

NOTICE WARNING CONCERNING COPYRIGHT RESTRICTIONS:
The copyright law of the United States (title 17, U.S. Code) governs the making of photocopies or other reproductions of copyrighted material. Any copying of this document without permission of its author may be prohibited by law.

ORDER OF VECTOR RECURRENCES WITH APPLICATIONS TO
NONLINEAR ITERATION, PARALLEL ALGORITHMS, AND THE POWER METHOD

Alan Feldstein
Department of Mathematics
Arizona State University
Tempe, Arizona

J. F. Traub
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pa.

December 1974

Part of the first author's work was done while in residence at the Naval Research Laboratory, Washington, D.C. and at ICASE, NASA-Langley Research Center under NASA Grant NGR 47-102-001. Part of the second author's work was performed under the auspices of the U.S. Atomic Energy Commission while consulting at Lawrence Livermore Laboratory. The second author was supported in part by the National Science Foundation under Grant GJ32111 and the Office of Naval Research under Contract N0014-67-A-0314-0010, NR 044-422.

ABSTRACT

The behavior of the vector recurrence $\underline{y}_{n+1} = M\underline{y}_n + \underline{w}_{n+1}$ is studied under very weak assumptions. Let $\lambda(M)$ denote the spectral radius of M and let $\lambda(M) \geq 1$. Then if the \underline{w}_n are bounded in norm and a certain subspace hypothesis holds, the root order of the \underline{y}_n is shown to be $\lambda(M)$. If one additional hypothesis on the dimension of the principal Jordan blocks of M holds, then the quotient order of the \underline{y}_n is also $\lambda(M)$. The behavior of the homogeneous recurrence is studied for all values of $\lambda(M)$.

These results are applied to the analysis of

- (1) Nonlinear iteration with application to iteration with memory and to parallel iteration algorithms
- (2) Order and efficiency of composite iteration
- (3) The power method.

1. INTRODUCTION*

We study the behavior of the vector recurrence

$$(1) \quad y_{n+1} = My_n + w_{n+1}$$

under very weak assumptions. We apply our results to the power method to the analysis of iterations for nonlinear equations and to the composition of such iterations. In particular our results can be used to study one-point iterations with memory and iterations for solving nonlinear equations on parallel computers.

Let $\| \cdot \|$ denote any convenient vector norm or the induced matrix norm. When the following limits exist, define the root order by

$$R(y_n) = \lim_{n \rightarrow \infty} \|y_n\|^{\frac{1}{n}}$$

and the quotient order by

$$Q(y_n) = \lim_{n \rightarrow \infty} \frac{\|y_{n+1}\|}{\|y_n\|}$$

Clearly, if the quotient order exists, then so does the root order (though not conversely), and they are equal.

Let U be a nonsingular matrix such that

$$M = U^{-1}JU,$$

where J is the direct sum of K Jordan block matrices,

$$J = J_1 \oplus J_2 \oplus \dots \oplus J_K.$$

*

The first author presented some of the material in this paper at an ICASE Colloquium in August, 1973.

Let λ_k be the eigenvalue corresponding to J_k and let the dimension of J_k be D_k . Let the K Jordan blocks of J be arranged so that

$$\begin{cases} |\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_K|, \text{ and} \\ |\lambda_1| = |\lambda_2| = \dots = |\lambda_L| \text{ implies } D_1 \geq D_2 \geq \dots \geq D_L. \end{cases}$$

J_1, J_2, \dots, J_L are called the principal Jordan blocks of M . Denote $D = D_1$, and $\lambda = \lambda_1$. Thus $|\lambda|$ is the spectral radius of M , which we shall sometimes write as $\lambda(M)$.

In order to draw the conclusions which follow we must assume that the initial vector \underline{y}_0 does not lie in a certain subspace. Since the statement of this hypothesis is given in equation (16) and involves certain quantities not defined until Section 5, we find it convenient to label this as the "subspace hypothesis". We now state our main result; the proof is given in Section 5.

THEOREM 1. Assume $\lambda(M) \geq 1$ and

1. $\|\underline{y}_n\| \leq w < \infty$ for all n ,
2. "Subspace hypothesis".

Then

- i. $R(\underline{y}_n) = \lambda(M)$.

If, in addition,

3. $D_k < D$ for $2 \leq k \leq L$, when $|\lambda_k| = |\lambda|$ with $\lambda_k \neq \lambda$, then

- ii. $Q(\underline{y}_n) = \lambda(M)$

Herzberger [74] has independently analyzed the order of (1). However his assumptions are far more restrictive than ours. Of course there are important applications of (1) where Herzberger's conditions hold. In our terminology Herzberger's main result may be stated as

THEOREM 1'. Assume

1. $\lim_{n \rightarrow \infty} \underline{w}_n = \underline{w} < \infty$
2. $\lambda(M) > 1$
3. M is a non-negative matrix
4. M is primitive

Then

$$R(\underline{y}_n) = Q(\underline{y}_n) = \lambda(M).$$

Because of his strong conditions, Herzberger does not distinguish between the existence of root and quotient order. Recall (Varga [62]) that primitive means both irreducible and the existence of exactly one eigenvalue of largest modulus. Herzberger does not include a subspace hypothesis although we believe one to be necessary. There are many interesting problems where Theorem 1 holds but Theorem 1' cannot be applied. See Examples 1, 5-8.

The example $\underline{y}_{n+1} = \lambda \underline{y}_n + \underline{w}$, $\lambda < 1$, $\underline{w} \neq 0$, shows that Theorem 1 need not hold if $\lambda < 1$. However the conclusions of Theorem 1 hold for all values of λ if we restrict ourselves to homogeneous recurrences, $\underline{w}_n = 0$ for all n . We have

THEOREM 2. Assume

1. $\underline{w}_n = \underline{0}$ for all n .

If $\lambda(M) = 0$,

- i. $R(\underline{y}_n) = \lambda(M)$.

If $\lambda(M) > 0$ and if

2. "Homogeneous subspace hypothesis" then

ii. $R(\underline{y}_n) = \lambda(M)$.

If, in addition,

3. $D_k < D$ for $2 \leq k \leq L$, when $|\lambda_k| = \lambda$ with $\lambda_k = \lambda$, then

iii. $Q(\underline{y}_n) = \lambda(M)$.

Observe that if $\lambda(M) = 0$, then no subspace condition is required. If $\lambda(M) > 0$, then the homogeneous subspace hypothesis is the classical condition that the initial vector \underline{y}_0 may not be an eigenvector corresponding to a subdominant eigenvalue. In the notation of this paper the homogeneous subspace hypothesis is (16) with $\underline{f} = \underline{0}$ and $\delta = 1$ for all λ (even $|\lambda| = 1$).

EXAMPLE 1. We give an example where root order exists but quotient order does not. Consider (1) with

$$M = M_1 = \begin{pmatrix} 0 & 2 \\ 8 & 0 \end{pmatrix}, \quad \underline{w}_n = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \text{and } \underline{y}_0 = \begin{pmatrix} u \\ v \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The eigenvalues of M are ± 4 , so $\lambda(M_1) = 4$. Notice that

$$\left. \begin{aligned} y_{2n} &= 4^{2n} \begin{pmatrix} u \\ v \end{pmatrix} \\ y_{2n+1} &= 4^{2n+1} \begin{pmatrix} \frac{1}{2} v \\ 2 u \end{pmatrix} \end{aligned} \right\} \text{ for } n = 0, 1, 2, \dots$$

$$\|y_{2n}\| \frac{1}{2^n} = 4 \left\| \begin{pmatrix} u \\ v \end{pmatrix} \right\| \frac{1}{2^n} \rightarrow 4 \text{ as } n \rightarrow \infty$$

$$\|y_{2n+1}\| \frac{1}{2^{n+1}} = 4 \left\| \begin{pmatrix} \frac{1}{2} v \\ 2 u \end{pmatrix} \right\| \frac{1}{2^{n+1}} \rightarrow 4 \text{ as } n \rightarrow \infty.$$

Clearly $R(y_n) = 4 = \lambda(M_1)$. Since $\lambda_1 = 4$ and $\lambda_2 = -4$ while $D_1 = D_2 = 1$, hypothesis 3 of Theorem 1 does not hold, and thus the quotient order part of Theorem 1 does not apply. Indeed M_1 does not have quotient order. Let

$$A_n = \|y_{n+1}\| / \|y_n\|.$$

$$a = \left\| \begin{pmatrix} u \\ v \end{pmatrix} \right\| \quad \text{and} \quad b = \left\| \begin{pmatrix} \frac{1}{2} v \\ 2 u \end{pmatrix} \right\|.$$

Then $A_{2n} = 4b/a$ and $A_{2n+1} = 4a/b$. Clearly quotient order exists if and only

if $a = b$. For ℓ_p norms, $a = b$ if and only if $2|u| = |v|$.

Let $M_2 = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}$ with w_n and y_0 as above. Then

$$y_n = 2^n \begin{pmatrix} u \\ \frac{n}{2} u + v \end{pmatrix} \quad \text{for } n = 0, 1, 2, \dots$$

Therefore

$$\|y_{n+1}\| / \|y_n\| \rightarrow 2 = \lambda(M_2). \quad \blacksquare$$

We summarize the rest of this paper. Applications are considered in the next three sections while the proof of the main theorem is deferred until the last two sections. Readers interested primarily in the proof should turn first to Section 5. Section 2 discusses utilization of the power method to

calculate the spectral radius. Section 3 discusses the matrix representation of nonlinear iteration and utilizes the representation in the analysis of parallel algorithms. New results on the order and efficiency of composite iteration are analyzed in Section 4. The main result is proved in Section 5. Proofs of estimates needed in the proof of the main theorem may be of independent interest and appear in Section 6.

2. THE POWER METHOD

If $w_{n+1} = 0$ for all n , then (1) becomes $y_{n+1} = My_n$ and the y_n are the iterates of the power method. From Theorem 2 we can then conclude that the root order is always $\lambda(M)$ provided only that the subspace hypothesis holds. This hypothesis is non-restrictive in practice. We shall not pursue here whether this is the basis of a practical algorithm for estimating $\lambda(M)$. Usually the quotient order is used to compute $\lambda(M)$.

Since Bernoulli's method for polynomial zeros is a special case of the power method with M a companion matrix, Theorem 2 can be applied to Bernoulli's method when it is used to calculate the modulus of the largest zero.

One application for the calculation of the spectral radius is in connection with the determination of optimal relaxation factors for SOR. Young [71, p. 206] points out that the power method can be used to determine $\lambda(\mathcal{L}_\omega)$ if $\omega < \omega_b$. (We are using Young's notation.) Theorem 2 shows that the root form of the power method can be used even if $\omega \geq \omega_b$, and thus (at least in theory) the root form of the power method can be used for all ω .

3. MATRIX REPRESENTATION OF NONLINEAR ITERATION:
APPLICATIONS TO ITERATION WITH MEMORY AND TO
PARALLEL ALGORITHMS

Let a sequence of vectors $\{\underline{x}_n\}$ be generated by the vector-valued function $\varphi: R^N \rightarrow R^N$ and

$$(2) \quad \underline{x}_{n+1} = \varphi(\underline{x}_n).$$

Assume that at least one component of \underline{x}_n converges to at least one component of the constant vector $\underline{\alpha}$. Let the components of \underline{x}_n , \underline{y}_n , and $\underline{\alpha}$ be labelled $x_{n,j}$, $y_{n,j}$, and α_j . Let

$$y_{n,j} = \log|x_{n,j} - \alpha_j|$$

Then $\|\underline{y}_n\| \rightarrow \infty$ as $n \rightarrow \infty$. For many important problems, the vectors \underline{y}_n satisfy (1). Examples are given below. We then call

(1) the logarithmic error equation (or simply the error equation) for the sequence $\{\underline{x}_n\}$ and call M the matrix representation of the iteration function φ .

If (1) is the error equation of (2) we define the root order of φ as

$$R(\varphi) = R(\underline{y}_n)$$

and the quotient order of φ as

$$Q(\varphi) = Q(\underline{y}_n).$$

A comprehensive discussion of the order of iterative processes may be found in Ortega and Rheinboldt [70]. We have confined ourselves here to definitions of root and quotient order sufficient for our purpose.

The important idea of matrix representation of nonlinear iteration is due to Rice [71]. His matrix representation seems unnecessarily complicated. Rice's analysis does not distinguish between root and quotient order.

We turn to a number of examples and applications.

EXAMPLE 2

Assume that α is a zero of a scalar function g . One-point iterations with memory are of the form

$$(3) \quad z_{n+1} = \varphi(z_n, z_{n-1}, \dots, z_{n-N+1})$$

with errors satisfying

$$(4) \quad z_{n+1} - \alpha = c_{n+1} (z_n - \alpha)^{b_1} \dots (z_{n-N+1} - \alpha)^{b_N},$$

where the b_i are non negative integers. Examples of iterations satisfying equations (3) and (4) are interpolatory iterations (Traub [64]) and more generally HIFs (Hermite interpolatory iteration functions) denoted by (b_1, b_2, \dots, b_N) . See Feldstein and Firestone [67] and [69], Hindmarsh [72].

Any iteration satisfying equations (3) and (4) may be cast into the form of equations (1) and (2) as follows. Let

$$\begin{cases} x_{n+1,1} = z_{n+1} \\ x_{n,j} = z_{n+1-j} \end{cases} \quad \text{for } j = 1, 2, \dots, N.$$

Hence
$$x_{n+1,j} = x_{n,j-1} \quad \text{for } j = 2, 3, \dots, N.$$

Then (3) may be written as

$$(5) \quad \begin{cases} x_{n+1,1} = \varphi(x_{n,1}, \dots, x_{n,N}) \\ x_{n+1,j} = x_{n,j-1} \end{cases} \quad \text{for } j = 2, 3, \dots, N.$$

Let all components of $\underline{\alpha}$ be α . Taking absolute values and the logarithm of (4) yields

(6) $\underline{y}_{n+1} = M\underline{y}_n + \underline{w}_{n+1}$ where

$$M = \begin{pmatrix} b_1 & b_2 & \dots & b_{N-1} & b_N \\ 1 & 0 & & & \\ 0 & 1 & & & \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & \\ 0 & 0 & & 1 & 0 \end{pmatrix}, \quad \underline{w}_{n+1} = \begin{pmatrix} \log|c_{n+1}| \\ 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}$$

Observe that M is the companion matrix for the indicial polynomial of the linear recurrence obtained from (4) by taking logarithms.

Conclusions i and ii of Theorem 1 hold for all HIFs. The quotient order was first established by Traub [64] for the equal information case $b_1 = b_2 = \dots = b_N$, and by Feldstein and Firestone [67] for arbitrary non-negative integers b_1, \dots, b_N using recurrence equation techniques. ■

EXAMPLE 3

Write (2) in components as

$$x_{n+1,i} = \varphi_i(x_{n,1}, \dots, x_{n,N}) \quad \text{for } i = 1, \dots, N.$$

Assume that the errors satisfy

$$x_{n+1,i} - \alpha_i = c_{n,i} \prod_{j=1}^N (x_{n,j} - \alpha_j)^{m_{i,j}}.$$

This equation holds if each φ_i is a HIF (Feldstein and Firestone [67]).

Then the error equation (1) holds with

$$\underline{w}_n = (\log|c_{n,1}|, \dots, \log|c_{n,N}|)^T$$

and with the elements of M given by the $m_{i,j}$.

EXAMPLE 4

We wish to calculate a simple zero α of a scalar function g on a parallel or vector computer. A number of authors (Feldstein and Firestone [67], Shedler [67], Miranker [69], Rice [71]) have suggested generating N estimates of α at each iterative step. N may be the number of processors of a parallel machine. Most iterations proposed for this problem have the form of Example 2 with all components of $\underline{\alpha}$ equal to α . Suppose that processor i uses the scalar algorithm φ_i which satisfies the error equation in Example 3 (such as when each processor uses a HIF). Then the error equation (1) holds with $M = (m_{i,j})$ and \underline{w}_n as given in Example 3. ■

EXAMPLE 5

Care must be taken as to how the matrix M is constructed for parallel processing. For instance, if there are three processors and the first one uses the secant method (1,1), the second one uses the HIF (2,2), and the third one uses Newton's method (2), then the matrix representation is

$$M = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 2 & 0 \\ 2 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \lambda(M) = 3.$$

On the other hand if the second and third processors are interchanged, then

$$M = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & 0 \\ 2 & 2 & 0 \end{pmatrix} \quad \text{and} \quad \lambda(M) = 2.$$

Suppose instead that there are two processors and the first one uses Muller's method (1,1,1) while the second one uses the HIF (2,2,2); then a 3 dimensional vector of iterates is needed at each step. Both processors work on all three

components. The output of the first processor is $x_{n+1,1}$ and of the second processor is $x_{n+1,2}$. There are three simple choices for $x_{n+1,3}$: $x_{n+1,3} = x_{n,1}$ or $x_{n,2}$ or $x_{n,3}$.

These choices result, respectively, in the following three matrix representations:

$$M_1 = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \lambda(M_1) = \frac{3+\sqrt{13}}{2} \approx 3.303$$

$$M_2 = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \lambda(M_2) = \frac{3+\sqrt{17}}{2} \approx 3.562$$

$$M_3 = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \lambda(M_3) = 3.$$

In both these cases the maximum order corresponds to that algorithm which supplies the best information to that processor which uses the highest order scalar method. ■

4. COMPOSITION AND EFFICIENCY

We turn to the order of a composite iteration. (We could also study the order of composite vector recurrences.) Let φ_1 and φ_2 be 2 iteration functions both of which map $R^N \rightarrow R^N$. (It is sufficient to consider the composition of just 2 iteration functions. Multiple composition is handled similarly.) Suppose that

$$x_{n+1}^{(1)} = \varphi_1(x_n^{(1)})$$

$$x_{n+1}^{(2)} = \varphi_2(x_n^{(2)})$$

Then the composite iteration function $\psi = \varphi_2 \circ \varphi_1$, where $\psi: R^N \rightarrow R^N$, and the composite iteration sequence $\{x_n\}$ are defined by

$$x_{n+1} = \psi(x_n) \equiv \varphi_2(\varphi_1(x_n)).$$

Let φ_1 and φ_2 have characteristic matrices M_1 and M_2 with logarithmic error equations

$$y_{n+1}^{(1)} = M_1 y_n^{(1)} + w_{n+1}^{(1)}$$

$$y_{n+1}^{(2)} = M_2 y_n^{(2)} + w_{n+1}^{(2)}$$

Let $M = M_2 M_1$ and $w_{n+1} = M_2 w_{n+1}^{(1)} + w_{n+1}^{(2)}$. Then ψ has the error equation

$$y_{n+1} = M y_n + w_{n+1}$$

EXAMPLE 6

Consider the two HIFs (1,2) and (2,1) along with their composition (1,2,1) as discussed by Feldstein and Firestone [67, 69], Hindmarsh [72], and [71]. It is possible to obtain a matrix representation for the composition (1,2,1) by application of Example 2. Observe the following:

<u>Algorithm</u>	<u>Logarithmic Error Equation</u>	<u>Matrix</u>
(2,1)	$\Leftrightarrow z_{n+1} = 2z_n + z_{n-1} + 0(1) \Leftrightarrow$	$\begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$
(1,2)	$\Leftrightarrow z_{n+2} = z_{n+1} + 2z_n + 0(1) \Leftrightarrow$	$\begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix}$

The latter equation can be rewritten with the index reduced by 2 as

$$z_n = z_{n-1} + 2z_{n-2} + 0(1)$$

Eliminate z_{n+1} and z_{n-1} from these three equations to obtain

$$(1,2) \circ (2,1) \Leftrightarrow z_{n+2} = 5z_n - 2z_{n-2} + 0(1) \Leftrightarrow \begin{pmatrix} 5 & -2 \\ 1 & 0 \end{pmatrix}$$

Thus, this matrix representation of the composite algorithm $(1,2) \circ (2,1)$ has a negative entry and Herzberger's result (Theorem 1') cannot be applied. ■

In the following theorem the use of D , D_k , λ , λ_k and the subspace hypotheses all refer to the matrix $M = M_2 M_1$. We state the main Composition Theorem. Its proof is an immediate consequence of Theorem 1 plus the fact that ψ has the error equation

$$y_{n+1} = M y_n + w_{n+1}.$$

Since we refer to convergent iteration functions in Theorem 3, we have included the hypothesis that $\lambda(M_2 M_1) \geq 1$. This hypothesis may be deleted for homogeneous composite vector recurrences.

THEOREM 3. Assume

1. $\|w_{n+1}\| \leq w < \infty$ for all n
2. "Subspace Hypothesis"
3. $\lambda(M_2 M_1) \geq 1$

Then

$$R(\varphi_2 \circ \varphi_1) = \lambda(M_2 M_1).$$

If, in addition,

4. $D_k < D$ for $2 \leq k \leq L$, when $|\lambda_k| = |\lambda|$ with $\lambda_k \neq \lambda$, then

$$Q(\varphi_2 \circ \varphi_1) = \lambda(M_2 M_1).$$

In the discussions that follow we do not distinguish between root and quotient order; either order is represented by $p(\varphi)$. We shall denote $p_i = p(\varphi_i)$ and $p_{i,j} = p(\varphi_i \circ \varphi_j)$.

EXAMPLE 7

Consider the three HIFs $\varphi_1 = (1,2)$, $\varphi_2 = (2,1)$, $\varphi_3 = (2)$. It is instructive to consider the possible composite algorithms $\psi_{i,j} = \varphi_i \circ \varphi_j$ with order $p_{i,j}$. The matrix representations and orders are

$$\begin{aligned} M_1 &= \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix} & \text{and} & p_1 = 2 \\ M_2 &= \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} & \text{and} & p_2 = 1+\sqrt{2} \\ M_3 &= \begin{pmatrix} 2 & 0 \\ 1 & 0 \end{pmatrix} & \text{and} & p_3 = 2 \end{aligned}$$

By Theorem 2

$$\begin{aligned} \psi_{1,2} &\Leftrightarrow \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 2 & 1 \end{pmatrix} & \text{and} & p_{1,2} = \frac{5+\sqrt{17}}{2} \\ \psi_{3,1} &\Leftrightarrow \begin{pmatrix} 2 & 0 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix} & \text{and} & p_{3,1} = 4 \\ \psi_{3,2} &\Leftrightarrow \begin{pmatrix} 2 & 0 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix} & \text{and} & p_{3,2} = 5 \end{aligned}$$

Then

$$\begin{aligned} 4.562 &\approx \frac{5+\sqrt{17}}{2} = p_{1,2} < p_1 p_2 = 2(1+\sqrt{2}) \approx 4.828 \\ 4 &= p_{3,1} = p_3 p_1 = 4 \\ 5 &= p_{3,2} > p_3 p_2 = 2(1+\sqrt{2}) \approx 4.828 \end{aligned}$$

Thus, a composite algorithm may have an order of convergence either less than, equal to, or greater than the product of the orders of the individual algorithms. ■

If the characteristic matrices M_1 and M_2 do not have the same dimension, the size of the smaller one can be increased in a fashion which does not alter its association with the underlying algorithm and which still permits the application of Theorem 3. Suppose M_1 is $N \times N$ in size. Define its $(N+1) \times (N+1)$ extension M_1^* as follows:

$$M_1^* = \left(\begin{array}{c|c} M_1 & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline \underbrace{0 \ \dots \ 0}_{N-1} & \begin{matrix} 1 \\ 0 \end{matrix} \end{array} \right)$$

Clearly $\lambda(M_1^*) = \lambda(M)$. Furthermore, if M_1 and M_2 are $N \times N$ matrices, then

$$\lambda(M_1 M_2) = \lambda(M_1^* M_2^*).$$

If M_1 represents the map from \underline{x}_{n-1} to the N vector $(x_{n,1}, \dots, x_{n,N})^T$, then M_1^* represents the map to the $(N+1)$ vector $(x_{n,1}, \dots, x_{n,N}, x_{n-1,N})^T$.

Since (Wilkinson [65]) $\lambda(M_1 M_2) = \lambda(M_2 M_1)$, then Theorem 2 implies that $p_{1,2} = p_{2,1}$ and we have

COROLLARY 1. Order is invariant under commuting of two compositions.

EXAMPLE 8

Corollary 1 is false for three compositions, because $\lambda(M_1 M_2 M_3)$ need not equal $\lambda(M_3 M_2 M_1)$. For example, let

$$M_1 = \begin{pmatrix} 2 & 0 \\ 1 & 0 \end{pmatrix}, M_2 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, M_3 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

Then

$$M_1 M_2 M_3 = \begin{pmatrix} 4 & 0 \\ 2 & 0 \end{pmatrix} \quad \text{and} \quad p_{1,2,3} = 4.$$

$$M_3 M_2 M_1 = \begin{pmatrix} 3 & 0 \\ 3 & 0 \end{pmatrix} \quad \text{and} \quad p_{3,2,1} = 3.$$

The iteration $M_3 M_2 M_1$ appears in Traub [64, Example 8-1] and is also considered by Rice. The iteration $M_1 M_2 M_3$ represents successive applications of Newton iteration. ■

If φ_1 and φ_2 are iterations with the same matrix representation M , then $\lambda(M_1 M_2) = \lambda(M^2) = \lambda^2(M)$. In particular this holds if $\varphi_1 = \varphi_2$. Thus we have the very useful

COROLLARY 2. Order multiplies under self-composition.

In addition to the case covered by Corollary 2, there is another important case for which order multiplies under composition. This is when φ is a scalar one-point iteration (Traub [64, Chapter 2]). In general order does not multiply under composition even in the case of scalar φ . (See the algorithms in Example 7.) This was first shown by Hindmarsh [72] who used difference equation techniques in his analysis, and by Rice [71] using matrix representation techniques.

To compare iterations we need the concept of an efficiency measure. Work on efficiency is reported in Brent [72], Feldstein [69], Feldstein and Firestone [67, 69], Hindmarsh [72], Kung [73], Kung and Traub [74], Ostrowski [66] and Traub [72, 74, 74a]. Let $c(\varphi) > 0$ be some "cost" associated with computing x_{n+1} from x_n . The efficiency of φ is defined, for $p > 1$, by

$$e(\varphi) = \frac{\log p(\varphi)}{c(\varphi)}$$

We wish to compare the efficiency of the composite iteration $\varphi_2 \circ \varphi_1$ with the efficiencies of φ_1 and φ_2 . Denote

$$c_{i,j} = c(\varphi_i \circ \varphi_j) \quad \text{and} \quad e_{i,j} = e(\varphi_i \circ \varphi_j).$$

We assume that $c_{1,2} = c_{2,1}$. Then by Corollary 1, $e_{1,2} = e_{2,1}$. It is reasonable to assume that $c_{2,1} \leq c_1 + c_2$. In the following Theorem assume that $p_1 = \lambda(M_1)$, $p_2 = \lambda(M_2)$, and $p_{2,1} = \lambda(M_2 M_1)$. The hypotheses of Theorems 1 and 3 are sufficient to guarantee this.

THEOREM 4

1. Let $c_{2,1} = c_1 + c_2$.

- a. If $\lambda(M_2 M_1) > \lambda(M_2)\lambda(M_1)$, then $\min(e_1, e_2) < e_{2,1}$.
- b. If $\lambda(M_2 M_1) = \lambda(M_2)\lambda(M_1)$, then $\min(e_1, e_2) \leq e_{2,1} \leq \max(e_1, e_2)$.
- c. If $\lambda(M_2 M_1) < \lambda(M_2)\lambda(M_1)$, then $e_{2,1} < \max(e_1, e_2)$.

2. On the other hand, let $c_{2,1} < c_1 + c_2$.

If $\lambda(M_2 M_1) \geq \lambda(M_2)\lambda(M_1)$, then $\min(e_1, e_2) < e_{2,1}$.

PROOF. We confine the proof to case 1b since the remaining cases are proven similarly.

$$\begin{aligned} e_{2,1} &= \frac{\log \lambda(M_2 M_1)}{c_{2,1}} = \frac{\log \lambda(M_2) + \log \lambda(M_1)}{c_2 + c_1} \\ &= \frac{c_2 e_2 + c_1 e_1}{c_2 + c_1} \end{aligned}$$

and the result follows. ■

This theorem gives conditions such that the efficiency of a composite iteration is greater than the minimum efficiencies of the component iterations. Of greater practical and theoretical interest is the possibility that the efficiency of a composite iteration may be greater than the maximum efficiencies of the component iterations. That is, we hope to obtain

$$(7a) \quad e_{2,1} > \max(e_1, e_2).$$

In view of Theorem 4 this can only happen if either of the following conditions holds:

$$(7b) \quad c_{2,1} < c_1 + c_2$$

or

$$(7c) \quad \lambda(M_2 M_1) > \lambda(M_2) \cdot \lambda(M_1).$$

Example 9 shows a situation where (7a,b,c) all hold. The possibility of using (7b) in order to increase efficiency was first pointed out by Feldstein and Firestone [67] and was exploited by them in Feldstein and Firestone [69]. Hindmarsh [72] demonstrated that (7c) might hold.

We can easily calculate sufficient conditions for (7a) to hold. For example, assume that $c_{2,1} = c_1 + c_2$ and that (7c) holds. Let

$$\eta = \log \lambda(M_2 M_1) - \log \lambda(M_2) - \log \lambda(M_1) > 0.$$

Then $e_{2,1} > \max(e_1, e_2)$ if

$$\eta > c_1(e_2 - e_1), \text{ for } e_2 > e_1$$

$$\eta > c_2(e_1 - e_2), \text{ for } e_2 < e_1$$

In general we are interested in iterations such that the quantities η and μ defined by

$$\eta = \log \lambda(M_2 M_1) - \log \lambda(M_2) - \log \lambda(M_1)$$

$$\mu = c(\varphi_1) + c(\varphi_2) - c(\varphi_2 \circ \varphi_1)$$

are positive and as large as possible. How to do this is an open question.

EXAMPLE 9

Use the notation and algorithms of Example 7. Let $c(\varphi) \equiv$ the number of new function or derivative evaluations per iteration step. Then

$$c_1 = c_2 = c_3 = 2$$

$$c_{1,2} = c_{3,1} = c_{3,2} = 3$$

Thus, (7b) holds. In Example 7 it was shown that $\lambda(M_3 M_2) > \lambda(M_2) \cdot \lambda(M_3)$ and thus (7c) holds, too. Furthermore,

$$e_2 = \frac{\log \lambda_2}{c_2} \approx .4407$$

$$e_3 = \frac{\log \lambda_3}{c_3} \approx .3466$$

$$e_{2,3} = \frac{\log \lambda_{2,3}}{c_{2,3}} \approx .5365$$

Thus $e_{2,3} > \max(e_2, e_3)$ and (7a) also holds. ■

5. PROOF OF THE MAIN THEOREM

The proof of Theorem 1 uses certain estimates (Lemmas 1 and 2, below) on the growth as $n \rightarrow \infty$ of each block J_k^n for $1 \leq k \leq K$. Since the proofs of these lemmas are rather long and since these lemmas may be of independent interest, we defer the proofs until Section 6 and confine ourselves here to a statement of the results needed for the proof of Theorem 1. We shall start from equation (1), written here for notational convenience with primes as

$$\underline{y}'_{n+1} = M\underline{y}'_n + \underline{w}'_{n+1}.$$

Write M in the Jordan form $M = U^{-1}JU$. Let

$$\underline{y}_n = U\underline{y}'_n \text{ and } \underline{w}_n = U\underline{w}'_n.$$

Then

$$\underline{y}_{n+1} = J\underline{y}_n + \underline{w}_{n+1}.$$

Let $C(a,b)$ denote a binomial coefficient. Let $\underline{w}_{n,k}$ be that portion of the vector \underline{w}_n which is associated with the Jordan block J_k . Let

$$(8) \quad \delta = \begin{cases} 0 & \text{if } |\lambda| = 1 \\ 1 & \text{otherwise,} \end{cases}$$

$$(9) \quad Q_{n,k} = J_k^n / (\lambda^n C(n, D-\delta)),$$

$$(10) \quad \underline{s}_{n,k} = \sum_{i=0}^n J_k^i \underline{w}_{n-i,k}.$$

When the limit exists (see equations (28) and (30)) define

$$\underline{f}_k(\underline{z}) = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-D} z^i \frac{C(n-i, D-1)}{C(n, D-\delta)} \underline{w}_{i,k}.$$

If $\|\underline{w}_n\| < w < \infty$ for all n , then we will show (30) that this limit exists and indeed that $\underline{f}_k(z)$ is analytic for $|z| \leq |\lambda^{-1}|$.

For $1 \leq k \leq K$ let W_k be the $D_k \times D_k$ matrix having ones on the superdiagonal and zeros elsewhere. Then $W_k^j = 0$ for $j \geq D_k$ and $W_k^{D_k-1}$ is a matrix with a one in the upper righthand corner and zeros elsewhere. If $D_k = 1$, define $W_k = 0$, $W_k^{D_k-1} = 1$. Usually the subscript of W_k will be clear from context, and we will simply write W . The symbol $O(n^{-1})$ will denote a scalar, a vector, or a matrix (according to context) each of whose entries is bounded in absolute value by n^{-1} times some nonnegative constant.

LEMMA 1. Let $\lambda(M) > 0$. The following hold for $1 \leq k \leq K$:

1. If $|\lambda_k| < |\lambda|$, then $Q_{n,k} = O(n^{-1})$.
2. If $D_k \leq D-1$, then $Q_{n,k} = O(n^{-1})$.
3. If $D_k = D$ and $|\lambda_k| = |\lambda|$, then

$$Q_{n,k} = \delta (\lambda_k/\lambda)^n \lambda_k^{1-D} W^{D-1} + O(n^{-1}).$$

LEMMA 2. Let $\lambda(M) \geq 1$ and $\|\underline{w}_n\| \leq w < \infty$ for all n . The following hold for $1 \leq k \leq K$:

1. If $|\lambda_k| < |\lambda|$, then $\underline{s}_{n,k}/(\lambda^n C(n, D-\delta)) = O(n^{-1})$.
2. If $D_k \leq D-1$, then $\underline{s}_{n,k}/(\lambda^n C(n, D-\delta)) = O(n^{-1})$.
3. If $|\lambda_k| = |\lambda|$ and $D_k = D$, then

$$\underline{s}_{n,k}/(\lambda^n C(n, D-\delta)) = (\lambda_k/\lambda)^n \lambda_k^{1-D} W^{D-1} \underline{f}_k(\lambda_k^{-1}) + O(n^{-1})$$

$$\lim_{n \rightarrow \infty} \|\underline{s}_{n,k}/(\lambda^n C(n, D-\delta))\| = |\lambda_k|^{1-D} \|W^{D-1} \underline{f}_k(\lambda_k^{-1})\|$$

Furthermore $\underline{f}_k(z)$ is analytic for $|z| \leq |\lambda^{-1}|$.

REMARK

In part 3 of Lemmas 1 and 2 notice that $\lim_{n \rightarrow \infty} (\lambda_k / \lambda)^n$ exists if and only if $\lambda_k = \lambda$. Hence the following limit results hold if and only if $\lambda_k = \lambda$:

$$\lim_{n \rightarrow \infty} Q_{n,k} = \delta \lambda_k^{1-D} W^{D-1}, \quad \lim_{n \rightarrow \infty} \frac{s_{n,k}}{(\lambda^n C(n, D-\delta))} = \lambda_k^{1-D} W^{D-1} \underline{f}_k(\lambda_k^{-1})$$

This point is the important key to hypothesis 3 which distinguishes between the root and quotient order results of Theorems 1 and 2, because in order for the quotient order result to hold, the limits above must hold. See also equation (17). ■

We now prove Theorem 1. Start from (1) in Jordan form

$$(11) \quad y_{n+1} = J y_n + w_{n+1}.$$

Let $y_{n,k}$ and $w_{n,k}$ be the portions of the vectors y_n and w_n corresponding to the Jordan block J_k . It may be easily verified from equations (10) and (11) that

$$(12) \quad y_{n,k} = J_k^n y_{0,k} + s_{n,k}$$

where $s_{n,k}$ was given by (10). Let

$$(13) \quad v_{n,k} = y_{n,k} / (\lambda^n C(n, D-\delta)).$$

Substitute (12) and (9) into (13). Then

$$(14) \quad v_{n,k} = Q_{n,k} y_{0,k} + \frac{s_{n,k}}{\lambda^n C(n, D-\delta)}.$$

Since $\|w'_n\|$ is bounded by hypothesis and U is a nonsingular matrix, $\|w_n\|$ is also bounded. Lemmas 1 and 2 may be applied. Parts 1 and 2 of the Lemmas clearly show that we need to consider only those components for which $|\lambda_k| = |\lambda|$ and $D_k = D$ both hold, for otherwise $v_{n,k} \rightarrow 0$. Apply part 3 of Lemmas 1 and 2 to those components to obtain

$$(15) \quad v_{n,k} = (\lambda_k / \lambda)^n \lambda_k^{1-D} W_k^{D-1} \{ \delta y_{0,k} + \underline{f}_k(\lambda_k^{-1}) \} + O(n^{-1}).$$

We are now ready to define the subspace hypothesis which is hypothesis 2 in Theorems 1, 2, and 3.

$$(16) \quad W_k^{D-1} \{ \delta y_{0,k} + \underline{f}_k(\lambda_k^{-1}) \} \neq \underline{0} \text{ for some } k \text{ such that } |\lambda_k| = |\lambda| \text{ and } D_k = D.$$

(Recall that $y_{0,k}$ and \underline{f}_k in (16) come from the Jordan transform of (1). Recall also that in Theorem 2 $\underline{f}_k = \underline{0}$ and $\delta = 1$, always.) Note

(16) implies that the full vector \underline{v}_n has at least one component which, as $n \rightarrow \infty$ is bounded away from zero (recall that $|\lambda_k| = |\lambda| \neq 0$). By (13), the same is true for the vector \underline{y}_n and thus also for \underline{y}'_n .

Since $\delta = 0$ when $\lambda(M) = 1$, the subspace hypothesis becomes

$$W_k^{D-1} \underline{f}_k(\lambda_k^{-1}) \neq 0$$

for some k such that $|\lambda_k| = 1$ and $D_k = D$. There are many nontrivial situations that yield $\underline{f}_k(\lambda_k^{-1}) = 0$. In such cases Theorem 1 cannot be applied although $R(\underline{y}_n)$ and $Q(\underline{y}_n) = 1$ may still hold. We shall not pursue this in the present paper.

To establish the root order result, take norms in (15).

$$\| \underline{v}_{n,k} \| \leq |\lambda|^{1-D} \| W \|^{D-1} \{ \delta \| \underline{y}_{0,k} \| + \| \underline{f}_k(\lambda_k^{-1}) \| \} + o(n^{-1})$$

Thus \underline{v}_n is bounded in norm. Since $C(n, D-\delta)/n^{D-\delta} \rightarrow 1/(D-\delta)!$ as $n \rightarrow \infty$, then

$$\underline{y}'_n = U^{-1} \underline{y}_n = \lambda^n n^{D-\delta} \underline{v}'_n$$

for some vector \underline{v}'_n which is bounded in norm and which by application of the subspace hypotheses is bounded away from zero for n sufficiently large.

Therefore,

$$\| \underline{y}'_n \| \frac{1}{n} = |\lambda| n^{\frac{D-\delta}{n}} \| \underline{v}'_n \| \frac{1}{n} \rightarrow |\lambda| \text{ as } n \rightarrow \infty.$$

Hence $R(\underline{y}'_n) = \lambda(M)$.

To establish the quotient order result, we need to consider only $k = 1$ (by the hypotheses of Theorem 1 and by Lemmas 1 and 2). Since $\lambda_1 = \lambda$, then equation (15) becomes (see the Remark following Lemma 2)

$$(17) \quad \underline{v}_{n,1} = \lambda^{1-D} W^{D-1} \{ \delta \underline{y}_{0,1} + \underline{f}_1(\lambda^{-1}) \} + O(n^{-1}).$$

An application of the subspace hypothesis shows that $\underline{v}_{n,1}$ has a nonzero limit as $n \rightarrow \infty$. Thus there is some vector \underline{v}'_n , with a nonzero limit, for which

$$\underline{y}'_n = \lambda^n n^{D-\delta} \underline{v}'_n$$

$$\frac{\|\underline{y}'_{n+1}\|}{\|\underline{y}'_n\|} = |\lambda| \left(\frac{n+1}{n} \right)^{D-\delta} \frac{\|\underline{v}'_{n+1}\|}{\|\underline{v}'_n\|} \rightarrow |\lambda| \text{ as } n \rightarrow \infty.$$

Hence $Q(\underline{y}'_n) = \lambda(M)$ which completes the proof of Theorem 1. ■

The proof of Theorem 2 is a minor modification of the proof of Theorem 1, with $\underline{s}_{n,k} = \underline{0}$, and is omitted.

6. PROOFS OF THE LEMMAS

The statements of Lemmas 1 and 2 were given in Section 5 and will not be repeated here. Fix k for $1 \leq k \leq K$. Let $\lambda_k \neq 0$, since the results are trivial otherwise. Recall that $\delta = 0$ if $|\lambda| = 1$ and $\delta = 1$ otherwise $n \geq D_k$. Then

$$J_k^n = (\lambda_k I + W)^n = \sum_{m=0}^{D_k-1} \lambda_k^{n-m} C(n,m) W^m$$

$$(18) \quad J_k^n = \lambda_k^n C(n, D-\delta) \sum_{m=0}^{D_k-1} \lambda_k^{-m} \frac{C(n,m)}{C(n, D-\delta)} W^m.$$

PROOF OF LEMMA 1. Substituting (18) into (9) yields

$$Q_{n,k} = \left(\frac{\lambda_k}{\lambda}\right)^n \sum_{m=0}^{D_k-1} \lambda_k^{-m} \frac{C(n,m)}{C(n, D-\delta)} W^m$$

Since $C(n,m)/C(n, D-\delta) = O(n^{m-D+\delta}) \leq O\left(\binom{D_k-D}{n}\right)$ for $0 \leq m \leq D_k-1$, then

$$\|Q_{n,k}\| \leq |\lambda_k/\lambda|^n O\left(\binom{D_k-D}{n}\right).$$

1. If $|\lambda_k| < |\lambda|$, then $|\lambda_k/\lambda|^n O\left(\binom{D_k-D}{n}\right) = O(n^{-1})$. Thus Q_n , as desired.
2. If $D_k \leq D-1$, then $O\left(\binom{D_k-D}{n}\right) \leq O(n^{-1})$ while $|\lambda_k/\lambda|^n \leq 1$. $Q_{n,k} = O(n^{-1})$, as desired.
3. If $D_k = D$ and $|\lambda_k| = |\lambda|$, then

$$\left(\frac{\lambda_k}{\lambda}\right)^n \sum_{m=0}^{D_k-2} \lambda_k^{-m} \frac{C(n,m)}{C(n, D-\delta)} W^m = O(n^{-1})$$

by the proof of part 2 above. Hence $Q_{n,k}$ is dominated by the term $m = D_k - 1$; that is,

$$Q_{n,k} = \left(\frac{\lambda_k}{\lambda}\right)^n \left\{ \lambda_k^{1-D} \frac{C(n, D-1)}{C(n, D-\delta)} W^{D-1} \right\} + O(n^{-1}).$$

Since $\delta = 0$ or $\delta = 1$, then

$$C(n, D-1)/C(n, D-\delta) = \delta + O(n^{-1}).$$

Therefore

$$Q_{n,k} = \delta (\lambda_k/\lambda)^n \lambda_k^{1-D} W^{D-1} + O(n^{-1}),$$

which completes the proof of Lemma 1. ■

PROOF OF LEMMA 2. Write

$$(19) \quad \underline{t}_{n,k} = \sum_{i=D_k}^n J_k^i \underline{w}_{n-i,k}.$$

Recall (10) and apply $\|\underline{w}_n\| \leq w$. Hence

$$\|\underline{s}_{n,k} - \underline{t}_{n,k}\| \leq w \sum_{i=0}^{D_k-1} \|J_k\|^i = O(1).$$

Hence

$$\lim_{n \rightarrow \infty} \|\underline{s}_{n,k} - \underline{t}_{n,k}\| / (\lambda^n C(n, D-\delta)) = 0.$$

Thus, it suffices to consider $\underline{t}_{n,k}$ instead of $\underline{s}_{n,k}$. Denote

$$(20) \quad \underline{q}_{n,k} = \underline{t}_{n,k} / (\lambda^n C(n, D-\delta)).$$

Substitute (18) with $n = i$ and (19) into (20). Interchange the order of summation to obtain

$$(21) \quad q_{n,k} = \sum_{m=0}^{D_k-1} W^m \left\{ \sum_{i=D_k}^n \lambda_k^{i-m} \lambda^{-n} \frac{C(i,m)}{C(n,D-\delta)} w_{n-i,k} \right\}.$$

1. Suppose that $|\lambda_k| < |\lambda|$. Consider $\tau_k \geq 1$ such that $|\lambda_k| < |\tau_k| \leq |\lambda|$.

(τ_k will be picked later.) Let $\underline{b}_{n,k}$ denote the vector.

$$(22) \quad \underline{b}_{n,k} = \sum_{m=0}^{D_k-1} \lambda_k^{-m} W^m \sum_{i=D_k}^n C(i,m) \left(\frac{\lambda_k}{\tau_k} \right)^i \tau_k^{i-n} w_{n-i,k}.$$

Then equation (21) may be written as

$$(23) \quad q_{n,k} = \left(\frac{\tau_k}{\lambda} \right)^n \underline{b}_{n,k} / C(n,D-\delta).$$

Consider the functions $\psi_n(z) = \sum_{i=0}^n z^i$ and $\psi(z) = \sum_{i=0}^{\infty} z^i$. Since $\lim_{n \rightarrow \infty} \psi_n(z) \equiv \psi(z)$

is analytic for $|z| < 1$, then $\lim_{n \rightarrow \infty} \psi_n^{(m)}(z) \equiv \psi^{(m)}(z)$ is also analytic, and thus also absolutely convergent for $|z| < 1$ (superscript denotes differentiation). Thus

$$|\psi_n^{(m)}(z)| \leq \psi^{(m)}(|z|) \text{ for } |z| < 1.$$

It is not hard to verify that

$$(24) \quad \sum_{i=m}^n C(i,m) z^i = \frac{1}{m!} z^m \psi_n^{(m)}(z).$$

Since $0 \leq m \leq D_k - 1$, $|\lambda_k/\tau_k| < 1$, $|\tau_k|^{i-n} \leq 1$ for $D_k \leq i \leq n$, and $\|w_{n-i,k}\| \leq w$, then the norm of (22) yields

$$(25) \quad \|b_{n,k}\| \leq w \sum_{m=0}^{D_k-1} |\lambda_k|^{-m} \|w^m\| \frac{1}{m!} \psi^{(m)}(|\lambda_k|/|\tau_k|) \equiv b.$$

Clearly, $b = O(1)$. Take norms in (23) to obtain

$$\|a_{n,k}\| \leq b \left| \frac{\tau_k}{\lambda} \right|^n / C(n, D-\delta).$$

If $|\lambda| > 1$, choose $\tau_k < |\lambda|$; in this case $a_{n,k} = O(n^{-1})$. If $|\lambda| = 1$, then $\delta = 0$; in this case $1/C(n, D-\delta) = O(n^{-1})$ and thus $a_{n,k} = O(n^{-1})$. This establishes part 1 of Lemma 2.

2. Suppose that $1 \leq D_k \leq D-1$. In view of part 1, it is only necessary to prove part 2 when $|\lambda_k| = |\lambda|$. Note that $0 \leq m \leq D_k-1 \leq D-2$.

If $|\lambda| = 1$, then

$$\sum_{i=D_k}^n \frac{C(i,m)}{C(n, D-\delta)} |\lambda|^{i-n} \leq \sum_{i=m}^n \frac{C(i,m)}{C(n, D)} = \frac{C(n+1, m+1)}{C(n, D)} \leq O(n^{m+1-D}) \leq O(n^{-1}).$$

If $|\lambda| > 1$, then

$$\sum_{i=D_k}^n \frac{C(i,m)}{C(n, D-\delta)} |\lambda|^{i-n} \leq \frac{C(n, D-2)}{C(n, D-1)} \cdot (1 + |\lambda|^{-1} + |\lambda|^{-2} + \dots) \leq O(n^{-1}).$$

Thus, for $|\lambda| \geq 1$ and for $0 \leq m \leq D-2$,

$$(26) \quad \sum_{i=D_k}^n \frac{C(i,m)}{C(n, D-\delta)} |\lambda|^{i-n} = O(n^{-1}).$$

Take norms in (21), apply $|\lambda_k| = |\lambda|$, $\|w_{n-i,k}\| \leq w$, and (26) to obtain

$\|a_{n,k}\| \leq O(n^{-1})$. This establishes part 2 of Lemma 2.

3. Suppose that $|\lambda_k| = |\lambda|$ and $D_k = D$. Then (21) becomes

$$\begin{aligned} \underline{q}_{n,k} &= \sum_{m=0}^{D-1} \lambda^{-n} \lambda_k^{n-m} W^m \sum_{i=D}^n \lambda_k^{i-n} \frac{C(i,m)}{C(n,D-\delta)} w_{n-i,k} \\ &= \left(\frac{\lambda_k}{\lambda}\right)^{nD-1} \sum_{m=0}^{nD-1} \lambda_k^{-m} W^m \sum_{i=0}^{n-D} \lambda_k^{-i} \frac{C(n-i,m)}{C(n,D-\delta)} w_{i,k}. \end{aligned}$$

We may write

$$(27) \quad \underline{q}_{n,k} = \left(\frac{\lambda_k}{\lambda}\right)^{nD-1} \sum_{m=0}^{nD-1} \lambda_k^{-m} W^m \underline{f}_{k,m,n}(\lambda_k^{-1})$$

where $\underline{f}_{k,m,n}(z)$ is the vector polynomial

$$(28) \quad \underline{f}_{k,m,n}(z) = \sum_{i=0}^{n-D} z^i \frac{C(n-i,m)}{C(n,D-\delta)} w_{i,k}.$$

Hence

$$(29) \quad \|\underline{f}_{k,m,n}(z)\| \leq w \sum_{i=D}^n |z|^{n-i} \frac{C(i,m)}{C(n,D-\delta)}$$

Let $z = \lambda_k^{-1}$. Since $|\lambda_k| = |\lambda| \geq 1$, we may apply (26) to (29) and obtain

$$\|\underline{f}_{k,m,n}(\lambda_k^{-1})\| = O(n^{-1}) \quad \text{for } 0 \leq m \leq D-2.$$

Thus $\underline{q}_{n,k}$ in (27) is dominated by the term with $m = D-1$. Evaluate (29) with $m = D-1$ for $|z| \leq |\lambda^{-1}|$. If $|\lambda| > 1$, then $\delta = 1$ and

$$\|\underline{f}_{k,D-1,n}(z)\| \leq w \frac{C(n,D-1)}{C(n,D-1)} \sum_{i=0}^n |\lambda|^{i-n}$$

$$\leq w(1+|\lambda|^{-1}+|\lambda|^{-2}+\dots) = \frac{w|\lambda|}{|\lambda|-1}$$

On the other hand if $|\lambda| = 1$, then $\delta = 0$ and

$$\begin{aligned} \|\underline{f}_{k,D-1,n}(z)\| &\leq w \left\{ \sum_{i=D-1}^n C(i,D-1) - 1 \right\} / C(n,D) \\ &= w \{C(n+1,D) - 1\} / C(n,D) \leq w(n+1)/(n+1-D) = O(1). \end{aligned}$$

Thus $\underline{f}_{k,D-1,n}(z)$ is bounded in norm uniformly in n for $|z| \leq |\lambda^{-1}|$. Hence

$$(30) \quad \underline{f}_k(z) \equiv \lim_{n \rightarrow \infty} \underline{f}_{k,D-1,n}(z)$$

is analytic for $|z| \leq |\lambda^{-1}|$. (In fact $\underline{f}_k(z)$ is analytic for $|z| \leq 1$ for $|\lambda| = 1$ and for $|z| < 1$ for $|\lambda| > 1$.) Equation (27) may be written as

$$\underline{q}_{n,k} = (\lambda_k/\lambda)^n \lambda_k^{1-D} W^{D-1} \underline{f}_k(\lambda_k^{-1}) + O(n^{-1}).$$

To complete the proof, take norms, recall that $|\lambda_k| = |\lambda|$, and take the limit as $n \rightarrow \infty$. ■

ACKNOWLEDGMENT

We are indebted to D. Heller, CMU, A. Hindmarsh, LLL, H. T. Kung, CMU, and H. Wozniakowski, CMU and the University of Warsaw, for major suggestions for improving this paper.

BIBLIOGRAPHY

- Brent [72] Brent, R., The Computational Complexity of Iterative Methods for Systems of Nonlinear Equations. Complexity of Computer Computations, IBM Symposium, edited by R. E. Miller and J. W. Thatcher, Plenum Press, New York and London (1972), 61-71.
- Feldstein [69] Feldstein, Alan, Bounds on Order and Ostrowski Efficiency for Interpolatory Iteration Algorithms. Report UCRL-72238, Lawrence Radiation Laboratory, Livermore, California (1969).
- Feldstein and Firestone [67] Feldstein, Alan and Firestone, R. M., Hermite Interpolatory Iteration Theory and Parallel Numerical Analysis. Applied Mathematics Department Report, Brown University, October, 1967.
- Feldstein and Firestone [69] Feldstein, Alan and Firestone, R. M., A Study of Ostrowski Efficiency for Composite Iteration Algorithms. Proc. Nat. Conf. Assoc. Comp. Mach. (1969), 147-155.
- Herzberger [74] Herzberger, J., Über Matrixdarstellungen für Iterationsverfahren bei nichtlinearen Gleichungen. Computing 12 (1974), 215-222.
- Hindmarsh [72] Hindmarsh, A. C., Optimality in a Class of Root-finding Algorithms. SIAM J. Numer. Anal. 9 (1972), 205-214.
- Kung [73] Kung, H. T., A Bound on the Multiplicative Efficiency of Iteration. J. of Computer and System Sciences 7 (1973), 334-342.
- Kung and Traub [74] Kung, H. T. and Traub, J. F., Complexity of Computation. SIAM-AMS Proceedings 7, edited by R. Karp, American Mathematical Society (1974), 149-160.
- Miranker [69] Miranker, W. L., Parallel Methods for Approximating the Root of a Function. IBM J. Res. Develop. 13 (1969) 297-301.
- Ortega and Rheinboldt [70] Ortega, J. M. and Rheinboldt, W. C., Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York and London (1970).

- Ostrowski [66] Ostrowski, A. M. Solution of Equations and Systems of Equations. Academic Press, New York and London (1966).
- Rice [71] Rice, J. R., Matrix Representations of Nonlinear Equations--Application to Parallel Computation. Mathematics of Computation 25 (1971), 639-647.
- Shedler [67] Shedler, G. S., Parallel Numerical Methods for the Solution of Equations. Comm. ACM 10 (1967), 286-291.
- Traub [64] Traub, J. F., Iterative Methods for the Solution of Equations. Prentice-Hall, Englewood Cliffs, N.J. (1964).
- Traub [72] Traub, J. F., Computational Complexity of Iterative Processes. SIAM J. Comput. 1 (1972), 167-179.
- Traub [74] Traub, J. F., An Introduction to Some Current Research in Numerical Computational Complexity. Proceedings of the Conference on the Influence of Computing on Mathematical Research and Education, edited by J. P. Lasalle, American Mathematical Society (1974), 47-55.
- Traub [74a] Traub, J. F., Theory of Optimal Algorithms. Software for Numerical Mathematics, edited by D. J. Evans, Academic Press, London and New York (1974), 1-13.
- Varga [62] Varga, R. S., Matrix Iterative Analysis. Prentice-Hall, Englewood Cliffs, N.J. (1962).
- Wilkinson [65] Wilkinson, J. H., The Algebraic Eigenvalue Problem. Clarendon Press, Oxford (1965).
- Young [71] Young, D. M., Iterative Solution of Large Linear Systems. Academic Press, New York and London (1971).

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ORDER OF VECTOR RECURRENCES WITH APPLICATION TO NONLINEAR ITERATION, PARALLEL ALGORITHMS AND THE POWER METHOD		5. TYPE OF REPORT & PERIOD COVERED Interim
7. AUTHOR(s) Alan Feldstein and J. F. Traub		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS Carnegie-Mellon University Computer Science Dept. Pittsburgh, PA 15213		8. CONTRACT OR GRANT NUMBER(s) N0014-67-A-0314-0010 NR 044-422
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington VA 22217		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE December 1974
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release; Distribution unlimited.		13. NUMBER OF PAGES 36
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
18. SUPPLEMENTARY NOTES		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) <p>The behavior of the vector recurrence $y_{n+1} = My_n + w_{n+1}$ is studied under very weak assumptions. Let $\lambda(M)$ denote the spectral radius of M and let $\lambda(M) \geq 1$. Then if the w_n are bounded in norm and a certain subspace hypothesis holds, the root order of the y_n is shown to be $\lambda(M)$. The behavior of the homogeneous recurrence is studied for all values of λ.</p> <p>These results are applied to the analysis of</p>		

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

- (1) Nonlinear iteration with application to iteration with memory and to parallel iteration algorithms
- (2) Order and efficiency of composite iteration
- (3) The power method.

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)