Yokohama Publishers
ISSN 1338-9151 ONLINE JOURNAL
Pacific Journal of Optimization
Vol. 1, No. 1, pp. 31–37, January 2005

SOME REMARKS ON FINITE TERMINATION OF DESCENT METHODS

Sien Deng

This paper is dedicated to Professor R.T. Rockafellar on the occasion of his 70th birthday.

Abstract: This paper establishes some sufficient conditions for finite termination of the proximal point method and the subgradient method. It is shown that the notion of boundedly weak sharp minima is sufficient for finite termination of these descent methods under mild conditions.

Key words: *boundedly weak sharp minima, descent methods, finite convergence*

Mathematics Subject Classification: *90C31, 90C25*

1 Introduction

Consider the convex minimization problem.

$$(\mathcal{P}) \quad \text{minimize } f(x) \text{ s.t. } x \in I\!R^n,$$

where f is a proper closed convex function on $I\!R^n$. We assume throughout that the optimal solution set S of $(\mathcal{P})$ is non-empty, and denote the optimal value of $(\mathcal{P})$ by f_{min}.

Among many descent methods to solve $(\mathcal{P})$, the proximal point (resolvent) method generates, with a given x_1, an iterative sequence $\{x_i\}$ as follows:

$$x_{i+1} = x_i - \lambda_{i+1} x^*_{i+1}, \tag{1}$$

where $x^*_{i+1} \in \partial f(x_{i+1})$, the step length $\lambda_i > 0$, and

$$x_{i+1} = \text{argmin}_{x \in I\!R^n} \left\{ f(x) + \frac{1}{2\lambda_{i+1}} \|x - x_i\|^2 \right\}.$$

Since the seminal work of [5], the proximal point method has been the subject of much attention. In [5], Rockafellar has shown the proximal point method exhibits the finite termination property when either $0 \in int\,(\partial f(\bar{x}))$ [5, Theorem 3] for some $\bar{x} \in S$ or f is a polyhedral convex function [5, Proposition 8], which includes linear programming as a special case. In [3], Ferris generalized Rockafellar's result on the finite termination property under the assumption that S is a set of weak sharp minima for f (see the formal definition in the next section).

In this paper, we study conditions under which the finite termination property of descent methods holds. Our approach is quite different from that used in [3]. Specifically, by establishing a bound on the sum of step lengths first, we show that boundedly weak sharp minimality is sufficient for finite termination of the proximal point method; we also show that the proximal point method will find an ϵ-optimal solution in finitely many iterations under only the assumption that S is non-empty. Furthermore, we extend our analysis on finite termination to subgradient methods.

Throughout this paper, the notation we will use is the same as that in Rockafellar [4].

2 Proximal Point Method

We begin with a lemma, which states that the magnitude of any subgradient from the set $\partial f(y)$ with $y \notin S$, is at least as large as the rise "$f(y) - f_{min}$" divided by the run "$dist(y, S)$."

Lemma 2.1 *Suppose that $y \in dom\ f \backslash S$. Then*

$$\|y^*\| \geq \frac{f(y) - f_{min}}{dist\ (y, S)}, \qquad \forall y^* \in \partial f(y), \tag{2}$$

where $||y^*||$ is the Euclidean norm of y^*, and $dist(y, S)$ denotes the Euclidean distance between y and S.

Proof. Let $\Pi_S(y)$ be the projection of y onto S. Then for any $y^* \in \partial f(y)$, we have, by the convexity of f,

$$f_{min} - f(y) = f(\Pi_S(y)) - f(y) \geq \langle y^*, \Pi_S(y) - y \rangle.$$

It follows that

$$\|y^*\| \text{dist } (y, S) \geq \langle y^*, y - \Pi_S(y) \rangle \geq f(y) - f_{min}.$$

Thus (2) holds. □

The first result of this note follows.

Theorem 2.1 *Consider $(\mathcal{P})$. Let $\{x_i\}$ be a sequence generated by the algorithm (1). Suppose that $x_i \notin S$ for $i = 1, ..., N$. Then*

$$\sum_{i=2}^{N} \lambda_i \leq \frac{f(x_1) - f_{min}}{{\tau_N}^2}, \tag{3}$$

where $\tau_N = \min_{2 \leq i \leq N} \left(\frac{f(x_i) - f_{min}}{dist(x_i, S)} \right)$.

Proof. By (1), we have that

$$f(x_i) - f(x_{i+1}) \geq \lambda_{i+1} \|x^*_{i+1}\|^2 \quad \text{for } i = 1, 2, ..., N-1.$$

By summing this expression over all indices $i \leq N - 1$, we obtain

$$\sum_{i=1}^{N-1} \lambda_{i+1} \|x^*_{i+1}\|^2 \leq f(x_1) - f(x_N) \leq f(x_1) - f_{min}. \tag{4}$$

By Lemma 2.1, $\|x_i^*\| \geq \frac{f(x_i)-f_{min}}{dist\ (x_i,S)}$ for $i = 2, ..., N$. This observation along with (4) yields

$$\tau_N \left(\sum_{i=2}^{N} \lambda_i \right) \leq \min_{2 \leq i \leq N} \left(\frac{f(x_i) - f_{min}}{dist(x_i, S)} \right)^2 \left(\sum_{i=2}^{N} \lambda_i \right) \leq f(x_1) - f_{min}.$$

This establishes (3), and the proof is complete. □

It is well-known that $(I + \lambda_{i+1}\partial f)^{-1}(x) = x$ if and only if $x \in S$, and the operator $(I + \lambda_{i+1}\partial f)^{-1}(\cdot)$ is non-expansive. This implies that $\|x_{i+1} - x\| \leq \|x_i - x\|$ for all $x \in S$. So, for any $x_k \in \text{dom } f$ and $z \in S$,

$$\|x_k - z\| \leq \|x_{k-1} - z\|. \tag{5}$$

By invoking (5) repeatedly for $k = 2, 3, \ldots, i$, we get

$$\|x_i - z\| \leq \|x_1 - z\|, \quad \forall z \in S. \tag{6}$$

As a consequence of (6),

$$dist(x_i, S) \leq \|x_i - \Pi_S(x_1)\| \leq \|x_1 - \Pi_S(x_1)\| = dist(x_1, S).$$

Therefore,

$$\tau_N = \min_{2 \leq i \leq N} \left(\frac{f(x_i) - f_{min}}{dist(x_i, S)} \right) \geq \min_{2 \leq i \leq N} \left(\frac{f(x_i) - f_{min}}{dist(x_1, S)} \right). \tag{7}$$

Theorem 2.1 has some interesting implications. Suppose that $\{x_i\}$ is an infinite sequence generated by the algorithm (1) with $x_i \notin S$ for all i. First, we observe that $\{\tau_N\}$ is a positive decreasing sequence of N, and the right side of (3) is bounded above for all N if and only if

$$\lim_{N \to \infty} \tau_N > 0.$$

Also, the inequality (6) implies that $\{x_i\}$ is a bounded sequence. So a sufficient condition for $\lim_{N\to\infty} \tau_N > 0$ is the following notion of boundedly weak sharp minima [1]: $\forall r > 0$ there is some $\alpha_r > 0$ such that

$$f(x) - f_{min} \geq \alpha_r dist(x, S), \qquad \forall x \in r\mathbb{B} \cap (dom\ f). \tag{8}$$

If (8) holds for $r = \infty$, we say that S is a set of weak sharp minima for f with modulus α_∞. In this case, we simply denote α_∞ by α. [1, Example 6.6] shows that the notion of boundedly weak sharp minima is weaker than that of weak sharp minima. Secondly, if

$$\sum_{i=1}^{N} \lambda_i \to +\infty \quad \text{as } N \to +\infty, \tag{9}$$

then by (3), $\lim_{N\to\infty} \tau_N = 0$; that is, (9) and $\lim_{N\to\infty} \tau_N > 0$ are not compatible. The above observations yield the following corollary. No proof is needed.

Corollary 2.1 *Consider ($\mathcal{P}$). Let x_1 be given. Suppose that the algorithm (1) is implemented such that (9) is satisfied. Suppose that S is a set of boundedly weak sharp minima. Then the algorithm (1) terminates in finitely many iterations.*

A few remarks on Corollary 2.1 and [3, Theorem 6] are in order. Corollary 2.1 improves [3, Theorem 6] in two ways: First, the weak sharp minimality is replaced by the boundedly weak sharp minimality; secondly, the assumption that $\lambda_i \geq \bar{\lambda} > 0$ is replaced by (9).

Recall that, for any $\epsilon > 0$, the set of ϵ-optimal solutions is defined as follows:

$$S_\epsilon = \{x \in I\!R^n |\ f(x) - f_{min} \leq \epsilon\}.$$

Another consequence is the following corollary on ϵ-optimal solutions.

Corollary 2.2 *Consider problem ($\mathcal{P}$). Let x_1 be given. Suppose that the algorithm (1) is implemented such that (9) is satisfied. Let $\epsilon > 0$ be given. Then, there is some N such that $x_N \in S_\epsilon$. If it is further assumed that $\lambda_i \geq \bar{\lambda} > 0$ for all i, then*

$$N \leq 1 + \frac{dist^2(x_1, S)(f(x_1) - f_{min})}{\bar{\lambda}\epsilon^2}. \tag{10}$$

Proof. Suppose that x_i is not an ϵ-optimal solution for $i = 0, 1, ..., N$. We have that $f(x_i) - f_{min} \geq \epsilon$ for all $i = 1, ..., N$. Relation (7) implies that

$$\tau_N = \min_{2 \leq i \leq N} \left(\frac{f(x_i) - f_{min}}{dist(x_i, S)} \right) \geq \frac{\epsilon}{dist(x_1, S)}.$$

Relation (3) along with the above estimate of τ_N yields

$$\sum_{i=2}^{N} \lambda_i \leq \frac{dist^2(x_1, S)(f(x_1) - f_{min})}{\epsilon^2}. \tag{11}$$

Clearly, both (9) and (11) cannot hold at the same time for N sufficiently large. So there is some N such that $x_N \in S_\epsilon$.

If $\lambda_i \geq \bar{\lambda}$ for all i, then, by (11), we have

$$\bar{\lambda}(N - 1) \leq \frac{dist^2(x_1, S)(f(x_1) - f_{min})}{\epsilon^2}.$$

This establishes (10). □

3 Subgradient Method

Another commonly used descent method for solving ($\mathcal{P}$) can be described as follows. Given x_1, let

$$x_{i+1} = x_i - t_i x_i^* \qquad \text{for } i = 1, 2, \ldots \tag{12}$$

where the step length $t_i > 0$ and $x_i^* \in \partial f(x_i)$. If f is differentiable, then $\partial f(x_i) = \{\nabla f(x_i)\}$, and the algorithm (12) is the well-known steepest descent method. It is well-known that $\{x_i\}$ generated by (12) may not converge without a proper line search. We shall impose the following sufficient descent condition at each iteration: there is some $0 < m \leq 1$ such that

$$f(x_{i+1}) \leq f(x_i) - m t_i ||x_i^*||^2. \tag{13}$$

By the non-emptiness assumption of S, and by summing (13) over all indices $i \leq N$ with $x_i \notin S$, we have the following inequality similar to (4):

$$\sum_{i=1}^{N} t_i ||x_i^*||^2 \leq \frac{f(x_1) - f(x_N)}{m} \leq \frac{f(x_1) - f_{min}}{m}. \tag{14}$$

Inspired by the related work on convergence analysis of descent methods for differentiable convex functions [6] and the bundle methods [2], we give below some convergence results on the algorithm (12). To make our presentation self-contained, we include the proofs here.

Proposition 3.1 *Suppose that $\{x_i\}$, generated by the algorithm (12), is an infinite sequence. Assume that $x_i \notin S$ for all i, and $\{x_i\}$ satisfies (13). Then the following is true.*

(a) If

$$\sum_{i=1}^{N} t_i \to \infty \qquad \text{as } N \to \infty, \tag{15}$$

then $\lim_{i\to\infty} f(x_i) = f_{min}$.

(b) If $\lim_{i\to\infty} f(x_i) > f_{min}$, *then* $\{x_i\}$ *is bounded.*

(c) If

$$\sum_{i=1}^{\infty} ||x_{i+1} - x_i||^2 < \infty, \tag{16}$$

then $\{x_i\}$ is bounded. Relation (16) holds whenever there is some $c > 0$ such that $t_i \le c$ for all i.

(d) If $\lim_{i\to\infty} f(x_i) = f_{min}$, *and (16) holds, then $\{x_i\}$ converges to a point of S.*

Proof. [Proof of (a):] We observe that $\{f(x_i)\}$ is decreasing. If the conclusion does not hold, then there is some $\delta > 0$ such that $f(x_i) \ge f_{min} + \delta$ for all i. Let $\bar{x} \in S$. By

$$||x_{i+1} - \bar{x}||^2 = ||x_i - \bar{x}||^2 + ||x_{i+1} - x_i||^2 + 2\langle x_{i+1} - x_i, x_i - \bar{x}\rangle,$$

and the convexity of f,

$$||x_{i+1} - \bar{x}||^2 \le ||x_i - \bar{x}||^2 + ||x_{i+1} - x_i||^2 + 2t_i(f_{min} - f(x_i)). \tag{17}$$

From (14), $\lim_{i\to\infty} t_i ||x_i^*||^2 = 0$. So there is some k such that

$$||x_{i+1} - x_i||^2 + 2t_i(f_{min} - f(x_i)) \le -\delta t_i \qquad \text{for } i \ge k. \tag{18}$$

It follows, by summing (17) over $k \le i \le N$, that

$$||x_{N+1} - \bar{x}||^2 \le ||x_k - \bar{x}||^2 - \delta \sum_{i=k}^{N} t_i.$$

By (15), the above inequality cannot hold for N sufficiently large. The contradiction proves (a).

[Proof of (b):] Since $\{f(x_i)\}$ is monotone decreasing, $\lim_{i\to\infty} f(x_i)$ exists. If $\lim_{i\to\infty} f(x_i) > f_{min}$, then there is some $\delta > 0$ such that $f(x_i) \ge f_{min} + \delta$. By a similar argument as in the proof of (a), we see that (17) and (18) hold, and the boundedness of $\{x_i\}$ follows.

[Proof of (c):] Again from (17), $||x_{i+1} - \bar{x}||^2 \le ||x_i - \bar{x}||^2 + ||x_{i+1} - x_i||^2$. So, for any N,

$$||x_N - \bar{x}||^2 \le ||x_1 - \bar{x}||^2 + \sum_{i=1}^{N} ||x_{i+1} - x_i||^2.$$

This shows that $\{x_i\}$ is bounded by (16). The last part follows from (14), $t_i \le c$, and $||x_{i+1} - x_i||^2 = t_i^2 ||x_i^*||^2$.

[Proof of (d):] By (c), $\{x_i\}$ is a bounded sequence. Let $\{x_{i_k}\}$ be a convergent subsequence with the limit $\bar{x}$. Then $\bar{x} \in S$ since $f(\bar{x}) = f_{min}$. For any $\epsilon > 0$, by $\lim_{k\to\infty} x_{i_k} = \bar{x}$, and by (16), there is some positive integer K such that $||x_{i_k} - \bar{x}||^2 \leq \frac{\epsilon^2}{2}$ for all $k \geq K$, and $\sum_{i=k}^{\infty} ||x_{i+1} - x_i||^2 \leq \frac{\epsilon^2}{2}$. Then for any $j > i_K$, by (17),

$$||x_j - \bar{x}||^2 \leq ||x_K - \bar{x}||^2 + \sum_{i=K}^{j} ||x_{i+1} - x_i||^2 \leq \epsilon^2.$$

This shows that $\{x_i\}$ converges to $\bar{x}$. □

When S is a set of weak sharp minima or boundedly weak sharp minima, we have the following sharp results.

Theorem 3.1 *Consider $(\mathcal{P})$. Let $x_1 \in dom\ (f)$ be given. Then the following holds.*

(a) Suppose that S is a set of weak sharp minima for f with modulus α. Suppose that the algorithm (12) is implemented such that (13) is satisfied. Then $\{x_i\}$ is a convergent sequence. If it is further assumed that (15) is satisfied, then the algorithm (12) finds a minimizer of $(\mathcal{P})$ in finitely many iterations.

(b) Suppose that S is a set of boundedly weak sharp minima. Suppose that the algorithm (12) is implemented such that (13) is satisfied. Then either $\lim_{i\to\infty} f(x_i) = f_{min}$ or $\{x_i\}$ is a convergent sequence. If it is further assumed that (15) is satisfied and $\{x_i\}$ is bounded, then $\{x_i\}$ is a finite sequence; that is, the algorithm (12) finds a minimizer of $(\mathcal{P})$ in finitely many iterations.

Proof. [Proof of (a):] By (12), $||x_{i+1} - x_i|| = t_i||x_i^*||$. Since S is a set of weak sharp minima for f with modulus α, and $x_i \notin S$, by Lemma 2.1 and (8) $||x_i^*|| \geq \alpha$ for all i. These facts along with (14) show that, for any N,

$$\sum_{i=1}^{N} ||x_{i+1} - x_i|| \leq \frac{f(x_1) - f_{min}}{\alpha m}. \tag{19}$$

So $\{x_i\}$ is a convergent sequence.

Relation (14) yields

$$\sum_{i=1}^{N} t_i \leq \frac{f(x_1) - f_{min}}{\alpha^2 m}. \tag{20}$$

When (15) holds, (20) cannot hold for N sufficiently large. So the algorithm (12) terminates in finitely many iterations.

[Proof of (b):] If $\lim_{i\to\infty} f(x_i) = f_{min}$, then we are done. Otherwise, $\{x_i\}$ is bounded by Part (b) of Proposition 3.1, Since S is a set of boundedly weak sharp minima, there is some $\alpha_r > 0$ such that (19) holds with α replaced by α_r. This shows that $\{x_i\}$ is convergent.

If (15) holds and $\{x_i\}$ is bounded, then there is some $\alpha_r > 0$ such that (20) holds with α replaced by α_r. The assertion on finite termination follows again from (15). □

For the algorithm (12), we can also obtain results similar to Corollary 2.2 on ϵ-optimal solutions. Since the analysis is quite similar, we omit these results here.

References

[1] J.V. Burke and S. Deng, Weak sharp minima revisited, Part I: basic theory, *Control and Cybernetics* 31 (2002) 439–469.

[2] R. Correa and C. Lemaréchal, Convergence of some algorithms for convex minimization, *Math. Program.* 62 (1993) 261–275.

[3] M.C. Ferris, Finite termination of the proximal point algorithm, *Math. Program.* 50 (1991) 359–366.

[4] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, N. J., 1970.

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

[6] Z. Wei, L. Qi, and H. Jiang, Some convergence properties of descent methods, *Journal of Optimization Theory and Applications* 95 (1997) 177–188.

Manuscript received 11 June 2004
revised 14 September 2004
accepted for publication 16 September 2004

Sien Deng
Department of Mathematical Sciences, Northern Illinois University, DeKalb, Illinois 60115, USA
E-mail address: deng@math.niu.edu