SIAM J. CONTROL
Vol. 8, No. 2, May 1970

A NONGRADIENT AND PARALLEL ALGORITHM FOR UNCONSTRAINED MINIMIZATION*

D. CHAZAN AND W. L. MIRANKER†

Abstract. The purpose of this paper is to describe an algorithm for unconstrained optimization which is suitable for execution on a parallel computer. A nongradient method similar in nature to Powell's method is used and it is shown that the algorithm terminates at the minimum for quadratics and converges for strictly convex twice continuously differentiable functions.

1. Introduction. In this paper we consider an algorithm suitable for unconstrained minimization of strictly convex functions. The method is a nongradient method since it requires no information about derivatives. In addition, it has been designed as a parallel method and may be executed simultaneously on a set of arithmetic processors. Parallelism has recently attracted some attention in various fields of computation and computer science, e.g., [4], [5], [6]. This paper gives a procedure for utilizing parallelism in the area of unconstrained minimization.

Algorithms for unconstrained minimization usually proceed by a sequence of univariate minimizations. The directions in which to make a univariate search typically depend on the gradient at the current point. Steepest descent [1], conjugate gradient [1], and variable metric [2] are examples of such methods. Since gradient computation is costly in certain computational situations, methods which do not compute gradients have been devised. Apart from those methods which estimate gradients by finite differences, such nongradient methods include pattern search [1] and a conjugate method devised by Powell [3]. This latter method has been the motivation for the study presented here.

We will now give a brief description of Powell's minimization method which requires no derivative computation. In m-space we are given a point p and m directions $v^1, \cdots, v^m$. Starting at p we make m univariate minimizations in sequence in the directions $v^1, \cdots, v^m$, respectively. This procedure produces a polygonal trajectory which terminates at a point, q, say. We now make one more univariate minimization starting at q and in the direction $v^{m+1} = q - p$. Let r be the point which this last minimization results in. This is one cycle in the algorithm. To execute the next cycle, we update p and $v^1, \cdots, v^m$ as follows: $r \to p$ and $v^{i+1} \to v^i$, $i = 1, \cdots, m$.

Figure 1.1 illustrates two cycles of the algorithm in 3-space. Arrows denote univariate minimizations and are labeled by their directions.

Powell tried out his algorithm and obtained convergence in several examples. However, he did not produce a proof of convergence for his algorithm. He does produce a discussion of the method when it is applied to a function $f(x)$ of the form

$$f(x) = xAx + bx + c, \tag{1.1}$$

where c is a scalar, b is an m-vector, and A an $m \times m$ positive, definite, symmetric

* Received by the editors September 24, 1968, and in revised form October 10, 1969.

† Thomas J. Watson Research Center, International Business Machines Corporation, Yorktown Heights, New York 10598.

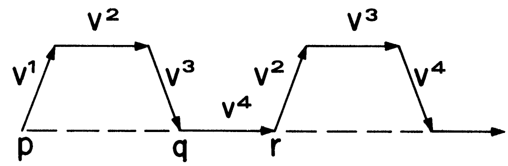

FIG. 1.1

matrix. In this discussion Powell claimed that his algorithm terminates in at most m cycles at the minimum of $f(x)$.

Except for a small gap, this discussion is a proof of convergence in this case. Subsequently, W. Zangwill [6] modified Powell's algorithm and filled the gap in Powell's argument for quadratic functions. Moreover, Zangwill produced a convergence proof for this modification in the general case when $f(x)$ is a strictly convex function. Zangwill's algorithm is identical to Powell's, except that each cycle of minimizations is augmented by an additional minimization in a coordinate direction. The coordinate directions are chosen in cyclical order.

The methods of Powell and of Zangwill, as well as most other minimization methods, proceed by means of a sequence of univariate minimizations. Thus they are, of necessity, sequential computations. If we have at our disposal a computer with a set of arithmetic processors capable of simultaneous operation, and if we seek to exploit this computer to improve the speed of execution of a typical sequential minimization algorithm, we will, in general, be unable to do so. The method discussed in this paper proceeds by simultaneous univariate minimization, with simultaneity of degree as high as the dimension of the problem, and so is appropriate for exploitation of the type of parallel computer in question.

We will now sketch our parallel algorithm for unconstrained minimizations.

First we choose a set U of m linearly independent unit vectors $u_1, \cdots, u_m$. At each cycle of the algorithm we will select a vector from this set in cyclical order. We are given a point p and $m - 1$ vectors $v^1, \cdots, v^{m-1}$. We select a vector from U and call it v^m. Starting at p, construct a polygonal path by stringing together the m vectors $v^1, \cdots, v^m$ (no minimizations). Then from each vertex of this polygonal path (excluding its initial point p), perform m simultaneous (in parallel) univariate minimizations in the common direction v^1. These minimizations produce a displacement $\alpha_i v^1$, $i = 1, \cdots, m$, from the vertex $p + \sum_{j \leqq i} v^j$, $i = 1, \cdots, m$, of the polygonal path. Now update p and $v^1, \cdots, v^{m-1}$ as follows. Let $p + v^1 + \alpha_1 v^1$ be the updated p. Let $v^{i+1} + (\alpha_{i+1} - \alpha_i)v^1$ be the updated v^i, $i = 1, \cdots, m - 1$. This completes one cycle of the algorithm.

Figure 1.2 illustrates one cycle in the algorithm. Line segments with arrowheads denote minimizations. The updated p and the updated v^i are given an asterisk. In the schematic, v^3 is chosen as u_j so that $(v^3)^*$ must be u_{j+1}. As above, parallel and nonparallel lines may appear nonparallel or parallel, respectively, in the schematic. For example, the broken line from p to p^* given by $p + v^1 + \alpha_1 v^1$ is, in fact, a straight-line segment.

It is seen that our algorithm uses features found in Powell's method as well as a feature resembling the modification introduced by Zangwill. However, it is quite different from either of these methods and has as its principal objective to increase the speed of a calculation by executing simultaneous or parallel

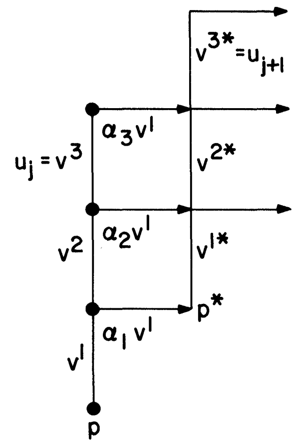

FIG. 1.2

minimizations. In the case that $f(x)$ is a quadratic in m-space our algorithm is finitely convergent and employs m^2 minimizations to locate the minimum. In the cases that they are finitely convergent, Powell's method requires m^2 minimizations also while Zangwill's requires $(m + 1)^2$. (It is a simple matter to modify all of these methods to cut these numbers in half.) However, the latter two methods are purely sequential. If we assume that a univariate functional minimization takes one unit of time, the latter two methods (suitably modified) require approximately m^2 time steps. The method in this paper is a parallel method and requires only m time steps to find the minimum of a quadratic. Thus there is a gain in speed of computation which is linear in the number of processors used for minimizing a quadratic function.

We have not produced an estimate of the increase in speed of calculation for our algorithm for the nonquadratic case. This remains an open question.

In § 2 we define the parallel minimization algorithm in m-space. We then consider the case of quadratic functions and show that the algorithm converges in finitely many steps. In § 3 we give an outline of the proof of convergence of the algorithm for the class of locally convex functions. In §§ 4, 5 and 6 the technical details of the proof are furnished.

Remark on parallel operation. It is possible that in one cycle of m simultaneous univariate minimizations one minimization is much more difficult than the $m - 1$ others. This is a cost parallelism which in practice may be reduced in various ways.

2. Description of algorithm and the case of quadratic functions. Let $f(x)$ denote the function to be minimized. Then the parallel nongradient minimization algorithm is given formally as follows.

Let $U = \{u_i, i = 1, \cdots, m\}$, where the u_i are linearly independent m-vectors.

Let $\beta_r, r = 1, \cdots,$ be a sequence of positive scalars tending to zero. Let $w_n^m = u_i$ if $n \equiv i \bmod m, i = 1, \cdots, m$.

Let $p_n^1, n = 1, \cdots,$ be a sequence of points in m-space, and let v_{n+j}^j, $j = 1, \cdots, m; n = 1, \cdots,$ be a sequence of m-vectors. For each $n, n = 1, \cdots,$ a step in the algorithm is given by a mapping

$$(p_n^1, v_{n+1}^1, v_{n+2}^2, \cdots, v_{n+m}^m) \to (p_{n+1}^1, v_{n+2}^1, v_{n+m}^2, \cdots, v_{n+m+1}^m) \tag{2.1}$$

defined as follows.

Determine the scalars α_{n+1}^j, $j = 1, \cdots, m$, by performing (simultaneously) the m univariate minimizations

$$\min_{\alpha_{n+1}^j} f\left(p_n^1 + \sum_{i \leq j} v_{n+i}^i + \alpha_{n+1}^j v_{n+1}^1\right), \qquad j = 1, \cdots, m. \tag{2.2}$$

Then the mapping (2.1) is defined by

$$\begin{aligned} p_{n+1}^1 &= p_n^1 + (1 + \alpha_{n+1}^1)v_{n+1}^1, \\ v_{n+j}^j &= (\alpha_n^{j+1} - \alpha_n^j)v_n^1 + v_{n+j}^{j+1}, \qquad j = 1, \cdots, m-1, \\ v_{n+m}^m &= \beta_{n+m} w_n^m. \end{aligned} \tag{2.3}$$

The algorithm is schematized in Fig. 2.1.

We may note that if v_n^i turns out to be zero v_n^1 would be zero and our algorithm (as defined) would simply imply that $p_{n+1}^1 = p_n^1$ and $v_{n+1}^j = v_n^j$.

If $f(x)$ is a quadratic given by (1.1) with A positive definite, it is easy to see that the algorithm converges in at most m steps. Consider to this end the following two lemmas found in [3].

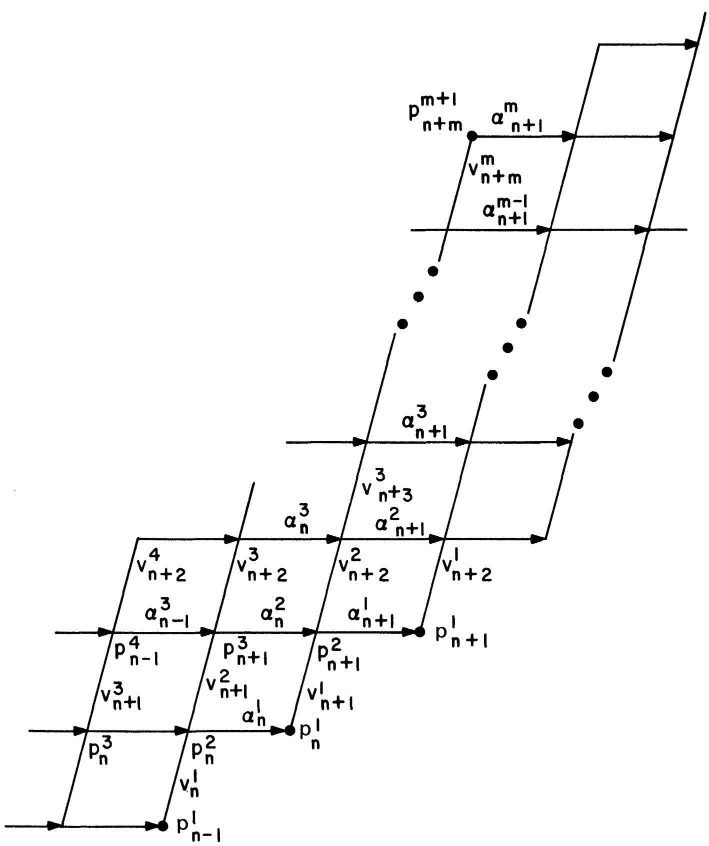

FIG. 2.1

LEMMA 2.1. *If $q^1, \cdots, q^k$, $k \leqq m$, are mutually conjugate directions, then the minimum of the quadratic function $f(x)$, where x is a general point in the k-dimensional space containing the directions $q^1, \cdots, q^k$, may be found by searching along each of the directions once only and, moreover, in any order.*

LEMMA 2.2. *If y and z are the locations of minima of $f(x)$ in a space containing the direction q, then the vector $y - z$ is the conjugate to q.*

Referring to Fig. 2.1, we see that the vector v_{n+1}^1 is conjugate to v_{n-j}^1, $j = 0, \cdots, m-2$. This follows from Lemma 2.2 since v_{n+1}^1 is determined by two points which are locations of minima of $f(x)$ in a space containing v_{n-j}^1, $j = 0, \cdots, m-2$. Thus any sequence of m vectors v_{r+j}^1, $j = 1, \cdots, m$, $r \geqq 0$, are mutually conjugate. Lemma 2.1 then implies that every point p_n^1, $n \geqq m$, lies at the minimum of $f(x)$.

Notice that the proof is identical to Powell's; but, whereas it is insufficient for his own algorithm, it is complete for ours.

3. Outline of the proof of convergence. We have seen that the parallel minimization scheme defined in the previous section locates the minimum of $f(x)$ in a finite number of steps if $f(x)$ is quadratic. The purpose of this and the following sections will be to demonstrate convergence of the sequence of iterates to the minimum of f when f is twice continuously differentiable, strictly convex, and tends to ∞ as $\|x\| \to \infty$. It is not at all clear that the strict convexity is essential for convergence. Indeed, a close relative of this method, the method of coordinate search, does not require such an assumption. This difficulty stems from the fact that at some point in the procedure we have to guarantee that the search directions keep spanning the whole space. This is always true for the method of coordinate search. To prove this fact for our scheme, we shall use the strict convexity of $f(x)$ to "localize" a finite sequence of steps around some point and use linearizations of the function near this point to obtain the spanning property.

Since the proof, while elementary, is composed of a large number of disjoint parts, we shall present here a short summary to serve as a guideline.

Let us start by noting that the sequence of points p_n^1 is a descending sequence; i.e., $f(p_{n+1}) \leqq f(p_n)$. Then proceeding by contradiction, we shall assume that $f(p_n^1) > a_1 > \min_p f(p)$. Thus there exists some value a_2 so that $f(p_n^1) \downarrow a_2$ and the sequence p_n^1 is bounded between the two contour surfaces $f(p) = a_2$ and $f(p) = a_2 + \varepsilon_n$ with $\varepsilon_n \to 0$. Clearly, the accumulation points of p_n^1 lie on $\{p : f(p) = a_2\}$. Let p_0 be one such point, and let g_0 be the gradient at p_0. Also, let g_n be the gradient at p_n^1.

The main part of the proof will consist of showing that there exist a positive integer j_0 and a positive scalar α so that for each m-vector v,

$$\max_{1 \leqq j \leqq j_0} \frac{\langle v, v_{n+j}^1 \rangle}{\|v\| \, \|v_{n+j}^1\|} > \alpha.$$

From this it will follow that any sequence of $j_0 \geqq m$ successive v_n^1 (search directions) spans the space. We combine this with the fact that

$$\frac{\langle v_n^1, g_n \rangle}{\|v_n^1\| \, \|g_n\|} = 0$$

and that g_{n+j}, $j = 1, \cdots, j_0$, can be made arbitrarily close to g_0 by continuity of the gradient to conclude that $\langle v_{n+j}^1, g_0 \rangle / (\|v_n^1\| \, \|g_0\|)$, $j = 1, \cdots, j_0$, has zero as a limit point. This contradicts the fact just cited with g_0 taking the place of v, that the normalized projection of any vector onto v_n^1 is greater than $\alpha > 0$.

As noted above, the critical part of the argument presented is the spanning property of the vectors v_{n+j}^1, $1 \leqq j \leqq j_0$, for some j_0. If $f(\cdot)$ were a positive definite quadratic form, the vectors v_{n+j}^1, $1 \leqq j \leqq m$, would be pairwise conjugate and would, therefore, certainly have the spanning property. In general, this is not the case. However, suppose we could localize a block of Fig. 2.1, i.e., a collection of points $\{p_{n+j}^i : 1 \leqq j \leqq j_0, 1 \leqq i \leqq m\}$, inside a neighborhood N of the accumulation point p_0. We could then approximate $f(\cdot)$ by a quadratic consisting of the first three terms in its Taylor series about p_0 and obtain approximate conjugacy statements.

Conjugacy here of $v_{n+j_1}^1$ and $v_{n+j_2}^1$ is taken to mean $\langle v_{n+j_1}^1, f_{xx}(p_0)v_{n+j_2}^1\rangle = 0$, where $f_{xx}(p_0)$ is the Hessian matrix of f at p_0. Thus, to conclude the proof, two facts have to be verified:

(a) A "block" of the p_j^i can be put into an arbitrarily small neighborhood of p_0;

(b) If the p_j^i are sufficiently close to p_0, the search directions v_{n+j}^1, $1 \leqq j \leqq m$, are approximately conjugate.

Statement (b) as it stands turns out not to be true. There may be freak circumstances where $v_{n+j_1}^1$ and $v_{n+j_1+j_2}^1$, $1 \leqq j_2 < m$, are not even approximately conjugate. It is, however, always true that v_n^1 is (approximately) conjugate to v_{n+1}^j, $1 \leqq j \leqq m$. For $j = 1$ this last property asserts that successive search directions are (approximately) conjugate. This, together with the fact that the v_n^m have the spanning property, is enough to induce the spanning property for v_{n+j}^1, $1 \leqq j \leqq 2m - 1$.

The proof of statement (a) requires the following assertions:

(a1) The sequence $\|v_n^i\| \to 0$ as $n \to \infty$, $1 \leqq i \leqq m$;

(a2) $\|p_{n+1}^1 - p_n^1\| \to 0$ as $n \to \infty$.

From (a2) and the fact that p_0 is a limit point of p_n^1, we are able to force the whole sequence p_{n+j}^1, $1 \leqq j \leqq j_0$, for any fixed j_0, into an arbitrarily small neighborhood of p_0 infinitely often. Combining this with (a1), we are assured that the p_{n+j}^i, $1 \leqq i \leqq m$, $1 \leqq j \leqq j_0$, lies in N also.

The next section will be devoted to the statement of an elementary theorem which will allow us to obtain (a1). This uses the fact that the $v_n^m \to 0$, which in turn follows from the fact that the $\beta_n \to 0$. This theorem will also be used in § 6 to obtain the approximate conjugacy of v_n^1 to the v_{n+1}^j, $1 \leqq j \leqq m$.

In § 5 we shall prove (a2) using the strict convexity of f and the fact that p_n is in the strip $\{p : a \leqq f(p) \leqq a + \varepsilon_n\}$.

In § 6 we shall combine all these facts together to obtain the convergence statement.

4. Proof of $\|v_n^i\| \to 0$ as $n \to \infty$. As noted above, the argument demonstrating convergence depends on the approximate conjugacy of successive minimization directions. We shall now state a lemma which will formalize the notion of approximate conjugacy, and apply it to deduce that $v_n^i \to 0$ (assertion (a1)).

We recall that the function $f(x)$ is twice continuously differentiable, strictly convex and tends to infinity as $\|x\| \to \infty$.

LEMMA 4.1. *Let $f(x)$ be a twice differentiable and strictly convex function of the m-vector x and let $f(x) \to \infty$ as $\|x\| \to \infty$. Let v be a given m-vector and let S be the subspace orthogonal to v. Let $g: S \to R^1$ be defined as follows:*

$$f(w + g(w)v) \leqq f(w + kv)$$

for all real k. Then

(a) *$g(w)$ is a well-defined differentiable function;*

(b) $\langle v, f_{ww}(w + g(w)v\Delta w\rangle + \langle v, f_{ww}(w + g(w)v)v\Delta g\rangle + h(\Delta w) = 0$, *where* $|h(\Delta w)| = o(\Delta w)$.

Furthermore, if S' is the surface, $S' = \{vg(w) + w : w \in R^m\}$ and q_1, q_2 are two points on S' which have the form $w_1 + g(w_1)v$ and $w_1 + g(w_2)v$, respectively, then

$$\text{(c)} \qquad \left| \frac{\langle v, f_{ww}(q_1)(q_1 - q_2)\rangle}{\|v\| \, \|w_1 - w_2\|} \right| \leqq O(\|w_1 - w_2\|).$$

We may note that this also implies that

$$\langle v, f_w(q_1)(q_1 - q_2)\rangle \leqq O(q_1 - q_2)\|q_1 - q_2\| \, \|v\|$$

since

$$\|q_1 - q_2\|^2 = \|w_1 - w_2\|^2 + \|v\|^2 g^2(w) = \|w_1 - w_2\|^2.$$

Proof. $f(w + kv)$ is strictly convex in k and tends to ∞ as $k \to \infty$. Thus it has a unique, well-defined minimum at $g(w)$. Let $G(w) = f_w(w)$. Then $g(w)$ satisfies the equation $\langle v, G(w + g(w)v)\rangle = 0$. Since $G_w(w + kv)$ is positive definite, it follows from the implicit function theorem that $g(w)$ is differentiable. This demonstrates (a). Now

$$0 = \delta\langle v, G(w + g(w)v)\rangle = \langle v, H(w + g(w)v) \cdot (\delta w + g_w(w)\delta wv)\rangle + h(\Delta w),$$

where $H(w) = G_w$ is the Hessian matrix of f and $\langle v, G(v + g(w)v)\rangle$ was expanded in a Taylor series. This demonstrates (b). Dividing through by $\|v\| \, \|w_1 - w_2\|$ demonstrates (c).

COROLLARY. *Let $\{q_n^1, q_n^2\}$ be two sequences of points in R^m, which are contained in a compact set. Let $q_n^1 - q_n^2 \to 0$ as $n \to \infty$. Then $q_n^1 + vg(q_n^1) - (q_n^2 + vg(q_n^2)) \to 0$.*

Proof. The statement follows immediately from the uniform continuity of g on a compact set.

We will now use this corollary to show that the network of Fig. 2.1 collapses in its vertical direction as it evolves, so that the whole forms a narrowing tube.

Let $q_k^1 = p_k^m$ and $q_k^2 = p_k^m + v_{k+1}^m$ (see Fig. 4.1). Since $\sum_{j=1}^{\infty} \beta_j < \infty$, the p_k^m do indeed lie in a compact set. Then the corollary allows us to conclude immediately that $v_{k+1}^{m-1} \to 0$ as $k \to \infty$. By induction, it follows similarly that $v_{k+1}^j \to 0$, $j = 1, \cdots, m$, as $k \to \infty$. In this way we obtain m sequences, all of which converge to zero. Thus for every ε, there exist $k_1, \cdots, k_m$ so that $\|v_k^i\| < \varepsilon$ for $k \geqq k_i$. Hence $\|v_k^i\| \leqq \varepsilon$ for all $k \geqq \max_{1 \leqq i \leqq m} k_i$.

5. Proof of $\|p_{n+1}^1 - p_n^1\| \to 0$ as $n \to \infty$. Let us note now that the sequence of points p_j^1 has the property:

$$f(p_{j+1}^1) \leqq f(p_j^1), \tag{5.1}$$

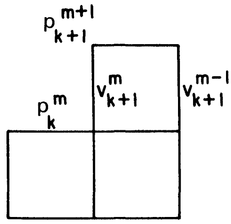

FIG. 4.1

which follows from the fact that p_{j+1}^1 is the point where the one-dimensional minimum of f in the direction v_{j+1}^1 starting from p_{j+1}^2 is obtained, and that the line $p_{j+1}^2 + \alpha v_{j+1}^1$ contains p_j^1. In particular, $f(p_j^1) \leqq f(p_1^1)$.

Let us also remark that $\{p : f(p) \leqq a\}$ is compact since $f(p) \to \infty$ as $\|p\| \to \infty$, and that the sequence p_j^1 must, therefore, lie in a compact set. We wish to show that any accumulation point p of the sequence p_j^1 has the property $f(p) \leqq f(q)$ for all points $q \in R^m$.

In this section we shall show that $\|p_{j+1}^1 - p_j^1\| \to 0$. Thus if p is an accumulation point of p_j^1 then $p_j^1, j_1 \leqq j \leqq j_1 + j_0$, can be made to stay within ε of p for any fixed j_0 and for infinitely many values of j_1. Consider now the following lemma.

LEMMA 5.1. *Let q_n, w_n be a sequence of points satisfying the following relation*:

$$a \leqq f(q_n + \alpha w_n) \leqq a + \varepsilon_n$$

with $\varepsilon_n \to 0, 0 \leqq \alpha \leqq 1$. If f is continuous, strictly convex, and $f(p) \to \infty$ as $\|p\| \to \infty$, then $\|w_n\| \to 0$.

Proof. Suppose otherwise. Then there exists a subsequence of $\|w_n\|$ with $\|w_{n_i}\| \geqq c > 0$. Since the set $S = \{q : f(q) \leqq a + \varepsilon_1\}$ is closed and bounded, it is compact. Similarly, the set $\{v : v = q_1 - q_2 ; q_1, q_2 \in S\}$ is also compact. Hence q_n, w_n lie in compact sets and there exists a subsequence m_i of n_i with $(q_{n_i}, w_{n_i}) \to (q, w)$. By continuity, $f(q_{n_i} + \alpha w_{m_i}) \to f(q + \alpha w)$. Clearly

$$a \leqq f(q + \alpha w) \leqq a + \varepsilon_n$$

for all n and

$$\|w\| \geqq c.$$

But then f is constant along w which contradicts the strict convexity of f.

Using this lemma, we obtain the stated conclusion, $\|p_{j+1}^1 - p_j^1\| \to 0$, by letting $q_j = p_j^1$ and $w_j = p_{j+1}^1 - p_j^1$.

6. Proof of convergence.

From the result of the last two sections, we may conclude that for every $\varepsilon > 0$ and $j_0 > 0$, there exists a $j \geqq 0$ so that $\|p_{j+k}^i - p\| < \varepsilon$, $0 \leqq k \leqq j_0, 1 \leqq i \leqq m$. This statement may be viewed geometrically as a collapsing of the array of Fig. 2.1 into an ε-neighborhood of f for durations of j_0 steps at a time. Of course, p is an accumulation point of the sequence $\{p_j^n\}$. We now wish to use these facts to show that for ε sufficiently small, it follows that for infinitely many j,

$$\max_{1 \leqq k \leqq 2m-1} \left| \left\langle f_{pp}(p)w, \frac{v_{j+k}^1}{\|v_{j+k}^1\|} \right\rangle \right| > \gamma$$

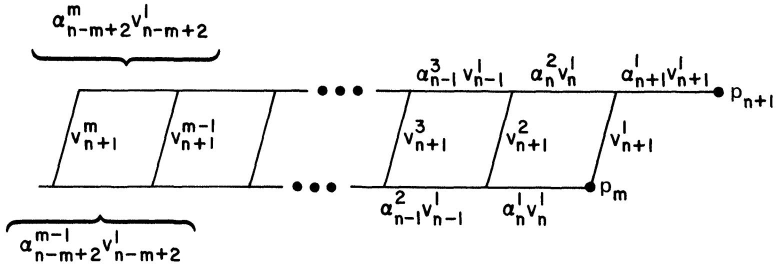

FIG. 6.1

for some $\gamma > 0$ and any w with $\|w\| = 1$. We shall accomplish this by noting that if $2m - 1$ successive search directions stay close to some subspace of R^m, then m successive v_j^m will also stay close to that subspace. This contradicts the fact that any m successive v_j^m are linearly independent and cannot collapse into a subspace.

Let $H = f_{pp}(p)$, where p is a limit point of the sequence $\{p_j^1\}$, and let $\|v\|_H^2 = \langle Hv, v\rangle$ for any symmetric H.

LEMMA 6.1. *There exists a positive γ so that for an ε sufficiently small, whenever $\|p_{j+k}^i - p\| < \varepsilon$, $1 \leqq k \leqq 2m - 1$, $1 \leqq i \leqq m$, we have*

$$\max_{1 \leqq k \leqq 2m-1} \left| \left\langle Hw, \frac{v_{j+k}^1}{\|v_{j+k}^1\|} \right\rangle \right| > \gamma$$

for any w.

Thus any $2m + 1$ successive search directions have the spanning property.

Proof. Suppose otherwise. Then for any ε, γ there exist j, w so that $\|p_{j+k}^i - p\| \leqq \varepsilon$, $1 \leqq k \leqq 2m - 1$, $1 \leqq i \leqq m$, but

$$\left| \left\langle Hw, \frac{v_{j+k}^1}{\|v_{j+k}^1\|} \right\rangle \right| \leqq \gamma, \qquad 1 \leqq k \leqq 2m - 1.$$

Let $z_i = (\alpha_{j+k+i}^{m-i} - \alpha_{j+k+i}^{m-i+1}) v_{j+k+i}^1$, $i = 1, \cdots, m - 1$, $z_m = v_{j+k+m}^1$. Then, by viewing Fig. 6.1, it is easily seen that

$$v_{j+k+m}^m = \sum_{i=1}^m z_i.$$

Since the z_i are all of the form $c \cdot v_{j+k}^1$ for some real c and some integer k, we know that

$$\left| \left\langle Hw, \frac{z_i}{\|z_i\|} \right\rangle \right| \leqq \gamma.$$

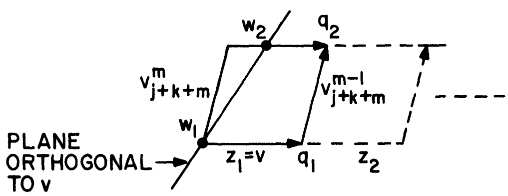

FIG. 6.2

We would like to conclude from this that

$$\left|\left\langle Hw, \frac{v^m_{j+k+m}}{\|v^m_{j+k+m}\|}\right\rangle\right|$$

is small for m successive values of k to contradict the independence of m successive v^m_{j+k+m}. To do so it is enough to show that $\|v^m_{j+k+m}\| \geqq c\|z_i\|$ for some constant c independent of j, ε. In this case,

$$\begin{aligned}\left|\left\langle Hw, \frac{v^m_{j+k+m}}{\|v^m_{j+k+m}\|}\right\rangle\right| &\leqq \left|\sum \left\langle Hw, \frac{z_i}{\|v^m_{j+k+m}\|}\right\rangle\right| \\ &\leqq \left|\sum \left\langle Hw, \frac{z_i}{c\|z_i\|}\right\rangle\right| \\ &\leqq \sum \frac{\gamma}{c} = \frac{m\gamma}{c}\end{aligned}$$

and the desired contradiction would follow.

We shall now show that v^m_{j+k+m} satisfies

$$\|v^m_{j+k+m}\|^2 \geqq c\|z_i\|^2(1 - O(\varepsilon))/\|H^{-1}\|\,\|H\|.$$

Indeed since

$$\|v^m_{j+k+m}\|^2 \geqq \|v^m_{j+k+m}\|^2_H/\|H\|$$

it suffices to show

$$\|v^m_{j+k+m}\|^2_H = \|z_1 + \sum_{i=2}^m z_i\|^2_H \geqq (1 - O(\varepsilon))\|z_i\|^2_H,$$

since $\|z_i\|^2_H \geqq \|z_i\|^2/\|H^{-1}\|$. Referring to Figs. 6.1 and 6.2, we have $\sum_{i=2}^m z_i = v^{m-1}_{j+k+m}$. Let $v = z_1$, let P be the orthogonal projection onto the plane orthogonal to v, let $w_1 = p^m_{j+k+m}$, $w_2 = w_1 + P(v^m_{j+k+m})$ and $q_1 = p^{m-2+i}_{j+k+m-2+i}$, $i = 1, 2$. Then $q_1 - q_2 = v^{m-1}_{j+k+m}$. Since v^{m-1}_{j+k+m} and v^m_{j+k+m} differ by a multiple of v, $p(v^{m-1}_{j+k+m}) = p(v^m_{j+k+m})$. Applying Lemma 4.1:

$$\begin{aligned}\langle z_1, f_{pp}(q_1)v^{m-1}_{j+k+m}\rangle &= O(w_1 - w_2)\|z_1\|\,\|w_1 - w_2\| \\ &= O(P(v^{m-1}_{j+k+m}))\|z_1\|_H\|Pv^{m-1}_{j+k+m}\|\,\|H^{-1}\| \\ &\leqq O(v^{m-1}_{j+k+m})\|z_1\|_H\|v^{m-1}_{j+k+m}\|_H.\end{aligned}$$

It follows that

$$\begin{aligned}\|v^m_{j+k+m}\|^2_H &= \left\|z_1 + \sum_{i=2}^m z_i\right\|^2_H = \|z_1 + v^{m-1}_{j+k+m}\|^2_H \\ &= \|z_1\|^2_H + \left\|\sum_{i=2}^m z_i\right\|^2_H + 2\langle z_1, Hv^{m-1}_{j+k+m}\rangle \\ &\geqq \max\left(\|z_1\|^2_H, \left\|\sum_{i=2}^m z_i\right\|^2_H\right) \\ &\quad -2\langle z_1, (f_{pp}(p^{m-1}_{j+k+m-1}) + O(\varepsilon))v^{m-1}_{j+k+m}\rangle\end{aligned}$$

$$\geqq \max\left(\|z_1\|_H^2, \left\|\sum_{i=2}^{m} z_i\right\|_H^2\right) - O(\varepsilon)\|z_1\|_H \left\|\sum_{i=2}^{m} z_i\right\|_H$$

$$\geqq \max\left(\|z_1\|_H^2, \left\|\sum_{i=2}^{m} z_i\right\|_H^2\right)(1 - O(\varepsilon)).$$

Continuing by induction by applying the same argument to $\sum_{i=k}^{m} z_i$, we obtain the desired result:

$$\|v_{j+k+m}^{m}\|_H^2 \geqq \max \|z_i\|_H^2(1 - O(\varepsilon)).$$

With the help of this lemma and the outline of the proof given in § 3, we obtain our final result.

THEOREM 6.1. *The sequence* p_j^1 *defined in* § 2 *converges iff* f *is twice continuously differentiable and strictly convex; i.e.,* $f_{pp} \geqq 0$ *and* $f(x) \to \infty$ *as* $x \to \infty$.

REFERENCES

[1] D. J. WILDE, *Optimum Seeking Methods*, Prentice-Hall, Englewood Cliffs, New Jersey, 1964.

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

[3] M. J. D. POWELL, *An efficient method for finding the minimum of a function of several variables without calculating derivatives*, Ibid., 7 (1964), pp. 155–162.

[4] D. CHAZAN AND W. L. MIRANKER, *Chaotic relaxation*, Linear Algebra and Its Applications, 2 (1969), pp. 199–222.

[5] R. M. KARP AND R. E. MILLER, *Parallel program schemata*, J. Comput. Systems Sci., 3 (1969), pp. 147–195.

[6] W. I. ZANGWILL, *Minimizing a function without calculating derivatives*, Comput. J., 10 (1967), pp. 293–296.