71-15,747

COHEN, Arthur Ira, 1945-
    RATE OF CONVERGENCE AND OPTIMALITY PROPERTIES
    OF ROOT FINDING AND OPTIMIZATION ALGORITHMS.

    University of California, Berkeley, Ph.D.,
    1970
    Engineering, electrical

University Microfilms, A XEROX Company, Ann Arbor, Michigan

Rate .of Convergence and Optimality Properties of Root Finding
and .Optimization Algorithms

By

Arthur Ira Cohen

B.S. (Cornell University) 1965
M.S. (Cornell University) 1967

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Engineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

. . . .

. . . .

. . . .

Committee in Charge

DEGREE CONFERRED SEPT. 18, 1970

Acknowledgements

Contents

Abstract

A class of optimal root finding algorithms are described which include the usual interpolatory methods such as the Newton-Raphson and secant algorithms. Rate of convergence is found for several specific algorithms. In particular a multipoint secant method is developed which uses no derivatives and yet converges to the root of systems of equations with the same rate as the Newton-Raphson method. Also rate of convergence is found for several conjugate gradient algorithms when they are used to minimize nonlinear, nonquadratic functions on R^n.

Root finding algorithms can be classified by the amount of information or data used at each step. If one assumes infinite precision, it is possible to encode past information (memory) into a data point. We have added a condition to the definition of rate which insures that encoding does not improve the rate. Using this new definition, the class of optimal algorithms is obtained.

The rate of convergence of the multipoint secant method is found with the help of an interpolatory error formula that is developed. The formula compares a function mapping R^n into R^n with an affine interpolation which agrees with the function at n + 1 points.

Rate of convergence of the conjugate gradient algorithms

is found by comparing them with the Newton-Raphson method.

It is shown that n steps of these algorithms approach the

minimum with rate two. This result is also shown to hold

for the case where the conjugate variable is reinitialized every

r steps (where r is greater or equal n).

I. Introduction

The classification, convergence and rate of convergence of root finding and unconstrained optimization algorithms have been studied extensively by Ostrowski[8], Traub[11], and others[6],[7]. These algorithms have been classified by the amount of new and old data (the old data used is called memory) used at each step. Rate of convergence has been defined and has been calculated for many algorithms. It has been assumed in these papers (and it shall be assumed here) that one knows exactly all the data, that is, there are no truncation or other errors. Winograd[†] has pointed out that this can lead to ambiguity in the notion of memory since instead of using old information explicitly at each step, one can use it implicitly by embedding it in other data. Since it is desired to find the best algorithms belonging to a certain class, this ambiguity must be dealt with. One can do this in two ways. The first is by putting strict conditions on the types of algorithms you will allow [11] and the second is by changing the definition of rate so that it will not be affected by implicit information. The second method is used in this paper.

After defining classes of iteration functions and rate of convergence, we will show that certain algorithms are best in a particular class. It turns out that the best or optimal algorithms are not unique, however it will be shown that the standard interpolatory algorithms are among the optimal ones.

[†] Lecture at the University of California, Berkeley during the Fall of 1968.

We will then be concerned with giving rates for specific algorithms. In particular an algorithm, called the multipoint secant method, has been developed which uses no derivatives and yet converges to the root of functions from R^n into R^n with the same rate as the Newton-Raphson algorithm.

Finally, we find the rate of convergence of several conjugate gradient algorithms when they are used to minimize a class of nonlinear, nonquadratic, real valued functions on R^n. Conjugate gradient algorithms use a one-dimensional minimization at each step which, if done exactly, would in general require an infinite number of function evaluations. For this reason we cannot compare it with methods that use a finite amount of data at each step. This section of the thesis is therefore a slight digression from the rest in that it does not attempt to give any optimality properties for these algorithms.

II.1 Notation

The notation used can best be described with the help of an example. Consider the secant method when it is used to find the roots of functions $f \in \mathcal{F}$ mapping R→R. It is

$$x_{k+1} = x_k - \left(\frac{x_k - x_{k-1}}{f(x_k) - f(x_{k-1})} \right) f(x_k) \quad .$$

We define the data at the kth step by $d(f, \underline{x}_k) = (x_k, x_{k-1}, f(x_k), f(x_{k-1}))$, where $\underline{x}_k = (x_k, x_{k-1})$, and let $\mathcal{D}(\mathcal{F})$ be the set of all data that can be obtained from $\mathcal{F}$. We are interested in the asymptotic behavior of the algorithm therefore we assume that the algorithm does not reach the root in a finite number of steps. Thus x_k does not equal x_{k-1} for all k since their being equal would imply that x_k was a root of f.

The algorithm can now be considered as a map Φ from $\mathcal{D}(\mathcal{F})$ into R, that is

$$x_{k+1} = \Phi(d(f, \underline{x}_k)) = \Phi(x_k, x_{k-1}, f(x_k), f(x_{k-1})) \quad .$$

Since the value of f at x_{k-1} needed at the kth step has been already calculated at the (k-1) step, we say that this algorithm has memory.

In general the following notation is used:

(i) x, x_k or x_k^j are points in R^n.

(ii) $\mathcal{F}$ is the set containing the functions whose roots (or extremum) we want to find.

(iii) $\underline{x}_k = (x_k^0, x_k^1, \ldots, x_k^s)$ is an $(s+1)$-tuple called the data point at the kth step. It consists of those points, at the kth step, for which it is necessary to know f. We shall assume all the elements of $\underline{x}_k$ are different.

(iv) $d(f,\underline{x}_k)$ is the data used at the kth step. We sometimes write $d(f,\underline{x}_k) = (d^0(f,\underline{x}_k), \overline{d}(f,\underline{x}_k))$ where $d^0(f,\underline{x}_k) = \underline{x}_k$ and $d(f,\underline{x}_k)$ is the rest of the data.

(v) $\mathcal{D}(\mathcal{F})$ is the set of all data, $d(f,\underline{x}_k)$, that can be obtained from $\mathcal{F}$.

Also:

(vi) $f'(x) = f^{(1)}(x) = Df(x) = D^1 f(x)$ is the derivative of f at x. $D_j f(x)$ is the partial derivative of f with respect to the jth component of x.

(vii) $f^{(j)}(x) = D^j f(x)$ is the jth derivative of f at x. If $f: R^n \to R^m, n>1$, we shall consider $D^j f(x)$ to be a multilinear map from $R^n x \ldots x R^n$ (j times) into R^m defined by

$$D^j f(x)(t_1, \ldots, t_j) = \sum_{i_1=1}^{n} \ldots \sum_{i_j=1}^{n} D_{i_1} \ldots D_{i_j} f(x) t_{1,i_1} \ldots t_{j,i_j}$$

where $t_k = (t_{k,1}, t_{k,2}, \ldots, t_{k,n})$. We also define
$$D^j f(x) t^j = D^j f(x)(\underbrace{t, \ldots, t})_{j \text{ elements}} .$$

(See Dieudonné [3] for justification of this notation).

We shall call $D^j f(x)$ nonsingular if $D^j f(x)(t_1,\ldots,t_j) = 0$

implies $t_k = 0$ for some $k\epsilon\{1,2,\ldots,j\}$.

(viii) $|x|$ denotes the absolute value of x if $x\epsilon R$ and the norm

of x when $x\epsilon R^n$. If the type of norm makes a difference

in a particular theorem then it shall be specified at

that time.

(ix) $B_r(\alpha) = \{x : |x-\alpha|<r\}$.

(x) $co(x_1,\ldots,x_n) = \{y = \sum_{i=1}^{n} \lambda_i x_i : \lambda_i \geq 0 \text{ for } i = 1,2,\ldots,n, \sum_{i=1}^{n} \lambda_i = 1\}$

(xi) $co(\underline{x}_k,z_1,z_2,\ldots,z_n) = co(x_k^0,\ldots,x_k^s,z_1,\ldots,z_n)$.

II.2 Classification of Algorithms

The types of data we will consider fall into four main

classes.

(i) One point: The data is made up of f and its derivatives

evaluated at one point x_k,

(e.g. $d(f,x_k) = (x_k,f(x_k),f^{(1)}(x_k),\ldots,f^{(q-1)}(x_k)))$.

(ii) One point with memory: The data is made up of f and

its derivatives evaluated at one point x_k and of

reused information from past times,

(e.g. $d(f,\underline{x}_k) = (x_k,x_{k-1},\ldots,x_{k-s},f(x_k),f^{(1)}(x_k),$

$\ldots,f^{(\gamma_0-1)}(x_k),f(x_{k-1}),\ldots,f^{(\gamma_1-1)}(x_k),\ldots,f(x_{k-s}),$

$\ldots,f^{(\gamma_s-1)}(x_{k-s})))$.

(iii) Multipoint: The data is made up of f and its derivatives

evaluated at x_k and at other points $z_1(f,x_k),\ldots,z_j(f,x_k)$. The $z_i(f,x_k)$ $i = 1,\ldots,j$ are given functions of f and its derivatives evaluated at x_k.

(iv) Multipoint with memory: The data is made up of evaluations of f and its derivatives at x_k, at other points $z_1(f,x_k),\ldots,z_j(f,x_k)$, and of reused information from past times.

A schematic diagram of an algorithm is given in Figure 1.

II.3 Definition of Algorithm

The properties of an algorithm are dependent on the class of functions $\mathcal{F}$ on which they act. We shall, unless stated otherwise, assume that $\mathcal{F} \subset C^{\infty}(R^n)$, the space of infinitely differentiable functions. This assumption is made to make the discussion more readable; it shall be clear from the context when it can be relaxed. We shall now define what we mead by an algorithm.

(1) Definition: $\Phi: \mathcal{D}(\mathcal{F}) \to R^n$ is an algorithm on $\mathcal{D}(\mathcal{F})$, denoted $\Phi \in \mathcal{A}(\mathcal{D}(\mathcal{F}))$, if for all f in $\mathcal{F}$ there exists a neighborhood T_f of one of the roots of f, $\alpha(f)$, such that for all initial conditions depending on data on T_f, the sequence $x_1, x_2, \ldots$ formed by $x_{i+1} = \Phi(d(f,\underline{x}_i))$ satisfies

(2)   (a) $x_i \in T_f$ for $i = 1,2,\ldots$

(3)   (b) $x_i \to \alpha(f)$ as $i \to \infty$.

We shall call such a sequence a solution sequence for f.

Remark: The initial data, $d(f, \underline{x}_0)$, is made up of f and its

derivatives evaluated at arbitrary distinct points in T_f. So

if, for example, $d(f, \underline{x}_k) = (x_k, x_{k-1}, z(f, x_k), f(x_k), f(x_{k-1}), f(z(f, x_k)))$

then $d(f, \underline{x}_0) = (x^0, x^1, z(f, x^0), f(x^0), f(x^1), f(z(f, x^0)))$ where

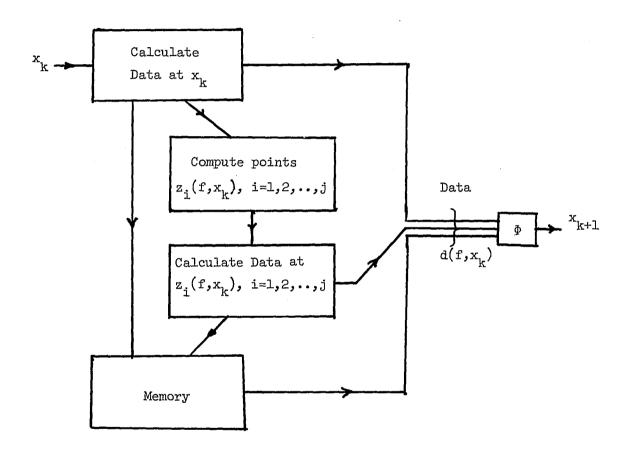x^0 and x^1 are arbitrary distinct elements of T_f.

Figure 1. A Schematic Diagram of an Algorithm

II.4 Implicit Use of Past Data and Rate of Convergence

We can implicitly use at step k past information, not given explicitly in $d(f,\underline{x}_k)$, by encoding it into the data point. For example, let $f:R\to R$ and let Φ be an algorithm on a data space $\mathcal{D}$ containing elements of the form $d(f,\underline{x}_k) = (x_k, x_{k-1}, x_{k-2}, f(x_k), f(x_{k-1}), f(x_{k-2}))$. We shall construct a new algorithm $\tilde{\Phi}$ which acts essentially the same as Φ but is defined on a data space $\widetilde{\mathcal{D}}$ containing elements of the form $\tilde{d}(f,\underline{x}_k) = (x_k, x_{k-1}, f(x_k), f(x_{k-1}))$. This shall be done with help of a map h from $R \times R \times R_u \times R_u \to R$ which we shall denote by $h(a,b,c,d) = [a.b.c.d]$ (where $R_u = \{x \in R : x<1\}$). H encodes the three points a,b,c, into one point. To define h exactly we must express a,b,c, and d in decimal notation:

$$a = \text{sgn}(a)\ a_p a_{p-1}\cdots a_0 . a_{-1} a_{-2}\cdots$$
$$b = \text{sgn}(b)\ b_q b_{q-1}\cdots b_0 . b_{-1} b_{-2}\cdots$$
$$c = \text{sgn}(c)\qquad\quad 0.0\ldots 0c_{-r} c_{-r-1}\cdots$$
$$d = \text{sgn}(d)\qquad\quad 0.0\ldots 0d_{-s} d_{-s-1}\cdots$$

Then define

$$[a.b.c.d] = \text{sgn}(a) a_p a_{p-1}\cdots a_0 . a_{-1}\cdots a_{-ms} b_{q+1} c_{r+1} a_{-ms-1} b_q c_r \cdots$$

where b_{q+1} (or c_{r+1}) is 1 if b (or c) is positive and is 0 if b(or c) is negative, and where m is specified below. Note that

if we are given [a.b.c.d] and d, we can exactly determine

a,b, and c. We now define

$$\tilde{\Phi}(x_k,x_{k-1},f(x_k),f(x_{k-1})) = [\Phi(a_k,x_{k-1},b_k,f(x_k),f(x_{k-1}),c_k).x_{k-1}.$$
$$f(x_{k-1}).f(x_k)]$$

where $x_k = [a_k.b_k.c_k.d_k]$. So if (x_2,x_1,x_0) are the initial

points used for Φ , producing the sequence $\{z_k\}$, and if

$([x_2.x_0.f(x_0).f(x_1)],x_1)$ is used for $\tilde{\Phi}$, producing the

sequence $\{x_k\}= \{[a_k.b_k.c_k.d_k]\}$, then $z_k = a_k$ for all k and

$|z_k - x_k| \leq 10^{-ms}$. So if $z_k \to \alpha$ with power p (power is defined

later in this section) we can, since $f(x_k) \to 0$, make m large

enough to insure that $x_k \to \alpha$ with power p also.

Clearly $\tilde{\Phi}$ cannot be compared to an algorithm such as the

secant method which does not use encoding. Therefore, in

order to answer "What is the best algorithm given a particular

amount of data?", we must eliminate from consideration

algorithms that encode. The most direct way to accomplish

this is to just consider algorithms that do not encode;

however, it is not clear how to mathematically describe this

class. We shall instead add a condition to the definition

of rate which insures that encoding does not increase the

rate.

The rate or power of an algorithm $\Phi \in \mathcal{A}(\mathcal{D}(\mathcal{F}))$ is

defined by Ostrowski [8] as the largest constant p>1

such that for all solution sequences $\{x_k\}$ and for all $f \in \mathcal{F}$

there exists a constant C so that

$$\varlimsup_{k \to \infty} \frac{|\Phi(d(f,x_k)) - \alpha(f)|}{|x_k - \alpha(f)|^p} \leq C < \infty \quad .$$

The condition we shall add requires that the error

$|x_k - \alpha(f)|$ is decreased by the same order for a large class

of functions whose data at step k is the same (see part (iii)

in the definition below). This will insure that information

not given explicitly in d will not increase the rate. It

should be emphasized that encoding does indeed increase the

speed at which an algorithm can converge and so there is

nothing wrong with the original definition; it prevents us,

however, from comparing algorithms meaningfully.

The new definition is:

(4) Definition: $\Phi \in \mathcal{A}(\mathcal{D}(\mathcal{F}))$ has uniform rate or

uniform power p if:

(i) for all $f \in \mathcal{F}$, there exists a constant C_1 such that

(5) $$\varlimsup_{k \to \infty} \frac{|\Phi(d(f,x_k)) - \alpha(f)|}{|x_k - \alpha(f)|^p} \leq C_1 < \infty$$

where $\{x_k\}$ is any solution sequence on f,

(ii) there exists an $h \in \mathcal{F}$ and a solution sequence,

$\{z_k\}$, on h such that

$$(6) \qquad \varliminf_{k \to \infty} \frac{|\Phi(d(h, \underline{z}_k)) - \alpha(h)|}{|z_k - \alpha(h)|^p} > 0 \quad ,$$

(iii) for all $h \in \mathcal{F}$ and solution sequences $\{z_k\}$ on h for which (6) occurs and for all sequences of functions $\{g_k\}$ in $\mathcal{F}$ satisfying

$(7) \qquad$ (a) $d(g_k, \underline{z}_k) = d(h, \underline{z}_k)$

$(8) \qquad$ (b) $g_k \to h$ uniformly at $\alpha(h)$, that is, for all $\varepsilon > 0$ there exists $\delta > 0$ and K such that $k \geq K$ and $a \in B_\delta(\alpha(h))$ implies

$$|g_k(z) - h(z)| + |Dg_k(z) - Dh(z)| < \varepsilon$$

(c) $\{D^j g_k(\xi_k)\}$ is a Cauchy sequence for all $j = 1, 2, \ldots$ and sequences $\{\xi_k\}$ where $\xi_k \in co(\underline{z}_k, \alpha(h))$,

there exists a constant $C_2(h)$ and roots $\alpha(g_k)$ of g_k such that

$$(9) \qquad \varlimsup_{k \to \infty} \frac{|\Phi(d(h, \underline{z}_k)) - \alpha(g_k)|}{|z_k - \alpha(g_k)|^p} \leq C_2 < \infty \quad .$$

(10) Remark: Condition (c) is required to insure that the new definition of rate does not break down when we are analyzing algorithms with memory. If we were just considering memoryless algorithms it would not be required.

(11) Remark: If we assume that $Dh(\alpha(h))$ is invertible, condition (b)

implies that g_k has a root $\alpha(g_k)$ such that $|z_k - \alpha(g_k)|$

goes to zero with the same rate as $|z_k - \alpha(h)|$. To be precise:

(12) Lemma: Suppose that for all functions $f \in \mathcal{F}$, $f \in C^1(R^n)$,

and $Df(\alpha(f))$ is invertible. Let $\{x_k\}$ be any sequence converging

to $\alpha(f)$ and let $\{g_k\}$ be a sequence of functions such that

$g_k(x_k) = f(x_k)$ and $g_k \to f$ uniformly at $\alpha(f)$, then there exists

m, M, K and a root $\alpha(g_k)$ of g_k such that $k \geq K$ implies

$$0 < m \leq \frac{|x_k - \alpha(f)|}{|x_k - \alpha(g_k)|} \leq M < \infty \quad .$$

The proof of this Lemma is found in Appendix A. This result

is used in many of our proofs, therefore we shall usually

assume that $Df(\alpha(f))$ is invertible.

II.5 Sufficient Algorithms

We shall now construct optimal algorithms. An algorithm

in $\mathcal{A}(\mathcal{D}(\mathcal{F}))$ will be called optimal if its power is greater

or equal the power of any other algorithm in $\mathcal{A}(\mathcal{D}(\mathcal{F}))$.

We will begin by looking at a particular subset of $\mathcal{F}$.

(13) Definition: $\widehat{\mathcal{F}} \subset \mathcal{F}$ is a sufficient class of functions for

$\mathcal{D}(\mathcal{F})$ if for all $\omega \in \mathcal{D}(\mathcal{F})$, there exists a unique

$\hat{f}_\omega \in \widehat{\mathcal{F}}$ such that $d(\hat{f}_\omega, \omega^0) = \omega$. In addition we require that:

(i) $\hat{f}_{d(f, \underline{x}_i)}$ converges pointwise to the same function

for all sequences of data points $\underline{x}_i \to \underline{x}$. If $\underline{x}$ is a data

point then $\hat{f}_{d(f,\underline{x}_i)} \to \hat{f}_{d(f,\underline{x})}$. If $\underline{x}$ is not a data point

then we define $\hat{f}_{d(f,\underline{x})} = \lim_{i \to \infty} \hat{f}_{d(f,\underline{x}_i)}$.

(ii) For all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\left| \hat{f}_{d(f,\underline{x})}(x) - f(x) \right| + \left| D\hat{f}_{d(f,\underline{x})}(x) - Df(x) \right| < \varepsilon$$

for all x and $\underline{x}$ such that $co(x,\underline{x}) \subset B_\delta(\alpha(f))$.

(iii) $D^j \hat{f}_{d(f,\underline{x})}(x)$ is continuous in $(\underline{x}, x)$ at $(\underline{\alpha(f)}, \alpha(f))$

for $j = 1, 2, \ldots$, where $\underline{\alpha(f)} = (\alpha(f), \ldots, \alpha(f))$.

(14) Remark: If $\{x_k\}$ is a solution sequence for f, then

$\{\hat{f}_{d(f,\underline{x}_k)}\}$ satisfies conditions (a), (b), and (c) required of

$\{g_k\}$ in definition of rate.

If we consider $\hat{f}_{d(f,\underline{x}_k)}$ as an approximation to f, then

we arrive naturally at the following method of obtaining

the root of f.

(15) Definition: $\hat{\Phi} \in \mathcal{A}(\mathcal{D}(\mathcal{F}))$ is a sufficient algorithm

if there exists a sufficient class of functions $\hat{\mathcal{F}}$ such that

$\hat{\Phi}(\omega) = \alpha(\hat{f}_\omega)$ for all $\omega \in \mathcal{D}(\mathcal{F})$, where $\alpha(\hat{f}_\omega)$ is some root of $\hat{f}_\omega$.

(16) Example: If $\mathcal{D}(\mathcal{F})$ contains data of the form

$\omega_0 = (x_0, f(x_0), Df(x_0))$ where $Df(x_0)$ is assumed invert ble,

then we can define the function $\hat{f}_{\omega_0}(x) = f(x_0) + Df(x_0)(x - x_0)$.

$\hat{\mathcal{F}} = \{\hat{f}_\omega : \omega \in \mathcal{D}(\mathcal{F})\}$ will be a sufficient class of

functions and $\hat{\Phi}(\omega_0) = x_0 - (Df(x_0))^{-1} f(x_0)$ will be a sufficient

algorithm. $\hat{\Phi}$ is the Newton-Raphson algorithm.

II.6 Optimal Sufficient Algorithms

Under certain conditons sufficient algorithms are optimal.
To show this, it is first necessary to give the relationship
between the data $\underline{x}_k = (x_k^0, \ldots, x_k^n)$ and x_{k+1} for a sufficient
algorithm used to find roots of functions from $R \to R$.

(17) Lemma: Let $\mathcal{F} \subset C^\infty(R)$, let $d(f,\underline{x}_k) = (x_k^0, \ldots, x_k^n,$
$f(x_k^0), \ldots, f^{(\gamma_0-1)}(x_k^0), f(x_k^1), \ldots, f^{(\gamma_n-1)}(x_k^n))$, and let $\widehat{\mathcal{F}}$ be
a sufficient class of functions for $\mathcal{D}(\mathcal{F})$. Let $\hat{\Phi}$ be
defined on $\widehat{\mathcal{F}}$, and let $\hat{f}_k \equiv \hat{f}_{d(f,\underline{x}_k)}$. Then there exists,
for each $t_0 \in R$, a $\xi \in co(\underline{x}_k, t_0)$ such that

$$(18) \qquad f(t_0) = \hat{f}_k(t_0) + \frac{1}{q!}\left[f^{(q)}(\xi) - \hat{f}_k^{(q)}(\xi) \right] \quad w(t_0)$$

where $q = \sum_{j=0}^{n} \gamma_j$ and $w(t) = \prod_{j=0}^{n} (t-x_k^j)^{\gamma_j}$. Also if $\hat{f}_k'(x) \neq 0$ on
$co(x_{k+1}, \alpha)$, then there exists an $\eta_{k+1} \in co(x_{k+1}, \alpha)$ and
$\xi_k \in co(\underline{x}_k, \alpha)$ such that

$$(19) \qquad |x_{k+1} - \alpha| = H_{k+1} \prod_{j=0}^{n} |x_k^j - \alpha|^{\gamma_j}$$

where $H_{k+1} = \frac{1}{q!} \left| \dfrac{f^{(q)}(\xi_k) - \hat{f}_k^{(q)}(\xi_k)}{\hat{f}_k'(\eta_{k+1})} \right|$.

Proof: Let $g(t) = f(t) - \hat{f}_k(t) - Mw(t)$ where M is chosen
so that

$$(20) \qquad f(t_0) = \hat{f}_k(t_0) + Mw(t_0), \text{ for some } t_0 \neq x_k^0, x_k^1, \ldots, x_k^n.$$

Clearly $g^{(1_j)}(x_k^j) = 0$ for $j = 0,1,\ldots,n$; $1_j = 0,1,\ldots,\gamma_j-1$, since $d(f,\underline{x}_k) = d(\hat{f}_k,\underline{x}_k)$. Also $g(t_0) = 0$, so by the mean value theorem there exists n+1 points $x_1^i \, \varepsilon \, co(x_k^i, x_k^{i+1})$, $i = 0,1,\ldots,n-1$, and $x_1^n \, \varepsilon \, co(x_k^n, t_0)$ such that $g'(x_1^i) = 0$ for $i = 0,1,\ldots,n$. Repeating this process, using points at which g' is zero, we get points where $g^{(2)}$ is zero. If we keep doing this, we eventually see that there exists a $\xi \, \varepsilon \, co(x_k^0, x_k^1, \ldots, x_k^n, t_0)$ such that $g^{(q)}(\xi) = 0$, where $q = \sum_{j=0}^{n} \gamma_j$. But

$$g^{(q)}(\xi) = f^{(q)}(\xi) - \hat{f}_k^{(q)}(\xi) - Mw^{(q)}(\xi)$$

and $w^{(q)}(\xi) = 1/q!$. Thus $M = \frac{1}{q!}\left(f^{(q)}(\xi) - \hat{f}_k^{(q)}(\xi)\right)$, or from (20)

$$f(t_0) = \hat{f}_k(t_0) + \frac{1}{q!}\left(f^{(q)}(\xi) - \hat{f}_k^{(q)}(\xi)\right) w(t_0).$$

To prove (19), let $t_0 = \alpha(f) = \alpha$, then

$$(21) \qquad \hat{f}_k(\alpha) = \frac{-1}{q!}\left(f^{(q)}(\xi_k) - \hat{f}_k^{(q)}(\xi_k)\right) w(\alpha),$$

where $\xi_k \, \varepsilon \, co(x_k^0, x_k^1, \ldots, x_k^n, \alpha)$. Now if $\hat{f}_k$ has a zero α_k,

$$\hat{f}_k(\alpha) = \hat{f}_k^!(\eta_{k+1}) \ (\alpha - \alpha_k)$$

where $\eta_{k+1} \in co(\alpha, \alpha_k)$. By definition $\hat{\Phi}(d(f, \underline{x}_k)) = x_{k+1} = \alpha_k$, so

$$x_{k+1} - \alpha = -\hat{f}_k(\alpha)/\hat{f}_k^!(\eta_{k+1}) \quad .$$

Using (21) we get

$$x_{k+1} - \alpha = \frac{1}{q!} \left[\frac{f^{(q)}(\xi_k) - \hat{f}_k^{(q)}(\xi_k)}{\hat{f}_k^!(\eta_{k+1})} \right] \prod_{j=0}^{n} (\alpha - x_k^j)^{\gamma_j}$$

from which (19) follows.

We are now ready to give conditions under which sufficient algorithms are optimal. We shall first consider the case where we are finding roots of functions from $R \to R$ and where the data is one-point with memory.

(22) Theorem: Let $\mathcal{D}(\mathcal{F})$ consist of data of the form $d(f, \underline{x}_k) = (x_k, x_{k-1}, \ldots, x_{k-n}, f(x_k), Df(x_k), \ldots, D^{\gamma_0 - 1} f(x_k), f(x_{k-1}), \ldots, D^{\gamma_n - 1} f(x_{k-n}))$ and let $q = \sum_{j=0}^{n} \gamma_j$. Suppose $f: R \to R$ $|D^q f(x)| \leq M$, and $Df(\alpha(f))$ is invertible for all $f \in \mathcal{F}$ and $x \in T_f$. Let $\hat{\Phi} \in \mathcal{A}(\mathcal{D}(\mathcal{F}))$ be a sufficient algorithm with power $\hat{p}$ defined on the sufficient class $\hat{\mathcal{F}}$ and let Φ be another algorithm in $\mathcal{A}(\mathcal{D}(\mathcal{F}))$ with power p. If there exists a function h_1 [h_2] and a solution sequence $\{x_k\}$ [$\{z_k\}$] on h_1 [h_2] formed by $\hat{\Phi}$ [Φ] such that

$$(23) \qquad \varliminf_{k\to\infty} \frac{|x_{k+1} - \alpha(h_1)|}{|x_k - \alpha(h_1)|^{\hat{p}}} \geq \hat{C}_1 > 0$$

$$(24) \qquad \lim_{k\to\infty} D^q \hat{f}_{d(h_1,\underline{x}_k)}(\alpha(h_1)) \neq D^q h_1(\alpha(h_1)) \quad ,$$

$$(25) \qquad \varliminf_{k\to\infty} \frac{|z_{k+1} - \alpha(h_2)|}{|z_k - \alpha(h_2)|^p} \geq C_1 > 0$$

and

$$(26) \qquad \lim_{k\to\infty} D^q \hat{f}_{d(h_2,\underline{z}_k)}(\alpha(h_2)) \neq D^q h_2(\alpha(h_2)) \quad .$$

then $p \leq \hat{p}$.

Proof: Since $Dh_1(\alpha(h_1))$ is invertible, property (ii) in Definition (13) of sufficient class implies there exists a K such that $D\hat{f}_{d(h_1,\underline{x}_k)}(\xi) \neq 0$ for $\xi \in co(x_{k+1},\alpha(h_1))$ if $k \geq K$. We can use Lemma (17) to obtain

$$(27) \qquad |x_{k+1} - \alpha(h_1)| = \hat{H}_{k+1} \prod_{j=0}^{n} |x_{k-j} - \alpha(h_1)|^{\gamma_j}$$

where $q = \sum_{j=0}^{n} \gamma_j$, $\hat{H}_{k+1} \doteq \frac{1}{q!} \left| \frac{D^q h_1(\hat{\xi}_k) - D^q \hat{f}_{d(h_1,\underline{x}_k)}(\hat{\xi}_k)}{D\hat{f}_{d(h_1,\underline{x}_k)}(\hat{\eta}_{k+1})} \right|$

$\hat{\xi}_k \in co(\underline{x}_k,\alpha(h_1))$ and $\hat{\eta}_{k+1} \in co(x_{k+1},\alpha(h_1))$. Also since $\hat{\Phi}$ has rate $\hat{p}$,

$$(28) \qquad \varlimsup_{k\to\infty} \frac{|x_{k+1} - \alpha(h_1)|}{|x_k - \alpha(h_1)|^{\hat{p}}} \leq \hat{C}_2 < \infty \qquad .$$

(23) and (28) imply there exists an $\hat{L}_1$ and a sequence $\{\hat{C}_k\}$ such that

$$(29) \qquad |x_{k+1} - \alpha(h_1)| = \hat{C}_k \, |x_k - \alpha(h_1)|^{\hat{p}}$$

where $\hat{C}_k$ is bounded away from zero and infinity for $k \geq \hat{L}_1$. By (24) and our assumptions on $\mathcal{F}$, there exists an $\hat{L}_2$ such that $\hat{H}_k$ is also bounded away from zero and infinity for $k \geq \hat{L}_2$. Now from (29), for $j = 0,1,\ldots,n$,

$$|x_{k-j} - \alpha(h_1)|^{\gamma_j} = \hat{C}_{k-j-1}^{\gamma_j} \, |x_{k-j-1} - \alpha(h_1)|^{\hat{p}\gamma_j}$$

$$= \hat{G}_{k,j} \, |x_{k-n} - \alpha(h_1)|^{\gamma_j \hat{p}^{n-j}}$$

So using (27), we get for $k \geq \hat{L} = \max(\hat{L}_1, \hat{L}_2) + n$,

$$(30) \qquad \frac{|x_{k+1} - \alpha(h_1)|}{|x_k - \alpha(h_1)|^{\hat{p}}} = \hat{G}_k \, |x_{k-n} - \alpha(h_1)|^{g(\hat{p})} \qquad ,$$

where $g(t) = \sum_{j=0}^{n} \gamma_j t^{n-j} - t^{n+1}$; because of our bounds on $\hat{H}_k$ and $\hat{C}_k$, $\hat{G}_k$ is bounded away from zero and infinity for $k \geq \hat{L}$. In order for (23),(28) and (30) to be compatible it must be the case that $g(\hat{p}) = 0$.

Now consider

$$\frac{|\hat{\Phi}(d(h_2, z_k)) - \alpha(h_2)|}{|z_k - \alpha(h_2)|^p} \qquad .$$

Again by Lemma (17) we can write

$$(31) \qquad |\hat{\Phi}(d(h_2,\underline{z}_k)) - \alpha(h_2)| = H_{k+1} \prod_{j=0}^{n} |z_{k-j} - \alpha(h_2)|^{\gamma_j}$$

$$\text{where } H_{k+1} = \frac{1}{q!} \left| \frac{D^q h_i(\xi_k) - D^q \hat{f}_{d(h_2,\underline{z}_k)}(\xi_k)}{D\hat{f}_{d(h_2,\underline{z}_k)}(\eta_{k+1})} \right| \quad .$$

Since Φ has power p, there exists a constant C_2 such that

$$(32) \qquad \overline{\lim_{k\to\infty}} \frac{|z_{k+1} - \alpha(h_2)|}{|z_k - \alpha(h_2)|^p} \leq C_2 < \infty \quad .$$

Using (26), (31) and (32) and the same arguments we used to obtain (30) we have that there exists an L such that for $k \geq L$,

$$(33) \qquad \frac{|\hat{\Phi}(d(h_2,\underline{z}_k)) - \alpha(h_2)|}{|z_k - \alpha(h_2)|^p} = G_k |z_{k-n} - \alpha(h_2)|^{g(p)}$$

where G_k is bounded away from zero and infinity for $k \geq L$.

By Descartes' rule of signs, $\hat{p}$ is the only real positive root of g; so since $g(t)\to-\infty$ as $t\to+\infty$, we have by continuity that $p>\hat{p}$ implies that $g(p)<0$. Therefore $p>\hat{p}$ implies, by (33), that

$$\overline{\lim_{k\to\infty}} \frac{|\hat{\Phi}(d(h_2,\underline{z}_k)) - \alpha(h_2)|}{|z_k - \alpha(h_2)|^p} \to +\infty \quad .$$

But

$$(34) \quad \varlimsup_{k \to \infty} \frac{|\hat{\Phi}(d(h_2, \underline{z}_k)) - \alpha(h_2)|}{|z_k - \alpha(h_2)|^p} \leq \varlimsup_{k \to \infty} \frac{|z_{k+1} - \alpha(h_2)|}{|z_k - \alpha(h_2)|^p}$$

$$+ \varlimsup_{k \to \infty} \frac{|z_{k+1} - \hat{\Phi}(d(h_2, \underline{z}_k))|}{|z_k - \alpha(h_2)|^p}$$

Since Φ has power p and since $\hat{\Phi}(d(h_2, \underline{z}_k))$ has the properties required of $\alpha(g_k)$ in the definition of rate, both terms on the right hand side are finite. We therefore have a contradiction.

(35) Remark: For interpolatory polynomials $D^q f_{d(\hat{f}, \underline{x}_k)}(x) \equiv 0$ for all $f \in \mathcal{F}$ and so (24) and (26) are trivially satisfied if $D^q h_i(\alpha(h_i)) \neq 0$, $i = 1, 2$.

This theorem can easily be extended to multipoint iteration functions with memory.

(36) Theorem: Let $\hat{\Phi} \in \mathcal{A}(\mathcal{D}(\mathcal{F}))$, with power $\hat{p}$, be a sufficient algorithm defined on the sufficient set $\hat{\mathcal{F}}$, where $\mathcal{D}(\mathcal{F})$ consists of data of the form $d(f, \underline{x}_k) = (x_k, z_1(f, x_k), \ldots, z_{s_0}(f, x_k), x_{k-1}, z_1(f, x_{k-1}), \ldots, z_{s_1}(f, x_{k-1}),$

$\ldots, z_{s_n}(f, x_{k-n}), f(x_k), \ldots, f^{(\gamma_0^0 - 1)}(x_k), f(z_1(f, x_k)), \ldots,$

$f^{(\gamma_0^1 - 1)}(z_1(f, x_k)), \ldots, f(z_{s_0}(f, x_k)), \ldots, f^{(\gamma_0^{s_0 - 1})}(z_{s_0}(f, x_k)),$

$\ldots, f^{(\gamma_n^{s_n - 1})}(z_{s_n}(f, x_{k-n})))$. Suppose for each $j = 1, 2, \ldots, s_0$, and $f \in \mathcal{F}$, $z_j(f, x)$ has the properties:

(37)   (i) there exists a data set $\widetilde{\mathcal{D}}(\mathcal{F})$ containing data

of the form $\tilde{d}(f,x) = (x, f(x), f^{(1)}(x), \ldots, f^{(t)}(x))$

such that if $f, g \in \mathcal{F}$ then $\tilde{d}(f,x) = \tilde{d}(g,x)$ implies $z_j(f,x) = z_j(g,x)$.

(ii) there exists an $M_j(f,x)$ continuous in x with the property that $|M_j(f,x))| \leq M < \infty$ for all $f \in \mathcal{F}$ and $x \in T_f$ and a positive integer β_j such that

$$(39) \qquad |z_j(f,x) - \alpha(f)| = M_j(f,x) \, |x - \alpha(f)|^{\beta_j} .$$

$$(40) \qquad (iii) \ x \neq z_j(f,x), \text{ for all } j = 1,2,\ldots, s_0 , \ f \in \mathcal{F}$$

and $x \in T_f - \{\alpha(f)\}$. Also, if for $f \in \mathcal{F}$

and $x \in T_f$, $z_j(f,x) \neq \alpha(f)$, for all j, then

$z_j(f,x) \neq z_i(f,x)$ for $i \neq j$, $i,j = 1,2,\ldots,s_0$.

Suppose

(a) the assumptions on $\mathcal{F}$ and $\hat{\mathcal{F}}$ given in Theorem (22) hold with $q = \sum\limits_{j=0}^{n} \sum\limits_{i=0}^{s_j} \gamma_j^i$,

(b) for the h_1, h_2 satisfying (23), (24), (25), and (26), $M_j(h_i,x)$ is bounded away from zero for all $x \in T_{h_i}$, for all j, $i = 1,2$,

then any other algorithm Φ will have power $p \leq \hat{p}$.

Proof: By Lemma (17)

$$(41) \qquad |x_{k+1} - \alpha(h_1)| = H_{k+1} \prod_{j=0}^{n} \prod_{i=0}^{s_j} |z_i(h_1, x_{k-j}) - \alpha(h_1)|^{\gamma_j^i}$$

where $z_0(h_1, x_{k-j}) \equiv x_{k-j}$ and $H_{k+1} = \frac{1}{q!} \left| \dfrac{D^q h_1(\xi_k) - D^q \hat{f}_{d(h_1, \underline{x}_k)}(\xi_k)}{D\hat{f}_{d(h_1, \underline{x}_k)}(\eta_{k+1})} \right|$,

$\xi_k \in co(\underline{x}_k, \alpha(h_1)), \eta_{k+1} \in co(x_{k+1}, \alpha(h_1))$. Using (39), (41) becomes

$$(42) \qquad |x_{k+1} - \alpha(h_1)| = H_{k+1} \prod_{j=0}^{n} \prod_{i=0}^{s_j} \left(M_i(h_1, x_{k-j}) \, |x_{k-j} - \alpha(h_1)|^{\beta_i} \right)^{\gamma_j^i}$$

$$= \tilde{H}_{k+1} \prod_{j=0}^{n} |x_{k-j} - \alpha(h_1)|^{\delta_j}$$

where $\tilde{H}_{k+1} = H_{k+1} \prod_{j=0}^{n} \prod_{i=0}^{s_j} [M_i(h_1, x_{k-j})]^{\gamma_j^i}$, $\beta_0 = 1$, $M_0 = 1$

and $\delta_j = \sum_{i=0}^{s_j} \beta_i \gamma_j^i$.

By our assumptions on $\mathcal{F}$, $\hat{\mathcal{F}}$ and M_i, $\tilde{H}_{k+1}$ is bounded away from zero and infinity for large k. Thus, (42) is of the same form as Equation (27). We can therefore again show, as in Theorem (22), that $\hat{p}$ must be the positive real root of

$\sum_{j=0}^{n} \delta_j t^{n-j} - t^{n+1}$. We can now proceed by exactly the same steps used in Theorem (22) to again arrive at a contradiction if we assume that $p > \hat{p}$.

The most obvious type of function $z(f,x)$ that satisfies (37), (38), and (40) is a sufficient algorithm defined on data depending just on x. However $z(f,x) = x + f(x)$, for

example, also works if we assume $|Df(x)| \leq R < \infty$ for all $f \in \mathcal{F}$, and $x \in T_f$. Clearly $z(f,x)$ satisfies (37). Also

$$\begin{aligned}
|z(f,x) - \alpha| &= |(x - \alpha) + f(x)| \\
&= |Df(\xi(x)) (x - \alpha) + (x - \alpha)| \text{ where } \xi(x) \in co(x,\alpha) \\
&= |(1 + Df(\xi(x))) (x - \alpha)| \\
&= M(f,x) |x - \alpha|.
\end{aligned}$$

$M(f,x)$ satisfies the conditions in (38). Also (40) is satisfied since for $x \in T_f$, $z(f,x) = \alpha$ if and only if $f(x) = 0$ if and only if $x = \alpha$.

II.7 The Power Of Sufficient Algorithms

This section gives conditions on $\mathcal{F}$ and $\hat{\mathcal{F}}$ to insure that algorithms defined on a sufficient class are (i) indeed algorithms, that is, are in $\mathcal{A}(\mathcal{D}(\mathcal{F}))$, and (ii) have certain power. We shall again consider the case where we are finding roots of functions from $R \to R$ and where the data is one-point memory.

(43) Theorem (Traub [11]): Let $\mathcal{D}(\mathcal{F})$ consist of data of the form
$d(f,\underline{x}_k) = (x_k, \ldots, x_{k-n}, f(x_k), \ldots, f^{(\gamma_0 - 1)}(x_k), \ldots, f(x_{k-n}),$
$\ldots, f^{(\gamma_n - 1)}(x_{k-n}))$ and let $q = \sum_{j=0}^{n} \gamma_j$. Suppose $f: R \to R$,
$|D^q f(x)| \leq R < \infty$ and $|f'(\alpha(f))| \neq 0$ for all $f \in \mathcal{F}$, and
$x \in R$. Let $\hat{\mathcal{F}}$ be a sufficient class, then there exists an

algorithm $\hat{\Phi}$ defined on it that is in $\mathcal{A}(\mathcal{D}(\mathcal{F}))$.

Proof: We need to find, for all $f \in \mathcal{F}$, an interval T_f containing $\alpha(f)$ such that $x_k, \ldots, x_{k-n} \in T_f$ implies (i) $x_{k+1} \in T_f$ and (ii) $x_k \to \alpha(f)$. Pick $f \in \mathcal{F}$ and let $\alpha \equiv \alpha(f)$ and $\hat{f}_k \equiv \hat{f}_{d(f, \underline{x}_k)}$, then to prove (i) we must show that $\hat{f}_k$ has a zero on T_f. Let $x \in \{x : |x - \alpha| < \Gamma_0\} = J_0$ imply $f'(x) \neq 0$ on J_0. Then α is the only zero on J_0. If the x_{k-j}, $j = 0, 1, \ldots, n$, bracket α, then there exists a root of $\hat{f}_k$ in J_0. This follows from the fact that $f'(x) \neq 0$ for $x \in J_0$ and that $f(x_{k-j}) = \hat{f}_k(x_{k-j})$ for $j = 0, 1, \ldots, n$. So suppose the x_{k-j} do no bracket α. Without loss of generality, let us assume $f'(x) \geqslant r_f > 0$ on J_0 and $\alpha < x_{k-j}$ for all $j = 0, 1, \ldots, n$. By Lemma (17)

$$(44) \qquad \hat{f}_k(\alpha - \Gamma) = f(\alpha - \Gamma) - \frac{1}{q!} [f^{(q)}(\xi) - \hat{f}_k^{(q)}(\xi)] \, w(\alpha - \Gamma)$$

where $\Gamma_0 > \Gamma > 0$, $\xi \in co(\underline{x}_k, \alpha - \Gamma)$ and $w(\alpha - \Gamma) = \prod_{j=0}^{n} (\alpha - \Gamma - x_{k-j})^{\gamma_j}$.

Let $J_\Gamma = \{x : |x - \alpha| \leq \Gamma\}$ and let $J_{\Gamma_0} \subset J_0$. Now

$$f(\alpha - \Gamma_0) = f(\alpha) - f'(\eta)\Gamma_0 = -f'(\eta)\Gamma_0$$

where $\eta \in co(\alpha, \alpha - \Gamma_0)$. Also if $x_{k-j} \in J_{\Gamma_0}$, for $j = 0, 1, \ldots, n$ then $w(\alpha - \Gamma_0) \leq \Gamma_0^q$, or

$$\left| \frac{1}{q!} \left[f^{(q)}(\xi) - \hat{f}_k^{(q)}(\xi) \right] w(\alpha - \Gamma) \right| \leq \frac{2R}{q!} \Gamma_0^q \quad .$$

Therefore, if we pick Γ_0 small enough so that $\left| f'(\eta) \Gamma_0 \right| \geq$
$\Gamma_0 r_f \geq \frac{2R}{q!} \Gamma_0^q$, then $\hat{f}_k(\alpha - \Gamma_0)$ will be negative by (44). Since
$\hat{f}_k(x_k) = f(x_k) > 0$, this implies the existence of a root
in J_{Γ_0}.

To prove (ii) we again use Lemma (17), to get

$$(45) \qquad \left| x_{k+1} - \alpha \right| = H_{k+1} \prod_{j=0}^{n} \left| x_{k-j} - \alpha \right|^{\gamma_j}$$

where $H_{k+1} = \frac{1}{q!} \left| \dfrac{f^{(q)}(\xi_k) - \hat{f}_k^{(q)}(\xi_k)}{\hat{f}_k'(\eta_{k+1})} \right| \quad ,$

$\xi_k \in co(\underline{x}_k, \alpha)$ and $\eta_{k+1} \in co(x_{k+1}, \alpha)$. For (45) to hold, we require
$\hat{f}_k'(x) \neq 0$ on $co(x_{k+1}, \alpha)$. Since $\hat{\mathcal{F}}$ is a sufficient class,
there exists a $\Gamma_1 < \Gamma_0$ such that $\left| \hat{f}_k'(x) - f'(x) \right| \leq \frac{r_f}{2}$ for all
$x \in J_{\Gamma_1}$; this implies $\hat{f}_k'(x) > \frac{r_f}{2}$ on J_{Γ_1}. Therefore (45) holds
for $co(\underline{x}_k, \alpha) \subset J_{\Gamma_1}$. Now $H_{k+j} \leq \frac{1}{q!} \frac{4R}{r_f} \equiv H$ for all j. Let
$\Gamma_2 \leq \Gamma_1$ be small enough so that $H\Gamma_2^{q-1} \equiv P < 1$, then by (45)
$co(\underline{x}_k, \alpha) \subset J_{\Gamma_2}$ (or equivalently $\left| x_{k-j} - \alpha \right| < \Gamma_2$) implies

$$\left| x_{k+1} - \alpha \right| \leq H\Gamma^{q-1} \left| x_k - \alpha \right| \leq P \left| x_k - \alpha \right|$$

or

$$|x_{k+j} - \alpha| \leq P^j |x_k - \alpha|.$$

Therefore $|x_k - \alpha| \to 0$ as $k \to \infty$ if $co(\underline{x}_0) \subset T_f \subset J_{\Gamma_2}$.

The next theorem shows how to obtain the power of certain algorithms.

(46) Theorem: Let $\mathcal{D}(\mathcal{F})$ contain data of the form $d(f,\underline{x}_k) = (x_k, x_{k-1}, \ldots, x_{k-n}, f(x_k), \ldots, f^{(s-1)}(x_k), \ldots, f^{(s-1)}(x_{k-n}))$ where $s(n+1) = q$. Suppose $Df(x) \neq 0$, $|D^q f(x)| \leq R$ for all $f \in \mathcal{F}$ and $x \in T_f$. Let $\hat{\phi} \in \mathcal{A}(\mathcal{D}(\mathcal{F}))$ be a sufficient algorithm defined on a sufficient set $\hat{\mathcal{F}}$. Suppose there exists an $h \in \mathcal{F}$ and a solution sequence on h, $\{z_k\}$, such that

$$\lim_{k \to \infty} D^q \hat{f}_{d(h,\underline{z}_k)}(\alpha(h)) \neq D^q h(\alpha(h)) ,$$

then $\hat{\phi}$ has power p, the real positive root of $g(t) = t^{n+1} - s \sum_{j=0}^{n} t^j$,

Proof: From Lemma (17), for any $f \in \mathcal{F}$, (with $\alpha \equiv \alpha(f)$ and $\hat{f}_k \equiv \hat{f}_{d(f,\underline{x}_k)}$),

$$(47) \qquad |x_{k+1} - \alpha| = H_{k+1} \prod_{j=0}^{n} |x_{k-j} - \alpha|^s$$

where $H_{k+1} = \dfrac{1}{q!} \left| \dfrac{D^q f(\xi_k) - D^q \hat{f}_k(\xi_k)}{D\hat{f}_k(\eta_{k+1})} \right| ,$

$\xi_k \in co(\underline{x}_k, \alpha)$ and $\eta_{k+1} \in co(x_{k+1}, \alpha)$. By our assumptions H_{k+1} converges to say $H < \infty$. Using the development of Traub [11], Chapter 3, we let $e_k = |x_k - \alpha|$ and $D_k = \log(e_k/e_{k-1}^{\hat{p}}) =$

$\log e_k - \hat{p} \log e_{k-1}$ where $\hat{p}$ is the positive real root of $g(t)$. Taking the log of both sides of (47) and substituting D_k we get

$$(48) \qquad \sum_{j=0}^{n} c_j \, D_{k+1-j} = J_k$$

where $J_k = \log H_k$ and $c_j = \hat{p}^j - s \sum_{i=0}^{j-1} \hat{p}^i$. Traub shows that all roots of the characteristic equation of (48) lie within the unit sphere. (This type of system is called asymptotically stable.) He also shows that if $J_k \to J$, then $D_k \to (J / \sum_{j=0}^{n} c_j)$, which implies that

$$\frac{e_{k+1}}{e_k^{\hat{p}}} \;\to\; (H)^{\sum_{j=0}^{n} c_j} \equiv G_1(f)$$

or

$$(49) \qquad \lim_{k \to \infty} \frac{|\hat{\phi}(d(f,\underline{x}_k)) - \alpha|}{|x_k - \alpha|^{\hat{p}}} = G_1(f) < \infty \quad .$$

The boundedness of $D^q f$ and (48) imply that (5) (in definition of rate) is satisfied. Our assumption on h, implies $G_1(h) > 0$, so (6) is satisfied. It remains to show (9). So suppose (6) occurs for $\overline{h} \in \mathcal{F}$ and solution sequence $\{z_k\}$ on $\overline{h}$. It is clear that $G_1(\overline{h}) > 0$. As in (47), we have

(50) $\qquad |z_{k+1} - \alpha(\overline{h})| = \overline{H}_{k+1} \prod_{j=0}^{n} |z_{k-j} - \alpha(\overline{h})|^s$.

$G_1(\overline{h}) > 0$ implies $\overline{H}_{k+1} \to \overline{H} > 0$. Suppose $\{g_k\}$ satisfies (7) and (8), then since $d(g_k, \underline{z}_k) = d(\overline{h}, \underline{z}_k)$, we can again use Lemma (17) to obtain

(51) $\qquad |\alpha(g_k) - \alpha(\overline{h})| = \tilde{H}_{k+1} \prod_{j=0}^{n} |z_{k-j} - \alpha(\overline{h})|^s$.

By (50) and (51),

(52) $\qquad |\alpha(g_k) - \alpha(\overline{h})| = \dfrac{\tilde{H}_{k+1}}{\overline{H}_{k+1}} |z_{k+1} - \alpha(\overline{h})|$.

Now from Lemma (12) and (52),

(53) $\qquad \dfrac{|z_{k+1} - \alpha(g_k)|}{|z_k - \alpha(g_k)|^p} \le \dfrac{M^p \, |z_{k+1} - \alpha(g_k)|}{|z_k - \alpha(\overline{h})|^p}$

$$\le \dfrac{M^p \, (|z_{k+1} - \alpha(\overline{h})| + |\alpha(g_k) - \alpha(\overline{h})|)}{|z_k - \alpha(\overline{h})|^p}$$

$$\le M^p \left(1 + \dfrac{\tilde{H}_{k+1}}{\overline{H}_{k+1}}\right) \dfrac{|z_{k+1} - \alpha(\overline{h})|}{|z_k - \alpha(\overline{h})|^p}$$.

We have already shown, in (49), that

$$\varlimsup_{k\to\infty} \frac{|z_{k+1} - \alpha(\bar{h})|}{|z_k - \alpha(\bar{h})|^p} \leq C_1 < \infty \quad .$$

So noting that $\tilde{H}_{k+1} \to \bar{H} > 0$, and $\tilde{H}_{k+1} \to \tilde{H} < \infty$, we see that condition (9) in definition of rate is satisfied.

We can extend Theorem (46) to the case of multipoint data with memory.

(54) Theorem. Let $\mathcal{D}(\mathcal{F})$ contain data of the form $d(f, \underline{x}_k) =$
$(x_k, z_1(f, x_k), \ldots, z_r(f, x_k), x_{k-1}, z_1(f, x_{k-1}), \ldots, z_r(f, x_{k-1}),$
$\ldots, x_{k-n}, z_1(f, x_{k-n}), \ldots, z_r(f, x_{k-n}), f(x_k), \ldots, f^{(s-1)}(x_k),$
$f(z_1(f, x_k)), \ldots, f^{(s-1)}(z_1(f, x_k)), \ldots, f(z_r(f, x_k)), \ldots,$
$f^{(s-1)}(z_r(f, x_k)), \ldots, f(x_{k-n}), \ldots, f^{(s-1)}(z_r(f, x_{k-n}))).$

Suppose for each i, $z_i(f, x)$ satisfies conditions (37), (38), and (40). Let $S = s \sum_{i=0}^{r} \beta_i$ and let $q = (n+1)(s)(r+1)$. Suppose $f: R \to R$, $|D^q f(x)| \leq R$ and $Df(x) \neq 0$ for all $f \in \mathcal{F}$ and $x \in T_f$. Let $\hat{\Phi} \in \mathcal{A}(\mathcal{D}(\mathcal{F}))$ be a sufficient algorithm defined on a sufficient class $\hat{\mathcal{F}} \subset \mathcal{F}$. Also suppose there exists an $h \in \hat{\mathcal{F}}$ such that $D^q \hat{f}_{d(h, \underline{x}_k)} \not\to D^q h(\alpha(h))$ and such that $M_i(h, x)$ is bounded away from zero for all i, and $x \in T_h$, then $\hat{\Phi}$ has power $\hat{p}$, the real positive root of $g(t) =$
$$t^{(n+1)(r+1)} - S \sum_{j=0}^{nr+r+n} t^j .$$

Proof: By Lemma (17), for any $f \in \hat{\mathcal{F}}$

$$(55) \qquad |x_{k+1} - \alpha| = H_{k+1} \prod_{j=0}^{n} \prod_{i=0}^{r} |z_i(f, x_{k-j}) - \alpha|^s$$

where $\alpha = \alpha(f)$, $z_0(f, x_{k-j}) = x_{k-j}$,

$$H_{k+1} = \frac{1}{q!} \left| \frac{D^q f(\hat{\xi}_k) - D^q \hat{f}_{d(f,\underline{x}_k)}(\hat{\xi}_k)}{D\hat{f}_{d(f,\underline{x}_k)}(\hat{\eta}_{k+1})} \right| \quad ,$$

$q = s(n+1)(r+1)$, $\hat{\xi}_k \in co(\underline{x}_k, \alpha)$, and $\hat{\eta}_{k+1} \in co(x_{k+1}, \alpha)$.

Using (39), (55) becomes

$$|x_{k+1} - \alpha| = H_{k+1} \prod_{j=0}^{n} \prod_{i=0}^{r} \left(M_i(f, x_{k-j}) \, |x_{k-j} - \alpha|^{\beta_i} \right)^s$$

$$= \tilde{H}_{k+1} \prod_{j=0}^{n} |x_{k-j} - \alpha|^S$$

where $\tilde{H}_{k+1} = H_{k+1} \prod_{j=0}^{n} \prod_{i=0}^{r} [M_i(f, x_{k-j})]^s$, $M_0 = 1$, $\beta_0 = 1$

and $S = s \sum_{i=0}^{r} \beta_i$. Since M_i is continuous in x, $f \in C^q(R)$,

$D^q \hat{f}_{d(f,\underline{x}_k)}(\xi_k) \to J_f$ and $D\hat{f}_{d(f,\underline{x}_k)}(\eta_{k+1}) \to Df(\alpha)$, $\tilde{H}_k$ will

converge to, say $\tilde{H}$. We can now follow the proof of (46)

exactly to obtain the proof of this theorem.

II.8 Examples of Sufficient Classes of Functions

It is well known that given data of the form $d(f, \underline{x}) =$

$(x^0, x^1, \ldots, x^n, f(x^0), \ldots, f^{(\gamma_0 - 1)}(x^0), \ldots, f(x^n), \ldots, f^{(\gamma_n - 1)}(x^n))$,

where $f: R \to R$, there exists a unique polynomial $P_{d(f,\underline{x})}$ of

degree q-1 (where $q = \sum_{j=0}^{n} \gamma_j$) such that $d(f, \underline{x}) = d(P_{d(f,\underline{x})}, \underline{x})$.

We claim that $\hat{\mathcal{F}} = \{P_{d(f,\underline{x})} : f \in \mathcal{F}, \underline{x} \text{ on } T_f\}$ is a sufficient class of functions. Inoorder to simplify notation we will only consider the equal information case (i.e., $\gamma_0 = \gamma_1 = \ldots = \gamma_n = s$).

(56) Theorem: Suppose $\mathcal{D}(\mathcal{F})$ contains data of the form

$$d(f,\underline{x}) = (x^0,\ldots,x^n,f(x^0),\ldots,f^{(s-1)}(x^0),\ldots,f(x^n),\ldots,f^{(s-1)}(x^n)).$$

If $P_{d(f,\underline{x})}$ is the unique polynomial of degree q-1 (where q = (n+1)s) such that $d(f,\underline{x}) = d(P_{d(f,\underline{x})},\underline{x})$, then $\hat{\mathcal{F}} = \{P_{d(f,\underline{x})} : f \in \mathcal{F}, \underline{x} \text{ on } T_f\}$ is a sufficient class of functions on $\mathcal{F}$.

Proof: Since $d(P_{d(f,\underline{x})},\underline{x}) = d(f,\underline{x})$ by definition, itrremains to show $P_{d(f,\underline{x})}$ satisfies (i), (ii) and (iii) in Definition (13). To show (i) we shall state without proof some well known results of interpolation. (See Ostrowski and Traub for details). Using the Newtonian formations

$$P_{d(f,\underline{x})}(t) = \sum_{j=0}^{n} \sum_{i=0}^{s-1} C_{i,j}^s(t) \, f[x^0,s;x^1,s;\ldots;x^j,i+1]$$

where $C_{i,j}^s(t) = (t - x^j)^i \prod_{k=0}^{j-1} (t - x^k)^s$,

$$f[x^0,s;\ldots;x^j,i+1] = \frac{1}{j(s-1)! \, i!} \frac{\partial^{j(s-1)+i}}{\partial^{s-1}_x 0 \cdots \partial^i_x j} f[x^0,\ldots,x^j] \quad,$$

$$f[x^0,\ldots,x^j] = \frac{f[x^0,\ldots,x^{j-1}] - f[x^0,\ldots,x^{j-2},x^j]}{x^{j-1} - x^j} \quad,$$

and $f[x^j] = f(x^j)$. Using this formulation, $P_{d(f,\underline{x})}$ is well defined for all $\underline{x}$ such that $co(\underline{x}) \subset T_f$, including points that are not data points. Also $P_{d(f,\underline{x}_i)} \to P_{d(f,\underline{x})}$ where $\underline{x}_i$ is a sequence on T_f converging to $\underline{x}$. Therefore property (i) holds.

To show (ii), consider the error formula derived in Lemma (17):

$$f(t) - P_{d(f,\underline{x})}(t) = \frac{f^{(q)}[\xi(t)]}{q!} \prod_{j=0}^{n} (t - x^j)^s$$

where $\xi(t) \in co(\underline{x},t)$. This formula holds for $\underline{x}$ not data points for $P_{d(f,\underline{x})}$ given by the Newtonian formulation. Now suppose $co(\underline{x},t) \subset J = \{x : |x - \alpha| < \delta\}$. By continuity there exists an M such that $|f^{(q)}(v)| \leq M\, q!$ for $v \in J$; thus if $M\, \delta^q < \frac{\varepsilon}{2}$, then $|f(t) - P_{d(f,\underline{x})}(t)| \leq \frac{\varepsilon}{2}$.

By a proof similar to the one in Lemma (17) we can show that there exists an $\eta(t) \in co(\underline{x},t)$ such that

$$Df(t) - DP_{d(f,\underline{x})}(t) = \frac{D^q f(\eta(t))}{q!} \frac{d}{dt} \prod_{k=0}^{n} (t - x^k)^s$$

$$= \frac{s}{q!} D^q f(\eta(t)) \sum_{k=0}^{n} (t - x_k)^{s-1} \prod_{\substack{j=1 \\ j \neq k}}^{n} (t - x^j)^s.$$

So if $co(\underline{x},t) \subset J$ and if $|D^q f(v)| \leq M \frac{q!}{s}$ for $v \in J$, then $|Df(t) - DP_{d(f,\underline{x})}(t)| \leq M\, \delta^{q-1}$. Therefore if δ is picked

such that $M \delta^{q-1} < \varepsilon/2$, then $|f(t) - P_{d(f,\underline{x})}(t)| +$

$|Df(t) - DP_{d(f,\underline{x})}(t)| < \varepsilon$ whenever $co(t,\underline{x}) \subset \{x : |x - \alpha| < \delta\}$.

We still must show $D^i P_{d(f,\underline{x})}(t)$ is continuous in $\underline{x}$

and t. However, looking at the Newtonian formulation

given above, this is clear since $f[x^0,\ldots,x^j]$ is continuous

in the x^k, $k = 0,1,\ldots,j$.

We can also use inverse interpolation to obtain a

sufficient class. If f' is nonzero on an interval J about

$\alpha(f)$ and if $f \in C^q(R)$, then f has an inverse F such that

$F \in C^q(R)$. Let the data be of the form given at the beginning

of this section, then if $y^j = f(x^j)$, there exists a unique

polynomial $Q_{d(F,\underline{y})}$ of degree q-1 such that $d(F,\underline{y}) = d(Q_{d(F,\underline{y})},\underline{y})$

where $\underline{y} = (y^0,y^1,\ldots,y^n)$. One can show that there exists

a $\delta_1 > 0$ such that $co(\underline{y},t) \subset J_1 = \{y : |y| < \delta_1\}$ implies

that $Q_{d(F,\underline{y})}(t)$ is nonzero on J_1. Let $\Theta_{d(F,\underline{x})}(t) = Q_{d(F,\underline{y})}^{-1}(t)$

for $t \in Q_{d(F,\underline{y})}(J_1)$. By arguments almost identical to that

used in the previous theorem we can show that $\{\Theta_{d(F,\underline{x})} : f \in \mathcal{F},$

$\underline{x}$ on $T_f\}$ is a sufficient class of functions. The algorithm

defined on this class is very simple to obtain, to wit:

$$\Phi(d(f,\underline{x})) = \alpha(\Theta_{d(f,\underline{x})}) = Q_{d(F,\underline{y})}(0).$$

III.1 Roots of Systems of Equations

We shall consider only memoryless algorithms for finding roots of systems of equations. The theorems we shall prove are extensions to n dimensions of what we were able to prove for one dimension. First we will give the relation between the data point x_k and the new point x_{k+1} for a sufficient algorithm.

(1) Lemma: Let $\mathcal{D}(\mathcal{F})$ contain data of the form $d(f,x_k) = (x_k, f(x_k), Df(x_k), \ldots, D^{q-1}f(x_k))$. Let $\hat{\Phi}$ be a sufficient algorithm defined on the sufficient class $\hat{\mathcal{F}}$. Then if we define $\hat{f}_k \equiv \hat{f}_{d(f,x_k)}$ we have for all $t_0 \varepsilon R^n$

$$(2) \qquad f(t_0) = \hat{f}_k(t_0) + \int_0^1 W_k(\xi, t_0) \, d\xi \, (t_0 - x_k)^q$$

where $W_k(\xi, t_0) = \dfrac{1}{(q-1)!} (1 - \xi)^{q-1} [D^q f(x_k + \xi(t_0 - x_k)) - D^q \hat{f}_k(x_k + \xi(t_0 - x_k))]$. Also if $\hat{f}_k^{-1}$ exists and is C^1 in a ball U containing both 0 and $\hat{f}_k(\alpha)$, then

$$(3) \qquad |x_{k+1} - \alpha| \leq H_{k+1} |x_k - \alpha|^q$$

where $H_{k+1} = \left| [\int_0^1 D\hat{f}_k^{-1}(\xi \hat{f}_k(\alpha)) \, d\xi] \, [\int_0^1 W_k(\xi, \alpha) \, d\xi] \right|$

Proof: By Taylor's theorem

$$f(t_0) = f(x_k) + Df(x_k) (t_0 - x_k) + \ldots +$$

$$\frac{1}{(q-1)!} D^{q-1} f(x_k) (t_0 - x_k)^{q-1} +$$

$$\int_0^1 \frac{(1-\xi)^{q-1}}{(q-1)!} D^q f(x_k + \xi(t_0 - x_k)) (t_0 - x_k)^q \, d\xi \quad .$$

We can write $\hat{f}_k$ in the same form. If we note that $D^i f(x_k) = D^i \hat{f}_k(x_k)$ $i = 0, 1, \ldots, q-1$, we see that the expansions of f and $\hat{f}_k$ about x_k differ only in the remainder terms. So

$$f(t_0) = \hat{f}_k(t_0) + \int_0^1 W_k(\xi, t_0) \, d\xi \, (t_0 - x_k)^q \quad .$$

To prove (3), let $t_0 = f^{-1}(0) = \alpha$ in (2). Since $\hat{f}_k$ has a root α_k, we have

$$\alpha_k - \alpha = \hat{f}_k^{-1}(0) - \hat{f}_k^{-1}(\hat{f}_k(\alpha))$$

$$= - \int_0^1 D\hat{f}_k^{-1}(\xi \hat{f}_k(\alpha)) (\hat{f}_k(\alpha)) \, d\xi \quad .$$

Using (2), and recalling $\alpha_k = x_{k+1}$, we get

$$|x_{k+1} - \alpha| = |[\int_0^1 D\hat{f}_k^{-1}(\xi \hat{f}_k(\alpha)) \, d\xi][\int_0^1 W_k(\xi, \alpha) \, d\xi] (\alpha - x_k)^q|$$

or

$$|x_{k+1} - \alpha| \leq H_{k+1} |x_k - \alpha|^q .$$

We can now show, under certain conditions on $\mathcal{F}$, that algorithms defined on a sufficient class of functions are elements of $\mathcal{A}(\mathcal{D}(\mathcal{F}))$.

(4) Theorem: Let $\mathcal{D}(\mathcal{F})$ contain data of the form $d(f,x_k) = (x_k, f(x_k), Df(x_k), \ldots, D^{q-1}f(x_k))$ where $q \geq 2$. Suppose $|D^q f(x)| \leq M$, $Df(\alpha(f))$ is invertible with $|Df(\alpha(f)))^{-1}| = r_f$ for all $f \in \mathcal{F}$ and $x \in R^n$. Let $\widehat{\mathcal{F}}$ be a sufficient class, then there exists a sufficient algorithm $\hat{\Phi}$ defined on it.

Proof: For each $f \in \mathcal{F}$, we need to find a neighborhood T_f of $\alpha(f)$ such that $x_k \in T_f$ implies (i) $x_{k+1} \in T_f$ and (ii) $x_{k+1} \to \alpha(f)$ as $i \to \infty$. Both of these follow from a lemma:

(5) Lemma: Suppose $Df(\alpha)$ is invertible, and $|(Df(\alpha))^{-1}| = r_f$, where $f(\alpha) = 0$, and for all $\epsilon > 0$, there exists a $\delta > 0$ such that

$$|\hat{f}_{d(f,y)}(x) - f(x)| + |D\hat{f}_{d(f,y)}(x) - Df(x)| < \epsilon$$

for $x, y \in B_\delta(\alpha)$, and $\hat{f}_{d(f,x)}(x) = f(x)$. Then there exists a neighborhood U of α, such that f and $\hat{f}_{d(f,y)}$ are injective and f^{-1} and $\hat{f}^{-1}_{d(f,y)}$ are in C^1 on U for $y \in U$. Also there exists a neighborhood W of 0 such that $W \subset f(U)$ and $W \subset \hat{f}_{d(f,y)}(U)$, $Y \in U$, and $|[D\hat{f}_{d(f,y)}(x)]^{-1}| \leq 2r_f$ for $x, y \in U$.

The proof of this lemma is the same as the proof of Lemma (A.15) in Appendix A.

Since $\hat{f}$, by definition, satisfies the conditions of Lemma (5), we have that there exists a neighborhood U of α such that $0 \in f(U) \cap \hat{f}_{d(f,y)}(U)$ for $y \in U$. Thus for $x_k \in U$, $\hat{f}^{-1}_{d(f,x_k)}(0) = x_{k+1} \in U$; We have therefore shown (i).

To prove (ii) we use Lemma (1). For $x_k \in U$, (by Lemma (5)) the assumptions in Lemma (1) are satisfied, so from (3), with $\hat{f}_k \equiv \hat{f}_{d(f,x_k)}$,

$$|x_{k+1} - \alpha| \leq H_{k+1} |x_k - \alpha|^q$$

where

$$H_{k+1} \leq \left(\sup_{\xi \in (0,1)} \left| [D\hat{f}^{-1}_k(\xi \hat{f}_k(\alpha))]\right| \right) \cdot$$
$$\left(\sup_{\xi \in (0,1)} |D^q f(x_k + \xi(\alpha - x_k)) - D^q \hat{f}_k(x_k + \xi(\alpha - x_k))| \right)$$
$$\leq 4 Mr_f = H .$$

So if $\{x : |x - \alpha| < \Gamma\} \subset U$ where $H\Gamma^{q-1} = P < 1$, then $|x_k - \alpha| < \Gamma$ implies $|x_{k+1} - \alpha| < P |x_k - \alpha|$ or $|x_{k+j} - \alpha| \leq P^j |x_k - \alpha|$ which in turn implies $x_{k+j} \to \alpha$ as $j \to \infty$.

Given the conditions on $\mathcal{F}$ in Theorem (4), we can show that $\hat{\phi}$ has rate q.

(6) Theorem: Suppose all the conditions on $\mathcal{D}(\mathcal{F})$ and $\mathcal{F}$ given in Theorem (4) hold. Also suppose there exists an $h \in \mathcal{F}$ such that $\left[D^q h(\alpha(h)) - D^q \hat{f}_{d(h,\alpha(h))}(\alpha(h)) \right]$ is non-singular, then sufficient algorithms in $\mathcal{A}(\mathcal{D}(\mathcal{F}))$ have rate q.

Proof: Let $\hat{\Phi}$ be a sufficient algorithm defined on the sufficient class $\hat{\mathcal{F}}$. Let $\{x_k\}$ be a solution sequence for $f \in \hat{\mathcal{F}}$ and let $\hat{f}_k \equiv \hat{f}_{d(f,x_k)}$. Let g_k satisfy (a), (b), and (c) in definition of rate, then $g_k \to f$ uniformly at $\alpha \equiv \alpha(f)$; also by Remark (2.14), $\hat{f}_k \to f$ uniformly at α. Therefore by Lemma (A.15), there exists a K and a neighborhood W of 0 such that $\hat{f}_k^{-1}$ and g_k^{-1} exist and are C^1 on W for $k \geq K$. Now we can apply (3) of Lemma (1) with α replaced by $\alpha(g_k)$:

(7) $$|x_{k+1} - \alpha(g_k)| \leq H_{k+1} |x_k - \alpha(g_k)|^q$$

where

$$
\begin{aligned}
H_{k+1} = \Bigg| &\left(\int_0^1 D\hat{f}_k^{-1}(\xi \hat{f}_k(\alpha(g_k))) \, d\xi \right) \cdot \\
&\left(\int_0^1 \frac{(1-\xi)^{q-1}}{(q-1)!} \left[D^q g_k(x_k + \xi(\alpha(g_k) - x_k)) \right. \right. \\
&\qquad\qquad \left. \left. - D^q \hat{f}_k(x_k + \xi(\alpha(g_k) - x_k)) \right] d\xi \right) \Bigg| .
\end{aligned}
$$

By our assumptions on $\hat{\mathcal{F}}$, $H_{k+1} \leq 4 r_f M = H$. So

$$|x_{k+1} - \alpha(g_k)| \leq H |x_k - \alpha(g_k)|^q$$

which implies $\hat{\Phi}$ has rate at least q.

Now by (3), for $\{x_k\}$ a solution sequence for h,

$$|x_{k+1} - \alpha(h)| = \left|\left(\int_0^1 D\hat{f}_{d(h,x_k)}^{-1}(\xi \hat{f}_{d(h,x_k)}(\alpha(h))) \, d\xi\right) \cdot\right.$$

$$\left(\int_0^1 \frac{(1-\xi)^{q-1}}{(q-1)!} \left[D^q h(x_k + \xi(\alpha(h) - x_k))\right.\right.$$

$$\left.\left. - D^q \hat{f}_{d(h,x_k)}(x_k + \xi(\alpha(h) - x_k))\right] d\xi\right) \cdot$$

$$\left.(x_k - \alpha(h))^q\right|$$

$$= \left|V_k(x_k - \alpha(h))^q\right| \quad .$$

Since $x_k \to \alpha(h)$, V_k converges to $V =$
$Dh^{-1}(0) \cdot \left[D^q h(\alpha(h)) - D^q \hat{f}_{d(h,\alpha(h))}(\alpha(h))\right]$. Thus

$$\frac{|x_{k+1} - \alpha(h)|}{|x_k - \alpha(h)|^q} = \frac{|V_k(x_k - \alpha(h))^q|}{|x_k - \alpha(h)|^q} = |V_k e_k^q|$$

where $e_k = \dfrac{x_k - \alpha(h)}{|x_k - \alpha(h)|}$, so $|e_k| = 1$ for all k. Since

$Dh^{-1}(0)$ is nonsingular, and by our assumption on
$D^q \hat{f}_{d(h,\alpha(h))}$, V is nonsingular, so

(8) $\quad \displaystyle\lim_{k \to \infty} \frac{|x_{k+1} - \alpha(h)|}{|x_k - \alpha(h)|^q} = \lim_{k \to \infty} |V_k e_k^q| \neq 0$,

therefore $\hat{\Phi}$ has rate q.

III.2 Optimal Sufficient Algorithm on R^n

Sufficient algorithms in R^n are also optimal.

(9) Theorem: Suppose the conditions on $\mathcal{D}(\mathcal{F})$ and $\mathcal{F}$ given in Theorem (4) hold. Let $\hat{\Phi} \in \mathcal{A}(\mathcal{D}(\mathcal{F}))$ be a sufficient algorithm with power q defined on $\hat{\mathcal{F}}$. If there exists an $h \in \mathcal{F}$ and a solution sequence $\{z_k\}$ on h formed by Φ such that

(10) $$\lim_{k \to \infty} \frac{|z_{k+1} - \alpha(h)|}{|z_k - \alpha(h)|^p} > 0$$

and

(11) $$\left[D^q h(\alpha(h)) - D^q \hat{f}_{d(h,\alpha(h))}(\alpha(h)) \right]$$

is non singular, then any other algorithm $\dot{\Phi} \in \mathcal{A}(\mathcal{D}(\mathcal{F}))$ has power $p \leq q$.

Proof: Condition (11) implies that

(12) $$\lim_{k \to \infty} \frac{|\hat{\Phi}(d(h,z_k)) - \alpha(h)|}{|z_k - \alpha(h)|^q} > 0 \quad .$$

This can easily be shown using the same proof that yielded (8) in Theorem (6). Therefore, if we assume $p > q$,

$$\lim_{k \to \infty} \frac{|\hat{\Phi}(d(h,z_k)) - \alpha(h)|}{|z_k - \alpha(h)|^p} = + \infty \quad .$$

However,

$$\varlimsup_{k \to \infty} \frac{\left| \hat{\phi}(d(h,z_k)) - \alpha(h) \right|}{\left| z_k - \alpha(h) \right|^p} \leq \varlimsup_{k \to \infty} \frac{\left| z_{k+1} - \alpha(h) \right|}{\left| z_k - \alpha(h) \right|^p}$$

$$+ \varlimsup_{k \to \infty} \frac{\left| z_{k+1} - \hat{\phi}(d(h,z_k)) \right|}{\left| z_k - \alpha(h) \right|^p} \quad .$$

By condition (2.9) in the definition of rate, both terms on the right hand side are finite and so we have a contradiction, therefore $p \leq q$.

III.3 Examples of Sufficient Classes of Functions

We shall consider only the one point case; that is, the data is of the form $d(f,x) = (x,f(x),Df(x),\ldots,D^{q-1}f(x))$, where q is greater or equal 2. The first q terms of the Taylor expansion for f,

$$P_{d(f,x)}(t) = f(x) + Df(x) \ (t - x) + \ldots +$$

$$\frac{1}{(q-1)!} \ D^{q-1}f(x) \ (t - x)^{q-1} \quad ,$$

has the property that $d(f,x) = d(P_{d(f,x)},x)$. Also

$$\left| P_{d(f,x)}(t) - f(t) \right| \leq \frac{1}{q!} \sup_{\lambda \in (0,1)} \left| D^q f(t + \lambda(x - t)) \right| \ \left| t - x \right|^q$$

and

$$\left| DP_{d(f,x)}(t) - Df(t) \right| \leq \frac{1}{(q-1)!} \sup_{\lambda \varepsilon (0,1)} \left| D^{q-1} f(t+\lambda(x-t)) \right| \; \left| t-x \right|^{q-1} .$$

It is therefore straight forward to show that $\{ P_{d(f,x)} : f \varepsilon \mathcal{F}, x \varepsilon T_f \}$ is a sufficient class for $\mathcal{F}$.

If $Df(x)$ is invertible, we can expand $F \equiv f^{-1}$ into a Taylor series, so let

$$Q_{d(F,y)}(s) = F(y) + Df(y)(s-y) + \ldots +$$

$$\frac{1}{(q-1)!} D^{q-1} F(y)(s-y)^{q-1}$$

where $y = f(x)$. Clearly $d(F,y) = d(Q_{d(F,y)}, y)$. Also

$$\left| Q_{d(F,y)}(s) - F(s) \right| \leq \frac{1}{q!} \sup_{\lambda \varepsilon (0,1)} \left| D^q F(s + \lambda(y-s)) \right| \; \left| s-y \right|^q$$

and

$$\left| DQ_{d(F,y)}(s) - DF(s) \right| \leq \frac{1}{(q-1)!} \sup_{\lambda \varepsilon (0,1)} \left| D^{q-1} F(s+\lambda(y-s)) \right| \; \left| s-y \right|^{q-1} .$$

Therefore, y small enough (or equivalently x close enough to α) implies that $Q_{d(F,y)}$ is invertible in a region containing 0 and y. Let $\Theta_{d(f,x)} = Q_{d(F,y)}^{-1}$, then $\{ \Theta_{d(f,x)} : f \varepsilon \mathcal{F}, x \varepsilon T_f \}$ can easily be shown to be a sufficient class.

43

III.4 Multipoint Algorithms in R^n

Our theory can be extended to include algorithms dependent on multipoint data of a particular form. We shall first give a relation between the error, $|x_k - \alpha(f)|$, at the k and k+1 steps for this algorithm.

(13) Lemma: Let $\mathcal{D}(\mathcal{F})$ contain data of the form $d(f,\underline{x}) = (x,z(f,x),f(x),Df(x),\ldots,D^{r-1}f(x),f(z(f,x)),Df(z(f,x)),\ldots, D^{s-1}f(z(f,x)))$ where we assume $r \geq s$. Suppose $Df(\alpha(f))$ is invertable and $|D^r f(x)| \leq R < \infty$ for all $f \in \mathcal{F}$ and $x \in B_f$ where B_f is a ball about $\alpha(f)$. Also suppose

(i) there exists a data set $\tilde{\mathcal{D}}(\mathcal{F})$ containing data of the form $\tilde{d}(f,x) = (x,f(x),\ldots,D^t f(x))$ such that $f,g \in \mathcal{F}$ and $\tilde{d}(f,x) = \tilde{d}(g,x)$ implies $z(f,x) = z(g,x)$,

(ii) there exists an $M(f,x)$ continuous in x, where $|M(f,x)| \leq M < \infty$ for all $f \in \mathcal{F}$ and $x \in B_f$, and a positive integer β such that $z(f,x) - \alpha(f) = M(f,x) (x - \alpha(f))^\beta$,

(iii) $z(f,x) \neq x$ for all $f \in \mathcal{F}$ and $x \in B_f - \{\alpha(f)\}$.

Let $q = s(\beta - 1) + r$, and let $\hat{\phi}$ be an algorithm defined on a sufficient class $\hat{\mathcal{F}}$. Let $f \in \hat{\mathcal{F}}$ and define $\hat{f}_k \equiv \hat{f}_{d(f,\underline{x}_k)}$ and $\alpha \equiv \hat{\phi}(f)$. If $\hat{f}_k^{-1}$ exists and is in C^1 in a ball containing 0 and $\hat{f}_k(\alpha)$, then

$$|x_{k+1} - \alpha| \leq H_{k+1} |x_k - \alpha|^q$$

where $H_{k+1} \leq 2R(1 + M)^r \sup\limits_{\xi \in (0,1)} \{|D\hat{f}_k^{-1}(\xi \hat{f}_k(\alpha))|\}$.

Proof: Let $x_k \in B_f$ and $z_k = z(f, x_k) \in B_f$, then by Lemma (1)

$$\hat{f}_k(\alpha) = \hat{f}_k(\alpha) - f(\alpha) = \int\limits_0^1 \frac{(1-\xi)^{s-1}}{(q-1)!} \left[D^s \hat{f}_k(z_k + \xi(\alpha - z_k)) - D^s f(z_k - \xi(\alpha - z_k)) \right] (\alpha - z_k) \, d\xi .$$

Also by Lemma (1), with $w_k = z_k + \xi(\alpha - z_k)$,

$$D^s \hat{f}_k(w_k) - D^s f(w_k)$$

$$= \frac{1}{(r-s-1)!} \int\limits_0^1 (1 - \eta)^{r-s-1} \left[D^r \hat{f}_k(x_k + \eta(w_k - x_k)) - D^r f(x_k + \eta(w_k - x_k)) \right] d\eta \, (w_k - x_k)^{r-s}$$

$$= H_{k+1}^1(\xi) \, (w_k - x_k)^{r-s} \quad .$$

Now

$$(w_k - x_k) = z_k + \xi(\alpha - z_k) - x_k$$
$$= (1 - \xi)(z_k - \alpha) - (x_k - \alpha)$$
$$= (1 - \xi) \left[M(f, x_k)(x_k - \alpha)^{\beta - 1} - I \right] (x_k - \alpha)$$
$$= H_{k+1}^2(\xi) \, (x_k - \alpha) \quad .$$

Therefore

$$(14) \qquad \hat{f}_k(\alpha) = \int_0^1 \frac{(1-\xi)^{q-1}}{(q-1)!} \left[H_{k+1}^1(\xi) \ [H_{k+1}^2(\xi) \ (x_k - \alpha)]^{r-s} \right] d\xi$$

$$\cdot \left[M(f,x_k) \ (x_k - \alpha)^\beta \right]^s \quad .$$

Now $|H_{k+1}^1(\xi)| \leq 2R$ and $|H_{k+1}^2(\xi)| \leq M+1$ for $|x_k - \alpha| \leq 1$, so

$$|\hat{f}_k(\alpha)| \leq 2R(M + 1)^{r-s} M^s \ |x_k - \alpha|^{s(\beta-1)+r}$$

$$\leq 2R(M + 1)^r \ |x_k - \alpha|^{s(\beta-1)+r} \qquad .$$

Since $\hat{f}_k^{-1}$ exists and is in C^1 in a ball U containing 0 and $\hat{f}_k(\alpha)$ we have

$$x_{k+1} - \alpha = \hat{f}_k^{-1}(0) - \hat{f}_k^{-1}(\hat{f}_k(\alpha))$$

$$= - \int_0^1 D\hat{f}_k^{-1}(\xi\hat{f}_k(\alpha)) \ \hat{f}_k(\alpha) \ d\xi \quad ,$$

or

$$|x_{k+1} - \alpha| \leq \sup_{\xi \in (0,1)} |D\hat{f}_k^{-1}(\xi\hat{f}_k(\alpha))| \ |\hat{f}_k(\alpha)|$$

$$\leq 2R(M + 1)^r \left(\sup_{\xi \in (0,1)} |D\hat{f}_k^{-1}(\xi\hat{f}_k(\alpha))| \right) |x_k - \alpha|^{s(\beta-1)+r}$$

$$\leq H_{k+1} \ |x_k - \alpha|^{s(\beta-1)+r} \qquad .$$

We can now show that sufficient algorithms $\hat{\Phi}$ defined in Lemma (12) satisfies (i) and (ii) in the definition of

$\mathcal{A}(\mathcal{D}(\mathcal{F}))$ and has power $s(\beta-1)+r$. To be precise:

(15) Theorem: Suppose all the assumptions in Lemma (12) are satisfied, then $\hat{\Phi} \in \mathcal{A}(\mathcal{D}(\mathcal{F}))$. If in addition there exists an $h \in \mathcal{F}$ such that $\left[D^r h(\alpha(h)) - D^r \hat{f}_{d(h,\alpha(h))}(\alpha(h)) \right]$ and $M(h,\alpha(h))$ are nonsingular, then $\hat{\Phi}$ has power $q = s(\beta-1)+r$.

The proof of this theorem is analigous to the proofs of Theorems (4) and (6) so it will not be given.

We can also prove the counterpart of Theorem (9) for this multipoint algorithm. Again the proof is analigous so it will not be given.

III.5 A Multipoint Secant Method in R^n.

The secant method to find the root of a function f mapping R into R works as follows: Given two points x_0 and x_1, let $p:R \to R$ be the affine function that has the property that $p(x_0) = f(x_0)$ and $p(x_1) = f(x_1)$. Let x_2 be the point where p is zero. Repeat using x_1 and x_2. This method can be extended to find the roots of functions mapping R^n into R^n as follows: Given n+1 points $x_0, x_1, \ldots, x_n$, let $p:R^n \to R^n$ be the affine function such that $p(x_i) = f(x_i)$ for i = 0,1,...,n. Let x_{n+1} be the point where p is the zero vector. Repeat using $x_1, x_2, \ldots, x_{n+1}$. This method has been suggested by Ostrowski[8] and Wolfe[12] but no convergence proof exists.

We shall derive an interpolation formula giving a relation between the function f and the affine function p for all $x \in R^n$. It shall be clear that convergence cannot be proven without

at least assuring that the vectors $(x_{i+n}-x_i),\ldots,(x_{i+1}-x_i)$ span R^n.

A multipoint variant of the secant method is: Given a point x_0, let $x_0 = x_0^0$ and $x_0^i = x_0 + \beta |f(x_0)| e_i$ for $i = 1,2,\ldots,n$ where β is a constant and $\{e_i: i = 1,\ldots,n\}$ is the standard basis. Construct a hyperplane through $(f(x_0),x_0),f(x_0^1),x_0^1),\ldots,(f(x_0^n),x_0^n)$, call the root of the hyperplane x_1 and then repeat. It is easy to show that this method has rate 2 when used to find the root of a function from R→R. (See Traub[11], pp. 178-179). We shall show that this method, when used on a function from R^n→R^n, also has rate 2, and that it is a sufficient algorithm.

Let us first derive a mathematical formula for the multipoint secant algorithhm (which we will denote Φ).

Let $p_k(t) = A_k t + b_k$ be the hyperplane that agrees with f at $(x_k^0, x_k^1, \ldots, x_k^n)$ where $x_k^0 = x_k$ and $x_k^i = x_k + \beta|f(x_k)| e_i$ for $i = 1,2,\ldots,n$. Let $f_k^i = f(x_k^i)$, then $A_k x_k^i + b_k = f_k^i$ or $A_k(x_k^i - x_k^0) = f_k^i - f_k^0$. Now

$$x_{k+1} = \Phi(x_k) = \alpha(p_k) = -A_k^{-1}b_k = -A_k^{-1}(f_k^0 - A_k x_k)$$

$$= x_k^0 - A_k^{-1}f_k^0 = x_k - A_k^{-1}f(x_k).$$

Let F_k be an n×n matrix whose i^{th} column is $f_k^i - f_k^0$, then noting that $x_k^i - x_k^0 = \beta|f(x_k)| e_i$, we obtain $A_k^{-1} = \beta|f(x_k)| F_k^{-1}$.

Therefore

$$(16) \qquad x_{k+1} = x_k - \beta \, |f(x_k)| \, F_k^{-1} f(x_k) = x_k - A_k^{-1} f(x_k)$$

where the ith column of A_k is $\dfrac{1}{\beta|f(x_k)|} \left(f(x_k + \beta|f(x_k)| \, e_i) - f(x_k) \right)$. Equation (16) looks strikingly like the Newton–Raphson method with the matrix $Df(x_k)$ approximated by A_k. We will show in the course of the proof that A_k indeed converges to $Df(\alpha(f))$. The multipoint secant algorithm given in (16) can be seen to be a slight variation of a minimization algorithm suggested by Goldstein and Price[5]. They approximate $D^2 f(x_k)$ by a matrix $\overline{A}_k$ whose i^{th} column is $\dfrac{1}{\varepsilon_k} (g(x_k + \varepsilon_k e_i) - g(x_k))$ where g is the gradient of the function to be minimized and where $\varepsilon_k = \beta|\overline{A}_{k-1}^{-1} g(x_{k-1})|$ (β is an arbitrary constant). This algorithm can be shown to have rate less than two by the arguments developed below.

In order to obtain the rate of the algorithm we first derive a relation between a function f and a hyperplane p. (17) Lemma: Let $f \in C^2(R^n, R)$ and let $p: R^n \to R$ be affine. Let $\{x_i : i = 0, n\}$ be points in R^n such that $\{x_i - x_0 : i = 1, n\}$ are linearly independent. Let $x \in R^n$, and define α_j $j = 1, \ldots, n$ so that

$$(18) \qquad x - x_0 = \sum_{i=1}^{n} \alpha_i (x_i - x_0) \; .$$

Suppose $p(x_i) = f(x_i)$ for $i = 0,1,\ldots,n$, then there exists
points $\{\eta_{m,k} : m, k = 1,2,\ldots,n , m \leq k\}$ in R^n (defined more
precisely in the proof) such that

$$(19) \qquad f(x) = p(x) - \sum_{m=1}^{n} \sum_{k=m}^{n} \alpha_{m-1} (x_k-x_{m-1})^T D^2 f(\eta_{m,k}) \alpha_k (x_k-x_{m-1})$$

where $\alpha_0 = 1 - \sum_{i=1}^{n} \alpha_i$.

Proof: Define $h_{j,i}$, $\lambda_{j,i}$, and A_j , for $j, i = 1,2,\ldots,n$,
$j \leq i$, as follows: Let $A_1 = \sum_{i=1}^{n} \alpha_i$ and $\lambda_{1,i} = \dfrac{\alpha_i}{A_1}$ for $i = 1,\ldots,n$.
Thus

$$x - x_0 = \sum_{i=1}^{n} \alpha_i (x_i - x_0)$$

$$= \sum_{i=1}^{n} \lambda_{1,i} A_1(x_i - x_0)$$

and $\sum_{i=1}^{n} \lambda_{1,i} = 1$. Let $h_{1,1} = A_1 x_i + (1 - A_1) x_0$, then

$$x - h_{1,1} = x - x_0 - A_1 (x_1 - x_0)$$
$$= \sum_{i=1}^{n} \lambda_{1,i} [A_1 (x_i - x_0) - A_1 (x_1 - x_0)]$$

$$= \sum_{i=2}^{n} \lambda_{1,i} A_1 (x_i - x_1) \quad .$$

In general for $j, i = 1,\ldots,n, j \leq i$, let $h_{j,i}, A_j, \lambda_{j,i}$ be
defined by:

$$h_{j,i} = A_j h_{j-1,i} + (1 - A_j) h_{j-1,j-1} \text{, where } h_{0,i} = x_i \text{ .}$$

Therefore

$$(20) \qquad x - h_{j,j} = \sum_{i=j+1}^{n} \lambda_{j+1,i} A_{j+1} (h_{j,i} - h_{j,j})$$

$$(21) \qquad\qquad = \sum_{i=j+1}^{n} \lambda_{j+1,i} A_{j+1} A_j \cdots A_1 (x_i - x_j)$$

where $A_j = \sum_{i=j}^{n} \lambda_{j-1,i}$ and $\lambda_{j,i} = \dfrac{\lambda_{j-1,i}}{A_j}$. Now from (20)

$$x - h_{n-1,n-1} = \lambda_{n,n} A_n (h_{n-1,n} - h_{n-1,n-1})$$

or

$$x = \lambda_{n,n} A_n h_{n-1,n} + (1 - \lambda_{n,n} A_n) h_{n-1,n-1} \text{ .}$$

But $\lambda_{n,n} = \dfrac{\lambda_{n-1,n}}{A_n} = 1$, thus

$$x = A_n h_{n-1,n} + (1 - A_n) h_{n-1,n-1} = h_{n,n} \text{ .}$$

Therefore $f(x) = f(h_{n,n})$.

In order to prove (19), it is now necessary to prove the following, by induction, for $i \geq j$:

(22) $f(h_{j,i}) = p(h_{j,i}) +$

$$\sum_{m=1}^{i} \left[\sum_{k=m}^{j-1} A_{m+1}A_{m+2}\cdots A_k(1 - A_{k+1})\beta_{m,k} + A_{m+1}\cdots A_j\beta_{m,i} \right]$$

where $\beta_{m,k} = \frac{1}{2}(h_{m,k} - h_{m-1,k}) \, D^2 f(\eta_{m,k}) \, (h_{m,k} - h_{m-1,m-1})$

and $\eta_{m,k} \in co(h_{m,k}, h_{m-1,k}, h_{m-1,m-1})$. (Note that $h_{m,k}$ is a

linear combination of $h_{m-1,k}$ and $h_{m-1,m-1}$, therefore the

$co(h_{m,k}, h_{m-1,k}, h_{m-1,m-1})$ is a line segment). To prove

(22), we need the following corollary of Lemma (2.17);

(23) Corollary: Suppose $\hat{f}, \hat{p} \in C^2(R)$, with $\hat{p}$ affine. If

$\hat{f}(t) = \hat{p}(t)$ for $t = 0$ and 1, then there exists a $\xi \in co(0,1,t)$

such that $\hat{f}(t) = \hat{p}(t) + \frac{1}{2}t \, \hat{f}''(\xi) \, (t - 1)$.

Now let

$$\hat{f}_{1,i}(t) = f(t \, x_i + (1 - t)x_0)$$

and

$$\hat{p}_{1,i}(t) = p(t \, x_i + (1 - t)x_0)$$
$$= t \, p(x_i) + (1 - t)p(x_0) \quad .$$

Then $\hat{f}_{1,i}(t) = \hat{p}_{1,i}(t)$ for $t = 0,1$, so there exists a

$\xi_{1,i} \in co(0,1,A_1)$ such that

(24) $\hat{f}_{1,i}(A_1) = \hat{p}_{1,i}(A_1) + \frac{1}{2}A_1\hat{f}''_{1,i}(\xi_{1,i}) \, (A_1 - 1)$

But $A_1 x_i + (1 - A_1)x_0 = h_{1,i}$, so (24) implies there exists

an $\eta_{1,i} \in co(x_0, x_1, h_{1,i})$ such that

$$(25) \qquad f(h_{1,i}) = p(h_{1,i}) + \frac{1}{2} A_1(x_i - x_0)\ f''(\eta_{1,i})(A_1 - 1)(x_i - x_0)$$

or

$$f(h_{1,i}) = p(h_{1,i}) + \frac{1}{2}(h_{1,i} - x_i)\ f''(\eta_{1,i})(h_{1,i} - x_0)$$

Similarly let

$$\hat{f}_{2,i}(t) = f(t\ h_{1,i} + (1 - t)\ h_{1,1})$$

and

$$\hat{p}_{2,i}(t) = t\ f(h_{1,i}) + (1 - t)\ f(h_{1,1})$$

Thus $\hat{f}_{2,i}(t) = \hat{p}_{2,i}(t)$ for $t = 0, 1$ and for $i \geq 2$. Now
$h_{2,i} = A_2\ h_{1,i} + (1 - A_2)\ h_{1,1}$, Therefore there exists
an $\eta_{2,i} \in co(h_{1,i}, h_{1,1}, h_{2,i})$ such that

$$f(h_{2,i}) = A_2 f(h_{1,i}) + (1 - A_2)f(h_{1,1})$$
$$+ \frac{1}{2}(h_{2,i} - h_{1,i})f''(\eta_{2,i})(h_{2,i} - h_{1,1})\ .$$

In general there exists an $\eta_{j,i} \in co(h_{j,i}, h_{j-1,i}, h_{j-1,j-1})$

such that

$$(26) \qquad f(h_{j,i}) = A_j\, f(h_{j-1,i}) + (1 - A_j)f(h_{j-1,j-1}) + \beta_{j,i}$$

where $\beta_{j,i} = \frac{1}{2}(h_{j,i} - h_{j-1,i})\, f''(\eta_{j,i})\, (h_{j,i} - h_{j-1,j-1})$.

We are now ready to prove (22). First, from (25), it holds

for $j = 1$. So suppose (22) holds for j, then by (26):

$$f(h_{j+1,i}) = A_{j+1}\, f(h_{j,i}) + (1 - A_{j+1})\, f(h_{j,j}) + \beta_{j+1,i}$$

$$= A_{j+1}\, p(h_{j,i}) + (1 - A_{j+1})\, p(h_{j,j})$$

$$+ \sum_{m=1}^{j} \Bigg[\sum_{k=m}^{j-1} A_{m+1} A_{m+2} \cdots A_k (1 - A_{k+1}) \beta_{m,k}$$

$$+ A_{j+1}(A_{m+1} \cdots A_j \beta_{m,i})$$

$$+ (1 - A_{j+1})(A_{m+1} \cdots A_j \beta_{m,j}) \Bigg] + \beta_{j+1,i}$$

$$= p(h_{j+1,i}) + \sum_{m=1}^{j+1} \Bigg[\sum_{k=m}^{j} A_{m+1} \cdots A_k (1 - A_{k+1}) \beta_{m,k}$$

$$+ A_{m+1} \cdots A_{j+1}\, \beta_{m,i} \Bigg]$$

We have therefore shown (22).

Now note that

$$(27) \quad A_k A_{k-1} \cdots A_1 = \sum_{i=k}^{n} \lambda_{k-1,i} \, A_{k-1} \cdots A_1$$

$$= \sum_{i=k}^{n} \frac{\lambda_{k-1,i}}{A_{k-1}} A_{k-1} \cdots A_1$$

$$= \sum_{i=k}^{n} \lambda_{k-2,i} \, A_{k-2} \cdots A_1$$

$$\vdots$$

$$= \sum_{i=k}^{n} \alpha_i$$

and

$$(1 - A_k) = \lambda_{k-1,k-1} \; ,$$

so

$$(28) \quad (1 - A_k) \, A_{k-1} \cdots A_1 = \alpha_{k-1} \quad .$$

Recall that

$$(h_{m,k} - h_{m,k}) = (A_{m,i} - 1) \, (h_{m-1,k} - h_{m-1,m-1})$$

and

$$(h_{m,k} - h_{m-1,m-1}) = A_{m} \, (h_{m-1,k} - h_{m-1,m-1}) \quad ,$$

therefore

$$(h_{m,k} - h_{m-1,k}) = (A_m - 1)A_{m-1} \ldots A_1(x_k - x_{m-1})$$

and

$$(h_{m,k} - h_{m-1,m-1}) = A_m A_{m-1} \ldots A_1(x_k - x_{m-1}) \qquad ,$$

By (22)

$$f(x) = f(h_{n,n}) = p(x) + \sum_{m=1}^{n} \left[\sum_{k=m}^{n-1} \frac{1}{2}(A_m - 1)A_{m-1} \ldots A_1(x_k - x_{m-1})^T \right.$$

$$\cdot D^2 f(\eta_{m,k}) \ (1 - A_{k+1})A_k \ldots A_1(x_k - x_{m-1})$$

$$+ \frac{1}{2}(A_m - 1)A_{m-1} \ldots A_1(x_n - x_{m-1})^T D^2 f(\eta_{m,n})$$

$$\left. \cdot A_n \ldots A_1(x_n - x_{m-1}) \right]$$

$$= p(x) - \sum_{m=1}^{n} \sum_{k=m}^{n} \alpha_{m-1}(x_k - x_{m-1})^T D^2 f(\eta_{m,k})\alpha_k \ (x_k - x_{m-1})$$

where $\alpha_0 = 1 - A_1 = 1 - \sum_{k=1}^{n} \alpha_k$.

A formula of the form of (19) can only be obtained when $x - x_0 \in sp\{x_i - x_0 : i = 1,2,\ldots,n\}$. In the regular secant method for R^n we cannot insure that this occurs and so we cannot use (19). In the multipoint secant method, however,

the vectors $\{x_i - x_0 : i = 1,2,\ldots,n\}$ always span the space.
We can therefore prove the following:

(29) Lemma: Let $f \in C^2(R^n, R^n)$ with a root $\alpha(f)$ and let $p:R^n \to R^n$
be affine. Assume $|D^2 f(x)| \leq M$, for all x. Given x_0, define
$x_i = x_0 + \beta |f(x_0)| e_i$. Suppose $p(x_i) = f(x_i)$ for $i = 0,1,\ldots,n$,
and that p has a root $\alpha(p)$. We can write $p(x) = Ax + b$.
Assume that A is invertible, then

$$|p(x) - f(x)| \leq Mn|x - x_0|\left(\frac{(n-1)}{2}|x - x_0| + \beta|f(x_0)|\right)$$

and

$$|\alpha(p) - \alpha(f)| \leq H |x_0 - \alpha(f)|^2$$

where $H = \frac{Mn(n-1)}{2}|A^{-1}|\left(1 + \sup_{t \in (0,1)} \{|Df[\alpha(f) + t(x_0 - \alpha(f))]|\}\right)$.

Proof: Define α_i, $i = 1,2,\ldots,n$ by

$$x - x_0 = \sum_{i=1}^{n} \alpha_i (x_i - x_0) .$$

Let $c = \beta|f(x_0)|$, and $\underline{\alpha} = (\alpha_1,\ldots,\alpha_n)^T$, so $x - x_0 = c\underline{\alpha}$,
or $\alpha_i = \frac{1}{c} <e_i, x - x_0>$. Now

$$\alpha_j(x_k - x_m) = <e_j, x - x_0> (e_k - e_m) \quad j,k,m = 1,2,\ldots,n , \quad k \neq m.$$

So

(30) $|\alpha_j(x_k - x_m)| \leq |x - x_0|$ $j,k,m = 1,2,\ldots,n$, $k \neq m$,

where $|x| = \sup |x^1|$. Also

$$\alpha_0(x_k - x_0) = (1 - \sum_{i=1}^{n} \alpha_i) \, (x_k - x_0)$$

$$= ce_k - (\sum_{i=1}^{n} \alpha_i) \, ce_k \quad .$$

But

$$(\sum_{i=1}^{n} \alpha_i) \, ce_k = \sum_{i=1}^{n} \langle e_i, \, x - x_0 \rangle \, e_k$$

$$= \sum_{i=1}^{n} \langle e_i, x - x_k \rangle \, e_k + \sum_{i=1}^{n} \langle e_i, x_k - x_0 \rangle \, e_k$$

$$= \sum_{i=1}^{n} \langle e_i, \, x - x_k \rangle e_k + ce_k \quad .$$

Thus

$$\alpha_0(x_k - x_0) = -\sum_{i=1}^{n} \langle e_i, x - x_k \rangle \, e_k$$

$$= -\langle (1,1,\ldots,1), \, x - x_k \rangle \, e_k$$

or

(31) $|\alpha_0(x_k - x_0)| \leq |x - x_k|$.

From (19) for $i = 1, 2, \ldots, n$, there exists points $\{\eta^i_{m,k} : i, m, k = 1, \ldots, n , m \leq k\}$ such that

$$f^i(x) = p^i(x) - \sum_{m=1}^{n} \sum_{k=m}^{n} \alpha_{m-1}(x_k - x_{m-1})^T D^2 f^i(\eta^i_{m,k}) \alpha_k (x_k - x_{m-1}) \; .$$

Using (30) and (31),

$$|f^i(x) - p^i(x)| \leq \sum_{m=2}^{n} \sum_{k=m}^{n} |D^2 f(\eta^i_{m,k})| \; |x - x_0|^2$$

$$+ \sum_{k=1}^{n} |D^2 f(\eta^i_{m,k})| \; |x - x_0| \; |x - x_k| \; .$$

Now $|x - x_k| \leq |x - x_0| + \beta |f(x_0)|$, and $|f(x_0)| \leq$

$\sup\limits_{t \in (0,1)} \{|Df(\alpha(f) + t(x_0 - \alpha(f)))|\} |x_0 - \alpha(f)| \equiv \overline{D} |x_0 - \alpha(f)|$.

So $|\alpha(f) - x_k| \leq (1 + \overline{D} |x_0 - \alpha(f)|)$. Therefore

$$|p^i(x) - f^i(x)| \leq \frac{Mn(n-1)}{2} |x_0 - x|^2 + nM\beta |f(x_0)| \; |x - x_0|$$

and

$$|p^i(\alpha(f))| \leq M(1 + \overline{D}) \frac{n(n-1)}{2} |x_0 - \alpha(f)|^2 \quad .$$

Now suppose $p(x) = Ax + b$ and A is invertible, then $\alpha(p) = -A^{-1}b$. So

$$|\alpha(p) - \alpha(f)| = |{-}A^{-1}b - \alpha(f)|$$

$$= |A^{-1}(b + A\alpha(f))|$$

$$= |A^{-1} \, p(\alpha(f))|$$

$$\leq |A^{-1}| \, |p(\alpha(f))| \quad .$$

Thus

$$|\alpha(p) - \alpha(f))| \leq H \, |x_0 - \alpha(f)|^2$$

where $H = \dfrac{Mn(n-1)}{2} \, |A^{-1}| \left(1 + \sup_{t\epsilon(0,1)} \{|Df(\alpha(f) + t(x_0 - \alpha(f)))|\}\right)$.

To show that the algorithm converges we need a relation between $Df(x)$ and $Dp(x)$.

(32) Lemma: Under the assumptions of Lemma (29),

$$|Df(x) - Dp(x)| \leq \frac{Mn}{2} \left[(n-1)|x - x_0| + \beta|f(x_0)|\right] \quad .$$

Proof: Analagously to the proof of Lemma (17), we can show that for each $i,j = 1,2,\ldots,n$, there exists $\eta_{m,k}^{i,j}$ such that

$$D_j f^i(x) = D_j p^i(x) - \frac{1}{2} \sum_{m=1}^{n} \sum_{k=m}^{n} \left\{ \frac{d\alpha_{m-1}}{dx_j} (x_k - x_{m-1})^T \right.$$

$$\cdot D^2 f^i(\eta_{m,k}^{i,j})\alpha_k(x_k - x_{m-1})$$

$$\left. + \alpha_{m-1}(x_k - x_{m-1})^T D^2 f^i(\eta_{m,k}^{i,j}) \frac{d\alpha_k}{dx_j} (x_k - x_{m-1}) \right\} \quad .$$

Now, $\alpha_j(x_k - x_m) = <e_j, x - x_0> (e_k - e_m)$, for $j \neq 0$. So

$\frac{d\alpha_j}{dx}(x_j - x_m) = e_k - e_m > <e_j$, therefore $\left|\frac{d\alpha_j}{dx}(x_j - x_m)\right| \leq 1$.

Also $\alpha_0(x_k - x_0) = -<(1,\ldots,1),(x - x_k)> e_k$ and so

$\frac{d\alpha_0}{dx}(x_k - x_0) = e_k > <(1,\ldots,1)$. Therefore $\left|\frac{d\alpha_0}{dx}(x_k - x_0)\right| \leq 1$.

We then have

$$|Df(x) - Dp(x)| \leq M \sum_{m=2}^{n} \sum_{k=m}^{n} |x - x_0| +$$

$$\frac{M}{2} \sum_{k=1}^{n} (|x - x_0| + |x - x_k|) .$$

But $|x - x_k| \leq |x - x_0| + \beta|f(x_0)|$, thus

$$|Df(x) - Dp(x)| \leq M \frac{n(n-1)}{2}|x - x_0| + \frac{Mn}{2}\beta|f(x_0)| .$$

The multipoint secant method defined above is a sufficient algorithm in $\mathcal{A}(\mathcal{D}(\mathcal{F}))$ and has power 2, to be precise:

(33) Theorem: Let $\mathcal{D}(\mathcal{F})$ contain data of the form $d(f,x) = (x, x + \beta|f(x)|e_1,\ldots,x + \beta|f(x)|e_n, f(x), f(x + \beta|f(x)|e_1),\ldots, f(x + \beta|f(x)|e_n))$, where $f \in C^2(\mathbb{R}^n)$, $Df(\alpha(f))$ is invertible and $|D^2f(x)| \leq M$ for all $f \in \mathcal{F}$ and $x \in \mathbb{R}^n$. Define $P_{d(f,\underline{x})}(t)$ as that affine function from $\mathbb{R}^n \rightarrow \mathbb{R}^n$ such that $d(P_{d(f,\underline{x})},\underline{x}) = d(f,\underline{x})$, then the set of functions $\hat{\mathcal{F}} = \{P_{d(f,\underline{x})} : f \in \mathcal{F}, \underline{x} \text{ on } T_f\}$ is a sufficient class and $\hat{\Phi}(d(f,\underline{x})) = \alpha(P_{d(f,\underline{x})})$ is a

sufficient algorithm with rate 2. Also $\hat{\Phi}$ is optimal in

$\mathcal{A}(\mathcal{D}(\mathcal{F}))$.

Proof: Let $f \in \mathcal{F}$ with root $\alpha(f)$ and let $P_{d(f,\underline{x})}(t) = At + b$

for a particular $x \in T_f$. Then $Ax^i + b = f(x^i) = f^i$ for

$i = 0,1,\ldots,n$ where $\underline{x} = (x^0,\ldots,x^n)$, so $A(x^i - x^0) = f^i - f^0$

or $A = F X^{-1}$ where $X^i = i^{th}$ column of $X = (x^i - x^0)$ and

$F^i = f^i - f^0$. Also $b = f^0 - Ax^0$, so $P_{d(f,\underline{x})}(t) =$

$A(t - x^0) + f^0 = A(t - x) + f(x)$. Since $\{x^i - x^0 : i = 1,\ldots,n\}$

is linearly independent at all data points, and since

F add X are continuous in $\underline{x}$, if $\underline{x}_i$ is a sequence of data

points converging to a data point $\underline{x}$, then $P_{d(f,\underline{x}_i)} \rightarrow P_{d(f,\underline{x})}$.

The only point, not a data point, that a sequence of data

points can converge to, is $\underline{\alpha(f)} = (\alpha(f),\ldots,\alpha(f))$. So

let $\underline{x}_i$ be a sequence of data points converging to $\underline{\alpha(f)}$, and

let A_i and b_i be such that $P_{d(f,\underline{x}_i)}(t) = A_i t + b_i$. The

convergence of A_i is sufficient to imply the convergence

of $P_{d(f,\underline{x}_i)}$; but $A = DP_{d(f,\underline{x}_i)}(t)$, and by Lemma (32)

$$\left| Df(t) - DP_{d(f,\underline{x}_i)}(t) \right| = \left| Df(t) - A_i \right|$$

$$\leq \frac{Mn}{2} \left[(n - 1) \, |t - x_i| + \beta |f(x_i)| \right] \quad ,$$

so

$$\lim_{\underline{x}_i \rightarrow \underline{\alpha}} \left| Df(\alpha) - A_i \right| = \lim_{\underline{x}_i \rightarrow \underline{\alpha}} \frac{Mn}{2} \left[(n - 1)|\alpha - x_i| + \beta |f(x_i)| \right] = 0 \quad ,$$

or $A_i \to Df(\alpha)$. So $P_{d(f,\underline{x})}$ satisfies condition (i) in definition (2.13). We must now show for all $\epsilon \gtrless 0$ there exists a $\delta > 0$ such that

$$|P_{d(f,\underline{x})}(t) - f(t)| + |Dp_{d(f,\underline{x})}(t) - Df(t)| < \epsilon$$

for $co(t,\underline{x}) \subset B_\delta(\alpha(f))$. It is clear from Lemmas (29) and (32) that this will be satisfied. This completes the proof showing $\widehat{\mathcal{F}}$ is a sufficient class of functions. It is obvious that $\hat{\Phi}$ is defined on $\widehat{\mathcal{F}}$.

The facts that $\hat{\Phi} \in \mathcal{A}(\mathcal{D}(\widehat{\mathcal{F}}))$ and has rate 2 follows from the sufficiency of $\widehat{\mathcal{F}}$ and Lemma (29). The proofs are identical to the proofs of Theorems (4) and (6). The proof that $\hat{\Phi}$ is optimal in $\mathcal{A}(\mathcal{D}(\widehat{\mathcal{F}}))$ is identical to the proof of Theorem (9).

IV.1 Rate of Convergence of Several Conjugate Gradient Algorithms

Conjugate gradient algorithms converge to the extermum of quadratic functions in a finite number of steps. There are several different forms of this method, most of which use, at each step, the value of the function to be extremized and its first derivative. This paper is concerned with convergence rates of several conjugate gradient algorithms when they are used to minimize a class of nonlinear, non-quadratic, real valued functions on R^n. The algorithms considered are given by Polak and Ribiere[9], Daniel[1], and Fletcher and Reeves[4].

Conjugate gradient algorithms are either used directly or with the conjugate variable reinitialized every r steps (where r is some constant bigger or equal to n). Rates of convergence will be found for both the Polak-Ribiere and Daniel algorithms when they are used both ways and for the Fletcher-Reeves algorithm only when the conjugate variable is reinitialized.

IV.2 Definitions and Preliminary Remarks.

We shall use the definition of rate of convergence given by Ostrowski[8]. Recall that it is defined as follows:

Definition: The power of an algorithm $\Phi: R^n \rightarrow R^n$ that produces a sequence z_0, z_1, ... (where z_0 is given and $z_{k+1} = \Phi(z_k)$ for $k = 0,1,...$) converging to α, is the largest real constant

$p > 1$ such that

$$\overline{\lim_{k \to \infty}} \frac{|\Phi(z_k) - \alpha|}{|z_k - \alpha|^p} \leq C < \infty .$$

Note that we are now considering Φ to be a map from $R^n \to R^n$, f is considered fixed. The power of the conjugate gradient algorithm will be determined by comparing it to the Newton-Raphson method. The Newton-Raphson method, denoted $\hat{\Phi}$, is for a particular function f:

$$\hat{\Phi}(x) = x - (D^2 f(x))^{-1} Df(x) .$$

It is well known that his method converges, with power $p = 2$, to the root α of Df for the set of functions in C^3 (R^n, R) with invertable second derivatives. The weaker result that

$$\overline{\lim_{z \to \alpha}} \frac{|\hat{\Phi}(z) - \alpha|}{|z - \alpha|^2} \leq C < \infty$$

is shown in Appendix B with the additional assumption (AS2) (given below).

Minimization Problem: The algorithms considered are to be used to solve the problem:

Minimize the function $f : R^n \to R$ where it is assumed

(AS1): $f \in C^3 (R^n, R)$

(AS2):   There exists $m > 0$ and M such that $H(z) \equiv D^2 f(z)$

satisfies $m \langle y, y \rangle \leq \langle y, H(z) y \rangle \leq M \langle y, y \rangle$ for all

z and y in R^n.

Algorithm: The conjugate gradient algorithms (PR), (D) and

(FP) (which denote respectively the Polak-Ribiere, Daniel

and Fletcher-Reeves algorithms) are:

(i)   Given initial condition z_0, let $h_0 = g_0 = -Df(z_0)$.

(ii)   Given z_k, h_k, g_k construct $z_{k+1}, h_{k+1}, g_{k+1}$ according to:

(1) $z_{k+1} = z_k + \lambda_k h_k$

(2) $g_{k+1} = -Df(z_{k+1})$

(3) $h_{k+1} = g_{k+1} + \gamma_k h_k$

where λ_k is the small positive constant such that

(4) $f(z_k + \lambda_k h_k) \leq f(z_k + \lambda h_k)$ for all $\lambda \geq 0$ and where

(5)(PR) $\gamma_k = \dfrac{\langle g_{k+1} - g_k, \ g_{k+1} \rangle}{\langle g_k, \ g_k \rangle}$

(5)(D) $\gamma_k = - \dfrac{\langle g_{k+1}, \ H_{k+1} h_k \rangle}{\langle h_k, \ H_{k+1} \ h_k \rangle}$ where $H_{k+1} = H(z_{k+1})$,

$$(5)(\text{FR}) \quad \gamma_k = \frac{<g_{k+1}, \; g_{k+1}>}{<g_k, \; g_k>} \quad .$$

h_k is called the conjugate gradient. Note that it can be shown that the γ_k are identical for the case where f is quadratic. The algorithm will be denoted $z_{k+1} = \psi(z_k)$, the dependence on f and h_k being understood. For the case where the conjugate variable is reinitialized, h_k will be set equal to g_k every r steps.

IV.3 Rate of Convergence

The main result of this section is that the algorithm formed by composing n steps of the conjugate gradient algorithms has power at least two. This is true for all three algorithms given above when the conjugate variable is reinitialized, denoted (w.r.), and for the Polak and Ribiere, and Daniel algorithms when it is not, denoted (w/o.r.)†. The precise results are given below.

(6) Theorem (w/o.r.): Suppose the conjugate gradient algorithm (PR) or (D) is used to solve the Minimization Problem. Let z_0, z_1, ... be the sequence formed, then there exists a constant C such that

$$(7) \qquad \overline{\lim_{k \to \infty}} \frac{|z_{k+n} - \alpha|}{|z_k - \alpha|^2} \leq C < \infty$$

---

†Daniel[2] has shown this result for the (D) algorithm (w/o.r.) .

where α is a stationary point (hence the minimum since f is convex) of the function to be minimized.

(8) Theorem (w.r.):

Suppose the conjugate gradient algorithm (PR), (D) or (FP) is used to solve the Minimization Problem where we reinitialize the conjugate variable every r steps (where $r \geq n$). Let $z_0, z_1, z_2, \ldots$ be the sequence formed, then there exists a constant C such that

$$(9) \qquad \overline{\lim_{k \to \infty}} \frac{|z_{kr+n} - \alpha|}{|z_{kr} - \alpha|^2} \leq C < \infty \quad .$$

Theorems (6) and (8) shall be proven simultaneously; any differences will be denoted (w.r.) or (w/o.r.). The proof follows in two steps, the first step is true for all three algorithms.

Step I: At each point z_k define the quadratic function $\hat{f}_k$ by

$$(10) \qquad \hat{f}_k(z) = f(z_k) + \langle Df(z_k), z - z_k \rangle + \frac{1}{2}\langle z - z_k, D^2 f(z_k)(z - z_k) \rangle.$$

Let $z_k^0 = z_k$, $z_k^1 = \psi_{\hat{f}_k}(z_k^0)$, $z_k^2 = \psi_{\hat{f}_k}(z_k^1)$, $\ldots, z_k^n = \psi_{\hat{f}_k}(z_k^{n-1})$,

where $\psi_{\hat{f}_k}$ is the conjugate gradient algorithm applied to $\hat{f}_k$ with the conjugate variable h_k^0 set to h_k . (Note for the case (w.r.), $h_k^0 = h_k = g_k$, for k an integer multiple of r). Then the n step conjugate gradient method (w/o.r.) will satisfy (7) if

there exists a constant C' such that

$$(11) \quad \varlimsup_{k\to\infty} \frac{|(z_{k+i+1} - z_{k+i}) - (z_k^{i+1} - z_k^i)|}{|z_k - \alpha|^2} \leqq C' < \infty$$

for i = 0,1,2,...,n-1 , and it (w.r.) will satisfy (9) if there

exists a constant C" such that

$$(12) \quad \varlimsup_{k\to\infty} \frac{|(z_{kr+i+1} - z_{kr+i}) - (z_{kr}^{i+1} - z_{kr}^i)|}{|z_{kr} - \alpha|^2} \leqq C'' < \infty$$

for i = 0,1,2,...,n-1 .

Proof of Step I:

Step I shall be proved for the case (w/o.r.), the case (w.r.)

then follows trivially.

First note that the stationary point of $\hat{f}_k(z)$, denoted $\alpha(\hat{f}_k)$,

will be

$$(13) \quad \alpha(\hat{f}_k) = z_k - (D^2 f(z_k))^{-1} Df(z_k) \quad .$$

But this is just the Newton-Raphson algorithm $\hat{\Phi}$ applied on

f at z_k, so:

$$\hat{\Phi}(z_k) = \alpha(\hat{f}_k) \quad .$$

Now

$$\varlimsup_{k\to\infty} \frac{|z_{k+n} - \alpha|}{|z_k - \alpha|^2} \leq \varlimsup_{k\to\infty} \frac{|z_{k+n} - \alpha(\hat{f}_k)| + |\alpha(\hat{f}_k) - \alpha|}{|z_k - \alpha|^2} \quad .$$

But

$$\varlimsup_{k\to\infty} \frac{|\alpha(\hat{f}_k) - \alpha|}{|z_k - \alpha|^2} = \varlimsup_{k\to\infty} \frac{|\hat{\Phi}(z_k) - \alpha|}{|z_k - \alpha|^2} \leq C_1 < \infty \quad ,$$

where C_1 is some constant, since $\hat{\Phi}$ is the Newton-Raphson algorithm. So to show that (7) holds, it is enough to show that there exists a constant C_2 for all z_0 in R^n, such that

$$(14) \qquad \varlimsup_{k\to\infty} \frac{|z_{k+n} - \alpha(\hat{f}_k)|}{|z_k - \alpha|^2} \leq C_2 < \infty \quad .$$

Now note that $z_k^n = \psi_{\hat{f}_k}(\psi_{\hat{f}_k}(\ldots \psi_{\hat{f}_k}(z_k)\ldots)) = \alpha(\hat{f}_k)$. This follows from the fact that the conjugate gradient algorithm will reach the minimum of a quadratic function in n steps. So

$$|z_{k+n} - \alpha(\hat{f}_k)| = |z_{k+n} - z_k^n|$$

But since $z_k = z_k^0$,

$$|z_{k+n} - z_k^n| = |\sum_{i=0}^{n-1} (z_{k+i+1} - z_{k+i}) - (z_k^{i+1} - z_k^i)|$$

$$\leq \sum_{i=0}^{n-1} |(z_{k+i+1} - z_{k+i}) - (z_k^{i+1} - z_k^i)| \quad ;$$

so the constant C_2 will exist in (14) if there exists C' such that

$$\overline{\lim_{k \to \infty}} \frac{|(z_{k+i+1} - z_{k+i}) - (z_k^{i+1} - z_k^i)|}{|z_k - \alpha|^2} \leq C' < \infty \quad i = 0,1,\ldots,n-1.$$

So Step I is proved.

Referring to (1) in the definition of the conjugate gradient algorithms we see that $(z_{k+i+1} - z_{k+i}) = \lambda_{k+i} h_{k+i}$ and $(z_k^{i+1} - z_k^i) = \lambda_k^i h_k^i$ so we can rewrite the condition in (11) as

$$(15) \qquad \overline{\lim_{k \to \infty}} \frac{|\lambda_{k+i} h_{k+i} - \lambda_k^i h_k^i|}{|z_k - \alpha|^2} \leq C' < \infty \quad i = 0,1,\ldots,n-1$$

and the condition in (12) as

$$(16) \qquad \overline{\lim_{k \to \infty}} \frac{|\lambda_{kr+i} h_{kr+i} - \lambda_{kr}^i h_{kr}^i|}{|z_{kr} - \alpha|^2} \leq C'' < \infty \quad i = 0,1,\ldots,n-1.$$

Step II. It remains to show (15) or (16) for the various conjugate gradient algorithms. The proof uses the fact that the

algorithms converge to the stationary point of f. The convergence of the algorithms with reinitialization follows from the convergence of the steepest decent algorithms since every r[th] step is a steepest decent step. The convergence of the (PR) and (D) algorithms without reinitialization is given in [8] and [1] respectfully.

In order to prove (15) and (16) it is necessary to first prove several lemmas[†]. These lemmas hold, unless otherwise noted, for all three methods and for all steps k.

(17) Lemma: $\langle g_k, h_k \rangle = \langle g_k, g_k \rangle$.

(18) Lemma: $\langle g_{k+1}, h_k \rangle = 0$.

Proof: $\langle g_0, h_0 \rangle = \langle g_0, g_0 \rangle$ since $h_0 = g_0$. Let $q_0(\lambda) = f(z_0 + \lambda h_0)$. The global minimum λ_0^* of q_0 will be positive since by (AS2) q_0 is convex and since $\dfrac{dq_0(0)}{d\lambda} = -\langle g_0, g_0 \rangle < 0$. Thus $\lambda_0 = \lambda_0^*$, and $\dfrac{dq_0(\lambda_0)}{d\lambda} = -\langle g_1, h_0 \rangle = 0$, and so $\langle g_1, h_1 \rangle = \langle g_1, g_1 + \gamma_0 h_0 \rangle = \langle g_1, g_1 \rangle$. Repeating this argument (17) and (18) are shown for all k.

(19) Lemma: $\lambda_k = \dfrac{\langle g_k, g_k \rangle}{\langle h_k, \hat{H}_k h_k \rangle}$, where $H(z) \equiv D^2 f(z)$

and

(20) $\hat{H}_k = \displaystyle\int_0^1 H(z_k + \xi\lambda_k h_k) \, d\xi$.

[†]Daniel[1] gives most of these lemmas and states (15) for the (D) algorithm (w/o.r.).

Proof: $-g_{k+1} = Df(z_k + \lambda_k h_k)$

$\qquad\qquad\quad = Df(z_k) + \lambda_k \hat{H}_k h_k$

(21) $\qquad\qquad = -g_k + \lambda_k \hat{H}_k h_k$.

By (18), $\langle g_{k+1}, h_k \rangle = 0$, so $\langle g_k, h_k \rangle - \lambda_k \langle h_k, \hat{H}_k h_k \rangle = 0$
or by (17)

$$\lambda_k = \frac{\langle g_k, g_k \rangle}{\langle h_k, \hat{H}_k h_k \rangle} \quad .$$

(22) Lemma: $|h_{k+1}|^2 = |g_{k+1}|^2 + \gamma_k^2 |h_k|^2$; follows from (18).

(23) Lemma: $|g_k| \leq |h_k|$ by (22).

(24) Lemma: $|\lambda_k| = \dfrac{|g_k|^2}{\langle h_k, \hat{H}_k h_k \rangle} \leq \dfrac{|g_k|^2}{m |h_k|^2} \leq \dfrac{1}{m}$ by (19), (23) and (AS2).

(25) Lemma: $|g_{k+1}| \leq (1 + \dfrac{M}{m}) |h_k|$.

Proof: $|g_{k+1}| \leq |g_{k+1} - g_k| + |g_k|$

$\qquad\qquad\quad \leq |\lambda_k \hat{H}_k h_k| + |h_k|$ by (21) and (23),

$\qquad\qquad\quad \leq \dfrac{1}{m} M |h_k| + |h_k|$ by (24) and (AS2).

(26)(PR) Lemma: $\gamma_k = - \dfrac{\langle g_{k+1}, \hat{H}_k h_k \rangle}{\langle h_k, \hat{H}_k h_k \rangle}$.

Proof: $\qquad \gamma_k = - \dfrac{\langle g_{k+1}, g_k - g_{k+1} \rangle}{\langle g_k, g_k \rangle}$ by (5)(PR),

$\qquad\qquad\quad = - \dfrac{\langle g_{k+1}, \lambda_k \hat{H}_k h_k \rangle}{\langle g_k, g_k \rangle}$ by (21)

$$\gamma_k = \frac{\langle g_k, \ g_k \rangle}{\langle h_k, \ \hat{\hat{H}}_k h_k \rangle} \ \frac{\langle -g_{k+1}, \ \hat{H}_k h_k \rangle}{\langle g_k, \ g_k \rangle} \qquad \text{by (19)} \quad .$$

(26)(D) Lemma: $\quad \gamma_k = -\dfrac{\langle g_{k+1}, \ H_{k+1} h_k \rangle}{\langle h_k, \ H_{k+1} h_k \rangle} \qquad$ by definition.

(26)(FR) Lemma: $\qquad \gamma_k = 1 - \dfrac{\langle g_{k+1} + g_k, \ \hat{H}_k h_k \rangle}{\langle h_k, \ \hat{H}_k h_k \rangle} \qquad .$

Proof: $\quad \gamma_k = \dfrac{\langle g_{k+1}, \ g_{k+1} \rangle}{\langle g_k, \ g_k \rangle}$

$$= \frac{\langle g_{k+1}, \ g_{k+1} - g_k \rangle + \langle g_{k+1}, \ g_k \rangle}{\langle g_k, \ g_k \rangle}$$

$$= -\frac{\langle g_{k+1}, \hat{H}_k h_k \rangle}{\langle h_k, \ \hat{H}_k h_k \rangle} + \frac{\langle g_{k+1}, \ g_k \rangle}{\langle g_k, \ g_k \rangle} \qquad \text{by (26)(PR)}.$$

Now

$$\frac{\langle g_{k+1}, \ g_k \rangle}{\langle g_k, \ g_k \rangle} = \frac{\langle g_k - \lambda_k \hat{H}_k h_k, \ g_k \rangle}{\langle g_k, \ g_k \rangle} \qquad \text{by (21)},$$

$$= 1 - \lambda_k \frac{\langle g_k, \ \hat{H}_k h_k \rangle}{\langle g_k, \ g_k \rangle}$$

$$= 1 - \frac{\langle g_k, \ \hat{H}_k h_k \rangle}{\langle h_k, \ \hat{H}_k h_k \rangle} \qquad \text{by (19)} \quad .$$

So

$$\gamma_k = 1 - \frac{<g_{k+1} + g_k, \ \hat{H}_k h_k>}{<h_k, \ \hat{H}_k h_k>} \ .$$

(27)(PR) Lemma: $|\gamma_k| \ |h_k| \leq \frac{M}{m} |g_{k+1}| \ .$

(28)(PR) Lemma: $|\gamma_k| \leq (1 + \frac{M}{m}) \frac{M}{m} \ .$

Proof: $|\gamma_k| = \left| \frac{<g_{k+1}, \ \hat{H}_k h_k>}{<h_k, \ \hat{H}_k h_k>} \right|$ \qquad by (26)(PR)

$\leq \frac{|g_{k+1}|}{|h_k|} \frac{M}{m}$ \qquad by (AS2)

$\leq (1 + \frac{M}{m}) \frac{M}{m}$ \qquad by (25) .

(27)(D) Lemma: $|\gamma_k| \ |h_k| \leq \frac{M}{m} |g_{k+1}| \ .$

(28)(D) Lemma: $|\gamma_k| \leq (1 + \frac{M}{m}) \frac{M}{m} \ .$

Proof: Same as for (27)(PR) and (28)(PR).

(28)(FR) Lemma: $|\gamma_k| \leq (1 + \frac{M}{m})^2 \ .$

Proof: $|\gamma_k| \leq \left| \frac{<g_{k+1}, \ \hat{H}_k h_k>}{<h_k, \ \hat{H}_k h_k>} \right| + \left| \frac{<g_{k+1}, \ g_k>}{<g_k, \ g_k>} \right|$ \qquad by (26)(FR),

or

$$|\gamma_k| \leq (1 + \frac{M}{m}) \frac{M}{m} + \left| \frac{<g_k, \; g_k> - \lambda_k <g_k, \; \hat{H}_k h_k>}{<g_k, \; g_k>} \right| \quad \text{by(27)(PR) and (21)}.$$

$$\leq (1 + \frac{M}{m}) \frac{M}{m} + 1 + |\lambda_k| \left| \frac{<g_k, \; \hat{H}_k h_k>}{<g_k, \; g_k>} \right|$$

$$\leq (1 + \frac{M}{m}) \frac{M}{m} + 1 + \left| \frac{<g_k, \; \hat{H}_k h_k>}{<h_k, \; \hat{H}_k h_k>} \right| \quad \text{by (19)},$$

$$\leq (1 + \frac{M}{m}) \frac{M}{m} + 1 + \frac{M|g_k| \; |h_k|}{m|h_k|^2} \quad \text{by (A1)},$$

$$\leq (1 + \frac{M}{m}) \frac{M}{m} + 1 + \frac{M}{m} = (1 + \frac{M}{m})^2 \quad \text{by (23)}.$$

(29)(PR)

(29)(D) $\left. \right\}$ Lemma: $|h_k| \leq (1 + \frac{M}{m}) \; |g_k|$.

Proof: $|h_{k+1}| \leq |g_{k+1}| + |\gamma_k| \; |h_k|$

$$\leq |g_{k+1}| + \frac{M}{m} |g_{k+1}| \quad \text{by} \quad \left\{ \begin{array}{l} (27)(PR) \\ \\ (27)(D) \end{array} \right.$$

$$\leq (1 + \frac{M}{m}) \; |g_{k+1}| .$$

Remark: The following lemma allows one to ogtain rate of convergence 6or the (PR) and (D) algorithms without reinitialization.

(30) Lemma: $|g_k| \leq M \; |z_k - \alpha|$.

Proof: $-g_k = Df(z_k) = Df(\alpha) + \int_0^1 H(\alpha + \xi(z_k - \alpha)) \, d\xi \, (z_k - \alpha)$

so

$$|g_k| \leq \int_0^1 |H(\alpha + \xi(z_k - \alpha))| \, d\xi \, |z_k - \alpha|$$

$$\leq M \, |z_k - \alpha| \quad .$$

(31) Lemma: $|h_{k+1}| \leq 2(1 + \frac{M}{m})^2 |h_k| \quad .$

Proof: $|h_{k+1}| \leq |g_{k+1}| + |\gamma_k| \, |h_k|$

$$\leq (1 + \frac{M}{m}) \, |h_k| + (1 + \frac{M}{m})^2 \, |h_k| \quad \text{by (25) and (27).}$$

From these lemmas, the following result follows for all three algorithms:

(32) Lemma: Let $\{z_k\}$ be the sequence produced when one of the three conjugate gradient algorithms is used to solve the Minimization Problem. Then if $h_k \to 0$, there exists K and C_3 such that

(i) for the case (w/o.r.):

(33)(w/o.r.) $\qquad |\lambda_{k+1}^i h_{k+1}^i - \lambda_k^i h_k^i| \leq C_3 \, |h_k|^2 \quad i = 0, 1, \ldots, n-1$

for $k \geq K$. Similarly

(ii) for the case (w.r.):

(33)(w.r.) $\qquad |\lambda_{kr+1}^i h_{kr+1}^i - \lambda_{kr}^i h_{kr}^i| \leq C_3 \, |h_{kr}|^2 \quad i = 0, 1, \ldots, n-1$

for $k \geq K$, where the conjugate variable h_{kr} is set

equal to g_{kr} for all k and where $r \geq n$.

The proof of Theorem (32) is given in Appendix C.

(15) and (16) follow easily from Lemma (32). First, for the case (w/o.r.), the algorithms (PR) and (D) satisfy

$$(34) \qquad |h_k| \leq M(1 + \frac{M}{m}) \ |z_k - \alpha|$$

by (29) and (30). Since the algorithm converges for the given conditions on f, $z_k \to \alpha$, and so the assumption $h_k \to 0$ holds. (33)(w/o.r.) then becomes

$$(35) \qquad |\lambda_{k+1} h_{k+1} - \lambda_k^i h_k^i| \leq C_3 M^2 (1 + \frac{M}{m})^2 \ |z_k - \alpha|^2$$

for $k \geq K$, which implies (15), the desired result.

For the case (w.r.), h_{kr} is set equal to g_{kr} for $k = 0,1,\ldots$. By Lemma (31), $|h_{k+1}| \leq 2(1 + \frac{M}{m}) \ |h_k|$, so

$$|h_{kr+i}| \leq \left[2(1 + \frac{M}{m})^2\right]^i |h_{kr}| \text{ for } i = 0,1,\ldots,n-1.$$

Since the algorithm converges, Lemma (30) implies $g_k \to 0$; therefore the assumption $h_k \to 0$ holds and so (33)(w.r.) becomes

$$(36) \qquad |\lambda_{kr+1} h_{kr+1} - \lambda_{kr}^i h_{kr}^i| \leq C_3 \ |h_{kr}|^2 = C_3 \ |g_{kr}|^2 .$$

Again using (30), (36) becomes

$$(37) \qquad |\lambda_{kr+i} h_{kr+i} - \lambda^i_{kr} h^i_{kr}| \leq C_3 M^2 (1 + \tfrac{M}{m}) \quad |z_{kr} - \alpha|^2$$

which in turn implies (16).

Appendix A

Proof of Lemma (2.11)

Lemma (2.11) is very important in the development of our theory. The proof requires us to investigate carefully the inverse function theorem in finite dimensional spaces. We shall first give a theorem on invertability of matrices.

(A.1) Lemma: Suppose A is an invertible nxn matrix. Let $|A^{-1}|^{-1} = a$. Then if $|B - A| = d < a$,

(A.2) (i) B is invertible

(A.3) (ii) $|B^{-1}| \leq \dfrac{1}{a - d}$

(A.4) (iii) $b = |B^{-1}|^{-1} \leq |B| \leq |A| + d$.

Proof: (i) Given in Rudin[10], Theorem 9.8 .

$$\begin{aligned}
\text{(ii)} \quad |B^{-1}| &\leq |A^{-1}| + |A^{-1} - B^{-1}| \\
&\leq |A^{-1}| + |A^{-1}(A - B) B^{-1}| \\
&\leq |A^{-1}| + |A^{-1}| \, |A - B| \, |B^{-1}| \\
&\leq |A^{-1}| / (1 - |A - B| \, |A^{-1}|) \\
&\leq 1 / (|A^{-1}|^{-1} - |A - B|) \\
&\leq 1 / (a - d) \ .
\end{aligned}$$

(iii) $|z| = |B^{-1}B \, z| \leq |B^{-1}| |B \, z|$

so

$$b = |B^{-1}|^{-1} \leq \frac{|B \, z|}{|z|} \leq |B| \ .$$

But

$$|B| \leq |A| + |A - B| = |A| + d$$

so

$$b \leq |B| \leq |A| + d \quad .$$

(A.5) Lemma: Let $h \in C^1(R^n)$ have the following properties for all $x \in U = B_{\delta_3}(\alpha)$:

(A.6) $\qquad |Dh(x) - Dh(\alpha)| \leq \dfrac{c}{2\sqrt{n}}$

(A.7) $\qquad Dh(\alpha)$ is invertible and $|D^{-1}h(\alpha)| \leq \dfrac{6}{5c}$ $\quad .^{\dagger}$

Then

(A.8) $\qquad h$ is injective on U

(A.9) $\qquad h^{-1}\big|_{h(U)}$ is continuous

(A.10) $\qquad V = B_{\sigma}(h(\alpha)) \subset h(U)$ for $\sigma = \dfrac{1}{6} c\delta_3$

(A.11) $\qquad h^{-1} \in C^1(V)$ and $Dh^{-1}(y) = D^{-1}h(h^{-1}(y))$ for $y \in V$.

---

†Notation: $D^{-1}h(\alpha) = \big(Dh(\alpha)\big)^{-1}$.

Proof: Let $x, y \in U$, then $x + t(y - x) \in U$.

Define $H(t) = h(x + t(y - x)) - t \, Dh(\alpha) (y - x)$. Then

$$(A.12) \quad |DH(t)| = |Dh(x + t(y - x)) (y - x) - Dh(\alpha) (y - x)|$$
$$\leq |Dh(x + t(y - x)) - Dh(\alpha)| \; |y - x|$$
$$\leq \frac{c}{2\sqrt{n}} |y - x|$$

by (A.6). But

$$|y - x| = |D^{-1}h(\alpha) \, Dh(\alpha) (y - x)|$$
$$\leq |D^{-1}h(\alpha)| \; |Dh(\alpha) (y - x)|$$
$$(A.13) \quad \leq \frac{6}{5c} |Dh(\alpha) (y - x)|$$

by (A.7). Combining (A.12) and (A.13)

$$|DH(t)| \leq \frac{3}{5\sqrt{n}} |Dh(\alpha) (y - x)| .$$

By the mean value theorem, there exists t_i for $i = 1, 2, \ldots, n$ such that

$$|H_i(0) - H_i(1)| = |DH_i(t_i)|$$
$$\leq |DH(t_i)|$$
$$\leq \frac{3}{5\sqrt{n}} |Dh(\alpha) (y - x)| ,$$

therefore

$$|H(1) - H(0)|^2 = \sum_{i=1}^{n} (H_i(1) - H_i(0))^2$$

$$\leq \frac{9}{25} |Dh(\alpha) \ (y - x)|^2 .$$

But

$$H(1) - H(0) = h(y) - h(x) - Dh(\alpha) \ (y - x)$$

so

$$|h(y) - h(x) - Dh(\alpha) \ (y - x)| \leq \frac{3}{5} |Dh(\alpha) \ (y - x)| .$$

Now

$$|Dh(\alpha) \ (y - x)| \leq |h(y) - h(x)| + |h(y) - h(x) - Dh(\alpha) \ (y - x)|$$

$$\leq |h(y) - h(x)| + \frac{3}{5} |Dh(\alpha) \ (y - x)|$$

or

$$(A.14) \qquad |h(y) - h(x)| \geq \frac{2}{5} |Dh(\alpha) \ (y - x)|$$

$$\geq \frac{2}{5} \cdot \frac{5c}{6} |y - x|$$

$$\geq \frac{c}{3} |y - x| .$$

(A.14) holds for all $x, y \in U$, this implies that h is injective
on U. Let $t = h(y)$ and $s = h(x)$, then

$$|t - s| \geq \frac{c}{3} \, |h^{-1}(t) - h^{-1}(s)|$$

for $t, s \in h(U)$, so h^{-1} is continuous on $h(U)$. (A.8) and (A.9) are therefore shown.

Let $y^* \in B_\sigma(h(\alpha))$. We will find an $x^* \in U$ such that $h(x^*) = y^*$; This will prove $B_\sigma(h(\alpha)) \subset h(U)$. Define $\phi(x) = |y^* - h(x)|$. We will show ϕ has a minimum of 0 on U. Since ϕ is continuous, it attains its minimum on $\overline{B_{\delta_3}(\alpha)}$. To show the minimum is in the interior, consider x on the boundary of $\overline{B_{\delta_3}(\alpha)}$, thus $|x - \alpha| = \delta_3$. Now

$$\frac{c\delta_3}{3} = \frac{c}{3} \, |x - \alpha| \leq |h(x) - h(\alpha)| \qquad \text{by (A.14)}$$
$$\leq |h(x) - y^*| + |h(\alpha) - y^*|$$
$$\leq \phi(x) + \phi(\alpha) \quad .$$

But $\phi(\alpha) < \sigma$, so recalling $\sigma = \frac{1}{6}c\delta_3$,

$$\frac{c\delta_3}{3} < \phi(x) + \sigma = \phi(x) + \frac{1}{6} \, c\delta_3$$

or

$$\phi(x) > \frac{c\delta_3}{6} \quad .$$

Thus

$$\phi(\alpha) < \sigma = \frac{1}{6} c\delta_3 < \phi(x).$$

So ϕ does not have a minimum on the boundary.

Let x^* be the point where ϕ is minimum on $\overline{B_\delta (\alpha)}$. We know $D\phi^2(x^*) = 0$ since $\phi^2 \in C^1$. Now

$$D\phi^2(x^*) = -2 \sum_{i=1}^{n} (y_i - h_i(x^*)) \ Dh_i(x^*) = 0$$

But (A.7) implies $\{Dh_i(x^*) : i = 1,2,\ldots,n\}$ are linearly independent, thus $y^* = h(x^*)$, so we have shown (A.10). (A.11) is proven in Theorem (9.17) in Rudin[10].

(A15) Lemma: Suppose

(A.16) $f \in C^1(R^n)$ and $g_k \in C^1(R^n)$ $k = 0,1,\ldots$

(A.17) $Df(\alpha)$ is invertible and $f(\alpha) = 0$

(A.18)   for all $\epsilon \geqslant 0$ there exists K and a $\delta > 0$ such that

$k \geqslant K$ implies

$$|g_k - f|_\delta = \sup_{x \in B_\delta(\alpha)} \{|g_k(x) - f(x)| + |Dg_k(x) - Df(x)|\} < \epsilon$$

(A.19)   there exists a sequence $\{x_k\}$ such that $x_k \to \alpha$

and $g_k(x_k) = f(x_k)$

then

(A.20)   there exists a K_1 and a neighborhood of α, U, such

that for $k \geq K_1$, f and g_k are injective on U.

(A.21)   there exists a K_2 and a neighborhood of 0, W, such

that for $k \geq K_2$, $W \subset f(U)$ and $W \subset g_k(U)$.

(A.22)   for all $\hat{\epsilon} > 0$ there exists a $\hat{K}$ and a $\hat{\delta} > 0$ such that

$k \geq \hat{K}$ implies

$$|g_k^{-1} - f^{-1}|_{\hat{\delta}} = \sup_{y \in B_{\hat{\delta}}(0)} \{|g_k^{-1}(y) - f^{-1}(y)| + |Dg_k^{-1}(y) - Df^{-1}(y)|\} < \epsilon .$$

Proof: Since $f \in C^1$, there exists a $\delta_1 > 0$ such that

$x \in B_{\delta_1}(\alpha)$ implies

(A.23) $|Df(x) - Df(\alpha)| < \dfrac{c}{6\sqrt{n}}$

where $c = |D^{-1}f(\alpha)|^{-1}$. Also, by (A.18), there exists a

$\delta_2 > 0$ and a K_1 such that $k \geq K_1$ implies

(A.24) $|Dg_k(x) - Df(x)| < \dfrac{c}{6\sqrt{n}}$

for $x \in B_{\delta_2}(\alpha)$. Lemma (A.1) and (A.23) imply that $Df(x)$ is

invertible for all $x \in B_{\delta_1}(\alpha)$ and that

$$(A.25) \qquad |D^{-1}f(x)|^{-1} \geq |D^{-1}f(\alpha)|^{-1} - |Df(\alpha) - Df(x)|$$

$$\geq \frac{c}{6\sqrt{n}} [6\sqrt{n} - 1]$$

$$\geq \frac{5c}{6\sqrt{n}} \quad .$$

So

$$(A.26) \qquad |Dg_k(x) - Df(x)| < \frac{c}{6\sqrt{n}} < |D^{-1}f(x)|^{-1}$$

for $x \in B_{\delta_1}(\alpha) \cap B_{\delta_2}(\alpha)$ and $k \geq K_1$ which, together with Lemma (A.1) implies that $Dg_k(x)$ is invertible for $x \in B_{\delta_3}(\alpha)$ and $k \geq K_1$ where $\delta_3 = \min(\delta_1, \delta_2)$. We also have, from (A.24) and (A.26) that $x \in B_{\delta_3}(\alpha)$ and $k \geq K_1$ implies

$$|Dg_k(x) - Dg_k(\alpha)| \leq |Dg_k(x) - Df(x)| + |Df(x) - Df(\alpha)|$$
$$+ |Df(\alpha) - Dg_k(\alpha)|$$
$$\leq \frac{c}{2\sqrt{n}}$$

and from Lemma (A.1)

$$|D^{-1}g_k(\alpha)| \leq \left(|D^{-1}f(\alpha)|^{-1} - |Dg_k(\alpha) - Df(\alpha)| \right)^{-1}$$

$$\leq [c - \frac{1}{6\sqrt{n}}c]^{-1}$$

$$\leq \frac{6}{5c} \quad .$$

To recapitulate, for $k \geq K_1$ and $x \in B_{\delta_3}(\alpha)$,

(A.27) $Df(x)$ and $Dg_k(x)$ are invertible

(A.28) $\left| D^{-1}f(\alpha) \right| = \frac{1}{c}$ and $\left| D^{-1}g_k(\alpha) \right| \leq \frac{6}{5c}$

(A.29) $\left| Df(x) - Df(\alpha) \right| \leq \frac{c}{6\sqrt{n}}$ and $\left| Dg_k(x) - Dg_k(\alpha) \right| \leq \frac{c}{2\sqrt{n}}$.

Let $U = B_{\delta_3}(\alpha)$. Then for $k \geq K_1$, both f and g_k have
properties (A.6), and (A.7) in Lemma (A.5). So both f and
g_k are injective on U, $f^{-1}\Big|_{f(U)}$ and $g_k^{-1}\Big|_{g_k(U)}$ are continuous,

(A.30) $B_\sigma(g_k(\alpha)) \subset g_k(U)$

and

(A.31) $B_\sigma(f(\alpha)) \subset f(U)$

for $\sigma = \frac{1}{6}c\delta_3$. Thus (A.20) is shown. To show (A.21), note that
(A.18) implies there exists a K' such that $k \geq K'$ implies

$$\left| g_k(\alpha) - f(\alpha) \right| = \left| g_k(\alpha) \right| < \frac{\sigma}{2} .$$

So $0 \in B_{\frac{\sigma}{2}}(g_k(\alpha))$ for all $k \geq K_2 = \max(K_1, K')$. Moreover,

(A.32) $\qquad B_{\frac{\sigma}{2}}(0) \subset B_\sigma(g_k(\alpha))$

for $k \geq K_2$, since if $y \in B_{\frac{\sigma}{2}}(0)$,

$$|y - g_k(\alpha)| \leq |y| + |g_k(\alpha)| \leq \frac{\sigma}{2} + \frac{\sigma}{2} \leq \sigma \quad .$$

Let $W = B_{\frac{\sigma}{2}}(0)$, then by (A.31), $W \subset f(U)$, and by (A.32), $W \subset g_k(U)$ for $k \geq K_2$. We have therefore shown (A.21).

In order to show (A.22), note that

$$|g_k^{-1}(y) - f^{-1}(y)| \leq |g_k^{-1}(y) - x_k| + |f^{-1}(y) - x_k|$$

where $\{x_k\}$ is given in (A.19) (so $f(x_k) = g_k(x_k)$). Since $x_k \to \alpha$ and $f(x_k) \to 0$, there exists a $\hat{K}_1$ such that $k \geq \hat{K}_1$ implies that $x_k \in U$ and $f(x_k) \in B_{\frac{c\hat{e}}{24}}(0)$. Let $y \in W$, then by (A.14) in Lemma (A.5) ,

$$|g_k^{-1}(y) - x_k| \leq \frac{3}{c}|y - g_k(x_k)|$$

$$\leq \frac{3}{c}[|y| + |g_k(x_k)|]$$

for $k \geq \hat{K}_2 = \max(K_1, \hat{K}_1)$, and

$$|f^{-1}(y) - x_k| \leq \frac{3}{c}[|y| + |f(x_k)|]$$

so

(A.33) $|g_k^{-1}(y) - f^{-1}(y)| \leq \dfrac{6}{c}[|y| + |f(x_k)|]$.

Now if $y \in B_{\hat{\delta}_1}(0)$ where $\hat{\delta}_1 = \min(\dfrac{\sigma}{2}, \dfrac{c\hat{\epsilon}}{24})$ and if $k \geq \hat{K}_2$, we have

(A.34) $|g_k^{-1}(y) - f^{-1}(y)| \leq \dfrac{6}{c}\left[\dfrac{c\hat{\epsilon}}{24} + \dfrac{c\hat{\epsilon}}{24}\right]$

$$\leq \dfrac{\hat{\epsilon}}{2} .$$

By (A.11), the functions f^{-1} and $\{g_k^{-1} : k \geq \hat{K}_2\}$ are in $C^1(W)$, $Df^{-1}(y) = D^{-1}f(f^{-1}(y))$ and $Dg_k^{-1}(y) = D^{-1}g_k(g_k^{-1}(y))$. Thus

(A.35) $|Df^{-1}(y) - Dg_k^{-1}(y)| \leq |D^{-1}f(f^{-1}(y)) - D^{-1}g_k(g_k^{-1}(y))|$

$$\leq |D^{-1}f(f^{-1}(y))| \; |D^{-1}g_k(g_k^{-1}(y))|$$

$$\cdot |Df(f^{-1}(y)) - Dg_k(g_k^{-1}(y))|$$

$$\leq \dfrac{36n}{c^2} |Df(f^{-1}(y)) - Dg_k(g_k^{-1}(y))|$$

since by (A.25), $|D^{-1}f(x)| \leq \dfrac{6\sqrt{n}}{c}$ for all $x \in B_{\delta_1}(\alpha) \subset U$ and by Lemma (A.1), (A.24), and (A.25):

$$|D^{-1}g_k(x)| \leq [|D^{-1}f(x)|^{-1} - |Dg_k(x) - Df(x)|]^{-1}$$

$$\leq \dfrac{c}{6\sqrt{n}}\left[(6\sqrt{n} - 1) - \dfrac{c}{6\sqrt{n}}\right]^{-1}$$

$$\leq \dfrac{6\sqrt{n}}{c}$$

for all $x \in B_{\hat{\delta}_1}(\alpha) \supset U$. Now

(A.36)
$$|Df(f^{-1}(y)) - Dg_k(g_k^{-1}(y))| \leq |Df(f^{-1}(y)) - Df(g_k^{-1}(y))|$$
$$+ |Df(g_k^{-1}(y)) - Dg_k(g_k^{-1}(y))| .$$

By continuity of Df and (A.33) there exists a $\hat{\delta}_2$ and a $\hat{K}_3$ such that $y \in B_{\hat{\delta}_2}(0)$ and $k \geq \hat{K}_3$ implies

(A.37)
$$|Df(f^{-1}(y)) - Df(g_k^{-1}(y))| \leq \frac{c^2}{36n} \cdot \frac{\hat{\epsilon}}{4} .$$

From (A.18), there exists a $\hat{\delta}_3$ and a $\hat{K}_4$ such that $x \in B_{\hat{\delta}_3}(0)$ and $k \geq \hat{K}_4$ imply

(A.38)
$$|Df(x) - Dg_k(x)| \leq \frac{c^2}{36n} \cdot \frac{\hat{\epsilon}}{4} .$$

Now by (A.14) in Lemma (A.5), (since g_k has properties (A.6) and (A.7))

(A.39)
$$|\tilde{g}_k^{-1}(y) - \alpha| \leq \frac{3}{c}|y - g_k(\alpha)| \leq \frac{3}{c}(|y| + |g_k(\alpha)|)$$

for $k \geq K_1$. By (A.18) there exists $\hat{K}_5$ such that $k \geq \hat{K}_5$, implies

(A.40)
$$|g_k(\alpha)| \leq \frac{\hat{\delta}_3}{2} \cdot \frac{c}{3} .$$

Let $\hat{\delta}_4 = \frac{c}{3} \cdot \frac{\hat{\delta}_3}{2}$. So if $y \in B_{\hat{\delta}_4}(0)$ and $k \geq \hat{K}_5$, then (A.39) and (A.40) give $|g_k^{-1}(y) - \alpha| \leq \hat{\delta}_3$ or

(A.41) $\qquad g_k^{-1}(y) \in B_{\hat{\delta}_3}(\alpha)$.

If $y \in B_{\hat{\delta}_4}(0)$ and $k \geq \max(\hat{K}_4, \hat{K}_5)$ then

(A.42) $\qquad |Df(g_k^{-1}(y)) - Dg_k(g_k^{-1}(y))| \leq \frac{c^2}{36n} \cdot \frac{\hat{\varepsilon}}{4}$

using (A.38) and (A.41). Let $\hat{\delta}_5 = \min(\hat{\delta}_2, \hat{\delta}_4)$ and $\hat{K}_6 = \max(\hat{K}_3, \hat{K}_4, \hat{K}_5)$, then by (A.37), (A.38) and (A.42), $y \in B_{\hat{\delta}_5}(0)$ and $k \geq \hat{K}_5$ imply that

$$|Df(f^{-1}(y)) - Dg_k(g_k^{-1}(y))| \leq \frac{c^2}{36n} \cdot \frac{\hat{\varepsilon}}{4} + \frac{c^2}{36n} \cdot \frac{\hat{\varepsilon}}{4}$$
$$\leq \frac{c^2}{36n} \cdot \frac{\hat{\varepsilon}}{2} \qquad .$$

So, $y \in B_{\hat{\delta}_5}(0)$, $k \geq \hat{K}_5$ implies

(A.43) $\qquad |Df^{-1}(y) - Dg_k^{-1}(y)| \leq \frac{\hat{\varepsilon}}{2}$

from (A.35). Let $\hat{\delta} = \min(\hat{\delta}_1, \hat{\delta}_5)$ and $\hat{K} = \max(\hat{K}_2, \hat{K}_5)$, then by (A.34) and (A.43),

$$|g_k^{-1}(y) - f^{-1}(y)| + |Dg_k^{-1}(y) - Df^{-1}(y)| < \hat{\varepsilon}$$

for $y \in B_{\hat{\delta}}(0)$, $k \geq \hat{K}$ or

$$|g_k^{-1} - f^{-1}|_{\hat{\delta}} \leq \hat{\epsilon} \text{ for } k \geq \hat{K} ,$$

so (A.22) is proven.

We are now able to prove Lemma (11).

(A.44) Theorem: Let $\mathcal{F} = \{f \in C^1(R^n) : f$ has a zero at which Df is invertable$\}$. Let $f \in \mathcal{F}$ and let α be the zero at which $Df(\alpha)$ is invertible. Let $\{x_k\}$ be a sequence such that $x_k \to \alpha$ and let $\{g_k\}$ be a sequence of functions in $\mathcal{F}$ such that $g_k(x_k) = f(x_k)$ and $g_k \to f$ uniformly at α. Then there exists an m, M and K and roots (g_k) of g_k such that k K implies

$$0 < m \leq \frac{|x_k - \alpha|}{|x_k - \alpha(g_k)|} \leq M .$$

Proof: let $|Df(\alpha)| = b$. By continuity of Df, there exists a ball N_1 about α such that $x \in N_1$ implies $|Df(x)| \leq 2b$. Let $|Df^{-1}(0)| = c$. By the inverse function theorem, there exists a neighborhood of 0 where Df^{-1} is continuous. Thus there exists a ball P_1 about 0 such that $y \in P_1$ implies $|Df^{-1}(y)| \leq 2c$. Since $f \in C^1$ and $g_k \in C^1$ for all k, we can apply the fundemental theorem of calculus and get

$$f(x_k) = f(\alpha) + \int_0^1 Df(\alpha + t(x_k - \alpha)) (x_k - \alpha) \, dt$$

or

$$(A.45) \qquad |f(x_k)| \leq \sup_{t \in (0,1)} \{|Df(\alpha + t(x_k - \alpha))|\} |x_k - \alpha|$$

Similarly,

$$(A.46) \qquad |g_k(x_k)| \leq \sup_{t \in (0,1)} \{|Dg_k(\alpha(g_k) + t(x_k - \alpha(g_k)))|\} |x_k - \alpha(g_k)|$$

where $\alpha(g_k)$ is the zero of g_k. Since the assumptions

of Lemma (A.15) are satisfied, the results (A.20), (A.21), and

(A.22) hold. Noting that (A.20), (A.21) and the inverse

function theorem imply that $f^{-1} \in C^1(W)$ and $g_k^{-1} \in C^1(W)$ for

$k \geq K$, we can again apply the fundemental theorem to get

for $y \in W$

$$(A.47) \qquad |f^{-1}(y) - f^{-1}(0)| \leq \sup_{t \in (0,1)} \{|Df^{-1}(ty)|\} |y|$$

and

$$(A.48) \qquad |g_k^{-1}(y) - g_k^{-1}(0)| \leq \sup_{t \in (0,1)} \{|Dg_k^{-1}(ty)|\} |y| .$$

Let $k \geq K_1$ imply that $x_k \in N_1$ and $f(x_k) \in P_1 \cap W$, then (A.45)

implies

$$(A.49) \qquad |f(x_k)| \leq 2b |x_k - \alpha| .$$

Let $y = f(x_k)$ in (A.47), then

(A.50) $\qquad |f^{-1}(f(x_k)) - f^{-1}(0)| = |x_k - \alpha| \leq 2c\,|y| = 2c\,|f(x_k)|$.

Since $g_k \to f$ uniformly at α (by assumption) and since $g_k^{-1} \to f^{-1}$ uniformly at 0 by (A.22), there exists K_2, δ_2 and $\hat{\delta}_2$ such that $k \geq K_2$, $x \in B_{\delta_2}(\alpha)$, and $y \in B_{\hat{\delta}_2}(0)$ implies $|Dg_k(x)| \leq 4b$ and $|Dg_k^{-1}(y)| \leq 4c$. Also, there exists a K_3 such that $k \geq K_3$ implies $|g_k^{-1}(0) - f^{-1}(0)| = |\alpha(g_k) - \alpha| < \delta_2$, which implies $\alpha(g_k) \in B_{\delta_2}(\alpha)$. Let $k \geq K_4$ imply that $x_k \in B_{\delta_2}(\alpha)$ and $f(x_k) \in B_{\hat{\delta}_2}(\alpha)$ then by (A.46), $k \geq K = \max(\hat{K}, K_1, K_2, K_3, K_4)$ implies

(A.51) $\qquad |g_k(x_k)| = |f(x_k)| \leq 4b\,|x_k - \alpha(g_k)|$.

By (A.48), letting $y = f(x_k) = g_k(x_k)$, $k \geq K$, we get

(A.52) $\qquad |x_k - \alpha(g_k)| \leq 4c\,|f(x_k)|$.

So combining (A.49), (A.50), (A.51), and (A.52) we have for $k \geq K$ that

$$0 < \frac{1}{8bc} \leq \frac{|x_k - \alpha|}{|x_k - \alpha(g_k)|} \leq 8bc \quad .$$

Appendix B

Rate of Convergence of the Newton–Raphson Algorithm

Theorem: Let f be a function from $R^n \to R$ that satisfies

assumptions (AS1) and (AS2), that is, $f \in C^3$ and there exists

$m > 0$ and M such that $m <y,y> \leq <y,D^2f(z)y> \leq M<y,y>$ $\forall y, z \in R^n$.

Then there exists a constant C such that

$$\overline{\lim_{x \to \alpha}} \frac{|\hat{\phi}(z) - \alpha|}{|z - \alpha|^2} \leq C < \infty$$

where $\hat{\phi}(z) = z - (D^2f(z))^{-1} Df(z)$ is the Newton–Raphson

algorithm and α is a stationary point of f.

Proof: Let $g = Df$. Then $g \in C^2 (R^n)$ and $\hat{\phi}(z) = z - (Dg(z))^{-1}g(z)$.

Since D^2g is continuous, there exists a B, such that

$|D^2g(z)| \leq B$ $\forall z \in B_1(\alpha)$. Also since $g(\alpha) = 0$ and

$|(Dg(z))^{-1}g(z)| \leq \frac{1}{m} |g(z)|$, there exists a ρ such that $z \in B_\rho(\alpha)$

implies $\hat{\phi}(z) \in B_1(\alpha)$.

So by Taylor's Theorem,

$$g(\hat{\phi}(z)) = g\left[z - (Dg(z))^{-1}g(z)\right]$$

$$= g(z) + Dg(z) (- (Dg(z))^{-1}g(z))$$

$$+ \int_0^1 (1-t) D^2g\left\{z-t(Dg(z))^{-1}g(z)\right\} \left[(Dg(z))^{-1}g(z)\right]^2 dt$$

or

$$|g(\hat{\phi}(z))| \leq \left[\int_0^1 (1-t) \ |D^2g\{z-t(Dg(z))^{-1}g(z)\}| \ dt\right] |[D(g(z))]^{-1}g(z)|^2$$

$$\leq \left[\int_0^1 (1-t) \ B \ dt\right]\frac{1}{m^2} \ |g(z)|^2$$

(B.1) $$\leq \frac{B}{2m^2} \ |g(z)|^2 \ .$$

for $z \in B_\rho(\alpha)$.

Now again by Taylor's Theorem:

$$|g(z)| = \left|g(\alpha) + \int_0^1 Dg(\alpha + t(z-\alpha)) \ (z-\alpha) \ dt\right|$$

$$\leq \int_0^1 |Dg(\alpha + t(z-\alpha))| dt |z-\alpha| \ ,$$

(B.2) $$\leq M \ |z-\alpha| \ .$$

Therefore by (B.1) and (B.2),

(B.3) $$|g(\hat{\phi}(z))| \leq \frac{BM^2}{2m^2} \ |z-\alpha|^2$$

for $z \in B_\rho(\alpha)$. Also

$$g(\hat{\phi}(z)) = g(\alpha) + \int_0^1 Dg[\alpha + t(\hat{\phi}(z) - \alpha)] \ (\hat{\phi}(z) - \alpha)dt,$$

so

$$\langle \hat{\Phi}(z)-\alpha, \ g(\hat{\Phi}(z))\rangle = \int_0^1 \langle \hat{\Phi}(z)-\alpha, \ Dg(\alpha+t(\hat{\Phi}(z)-\alpha)) \ (\hat{\Phi}(z)-\alpha)\rangle \ dt.$$

We then obtain

$$m|\hat{\Phi}(z) - \alpha|^2 \leq \langle \hat{\Phi}(z) - \alpha, \ g(\hat{\Phi}(z))\rangle \leq |\hat{\Phi}(z) - \alpha| \ |g(\hat{\Phi}(z))|$$

or

$$(B.4) \qquad |\hat{\Phi}(z) - \alpha| \leq \frac{1}{m} |g(\hat{\Phi}(z))| \quad .$$

Therefore by (B.3) and (B.4),

$$(B.5) \qquad |\hat{\Phi}(z) - \alpha| \leq \frac{BM^2}{2m^3} |z - \alpha|^2$$

$$\leq C \ |z - \alpha|^2 \quad \text{for } z \ \varepsilon \ B_\rho(\alpha) \quad ,$$

which in turn implies the desired result.

Appendix C

Proof of Lemma (4.32)

We shall only prove (4.33)(w/o.r.) since the proof of
(4.33)(w.r.) will follow if we assume k is a multiple of r.
The proof is by induction and requires one to show for
i = 0,1,...,n-1 that

(C.1) $\quad |g_{k+1} - g_k^i| \leqslant 0(|h_k|^2)$

(C.2) $\quad |h_{k+1} - h_k^i| \leq 0(|h_k|^2)$

(C.3) $\quad |\lambda_{k+1}h_{k+1} - \lambda_k^i h_k^i| \leqslant 0(|h_k|^2)$

where $0 : R \to R$ has the property that there exists an r and c
such that

$\quad |0(\xi)| \leq c|\xi|$ for all $\xi \in B_r(0)$.

We use the fact that the algorithms converge, that is,

$z_k \to \alpha$ where $Df(\alpha) = 0$. This implies, since $f \in C^3$, that there exists K_1 and B such that $|D^3 f(z_k)| \leq B$ for all $k \geq K_1$.

It is now convenient to prove several lemmas.

(C.4) Lemma: $|H_{k+1} - H_k| \leq 0(|h_k|)$.

Proof: $|H_{k+1} - H_k| \leq \displaystyle\sum_{\ell=0}^{i-1} |H_{k+\ell+1} - H_{k+\ell}|$

Now

$$|H_{k+\ell+1} - H_{k+\ell}| = |H(z_{k+\ell} + \lambda_{k+\ell} h_{k+\ell}) - H(z_{k+\ell})|$$

$$= \left| \int_0^1 D^3 f(z_{k+\ell} + \xi \lambda_{k+\ell} h_{k+\ell}) \lambda_{k+\ell} h_{k+\ell} d\xi \right|$$

$$\leq \sup_{\xi \in (0,1)} \left\{ |D^3 f(z_{k+\ell} + \xi \lambda_{k+\ell} h_{k+\ell})| \right\} |\lambda_{k+\ell}| |h_{k+\ell}|$$

$$\leq \frac{B}{m} |h_{k+\ell}| \quad \text{for } k \geq K_1, \text{ by (4.24)}.$$

$$\leq 0(|h_k|) \qquad\qquad \text{by (4.31)}.$$

Therefore

$$|H_{k+1} - H_k| \leq 0(|h_k|).$$

(C.5) Lemma: $|\hat{H}_{k+1} - H_k| \leq 0(|h_k|)$.

Proof: $|\hat{H}_{k+1} - H_k| \leq |\hat{H}_{k+1} - H_{k+1}| + |H_{k+1} - H_k|$.

Now

$$|\hat{H}_{k+1} - H_{k+1}| = |\int_0^1 H(z_{k+1} + \xi\lambda_{k+1}h_{k+1})d\xi - H_{k+1}|$$

$$(C.6) \qquad \le \int_0^1 |H(z_{k+1} + \xi\lambda_{k+1}h_{k+1}) - H_{k+1}|d\xi$$

and

$$|H(z_{k+1} + \xi\lambda_{k+1}h_{k+1}) - H_{k+1}| = |\int_0^1 D^3f(z_{k+1} + \eta\xi\lambda_{k+1}h_{k+1})\ \xi\lambda_{k+1}h_{k+1}d\eta|$$

$$\le \sup_{\eta\varepsilon(0,1)}\{|D^3f(z_{k+1} + \eta\lambda_{k+1}h_{k+1})|\}|\lambda_{k+1}||h_{k+1}|$$

$$\text{since } \xi\varepsilon(0,1),$$

$$\le \frac{B}{m}|h_{k+1}|.$$

Thus

$$|\hat{H}_{k+I} - H_{k+1}| \le \frac{B}{m}|h_{k+1}| \le 0(|h_k|) \qquad \text{by (4.31)}.$$

This and Lemma (C.4), gives the desired result.

The expression for γ_k, given in (4.26), is different for the three algorithms. The following lemma, however, will be proved in detail only for (PR) since the proofs for the other algorithms are almost identical.

(C.7)(PR)
(C.7)(D) } Lemma: $|h_{k+i+1} - h_k^{i+1}| \leq 0_1(|h_{k+i} - h_k^i|)$

$$+ 0_2(|g_{k+i+1} - g_k^{i+1}|) + 0_3(|h_k|^2).$$

(C.7)(FR)Lemma: $|h_{k+i+1} - h_k^{i+1}| \leq 0_1(|h_{k+i} - h_k^i|) + 0_2(|g_{k+i+1} - g_k^{i+1}|)$

$$+ 0_3(|g_{k+i} - g_k^i|) + 0_4(|h_k|^2).$$

Proof (for (PR)): $|h_{k+i+1} - h_k^{i+1}| = |g_{k+i+1} + \gamma_{k+i} h_{k+i} - g_k^{i+1} - \gamma_k^i h_k^i|$

$$\leq |g_{k+i+1} - g_k^{i+1}| + |\gamma_{k+i} h_{k+i} - \gamma_k^i h_k^i|.$$

Now by (4.26)(PR)

$$|\gamma_{k+i} h_{k+i} - \gamma_k^i h_k^i| = \left| \frac{\langle g_{k+i+1}, \hat{H}_{k+i} h_{k+i} \rangle \cdot h_{k+i}}{\langle h_{k+i}, \hat{H}_{k+i} h_{k+i} \rangle} - \frac{\langle g_k^{i+1}, H_k h_k^i \rangle \cdot h_k^i}{\langle h_k^i, H_k h_k^i \rangle} \right|.$$

Let

(C.8) $c_{k+i} = \langle h_{k+i}, \hat{H}_{k+i} h_{k+i} \rangle \langle h_k^i, H_k h_k^i \rangle,$

so

$$|\gamma_{k+i} h_{k+i} - \gamma_k^i h_k^i| = \frac{1}{c_{k+i}} \Big| \langle g_{k+i+1}, \hat{H}_{k+i} h_{k+i} \rangle \langle h_k^i, H_k h_k^i \rangle h_{k+i}$$

$$- \langle g_k^{i+1}, H_k h_k^i \rangle \langle h_{k+i}, \hat{H}_{k+i} h_{k+i} \rangle h_k^i \Big|$$

$$\le \frac{1}{c_{k+1}} \Bigg\} \; \Big| <g_{k+i+1}, \; \hat{H}_{k+1}(h_{k+1} - h_k^i)> <h_k^i, \; H_k h_k^i> \; h_{k+1} \Big|$$

$$+ \Big| <g_{k+i+1}, \; \hat{H}_{k+1} h_k^i> <h_k^i - h_{k+1}, \; H_k h_k^i> \; h_{k+1} \Big|$$

$$+ \Big| <g_{k+i+1} - g_k^{i+1}, \; \hat{H}_{k+1} h_k^i> <h_{k+1}, \; H_k h_k^i> \; h_{k+1} \Big|$$

$$+ \Big| <g_k^{i+1}, \; \hat{H}_{k+1} h_k^i> <h_{k+1}, \; H_k(h_k^i - h_{k+1})> \; h_{k+1} \Big|$$

$$+ \Big| <g_k^{i+1}, \; (\hat{H}_{k+1} - H_k)h_k^i> <h_{k+1}, \; H_k h_{k+1}> \; h_{k+1} \Big|$$

$$+ \Big| <g_k^{i+1}, \; H_k h_k^i> <h_{k+1}, \; (H_k - \hat{H}_{k+1}) \; h_{k+1}> \; h_{k+1} \Big|$$

$$+ \Big| <g_k^{i+1}, \; H_k h_k^i> <h_{k+1}, \; \hat{H}_{k+1} h_{k+1}> \; (h_{k+1} - h_k^i) \Big| \Bigg\} \; .$$

Noting that

$$\frac{1}{c_{k+1}} \le \frac{1}{m^2 \; |h_{k+1}|^2 \; |h_k^i|^2} \; ,$$

and using Lemmas (4.25), (4.31) and (6.5), the result is obtained.

(C.9) Lemma: $\quad |g_{k+i+1} - g_k^{i+1}| \le |g_{k+i} - g_k^i| + O(|h_k|^2)$

$$+ M \; |\lambda_{k+i} h_{k+i} - \lambda_k^i h_k^i| \; .$$

Proof: $\quad |g_{k+i+1} - g_k^{i+1}| = |g_{k+i} - \lambda_{k+i} \hat{H}_{k+1} h_{k+1} - g_k^i + \lambda_k^i H_k^i h_k^i| \;$ by (4.21).

$$\le |g_{k+i} - g_k^i| + |(H_k - \hat{H}_{k+1})\lambda_k^i h_k^i| + |\hat{H}_{k+1}(\lambda_k^i h_k^i - \lambda_{k+i} h_{k+1})|$$

$$\stackrel{\leq}{=} |g_{k+1} - g_k^i| + \frac{1}{m}|H_k - \hat{H}_{k+1}||h_k^i| + M|\lambda_k^i h_k^i - \lambda_{k+1} h_{k+1}| \quad \text{by}(4.24)$$

$$\stackrel{\leq}{=} |g_{k+1} - g_k^i| + 0(|h_k|^2) + M|\lambda_k^i h_k^i - \lambda_{k+1} h_{k+1}| \quad \text{by}(4.31) \text{ and } (C.4).$$

(C.10) Lemma: $\quad |\lambda_{k+1} h_{k+1} - \lambda_k^i h_k^i| \leqslant 0_1(|g_{k+1} - g_k^i|)$

$$+ 0_2(|h_{k+1} - h_k^i|) + 0_3(|h_k|^2) .$$

Proof: $\quad |\lambda_{k+1} h_{k+1} - \lambda_k^i h_k^i| = \left| \dfrac{|g_{k+1}|^2 h_{k+1}}{\langle h_{k+1}, \hat{H}_{k+1} h_{k+1}\rangle} - \dfrac{|g_k^i|^2 h_k^i}{\langle h_k^i, H_k h_k^i\rangle} \right|$

$$= \frac{1}{c_{k+1}} \left| |g_{k+1}|^2 \langle h_k^i, H_k h_k^i\rangle h_{k+1} - |g_k^i|^2 \langle h_{k+1}, \hat{H}_{k+1} h_{k+1}\rangle h_k^i \right|$$

$$\stackrel{\leq}{=} \frac{1}{c_{k+1}} \left\{ \left| \langle g_{k+1}, g_{k+1} - g_k^i\rangle \langle h_k^i, H_k h_k^i\rangle h_{k+1} \right| \right.$$

$$+ \left| \langle g_{k+1}, g_k^i\rangle \langle h_k^i - h_{k+1}, H_k h_k^i\rangle h_{k+1} \right|$$

$$+ \left| \langle g_{k+1} - g_k^i, g_k^i\rangle \langle h_{k+1}, H_k h_k^i\rangle h_{k+1} \right|$$

$$+ \left| |g_k^i|^2 \langle h_{k+1}, H_k(h_k^i - h_{k+1})\rangle h_{k+1} \right|$$

$$+ \left| |g_k^i|^2 \langle h_{k+1}, H_k h_{k+1}\rangle (h_{k+1} - h_k^i) \right|$$

$$+ \left. \left| |g_k^i|^2 \langle h_{k+1}, (H_k - \hat{H}_{k+1}) h_{k+1}\rangle h_k^i \right| \right\} .$$

where c_{k+1} is given in (C.8). Noting that $\dfrac{1}{c_{k+1}} \stackrel{\leq}{=} \dfrac{1}{m^2 |h_{k+1}|^2 |h_k^i|^2}$,

$|g_{k+1}| \stackrel{\leq}{=} |h_{k+1}|$ and $|g_k^i| \stackrel{\leq}{=} |h_k^i|$ (by Lemma (4.23)),

$|H_k - \hat{h}_{k+1}| \leq 0(|h_k|)$ (by (C.5)), and $|h_k^i| \leq 0(|h_k|)$ (by Lemma (4.31)), the desired result is readily obtained.

The proof of (C.1) through (C.3) now follows easily. First for $i = 0$:

(C.11) $|g_k - g_k^0| = 0$ since $g_k = g_k^0$,

(C.12) $|h_k - h_k^0| = 0$ since $h_k = h_k^0$, and

(C.13) $|\lambda_k h_k - \lambda_k^0 h_k^0| \leq 0(|h_k|^2)$ by (C.10).

So (C.1) through (C.3) hold for $i = 0$. Now assuming they hold for i, that is:

(C.14) $|g_{k+1} - g_k^i| \leq 0(|h_k|^2)$

(C.15) $|h_{k+1} - h_k^i| \leq 0(|h_k|^2)$

(C.16) $|\lambda_{k+1} h_{k+1} - \lambda_k^i h_k^i| \leq 0(|h_k|^2)$,

it is necessary to show they hold for $i + 1$. By (C.9),

(C.17) $|g_{k+1+1} - g_k^{i+1}| \leq |g_{k+1} - g_k^i| + 0(|h_k|^2)$

$$+ M |\lambda_{k+1} h_{k+1} - \lambda_k^i h_k^i|$$

$$\leq 0(|h_k|^2) \text{ using (C.14) and (C.16)}.$$

By (C.7)(PR) or (D)

$$|h_{k+i+1} - h_k^{i+1}| \leqslant 0_1(|h_{k+i} - h_k^i|) + 0_2(|g_{k+i+1} - g_k^{i+1}|) + 0_3(|h_k|^2)$$

and by (C.7)(FR)

$$|h_{k+i+1} - h_k^{i+1}| \leqslant 0_1(|h_{k+i} - h_k^i|) + 0_2(|g_{k+i+1} - g_k^{i+1}|)$$
$$+ 0_3(|g_{k+i} - g_k^i|) + 0_4(|h_k|^2) \quad .$$

So

(C.18) $\qquad |h_{k+i+1} - h_k^{i+1}| \leqslant 0(|h_k|^2) \quad$ by (C.14), (C.15) and (C.16).

And finally by (C.10),

(C.19) $\qquad |\lambda_{k+i+1} h_{k+i+1} - \lambda_k^{i+1} h_k^{i+1}|$

$\qquad\qquad \leqslant 0_1(|g_{k+i+1} - g_k^{i+1}|) + 0_2(|h_{k+i+1} - h_k^{i+1}|) + 0_3(|h_k|^2)$

$\qquad\qquad \leqslant 0(|h_k|^2) \quad$ using (C.17) and (C.18).

The induction proof is therefore complete.

References

1. Daniel, J.W., "The Conjugate Gradient Method for Linear and Nonlinear Operator Equations," Siam. J. Num. Anal.. vol.4, no11(1967), pp. 10-26.

2. Daniel, J.W., "Convergence of the Conjugate Gradient Method with Computationally Convenient Modifications," Numerische Mathematik, 10(1967), pp. 125-131.

3. Dieudonné, J., Foundations of Modern Analysis, Academic Press, New York, 1960.

4. Fletcher, R. and Reeves, C.M., "Function Minimization by Conjugate Gradients," Comput. J., 7(1964), pp.145-154.

5. Goldstein, A.A. and Price, J.F., "An Effective Algorithm for Minimization," Numerische Mathematik, 10(1967),pp. 184-189.

6. Jeeves, T.A., "Secant Modification of Newton's Method," Comm. ACM11, 8(1958), pp. 9-10.

7. Kincaid, W.M., "Solution of Equations by Interpolation," Ann. Math. Statist., 19(1958), pp. 207-219.

8. Ostrowski, A., Solution of Equations and Systems of Equations, Academic Press, New York, 1966.

9. Polak, E. and Ribiere, G., "Note Sur La Cönvergence de Methods de Direction Conjuguees," R.I.R.O., 16(1969).

10. Rudin, W., Principles of Mathematical Analysis, McGraw Hill, New York, 1964.

11. Traub, J.F., Iterative Methods for the Solution of Equations, Prentice-Hall, Englewood Cliffs, New Jersey, 1964.

12. Wolfe, P., "The Secant Method for Simultaneous Nonlinear

  Equations"," Comm. ACM 2, 12 (1959), pp. 12-13.