

A Memoryless Augmented Gauss-Newton Method
For
Nonlinear Least-Squares Problems

by

J.E. Dennis, Jr.¹,

Sheng Songbai²

and

Phuong Anh Vu³

Technical Report 85-1, February 1985.

¹Mathematical Sciences Department, Rice University, Houston, Texas, USA 77251. Research sponsored by DOE DE-AS05-82ER13016, ARO DAAG-29-83-K-0035, NSF MCS81-16779.

²Department of Mathematics, Nanjing University, Nanjing, Peoples Republic of China.

³International Mathematical and Statistical Libraries, 7500 Bellaire Blvd, Houston, Texas, USA

Abstract

In this paper, we develop, analyze, and test a new algorithm for nonlinear least-squares problems. The algorithm uses a BFGS update of the Gauss-Newton Hessian when some heuristics indicate that the Gauss-Newton method may not make a good step. Some important elements are that the secant or quasi-Newton equations considered are not the obvious ones, and the method does not build up a Hessian approximation over several steps. The algorithm can be implemented easily as a modification of any Gauss-Newton code, and it seems to be useful for large residual problems.

Key words

Nonlinear least squares, quasi-Newton methods, least-change secant update methods, variable metric methods.

1. Introduction

Nonlinear least-squares problems are frequently encountered in practical optimization, and they are also of interest to the algorist because of their highly structured nature. In this paper, we suggest another way to use this structure in an attempt to increase the efficiency of the trust-region-Gauss-Newton or Levenberg-Marquardt algorithm, Moré (1977), Dennis-Schnabel (1983).

The algorithm presented here is inspired by NL2SOL, Dennis, Gay, Welsch (1981a,b), in that it chooses at each iteration whether to use a Gauss-Newton quadratic model or a variable metric augmentation of the Gauss-Newton model to define the next iterate. The difference is that the variable metric augmentation used here requires less storage, less algebra, and less code than NL2SOL. However, it seems to have no better theoretical justification than the Gauss-Newton method. Still, it seems to use fewer residual and Jacobian computations than the Gauss-Newton for some large residual problems and to require little additional arithmetic at each iteration. Conversation with NL2SOL users encouraged us to undertake this research, and we publish it now in hopes that they will find it helpful and that our colleagues will find it an interesting use of secant updating ideas.

Section 2 explains the augmented local model in its various forms, and points out some overlap between our ideas and those of Al-Baali and Fletcher (1983). Section 3 contains a unified local convergence proof under standard Gauss-Newton-type assumptions for all combinations of the methods presented here. Section 4 describes a model-switching strategy and the resulting hybrid algorithm that adaptively decides whether to use the Gauss-Newton model or an augmentation at each iteration. Section 5 compares an experimental implementation of the algorithms suggested here to the LMDER implementation of the Gauss-Newton method and the NL2S1 routine from NL2SOL.

2. The Augmented Model.

Let $F: \Omega \subset \mathbf{R}^p \rightarrow \mathbf{R}^n$ be continuously differentiable, and consider the nonlinear least-squares problem of finding a local minimizer x_* for

$$\phi(x) = \frac{1}{2} F(x)^T F(x) = \frac{1}{2} \sum_{i=1}^n (f_i(x))^2. \quad (2.1)$$

The classical algorithm for this problem is the Gauss-Newton method which can be thought of in two ways:

First, we can linearize $F(x) - F(x_c)$ about the current parameter vector x_c to obtain the local affine model for $F(x)$,

$$F(x) \approx F(x_c) + J_c(x - x_c),$$

where $J_c = J(x_c) = F'(x_c) = \left(\frac{\partial f_i}{\partial x_j}(x_c) \right)$. Then we can seek to improve x_c by taking the next estimate x_- to be the value of the parameter vector that solves the linear least-square problem defined by the local affine model.

The sum-of-squares-of-residuals of this model is

$$\frac{1}{2} [F(x_c) + J_c(x - x_c)]^T [F(x_c) + J_c(x - x_c)] , \quad (2.2)$$

and it can be viewed as a local quadratic model of $\phi(x)$ of the form

$$\phi(x) \approx m_c^{GN}(x) \equiv \phi(x_c) + \nabla \phi(x_c)^T (x - x_c) + \frac{1}{2} (x - x_c)^T J_c^T J_c (x - x_c) . \quad (2.3)$$

A second way to view this local quadratic model is as an approximation to the Newton model

$$m_c^N(x) = \phi(x_c) + \nabla \phi(x_c)^T (x - x_c) + \frac{1}{2} (x - x_c)^T \nabla^2 \phi(x_c) (x - x_c) \quad (2.4)$$

where

$$\nabla^2\phi(x_c) - J_c^T J_c = \sum_{i=1}^n f_i(x_c) \nabla^2 f_i(x_c) \equiv S(x_c) \quad (2.5)$$

is approximated by the zero matrix. It is easy to reason from either derivation that the difference between the two models depends on the size of the residuals $F(x_c)$ and on how nearly affine F is in a neighborhood of x_c .

Aside from the obvious advantage that the Gauss-Newton method has of not having to compute or make assumptions about the $n \times p$ Hessians $\nabla^2 f_i(x_c)$, $i = 1, \dots, n$, it also is guaranteed to generate descent directions as long as J_c is of full rank. This happens because $\nabla^2 m_c^{GN}(x_c) = J_c^T J_c$ is positive definite, and it means that the next iterate x_+^{GN} can be calculated by solving the linear least-squares problem associated with (2.2) for $s^{GN} = x_+^{GN} - x_c$. This leads to a useful simplification over Newton's method of the problem of proceeding from a poor initial guess when $\nabla^2\phi(x_c)$ may not be positive definite even though $\nabla^2\phi(x_*)$ is.

The major disadvantage of (2.3) with respect to (2.4) is that neglecting $S(x_c)$ often leads to a significantly less accurate local model. It is not surprising that this would cause slower convergence near a local minimizer of a large residual nonlinear problem; however far from x_* , it results often either in x_+^{GN} not being an acceptable next iterate when x_+^N is, or in a smaller residual reduction than Newton's method provides.

NL2SOL essentially retained the Newton advantages without having to compute $S(x_c)$. It did this by using a variable metric update method and an adaptive modeling technique to switch between the Gauss-Newton model and an augmented model of the form

$$m_c^A(x) = \phi(x_c) + \nabla\phi(x_c)^T(x - x_c) + \frac{1}{2}(x - x_c)^T [J_c^T J_c + S_c](x - x_c) \quad (2.6)$$

where S_c is a variable metric *approximation* to $S(x_c)$ and $J_c = J(x_c)$. Here as in NL2SOL, the decision is made at the end of the current iteration whether to use the Gauss-Newton or its variable metric augmentation to make the next step. This decision is based on a simple comparison of the actual residual reduction $\phi(x_c) - \phi(x_+)$ to the predictions $\phi(x_c) - m_c^{GN}(x_+)$ and $\phi(x_c) - m_c^A(x_+)$. The algorithm suggested here does not attempt to build up a good approximation to $S(x_c)$ over several iterations; it temporizes a cheap rank-2 approximation if a given iteration seems to call for

it.

We complete this section with a description of the way we suggest defining S_c in (2.6) and with some basic facts about this definition. We postpone a discussion of implementational details until Section 4.

Given $x_c, x_-, J_c, J_-, F_c, F_-$, information at the current and previous iterate,

$$\text{Set } s_- = x_c - x_- \text{ and } 0 \leq \alpha_c \leq \min\{1, \|J_c^T F_c\|\}; \quad (2.7)$$

Set either

$$y_- = J_c^T [J_c s_- + \alpha_c (F_c - F_-)] \quad (2.8a)$$

or

$$y_- = J_c^T J_c s_- + \alpha_c (J_c^T F_c - J_-^T F_-), \quad (2.8b)$$

or

$$y_- = J_c^T J_c s_- + \alpha_c (J_c^T - J_-^T) F_c, \quad (2.8c)$$

and either

$$S_c^{BFGS} = \frac{y_- y_-^T}{s_-^T y_-} - \frac{J_c^T J_c s_- s_-^T J_c^T J_c}{s_-^T J_c^T J_c s_-}, \quad (2.9a)$$

or

$$S_c^{DFP} = \frac{(y_- - J_c^T J_c s_-) y_-^T + y_- (y_- - J_c^T J_c s_-)^T}{s_-^T y_-} - \frac{s_-^T (y_- - J_c^T J_c s_-) y_- y_-^T}{(s_-^T y_-)^2}. \quad (2.9b)$$

Note that with any pair (2.8), (2.9), we have

$$(J_c^T J_c + S_c) s_- = y_-, \quad (2.10)$$

and that $\alpha_c = 0$ recovers the Gauss-Newton method. If we take $\alpha_c = 1$, then we obtain the two methods that Al-Baali and Fletcher (1983) call the GN-BFGS and GN-DFP methods. Our proof techniques can not support always taking $\alpha_c = 1$, and so our analysis does not apply to their methods. Our numerical results suggest that $\alpha_c = 1$ is not always the best choice.

It is tedious, but not difficult, to show that if J_c has full rank and $s_-^T y_- > 0$, then $H_c \equiv J_c^T J_c + S_c$ is positive definite and

$$(H_c^{BFGS})^{-1} \equiv (J_c^T J_c + S_c^{BFGS})^{-1} = (J_c^T J_c)^{-1} + \frac{[\mathbf{s}_- - (J_c^T J_c)^{-1} \mathbf{y}_-] \mathbf{s}_-^T + \mathbf{s}_- [\mathbf{s}_- - (J_c^T J_c)^{-1} \mathbf{y}_-]^T}{\mathbf{s}_-^T \mathbf{y}_-} - \frac{\mathbf{y}_-^T [\mathbf{s}_- - (J_c^T J_c)^{-1} \mathbf{y}_-] \mathbf{s}_- \mathbf{s}_-^T}{(\mathbf{s}_-^T \mathbf{y}_-)^2}. \quad (2.11)$$

In fact, if any symmetric nonsingular matrix A replaces $J_c^T J_c$ in (2.9), then

$$(A + S_c^{BFGS})^{-1} = \left(I - \frac{\mathbf{y}_- \mathbf{s}_-^T}{\mathbf{s}_-^T \mathbf{y}_-} \right)^T A^{-1} \left(I - \frac{\mathbf{y}_- \mathbf{s}_-^T}{\mathbf{s}_-^T \mathbf{y}_-} \right) + \frac{\mathbf{s}_- \mathbf{s}_-^T}{\mathbf{s}_-^T \mathbf{y}_-}, \quad (2.12a)$$

and

$$A + S_c^{DFP} = \left(I - \frac{\mathbf{y}_- \mathbf{s}_-^T}{\mathbf{s}_-^T \mathbf{y}_-} \right) A \left(I - \frac{\mathbf{y}_- \mathbf{s}_-^T}{\mathbf{s}_-^T \mathbf{y}_-} \right)^T + \frac{\mathbf{y}_- \mathbf{y}_-^T}{\mathbf{s}_-^T \mathbf{y}_-}. \quad (2.12b)$$

3. Local Convergence.

In this section, we will present a local convergence analysis for the quasi-Newton or Newton-like method based on the augmented local model discussed in the previous section. Our major result, Theorem 3.4, will be the same as the standard local result for the Gauss-Newton method (see Dennis (1977)); indeed, if we always choose $\alpha=0$, then the new method is easily seen to be the Gauss-Newton method. It will be convenient to collect some useful bounds before we start the main proof. We will always use the l_2 norm for both vectors and matrices, and for $\mathbf{x} \in \mathbf{R}^n$, and $\epsilon \geq 0$, $N(\mathbf{x}; \epsilon) \equiv \{\tilde{\mathbf{x}} \in \mathbf{R}^n : \|\tilde{\mathbf{x}} - \mathbf{x}\| \leq \epsilon\}$. For any set $D \subseteq \mathbf{R}^n$, \bar{D} will denote the norm closure of D .

Lemma 3.1. If \mathbf{x}, \mathbf{y} are two vectors of order n such that $\mathbf{x}^T \mathbf{y} = 1$, I is unit matrix of order n then

$$\|I - \mathbf{x} \mathbf{y}^T\| = \|\mathbf{x}\| \cdot \|\mathbf{y}\|.$$

Proof. Straightforward.

Lemma 3.2. Let $F: \mathbf{R}^n \rightarrow \mathbf{R}^m$ be continuously differentiable in an open convex set $D \subset \mathbf{R}^n$, and let J be Lipschitz continuous in D :

$$\|J(\mathbf{x}) - J(\bar{\mathbf{x}})\| \leq \gamma \|\mathbf{x} - \bar{\mathbf{x}}\|, \quad (3.1a)$$

for any $\mathbf{x}, \bar{\mathbf{x}} \in D$. Then,

$$\|J(x)^T - J(\bar{x})^T\| \leq \gamma \|x - \bar{x}\|, \quad (3.1b)$$

$$\|F(x) - F(\bar{x}) - J(\bar{x})(x - \bar{x})\| \leq \frac{\gamma}{2} \|x - \bar{x}\|^2, \quad (3.2)$$

and if \bar{D} is compact, then for $\sigma_M \geq \max_{\bar{x} \in \bar{D}} \|J(\bar{x})\|$,

$$\|J(x)^T J(x) - J(\bar{x})^T J(\bar{x})\| \leq \gamma_1 \|x - \bar{x}\|, \quad (3.3)$$

$$\|J(x)^T F(x) - J(\bar{x})^T F(\bar{x})\| \leq \gamma_2 \|x - \bar{x}\|, \quad (3.4)$$

for $\gamma_1 \equiv 2\gamma \cdot \sigma_M$ and $\gamma_2 \equiv \sigma_M^2 + \gamma \cdot \max_{\bar{x} \in \bar{D}} \|F(\bar{x})\|$.

Proof. See Dennis and Schnabel (1983) pg.75 for (3.2). The Lipschitz condition (3.1b) follows directly from the fact that the l_2 norm of a matrix and its transpose are the same. The Lipschitz condition (3.3) follows because

$\|J(\bar{x})^T J(\bar{x}) - J(x)^T J(x)\| \leq \|J(\bar{x})^T [J(\bar{x}) - J(x)]\| + \|[J(\bar{x}) - J(x)]^T J(x)\| \leq 2\gamma\sigma_M \|\bar{x} - x\|$,
where σ_M exists because \bar{D} is compact. Finally, to get (3.4),

$$\begin{aligned} \|J(x)^T F(x) - J(\bar{x})^T F(\bar{x})\| &\leq \|J(\bar{x})^T [F(x) - F(\bar{x})]\| + \|[J(x) - J(\bar{x})]^T F(x)\| \\ &\leq \sigma_M \int_0^1 \|J(\bar{x} + \Theta(x - \bar{x}))\| \cdot \|x - \bar{x}\| d\Theta + \gamma \|x - \bar{x}\| \cdot \|F(x)\| \leq \gamma_2 \|x - \bar{x}\|. \end{aligned}$$

Lemma 3.3. Let the hypothesis of Lemma 3.2 hold and let $x_* \in D$. If $J(x_*)^T J(x_*)$ is positive definite with smallest eigenvalue λ_* , then for any $\rho \in (0, 1)$, there exists $\epsilon > 0$ such that for $\bar{x}, x \in N(x_*; \epsilon)$ and $s = \bar{x} - x$,

$$s^T J(\bar{x})^T [F(\bar{x}) - F(x)] > \frac{\rho \lambda_*}{2} s^T s. \quad (3.5)$$

Also, $J(x)^T J(x)$ is positive definite with smallest eigenvalue greater than $\rho \cdot \lambda_*$ and satisfies the Lipschitz condition:

$$\|[J(\bar{x})^T J(\bar{x})]^{-1} - [J(x)^T J(x)]^{-1}\| \leq \gamma_3 \cdot \|\bar{x} - x\|, \quad (3.6)$$

for some $\gamma_3 \leq \frac{\gamma_1}{(\rho \cdot \lambda_*)^2}$.

Proof. Because $J(x)^T J(x)$ is continuous at x_* , there exists $\delta > 0$ such that the smallest eigenvalue of $J(x)^T J(x)$ is larger than $\rho \cdot \lambda_*$ for all $\|x - x_*\| \leq \delta$. Thus, $J(x)^T J(x)$ is invertible and

$$\| [J(x)^T J(x)]^{-1} \| \leq \frac{1}{\rho \cdot \lambda_*}. \quad (3.7a)$$

For any $\bar{x}, x \in N(x_*, \delta)$,

$$\begin{aligned} \| [J(\bar{x})^T J(\bar{x})]^{-1} - [J(x)^T J(x)]^{-1} \| &\leq \| [J(\bar{x})^T J(\bar{x})]^{-1} \| \cdot \| J(\bar{x})^T J(\bar{x}) - J(x)^T J(x) \| \cdot \| [J(x)^T J(x)]^{-1} \| \\ &\leq \left(\frac{1}{\rho \cdot \lambda_*} \right)^2 \cdot \gamma_1 \| \bar{x} - x \| \end{aligned}$$

by (3.7) and (3.3).

Now set $\epsilon < \min\left\{ \delta, \frac{\rho \cdot \lambda_*}{2 \cdot \gamma \cdot \sigma_M} \right\}$. So for any $p \in \mathbf{R}^n$, $x \in N(x_*, \epsilon)$, the inequality

$$\rho \cdot \lambda_* \cdot \| p \|^2 \leq p^T J(x)^T J(x) p \quad (3.7b)$$

holds. Now suppose $\bar{x}, x \in N(x_*, \epsilon)$, then from the previous lemmas and the Cauchy-Schwartz-Bunyakovskii inequality, we have

$$\begin{aligned} \|[J(\bar{x})s]^T [F(\bar{x}) - F(x)] - [J(\bar{x})s]^T [J(\bar{x})s]\| &\leq \sigma_M \|s\| \frac{1}{2} \gamma \cdot \|s\|^2 \\ &\leq \frac{1}{2} \sigma_M \gamma [\|\bar{x} - x_*\| + \|x - x_*\|] \|s\|^2 \\ &\leq \gamma \sigma_M \epsilon \|s\|^2 \leq \frac{\rho \cdot \lambda_*}{2} \cdot \|s\|^2. \end{aligned}$$

Therefore,

$$[J(\bar{x})s]^T [J(\bar{x})s] - \frac{\rho \cdot \lambda_*}{2} \|s\|^2 \leq [J(\bar{x})s]^T [F(\bar{x}) - F(x)] \leq [J(\bar{x})s]^T [J(\bar{x})s] + \frac{\rho \cdot \lambda_*}{2} \|s\|^2.$$

Hence

$$s^T J(\bar{x})^T [F(\bar{x}) - F(x)] \geq \rho \cdot \lambda_* \|s\|^2 - \frac{\rho \cdot \lambda_*}{2} \|s\|^2 = \frac{\rho \cdot \lambda_*}{2} \cdot \|s\|^2 > 0.$$

Theorem 3.4. Let the hypothesis of Lemma 3.3 hold and assume that x_* is the only local minimizer of ϕ which is convex in D . Assume that for some $\gamma_* < \lambda_*$ and every $x \in D$,

$$\| [J(x) - J(x_*)]^T F(x_*) \| \leq \gamma_* \|x - x_*\|. \quad (3.8)$$

Let $r \in \left(\frac{\gamma_*}{\lambda_*}, 1 \right)$. Then there exists $\epsilon > 0$, such that for all $x_0 \in N(x_*, \epsilon)$, the following sequence generated by any combination of the Gauss-Newton method with its BFGS and DFP augmentations is well defined and converges to x_* with $\|x_{k+1} - x_*\| \leq r \|x_k - x_*\|$:

$$x_1 = x_0 - (J_0^T J_0)^{-1} J_0^T F_0$$

For $k = 1, 2, 3, \dots$; DO

$$s_{k-1} = x_k - x_{k-1}, \text{ choose } 0 \leq \alpha_k \leq \min\{1, \|J_k^T F_k\|\};$$

$$y_{k-1} = \begin{cases} J_k^T [J_k s_{k-1} + \alpha_k (F_k - F_{k-1})], \\ \text{or} \\ J_k^T J_k s_{k-1} + \alpha_k (J_k^T F_k - J_{k-1}^T F_{k-1}), \\ \text{or} \\ J_k^T J_k s_{k-1} + \alpha_k (J_k^T - J_{k-1}^T) F_k; \end{cases}$$

$$S_k = S_k^{BFGS} = \frac{y_{k-1} y_{k-1}^T}{s_{k-1}^T y_{k-1}} - \frac{J_k^T J_k s_{k-1} s_{k-1}^T J_k^T J_k}{s_{k-1}^T J_k^T J_k s_{k-1}},$$

or

$$S_k = S_k^{DFP} = \frac{(y_{k-1} - J_k^T J_k s_{k-1}) y_{k-1}^T + y_{k-1} (y_{k-1} - J_k^T J_k s_{k-1})^T}{s_{k-1}^T y_{k-1}} - \frac{s_{k-1}^T (y_{k-1} - J_k^T J_k s_{k-1}) y_{k-1} y_{k-1}^T}{(s_{k-1}^T y_{k-1})^2};$$

$$H_k = J_k^T J_k + S_k;$$

$$x_{k+1} = x_k - H_k^{-1} J_k^T F_k;$$

END.

Furthermore, if $\gamma_* = 0$ or $\|F(x_*)\| = 0$, then the convergence is at least q-quadratic.

Proof. First we establish some more useful inequalities. Choose

$$\rho \in \left(\frac{\gamma_*}{\lambda_*}, 1 \right),$$

and take $\epsilon_1 > 0$ to be as chosen by Lemma 3.3. From (3.4) and the fact that $J(x_*)^T F(x_*) \equiv J_*^T F_* = 0$, we see that if $x \in N(x_*, \epsilon_1)$, then

$$\|J(x)^T F(x)\| \leq \gamma_2 \|x - x_*\|. \quad (3.9)$$

Furthermore, by (3.2), (3.3), (3.8),

$$\begin{aligned}
\|J(\mathbf{x})^T F(\mathbf{x}) - J_s^T J_s(\mathbf{x} - \mathbf{x}_s)\| &\leq \|J(\mathbf{x})\| \cdot \|F_s - F(\mathbf{x}) - J(\mathbf{x})(\mathbf{x}_s - \mathbf{x})\| \\
&\quad + \|J(\mathbf{x})^T J(\mathbf{x}) - J_s^T J_s\| \cdot \|\mathbf{x} - \mathbf{x}_s\| + \|(J(\mathbf{x}) - J_s)^T F_s\| \\
&\leq \sigma_M \frac{\gamma}{2} \|\mathbf{x} - \mathbf{x}_s\|^2 + \gamma_1 \|\mathbf{x} - \mathbf{x}_s\|^2 + \gamma_s \|\mathbf{x} - \mathbf{x}_s\| \\
&= (\gamma_4 \|\mathbf{x} - \mathbf{x}_s\| + \gamma_s) \cdot \|\mathbf{x} - \mathbf{x}_s\| ,
\end{aligned} \tag{3.10}$$

where $\gamma_4 \equiv \sigma_M \frac{\gamma}{2} + \gamma_1$.

Similarly, for $\mathbf{x}_-, \mathbf{x}_c \in N(\mathbf{x}_s; \epsilon_1)$ and $\mathbf{s}_-, \alpha_c, \mathbf{y}_-$ defined as in (2.7), (2.8a), we get from (3.3), (3.9), and the choice of $\alpha_c \leq \|J_c^T F_c\|$, that for $\gamma_5 \equiv \gamma_1 + \gamma_2 \max\{\sigma_M^2, \gamma_2\}$,

$$\begin{aligned}
\|\mathbf{y}_- - J_s^T J_s \mathbf{s}_-\| &= \|J_c^T J_c \mathbf{s}_- + \alpha_c J_c^T (F_c - F_-) - J_s^T J_s \mathbf{s}_-\| \\
&\leq \|J_c^T J_c - J_s^T J_s\| \cdot \|\mathbf{s}_-\| + \alpha_c \cdot \|J_c^T (F_c - F_-)\| \\
&\leq \|J_c^T J_c - J_s^T J_s\| \cdot \|\mathbf{s}_-\| + \|J_c^T F_c\| \cdot \|J_c^T (F_c - F_-)\| \\
&\leq \gamma_1 \|\mathbf{s}_-\| \cdot \|\mathbf{x}_c - \mathbf{x}_s\| + \gamma_2 \sigma_M^2 \|\mathbf{s}_-\| \cdot \|\mathbf{x}_c - \mathbf{x}_s\| \leq \gamma_5 \|\mathbf{s}_-\| \cdot \|\mathbf{x}_c - \mathbf{x}_s\| ,
\end{aligned} \tag{3.11a}$$

and for \mathbf{y}_- defined by (2.8b),

$$\begin{aligned}
\|\mathbf{y}_- - J_s^T J_s \mathbf{s}_-\| &= \|J_c^T J_c \mathbf{s}_- + \alpha_c (J_c^T F_c - J_-^T F_-) - J_s^T J_s \mathbf{s}_-\| \\
&\leq \|J_c^T J_c - J_s^T J_s\| \cdot \|\mathbf{s}_-\| + \|J_c^T F_c\| \cdot \|J_c^T F_c - J_-^T F_-\| \\
&\leq \gamma_1 \|\mathbf{s}_-\| \cdot \|\mathbf{x}_c - \mathbf{x}_s\| + \gamma_2^2 \|\mathbf{s}_-\| \cdot \|\mathbf{x}_c - \mathbf{x}_s\| \leq \gamma_5 \|\mathbf{s}_-\| \cdot \|\mathbf{x}_c - \mathbf{x}_s\| ,
\end{aligned} \tag{3.11b}$$

and for \mathbf{y}_- defined by (2.8c),

$$\begin{aligned}
\|\mathbf{y}_- - J_s^T J_s \mathbf{s}_-\| &= \|J_c^T J_c \mathbf{s}_- + \alpha_c (J_c^T - J_-^T) F_c - J_s^T J_s \mathbf{s}_-\| \\
&\leq \|J_c^T J_c - J_s^T J_s\| \cdot \|\mathbf{s}_-\| + \|J_c^T F_c\| \cdot \|(J_c^T - J_-^T) F_c\| \\
&\leq \gamma_1 \|\mathbf{s}_-\| \cdot \|\mathbf{x}_c - \mathbf{x}_s\| + \gamma \|\mathbf{s}_-\| \cdot \|F_c\| \cdot \gamma_2 \cdot \|\mathbf{x}_c - \mathbf{x}_s\| \\
&\leq \gamma_1 \|\mathbf{s}_-\| \cdot \|\mathbf{x}_c - \mathbf{x}_s\| + \gamma_2^2 \|\mathbf{s}_-\| \cdot \|\mathbf{x}_c - \mathbf{x}_s\| \leq \gamma_5 \|\mathbf{s}_-\| \cdot \|\mathbf{x}_c - \mathbf{x}_s\| .
\end{aligned} \tag{3.11c}$$

Also, $\alpha_c \leq 1$, so for the three different \mathbf{y}_- definitions and $\gamma_6 \equiv \sigma_M^2 + \max\{\sigma_M^2, \gamma_2\}$,

$$\|\mathbf{y}_-\| \leq \|J_c^T J_c\| \cdot \|\mathbf{s}_-\| + \alpha_c \|J_c^T [F_c - F_-]\| \leq \sigma_M^2 \|\mathbf{s}_-\| + \sigma_M^2 \|\mathbf{s}_-\| \leq \gamma_6 \|\mathbf{s}_-\| , \tag{3.12a}$$

$$\|\mathbf{y}_-\| \leq \|J_c^T J_c\| \cdot \|\mathbf{s}_-\| + \alpha_c \|J_c^T F_c - J_-^T F_-\| \leq \sigma_M^2 \|\mathbf{s}_-\| + \gamma_2 \|\mathbf{s}_-\| \leq \gamma_6 \|\mathbf{s}_-\| , \tag{3.12b}$$

and

$$\|\mathbf{y}_-\| \leq \|J_c^T J_c\| \cdot \|\mathbf{s}_-\| + \alpha_c \|(J_c^T - J_-^T) F_c\| \leq \sigma_M^2 \|\mathbf{s}_-\| + \gamma_2 \|\mathbf{s}_-\| \leq \gamma_6 \|\mathbf{s}_-\| . \tag{3.12c}$$

Finally, for the three \mathbf{y}_- definitions, we have

$$\mathbf{s}_-^T \mathbf{y}_- = \mathbf{s}_-^T J_c^T J_c \mathbf{s}_- + \alpha_c \mathbf{s}_-^T J_c^T (F_c - F_-) \geq \mathbf{s}_-^T J_c^T J_c \mathbf{s}_- \geq \rho \lambda_s \|\mathbf{s}_-\|^2 \tag{3.13a}$$

from (3.5) and (3.7b), and

$$\mathbf{s}_-^T \mathbf{y}_- = \mathbf{s}_-^T J_c^T J_c \mathbf{s}_- + \alpha_c \mathbf{s}_-^T [J_c^T F_c - J_-^T F_-] \geq \mathbf{s}_-^T J_c^T J_c \mathbf{s}_- \geq \rho \lambda_* \|\mathbf{s}_-\|^2, \quad (3.13b)$$

since ϕ is convex on D . See Ortega-Rheinboldt (1970), pg 86. In the same manner,

$$\begin{aligned} \mathbf{s}_-^T \mathbf{y}_- &= \mathbf{s}_-^T J_c^T J_c \mathbf{s}_- + \alpha_c \mathbf{s}_-^T [J_c^T - J_-^T] F_c \pm \alpha_c \mathbf{s}_-^T J_-^T F_- \\ &= \mathbf{s}_-^T J_c^T J_c \mathbf{s}_- + \alpha_c \mathbf{s}_-^T [J_c^T F_c - J_-^T F_-] + \alpha_c \mathbf{s}_-^T J_-^T (F_c - F_-) \\ &\geq \mathbf{s}_-^T J_c^T J_c \mathbf{s}_- \geq \rho \lambda_* \|\mathbf{s}_-\|^2. \end{aligned} \quad (3.13c)$$

These, (3.12), and Lemma 3.1 allow us to say in either case that

$$\|I - \frac{\mathbf{y}_- \mathbf{s}_-^T}{\mathbf{s}_-^T \mathbf{y}_-}\| = \|\frac{\mathbf{y}_- \mathbf{s}_-^T}{\mathbf{s}_-^T \mathbf{y}_-}\| = \|\mathbf{y}_-\| \cdot \|\frac{\mathbf{s}_-}{\mathbf{s}_-^T \mathbf{y}_-}\| = \frac{\|\mathbf{y}_-\| \cdot \|\mathbf{s}_-\|}{\mathbf{s}_-^T \mathbf{y}_-} \leq \frac{\gamma_6}{\rho \lambda_*} \equiv \gamma_7. \quad (3.14)$$

It will also be useful to have $\gamma_8 \equiv \gamma_7^2 \gamma_1 + \gamma_7 \gamma_5 (1 + \gamma_7)$.

Now we prove the theorem by induction. Let $r \in (\frac{\gamma_*}{\lambda_*}, 1)$ and take

$$\epsilon < \min \left\{ \epsilon_1, \frac{r - \frac{\gamma_*}{\lambda_*}}{\frac{\gamma_4}{\lambda_*} + \gamma_3 \gamma_2}, \frac{r - \frac{\gamma_*}{\lambda_*}}{\frac{\gamma_4}{\lambda_*} + \gamma_7^2 \gamma_3 \gamma_2 + \gamma_5 \frac{\gamma_2}{\rho \lambda_*^2} (1 + \gamma_7)}, \frac{\lambda_*}{2 \gamma_8}, \frac{r - 2 \frac{\gamma_*}{\lambda_*}}{2 \cdot \frac{\gamma_4 + \gamma_8}{\lambda_*}} \right\}.$$

Given any $\mathbf{x}_0 \in N(\mathbf{x}_*; \epsilon)$, $J_0^T J_0$ is nonsingular by the choice of ϵ_1 , so \mathbf{x}_1 is well defined and from (3.10), (3.6) and (3.9) we have

$$\begin{aligned} \|\mathbf{x}_1 - \mathbf{x}_*\| &= \|\mathbf{x}_0 - \mathbf{x}_* - (J_0^T J_0)^{-1} J_0^T F_0\| \\ &\leq \|(J_0^T J_0)^{-1}\| \cdot \|J_0^T F_0 - J_0^T J_0 (\mathbf{x}_0 - \mathbf{x}_*)\| + \|(J_0^T J_0)^{-1} - (J_*^T J_*)^{-1}\| \cdot \|J_0^T F_0\| \\ &\leq \frac{1}{\lambda_*} \cdot [(\gamma_4 \|\mathbf{x}_0 - \mathbf{x}_*\|) + \gamma_*] \|\mathbf{x}_0 - \mathbf{x}_*\| + \gamma_3 \|\mathbf{x}_0 - \mathbf{x}_*\| \gamma_2 \|\mathbf{x}_0 - \mathbf{x}_*\| \\ &\leq [(\frac{\gamma_4}{\lambda_*} + \gamma_3 \gamma_2) \epsilon + \frac{\gamma_*}{\lambda_*}] \|\mathbf{x}_0 - \mathbf{x}_*\| \leq r \|\mathbf{x}_0 - \mathbf{x}_*\| < \epsilon. \end{aligned}$$

Suppose $\|\mathbf{x}_j - \mathbf{x}_*\| \leq r \|\mathbf{x}_{j-1} - \mathbf{x}_*\|$ for $j = 1, 2, \dots, k$. Then by Lemma 3.3, $J_k^T J_k$ is positive definite and this together with (3.13) ensures that H_k is positive definite in either case; so \mathbf{x}_{k+1} exists. Since either augmentation gives the Gauss-Newton step when $\alpha_k = 0$, we can concentrate on the augmented steps. First, we will carry out the induction step for the BFGS augmentation. By (2.10), (2.11), (2.12), (3.14), (3.10), (3.9), and (3.11),

$$\begin{aligned}
\|x_{k+1} - x_*\| &= \|x_k - x_* - H_k^{-1} J_k^T F_k \pm (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}})^T [(J_k^T J_k)^{-1} (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}) J_k^T F_k]\| \\
&= \|x_k - x_* - \left\{ (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}})^T [(J_k^T J_k)^{-1} \pm (J_*^T J_*)^{-1}] (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}) + \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}} \right\} J_k^T F_k\| \\
&= \|x_k - x_* - (J_*^T J_*)^{-1} J_k^T F_k - \left\{ (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}})^T [(J_k^T J_k)^{-1} - (J_*^T J_*)^{-1}] (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}) \right. \\
&\quad \left. + \frac{s_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}} + (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}})^T (J_*^T J_*)^{-1} (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}) - (J_*^T J_*)^{-1} \right\} J_k^T F_k\| \\
&\leq \|x_k - x_* - (J_*^T J_*)^{-1} J_k^T F_k\| + \left\| (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}})^T [(J_k^T J_k)^{-1} - (J_*^T J_*)^{-1}] (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}) \right\| \cdot \|J_k^T F_k\| \\
&\quad + \left\| \frac{[s_{k-1} - (J_*^T J_*)^{-1} y_{k-1}] s_{k-1}^T}{s_{k-1}^T y_{k-1}} + \frac{s_{k-1} [s_{k-1} - (J_*^T J_*)^{-1} y_{k-1}]^T}{s_{k-1}^T y_{k-1}} (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}) \right\| \cdot \|J_k^T F_k\| \\
&\leq \frac{\gamma_4 \|x_k - x_*\| + \gamma_* \|x_k - x_*\| + \gamma_7^2 \gamma_3 \gamma_2 \|x_k - x_*\|^2}{\lambda_*} \\
&\quad + \frac{1}{\lambda_*} \left[\frac{\gamma_5 \|s_{k-1}\| \cdot \|x_k - x_*\| \cdot \|s_{k-1}\|}{\rho \lambda_* \|s_{k-1}\|^2} + \frac{\|s_{k-1}\| \gamma_6 \|s_{k-1}\| \cdot \|x_k - x_*\|}{\rho \lambda_* \|s_{k-1}\|^2} \gamma_7 \gamma_2 \|x_k - x_*\| \right] \\
&\leq \left[\frac{\gamma_*}{\lambda_*} + \left\{ \frac{\gamma_4}{\lambda_*} + \gamma_7^2 \gamma_3 \gamma_2 + \gamma_5 \frac{\gamma_2}{\lambda_*^2 \rho} (1 + \gamma_7) \right\} \epsilon \right] \|x_k - x_*\| \\
&\leq r \|x_k - x_*\|.
\end{aligned} \tag{3.15a}$$

To complete the induction, let us consider the case when H_k is the DFP augmentation of $J_k^T J_k$. From (2.12b), (3.3), (3.12), (3.13), and (3.14),

$$\begin{aligned}
\|H_k - J_*^T J_*\| &= \left\| (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}) J_k^T J_k (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}})^T + \frac{y_{k-1} y_{k-1}^T}{s_{k-1}^T y_{k-1}} \right. \\
&\quad \left. \pm (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}) J_*^T J_* (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}})^T - J_*^T J_* \right\| \\
&= \left\| (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}}) (J_k^T J_k - J_*^T J_*) (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}})^T \right. \\
&\quad \left. + \frac{(y_{k-1} - J_*^T J_* s_{k-1}) y_{k-1}^T}{s_{k-1}^T y_{k-1}} + \frac{y_{k-1} (y_{k-1} - J_*^T J_* s_{k-1})^T}{s_{k-1}^T y_{k-1}} (I - \frac{y_{k-1} s_{k-1}^T}{s_{k-1}^T y_{k-1}})^T \right\| \\
&\leq \gamma_7^2 \|J_k^T J_k - J_*^T J_*\| + \frac{\|y_{k-1}\| \|y_{k-1} - J_*^T J_* s_{k-1}\|}{s_{k-1}^T y_{k-1}} \left(1 + \frac{\|s_{k-1}\| \|y_{k-1}\|}{s_{k-1}^T y_{k-1}} \right) \\
&\leq (\gamma_7^2 \gamma_1 + \gamma_7 \gamma_6 (1 + \gamma_7)) \|x_k - x_*\| \equiv \gamma_8 \|x_k - x_*\|.
\end{aligned}$$

Thus,

$$\|H_k - J_s^T J_s\| \leq \gamma_8 \epsilon < \frac{\lambda_s}{2}$$

and

$$\|(J_s^T J_s)^{-1}[H_k - J_s^T J_s]\| \leq \frac{1}{\lambda_s} \frac{\lambda_s}{2} = \frac{1}{2}.$$

Hence by the Banach Perturbation Lemma (Dennis and Schnabel (1983), pg.45),

$$\|H_k^{-1}\| \leq \frac{\|(J_s^T J_s)^{-1}\|}{1 - \|(J_s^T J_s)^{-1}[H_k - J_s^T J_s]\|} \leq \frac{1}{\lambda_s} \cdot \frac{1}{1 - \frac{1}{2}} = \frac{2}{\lambda_s}$$

Finally,

$$\begin{aligned} \|\mathbf{x}_{k+1} - \mathbf{x}_s\| &= \|\mathbf{x}_k - \mathbf{x}_s - H_k^{-1} J_k^T F_k\| \\ &\leq \| -H_k^{-1} \| \cdot \| J_k^T F_k - J_s^T J_s(\mathbf{x}_k - \mathbf{x}_s) - (H_k - J_s^T J_s)(\mathbf{x}_k - \mathbf{x}_s) \| \\ &\leq \frac{2}{\lambda_s} [(\gamma_4 \|\mathbf{x}_k - \mathbf{x}_s\| + \gamma_s) \|\mathbf{x}_k - \mathbf{x}_s\| + \gamma_8 \|\mathbf{x}_k - \mathbf{x}_s\|^2] \\ &\leq \frac{2}{\lambda_s} [\gamma_s + (\gamma_4 + \gamma_8)\epsilon] \cdot \|\mathbf{x}_k - \mathbf{x}_s\| \leq r \|\mathbf{x}_k - \mathbf{x}_s\|. \end{aligned} \tag{3.15b}$$

This completes the proof of q-linear convergence. If $\gamma_s = 0$, then q-quadratic convergence follows immediately from (3.15). Of course, if $\|F(\mathbf{x}_s)\| = 0$, then $\gamma_s = 0$.

4. The Hybrid Algorithm.

In this section, we will describe a model switching strategy and give details of a hybrid algorithm that adaptively decides at each iteration whether to use the Gauss-Newton model (2.3) or a BFGS augmented model (2.6), (2.9a). Much of the implementation follows MINPACK and NL2SOL. Except for model switching, the basic algorithm is essentially the same as in the subroutine LMDER of MINPACK. The model switching decision is borrowed from the ideas in the subroutine NL2S1 of NL2SOL. Among the major differences between our hybrid algorithm and the one used in NL2SOL are:

- (1) No doubling of the trust radius is tried internal to an iteration.
- (2) If five unsuccessful steps are attempted in an iteration with the currently preferred model, then the algorithm will start the next iteration with the other model.

Before presenting the algorithm, we make some preliminary remarks that will also serve to define the variables in the algorithm. The actual reduction *ared*, the predicted reduction *pred*, and the directional derivative *dirder* are computed as in the subroutine LMDER. Moreover, the updating strategies for the step bound Δ_k and the Levenberg-Marquardt parameter λ are also taken from that routine.

The extra work involved in trying the alternate model in any iteration after the currently preferred model step has been computed is not significant. To see this, suppose that the alternative is the BFGS augmented model. We would already have the factorization

$$J_k = Q \cdot R ,$$

and if we follow Goldfarb (1976), Dennis-Schnabel (1981) or (1983, pp.200-201), and define

$$J_{BFGS} = J_k + uv^T ,$$

$$\text{where } u = \frac{J_k s_-}{\|J_k s_-\|} , v = \frac{(y_- - t \cdot J_k^T J_k s_-)}{\sqrt{y_-^T s_-}} , \text{ and } t = \frac{\sqrt{y_-^T s_-}}{\sqrt{s_-^T J_k^T J_k s_-}} , \quad (4.1)$$

then $H_k^{BFGS} = J_{BFGS}^T J_{BFGS}$, and it is cheap to obtain the QR decomposition of J_{BFGS} by the following well known means:

$$J_{BFGS} = Q \cdot R + uv^T \equiv Q(R + uv^T) = Q \cdot \bar{Q} \cdot \bar{R} \equiv \bar{Q} \cdot \bar{R} .$$

Similarly, in the case that the iteration starts with the BFGS augmented model and the alternate is the Gauss-Newton model, the QR factorization of J_k can be obtained easily from the QR factorization of J_{BFGS} . For complete details, see Dennis-Schnabel (1983, pg.57). In the algorithm below, we will use the notation J_a and J_c to denote the Jacobian or J_{BFGS} as is appropriate depending on which is the alternate and which is the currently preferred model. We will use m_a and m_c for the quadratic models themselves.

The Hybrid Algorithm

- (1) Initialize $k = 1$, $\lambda_k = 0$, $s_0 = 0 = y_0$, and Δ_k .
 Set Current_model = Gauss-Newton; Alternate_model = BFGS.
 Compute $F_k = F(x_k)$, $Fnorm = \|F_k\|$, and $phik = \|F_k\|^2$.

- (2) Set $first = .TRUE.$, $ifail = 0$ and compute $J_k = J(x_k)$.
 Rescale D_k if necessary.
If $Current_model = Gauss\text{-}Newton$ **Then** $J_c = J_k$
Else $J_c = J_{BFGS}$
End If
- (3) Solve for s to minimize $\|F_k + J_k s\|$, subject to $\|D_k s\| \leq \Delta_k$.
- (4) Set $x_{k+1}^p = x_k + s$ and $pnorm = \|D_k s\|$.
 Compute $F(x_{k+1}^p)$ and $Fnorm1 = \|F(x_{k+1}^p)\|$, $phikp = \|F(x_{k+1}^p)\|^2$.
- (5) Compute:
- $$ared = 1 - (Fnorm1/Fnorm)^2;$$
- $$pred = (\|J_k s\|/Fnorm)^2 + 2 \cdot (\lambda_k^{1/2} \cdot pnorm / Fnorm)^2;$$
- $$dirder = -[(\|J_k s\|/Fnorm)^2 + (\lambda_k^{1/2} \cdot pnorm / Fnorm)^2];$$
- $$rho = ared / pred.$$

- (6) **If** $\rho \geq .25$ **Then** (* Probably a good step.)
 If $\rho \geq .75$ or $\lambda_k = 0$ **Then** (* It was a good step.) $\Delta_{k+1} = 2 \cdot \Delta_k$, $\lambda_{k+1} = \lambda_k / 2$
 Else $\Delta_{k+1} = \Delta_k$, $\lambda_{k+1} = \lambda_k$
 End If
Else (* Probably a bad step.)
 If $first = .TRUE.$ **Then**
 $first = .FALSE.$,
 Compute J_k (* See the discussion before the algorithm.)
 If $|m_c(x_{k+1}^p) - phikp| > 1.5 \cdot |m_a(x_{k+1}^p) - phikp|$ **Then**
 Solve for s to minimize $\|F_k + J_a s\|$
 subject to $\|D_k s\| \leq \Delta_k$.
 Set $x_{k+1}^p = x_k + s$ and $pnorm2 = \|D_k s\|$.
 Compute $F(x_{k+1}^p)$, $Fnorm2 = \|F(x_{k+1}^p)\|$,
 and $phikpp = \|F(x_{k+1}^p)\|^2$.
 If $phikpp < phikp$ **Then**
 $x_{k+1}^p = x_{k+1}^p$,
 $F(x_{k+1}^p) = F(x_{k+1}^p)$, $J_k = J_a$,
 $pnorm = pnorm2$, $Fnorm1 = Fnorm2$, $phikp = phikpp$.
 Go to (5) (* Check fit of alternate model.)
 End If.
 End If.
 End If.
 If $ared \geq 0$ **Then** $temp = .5$
 Else $temp = .5 \cdot dirder / (dirder + .5 \cdot ared)$
 End If.
 If $(Fnorm1 \geq 10 \cdot Fnorm$ or $temp < .1)$ **Then** $temp = .1$
 End If.
 $\Delta_{k+1} = temp \cdot \min\{\Delta_k, 10 \cdot pnorm\}$
 $\lambda_{k+1} = \lambda_k / temp$
 End If.
- (7) Check for convergence.

- (8) **If** $\rho < 10^{-4}$ **Then** (* Bad step; compute another.)
 $ifail = ifail + 1$
 If $ifail = 5$ **Then**
 Exchange model preferences for next iteration
 Go to (2)
 End If.
 Go to (3).
End If.
- (9) Set $x_{k+1} = x_{k+1}^p$, $F_{k+1} = F(x_{k+1}^p)$, $Fnorm = Fnorm1$, $phik = phikp$.
If $|m_c(x_{k+1}) - phikp| > 1.5 \cdot |m_s(x_{k+1}) - phikp|$ **Then**
 Exchange model preferences for next iteration
End If.
- (10) Set $k = k + 1$, Go to (2).

5. Numerical Results.

In this section, we will try to compare a preliminary implementation of our hybrid algorithm NONLSQ with the MINPACK subroutine LMDER and the NL2SOL subroutine NL2S1. Our implementation is as close as possible to LMDER since we felt that the hybrid algorithm should be viewed as a way to modify the Gauss-Newton method on large residual problems. In fact, we didn't even tune any of the constants involved in the LMDER code to find values that gave better performance with our hybrid algorithm. All the runs were done in double precision on the Data General MV-1000 at IMSL using stopping criteria of $1.0D-8$ on the relative stepsize or the relative change in the sum of squares. The 26 problems used in our test were:

Problems 1-18 were the MINPACK-1 nonlinear least-squares test problems.

Problems 19-26 were the NL2SOL test problems that were not included in the MINPACK test set. They were: 19.) Woods' function, 20.) Zangwill's function, 21.) Engvall's function, 22.) Branin's function, 23.) Beale's function, 24.) Cragg's and Levy's function, 25.) the Davidon 1 function, 26.) Madsen's function.

The results for LMDER and NONLSQ were identical for most of the problems and so we only present the results for Problems 7, 8, 9, 14, 15, 18, 26. These results are summarized in Tables 1-8 where we follow the practice of using the 'standard' starting point for the problem the first time the problem is listed, ten times that standard point if the problem is listed again, one hundred times the standard starting point if there is a third listing, etc. For the problems listed, we continue multiplying the standard initial guess by powers of ten until the initial point is too large for us to even be able to compute the initial residuals. Convergence was never a problem.

These numerical results support the conclusions that for poor initial guesses and nonzero residuals, the method suggested here may enjoy some advantages over the Levenberg-Marquardt method. Therefore, further investigation seems justified; for example, we do not allow internal doubling, and we have made use of the freedom in the parameter α_c only enough to show that it does affect performance.

The results in Table 7 show that (2.8c) with $\alpha_c = \min\{1, \|J_c^T F_c\|\}$ is probably the best choice of secant condition. We included $\alpha_c = 1$ because of the relationship to ideas of Al-Baali

and Fletcher (1983) mentioned in Section 2. The comparison of NONLSQ to NL2S1 is much harder due to the differences in the implementation details, and so we will just present the NL2S1 results in Table 2 for completeness.

We hope that refinement of this simple and computationally convenient idea will lead to improved algorithms for large residual nonlinear least squares.

Table 1: Gauss-Newton Results

NPROB	N	M	NFEV	NJEV	FINAL L2 NORM
7	2	2	14	8	.6998875E+01
7	2	2	19	12	.6998875E+01
7	2	2	24	17	.6998875E+01
7	2	2	31	25	.6998875E+01
7	2	2	36	29	.6998875E+01
8	3	15	6	5	.9063596E-01
8	3	15	37	36	.4174769E+01
8	3	15	14	13	.4174769E+01
8	3	15	5	4	.4174769E+01
8	3	15	5	4	.4174769E+01
9	4	11	18	16	.1753584E-01
9	4	11	78	70	.3205219E-01
9	4	11	524	390	.1753584E-01
9	4	11	804	692	.1753584E-01
9	4	11	2112	1628	.3073494E-01
14	4	20	266	247	.2929543E+03
14	4	20	56	44	.2929543E+03
14	4	20	253	236	.2929543E+03
14	4	20	226	211	.2929543E+03
14	4	20	495	475	.2929543E+03
15	1	8	1	1	.1886238E+01
15	1	8	29	28	.1884248E+01
15	1	8	47	46	.1884248E+01
15	1	8	64	63	.1884248E+01
15	1	8	81	80	.1884248E+01
18	11	65	17	13	.2003440E+00
18	11	65	13	10	.1337839E+01
21	3	5	25	15	.2559996E+01
21	3	5	31	19	.2559996E+01
21	3	5	2794	2741	.2559997E+01
26	2	3	29	27	.8793174E+00
26	2	3	27	25	.8793174E+00
26	2	3	39	37	.8793174E+00
26	2	3	27	25	.8793174E+00
26	2	3	49	42	.8793174E+00

Table 2: NL2SOL Results

NPROB	N	M	NFEV	NJEV	FINAL L2 NORM
7	2	2	10	9	.6998875E+01
7	2	2	21	16	.6998875E+01
7	2	2	37	27	.6998875E+01
7	2	2	52	39	.6998875E+01
7	2	2	2	2	.1131484E+14
8	3	15	7	7	.9063596E-01
8	3	15	43	29	.4174773E+01
8	3	15	23	15	.4174769E+01
8	3	15	11	7	.4174769E+01
8	3	15	72	33	.9063596E-01
9	4	11	11	10	.1753584E-01
9	4	11	51	38	.1753584E-01
9	4	11	152	108	.1753584E-01
9	4	11	228	172	.1753584E-01
9	4	11	191	135	.3073309E-01
14	4	20	21	15	.2929543E+03
14	4	20	29	19	.2929543E+03
14	4	20	32	21	.2929543E+03
14	4	20	38	24	.2929543E+03
14	4	20	2	2	.6188207E+12
15	1	8	1	1	.1886238E+01
15	1	8	60	48	.1884248E+01
15	1	8	2	1	.1180887E+19
15	1	8	2	1	.1269793E+27
15	1	8	2	1	.1278976E+35
18	11	65	20	14	.2003440E+00
18	11	65	19	13	.1337839E+01
21	3	5	13	11	.2559996E+01
21	3	5	30	21	.2559996E+01
21	3	5	48	33	.2559996E+01
26	2	3	13	11	.8793174E+00
26	2	3	19	13	.8793174E+00
26	2	3	34	22	.8793174E+00
26	2	3	35	23	.8793174E+00
26	2	3	48	30	.8793174E+00

Table 3: NONLSQ with (2.8a), $\alpha = \min(1, |JtF|)$

NPROB	N	M	NFEV	NJEV	FINAL L2 NORM
7	2	2	14	8	.6998875E+01
7	2	2	24	16	.6998875E+01
7	2	2	24	17	.6998875E+01
7	2	2	31	25	.6998875E+01
7	2	2	37	29	.6998875E+01
8	3	15	6	5	.9063596E-01
8	3	15	37	36	.4174769E+01
8	3	15	13	12	.4174769E+01
8	3	15	5	4	.4174769E+01
8	3	15	5	4	.4174769E+01
9	4	11	18	16	.1753584E-01
9	4	11	78	70	.3205219E-01
9	4	11	440	374	.1753584E-01
9	4	11	790	686	.1753584E-01
9	4	11	2038	1578	.3073494E-01
14	4	20	328	309	.2929543E+03
14	4	20	65	53	.2929543E+03
14	4	20	265	247	.2929543E+03
14	4	20	222	206	.2929543E+03
14	4	20	468	439	.2929543E+03
15	1	8	1	1	.1886238E+01
15	1	8	29	28	.1884248E+01
15	1	8	47	46	.1884248E+01
15	1	8	64	63	.1884248E+01
15	1	8	81	80	.1884248E+01
18	11	65	17	13	.2003440E+00
18	11	65	15	12	.1337839E+01
21	3	5	24	15	.2559996E+01
21	3	5	46	32	.2559996E+01
21	3	5	2675	2568	.2559996E+01
26	2	3	29	27	.8793174E+00
26	2	3	27	25	.8793174E+00
26	2	3	40	35	.8793174E+00
26	2	3	46	40	.8793174E+00
26	2	3	48	42	.8793174E+00

Table 4: NONLSQ with (2.8a) and $\alpha = 1.0$

NPROB	N	M	NFEV	NJEV	FINAL L2 NORM
7	2	2	17	11	.6998875E+01
7	2	2	22	15	.6998875E+01
7	2	2	24	17	.6998875E+01
7	2	2	32	25	.6998875E+01
7	2	2	37	29	.6998875E+01
8	3	15	6	5	.9063596E-01
8	3	15	37	36	.4174769E+01
8	3	15	13	12	.4174769E+01
8	3	15	5	4	.4174769E+01
8	3	15	5	4	.4174769E+01
9	4	11	12	10	.1753584E-01
9	4	11	88	73	.3205219E-01
9	4	11	395	350	.1753584E-01
9	4	11	706	639	.1753584E-01
9	4	11	1946	1533	.3073494E-01
14	4	20	328	309	.2929543E+03
14	4	20	65	53	.2929543E+03
14	4	20	265	247	.2929543E+03
14	4	20	222	206	.2929543E+03
14	4	20	468	439	.2929543E+03
15	1	8	1	1	.1886238E+01
15	1	8	29	28	.1884248E+01
15	1	8	47	46	.1884248E+01
15	1	8	64	63	.1884248E+01
15	1	8	81	80	.1884248E+01
18	11	65	18	14	.2003440E+00
18	11	65	17	14	.1337839E+01
21	3	5	24	15	.2559996E+01
21	3	5	46	32	.2559996E+01
21	3	5	2675	2568	.2559996E+01
26	2	3	29	27	.8793174E+00
26	2	3	27	25	.8793174E+00
26	2	3	27	22	.8793174E+00
26	2	3	34	28	.8793174E+00
26	2	3	48	42	.8793174E+00

Table 5: NONLSQ with (2.8b), $\alpha = \min(1, |JtF|)$

NPROB	N	M	NFEV	NJEV	FINAL L2 NORM
7	2	2	15	8	.6998875E+01
7	2	2	19	12	.6998875E+01
7	2	2	25	17	.6998875E+01
7	2	2	32	25	.6998875E+01
7	2	2	37	29	.6998875E+01
8	3	15	6	5	.9063596E-01
8	3	15	37	36	.4174769E+01
8	3	15	14	13	.4174769E+01
8	3	15	5	4	.4174769E+01
8	3	15	5	4	.4174769E+01
9	4	11	18	16	.1753584E-01
9	4	11	78	70	.3205219E-01
9	4	11	388	334	.1753584E-01
9	4	11	800	693	.1753584E-01
9	4	11	2112	1628	.3073494E-01
14	4	20	121	81	.2929543E+03
14	4	20	49	32	.2929543E+03
14	4	20	233	176	.2929543E+03
14	4	20	186	141	.2929543E+03
14	4	20	450	404	.2929543E+03
15	1	8	1	1	.1886238E+01
15	1	8	29	28	.1884248E+01
15	1	8	47	46	.1884248E+01
15	1	8	64	63	.1884248E+01
15	1	8	81	80	.1884248E+01
18	11	65	17	13	.2003440E+00
18	11	65	13	10	.1337839E+01
21	3	5	30	17	.2559996E+01
21	3	5	33	22	.2559996E+01
21	3	5	2722	2619	.2559997E+01
26	2	3	29	26	.8793174E+00
26	2	3	33	31	.8793174E+00
26	2	3	36	34	.8793174E+00
26	2	3	27	25	.8793174E+00
26	2	3	56	49	.8793174E+00

Table 6: NONLSQ with (2.8b) and $\alpha = 1.0$

NPROB	N	M	NFEV	NJEV	FINAL L2 NORM
7	2	2	17	9	.6998875E+01
7	2	2	20	12	.6998875E+01
7	2	2	26	17	.6998875E+01
7	2	2	33	25	.6998875E+01
7	2	2	37	29	.6998875E+01
8	3	15	6	5	.9063596E-01
8	3	15	37	36	.4174769E+01
8	3	15	14	13	.4174769E+01
8	3	15	5	4	.4174769E+01
8	3	15	5	4	.4174769E+01
9	4	11	11	9	.1753584E-01
9	4	11	81	71	.3205219E-01
9	4	11	175	132	.3072655E-01
9	4	11	926	775	.1753584E-01
9	4	11	2159	1661	.3073494E-01
14	4	20	121	81	.2929543E+03
14	4	20	49	32	.2929543E+03
14	4	20	233	176	.2929543E+03
14	4	20	186	141	.2929543E+03
14	4	20	450	404	.2929543E+03
15	1	8	1	1	.1886238E+01
15	1	8	29	28	.1884248E+01
15	1	8	47	46	.1884248E+01
15	1	8	64	63	.1884248E+01
15	1	8	81	80	.1884248E+01
18	11	65	17	13	.2003440E+00
18	11	65	13	10	.1337839E+01
21	3	5	31	16	.2559996E+01
21	3	5	37	25	.2559996E+01
21	3	5	2618	2531	.2559996E+01
26	2	3	15	11	.8793174E+00
26	2	3	18	16	.8793174E+00
26	2	3	23	21	.8793174E+00
26	2	3	24	22	.8793174E+00
26	2	3	42	35	.8793174E+00

Table 7: NONLSQ with (2.8c), $\alpha = \min(1, |JtF|)$

NPROB	N	M	NFEV	NJEV	FINAL L2 NORM
7	2	2	18	10	.6998875E+01
7	2	2	23	13	.6998875E+01
7	2	2	31	22	.6998875E+01
7	2	2	32	26	.6998875E+01
7	2	2	44	34	.6998875E+01
8	3	15	7	6	.9063596E-01
8	3	15	37	36	.4174769E+01
8	3	15	14	13	.4174769E+01
8	3	15	5	4	.4174769E+01
8	3	15	5	4	.4174769E+01
9	4	11	18	16	.1753584E-01
9	4	11	79	70	.3205219E-01
9	4	11	388	334	.1753584E-01
9	4	11	801	689	.1753584E-01
9	4	11	2113	1628	.3073494E-01
14	4	20	72	49	.2929543E+03
14	4	20	47	33	.2929543E+03
14	4	20	189	151	.2929543E+03
14	4	20	190	148	.2929543E+03
14	4	20	336	303	.2929543E+03
15	1	8	1	1	.1886238E+01
15	1	8	29	28	.1884248E+01
15	1	8	47	46	.1884248E+01
15	1	8	64	63	.1884248E+01
15	1	8	81	80	.1884248E+01
18	11	65	17	13	.2003440E+00
18	11	65	13	10	.1337839E+01
21	3	5	25	14	.2559996E+01
21	3	5	46	30	.2559996E+01
21	3	5	1587	1504	.2559997E+01
26	2	3	29	27	.8793174E+00
26	2	3	25	24	.8793174E+00
26	2	3	37	35	.8793174E+00
26	2	3	42	39	.8793174E+00
26	2	3	40	39	.8793174E+00

Table 8: NONLSQ with (2.8c) and $\alpha = 1.0$

NPROB	N	M	NFEV	NJEV	FINAL L2 NORM
7	2	2	19	10	.6998875E+01
7	2	2	25	13	.6998875E+01
7	2	2	32	22	.6998875E+01
7	2	2	33	26	.6998875E+01
7	2	2	47	34	.6998875E+01
8	3	15	7	6	.9063596E-01
8	3	15	37	36	.4174769E+01
8	3	15	14	13	.4174769E+01
8	3	15	5	4	.4174769E+01
8	3	15	5	4	.4174769E+01
9	4	11	13	11	.1753584E-01
9	4	11	101	84	.3205219E-01
9	4	11	510	424	.1753584E-01
9	4	11	865	742	.1753584E-01
9	4	11	2126	1637	.3073494E-01
14	4	20	72	49	.2929543E+03
14	4	20	47	33	.2929543E+03
14	4	20	189	151	.2929543E+03
14	4	20	190	148	.2929543E+03
14	4	20	336	303	.2929543E+03
15	1	8	1	1	.1886238E+01
15	1	8	29	28	.1884248E+01
15	1	8	47	46	.1884248E+01
15	1	8	64	63	.1884248E+01
15	1	8	81	80	.1884248E+01
18	11	65	18	14	.2003440E+00
18	11	65	13	10	.1337839E+01
21	3	5	27	16	.2559996E+01
21	3	5	46	30	.2559996E+01
21	3	5	1442	1366	.2559996E+01
26	2	3	29	27	.8793174E+00
26	2	3	16	15	.8793174E+00
26	2	3	37	35	.8793174E+00
26	2	3	30	27	.8793174E+00
26	2	3	28	27	.8793174E+00

References

- (1) M.Al-Baali and R.Fletcher (1983), "Variational methods for nonlinear least squares", Mathematics Report NA/71, University of Dundee.
- (2) J.E.Dennis Jr. (1977), "Nonlinear least squares and equations", in *The State of the Art in Numerical Analysis*, D.Jacobs, Ed., Academic Press, London, pp.269-312.
- (3) J.E.Dennis Jr., D.M.Gay, and R.E.Welsch (1981a), "An adaptive nonlinear least-squares algorithm", *TOMS* 7, pp.348-368.
- (4) J.E.Dennis Jr., D.M.Gay, and R.E.Welsch (1981b), "Algorithm 573 NL2SOL - An adaptive nonlinear least-squares algorithm [E4]", *TOMS* 7, pp.369-383.
- (5) J.E. Dennis Jr. and R.B. Schnabel (1981), "A new derivation of symmetric positive definite secant updates," in *Nonlinear Programming 4*, ed. by O.L.Mangasarian, R.R.Meyer and S.M.Robinson, Academic Press, NY, pp167-199.
- (6) J.E.Dennis Jr. and R.B.Schnabel (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ.
- (7) D.Goldfarb (1976), "Factorized variable metric methods for unconstrained optimization", *Math. Comp.* 30, pp.796-811.
- (8) J.J.More (1977), "The Levenberg-Marquardt algorithm: implementation and theory", in *Numerical Analysis*, G.A.Watson, Ed., Lecture Notes in Math. 630, Springer Verlag, Berlin, pp.105-116.