SIAM REVIEW
Vol. 13, No. 2, April 1971

Downloaded 04/19/13 to 129.11.21.2. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

CONVERGENCE CONDITIONS FOR ASCENT METHODS. II: SOME CORRECTIONS*

PHILIP WOLFE†

Abstract. Some corrections and amplifications are appended to *Convergence conditions for ascent methods*, this Review, 11 (1969), pp. 226–235.

My paper [1] on "ascent methods" has two misstatements I would like to set right, particularly since their correction makes in itself a contribution to its subject matter. The paper dealt with the familiar attack on the problem of finding the maximum of a function of several variables by procedures which, beginning with the point x_0, repeat these two steps for $n = 0, 1, \cdots$: (i) Choose a direction y_n in the space of the variables in which the function can increase; (ii) Find some approximation to $t_n = t$ solving $\max\{f(x_n + ty_n): t \geqq 0\}$, and define $x_{n+1} = x_n + t_n y_n$.

The first correction is to the statement that for an ascent method satisfying certain direction and step-size conditions "convergence to a stationary point which is not a relative maximum is unlikely" [1, p. 226]. Theorem 4 showed that when steepest ascent is used in the maximization of a quadratic form it usually behaves decently: if the form is concave, it converges to the maximum; and if not—in which case there is no maximum, only a stationary point—it will tend to $+\infty$ unless the starting point happens to lie in the subspace spanned by the negative-valued eigenvectors. Although no results were given for nonquadratic functions, I felt "safe in asserting that in order for the sequence to converge to a point other than a relative maximum, the directions and step lengths must be chosen with great care" [1, p. 231]. The feeling of safety was illusory: there is a simple counterexample.

Let $f(x, y) = -\frac{1}{3}x^3 - \frac{1}{2}y^2$; some of its contours $f(x, y) = c$ are sketched in Fig. 1. The slope of the contour passing through (x, y) is $-x^2/y$, and the function is strictly concave for $x > 0$. Let the point (x, y) have $0 < x < 1$, $y > 0$. The ray $\{(x, y) + t\nabla f(x, y): t \geqq 0\} = \{(x - tx^2, y - ty): t \geqq 0\}$ crosses the y-axis for $t = 1/x > 1$, when $y - ty < 0$, and is thus tangent to some contour of f in the fourth quadrant, at the point (x^*, y^*) which is the steepest ascent successor of (x, y), having $0 < x^* < x$, $y^* < 0$. Similarly, if $0 < x < 1$, $y < 0$, then the successor (x^*, y^*) of (x, y) has $0 < x^* < x$, $y^* > 0$. In any case, the steepest ascent step maps the strip $0 < x < 1$ into itself, so that $x > 0$ always, whence $f(x, y) < 0$ for all points of the ascent sequence. It follows from Theorem 2 that $\nabla f(x, y)$ tends to zero, so the ascent sequence converges to $(0, 0)$, a stationary point which is *not* a local minimum, whenever the starting point satisfies $0 < x < 1$.

The possibility of such an example was pointed out by Ralph Gomory, who suggested constructing an example for which the differential equation $dx/dt = \nabla f(x)$ would stall in a stationary nonminimum point. The example above does

* Received by the editors February 3, 1970.

† IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598. This work was supported in part by the Office of Naval Research under Contract Nonr-3775(00).

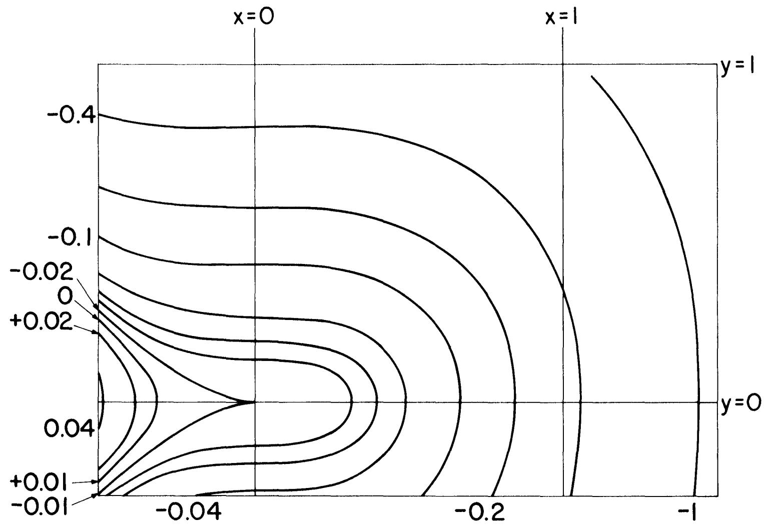

FIG. 1. *Contours of* $-\frac{1}{3}x^3 - \frac{1}{2}y^2$

just that. It is even more obdurate: The powerful Davidon–Fletcher–Powell algorithm [2] also converges to (0, 0) when started sufficiently close to it. Studying the behavior of an ascent procedure on a problem by studying the trajectories of the related differential equation is a surprisingly sturdy heuristic; I know of no case in which it is completely misleading—unlike inferring properties of general functions from properties of their quadratic approximations. The difficulty with our example here lies, of course, in the term x^3: the best quadratic approximation to $f(x, y)$ is just $-\frac{1}{2}y^2$, which is not good enough. I still "feel safe" in believing that steepest ascent will almost never converge to a stationary point at which the Hessian of f is nonsingular and not negative semidefinite. The question is interesting, but requires some new ideas before being resolved.

The second correction regards the proposition I quoted on p. 231 to the effect that if in part (ii) of the ascent process described above one satisfies the condition of Curry [3] that t_n shall give the first stationary value of f for $t > 0$ (in which case we shall call $\{x_n\}$ a *Curry sequence*), and the series $\sum_n \cos \theta_n$ diverges (where θ_n denotes the acute angle between y_n and the gradient $\nabla f(x_n)$), then any limit point of the Curry sequence must be a stationary point of f (i.e., satisfy $\nabla f(x) = 0$). Prompted by a comment of Michael J. D. Powell to look at the proposition more closely, I found it false. As suitable propositions dealing with the same ideas, I offer Theorems 5 and 6 below (continuing the numeration of [1]). Theorem 5 gives the desired sufficient condition, which turns out to involve $\sum_n \cos^2 \theta_n$, and Theorem 6 in part shows it to be necessary for certain functions. (Since this sort of analysis can be framed in terms of using "a different ε at each step" [1, p. 228], I no longer "doubt that (such an extension) would be useful." Indeed, a detailed study of the ascent process using the ideas required in stating the theorems below has turned out to be essential to the study of rates of convergence of some optimization procedures [4], a task which is much more important than that of establishing convergence.)

Downloaded 04/19/13 to 129.11.21.2. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 04/19/13 to 129.11.21.2. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

We will say that $f \in C^2$ is *of bounded curvature* if there is a constant A such that $A \geqq |\lambda|$ for every eigenvalue λ of the Hessian $H(x) = [\partial^2 f(x)/\partial x_i \partial x_j]$ for every x, and that f is *strongly concave* if the inequality $\lambda \leqq -a$ is similarly satisfied for some $a > 0$.

THEOREM 5. *If f is of bounded curvature, $\{x_n\}$ is a Curry sequence, and $\sum_n \cos^2 \theta_n = \infty$, then either $\lim_{n\to\infty} f(x_n) = +\infty$ or $\liminf_{n\to\infty} |\nabla f(x_n)| = 0$.*

THEOREM 6. (a) *If f is strongly concave, then it assumes its maximum value M at some point $\bar{x}$. If f is also of bounded curvature and $\{x_n\}$ is a Curry sequence, then* (b) $\lim_{n\to\infty} x_n = \bar{x}$ *if* $\sum_n \cos^2 \theta_n = \infty$, *while* (c) *if* $\sum_n \cos^2 \theta_n < \infty$, *then* $\liminf_{n\to\infty} f(x_n) < M$ *and* $\liminf_{n\to\infty} |\nabla f(x_n)| > 0$.

Proofs. Normalize so that $|y_n| = 1$ and let $g(t) = f(x_n + ty_n) - f(x_n)$. Then, if $t_n = t$ gives the first zero of $g'(t) = \nabla f(x_n + ty_n)^T y_n$, we have $x_{n+1} = x_n + t_n y_n$. Since $g''(t) = y_n^T H(x_n + ty_n) y_n \geqq -A$, we have $g'(t) \geqq g'(0) - At$ and $g(t) \geqq g'(0)t - \frac{1}{2}At^2$ for all $t \geqq 0$. Consequently $t_n \geqq g'(0)/A$ so that $f(x_{n+1}) - f(x_n) = g(t_n) \geqq g(g'(0)/A) \geqq g'(0)^2/2A = (\nabla f(x_n)^T y_n)^2/2A = |\nabla f(x_n)|^2 (\cos^2 \theta_n)/2A$, or

$$f(x_{n+1}) - f(x_n) \geqq |\nabla f(x_n)|^2 (\cos^2 \theta_n)/2A. \tag{1}$$

Now the sequence $\{f(x_n)\}$ is monotone nondecreasing. If it is bounded, so that $\sum_n [f(x_{n+1}) - f(x_n)]$ converges, while $\sum_n \cos^2 \theta_n$ diverges, the sequence $|\nabla f(x_n)|$ cannot remain bounded away from zero, and Theorem 5 is proved.

If f is strongly concave, we can proceed as above from the relation $g''(t) \leqq -a$ to show that

$$g(t) \leqq g'(0)t - \tfrac{1}{2}at^2 \tag{2}$$

and thus

$$f(x_{n+1}) - f(x_n) \leqq |\nabla f(x_n)|^2 (\cos^2 \theta_n)/2a. \tag{3}$$

Formula (2) tells us that $f(x_n + ty_n) < f(x_n)$ for any y_n and t with $|ty_n| > 2g'(0)/a$, so that all the values of f are dominated by those taken on the sphere of radius $2g'(0)/a$ about x_n, whence f does have a maximum value M assumed at some point $\bar{x}$ of that sphere, proving (a) of Theorem 6.

Formula (2) also applies if we set $x_n = \bar{x}$ and let y_n have arbitrary direction, from which it follows that the point at which the maximum is assumed is unique. Relation (3) shows (letting $x_n = y$) that $\nabla f(y) = 0$ implies $y = \bar{x}$, whence the solution of $\nabla f(y) = 0$ is unique. In the case that $\sum_n \cos^2 \theta_n = \infty$, we may apply Theorem 5, finding that any point of accumulation y of the bounded Curry sequence $\{x_n\}$ satisfies $\nabla f(y) = 0$. Since that point of accumulation is unique, $\lim_{n\to\infty} x_n = \bar{x}$, proving (b).

Applying (1) with the choice $\theta_n = 0$ gives the bound

$$M - f(x_n) \geqq |\nabla f(x_n)|^2/2A. \tag{4}$$

Dividing (3) by (4) gives $[f(x_{n+1}) - f(x_n)]/[M - f(x_n)] \leqq (A/a) \cos^2 \theta_n$, or

$$\frac{M - f(x_{n+1})}{M - f(x_n)} \geqq 1 - \frac{A}{a} \cos^2 \theta_n.$$

Assuming $\sum_n \cos^2 \theta_n < \infty$, let N be so large that $\cos^2 \theta_n < a/A$ for $n \geqq N$. Then

$$M - f(x_n) \geqq [M - f(x_N)] \prod_{k=N}^{n-1} \left(1 - \frac{A}{a} \cos^2 \theta_n\right)$$

for $n \geqq N$. As a consequence of the theory of infinite products (see, e.g., [5]), the term on the right has a limit $B > 0$, so that $\lim_{n \to \infty} f(x_n) = M - B < M$. Finally, by choosing $y_n = (\bar{x} - x_n)/|\bar{x} - x_n|$ for (3), in which case $x_{n+1} = \bar{x}$, (3) implies $|\nabla f(x_n)|^2 \geqq 2a[M - f(x_n)] \geqq 2aB$ for all n, which completes the proof of part (c).

Since the paper [1] appeared, several people sent me copies of closely related work. The most relevant of these is Richard Elkin's dissertation [6], which covers considerably more ground than my paper does, dealing with the choice of direction for certain problems as well as with the choice of step length. In particular, his Theorem 3.2.1 gives roughly the same result as Theorem 2 of [1]. I understand that even more general results will be given in a book to be published by J. M. Ortega and W. Rheinboldt in 1970. For some versions of the method of conjugate gradients (extended to the extremization of nonquadratic functions), James Daniel [7] has established convergence and rate-of-convergence estimates under fairly loose step-size requirements. Ljubic [8] has made a very interesting study of convergence conditions in the case of a quadratic function; Theorem 6 above turns out to be an extension of his Corollary 1. Finally, it appears that Poljak was first to use the term "strong convexity" as done here. His paper [9] develops existence theorems for the minimization of a strongly convex function on a Banach space, and I understand that his more recent work is concerned with the same kind of problem studied here.

REFERENCES

[1] PHILIP WOLFE, *Convergence conditions for ascent methods*, this Review, 11 (1969), pp. 226–235.

[2] R. FLETCHER AND M. J. D. POWELL, *A rapidly convergent descent method for minimization*, Comput. J., 6 (1963), pp. 163–168.

[3] HASKELL B. CURRY, *The method of steepest descent for non-linear minimization problems*, Quart. Appl. Math., 2 (1944), pp. 258–261.

[4] PHILIP WOLFE, *Convergence theory in nonlinear programming*, Integer and Nonlinear Programming, J. Abadie, ed., North Holland, Amsterdam, 1970.

[5] KONRAD KNOPP, *The Theory and Application of Infinite Series*, Hafner, New York, 1950.

[6] RICKARD M. ELKIN, *Convergence theorems for Gauss-Seidel and other minimization algorithms*, Tech. Rep. 68–59, Computer Science Center, University of Maryland, College Park, 1968, 121 pp.

[7] JAMES W. DANIEL, *Convergence of the conjugate gradient method with computationally convenient modifications*, Numer. Math., 10 (1967), pp. 125–131.

[8] JU. I. LJUBIC, *General theorems on quadratic relaxation*, Dokl. Akad. Nauk SSSR, 161 (1965), pp. 1274–1277 = Soviet Math. Dokl., 6 (1965), pp. 588–591.

[9] B. T. POLJAK, *Existence theorems and convergence of minimizing sequences in extremum problems with restrictions*, Dokl. Akad. Nauk SSSR, 166 (1966), pp. 287–290 = Soviet Math. Dokl., 7 (1966), pp. 72–75.

Downloaded 04/19/13 to 129.11.21.2. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php