Pertanika J. Sci. & Technol. **10(2)**: 145 - 152 (2002)

ISSN: 0128-7680
© Universiti Putra Malaysia Press

Convergence of the Steepest Descent Method for Minimizing Convex Functions

Malik Hj. Abu Hassan, Mansor b. Monsi & Leong Wah June
Department of Mathematics
Universiti Putra Malaysia
43400, UPM Serdang,
Selangor

Received: 28 June 1999

ABSTRAK

Kiwiel dan Murty (1996) membincangkan sifat penumpuan bagi suatu kelas algoritma penurunan tercuram untuk meminimumkan fungsi kuasi cembung yang selanjar dan boleh beza f atas $\Re^n$. Di bawah syarat sederhana kita buktikan bahawa had infimum bagi $\|\nabla f(x_k)\|$ adalah sifar dan penumpuan palsu tidak berlaku walaupun bila cembung.

ABSTRACT

Kiwiel and Murty (1996) discuss the convergence properties of a class of steepest descent algorithm for minimizing a continuously differentiable quasiconvex function f on $\Re^n$. Under mild conditions, we prove that the limit infimum of $\|\nabla f(x_k)\|$ is zero and that false convergence does not occur even when f is convex.

Keywords: Convergence, steepest descent method, convex functions, minimization

INTRODUCTION

Consider the following unconstrained minimization problem:

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

when f is assumed continuously differentiable on $\Re^n$.

Descent algorithms for solving (1) usually generate a sequence $\{x_k\}$ such that $f(x_{k+1}) < f(x_k)$ for all k. However, such a procedure does not always guarantee that $f(x_k)$ converges to the infimum of f on $\Re^n$, even if f is a convex function and $\lim_{k\to\infty} \nabla f(x_k) = 0$. In fact, Rockafellar (1970), Todd (1989), and Auslender and Crouzeix (1989) have given examples to confirm the above phenomenon, which has been called false convergence.

The Todd example has the following properties:

(i) f is convex and continuously differentiable;
(ii) the sequence $\{f(x_k)\}$ is monotonically decreasing and $\lim_{k\to\infty} \nabla f(x_k) = 0$;
(iii) $\lim_{k\to\infty} \nabla f(x_k) > \inf_{x\in\Re^n} f(x)$.

Steepest descent method with Armijo's stepsizes (1966) generates a sequence $\{x_k\}$ via

$$x_{k+1} = x_k + t_k d_k, \quad k = 0, 1, \ldots \tag{2}$$

where

$$d_k = -\nabla f(x_k) \tag{3}$$

and

$$t_k = \arg\max \{t: f(x^k + t_k d_k) \leq f(x^k) + \alpha t \nabla f(x_k)^T d_k,\ t = 2^{-i},\ i = 0, 1, \ldots\}, \tag{4}$$

with $\alpha \in (0,1)$.

Under the following standing assumption that generalizes Armijo's condition (4),

Assumption 1.1. Let $\phi : \Re_+ \rightarrow \Re_+$ be a function such that:

(A1) $\exists \alpha \in (0,1),\ \tau_\alpha > 0,\ \forall t \in (0, \tau_\alpha] : \phi(t) \leq \alpha t$,
(A2) $\exists \beta > 0,\ \tau_\beta \in (0, \infty),\ \forall t \in (0, \tau_\beta] \cap \Re : \phi(t) \geq \beta t^2$,
(A3) $\forall k,\ f(x_k + t_k d_k) \leq f(x_k) + \phi(t) \nabla f(x_k)^T d_k$ and $0 < t_k < \tau_\beta$ in (1),
(A4) $\exists \gamma > 1,\ \tau_\gamma > 0,\ \forall k$: $t_k \geq \tau_\gamma$ or $[\exists \bar{t}_k \in [t_k, \gamma t_k]: f(x_k + \bar{t}_k d_k) \geq f(x_k) + \phi(\bar{t}_k) \nabla f(x_k)^T d_k]$.

Kiwiel and Murty have proven that for the steepest descent method, the false convergence does not happen if f is quasiconvex. We present our global convergence results in the next section without the quasiconvexity restriction.

GLOBAL CONVERGENCE PROPERTIES

Theorem 2.1. Suppose that Assumption 1.1 holds, then:

(i) either $f(x_k) \rightarrow -\infty$ or $\lim\inf_{k\rightarrow\infty} \|\nabla f(x_k)\| = 0$;
(ii) either $f(x_k) \rightarrow -\infty$ or $\lim_{k\rightarrow\infty} \|\nabla f(x_k)\| = 0$,
if ∇f is Hölder continuous on $\Re^n$; i.e., there exist two positive scalar $l > 0$ and $M > 0$ such that, for all $x, y \in \Re^n$,

$$\|\nabla f(x) - \nabla f(y)\| \leq M \|x - y\|^l. \tag{5}$$

Proof. Since for all k,

$$\nabla f(x_k)^T d_k = -\|\nabla f(x_k)\|^2 < 0,$$

we have

$$f(x_{k+1}) < f(x_k),$$

which implies that $\{f(x_k)\}$ is a monotonically decreasing sequence. If $f(x_k)$ tends to $-\infty$, then we complete the proof. Therefore, in the following discussion, we assume that $\{f(x_k)\}$ is a bounded set, i.e.,

$$f(x_0) \geq f(x_k) \geq f(\tilde{x}), \text{ for some fixed } \tilde{x} \text{ and all k.}$$

(I) Suppose that (i) is not true. Then, there exists $\varepsilon > 0$ such that, for all k,

$$\|\nabla f(x_k)\| \geq \varepsilon. \tag{6}$$

It follows from (3), (A2) and (A3) that

$$f(x_{k+1}) - f(x_k) \leq -\beta t_k^2 \|\nabla f(x_k)\|^2. \tag{7}$$

The above inequality, (6), and the boundedness of $\{f(x_k)\}$ yield

$$\sum_{k=0}^{\infty} t_k^2 \|\nabla f(x_k)\|^2 \leq (f(\tilde{x}) - f(x_0)) / \beta,$$

and imply that

$$\sum_{k=0}^{\infty} t_k^2 \|\nabla f(x_k)\|^2 < +\infty. \tag{8}$$

By using (2) and (3), we obtain that, for any k,

$$\|x_{k+1} - x_k\|^2 = t_k^2 \|\nabla f(x_k)\|^2.$$

Then, (8) implies that

$$\sum_{k=0}^{\infty} \|x_{k+1} - x_k\|^2 \leq +\infty,$$

which yields that $\{x_k\}$ is convergent, say to a point x^*. From (6), (8), (A2) and (A3), we have

$$\lim_{k \to \infty} t_k = 0. \tag{9}$$

Without loss of generality, we may assume that there exists an index set $\mathbf{K}$ such that

$$\lim_{k \to \infty,\, k \in \mathbf{K}} d_k = d^*.$$

Then from (A1) and (A4), we deduce that, for $k \in \mathbf{K}$,

$$f(x_k + t_k d_k) - f(x_k) \geq \alpha t_k \nabla f(x_k)^T d_k.$$

Hence, for all k∈K, we have

$$[f(x_k + t_k d_k) - f(x_k)]/t_k \geq \alpha \nabla f(x_k)^T d_k.$$

Taking the limit for k∈K and by (9), we have

$$\nabla f(x^*)^T d^* \geq -\alpha \nabla f(x^*)^T d^* \tag{10}$$

which contradict (A1). (recall that $\nabla f(x_k)^T d_k < 0$.)
Therefore, Assumption 1.1, (10) and (3) imply that $\|\nabla f(x^*)\| = 0$. This completes the proof of (i).

(II) Suppose that there exists an infinite index set **K** and a positive scalar $\varepsilon > 0$ such that, for all k∈**K**,

$$\|\nabla f(x_k)\| \geq \varepsilon. \tag{11}$$

Analogous to the proof of (I), it is easy to prove that

$$\lim_{k\to\infty,\, k\in K} t_k = 0.$$

and

$$\lim_{k\to\infty,\, k\in K} t_k^2 \|d_k\|^2 = \lim_{k\to\infty,\, k\in K} t_k^2 \|\nabla f(x_k)\|^2 = 0. \tag{12}$$

Therefore, for all k∈K,

$$f(x_k + t_k d_k) - f(x_k) \geq \alpha t_k \nabla f(x_k)^T d_k.$$

Using (5) and the Taylor expansion formula, we have

$$\begin{aligned} & f(x_k + t_k d_k) - f(x_k) \\ &= t_k \nabla f(x_k)^T d_k + \int_0^1 [\nabla f(x_k + \theta t_k d_k) - \nabla f(x_k)]^T (t_k d_k)\, d\theta \\ &\leq t_k \nabla f(x_k)^T d_k + M \|t_k d_k\|^{1+l}. \end{aligned}$$

The above two inequalities and (2) yield

$$0 \geq (1-\alpha) t_k \nabla f(x_k)^T d_k \geq -M \|t_k d_k\|^{1+l}.$$

Dividing the above inequality by $t_k\|d_k\|$, and taking the limit as $k \to \infty$, k∈**K**, we obtain by (12)

$$\lim_{k\to\infty,\, k\in K} (1-\alpha) \nabla f(x_k)^T d_k / \|d_k\| = 0,$$

which contradicts (11), by (3) and Assumption 1.1.

Let

$$f^* = \inf \{f(x) : x \in \Re^n\}, \ \tilde{f} = \lim_{k \to \infty} f(x_k).$$

The following results, given by Wei, Qi and Jiang (1997) show that our algorithm cannot exhibit the phenomenon of false convergence.

Theorem 2.2. If f is a convex function on $\Re^n$ and the algorithm of (2), (3) and Assumption 1.1 is used, then:

(i) $f(x_k) \to f^*$;
(ii) $\{x_k\}$ is an unbounded set if and only if f has an empty set of minima;
(iii) if f has a nonempty set of minima, then x_k converges to a minimal point of f.

Proof. Note that, for all x and all k,

$$\|x_{k+1} - x\|^2 = \|x_k - x\|^2 + \|x_{k+1} - x_k\|^2 + 2(x_{k+1} - x_k)^T(x - x_k), \tag{13}$$
$$f(x) - f(x_k) \geq \nabla f(x_k)^T(x - x_k),$$

by convexity of f. It follows from (2) and (3) that, for all $x \in \Re^n$ and all k,

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 + \|x_{k+1} - x_k\|^2 + 2t_k(f(x) - f(x_k)). \tag{14}$$

(i) We prove this conclusion by the following three cases (ia), (ib), (ic).
(ia) $f^* = f$. This case is trivial, since $f^* = \tilde{f} = \lim_{k\to\infty} f(x_k)$.
(ib) $\{x_k\}$ is bounded. From the fact that $\{f(x_k)\}$ is a monotonically decreasing sequence, we have that

$$\lim_{k\to\infty} f(x_k) = \tilde{f} > -\infty,$$

which combined with (i) of Theorem 2.1 implies that there exists an index set **K** and a point $x^{**} \in \Re n$ such that

$$\lim_{k\to\infty,\, k\in \mathbf{K}} x_k = x^{**},$$

$$\nabla f(x^{**}) = \lim_{k\to\infty, k\in\mathbf{K}} \nabla f(x_k) = 0.$$

the convexity of f implies that x^{**} is a minimal point of f.
Therefore, $f^* = f(x^{**}) = \tilde{f}$.

(ic) We now assume that $\tilde{f} > -\infty$ and $\{x_k\}$ is unbounded. Suppose that there exists $\tilde{x} \in \Re^n$, $\varepsilon > 0$, and k_1 such that, for all $k \geq k_1$,

$$f(x_k) \geq f(\tilde{x}) + \varepsilon. \tag{15}$$

Setting $x = \tilde{x}$ in (14), we have

$$\|x_{k+1} - \tilde{x}\|^2 \leq \|x_k - \tilde{x}\|^2 + t_k[t_k\|\nabla f(x_k)\|^2 - 2\varepsilon]. \quad (16)$$

Therefore, the fact that $\{f(x_k)\}$ is bounded from below and the inequality

$$f(x_{k+1}) - f(x_k) \geq \alpha t_k\|\nabla f(x_k)\|^2$$

imply that

$$\sum_{k=0}^{\infty} t_k^2\|\nabla f(x_k)\|^2 < +\infty;$$

hence,

$$t_k^2\|\nabla f(x_k)\|^2 \to 0.$$

Then, (16) implies that $\{\|x_k - \tilde{x}\|^2\}$ is a descent sequence for sufficiently large k. It follows that $\{\|x_k\|\}$ is bounded, which contradicts our assumption.

(ii) [$\Rightarrow$]. Assume that f has an optimal solution point x^*. Setting $x = x^*$ in (14), and noting that $f(x^*) \leq f(x_k)$, we obtain

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 + \|x_{k+1} - x_k\|^2. \quad (17)$$

By using Assumption 1.1, we have

$$\sum_{k=0}^{\infty} t_k^2\|\nabla f(x_k)\|^2 < +\infty.$$

Therefore, (2) and (3) yield

$$\sum_{k=0}^{\infty} \|x_{k+1} - x_k\|^2 \leq +\infty. \quad (18)$$

The inequality (17) implies that, for any k,

$$\sum_{i=0}^{k} (\|x_{i+1} - x^*\|^2 - \|x_i - x^*\|^2) \leq \sum_{i=0}^{k} \|x_{i+1} - x_i\|^2.$$

Hence, for any k,

$$0 \leq \|x_{k+1} - x^*\|^2 \leq \|x_0 - x^*\|^2 + \sum_{i=0}^{\infty} \|x_{i+1} - x_i\|^2,$$

which combined with (18) implies that $\{x_k\}$ is bounded. This is a contradiction.

(ii) [$\Leftarrow$]. Suppose that $\{x_k\}$ is bounded. By using the proof of (ib), we can deduce that there exists an accumulation point x^{**} of $\{x_k\}$ such that x^* is a minimal point of f. This contradicts the assumption.

(iii) Suppose that x^* is any fixed optimal solution of (1). Similar to the proof of (ii), (17)-(18) still hold. Then, we deduce that there exists a constant c such that

$$\lim_{k\to\infty} \|x_k - x^*\|^2 = c < +\infty \tag{19}$$

and that the sequence $\{x_k\}$ is bounded. Similar to the proof of (ii) [$\Leftarrow$], we see that $\{x_k\}$ has an accumulation point, which is an optimal solution of (1). By the monotonicity property of $\{f(x_k)\}$, any accumulation point of $\{x_k\}$ is an optimal solution of (1). Suppose that $\{x_k\}$ has two different accumulation points y_1 and y_2. It is clear that y_1 and y_2 are optimal solutions of (1) from the above argument. It follows from (19) that there exist constants c_1 and c_2 such that

$$\lim_{k\to\infty} \|x_k - y_i\|^2 = c_i < +\infty, \quad i = 1, 2. \tag{20}$$

It is easy to see by (2), (3) and (18) that $c_1 = c_2 = 0$ and $\|y_1 - y_2\| = 0$. This is a contradiction. Therefore, $\{x_k\}$ is convergent to an optimal solution point of (1).

DISCUSSION

First, the proof of Theorem 2.1 can be extended to more general search direction

$$d_k = -H_k \nabla f(x_k), \tag{21}$$

where H_k is an n×n symmetric positive-definite matrix, satisfied the following assumption:

(A5) There exist constants $\lambda_1, \lambda_2 > 0$ such that

$$\|H_k \nabla f(x_k)\| \le \lambda_1 \|\nabla f(x_k)\|$$

and

$$\nabla f(x_k)^T H_k \nabla f(x_k) \ge \lambda_2 \|\nabla f(x_k)\|^2,$$

for any integer k.

The conditions on $\{H_k\}$ are not very restrictive. In fact, the commonly used generating formulas of H_k, i.e., quasi-Newton updating methods in convex minimization, can guarantee Condition (A5). We refer the reader to the recent (Byrd and Nocedal 1989).

Second, by slightly modifying the proof of Theorem 3.1 in (Wu 1992), we can obtain the following results, which show that the algorithm with (21) and Assumption 1.1 cannot exhibit the phenomenon of false convergence.

Theorem 3.1. Suppose that (A5) holds. If f is convex function on $\Re^n$, then

$$f^* = \tilde{f}. \tag{22}$$

Third, if f is quasiconvex, Theorem 9.2.4 in (Mangasarian 1969) ensures that case (ii) in Theorem 2.1 cannot happen. Furthermore, by the quasiconvexity of f (Theorem 9.1.4 in (Mangasarian 1969)),

$$\nabla f(x_k)^T(x - x_k) \leq 0.$$

Then, the inequality (13) is reduced to

$$\begin{aligned} \|x_{k+1} - x\|^2 &\leq \|x_k - x\|^2 + \|x_{k+1} - x_k\|^2 \\ &\leq \|x_k - x\|^2 + t_k^2\|\nabla f(x_k)\|^2, \end{aligned}$$

so that

$$\|x_l - x\|^2 \leq \|x_k - x\|^2 + \sum_{j=k} t_j^2\|\nabla f(x_j)\|^2 < +\infty,$$

if $l > k$. Hence, $\{\|x_k\|\}$ is bounded and therefore, the case where $\{x_k\}$ is an unbounded set is not considered in Theorem 2.2.

REFERENCES

ARMIJO, L. 1966. Minimization of functions having continuous partial derivatives. *Pacific Journal of Mathematics* **16**: 1-3.

AUSLENDER, A. and J. P. CROUZEIX. 1989. *Well-behaved Asymptotical Convex Functions*, p. 101-122. Analyse Non-Linéare, Gauthier-Villars, Paris.

BYRD, R. H. and J. NOCEDAL. 1989. A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. *SIAM Journal on Numerical Analysis* **26:** 727-739.

KIWIEL, K. C. and K. MURTY. 1996. Convergence of the steepest descent method for minimizing quasiconvex functions. *Journal of Optimization Theory and Applications* **89**: 221-226.

MANGASARIAN, O. L. 1969. *Nonlinear Programming*. New York: Mc-Graw-Hill.

ROCKAFELLAR, R. L. 1970. *Convex Analysis*. Princeton, New Jersey: Princeton University Press.

TODD, M. J. 1989. On convergence properties of algorithms for unconstrained minimization. *IMA Journal of Numerical Analysis* **9**: 435-441.

WEI, Z., L. QI and H. JIANG. 1997. Some convergence properties of descent methods. *Journal of Optimization Theory and Applications* **95**: 177-188.

WU, S. Q. 1992. Convergence properties of descent methods for unconstrained minimization. *Optimization* **26**: 229-237.