Appl Math Optim 38:141–158 (1998)

Applied Mathematics & Optimization

A General Approach to Convergence Properties of Some Methods for Nonsmooth Convex Optimization*

J. R. Birge,[1] L. Qi,[2] and Z. Wei[2]

[1]Department of Industrial and Operations Engineering,
University of Michigan, Ann Arbor, MI 48109, USA

[2]School of Mathematics, University of New South Wales,
Sydney, NSW 2052, Australia

Communicated by P. Kall

Dedicated to Roger Wets on the event of his 60th birthday

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) the 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 Classification. 90C25, 90C30, 90C33.

* This research was supported by the National Science Foundation under Grant DDM-9215921 and the Australian Research Council.

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 presented by 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: $R^n \to R \cup \{\infty\}$ and to establish the convergence properties without any additional assumption on f. We 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 to our knowledge.

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

$$\partial f(x) = \{g : g \in R^n,\ f(y) \ge f(x) + \langle g, y - x\rangle,\ \text{for all } y \in R^n\}. \tag{1.1}$$

For any $\varepsilon \ge 0$, let

$$\partial_\varepsilon f(x) = \{g : g \in R^n,\ f(y) \ge f(x) + \langle g, y - x\rangle - \varepsilon,\ \text{for all } y \in R^n\}. \tag{1.2}$$

If a real number f_ε^* satisfies

$$f(x) \ge f_\varepsilon^* - \varepsilon, \qquad \text{for all} \quad x \in R^n, \tag{1.3}$$

then we say it is an ε-minimum value of f. If $x^* \in R^n$ satisfies

$$f(x) \ge f(x^*) - \varepsilon, \qquad \text{for all} \quad x \in R^n, \tag{1.4}$$

then we say that x^* is an ε-minimum point of f. Let $f^* = f_0^*$, the infimal value of f, $R^+ = \{\alpha \in R : \alpha > 0\}$, and $R_0^+ = R^+ \cup \{0\}$. 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 $(\varepsilon_{1,k}, \varepsilon_{2,k}, t_k, x_{k+1}, g_{k+1}) \in R_0^+ \times R_0^+ \times R^+ \times R^n \times R^n$ and $g_{k+1} \in \partial_{\varepsilon_{2,k}} f(x_{k+1})$ satisfying the following inequality:

$$f(x_{k+1}) \leq f(x_k) - t_k \langle g_{k+1}, g_{k+1} \rangle + \varepsilon_{1,k}. \tag{1.5}$$

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 assumptions 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

The following lemma, given in [17], is used in the global convergence analysis of Algorithm 1.

Lemma 2.1 (Lemma 3.1 of [17]). *If* $\tau_k > 0$ $(k = 0, 1, 2, \ldots)$ *and* $\sum_{k=0}^{\infty} \tau_k = +\infty$, *then* $\sum_{k=0}^{\infty} (\tau_k / S_k) = +\infty$, *where* $S_k = \sum_{i=0}^{k} \tau_i$.

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

(i) *Either* $\liminf_{k\to\infty} f(x_k) = -\infty$ *or* $\{f(x_k)\}$ *is a convergent sequence.*

(ii) *If* $\{f(x_k)\}$ *is convergent, then* $\liminf_{k\to\infty} \|g_k\| = 0$. *In particular, if* $\{x_k\}$ *is bounded, then any of its accumulation points,* $x^* \in R^n$, *is an* ε^**-minimum point of* f, $f_{\varepsilon^*}^* = f(x^*)$, *and* $\lim_{k\to\infty} f(x_k) = f_{\varepsilon^*}^*$, *where*

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

(iii) *If* $\{f(x_k)\}$ *is convergent and* $\inf\{t_k\} > 0$, *then* $\lim_{k\to\infty} \|g_k\| = 0$. *In this case, every accumulation point of* $\{x_k\}$ (*if one exists*) *is an* $\bar{\varepsilon}^*$*-minimum point of* f, *where*

$$\bar{\varepsilon}^* = \sup \left\{ \limsup_{k \in K_2, k\to\infty} \varepsilon_{2,k-1} \colon\ \textit{whenever} \ \{x_k \colon k \in K_2\} \ \textit{is a convergent subsequence} \right\}.$$

(iv) *If* $\varepsilon_{2,k} \to 0$, $\sum_{k=1}^{\infty} t_k \varepsilon_{2,k} < +\infty$, *and there exists a positive number* $m > 0$ *such that, for all large* k,

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

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. (i) Suppose that a third case happens. By (1.5), $\{f(x_k)\}$ is bounded from above. Then the only possibility is that $\{f(x_k)\}$ has a cluster point $\bar{f}_*$ and $\liminf_{k\to\infty} f(x_k) \leq \bar{\bar{f}}_* < \bar{f}_*$. This implies that there exist two positive integer number $k_2 > k_1$ such that

$$f(x_{k_2}) > \bar{f}_* - \tfrac{1}{3}(\bar{f}_* - \bar{\bar{f}}_*),$$

$$f(x_{k_1}) < \bar{\bar{f}}_* + \tfrac{1}{3}(\bar{f}_* - \bar{\bar{f}}_*),$$

and

$$\sum_{k=k_1}^{\infty} \varepsilon_{1,k} < \tfrac{1}{3}(\bar{f}_* - \bar{\bar{f}}_*).$$

So, we have

$$\begin{aligned} \bar{f}_* - \tfrac{1}{3}(\bar{f}_* - \bar{\bar{f}}_*) &< f(x_{k_2}) \\ &< f(x_{k_1}) + \sum_{k=k_1}^{k_2} \varepsilon_{1,k} \\ &< \bar{\bar{f}}_* + \tfrac{1}{3}(\bar{f}_* - \bar{\bar{f}}_*) + \tfrac{1}{3}(\bar{f}_* - \bar{\bar{f}}_*). \end{aligned}$$

Hence,

$$\tfrac{2}{3}\bar{f}_* + \tfrac{1}{3}\bar{\bar{f}}_* < \tfrac{2}{3}\bar{f}_* + \tfrac{1}{3}\bar{\bar{f}}_*,$$

which is impossible. This excludes the third case. Hence, either $\liminf_{k\to\infty} f(x_k) = -\infty$ or $\{f(x_k)\}$ is a convergent sequence.

(ii) In this case we have

$$\sum_{k=1}^{\infty}[f(x_{k+1}) - f(x_k)] > -\infty, \tag{2.4}$$

which combined with (1.5) implies that

$$\liminf_{k\to\infty}\|g_k\| = 0. \tag{2.5}$$

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 $\{x_k : k \in K_1\}$ 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,

$$f(x) \geq f(x_k) + \langle g_k, x - x_k\rangle - \varepsilon_{2,k-1}, \tag{2.6}$$

and using (2.5), we have that, for all $x \in R^n$,

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

This implies that $f(x^*) = f^*_{\varepsilon^*}$. Now, let x^{**} be an arbitrary accumulation point of $\{x_k\}$. Since $\{f(x_k\}$ is convergent, we have $f(x^*) = \lim_{k\to\infty} f(x_k) = f(x^{**})$. Thus, the above conclusion is also true if we replace x^* by x^{**}. This proves (ii).

(iii) In this case, combining (2.4) with (1.5), we have

$$\lim_{k\to\infty}\|g_k\| = 0. \tag{2.7}$$

Let $\{x_k : k \in K\}$ be any convergent subsequence of $\{x_k\}$, i.e., $\lim_{k\in K,k\to\infty}\|x_k - x^*\| = 0$. By (2.6) and (2.7), x^* is an $\bar{\varepsilon}^*$-minimum point of f.

(iv) From (ii) and (iii), it suffices to consider the case where $\{f(x_k)\}$ 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 $k \geq k_0$,

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

Inequality (1.5) implies that

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq -t_k\|g_{k+1}\|^2 + \varepsilon_{1,k} \\ &\leq t_k\|g_{k+1}\|\frac{\langle g_{k+1}, \bar{x} - x_{k+1}\rangle}{\|\bar{x} - x_{k+1}\|} + \varepsilon_{1,k}. \end{aligned}$$

Therefore, we have that, for all k,

$$f(x_{k+1}) - f(x_k) \leq -\tau t_k\frac{\|g_{k+1}\|}{\|\bar{x} - x_{k+1}\|} + \varepsilon_{1,k}. \tag{2.8}$$

By (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

$$\sum_{k=1}^{\infty} t_k\|g_{k+1}\| = +\infty \tag{2.9}$$

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 $k > k_1$,

$$\|x_{k+1} - \bar{x}\| \leq \|x_1 - \bar{x}\| + \sum_{i=1}^{k} \|x_{i+1} - x_i\| \leq 2m^{-1} \sum_{i=1}^{k} t_i \|g_{i+1}\|. \tag{2.10}$$

By (2.8) and (2.10), we have

$$f(x_{k+1}) - f(x_k) \leq -\frac{m\tau}{2} \frac{t_k \|g_{k+1}\|}{\sum_{i=1}^{k} t_i \|g_{i+1}\|} + \varepsilon_{1,k}. \tag{2.11}$$

This inequality, (2.9), and Lemma 2.1 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 $x \in R^n$,

$$\limsup_{k \to \infty} \langle g_k, x - x_k \rangle \geq 0. \tag{2.12}$$

The ε-subgradient inequality,

$$f(x) - f(x_k) \geq \langle g_k, x - x_k \rangle - \varepsilon_{2,k-1}$$

and (2.12) yield that, for all $x \in R^n$,

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

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

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

$$\|x_k - x\|^2 = \|x_{k+1} - x\|^2 + \|x_k - x_{k+1}\|^2 + 2\langle x_k - x_{k+1}, x_{k+1} - x \rangle. \tag{2.13}$$

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

$$\langle x_k - x_{k+1}, x_{k+1} - x \rangle \geq t_k [f(x_{k+1}) - f(x) - \varepsilon_{2,k}].$$

Combining it with (2.13), we have

$$\|x_k - x\|^2 \geq \|x_{k+1} - x\|^2 + t_k^2 \|g_{k+1}\|^2 + 2t_k [f(x_{k+1}) - f(x) - \varepsilon_{2,k}]. \tag{2.14}$$

Using any given minimum point x^* in (2.14), we have

$$\|x_k - x^*\|^2 \geq \|x_{k+1} - x^*\|^2 - 2t_k \varepsilon_{2,k}.$$

This implies that

$$\|x_{k+1} - x^*\|^2 \leq \|x_1 - x^*\|^2 + \sum_{i=1}^{k} 2t_i \varepsilon_{2,i} < +\infty.$$

Thus, $\{\|x_k - x^*\|\}$ is bounded. This further implies that $\{\|x_k\|\}$ is also bounded. With an argument similar to the proof of (i) of this theorem, we can show that

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

Suppose there are two accumulation points x_1^* and x_2^* of $\{x_k\}$. From (ii), x_1^* and x_2^* are two $\bar{\varepsilon}^*$-minimum points of f. Since $\varepsilon_{2,k} \to 0$, by the definition of $\bar{\varepsilon}^*$, $\bar{\varepsilon}^* = 0$. Hence, both x_1^* and x_2^* are minimum points of f. By the above arguments, we have

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

Since x_1^* and x_2^* are accumulation points of $\{x_k\}$, $l_i = 0$ for $i = 1, 2$. Hence, $x_1^* = x_2^*$ and $\{x_k\}$ converges to it. $\square$

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

$$a_k = \frac{\varepsilon_{1,k}}{\|g_{k+1}\|^2}, \qquad b_k = \frac{\varepsilon_{2,k}}{\|g_{k+1}\|},$$

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* x^* *minimizes* f *and* $\lim_{k\to\infty} x_k = x^*$. *If, for all* k, $t_k > a_k$, *then*

$$\frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} \le \frac{\sqrt{1+4c_k} - 1}{2c_k}. \tag{3.1}$$

Consequently,

(i) *if, for all* k, $c_k \ge c^* \in (0, +\infty)$, *then*

$$\overline{\lim_{k\to\infty}} \frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} \le \frac{\sqrt{1+4c^*} - 1}{2c^*} < 1;$$

(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 $g_{k+1} \in \partial_{\varepsilon_{2,k}} f(x_{k+1})$, we have

$$\begin{aligned} f(x^*) &\geq f(x_{k+1}) + g_{k+1}^T(x^* - x_{k+1}) - \varepsilon_{2,k} \\ &\geq f(x_{k+1}) - \|g_{k+1}\|\|x^* - x_{k+1}\| - b_k\|g_{k+1}\|. \end{aligned}$$

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

$$\begin{aligned} f(x_k) &\geq f(x_{k+1}) + t_k\|g_{k+1}\|^2 - \varepsilon_{1,k} \\ &= f(x_{k+1}) + t_k\|g_{k+1}\|^2 - a_k\|g_{k+1}\|^2 \\ &\geq f(x_{k+1}) + \frac{t_k - a_k}{(b_k + \|x_{k+1} - x^*\|)^2}[f(x_{k+1}) - f(x^*)]^2. \end{aligned}$$

Hence

$$f(x_k) - f(x^*) \geq [f(x_{k+1}) - f(x^*)]\left[1 + \frac{(t_k - a_k)(f(x_{k+1}) - f(x^*))}{(b_k + \|x_{k+1} - x^*\|)^2}\right].$$

Therefore

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

Thus, (3.1) follows.

Since $\sqrt{1+4t}/2t$ $(t > 0)$ 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* $\varepsilon_{2,k} \to 0$ *and* $\sum_{k=1}^{\infty}\sqrt{\varepsilon_{1,k}} < +\infty$. *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,* $\{\|x_k\|^2/k\}$ *is bounded*.

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

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

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

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

Hence

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

This inequality and our assumptions on t_k and $\varepsilon_{1,k}$ 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 a geometric rate.

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

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

which implies

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

On the other hand,

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

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

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

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

- *A family of variable metric proximal methods proposed in* [4].
- *The convex minimization methods given in* [6].
- *The 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

$$\varphi_B(z) := f(z) + \tfrac{1}{2}\langle B(z-x), z-x\rangle, \tag{4.1}$$

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

$$\delta_k := f(x_k) - f(x_k^p) - \tfrac{1}{2}\langle g_k^p, W_k g_k^p\rangle, \tag{4.3}$$

where $W_k = B_k^{-1}$ and g_k^p is a subgradient in $\partial f(x_k^p)$, satisfying

$$x_k^p = x_k - W_k g_k^p. \tag{4.4}$$

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

$$f(x_{k+1}) \le f(x_k) - m_0\delta_k. \tag{4.5}$$

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* 4.1. *Let*

$$\varepsilon_{2,k} = f(x_{k+1}) - f(x_k^p) - \langle g_k^p, x_{k+1} - x_k^p\rangle, \tag{4.6}$$

then $\varepsilon_{2,k} \ge 0$,

$$g_k^p \in \partial_{\varepsilon_{2,k}} f(x_{k+1}) \tag{4.7}$$

and

$$f(x_{k+1}) \le f(x_k) - \frac{m_0}{2}\langle g_k^p, W_k g_k^p\rangle. \tag{4.8}$$

Thus, (1.5) *holds with* $t_k = (m_0/2)\lambda_{\min}(W_k)$, *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 $x \in R^N$,

$$\begin{aligned} f(x) &\geq f(x_k^p) + \langle g_k^p, x - x_k^p \rangle \\ &= f(x_{k+1}) + \langle g_k^p, x - x_{k+1} \rangle - [f(x_{k+1}) - f(x_k^p) - \langle g_k^p, x_{k+1} - x_k^p \rangle]. \end{aligned} \tag{4.9}$$

By the convexity of f, we see $\varepsilon_{2,k} \geq 0$ and that (4.7) follows (4.9).

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

$$f(x_k^p) + \tfrac{1}{2}\langle B_k(x_k^p - x_k), x_k^p - x_k \rangle \leq f(x) + \tfrac{1}{2}\langle B_k(x - x_k), x - x_k \rangle.$$

Setting $x = x_k$, we have

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

by (4.4). Relations (4.10), (4.3), and (4.5) imply (4.8). □

Conclusion (a) in the Theorem 4.1 is the global convergence result of [4]. We give a simple proof here using our general results. Conclusion (b) is new for this algorithm.

Theorem 4.1.

(a) (*Theorem* 2.3 *of* [4].) *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 (4.8) and (4.4). (In fact, this conclusion follows due to Theorem 2.3 of [4].) Using (4.5), we have

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

By (4.8), we obtain

$$\langle g_k^p, W_k g_k^p \rangle \to 0. \tag{4.14}$$

Hence,

$$f(x_k) - f(x_k^p) - \langle g_k^p, W_k g_k^p \rangle = \delta_k - \tfrac{1}{2}\langle g_k^p, W_k g_k^p \rangle \to 0. \tag{4.15}$$

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

By (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 (4.8) and (4.11), we have $\liminf_{k\to\infty}\|g_k^p\| = 0$.

(b) From (4.12) and (4.4), we have, for all large k, that

$$\frac{1}{\bar{t}_k\|W_k\|}\|x_{k+1} - x_k\| \le \|g_k^p\|.$$

Using (4.4), (4.15), and the fact that $f(x_{k+1}) < f(x_k)$, we have, for all k,

$$\varepsilon_{2,k} < f(x_k) - f(x_k^p) - \langle g_k^p, W_k g_k^p\rangle < \delta_k.$$

On the other hand, from (4.5), we have

$$\sum_{k=1}^{\infty}\delta_k \le m_0^{-1}(f(x_1) - f_0^*).$$

Therefore

$$\sum_{k=1}^{\infty}\varepsilon_{2,k} \le m_0^{-1}(f(x_1) - f_0^*) < +\infty. \tag{4.16}$$

Let $m = 1/(\bar{t}\sup\{\|W_k\|\})$ and let $t_k = 1$ for all large k, then (4.16), the assumptions in (b), and the result (iv) in Theorem 2.1 yield that $f(x_k) \to f_0^*$. □

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* $\{\|x_k\|^2/k\}$ *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 include (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 $\{x_k\}$ is constructed with the following formulas:

$$x_{k+1} = x_k - \tau_k\gamma_k, \tag{4.17}$$

$$\gamma_k \in \partial_{\varepsilon_{3,k}} f(x_k), \tag{4.18}$$

$$f(x_{k+1}) \leq f(x_k) - m_1 \tau_k \|\gamma_k\|^2, \tag{4.19}$$

where $\varepsilon_{3,k}$ is nonnegative, $\tau_k > 0$ is the stepsize, and m_1 is a positive constant.

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

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

$$t_k = m_1 \tau_k, \tag{4.20}$$

$$\varepsilon_{2,k} = \max\{0, \varepsilon_{3,k} + (1 - m_1)\tau_k \|\gamma_k\|^2\}. \tag{4.21}$$

Then

$$\gamma_k \in \partial_{\varepsilon_{2,k}} f(x_{k+1}) \tag{4.22}$$

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

Proof. It suffices to prove that (4.22) holds. By (4.18), we have, for all $x \in R^n$,

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

This inequality, (4.17), and (4.19) imply that

$$f(x) \geq f(x_{k+1}) + \langle \gamma_k, x - x_{k+1} \rangle - [\varepsilon_{3,k} + (1 - m_1)\tau_k \|\gamma_k\|^2].$$

So (4.22) follows. □

The following is a main result of [6].

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

(i) *Assume that*

$$\sum_{k=1}^{\infty} \tau_k = +\infty \tag{4.23}$$

and

$$\varepsilon_{3,k} \to 0. \tag{4.24}$$

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

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

$$\sum_{k=1}^{\infty} \varepsilon_{3,k} < +\infty, \tag{4.25}$$

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 (4.19), we have

$$\sum_{k=1}^{\infty} \tau_k \|\gamma_k\|^2 < +\infty. \tag{4.26}$$

This and (4.21) imply that $\{\varepsilon_{3,k}\}$ tends to zero if and only if $\{\varepsilon_{2,k}\}$ tends to zero. So the conclusion follows by Lemma 4.2 and (iv) of Theorem 2.1.

(ii) Suppose f has a minimum point. Then $\{f(x_k)\}$ is bounded from below. Thus, (4.26) holds. Inequalities (4.26) and (4.25) imply that

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

by the boundedness of $\{\tau_k\}$ and (4.20). The results of (iv) in Theorem 2.1 imply that $\{x_k\}$ converges to a minimum point of f. □

Note that from (4.17) and Theorem 3.2, we obtain the following new convergence rate for Algorithm 4.2.

Theorem 4.4. *Let* $\{(\varepsilon_{3,k}, \tau_k, x_{k+1}, \gamma_k)\}$ *be generated by Algorithm* 4.1 *with* (4.23) *and* (4.24). *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* $\{\|x_k\|^2/k\}$ *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

$$\operatorname{dist}(0, S_k(x_{k+1})) \le \frac{\sigma_k}{\lambda_k}\|x_{k+1} - x_k\|, \tag{4.27}$$

where

$$\sum_{k=1}^{\infty} \sigma_k < \infty, \tag{4.28}$$

and

$$S_k(x) = \partial f(x) + \frac{1}{\lambda_k}(x - x_k). \tag{4.29}$$

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

$$\sigma_k \in [0, \tfrac{1}{2}]. \tag{4.30}$$

Lemma 4.3. *Suppose that* $\{(\sigma_k, \lambda_k, x_{k+1}, g_{k+1})\} \in R_0^+ \times R^+ \times R^n \times R^n$ *is generated by Algorithm* 4.3. *Let*

$$t_k = \frac{1-\sigma_k}{(1+\sigma_k)^2}\lambda_k, \tag{4.31}$$

then (1.5) *holds for* $\varepsilon_{1,k} = \varepsilon_{2,k} = 0$. *Furthermore,*

$$\frac{(1-\sigma_k)^2}{(1+\sigma_k)^2}\|x_{k+1} - x_k\| \le t_k\|g_{k+1}\|. \tag{4.32}$$

Proof. By (4.27), (4.30), and (4.29), we have

$$\left\| g_{k+1} + \frac{1}{\lambda_k}(x_{k+1} - x_k) \right\| \le \frac{\sigma_k}{\lambda_k}\|x_{k+1} - x_k\|. \tag{4.33}$$

The inequality,

$$\left\langle g_{k+1} + \frac{1}{\lambda_k}(x_{k+1} - x_k), x_{k+1} - x_k \right\rangle \le \left\| g_{k+1} + \frac{1}{\lambda_k}(x_{k+1} - x_k) \right\| \|x_{k+1} - x_k\|,$$

and (4.33) imply that

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

Therefore,

$$\langle g_{k+1}, x_k - x_{k+1} \rangle \ge \frac{1-\sigma_k}{\lambda_k}\|x_{k+1} - x_k\|^2. \tag{4.34}$$

On the other hand, by (4.33), we have

$$\|g_{k+1}\| \le \frac{1+\sigma_k}{\lambda_k}\|x_{k+1} - x_k\|. \tag{4.35}$$

Inequalities (4.34) and (4.35) yield

$$\langle g_{k+1}, x_k - x_{k+1} \rangle \ge t_k \langle g_{k+1}, g_{k+1} \rangle. \tag{4.36}$$

Applying the subgradient inequality for convex functions, we have

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

Hence, (1.5) follows.

By (4.33), we have

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

which implies that (4.32) 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 a point exists*.

For Algorithm 4.3, we obtain the following two basic convergence rate results from Theorems 3.1 and 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* $\|x - x^*\| \le r$,

$$f(x) - f(x^*) \ge M\|x - x^*\|^2. \tag{4.37}$$

(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* $f_0^* > -\infty$. *The rate of* $\sqrt{f(x_k) - f_0^*}$ *converging to zero is then less than geometric and* $\{\|x_k\|^2/k\}$ *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 (4.30), that

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

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

(b) From the definition of m in (2.2), for Algorithm 4.3, we can choose $m = \frac{1}{9}$. In this case,

$$\frac{(1-\sigma_k)^2}{(1+\sigma_k)^2} \geq m$$

by (4.30). This implies that the results of (b) hold by using Theorem 3.2. □

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 line 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.

Acknowledgments

The authors are thankful to Peter Kall and two referees for their helpful comments. Indeed, Theorem 2.1, Lemma 4.1, and their proofs were dramatically improved by one referee's suggestions.

References

1. Auslender A (1993) Convergence of stationary sequences for variational inequalities with maximal monotone operators, Applied Mathematics and Optimization 28: 161–172
2. Auslender A, Cominetti R, Crouzeix JP (1993) Convex functions with unbounded level sets and applications to duality theory, SIAM Journal on Optimization 3: 669–687
3. Auslender A, Crouzeix JP (1989) Well behaved asymptotical convex functions, in: Analyse Non-linéare (Perpignan, 1987), Université de Montréal, Montréal, pp 101–122
4. Bonnans JF, Gilbert JC, Lemaréchal C, Sagastizábal CA (1995), A family of variable metric proximal methods, Mathematical Programming 68: 15–47
5. Chou CC, Ng KF, Pang JS (1995) Minimizing and stationary sequences for optimization problems, Manuscript, Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD
6. Correa R, Lemaréchal C (1993) Convergence of some algorithms for convex minimization, Mathematical Programming 62: 261–275
7. Fukushima F, Pang JS (1995) Minimizing and stationary sequences of merit functions for complementarity problems and variational inequalities, Manuscript, Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD
8. Güler O (1991) On the convergence of the proximal point algorithm for convex minimization, SIAM Journal on Control and Optimization 29: 403–419

9. Kiwiel KC (1983) An aggregate subgradient method for nonsmooth convex minimization, Mathematical Programming 27: 320–341
10. Lemaire B (1992) About the convergence of the proximal method, in: D. Pallaschke, ed, Advances in Optimization. Lecture Notes in Economics and Mathematics Systems, Springer-Verlag, Berlin, pp 39–51
11. Poljak BT (1967) A general methods for solving extremum problems, Soviet Mathematics Doklady 8: 593–597
12. Rockafellar RT (1970) Convex Analysis, Princeton University Press, Princeton, NJ
13. Rockafellar RT (1976) Monotone operators and proximal point algorithm, SIAM Journal on Control and Optimization 14: 877–898
14. Todd MJ (1989) On convergence properties of algorithms for unconstrained minimization, IMA Journal of Numerical Analysis 9: 435–441
15. Wei Z, Qi L (1996) Convergence analysis of a proximal Newton method, Numerical Functional Analysis and Optimization 17: 463–472
16. Wei Z, Qi L, Jiang J (to appear) Some convergence properties of descent methods, Journal of Optimization: Theory and Applications
17. Wu SQ (1992) Convergence properties of descent methods for unconstrained minimization, Optimization 26: 229–237

Accepted 15 *October* 1996