A Convergence Theory for the Structured
BFGS Secant Method with an Application
to Nonlinear Least Squares[1,2,3]

by

J.E. Dennis[4]
H.J. Martinez[5]
R.A. Tapia[4]

Technical Report 87-15, May 1987
(Revised July, 1988)
TO APPEAR JOTA 61(1989), 159 – 176

---

[1]Research sponsored by SDIO/IST/ARO, AFOSR 85-0243, and DOE DEFG05-86ER 25017.

[2]A portion of this work is contained in the second author's doctoral thesis under the supervision of the other two authors in the Department of Mathematical Sciences, Rice University. The second author would like to thank Universidad del Valle for support during his graduate studies.

[3]An early draft of this work was presented at the SIAM 35th Anniversary Meeting, October 12-15, 1987, Denver, CO.

[4]Professors, Department of Mathematical Sciences, Rice University, Houston. TX 77251-1892.

[5]Assistant Professor, Departamento de Matemáticas, Universidad del Valle, A.A. 2188, Cali, Colombia.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **JUL 1988**	2. REPORT TYPE	3. DATES COVERED **00-00-1988 to 00-00-1988**

4. TITLE AND SUBTITLE **A Convergence Theory for th eStructured BFGS Secant Method with an Application to Nonlinear Least Squares**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Computational and Applied Mathematics Department ,Rice University,6100 Main Street MS 134,Houston,TX,77005-1892**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**31**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

A Convergence Theory for the Structured BFGS Secant Method with an Application to Nonlinear Least Squares[1,2,3]

J.E. Dennis Jr.[4], H.J. Martinez[5] and R.A. Tapia[4]

Abstract. In 1981, Dennis and Walker developed a convergence theory for structured secant methods which included the PSB and the DFP secant methods, but not the straightforward structured version of the BFGS secant method. Here we fill this gap in the theory by establishing a convergence theory for the structured BFGS secant method. A direct application of our new theory gives the first proof of local and q-superlinear convergence of the important structured BFGS secant method for the nonlinear least-squares problem which is used by Dennis, Gay and Welsh in the current version of the popular and successful NL2SOL code.

Key words. secant, quasi-Newton, least-squares, superlinear convergence, bounded deterioration.

1980 Mathematics Subject Classification: Primary 49D15, G5K05.

---

[1]Research sponsored by SDIO/IST/ARO, AFOSR 85-0243, and DOE DEFG05-86ER 25017.

[2]A portion of this work is contained in the second author's doctoral thesis under the supervision of the other two authors in the Department of Mathematical Sciences, Rice University. The second author would like to thank Universidad del Valle for support during his graduate studies.

[3]An early draft of this work was presented at the SIAM 35th Anniversary Meeting, October 12-15, 1987, Denver, CO.

[4]Professor, Department of Mathematical Sciences, Rice University, Houston, TX 77251-1892.

[5]Assistant Professor, Departamento de Matemáticas, Universidad del Valle, A.A. 2188, Cali, Colombia.

1. Introduction.

Historically, by a secant method for the unconstrained optimization problem

$$\textit{minimize}\ f(x) \tag{1}$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, we mean the iterative procedure

$$\begin{aligned} x_+ &= x + s \\ B_+ &= \mathbb{B}(x, s, y, B) \end{aligned} \tag{2}$$

where s and y are defined by

$$Bs = -\nabla f(x) \tag{3}$$

$$y = \nabla f(x_+) - \nabla f(x) \tag{4}$$

and the update B_+ must satisfy the secant equation

$$B_+ s = y\ . \tag{5}$$

We interpret B_+ as an approximation to $\nabla^2 f(x_+)$ and y as an approximation to $\nabla^2 f(x_+)s$. Most interesting secant updates can be written in the form

$$B_+ = B + \Delta(s, y, B, v) \tag{6}$$

where

$$\Delta(s, y, B, v) = \frac{(y - Bs)v^T + v(y - Bs)^T}{v^T s} - \frac{(y - Bs)^T s}{(v^T s)^2} vv^T \tag{7}$$

for some choice of the vector v. Following Dennis and Walker [Ref. 1] we call v the *scale* of the particular secant update in question. The scale v will usually depend on s, y or B as is the case for the following well-known updates:

$$PSB \qquad v = s \tag{8}$$

$$DFP \qquad v = y \tag{9}$$

$$BFGS \qquad v = y + \left(\frac{y^T s}{s^T B s} \right)^{\frac{1}{2}} Bs \ . \tag{10}$$

In what follows, we will write $v(s, y, B)$ when the use of v alone may cause confusion.

Often, in practice, a part of $\nabla^2 f(x)$ is available and we need only to approximate the remaining part. Suppose that

$$\nabla^2 f(x) = S(x) + C(x)$$

where $C\colon \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$, the available part of $\nabla^2 f(x)$, is symmetric. In several important applications, e.g. nonlinear least-squares, $C(x)$ is composed of the first-order information and $S(x)$ requires second-order information.

By a *structured approximation* of $\nabla^2 f(x)$ we mean an approximation of the form

$$B = A + C(x)$$

where A is an approximation to $S(x)$. Moreover, if B is updated according to the formula

$$B_+ = A_+ + C(x_+)$$

where

$$A_+ = A + \Delta(s, y^{\#}, A, v) \tag{11}$$

and $y^{\#}$ is an approximation to $S(x_+)s$, then we call B_+ a *structured* Δ *approximation* of $\nabla^2 f(x_+)$. Observe that the update (11) satisfies the secant equation

$$A_+ s = y^{\#} \ . \tag{12}$$

We obtain a *structured* Δ *method* for problem (1) if in (2) we use $B_+ = A_+ + C(x_+)$ where A_+ is given by (11).

Historically, the major issues one faces in a particular application of structure is the choice of $y^{\#}$ and the choice of scale v in (11). In an effort to give a choice for $y^{\#}$, which could be used when the structure did not suggest a better choice, Dennis and Walker [Ref. 1] proposed the default choice

$$y^{\#} = y - C(x_+)s$$

where y is given by (4). The rationale for the default choice is quite straightforward. This choice of $y^{\#}$ leads to an update $B_+ = A_+ + C(x_+)s$ which satisfies the standard (unstructured) form of the secant equation. To see this observe that if $A_+s = y - C(x_+)s$, then $B_+s = y$.

The primary criticism of the default choice is that it does not take full advantage of structure; the quantity $y = \nabla f(x_+) - \nabla f(x)$ does not exploit structure. It has been our experience that each application of structure suggests a choice for $y^{\#}$ which takes advantage of structure and is superior to the default choice.

While the ambiguity in the choice for $y^{\#}$ has not created serious problems in the application of structured secant methods, the ambiguity in the choice of scale has been the major detriment to development of successful structured BFGS secant methods and a general convergence theory for such methods. For this reason, we first present a fairly complete historical development of the choice of scale for structured secant methods and then present a general rule for choosing the scale in structured secant updates.

The primary application for structured secant methods has been the nonlinear least-squares problem (see Section 4). Work in this area includes Brown and Dennis [Ref. 2], Dennis [Refs. 3, 4, 5], Betts [Ref. 6], Bartholomew-Biggs [Ref. 7], Dennis and Welsch [Ref. 8], Dennis, Gay and Welsch [Ref. 9, 10], Dennis and Walker [Ref. 1], Dennis and Schnabel [Ref. 11], Al-Baali and Fletcher [Ref.

12]. Xu [Ref. 13], Mahdavi and Bartels [Ref. 14] and Toint [Ref. 15]. Several of these works considered the structured PSB update. For the PSB update the issue of the proper choice of scale does not arise, since the scale is $v = s$.

In all these works, the general problem of how should the scale be modified when one decides to utilize structure in the secant update is not considered. Indeed, it is interesting that in these works only Al-Baali and Fletcher [Ref. 12] actually carried the structure into the choice of scale. However, they gave no convergence analysis for their structured BFGS secant method. The only works that contain a convergence analysis are Dennis and Walker [Ref. 1] and Xu [Ref. 13]. Dennis and Walker, as an application of their general theory, established local and superlinear convergence for the structured PSB and DFP methods in general and for the nonlinear least-squares problem in particular. Their theory does not include structured BFGS secant methods. On the other hand, the BFGS secant method for the nonlinear least-squares problem studied by Xu in Ref. 13 (see Ref. 12) is only mildly structured in that it utilizes structure in the choice of y as an approximation to $\nabla^2 f(x_+)s$, but not in B_+. As such, q-superlinear convergence follows from the standard theory by viewing their choice for y as a perturbation of the standard unstructured choice for y.

The nonlinear least-squares problem is an important problem and the use of structure is a significant part of the formulation of any secant method for this problem. These two facts have been reinforced by the popularity and success of the NL2SOL code of Dennis, Gay and Welsh [Ref. 9, 10]. This code originally used the structured DFP secant update analyzed by Dennis and Walker in Ref. 1, but now uses the structured BFGS secant update suggested by Al-Baali and Fletcher in Ref. 12. While the authors report improved numerical results, there is no local convergence theory for the new version of the algorithm. This lack of theory played a major role in motivating the present work.

Another important application of structured secant methods was given recently by Tapia [Ref. 17]. He extended the class of secant updates given by (6)-(7) to updates for equality constrained optimization which utilize the structure present in the Hessian of the augmented Lagrangian. Local and q-superlinear convergence for the DFP and BFGS versions of these structured secant methods was established under standard assumptions.

A close look at the ingredients in Tapia's theory [Ref. 17] reveals a structure principle which we can extract and use to formulate a general rule for defining the scale in any structured secant update. This structure principle also provides an insightful way of viewing the structured secant approximation when formulating our convergence theory.

Structure Principle. Assume that $\nabla^2 f(x) = S(x) + C(x)$. Given

$$B = A + C(x)$$

as an approximation to $\nabla^2 f(x)$ we want

$$B_+ = A_+ + C(x_+)$$

as an approximation to $\nabla^2 f(x_+)$, where $x_+ = x + s$.

Compute B_+ as an update of $A + C(x)$. Toward this end consider

$$y^S = y^\# + C(x_+)s \ ,$$

as an approximation to $\nabla^2 f(x_+)s$ and let

$$B^S = A + C(x_+).$$

The secant update of B^S is

$$B_+ = B^S + \Delta(s, y^S, B^S, v(s, y^S, B^S)).$$

Now, observe that for any v

$$\Delta(s, y^S, B^S, v) = \Delta(s, y^\#, A, v)$$

so that we can write

$$B_+ = A + C(x_+) + \Delta(s, y^\#, A, v(s, y^S, B^S)).$$

It now seems reasonable to define

$$A_+ = A + \Delta(s, y^\#, A, v(s, y^S, B^S)). \tag{13}$$

and call it *the structured secant update* of A.

Remark 1.1. Clearly A_+ given by (13) satisfies the secant equation (12).

Remark 1.2. In essence the structure principle is saying that the scale should take structure and the complete problem into account.

Remark 1.3 For the PSB update the structure principle leaves the scale unchanged.

Remark 1.4. In his application [Ref. 17], Tapia calls the update A_+ which results from the structure principle the augmented scale secant update.

Remark 1.5. In the remainder of this paper, when we refer to the structured BFGS secant update, we will assume that in (7) the scale is $v(s, y^S, B^S)$ where $v(s, y, B)$ is given by (10).

Remark 1.6. For the nonlinear least-squares problem our structured BFGS secant update is the same as that suggested by Al-Baali and Fletcher in Ref. 12.

In our analysis, we will use several different matrix norms. The Frobenius norm will be denoted by $\|\cdot\|_F$, the Frobenius norm weighted by $\nabla^2 f(x_*)$ will be denoted by $\|\cdot\|_*$, i.e. $\|\cdot\|_* = \|\nabla^2 f(x_*)^{-1/2}(\cdot)\nabla^2 f(x_*)^{-1/2}\|_F$ and the l_2-operator norm will be denoted by $\|\cdot\|$. The only vector norm that will be used is the Euclidean norm, and it will be denoted by $\|\cdot\|$.

The standard assumptions for problem (1) are:

A1: Problem (1) has a local solution x_*

A2: The function $f \in C^2$, and $\nabla^2 f$ and C are locally Lipschitz continuous at x_*, i.e., there exist constants $L \geq 0$, $L_C \geq 0$ and $\epsilon_1 > 0$ such that

$$||\nabla^2 f(x) - \nabla^2 f(x_*)|| \leq L \, ||x - x_*|| \tag{14}$$

and

$$||C(x) - C(x_*)|| \leq L_C \, ||x - x_*|| \tag{15}$$

for $x \in D_1 = \{x: ||x - x_*|| < \epsilon_1\}$.

A3: The matrix $\nabla^2 f(x_*)$ is positive definite, i.e., there exist positive constants m and M such that

$$m \, ||z||^2 \leq z^T \nabla^2 f(x_*) z \leq M \, ||z||^2 \quad \textit{for all} \;\; z \in \mathbb{R}^n . \tag{16}$$

In this paper we will consider only the structured BFGS secant method. In Section 2 we prove that the structured BFGS approximations to the Hessian satisfy a surprising and strong form of bounded deterioration. In Section 3 we establish local q-superlinear convergence for the structured BFGS secant method using the Broyden, Dennis and Moré, Dennis and Moré and Griewank and Toint theories [Refs. 18, 19, 20]. Finally, in Section 4 we use this theory to prove the local q-superlinear convergence of the structured BFGS secant method used in the current version of the popular NL2SOL code for nonlinear least-squares problems.

2. Bounded Deterioration for the Structured BFGS Update.

Our objective in this section is to demonstrate that the structured BFGS approximations to the Hessian satisfy the bounded deterioration principle given by Dennis in Ref. 21 and popularized by Broyden, Dennis and Moré in Ref. 18. Moreover, we will prove that the BFGS secant updates satisfy a surprising and stronger form of this principle. Specifically they satisfy, for x and B sufficiently close to x_* and $\nabla^2 f(x_*)$ respectively, the condition

$$||B_+ - \nabla^2 f(x_*)||_* \leq ||B - \nabla^2 f(x_*)||_* + \alpha\sigma(x, x_+) \tag{17}$$

where $\sigma(u, v) = \max\{||u - x_*||, ||v - x_*||\}$, and α is a nonegative constant.

This fact will allow us to use the Broyden-Dennis-Moré theory to establish that under the standard conditions the sequence $\{x_k\}$ generated by a structured BFGS secant method is locally q-linearly convergent to x_*. The q-superlinear convergence will then follow from Proposition 4 of Griewank and Toint [Ref. 20] and the Dennis-Moré characterization [Ref. 19].

The bounds needed to prove inequality (17) when the structure in the Hessian is not used follow from the fact that y is a good approximation to $\nabla^2 f(x_*)s$ and Assumption A3. We formalize this fact in the following proposition.

PROPOSITION 2.1. *Assume that Standard Assumption A3 holds and let D be a neighborhood of x_*. For $x_1, x_2 \in D$ define $s = x_2 - x_1$ and let y be an approximation to $\nabla^2 f(x_*)s$. If there exists $K_1 > 0$ such that*

$$||y - \nabla^2 f(x_*)s|| \leq K_1\sigma(x_1, x_2)||s|| \tag{18}$$

for all $x_1, x_2 \in D$, then the following inequalities hold:

$$||y|| \leq (M + K_1\sigma(x_1, x_2))||s|| \tag{19a}$$

$$y^T s \leq (M + K_1\sigma(x_1, x_2))||s||^2 \tag{19b}$$

where M is given in Standard Assumption A3. Moreover, there exist positive constants ϵ_2, and β such that the following inequalities hold:

$$y^T s \geq \beta\, ||s||^2 \tag{20a}$$

$$\frac{\|y\|\,\|s\|}{y^T s} \le \frac{M}{\beta} + \frac{K_1}{\beta}\sigma(x_1, x_2), \quad s \ne 0 \tag{20b}$$

for $x_1, x_2 \in D_2 = \{x:\ \|x - x_*\| \le \epsilon_2\} \subset D$.

Proof. Let $z = y - \nabla^2 f(x_*)s$ and $x_1, x_2 \in D$. Then (19) follows directly from inequality (18) and Standard Assumption A3. To define D_2, choose ϵ_2 so that $K_1 \epsilon_2 < m$ and $D_2 \subset D$, where m is given in Standard Assumption A3. If $x_1, x_2 \in D_2$, then (20a) follows from Standard Assumption A3 with $\beta = m - K_1\epsilon_2$. Finally, notice that for $s \ne 0$

$$\frac{\|y\|\,\|s\|}{y^T s} = \frac{\|y\|}{\|s\|}\,\frac{\|s\|^2}{y^T s};$$

so that (20b) follows from inequalities (19a) and (20a). •

Similarly, when the structure in the Hessian is used, the bounds needed to establish bounded deterioration (17) follow from Standard Assumption A2 and A3, and the fact that $y^\#$ is a "good" approximation to $S(x_*)s$. We formulate this fact in the next proposition.

PROPOSITION 2.2. *Assume that Standard Assumption A2 holds and let* D *be a neighborhood of* x_*. *For* $x_1, x_2 \in D$ *define* $s = x_2 - x_1$ *and let* $y^\#$ *be an approximation to* $S(x_*)s$. *If there exists* $K_2 > 0$ *such that*

$$\|y^\# - S(x_*)s\| \le K_2 \sigma(x_1, x_2)\,\|s\| \tag{21}$$

for all $x_1, x_2 \in D$, *then there exists* $K_3 > 0$ *such that* $y^S = y^\# + C(\bar{x})s$ *for any* $\bar{x} \in [x_1, x_2]$ *satisfies*

$$\|y^S - \nabla^2 f(x_*)s\| \le K_3 \sigma(x_1, x_2)\,\|s\| \tag{22}$$

for all $x_1, x_2 \in D_1 \cap D$ *where* D_1 *is given in Standard Assumption A2.*

Proof. Let $x_1, x_2 \in D_1 \cap D$. Taking advantage of the structure in y^S and in the Hessian, we can write

$$\begin{aligned} \|y^S - \nabla^2 f(x_*) s\| &\le \|y^\# - S(x_*) s\| + \|[C(\bar{x}) - C(x_*)] s\| \\ &\le K_2 \sigma(x_1, x_2) \|s\| + L_C \|\bar{x} - x_*\| \|s\| \\ &\le (K_2 + L_C) \sigma(x_1, x_2) \|s\| . \qquad \bullet \end{aligned}$$

The next lemma is very useful when dealing with weighted Frobenius norms. Particular cases of it were established by Powell and by Griewank and Toint [Ref. 22, 20].

LEMMA 2.3. *Consider a symmetric matrix* $\bar{B} \in \mathbb{R}^{n \times n}$ *and vectors* $u, z \in \mathbb{R}^n$. *Suppose that*

$$u^T u = 1 \quad \text{and} \quad u^T \bar{B} u = (u^T z)^2 . \tag{23}$$

If we define

$$\bar{B}' = \bar{B} + u u^T - z z^T , \tag{24}$$

then

$$\|\bar{B}' - I\|^2_F = \|\bar{B} - I\|^2_F - \{(1 - z^T z)^2 + 2(z^T \bar{B} z - (z^T z)^2)\} . \tag{25}$$

Moreover, if $\bar{B}$ *is symmetric and positive definite,* $u = \dfrac{v}{\|v\|}$ *and* $z = \dfrac{\bar{B} v}{\sqrt{v^T \bar{B} v}}$ *for some vector* $v \in \mathbb{R}^n$, $v \neq 0$, *then*

$$\|\bar{B}' - I\|_F \le \|\bar{B} - I\|_F . \tag{26}$$

Proof. The first part, (25), is a straightforward application of $\|A\|^2_F = \mathit{trace}\,(A^T A)$, $\mathit{trace}\,(A + B) = \mathit{trace}\,(A) + \mathit{trace}\,(B)$, and $\mathit{trace}\,(x y^T) = x^T y$. Observe that

$$(\overline{B}'-I)^T(\overline{B}'-I) = (\overline{B}-I)^T(\overline{B}-I)+(\overline{B}-I)\,uu^T+uu^T(\overline{B}-I)$$
$$-(\overline{B}-I)\,zz^T - zz^T(\overline{B}-I)+(u^Tu)\,uu^T$$
$$+(z^Tz)\,zz^T-(u^Tz)\,uz^T-(z^Tu)\,zu^T$$

and so

$$trace\ (\overline{B}'-I)^T(\overline{B}'-I) = trace\ (\overline{B}-I)^T(\overline{B}-I)+2u^T(\overline{B}-I)\,u$$
$$-2z^T(\overline{B}-I)\,z+(u^Tu)^2+(z^Tz)^2-2(u^Tz)^2\,.$$

Finally, we obtain (25) using (23).

To demonstrate (26), notice that the given u and z satisfy (23) for any vector $v \neq 0$. Therefore, (26) will be true if $z^T\overline{B}z-(z^Tz)^2 \geq 0$.

Using the definition of z we have

$$z^T\overline{B}z-(z^Tz)^2 = \frac{v^T\overline{B}^3v}{v^T\overline{B}v}-\left(\frac{v^T\overline{B}^2v}{v^T\overline{B}v}\right)^2 = \frac{v^T\overline{B}^3v\cdot v^T\overline{B}v-(v^T\overline{B}v)^2}{(v^T\overline{B}v)^2}\,.$$

We will now show that the numerator of the last expression is positive. From the Cauchy-Schwarz inequality we have

$$v^T\overline{B}^3v\cdot v^T\overline{B}v = ||\overline{B}^{3/2}v\,||^2\,||\overline{B}^{1/2}v\,||^2 = \left[||\overline{B}^{3/2}v\,||\,||\overline{B}^{1/2}v\,||\right]^2$$
$$\geq \left[(\overline{B}^{3/2}v)^T(\overline{B}^{1/2}v)\right]^2 = \left[v^T\overline{B}^2v\right]^2. \quad \bullet$$

Now we establish the bounded deterioration principle for the (unstructured) BFGS secant approximations. The proof is based on the approach used by Griewank and Toint [Ref. 20] for the Broyden convex class of secant updates. However, our result is stronger than the specialization to the BFGS of their result (we obtain a sharper bounded deterioration inequality). Moreover, in order to fully expose the ideas involved, we will not assume that the problem has been transformed so that the Hessian at x_* is the identity matrix.

THEOREM 2.4. *Suppose that Standard Assumption A3 holds. Let B_+ be the (unstructured) BFGS secant update, i.e.*

$$B_+ = B + \Delta(s, y, B, v) \tag{27}$$

where $s = x_+ - x$, the scale v is given by (10) and y is an approximation to $\nabla^2 f(x_*)s$. If y satisfies inequality (18), then the bounded deterioration inequality

$$\|B_+ - \nabla^2 f(x_*)\|_* \leq \|B - \nabla^2 f(x_*)\|_* + \alpha_1 \sigma(x, x_+) \tag{28}$$

holds whenever $x, x_+ \in D_2$, where D_2 is given in Proposition 2.1.

Proof. Let $B^* = \nabla^2 f(x_*)$ and $x, x_+ \in D_2$. Recall that the BFGS secant correction, (7) with (10), can also be written as

$$BFGS(s, y, B) = \frac{yy^T}{y^T s} - \frac{Bss^T B}{s^T Bs} \tag{29}$$

(see Chapter 9 of Dennis and Schnabel [Ref. 11]). Define

$$B' = B + BFGS\,(s, B^* s, B)\,. \tag{30}$$

The idea of the proof is to determine bounds on $\|B_+ - B'\|_*$ and $\|B' - B^*\|_*$ in terms of $\|B - B^*\|_*$ and then apply the triangle inequality to obtain (28). Notice below that the strong form of bounded deterioration given by (28) is a consequence of the fact that the difference between B_+ and B' does not depend on B.

The bound on $\|B' - B^*\|_*$ follows from (26) in Lemma 2.3. If $\bar{B}' = B^{*-1/2} B' B^{*-1/2}$, $\bar{B} = B^{*-1/2} B B^{*-1/2}$ and $v = B^{*1/2} s$ we can write

$$
\begin{aligned}
\|B' - B^*\|_* &= \|B^{*-1/2}(B' - B^*)B^{*-1/2}\|_F = \|\bar{B}' - I\|_F \\
&= \left\|B^{*-1/2}\left(B - B^* + \frac{B^* s s^T B^*}{s^T B^* s} - \frac{B s s^T B}{s^T B s}\right)B^{*-1/2}\right\|_F \\
&= \left\|B^{*-1/2}(B - B^*)B^{*-1/2} + \frac{B^{*1/2} s s^T B^{*1/2}}{\|B^{*1/2} s\|^2} - \frac{B^{*-1/2} B s s^T B B^{*-1/2}}{s^T B s}\right\|_F \\
&= \left\|\bar{B} - I + \frac{v v^T}{\|v\|^2} - \frac{\bar{B} v (\bar{B} v)^T}{v^T \bar{B} v}\right\|_F .
\end{aligned}
$$

Therefore, by (26)

$$
\|B' - B^*\|_* \le \|B - B^*\|_* . \tag{31}
$$

To derive a bound on $\|B_+ - B'\|_*$, observe that

$$
\begin{aligned}
B_+ - B' &= \frac{y y^T}{y^T s} - \frac{B^* s s^T B^*}{s^T B^* s} \\
&= \frac{y (y - B^* s)^T}{y^T s} + \left[\frac{1}{y^T s} - \frac{1}{s^T B^* s}\right] y s^T B^* + \frac{(y - B^* s) s^T B^*}{s^T B^* s} .
\end{aligned}
$$

Using A3, (18), (19), (20) and $\|x y^T\|_F = \|x\| \, \|y\|$, we have

$$
\begin{aligned}
\|B_+ - B'\|_F &\le \frac{\|y\|\,\|y - B^* s\|}{y^T s} + \frac{\|y\|\,\|B^* s\|\,\|y - B^* s\|\,\|s\|}{y^T s \cdot s^T B^* s} + \\
&\quad + \frac{\|y - B^* s\|\,\|B^* s\|}{s^T B^* s} \\
&\le \frac{\|y\|\,\|s\|}{y^T s} \cdot \frac{\|y - B^* s\|}{\|s\|} + \\
&\quad + \frac{\|y\|\,\|s\|}{y^T s} \cdot \frac{\|B^*\|\,\|s\|^2}{s^T B^* s} \cdot \frac{\|y - B^* s\|}{\|s\|} + \\
&\quad + \frac{\|y - B^* s\|}{\|s\|} \cdot \frac{\|B^*\|\,\|s\|^2}{s^T B^* s} \\
&\le \left[\frac{M + K_1 \sigma(x, x_+)}{\beta} + \frac{M + K_1 \sigma(x, x_+)}{\beta} \cdot \frac{M}{m} + \frac{M}{m} \right] K_1 \sigma(x, x_+) \\
&\le \left[\frac{M + K_1 \epsilon_2}{\beta} + \frac{M + K_1 \epsilon_2}{\beta} \frac{M}{m} + \frac{M}{m} \right] K_1 \sigma(x, x_+) \,.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\|B_+ - B'\|_* &\le \|B^{*^{-1/2}}\|^2 \, \|B_+ - B'\|_F \\
&\le \alpha_1 \sigma(x, x_+) \,,
\end{aligned}
\tag{32}
$$

where

$$
\alpha_1 = \frac{K_1}{m} \left(\frac{M + K_1 \epsilon_2}{\beta} (1 + \frac{M}{m}) + \frac{M}{m} \right).
$$

Finally, using the triangle inequality, (31), and (32), we have

$$
\begin{aligned}
\|B_+ - B^*\|_* &\le \|B_+ - B'\|_* + \|B' - B^*\|_* \\
&\le \alpha_1 \sigma(x, x_+) + \|B - B^*\|_* \,,
\end{aligned}
\tag{33}
$$

which is the strong form of bounded deterioration that we set out to prove. •

Finally, we prove an analogous result for the structured BFGS secant approximations.

THEOREM 2.5. *Suppose that Standard Assumptions A2 and A3 hold. Let* B_+ *be the structured BFGS secant update, i.e.,*

$$B_+ = A_+ + C(x_+) \tag{34a}$$

where

$$A_+ = A + \Delta(s, y^{\#}, A, v(s, y^S, B^S)), \tag{34b}$$

$s = x_+ - x$, *the scale* v *is given by (10), and* y^S *and* $y^{\#}$ *are approximations to* $\nabla^2 f(x_*)s$ *and* $S(x_*)s$ *respectively such that* $y^S - y^{\#} = C(\bar{x})$ *for any* $\bar{x} \in [x, x_+]$. *If* $y^{\#}$ *satisfies inequality (21), then there exists a neighborhood* D_3 *of* x_* *such that*

$$||B_+ - \nabla^2 f(x_*)||_* \leq ||B - \nabla^2 f(x_*)||_* + \alpha_2 \sigma(x, x_+) \tag{35}$$

holds whenever $x, x_+ \in D_3 = D_1 \cap D_2$, *where* D_1 *and* D_2 *are given in A2 and Proposition 2.1 respectively.*

Proof. Let $B^* = \nabla^2 f(x_*)$, $B^S = A + C(\bar{x})$ and restrict D_3 as needed so that B^S is positive definite.

Now, using (34) and the following simple observation (which we commented about in Section 1):

$$\Delta(s, y^{\#}, A, v) = \Delta(s, y^S, B^S, v)$$

we have for $x, x_+ \in D_3$

$$
\begin{aligned}
B_+ &= A_+ + C(x_+) \\
&= A + \Delta(s, y^{\#}, A, v) + C(x_+) \\
&= A + \Delta(s, y^S, B^S, v) + C(x_+) \\
&= B^S - C(x) + \Delta(s, y^S, B^S, v) + C(x_+) \\
&= B^S + \Delta(s, y^S, B^S, v) + C(x_+) - C(x)\,.
\end{aligned}
\tag{36}
$$

Since Proposition 2.2 allows us to use Theorem 2.4, and $B^S = B + C(\bar{x}) - C(x)$, we can write

$$
\begin{aligned}
\|B_+ - B^*\|_* &\le \|B^S + \Delta(s, y^S, B^S, v) - B^*\|_* + \|C(x_+) - C(x)\|_* \\
&\le \|B^S - B^*\|_* + \alpha_1 \sigma(x, x_+) + \\
&\quad + \sqrt{n}\, L_C \, \|B^{*^{-1/2}}\|^2 (\|x_+ - x_*\| + \|\bar{x} - x_*\|) \\
&\le \|B - B^*\|_* + \|C(\bar{x}) - C(x)\|_* + \alpha_1 \sigma(x, x_+) + \frac{2\sqrt{n}\, L_C}{m} \sigma(x, x_+) \\
&\le \|B - B^*\|_* + [\alpha_1 + \frac{4\sqrt{n} L_C}{m}] \sigma(x, x_+)\,,
\end{aligned}
$$

which is (35) with $\alpha_2 = \alpha_1 + \dfrac{4\sqrt{n} L_C}{m}$ where α_1 is given in Theorem 2.4. •

3. Local Convergence Theory.

In this section we will establish the local and q-superlinear convergence of the structured BFGS secant method defined in Section 1. Our approach will be to use the results of Section 2 and the Broyden-Dennis-Moré theory to prove local q-linear convergence. Then we use (36), Proposition 4 of Griewank and Toint [Ref. 20] and the Dennis-Moré characterization [Ref. 19] to obtain q-superlinear convergence. For completeness we restate the Griewank-Toint proposition as follows.

PROPOSITION 3.1 (Griewank and Toint [Ref. 20]). *Suppose that Standard Assumptions A1, A2 and A3 hold. Let* $\{x_k\}$ *be a sequence which converges to* x_* *and satisfies*

$$\sum_{k>0} \|x_k - x_*\| < \infty . \tag{37}$$

Also, let $\{B_k\}$*, the approximations to the Hessian, be generated by (2) and (6)-(7) starting with a symmetric positive definite matrix* B_0*. Then*

$$\lim_k \frac{\|(B_k - \nabla^2 f(x_*))s_k\|}{\|s_k\|} = 0 . \tag{38}$$

The next theorem gives sufficient conditions to insure local q-superlinear convergence for the structured BFGS secant method.

THEOREM 3.2. *Suppose that Standard Assumptions A1, A2 and A3 hold. If* $s = x_1 - x_2$, y^S *and* $y^\#$ *are approximations to* $\nabla^2 f(x_*)s$ *and* $S(x_*)s$ *respectively such that* $y^S - y^\# = C(\bar{x})s$ *for some* $\bar{x} \in [x, x_+]$*, and* $y^\#$ *satisfies*

$$\|y^\# - S(x_*)s\| \leq K_2 \sigma(x_1, x_2) \|s\|$$

for $x_1, x_2 \in D$ *and some* $K_2 > 0$*, then there exist positive constants* ϵ, δ *such that for* $x_0 \in \mathbb{R}^n$ *and symmetric* $A_0 \in \mathbb{R}^{n \times n}$ *satisfying* $\|x_0 - x_*\| < \epsilon$ *and* $\|A_0 - S(x_*)\| < \delta$*, then sequence* $\{x_k\}$ *generated by the structured BFGS secant method for problem (1) is* q*-superlinearly convergent to* x_**.*

Proof. As was the case in Dennis and Walker [Ref. 1] the local q-linear convergence is a straightforward application of bounded deterioration (Theorem 2.5 in this case) and the standard Broyden-Dennis-Moré theory. Let $B^* = \nabla^2 f(x_*)$ and $A_* = S(x_*)$.

Since B^* is positive definite, there exist neighborhoods N_1 of x_* and N_2 of B^* which are sufficiently small so that $N_1 \subset D_3$, N_2 contains only positive definite matrices and $x_+ \in D_3$ for every $(x, B) \in N_1 \times N_2$. Now, choose a neighborhood N_3 of A_* and restrict N_1 as needed so that $(x, A) \in N = N_1 \times N_3$ implies that $A + C(x) \in N_2$.

Theorem 2.5 allows us to use Theorem 3.2 of Broyden, Dennis and Moré [Ref. 18] to prove that $\{x_k\}$ converges q-linearly to x_*. Now since the difference between B_+, the structured BFGS secant update, and an (unstructured) BFGS secant update is of size $\sigma(x, x_+)$ (see (36)), we can use Proposition 3.1 to prove that the sequence of structured BFGS secant updates satisfies Limit (3.2). Finally, from Theorem 2.2 of Dennis and Moré [Ref. 19] we conclude that the rate of convergence is q-superlinear. •

4. Application to Nonlinear Least Squares.

In this section we apply the result of Section 3 to establish the local and q-superlinear convergence of the structured BFGS secant method for the nonlinear least-squares problem and implemented in the current version of the NL2SOL code given in Refs. 9 and 10. Our presentation of the nonlinear least-squares problem follows Chapter 10 of Dennis and Schnabel [Ref. 11].

The nonlinear least-squares problem is

$$\text{minimize } f(x) = \frac{1}{2} R(x)^T R(x) = \frac{1}{2} \sum_{i=1}^{m} r_i(x)^2 \tag{39}$$

where $m \geq n$, the residual function $R : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is nonlinear and $r_i(x)$ denotes the i^{th} component function of $R(x)$. Straightforward calculations show that the gradient of f is given by

$$\nabla f(x) = J(x)^T R(x) \tag{40}$$

where $J(x)$ denotes the Jacobian of R at x, and the Hessian of f is given by

$$\nabla^2 f(x) = C(x) + S(x) \tag{41}$$

where

$$C(x) = J(x)^T J(x), \tag{42a}$$

$$S(x) = \sum_{i=1}^{m} r_i(x) \nabla^2 r_i(x), \tag{42b}$$

and $\nabla^2 r_i(x)$ is the Hessian of r_i at x.

As we mentioned in the introduction, the use of structure is an important part of the formulation of any secant method for the least-squares problem (39). Among all the suggested secant formulations for this problem, one of the most popular and successful is the NL2SOL code of Dennis, Gay and Welsch [Refs. 9, 10]. The choice of $y^\#$ used in this code is

$$y^\# = \left[J(x_+) - J(x)\right]^T R(x_+) \tag{43}$$

which was given by Dennis [Ref. 4] and, independently, by Bartholomew-Biggs [Ref. 7].

The NL2SOL code originally used the structured DFP secant update ((7)-(9)) suggested by Dennis and Welsh [Ref. 8] and analyzed by Dennis and Walker [Ref. 1], but now it uses the structured BFGS secant update ((7)-(10)) suggested by Al-Baali and Fletcher in Ref. 12. While the authors report improved numerical results, there is no local convergence theory for the new version of the algorithm. We will establish such theory in the next paragraphs.

Consider the following standard assumptions for problem (39).

A1: Problem (39) has a local solution x_*.

A2: The function $f \in C^2$ and J and $\nabla^2 f$ are locally Lipschitz continuous at x_*, i.e., there exist L_1, L_2, and ϵ such that

$$||J(x) - J(x_*)|| \le L_1 ||x - x_*|| \tag{44a}$$

and

$$||\nabla^2 f(x) - \nabla^2 f(x_*)|| \le L_2 ||x - x_*|| \tag{44b}$$

for $x \in D = \{x : ||x - x_*|| < \epsilon\}$.

A3: The matrix $\nabla^2 f(x_*)$ is positive definite.

The following lemma serves as the foundation of our convergence result.

LEMMA 4.1. *Suppose that the standard assumptions for problem (39) hold. Then there exists a positive constant K such that*

$$||y^{\#} - S(x_*)s|| \le K\sigma(x, x_+) ||s|| \tag{45}$$

where $y^{\#}$ is given by (43), $x, x_+ \in D$, and $s = x_+ - x$.

Proof. Observe that by adding and subtracting the appropriate term we have

$$\begin{aligned} y^{\#} - S(x_*)s &= J(x_+)^T R(x_+) - J(x)^T R(x_+) - S(x_*)s \\ &= \nabla f(x_+) - \nabla f(x) - J(x)^T \left[R(x_+) - R(x) - J(x_*)s\right] \\ &\quad - \left[J(x) - J(x_*)\right]^T J(x_*)s - \nabla^2 f(x_*)s \ . \end{aligned} \tag{46}$$

From (44) and Lemma 4.1.15 in Dennis and Schnabel [Ref. 11] we have

$$||\nabla f(x_+) - \nabla f(x) - \nabla^2 f(x_*)s|| \le L_2 \sigma(x, x_+) ||s|| \tag{47a}$$

and

$$||R(x_+) - R(x) - J(x_*)s\,|| \leq L_1 \sigma(x, x_+)\,||s\,|| \tag{47b}$$

Therefore, using (46) and (47)

$$\begin{aligned} ||y^{\#} - S(x_*)s\,|| &\leq L_2 \sigma(x, x_+)\,||s\,|| + ||J(x)\,||L_1 \sigma(x, x_+)\,||s\,|| \\ &\quad + ||J(x_*)\,||L_1\,||x - x_*\,||\,||s\,|| \\ &\leq \Big[L_2 + (L_1 \epsilon + L_*)L_1 + L_* L_1\Big] \sigma(x, x_+)\,||s\,|| \end{aligned}$$

where $L_* = ||J(x_*)\,||$. •

THEOREM 4.2. *Suppose that the standard assumptions for problem (39) hold. Then there exist positive constants* ϵ, δ *such that for* $x_0 \in \mathbb{R}^n$ *and symmetric* $A_0 \in \mathbb{R}^n$ *satisfying* $||x_0 - x_*\,|| < \epsilon$ *and* $||A_0 - S(x_*)\,|| < \delta$, *the iteration sequence* $\{x_k\}$ *generated by the structured BFGS secant method for problem (39) is* q*-superlinearly convergent to* x_*.

Proof. The proof of this theorem is a straightforward application of Theorem 3.2 and Lemma 4.1. •

5. Conclusions and Summary.

In this paper we have defined, and established the local and q-superlinear convergence of, the structured BFGS secant method for unconstrained optimization. Moreover, we have introduced the structure principle as a tool for formulating the appropriate scale in any structured secant update for a particular problem or application. Indeed, in his Ph.D. thesis [Ref. 23], Martinez defined, and derived a convergence theory for, structured secant methods generated from the entire Broyden convex class, using this structure principle to extend the definition of the scale from unstructured to structured applications.

Although additional work is needed to develop a global convergence theory for these structured algorithms, we think that the surprising and stronger form of bounded deterioration proved here may be useful in the development of this global theory, especially in the constext of a trust region globalization strategy is used.

Finally, as a direct application of the theory given in this paper, we gave the first proof of local and q-superlinear convergence of the important structured BFGS secant method for the nonlinear least-squares problem which is used by Dennis, Gay and Welsh in the current version of the popular NL2SOL code.

REFERENCES

1. DENNIS, J.E. Jr. and H.F. WALKER, *Convergence theorems for least-change secant update methods,* SIAM Journal Numerical Analysis, Vol. 18, pp. 949-987, 1981.

2. BROWN, K.M. AND DENNIS, J.E. Jr., *A new algorithm for nonlinear least-squares curve fitting,* Mathematical Software, Edited by J. Rice, Academic Press, New York (1971), pp. 391-396, 1971.

3. DENNIS, J.E. Jr., *Some computational techniques for the nonlinear least squares problem,* Numerical Solution of Systems of Nonlinear Algebraic Equations, Edited by G.D. Byrne and C.D. Hall, New York Academic Press, 1973.

4. DENNIS, J.E. Jr., *A brief survey of convergence results for quasi-Newton methods,* Nonlinear Programming, Edited by R. Cottle and C. Lemke, SIAM-AMS Proceedings, Vol. 9, 1976.

5. DENNIS, J.E. Jr., *Nonlinear least squares and equations,* The State of the Art in Numerical Analysis, Edited by D. Jacobs, Academic Press, London, pp. 269-312, 1977.

6. BETTS, J.T., *Solving the nonlinear least squares problem: application of a general method,* Journal of Optimization Theory and Applications, Vol. 18, 469-484, 1976.

7. BARTHOLOMEW-BIGGS, M.C., *The estimation of the Hessian matrix in nonlinear least squares problems with non-zero residuals,* Mathematical Programming, Vol. 12, pp. 67-80, 1977.

8. DENNIS, J.E., Jr. and R.E. WELSCH, *Techniques for nonlinear least-squares and robust regression,* Commun. Statist. Simula. Comput., B7(4), pp. 345-359, 1978.

9. DENNIS, J.E. Jr., D.M. GAY and R.E. WELSCH, *An adaptive nonlinear least-squares algorithm,* TOMS, Vol. 7, pp. 348-368, 1981.

10. DENNIS, J.E. Jr., D.M. GAY and R.E. WELSCH, *Algorithm 573 NL2SOL - An adaptive nonlinear least-squares algorithm [E4],* TOMS, Vol. 7, pp. 369-383, 1981.

11. DENNIS, J.E. Jr. and R.B. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations,* Prentice-Hall, Englewood Cliffs, New Jersey, 1983.

12. AL-BAALI, M. and R. FLETCHER, *Variational methods for nonlinear least squares,* J. Opl. Res. Soc., 36, pp. 405-421, 1985.

13. XU, C., *Hybrid methods for nonlinear least squares and related problems,* University of Dundee, PhD Thesis, 1986.

14. MAHDAVI-AMIRI, N. and R.H. BARTLES, *Constrained nonlinear least squares: An exact penalty approach with projected structured quasi-Newton updates,* unpublished manuscript.

15. TOINT, PH. L., *On large scale nonlinear least squares calculations,* SIAM Journal of Scientific and Statistical Computations, Vol. 8, pp. 416-435, 1987.

16. FLETCHER, R. and C. XU, *Hybrid methods for nonlinear least squares,* IMA Journal Numerical Analysis, Vol 7, pp. 371-389, 1987.

17. TAPIA, R.A., *On secant updates for use in general constrained optimization,* Rice University, Mathematical Sciences, Report No. TR84-3, Revised January, 1987, To appear in Mathematics of Computation.

18. BROYDEN, C.G., J.E. DENNIS Jr., and J.J. MORE, *On the local and superlinear convergence of quasi-Newton methods,* IMA Journal of Applyed Mathematics, Vol. 12, pp. 223-246, 1973.

19. DENNIS, J.E Jr. and J.J. MORE, *A characterization of superlinear convergence and its application to quasi-Newton methods,* Mathematics of Computation, Vol. 28, pp. 549-560, 1974.

20. GRIEWANK, A.O. and PH. L. Toint, *Local convergence analysis for partitioned quasi-Newton updates,* Numer. Math., Vol. 39, pp. 429-448, 1982.

21. DENNIS, J.E. Jr., *Toward a unified convergence theory for Newton-like methods,* Non-linear functional analysis and applications, Edited by L. B. Rall, New York Academic Press, 1971.

22. POWELL, M.J.D., *The convergence of variable metric methods for nonlinearly constrained optimization problems,* Nonlinear Programming, 3, Edited by O.L. Mangasarian, R. Meyer, S. Robinson, Academic Press, New York, pp. 27-63, 1978.

23. MARTINEZ, H.J., *Local and superlinear convergence of structured secant methods from the convex class,* Rice University, PhD Thesis, 1988.

REFERENCES

1. DENNIS, J.E., Jr., and WALKER, H.F., *Convergence Theorems for Least Change Secant Update Methods*, SIAM Journal on Numerical Analysis, Vol. 18, pp. 949-987, 1981.

2. BROWN, K.M., and DENNIS, J.E., Jr., *A New Algorithm for Nonlinear Least Squares Curve Fitting*, Mathematical Software, Edited by J. Rice, Academic Press, New York, New York, pp. 391-396, 1971.

3. DENNIS, J.E., Jr., *Some Computational Techniques for the Nonlinear Least Squares Problem*, Numerical Solution of Systems of Nonlinear Algebraic Equations, Edited by G.D. Byrne and C.D. Hall, Academic Press, New York, New York, 1973.

4. DENNIS, J.E., Jr., *A Brief Survey of Convergence Results for Quasi-Newton Methods*, Nonlinear Programming, Edited by R. Cottle and C. Lemke, SIAM-AMS Proceedings, Vol. 9, pp. 185-199, 1976.

5. DENNIS, J.E., Jr., *Nonlinear Least Squares and Equations*, The State of the Art in Numerical Analysis, Edited by D. Jacobs, Academic Press, London, England, pp. 269-312, 1977.

6. BETTS, J.T., *Solving the Nonlinear Least Squares Problem: Application of a General Method*, Journal of Optimization Theory and Applications, Vol. 18, pp. 469-484, 1976.

7. BARTHOLOMEW-BIGGS, M.C., *The Estimation of the Hessian Matrix in Nonlinear Least Squares Problems with Nonzero Residuals*, Mathematical Programming, Vol. 12, pp. 67-80, 1977.

8. DENNIS, J.E., Jr., and WELSCH, R.E., *Techniques for Nonlinear Least Squares and Robust Regression*, Communications in Statistics, Vol. B7, pp. 345-359, 1978.

9. DENNIS, J.E., Jr., GAY, D.M., and WELSCH, R.E., *An Adaptive Nonlinear Least Squares Algorithm*, Transactions on Mathematical Software, Vol. 7, pp. 348-368, 1981.

10. DENNIS, J.E., Jr., GAY, D.M., and WELSCH, R.E., *Algorithm 573 NL2SOL – An Adaptive Nonlinear Least Squares Algorithm*, Transactions on Mathematical Software, Vol. 7, pp. 369-383, 1981.

11. DENNIS, J.E., Jr., and SCHNABEL, R.B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, New Jersey, 1983.

12. AL-BAALI, M., and FLETCHER, R., *Variational Methods for Nonlinear Least Squares*, Journal of the Operational Research Society, 36, pp. 405-421, 1985.

13. XU, C., *Hybrid Methods for Nonlinear Least Squares and Related Problems*, University of Dundee, PhD Thesis, 1986.

14. MAHDAVI-AMIRI, N., and BARTLES, R.H., *Constrained Nonlinear Least Squares: An Exact Penalty Approach with Projected Structured Quasi-Newton Updates*, Unpublished Manuscript, 1988.

15. TOINT, PH.L., *On Large Scale Nonlinear Least Squares Calculations*, SIAM Journal of Scientific and Statistical Computations, Vol. 8, pp. 416-435, 1987.

16. FLETCHER, R., and XU, C., *Hybrid Methods for Nonlinear Least Squares*, IMA Journal on Numerical Analysis, Vol. 7, pp. 371-389, 1987.

17. TAPIA, R.A., *On Secant Updates for Use in General Constrained Optimization*, Rice University, Mathematical Sciences, Report No. TR84-3, Revised January, 1987 (to appear in Mathematics of Computation).

18. BROYDEN, C.G., DENNIS, J.E., Jr., and MORE, J.J., *On the Local and Superlinear Convergence of Quasi-Newton Methods*, IMA Journal of Applied Mathematics, Vol. 12, pp. 223-246, 1973.

19. DENNIS, J.E., Jr., and MORE, J.J., *A Characterization of Superlinear Convergence and Its Application to Quasi-Newton Methods*, Mathematics of Computation, Vol. 28, pp. 549-460, 1974.

20. GRIEWANK, A.O., and TOINT, PH.L., *Local Convergence Analysis for Partitioned Quasi-Newton Updates*, Numerisch Mathematik, Vol. 39, pp. 429-448, 1982.

21. DENNIS, J.E., Jr., *Toward a Unified Convergence Theory for Newton-Like Methods*, Nonlinear Functional Analysis and Applications, Edited by L.B. Rall, Academic Press, New York, New York, 1971.

22. POWELL, M.J.D., *The convergence of Variable Metric Methods for Nonlinearly Constrained Optimization Problems*, Nonlinear Programming 3, Edited by O.L. Mangasarian, R. Meyer, and S. Robinson, Academic Press, New York, New York, pp. 27-63, 1978.

23. MARTINEZ, H.J., *Local and Superlinear Convergence of Structured Secant Methods from the Convex Class*, Rice University, PhD Thesis, 1988.