JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 89, No. 1, pp. 221-226, APRIL 1996

TECHNICAL NOTE

Convergence of the Steepest Descent Method for Minimizing Quasiconvex Functions[1,2]

K. C. KIWIEL[3] AND K. MURTY[4]

Communicated by O. L. Mangasarian

Abstract. To minimize a continuously differentiable quasiconvex function $f: \mathbb{R}^n \to \mathbb{R}$, Armijo's steepest descent method generates a sequence $x^{k+1} = x^k - t_k \nabla f(x^k)$, where $t_k > 0$. We establish strong convergence properties of this classic method: either $x^k \to \bar{x}$, s.t. $\nabla f(\bar{x}) = 0$; or $\arg\min f = \emptyset$, $\|x^k\| \to \infty$, and $f(x^k) \downarrow \inf f$. We also discuss extensions to other line searches.

Key Words. Steepest descent methods, convex programming, Armijo's line search.

1. Introduction

To minimize a continuously differentiable quasiconvex function $f: \mathbb{R}^n \to \mathbb{R}$, Cauchy's steepest descent method (Ref. 1) with Armijo's stepsizes (Ref. 2) generates a sequence $\{x^k\}$ via

$$x^{k+1} = x^k - t_k g^k, \quad g^k = \nabla f(x^k), \qquad k = 0, 1, \ldots, \tag{1}$$

[1]The research of the first author was supported by the Polish Academy of Sciences. The second author acknowledges the support of the Department of Industrial Engineering, Hong Kong University of Science and Technology.

[2]We wish to thank two anonymous referees for their valuable comments. In particular, one referee has suggested the use of quasiconvexity instead of convexity of f.

[3]Professor, Systems Research Institute, Warsaw, Poland.

[4]Professor, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan.

where

$$t_k = \arg\max\{t: f(x^k - tg^k) \le f(x^k) - \alpha t\|g^k\|^2, t = 2^{-i}, i = 0, 1, \ldots\}, \tag{2}$$

with $\alpha \in (0, 1)$. We prove in Section 2 the following strong convergence result.

Theorem 1.1. Global Convergence. Either $x^k \to \bar{x} \in X := \{x: \nabla f(x) = 0\}$, or $\check{X} := \arg\min f = \emptyset$, $\|x^k\| \to \infty$, and $f(x^k) \downarrow \inf f$.

A closely related result appeared in Ref. 3 after an earlier version of this note was accepted. The present version provides a considerably simpler convergence proof that permits generalization to the quasiconvex case. Other related results for nondifferentiable optimization methods are given in Ref. 4 and Ref. 5, Remark 3.2. These relations and extensions are discussed in Section 3.

2. Global Convergence of Steepest Descent

We make the following standing assumption that generalizes Armijo's condition (2).

Assumption 2.1. Let $\phi: \mathbb{R}_+ \to \mathbb{R}_+$ be a function such that:

(A1) $\exists \alpha \in (0, 1)$, $\tau_\alpha > 0$, $\forall t \in (0, \tau_\alpha]$: $\phi(t) \le \alpha t$,

(A2) $\exists \beta > 0$, $\tau_\beta \in (0, \infty]$, $\forall t \in (0, \tau_\beta] \cap \mathbb{R}$: $\phi(t) \ge \beta t^2$,

(A3) $\forall k, f(x^{k+1}) \le f(x^k) - \phi(t_k)\|g^k\|^2$ and $0 < t_k \le \tau_\beta$ in (1),

(A4) $\exists \gamma > 1$, $\tau_\gamma > 0$, $\forall k$: $t_k \ge \tau_\gamma$ or $[\exists \tilde{t}_k \in [t_k, \gamma t_k]: f(x^k - \tilde{t}_k g^k) \ge f(x^k) - \phi(\tilde{t}_k)\|g^k\|^2]$.

Note that (2) corresponds to

$$\phi(t) = \alpha t, \qquad \beta = \alpha, \qquad \gamma = 2, \qquad \tau_\alpha = \tau_\beta = \tau_\gamma = 1.$$

As in Ref. 4, we start by considering the condition

$$f(x^k) \ge f(\tilde{x}), \qquad \text{for some fixed } \tilde{x} \text{ and all } k, \tag{3}$$

which holds if $\check{X} \ne \emptyset$ or $\tilde{x}$ is a cluster point of $\{x^k\}$.

Lemma 2.1. If (3) holds, then

$$\sum_{k=0}^{\infty} t_k^2 \|g^k\|^2 \le [f(x^0) - f(\tilde{x})]/\beta. \tag{4}$$

Moreover, $x^k \to \bar{x}$ for some $\bar{x}$.

Proof. By (A2)–(A3),

$$\beta t_k^2 \|g^k\|^2 \leq \phi(t_k)\|g^k\|^2 \leq f(x^k) - f(x^{k+1});$$

adding these inequalities yields (4). Next, since $\langle g^k, \tilde{x} - x^k \rangle \leq 0$ by (3) and quasiconvexity of f [Ref. 6, Theorem 9.1.4], and since $x^k - x^{k+1} = t_k g^k$, we deduce that

$$\begin{aligned} \|\tilde{x} - x^{k+1}\|^2 &= \|\tilde{x} - x^k\|^2 + 2\langle \tilde{x} - x^k, x^k - x^{k+1}\rangle + \|x^{k+1} - x^k\|^2 \\ &\leq \|\tilde{x} - x^k\|^2 + t_k^2 \|g^k\|^2, \end{aligned}$$

so that

$$\|\tilde{x} - x^l\|^2 \leq \|\tilde{x} - x^k\|^2 + \sum_{j=k}^{\infty} t_j^2 \|g^j\|^2 < \infty,$$

if $l > k$. Hence, $\{x^k\}$ is bounded and has a cluster point $\bar{x}$, so we may set $\tilde{x} = \bar{x}$ above to deduce from (4) for any $\epsilon > 0$ the existence of k such that $\|\bar{x} - x^k\|^2 \leq \epsilon/2$ and

$$\sum_{j=k}^{\infty} t_j^2 \|g^j\|^2 \leq \epsilon/2;$$

thus, $\|\bar{x} - x^l\|^2 \leq \epsilon$ for all $l > k$, i.e., $x^k \to \bar{x}$. □

Lemma 2.2. If $\bar{x}$ is a cluster point of $\{x^k\}$, then $\bar{x} \in X$, i.e., $\nabla f(\bar{x}) = 0$.

Proof. Suppose that $x^k \xrightarrow{K} \bar{x}$, but $\bar{g} := \nabla f(\bar{x}) \neq 0$. Then, $t_k \xrightarrow{K} 0$ from [cf. (A2)–(A3)]

$$0 \leq \beta t_k^2 \|g^k\|^2 \leq f(x^k) - f(x^{k+1}) \xrightarrow{K} 0,$$

with $g^k \xrightarrow{K} \bar{g} \neq 0$ and $f(x^k) \downarrow f(\bar{x})$ by continuity. Thus, for all large $k \in K$,

$$f(x^k - \tilde{t}_k g^k) - f(x^k) \geq -\phi(\tilde{t}_k)\|g^k\|^2 \geq -\alpha \tilde{t}_k \|g^k\|^2, \tag{5}$$

by (A4) and (A1), where the left side equals $-\tilde{t}_k \langle g^k, \nabla f(x^k - \breve{t}_k g^k)\rangle$ for some $\breve{t}_k \in [0, \tilde{t}_k]$ by the mean-value theorem, and by (A4), $0 \leq \tilde{t}_k \leq \gamma t_k \xrightarrow{K} 0$. Hence, dividing (5) by $\tilde{t}_k$ and letting $k \xrightarrow{K} \infty$ yields $-\|\bar{g}\|^2 \geq -\alpha \|\bar{g}\|^2$, a contradiction with $\alpha < 1$ [cf. (A1)]. □

We can now prove Theorem 1.1 under Assumption 2.1 that generalizes (2).

Proof of Theorem 1.1. If (3) holds, e.g., $\check{X} \neq \emptyset$ or $\{x^k\}$ has a cluster point, then the preceding results yield $x^k \to \bar{x} \in X$. If $\|x^k\| \not\to \infty$, then $\{x^k\}$ has a cluster point. If $\lim_{k\to\infty} f(x^k) > \inf f$, then (3) holds. □

3. Discussion of Other Line Searches

First, suppost that $\alpha \in (\frac{1}{2}, 1)$ and f is convex. Then, the proof of Lemma 2.2 simplifies, since

$$\langle g^k, x^k - \tilde{x}\rangle \geq f(x^k) - f(\tilde{x}) \geq f(x^k) - f(x^{k+1}) \geq \alpha t_k \|g^k\|^2,$$

$$\begin{aligned}\|\tilde{x} - x^{k+1}\|^2 - \|\tilde{x} - x^k\|^2 &\leq -2\alpha t_k^2 \|g^k\|^2 + t_k^2 \|g^k\|^2 \\ &= -(2\alpha - 1)\|x^{k+1} - x^k\|^2 \leq 0.\end{aligned}$$

This observation is used in Ref. 7 to prove that $x^k \to \bar{x} \in \check{X}$ if $\check{X} \neq \emptyset$ and ∇f is Lipschitz continuous; thus, our result improves that of Ref. 7.

Second, it is easy to verify Theorem 1.1 for any line search for which Lemma 2.3 holds and for all k,

$$f(x^{k+1}) \leq f(x^k) - \alpha t_k \|g^k\|^2$$

and

$$t_k \in (0, t_{\max}], \qquad \text{for some fixed } t_{\max} > 0.$$

Such stepsizes may be found by many procedures [Refs. 8–12]. Note that exact line searches are not admissible, but one may use, as in Ref. 12, Section 10.7.2,

$$t_k \approx \arg\min\{f(x^k - tg^k) : f(x^k - tg^k) \leq f(x^k) - \alpha t \|g^k\|^2,\ 0 < t \leq t_{\max}\}.$$

Third, under (A1)–(A2) to satisfy (A3)–(A4), one may let (cf. the proof of Lemma 2.3)

$$\begin{aligned} t_k = \arg\max\{t : f(x^k - tg^k) \leq f(x^k) - \phi(t)\|g^k\|^2, \\ t = 2^{-i}\min[\tau_\alpha, \tau_\beta],\ i = 0, 1, \ldots\}. \end{aligned} \tag{6}$$

We note that (6) with $\phi(t) = \alpha t^2$ was used in Ref. 13. Again, the Armijo-type search (6) may be relaxed as in the preceding paragraph. In particular, one may use

$$t_k \approx \check{t}_k := \arg\min\{f(x^k - tg^k) + \alpha t^2 \|g^k\|^2 : t > 0\}.$$

If f is pseudoconvex, then $X = \check{X}$ [Ref. 6, Theorem 9.3.3]; so if $\check{X} \neq \emptyset$ and $t_k = \check{t}_k$ for all k, then $x^k \to \bar{x} \in X$; thus, we recover the result of Ref. 14.

Fourth, one may verify Assumption 2.1 for the algorithms of Ref. 3; in their notation, let $\phi(t)=\beta t^2$ with $\beta = L\delta_2/2(1-\delta_2)$ for Algorithm A, $\phi=\psi$ for Algorithm B. Theorem 1.1 is stronger than Theorem 3 of Ref. 3, and our proof is simpler.

We note that quasiconvexity of f is necessary for Lemma 2.2, and consequently Theorem 1.1. For example, let

$$n=2, \qquad f(x)=e^{x_1}-x_2^2, \qquad x^0=(0,0)^T.$$

Each of the above methods generates

$$x^k=(x_1^k, 0)^T, \text{ with } x_1^k \downarrow -\infty \text{ and } f(x^k) \downarrow 0, \text{ while } \inf f=-\infty.$$

References

1. Cauchy, A., *Méthode Générale pour la Résolution des Systèmes d'Équations Simultanées*, Comptes Rendus de Académie des Sciences, Paris, Vol. 25, pp. 536–538, 1847.
2. Armijo, L., *Minimization of Functions Having Continuous Partial Derivatives*, Pacific Journal of Mathematics, Vol. 16, pp. 1–3, 1966.
3. Burachik, R., Graña Drummond, L. M., Iusem, A. N., and Svaiter, B. F., *Full Convergence of the Steepest Descent Method with Inexact Line Searches*, Optimization, Vol. 32, pp. 137–146, 1995.
4. Kiwiel, K. C., *An Aggregate Subgradient Method for Nonsmooth Convex Minimization*, Mathematical Programming, Vol. 27, pp. 320–341, 1983.
5. Kiwiel, K. C., *A Direct Method of Linearizations for Continuous Minimax Problems*, Journal of Optimization Theory and Applications, Vol. 55, pp. 271–287, 1987.
6. Mangasarian, O. L., *Nonlinear Programming*, Mc-Graw-Hill, New York, New York, 1969; Reprinted by SIAM, Philadelphia, Pennsylvania, 1994.
7. Bereznev, V. A., Karmanov, V. G., and Tretyakov, A. A., *On the Stabilizing Properties of the Gradient Method*, Zhurnal Vychislitelnoi Matematiki i Matematicheskoi Fiziki, Vol. 26, pp. 134–137, 1986 (in Russian).
8. Polak, E., *Computational Methods in Optimization*, Academic Press, New York, New York, 1971.
9. Bazaraa, M. S., Sherali, H. D., and Shetty, C. M., *Nonlinear Programming: Theory and Algorithms*, 2nd Edition, Wiley, New York, New York, 1993.
10. Dennis, J. E., Jr., and Schnabel, R. B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, Englewood Cliffs, New Jersey, 1983.
11. Fletcher, R., *Practical Methods of Optimization*, 2nd Edition, Wiley, Chichester, England, 1987.
12. Murty, K. G., *Linear Complementarity, Linear and Nonlinear Programming*, Heldermann Verlag, Berlin, Germany, 1988.

13. KIWIEL, K. C., *A Linearization Method for Minimizing Certain Quasidifferentiable Functions*, Mathematical Programming Study, Vol. 29, pp. 85–94, 1986.
14. IUSEM, A. N., and SVAITER, B. F., *A Proximal Regularization of the Steepest Descent Method*, RAIRO Rechérche Opérationelle (to appear).