima and J-S. Pang, Minimizing and stationary sequences of merit functions for complementarity problems and variational inequalities, manuscript, Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, Maryland, August 1995.

[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] K. C. Kiwiel, An aggregate subgradient method for nonsmooth convex minimization, *Mathematical Programming* 27 (1983) 320-341.

[10] B. Lemaire, About the convergence of the proximal method, in: D. Pallaschke, ed., *Advances in Optimization. Lecture Notes in Economics and Mathematics System*, (Springer,Berlin, 1992), 39-51.

[11] B. T. Poljak, A general methods for solving extremum problems, *Soviet Mathematics Doklady* 8 (1967) 593-597.

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

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

[14] M. J. Todd, On convergence properties of algorithms for unconstrained minimization, *IMA Journal of Numerical Analysis* 9 (1989) 435-441.

[15] Z. Wei and L. Qi, Convergence analysis of a proximal Newton method, AMR 95/33, Applied Mathematics Report, University of New South Wales, 1995.

[16] Z. Wei, L. Qi and H. Jiang, Some convergence properties of decent methods, AMR 95/34, Applied Mathematics Report, University of New South Wales, 1995.

[17] S. Q. Wu, Convergence properties of descent methods for unconstrained minimization, *Optimization* 26 (1992) 229-237.