SIAM REVIEW
Vol. 11, No. 2, April 1969

CONVERGENCE CONDITIONS FOR ASCENT METHODS*

PHILIP WOLFE†

Abstract. Liberal conditions on the steps of a "descent" method for finding extrema of a function are given; most known results are special cases.

1. Introduction. An ascent method for maximizing a real function f generates a sequence $x_0, x_1, \cdots$ of points for which $f(x_{n+1}) \geqq f(x_n)$ in this way: Given x_n, a direction y_n is chosen, if possible, so that $f(x_n + ty_n) > f(x_n)$ for some $t > 0$; a suitable value $t = t_n$ is then selected, and the succeeding point x_{n+1} is defined to be $x_n + t_n y_n$. The direction y_n usually depends on the gradient ∇f of f at x_n, but may also depend on other present or past information about the process.

If f has a maximum at a point x at which the gradient exists, then $\nabla f(x) = 0$. Whether the ascent method helps in finding such an x depends on the choice of y_n. Until that choice is specified, we can only answer questions about the rate of change of f along the ascent path (the polygonal path whose vertices are x_n), which we pose as questions about the directional derivative

$$D_{y_n} f(x_n) = \left[\frac{d}{dt} f(x_n + ty_n)\right]_{t=0+}.$$

The answers we obtain can then be applied to the original problem.

Our principal aim in this note is to develop the weakest conditions we can find on individual step lengths t_n so as to obtain a result of the following kind: If an infinite number of steps satisfy the conditions, then either $f(x_n) \to \infty$ or, for those steps, $D_{y_n} f(x_n) \to 0$.

After some preliminaries in § 2, we set forth the conditions we have found in § 3. They may seem complicated; the reader might well concentrate on (iv) and (v), probably the most useful, which together constitute a loosening of the usual "optimal gradient" step prescription.

The main result on convergence is given in terms of directional derivatives in § 4. Only modest continuity properties on the derivatives are required, and the result holds in any normed linear space. In § 5 we see how this leads to finding stationary points and give a partial result indicating that convergence to a stationary point which is not a relative maximum is unlikely.

Section 6 presents some examples of the vagaries of the path taken by the ascent process. Finally, in § 7 we pose some open questions about ascent methods and comment on their usefulness.

2. Some preliminaries. The segment $[x, x + Ty]$ is the set of all points of the form $z = x + ty$ for $0 \leqq t \leqq T$. We suppose $|y| = 1$, in which case the direc-

* Received by the editors December 26, 1967, and in revised form November 27, 1968.

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

tional derivative of f at z in the direction y is just

$$D_y f(z) = \frac{d}{dt} f(x + ty),$$

the right-hand (left-hand) derivative being taken if $t = 0$ ($t = T$). (The function f is called differentiable at z if $|w|D_{w/|w|}f(z)$ is a continuous linear function of w at 0; then the function is the Fréchet derivative, denoted $\nabla f(z)$. In the finite-dimensional case, of course, the function may be written $w \cdot \nabla f(z)$, the inner product of the vector w and the vector $\nabla f(z)$ of first partial derivatives of f; existence of these derivatives is necessary for the differentiability of f, but not sufficient; but their continuity is sufficient.)

We shall assume that $D_{y_n}f(z)$ exists whenever z is on the ascent path, that is, for $z \in [x_n, x_{n+1}]$ and $y_n = (x_{n+1} - x_n)/|x_{n+1} - x_n|$, and use the abbreviations

$$D_n f = D_{y_n} f, \qquad \bar{D}_n = D_{y_n} f(x_n).$$

The directional derivative must have some kind of continuity for our results. The weakest statement seems to be the following.

DEFINITION. $D_n f$ is said to be *sufficiently continuous* on the subsequence $x_{n'}$ of points of the ascent path if, for any $e > 0$, there exist N and $d > 0$ such that $n' \geqq N$ and $s \leqq d$ imply $|D_{n'}f(x_{n'} + sy_{n'}) - D_{n'}| < e$ (that is, if $D_{n'}f$ is continuous at $x_{n'}$, uniformly in n').

Sufficient continuity is just a specialization of uniform continuity on the ascent path, which is assured if the path lies in a region in which ∇f is uniformly continuous. (Often one makes the stronger assumption that f has bounded second derivatives.) It is thus assured if ∇f is continuous on a compact set containing the ascent path. Lemma 1 shows that it is also assured if the sequence $x_{n'}$ converges to a point of continuity of ∇f (which does not follow from the previous observation, since neighborhoods are not necessarily compact).

LEMMA 1. *If ∇f is continuous at x' and $x_k \to x'$, then for any $e > 0$ there exist $d > 0$ and K so that*

$$|D_k f(z) - D_k f(x_k)| < e \quad \textit{if} \quad |z - x_k| < d \quad \textit{and} \quad k \geqq K.$$

Proof. Choose d so that $|z - x'| < 2d$ implies $|\nabla f(z) - \nabla f(x')| < e$ and K so that, for $k \geqq K$, we have $|x_k - x'| < d$ and $|\nabla f(x') - \nabla f(x_k)| < e$. Then if $|z - x_k| < d$ and $k \geqq K$, we have $|z - x'| \leqq |z - x_k| + |x_k - x'| < 2d$, so that

$$\begin{aligned} |D_k f(z) - D_k f(x_k)| &= |y_k \cdot [\nabla f(z) - \nabla f(x_k)]| \\ &\leqq |y_k| \, |\nabla f(z) - f(x_k')| \\ &\leqq |\nabla f(z) - \nabla f(x')| + |\nabla f(x') - \nabla f(x_k)| < e. \end{aligned}$$

3. The ascent sequence. The ascent sequence is defined by

$$x_{n+1} = x_n + t_n y_n,$$

with $t_n > 0$, $|y_n| = 1$, $\bar{D}_n = D_n f(x_n) \geqq 0$ and $f(x_{n+1}) \geqq f(x_n)$ for all n, with the following exception: the sequence will be terminated if $\bar{D}_n = 0$ or if it is found that $\lim_{t\to\infty} f(x_n + ty_n) = \infty$; but we do not insist on testing for the latter possibility. Define

$$\Delta x_n = x_{n+1} - x_n, \qquad \Delta f_n = f(x_{n+1}) - f(x_n).$$

We think of the nth step of the ascent procedure as characterized by all of these quantities associated with n.

We distinguish five different, but not mutually exclusive nor compatible, properties that a step may have. They are:

(i) $\Delta f_n \geqq \varepsilon$;
(ii) $|\Delta x_n| \geqq \varepsilon \min\{1, \bar{D}_n\}$;
(iii) $\Delta f_n \geqq \varepsilon|\Delta x_n|\bar{D}_n$;
(iv) $D_n f(z_n) \leqq (1 - \varepsilon)\bar{D}_n$ for some $z_n \in [x_n, x_{n+1}]$ such that f is nondecreasing from x_n to z_n;
(v) with z_n as in (iv), $f(x_{n+1}) \geqq f(z_n)$.

The quantity ε is fixed, $0 < \varepsilon < 1$.

The properties will be used in such combinations that, if Δf_n is less than ε, then $\bar{D}_n$ must be small. Property (i) can of course be used alone. If (ii) and (iii) both hold, then we cannot have Δf_n small and $\bar{D}_n$ large. The useful combinations are collected in the following.

DEFINITION. A step will be called *serious* if one of these four sets of conditions is satisfied: (i) alone; (ii) and (iii); (iii) and (iv); or (iv) and (v).

This definition constitutes a considerable weakening of the properties ordinarily required of an ascent sequence. For example, Curry's [1] requirement is that x_{n+1} give the first relative maximum of $f(x_n + ty_n)$ for $t \geqq 0$, that is, the first zero of $D_n f(x_n + ty_n)$, which is subsumed under (iv), (v) if z_n yields the first zero and $z_n = x_{n+1}$. Cauchy's requirement [2], that x_{n+1} yield the absolute maximum of $f(x_n + ty_n)$, is covered by (iv), (v) by taking z_n again as the first relative maximum. Actually, using (iv) with $\varepsilon > 0$ eliminates a nuisance that might arise under the Curry or the Cauchy requirements: it could happen, as with $f(x_n + ty_n) = t/(t - 1)$, that no point satisfying their requirements could be found; yet f would remain bounded. Collectively, properties (ii)–(iv) are similar to those used by Goldstein [3], but a little more general. It should be noted that a different ε might be used in each of (i)–(iv) without changing any of the results below. It would also be possible to use a different ε at each step; for example, if $\Delta f_n \geqq \varepsilon_n$ always held, and $\sum \varepsilon_n$ were divergent, we should have the needed result. Such extensions are not entirely trivial, but we doubt that they would be useful.

A different type of requirement has been used by Goldstein [4] and Ostrowski [5]. Suppose that the Hessian $H(x)$ of $f(x)$ is negative definite in the region of interest, $-K < 0$ being a uniform upper bound for its eigenvalues. Choose s_n so that $s_n \leqq 1 - \varepsilon$ (the ε of (iii), if it is necessary to distinguish) and also so that $s_n \geqq \varepsilon/2K$ (the ε of (ii)). Let $t_n = 2s_nD_n/K$. Property (ii) trivially holds.

By Taylor's theorem there exists z on the segment $[x_n, x_{n+1}]$ such that

$$\begin{aligned}\Delta f_n &= \Delta x_n \cdot \nabla f(x_n) + \tfrac{1}{2}\Delta x_n H(z)\Delta x_n \\ &= t_n\bar{D}_n + \tfrac{1}{2}t_n^2 y_n H(z) y_n \geqq t_n[\bar{D}_n - \tfrac{1}{2}Kt_n] \\ &= t_n\bar{D}_n[1 - s_n] \geqq \varepsilon t_n \bar{D}_n,\end{aligned}$$

so that property (iii) holds, and the step is serious.

A given step cannot be both serious and arbitrarily short, for when $D_n \neq 0$, the properties (i), (ii) and (iv) all fail for sufficiently small t_n. On the other hand, when $D_n \neq 0$ it is always possible to take either of two kinds of serious steps. This is the content of the next lemma.

LEMMA 2. *Let $D_n \neq 0$. (a) If $\varepsilon < 1$, then t_n can be chosen so that the step satisfies either* (i), *or* (iv) *and* (v). (b) *If $\varepsilon \leqq \frac{1}{2}$, then t_n can be chosen so that the step satisfies either* (ii) *and* (iii) *or* (iii) *and* (iv).

Proof. (a) If property (i) fails for all t, then $[f(x_n + ty_n) - f(x_n)]/t \to 0$ as $t \to \infty$; choosing t_n so that the difference quotient is less than $(1 - \varepsilon)\bar{D}_n$, we see that the theorem of the mean yields z_n as required for (iv), and (v) is satisfied with $x_{n+1} = z_n$.

(b) Since $\bar{D}_n$ is the limit of the difference quotient above as $t \to 0+$, property (iii) holds for t_n sufficiently small. If it also holds for $t_n = \varepsilon \min\{1, D_n\}$, then (ii) and (iii) are simultaneously satisfied; if it does not hold for that value of t_n, then by continuity there is some smaller value of t_n for which $\Delta f_n = \varepsilon t_n \bar{D}_n$, that is, for which $\Delta f_n/t_n = \varepsilon \bar{D}_n$, whence by the theorem of the mean there exists z so that $D_n f(z) = \varepsilon \bar{D}_n$. Letting z_n be the closest such z to x_n, we see that (iv) is satisfied, since $\varepsilon\bar{D}_n \leqq (1 - \varepsilon)\bar{D}_n$. Thus (iii) and (iv) both hold.

4. The convergence theorem. The conditions imposed on the steps, and the proof of the main result, can be clarified through considering these two (exhaustive, but not exclusive) cases that might arise.

Case 1. There is a subsequence n' and $d > 0$ such that $|\Delta x_{n'}| \geqq d$. In this case if, for example, f is nondecreasing from $x_{n'}$ to $x_{n'+1}$ for all n' (and this requirement can be considerably weakened provided we do not do something really silly), then our conclusion is obvious: if f does not increase to infinity, the derivatives must tend to zero.

Case 2. $|\Delta x_{n'}| \to 0$ for some subsequence n'. In this the burden falls squarely on the method for choosing $|\Delta x_n|$, which had better be such that $D_{y_{n'}} f(x_{n'}) \to 0$. We have essentially only one way for assuring that, expressed as property (iv) in § 3: there must be a definite decrease ("percentagewise") in the rate of change of f as we proceed along the segment joining x_n to x_{n+1}. If that is true, and the segments shrink to zero, then the rate of change must tend to zero.

LEMMA 3. *If* (iii) *holds for a subsequence n' of steps for which $|\Delta x_{n'}| \geqq d > 0$, then either $f(x_n) \to \infty$ or $\bar{D}_{n'} \to 0$.*

LEMMA 4. *If* (iv) *holds for a subsequence n' of steps for which $|z_{n'} - x_{n'}| \to 0$ and $D_n f$ is sufficiently continuous on the subsequence, then $\bar{D}_{n'} \to 0$.*

Proof. (Lemma 3 is obvious.) By property (iv), $D_{n'}f(z_{n'}) - \bar{D}_{n'} \leqq -\varepsilon\bar{D}_{n'}$. Since the left-hand side of the inequality tends to zero, the right-hand side, being nonpositive, does too.

THEOREM 1. *Let the ascent sequence have a subsequence n' of serious steps. If $D_n f$ is sufficiently continuous on the subsequence, then either $f(x_n) \to \infty$ or $\bar{D}_{n'} \to 0$.*

Proof. We shall show that from any subsequence n'' of n' we can extract a further subsequence n''' such that either $f(x_{n'''}) \to \infty$ or $\bar{D}_{n'''} \to 0$. This will prove the theorem since, if it were false, $f(x_{n''})$ would be bounded for any subsequence n'', which could be chosen so that $\bar{D}_{n''} \geqq \varepsilon > 0$.

If (i) holds for infinitely many steps of n'', the conclusion is obvious.

If (ii) and (iii) hold for infinitely many steps of n'', then let n''' consist of those steps. If f remains bounded, then $\Delta f(x_n) \to 0$, so that

$$\Delta f_{n'''} \geqq \varepsilon|\Delta x_{n'''}|\bar{D}_{n'''} \geqq \varepsilon^2 \min\{D_{n'''}, D_{n'''}^2\}$$

yields $D_{n'''} \to 0$.

If (iii) and (iv) hold for infinitely many steps of n'', then we may either choose n''' so that $|\Delta x_{n'''}| \geqq d > 0$ for some d, in which case Lemma 3 gives the conclusion, or we may choose it so that $|\Delta x_{n'''}| \to 0$, in which case Lemma 4 gives the conclusion.

Finally, assume that (iv) and (v) hold for infinitely many n''. If there exists a subsequence n''' for which $\bar{D}_{n'''} \to 0$, then we are through at once; and if such exists for which $|z_{n'''} - x_{n'''}| \to 0$, then we are through by Lemma 4. Otherwise there exists $\varepsilon > 0$ such that $|z_{n''} - x_{n''}| > \varepsilon$ and $\bar{D}_{n''} > 2\varepsilon$ for all sufficiently large n''. The sufficient continuity of $D_n f$ implies that $D_{n''}f(z) \geqq \bar{D}_{n''} - |D_{n''}f(z) - \bar{D}_{n''}| \geqq 2\varepsilon - \varepsilon = \varepsilon$ for all z on the segment $[x_{n''}, x_{n''+1}]$ such that $|z - x_{n''}| < d$ for suitable $d > 0$. Using (v) and the monotonicity of f between $x_{n''}$ and $z_{n''}$, we have

$$\Delta f_{n''} \geqq f(z_{n''}) - f(x_{n''}) \geqq \int_0^d D_{n''}(x_{n''} + sy_{n''})\,ds \geqq \varepsilon d.$$

It follows that $f(x_{n''}) \to \infty$.

5. Convergence to stationary points. Theorem 1 establishes that, unless $f(x_n) \to \infty$ on the ascent sequence, the directional derivatives at the serious steps tend to zero. It remains to prescribe the choice of the directions y_n so that we can draw conclusions about the stationarity of f.

The only general prescription we shall use is that y_n must be chosen so that its angle with $\nabla f(x_n)$ is uniformly bounded away from 90°, that is, so that for some $\delta > 0$,

$$y_n \cdot \nabla f(x_n) \geqq \delta|\nabla f(x_n)|.$$

Then, since the left-hand side is just $\bar{D}_n$, from $\bar{D}_n \to 0$ we can conclude $\nabla f(x_n) \to 0$. Calling a serious step having the directional property above a *very serious* step, and recalling that uniform continuity of f implies the sufficient continuity of $D_n f$, we have established the main theorem on convergence.

THEOREM 2. *Let the ascent sequence have a subsequence n' of very serious steps and let ∇f be uniformly continuous on the ascent path. Then either $f(x_n) \to \infty$ or $\nabla f(x_{n'}) \to 0$.*

Lemma 1 and the above remarks yield Theorem 2 immediately.

THEOREM 3. *If x is a limit point of very serious steps of the ascent sequence, then $\nabla f(x) = 0$.*

Zoutendijk [6] has used a more liberal prescription for y_n and has shown that any limit point of an ascent sequence for which t_n satisfies Curry's condition (cf. § 3) and for which $\sum \cos \angle(y_n, \nabla f(x_n))$ diverges must be a stationary point; he uses this result to prove the convergence of the conjugation procedure of Fletcher and Reeves [7]. As mentioned before, we consider here only conditions which may be imposed uniformly on single steps, not those which depend on the whole sequence.

Theorem 3 says that if the ascent sequence converges, it converges to a stationary point of f; it says nothing about whether that point gives, as desired, a relative maximum. As a partial answer to that question, we offer Theorem 4 below, which asserts that if f is quadratic and the direction of each step is that of the gradient and the starting point x_0 is taken in general position, convergence to other than a relative maximum cannot occur. The discussion for nonquadratic f seems hard to carry out; but, to the degree that f can be approximated by a quadratic function in the neighborhood of a limit point of the ascent sequence, we feel 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.

THEOREM 4. *Let $f(x) = px + xQx$, and let M be the subspace of all eigenvectors of the symmetric matrix Q associated with negative eigenvalues. Let the ascent sequence $x_0, x_1, \cdots$, with $y_n = \nabla f(x_n)/|\nabla f(x_n)|$, have infinitely many serious steps. Then the sequence converges if and only if $\nabla f(x_0) \in M$; and if it does not converge, then $f(x_n) \to \infty$.*

Proof. Let a complete orthonormal set of eigenvectors of Q be chosen: $v_1, \cdots$, spanning M, associated with the negative eigenvalues $\lambda_1, \cdots$, and $w_1, \cdots$, associated with the nonnegative eigenvalues $\mu_1, \cdots$. Note that

$$\nabla f(x_{n+1}) = p + 2Qx_{n+1} = p + 2Qx_n + 2t_nQy_n$$

$$= \nabla f(x_n) + 2t_nQy_n = \left(I + \frac{2t_n}{|\nabla f(x_n)|}Q\right)\nabla f(x_n).$$

Assume $\nabla f(x_0) \in M$. Since $QM \subseteq M$, it follows from the above formula that $\nabla f(x_n) \in M$ for all n, so that $x_n - x_0 \in M$ for all n. Writing $x_n - x_0 = \sum_j s_j^n v_j$ and $\nabla f(x_0) = p + 2Qx_0 = \sum_j r_j v_j$, we have

$$\begin{aligned} f(x_n) - f(x_0) &= px_n - px_0 + x_nQx_n - x_0Qx_0 \\ &= (p + 2Qx_0)(x_n - x_0) + (x_n - x_0)Q(x_n - x_0) \end{aligned}$$

$$= \sum_j [r_j s_j^n + \lambda_j s_j^{n2}]$$

$$= \sum_j \left[\lambda_j\left(\frac{s_j^n + r_j}{2\lambda_j}\right)^2 - \frac{r_j^2}{4\lambda_j}\right]$$

$$\leqq -\tfrac{1}{4}\sum_j \frac{r_j^2}{\lambda_j}.$$

Thus $f(x_n)$ is bounded above. Since each serious step is very serious, Theorem 2 yields $\nabla f(x_n) \to 0$. Since

$$\nabla f(x_n) = \nabla f(x_0) + 2Q(x_n - x_0) = \nabla f(x_0) + \sum_j s_j^n \lambda_j v_j,$$

each sequence s_j^n converges, so that the sequence x_n converges.

If $\nabla f(x_0) \notin M$, then $w_j \cdot \nabla f(x_0) \neq 0$ for some w_j. Further, $w_j \cdot \nabla f(x_{n+1}) = (1 + 2t_n\lambda_j/|\nabla f(x_n)|)w_j \cdot \nabla f(x_n)$. Since the quantity in parentheses is at least one, $\nabla f(x_n)$ cannot tend to zero; Theorem 2 then yields $f(x_n) \to \infty$.

6. Some ascent paths. Even for the most familiar ascent method, and the most precisely defined, in which y_n has the direction of $\nabla f(x_n)$ and x_{n+1} is chosen as yielding the first relative maximum of f in that direction, not much can be said to characterize the ascent path or the set of limit points of the ascent sequence. The examples below will show the variety of behavior that can arise. All our examples are two-dimensional, and the two coordinates will be called x, y. Such examples have a peculiarity when treated by the above ascent method: since x_{n+1} is obtained through the relation

$$0 = D_n f(x_{n+1}) = \nabla f(x_{n+1}) \cdot \nabla f(x_n)/|\nabla f(x_n)|,$$

the successive directions are always orthogonal, and the ascent method reduces to maximizing alternately in each of two fixed orthogonal directions; by rotating the coordinate system appropriately, it could be done by changing just one variable at a time.

EXAMPLE 1 (The usual example). $f(x, y) = -(ax^2 + by^2)$, with $a, b > 0$. Unless one of the starting coordinates is zero, the point (x, y) leads to the point (c^2x, c^2y) after two iterations, where c is fixed and not greater than $|a - b|/(a + b)$.

EXAMPLE 2. $f(x, y) = y - \frac{1}{2}x^2$. The successor of the point (x, y) is $(-x^{-1}, 1 + y + x^{-2})$, and, in particular, the nth successor of $(1, 0)$ is $((-1)^n, 2n)$. Since $\nabla f(x, y) = (-x, 1)$, f tends to infinity and ∇f does not tend to zero.

EXAMPLE 3. $f(x, y) = [\frac{1}{2}x^2 - y]^{-1}$. Since, in general, $\nabla g(f(x, y)) = g'(f(x, y)) \cdot \nabla f(x, y)$, this example has the same ascent path as Example 2; but both f and ∇f tend to zero.

EXAMPLE 4. $f(x, y) = \log(y - \frac{1}{2}x^2)$. This example has again the same ascent path as Example 2, but f tends to infinity, while ∇f tends to zero.

EXAMPLE 5. $f(x, y) = -x^3[1 + (y - \cos(1/x))^2]$.

$$\frac{\partial f}{\partial x} = -3x^2\left[1 + \left(y - \cos\frac{1}{x}\right)^2\right] + 2x\left(y - \cos\frac{1}{x}\right)\sin\frac{1}{x};$$

$$\frac{\partial f}{\partial y} = -2x^3\left(y - \cos\frac{1}{x}\right).$$

We suppose the starting point is chosen so that the initial gradient is horizontal, that is, so as to lie on the curve $C = \{(x, y) : y = \cos(1/x)\}$. Then any even-numbered point of the ascent sequence lies on C, and the next point lies to its left. ∇f is continuous, and vanishes only when $x = 0$.

For each $k = 1, 2, \cdots$ draw the vertical ray R_k lying in the line $x_k = 1/(k\pi)$ and having the endpoint $1/(k\pi) \pm (1 - [(k/(k - 1))^3 - 1])^{1/2}$. For k even "+" is used, and the ray points downward; for k odd, both are reversed.

Let $(x_k, y) \in R_k$ and $(x, y) \in C$, with $x_k < x < x_{k-1}$. It is easy to check that $f(x_k, y) < f(x, y)$, so that if (x, y) is an even-numbered point of the ascent sequence, its successor lies between (x, y) and R_k. On the other hand, the successor of any odd-numbered point lies on a vertical line through that point. Thus the ascent path does not cross any R_k.

Since $\nabla f \to 0$, $x \to 0$ on the ascent sequence. In order to avoid $\bigcup_k R_k$, there must be points of the sequence assuming y-coordinates arbitrarily close to ± 1. Thus the vertical segment from $(0, -1)$ to $(0, 1)$ is the set of all limit points of the ascent path.

Note that in Example 5, except for the endpoints, we have not found just what points of the interval are limit points of the ascent *sequence*, although it seems likely that the entire interval consists of limit points. Some further examples we have not been able to analyze completely can be indicated. Using polar coordinates, we see that the function

$$f(r, \theta) = -(r - 1)^3[2 + \cos(\theta - 1/(r - 1))]$$

should behave like Example 5, the spiral $\theta - 1/(r - 1) = -\pi/2$ playing a role like that of the curve C, and the circle $r = 1$ constituting the set of all limit points. (This function seems to fit the example outlined by Curry [1].) Finally, by drawing suitable (convex) contours, one can convince himself that an ascent path may have precisely two or precisely four limit points and that the function f may be concave; but we do not have any formulas in support of this.

7. Problems and warning. The examples of the previous section indicate that the ascent path can have many different kinds of behavior. A fair open question thus seems to be, What are some nontrivial properties of an ascent path? Some more precise questions can be asked, the first two arising from the incomplete remarks of the previous section: (i) Are the limit points connected with the function $f(r, \theta)$ as stated? (ii) Can you exhibit a concave or quasi-concave function whose ascent sequence has exactly two or four limit points? (iii) Is there any restriction on the cardinality of the set of limit points of an ascent sequence? (iv) If f is concave, and $\sup_x f(x) = M$, and x_n is the ordinary steepest ascent sequence, does $f(x_n) \to M$? (This problem is due to Oettli [8]. Of course, if the sequence is bounded, the answer is "yes," and also if there are only two dimensions.)

The number of open questions concerning steepest ascent, its ubiquity in the literature of optimization, and its continued use in computation might lead the unwary to think that it was a good thing to do in practice; but we think that in the art of computation it should be considered as a last resort, *faute d. mieux*, as Cauchy [2] might have said. An estimate of the rate of convergence was first given by Kantorovich [9] who showed that ordinary steepest ascent, applied to a quadratic form having maximum and minimum eigenvalues M and m, respectively, satisfies the relationship $f(x_{n+1}) \leqq f(x_n)(M - m)^2/(M + m)^2$. If, as is overwhelmingly often the case, m is only a few per cent of M, this bound is quite discouraging. What is worse, Akaike [10] has shown, loosely speaking, that the inequality of the above relationship can be replaced by approximate equality: the worst almost always happens. There are, on the other hand, methods such as that of Fletcher and Powell [11] which are ascent methods, but make far more clever choices of the ascent direction, and these are quite useful in practice. The general results of the present paper apply to some of these, establishing their convergence under certain circumstances; but in the present state of our knowledge of them, we can only judge them by empirical studies like those of Box [12], which we recommend to those who require numerical answers.

Ascent methods are of course used in solving constrained optimization problems as well as the problem studied here, although the modifications needed for such problems considerably narrow the conditions under which convergence can be demonstrated. Topkis and Veinott [13] have given conditions which apply to a variety of problems, and we have shown elsewhere [14] that the most obvious (and popular) modification of the method of steepest ascent to handle constraints is not generally convergent.

REFERENCES

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

[2] AUGUSTINE CAUCHY, *Méthode générale pour la résolution des systèmes d'équations simultanées*, C.R. Acad. Sci., 25 (1847), pp. 536–538.

[3] A. A. GOLDSTEIN, *Minimizing functionals on normed-linear spaces*, J. SIAM Control, 4 (1966), pp. 81–89.

[4] ———, *Cauchy's method of minimization*, Numer. Math., 4 (1962), pp. 146–150.

[5] ALEXANDER OSTROWSKI, *Solution of Equations and Systems of Equations*, Academic Press, New York, 1966.

[6] G. ZOUTEHDIJK, Private communication, 1967.

[7] R. FLETCHER AND C. M. REEVES, *Function minimization by conjugate gradients*, Comput. J., 7 (1964), pp. 149–154.

[8] W. OETTLI, Private communication, 1967.

[9] L. V. KANTOROVICH AND G. P. AKILOV, *Functional Analysis on Normed Spaces*, Macmillan, New York, 1964, Chap. 15.

[10] HIROTUGU AKAIKE, *On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method*, Ann. Inst. Statist. Math., 11 (1959), pp. 1–16.

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

[12] M. J. BOX, *A comparison of several current optimization methods, and the use of transformations in constrained problems*, Ibid., 9 (1966), pp. 67–77.

[13] DONALD M. TOPKIS AND ARTHUR F. VEINOTT, *On the convergence of some feasible direction algorithms for nonlinear programming*, J. SIAM Control, 5 (1967), pp. 268–274.

[14] PHILIP WOLFE, *On the convergence of gradient methods under constraints*, Rep. RZ-204, IBM Watson Research Center, Yorktown Heights, New York, 1966.