The Pennsylvania State University

The Graduate School

N-STEP QUADRATIC CONVERGENCE IN THE CONJUGATE GRADIENT METHOD

A Thesis in

Mathematics

by

Holly Peters Hirst

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 1989

We approve the Thesis of Holly Peters Hirst.

Date of Signature

William W. Hager — March 17, 1989

William W. Hager
Professor of Mathematics
Thesis Advisor
Chair of Committee

Allan M. Krall — March 20, 1989

Allan M. Krall
Professor of Mathematics

Pan. M. Pardalos — March 17, 1989

Panayote Pardalos
Assistant Professor of
Computer Science

L. Ridgway Scott — 22 March 1989

L. Ridgway Scott
Professor of Mathematics
Professor of Computer Science

W. Dale Brownawell — March 20, 1989

W. Dale Brownawell
Professor of Mathematics
Acting Chair of the Department
of Mathematics

ABSTRACT

For the past twenty years, conjugate gradient techniques have been used with great success to find the unconstrained minimum of a real valued function. They have several attractive features to which their popularity can be attributed; the values of the function and its first derivatives are used, but the second derivatives are not necessary. There is no matrix to store. The sequences generated by these iterative techniques show rapid convergence.

The convergence properties of the basic Fletcher-Reeves (FR) version of the algorithm with restart are well known. We will concern ourselves with the rate of convergence of this algorithm and simplifications which decrease the computational complexity, yet still preserve the desirable convergence rate. We will see that the basic algorithm with restart is "n-step" quadratically convergent under certain conditions, including a condition on the convexity of the Hessian of the function, as shown by A. I. Cohen [SIAM J. Num. Anal. **9** (1972)]. Even without the time consuming exact line searches, the algorithm can be implemented to retain this desirable convergence rate.

The main result of this work involves a particular simplification of the line search based on quadratic interpolation which is easier to implement than others previously suggested in the literature. In the main theorem, which is based upon work by M. L. Lenard [Math. Prog. **10** (1976)], local n-step quadratic convergence for this algorithm is proven under certain conditions, and several standard variants of the algorithm are examined. Results from the numerical testing provide insight into the behavior of the algorithm when applied to problems which violate the conditions of the rate-of-convergence theorem. Even when the convexity condition is violated, the algorithm exhibits super-linear convergence.

TABLE OF CONTENTS

Chapter 1
THE CONJUGATE GRADIENT METHOD

For the past twenty years, conjugate gradient techniques have been used with great success to find the unconstrained minimum of a real valued function. They have several attractive features to which their popularity can be attributed; the values of the function and its first derivatives are used, but the second derivatives are not necessary. There is no matrix to store. The sequences generated by these iterative techniques show rapid convergence.

1.1. History of the Algorithm

The method of conjugate gradients was first presented in 1952 by Hestenes and Stiefel [12] as a technique for solving systems of linear equations. It was then applied to unconstrained minimization in 1964 by Fletcher and Reeves [8], which led to different implementations by Polak and Ribiere [16] and Daniel [4], among others. These algorithms are presently used extensively in constrained minimization solvers such as multiplier methods, where sequences of unconstrained minimizations of the augmented Lagrangian function are performed. As these unconstrained minimizations are the most time consuming part of this type of constrained optimization solver, an efficient simplification of the conjugate gradient method which retains a desirable convergence rate is attractive. For examples, see Bertsekas [1] and Hager [9, 10].

The convergence properties of the basic algorithm are well known. Our work will focus upon the rate of convergence and simplifications in the algorithm which decrease the computational complexity, yet preserve the desirable convergence rate.

We will see that the basic algorithm with restart has local "n-step" quadratic convergence under certain conditions, and that, even without the time consuming exact line searches, the algorithm can be implemented to retain this rate. We will present a simplification of the algorithm which is easier to implement than others previously discussed in the literature, and prove that this variation is n-step quadratically convergent. It is also of interest to note that the locally n-step quadratically convergent algorithm we present can be converted into a globally quadratically convergent algorithm by following the procedure presented by Hager in [11].

1.2. Basic Notation and Definitions

We will use the following notation.

- $\mathbf{R}^n$:= Euclidean n-space
- $\| \ \|$:= Euclidean norm
- ∇f := column vector gradient of f
- $\mathbf{H}(\mathbf{x})$:= Hessian matrix $\nabla^2 f(\mathbf{x})$
- $\mathbf{C}^k(\mathbf{R}^n)$:= space of all k times continuously differentiable functions from $\mathbf{R}^n$ to $\mathbf{R}$
- $B_\delta(\alpha)$:= $\{x : \|x-\alpha\| < \delta\}$
- superscript T denotes vector or matrix transpose
- let $s : \mathbf{R} \to \mathbf{R}$, then $s(\xi) = O(\xi)$ implies $\exists r, c < \infty$ s.t. $\forall \xi \in B_r(0)$ $|s(\xi)| \le c\,|\xi|$

The conjugate gradient method is a variant of the method of steepest descent. The basic algorithm proceeds as follows.

(1.1) <u>Basic Conjugate Gradient Algorithm</u>

Given arbitrary $\mathbf{x}_0 \in \mathbf{R}^n$:

1. $\mathbf{g}_0 = \nabla f(\mathbf{x}_0)$

2. $\mathbf{d}_0 = -\mathbf{g}_0$

3. for k = 0, 1, 2, 3, . . .

   i. $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$

   where $\alpha_k = \underset{\alpha \geq 0}{\arg\min} f(\mathbf{x}_k + \alpha \mathbf{d}_k)$

   ii. $\mathbf{g}_{k+1} = \nabla f(\mathbf{x}_{k+1})$; check stopping criterion

   iii. $\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k$

   where β_k depends upon the specific method used.

There are four main variations of the conjugate gradient method. They differ from each other in the choice of β_k. In the Fletcher-Reeves (FR) method:

$$\beta_k = \frac{-\mathbf{g}_{k+1}^T \mathbf{g}_{k+1}}{\mathbf{g}_k^T \mathbf{g}_k}.$$

In the Hestenes-Steifel (HS) method:

$$\beta_k = \frac{(\mathbf{g}_{k+1} - \mathbf{g}_k)^T \mathbf{g}_{k+1}}{(\mathbf{g}_{k+1} - \mathbf{g}_k)^T \mathbf{d}_k}.$$

In the Polak-Ribiere (PR) method:

$$\beta_k = \frac{(\mathbf{g}_{k+1} - \mathbf{g}_k)^T \mathbf{g}_{k+1}}{\mathbf{g}_k^T \mathbf{g}_k}.$$

In the Daniel (D) method:

$$\beta_k = \frac{\mathbf{g}_{k+1}^T \mathbf{H}_{k+1} \mathbf{d}_k}{\mathbf{d}_k^T \mathbf{H}_{k+1} \mathbf{d}_k}$$

where $\mathbf{H}_{k+1} = \mathbf{H}(\mathbf{x}_{k+1})$.

The algorithm was originally developed for the quadratic case:

$$\text{minimize } f(\mathbf{x}) \text{ over } \mathbf{x} \in \mathbf{R}^n$$

where

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \mathbf{b}^T\mathbf{x}$$

and **Q** is a symmetric positive definite n-dimensional matrix. In this case, the algorithm reduces to:

(1.2) <u>Reduced Conjugate Gradient Algorithm</u>

For k = 1, 2, 3, . . .

1. $\mathbf{g}_k = \mathbf{Q}\mathbf{x}_k - \mathbf{b}$

2. $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$, where $\alpha_k = \dfrac{-\mathbf{g}_k^T \mathbf{d}_k}{\mathbf{d}_k^T \mathbf{Q} \mathbf{d}_k}$

3. $\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k$, where $\beta_k = \dfrac{\mathbf{g}_{k+1}^T \mathbf{Q} \mathbf{d}_k}{\mathbf{d}_k^T \mathbf{Q} \mathbf{d}_k}$

Note that each of the formulas for β_k reduces to this simplified form when applied to the quadratic problem above. In fact, the method was originally extended to nonquadratics by reversing this process; the β_k term was generalized.

The convergence properties of the conjugate gradient method are well known. We refer the reader to Luenberger [**15**, Chapter 8] for proofs of the following theorems, and a more in-depth treatment of the theory.

(1.3) DEFINITION. *Given a symmetric matrix* **Q**, *two vectors* $\mathbf{d}_1$ *and* $\mathbf{d}_2$ *are called* **Q**-*conjugate iff* $\mathbf{d}_1^T\mathbf{Q}\mathbf{d}_2 = 0$.

(1.4) THEOREM. (Conjugate Directions) *Let* $\mathbf{Q}$ *be a symmetric positive definite n-dimensional matrix,* $\mathbf{b} \in \mathbf{R}^n$, $\mathbf{b} \neq 0$, *and let* $\{\mathbf{d}_i\}_0^{n-1}$ *be a set of non-zero* $\mathbf{Q}$*-conjugate vectors; for any* $\mathbf{x}_0 \in \mathbf{R}^n$, *the sequence* $\{\mathbf{x}_k\}$ *generated according to*

1. $\mathbf{g}_k = \mathbf{Q}\mathbf{x}_k - \mathbf{b}$

2. $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$, where $\alpha_k = \dfrac{-\mathbf{g}_k^T \mathbf{d}_k}{\mathbf{d}_k^T \mathbf{Q} \mathbf{d}_k}$

converges to the unique solution $\mathbf{x}^*$ *of* $\mathbf{Q}\mathbf{x}=\mathbf{b}$ *after at most n steps; ie., for some* $i \leq n$, $\mathbf{x}_i = \mathbf{x}^*$, *which is also the unique minimum of the quadratic,* $\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \mathbf{b}^T\mathbf{x}$.

(1.5) THEOREM. (Conjugate Gradients) *Each of the conjugate gradient methods described above is a conjugate direction method. If it does not terminate at* $\mathbf{x}_k$, *then:*

1. $\text{span}\,\{\mathbf{g}_0, \cdots, \mathbf{g}_k\} = \text{span}\,\{\mathbf{g}_0, \mathbf{Q}\mathbf{g}_0, \cdots, \mathbf{Q}^k\mathbf{g}_0\}$

2. $\text{span}\,\{\mathbf{d}_0, \cdots, \mathbf{d}_k\} = \text{span}\,\{\mathbf{g}_0, \mathbf{Q}\mathbf{g}_0, \cdots, \mathbf{Q}^k\mathbf{g}_0\}$

3. $\forall i \leq k-1,\ \mathbf{d}_k^T \mathbf{Q} \mathbf{d}_i = 0$

4. $\alpha_k = \dfrac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{d}_k^T \mathbf{Q} \mathbf{d}_k}$

5. $\beta_k = \dfrac{\mathbf{g}_{k+1}^T \mathbf{g}_{k+1}}{\mathbf{g}_k^T \mathbf{g}_k}$.

When extending the theory to non-quadratic functions, f, we no longer have a direct formula for α_k at any step; therefore, we perform the line search

$$\min_{\alpha>0} f(\mathbf{x}_k + \alpha \mathbf{d}_k),$$

setting α_k equal to the minimizer. This is often the most costly step in the algorithm. We also lose the guaranteed n-step termination from Theorem 1.4. Thus we can continue on finding new search directions, $\mathbf{d}_k$, and stop when some criterion is satisfied. However, since Q-conjugacy of the direction vectors is dependent upon

$$\mathbf{d}_0 = -\nabla f(\mathbf{x}_0),$$

another modification ensuring convergence is to restart after every $r \geq n$ steps by setting

$$\mathbf{d}_{mr} = -\nabla f(\mathbf{x}_{mr})$$

for every integer m. This means that every r-th step will be a steepest descent step, and thus we have global convergence. Every r-th step will decrease the function value, and all of the intermediate steps will be designed not to increase the functional value, ie.,

$$\forall\, k,\, f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k).$$

In what follows, we will assume that restart does occur for some $r \geq n$ unless specifically stated otherwise.

Chapter 2
N-STEP QUADRATIC CONVERGENCE

The first rigorous proofs of convergence rates for conjugate gradient algorithms were those of Powell [18] and Broyden et al. [2], in the early 1970's. These papers contained proofs of superlinear convergence for certain algorithms. However, they didn't make use of the following quadratic termination property.

(QTP) The iterations will terminate in $\le n$ steps if applied to a convex quadratic function for any starting point.

Most experts believed that these methods should also converge "very fast" for sufficiently smooth non-quadratic functions, as they can be approximated arbitrarily closely by quadratics in a neighborhood of the minimum. So we would expect to find a bound of the form:

$$\frac{\| \mathbf{x}_{i+n} - \mathbf{x}^* \|}{\| \mathbf{x}_i - \mathbf{x}^* \|^2} < \infty,$$

where $\mathbf{x}^*$ is the minimizer of the function. We will refer to methods which satisfy such a bound as being "n-step" quadratically convergent. We will present proofs for this type of bound for all of the methods, and show that even when the line search is approximated, this convergence can be retained.

2.1. Exact Line Search

Many initial attempts at proving n-step quadratic convergence were incorrect. Daniel [4, 5, 6] attempted a proof for his own version of the algorithm in 1969; the proof contained several errors. Polyak [17] also attempted a proof for the (PR) algorithm in 1969 with no success. Finally, in 1974, after an initial attempt in his dissertation, Cohen [3] provided a proof of this type of bound for the three methods (D), (PR) and (FR). The assumption of restart discussed in Chapter 1 was used in his proof of this result. We present Cohen's proof and extend it to include the (HS) method.

2.1.1. Cohen's Proof for FR, PR and D

We consider the following minimization problem (MP).

$$\textit{minimize } f(\mathbf{x})$$

where:

(AS1) $f \in C^3(\mathbf{R}^n)$

(AS2) $\exists\, 0 < m < M < \infty \text{ s.t. } \forall\, x, y \in \mathbf{R}^n,\ m\,\mathbf{y}^T\mathbf{y} \le \mathbf{y}^T\mathbf{H}(\mathbf{x})\mathbf{y} \le M\,\mathbf{y}^T\mathbf{y}.$

The algorithm with restart can be stated as follows.

(2.1) Basic Conjugate Gradient Algorithm with Restart

Given arbitrary $\mathbf{x}_0 \in \mathbf{R}^n$:

1. $\mathbf{g}_0 = \nabla f(\mathbf{x}_0)$
2. $\mathbf{d}_0 = -\mathbf{g}_0$
3. for $k = 0$ to $n-1$ do

   i. $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$

   where $\alpha_k = \arg\min_{\alpha \ge 0} f(\mathbf{x}_k + \alpha \mathbf{d}_k)$

   ii. $\mathbf{g}_{k+1} = \nabla f(\mathbf{x}_{k+1})$; check stopping criterion

iii. $\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k$

where β_k is chosen to be 0 if $k = n-1$, otherwise the appropriate method-dependent formula is used.

4. if convergence has not been reached:

i. $\mathbf{x}_0 \leftarrow \mathbf{x}_n$

ii. $\mathbf{g}_0 \leftarrow \mathbf{g}_n$

iii. $\mathbf{d}_0 \leftarrow \mathbf{d}_n$

iv. go to 3.

This algorithm will be denoted $\mathbf{x}_{k+1} = \Psi_f(\mathbf{x}_k, \mathbf{d}_k)$. For notational convenience, we will usually suppress the second argument, usually writing $\mathbf{x}_{k+1} = \Psi_f(\mathbf{x}_k)$, where the dependence on $\mathbf{d}_k$ is understood. We consider the following theorem as stated and proved by Cohen [3]:

(2.2) THEOREM. *Suppose that the conjugate gradient method (PR), (FR) or (D) is used to solve the minimization problem (MP), where we reinitialize every $r \geq n$ steps; then $\exists c \in \mathbf{R}$ s.t.*

$$\limsup_{k \to \infty} \frac{\| \mathbf{x}_{kr+n} - \mathbf{x}^* \|}{\| \mathbf{x}_{kr} - \mathbf{x}^* \|^2} \leq c < \infty. \qquad (1)$$

The main idea behind Cohen's proof is to compare the conjugate gradient method to Newton's method. Newton's method, which we will denote $\mathbf{x}_{k+1} = \Phi(\mathbf{x}_k)$, is, for a function f,

$$\Phi(\mathbf{x}_k) = \mathbf{x}_k - \mathbf{H}(\mathbf{x}_k)^{-1} \nabla f(\mathbf{x}_k).$$

Newton's method converges locally quadratically to the root $\mathbf{x}^*$ of $\nabla f(\mathbf{x}) = 0$, for the set of functions satisfying (AS1) with H(x) invertible. Therefore, if we assume (AS1) and (AS2), we have

$$\| \Phi(\mathbf{x}) - \mathbf{x}^* \| = O(\| \mathbf{x} - \mathbf{x}^* \|^2). \tag{2}$$

PROOF OF THEOREM 2.2: The proof will be completed in two steps. For more detail, see [3].

STEP 1: At each point $\mathbf{x}_{kr}$, define the quadratic:

$$\hat{f}_{kr}(\mathbf{x}) = f(\mathbf{x}_{kr}) + \nabla f(\mathbf{x}_{kr})^T(\mathbf{x}-\mathbf{x}_{kr}) + \frac{1}{2}(\mathbf{x}-\mathbf{x}_{kr})^T \mathbf{H}_{kr}(\mathbf{x}-\mathbf{x}_{kr}),$$

where $\mathbf{H}_{kr} = \mathbf{H}(\mathbf{x}_{kr})$. Let

$$\mathbf{x}_{kr}^0 = \mathbf{x}_{kr}; \quad \mathbf{x}_{kr}^1 = \Psi_{\hat{f}}(\mathbf{x}_{kr}^0); \quad . \ . \ . \quad ; \quad \mathbf{x}_{kr}^n = \Psi_{\hat{f}}(\mathbf{x}_{kr}^{n-1}),$$

where $\Psi_{\hat{f}}$ is the conjugate gradient method applied to $\hat{f}_{kr}$ with $\alpha_{kr}^0 = \alpha_{kr}$, $\mathbf{d}_{kr}^0 = \mathbf{d}_{kr}$ and $\mathbf{g}_{kr}^0 = \mathbf{g}(\mathbf{x}_{kr})$. Then the conjugate gradient algorithm satisfies (1) if

$$\| \alpha_{kr+i}\mathbf{d}_{kr+i} - \alpha_{kr}^i \mathbf{d}_{kr}^i \| = O(\|\mathbf{x}_{kr} - \mathbf{x}^* \|^2) \tag{3}$$

for $i = 0, 1, ..., j(k)-1$, where $j(k)$ is the integer $\leq n$ s.t.

$$\mathbf{x}_{kr}^{j(k)} = \alpha(\hat{f}_{kr}) = \arg\min_{\mathbf{x}} \hat{f}_{kr}(\mathbf{x})$$

and

$$\mathbf{x}_{kr}^{j(k)-1} \neq \alpha(\hat{f}_{kr}).$$

PROOF OF STEP 1: First note that $j(k)$ exists because $\hat{f}_{kr}$ is quadratic with positive definite $\mathbf{H}(\mathbf{x}_{kr})$, and the algorithm will reach the minimum of such functions in $\leq n$ steps. Since $\forall\, k$, $f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k)$, we know that

$$f(\mathbf{x}_{k+1}) - f(\mathbf{x}^*) \leq f(\mathbf{x}_k) - f(\mathbf{x}^*).$$

Expanding in a Taylor series gives

$$f(\mathbf{x}_k) - f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^T(\mathbf{x}_k - \mathbf{x}^*) + (\mathbf{x}_k - \mathbf{x}^*)^T\mathbf{H}(\eta)(\mathbf{x}_k - \mathbf{x}^*)$$

$$= (\mathbf{x}_k - \mathbf{x}^*)^T\mathbf{H}(\eta)(\mathbf{x}_k - \mathbf{x}^*),$$

because $\nabla f(\mathbf{x}^*) = 0$, where $\eta = \mathbf{x}_k + \lambda(\mathbf{x}^* - \mathbf{x}_k)$ and $\lambda \in (0, 1)$. Using (AS2), we see that

$$m \| \mathbf{x}_k - \mathbf{x}^* \|^2 \le f(\mathbf{x}_k) - f(\mathbf{x}^*) \le M \| \mathbf{x}_k - \mathbf{x}^* \|^2.$$

Thus

$$m \| \mathbf{x}_{k+1} - \mathbf{x}^* \|^2 \le \| f(\mathbf{x}_{k+1}) - f(\mathbf{x}^*) \| \le \| f(\mathbf{x}_k) - f(\mathbf{x}^*) \| \le M \| \mathbf{x}_k - \mathbf{x}^* \|^2,$$

which in turn implies

$$\| \mathbf{x}_{k+1} - \mathbf{x}^* \| \le \sqrt{\frac{M}{m}} \| \mathbf{x}_k - \mathbf{x}^* \|.$$

Using the triangle inequality repeatedly yields

$$\| \mathbf{x}_{kr+n} - \mathbf{x}^* \| \le \left[\frac{M}{m}\right]^{\frac{(n-j(k))}{2}} \| \mathbf{x}_{kr+j(k)} - \mathbf{x}^* \|.$$

$$\le \left[\frac{M}{m}\right]^{\frac{n}{2}} \| \mathbf{x}_{kr+j(k)} - \mathbf{x}^* \|.$$

Thus to prove (1), it suffices to show

$$\| \mathbf{x}_{kr+j(k)} - \mathbf{x}^* \| = O(\| \mathbf{x}_{kr} - \mathbf{x}^* \|^2). \tag{4}$$

Now,

$$\| \mathbf{x}_{kr+j(k)} - \mathbf{x}^* \| \le \| \mathbf{x}_{kr+j(k)} - \alpha(\hat{f}_{kr}) \| + \| \alpha(\hat{f}_{kr}) - \mathbf{x}^* \|. \tag{5}$$

Note that:

$$\alpha(\hat{f}_{kr}) = \mathbf{x}_{kr} - (\mathbf{H}(\mathbf{x}_{kr}))^{-1}\nabla f(\mathbf{x}_{kr}).$$

This is just Newton's method applied to f at $\mathbf{x}_{kr}$. Thus:

$$\|\alpha(\hat{f}_{kr}) - \mathbf{x}^*\| = \|\Phi(\mathbf{x}_{kr}) - \mathbf{x}^*\| = O(\|\mathbf{x}_{kr} - \mathbf{x}^*\|^2).$$

Plugging this into (5) yields:

$$\|\mathbf{x}_{kr+j(k)} - \mathbf{x}^*\| \le O(\|\mathbf{x}_{kr} - \mathbf{x}^*\|^2) + \|\mathbf{x}_{kr+j(k)} - \alpha(\hat{f}_{kr})\|.$$

So we have reduced the problem to showing:

$$\|\mathbf{x}_{kr+j(k)} - \alpha(\hat{f}_{kr})\| = \|\mathbf{x}_{kr+j(k)} - \mathbf{x}_{kr}^{j(k)}\| = O(\|\mathbf{x}_{kr} - \mathbf{x}^*\|^2).$$

Since $\mathbf{x}_{kr} = \mathbf{x}_{kr}^0$, we have:

$$\begin{aligned}\|\mathbf{x}_{kr+j(k)} - \mathbf{x}_{kr}^{j(k)}\| &\le \|\sum_{i=0}^{j(k)-1} [(\mathbf{x}_{kr+i+1} - \mathbf{x}_{kr+i}) - (\mathbf{x}_{kr}^{i+1} - \mathbf{x}_{kr}^{i})]\| \\ &\le \|\sum_{i=0}^{j(k)-1} [(\alpha_{kr+i}\mathbf{d}_{kr+i} - \alpha_{kr}^{i}\mathbf{d}_{kr}^{i})\| \\ &\le \sum_{i=0}^{j(k)-1} \|\alpha_{kr+i}\mathbf{d}_{kr+i} - \alpha_{kr}^{i}\mathbf{d}_{kr}^{i}\|.\end{aligned}$$

But by (3), each of these terms is equivalent to $O(\|\mathbf{x}_{kr} - \mathbf{x}^*\|^2)$. Thus, (3) implies (1), and the proof of step 1 is complete.

STEP 2: It remains to show that (3) holds for the various conjugate gradient algorithms. We will use the fact that with reinitialization, these algorithms do converge to stationary points of f. The following lemmas are needed and hold for all values of $k \neq r$ given (AS1) and (AS2). Lemmas 2.7, 2.10, 2.11 and 2.12 below are the only lemmas actually cited in the proof of Step 2. However, each lemma builds upon the previous ones, and we will need to see this relationship more clearly when we extend the proof of the theorem to method HS. For proofs of these lemmas, see [3].

(2.3) LEMMA. *a*) $\mathbf{g}_k^T\mathbf{d}_k = -\mathbf{g}_k^T\mathbf{g}_k$

b) $\mathbf{g}_{k+1}^T\mathbf{d}_k = 0$

c) $\alpha_k = \dfrac{\mathbf{g}_k^T\mathbf{g}_k}{\mathbf{d}_k^T\hat{\mathbf{H}}_k\mathbf{d}_k}$ *where* $\hat{\mathbf{H}}_k = \int_0^1 \mathbf{H}(\mathbf{x}_k + \xi\alpha_k\mathbf{d}_k)d\xi.$

(2.4) LEMMA. *a*) $\| \mathbf{d}_{k+1} \|^2 = \| \mathbf{g}_{k+1} \|^2 + \beta_k^2 \| \mathbf{d}_k \|^2$

b) $\| \mathbf{g}_k \| \leq \| \mathbf{d}_k \|$

c) $\| \alpha_k \| = \dfrac{\| \mathbf{g}_k \|^2}{\mathbf{d}_k^T\hat{\mathbf{H}}_k\mathbf{d}_k} \leq \dfrac{\| \mathbf{g}_k \|^2}{m \| \mathbf{d}_k \|^2} \leq \dfrac{1}{m}$

d) $\| \mathbf{g}_{k+1} \| \leq \left[1 + \dfrac{M}{m}\right] \| \mathbf{d}_k \|.$

(2.5) LEMMA. (*PR*) $\beta_k = \dfrac{\mathbf{g}_{k+1}^T\hat{\mathbf{H}}_k\mathbf{d}_k}{\mathbf{d}_k^T\hat{\mathbf{H}}_k\mathbf{d}_k}$

(*FR*) $\beta_k = 1 + \dfrac{(\mathbf{g}_{k+1} + \mathbf{g}_k)^T\hat{\mathbf{H}}_k\mathbf{d}_k}{\mathbf{d}_k^T\hat{\mathbf{H}}_k\mathbf{d}_k}$

(*D*) $\beta_k = \dfrac{\mathbf{g}_{k+1}^T\mathbf{H}_{k+1}\mathbf{d}_k}{\mathbf{d}_k^T\mathbf{H}_{k+1}\mathbf{d}_k}.$

(2.6) LEMMA. (*PR*) and (*D*) $| \beta_k | \leq \left[1 + \dfrac{M}{m}\right]\dfrac{M}{m}$

(*FR*) $| \beta_k | \leq \left[1 + \dfrac{M}{m}\right]^2.$

(2.7) LEMMA. *a*) $\| \mathbf{g}_k \| \le M \| \mathbf{x}_k - \mathbf{x}^* \|$

$$b)\ \| \mathbf{d}_{k+1} \| \le 2 \left[1 + \frac{M}{m} \right]^2 \| \mathbf{d}_k \|.$$

It should be noted that these lemmas also hold for the conjugate gradient method operating on $\hat{f}_{kr}$ with $\mathbf{x}_k$, $\mathbf{x}_{k+1}$, $\mathbf{g}_k$, $\mathbf{d}_k$, α_k, β_k, ... replaced by $\mathbf{x}_{kr}^i$, $\mathbf{x}_{kr}^{i+1}$, $\mathbf{g}_{kr}^i$, $\mathbf{d}_{kr}^i$, α_{kr}^i, β_{kr}^i, ..., respectively. Also note that

$$\mathbf{H}_{kr}^i = \hat{\mathbf{H}}_{kr}^i = \mathbf{H}_{kr}$$

since $\nabla^2 \hat{f}_{kr}$ is constant.

PROOF OF STEP 2: We start with the following assertion;

(2.8) CLAIM. *Let* $\{\mathbf{x}_k\}$ *be the sequence produced when one of the three conjugate gradient methods is used to solve the minimization problem (MP). Then*

$$\mathbf{d}_k \to 0$$

and

$$\| \alpha_{kr+i} \mathbf{d}_{kr+i} - \alpha_{kr}^i \mathbf{d}_{kr}^i \| = O(\| \mathbf{d}_{kr} \|^2) \tag{6}$$

for $i = 0, 1, ..., j(k)-1$, *where,* $\forall\, k$, *the convergence variable* $\mathbf{d}_{kr}$ *is set to* $-\mathbf{g}_{kr}$.

Note that if this claim is true, then since, by Lemma 2.7a, $\| \mathbf{g}_{kr} \| \le M \| \mathbf{x}_{kr} - \mathbf{x}^* \|$, we will have proven (3). The proof is given in detail by Cohen in an appendix to [3].

PROOF OF CLAIM 2.8. By Lemma 2.7b, $\| \mathbf{d}_{k+1} \| \le 2 \left[1 + \frac{M}{m} \right]^2 \| \mathbf{d}_k \|$, so

$$\| \mathbf{d}_{kr+i} \| \le \left\{ 2 \left[1 + \frac{M}{m} \right]^2 \right\}^i \| \mathbf{d}_{kr} \|$$

for $i = 0, \ldots, j(k)-1$. Since this algorithm converges, Lemma 2.7a implies that $\mathbf{g}_k \to 0$; thus $\mathbf{d}_{kr} \to 0$ and by Lemma 2.7b, $\mathbf{d}_k \to 0$. To show (6), assume that k is a multiple of r. The proof is by induction, and requires one to show:

$$\| \alpha_{k+i} \mathbf{d}_{k+i} - \alpha_k^i \mathbf{d}_k^i \| = O(\| \mathbf{d}_k \|^2) \tag{7a}$$

$$\| \mathbf{d}_{k+i} - \mathbf{d}_k^i \| = O(\| \mathbf{d}_k \|^2) \tag{7b}$$

$$\| \mathbf{g}_{k+i} - \mathbf{g}_k^i \| = O(\| \mathbf{d}_k \|^2) \tag{7c}$$

for $i = 0, \ldots, j(\frac{k}{r})-1$. We use the fact that the algorithms converge, ie., $\mathbf{x}_k \to \mathbf{x}^*$, where $\nabla^2 f(\mathbf{x}^*) = 0$. We need the following lemmas from the appendix in Cohen's paper [3] to continue.

(2.9) LEMMA. *a*) $\| \mathbf{H}_{k+i} - \mathbf{H}_k \| = O(\| \mathbf{d}_k \|)$

b) $\| \hat{\mathbf{H}}_{k+i} - \mathbf{H}_k \| = O(\| \mathbf{d}_k \|)$.

(2.10) LEMMA. (*PR*) and (*D*) $\| \mathbf{d}_{k+i+1} - \mathbf{d}_k^{i+1} \| = O(\| \mathbf{d}_{k+i} - \mathbf{d}_k^i \|) + O(\| \mathbf{d}_k \|^2)$

$$+ O(\| \mathbf{g}_{k+i+1} - \mathbf{g}_k^{i+1} \|)$$

(*FR*) $\| \mathbf{d}_{k+i+1} - \mathbf{d}_k^{i+1} \| = O(\| \mathbf{d}_{k+i} - \mathbf{d}_k^i \|) + O(\| \mathbf{d}_k \|^2)$

$$+ O(\| \mathbf{g}_{k+i+1} - \mathbf{g}_k^{i+1} \|) + O(\| \mathbf{g}_{k+i} - \mathbf{g}_k^i \|).$$

(2.11) LEMMA. $\| \mathbf{g}_{k+i+1} - \mathbf{g}_k^{i+1} \| \le O(\| \mathbf{g}_{k+i} - \mathbf{g}_k^i \|) + O(\| \mathbf{d}_k \|^2)$

$$+ M \| \alpha_{k+i} \mathbf{d}_{k+i} - \alpha_k^i \mathbf{d}_k^i \|.$$

(2.12) LEMMA. $\| \alpha_{k+i} \mathbf{d}_{k+i} - \alpha_k^i \mathbf{d}_k^{i+1} \| = O(\| \mathbf{g}_{k+i} - \mathbf{g}_k^i \|) + O(\| \mathbf{d}_k \|^2)$

$$+ O(\| \mathbf{d}_{k+i} - \mathbf{d}_k^i \|).$$

Now we continue with the proof by induction. For the base step $i = 0$, we have

$$\| \alpha_k \mathbf{d}_k - \alpha_k^0 \mathbf{d}_k^0 \| = 0$$

$$\| \mathbf{d}_k - \mathbf{d}_k^0 \| = 0$$

$$\| \mathbf{g}_k - \mathbf{g}_k^0 \| = 0$$

because $k = rl$ for some integer l. Now assume that (7a), (7b) and (7c) hold for i; we must prove them for $i+1$. By Lemma 2.11 and the induction hypothesis,

$$\| \mathbf{g}_{k+i+1} - \mathbf{g}_k^{i+1} \| \leq O(\| \mathbf{g}_{k+i} - \mathbf{g}_k^i \|) + O(\| \mathbf{d}_k \|^2) + M \| \alpha_{k+i} \mathbf{d}_{k+i} - \alpha_k^i \mathbf{d}_k^i \|$$

$$= O(\| \mathbf{d}_k \|^2).$$

By Lemma 2.10, we have

$$\| \mathbf{d}_{k+i+1} - \mathbf{d}_k^{i+1} \| = O(\| \mathbf{d}_{k+i} - \mathbf{d}_k^i \|) + O(\| \mathbf{d}_k \|^2) + O(\| \mathbf{g}_{k+i+1} - \mathbf{g}_k^{i+1} \|)$$

and

$$\| \mathbf{d}_{k+i+1} - \mathbf{d}_k^{i+1} \| = O(\| \mathbf{d}_{k+i} - \mathbf{d}_k^i \|) + O(\| \mathbf{d}_k \|^2) + O(\| \mathbf{g}_{k+i+1} - \mathbf{g}_k^{i+1} \|)$$

$$+ O(\| \mathbf{g}_{k+i} - \mathbf{g}_k^i \|),$$

the first for (PR) and (D), the second for (FR). So, by induction, we have

$$\| \mathbf{d}_{k+i+1} - \mathbf{d}_k^{i+1} \| = O(\| \mathbf{d}_k \|^2)$$

and finally, by Lemma 2.12 and induction:

$$\| \alpha_{k+i+1} \mathbf{d}_{k+i+1} - \alpha_k^{i+1} \mathbf{d}_k^{i+1} \| = O(\| \mathbf{g}_{k+i+1} - \mathbf{g}_k^{i+1} \|) + O(\| \mathbf{d}_{k+i+1} - \mathbf{d}_k^{i+1} \|)$$

$$+ O(\| \mathbf{d}_k \|^2)$$

$$= O(\| \mathbf{d}_k \|^2).$$

////

2.1.2. Cohen's Theorem Extended for HS

The proof given above can be extended easily to include the HS method. Remember that the proof is completed in two steps. The first step is contingent upon known results from the quadratic case of the algorithm. The place where the specific formula for β comes into play is in the second step, in Lemma 2.5. The following addition to Lemma 2.5 will allow the proof of step 2 to be extended to HS.

(2.5*) LEMMA. (*HS*) $\beta_k = \dfrac{\mathbf{g}_{k+1}^T \hat{\mathbf{H}}_k \mathbf{d}_k}{\mathbf{d}_k^T \hat{\mathbf{H}}_k \mathbf{d}_k}$ *where* $\hat{\mathbf{H}}_k = \int_0^1 \mathbf{H}(\mathbf{x}_k + \xi \alpha_k \mathbf{d}_k) d\xi$.

PROOF OF LEMMA (2.5*): By definition of the HS method,

$$\beta_k = \frac{(\mathbf{g}_{k+1} - \mathbf{g}_k)^T \mathbf{g}_{k+1}}{(\mathbf{g}_{k+1} - \mathbf{g}_k)^T \mathbf{d}_k} = \frac{\mathbf{g}_{k+1}^T (\mathbf{g}_{k+1} - \mathbf{g}_k)}{\mathbf{d}_k^T (\mathbf{g}_{k+1} - \mathbf{g}_k)}.$$

By Taylor's Theorem,

$$\mathbf{g}_{k+1} = \mathbf{g}_k - \alpha_k \hat{\mathbf{H}}_k \mathbf{d}_k .$$

Therefore,

$$\beta_k = \frac{\mathbf{g}_{k+1}^T (-\alpha_k \hat{\mathbf{H}}_k \mathbf{d}_k)}{\mathbf{d}_k^T (-\alpha_k \hat{\mathbf{H}}_k \mathbf{d}_k)} = \frac{\mathbf{g}_{k+1}^T \hat{\mathbf{H}}_k \mathbf{d}_k}{\mathbf{d}_k^T \hat{\mathbf{H}}_k \mathbf{d}_k} \qquad ////$$

Note that the conclusion of this lemma for HS is identical to the conclusion for PR in Lemma 2.5. The rest of the proof of Step 2 for HS proceeds in the same manner as that for the PR method. We have now established n-step quadratic convergence for the conjugate gradient methods PR, FR, D and HS with exact line search, under conditions (AS1) and (AS2).

2.2. Inexact Line Search

The most time consuming step of any conjugate gradient algorithm is the line search performed to update **x**:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k, \text{ where } \alpha_k = \arg\min_{\alpha > 0} f(\mathbf{x}_k + \alpha \mathbf{d}_k).$$

Since this minimum is rarely attained computationally, the effects of inaccuracy here are of great concern. We will present an analysis which will answer the question, "How accurate do these line searches have to be in order to preserve n-step quadratic convergence?"

In 1976, Lenard presented a proof of n-step quadratic convergence for the conjugate gradient methods FR and HS with inaccurate line searches [14]. The conditions imposed on the function to be minimized are slightly different than those of Cohen. (AS1) is weakened to

(AS1W) $\quad f \in \mathbf{C}^2(\mathbf{R}^n)$;

(AS2) is retained as stated previously:

(AS2) $\quad \exists\, 0 < m < M < \infty \; s.t. \; \forall\, x, y \in \mathbf{R}^n,\; m\,\mathbf{y}^T\mathbf{y} \le \mathbf{y}^T\mathbf{H}(\mathbf{x})\mathbf{y} \le M\,\mathbf{y}^T\mathbf{y}$;

and a new condition on the Hessian is added:

(AS3) $\quad \exists\, L \in \mathbf{R} \; s.t. \; \forall\, \mathbf{x} \in \mathbf{R}^n \; \|\mathbf{H}(\mathbf{x}) - \mathbf{H}(\mathbf{x}^*)\| \le L\, \|\mathbf{x} - \mathbf{x}^*\|$.

This last condition is refered to as a second derivative Lipschitz condition. The algorithm proceeds from $\mathbf{x}_k$ to $\mathbf{x}_{k+1}$ as follows:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k,$$

where α_k approximates what we will henceforth call α_k^*:

$$\alpha_k^* = \arg\min_{\alpha > 0} f(\mathbf{x}_k + \alpha \mathbf{d}_k).$$

The method then updates

$$\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k$$

as before, where β_k is selected using either the FR or HS formula. We assume that the procedure is restarted every $r \geq n$ steps. Lenard [14] proves that given (AS1W), (AS2), (AS3) and certain conditions on the line search, the conjugate gradient methods FR and HS are n-step quadratically convergent. We present her proof in the next section and then extend the result to include the PR and D methods.

2.2.1. Lenard's Proof for FR and HS

The method of proof will be to compare n steps of the conjugate gradient method with inaccurate line searches applied to f, to n steps of the same method with exact line searches applied to the quadratic approximation

$$Q(\mathbf{x}) = f(\mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T \mathbf{H}(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*).$$

Note that

$$\| \mathbf{H}(\mathbf{x}) - \nabla_2 Q(\mathbf{x}) \| \leq L \| \mathbf{x} - \mathbf{x}^* \|.$$

We will use the following notation to distinguish between the method applied to the original function and the quadratic approximation. We set $\mathbf{x}_0 = \mathbf{x}_0^Q = \mathbf{x}_0^f$.

For the quadratic, Q, let

$$\mathbf{g}_k^Q = \nabla Q(\mathbf{x}_k^Q)$$

$$\mathbf{x}_{k+1}^Q = \mathbf{x}_k^Q + \alpha_k^Q \mathbf{d}_k^Q$$

$$\mathbf{d}_{k+1}^Q = -\mathbf{g}_{k+1}^Q + \beta_k^Q \mathbf{d}_k^Q$$

Note that $g_k^{Q^T} d_k^Q < 0$ and $g_{k+1}^{Q^T} d_k^Q = 0$ because Q is quadratic and the line searches are exact.

For the original function, f,

$$g_k^f = \nabla f\,(x_k^f)$$

$$x_{k+1}^f = x_k^f + \alpha_k^f d_k^f$$

$$d_{k+1}^f = -g_{k+1}^f + \beta_k^f d_k^f.$$

We will need the following lemmas from Lenard's papers **[13, 14]** to present the proof of n-step quadratic convergence.

(2.13) LEMMA. *Given (AS1W) and (AS2),* $\forall\; x, x' \in R^n$

$$m \,\| x - x' \| \le \| \nabla f\,(x) - \nabla f\,(x') \| \le M \,\| x - x' \|.$$

(2.14) LEMMA. *If* x^* *is the minimizer of* f *satisfying (AS1W) and (AS2), then*

$$\frac{\| \nabla f\,(x) \|^2}{2M} \le f\,(x) - f\,(x^*) \le \frac{\| \nabla f\,(x) \|^2}{2m}.$$

(2.15) LEMMA. *If* x^* *is the minimizer of* f *satisfying (AS1W) and (AS2), and if for two points* x, x', $f\,(x') < f\,(x)$, *then*

$$\| x' - x^* \| < \left[\frac{M}{m} \right]^{\frac{3}{2}} \| x - x^* \|.$$

(2.16) LEMMA. *Suppose that* $g_k^{f^T} d_k^f < 0$ *and* $\theta_k = \dfrac{g_{k+1}^{f\,T} d_k^f}{g_k^{f^T} d_k^f}$ *for* $0 \le \theta_k \le 1 - \eta$, *where* $0 < \eta < 1$. *Let* ω *be s.t.* $\| x_k^Q - x^* \| \le \omega$ *and* $\| x_k^f - x^* \| \le \omega$, *and suppose that* $\exists\, a > 0$ *s.t.* $\| d_k^Q \| \ge a \,\| g_k^Q \|$, *then*

$$\| \mathbf{d}^Q_{k+1} - \mathbf{d}^f_{k+1} \| \le B_1 \| \mathbf{x}^Q_k - \mathbf{x}^f_k \| + B_2 \| \mathbf{d}^Q_k - \mathbf{d}^f_k \| + B_3 \omega^2 + B_4 \theta_k \omega$$

for $B_i > 0$ depending only upon m, M and L.

Using these lemmas it can be proven that the conjugate gradient methods HS and FR are n-step quadratically convergent with inaccurate line searches, provided that the line searches become more accurate [14]:

(2.17) THEOREM. *If $\theta_k \le K \| \nabla f(\mathbf{x}_0) \|$, $k = 0, 1, \ldots, n-1$, with $0 \le \theta_k \le 1 - \eta$, $0 < \eta < 1$ for $\mathbf{x}_0$ in a neighborhood of the minimizer, $\mathbf{x}^*$, of f, where K and η are positive constants independent of k; and if the procedure satisfies:*

i) *A quadratic problem is solved in at most n steps with exact line search.*

ii) *$\exists a > 0$ depending only upon m and M s.t.*

$$\| d^Q_k \| \ge a \| \mathbf{g}^Q_k \|.$$

iii) *Under the hypotheses of Lemma 2.16,*

$$\| \mathbf{d}^Q_{k+1} - \mathbf{d}^f_{k+1} \| \le C_1 \| \mathbf{x}^Q_k - \mathbf{x}^f_k \| + C_2 \| \mathbf{d}^Q_k - \mathbf{d}^f_k \| + C_3 \omega^2 + C_4 \theta_k \omega,$$

for positive constants C_i depending only upon m, M and L.

Then:

$$\| \mathbf{x}^f_n - \mathbf{x}^* \| \le \delta \| \mathbf{x}_0 - \mathbf{x}^* \|^2$$

where δ is a constant depending only upon m, M, L and K.

First we prove the theorem in general, and then we check the various hypotheses for methods FR and HS.

PROOF OF THEOREM (2.17): Let $w = (\mathbf{x}_0^T, \mathbf{d}_0^T)^T$ denote the starting point with $\mathbf{d}_0 = -\mathbf{g}_0$. Consider two other starting points, u and v with $u = (\mathbf{x}_u^T, \mathbf{d}_u^T)^T$ and $v = (\mathbf{x}_v^T, \mathbf{d}_v^T)^T$, where

$$\mathbf{d}_u = -\mathbf{g}_u = -\nabla f(\mathbf{x}_u),$$

$$\mathbf{d}_v = -\mathbf{g}_v = -\nabla f(\mathbf{x}_v)$$

and

$$\| \mathbf{x}_u - \mathbf{x}^* \| \le \| \mathbf{x}_0 - \mathbf{x}^* \|,$$

$$\| \mathbf{x}_v - \mathbf{x}^* \| \le \| \mathbf{x}_0 - \mathbf{x}^* \|.$$

Two sequences of points can be generated:

$$u_{k+1} = \Psi_f(u_k^f) \text{ and } v_{k+1}^Q = \Psi_Q(u_k^Q),$$

with Ψ defined as before, as one step of the given conjugate gradient method applied to the function noted in the subscript, letting $u_0^f = u$ and $v_0^Q = v$. To distinguish the two parts of the 2-n vector, we use the following notation:

$$u_k^f = (\mathbf{x}_k^f(u)^{\mathrm{T}}, \mathbf{d}_k^f(u)^{\mathrm{T}})^{\mathrm{T}} \text{ and } v_k^Q = (\mathbf{x}_k^Q(v)^{\mathrm{T}}, \mathbf{d}_k^Q(u)^{\mathrm{T}})^{\mathrm{T}}.$$

If $A = \left[\frac{M}{m}\right]^{\frac{3}{2}}$, $M \ge m$ implies $A \ge 1$. Lemma 2.16 and iii) applied to u and v give:

$$\| \mathbf{x}_f^f(u) - \mathbf{x}_f^Q(v) \| + \| \mathbf{d}_f^f(u) - \mathbf{d}_f^Q(v) \| \le$$

$$D_1 \| \mathbf{x}_u - \mathbf{x}_v \| + D_1 \| \mathbf{d}_u - \mathbf{d}_v \| + A^2 D_3 \| \mathbf{x}_0 - \mathbf{x}^* \|^2 + A D_4 \theta_0 \| \mathbf{x}_0 - \mathbf{x}^* \| \qquad (8)$$

where $D_i = B_i + C_i$. From the hypothesis and Lemma 2.13, we have:

$$\theta_0 \le KM \| \mathbf{x}_0 - \mathbf{x}^* \|. \qquad (9)$$

So,

$$\| \mathbf{x}_f^f(u) - \mathbf{x}_f^Q(v) \| + \| \mathbf{d}_f^f(u) - \mathbf{d}_f^Q(v) \| \le$$

$$\leq R(\| \mathbf{x}_u - \mathbf{x}_v \| + \| \mathbf{d}_u - \mathbf{d}_v \|) + S \| \mathbf{x}_0 - \mathbf{x}^* \|^2, \tag{10}$$

where $R = D_1 + D_2$, and $S = A^2 D_3 + AKMD_4$. We show by induction that

$$\| \mathbf{x}_{k+1}^f(u) - \mathbf{x}_{k+1}^g(v) \| + \| \mathbf{d}_{k+1}^f(u) - \mathbf{d}_{k+1}^g(v) \| \leq$$

$$\leq R^{k+1}(\| \mathbf{x}_u - \mathbf{x}_v \| + \| \mathbf{d}_u - \mathbf{d}_v \|) + S \| \mathbf{x}_0 - \mathbf{x}^* \|^2 (1 + R + \ldots + R^k). \tag{11}$$

When $k = 0$, (11) is true by (10). Assume (11) for k. We must show that (11) is also true for $k+1$.

By Lemma 2.15, we have

$$\| \mathbf{x}_k^f(u) - \mathbf{x}^* \| \leq A \| \mathbf{x}_u - \mathbf{x}^* \| \leq A \| \mathbf{x}_0 - \mathbf{x}^* \|$$

$$\| \mathbf{x}_k^g(v) - \mathbf{x}^* \| \leq A \| \mathbf{x}_v - \mathbf{x}^* \| \leq A \| \mathbf{x}_0 - \mathbf{x}^* \|.$$

Then, by ii), iii) and Lemma 2.16:

$$\| \mathbf{x}_{k+1}^f(u) - \mathbf{x}_{k+1}^g(v) \| + \| \mathbf{d}_{k+1}^f(u) - \mathbf{d}_{k+1}^g(v) \| \leq$$

$$\leq R(\| \mathbf{x}_u - \mathbf{x}_v \| + \| \mathbf{d}_u - \mathbf{d}_v \|) + S \| \mathbf{x}_0 - \mathbf{x}^* \|^2.$$

By induction:

$$\| \mathbf{x}_{k+1}^f(u) - \mathbf{x}_{k+1}^g(v) \| + \| \mathbf{d}_{k+1}^f(u) - \mathbf{d}_{k+1}^g(v) \| \leq$$

$$\leq S \| \mathbf{x}_0 - \mathbf{x}^* \|^2 + R\{S \| \mathbf{x}_0 - \mathbf{x}^* \|^2 \left[1 + R + \ldots + R^{k-1}\right]\}$$

$$+ R\{R^k(\| \mathbf{x}_u - \mathbf{x}_v \| + \| \mathbf{d}_u - \mathbf{d}_v \|)\}$$

$$= S \| \mathbf{x}_0 - \mathbf{x}^* \|^2 (1 + R + \ldots + R^k) + R^{k+1}(\| \mathbf{x}_u - \mathbf{x}_v \| + \| \mathbf{d}_u - \mathbf{d}_v \|),$$

which is exactly (11) for $k+1$! Thus (11) hold for all k. Now let $k = n-1$, $u = v = w$; using i):

$$\| \mathbf{x}_n^f(w) - \mathbf{x}^* \| \leq R^n(\| \mathbf{g}_0^g - \mathbf{g}_0^f \|) + S \| \mathbf{x}_0 - \mathbf{x}^* \|^2 (1 + R + \ldots + R^{n-1}), \tag{12}$$

and, finally,

$$\| g_0^Q - g_0^f \| = \| \nabla^2 Q (x_0 - x^*) - \tilde{H}_f (x_0 - x^*) \|$$
$$\leq \| \nabla^2 Q - \tilde{H}_f \| \| x_0 - x^* \|$$

where $\tilde{H}_f = \nabla^2 f(\tilde{x})$ for $\tilde{x} = \lambda x_0 + (1-\lambda)x^*$ with $\lambda \in [0,1]$. Using the Lipschitz condition:

$$\| g_0^Q - g_0^f \| \leq L \| x_0 - x^* \|^2.$$

Thus (12) becomes:

$$\| x_n^f - x^* \| \leq \delta \| x_0 - x^* \|^2$$

for x_0 in a neighborhood of x^*, where $\delta = S(1 + R +, \ldots, + R^{n-1}) + R^n L.$ ////

Now that the theorem is proven, we can use it to show that the FR and HS methods are n-step quadratically convergent. For further detail, see [14]. We check the hypotheses for each method.

i) This hypothesis is satisfied by both since they each will solve the quadratic problem with exact line searches in $\leq n$ steps.

ii) This is also met by both methods. Since $d_{k+1}^Q = -g_{k+1}^Q + \alpha_k^Q d_k^Q$, we have

$$\| d_{k+1}^Q \|^2 \geq \| g_{k+1}^Q \|^2 + \| \alpha_k^Q \|^2 \| d_k^Q \|^2,$$

from which it follows that

$$\| d_{k+1}^Q \|^2 \geq \| g_{k+1}^Q \|^2.$$

iii) We need the condition in Lemma 2.16 to be satisfied, namely $g_k^{f\,T} d_k^f < 0$. This condition guarantees that each search direction is a descent direction. For FR, we need $\theta_k \geq 0$ which is one of the conditions of Theorem 2.17. For HS, we need the added condition,

$$\theta_k \mathbf{g}_{k+1}^{f\,\mathrm{T}} \mathbf{d}_k^f \le \| \mathbf{g}_{k+1}^f \|^2.$$

Next we verify iii) for each method. For both methods, we have

$$\| \mathbf{d}_{k+1}^{Q} - \mathbf{d}_{k+1}^{f} \| \le \| \mathbf{g}_{k+1}^{Q} - \mathbf{g}_{k+1}^{f} \| + \| \beta_k^{Q} \mathbf{d}_k^{Q} - \beta_k^{f} \mathbf{d}_k^{f} \|$$

$$\le \| \mathbf{g}_{k+1}^{Q} - \nabla Q(\mathbf{x}_{k+1}^{f}) \| + \| \nabla Q(\mathbf{x}_{k+1}^{f}) - \mathbf{g}_{k+1}^{f} \|$$

$$+ | \beta_k^{Q} | \| \mathbf{d}_k^{Q} - \mathbf{d}_k^{f} \| + | \beta_k^{Q} - \beta_k^{f} | \| \mathbf{d}_k^{f} \|.$$

Applying the Lipschitz condition and Lemmas 2.13 and 2.16, we get

$$\| \mathbf{d}_{k+1}^{Q} - \mathbf{d}_{k+1}^{f} \| \le M \| \mathbf{x}_{k+1}^{Q} - \mathbf{x}_{k+1}^{f} \| + L \left[\frac{M}{m} \right]^3 \| \mathbf{x}_0 - \mathbf{x}^* \|^2$$

$$+ | \beta_k^{Q} | \| \mathbf{d}_k^{Q} - \mathbf{d}_k^{f} \| + | \beta_k^{Q} - \beta_k^{f} | \| \mathbf{d}_k^{f} \|. \qquad (13)$$

Now, we will consider each of the terms in (13) individually.

We can bound $\| \mathbf{x}_{k+1}^{Q} - \mathbf{x}_{k+1}^{f} \|$ using Lemma 2.16 to get:

$$\| \mathbf{d}_{k+1}^{Q} - \mathbf{d}_{k+1}^{f} \| \le M (B_1 \| \mathbf{x}_k^{Q} - \mathbf{x}_k^{f} \| + B_2 \| \mathbf{d}_k^{Q} - \mathbf{d}_k^{f} \| + B_3 \omega^2 + B_4 \theta_k \omega)$$

$$+ L \left[\frac{M}{m} \right]^3 \| \mathbf{x}_0 - \mathbf{x}^* \|^2 + | \beta_k^{Q} | \| \mathbf{d}_k^{Q} - \mathbf{d}_k^{f} \| + | \beta_k^{Q} - \beta_k^{f} | \| \mathbf{d}_k^{f} \|. \qquad (14)$$

Next, to see that $| \beta_k^{Q} |$ is bounded, first note that,

$$\beta_k^{Q} = \frac{\mathbf{g}_{k+1}^{Q\,\mathrm{T}} \nabla^2 Q \, \mathbf{d}_k^{Q}}{\mathbf{d}_k^{Q\,\mathrm{T}} \nabla^2 Q \, \mathbf{d}_k^{Q}}.$$

This follows for both methods from the definition of β_k because Q is quadratic. Using ii) and Lemma 2.14 we have,

$$|\beta_k^{\mathcal{Q}}| \le \frac{M\,\|g_{k+1}^{\mathcal{Q}}\|\,\|g_k^{\mathcal{Q}}\|}{m\,\|d_k^{\mathcal{Q}}\|\,\|g_k^{\mathcal{Q}}\|} \le \frac{1}{a}\left[\frac{M}{m}\right]^{\frac{3}{2}}. \tag{15}$$

Now we consider $|\beta_k^{\mathcal{Q}} - \beta_k^{\mathcal{L}}|\,\|d_k^{\mathcal{L}}\|$ for methods FR and HS.

The formula for β_k in the HS method yields

$$\beta_k^{\mathcal{Q}} - \beta_k^{\mathcal{L}} = \left[\frac{g_{k+1}^{\mathcal{Q}\,T}(g_{k+1}^{\mathcal{Q}} - g_k^{\mathcal{Q}})}{d_k^{\mathcal{Q}T}(g_{k+1}^{\mathcal{Q}} - g_k^{\mathcal{Q}})} - \frac{g_{k+1}^{\mathcal{L}\,T}(g_{k+1}^{\mathcal{L}} - g_k^{\mathcal{L}})}{d_k^{\mathcal{L}T}(g_{k+1}^{\mathcal{L}} - g_k^{\mathcal{L}})}\right]$$

Further,

$$\beta_k^{\mathcal{Q}} - \beta_k^{\mathcal{L}} = \left[\frac{g_{k+1}^{\mathcal{Q}\,T}H_{k+1}^{\mathcal{Q}}d_k^{\mathcal{Q}}}{d_k^{\mathcal{Q}T}H_{k+1}^{\mathcal{Q}}d_k^{\mathcal{Q}}} - \frac{g_{k+1}^{\mathcal{L}\,T}\tilde{H}_{k+1}^{\mathcal{L}}d_k^{\mathcal{L}}}{d_k^{\mathcal{L}T}\tilde{H}_{k+1}^{\mathcal{L}}d_k^{\mathcal{L}}}\right]$$

where $\tilde{H}_{k+1}^{\mathcal{L}} = \nabla^2 f(\tilde{x})$, $\tilde{x} = \lambda x_k^{\mathcal{L}} + (1-\lambda)x_{k+1}^{\mathcal{L}}$ for some λ, $0 \le \lambda \le 1$. Norming both sides and clearing denominators yields:

$$|\beta_k^{\mathcal{Q}} - \beta_k^{\mathcal{L}}|\,(m^2\|d_k^{\mathcal{Q}}\|^2\|d_k^{\mathcal{L}}\|^2) \le$$

$$\le \|g_{k+1}^{\mathcal{Q}\,T}H_{k+1}^{\mathcal{Q}}d_k^{\mathcal{Q}}d_k^{\mathcal{L}T}\tilde{H}_{k+1}^{\mathcal{L}}d_k^{\mathcal{L}} - g_{k+1}^{\mathcal{L}\,T}\tilde{H}_{k+1}^{\mathcal{L}}d_k^{\mathcal{L}}d_k^{\mathcal{Q}T}H_{k+1}^{\mathcal{Q}}d_k^{\mathcal{Q}}\|$$

$$\le \|g_{k+1}^{\mathcal{Q}\,T}H_{k+1}^{\mathcal{Q}}d_k^{\mathcal{Q}}(d_k^{\mathcal{L}} - d_k^{\mathcal{Q}})^T\tilde{H}_{k+1}^{\mathcal{L}}d_k^{\mathcal{L}}\| + \|g_{k+1}^{\mathcal{Q}\,T}H_{k+1}^{\mathcal{Q}}(d_k^{\mathcal{Q}} - d_k^{\mathcal{L}})d_k^{\mathcal{Q}T}\tilde{H}_{k+1}^{\mathcal{L}}d_k^{\mathcal{L}}\|$$

$$+ \|g_{k+1}^{\mathcal{Q}\,T}H_{k+1}^{\mathcal{Q}}d_k^{\mathcal{L}}d_k^{\mathcal{Q}T}\tilde{H}_{k+1}^{\mathcal{L}}(d_k^{\mathcal{L}} - d_k^{\mathcal{Q}})\| + \|(g_{k+1}^{\mathcal{Q}} - g_{k+1}^{\mathcal{L}})^T H_{k+1}^{\mathcal{Q}}d_k^{\mathcal{L}}d_k^{\mathcal{Q}T}\tilde{H}_{k+1}^{\mathcal{L}}d_k^{\mathcal{Q}}\|$$

$$+ \|g_{k+1}^{\mathcal{L}\,T}(H_{k+1}^{\mathcal{Q}} - \tilde{H}_{k+1}^{\mathcal{L}})d_k^{\mathcal{L}}d_k^{\mathcal{Q}T}\tilde{H}_{k+1}^{\mathcal{L}}d_k^{\mathcal{Q}}\| + \|g_{k+1}^{\mathcal{L}\,T}\tilde{H}_{k+1}^{\mathcal{L}}d_k^{\mathcal{L}}d_k^{\mathcal{Q}T}(\tilde{H}_{k+1}^{\mathcal{L}} - H_{k+1}^{\mathcal{Q}})d_k^{\mathcal{Q}}\|$$

$$\le 3M^2\|g_{k+1}^{\mathcal{Q}}\|\,\|d_k^{\mathcal{Q}}\|\,\|d_k^{\mathcal{L}}\|\,\|d_k^{\mathcal{Q}} - d_k^{\mathcal{L}}\| + M^2\|d_k^{\mathcal{Q}}\|^2\|d_k^{\mathcal{L}}\|\,\|g_{k+1}^{\mathcal{Q}} - g_{k+1}^{\mathcal{L}}\|$$

$$+ 2M\,\|g_{k+1}^{\mathcal{L}}\|\,\|d_k^{\mathcal{L}}\|\,\|d_k^{\mathcal{Q}}\|^2\|H_{k+1}^{\mathcal{Q}} - \tilde{H}_{k+1}^{\mathcal{L}}\|.$$

This implies that

$$|\beta_k^{Q} - \beta_k^{f}|\,\|\mathbf{d}_k^{f}\| \le 3\left[\frac{M}{m}\right]^2 \frac{\|\mathbf{g}_{k+1}^{Q}\|}{\|\mathbf{d}_k^{Q}\|}\,\|\mathbf{d}_k^{Q} - \mathbf{d}_k^{f}\|$$

$$+ \left[\frac{M}{m}\right]^2 \|\mathbf{g}_{k+1}^{Q} - \mathbf{g}_{k+1}^{f}\| + \frac{2M}{m^2}\,\|\mathbf{g}_{k+1}^{f}\|\,\|\mathbf{H}_{k+1}^{Q} - \tilde{\mathbf{H}}_{k+1}^{f}\|. \quad (16)$$

Using the Lipschitz condition, Lemma 2.13 and Lemma 2.14 gives:

$$\|\mathbf{g}_{k+1}^{Q} - \mathbf{g}_{k+1}^{f}\| \le M\,\|\mathbf{x}_{k+1}^{Q} - \mathbf{x}_{k+1}^{f}\| + L\left[\frac{M}{m}\right]^3 \|\mathbf{x}_0 - \mathbf{x}^*\|^2. \quad (17)$$

By Lemma 2.14 and ii),

$$\frac{\|\mathbf{g}_{k+1}^{Q}\|}{\|\mathbf{d}_k^{Q}\|} = \frac{\|\mathbf{g}_{k+1}^{Q}\|\,\|\mathbf{g}_k^{Q}\|}{\|\mathbf{g}_k^{Q}\|\,\|\mathbf{d}_k^{Q}\|} \le \frac{1}{a}\left[\frac{M}{m}\right]^{\frac{1}{2}}. \quad (18)$$

By Lemmas 2.13 and 2.15,

$$\|\mathbf{g}_{k+1}^{f}\| \le M\,\|\mathbf{x}_{k+1}^{f} - \mathbf{x}^*\| \le M\left[\frac{M}{m}\right]^{\frac{3}{2}} \|\mathbf{x}_0 - \mathbf{x}^*\|. \quad (19)$$

Finally, by the Lipschitz condition and Lemma 2.15,

$$\|\mathbf{H}_{k+1}^{Q} - \tilde{\mathbf{H}}_{k+1}^{f}\| \le L\,\|\tilde{\mathbf{x}} - \mathbf{x}^*\| \le L(\|\mathbf{x}_k^{f} - \mathbf{x}^*\| + \|\mathbf{x}_{k+1}^{f} - \mathbf{x}^*\|)$$

$$\le L\left[1 + \left[\frac{M}{m}\right]^{\frac{3}{2}}\right]\|\mathbf{x}_0 - \mathbf{x}^*\|. \quad (20)$$

Thus, substituting (17), (18), (19), and (20) into (16), gives:

$$|\beta_k^{Q} - \beta_k^{f}|\,\|\mathbf{d}_k^{f}\| \le 3\left[\frac{M}{m}\right]^{\frac{5}{2}} \frac{1}{a}\,\|\mathbf{d}_k^{Q} - \mathbf{d}_k^{f}\| + \frac{M^3}{m^2}\,\|\mathbf{x}_{k+1}^{Q} - \mathbf{x}_{k+1}^{f}\|$$

$$+ L \left[\frac{M}{m} \right]^5 \| \mathbf{x}_0 - \mathbf{x}^* \|^2 + \frac{2ML}{m^2} \left[\frac{M}{m} \right]^{\frac{3}{2}} \left[1 + \left[\frac{M}{m} \right]^{\frac{3}{2}} \right] \| \mathbf{x}_0 - \mathbf{x}^* \|^2. \qquad (21)$$

Combining (14), (15), (21) and Lemma 2.16, we get our result for HS.

We investigate $| \beta_k^Q - \beta_k^f | \, \| \mathbf{d}_k^f \|$ for FR. By definition,

$$\beta_k^Q - \beta_k^f = \left[\frac{\mathbf{g}_{k+1}^{Q\,\mathrm{T}} \mathbf{g}_{k+1}^Q}{\mathbf{g}_k^{Q\mathrm{T}} \mathbf{g}_k^Q} - \frac{\mathbf{g}_{k+1}^{f\,\mathrm{T}} \mathbf{g}_{k+1}^f}{\mathbf{g}_k^{f\mathrm{T}} \mathbf{g}_k^f} \right],$$

which gives

$$| \beta_k^Q - \beta_k^f | \, \| \mathbf{g}_k^Q \|^2 \| \mathbf{g}_k^f \|^2 \le$$

$$\le \| \mathbf{g}_{k+1}^{Q\,\mathrm{T}} \mathbf{g}_{k+1}^Q \mathbf{g}_k^{f\mathrm{T}} \mathbf{g}_k^f - \mathbf{g}_{k+1}^{f\,\mathrm{T}} \mathbf{g}_{k+1}^f \mathbf{g}_k^{Q\mathrm{T}} \mathbf{g}_k^Q \|$$

$$\le \| \mathbf{g}_{k+1}^{Q\,\mathrm{T}} \mathbf{g}_{k+1}^Q (\mathbf{g}_k^f - \mathbf{g}_k^Q)^\mathrm{T} \mathbf{g}_k^f \| + \| (\mathbf{g}_{k+1}^Q - \mathbf{g}_{k+1}^f)^\mathrm{T} \mathbf{g}_{k+1}^Q \mathbf{g}_k^{Q\mathrm{T}} \mathbf{g}_k^f \|$$

$$+ \| \mathbf{g}_{k+1}^{f\,\mathrm{T}} \mathbf{g}_{k+1}^Q \mathbf{g}_k^{Q\mathrm{T}} (\mathbf{g}_k^f - \mathbf{g}_k^Q) \| + \| \mathbf{g}_{k+1}^{f\,\mathrm{T}} (\mathbf{g}_{k+1}^Q - \mathbf{g}_{k+1}^f) \mathbf{g}_k^{Q\mathrm{T}} \mathbf{g}_k^Q \|$$

$$\le \| \mathbf{g}_{k+1}^Q \|^2 \| \mathbf{g}_k^f \| \| \mathbf{g}_k^f - \mathbf{g}_k^Q \| + \| \mathbf{g}_k^Q \| \| \mathbf{g}_{k+1}^Q \| \| \mathbf{g}_k^f \| \| \mathbf{g}_{k+1}^Q - \mathbf{g}_{k+1}^f \|$$

$$+ \| \mathbf{g}_{k+1}^f \| \| \mathbf{g}_{k+1}^Q \| \| \mathbf{g}_k^Q \| \| \mathbf{g}_k^f - \mathbf{g}_k^Q \| + \| \mathbf{g}_{k+1}^f \| \| \mathbf{g}_{k+1}^Q \|^2 \| \mathbf{g}_{k+1}^Q - \mathbf{g}_{k+1}^f \|$$

This implies that

$$| \beta_k^Q - \beta_k^f | \, \| \mathbf{g}_k^f \| \le \left[\frac{\| \mathbf{g}_{k+1}^Q \|^2}{\| \mathbf{g}_k^Q \|^2} + \frac{\| \mathbf{g}_{k+1}^Q \|}{\| \mathbf{g}_k^Q \|} \right]$$

$$\times \left[\frac{\| \mathbf{g}_{k+1}^Q \|}{\| \mathbf{g}_k^Q \|} \| \mathbf{g}_k^Q - \mathbf{g}_k^f \| + \| \mathbf{g}_{k+1}^Q - \mathbf{g}_{k+1}^f \| \right].$$

By Lemmas 2.13 and 2.14, and (17),

$$|\beta_k^Q - \beta_k^f| \, \| g_k^f \| \le 2 \left[\frac{M}{m} \right]^{\frac{1}{2}}$$

$$\times \left[M \left[\frac{M}{m} \right]^{\frac{1}{2}} \| x_k^Q - x_k^f \| + M \| x_{k+1}^Q - x_{k+1}^f \| + L \left[\frac{M}{m} \right]^3 \| x_0 - x^* \|^2 \right]. \qquad (22)$$

For both Q and f superscripts, we have

$$d_k = -g_k + \beta_{k-1} d_{k-1}.$$

So

$$\| d_k \|^2 = \| g_k \|^2 + \beta_{k-1}^2 \| d_{k-1} \|^2 - 2\beta_{k-1} g_k^T d_{k-1}$$

which implies that

$$\frac{\| d_k \|^2}{\| g_k \|^2} = 1 - 2\,\beta_{k-1} \frac{g_k^T d_{k-1}}{\| g_k \|^2} + \beta_{k-1}^2 \frac{\| d_{k-1} \|^2}{\| g_k \|^2}$$

$$= 1 + 2T \cos\phi + T^2,$$

where $T = -\dfrac{\| d_{k-1} \|}{\| g_k \|} \beta_{k-1}$ and $\cos\phi = \dfrac{g_k^T d_{k-1}}{\| g_k \|^2} \| d_{k-1} \|$. Thus we have,

$$\frac{\| d_k \|^2}{\| g_k \|^2} \le 1 + 2|T| \, |\cos\phi| + |T|^2 \le (1 + |T|)^2 \qquad (23)$$

or, using the definition of FR's β_{k-1},

$$\frac{\| d_k \|}{\| g_k \|} \le 1 + |T| = 1 + \frac{\| g_k \|^2}{\| g_{k-1} \|^2} \frac{\| d_{k-1} \|}{\| g_k \|}.$$

By Lemma 2.14

$$\frac{\| \mathbf{d}_k \|}{\| \mathbf{g}_k \|} \le 1 + |T| \le 1 + \left[\frac{M}{m}\right] \frac{\| \mathbf{d}_{k-1} \|}{\| \mathbf{g}_{k-1} \|}.$$

If $\mathbf{d}_0 = -\mathbf{g}_0$, then this recurrence relation gives:

$$\frac{\| \mathbf{d}_k \|}{\| \mathbf{g}_k \|} \le \sum_{j=0}^{k-1} \left[\frac{M}{m}\right]^{\frac{j}{2}} \le \sum_{j=0}^{n} \left[\frac{M}{m}\right]^{\frac{j}{2}} = R,$$

for $k = 0, 1, \ldots, n$. So we have,

$$\| \mathbf{d}_k \| \le R \, \| \mathbf{g}_k \|.$$

Using these inequalities (with superscripts) yields:

$$|\beta_k^{\varrho} - \beta_k^{\ell}| \frac{1}{R} \| \mathbf{d}_k^{\ell} \| \le \| \beta_k^{\varrho} - \beta_k^{\ell} \| \, \| \mathbf{g}_k^{\ell} \|$$

and thus, using (22),

$$|\beta_k^{\varrho} - \beta_k^{\ell}| \, \| \mathbf{d}_k^{\ell} \| \le 2R \left[\frac{M}{m}\right]^{\frac{1}{2}} \left\{ M \left[\frac{M}{m}\right]^{\frac{1}{2}} \| \mathbf{x}_k^{\varrho} - \mathbf{x}_k^{\ell} \| \right.$$

$$\left. + \, M \, \| \mathbf{x}_{k+1}^{\varrho} - \mathbf{x}_{k+1}^{\ell} \| + L \left[\frac{M}{m}\right]^{3} \| \mathbf{x}_0 - \mathbf{x}^* \| \right\}. \qquad (24)$$

Finally, combining (14), (15), (24) and Lemma 2.16 gives us our result for FR.

2.2.2. Lenard's Theorem Extended for PR and D

In order to show that the theorem holds for the PR and D methods, we must also check the hypotheses of the theorem for them. First note that i) and ii) are satisfied by PR and D using the same reasoning as for the FR and HS methods in the previous section. The real difference comes in iii). We first need to check that each direction is a descent direction. Here, as for HS, we need additional conditions on the line searches.

We ignore the superscript f momentarily. In general, to check that $\mathbf{g}_k^T\mathbf{d}_k < 0$, consider:

$$\begin{aligned} \mathbf{g}_k^T\mathbf{d}_k &= \mathbf{g}_k^T(-\mathbf{g}_k + \beta_{k-1}\mathbf{d}_{k-1}) \\ &= -\mathbf{g}_k^T\mathbf{g}_k + \beta_{k-1}\mathbf{g}_k^T\mathbf{d}_{k-1} \\ &= -\mathbf{g}_k^T\mathbf{g}_{k-1} + \beta_{k-1}\theta_{k-1}\mathbf{g}_{k-1}^T\mathbf{d}_{k-1}. \end{aligned} \tag{25}$$

We will use induction. For $k = 0$, we have

$$\mathbf{g}_0^T\mathbf{d}_0 = -\mathbf{g}_0^T\mathbf{g}_0 < 0$$

for both the PR and D methods. Now we assume (25) for k and prove it for $k+1$:

$$\mathbf{g}_{k+1}^T\mathbf{d}_{k+1} = -\mathbf{g}_{k+1}^T\mathbf{g}_{k+1} + \beta_k\theta_k\mathbf{g}_k^T\mathbf{d}_k.$$

For D, we have

$$\mathbf{g}_{k+1}^T\mathbf{d}_{k+1} = -\|\mathbf{g}_{k+1}\|^2 + \frac{\mathbf{g}_{k+1}^T\mathbf{H}_{k+1}\mathbf{d}_k}{\mathbf{d}_k^T\mathbf{H}_{k+1}\mathbf{d}_k}\theta_k\mathbf{g}_k^T\mathbf{d}_k.$$

It is easily verified that the added condition $\mathbf{g}_{k+1}^T\mathbf{H}_{k+1}\mathbf{d}_k\theta_k \geq 0$ guarantees descent.

For PR, we have

$$g_{k+1}^T d_{k+1} = - \|g_{k+1}\|^2 + \frac{(g_{k+1} - g_k)^T g_{k+1}}{g_k^T g_k} \theta_k g_k^T d_k .$$

Here, again, the added condition $\theta_k (g_{k+1} - g_k)^T g_{k+1} \geq \frac{\| g_{k+1} \|^2 \| g_k \|^2}{g_k^T d_k}$ guarantees descent. In her paper **[14]**, Lenard suggests the condition on the line search for PR, but does not show the proof of n-step quadratic convergence. Next we need to show that iii) holds for PR and D. Our proof will proceed in essentially the same fashion as Lenard's proof for HS and FR given above. First note that the inequalities (13) and (14) also hold for PR and D using the same reasoning as for FR and HS. It remains to bound $| \beta_k^Q - \beta_k^F | \, \| d_k^F \|$ appropriately for PR and D.

For D,

$$\beta_k^Q - \beta_k^F = \left[\frac{g_{k+1}^{Q\,T} H_{k+1}^Q d_k^Q}{d_k^{Q\,T} H_{k+1}^Q d_k^Q} - \frac{g_{k+1}^{F\,T} H_{k+1}^F d_k^F}{d_k^{F\,T} H_{k+1}^F d_k^F} \right].$$

Using (AS2) on H_{k+1}^Q and H_{k+1}^F and clearing denominators, we get

$$| \beta_k^Q - \beta_k^F | \, (m^2 \| d_k^Q \|^2 \| d_k^F \|^2) \leq$$

$$\leq \| g_{k+1}^Q{}^T H_{k+1}^Q d_k^Q d_k^F{}^T H_{k+1}^F d_k^F - g_{k+1}^F{}^T H_{k+1}^F d_k^F d_k^Q{}^T H_{k+1}^Q d_k^Q \|$$

$$\leq \| g_{k+1}^Q{}^T H_{k+1}^Q d_k^Q (d_k^F - d_k^Q)^T H_{k+1}^F d_k^F \| + \| g_{k+1}^Q{}^T H_{k+1}^Q (d_k^Q - d_k^F) d_k^Q{}^T H_{k+1}^F d_k^F \|$$

$$+ \| g_{k+1}^Q{}^T H_{k+1}^Q d_k^F d_k^Q{}^T H_{k+1}^F (d_k^F - d_k^Q) \| + \| (g_{k+1}^Q - g_{k+1}^F)^T H_{k+1}^Q d_k^F d_k^Q{}^T H_{k+1}^F d_k^Q \|$$

$$+ \| g_{k+1}^F{}^T (H_{k+1}^Q - H_{k+1}^F) d_k^F d_k^Q{}^T H_{k+1}^F d_k^Q \| + \| g_{k+1}^F{}^T H_{k+1}^F d_k^F d_k^Q{}^T (H_{k+1}^F - H_{k+1}^Q) d_k^Q \|$$

$$\leq 3M^2 \| g_{k+1}^Q \| \, \| d_k^Q \| \, \| d_k^F \| \, \| d_k^Q - d_k^F \| + M^2 \| d_k^Q \|^2 \| d_k^F \| \, \| g_{k+1}^Q - g_{k+1}^F \|$$

$$+ 2M \| g_{k+1}^F \| \, \| d_k^F \| \, \| d_k^Q \|^2 \| H_{k+1}^Q - H_{k+1}^F \| .$$

This implies that

$$|\beta_k^{Q} - \beta_k^{f}|\,\|\mathbf{d}_k^{f}\| \le 3\left[\frac{M}{m}\right]^2 \frac{\|\mathbf{g}_{k+1}^{Q}\|}{\|\mathbf{d}_k^{Q}\|}\,\|\mathbf{d}_k^{Q} - \mathbf{d}_k^{f}\|$$

$$+\left[\frac{M}{m}\right]^2 \|\mathbf{g}_{k+1}^{Q} - \mathbf{g}_{k+1}^{f}\| + \frac{2M}{m^2}\,\|\mathbf{g}_{k+1}^{f}\|\,\|\mathbf{H}_{k+1}^{Q} - \mathbf{H}_{k+1}^{f}\|. \qquad (26)$$

Using the same reasoning as for HS, we have that (17), (18) and (19) also hold for D. Also, by the Lipschitz condition and Lemma 2.15, we have

$$\|\mathbf{H}_{k+1}^{Q} - \mathbf{H}_{k+1}^{f}\| \le L\,\|\mathbf{x}_{k+1}^{f} - \mathbf{x}^*\| \le L\left[\frac{M}{m}\right]^{\frac{3}{2}} \|\mathbf{x}_k^{f} - \mathbf{x}^*\|$$

$$\le L\left[\frac{M}{m}\right]^{\frac{3}{2}} \omega. \qquad (27)$$

Thus, substituting (17), (18), (19), and (27) into (26), gives:

$$|\beta_k^{Q} - \beta_k^{f}|\,\|\mathbf{d}_k^{f}\| \le 3\left[\frac{M}{m}\right]^{\frac{5}{2}} \frac{1}{a}\,\|\mathbf{d}_k^{Q} - \mathbf{d}_k^{f}\| + \frac{M^3}{m^2}\,\|\mathbf{x}_{k+1}^{Q} - \mathbf{x}_{k+1}^{f}\|$$

$$+\left[\frac{M}{m}\right]^5 L\,\|\mathbf{x}_0 - \mathbf{x}^*\|^2 + LM\left[\frac{M}{m}\right]^3 \omega\,\|\mathbf{x}_0 - \mathbf{x}^*\|. \qquad (28)$$

Combining (14), (15), (28) and Lemma 2.16, we get our result for method D.

Lastly, we must investigate $|\beta_k^{Q} - \beta_k^{f}|\,\|\mathbf{d}_k^{f}\|$ for PR. By definition,

$$\beta_k^{Q} - \beta_k^{f} = \left[\frac{(\mathbf{g}_{k+1}^{Q} - \mathbf{g}_k^{Q})^{\mathrm{T}}\mathbf{g}_{k+1}^{Q}}{\mathbf{g}_k^{Q\mathrm{T}}\mathbf{g}_k^{Q}} - \frac{(\mathbf{g}_{k+1}^{f} - \mathbf{g}_k^{f})^{\mathrm{T}}\mathbf{g}_{k+1}^{f}}{\mathbf{g}_k^{f\mathrm{T}}\mathbf{g}_k^{f}}\right],$$

which gives

$$|\beta_k^{Q} - \beta_k^{f}|\,\|\mathbf{g}_k^{Q}\|^2\,\|\mathbf{g}_k^{f}\|^2 \le$$

$$\leq \| g_{k+1}^{Q\,T} d_{k+1}^{Q} - g_k^{Q\,T} g_{k+1}^{Q} \| \| g_k^f \|^2 + \| g_{k+1}^{f\,T} g_{k+1}^{f} - g_k^{f\,T} g_{k+1}^{f} \| \| g_k^Q \|^2$$

$$\leq \| g_{k+1}^Q \|^2 \| g_k^f \|^2 + \| g_k^Q \| \| g_{k+1}^Q \| \| g_k^f \|^2 + \| g_{k+1}^f \|^2 \| g_k^Q \|^2 + \| g_k^f \| \| g_{k+1}^f \| \| g_k^Q \|^2.$$

This implies that

$$| \beta_k^Q - \beta_k^f | \| g_k^f \| \leq \| g_k^f \| \left[\frac{\| g_{k+1}^Q \|^2}{\| g_k^Q \|^2} + \frac{\| g_{k+1}^Q \|}{\| g_k^Q \|} \right] + \| g_{k+1}^f \| \left[\frac{\| g_{k+1}^f \|}{\| g_k^f \|} + 1 \right].$$

By Lemmas 2.13 and 2.14, both $\frac{\| g_{k+1}^f \|}{\| g_k^f \|}$ and $\frac{\| g_{k+1}^Q \|}{\| g_k^Q \|}$ are $\leq \left[\frac{M}{m} \right]^{\frac{1}{2}}$. Also, by Lemmas 2.13 and 2.15,

$$\| g_k^f \| \leq M \| x_k - x^* \| \leq M \left[\frac{M}{m} \right]^{\frac{3}{2}} \| x_0 - x^* \|$$

and

$$\| g_{k+1}^f \| \leq M \| x_{k+1} - x^* \| \leq M \left[\frac{M}{m} \right]^{\frac{3}{2}} \| x_0 - x^* \|.$$

Therefore,

$$| \beta_k^Q - \beta_k^f | \| g_k^f \| \leq C_1 \| x_0 - x^* \| \tag{29}$$

where $C_1 = M \left[\frac{M}{m} \right]^{\frac{3}{2}} \left[\frac{M}{m} + 2 \left[\frac{M}{m} \right]^{\frac{1}{2}} + 1 \right].$

Using the same reasoning as for FR, (23) also holds for PR:

$$\frac{\| d_k \|^2}{\| g_k \|^2} \leq 1 + 2|T|\,|\cos\phi| + |T|^2 \leq (1+|T|)^2.$$

Using the PR formula for β_{k-1}, we have

$$\frac{\| \mathbf{d}_k \|}{\| \mathbf{g}_k \|} \le 1 + |T| = 1 + \frac{\| \mathbf{g}_k - \mathbf{g}_{k-1} \| \| \mathbf{g}_k \|}{\| \mathbf{g}_{k-1} \|^2} \frac{\| \mathbf{d}_{k-1} \|}{\| \mathbf{g}_k \|}$$

$$\le 1 + \frac{\| \mathbf{g}_k \|}{\| \mathbf{g}_{k-1} \|} \frac{\| \mathbf{d}_{k-1} \|}{\| \mathbf{g}_{k-1} \|} + \frac{\| \mathbf{d}_{k-1} \|}{\| \mathbf{g}_{k-1} \|}.$$

As with FR, if $\mathbf{d}_0 = -\mathbf{g}_0$, then this recurrence relation gives:

$$\frac{\| \mathbf{d}_k \|}{\| \mathbf{g}_k \|} \le \sum_{j=0}^{k-1} \left[\frac{M}{m} \right]^{\frac{j}{2}} \le \sum_{j=0}^{n} \left[\frac{M}{m} \right]^{\frac{j}{2}},$$

for $k = 0, 1, \ldots, n$. So we have,

$$\| \mathbf{d}_k \| \le \tilde{R} \, \| \mathbf{g}_k \|$$

where $\tilde{R} = 1 + \left[\frac{M}{m} \right]^{\frac{1}{2}} \sum_{j=0}^{n} \left[\frac{M}{m} \right]^{\frac{j}{2}} + \sum_{j=0}^{n} \left[\frac{M}{m} \right]^{\frac{j}{2}}$. Using these inequalities (with superscripts) yields:

$$| \beta_k^{Q} - \beta_k^{f} | \frac{1}{\tilde{R}} \| \mathbf{d}_k^{f} \| \le \| \beta_k^{Q} - \beta_k^{f} \| \| \mathbf{g}_k^{f} \| \le C_1 \| \mathbf{x}_0 - \mathbf{x}^* \|$$

and

$$| \beta_k^{Q} - \beta_k^{f} | \| \mathbf{d}_k^{f} \| \le C_1 \tilde{R} \, \| \mathbf{x}_0 - \mathbf{x}^* \|. \tag{30}$$

Finally, (14), (15), (30) and Lemma 2.16 give us our result for PR. We have now shown that the conjugate gradient methods FR, PR, HS and D applied to a function satisfying (AS1W), (AS2) and (AS3) are n-step quadratically convergent provided that the line search satisfies the condition $\forall\, k = 1, 2, \ldots, n-1$, $\theta_k \le K \| \mathbf{g}_0 \|$, where $0 \le \theta_k \le \eta$, $0 < \eta < 1$ with η and K constants independent of k.

Chapter 3
A REPLACEMENT FOR THE LINE SEARCH

There are several difficulties with Lenard's theory for conjugate gradient methods with inaccurate line searches. Finding a step-size which satisfies the condition on the line search may become very complicated, requiring several trial step-sizes. Each of these trials involves calculating the gradient, evaluations which can make even inexact line searches very expensive. We will present a different condition on the line search and a particular simplification which preserves the n-step quadratic convergence.

3.1. Justification for the Replacement

Consider the following proposition [11].

(3.1) PROPOSITION. *Let $\phi(s) \in C^3([0,\infty))$ and let $m(s)$ and $M(s)$ denote the minimum and maximum of ϕ'' on $[0,s]$. If $\phi'(0) \le 0$ and $m(b)>0$ for some $b>0$ with $\phi(b) \le \phi(0)$, then the quadratic that interpolates the value and the derivative of ϕ at $s=0$ and the value of ϕ at $s=b$ has a minimizer α. If, in addition, ϕ itself has a local minimizer, $\alpha^* > 0$ with $m(\alpha^*) > 0$, and*

$$\frac{M(\max\{b,\alpha^*\})}{m(\max\{b,\alpha^*\})} \le 4,$$

then

$$|\alpha - \alpha^*| \le \frac{3(\alpha^*)^2}{m(\max\{b,\alpha^*\})} \cdot \max_{0 \le s \le \max\{b,\alpha\}} |\phi'''(s)|.$$

This proposition implies that under the appropriate assumptions, the approximation, α_k, to the exact minimizer, α_k^*, found by one quadratic fit for

$\phi(s) = f(\mathbf{x}_{k-1} + s\,\mathbf{d}_{k-1})$ would satisfy the condition

$$|\alpha_k - \alpha_k^*| \le \frac{3(\alpha_k^*)^2}{m\,\|\mathbf{d}_{k-1}\|^2} \cdot \max_{0 \le s \le \max\{a,\alpha_k\}} |\phi'''(s)|.$$

Note that

$$|\phi'''(s)| = |\sum_{i,j,l=1}^{n} \frac{\partial^3 f(\mathbf{x}_k + s\,\mathbf{d}_k)}{\partial x_i \partial x_j \partial x_l} \cdot d_i d_j d_l|$$

$$\le \sum_{i,j,l} |\frac{\partial^3 f(\mathbf{x}_k + s\,\mathbf{d}_k)}{\partial x_i \partial x_j \partial x_l}| \cdot |d_i|\,|d_j|\,|d_l|,$$

assuming that $\mathbf{d}_k = (d_1, d_2, \ldots, d_n)^{\mathrm{T}}$ and $\mathbf{x}_k = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$. But we know that $|d_i| \le \|\mathbf{d}_k\|$, and since

$$|\frac{\partial^3 f(\mathbf{x}_k + s\,\mathbf{d}_k)}{\partial x_i\, \partial x_j\, \partial x_l}| \le \nu < \infty,$$

we have

$$|\phi'''(s)| \le 3n\nu\,\|\mathbf{d}_k\|^3.$$

This implies

$$|\alpha_k - \alpha_k^*| \le \frac{9(\alpha_k^*)^2 n\nu\,\|\mathbf{d}_k\|^3}{m\,\|\mathbf{d}_k\|^2},$$

or

$$\|(\alpha_k - \alpha_k^*)\mathbf{d}_k\| \le c\,\|\alpha^*\,\mathbf{d}_k\|^2$$

where $c \in \mathbf{R}$, $c > 0$.

We propose to use this as a general criterion on the line search in the conjugate gradient method. Since the error here is squared, as suggested by Proposition

3.1, the overall n-step quadratic convergence of the method should be preserved. We will show that with this condition, the following algorithm will converge n-step quadratically.

(3.2) <u>The Modified FR Conjugate Gradient Method</u>

Given arbitrary $\mathbf{x}_0 \in \mathbf{R}^n$:

1. $\mathbf{g}_0 = \nabla f(\mathbf{x}_0)$
2. $\mathbf{d}_0 = -\mathbf{g}_0$
3. for $k = 0$ to $n-1$ do

   i. $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$

   where α_k is found by one step of any univariate minimization scheme with $\| (\alpha_k - \alpha^*)\mathbf{d}_k \| \leq c \, \|\alpha^* \mathbf{d}_k\|^2$

   ii. $\mathbf{g}_{k+1} = \nabla f(\mathbf{x}_{k+1})$; check stopping criterion

   iii. $\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k$

   where β_k is chosen to be 0 if $k = n-1$, otherwise

   $$\beta_k = \frac{\mathbf{g}_{k+1}^T \mathbf{g}_{k+1}}{\mathbf{g}_k^T \mathbf{g}_k}$$

4. if convergence has not been reached:

   i. $\mathbf{x}_0 \leftarrow \mathbf{x}_n$

   ii. $\mathbf{g}_0 \leftarrow \mathbf{g}_n$

   iii. $\mathbf{d}_0 \leftarrow \mathbf{d}_n$

   iv. go to 3.

3.2. N-Step Quadratic Convergence

Before proceeding to the proof of n-step quadratic convergence, we will generalize Lenard's work with inaccurate line searches presented in Chapter 2. Again we will assume that the function, f, to be minimized satisfies conditions (AS1W), (AS2) and (AS3), and Lemmas 2.13, 2.14, and 2.15. We will prove the following variations of Lemma 2.16 and Theorem 2.17.

(3.3) LEMMA. (Generalization of 2.16.) *Suppose that* $g_k^{\ell\,T} d_k^\ell < 0$ *and* $\theta_k = \dfrac{g_{k+1}^{\ell\,T} d_k^\ell}{g_k^{\ell\,T} d_k^\ell}$ *for* $|\theta_k| \le 1-\eta$, *where* $0<\eta<1$. *Let* ω *be s.t.* $\| x_k^\varrho - x^* \| \le \omega$ *and* $\| x_k^\ell - x^* \| \le \omega$, *and suppose that* $\exists a > 0$ *s.t.* $\| d_k^\varrho \| \ge a \| g_k^\varrho \|$, *then*

$$\| d_{k+1}^\varrho - d_{k+1}^\ell \| \le B_1 \| x_k^\varrho - x_k^\ell \| + B_2 \| d_k^\varrho - d_k^\ell \| + B_3 \omega^2 + B_4 |\theta_k| \omega$$

for $B_i > 0$ *depending only upon* m, M *and* L.

(3.4) THEOREM. (Generalization of 2.17.) *If* $|\theta_k| \le K \| \nabla f(x_0) \|$, $k = 0, 1, \ldots, n-1$, *with* $|\theta_k| \le 1-\eta$, $0<\eta<1$ *for* x_0 *in a neighborhood of the minimizer,* x^*, *of* f, *(with* K *and* η *positive constants independent of* k*;) and if the procedure satisfies:*

i) *A quadratic problem is solved in at most n steps with exact line search.*

ii) $\exists a > 0$ *depending only upon* m *and* M *s.t.*

$$\| d_k^\varrho \| \ge a \| g_k^\varrho \|.$$

iii*) *Under the hypotheses of Lemma 3.3 and ii),*

$$\| d_{k+1}^\varrho - d_{k+1}^\ell \| \le C_1 \| x_k^\varrho - x_k^\ell \| + C_2 \| d_k^\varrho - d_k^\ell \| + C_3 \omega^2 + C_4 |\theta_k| \omega,$$

for positive constants C_i *depending only upon* m, M *and* L.

Then:

$$\| x_n^f - x^* \| \le \delta \| x_0 - x^* \|^2$$

where δ is a constant depending only upon m, M, L and K.

Note that the only changes in the statement of the lemma and the theorem concern θ_k. Lenard's condition that $0 \le \theta_k \le 1-\eta$, $0 < \eta < 1$ forces the line search to approach the minimizer from the left side only. We choose not to limit the line search in this way, requiring only that $|\theta_k| < 1-\eta$. The proofs follow Lenard's [14] very closely, the major differences occuring in the proof of the lemma. We will prove Lemma 3.3 in detail and highlight the differences between Theorems 2.17 and 3.4.

PROOF OF LEMMA 3.3: From the definition of x_{k+1},

$$\| x_{k+1}^Q - x_{k+1}^f \| = \| x_k^Q + \alpha_k^Q d_k^Q - x_k^f - \alpha_k^f d_k^f \|$$

$$\le \| x_k^Q - x_k^f \| + |\alpha_k^Q| \| d_k^Q - d_k^f \| + |\alpha_k^Q - \alpha_k^f| \| d_k^f \|.$$

Then

$$\| x_{k+1}^Q - x_{k+1}^f \| \le \| x_k^Q - x_k^f \| + |\alpha_k^Q| \| d_k^Q - d_k^f \|$$

$$+ (|\alpha_k^Q - \alpha_k^*| + |\alpha_k^* - \alpha_k^f|) \| d_k^f \|. \tag{31}$$

We will consider each term in (31) separately. Note that α_k^Q is determined by the fact that $g_{k+1}^{Q\,T} d_k^Q = 0$. But, by Taylor's Theorem,

$$g_{k+1}^Q = g_k^Q + \alpha_k^Q \nabla^2 Q d_k^Q.$$

So

$$\alpha_k^Q = \frac{-g_k^{Q\,T} d_k^Q}{d_k^{Q\,T} \nabla^2 Q \, d_k^Q} \le \frac{1}{ma}.$$

Similarly, letting $\tilde{x}_{k+1} = x_k^f + \alpha_k^* d_k^f$, (i.e., $\nabla f(\tilde{x}_{k+1})^T d_k = 0$,) we have

$$0 = \nabla f(x_k^f + \alpha_k^* d_k^f) = \nabla f(x_k^f) + \alpha_k^* \nabla^2 f(\bar{x})^T d_k^f,$$

where $\bar{x} = (1-\lambda)x_k^f + \lambda \tilde{x}_{k+1}$ for some $\lambda \in [0, 1]$. Therefore, we have

$$\alpha_k^f = \frac{-g_k^{fT} d_k^f}{d_k^{fT} H(\bar{x}) d_k^f}.$$

Using Lemma 2.15 and the hypothesis on ω,

$$\| \bar{x} - x^* \| \le \| x_k^f - x^* \| + \| x_{k+1}^f - x^* \| \le \omega + \left[\frac{M}{m}\right]^{\frac{3}{2}} \| x_k^f - x^* \|$$

$$\le \omega + \left[\frac{M}{m}\right]^{\frac{3}{2}} \omega = \left[1 + \left[\frac{M}{m}\right]^{\frac{3}{2}}\right] \omega. \tag{32}$$

Now,

$$\alpha_k^Q - \alpha_k^* = \frac{g_k^{fT} d_k^f}{d_k^{fT} H(\bar{x}) d_k^f} - \frac{g_k^{QT} d_k^Q}{d_k^{QT} \nabla^2 Q\, d_k^Q},$$

which implies that

$$(d_k^{fT} H(\bar{x}) d_k^f)(d_k^{QT} \nabla^2 Q\, d_k^Q)(\alpha_k^Q - \alpha_k^*) = g_k^{fT} d_k^f d_k^{QT} \nabla^2 Q\, d_k^Q - g_k^{QT} d_k^Q d_k^{fT} H(\bar{x}) d_k^f,$$

which implies that

$$m^2 \| d_k^f \|^2 \| d_k^Q \|^2 | \alpha_k^Q - \alpha_k^* | \le | g_k^{fT} d_k^f d_k^{QT} \nabla^2 Q\, d_k^Q - g_k^{QT} d_k^Q d_k^{fT} H(\bar{x}) d_k^f |$$

$$= | g_k^{QT} d_k^Q d_k^{fT} H(\bar{x}) d_k^f - g_k^{fT} d_k^f d_k^{QT} \nabla^2 Q\, d_k^Q \pm g_k^{QT} d_k^Q d_k^{fT} H(\bar{x}) d_k^Q$$

$$\pm g_k^{QT} d_k^f d_k^{fT} H(\bar{x}) d_k^Q \pm g_k^{QT} d_k^f d_k^{QT} H(\bar{x}) d_k^Q$$

$$\pm g_k^{fT} d_k^f d_k^{QT} H(\bar{x}) d_k^Q \pm g_k^{fT} d_k^f d_k^{QT} \nabla^2 Q\, d_k^Q |$$

$$\le | g_k^{QT} d_k^Q d_k^{fT} H(\bar{x})(d_k^Q - d_k^f) | + | g_k^{QT} (d_k^Q - d_k^f) d_k^{fT} H(\bar{x}) d_k^Q |$$

$$+ \; | g_k^{Q\mathrm{T}} d_k^f (d_k^f - d_k^Q)^{\mathrm{T}} H(\bar{x}) d_k^Q | + | (g_k^Q - g_k^f)^{\mathrm{T}} d_k^f d_k^{Q\mathrm{T}} H(\bar{x}) d_k^Q |$$

$$+ \; | g_k^{f\mathrm{T}} d_k^f d_k^{Q\mathrm{T}} (H(\bar{x}) - \nabla^2 Q) d_k^Q |$$

$$\le 3M \| g_k^Q \| \| d_k^Q \| \| d_k^f \| \| d_k^Q - d_k^f \| + M \| d_k^f \| \| d_k^Q \|^2 \| g_k^Q - g_k^f \|$$

$$+ L \| g_k^f \| \| d_k^f \| \| d_k^Q \|^2 \| \bar{x} - x^* \| .$$

This implies that

$$m^2 \| d_k^f \| | \alpha_k^Q - \alpha_k^* | \le 3M \frac{\| g_k^Q \|}{\| d_k^Q \|} \| d_k^Q - d_k^f \| + M \| g_k^Q - g_k^f \|$$

$$+ L \| g_k^f \| \| \bar{x} - x^* \| .$$

Using Lemma 2.13 and the hypothesis,

$$m^2 \| d_k^f \| | \alpha_k^Q - \alpha_k^* | \le \frac{3M}{a} \| d_k^Q - d_k^f \| + M^2 \| x_k^Q - x_k^F \|$$

$$+ ML \| x_k^f - x^* \| \| \bar{x} - x^* \| .$$

By (32), we have

$$| \alpha_k^Q - \alpha_k^* | \| d_k^f \| \le A_1 \| d_k^Q - d_k^f \| + A_2 \| x_k^Q - x_k^f \| + A_3 \omega^2, \qquad (33)$$

where $A_1 = \dfrac{3M}{am^2}$, $A_2 = M^2$ and $A_3 = ML \left[1 + \left[\dfrac{M}{m} \right]^{\frac{3}{2}} \right]$.

Finally, we consider $| \alpha_k^* - \alpha_k^f | \| d_k^f \|$. Let $\phi(\alpha) = f(x_k^f + \alpha d_k^f)$; then by Taylor,

$$\phi'(\alpha_k^*) = \phi'(\alpha_k^f) + \int_{\alpha_k^f}^{\alpha_k^*} \phi''(t) dt .$$

But

$$\phi'(\alpha_k^*) = \nabla f(\mathbf{x}_k^\ell + \alpha_k^* \mathbf{d}_k^\ell)^{\mathrm{T}} \mathbf{d}_k^\ell = 0$$

and

$$-\phi'(\alpha_k^\ell) = -\mathbf{g}_{k+1}^{\ell\,\mathrm{T}} \mathbf{d}_k^\ell = -\theta_k \mathbf{g}_k^{\ell\mathrm{T}} \mathbf{d}_k^\ell .$$

So,

$$-\theta_k \mathbf{g}_k^{\ell\mathrm{T}} \mathbf{d}_k^\ell = \int_{\alpha_k^\ell}^{\alpha_k^*} \phi''(t)dt ,$$

which implies that

$$|\theta_k \mathbf{g}_k^{\ell\mathrm{T}} \mathbf{d}_k^\ell| = \left| \int_{\alpha_k^\ell}^{\alpha_k^*} \phi''(t)dt \right| \geq |m(\alpha_k^* - \alpha_k^\ell) \| \mathbf{d}_k^\ell \|^2 |$$

or

$$|\theta_k| \| \mathbf{g}_k^\ell \| \| \mathbf{d}_k^\ell \| \geq m |\alpha_k^* - \alpha_k^\ell| \| \mathbf{d}_k^\ell \|^2 .$$

This implies that

$$|\alpha_k^* - \alpha_k^\ell| \| \mathbf{d}_k^\ell \| \leq \frac{|\theta_k|}{m} \| \mathbf{g}_k^\ell \| .$$

By Lemma 2.13, we have

$$|\alpha_k^\ell - \alpha_k^*| \| \mathbf{d}_k^\ell \| \leq \frac{|\theta_k|}{m} M \omega ,$$

which, combined with (31) and (33), implies that

$$\| \mathbf{x}_{k+1}^Q - \mathbf{x}_{k+1}^\ell \| \leq \| \mathbf{x}_k^Q - \mathbf{x}_k^\ell \| + \frac{1}{ma} \| \mathbf{d}_k^Q - \mathbf{d}_k^\ell \| + A_1 \| \mathbf{d}_k^Q - \mathbf{d}_k^\ell \|$$

$$+ A_2 \| \mathbf{x}_k^Q - \mathbf{x}_k^\ell \| + A_3 \omega^2 + \frac{M}{m} |\theta_k| \omega$$

$$= B_1 \| x_k^Q - x_k^f \| + B_2 \| d_k^Q - d_k^f \| + B_3 \omega^2 + B_4 \omega | \theta_k |. \qquad ////$$

Now we examine the changes imposed on the proof of Theorem 2.17 when the hypothesis is changed, and Lemma 3.3 is used instead of 2.16.

PROOF OF THEOREM (3.4): We begin in the same fashion, letting $w = (\mathbf{x}_0^T, \mathbf{d}_0^T)^T$ denote the starting point with $\mathbf{d}_0 = -\mathbf{g}_0$, and considering two other starting points, u and v with $u = (\mathbf{x}_u^T, \mathbf{d}_u^T)^T$ and $v = (\mathbf{x}_v^T, \mathbf{d}_v^T)^T$. Now we apply Lemma 3.3 and iii*) to u and v to give:

$$\| \mathbf{x}_1^f(u) - \mathbf{x}_1^Q(v) \| + \| \mathbf{d}_1^f(u) - \mathbf{d}_1^Q(v) \| \le D_1 \| \mathbf{x}_u - \mathbf{x}_v \| + D_2 \| \mathbf{d}_u - \mathbf{d}_v \|$$

$$+ A^2 D_3 \| \mathbf{x}_0 - \mathbf{x}^* \|^2 + A D_4 | \theta_0 | \| \mathbf{x}_0 - \mathbf{x}^* \| \qquad (34)$$

where $D_i = B_i + C_i$. This is the same inequality as (8) in the proof of Theorem 2.17, except for the addition of the absolute values around θ_0. The proof continues on as before, now with

$$| \theta_0 | \le KM \| \mathbf{x}_0 - \mathbf{x}^* \|. \qquad (35)$$

This is just equation (9) from the proof of Theorem 2.17, again with the additional absolute values. The proof continues, using (35) to substitute $KM \| \mathbf{x}_0 - \mathbf{x}^* \|$ into (34) for $| \theta_0 |$ just as in the proof of Theorem 2.17. The rest of the proof of Theorem 3.4 is exactly the same as that of Theorem 2.17. ////

Now consider the following Theorem.

(3.5) THEOREM. *When used to minimize a function which satisfies (AS1W), (AS2) and (AS3), the conjugate gradient algorithm 3.2 is n-step quadratically convergent in a neighborhood of the minimizer.*

The proof of this result will rely on the following technical lemmas. For all of these lemmas, we assume that we are using Algorithm 3.2 to minimize a function satisfying (AS1W), (AS2) and (AS3).

(3.6) LEMMA [11]. *If* $\phi \in C^2([0,\infty))$, $\phi'(\alpha_k^*) = 0$, $\phi''(\alpha_k^*) > 0$, *then* $\exists$ *a neighborhood of* α_k^* *with the property that for each* α *in that neighborhood with*

$$2|\alpha - \alpha_k^*| \leq |\alpha_k^*|$$

we have

$$\phi(\alpha) < \phi(0).$$

PROOF OF LEMMA 3.6: Let I be a closed interval containing α_k^* and 0, $\tilde{m} = \min_{\alpha \in I} \phi''(\alpha)$ and $\tilde{M} = \max_{\alpha \in I} \phi''(\alpha)$. Then, by Taylor's Theorem,

$$\phi(\alpha) \leq \phi(\alpha_k^*) + \frac{\tilde{M}}{2}(\alpha - \alpha_k^*)^2$$

and

$$\phi(0) \geq \phi(\alpha_k^*) + \frac{\tilde{m}}{2}(0 - \alpha_k^*)^2.$$

Thus

$$\phi(\alpha) - \phi(0) \leq \frac{\tilde{M}}{2}(\alpha - \alpha_k^*)^2 - \frac{\tilde{m}}{2}(\alpha_k^*)^2.$$

By the hypothesis, we have $(\alpha - \alpha_k^*)^2 \leq \frac{1}{4}(\alpha_k^*)^2$, which implies that

$$\phi(\alpha) - \phi(0) \leq \left[\frac{\tilde{M}}{8} - \frac{\tilde{m}}{2}\right](\alpha_k^*)^2.$$

So, if I is small enough that $\frac{\tilde{M}}{\tilde{m}} < 4$, we have our result. ////

(3.7) LEMMA. $|\alpha_k - \alpha_k^*| \le \frac{c}{m} \| g_k \| |\alpha_k^*|$

PROOF OF LEMMA 3.7: $\| (\alpha_k - \alpha_k^*)\mathbf{d}_k \| \le c \| \alpha_k^* \mathbf{d}_k \|^2$ implies

$$|\alpha_k - \alpha_k^*| \le c\, |\alpha_k^*| \cdot |\alpha_k^*| \, \| \mathbf{d}_k \|.$$

Using Taylor's Theorem, setting $\phi(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{d}_k)$,

$$\phi'(\alpha_k^*) = \phi'(0) + \int_0^{\alpha_k^*} \phi''(t)dt$$

or

$$\phi'(\alpha_k^*) \ge \phi'(0) + m\, |\alpha_k^*| \, \| \mathbf{d}_k \|^2.$$

But $\phi'(\alpha_k^*) = 0$ and $\phi'(0) = \mathbf{g}_k^T \mathbf{d}_k$ so

$$\frac{-\mathbf{g}_k^T \mathbf{d}_k}{m \| \mathbf{d}_k \|} \ge |\alpha_k^*| \, \| \mathbf{d}_k \|$$

or

$$|\alpha_k^*| \, \| \mathbf{d}_k \| \le \frac{\| \mathbf{g}_k \|}{m}.$$ ////

(3.8) LEMMA. *If* $\mathbf{x}_0$ *is sufficiently close to* $\mathbf{x}^*$ *then*

$$\| \mathbf{g}_k \| \le C_1 \| \mathbf{g}_{k-1} \| \quad \textit{and} \quad \| \mathbf{g}_k \| \le C_2 \| \mathbf{g}_0 \|,$$

where $C_1 = 1 + \frac{2M}{m}$ *and* $C_2 = \left[1 + \frac{2M}{m}\right]^n$.

PROOF OF LEMMA 3.8: We prove this result by induction. We'll need to assume that $\mathbf{x}_0$ is contained in a neighborhood of $\mathbf{x}^*$ where

$$\| \mathbf{g}_0 \| \le \frac{m}{2c} \cdot \left[1 + \frac{2M}{m}\right]^{-n}.$$

First we must prove the conclusion of the lemma for $k = 0$.

By (AS2) and the Mean Value Theorem, we have: $\| g_1 - g_0 \| \le M \| x_1 - x_0 \|$, which implies

$$\| g_1 \| \le \| g_0 \| + M \| x_1 - x_0 \| = \| g_0 \| + M \alpha_0 \| d_0 \|. \tag{36}$$

Let $\phi(\alpha) = f(x_0 + \alpha d_0)$; then by Taylor's Theorem,

$$\phi(\alpha_1) = \phi(0) + \alpha_1 \phi'(0) + \frac{(\alpha_1)^2}{2} \phi''(\tilde{\alpha})$$

where $\tilde{\alpha} = \lambda \alpha_1$, $\lambda \in [0, 1]$. We want to apply Lemma 3.6, so we need

$$2 | \alpha_1 - \alpha_1^* | \le |\alpha_1^*|. \tag{37}$$

By Lemma 3.7, $| \alpha_1 - \alpha_1^* | \le \frac{c}{m} | \alpha_1^* | \| g_0 \|$, and by our assumption on the size of the neighborhood above, we know that $\frac{c}{m} \| g_0 \| \le \frac{1}{2}$; thus we have (37). By lemma 3.6, we have

$$\phi(0) > \phi(\alpha_1) \ge \phi(0) + \alpha_1 g_0^T d_0 + \frac{(\alpha_1)^2}{2} m \| d_0 \|^2,$$

which implies that

$$\frac{-2\alpha_1 g_0^T d_0}{m \alpha_1 \| d_0 \|} \ge \alpha_1 \| d_0 \| \quad \text{or} \quad \alpha_1 \| d_0 \| \le \frac{2}{m} \| g_0 \|.$$

Again, to use Lemma 3.6, we must assume that x_0 is contained in a small enough neighborhood of x^* so that $\frac{\tilde{M}}{\tilde{m}} < 4$. Note that $\phi''(\alpha) = d_0^T H(x_0 + \alpha d_0) d_0$ and by (AS2), $\forall\, x \in R^n$

$$m \| d_0 \|^2 \le d_0^T H(x) d_0 \le M \| d_0 \|^2.$$

Also, max $\phi''(\alpha) = \tilde{M} \le M$, and min $\phi''(\alpha) = \tilde{m} \ge m$; thus, if $\mathbf{x}_0$ is in a sufficiently small neighborhood of $\mathbf{x}^*$,

$$\frac{\max \mathbf{d}_k^T \mathbf{H}(\mathbf{x}_k + \alpha \mathbf{d}_k)\mathbf{d}_k}{\min \mathbf{d}_k^T \mathbf{H}(\mathbf{x}_k + \alpha \mathbf{d}_k)\mathbf{d}_k} \le 4 \tag{38}$$

where $k = 0$ and the max and min are taken over the neighborhood of $\mathbf{x}^*$ containing $\mathbf{x}_0$.

Plugging into (36) yields

$$\| \mathbf{g}_1 \| \le \| \mathbf{g}_0 \| \left[1 + \frac{2M}{m} \right],$$

and the base step of our proof by induction is complete.

Now we assume the result for k and prove for k+1. By (AS2) and the Mean Value Theorem, as before, we have

$$\| \mathbf{g}_{k+1} \| \le \| \mathbf{g}_k \| + M \alpha_k \| \mathbf{d}_k \| \tag{39}$$

Letting $\phi(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{d}_k)$ and applying Taylor's Theorem yields

$$\phi(\alpha_k) = \phi(0) + \alpha_k \phi(0) + \frac{(\alpha_k)^2}{2} \phi''(\bar{\alpha})$$

where $\bar{\alpha} = \lambda \alpha_k$, $\lambda \in [0, 1]$.

Again, as in the base step, we want to apply lemma 3.6, so we must obtain the hypothesis, $2 | \alpha_{k+1} - \alpha_k^* | \le | \alpha_k^* |$. By Lemma 3.7,

$$| \alpha_k - \alpha_k^* | \le \frac{c}{m} | \alpha_k^* | \| \mathbf{g}_k \|.$$

Thus, by induction we have:

$$| \alpha_k - \alpha_k^* | \le \hat{c} \frac{c}{m} | \alpha_k^* | \| \mathbf{g}_0 \|,$$

where $\hat{c} = \left[1 + \frac{M}{m}\right]^n$. By our assumption,

$$\| \mathbf{g}_0 \| \le \frac{m}{2c}\left[1 + \frac{2M}{m}\right]^{-n},$$

which implies $2|\alpha_{k+1} - \alpha^*_{k+1}| \le |\alpha^*_{k+1}|$. Thus, by Lemma 3.6

$$\phi(0) > \phi(\alpha_k) \ge \phi(0) + \alpha_k \phi'(0) + \frac{(\alpha_k)^2}{2}\phi''\phi(\tilde{\alpha}),$$

which implies

$$-\alpha_k \mathbf{g}_k^T \mathbf{d}_k \ge \frac{(\alpha_k)^2}{2} m \| \mathbf{d}_k \|^2.$$

Note that, as in the base step, we had to assume (38) in order to use Lemma 3.6. Simplifying gives

$$\alpha_k \| \mathbf{d}_k \| \le \frac{2}{m} \| \mathbf{g}_k \|.$$

Plugging this into (39) gives

$$\| \mathbf{g}_{k+1} \| \le \left[1 + \frac{2M}{m}\right] \| \mathbf{g}_k \| \le \ldots \le \left[1 + \frac{2M}{m}\right]^{k+1} \| \mathbf{g}_0 \|$$

$$\le \left[1 + \frac{2M}{m}\right]^{n} \| \mathbf{g}_0 \|$$

$\forall k = 0, 1, \ldots, n-1.$ ////

(3.9) LEMMA. *If* $\mathbf{x}_0$ *is sufficiently close to* $\mathbf{x}^*$, *then*

$$\| \mathbf{d}_k \| \le C_3 \| \mathbf{g}_k \| \quad \textit{and} \quad \| \mathbf{d}_k \| \le C_4 \| \mathbf{g}_0 \|$$

for $C_i > 0$ *depending only upon* M, m *and* L.

PROOF OF LEMMA 3.9: We will use Lemma 3.8, so we will need the same restrictions on $\mathbf{x}_0$ as before. Note that $\mathbf{d}_k = -\mathbf{g}_k + \beta_{k-1}\mathbf{d}_{k-1}$. Using Lemma 3.8 gives:

$$\frac{\|\mathbf{d}_k\|}{\|\mathbf{g}_k\|} \le 1 + \frac{\|\mathbf{g}_k\|}{\|\mathbf{g}_{k-1}\|}\cdot\frac{\|\mathbf{d}_{k-1}\|}{\|\mathbf{g}_k\|} \le 1 + C_1\frac{\|\mathbf{d}_{k-1}\|}{\|\mathbf{g}_{k-1}\|}$$

Since $\frac{\|\mathbf{d}_0\|}{\|\mathbf{g}_0\|} = \frac{\|\mathbf{g}_0\|}{\|\mathbf{g}_0\|} = 1$, we have

$$\frac{\|\mathbf{d}_k\|}{\|\mathbf{g}_k\|} \le \sum_{j=0}^{k}(C_1)^j \le \sum_{j=0}^{n}(C_1)^j = C_3.$$

Thus

$$\|\mathbf{d}_k\| \le C_3\|\mathbf{g}_k\| \le C_3C_2\|\mathbf{g}_0\| = C_4\|\mathbf{g}_0\|. \qquad ////$$

(3.10) LEMMA. *If* $\|(\alpha_k - \alpha_k^*)\mathbf{d}_k\| \le c\,\|\alpha_k^*\mathbf{d}_k\|^2$ *then for some* $C_5 > 0$ *independent of* k*, we have*

$$|\alpha_k - \alpha_k^*|\,\|\mathbf{d}_k\| \le C_5\|\mathbf{g}_k\|^2.$$

PROOF OF LEMMA 3.10: From the hypothesis, we have

$$|\alpha_k - \alpha_k^*|\,\|\mathbf{d}_k\| \le c\,|\alpha_k^*|^2\|\mathbf{d}_k\|^2.$$

Let $\phi(\alpha) = f(\mathbf{x}_k + \alpha\mathbf{d}_k)$. Then $\phi'(\alpha_k^*) = 0$ and $\phi'(0) = \mathbf{g}_k^T\mathbf{d}_k$. By Taylor's Theorem, we have

$$0 = \phi'(\alpha_k^*) = \phi'(0) + \int_0^{\alpha_k^*}\phi''(t)dt.$$

This implies that

$$-\mathbf{g}_k^T\mathbf{d}_k = \int_0^{\alpha_k^*}\phi''(t)dt.$$

By (AS2),

$$-\mathbf{g}_k^T\mathbf{d}_k \geq \alpha_k^* m \, \| \mathbf{d}_k \|^2,$$

which implies that

$$\alpha_k^* \leq \frac{-\mathbf{g}_k^T\mathbf{d}_k}{m \, \| \mathbf{d}_k \|^2}.$$

So

$$| \alpha_k - \alpha_k^* | \, \| \mathbf{d}_k \| \leq c \frac{| -\mathbf{g}_k^T\mathbf{d}_k |^2}{| m \, \| \mathbf{d}_k \|^2 |^2} \| \mathbf{d}_k \|^2 \leq \frac{c}{m^2} \| \mathbf{g}_k \|^2 \leq C_5 \| \mathbf{g}_k \|^2. \qquad ////$$

(3.11) LEMMA. *If* $\mathbf{x}_0$ *is chosen sufficiently close to* $\mathbf{x}^*$, *then* $\mathbf{d}_{k+1}$ *is a descent direction, i.e.,* $\mathbf{g}_{k+1}^T\mathbf{d}_{k+1} < 0$.

PROOF OF LEMMA 3.11: We know that $\mathbf{d}_{k+1}$ is a descent direction if $\mathbf{g}_{k+1}^T\mathbf{d}_k = 0$, because then

$$\mathbf{g}_{k+1}^T\mathbf{d}_{k+1} = -\mathbf{g}_{k+1}^T\mathbf{g}_{k+1} + \beta_k \mathbf{g}_{k+1}^T\mathbf{d}_k = -\mathbf{g}_{k+1}^T\mathbf{g}_{k+1} < 0.$$

With an approximate α_k, $\mathbf{g}_{k+1}^T\mathbf{d}_k \neq 0$, which implies that $\mathbf{g}_{k+1}^T\mathbf{d}_{k+1} \neq -\mathbf{g}_{k+1}^T\mathbf{g}_{k+1}$. But it would be sufficient to have $\mathbf{g}_{k+1}^T\mathbf{d}_k \leq \varepsilon \, \mathbf{g}_k^T\mathbf{g}_k$ for $0 < \varepsilon < 1$, because then we would have

$$\begin{aligned} \mathbf{g}_{k+1}^T\mathbf{d}_{k+1} &= \beta_k \mathbf{g}_{k+1}^T\mathbf{d}_k - \mathbf{g}_{k+1}^T\mathbf{g}_{k+1} \\ &\leq \beta_k \, \varepsilon \, \mathbf{g}_k^T\mathbf{g}_k - \mathbf{g}_{k+1}^T\mathbf{g}_{k+1}. \end{aligned}$$

Thus, using the FR definition of β_k,

$$\mathbf{g}_{k+1}^T\mathbf{d}_{k+1} \leq (\varepsilon - 1) \| \mathbf{g}_{k+1} \|^2 < 0.$$

To show that $g_{k+1}^T d_k \le \varepsilon \| g_k \|^2$, let $\phi(\alpha) = f(x_k + \alpha d_k)$; then by Taylor's Theorem

$$\phi'(\alpha_k) = \phi'(\alpha_k^*) + \int_{\alpha_k^*}^{\alpha_k} \phi''(t)dt,$$

which implies that

$$\phi'(\alpha_k) \le (\alpha_k - \alpha_k^*)M \| d_k \|^2 = (\alpha_k - \alpha_k^*)\| d_k \| M \| d_k \|$$

and by Lemma 3.10,

$$\phi'(\alpha_k) \le C_5 \| g_k \|^2 M \| d_k \|.$$

But by Lemma 3.9,

$$C_5 M \| d_k \| \le C_5 M C_4 \| g_0 \|.$$

Once again, if x_0 is sufficiently close to x^*, then

$$\| g_0 \| \le \frac{\varepsilon}{C_4 C_5 M},$$

which implies that

$$\phi'(\alpha_k) = g_{k+1}^T d_k \le \varepsilon \| g_k \|^2. \qquad ////$$

Now we are prepared to prove Theorem 3.5.

PROOF OF THEOREM 3.5: We will use Lemma 3.4 and Theorem 2.17 to show our result. First we need to verify the hypotheses i), ii) and iii*).

i) This condition is satisfied because algorithm 3.2 is just the FR method when applied with exact line searches to a quadratic.

ii) Since $d_k^Q = -g_k^Q + \beta_{k-1}^Q d_{k-1}^Q$ we have $\| d_k^Q \|^2 = \| g_k^Q \|^2 + | \beta_{k-1}^Q |^2 \| d_{k-1}^Q \|^2$, which implies that

$$\| \mathbf{d}_k^{\mathcal{Q}} \|^2 \geq \| \mathbf{g}_k^{\mathcal{Q}} \|^2.$$

iii*) By Lemma 3.11, we have that $\mathbf{g}_k^{\mathcal{L}T} \mathbf{d}_k^{\mathcal{L}} < 0$. In general, we have

$$\| \mathbf{d}_{k+1}^{\mathcal{Q}} - \mathbf{d}_{k+1}^{\mathcal{L}} \| \leq \| \mathbf{g}_{k+1}^{\mathcal{Q}} - \mathbf{g}_{k+1}^{\mathcal{L}} \| + \| \beta_k^{\mathcal{Q}} \mathbf{d}_k^{\mathcal{Q}} - \beta_k^{\mathcal{L}} \mathbf{d}_k^{\mathcal{L}} \|$$

$$\leq \| \mathbf{g}_{k+1}^{\mathcal{Q}} - \nabla \mathcal{Q}(\mathbf{x}_{k+1}^{\mathcal{L}}) \| + \| \nabla \mathcal{Q}(\mathbf{x}_{k+1}^{\mathcal{L}}) - \mathbf{g}_{k+1}^{\mathcal{L}} \|$$

$$+ | \beta_k^{\mathcal{Q}} | \| \mathbf{d}_k^{\mathcal{Q}} - \mathbf{d}_k^{\mathcal{L}} \| + | \beta_k^{\mathcal{Q}} - \beta_k^{\mathcal{L}} | \| \mathbf{d}_k^{\mathcal{L}} \|$$

Applying Lemmas 2.13, 2.15 and 3.3

$$\| \mathbf{d}_{k+1}^{\mathcal{Q}} - \mathbf{d}_{k+1}^{\mathcal{L}} \| \leq M \| \mathbf{x}_{k+1}^{\mathcal{Q}} - \mathbf{x}_{k+1}^{\mathcal{L}} \| + L \left[\frac{M}{m} \right]^3 \| \mathbf{x}_0 - \mathbf{x}^* \|^2$$

$$+ | \beta_k^{\mathcal{Q}} | \| \mathbf{d}_k^{\mathcal{Q}} - \mathbf{d}_k^{\mathcal{L}} \| + | \beta_k^{\mathcal{Q}} - \beta_k^{\mathcal{L}} | \| \mathbf{d}_k^{\mathcal{L}} \|. \tag{40}$$

Note that this is exactly statement (13) from Chapter 2 which was used to show that Lenard's version of the FR method with inaccurate line searches satisfies condition iii) of Theorem 2.17. The proof of iii*) proceeds in exactly the same manner as that of iii) to give:

$$| \beta_k^{\mathcal{Q}} | \leq \frac{M \| \mathbf{g}_{k+1}^{\mathcal{Q}} \| \| \mathbf{g}_k^{\mathcal{Q}} \|}{m \| \mathbf{g}_k^{\mathcal{Q}} \| \| \mathbf{d}_k^{\mathcal{Q}} \|} \leq \frac{1}{a} \left[\frac{M}{m} \right]^{\frac{3}{2}} \tag{41}$$

and

$$\| \beta_k^{\mathcal{Q}} - \beta_k^{\mathcal{L}} \| \| \mathbf{d}_k^{\mathcal{L}} \| \leq 2R \left[\frac{M}{m} \right]^{\frac{1}{2}} \times$$

$$\left\{ M \left[\frac{M}{m} \right]^{\frac{1}{2}} \| \mathbf{x}_k^{\mathcal{Q}} - \mathbf{x}_k^{\mathcal{L}} \| + M \| \mathbf{x}_{k+1}^{\mathcal{Q}} - \mathbf{x}_{k+1}^{\mathcal{L}} \| + L \left[\frac{M}{m} \right]^3 \| \mathbf{x}_0 - \mathbf{x}^* \| \right\}. \tag{42}$$

Combining (40), (41), (42) and Lemma 3.3 gives iii*) for algorithm 3.2.

To finish the proof of Theorem 3.5, we must prove the following claim.

(3.12) CLAIM. *If the conjugate gradient algorithm 3.2 is used to minimize a function which satisfies (AS1W), (AS2) and (AS3), then for* $k = 0, 1, \ldots, n-1$,

$$|\theta_k| \le K \| g_0 \|$$

for K *independent of* k *and* x_0 *in a neighborhood of* x^*.

PROOF OF CLAIM 3.12: We'll prove the result by induction. For $k = 0$, we have

$$\theta_0 = \frac{g_1^T d_0}{g_0^T d_0} = \frac{-g_1^T g_0}{-g_0^T g_0} = \frac{g_1^T g_0}{\| g_0 \|^2}$$

Now let $\phi(\alpha) = f(x_0 + \alpha d_0)$. Then, $\phi'(\alpha_0^*) = 0$ and, by Lemma 3.10,

$$|\alpha_0 - \alpha_0^*| \, \| d_0 \| \le C_5 \| g_0 \|^2$$

which implies that

$$|\alpha_0 - \alpha_0^*| \le C_5 \| g_0 \|.$$

By Taylor's Theorem, we have

$$0 = \phi'(\alpha_0^*) = \phi'(\alpha_0) + \int_{\alpha_0}^{\alpha_0^*} \phi''(t)\,dt,$$

$$|-g_1^T d_0| = \left| \int_{\alpha_0}^{\alpha_0^*} \phi''(t)\,dt \right|$$

$$|g_1^T d_0| \le |(\alpha_0^* - \alpha_0) M \| d_0 \|^2| = |\alpha_0^* - \alpha_0| M \| d_0 \|^2 \le C_5 M \| g_0 \|^3.$$

Thus,

$$|\theta_0| = \frac{|\mathbf{g}_1^T \mathbf{d}_0|}{\|\mathbf{g}_0\|^2} \leq K \|\mathbf{g}_0\|,$$

where $K = MC_5$. The proof of the base step is complete.

Next we assume our result for $k-1$ and prove it for k. Note that we can assume that

$$\|\mathbf{g}_{k-1}\| > \tilde{C} \|\mathbf{g}_0\|^2, \tag{43}$$

for some $\tilde{C} = \delta K C_4$ (fixed $\delta > 1$,) because if $\exists \tilde{C} > 0$ such that $\|\mathbf{g}_{k-1}\| \leq \tilde{C} \|\mathbf{g}_0\|^2$, then the quadratic convergence we are looking for would already have been obtained before step k, and we'd be finished. We will use this fact twice in the remainder of the proof.

Let $\phi(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{d}_k)$; then

$$\phi'(\alpha_k^*) = 0 \quad \text{and} \quad \phi'(\alpha_k) = \mathbf{g}_{k+1}^T \mathbf{d}_k$$

By Taylor's Theorem, we know

$$\phi'(\alpha_k) = \phi'(\alpha_k^*) + \int_{\alpha_k^*}^{\alpha_k} \phi''(t)dt.$$

So

$$|\mathbf{g}_{k+1}^T \mathbf{d}_k| = \left| \int_{\alpha_k^*}^{\alpha_k} \phi''(t)dt \right|.$$

By (AS2) and Lemma 3.10, we have

$$|\mathbf{g}_{k+1}^T \mathbf{d}_k| \leq |(\alpha_k - \alpha_k^*) M \|\mathbf{d}_k\|^2| = |\alpha_k - \alpha_k^*| M \|\mathbf{d}_k\|^2 \leq C_5 M \|\mathbf{g}_k\|^2 \|\mathbf{d}_k\|. \tag{44}$$

Then by Lemma 3.9 and the definition of θ_k, we have

$$|\theta_k| = \frac{|\mathbf{g}_{k+1}^T \mathbf{d}_k|}{|\mathbf{g}_k^T \mathbf{d}_k|} \le \frac{C_5 M \|\mathbf{g}_k\|^2 \|\mathbf{d}_k\|}{|\mathbf{g}_k^T \mathbf{d}_k|}$$

$$= \frac{C_5 M \|\mathbf{g}_k\|^2 \|\mathbf{d}_k\|}{\left| \|\mathbf{g}_k\|^2 + \frac{\|\mathbf{g}_k\|^2}{\|\mathbf{g}_{k-1}\|^2} \mathbf{g}_k^T \mathbf{d}_{k-1} \right|}$$

$$= \frac{C_5 M \|\mathbf{d}_k\|}{\left| 1 + \frac{\mathbf{g}_k^T \mathbf{d}_{k-1}}{\|\mathbf{g}_{k-1}\|^2} \right|}$$

$$\le \frac{C_4 C_5 M \|\mathbf{g}_0\|}{\left| 1 + \frac{\theta_{k-1} \mathbf{g}_{k-1}^T \mathbf{d}_{k-1}}{\|\mathbf{g}_{k-1}\|^2} \right|}.$$

Now, note that

$$\frac{\theta_{k-1} \mathbf{g}_{k-1}^T \mathbf{d}_{k-1}}{\|\mathbf{g}_{k-1}\|^2} \le \frac{|\theta_{k-1}| \|\mathbf{d}_{k-1}\|}{\|\mathbf{g}_{k-1}\|};$$

by (43) and induction, we have

$$|\theta_{k-1}| \|\mathbf{d}_{k-1}\| \le K C_4 \|\mathbf{g}_0\|^2 \le \tilde{C} \|\mathbf{g}_0\|^2 < \|\mathbf{g}_{k-1}\|,$$

which implies that

$$\frac{|\theta_{k-1} \mathbf{g}_{k-1}^T \mathbf{d}_{k-1}|}{\|\mathbf{g}_{k-1}\|^2} \le \frac{|\theta_{k-1}| \|\mathbf{d}_{k-1}\|}{\|\mathbf{g}_{k-1}\|} < 1.$$

By the Triangle Inequality, we have

$$\left| 1 + \frac{\theta_{k-1} \mathbf{g}_{k-1}^T \mathbf{d}_{k-1}}{\|\mathbf{g}_{k-1}\|^2} \right| \quad \ge \quad 1 - \frac{|\theta_{k-1} \mathbf{g}_{k-1}^T \mathbf{d}_{k-1}|}{\|\mathbf{g}_{k-1}\|^2}.$$

Thus,

$$|\theta_k| \le \frac{C_4C_5M\,\|g_0\|}{\left|1 + \frac{\theta_{k-1}g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2}\right|} \le \frac{C_4C_5M\,\|g_0\|}{1 - |\theta_{k-1}|\frac{\|d_{k-1}\|}{\|g_{k-1}\|}},$$

and then using Lemma 3.9, induction and (43),

$$|\theta_k| \le \frac{C_4C_5M\,\|g_0\|}{1 - \frac{KC_4\|g_0\|^2}{\|g_{k-1}\|}} \le \frac{C_4C_5M\,\|g_0\|}{1 - \frac{KC_4\|g_0\|^2}{\tilde{C}\,\|g_0\|^2}} \le \frac{C_4C_5M}{1-\varepsilon}\|g_0\|,$$

with $\varepsilon = \frac{1}{\delta} < 1$, since $\tilde{C} = \delta K C_4$. Thus

$$|\theta_k| \le K\,\|g_0\|. \qquad ////$$

Thus we have proven that when algorithm 3.2 is used to minimize a function which satisfies (AS1W), (AS2) and (AS3), n-step quadratic convergence is attained provided that x_0, the initial guess, is sufficiently close to x^*, the minimizer of the function.

3.3. Other Choices of Beta

We next extend the proof of Theorem 3.5 for the other forms of β_k. First note that Lemmas 3.6, 3.7, 3.8, 3.9 and 3.10 all hold for the D, PR and HS methods with the condition

$$\|(\alpha_k - \alpha_k^*)d_k\| \le c\,\|\alpha_k^* d_k\|^2$$

with no changes to their proofs. The difficulties occur in the proofs of Lemma 3.11 and Theorem 3.5. We present an alternate version of each of these. We will need to assume an additional condition for each of the methods D, HS and PR.

(3.13) LEMMA. (Alternate for Lemma 3.11.) *Assume that Algorithm 3.2 (with the form of* β_k *changed to that of the D, PR or HS methods,) is used to minimize a function , f , which satisfies (AS1W), (AS2) and (AS3). If* $\mathbf{x}_0$ *is sufficiently close to* $\mathbf{x}^*$, *then* $\mathbf{d}_{k+1}$ *is a descent direction, i.e.,* $\mathbf{g}_{k+1}^{\mathrm{T}}\mathbf{d}_{k+1} < 0$.

PROOF OF LEMMA 3.13: We know that $\mathbf{d}_{k+1}$ is a descent direction if $\mathbf{g}_{k+1}^{\mathrm{T}}\mathbf{d}_k = 0$, because then

$$\mathbf{g}_{k+1}^{\mathrm{T}}\mathbf{d}_{k+1} = -\mathbf{g}_{k+1}^{\mathrm{T}}\mathbf{g}_{k+1} + \beta_k\mathbf{g}_{k+1}^{\mathrm{T}}\mathbf{d}_k = -\mathbf{g}_{k+1}^{\mathrm{T}}\mathbf{g}_{k+1} < 0.$$

With an approximate α_k, $\mathbf{g}_{k+1}^{\mathrm{T}}\mathbf{d}_k \neq 0$, which implies that $\mathbf{g}_{k+1}^{\mathrm{T}}\mathbf{d}_{k+1} \neq -\mathbf{g}_{k+1}^{\mathrm{T}}\mathbf{g}_{k+1}$. For methods HS and D, it would be sufficient to have $\mathbf{g}_{k+1}^{\mathrm{T}}\mathbf{d}_k \le \varepsilon\frac{m}{M}\,\|\mathbf{g}_{k+1}\|\,\|\mathbf{d}_k\|$ for $0 < \varepsilon < 1$, because then we would have

$$\begin{aligned}\mathbf{g}_{k+1}^{\mathrm{T}}\mathbf{d}_{k+1} &= \beta_k\mathbf{g}_{k+1}^{\mathrm{T}}\mathbf{d}_k - \mathbf{g}_{k+1}^{\mathrm{T}}\mathbf{g}_{k+1}\\ &\le \varepsilon\frac{m}{M}\beta_k\,\|\mathbf{g}_{k+1}\|\,\|\mathbf{d}_k\| - \mathbf{g}_{k+1}^{\mathrm{T}}\mathbf{g}_{k+1}. \end{aligned} \tag{44}$$

From Lemma 2.5* (for HS) and the definition of the D form of β_k, we know that

$$(D) \qquad \beta_k = \frac{\mathbf{g}_{k+1}^{\mathrm{T}}\mathbf{H}_{k+1}\mathbf{d}_k}{\mathbf{d}_k^{\mathrm{T}}\mathbf{H}_{k+1}\mathbf{d}_k}$$

and

$$(HS) \qquad \beta_k = \frac{\mathbf{g}_{k+1}^{\mathrm{T}}\hat{\mathbf{H}}_k\mathbf{d}_k}{\mathbf{d}_k^{\mathrm{T}}\hat{\mathbf{H}}_k\mathbf{d}_k}.$$

Thus, for both of these methods, we have

$$\beta_k \le \frac{M\,\|\mathbf{g}_{k+1}\|\,\|\mathbf{d}_k\|}{m\,\|\mathbf{d}_k\|^2} \le \frac{M\,\|\mathbf{g}_{k+1}\|}{m\,\|\mathbf{d}_k\|} \le \frac{M\,\|\mathbf{g}_{k+1}\|^2}{m\,\|\mathbf{d}_k\|\,\|\mathbf{g}_{k+1}\|}.$$

Plugging into (44) yields

$$g_{k+1}^T d_{k+1} \le \varepsilon\, g_{k+1}^T g_{k+1} < 0.$$

Thus we will assume that, for methods HS and D,

$$g_{k+1}^T d_k \le \frac{\varepsilon m}{M} \| g_{k+1} \| \| d_k \|. \tag{45}$$

For PR, we take a different approach. From Section 2.2.2. we saw that

$$\theta_k (g_{k+1} - g_k)^T g_{k+1} \ge \frac{\| g_{k+1} \|^2 \| g_k \|^2}{g_k^T d_k}$$

is a sufficient condition for descent for the PR method. Here we will strengthen this condition to

$$\theta_k (g_{k+1} - g_k)^T g_{k+1} > C_6 \frac{\| g_k \|^2 \| g_{k+1} \|^2}{g_k^T d_k} \tag{46}$$

where $C_6 > 1$. ////

Condition (46) is stronger than is necessary to guarantee descent for the PR method. However, we will need this condition to finish the proof of Corollary 3.14. A less stringent descent condition exists, but we were not able to prove the corollary without condition (46). We will assume that these conditions are true for the rest of this section.

To finish the proof of n-step quadratic convergence for algorithm 3.2 with the D, HS and PR forms of β_k, we need to verify i), ii), iii*) and Claim 3.12. The first three are obvious; the proofs of i), ii) and iii*) follow directly from the proofs of i), ii), iii) and iii*) in Theorems 2.17 and 3.5. Lastly, we present as a corollary the alternate version of Claim 3.12 necessary to complete the proof.

(3.14) COROLLARY. *If the conjugate gradient algorithm 3.2 (with the form of* β_k *changed to that of the D, HS or PR methods,) is used to minimize a function which satisfies (AS1W), (AS2) and (AS3), then for* $k = 0, 1, \ldots, n-1$,

$$|\theta_k| \leq K \, \| \mathbf{g}_0 \|$$

for K *independent of* k *and* $\mathbf{x}_0$ *in a neighborhood of* $\mathbf{x}^*$.

PROOF OF COROLLARY 3.14: For methods D and HS, induction is not needed. Let $\phi(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{d}_k)$; then

$$\phi'(\alpha_k^*) = 0 \quad \text{and} \quad \phi'(\alpha_k) = \mathbf{g}_{k+1}^T \mathbf{d}_k.$$

By Taylor's Theorem, we know

$$\phi'(\alpha_k) = \phi'(\alpha_k^*) + \int_{\alpha_k^*}^{\alpha_k} \phi''(t)\,dt.$$

So

$$|\mathbf{g}_{k+1}^T \mathbf{d}_k| = \left| \int_{\alpha_k^*}^{\alpha_k} \phi''(t)\,dt \right|.$$

By (AS2) and Lemma 3.10, we have

$$|\mathbf{g}_{k+1}^T \mathbf{d}_k| \leq |(\alpha_k - \alpha_k^*) M \, \| \mathbf{d}_k \|^2| = |\alpha_k - \alpha_k^*| M \, \| \mathbf{d}_k \|^2$$

$$\leq C_5 M \, \| \mathbf{g}_k \|^2 \, \| \mathbf{d}_k \| \leq M C_4 C_5 \, \| \mathbf{g}_0 \| \, \| \mathbf{g}_k \|^2.$$

Then by Lemma 3.9 and the definition of θ_k, we have

$$|\theta_k| = \frac{|\mathbf{g}_{k+1}^T \mathbf{d}_k|}{|\mathbf{g}_k^T \mathbf{d}_k|} \leq \frac{C_4 C_5 M \, \| \mathbf{g}_k \|^2 \, \| \mathbf{g}_0 \|}{|\mathbf{g}_k^T \mathbf{d}_k|}.$$

But

$$|\mathbf{g}_k^T\mathbf{d}_k| = |\,\|\mathbf{g}_k\|^2 + \beta_{k-1}\mathbf{g}_k^T\mathbf{d}_{k-1}|.$$

We know from the proof of Lemma 3.13 that for D and HS

$$\beta_{k-1} = \frac{\mathbf{g}_k^T\mathbf{H}\mathbf{d}_{k-1}}{\mathbf{d}_{k-1}^T\mathbf{H}\mathbf{d}_{k-1}} \geq \frac{m\,\|\mathbf{g}_k\|\,\|\mathbf{d}_{k-1}\|}{M\,\|\mathbf{d}_{k-1}\|^2}$$

where $\mathbf{H}$ is $\mathbf{H}_{k+1}$ for D and $\hat{\mathbf{H}}_k$ for HS. Thus

$$|\mathbf{g}_k^T\mathbf{d}_k| \geq |\,\|\mathbf{g}_k\|^2 + \frac{m\,\|\mathbf{g}_k\|(\mathbf{g}_k^T\mathbf{d}_{k-1})}{M\,\|\mathbf{d}_{k-1}\|}|$$

$$= \|\mathbf{g}_k\|^2|\,1 + \frac{m\,(\mathbf{g}_k^T\mathbf{d}_{k-1})}{M\,\|\mathbf{g}_k\|\,\|\mathbf{d}_{k-1}\|}|.$$

But

$$|\frac{m\,(\mathbf{g}_k^T\mathbf{d}_{k-1})}{M\,\|\mathbf{g}_k\|\,\|\mathbf{d}_{k-1}\|}| \leq \frac{m\,\|\mathbf{g}_k\|\,\|\mathbf{d}_{k-1}\|}{M\,\|\mathbf{g}_k\|\,\|\mathbf{d}_{k-1}\|} < 1;$$

thus

$$|\,1 + \frac{m\,\mathbf{g}_k^T\mathbf{d}_{k-1}}{M\,\|\mathbf{g}_k\|\,\|\mathbf{d}_{k-1}\|}| \geq 1 - \frac{m\,|\mathbf{g}_k^T\mathbf{d}_{k-1}|}{M\,\|\mathbf{g}_k\|\,\|\mathbf{d}_{k-1}\|} \geq 1 - \frac{m}{M}$$

which gives

$$|\mathbf{g}_k^T\mathbf{d}_k| \geq \|\mathbf{g}_k\|^2(1 - \frac{m}{M}).$$

Thus

$$|\theta_k| \leq \frac{MC_4C_5\|\mathbf{g}_0\|}{(1 - \frac{m}{M})}.$$

So

$$|\theta_k| \leq K\,\|\mathbf{g}_0\|$$

for K independent of k. We have completed the proof of the corollary for methods D and HS. It remains to present the proof for PR.

For PR, we have

$$|\theta_k| = \frac{|\mathbf{g}_{k+1}^T\mathbf{d}_k|}{|\mathbf{g}_k^T\mathbf{d}_k|} \le \frac{MC_4C_5\|\mathbf{g}_k\|^2\|\mathbf{g}_0\|}{|\mathbf{g}_k^T\mathbf{d}_k|}.$$

From the definition of $\mathbf{d}_k$,

$$|\mathbf{g}_k^T\mathbf{d}_k| = |-\|\mathbf{g}_k\|^2 + \beta_{k-1}\mathbf{g}_k^T\mathbf{d}_{k-1}| = \|\mathbf{g}_k\|^2 \left| \frac{\beta_{k-1}\mathbf{g}_k^T\mathbf{d}_{k-1}}{\|\mathbf{g}_k\|^2} - 1 \right|.$$

From the definition of β_k,

$$\frac{\beta_{k-1}\mathbf{g}_k^T\mathbf{d}_{k-1}}{\|\mathbf{g}_k\|^2} = \frac{(\mathbf{g}_k - \mathbf{g}_{k-1})^T\mathbf{g}_k\,\theta_{k-1}\mathbf{g}_{k-1}^T\mathbf{d}_{k-1}}{\|\mathbf{g}_k\|^2\|\mathbf{g}_{k-1}\|^2}$$

$$> \frac{C_6\mathbf{g}_{k-1}^T\mathbf{d}_{k-1}\|\mathbf{g}_k\|^2\|\mathbf{g}_{k-1}\|^2}{\mathbf{g}_{k-1}^T\mathbf{d}_{k-1}\,\|\mathbf{g}_k\|^2\|\mathbf{g}_{k-1}\|^2} > C_6 > 1,$$

by (46). Thus

$$\left| \frac{\beta_{k-1}\mathbf{g}_k^T\mathbf{d}_{k-1}}{\|\mathbf{g}_k\|^2} - 1 \right| = \varepsilon > 0,$$

which implies that

$$|\mathbf{g}_k^T\mathbf{d}_k| = \varepsilon\|\mathbf{g}_k\|^2$$

or

$$|\theta_k| \le \frac{MC_4C_5}{\varepsilon}\|\mathbf{g}_0\|. \qquad ////$$

Thus we have proven Corollary 3.14 for methods D, HS and PR.

In this chapter, we have shown that the conjugate gradient methods D, FR, HS and PR are locally n-step quadratically convergent even when the exact line searches are replaced by any method which satisfies the condition

$$\| (\alpha_k - \alpha_k^*)\mathbf{d}_k \| \leq c \, \| \alpha_k^* \mathbf{d}_k \|^2.$$

Chapter 4
EXPLICIT DANIEL METHOD

The four conjugate gradient methods analyzed in Chapters 1, 2 and 3, namely D, FR, HS and PR, are methods in which the technique for finding α_k of the line search is not the defining factor. These methods are "implicit" in some sense, because α_k is found via minimization. In 1967, Daniel [4] proposed a different, "explicit" method, which we will call the Explicit Daniel Method (ED). Here the line search is eliminated, and α_k is explicitly given by

$$\alpha_k = \frac{-\mathbf{g}_k^T \mathbf{d}_k}{\mathbf{d}_k^T \mathbf{H}_k \mathbf{d}_k}. \tag{47}$$

Daniel attempted a proof of the local n-step quadratic convergence of this method with no success [4, 5, 6]. In this chapter, we will show that this form of α_k implies the condition on the line search in Algorithm 3.2, and thus is locally n-step quadratically convergent with any of the choices of β_k.

To show that this version of α_k fits into our hypothesis on the line search:

$$\| (\alpha_k - \alpha_k^*) \mathbf{d}_k \| \le c \, \| \alpha_k^* \mathbf{d}_k \|^2,$$

we need to assume (AS1) instead of (AS1W), i.e., we need three continuous derivatives. Note that if $\phi(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{d}_k)$, then (47) becomes

$$\alpha_k = \frac{-\phi'(0)}{\phi''(0)}.$$

Subtracting α_k^* from both sides gives

$$\alpha_k - \alpha_k^* = \frac{\phi'(0)}{\phi''(0)} - \alpha_k^*.$$

If we expand $\phi'(0)$ in a Taylor Series around α_k^*, we get

$$\phi'(0) = \phi'(\alpha_k^*) - \alpha_k^* \phi''(\lambda\alpha_k^*),$$

where $\lambda \in [0,1]$. This implies that

$$\alpha_k - \alpha_k^* = \alpha_k^* \left[\frac{\phi''(\lambda\alpha_k^*)}{\phi''(0)} - 1 \right]$$

or

$$(\alpha_k - \alpha_k^*)\phi''(0) = \alpha_k^*(\phi''(\lambda\alpha_k^*) - \phi''(0)).$$

But we can also expand $\phi''(0)$ in a Taylor Series around $\lambda\alpha_k^*$ to give

$$\phi''(0) = \phi''(\lambda\alpha_k^*) - \lambda\alpha_k^*\phi'''(\eta),$$

where $\eta = \delta(\lambda\alpha_k^*)$, $\delta \in [0,1]$. So

$$\begin{aligned}(\alpha_k - \alpha_k^*)\phi''(0) &= \alpha_k^*(\phi''(\lambda\alpha_k^*) - \phi''(\lambda\alpha_k^*) + \lambda\alpha_k^*\phi'''(\eta)) \\ &= \lambda(\alpha_k^*)^2\phi'''(\eta).\end{aligned}$$

Taking absolute values of both sides and applying (AS2) gives

$$m\,|\,\alpha_k - \alpha_k^*\,|\,\|\,\mathbf{d}_k\,\|^2 \le \lambda(\alpha_k^*)^2\,|\,\phi'''(\eta)\,|. \tag{48}$$

To bound $|\,\phi'''(\eta)\,|$, consider

$$\begin{aligned}|\,\phi'''(\eta)\,| &= \Big|\sum_{i,j,l=1}^{n} \frac{\partial^3 f(\mathbf{x}_k + \eta\mathbf{d}_k)}{\partial x_i \partial x_j \partial x_l} \cdot d_i d_j d_l\,\Big| \\ &\le \sum_{i,j,l} \Big|\frac{\partial^3 f(\mathbf{x}_k + \eta\mathbf{d}_k)}{\partial x_i \partial x_j \partial x_l}\Big| \cdot |\,d_i\,|\,|\,d_j\,|\,|\,d_l\,|,\end{aligned}$$

assuming that $\mathbf{d}_k = (d_1, d_2, \ldots, d_n)^{\mathrm{T}}$ and $\mathbf{x}_k = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$. But we know that $|\,d_i\,| \le \|\,\mathbf{d}_k\,\|$, and since (AS1) gives

$$\left| \frac{\partial^3 f(\mathbf{x}_k + \eta \mathbf{d}_k)}{\partial x_i \, \partial x_j \, \partial x_l} \right| \le \nu < \infty,$$

we have

$$|\phi'''(\eta)| \le 3n\nu \| \mathbf{d}_k \|^3.$$

Plugging into (48) gives

$$|\alpha_k - \alpha_k^*| \| \mathbf{d}_k \|^2 \le \frac{\lambda}{m} (\alpha_k^*)^2 3n\nu \| \mathbf{d}_k \|^3$$

or

$$|\alpha_k - \alpha_k^*| \| \mathbf{d}_k \| \le \frac{3n\nu\lambda}{m} (\alpha_k^*)^2 \| \mathbf{d}_k \|^2 = c \, \| \alpha_k^* \mathbf{d}_k \|^2,$$

which is exactly our condition on the line search. Therefore, the form of α_k from method ED satisfies the conditions of Algorithm 3.2 and Theorem 3.5, and thus converges locally n-step quadratically.

Chapter 5
COMPUTATIONAL RESULTS

We perform the conjugate gradient method using the following estimate for α_k of the line search. Let $\phi(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{d}_k)$, and s be such that $0 < s \leq 1$ with $\phi(s) \leq \phi(0)$, i.e., $f(\mathbf{x}_k + s\,\mathbf{d}_k) \leq f(\mathbf{x}_k)$. Remember that $\phi'(0) < 0$. Thus the quadratic, p, that interpolates $\phi(\alpha)$ at $\alpha = 0$ and $\alpha = s$ with $\phi'(0) = p'(0)$, has a local minimizer near the minimizer, α^*, of the line search. This local minimizer, $\hat{\alpha}_k$, will be our estimate for α_k.

To derive a formula for $\hat{\alpha}_k$ consider

$$p(\alpha_k) = a\,\alpha_k^2 + b\,\alpha_k + c$$

then

$$p(0) = \phi(0) = f(\mathbf{x}_k) = c$$

$$p'(0) = \phi'(0) = \mathbf{g}_k^T \mathbf{d}_k = b$$

$$p(s) = \phi(s) = f(\mathbf{x}_k + s\,\mathbf{d}_k) = a\,s^2 + b\,s + c$$

$$= a\,s^2 + s\,\mathbf{g}_k^T \mathbf{d}_k + f(\mathbf{x}_k)$$

which implies that

$$a = \frac{f(\mathbf{x}_k + s\,\mathbf{d}_k) - s\,\mathbf{g}_k^T \mathbf{d}_k - f(\mathbf{x}_k)}{s^2}.$$

The minimum of $p(\alpha)$ occurs at $\alpha = \dfrac{-b}{2a}$, thus:

$$\hat{\alpha}_k = \frac{-s^2 \mathbf{g}_k^T \mathbf{d}_k}{2(f(\mathbf{x}_k + s\,\mathbf{d}_k) - f(\mathbf{x}_k) - s\,\mathbf{g}_k^T \mathbf{d}_k)}. \tag{49}$$

Algorithm 3.2 with α_k calculated directly using (49) was used to solve the five problems given below. The properties of each problem are summarized, including information on whether conditions (AS1W), (AS2) and (AS3) are satisfied, as well as starting points and exact solutions. The results for each of the choices of β_k are summarized, where an iteration is a block of n conjugate gradient steps. The number of steps performed in the last iteration is recorded in parentheses. Computations were performed in Fortran77 on an IBM 3081 in double precision. The stopping criterion was chosen to be $\| g_k \|_\infty = \| \nabla f(\mathbf{x}_k) \|_\infty \le 10^{-8}$.

(5.1) PROBLEM. (Perturbed Quadratic)

$$f(\mathbf{x}_k) = \mathbf{x}_k^T \mathbf{D} \mathbf{x}_k + \sum_{i=1}^{10} x_i^4$$

where **D** is the diagonal matrix *diag* [1, 2, ..., 10], and x_i is component i of $\mathbf{x}_k$.

n	AS1W	AS2	AS3	$\mathbf{x}_0$	$\mathbf{x}^*$	$f(\mathbf{x}^*)$
10	yes	yes	yes	$(1,1,\ldots,1)^T$	$(0,0,\ldots,0)^T$	0

D Algorithm.

iteration	$f(\mathbf{x}_k)$	$\| g_k \|_\infty$	$\| \mathbf{x}_k - \mathbf{x}^* \|_\infty$
0	65.0	24.0	1.0
1	0.5635×10^{-5}	0.4607×10^{-2}	0.1751×10^{-2}
2(10)	0.8983×10^{-20}	0.4658×10^{-9}	0.5963×10^{-10}

ED Algorithm.

iteration	$f(\mathbf{x}_k)$	$\| g_k \|_\infty$	$\| \mathbf{x}_k - \mathbf{x}^* \|_\infty$
0	65.0	24.0	1.0
1	0.5933×10^{-5}	0.7189×10^{-2}	0.8942×10^{-3}
2(10)	0.1275×10^{-20}	0.2224×10^{-9}	0.1112×10^{-10}

FR Algorithm.

iteration	$f(\mathbf{x}_k)$	$\| g_k \|_\infty$	$\| \mathbf{x}_k - \mathbf{x}^* \|_\infty$
0	65.0	24.0	1.0
1	0.9010×10^{-5}	0.5176×10^{-2}	0.2588×10^{-2}
2(10)	0.2095×10^{-17}	0.5691×10^{-8}	0.1134×10^{-8}

HS Algorithm.

iteration	$f(\mathbf{x}_k)$	$\| g_k \|_\infty$	$\| \mathbf{x}_k - \mathbf{x}^* \|_\infty$
0	65.0	24.0	1.0
1	0.1667×10^{-5}	0.3690×10^{-2}	0.7451×10^{-3}
2(10)	0.2520×10^{-21}	0.7928×10^{-10}	0.9725×10^{-11}

PR Algorithm.

iteration	$f(\mathbf{x}_k)$	$\lVert g_k \rVert_\infty$	$\lVert \mathbf{x}_k - \mathbf{x}^* \rVert_\infty$
0	65.0	24.0	1.0
1	0.9704×10^{-6}	0.3285×10^{-2}	0.6116×10^{-3}
2(10)	0.1318×10^{-21}	0.5846×10^{-10}	0.6782×10^{-11}

(5.2) PROBLEM. (Perturbed Quadratic)

$$f(\mathbf{x}_k) = \mathbf{x}_k^T \mathbf{D} \mathbf{x}_k + \sum_{i=1}^{10} x_i^4$$

where **D** is now the diagonal matrix *diag* [1, 2, 3, 40, 50, 60, 700, 800, 900, 1000].

n	AS1W	AS2	AS3	$\mathbf{x}_0$	$\mathbf{x}^*$	$f(\mathbf{x}^*)$
10	yes	yes	yes	$(1,1,\ldots,1)^T$	$(0,0,\ldots,0)^T$	0

D Algorithm.

iteration	$f(\mathbf{x}_k)$	$\Vert g_k \Vert_\infty$	$\Vert \mathbf{x}_k - \mathbf{x}^* \Vert_\infty$
0	3566.0	2004.0	1.0
4	0.1659×10^{-3}	0.4057	0.5006×10^{-2}
5	0.2063×10^{-7}	0.3714×10^{-2}	0.8161×10^{-4}
6	0.5430×10^{-11}	0.5840×10^{-5}	0.1593×10^{-5}
7(10)	0.2298×10^{-21}	0.7523×10^{-9}	0.5373×10^{-12}

ED Algorithm.

iteration	$f(\mathbf{x}_k)$	$\Vert g_k \Vert_\infty$	$\Vert \mathbf{x}_k - \mathbf{x}^* \Vert_\infty$
0	3566.0	2004.0	1.0
5	0.7685×10^{-7}	0.8511×10^{-3}	0.1418×10^{-3}
6	0.1737×10^{-9}	0.5710×10^{-3}	0.5615×10^{-5}
7	0.5820×10^{-18}	0.3953×10^{-7}	0.2196×10^{-10}
8(1)	0.5972×10^{-20}	0.3718×10^{-8}	0.2655×10^{-11}

FR Algorithm.

iteration	$f(\mathbf{x}_k)$	$\| g_k \|_\infty$	$\| \mathbf{x}_k - \mathbf{x}^* \|_\infty$
0	3566.0	2004.0	1.0
5	0.2529×10^{-6}	0.3677×10^{-2}	0.3112×10^{-3}
6	0.9360×10^{-9}	0.3009×10^{-3}	0.1750×10^{-4}
7	0.6298×10^{-16}	0.4154×10^{-6}	0.2597×10^{-9}
8(2)	0.2387×10^{-20}	0.1435×10^{-8}	0.2089×10^{-10}

HS Algorithm.

iteration	$f(\mathbf{x}_k)$	$\| g_k \|_\infty$	$\| \mathbf{x}_k - \mathbf{x}^* \|_\infty$
0	3566.0	2004.0	1.0
4	0.2390×10^{-5}	0.5093×10^{-2}	0.9945×10^{-3}
5	0.1890×10^{-8}	0.1377×10^{-2}	0.7013×10^{-5}
6	0.4963×10^{-16}	0.3918×10^{-6}	0.2450×10^{-10}
7(2)	0.2018×10^{-20}	0.7698×10^{-9}	0.2648×10^{-10}

PR Algorithm.

iteration	$f(\mathbf{x}_k)$	$\|g_k\|_\infty$	$\|\mathbf{x}_k - \mathbf{x}^*\|_\infty$
0	3566.0	2004.0	1.0
5	0.2862×10^{-6}	0.1154×10^{-1}	0.3223×10^{-3}
6	0.8248×10^{-9}	0.5828×10^{-4}	0.2230×10^{-4}
7	0.3273×10^{-16}	0.2461×10^{-6}	0.1538×10^{-9}
8(2)	0.5954×10^{-20}	0.2743×10^{-8}	0.1273×10^{-10}

(5.3) PROBLEM. (Davidon's Simple Quadratic [7])

$$f(\mathbf{x}_k) = x_1^2 - 2x_1x_2 + 2x_2^2$$

where x_i is component i of $\mathbf{x}_k$.

n	AS1W	AS2	AS3	$\mathbf{x}_0$	$\mathbf{x}^*$	$f(\mathbf{x}^*)$
2	yes	yes	yes	$(-4,2)^T$	$(0,0)^T$	0

D Algorithm.

iteration	$f(\mathbf{x}_k)$	$\|g_k\|_\infty$	$\|\mathbf{x}_k - \mathbf{x}^*\|_\infty$
0	40.0	16.0	1.0
1(2)	0.3944×10^{-30}	0.8882×10^{-15}	0.8882×10^{-15}

ED Algorithm.

iteration	$f(\mathbf{x}_k)$	$\|g_k\|_\infty$	$\|\mathbf{x}_k - \mathbf{x}^*\|_\infty$
0	40.0	16.0	1.0
1(2)	0.4930×10^{-31}	0.4441×10^{-15}	0.2220×10^{-15}

FR Algorithm.

iteration	$f(\mathbf{x}_k)$	$\|g_k\|_\infty$	$\|\mathbf{x}_k - \mathbf{x}^*\|_\infty$
0	40.0	16.0	1.0
1(2)	0.2021×10^{-29}	0.3997×10^{-14}	0.1110×10^{-14}

HS Algorithm.

iteration	$f(\mathbf{x}_k)$	$\|\mathbf{g}_k\|_\infty$	$\|\mathbf{x}_k - \mathbf{x}^*\|_\infty$
0	40.0	16.0	1.0
1(2)	0.7889×10^{-30}	0.1776×10^{-14}	0.8882×10^{-15}

PR Algorithm.

iteration	$f(\mathbf{x}_k)$	$\|\mathbf{g}_k\|_\infty$	$\|\mathbf{x}_k - \mathbf{x}^*\|_\infty$
0	40.0	16.0	1.0
1(2)	0.7889×10^{-30}	0.1776×10^{-14}	0.8882×10^{-15}

(5.4) PROBLEM. (Rosenbrock [7])

$$f(\mathbf{x}_k) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

where x_i is component i of $\mathbf{x}_k$.

n	AS1W	AS2	AS3	$\mathbf{x}_0$	$\mathbf{x}^*$	$f(\mathbf{x}^*)$
2	yes	no	yes	$(-1.2,1)^T$	$(1,1)^T$	0

D Algorithm.

iteration	$f(\mathbf{x}_k)$	$\lVert g_k \rVert_\infty$	$\lVert \mathbf{x}_k - \mathbf{x}^* \rVert_\infty$
0	24.2	215.6	2.2
10	0.1273×10^{-2}	1.3789	0.1687×10^{-1}
11	0.4011×10^{-6}	0.2521×10^{-1}	0.1386×10^{-3}
12	0.1139×10^{-14}	0.1343×10^{-5}	0.7621×10^{-8}
13(1)	0.1204×10^{-16}	0.2772×10^{-8}	0.6949×10^{-8}

ED Algorithm.

iteration	$f(\mathbf{x}_k)$	$\lVert g_k \rVert_\infty$	$\lVert \mathbf{x}_k - \mathbf{x}^* \rVert_\infty$
0	24.2	215.6	2.2
14	0.6163×10^{-2}	3.1631	0.1830×10^{-1}
15	0.1583×10^{-5}	0.5030×10^{-1}	0.1450×10^{-3}
16	0.1240×10^{-13}	0.4454×10^{-5}	0.1100×10^{-7}
17(1)	0.1922×10^{-16}	0.3504×10^{-8}	0.8779×10^{-8}

FR Algorithm.

iteration	$f(\mathbf{x}_k)$	$\| \mathbf{g}_k \|_\infty$	$\| \mathbf{x}_k - \mathbf{x}^* \|_\infty$
0	24.2	215.6	2.2
15	0.6673×10^{-5}	0.3799×10^{-1}	0.4802×10^{-2}
16	0.7573×10^{-8}	0.3474×10^{-2}	0.1403×10^{-4}
17	0.7543×10^{-13}	0.1098×10^{-4}	0.3014×10^{-7}
18(1)	0.1516×10^{-15}	0.9841×10^{-8}	0.2465×10^{-7}

HS Algorithm.

iteration	$f(\mathbf{x}_k)$	$\| \mathbf{g}_k \|_\infty$	$\| \mathbf{x}_k - \mathbf{x}^* \|_\infty$
0	24.2	215.60	2.2
6	0.1554×10^{-1}	4.4786	0.8781×10^{-1}
7	0.1883×10^{-3}	0.5435	0.3867×10^{-2}
8	0.1052×10^{-9}	0.3943×10^{-3}	0.5774×10^{-5}
9(2)	0.2940×10^{-20}	0.2152×10^{-8}	0.1469×10^{-10}

PR Algorithm.

iteration	$f(\mathbf{x}_k)$	$\| g_k \|_\infty$	$\| \mathbf{x}_k - \mathbf{x}^* \|_\infty$
0	24.2	215.60	2.2
6	0.1160×10^{-1}	3.8620	0.7897×10^{-1}
7	0.1217×10^{-3}	0.4372	0.3055×10^{-2}
8	0.2937×10^{-10}	0.2048×10^{-3}	0.3596×10^{-5}
9(2)	0.5224×10^{-21}	0.9080×10^{-9}	0.5793×10^{-11}

(5.5) PROBLEM. (Wood [7])

$$f(\mathbf{x}_k) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 90(x_4 - x_3^2)^2 + (1 - x_3)^2$$
$$+ 10.1[(x_2 - 1)^2 + (x_4 - 1)^2] + 19.8(x_2 - 1)(x_4 - 1)$$

where x_i is component i of $\mathbf{x}_k$.

n	AS1W	AS2	AS3	$\mathbf{x}_0$	$\mathbf{x}^*$	$f(\mathbf{x}^*)$
4	yes	no	yes	$(0,0,0,0)^T$	$(1,1,1,1)^T$	0

D Algorithm.

iteration	$f(\mathbf{x}_k)$	$\lVert g_k \rVert_\infty$	$\lVert \mathbf{x}_k - \mathbf{x}^* \rVert_\infty$
0	42.0	40.0	1.0
1	27.9301	100.8402	0.8467
4	0.1417×10^{-3}	0.4648	0.1405×10^{-2}
5	0.1912×10^{-5}	0.1704×10^{-2}	0.1465×10^{-2}
6	0.2639×10^{-10}	0.1402×10^{-3}	0.2373×10^{-5}
7(4)	0.2707×10^{-19}	0.6050×10^{-8}	0.7678×10^{-11}

ED Algorithm.

iteration	$f(\mathbf{x}_k)$	$\lVert g_k \rVert_\infty$	$\lVert \mathbf{x}_k - \mathbf{x}^* \rVert_\infty$
0	42.0	40.0	1.0
14	0.7332×10^{-5}	0.8550×10^{-1}	0.1713×10^{-2}
15	0.7929×10^{-8}	0.3557×10^{-2}	0.5733×10^{-5}
16	0.4104×10^{-14}	0.2557×10^{-5}	0.2294×10^{-8}
17(1)	0.8791×10^{-18}	0.1029×10^{-8}	0.9965×10^{-9}

FR Algorithm.

iteration	$f(\mathbf{x}_k)$	$\| \mathbf{g}_k \|_\infty$	$\| \mathbf{x}_k - \mathbf{x}^* \|_\infty$
0	42.0	40.0	1.0
1	34.4408	64.8834	1.0100
9	0.4059×10^{-5}	0.7347×10^{-1}	0.5603×10^{-3}
10	0.1118×10^{-7}	0.3891×10^{-2}	0.2677×10^{-4}
11	0.9400×10^{-13}	0.1161×10^{-4}	0.1076×10^{-7}
12(3)	0.1675×10^{-16}	0.4206×10^{-8}	0.4345×10^{-8}

HS Algorithm.

iteration	$f(\mathbf{x}_k)$	$\| \mathbf{g}_k \|_\infty$	$\| \mathbf{x}_k - \mathbf{x}^* \|_\infty$
0	42.0	40.0	1.0
1	35.9168	96.9042	1.0312
5	0.1571×10^{-3}	0.1528	0.1253×10^{-1}
6	0.1462×10^{-7}	0.1019×10^{-2}	0.3703×10^{-4}
7	0.3268×10^{-13}	0.4719×10^{-5}	0.1938×10^{-7}
8(3)	0.5011×10^{-17}	0.1704×10^{-8}	0.2361×10^{-8}

PR Algorithm.

iteration	$f(\mathbf{x}_k)$	$\| g_k \|_\infty$	$\| \mathbf{x}_k - \mathbf{x}^* \|_\infty$
0	42.0	40.0	1.0
1	34.1806	61.0045	1.0031
8	0.4937×10^{-7}	0.4396×10^{-2}	0.2015×10^{-3}
9	0.1760×10^{-10}	0.1541×10^{-3}	0.2130×10^{-6}
10	0.7527×10^{-16}	0.3266×10^{-6}	0.7025×10^{-9}
11(1)	0.4299×10^{-18}	0.8561×10^{-8}	0.6473×10^{-9}

To summarize, Problems 1, 2 and 3 all satisfy (AS1W), (AS2) and (AS3), and all three converge locally n-step quadratically. Problem 2 requires more iterations than Problem 1 because the neighborhood of convergence is smaller due to the ill conditioning of the matrix D. Problems 4 and 5 both fail to satisfy (AS2), but Problem 4 converges locally n-step quadratically and Problem 5 converges better than linearly.

The same problems were run through the same algorithms without restart, and all of the problems had difficulties. Problem 4 failed to converge after 150 iterations, and the other problems all took a significantly greater number of iterations to converge.

When individual methods are compared with each other, the HS method usually converges in fewer iterations than the other methods on these five test problems.

REFERENCES

1. Bertsekas, D. P., *Constrained Optimization and Multiplier Methods,* Academic Press, New York, 1982.

2. Broyden, C. G., J. E. Dennis, Jr., J. J. Moré, *On the Local and Superlinear Convergence of Quasi-Newton Methods,* J. IMA **12** (1973), p. 223.

3. Cohen, A. I., *Rate of Convergence of Several Conjugate Gradient Methods,* SIAM J. Num. Anal. **9** (1979), p. 248.

4. Daniel, J. W., *The Conjugate Gradient Method for Linear and Nonlinear Operator Equations,* SIAM J. Num. Anal. **4** (1967), p. 10.

5. __________, *Convergence of the Conjugate Gradient Method with Computationally Convenient Modifications,* Num. Math. **10** (1967), p. 125.

6. __________, *A Correction Concerning the Convergence Rate for the Conjugate Gradient Method,* SIAM J. Num. Anal. **7** (1970), p. 277.

7. Fletcher, R., M. J. D. Powell, *A Rapidly Convergent Descent Method for Minimization,* Comp. J. **6** (1963), p. 163.

8. Fletcher, R., C. M. Reeves, *Function Minimization by Conjugate Gradients,* Comp. J. **7** (1964), p. 149.

9. Hager, W. W., *Approximations to the Multiplier Method,* SIAM J. Num. Anal. **22** (1985), p. 16.

10. __________, *Dual Techniques for Constrained Optimization,* J. Optim. Theory Appl. **55** (1987), p. 37.

11. __________, *A Derivative Based Bracketing Scheme for Univariate Minimization,* to appear in Comp. Math. Appl.

12. Hestenes, M., E. Stiefel, *Method of Conjugate Gradients for Solving Linear Systems,* J. Res. Nat. B. of Stds. **49** (1952), p. 409.

13. Lenard, M. L., *Practical Convergence Conditions for Unconstrained Minimization,* Math. Prog. **4** (1973), p. 309.

14. __________, *Convergence Conditions for Restarted Conjugate Gradient Methods with Inaccurate Line Searches,* Math. Prog. **10** (1976), p. 32.

15. Luenberger, D. G., *Introduction to Linear and Nonlinear Programming,* Second Edition, Addison Wesley, New York, 1984.

16. Polak, E., G. Ribiere, *Note Sur la Convergence de Méthodes de directions Conjugées,* Revue Français d'Automatique Info. et Recherche Oper. **16-R1** (1969), p. 35.

17. Polyak, B. T., *The Conjugate Gradient Method in Extremal Problems,* USSR Comp. Math. and Math. Phys. **9** (1969), p. 94.

18. Powell, M. J, D., *On the Convergence of the Variable Metric Algorithm,* J. IMA **7** (1971), p. 21.