Pacific Journal of Mathematics

MINIMIZATION OF FUNCTIONS HAVING LIPSCHITZ CONTINUOUS FIRST PARTIAL DERIVATIVES

LARRY ARMIJO

Vol. 16, No. 1 November 1966

PACIFIC JOURNAL OF MATHEMATICS
Vol. 16, No. 1, 1966

MINIMIZATION OF FUNCTIONS HAVING LIPSCHITZ CONTINUOUS FIRST PARTIAL DERIVATIVES

LARRY ARMIJO

A general convergence theorem for the gradient method is proved under hypotheses which are given below. It is then shown that the usual steepest descent and modified steepest descent algorithms converge under the some hypotheses. The modified steepest descent algorithm allows for the possibility of variable stepsize.

For a comparison of our results with results previously obtained, the reader is referred to the discussion at the end of this paper.

Principal conditions. Let f be a real-valued function defined and continuous everywhere on E^n (real Euclidean n-space) and bounded below E^n. For fixed $x_0 \in E^n$ define $S(x_0) = \{x : f(x) \leqq f(x_0)\}$. The function f satisfies: condition I if there exists a *unique* point $x^* \in E^n$ such that $f(x^*) = \inf_{x \in E^n} f(x)$; Condition II at x_0 if $f \in C^1$ on $S(x_0)$ and $\nabla f(x) = 0$ for $x \in S(x_0)$ if and only if $x = x^*$; Condition III at x_0 if $f \in C^1$ on $S(x_0)$ and ∇f is Lipschitz continuous on $S(x_0)$, i.e., there exists a Lipschitz constant $K > 0$ such that $|\nabla f(y) - \nabla f(x)| \leqq K|y - x|$ for every pair $x, y \in S(x_0)$; Condition IV at x_0 if $f \in C^1$ on $S(x_0)$ and if $r > 0$ implies that $m(r) > 0$ where $m(r) = \inf_{x \in S_r(x_0)} |\nabla f(x)|$, $S_r(x_0) = S_r \cap S(x_0)$, $S_r = \{x : |x - x^*| \geqq r\}$, and x^* is any point for which $f(x^*) = \inf_{x \in E^n} f(x)$. (If $S_r(x_0)$ is void, we define $m(r) = \infty$.)

It follows immediately from the definitions of Conditions I through IV that Condition IV implies Conditions I and II, and if $S(x_0)$ is bounded, then Condition IV is equivalent to Conditions I and II.

2. The convergence theorem. In the convergence theorem and its corollaries, we will assume that f is a real-valued function defined and continuous everywhere on E^n, bounded below on E^n, and that Conditions III and IV hold at x_0.

THEOREM. *If* $0 < \delta \leqq 1/4K$, *then for any* $x \in S(x_0)$, *the set*

$$(1) \quad S^*(x, \delta) = \{x_\lambda : x_\lambda = x - \lambda \nabla f(x),\ \lambda > 0,\ f(x_\lambda) - f(x) \leqq -\delta |\nabla f(x)|^2\}$$

Received January 30, 1964. The research for this paper was supported in part by General Dynamics/Astronautics, San Diego, California, Rice University, Houston, Texas, and the Martin Company, Denver, Colorado. The author is currently employed by the National Engineering Science Company, Houston, Texas.

is a nonempty subset of $S(x_0)$ *and any sequence* $\{x_k\}_{k=0}^{\infty}$ *such that* $x_{k+1} \in S^*(x_k, \delta)$, $k = 0, 1, 2, \cdots$, *converges to the point* x^* *which minimizes* f.

Proof. If $x \in S(x_0)$, $x_\lambda = x - \lambda \nabla f(x)$ and $0 \leqq \lambda \leqq 1/K$, Condition III and the mean value theorem imply the inequality $f(x_\lambda) - f(x) \leqq -(\lambda - \lambda^2 K)\,|\nabla f(x)|^2$ which in turn implies that $x_\lambda \in S^*(x, \delta)$ for

$$\lambda_1 \leqq \lambda \leqq \lambda_2\,, \qquad \lambda_i = \frac{1}{2K}[1 + (-1)^i \sqrt{1 - 4\delta K}]\,,$$

so that $S^*(x, \delta)$ is a nonempty subset of $S(x_0)$. If $\{x_k\}_{k=0}^{\infty}$ is any sequence for which $x_{k+1} \in S^*(x_k, \delta)$, $k = 0, 1, 2, \cdots$, then (1) implies that sequence $\{f(x_k)\}_{k=0}^{\infty}$, which is bounded below, is monotone nonincreasing and hence that $|\nabla f(x_k)| \to 0$ as $k \to \infty$. The remainder of the theorem follows from Condition IV.

COROLLARY 1. (*The Steepest Descent Algorithm*) *If*

$$x_{k+1} = x_k - \frac{1}{2K}\nabla f(x_k),\ k = 0, 1, 2, \cdots$$

then the sequence $\{x_k\}_{k=0}^{\infty}$ *converges to the point* x^* *which minimizes* f.

Proof. It follows from the proof of the convergence theorem that the sequence $\{x_k\}_{k=0}^{\infty}$ defined in the statement of Corollary 1 is such that $x_{k+1} \in S^*(x_k, 1/4K)$, $k = 0, 1, 2, \cdots$.

COROLLARY 2. (*The Modified Steepest Descent Algorithm*) *If* α *is an arbitrarily assigned positive number,* $\alpha_m = \alpha/2^{m-1}$, $m = 1, 2, \cdots$, *and* $x_{k+1} = x_k - \alpha_{m_k}\nabla f(x_k)$ *where* m_k *is the smallest positive integer for which*

$$(2) \qquad f(x_k - \alpha_{m_k}\nabla f(x_k)) - f(x_k) \leqq -\frac{1}{2}\alpha_{m_k}\,|\nabla f(x_k)|^2\,,$$

$k = 0, 1, 2, \cdots$, *then the sequence* $\{x_k\}_{k=0}^{\infty}$ *converges to the point* x^* *which minimizes* f.

Proof. It follows from the proof of the convergence theorem that if $x \in S(x_0)$ and $x_\lambda = x - \lambda \nabla f(x)$, then $f(x_\lambda) - f(x) \leqq -(1/2)\lambda\,|\nabla f(x)|^2$ for $0 \leqq \lambda \leqq 1/2K$. If $\alpha \leqq 1/2K$, then for the sequence $\{x_k\}_{k=0}^{\infty}$ in the statement of Corollary 2, $m_k = 1$ and $x_{k+1} \in S^*(x_k, (1/2)\alpha)$, $k = 0, 1, 2, \cdots$. If $\alpha > 1/2K$, then the integers m_k exist and $\alpha_{m_k} > 1/4K$ so that $x_{k+1} \in S^*(x_k, 1/8K)$, $k = 0, 1, 2, \cdots$.

3. Discussion. The convergence theorem proves convergence under hypotheses which are more restrictive than those imposed by Curry [1] but less restrictive than those imposed by Goldstein [2]. However, both the algorithms which we have considered would be considerably easier to apply than the algorithm proposed by Curry since his algorithm requires the minimization of a function of one variable at each step. The method of Goldstein requires the assumption that $f \in C^2$ on $S(x_0)$ and that $S(x_0)$ be bounded. It also requires knowledge of a bound for the norm of the Hessian matrix of f on $S(x_0)$, but yields an estimate for the ultimate rate of convergence of the gradient method. It should be pointed out that the modified steepest descent algorithm of Corollary 2 allows for the possibility of variable stepsize and does not require knowledge of the value of the Lipschitz constant K.

The author is indebted to the referee for his comments and suggestions.

REFERENCES

1. H. B. Curry, *The method of steepest descent for nonlinear minimization problems*, Quart. Appl. Math. **2** (1944), 258-263.

2. A. A. Goldstein, *Cauchy's method of minimization*, Numer. Math. **4** (2), (1962), 146-150.

PACIFIC JOURNAL OF MATHEMATICS

EDITORS

H. SAMELSON
Stanford University
Stanford, California

*J. DUGUNDJI
University of Southern California
Los Angeles, California 90007

R. M. BLUMENTHAL
University of Washington
Seattle, Washington 98105

RICHARD ARENS
University of California
Los Angeles, California 90024

ASSOCIATE EDITORS

E. F. BECKENBACH B. H. NEUMANN F. WOLF K. YOSIDA

SUPPORTING INSTITUTIONS

UNIVERSITY OF BRITISH COLUMBIA
CALIFORNIA INSTITUTE OF TECHNOLOGY
UNIVERSITY OF CALIFORNIA
MONTANA STATE UNIVERSITY
UNIVERSITY OF NEVADA
NEW MEXICO STATE UNIVERSITY
OREGON STATE UNIVERSITY
UNIVERSITY OF OREGON
OSAKA UNIVERSITY
UNIVERSITY OF SOUTHERN CALIFORNIA
STANFORD UNIVERSITY
UNIVERSITY OF TOKYO
UNIVERSITY OF UTAH
WASHINGTON STATE UNIVERSITY
UNIVERSITY OF WASHINGTON

* * *

AMERICAN MATHEMATICAL SOCIETY
CHEVRON RESEARCH CORPORATION
TRW SYSTEMS
NAVAL ORDNANCE TEST STATION

Printed in Japan by International Academic Printing Co., Ltd., Tokyo Japan

Pacific Journal of Mathematics
Vol. 16, No. 1 November, 1966

Larry Armijo, *Minimization of functions having Lipschitz continuous first partial derivatives* 1
Edward Martin Bolger and William Leonard Harkness, *Some characterizations of exponential-type distributions* 5
James Russell Brown, *Approximation theorems for Markov operators* 13
Doyle Otis Cutler, *Quasi-isomorphism for infinite Abelian p-groups* 25
Charles M. Glennie, *Some identities valid in special Jordan algebras but not valid in all Jordan algebras* 47
Thomas William Hungerford, *A description of* $\mathrm{Mult}_i\,(A^1, \cdots, A^n)$ *by generators and relations* 61
James Henry Jordan, *The distribution of cubic and quintic non-residues* 77
Junius Colby Kegley, *Convexity with respect to Euler-Lagrange differential operators* 87
Tilla Weinstein, *On the determination of conformal imbedding* 113
Paul Jacob Koosis, *On the spectral analysis of bounded functions* 121
Jean-Pierre Kahane, *On the construction of certain bounded continuous functions* 129
V. V. Menon, *A theorem on partitions of mass-distribution* 133
Ronald C. Mullin, *The enumeration of Hamiltonian polygons in triangular maps* 139
Eugene Elliot Robkin and F. A. Valentine, *Families of parallels associated with sets* 147
Melvin Rosenfeld, *Commutative F-algebras* 159
A. Seidenberg, *Derivations and integral closure* 167
S. Verblunsky, *On the stability of the set of exponents of a Cauchy exponential series* 175
Herbert Walum, *Some averages of character sums* 189