

Generation of Conjugate Directions for Unconstrained Minimization without Derivatives*

By Larry Nazareth

Abstract. We analyze a technique for unconstrained minimization without derivatives. This stems from two theorems proved by M. J. D. Powell. A particular version, which we consider in detail, is related to the Jacobi process for finding the eigensystem of a symmetric matrix, and the two processes, although different, help to illuminate one another. We study convergence of the search directions to mutual conjugacy, cases when cycling occurs and identify a broad class of 'cyclic patterns' for which convergence to mutual conjugacy is proven.

1. Introduction. Given a method for unconstrained minimization that requires knowledge of derivatives, it is usually possible to derive a derivative-free method by suitably approximating derivatives by differences. However, many methods have been proposed which are not of this type; and of these, the method of Powell [1] seems to be the most widely used. This maintains a set of n search directions (where n is the dimension of the problem), and these are suitably updated so as to converge to a mutually conjugate set in a bounded number of steps. The method thus has the advantage of quadratic termination but suffers two drawbacks. Firstly, there is a danger of linear dependence in the search directions even for a quadratic; and precautions taken to ensure that the directions span the complete space of variables often adversely affect efficiency. Zangwill [2] and Brent [3] have both proposed techniques to overcome this. Secondly, generation of conjugate directions is tied to accuracy of the line searches, which must, in theory, be exact.

Brodie's method [4] gets around both these difficulties. The search directions used are always orthogonal and are updated as the algorithm progresses, so that they converge for a quadratic to mutual conjugacy. This updating process is not dependent upon accuracy of line searches. Brodie proves that the algorithm, applied to the quadratic $\psi(x)$ with Hessian A , *exactly* parallels a cyclic Jacobi process in the sense that successive approximations to the set of eigenvectors of A are identical. Unlike Powell's method, Brodie's algorithm does not possess a quadratic termination property, though numerical evidence has shown it to be satisfactory in practice. Seeking the set of eigenvectors of A as Brodie's algorithm does would also seem unnecessarily costly, since the directions they define are unique when eigenvalues of A are distinct.

Here we discuss some theoretical aspects of an algorithm that relaxes the requirement that a unique set of directions be generated but shares the advantages of Brodie's

Received July 9, 1974; revised January 27, 1975.

AMS (MOS) subject classifications (1970). Primary 65K05, 90C30.

*The material presented here is taken from Part II of the author's Ph. D dissertation written under the supervision of Dr. B. N. Parlett. The research was supported by the Office of Naval Research Contract N00014-A-0200-1017.

Copyright © 1976, American Mathematical Society

method. It is based upon a technique for generating conjugate directions developed by Powell [5]. The two theorems that form the basis of this technique are stated in Section 2, and we develop a particular algorithm that bears some similarity to the cyclic Jacobi process. This is a straightforward generalization of the two-variable case considered by Powell. In subsequent sections we study certain central issues which pertain to this technique, namely, convergence of the search directions to mutual conjugacy and cases when cycling occurs; finally we develop a class of 'cyclic patterns' for which convergence to mutual conjugacy is proven.

2. The Algorithm.

2.1. *Powell's Theorems.* We use the notation ΔD to mean the absolute value of the determinant of the matrix D .

THEOREM 2.1 (POWELL, 1964). *Given a quadratic function $\psi(x) = a + b^T x + \frac{1}{2} x^T A x$ where A is positive definite and symmetric, let (d_1, \dots, d_n) be any set of n directions satisfying the normalization conditions*

$$(2.1a) \quad d_i^T A d_i = 1, \quad i = 1, 2, \dots, n,$$

i.e., d_i are defined to be of unit length in the A -norm. If D is the matrix whose columns are the directions (d_1, \dots, d_n) , then the maximum value of ΔD is attained if and only if the directions d_i ($i = 1, \dots, n$) are mutually conjugate. \square

THEOREM 2.2 (POWELL, 1972). *Let (d_1, \dots, d_n) be any set of n directions normalized to satisfy (2.1a). Let D be the matrix whose columns are the directions (d_1, \dots, d_n) , and let Ω be any orthogonal matrix.*

Let the columns of the matrix D given by $\bar{D} = D\Omega$, define a new set of directions $(\bar{d}_1, \dots, \bar{d}_n)$.

Normalize each of the directions \bar{d}_i so that each is of unit length in the A -norm, thus obtaining directions d_i^ and matrix $D^* = (d_1^*, \dots, d_n^*)$, where $d_i^* = \bar{d}_i / (\bar{d}_i^T A \bar{d}_i)^{1/2}$. Then*

$$(2.1b) \quad \Delta D^* \geq \Delta D$$

where Δ is defined as above. \square

For proofs see Powell [5].

2.2. *Normalization.* Suppose $(\bar{d}_1, \dots, \bar{d}_n)$ are a set of unnormalized search directions at the current iterate x^c . In order to satisfy the normalization condition (2.1a), each direction \bar{d}_i must be divided by $(\bar{d}_i^T A \bar{d}_i)^{1/2}$. For a quadratic $\psi(x)$ this may be obtained by estimating the value of the second derivative of $\psi(x)$ at x^c in the direction \bar{d}_i , since:

$$\psi(x^c + \lambda \bar{d}_i) = \psi(x^c) + \lambda(b + A x^c)^T \bar{d}_i + (\lambda^2/2) \bar{d}_i^T A \bar{d}_i$$

and

$$\psi(x^c - \lambda \bar{d}_i) = \psi(x^c) - \lambda(b + A x^c)^T \bar{d}_i + (\lambda^2/2) \bar{d}_i^T A \bar{d}_i.$$

Thus

$$\bar{d}_i^T A \bar{d}_i = \frac{\psi(x^c + \lambda \bar{d}_i) - 2\psi(x^c) + \psi(x^c - \lambda \bar{d}_i)}{\lambda^2}.$$

Similar results hold approximately for a general smooth function $\phi(x)$ with A replaced by $A(x^c)$ and $0 < \lambda \ll 1$.

2.3. *The Resulting Algorithm and Questions to be Discussed.* The algorithm derived by Powell from these two theorems is as follows: A set of n search directions is maintained. At any iteration a search is conducted in sequence along each direction of this set, and the current estimate of the minimum improved in some way. It is an easy matter to estimate second derivatives along each direction. Thus, each direction may be normalized and the set of search directions revised by post-multiplying by a suitably chosen orthogonal transformation. This completes an iteration of the algorithm.

We shall denote the unnormalized search directions at the start of iteration k by the columns of $\bar{D}^{(k)}$, the *normalized* search directions by columns of $D^{(k)}$ and the orthogonal transformation used by $\Omega^{(k)}$. In general, knowledge of the off-diagonal elements of $D^{(k)T}AD^{(k)}$ is not explicitly available without further work. When the above algorithm is applied to a quadratic function $\psi(x)$, convergence of $D^{(k)}$ to mutual conjugacy is not assured. The optimal choice of $\Omega^{(k)}$ are the eigenvectors of $D^{(k)T}AD^{(k)}$. These are expensive to obtain. Thus one must restrict attention to a well chosen class of orthogonal transformations from which $\Omega^{(k)}$ is selected. Certain questions then arise quite naturally. Provided that no search direction is neglected, will convergence of the search directions to mutual conjugacy for a quadratic always occur? If not, can cases be exhibited for which one obtains nonconvergence or cycling? Can convergence always be assured by judiciously choosing, at each iteration, an orthogonal transformation from the class? What is the ultimate rate of convergence? Settling these questions is crucial to understanding the algorithm's behavior, particularly in the neighborhood of a minimum, where it will usually be well approximated by a quadratic.

2.4. *The Algorithm Arising From Use of Plane Rotations.* Let the orthogonal transformation $\Omega^{(k)}$ be selected from the class of plane rotations. This is a worthwhile context within which to investigate some of the above questions in the light of Brodlie's algorithm, and because this will clarify the use of more general classes of orthogonal transformations.

Suppose we are dealing with a quadratic function $\psi(x)$ and are given at a typical step k , a set of search directions $d_1^{(k)}, \dots, d_n^{(k)}$ normalized to be of unit length in the A -norm.

Now, revising $D^{(k)}$ by means of a plane rotation say in the (p, q) plane through angle θ , gives unnormalized directions $\bar{D}^{(k+1)}$ satisfying

$$\begin{aligned} \bar{d}_p^{(k+1)} &= d_p^{(k)} \cos \theta + d_q^{(k)} \sin \theta, \\ \bar{d}_q^{(k+1)} &= -d_p^{(k)} \sin \theta + d_q^{(k)} \cos \theta, \\ \bar{d}_r^{(k+1)} &= d_r^{(k)} \quad \text{for all } r \neq p \text{ or } q. \end{aligned} \tag{2.4a}$$

The best θ to choose is the angle that makes $d_p^{(k+1)}$ and $d_q^{(k+1)}$ conjugate. Thus

$$\tan 2\theta = 2d_p^{(k)}Ad_q^{(k)} / (d_p^{(k)}Ad_p^{(k)} - d_q^{(k)}Ad_q^{(k)}). \tag{2.4b}$$

Since $d_p^{(k)}$ and $d_q^{(k)}$ are normalized, the denominator $\neq 0$. Thus $\theta = \pm \pi/4$ unless $d_p^{(k)}Ad_q^{(k)} = 0$, in which case any θ will do.

We note that knowledge of the magnitude of $d_p^{(k)} A d_q^{(k)}$ is not needed.

Thus, revising $d_p^{(k)}$ and $d_q^{(k)}$ corresponds to post-multiplication by a member of the two element set of fixed orthogonal matrices $\{U_{pq}, U_{pq}^T\}$ where

$$(2.4c) \quad \begin{bmatrix} u_{pp} & u_{pq} \\ u_{qp} & u_{qq} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix},$$

$$u_{ii} = 1 \quad \forall i \neq p \text{ or } q,$$

$$u_{ij} = 0 \quad \text{otherwise.}$$

Thus, the class of orthogonal transformations used is given by

$$(2.4d) \quad S = \{U_{pq}, U_{pq}^T : \forall (p, q) \text{ s.t. } 1 \leq p < q \leq n\}.$$

Drawing an analogy with the classical Jacobi eigenvalue process would suggest that, at each iteration, one choose the pair (p, q) to be revised that corresponds to the maximum off-diagonal element of $D^{(k)T} A D^{(k)}$. From a practical standpoint, however, this is not feasible since off-diagonal elements of $D^{(k)T} A D^{(k)}$ are not available without much additional work. Henceforth, as in the cyclic Jacobi process we consider the *Cyclic Selection Rule*. This selects successive members of a *cyclic pattern*, i.e., a permutation of the $n(n-1)/2$ pairs $(1, 2), (1, 3), \dots, (1, n), (2, 3), \dots, (2, n), \dots, (n-1, n)$. A sweep of $n(n-1)/2$ iterations completes a cycle, and a fresh cycle is then started. In this case, knowledge of off-diagonal elements of $D^{(k)T} A D^{(k)}$ is not required for choosing (p, q) or $\Omega^{(k)}$, in contrast to Brodli's algorithm.

2.4.1. The algorithm outlined in 2.3 then specializes to Algorithm C below, whose k th iteration, initiated with a set of directions $(\bar{d}_1^{(k)}, \dots, \bar{d}_n^{(k)})$, is as follows:

Algorithm C. (i) Choose a pair (p, q) (called the *current pair*) according to some cyclic pattern.

(ii) Conduct a linear search, not necessarily exact, in sequence along the p th and q th search directions and improve the estimate of the minimum. Normalize these two directions by estimating the second derivative along each direction during the search (as discussed in 2.2) thus obtaining directions $d_p^{(k)}$ and $d_q^{(k)}$. All other directions remain unaltered.

(iii) Post-multiply the matrix whose columns consist of the set of search directions by $\Omega^{(k)} = U_{pq}$ or U_{pq}^T . This gives directions

$$(2.4e) \quad \bar{d}_p^{(k+1)} = \frac{1}{\sqrt{2}} (d_p^{(k)} \mp d_q^{(k)}), \quad \bar{d}_q^{(k+1)} = \frac{1}{\sqrt{2}} (\pm d_p^{(k)} + d_q^{(k)}),$$

and all other directions remain unaltered. Then start iteration $(k+1)$. \square

The algorithm is initiated with a set of linearly independent directions $\bar{d}_1^{(1)}, \dots, \bar{d}_n^{(1)}$ and terminates using, at step (ii), some suitable criterion, based upon change in current estimate of the minimum value and change in function value during a complete cycle. See e.g. Brodli [4].

2.5. *Basic Relations.* Consider Algorithm C applied to a quadratic $\psi(x)$. Let $D^{(1)}$ be the initial normalized set of search directions and $D^{(k)}$ the normalized directions at the start of iteration k . These are not explicitly maintained by Algorithm C but are

introduced for the purpose of analysis. Let us postulate also a diagonal normalizing matrix $N^{(k)}$, again not explicitly maintained, which restores the search directions at the end of iteration k to unit length in the A -norm.

Then

$$(2.4f) \quad D^{(k)} = D^{(1)}\Omega^{(1)}N^{(1)}\Omega^{(2)}N^{(2)} \dots \Omega^{(k-1)}N^{(k-1)}$$

and

$$(2.4g) \quad A^{(k)} = D^{(k)T}AD^{(k)} = N^{(k-1)}\Omega^{(k-1)T} \dots N^{(1)}\Omega^{(1)T}D^{(1)T}AD^{(1)}\Omega^{(1)}N^{(1)} \dots \Omega^{(k-1)}N^{(k-1)}.$$

We want $A^{(k)}$ in (2.4g) to converge to the unit matrix I , whence $D^{(k)}$ must tend, as $k \rightarrow \infty$, to a matrix whose columns are mutually conjugate. Henceforth, when we say that the set of search directions converge to mutual conjugacy we mean that $A^{(k)} \rightarrow I$. Note that this does *not* imply that the search directions converge to a *fixed* set of mutually conjugate directions.

After carrying out the k th iteration of Algorithm C with (p, q) the current pair for a quadratic $\psi(x) = a + b^T x + \frac{1}{2}x^T Ax$, where $A > 0$

$$(2.4h) \quad d_p^{(k+1)} = \frac{d_p^{(k)} \pm d_q^{(k)}}{\|d_p^{(k)} \pm d_q^{(k)}\|_A} \quad \text{and} \quad d_q^{(k+1)} = \frac{\pm d_p^{(k)} + d_q^{(k)}}{\|\pm d_p^{(k)} + d_q^{(k)}\|_A},$$

where both upper or both lower signs are taken together. Also,

$$d_r^{(k+1)} = d_r^{(k)} \quad \text{for all } r \neq p \text{ or } q.$$

Given any such r , we have

$$(2.4i) \quad \begin{aligned} |d_p^{(k+1)}Ad_r^{(k+1)}| &= (d_p^{(k)}Ad_r^{(k)} \pm d_q^{(k)}Ad_r^{(k)})/(\|d_p^{(k)} \pm d_q^{(k)}\|_A), \\ |d_q^{(k+1)}Ad_r^{(k+1)}| &= (\pm d_p^{(k)}Ad_r^{(k)} + d_q^{(k)}Ad_r^{(k)})/(\|\pm d_p^{(k)} + d_q^{(k)}\|_A). \end{aligned}$$

Definition 1. We call $|d_i^{(k)}Ad_j^{(k)}|$ the *weight* of pair (i, j) at iteration k . Note that in defining the weight we are using the *normalized* search directions.

Definition 2. If (p, q) is the current pair at the k th iteration, then for any $r \neq p$ or q , we say that (p, r) and (q, r) are *linked* during the k th iteration.

3. Convergence Using an Arbitrary Cyclic Pattern.

THEOREM 3.1. Consider Algorithm C applied to a quadratic $\psi(x)$ using an arbitrary cyclic pattern. Then there always exists a sequence $\Omega^{(1)}, \Omega^{(2)}, \dots, \Omega^{(k)}, \dots$, chosen from S (with $\Omega^{(k)} \in \{U_{pq}, U_{pq}^T\}$ if (p, q) is the current pair) s.t. the search directions converge to mutual conjugacy.

Two lemmas stated below are necessary. Proofs of the lemmas and of Theorem 3.1 may be found in Nazareth [6] and, being quite straightforward, are not included here. Note that Theorem 3.1 is nonconstructive since it does not enable one to know beforehand whether a particular policy for choosing $\Omega^{(k)}$ will succeed.

LEMMA 3.1. Let (p, q) be the current pair at iteration k . Suppose $|d_p^{(k)}Ad_r^{(k)}| \geq \mu$ for some pair (p, r) , $r \neq q$.

Then at least one of the two choices, $\Omega^{(k)} = U_{pq}$ and $\Omega^{(k)} = U_{pq}^T$ must give directions that satisfy $|d_p^{(k+1)}Ad_r^{(k+1)}| \geq \mu/2$.

COROLLARY. *If the weight on pair (p, r) exceeds μ and the pair (p, q) is revised, then the weight on at least one of the pairs (p, r) and (q, r) must exceed $\mu/2$ after the revision, for $\Omega^{(k)} \in \{U_{pq}, U_{pq}^T\}$.*

LEMMA 3.2. *Suppose at iteration t the current pair is (p, r) , $\Omega^{(t)} \in \{U_{pr}, U_{pr}^T\}$ and $|d_p^{(t)} A d_r^{(t)}| = \gamma$, where $0 \leq \gamma < 1$. Then*

$$(3.1a) \quad \Delta D^{(t+1)} = \Delta D^{(t)} / (1 - \gamma^2)^{1/2},$$

where $\Delta D^{(t)}$ denotes the absolute value of the determinant of $D^{(t)}$.

4. Cycling in Algorithm C. A misguided policy for choosing $\Omega^{(k)}$ can lead to cycling of the elements of $D^{(k)T} A D^{(k)}$. We exhibit an example of this related to one published by Hansen [7] for the Jacobi eigenvalue process. Cycling in Algorithm C has certain distinctive features not shared by the Jacobi process.

4.1. Example of Cycling. Given a quadratic function in four variables, let us seek its minimum using Algorithm C, with cyclic pattern

$$(4.1a) \quad (2, 3), (1, 4), (1, 3), (2, 4), (1, 2), (3, 4).$$

Suppose that the initial normalized search directions $d_1^{(1)}, \dots, d_4^{(1)}$ satisfy

$$A^{(1)} = D^{(1)T} A D^{(1)} = \begin{bmatrix} 1 & 0 & x & 0 \\ 0 & 1 & 0 & x \\ x & 0 & 1 & 0 \\ 0 & x & 0 & 1 \end{bmatrix}$$

with $(2, 3)$ the current pair.

Use the following sequence of orthogonal transformations chosen from the set S :

$$(4.1b) \quad U_{23} U_{14} U_{13} U_{24} U_{12} U_{34} U_{23}^T U_{14} U_{13} U_{24}^T U_{12}^T U_{34}^T.$$

These are then repeated in sweeps of 12 iterations.

We find that after six iterations

$$(4.1c) \quad A^{(7)} = D^{(7)T} A D^{(7)} = \begin{bmatrix} 1 & 0 & x & 0 \\ 0 & 1 & 0 & -x \\ x & 0 & 1 & 0 \\ 0 & -x & 0 & 1 \end{bmatrix},$$

and after 12 iterations $D^{(13)T} A D^{(13)} = D^{(1)T} A D^{(1)}$.

Thus, the search directions generated by the sequence (4.1b) do not converge to mutual conjugacy.

Note, however, that $D^{(1)}$ and $D^{(13)}$ may be distinct. It is possible that the search directions will also cycle.

If directions that are already conjugate need not necessarily be revised, examples of cycling may be constructed for any cyclic pattern containing the subsequence

$$(4.1d) \quad (i, j), (l, m), (j, l), (i, m), (i, l), (j, m).$$

Initial normalized search directions are chosen such that $d_j^{(1)}Ad_i^{(1)} = d_i^{(1)}Ad_m^{(1)} = x$, with all other pairs of directions mutually conjugate.

When the current pair is not in the above subsequence, then use $\Omega^{(k)} = I$. Such pairs remain conjugate throughout. Orthogonal transformations for pairs that are in (4.1d) are chosen analogously to the sequence of transformations (4.1b).

The above example is somewhat unsatisfactory in that search directions $d_p^{(k)}$ and $d_q^{(k)}$ that are already conjugate are nevertheless revised. This may be justified on the grounds that the information needed to recognize this is not available to Algorithm C without further function evaluations. In Nazareth [6] somewhat more satisfactory examples are discussed for which the current pair always has nonzero weight. We also discuss the implication of these examples for a threshold policy analogous to that used in the cyclic Jacobi process.

The proofs of Theorem 3.2 and the examples of 4.1 require the use of both positive and negative plane rotations through $\pi/4$ as equally valid choices for Ω^k . Both would be rendered invalid by changing the policy to allow only positive (or only negative) rotations through $\pi/4$. However, even for the policy $\Omega^{(k)} = U_{pq}$, an example exists for which there is no improvement in overall conjugacy during a *single* complete cycle. We refer the reader back to 4.1, where we see that $\Delta D^{(7)} = \Delta D^{(1)}$.

We may easily construct an example which uses a cyclic pattern containing a subsequence of the form

$$(4.1e) \quad (i, j), (l, m), (l, j), (i, m)$$

and for which the policy $\Omega^{(k)} = U_{pq}$ given an arbitrarily small increase in $\Delta D^{(k)}$ during one complete cycle of iterations and no current pair is mutually conjugate during this cycle. This means that most of the weight associated with off-diagonal elements of $D^{(k)T}AD^{(k)}$ is pushed round ahead of the current pair during this cycle.

It is for reasons given in this section that proofs of convergence of $D^{(k)}$ to mutual conjugacy, using an *arbitrary* cyclic pattern, are difficult to obtain.

5. Convergence Proofs for a Restricted Class of Cyclic Patterns. We now show that convergence of the search directions to mutual conjugacy can be proven for all cyclic patterns within a certain class P .

The motivation behind choosing this class P is to exclude patterns containing a subsequence of the form (4.1e). Note, however, that there exist cyclic patterns which do not contain a subsequence (4.1e) and are also not in P .

5.1. *Definition of the Class of Cyclic Patterns P.*

5.1.1. The class P of cyclic patterns is defined recursively using the following procedure. This is called with input consisting of a set of directions G and returns with a cyclic pattern in L . After the formal definition a detailed illustration is given.

Procedure P(G, L). Local Variables H, T .

Step A. Given a set of directions G , partition them into two groups, G_1 and G_2 . If G contains only a single member, *return*.

Step B. Form a list of pairs as follows: Either

- (i) Pick one member of the first group G_1 and pair it with every member of the

second group G_2 taken in any order. Repeat until all members of G_1 are exhausted; or

(ii) Carry out (i) with G_1 and G_2 interchanged.

Step C. Call $P(G_1, H)$ i.e., carry out a recursive call to Procedure P with the set of directions G_1 (in place of G). The list of pairs formed is returned from this recursive call in H .

Step D. Call $P(G_2, T)$ i.e., call P recursively with the set G_2 . The list from this call is returned in T .

Step E. Place the list of pairs H obtained in Step C at the head of the list formed in Step B. Similarly, place the list T at the tail of the list formed in Step B. *Return* with the complete list in L . \square

Any cyclic pattern returned by the above procedure when called with directions d_1, d_2, \dots, d_n is contained in the class of cyclic patterns P . An example of a cyclic pattern obtained is given in Figure 5.1, using seven directions. If all pairs of directions in this example are written as in Figure 5.2, then the cyclic pattern defined in the example is given by taking each pair of Figure 5.2 in the order determined by the number associated with it.

5.2. Proof of Convergence. We have not been able to avoid a certain degree of technical complexity in this proof. Therefore, to help guide the reader, its principal features are first outlined.

The proof is by contradiction. Suppose, when Algorithm C is applied to a quadratic $\psi(x) = a + b^T x + \frac{1}{2} x^T A x$, $A > 0$, the search directions $D^{(k)}$ do not converge to mutual conjugacy for some cyclic pattern $\in P$. Since, from Powell's second theorem (cf. 2.1) $D^{(k)}$ is a monotonically nondecreasing bounded sequence, it must tend to a limit as $k \rightarrow \infty$.

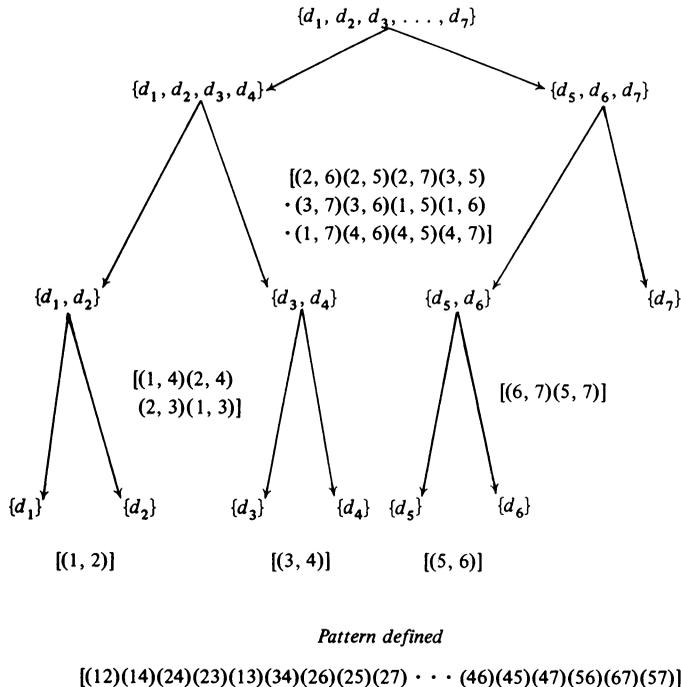


FIGURE 5.1. Example Illustrating Procedure P

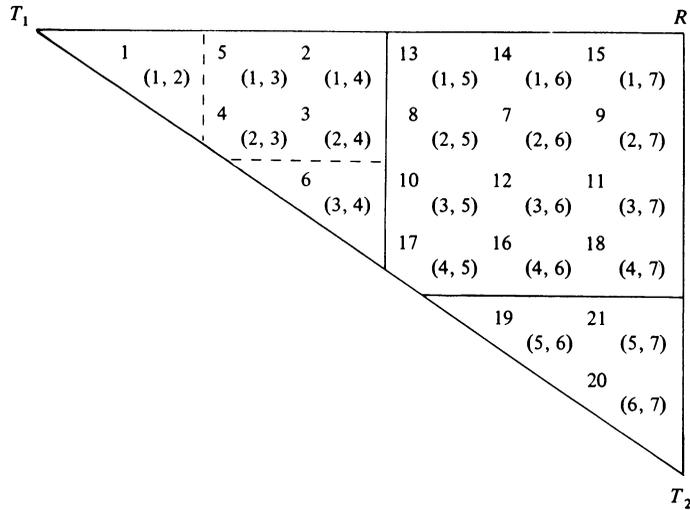


FIGURE 5.2

Now, by Powell's first theorem (cf. 2.1) $(\Delta A)^{-1/2}$ is the absolute value of the determinant of any matrix whose columns form a normalized set of mutually conjugate directions. Since we have assumed that the search directions given by columns of $D^{(k)}$ do not converge to mutual conjugacy, $\Delta D^{(k)}$ must tend to some limit strictly less than $(\Delta A)^{-1/2}$, say $(\Delta A)^{-1/2} - \delta$, where $\delta > 0$.

Further, there must exist a number $\beta > 0$ s.t. some pair of directions, at every iteration, has weight $\geq \beta$. (With A fixed, this β is dependent only on δ .) If not, then $\Delta D^{(k)}$ could be made arbitrarily close to $(\Delta A)^{-1/2}$, contradicting our assumption.

Assume also that a sufficiently large number of iterations have been carried out, so that $\Delta D^{(k)} \geq (\Delta A)^{-1/2} - \delta - \epsilon$ for some positive ϵ as small as we wish, and that we are at the start of a fresh cycle of iterations.

If we can claim that in proceeding through this cycle of iterations we must come across some current pair, say (p, q) with weight $\geq M(n)\beta$, (where $M(n)$ is a fraction dependent only on n), then we are practically home; because when pair (p, q) is revised, we can obtain a contradiction to our assumption $\Delta D^{(k)} \rightarrow (\Delta A)^{-1/2} - \delta$, from below. This follows from Lemma 3.2 and from $\Delta D^{(k)} \geq (\Delta A)^{-1/2} - \delta - \epsilon$; by taking ϵ sufficiently small we may show that $\Delta D^{(k)} > (\Delta A)^{-1/2} - \delta$, after revising the current pair (p, q) . \square

Most of our effort, therefore, goes into verifying the above claim. We utilize a series of lemmas leading to this result. Before each lemma we try to give some motivation for it. Detailed proofs are given in Appendix A. A closely related proof has also been used to show convergence of the cyclic Jacobi process for cyclic patterns in P , Nazareth [8].

THEOREM 5.1. *Suppose Algorithm C is applied to a quadratic and uses some cyclic pattern $\in P$. The current pair (p, q) , at iteration k , is selected according to this cyclic pattern, with $\Omega^{(k)}$ either U_{pq} or U_{pq}^T (if $d_p^{(k)} A d_q^{(k)} = 0$ the directions may either be revised or left unaltered).*

Then $D^{(k)}$ converges to mutual conjugacy, as $k \rightarrow \infty$.

To prove this theorem we shall need the following notation and several lemmas.

Notation. (i) When it is unnecessary to specify the iteration number, we shall denote the i th direction in the set by $d_i^{(\cdot)}$ and the matrix whose columns from the set of directions by $D^{(\cdot)}$.

(ii) Given a set of search directions $d_h^{(\cdot)}, d_{h+1}^{(\cdot)}, \dots, d_l^{(\cdot)}, \dots, d_m^{(\cdot)}$ Figure 5.3 develops the notation for certain sets of pairs of these directions.

The following simple lemma will prove useful later.

LEMMA 5.1. Consider normalized search directions $d_h^{(k+m)}, \dots, d_l^{(k+m)}, \dots, d_{l+m}^{(k+m)}$ obtained after $(m + 1)$ iterations of Algorithm C. These $(m + 1)$ iterations consist of successively revising pairs $(h, l), (h, l + 1), \dots, (h, l + m)$, i.e., from $\beta[h, l, m]$.

Assume during these iterations that no current pair had weight $> 1/2$; suppose further that $|d_\lambda^{(k+m)} Ad_\rho^{(k+m)}| > \delta$ ($0 < \delta < 1$) for some pair $(\lambda, \rho) \in Y[h, l, m]$. Then at iteration k

$$|d_\mu^{(k)} Ad_\nu^{(k)}| \geq \delta/2^{(m+1)}$$

for some pair $(\mu, \nu) \in Y[h, l, m]$.

LEMMA 5.2. Suppose at iteration k of Algorithm C the current pair is (p, q) and

$$(5.2a) \quad |d_q^{(k)} Ad_r^{(k)}| \geq \epsilon \quad (0 < \epsilon \leq 1).$$

If $|d_p^{(k)} Ad_r^{(k)}| \leq \epsilon/2$ then, after revising (p, q) , $|d_q^{(k+1)} Ad_r^{(k+1)}| \geq \epsilon/4$.

Referring to Figure 5.4, Lemma 5.2 states that, under certain specified conditions, the weight on (r, q) cannot all be transferred to (p, r) , when revising the current pair (p, q) . The following lemma is a generalization of this. Referring to Figure 5.3, it says that under certain specified conditions the total weight on $Z[h, l, m]$ cannot all be transferred to $\alpha[h, l]$ by successively revising pairs in $\beta[h, l, m]$.

LEMMA 5.3. Given a set of search directions $d_h^{(k)}, \dots, d_l^{(k)}, \dots, d_{l+m}^{(k)}$ suppose that for some pair $(\lambda, \rho) \in Z[h, l, m]$

$$(5.2b) \quad |d_\lambda^{(k)} Ad_\rho^{(k)}| \geq \epsilon, \quad 0 < \epsilon < 1.$$

Suppose that $(m + 1)$ further steps of Algorithm C are carried out, using successive pairs $(h, l), (h, l + 1), \dots, (h, l + m)$, i.e. from $\beta[h, l, m]$.

Then there exist nonzero fractions $K_1(m), K_2(m)$ and $K_3(m)$ dependent only upon m and monotonically nonincreasing with m , which satisfy the following proposition:

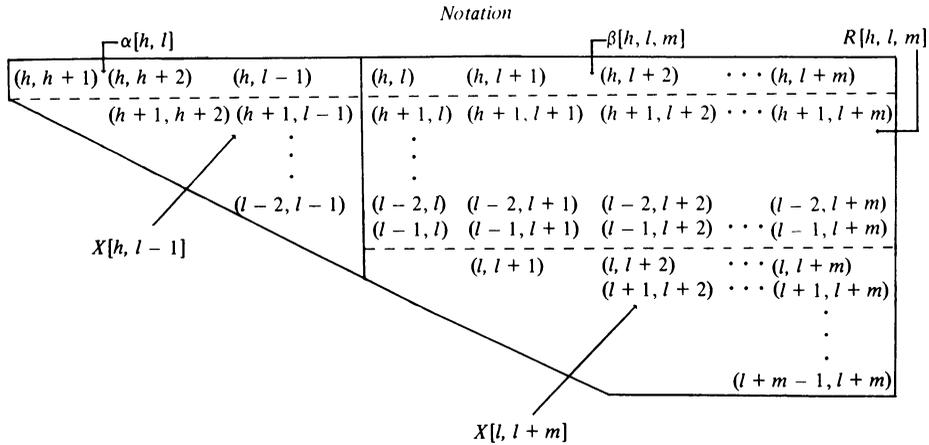
If

$$|d_h^{(k)} Ad_j^{(k)}| < K_1(m)\epsilon \quad \text{for all } (h, j) \in \alpha[h, l],$$

then at least one of the following two statements is true:

- (i) There exists a pair $(\mu, \nu) \in \gamma[h, l, m]$ such that $|d_\mu^{(k+m)} Ad_\nu^{(k+m)}| \geq K_2(m)\epsilon$.
- (ii) Some current pair during these $(m + 1)$ iterations has weight $\geq K_3(m)\epsilon$.

COROLLARY. The directions $d_l^{(k)}, d_{l+1}^{(k)}, \dots, d_{l+m}^{(k)}$ may be permuted arbitrarily. It is clear therefore that Lemma 5.3 holds with the pairs in $\beta[h, l, m]$ revised in any order.



- $\alpha[h, l] = \{(h, j): h < j < l\}$
- $\beta[h, l, m] = \{(h, j): l \leq j \leq l + m\}$
- $\gamma[h, l, m] = \{(i, j): (i < j) \ \& \ (h < i < l + m) \ \& \ (l \leq j \leq l + m)\}$
- $Y[h, l, m] = \alpha[h, l] \cup \beta[h, l, m] \cup \gamma[h, l, m]$
- $Z[h, l, m] = \beta[h, l, m] \cup \gamma[h, l, m]$
- $X[a, b] = \{(i, j) | a \leq i < j \leq b\}$
- $R[h, l, m] = \{(i, j) | (i < j) \ \& \ (h \leq i \leq l - 1) \ \& \ (l \leq j \leq l + m)\}$

FIGURE 5.3

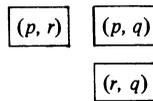


FIGURE 5.4

The next lemma is a generalization of Lemma 5.3. Referring to Figure 5.3 and using the notation developed there, Lemma 5.4 states that under certain specified conditions the total weight on $Z[h, l, m]$ cannot all be transferred to $X[h, l - 1]$ by revising pairs in $R[h, l, m]$.

LEMMA 5.4. *Given directions $d_h^{(k)}, \dots, d_l^{(k)}, \dots, d_{l+m}^{(k)}$ suppose that for some $(\lambda, \rho) \in Z[h, l, m]$*

$$(5.2c) \quad |d_\lambda^{(k)} A d_\rho^{(k)}| \geq \epsilon.$$

Suppose that $(l - h)(m + 1)$ further steps of Algorithm C are carried out, using pairs selected in sequence by rows from $R[h, l, m]$, i.e., $(h, l), \dots, (h, l + m), (h + 1, l), \dots, (h + 1, l + m), \dots, (l - 1, l), \dots, (l - 1, l + m)$.

Suppose that for all $(i, j) \in X[h, l - 1]$

$$(5.2d) \quad |d_i^{(k)} A d_j^{(k)}| < K_1(m) K_2(m)^{l-h-1} \epsilon.$$

Then at least one of the following two statements is true:

- (i) At the end of this process some pair $(\mu, \nu) \in X[l, l + m]$ has weight $\geq K_2(m)^{l-h}\epsilon$.
- (ii) Some current pair during the above $(l - h)(m + 1)$ iterations has weight $\geq K_3(m)K_2(m)^{l-h-1}\epsilon$.

COROLLARY 1. *Given two sets of directions $G_1 = \{d_n^{(\cdot)}, \dots, d_{l-1}^{(\cdot)}\}$ and $G_2 = \{d_l^{(\cdot)}, \dots, d_{l+m}^{(\cdot)}\}$ form a list of pairs L as in Step B(i) of Procedure P. Then Lemma 5.4 holds when successive pairs from $R[h, l, m]$ are revised in sequence given by list L . This follows immediately from application of the above proof to an appropriately permuted set of initial directions and use of the Corollary to Lemma 5.3.*

COROLLARY 2. *A similar result to Lemma 5.4 holds when the pairs in $R[h, l, m]$ are selected in sequence given by a list formed as in Step B(ii) of Procedure P.*

Our main lemma then follows:

LEMMA 5.5. *Consider Algorithm C applied to a quadratic using a cyclic pattern $\in P$. Assume that the search directions $d_1^{(k)}, \dots, d_n^{(k)}$, at the start of a fresh cycle, satisfy*

$$(5.2e) \quad |d_\lambda^{(k)} A d_\rho^{(k)}| \geq \epsilon$$

for some pair $(\lambda, \rho) \in X[1, n]$.

Then there exists a nonzero fraction $M(n)$ dependent only on n , such that the weight on some current pair during this cycle is $\geq M(n)\epsilon$.

Theorem 5.1 is proved as outlined earlier in subsection 5.2 using these lemmas. See Appendix A.

6. Conclusion. In earlier sections we have studied an attractive and simple way to revise the search directions that does not require knowledge of off-diagonal elements of $D^{(k)T}AD^{(k)}$. One potential disadvantage is as follows. Assume convergence to mutual conjugacy and suppose the off-diagonal elements of $D^{(k)T}AD^{(k)}$ are $O(\epsilon)$, where ϵ is small. Let us consider a pair (p, r) which has just been revised, so that its weight has been reduced from $O(\epsilon)$ to zero. The next time a pair which involves p or r is revised, the weight on (p, r) could build up again to $O(\epsilon)$. In order to curb this buildup it may be worthwhile, in the later stages of the iterative process, to incur the additional function evaluations needed to estimate $d_p^{(k)T}AD_q^{(k)}$. If we then use (2.4a) and (2.4b) to revise the current pair and renormalize p and q to be of different lengths in the A -norm, we observe the following. When the off-diagonal elements of $D^{(k)T}AD^{(k)}$ are $O(\epsilon)$, then $|\sin \theta|$ is also $O(\epsilon)$. A formal argument establishes that the total buildup of weight of a revised pair during a complete cycle of iterations is $O(\epsilon^2)$. The convergence is thus ultimately quadratic. Ultimate quadratic convergence of the cyclic Jacobi process applied to a matrix A with distinct eigenvalues is proven in a similar way, Wilkinson [9]. An important difference between the two processes, however, is that regardless of eigenvalues we can *always* bound θ , by ensuring that the denominator in (2.4b) does not vanish. In the Jacobi process renormalizations are not possible, since they would change the eigenvalues.

We have seen why it may be profitable to switch in later stages of the iteration from using fixed matrices chosen from S to using more general plane rotations. The

set S can be generalized, the matrix

$$\frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$$

is a typical example for the case when four directions are simultaneously revised. Let us represent the class of such matrices, when n directions are simultaneously revised, by $S[n]$ where n is even. Thus $S = S[2]$.

By taking products of certain matrices in S one may obtain a matrix in $S[n]$. For example, a matrix in $S[4]$ is given by $U_{12}U_{34}U_{13}U_{24}$ corresponding to the ordering (1, 2)(3, 4)(1, 3)(2, 4). From the discussion of cycling this implies that for orthogonal transformations chosen from $S[n]$ convergence of the search directions to mutual conjugacy need not occur. Also the conclusions arrived at earlier about ultimate rate of convergence apply to any fixed class of orthogonal matrices. The use of classes of orthogonal transformations given by $S[n]$ may, however, lead to fewer normalizations and hence fewer function evaluations.

Acknowledgement. I would like to thank Dr. B. N. Parlett for suggesting this topic of research and am deeply grateful for his encouragement, guidance and advice. Some useful comments by the referee are also gratefully acknowledged.

Appendix A.

Proof. Lemma 5.1. Suppose at some iteration the current pair is (h, q) . By assumption $|d_h^{(\cdot)}Ad_q^{(\cdot)}| \leq \frac{1}{2}$. Thus

$$(A.1) \quad \|d_h^{(\cdot)} \pm d_q^{(\cdot)}\|_A \geq 1.$$

With (h, q) the current pair, consider two linked pairs (h, r) and (q, r) each having weight $\leq w$. (2.4i) and (A.1) together imply that after revising (h, q) the weights on (h, r) and (q, r) are each $\leq 2w$.

Suppose, therefore, $|d_i^{(k)}Ad_j^{(k)}| < \delta/2^{(m+1)}$ for all $(i, j) \in Y[h, l, m]$.

After one iteration no pair can have weight exceeding $\delta/2^m$. Similarly, after $(m + 1)$ iterations no pair can have weight exceeding δ . This contradicts our assumption.

Therefore, there must exist a pair (μ, ν) , at iteration k , for which $|d_\mu^{(k)}Ad_\nu^{(k)}| \geq \delta/2^{(m+1)}$. \square

Proof. Lemma 5.2. Straightforward.

Proof. Lemma 5.3. With h and l fixed, let us use induction on m .

Suppose the lemma is true for all values up to $(m - 1)$. We must show it to be true for m .

1. If $\rho < l + m$ in (5.2b), then using the induction hypothesis and noting that the elements of $\gamma[h, l, m - 1]$ are unaffected by revising pair $(h, l + m)$, we see that the lemma holds with

$$K_i(m) = K_i(m - 1), \quad i = 1, 2, 3.$$

2. If $\rho = l + m$, then after revising pairs $(h, l), \dots, (h, l + m - 1)$ the weight of some element in column $l + m$, the last column of Figure 5.3, must exceed $\epsilon/2^m$. This follows from the Corollary to Lemma 3.1.

2.1. If this element is $(h, l + m)$, then (i) holds in the statement of Lemma 5.3 with $K_3(m) = 1/2^m$.

2.2. Suppose not. Say then that this element is $(i, l + m)$ where $h < i < l + m$.

2.2.1. After revising $(h, l + m)$ suppose the weight on $(i, l + m) > \frac{1}{4}(\epsilon/2^m)$.

Then (i) holds in Lemma 5.3 with $K_2(m) = 1/2^{m+2}$.

2.2.2. Suppose the assumption in 2.2.1 does not hold, i.e., after revising $(h, l + m)$ suppose

$$|d_i^{(k+m)} A d_{l+m}^{(k+m)}| < \frac{1}{4}(\epsilon/2^m).$$

It then follows from Lemma 5.2 that prior to revising $(h, l + m)$, the weight on pair (h, i) must have been $\geq \frac{1}{2}(\epsilon/2^m)$. From Lemma 5.1 it follows that at the start of the process (i.e., at iteration k) some element in $Y[h, l, m]$ has weight $\geq \frac{1}{2}(\epsilon/2^m)1/2^m$. This pair must be in $Z[h, l, m]$ provided we define $K_1(m) = K_1(m-1)/2^{2m+1}$. It follows from the induction hypothesis that after revising $(h, l), \dots, (h, l + m - 1)$ at least one of the following two statements must be true:

(i) for some $(\mu, \nu) \in \gamma(h, l, m - 1)$

$$|d_\mu^{(k+m-1)} A d_\nu^{(k+m-1)}| \geq K_2(m-1)(\epsilon/2^{2m+1});$$

(ii) some current pair has weight $\geq K_3(m-1)(\epsilon/2^{2m+1})$.

Furthermore, the pairs in $\gamma(h, l, m - 1)$ are unaffected when revising the pair $(h, l + m)$.

3. We see, therefore, for all cases in 1 and 2 above, that suitable values for $K_1(m)$, $K_2(m)$ and $K_3(m)$ are

$$K_1(m) = \frac{K_1(m-1)}{2^{2m+1}}, \quad K_2(m) = \frac{K_2(m-1)}{2^{2m+1}}, \quad K_3(m) = \frac{K_3(m-1)}{2^{2m+1}}.$$

4. To complete the argument by induction we must show that Lemma 5.3 is true when $m = 0$. Suppose some element in $Z[h, l, 0]$ has weight $\geq \epsilon$. If this is the current pair, then the lemma holds with $K_3(0) = 1$. If not, then by setting $K_1(0) = \frac{1}{2}$ and $K_2(0) = \frac{1}{4}$ we see that Lemma 5.3 is equivalent to Lemma 5.2. Therefore, Lemma 5.3 holds for $m = 0$.

This completes the inductive argument.

Proof. Lemma 5.4. With l and m fixed, the proof is by induction on h .

Suppose the lemma is true for directions $d_{h+1}^{(\cdot)}, \dots, d_l^{(\cdot)}, \dots, d_{l+m}^{(\cdot)}$. We must show it to be true for directions $d_h^{(\cdot)}, d_{h+1}^{(\cdot)}, \dots, d_l^{(\cdot)}, \dots, d_{l+m}^{(\cdot)}$. Consider revising pairs $(h, l), (h, l + 1), \dots, (h, l + m)$. By assumption all pairs in $X[h, l - 1]$ satisfy (5.2d). Thus, since $K_2(m) \leq 1$, all pairs in set $\alpha[h, l]$ have weight $< K_1(m)\epsilon$. Then, by Lemma 5.3 at least one of the following statements is true.

(a) some element in $\gamma[h, l, m]$ has weight $\geq K_2(m)\epsilon$.

(b) some current pair has weight $\geq K_3(m)\epsilon$.

Now, if (b) holds then for $d_h^{(\cdot)}, \dots, d_{l+m}^{(\cdot)}$, statement (ii) of Lemma 5.4 is true since

$K_3(m) \geq K_3(m)[K_2(m)]^{l-h-1}$. Suppose, therefore, (b) does not hold. Then (a) must hold. Furthermore no pair in $X[h + 1, l - 1]$ is affected when revising pairs in $\beta[h, l, m]$ since directions $d_{h+1}^{(\cdot)}, \dots, d_{l-1}^{(\cdot)}$ remain unaltered. Then, by the induction hypothesis, one of the following two statements is true:

- (i) at the end of the process some pair $(\mu, \nu) \in X[l, l + m]$ has weight $\geq K_2(m)^{l-h-1}(K_2(m)\epsilon)$;
- (ii) some current pair has weight $\geq K_3(m)K_2(m)^{l-h-2}(K_2(m)\epsilon)$. Therefore, in all cases Lemma 5.4 is true for directions $d_h^{(\cdot)}, \dots, d_l^{(\cdot)}, \dots, d_{l+m}^{(\cdot)}$.

To complete the inductive argument, we need only observe that for directions $d_{l-1}^{(\cdot)}, d_l^{(\cdot)}, \dots, d_{l+m}^{(\cdot)}$ Lemma 5.4 is equivalent to Lemma 5.3; and, therefore, the induction hypothesis holds for $h = l - 1$. \square

Proof. Lemma 5.5. The proof is by induction on the number of directions.

Suppose Lemma 5.5 is true for up to $(n - 1)$ directions. We must show it to be true for n directions.

Consider the first partition used to define the cyclic pattern employed. Say it is $\{d_1^{(\cdot)}, \dots, d_{l-1}^{(\cdot)}\}$ and $\{d_l^{(\cdot)}, \dots, d_n^{(\cdot)}\}$.

1. Suppose some element in $X[1, l - 1]$ has weight exceeding

$$(A.2) \quad \omega = K_1(n - l)[K_2(n - l)]^{l-2}\epsilon/2^I,$$

where $I = l(l - 1)/2$.

Then by the induction hypothesis some current pair has weight

$$\geq M(l - 1)\{K_1(n - l)[K_2(n - l)]^{l-2}/2^I\}\epsilon.$$

Thus Lemma 5.5 is true with

$$M(n) = M(l - 1)\{K_1(n - l)[K_2(n - l)]^{l-2}/2^I\}.$$

2. Suppose, therefore, that no element in $X[1, l - 1]$ has weight $\geq \omega$ as defined by (A.2). After revising all elements in $X[1, l - 1]$ in sequence given by the cyclic pattern, no element in $X[1, l - 1]$ has weight $\geq K_1(n - l)[K_2(n - l)]^{l-2}\epsilon$. This follows from Lemma 5.1. It follows also that the pair (λ, ρ) in (5.2e) must be $\in Z[1, l, n - l]$.

Then by Lemma 5.4, one of the following two statements is true:

- (i) after revising all elements in $R[1, l, n - l]$ in sequence given by the cyclic pattern, some element in $X[l, n]$ has weight $\geq [K_2(n - l)]^{l-1}\epsilon$;
- (ii) some current pair has weight $\geq K_3(n - l)[K_2(n - l)]^{l-2}\epsilon$.

- 2.1. If (ii) holds, then Lemma 5.5 is true with

$$M(n) = K_3(n - l)[K_2(n - l)]^{l-2}.$$

- 2.2. If (i) holds, then by the induction hypothesis applied to the directions $d_l^{(\cdot)}, \dots, d_n^{(\cdot)}$ some current pair has weight $\geq M(n - l + 1)[K_2(n - l)]^{l-1}\epsilon$. Again Lemma 5.5 holds with

$$M(n) = M(n - l + 1)[K_2(n - l)]^{l-1}.$$

3. Therefore, taking

$$M(n) = \{M(n-1)K_1(n)K_3(n)[K_2(n)]^n/2^N\}$$

where $N = n(n-1)/2$ covers all cases in 1 and 2 above. We see that Lemma 5.5 is true for n directions.

4. Trivially the lemma holds for two directions, with $M(2) = 1$, completing the proof by induction.

Proof. Theorem 5.1. Using Lemma 5.5, the proof of Theorem 5.1 is straightforward.

Suppose the search directions do not converge to a mutually conjugate set.

As $k \rightarrow \infty$,

$$\Delta D^{(k)} \rightarrow \Delta A^{-1/2} - \delta \quad (\delta > 0).$$

Given ϵ such that $0 < \epsilon \ll 1$, let $\Delta D^{(k)} > (\Delta A^{-1/2} - \delta - \epsilon)$ for all $k > K$.

We are assuming that the search directions do not converge to a mutually conjugate set. Then there exists $\beta > 0$ such that for any k , some pair (i, j) dependent on k , has weight $\geq \beta$.

Assume also we are at the start of a fresh cycle. Proceeding through a complete cycle of $n(n-1)/2$ iterations, Lemma 5.5 implies that some current pair, (p, q) has weight $\geq M(n)\beta = \gamma$ say, at iteration t .

Then, using Lemma 3.2,

$$\Delta D^{(t+1)} = \frac{\Delta D^{(t)}}{(1-\gamma^2)^{1/2}} > \frac{\Delta A^{-1/2} - \delta - \epsilon}{(1-\gamma^2)^{1/2}}.$$

If ϵ is chosen so that

$$\epsilon < (\Delta A^{-1/2} - \delta)[1 - (1 - \gamma^2)^{1/2}],$$

then $\Delta D^{(t+1)} > \Delta A^{-1/2} - \delta$.

This contradicts our assumption that $\Delta D^{(k)} \rightarrow \Delta A^{-1/2} - \delta$, from below.

Therefore, the n search directions $d_1^{(k)}, \dots, d_n^{(k)}$ must converge to a mutually conjugate set. \square

Applied Mathematics Division
Argonne National Laboratory
Argonne, Illinois 60439

1. M. J. D. POWELL, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *Comput. J.*, v. 7, 1964, pp. 155-162. MR 32 #4828.
2. W. I. ZANGWILL, "Minimizing a function without calculating derivatives," *Comput. J.*, v. 10, 1967, pp. 293-296. MR 38 #2930.
3. R. P. BRENT, *Algorithms for Minimization Without Derivatives*, Prentice-Hall Ser. in Automatic Computation, Prentice-Hall, Englewood Cliffs, N. J., 1973. MR 49 #4251.
4. N. W. BRODLIE, *A New Method for Unconstrained Minimization without Evaluating Derivatives*, IBM Report UKSC-0019-0019, IBM, Peterlee, United Kingdom, 1972.
5. M. J. D. POWELL, *Unconstrained Minimization Algorithms Without Computation of Derivatives*, UK. A. E. A. Harwell Report HL72/1713, 1972.
6. J. L. NAZARETH, Part I: *Unified Approach to Unconstrained Minimization*. Part II: *Generation of Conjugate Directions for Unconstrained Minimization Without Derivatives*, Dept. of Computer Science, Univ. of California, Berkeley, Rep. No. 23, 1973.
7. ELDON R. HANSEN, "On cyclic Jacobi methods," *J. Soc. Indust. Appl. Math.*, v. 11, 1963, pp. 448-459. MR 27 #6381.
8. J. L. NAZARETH, *On the Convergence of the Cyclic Jacobi Process*, Argonne National Lab. Tech. Memo, ANL-AMD, 1974.

9. J. H. WILKINSON, "Note on the quadratic convergence of the cyclic Jacobi process," *Numer. Math.*, v. 4, 1962, pp. 296–300. MR 26 #4473.
10. J. H. WILKINSON, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965. MR 32 #1894.
11. G. E. FORSYTHE & P. HENRICI, "The cyclic Jacobi method for computing the principal values of a complex matrix," *Trans. Amer. Math. Soc.*, v. 94, 1960, pp. 1–23. MR 22 #710.
12. PETER HENRICI, "On the speed of convergence of cyclic and quasicyclic Jacobi methods for computing eigenvalues of Hermitian matrices," *J. Soc. Indust. Appl. Math.*, v. 6, 1958, pp. 144–162. MR 20 #2084.
13. H. P. M. VAN KEMPEN, "On the convergence of the classical Jacobi method for real symmetric matrices with non-distinct eigenvalues," *Numer. Math.*, v. 9, 1966, pp. 11–18. MR 34 #2163.
14. J. N. FRANKLIN, *Matrix Theory*, Prentice-Hall, Englewood Cliffs, N. J., 1968. MR 38 #5798.