EFFICIENT IMPLEMENTATIONS OF THE POLAK-RIBIÈRE CONJUGATE GRADIENT ALGORITHM

by

R. Klessig and E. Polak

Memorandum No. ERL-M279

6 August 1970

Electronics Research Laboratory

College of Engineering
University of California, Berkeley
94720

EFFICIENT IMPLEMENTATIONS OF THE POLAK-RIBIÈRE CONJUGATE GRADIENT ALGORITHM*

R. Klessig and E. Polak†

I. INTRODUCTION

Quite commonly, theoretical algorithms are stated as a recurrence relation of the form $x_{i+1} \in A(x_i)$, $i = 0, 1, 2, \ldots$, where $A(\cdot)$ is a set valued function, and the sequence $\{x_i\}$ converges to a solution point. Also quite commonly, to compute a vector x_{i+1} in the set $A(x_i)$, we must bring in a subalgorithm which starts out by setting $y_0 = x_0$, and then constructs an infinite sequence $y_0, y_1, y_2, \ldots$ which converges to a point x_{i+1} in $A(x_i)$. Consequently, from a constructive point of view, such an algorithm ($x_{i+1} \in A(x_i)$) is not well defined because it is doubly infinitely iterative. The problem of implementing an algorithm of the form $x_{i+1} \in A(x_i)$ is that of finding an approximation map $\tilde{A}(.,.)$, possibly depending on a parameter ε, such that, (i) the computation of a point $\tilde{x}_{i+1} \in \tilde{A}(\varepsilon,\tilde{x}_i)$ can be carried out without constructing an infinite sequence $\{y_i\}$, and (ii) when the parameter ε is appropriately manipulated, the sequence $\{\tilde{x}_i\}$ has the same convergence properties as the sequence $\{x_i\}$.

In practice, an implementation of a doubly iterative algorithm is obtained by truncating the construction of the sequence $\{y_i\}$ after a finite number of elements have been obtained. A theoretical basis for this practice in certain cases can be found in Sec. 1.3 of [1]. From the results in [1], as well as from empirical knowledge, it is clear that if the

*Received by the editors

†Department of Electrical Engineering and Computer Sciences and Electronics Research Laboratory, University of California, Berkeley, California 94720

Research sponsored by National Aeronautics and Space Administration, Grant NGL-05-003-016(Sup. 7).

construction of the sequence $\{y_i\}$ is terminated too early, convergence or rate of convergence, for the sequence $\{\tilde{x}_i\}$ may be lost, while if the construction of the y_i is allowed to continue for too long, the computation becomes unduly expensive. Thus, the problem of constructing an efficient implementation is far from trivial.

In this paper we shall present two efficient implementations of the Polak-Ribière conjugate gradient algorithm [2] which was introduced in 1969. This <u>theoretical</u> algorithm solves the following problem

(1.1) $$\min \{f(z) \mid z \in \mathbb{R}^n\},$$

where $f: \mathbb{R}^n \to \mathbb{R}^1$ is strictly convex and twice continuously differentiable.

We now state this algorithm for future reference.

(1.2) <u>The Polak-Ribière Conjugate Gradient Algorithm[2]</u>.

<u>Step 0</u>: Select a $z_0 \in \mathbb{R}^n$.

<u>Step 1</u>: If $\nabla f(z_0) = 0$, stop; else set $i = 0$, set $g_0 = h_0 = -\nabla f(z_0)$ and go to Step 2.

<u>Step 2</u>: Compute $\lambda_i \geq 0$ such that

(1.3) $$f(z_i + \lambda_i h_i) = \min_{\lambda \geq 0} f(z_i + \lambda h_i).$$

<u>Step 3</u>: Set

(1.4) $$z_{i+1} = z_i + \lambda_i h_i$$

Step 4: Compute $\nabla f(z_{i+1})$

Step 3: If $\nabla f(z_{i+1}) = 0$, stop; else, set

(1.5) $$g_{i+1} = - \nabla f(z_{i+1})$$

(1.6) $$\gamma_i = \langle g_{i+1} - g_i, g_{i+1} \rangle / \|g_i\|^2 ,$$

(1.7) $$h_{i+1} = g_{i+1} + \gamma_i h_i ,$$

set i = i+1 and go to Step 2. □

The following convergence result was established in [2].

(1.8) Theorem (Polak-Ribière [2]): If there exist $0 < m \leq M < \infty$ such that

(1.9) $$m\|y\|^2 \leq \langle y, H(z)y \rangle \leq M\|y\|^2 \text{ for all } y,z \in \mathbb{R}^n \dagger,$$

where $H(z) \equiv \frac{\partial^2 f(z)}{\partial z^2}$, then there exists a $\rho \in (0,1)$ such that the sequences $\{g_i\}$ and $\{h_i\}$ constructed by (1.2), in the process of solving (1.1), satisfy

(1.10) $$\langle g_i, h_i \rangle \geq \rho \|g_i\| \|h_i\|, \; i = 0, 1, 2, \ldots$$

and the sequence $\{z_i\}$ converges to $\hat{z}$, the unique minimizer of $f(\cdot)$. □

The operation in algorithm (1.2) which requires us to use an infinite subprocedure is the minimization on a half line in (1.3). From this point of view, relation (1.10) describes a very important property of the Polak-Ribière algorithm (not shared, for example by the Fletcher-Reeves method [3]) for it makes the algorithm rather insensitive to an

† Note that under this assumption the function $f(\cdot)$ has a minimum which is achieved at a unique minimizer $\hat{z}$.

accumulation of errors in the approximate calculation of the minimum in (1.3), provided that the approximation is carried out intelligently, as we shall see in the next section.

In 1970, A. Cohen [4] established a bound on the rate of convergence for the theoretical algorithm (1.2), modified to reinitialize as shown below:

(1.11) <u>Definition</u>: Let v be an integer satisfying $v \geq n$ and let $I_v = \{0, v, 2v, \ldots\}$. For $i = 0, 1, 2, \ldots$ let $\omega(i) = 0$ if $i \in I_v$ and let $\omega(i) = 1$ otherwise. Suppose that algorithm (1.2) is modified by replacing (1.6) with

(1.12) $$\gamma_i = \omega(i+1) < g_{i+1} - g_i, g_{i+1} > / \|g_i\|^2.$$

We shall call the resulting algorithm the <u>Polak-Ribière algorithm with reinitialization</u>. $\square$

(1.13) <u>Theorem (Cohen[4])</u>: Suppose that $\{z_i\}_{i=0}^{\infty}$ is a sequence constructed by the Polak-Ribière algorithm with reinitialization in solving problem (1.1). If $f(\cdot)$ is three times continuously differentiable, (1.9) holds, and $z_i \longrightarrow \hat{z}$ as $i \longrightarrow \infty$, then there exist an integer $k \geq 0$ and a constant $q \in (0,\infty)$ such that

(1.14) $$\|z_{i+n} - \hat{z}\| \leq q\|z_i - \hat{z}\|^2 \text{ for all } i \geq k,\ i \in I_v. \quad \square$$

In the next two sections we shall construct implementations for algorithm (1.2), the first of which preserve the relation (1.10), while the second one preserves relations (1.10) and (1.14). Our main theoretical results are given in theorems (2.11), (2.53), (3.13) and (3.96). Convergence is established in the first two theorems, while superlinear rate of convergence is obtained in the last two.

2. A CONVERGENT IMPLEMENTATION OF THE POLAK-RIBIÈRE ALGORITHM.

In this section we shall construct a convergent implementation of the Polak-Ribière algorithm. In the next section we shall modify this implementation so as to ensure that the relation (1.14) is satisfied.

(2.1) Implementation of Polak-Ribière Algorithm I

Step 0: Select a $z_0 \in \mathbb{R}^n$ and parameters $\delta_0 \in (0,1)$, $\rho_0 \in (0,1)$, $\beta \in (0,1)$, $\beta' \in (0,1)$, $\beta'' \in (0,1)$.[†]

Step 1: Set $g_0 = h_0 = -\nabla f(z_0)$; set $i = 0$.

Step 2: If $g_0 = 0$, stop; else, go to Step 3.

Step 3: Set $z = z_i$, $h = \frac{1}{\|h_i\|} h_i$.

Step 4: Define $\theta: \mathbb{R}^1 \to \mathbb{R}^1$ by

$$\theta(x) = f(z + xh) - f(z). \tag{2.2}$$

Comment: To compute the step size, we shall apply to $\theta(\cdot)$ several iterations of a gradient method due to Armijo [5]. The exact number of iterations required will be determined by the test in Step 13.

Step 5: Set $x_0 = 0$, $\ell = 0$.

Step 6: Compute

$$\theta'(x_\ell) = \langle \nabla f(z + x_\ell h), h \rangle . \tag{2.3}$$

Step 7: If $\theta'(x_\ell) = 0$, set $x = x_\ell$ and go to Step 15; else go to Step 8.

Step 8: Compute the smallest non-negative integer, $j(x_\ell)$, which satisfies

$$\theta(x_\ell - \beta^{j(x_\ell)} \theta'(x_\ell)) - \theta(x_\ell) + \frac{\beta^{j(x_\ell)}}{2} \theta'(x_\ell)^2 \leq 0 . \tag{2.4}$$

[†]The authors have found $\delta_0 = \cos 85°$, $\rho_0 = \cos 5°$, $\beta = 0.6$, $\beta' = \beta'' = 0.8$ to be a good choice in a number of problems.

Step 9: Set $x = x_\ell - \beta^{j(x_\ell)} \theta'(x_\ell)$.

Step 10: Compute $\nabla f(z + xh)$.

Step 11: If $\nabla f(z + xh) = 0$, set $z_{i+1} = z + xh$ and stop; else, go to Step 12.

Step 12: Compute $\theta'(x)$ according to (2.3) and set $\cos = \theta'(x)/\|\nabla f(z + xh)\|$.

Step 13: If $|\cos| \leq \delta_i$, go to Step 15; else, set $x_{\ell+1} = x$ and go to Step 14.

Step 14: Set $\ell = \ell + 1$ and go to Step 8.

Step 15: Set

$$z_{i+1} = z + xh \tag{2.5}$$

$$g_{i+1} = - \nabla f(z + xh) \tag{2.6}$$

$$\gamma_i = \langle g_{i+1} - g_i, g_{i+1} \rangle / \|g_i\|^2 \tag{2.7}$$

$$h_{i+1} = g_{i+1} + \gamma_i h_i \; . \tag{2.8}$$

Step 16: If $\langle g_{i+1}, h_{i+1} \rangle \geq \rho_i \|g_{i+1}\| \|h_{i+1}\|$, set $\rho_{i+1} = \rho_i$, $\delta_{i+1} = \delta_i$, and go to Step 17; else, set $\rho_{i+1} = \beta''\rho_i$, $\delta_{i+1} = \beta'\delta_i$, and go to Step 17.

Step 17: Set $i = i + 1$ and go to Step 3. $\square$

(2.9) Lemma: Suppose that (1.9) holds and that $\nabla f(z + xh) \neq 0$ for all $x \in \mathbb{R}^1$. Then algorithm (2.1) cannot cycle indefinitely in the loop contained between Steps 8 and 14 (i.e. it can jam up at a point $z = z_i$ only if the minimizer $\hat{z}$ of $f(\cdot)$ is on the line $\{z' | z' = z + xh, x \in \mathbb{R}^1\}$).

Proof: Since (1.9) is satisfied, $\theta(\cdot)$ has a minimum on $\mathbb{R}^1$. Suppose that $\hat{x}$ is the minimizer of $\theta(\cdot)$. Then, since $\nabla f(z + \hat{x}h) \neq 0$, we have $\theta'(\hat{x})/\|\nabla f(z + \hat{x}h)\| = 0$. Consequently, by continuity, it follows that there exists an $\varepsilon_i > 0$ such that for a given $\delta_i > 0$,

$$(2.10) \qquad |\theta'(x)/\|\nabla f(z + xh)\|| \leq \delta_i,$$

for all $\|x - \hat{x}\| \leq \varepsilon_i$. But $x_\ell \to \hat{x}$ as $\ell \to \infty$ (see Sec. 2.1 of [1]) and hence there exists a finite integer k such that $\|x_\ell - \hat{x}\| \leq \varepsilon_i$ for all $\ell \geq k$. Therefore (2.10) is satisfied for $x = x_\ell$, for all $\ell \geq k$. $\square$

(2.11) Theorem: Suppose that (1.9) holds and consider the sequences $\{z_i\}$, $\{g_i\}$, $\{h_i\}$, $\{\rho_i\}$ constructed by algorithm (2.1) in the process of solving problem (1.1). If there exists a $\rho \in (0,1)$ such that $\rho_i \geq \rho$ for $i = 0, 1, 2, \ldots$, in the test in Step 16 of algorithm (2.1), i.e., $\langle g_i, h_i \rangle \geq \rho \|g_i\| \|h_i\|$ for $i = 0, 1, 2, \ldots$, then either the algorithm jams up at a point $z = z_k$ and (Steps 5 to 14) constructs a sequence $\{x_\ell\}_{\ell=0}^{\infty}$ such that $(z + x_\ell h) \to \hat{z}$, as $\ell \to \infty$, where $\hat{z}$ is the unique minimizer of $f(\cdot)$ over $\mathbb{R}^n$, or $\{z_i\}$ is infinite and $z_i \to \hat{z}$ as $i \to \infty$.

Proof: In view of Lemma (2.9), the first part of the theorem is trivial. Hence, let us assume that the sequence $\{z_i\}$ is infinite.

We shall now compute a bound on $j(0) = j(x_0)$ for (2.4). Making use of Taylor's formula for second order expansions, of (1.9), and of the $\rho > 0$ which we have assumed to exist, we obtain (c.f.(2.4) with $x_\ell = 0$, $\beta^j(x_\ell)$ replaced by λ, and θ, θ' replaced by the expressions (2.2) and (2.3)), (2.12) $$f(z - \lambda \langle \nabla f(z),h \rangle h) - f(z) + \frac{\lambda}{2} \langle \nabla f(z),h \rangle^2$$

$$= \langle \nabla f(z),h \rangle^2 \left[-\frac{\lambda}{2} + \lambda^2 \int_0^1 (1 - t) \langle H(z - t \langle \nabla f(z),h \rangle h) h,h \rangle dt \right.$$

(2.13) $$\leq \frac{\lambda}{2} \langle \nabla f(z),h \rangle^2 (-1 + \lambda M) .$$

Since for $\ell = 0$ ($x_0 = 0$), $j(0)$ is chosen so as to make the left hand side of (2.4) non-positive, we see from (2.12) that $j(0) \leq \hat{j}$, where $\hat{j}$ is the smallest integer such that $-1 + \beta^{\hat{j}} M \leq 0$.

Consequently, from (2.4), we obtain (since $\theta(x_0) = 0$ and $\theta(x_{\ell+1}) < \theta(x_\ell)$ for $\ell = 0, 1, 2, \ldots$) that for some $\ell \geq 0$

(2.14) $$f(z_{i+1}) - f(z_i) = \theta(x_\ell - \beta^{j(x_\ell)} \theta'(x_\ell)) \leq \theta(x_0 - \beta^{j(x_0)} \theta'(x_0)) \leq$$
$$\leq -\frac{\beta^{j(x_0)}}{2\|h_i\|^2} \langle \nabla f(z_i),h_i \rangle^2 \leq -\frac{\beta^{\hat{j}} \rho^2}{2} \|\nabla f(z_i)\|^2 < 0,$$
$$i = 0, 1, 2, \ldots$$

Now, because of (1.9), the level set $\{z | f(z) \leq f(z_0)\}$ is compact, and hence the sequence $\{z_i\}$ must have accumulation points. Suppose that $z_i \to z^*$ as $i \to \infty$ for $i \in K \subset \{0, 1, 2, \ldots\}$, and that $\nabla f(z^*) \neq 0$. Since $f(\cdot)$ is continuously differentiable by assumption, there exists an integer k such that $\|\nabla f(z_i)\|^2 \geq \frac{\|\nabla f(z^*)\|^2}{2}$ for all $i \geq k$, $i \in K$. Suppose that i and $i + j$ are two consecutive indices in K, with $i \geq k$, then, because of (2.14),

(2.15) $$f(z_{i+j}) - f(z_i) = [f(z_{i+j}) - f(z_{i+j-1})] + \cdots$$
$$+ [f(z_{i+1}) - f(z_i)] \leq -\frac{\beta^{\hat{j}} \rho^2}{4} \|\nabla f(z^*)\|^2 ,$$

which shows that the sequence $\{f(z_i)\}_{i \in K}$ is not Cauchy. But $\{f(z_i)\}_{i \in K}$

must converge to $f(z^*)$ because $f(\cdot)$ is continuous, and hence we have a contradiction. Consequently, $\nabla f(z^*) = 0$. Since there is only one point $\hat{z}$ in $\mathbb{R}^n$ such that $\nabla f(\hat{z}) = 0$, we conclude that $z_i \to \hat{z}$ as $i \to \infty$, which completes our proof. □

We shall now show that there exists indeed a $\rho \in (0,1)$ such that $\rho_i \geq \rho$ for $i = 0, 1, 2, \ldots$, in the test in Step 16 of algorithm, (2.1), i.e., that for some $\rho \in (0,1)$, the sequences $\{g_i\}$, $\{h_i\}$ constructed by (2.1) satisfy

(2.16) $$\langle g_i, h_i \rangle \geq \rho \, \|g_i\| \, \|h_i\|, \; i = 0, 1, 2, \ldots,$$

(2.17) Lemma: Consider the sequences $\{g_i\}$, $\{h_i\}$ and $\{\delta_i\}$ constructed by algorithm (2.1) in the process of solving the problem (1.1) (see (2.6), (2.7), (2.8), and the instructions in Steps 12 and 13 of (2.1)). Then

(2.18) $$\langle \bar{g}_{i+1}, \bar{h}_{i+1} \rangle \geq \frac{\|g_{i+1}\|}{\|h_{i+1}\|} (1-\delta_i^2) - \delta_i, \text{ for } i = 0, 1, 2, \ldots$$

where $\bar{g}_{i+1} = \frac{1}{\|g_{i+1}\|} g_{i+1}$ and $\bar{h}_{i+1} = \frac{1}{\|h_{i+1}\|} h_{i+1}$, $i = -1, 0, 1, 2, \ldots$.

Proof: From (2.8),

(2.19) $$\langle h_{i+1}, h_i \rangle = \langle g_{i+1}, h_i \rangle + \gamma_i \|h_i\|^2,$$

and hence

(2.20) $$|\gamma_i| = |\langle h_{i+1}, h_i \rangle - \langle g_{i+1}, h_i \rangle| / \|h_i\|^2 \leq$$
$$\leq (\|h_i\| \, \|h_{i+1}\| + |\langle g_{i+1}, h_i \rangle|) / \|h_i\|^2.$$

Now making use of the fact that by construction (see Steps 12 and 13 of (2.1)),

$$|\langle \bar{g}_{i+1}, \bar{h}_i \rangle| \le \delta_i, \; i = 0, 1, 2, \ldots \tag{2.21}$$

we obtain from (2.20) that

$$|\gamma_i| \le (\|h_{i+1}\| + \|g_{i+1}\|\delta_i)/\|h_i\| \; . \tag{2.22}$$

Also from (2.8), and making use of (2.21) and (2.22),

$$\begin{aligned} \langle \bar{g}_{i+1}, \bar{h}_{i+1} \rangle &= (\|g_{i+1}\|^2 + \gamma_i \langle g_{i+1}, h_i \rangle)/\|g_{i+1}\| \, \|h_{i+1}\| \\ &\ge (\|g_{i+1}\|^2 - |\gamma_i| \, |\langle g_{i+1}, h_i\rangle|)/\|g_{i+1}\| \, \|h_{i+1}\| \\ &\ge \frac{\|g_{i+1}\|}{\|h_{i+1}\|}(1 - \delta_i^2) - \delta_i \;, \quad i = 0, 1, 2, \ldots , \end{aligned} \tag{2.23}$$

and hence we are done. $\square$

(2.24) Definition: Consider the sequences $\{z_i\}$ and $\{h_i\}$ constructed by algorithm (2.1) in the process of solving problem (1.1). We define the sequences $\{\lambda_i\}$ and $\{H_i\}$ by the following relations (see (2.5))

$$z_{i+1} = z_i + \lambda_i h_i, \; i = 0, 1, 2, \ldots , \tag{2.25}$$

$$H_i = \int_0^1 \frac{\partial^2}{\partial z^2} f(z_i + t\lambda_i h_i) \, dt, \; i = 0, 1, 2, \ldots \tag{2.26}$$

(Note that when (1.9) is satisfied, $m\|y\|^2 \le \langle y, H_i y \rangle \le M\|y\|^2$ and $\|H_i\| \le M$) $\square$

(2.27) Lemma: Suppose that (1.9) holds, and consider the sequences

$\{g_i\}$, $\{h_i\}$, $\{\delta_i\}$ and $\{\lambda_i\}$ constructed by algorithm (2.1) in the process of solving problem (1.1). Then

$$(2.28) \qquad |\lambda_i| \leq \frac{1}{m\|h_i\|^2} [\|g_i\|^2 (1 + \delta_{i-1}^2) + \|g_i\| \|h_i\|\delta_{i-1} + \|g_{i+1}\| \|h_i\|\delta_i]$$

for $i = 1, 2, \ldots$

Proof: To begin with, note that because of (2.21) and (2.22),

$$(2.29) \qquad \begin{aligned} |\langle g_i,h_i\rangle| &= |\|g_i\|^2 + \gamma_{i-1}\langle g_i,h_{i-1}\rangle| \\ &\leq \|g_i\|^2 + |\gamma_{i-1}| \, |\langle g_i,h_{i-1}\rangle| \\ &\leq \|g_i\|^2 + \frac{1}{\|h_{i-1}\|} (\|h_i\| + \|g_i\|\delta_{i-1}) \, \delta_{i-1}\|g_i\| \|h_{i-1}\| \\ &= \|g_i\|^2 (1 + \delta_{i-1}^2) + \|g_i\| \|h_i\| \delta_{i-1} \end{aligned}$$

Now, making use of (2.6), (2.26) and of the Taylor formula for first order expansions, we obtain

$$(2.30) \qquad \begin{aligned} - g_{i+1} &= - g_i + \lambda_i \left(\int_0^1 \frac{\partial^2}{\partial z^2} f(z_i + t\lambda h_i) \, dt\right) h_i \\ &= - g_i + \lambda_i H_i h_i, \quad i = 0, 1, 2, \ldots \end{aligned}$$

Consequently, for $i = 0, 1, 2, \ldots,$

$$(2.31) \qquad - \langle g_{i+1},h_i\rangle = - \langle g_i,h_i\rangle + \lambda_i \langle h_i,H_i h_i\rangle$$

and hence, for $i = 0, 1, 2, \ldots,$

$$(2.32) \qquad \lambda_i = (\langle g_i,h_i\rangle - \langle g_{i+1},h_i\rangle)/\langle h_i,H_i h_i\rangle \, .$$

Finally, making use of (2.32), (2.29), (2.21), and (1.9) (which implies that $m\|h_i\|^2 \leq \langle h_i, H_i h_i \rangle \leq M\|h_i\|^2$ and hence that $m \leq \|H_i\| \leq M$), we obtain for $i = 1, 2, \ldots,$

$$(2.33) \qquad |\lambda_i| \leq \frac{1}{m\|h_i\|^2}(|\langle g_i, h_i \rangle| + |\langle g_{i+1}, h_i \rangle|)$$

$$\leq \frac{1}{m\|h_i\|^2}[\|g_i\|^2 (1 + \delta_{i-1}^2) + \|g_i\| \|h_i\| \delta_{i-1} + \|g_{i+1}\| \|h_i\| \delta_i]$$

which is the desired result. □

(2.34) <u>Lemma</u>: Suppose that (1.9) holds, and consider the sequences $\{g_i\}$, $\{h_i\}$, and $\{\delta_i\}$ constructed by algorithm (2.1) in the process of solving problem (1.1). Then

$$(2.35) \qquad \frac{\|h_{i+1}\|}{\|g_{i+1}\|} \leq 1 + \frac{M}{m}(1 + \delta_{i-1}^2) + \frac{M}{m}\left(\delta_{i-1} + \frac{\|g_{i+1}\|}{\|g_i\|}\delta_i\right)\frac{\|h_i\|}{\|g_i\|}, \qquad i = 1, 2, \ldots$$

<u>Proof</u>: From (2.7), (2.26), (2.28), and (2.30), we obtain

$$(2.36) \qquad |\gamma_i| = \left|\frac{\langle g_{i+1} - g_i, g_{i+1} \rangle}{\|g_i\|^2}\right| = \frac{|\lambda_i| \langle H_i h_i, g_{i+1} \rangle}{\|g_i\|^2} \leq$$

$$\leq \frac{M \|h_i\| \|g_{i+1}\|}{\|g_i\|^2} \cdot \frac{1}{m\|h_i\|^2} [\|g_i\|^2 (1 + \delta_{i-1}^2) +$$

$$+ \|g_i\| \|h_i\| \delta_{i-1} + \|g_{i+1}\| \|h_i\| \delta_i] =$$

$$= \frac{M}{m}\left[\frac{\|g_{i+1}\|}{\|h_i\|}(1 + \delta_{i-1}^2) + \frac{\|g_{i+1}\|}{\|g_i\|}\delta_{i-1} + \frac{\|g_{i+1}\|^2}{\|g_i\|^2}\delta_i\right]$$

Next, from (2.8) and from (2.36) we obtain

$$(2.37) \qquad \|h_{i+1}\| \le \|g_{i+1}\| + |\gamma_i| \, \|h_i\|$$

$$\le \|g_{i+1}\| + \frac{M}{m}\left[\|g_{i+1}\| \, (1 + \delta_{i-1}^2) + \frac{\|g_{i+1}\| \, \|h_i\|}{\|g_i\|} \, \delta_{i-1} + \right.$$

$$\left. + \frac{\|g_{i+1}\|^2 \, \|h_i\|}{\|g_i\|^2} \, \delta_i\right]$$

Since (2.35) follows from (2.37) by inspection, we are done. □

(2.38) Lemma: Suppose that (1.9) holds and consider the sequences $\{g_i\}$ and $\{h_i\}$ constructed by algorithm (2.1) in the process of solving problem (1.1). Then

$$(2.39) \qquad \frac{\|g_{i+1}\|}{\|g_i\|} \le 1 + 2\,\frac{M}{m} \text{ for } i = 0, 1, 2, \ldots$$

Proof: First, by the Taylor formula, together with (2.6), and (1.9), we obtain, since by construction $f(z_{i+1}) - f(z_i) < 0$, for $i = 0, 1, 2, \ldots$,

$$(2.40) \qquad 0 > f(z_{i+1}) - f(z_i) = f(z_i + \lambda_i h_i) - f(z_i) =$$

$$= -\lambda_i \langle g_i, h_i \rangle + \lambda_i^2 \int_0^1 (1 - t) \langle h_i, \frac{\partial^2}{\partial z^2} f(z_i + t\lambda_i h_i) h_i \rangle \, dt$$

$$\ge -|\lambda_i| \, \left|\langle g_i, h_i \rangle\right| + \lambda_i^2 \int_0^1 (1 - t) \, m\|h_i\|^2 \, dt$$

$$\ge -|\lambda_i| \, \|g_i\| \, \|h_i\| + \frac{\lambda_i^2}{2} m\|h_i\|^2, \; i = 0, 1, 2, \ldots$$

consequently, for $i = 0, 1, 2, \ldots$

$$(2.41) \qquad |\lambda_i| \leq \frac{2}{m} \frac{\|g_i\|}{\|h_i\|} .$$

Next, from (2.30) and making use of the fact that $\|H_i\| \leq M$,

$$(2.42) \qquad \|g_{i+1}\| \leq \|g_i\| + |\lambda_i| \, \|H_i\| \, \|h_i\| \leq (1 + \frac{2M}{m}) \|g_i\| ,$$

from which (2.39) follows directly. $\square$

(2.43) Lemma: Suppose that (1.9) is satisfied and consider the sequences $\{g_i\}$, $\{h_i\}$ and $\{\delta_i\}$ constructed by algorithm (2.1) in the process of solving problem (1.1). Let $\mu \in (0, 1)$. If there exists an integer N such that

$$(2.44) \qquad \delta_i \leq \delta \triangleq \frac{1}{2} (\frac{m}{m + M}) \frac{m}{M} \mu \text{ for } i = N-1, N, N+1, \ldots$$

then there exists an $L \in (0, \infty)$ such that

$$(2.45) \qquad \frac{\|h_i\|}{\|g_i\|} \leq L \text{ for } i = 0, 1, 2, \ldots$$

Proof: First, making use of (2.39) and (2.44) we obtain that

$$(2.46) \qquad \frac{M}{m}(\delta_{i-1} + \frac{\|g_{i+1}\|}{\|g_i\|} \delta_i) \leq \frac{M}{m} (\delta_{i-1} + \delta_i + \frac{2M}{m} \delta_i)$$

$$\leq \frac{2M}{m} (1 + \frac{M}{m}) \delta = \mu < 1,$$

$$\text{for } i = N, N+1, \ldots$$

Next, substituting from (2.46) into (2.35), we obtain, since $\delta_i \in (0,1)$ for $i = 0, 1, 2, \ldots$,

$$(2.47)\qquad \frac{\|h_{i+1}\|}{\|g_{i+1}\|} \leq [1 + \frac{M}{m}(1 + \delta^2)] + \mu \frac{\|h_i\|}{\|g_i\|}, \text{ for } i = N, N+1, \ldots$$

Let

$$(2.48)\qquad \nu = 1 + \frac{M}{m}(1 + \delta^2) \cdot$$

Then, from (2.47), for any $i \in \{N + 1, N + 2, \ldots\}$, and since $\mu \in (0, 1)$, we have

$$(2.49)\qquad \frac{\|h_i\|}{\|g_i\|} \leq \nu + \mu \frac{\|h_{i-1}\|}{\|g_{i-1}\|}$$

$$\leq \nu \sum_{j=0}^{i-N-1} \mu^j + \mu^{i-N} \frac{\|h_N\|}{\|g_N\|}$$

$$\leq \nu \sum_{j=0}^{\infty} \mu^j + \mu^{i-N} \frac{\|h_N\|}{\|g_N\|} \leq \frac{\nu\mu}{1-\mu} + \frac{\|h_N\|}{\|g_N\|},$$

Now since $\delta_i \in (0, 1]$ for $i = 0, 1, 2, \ldots$, and because of (2.35) and (2.39),

$$(2.50)\qquad \frac{\|h_{i+1}\|}{\|g_{i+1}\|} \leq (1 + \frac{2M}{m}) + \frac{2M}{m}(1 + \frac{M}{m}) \frac{\|h_i\|}{\|g_i\|} \text{ for } i = 0, 1, 2\ldots$$

Consequently, since $h_0 = g_0$, (and since $(1 + \frac{2M}{m}) > 1$),

$$(2.51)\qquad \frac{\|h_i\|}{\|g_i\|} \leq \left(1 + \frac{2M}{m}\right) \sum_{j=0}^{i-1} \left[\frac{2M}{m}\left(1 + \frac{M}{m}\right)\right]^j + \left[\frac{2M}{m}\left(1 + \frac{M}{m}\right)\right]^i$$

$$\leq \left(1 + \frac{2M}{m}\right) \sum_{j=0}^{i} \left[\frac{2M}{m}\left(1 + \frac{M}{m}\right)\right]^j \leq \left(1 + \frac{2M}{m}\right) \sum_{j=0}^{N} \left[\frac{2M}{m}\left(1 + \frac{M}{m}\right)\right]^j$$

for $i = 0, 1, 2, \ldots, N$.

Consequently, combining (2.51) and (2.49), we obtain, for $i = 0, 1, 2, \ldots,$

$$(2.52)\qquad \frac{\|h_i\|}{\|g_i\|} \leq \frac{\nu\mu}{1-\mu} + \left(1 + \frac{2M}{m}\right)\sum_{j=0}^{N}\left[\frac{2M}{m}\left(1 + \frac{M}{m}\right)\right]^j \triangleq L < \infty . \qquad \square$$

(2.53) <u>Theorem</u>: Suppose that (1.9) holds and consider the sequence $\{\rho_i\}$ constructed by algorithm (2.1) in the process of solving problem (1.1). Then there exists a $\rho \in (0, 1]$ such that $\rho_i \geq \rho$ for $i = 0, 1, 2, \ldots$.

<u>Proof</u>: If algorithm (2.1) jams up after a finite number of iterations, then the ρ_i are obviously bounded. Hence we only need to consider the case when the sequence $\{\rho_i\}$ is infinite.

Let $\bar{g}_i = \frac{1}{\|g_i\|} g_i$, $\bar{h}_i = \frac{1}{\|h_i\|} h_i$, $i = 0, 1, 2, \ldots,$ and suppose that $\delta_i \to 0$ as $i \to \infty$. We shall show that this leads to a contradiction.

Since $\delta_i \to 0$, as $i \to \infty$, the conditions of Lemma (2.43) are satisfied and hence there exists an $L \in (0, \infty)$ such that $\|h_i\|/\|g_i\| \leq L$ for $i = 0, 1, 2, \ldots$. Therefore, from (2.18), we obtain

$$(2.54)\qquad \langle \bar{g}_i, \bar{h}_i \rangle \geq \frac{\|g_i\|}{\|h_i\|}(1 - \delta_{i-1}^2) - \delta_{i-1}$$

$$\geq \frac{1}{L}(1 - \delta_{i-1}^2) - \delta_{i-1},\ i = 1, 2, \ldots .$$

Consequently, since $\delta_i \to 0$ as $i \to \infty$, there exists an integer $N' \geq 0$ such that

$$(2.55)\qquad \langle \bar{g}_i, \bar{h}_i \rangle \geq \frac{1}{2L} = \hat{\rho} > 0 \text{ for } i = N', N'+1, \ldots$$

Now (see Step 16 of (2.1)), (2.55) implies that for $i \geq N'$, $\rho_i \geq \beta''^j \rho_0 > 0$, where j is the smallest positive integer such that $\beta''^j \rho_0 \leq \hat{\rho}$. But

$\delta_i \to 0$ as $i \to \infty$ if and only if $\rho_i \to 0$ as $i \to \infty$, according to the instruction in Step 16 of (2.1). Hence we have a contradiction, and therefore $\{\delta_i\}$ does not converge to zero. Consequently, $\{\rho_i\}$ does not converge to zero, therefore the existence of a $\rho > 0$ such that $\rho_i \geq \rho$ for $i = 0, 1, 2 \ldots$, has been established. $\square$

Consequently, the assumptions of Theorem (2.11) are satisfied by algorithm (2.1).

3. A SUPERLINEARLY CONVERGING IMPLEMENTATION OF THE POLAK-RIBIÈRE ALGORITHM

The convergent implementation (2.1) has the very nice feature that it maintains within fixed limits the precision with which the minimization of $\theta(x)$ (see (2.2)) is carried out. This fixed precision is defined by the tests in Steps 13 and 16, and results from the fact, established in Theorem (2.53), that $\rho_i \geq \rho > 0$ and $\delta_i \geq \delta > 0$ for $i = 0, 1, 2\ldots$. However, if we wish to ensure that the result (1.14) be valid for the sequence $\{z_i\}$ constructed, then we must reinitialize as in (1.11) and make the minimization of $\theta(x)$ (see (2.2)) progressively more precise, as shown in Steps 13, 15, 16 of the algorithm below (cf. (2.1))

(3.1) Implementation of Polak-Ribière Algorithm II.

Step 0: Select a $z_0 \in \mathbb{R}^n$ and parameters $\tilde{\delta}_0 \in (0,1]$, $\rho_0 \in (0,1]$, $\beta \in (0,1)$, $\beta' \in (0,1)$, $\beta'' \in (0,1)$.

Step 1: Set $g_0 = h_0 = -\nabla f(z_0)$; set $i = 0$.

Step 2: If $g_0 = 0$, stop; else go to Step 3.

Step 3: Set $z = z_i$, $h = \frac{1}{\|h_i\|} h_i$

Step 4: Define θ: $\mathbb{R}^1 \to \mathbb{R}^1$ by

(3.2) $$\theta(x) = f(z + xh) - f(z)$$

Step 5: Set $x_0 = 0$, set $\ell = 0$.

Step 6: Compute

(3.3) $$\theta'(x_\ell) = \langle \nabla f(z + x_\ell h), h \rangle$$

Step 7: If $\theta'(x_\ell) = 0$, go to Step 15; else go to Step 8.

Step 8: Compute the smallest non-negative integer $j(x_\ell)$ which satisfies

(3.4) $$\theta(x_\ell - \beta^{j(x_\ell)}\theta'(x_\ell)) - \theta(x_\ell) + \frac{\beta^{j(x_\ell)}}{2}\theta'(x_\ell)^2 \leq 0$$

Step 9: Set $x = x_\ell - \beta^{j(x_\ell)}\theta'(x_\ell)$.

Step 10: Compute $\nabla f(z + xh)$.

Step 11: If $\nabla f(z + xh) = 0$, set $z_{i+1} = z + xh$ and stop; else go to Step 12.

Step 12: Compute $\theta'(x)$ according to (3.3) and set cos = $\theta'(x)/\|\nabla f(z + xh)\|$.

Step 13: If $|\cos| \leq \delta_i \triangleq \min\{\tilde{\delta}_i, \|g_i\|\}$, go to Step 15; else, set $x_{\ell+1} = x$ and go to Step 14.

Step 14: Set $\ell = \ell + 1$ and go to Step 8.

Step 15: Set

(3.5) $$z_{i+1} = z + xh$$

(3.6) $$g_{i+1} = -\nabla f(z + xh)$$

(3.7) $$\gamma_i = \omega(i+1)\langle g_{i+1} - g_i, g_{i+1}\rangle / \|g_i\|^2$$

(3.8) $$h_{i+1} = g_{i+1} + \gamma_i h_i ,$$

where $\omega(i+1)$ is as in (1.11).

Step 16: If $\langle g_{i+1}, h_{i+1} \rangle \geq \rho_i \| g_{i+1} \| \| h_{i+1} \|$, set $\rho_{i+1} = \rho_i$, $\tilde{\delta}_{i+1} = \tilde{\delta}_i$ and go to Step 17; else, set $\rho_{i+1} = \beta'' \rho_i$, $\tilde{\delta}_{i+1} = \beta' \tilde{\delta}_i$ and go to Step 17.

Step 17: Set $i = i + 1$ and go to Step 3. $\square$

We begin by noting that if $\{z_i\}$ is an infinite sequence constructed by algorithm (3.1) in the process of solving problem (1.1), then $z_i \longrightarrow \hat{z}$ as $i \longrightarrow \infty$, where $\hat{z}$ is the unique minimizer of $f(\cdot)$. To see this, note that for $i \in I_v$, (see(1.11)), $h_i = - \nabla f(z_i)$, and from (2.14) with $\rho = 1$, $f(z_{i+1}) - f(z_i) \leq - \frac{\beta^{\hat{j}}}{2} \| \nabla f(z_i) \|^2$ for $i \in I_v$. Since $f(\cdot)$ is bounded from below and $\{f(z_i)\}$ is a monotically decreasing sequence, we conclude that $\nabla f(z_i) \longrightarrow 0$ and that $f(z_i) \longrightarrow f(\hat{z})$, as $i \longrightarrow \infty$. Since the level sets of $f(\cdot)$ are compact and since $\hat{z}$ is the unique minimizer of $f(\cdot)$, we conclude that $\hat{z}$ must be the unique accumulation point of $\{z_i\}$, i.e. that $z_i \longrightarrow \hat{z}$ as $i \longrightarrow \infty$. When the sequence is not infinite, lemma (2.9) applies.

To establish that (1.14) holds, we proceed essentially as in [4], following a pattern of proof first used by J. Daniel [6] in conjunction with yet another theoretical, conjugate gradient algorithm. Basically, the approach consists in establishing the rate of convergence of our algorithm by means of a suitable comparison with the Newton-Raphson method which uses the recursion formula

(3.9) $$z_{i+1} = z_i - H(z_i)^{-1} \nabla f(z_i), \ i = 0, 1, \ldots$$

in minimizing the twice continuously differentiable function $f(\cdot)$. In (3.9), $H(z_i) = \dfrac{\partial^2 f(z_i)}{\partial z^2}$, as before. For the purpose of this comparison, we introduce the following sequence of approximating functions.

(3.10) Definition: Consider the sequence $\{z_i\}$ generated by algorithm (3.1) in the process of solving problem (1.1). Then, for $i \in I_v$(see (1.11)) we define the functions f_i: $\mathbb{R}^n \to \mathbb{R}^1$ by

$$(3.11)\quad f_i(z) = f(z_i) + \langle \nabla f(z_i), z - z_i \rangle + \frac{1}{2}\langle z - z_i, H(z_i)(z - z_i)\rangle. \;\square$$

Since the functions $f_i(\cdot)$ are quadratic, the Polak-Ribière algorithm (1.2) finds their minimizers in at most n iterations.

To ensure that we do not confuse the various sequences constructed in minimizing the $f_i(\cdot)$ with the sequences constructed in minimizing $f(\cdot)$, we shall designate sequences associated with $f_i(\cdot)$, by a subscript i and an overscript j, e.g., $\overset{j}{z}_i$, $\overset{j}{\lambda}_i$, etc. The overscript will be the running index.

(3.12) Definition: Consider the sequence $\{z_i\}$ generated by algorithm (3.1) in the process of solving problem (1.1). For $i \in I_v$, we shall denote by $\overset{j}{z}_i$, $\overset{j}{g}_i$, $\overset{j}{h}_i$, $\overset{j}{\lambda}_i$, $\overset{j}{\gamma}_i$, $j = 0, 1, 2, \ldots, n$, the quantities constructed by the Polak-Ribiere algorithm (1.2) in the process of minimizing the function $f_i(\cdot)$, with (1.2) being initialized with: $\overset{0}{z}_i = z_i$.
Note that for $i \in I_v$, $\overset{0}{h}_i = \overset{0}{g}_i = h_i = g_i = -\nabla f(z_i)$. $\square$

(3.13) Theorem: Suppose that the function $f(\cdot)$ in (1.1) is three times continuously differentiable and that (1.9) holds. Consider the sequences $\{z_i\}$, $\{h_i\}$ and $\{\lambda_i\}$ constructed by algorithm (3.1) in the process of solving problem (1.1) and let $\hat{z}$ be the limit point of $\{z_i\}$. If there exists a $q \in (0,\infty)$ and an integer N' such that

(3.14) $$\|\lambda^j_{i+j} h_{i+j} - \lambda^j_i h^j_i\| \leq q\|z_i - \hat{z}\|^2, \text{ for all } i \geq N', \; i \in I_v$$
$$j = 0, 1, 2, \ldots, \nu(i) - 1,$$

where $\nu(i) \leq n$ is such that $h_i^{\nu(i)} = 0$, $h_i^{\nu(i)-1} \neq 0$,

then there exist a $\hat{q} \in (0, \infty)$ and an integer N'' such that

(3.15) $$\|z_{i+n} - \hat{z}\| \leq \hat{q}\,\|z_i - \hat{z}\|^2 \text{ for all } i \geq N'', \; i \in I_v.$$

<u>Proof</u>: Since the functions $f_i(\cdot)$, $i \in I_v$, are quadratic, the point $z_i^{\nu(i)}$, $i \in I_v$, (see (3.12)) minimizes the function $f_i(\cdot)$ over $\mathbb{R}^n$. Since $H(z_i)$ in (3.11) is nonsingular because of (1.9), the Newton-Raphson method (3.9) can be applied to the minimization of $f_i(\cdot)$, and, since $f_i(\cdot)$ is quadratic, it computes $z_i^{\nu(i)}$ (the unique minimizer of $f_i(\cdot)$) in one iteration. Thus,

(3.16) $$\begin{aligned} z_i^{\nu(i)} &= z_i^0 - H(z_i)^{-1} \nabla f_i(z_i^0) \\ &= z_i - H(z_i)^{-1} \nabla f(z_i), \; i \in I_v \end{aligned}$$

(compare the second part of (3.16) with (3.9) !)

Let $a(z) \triangleq z - H(z)^{-1} \nabla f(z)$ denote the Newton-Raphson iteration function. Then, since by assumption $f(\cdot)$ is three times continuously differentiable, there exist an $\varepsilon > 0$ and a $q' \in (0, \infty)$ such that

(3.17) $$\|a(z) - \hat{z}\| \leq q'\,\|z - \hat{z}\|^2$$

for all z such that $\|z - \hat{z}\| < \varepsilon$, where $\hat{z}$ is the minimizer of $f(\cdot)$ (see Theorem (6.2.17) in [1]). Consequently, since $z_i \to \hat{z}$ and because of the second part of (3.16), there exists an integer $\tilde{N}'' > N'$ such that

(3.18) $\qquad \| \overset{\nu(i)}{z_i} - \hat{z}\| \le q' \|z_i - \hat{z}\|^2$ for all $i \ge N''$, $i \in I_v$

Now, since $\overset{0}{z_i} = z_i$, $i \in I_v$, because of (3.14), for $i \in I_v$

$$(3.19)\quad \|z_{i+\nu(i)} - \overset{\nu(i)}{z_i}\| = \|[(z_{i+\nu(i)} - z_{i+\nu(i)-1}) + (z_{i+\nu(i)-1} - z_{i+\nu(i)-2}) + \dots$$

$$+ (z_{i+1} - z_i)](\overset{\nu(i)}{z_i} - \overset{\nu(i)-1}{z_i}) + \dots + (\overset{1}{z_i} - \overset{0}{z_i})]\| \le$$

$$\le \sum_{j=0}^{\nu(i)-1} \|(z_{i+j+1} - z_{i+j}) - (\overset{j+1}{z_i} - \overset{j}{z_i})\|$$

$$= \sum_{j=0}^{\nu(i)-1} \|\lambda_{i+j} h_{i+j} - \overset{j}{\lambda_i} \overset{j}{h_i}\|$$

$$\le n\, q\|z_i - \hat{z}\|^2 \text{ for all } i \ge N',\ i \in I_v.$$

Consequently, because of (3.18) and (3.19), and because $z_i \to \hat{z}$ at least linearly (see Sec. 6.1 in [1]), there exists an integer $N'' \ge \tilde{N}''$ such that

$$(3.20)\quad \|z_{i+n} - \hat{z}\| \le \|z_{i+\nu(i)} - \hat{z}\| \le \|z_{i+\nu(i)} - \overset{\nu(i)}{z_i}\| + \|\overset{\nu(i)}{z_i} - \hat{z}\| \le$$

$$\le (n\, q + q')\, \|z_i - \hat{z}\|^2$$

$$\triangleq \hat{q}\, \|z_i - \hat{z}\|^2, \text{ for all } i \ge N'',\ i \in I_v$$

which completes our proof. $\Box$

The verification that (3.14) is satisfied by Algorithm (3.1) is quite laborious and requires a number of preliminary results which we shall now establish.

To simplify the statements of the lemmas and theorems to follow, we shall assume from now on without loss of generality that $\nu(i) = n$ for $i = 0,1,2,\dots$, that (1.9) is satisfied, that $f(\cdot)$ is three times continuously differentiable (whether required by a specific lemma or not), and that we are dealing with a set of infinite sequences $\{z_i\}$, $\{g_i\}$, $\{h_i\}$, $\{\lambda_i\}$, $\{\gamma_i\}$, $\{\delta_i\}$, $\{\tilde{\delta}_i\}$ and $\{\rho_i\}$ constructed by algorithm (3.1) in the process of solving problem (1.1) from a given initial point z_0. Corresponding sequences $\{z_i^j\}$, $\{g_i^j\}$, $\{h_i^j\}$, $\{\lambda_i^j\}$ and $\{\gamma_i^j\}$ are as defined in (3.12), for $i \in I_v$ and $j = 0, 1, 2, \ldots, n$.

(3.21) <u>Lemma</u>: There exists a $q_1 \in (0, \infty)$ such that

(3.22) $$|\gamma_i| \leq q_1 \text{ for } i = 0, 1, 2, \ldots$$

<u>Proof</u>: By (1.9), (3.7), (2.30), (2.41) and (2.39),

(3.23) $$\begin{aligned} |\gamma_i| &= \omega(i+1) \,|\langle g_{i+1} - g_i, g_{i+1} \rangle|\, / \|g_i\|^2 \\ &\leq |\lambda_i| \,|\langle H_i h_i, g_{i+1} \rangle|\, / \|g_i\|^2 \\ &\leq \frac{2}{m} \frac{\|g_i\|}{\|h_i\|} \frac{M \|h_i\| \, \|g_{i+1}\|}{\|g_i\|^2} = \frac{2M}{m} \frac{\|g_{i+1}\|}{\|g_i\|} \leq \\ &\leq \frac{2M}{m} \left(1 + \frac{2M}{m}\right) \triangleq q_1, \text{ for } i = 0, 1, 2, \cdots \qquad \square \end{aligned}$$

(3.24) <u>Lemma</u>: There exist an integer $N \geq 0$ and q_2, q_3 in $(0, \infty)$ such that

(3.25) $$q_3 \|g_i\| \leq \|h_i\| \leq q_2 \|g_i\| \text{ for all } i \geq N .$$

<u>Proof</u>: Making use of (3.23) and (3.8), we conclude that

$$\|h_{i+1}\| \leq \|g_{i+1}\| + |\gamma_i| \, \|h_i\| \leq \|g_{i+1}\| \left(1 + \frac{2M}{m} \frac{\|h_i\|}{\|g_i\|}\right) \; i = 0, 1, \ldots$$

Consequently,

$$\|h_{i+1}\| / \|g_{i+1}\| \leq 1 + \frac{2M}{m}\left(\|h_i\| / \|g_i\|\right) \quad i = 0, 1, 2, \ldots .$$

But $h_i = g_i$ for $i \in I_v$, and hence, for $i \in I_v$ and $j = 0, 1, 2, \ldots, v-1$,

$$\|h_{i+j}\| / \|g_{i+j}\| \leq \sum_{k=0}^{j} \left(\frac{2M}{m}\right)^k \leq \sum_{k=0}^{v-1} \left(\frac{2M}{m}\right)^k \triangleq q_2$$

i.e. the right hand side of (3.25) holds.

We now establish the second half of (3.25).

From (3.8), (3.23) and because by Step 13 of (3.1) $|\langle \bar{g}_{i+1},\bar{h}_i\rangle| \leq \|g_i\|$ for $i = 0, 1, 2, \ldots$, $(\bar{g}_i = \frac{1}{\|g_i\|} g_i, \bar{h}_i = \frac{1}{\|h_i\|} h_i)$, we obtain

$$\begin{aligned}
(3.26) \qquad \|h_i\|^2 &= \|g_i + \gamma_{i-1}h_{i-1}\|^2 \\
&= \|g_i\|^2 + \gamma_{i-1}^2\|h_{i-1}\|^2 + 2\gamma_{i-1}\langle g_i,h_{i-1}\rangle \\
&\geq \|g_i\|^2 - 2|\gamma_{i-1}||\langle g_i,h_{i-1}\rangle| \\
&\geq \|g_i\|^2 - 2|\gamma_{i-1}|\|g_{i-1}\| \|g_i\| \|h_{i-1}\| \\
&\geq \|g_i\|^2 \left(1 - \frac{4M}{m}\|h_{i-1}\|\right) \quad i = 1, 2, 3, \ldots
\end{aligned}$$

Now, since $g_i \to 0$ as $i \to \infty$, it follows from the first part of our proof that $\|h_i\| \to 0$ as $i \to \infty$. Hence there exists an integer N such that

$$(3.27) \qquad 1 - \frac{4M}{m}\|h_{i-1}\| \geq \frac{1}{2} \text{ for all } i \geq N$$

and therefore, from (3.26),

$$\|h_i\|^2 \geq \frac{1}{2}\|g_i\|^2 \text{ for all } i \geq N . \quad \square$$

(3.28) Corollary: There exist an integer N and q_4, q_5 in $(0, \infty)$ such that

$$(3.29) \qquad |\lambda_i| \leq q_4 \quad \text{for all } i \geq N ,$$

and

(3.30) $$\|g_{i+1}\| \leq q_5 \|h_i\| \text{ for all } i \geq N.$$

Proof: The inequality (3.29) follows from (2.41) and (3.25). Next, with N as in (3.25), it follows from (2.30), (1.9), (3.25) and (3.29) that

(3.31) $$\begin{aligned} \|g_{i+1}\| &\leq \|g_{i+1} - g_i\| + \|g_i\| \\ &= |\lambda_i| \|H_i h_i\| + \|g_i\| \\ &\leq (q_4 M + \frac{1}{q_3}) \|h_i\| \triangleq q_5 \|h_i\|, \; i = N, N+1, \ldots \quad \square \end{aligned}$$

The following two lemmas can be obtained by making use of (1.9), (3.22), (3.25), (3.29), (3.30) and of the fact that $\langle \overset{j+1}{g_i}, \overset{j}{g_i} \rangle = 0$ and that $\langle \overset{j+1}{h_i}, H(z_i)\overset{j}{h_i} \rangle = 0$ for $j = 0, 1, 2, .., n-1$ and $i \in I_v$ (see (6.3.20) - (6.3.31) in [1]).

(3.32) Lemma: For $i \in I_v$ and $j = 0, 1, \ldots, n-1$,

(3.33) $$\overset{j}{\lambda_i} = \langle \overset{j}{g_i}, \overset{j}{h_i} \rangle / \langle \overset{j}{h_i}, H(z_i)\overset{j}{h_i} \rangle = \|\overset{j}{g_i}\|^2 / \langle \overset{j}{h_i}, H(z_i) \overset{j}{h_i} \rangle$$

(3.34) $$\begin{aligned} \overset{j}{\gamma_i} &= \langle \overset{j+1}{g_i} - \overset{j}{g_i}, \overset{j+1}{g_i} \rangle / \|\overset{j}{g_i}\|^2 \\ &= - \langle \overset{j+1}{g_i}, H(z_i)\overset{j}{h_i} \rangle / \langle \overset{j}{h_i}, H(z_i) \overset{j}{h_i} \rangle \quad \square \end{aligned}$$

(3.35) <u>Lemma</u>: For $i \in I_v$ and $j = 0, 1, 2, \ldots, n-1$, there exist $\bar{q}_1, \bar{q}_2, \bar{q}_3, \bar{q}_4, \bar{q}_5$ in $(0,\infty)$ such that

$$\bar{q}_3 \| g_i^j \| \le \| h_i^j \| \le \bar{q}_2 \| g_i^j \|, \tag{3.36}$$

$$\| g_i^{j+1} \| \le \bar{q}_5 \| h_i^j \| \tag{3.37}$$

$$\lambda_i^j \le \bar{q}_4 \tag{3.38}$$

$$|\gamma_i^j| \le \bar{q}_1 \tag{3.39}$$

(3.40) <u>Lemma</u>: There exist an integer N and a $q_6 \in (0, \infty)$ such that

$$\| g_{i+j} \| \le q_6 \| h_i \|, \quad i \ge N, \quad j = 0, 1, \ldots, n \tag{3.41}$$

$$\| h_{i+j} \| \le q_6 \| h_i \|, \quad i \ge N, \quad j = 0, 1, \ldots, n \tag{3.42}$$

$$\| g_i^j \| \le q_6 \| h_i \| \qquad i \ge N,\ i \in I_v, \quad j = 0, 1, \ldots, n \tag{3.43}$$

$$\| h_i^j \| \le q_6 \| h_i \| \qquad i \ge N,\ i \in I_v, \quad j = 0, 1, \ldots, n \tag{3.44}$$

<u>Proof</u>: Making use of (3.25) and of (3.30) we obtain, for $i \ge N$ and $j = 1, 2, \ldots, n$,

$$\| g_{i+j} \| \le q_5 \| h_{i+j-1} \| \le q_5 q_2 \| g_{i+j-1} \| \le q_5 (q_2 q_5)^{j-1} \| h_i \| . \tag{3.45}$$

Combining (3.45) with (3.25), we obtain an inequality of the form of (3.41).

Since $h_i^0 = h_i = g_i$ for $i \in I_v$ an inequality of the form of (3.43) follows similarly from (3.36) and (3.37); an inequality of the form of (3.42) now follows from (3.25) and (3.41), and an inequality of the form of (3.44) from (3.36) and (3.43). Setting q_6 to be the largest of the constants in these inequalities, we see that the lemma holds. $\square$

Notation: We shall denote by $0_\ell(\cdot)$, $\ell = 1, 2, 3, \ldots$, functions from $\mathbb{R}^1$ into $\mathbb{R}^1$ with the property that for $\ell = 1, 2, 3, \ldots$, there exists an $\varepsilon_\ell > 0$ and an $r_\ell > 0$ such that

$$(3.46) \qquad |0_\ell(x)/x| \le r_\ell \text{ for all } |x| < \varepsilon_\ell, \; (\varepsilon_\ell > 0, \; r_\ell > 0).$$

(3.47) Lemma: There exists an integer N and a function $0_1 : \mathbb{R}^1 \to \mathbb{R}^1$ (satisfying (3.46) for $\ell = 1$) such that for all $i \ge N$ and $j = 0, 1, 2, \ldots, n-1$,

$$(3.48) \qquad \|H(z_{i+j}) - H(z_i)\| \le 0_1(\|h_i\|),$$

where, as before, $H(z) \triangleq \dfrac{\partial^2 f(z)}{\partial z^2}$.

Proof: First, note that for $j \in \{0, 1, 2, \ldots, n-1\}$

$$(3.49) \qquad \|H(z_{i+j}) - H(z_i)\| \le \sum_{k=0}^{j-1} \|H(z_{i+k+1}) - H(z_{i+k})\| \,.$$

Next, since $f(\cdot)$ is three times continuously differentiable, we obtain from the Taylor formula

(3.50) $$\|H(z_{i+k+1}) - H(z_{i+k})\| \le \int_0^1 \|DH(z_{i+k} + t\lambda_{i+k}h_{i+k})\,(\lambda_{i+k}h_{i+k})\|\,dt,$$

where $DH(\cdot)(\cdot)$ denotes the third derivative of $f(\cdot)$. Since $f(\cdot)$ is three times continuously differentiable, and since $z_i \to \hat{z}$ as $i \to \infty$, there exists a $b \in (0, \infty)$ such that for $i+k = 0, 1, 2, \ldots$.

(3.51) $$\|DH(z_{i+k} + t\lambda_{i+k}h_{i+k})\| \le b \text{ for all } t \in [0,1].$$

Consequently, because of (3.50), (3.29) and (3.42),

$$\|H(z_{i+j}) - H(z_i)\| \le n\,bq_4q_6\|h_i\| \triangleq 0_1(\|h_i\|) \text{ for all } i \ge N$$

where N is such that (3.29) and (3.42) hold. $\square$

(3.52) <u>Lemma</u>: There exists an integer N and a function 0_2: $\mathbb{R}^1 \to \mathbb{R}^1$ (satisfying (3.46) for $\ell = 2$) such that for all $i \ge N$ and $j = 0, 1, 2, \ldots, n-1$,

(3.53) $$\|H_{i+j} - H(z_i)\| \le 0_2(\|h_i\|),$$

where H_i was defined in (2.26).

<u>Proof</u>: First, making use of (3.48) for $j = 0, 1, 2, \ldots, n-1$ and $i \ge N$, we obtain

(3.54) $$\begin{aligned}\|H_{i+j} - H(z_i)\| &\le \|H_{i+j} - H(z_{i+j})\| + \|H(z_{i+j}) - H(z_i)\| = \\ &= \|H_{i+j} - H(z_{i+j})\| + 0_1(\|h_i\|)\end{aligned}$$

Next, making use of (2.26) and the mean value theorem we obtain,

(3.55) $$\begin{aligned}\|H_{i+j} - H(z_{i+j})\| &\le \int_0^1 \|\,H(z_{i+j} + t\lambda_{i+j}h_{i+j}) - H(z_{i+j})\,\|\,dt \\ &= \|H(z_{i+j} + \xi\lambda_{i+j}h_{i+j}) - H(z_{i+j})\|\end{aligned}$$

with $\xi \in [0,1]$. Then, proceeding as in the proof of Lemma (3.47) we conclude that

$$\|H_{i+j} - H(z_i)\| \leq o_1(\|h_i\|) + n\, b q_4 q_6 \|h_i\| \triangleq o_2(\|h_i\|) \quad \text{for } i \geq N \tag{3.56}$$

where N is such that (3.48) holds. $\square$

(3.57) Lemma: There exists an integer N and functions $o_3(\cdot)$, $o_4(\cdot)$ and $o_5(\cdot)$, from $\mathbb{R}^1$ into $\mathbb{R}^1$ and satisfying (3.46), such that for $i \geq N$, $i \in I_v$, and $j = 0, 1, 2, \ldots, n-1$,

$$\|h_{i+j+1} - h_i^{j+1}\| \leq o_3(\|h_{i+j} - h_i^{j}\|) + o_4(\|g_{i+j+1} - g_i^{j+1}\|) + o_5(\|h_i\|^2) \,. \tag{3.58}$$

Proof: In what follows, we assume that $i \geq N$, where N is an integer sufficiently large for all lemmas used to apply. First, suppose that $v = n$ and $j = n-1$. Then, since $\omega(i+n) = 0$ for $i \in I_n$, it follows from (3.7) and and (3.8) that

$$h_{i+n} = g_{i+n}, \quad i \in I_n \tag{3.59}$$

and hence, since $h_i^n = g_i^n = 0$, for $i \in I_n$,

$$\|h_{i+n} - h_i^n\| = \|g_{i+n} - g_i^n\|, \quad i \in I_n \,. \tag{3.60}$$

Consequently, (3.58) is satisfied for $i \in I_v$ and $j = n-1$ when $v = n$. In what follows, we assume that either $v > n$ and $j \in \{1, 2, \ldots, n-1\}$ or that $v = n$ and $j \in \{1, 2, \ldots, n-2\}$. Now,

$$\begin{aligned} \|h_{i+j+1} - h_i^{j+1}\| &= \|g_{i+j+1} + \gamma_{i+j} h_{i+j} - g_i^{j+1} - \gamma_i^{j} h_i^{j}\| \\ &\leq \|g_{i+j+1} - g_i^{j+1}\| + \|\gamma_{i+j} h_{i+j} - \gamma_i^{j} h_i^{j}\| \,. \end{aligned} \tag{3.61}$$

We shall now obtain a bound on $\|\gamma_{i+j}h_{i+j} - \gamma_i^j h_i^j\|$ in (3.61). Thus, by (3.7) and (2.30),

$$\text{(3.62)} \qquad \gamma_{i+j} = \frac{\langle g_{i+j+1} - g_{i+j},\ g_{i+j+1} \rangle}{\|g_{i+j}\|^2}$$

$$= - \frac{\langle g_{i+j+1},\ H_{i+j}h_{i+j} \rangle}{\|g_{i+j}\|^2} \lambda_{i+j}$$

$$= - \frac{\langle g_{i+j+1},\ H_{i+j}\ h_{i+j} \rangle}{\|g_{i+j}\|^2} \ \frac{\langle g_{i+j} - g_{i+j+1},\ h_{i+j} \rangle}{\langle h_{i+j},\ H_{i+j}h_{i+j} \rangle}$$

$$= - \frac{\langle g_{i+j+1},\ H_{i+j}h_{i+j} \rangle}{\langle h_{i+j},\ H_{i+j}h_{i+j} \rangle} -$$

$$- \gamma_{i+j-1} \frac{\langle g_{i+j+1},\ H_{i+j}h_{i+j} \rangle \langle g_{i+j},\ h_{i+j-1} \rangle}{\|g_{i+j}\|^2 \langle h_{i+j},\ H_{i+j}h_{i+j} \rangle} +$$

$$+ \frac{\langle g_{i+j+1},\ H_{i+j}h_{i+j} \rangle \langle g_{i+j+1},\ h_{i+j} \rangle}{\|g_{i+j}\|^2 \langle h_{i+j},\ H_{i+j}\ h_{i+j} \rangle} \quad .$$

From (3.34),

$$\text{(3.63)} \qquad \gamma_i^j = - \frac{\langle g_i^{j+1},\ H(z_i)h_i^j \rangle}{\langle h_i^j,\ H(z_i)h_i^j \rangle} \quad .$$

Consequently,

$$\text{(3.64)} \qquad \|\gamma_{i+j}h_{i+j} - \gamma_i^j h_i^j\| \leq \left\| - \frac{\langle g_{i+j+1},\ H_{i+j}h_{i+j} \rangle}{\langle h_{i+j},\ H_{i+j}h_{i+j} \rangle} h_{i+j} + \frac{\langle g_i^{j+1},\ H(z_i)h_i^j \rangle}{\langle h_i^j,\ H(z_i)h_i^j \rangle} h_i^j \right\|$$

$$+ \left\| \frac{\gamma_{i+j-1} \langle g_{i+j+1}, H_{i+j} h_{i+j} \rangle \langle g_{i+j}, h_{i+j-1} \rangle}{\|g_{i+j}\|^2 \langle h_{i+j}, H_{i+j} h_{i+j} \rangle} h_{i+j} \right\|$$

$$+ \left\| \frac{\langle g_{i+j+1}, H_{i+j} h_{i+j} \rangle \langle g_{i+j+1}, h_{i+j} \rangle}{\|g_{i+j}\|^2 \langle h_{i+j}, H_{i+j} h_{i+j} \rangle} h_{i+j} \right\| .$$

Next, since by construction $|\langle g_{i+j}, h_{i+j-1} \rangle| \leq \|g_{i+j-1}\| \, \|g_{i+j}\| \, \|h_{i+j-1}\|$, and making use of (1.9), (3.22), (3.25), (3.30) and (3.42), we obtain (see second term in (3.64))

$$p_{i+j} \triangleq \frac{|\gamma_{i+j-1}| \, |\langle g_{i+j+1}, H_{i+j} h_{i+j} \rangle| \, |\langle g_{i+j}, h_{i+j-1} \rangle| \, \|h_{i+j}\|}{\|g_{i+j}\|^2 \langle h_{i+j}, H_{i+j} h_{i+j} \rangle} \tag{3.65}$$

$$\leq \frac{q_1 \|g_{i+j+1}\| M \|h_{i+j}\| \, \|g_{i+j}\| \, \|h_{i+j-1}\|^2 \, \|h_{i+j}\|}{q_3 \|g_{i+j}\|^2 \, \|h_{i+j}\|^2 \, m}$$

$$\leq \frac{q_1 q_6^2 M}{q_3 m} \frac{\|g_{i+j+1}\|}{\|g_{i+j}\|} \|h_i\|^2 \leq \frac{q_1 q_2 q_5 q_6^2 M \|h_i\|^2}{q_3 m}$$

Similarly (see third term in (3.64)),

$$s_{i+j} \triangleq \frac{|\langle g_{i+j+1}, H_{i+j} h_{i+j} \rangle| \, |\langle g_{i+j+1}, h_{i+j} \rangle| \, \|h_{i+j}\|}{\|g_{i+j}\|^2 \langle h_{i+j}, H_{i+j} h_{i+j} \rangle} \tag{3.66}$$

$$\leq \frac{M \|g_{i+j+1}\| \, \|h_{i+j}\| \, \|g_{i+j}\| \, \|g_{i+j+1}\| \, \|h_{i+j}\|^2}{\|g_{i+j}\|^2 \, m \|h_{i+j}\|^2}$$

$$\leq \frac{q_2 \, q_6^2 \, M \, \|h_i\|^2}{m} .$$

Now, let

$$t_{i+j} \triangleq \langle h_{i+j}, H_{i+j} h_{i+j} \rangle \langle h_i^j, H(z_i^j) h_i^j \rangle . \tag{3.67}$$

Then (see first term in (3.64)), by adding and subtracting terms, we obtain

$$(3.68)\qquad v_{i+j} \triangleq \left\| \frac{\langle g_i^{j+1}, H(z_i)\, h_i^j \rangle}{\langle h_i^j, H(z_i)\, h_i^j \rangle}\, h_i^j - \frac{\langle g_{i+j+1}, H_{i+j} h_{i+j} \rangle}{\langle h_{i+j}, H_{i+j} h_{i+j} \rangle}\, h_{i+j} \right\|$$

$$= \frac{1}{t_{i+j}} \| \langle h_{i+j}, H_{i+j} h_{i+j} \rangle \langle g_i^{j+1}, H(z_i) h_i^j \rangle\, h_i^j -$$

$$- \langle h_i^j, H(z_i) h_i^j \rangle \langle g_{i+j+1}, H_{i+j} h_{i+j} \rangle\, h_{i+j} \|$$

$$\le \frac{1}{t_{i+j}} \{ \| \langle h_{i+j}, H_{i+j} h_{i+j} \rangle \langle g_i^{j+1}, H(z_i)\, (h_i^j - h_{i+j}) \rangle h_i^j \|$$

$$+ \| \langle h_{i+j}, H_{i+j} (h_{i+j} - h_i^j) \rangle \langle g_i^{j+1}, H(z_i) h_{i+j} \rangle h_i^j \|$$

$$+ \| \langle h_{i+j}, H_{i+j}\, h_i^j \rangle \langle g_i^{j+1} - g_{i+j+1}, H(z_i) h_{i+j} \rangle h_i^j \|$$

$$+ \| \langle h_{i+j} - h_i^j, H(z_i) h_i^j \rangle \langle g_{i+j+1}, H(z_i) h_{i+j} \rangle h_i^j \|$$

$$+ \| \langle h_i^j, (H_{i+j} - H(z_i)) h_i^j \rangle \langle g_{i+j+1}, H(z_i) h_{i+j} \rangle h_i^j \|$$

$$+ \| \langle h_i^j, H(z_i) h_i^j \rangle \langle g_{i+j+1}, (H(z_i) - H_{i+j}) h_{i+j} \rangle h_i^j \|$$

$$+ \| \langle h_i^j, H(z_i) h_i^j \rangle \langle g_{i+j+1}, H_{i+j} h_{i+j} \rangle\, (h_i^j - h_{i+j}) \| \}\,.$$

Since, by (1.9),

$$(3.69)\qquad \frac{1}{t_{i+j}} \le \frac{1}{m^2 \|h_{i+j}\|^2 \|h_i^j\|^2}\,,$$

and because of (1.9) and (3.56), we obtain from (3.67) that

$$(3.70)\qquad v_{i+j} \le \frac{1}{m^2}\Big\{ \frac{M^2\|\overset{j+1}{g_i}\|\,\|\overset{j}{h_i} - h_{i+j}\|}{\|\overset{j}{h_i}\|} + $$

$$+ \frac{M^2\|\overset{j+1}{g_i}\|\,\|\overset{j}{h_i} - h_{i+j}\|}{\|\overset{j}{h_i}\|} + M^2\|\overset{j+1}{g_i} - g_{i+j+1}\| + $$

$$+ \frac{M^2\,\|g_{i+j+1}\|\,\|\overset{j}{h_i} - h_{i+j}\|}{\|h_{i+j}\|} + \frac{M\|g_{i+j+1}\|O_2(\|h_i\|)\|\overset{j}{h_i}\|}{\|h_{i+j}\|}$$

$$+ \frac{M\|g_{i+j+1}\|O_2(\|h_i\|)\|\overset{j}{h_i}\|}{\|h_{i+j}\|} + \frac{M^2\|g_{i+j+1}\|\,\|\overset{j}{h_i} - h_{i+j}\|}{\|h_{i+j}\|}\Big\}$$

Finally, making use of (3.37), (3.30) and (3.44), we obtain,

$$(3.71)\qquad v_{i+j} \le \frac{M}{m^2}\{2\bar{q}_5 M\|\overset{j}{h_i} - h_{i+j}\| + M\|\overset{j+1}{g_i} - g_{i+j+1}\|$$

$$+ 2Mq_5\|\overset{j}{h_i} - h_{i+j}\| + 2q_5q_6O_2(\|h_i\|)\|h_i\|\}$$

$$= \frac{2M^2}{m^2}[\bar{q}_5 + q_5]\|\overset{j}{h_i} - h_{i+j}\| + \frac{M^2}{m^2}\|\overset{j+1}{g_i} - g_{i+j+1}\|$$

$$+ \frac{M}{m^2}q_5q_6O_2(\|h_i\|)\|h_i\| .$$

Now, from (3.61), (3.62) (3.64), (3.65), (3.66) and (3.68) we obtain,

$$(3.72)\qquad \|h_{i+j+1} - \overset{j+1}{h_i}\| \le \|g_{i+j+1} - \overset{j+1}{g_i}\| + \|\overset{j}{\gamma_i}\overset{j}{h_i} - \gamma_{i+j}h_{i+j}\|$$

$$\leq \|g_{i+j+1} - \overset{j+1}{g}_i\| + v_{i+j} + p_{i+j} + s_{i+j}$$

$$\leq 0_3(\|h_{i+j} - \overset{j}{h}_i\|) + 0_4(\|g_{i+j+1} - \overset{j+1}{g}_i\|)$$

$$+ 0_5(\|h_i\|^2),$$

where 0_3, 0_4 and 0_5 are defined in an obvious way from the relations (3.65), (3.66) and (3.71). Hence the lemma is true for $i \in I_v$ and $j = 1, 2, \ldots, n-1$.

Finally, for $i \in I_v$ and $j = 0$, $h_{i+j} = g_{i+j}$. Consequently, (see (3.62)) for $i \in I_v$,

$$\gamma_i = - \frac{\langle g_{i+1}, H_i\, h_i \rangle}{\langle h_i, H_i\, h_i \rangle} \qquad + \qquad \frac{\langle g_{i+1}, H_i\, h_i \rangle \langle g_{i+1}, h_i \rangle}{\|g_i\|^2 \ \langle h_i, H_i\, h_i \rangle} .$$

As a result, proceeding as before we can show that $\|h_{i+1} - \overset{1}{h}_i\| \leq \|g_{i+1} - \overset{j}{g}_i\| + v_i + s_i$, where v_i and s_i are defined as in (3.66) and (3.68) respectively, for $i = 0$. Since $p_i \geq 0$, it follows that (3.72) is also true for $j = 0$. $\square$

(3.73) Lemma: There exists an integer N and a function 0_6: $\mathbb{R}^1 \to \mathbb{R}^1$, satisfying (3.46), such that for all $i \geq N$, $i \in I_v$, $j = 0, 1, \ldots, n-1$,

$$(3.74) \qquad \|g_{i+j+1} - \overset{j+1}{g}_i\| \leq \|g_{i+j} - \overset{j}{g}_i\| + 0_6(\|h_i\|^2) + M\|\lambda_{i+j}h_{i+j} - \overset{j}{\lambda}_i\overset{j}{h}_i\| .$$

Proof: Since by (2.30), $- g_{i+j+1} = - g_{i+j} + \lambda_{i+j}\, H_{i+j}h_{i+j}$ and, similarly, $- \overset{j+1}{g}_i = - \overset{j}{g}_i + \overset{j}{\lambda}_i H(\overset{j}{z}_i)\overset{j}{h}_i$, it follows that for $i \in I_v$ and $j \in \{0,1,2,\ldots, n-1\}$

$$(3.75)\qquad \|g_{i+j+1} - g_i^{j+1}\| \le \|g_{i+j} - g_i^{j}\| + \|H_{i+j}\lambda_{i+j}h_{i+j} - H(z_i)\lambda_i^{j}h_i^{j}\|$$

$$\le \|g_{i+j} - g_i^{j}\| + \|(H_{i+j} - H(z_i))\lambda_i^{j}h_i^{j}\| +$$

$$+ \|(H_{i+j}(\lambda_{i+j}h_{i+j} - \lambda_i^{j}h_i^{j})\|$$

$$\le \|g_{i+j} - g_i^{j}\| + \bar{q}_4 q_6 o_2(\|h_i\|)\ \|h_i\| +$$

$$+ M\|\lambda_{i+j}h_{i+j} - \lambda_i^{j}h_i^{j}\| \ ,$$

where we have made use of (1.9), (3.38), (3.44) and (3.53). Setting $o_6(\|h_i\|^2) \equiv \bar{q}_4 q_6 o_2(\|h_i\|)\|h_i\|$, we obtain (3.74) from (3.75) . □

(3.76) <u>Lemma</u>: There exists an integer N and functions $o_7(\cdot)$, $o_8(\cdot)$ and $o_9(\cdot)$, from $\mathbb{R}^1$ into $\mathbb{R}^1$ and satisfying (3.46), such that for all $i \ge N$, $i \in I_v$, and $j = 0, 1, \ldots, n-1$,

$$(3.77)\qquad \|\lambda_{i+j}h_{i+j} - \lambda_i^{j}h_i^{j}\| \le o_7(\|g_{i+j} - g_i^{j}\|) + o_8(\|h_{i+j} - h_i\|) + o_9(\|h_i\|^2) \ .$$

<u>Proof</u>: Suppose that $i \in I_v$, $j \in \{0,1,\ldots, n-1\}$. Making use of (2.32) and (3.33) we obtain

$$(3.78)\qquad \|\lambda_{i+j}h_{i+j} - \lambda_i^{j}h_i^{j}\| \le \left\| \frac{\langle g_{i+j}, h_{i+j}\rangle}{\langle h_{i+j}, H_{i+j}h_{i+j}\rangle}\, h_{i+j} - \frac{\langle g_i^{j}, h_i^{j}\rangle}{\langle h_i^{j}, H(z_i)h_i^{j}\rangle}\, h_i^{j}\right\|$$

$$+ \left\| \frac{\langle g_{i+j+1}, h_{i+j}\rangle}{\langle h_{i+j}, H_{i+j}h_{i+j}\rangle}\, h_{i+j}\right\| \ .$$

Now, making use of (1.9), of the test in Step 13 of (3.1), (3.41) and (3.42), we obtain

$$(3.79)\qquad \frac{|\langle g_{i+j+1}, h_{i+j}\rangle|\ \|h_{i+j}\|}{\langle h_{i+j}, H_{i+j}h_{i+j}\rangle} \le \frac{\|g_{i+j}\|\ \|g_{i+j+1}\|\ \|h_{i+j}\|^2}{m\|h_{i+j}\|^2} \le \frac{q_6^2\|h_i\|^2}{m}\ .$$

Next, defining t_{i+j} as in (3.67), and first adding and subtracting terms and then making use of (1.9), (3.25), (3.36) and (3.53), we obtain

$$(3.80)\qquad \left\| \frac{\langle g_{i+j}, h_{i+j}\rangle}{\langle h_{i+j}, H_{i+j}h_{i+j}\rangle}\ h_{i+j} - \frac{\langle g_i^j, h_i^j\rangle}{\langle h_i^j, H(z_i)h_i^j\rangle}\ h_i^j\right\| =$$

$$\frac{1}{t_{i+j}}\ \|\langle g_{i+j}, h_{i+j}\rangle\ \langle h_i^j, H(z_i)h_i^j\rangle\ h_{i+j} -$$

$$\langle g_i^j, h_i^j\rangle\ \langle h_{i+j}, H_{i+j}h_{i+j}\rangle\ h_i^j\| \le$$

$$\le \frac{1}{t_{i+j}}\ \{\ \|\langle g_{i+j}, h_{i+j} - h_i^j\rangle\ \langle h_i^j, H(z_i)\ h_i^j\rangle\ h_{i+j}\|$$

$$+ \|\langle g_{i+j}, h_i^j\rangle\ \langle h_i^j - h_{i+j}, H(z_i)h_i^j\rangle\ h_{i+j}\|$$

$$+ \|\langle g_{i+j} - g_i^j, h_i^j\rangle\ \langle h_{i+j}, H(z_i)h_i^j\rangle\ h_{i+j}\|$$

$$+ \|\langle g_i^j, h_i^j\rangle\ \langle h_{i+j}, H(z_i)\ (h_i^j - h_{i+j})\rangle\ h_{i+j}\|$$

$$+ \|\langle g_i^j, h_i^j\rangle\ \langle h_{i+j}, (H(z_i) - H_{i+j})\ h_{i+j}\rangle\ h_{i+j}\|$$

$$+ \|\langle g_i^j, h_i^j\rangle\ \langle h_{i+j}, H_{i+j}h_{i+j}\rangle\ (h_{i+j} - h_i^j)\|\}$$

$$\le \frac{1}{m^2}\ \{\ \frac{M\|g_{i+j}\|\ \|h_{i+j} - h_i^j\|}{\|h_{i+j}\|} + \frac{M\|g_{i+j}\|\ \|h_{i+j} - h_i^j\|}{\|h_{i+j}\|}$$

$$+ M\|g_{i+j} - \overset{j}{g}_i\| + \frac{M\|\overset{j}{g}_i\| \, \|h_{i+j} - \overset{j}{h}_i\|}{\|\overset{j}{h}_i\|}$$

$$+ \frac{\|\overset{j}{g}_i\| \, \|h_{i+j}\| \, 0_2(\|h_i\|)}{\|\overset{j}{h}_i\|} + \frac{M\|\overset{j}{g}_i\|\|h_{i+j} - \overset{j}{h}_i\|}{\|\overset{j}{h}_i\|} \}$$

$$\leq \frac{1}{m^2} \{2M(\frac{1}{q_3} + \frac{1}{\bar{q}_3}) \|h_{i+j} - \overset{j}{h}_i\| + M\|g_{i+j} - \overset{j}{g}_i\| +$$

$$+ \frac{q_6}{\bar{q}_3} 0_2(\|h_i\|) \|h_i\| \} \; .$$

Since the existence, of functions $0_7(\cdot)$, $0_8(\cdot)$ and $0_9(\cdot)$ satisfying (3.77), follows directly from (3.78), (3.79) and (3.80), we are done. $\square$

(3.81) <u>Lemma</u>: There exists an integer N and functions $0_{10}(\cdot)$, $0_{11}(\cdot)$ and $0_{12}(\cdot)$ from $\mathbb{R}^1$ into $\mathbb{R}^1$ and satisfying (3.46) such that for $i \geq N$ $i \in I_v$, and $j = 0, 1, 2, \ldots, n-1$,

(3.82) $$\|g_{i+j} - \overset{j}{g}_i\| \leq 0_{10}(\|h_i\|^2)$$

(3.83) $$\|h_{i+j} - \overset{j}{h}_i\| \leq 0_{11}(\|h_i\|^2)$$

(3.84) $$\|\lambda_{i+j}h_{i+j} - \overset{j}{\lambda}_i\overset{j}{h}_i\| \leq 0_{12}(\|h_i\|^2) \; .$$

<u>Proof:</u> We make use of induction. Let N be an integer for which the conclusions of Lemmas (3.73), (3.57) and (3.77) hold, and let i be any positive integer satisfying $i \geq N$ and $i \in I_v$. Then, for $j = 0$, $\overset{0}{g}_i = g_i =$ $\overset{0}{h}_i = h_i$ and hence

$$\|g_i - \overset{0}{g}_i\| = \|h_i - \overset{0}{h}_i\| = 0, \tag{3.85}$$

and

$$\|\lambda_i h_i - \overset{0}{\lambda}_i \overset{0}{h}_i\| \le O_9(\|h_i\|^2) \; . \tag{3.86}$$

Now, for any $j \in \{0, 1, 2, \ldots, n-2\}$, suppose that there exist functions $\overset{j}{O}_{10}(\cdot)$, $\overset{j}{O}_{11}(\cdot)$ and $\overset{j}{O}_{12}(\cdot)$, satisfying (3.46), such that

$$\|g_{i+j} - \overset{j}{g}_i\| \le \overset{j}{O}_{10}(\|h_i\|^2) \tag{3.87}$$

$$\|h_{i+j} - \overset{j}{h}_i\| \le \overset{j}{O}_{11}(\|h_i\|^2) \tag{3.88}$$

$$\|\lambda_{i+j} h_{i+j} - \overset{j}{\lambda}_i \overset{j}{h}_i\| \le \overset{j}{O}_{12}(\|h_i\|^2) \; . \tag{3.89}$$

Then, from (3.73), (3.87) and (3.89),

$$\begin{aligned} \|g_{i+j+1} - \overset{j+1}{g}_i\| &\le \overset{j}{O}_{10}(\|h_i\|^2) + O_6(\|h_i\|^2) + M\overset{j}{O}_{12}(\|h_i\|^2) \\ &\triangleq \overset{j+1}{O}_{10}(\|h_i\|^2), \end{aligned} \tag{3.90}$$

where, obviously, $\overset{j+1}{O}_{10}(\cdot)$ is a function satisfying (3.46).

Next, from (3.58), (3.88) and (3.90),

$$\begin{aligned} \|h_{i+j+1} - \overset{j+1}{h}_i\| \le{} & [O_3(\|h_{i+j} - \overset{j}{h}_i\|)/\|h_{i+j} - \overset{j}{h}_i\|]\overset{j}{O}_{11}(\|h_i\|^2) \\ & + [O_4(\|g_{i+j+1} - \overset{j+1}{g}_i\|)/\|g_{i+j+1} - \overset{j+1}{g}_i\|]\overset{j}{O}_{10}(\|h_i\|^2) \\ & + O_5(\|h_i\|^2) \; . \end{aligned} \tag{3.92}$$

Since (3.88) and (3.90) hold, it follows from (3.92) that there exists a function $\overset{j+1}{O}_{11}(\cdot)$ satisfying (3.46) such that

(3.93) $$\| h_{i+j+1} - \overset{j+1}{h}_i \| \leq \overset{j+1}{O}_{11}(\| h_i \|^2).$$

Finally, making use of (3.77), (3.90) and (3.93) we obtain

(3.94) $$\| \lambda_{i+j+1} h_{i+j+1} - \overset{j+1}{\lambda}_i \overset{j+1}{h}_i \| \leq [O_7(\| g_{i+j+1} - \overset{j+1}{g}_i \|)/\| g_{i+j+1} - \overset{j+1}{g}_i \|] \cdot$$
$$\cdot \overset{j+1}{O}_{10}(\| h_i \|^2) + [O_8(\| h_{i+j+1} - \overset{j+1}{h}_i \|)/\| h_{i+j+1} - \overset{j+1}{h}_i \|] \overset{j+1}{O}_{11}(\| h_i \|^2)$$
$$+ O_9(\| h_i \|^2) .$$

Since (3.90) and (3.93) are satisfied, (3.94) implies that there exists a function $\overset{j+1}{O}_{12}(\cdot)$ satisfying (3.46) such that

(3.95) $$\| \lambda_{i+j+1} h_{i+j+1} - \overset{j+1}{\lambda}_i \overset{j+1}{h}_i \| \leq \overset{j+1}{O}_{12}(\| h_i \|^2).$$

Since (3.87)-(3.89) are true for $j = 0$, we see that they must also be true for $j = 1, 2, \ldots, n-1$. To complete the proof, we set $O_{10}(\cdot) = \max_j \overset{j}{O}_{10}(\cdot)$, $O_{11}(\cdot) = \max_j \overset{j}{O}_{11}(\cdot)$ and $O_{12}(\cdot) = \max_j \overset{j}{O}_{12}(\cdot)$. □

We are finally ready to establish our main result.

(3.96) <u>Theorem</u>: There exists an integer N' and a $q \in (0, \infty)$, such that for all $i \geq N'$, $i \in I_v$, and $j = 0, 1, 2, \ldots, n-1$,

$$(3.97)\qquad \|\lambda_{i+j}h_{i+j} - \lambda_i^j h_i^j\| \leq q\|z_i - \hat{z}\|,$$

where $\hat{z}$ is the unique minimizer of $f(\cdot)$.

Proof: First, by the Taylor formula,

$$(3.98)\qquad -g_i = \nabla f(z_i) = \nabla f(\hat{z}) + \int_0^1 H(z_i + t(z_i - \hat{z}))\,dt\,(z_i - \hat{z})$$

and hence, because of (1.9) and because $\nabla f(\hat{z}) = 0$,

$$(3.99)\qquad \|g_i\| \leq M\,\|z_i - \hat{z}\|.$$

Now, since $g_i \to 0$ as $i \to \infty$, because $z_i \to \hat{z}$ as $i \to \infty$, it follows from (3.25) that $h_i \to 0$ as $i \to \infty$. Consequently because of (3.84), (3.46), (3.25) and (3.99) there exists an integer $N' \geq N$ such that for all $i \geq N'$, $i \in I_v$, and $j = 0, 1, 2, \ldots, n-1$,

$$(3.100)\qquad \begin{aligned} \|\lambda_{i+j}\,h_{i+j} - \lambda_i^j\,h_i^j\| &\leq O_{12}(\|h_i\|^2) \leq r_{12}\|h_i\|^2 \\ &\leq r_{12}q_2^2\|g_i\|^2 \leq r_{12}q_2^2\,M^2\|z_i - \hat{z}\|^2 \\ &\triangleq q\|z_i - \hat{z}\|^2, \end{aligned}$$

where $r_{12} > 0$ is such that (3.46) holds for $\ell = 12$. This completes our proof.

Thus, the assumptions of Theorem (3.13) are indeed satisfied.

Conclusion:

We have shown in this paper that it is possible to construct a superlinearly convergent, conjugate gradient algorithm which does not include the minimization of a function along a line as a subprocedure. A most important consequence of this is that unlike its "theoretical" predecessors, our algorithm is directly implementable, i.e. it can be programmed as stated, without any need for heuristics to circumvent non-implementable operations. Of the two versions stated in this paper, we have programmed the first one, (2.1), and have tested it against a few standard problems such as the Rosenbrock's valley. The empirical results show that this version converges at about the same rate as the more complex version (3.1) (i.e. superlinearly), and hence this is the version which we would normally recommend.

The reason, for which we have used a gradient type subprocedure in the step size calculation, is that we wanted to make sure that the algorithm could be used in the minimization of nonconvex functions as well. Although there is no theory to justify such a practice, empirical results show that the algorithm does work for the nonconvex case (e.g. Rosenbrock's valley problem). For the convex case, the step size calculations could be carried out by using an adaptation of the golden section rule, which results in a more efficient algorithm.

References

[1] E. Polak, Computational methods in optimization: a unified approach, Academic Press, 1971.

[2] E. Polak and G. Ribière, Note sur la convergence de methodes de directions conjuguees, Revue Francaise d'Informatique et de Recherche Operationelle, no. 16-R1, 1969, pp. 35-43.

[3] R. Fletcher and C. M. Reeves, Function minimization by conjugate gradients, British Computer Journal, vol. 7, no. 2, 1964, pp. 149-154.

[4] A. Cohen, Rate of convergence and optimality conditions for root finding and optimization algorithms, Ph.D Dissertation, University of California, Berkeley, 1970.

[5] L. Armijo, Minimization of functions having first partial derivatives, Pacific J. Math. Vol 16,1966, pp. 1-3.

[6] J. Daniel, The conjugate gradient method for linear and nonlinear operator equations, SIAM J. Numer. Anal., vol. 4, no. 1, 1967, pp. 10-26.