TENSOR METHODS FOR
NONLINEAR EQUATIONS

Robert B. Schnabel*
Paul D. Frank

Department of Computer Science
University of Colorado
Boulder, Colorado 80309

CU-CS-243-83 April, 1983

*This research supported by ARO contract DAAG 29-81-K-0108
Abstract

A new class of methods for solving systems of nonlinear equations, called tensor methods, is introduced. Tensor methods are general purpose methods intended especially for problems where the Jacobian matrix at the solution is singular or ill-conditioned. They base each iteration on a quadratic model of the nonlinear function, the standard linear model augmented by a simple second order term. The second order term is selected so that the model interpolates function values from several previous iterations, as well as the current function value and Jacobian. The tensor method requires no more function and derivative information per iteration, and hardly more storage or arithmetic per iteration, than a standard method based on Newton’s method. In extensive computational tests, a tensor algorithm is significantly more efficient than a similar algorithm based on the standard linear model, both on standard nonsingular test problems and on problems where the Jacobian at the solution is singular.
1. Introduction

This paper introduces a new class of methods, tensor methods, for solving the nonlinear equations problem

\[ F : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad \text{find } x^* \in \mathbb{R}^n \text{ such that } F(x^*) = 0 \quad (1.1) \]

where it is assumed that \( F(x) \) is at least once continuously differentiable. The novel feature of these methods is that they base each iteration on a quadratic model of \( F(x) \) whose second order term has a special, restricted, form. Tensor methods are especially intended to improve upon the performance of standard methods on problems where the Jacobian matrix of \( F \) at \( x^* \), \( F'(x^*) \in \mathbb{R}^{n \times n} \), is singular or ill-conditioned. At the same time, they are intended to be at least as efficient as standard methods on problems where \( F'(x^*) \) is nonsingular. Their storage requirements and arithmetic operations per iteration are not significantly higher than the requirements of standard methods.

Standard methods for solving (1.1) base each iteration upon a linear model \( M(x) \) of \( F(x) \) around the current iterate \( x_c \in \mathbb{R}^n \),

\[ M(x_c + d) = F(x_c) + J_c d \quad (1.2) \]

where \( d \in \mathbb{R}^n \), \( J_c \in \mathbb{R}^{n \times n} \). These methods can be divided into two classes, those where \( J_c \) is the current Jacobian matrix \( F'(x_c) \) or a finite difference approximation to it, and those where \( J_c \) is a secant (quasi-Newton) approximation to the Jacobian. In this paper we propose extensions to the first type of methods, those that use analytic or finite difference Jacobians, because this is the most basic setting in which to study the new ideas of this paper. In subsequent papers, we intend to extend tensor methods to secant methods for nonlinear equations, and to unconstrained optimization.

When the analytic Jacobian is available, the linear model (1.2) becomes

\[ M(x_c + d) = F(x_c) + F'(x_c) d \quad (1.3) \]

The most basic method for nonlinear equations, Newton's method, is defined
when $F'(x_c)$ is nonsingular, and consists of setting the next iterate $x_i$ to the root of (1.3),

$$x_i = x_c - F'(x_c)^{-1}F(x_c).$$

(1.4)

The distinguishing feature of Newton’s method is that if $F'(x_c)$ is Lipschitz continuous in a neighborhood containing the root $x_*$ and $F'(x_*)$ is nonsingular, then the sequence of iterates produced by (1.4) converges locally and q-quadratically to $x_*$. This means that there exist $\delta > 0$ and $c \geq 0$ such that the sequence of iterates $\{x_k\}$ produced by Newton’s method obeys

$$||x_{k+1} - x_*|| \leq c ||x_k - x_*||^q$$

if $||x_0 - x_*|| \leq \delta$. In practice, local q-quadratic convergence means eventual fast convergence.

Newton’s method is not usually quickly locally convergent, however, if $F'(x_*)$ is singular. For example when applied to one equation in one unknown ($n=1$) where $f'(x_*)=0$ but $f''(x_*) \neq 0$, Newton’s method is locally q-linearly convergent with constant converging to $\frac{1}{2}$, meaning that the sequence of iterates $\{x_k\}$ obeys

$$|x_{k+1} - x_*| = c_k |x_k - x_*|, \quad \lim_{k \to \infty} c_k = \frac{1}{2}$$

if $|x_0 - x_*|$ is sufficiently small. For systems of equations, the situation is more complex and has been analyzed by many authors, including Decker and Kelley [1980a, 1980b, 1982], Decker, Keller, and Kelley [1982], Griewank [1980a, 1980b, 1983], Griewank and Osborne [1981], Keller [1970], Kelley and Suresh [1982], Rall [1966], and Reddien [1978, 1980]. In summary, their papers show that from many starting points, Newton’s method for systems of equations also is locally q-linearly convergent with constant converging to $\frac{1}{2}$ although from some starting points arbitrarily close to $x_*$, (1.4) may be repulsive. In practice, Newton’s method usually exhibits local linear convergence with constant $\approx \frac{1}{2}$ on singular problems (see Table 6.6), much slower convergence than one would like.
Several of the above mentioned papers, for example Decker, Keller, and Kelley [1982] and Griewank [1980a, 1983], propose methods that are rapidly convergent on some singular problems. Most of these methods are related to the one dimensional acceleration technique of taking $j$ times the Newton step if one has a $j^{th}$ order singularity. This requires deciding whether the problem is singular, which probably makes such methods unsuitable for general purpose use. The major aim of this paper is to provide a general purpose method that has rapid local convergence even when $F'(x_*)$ is singular. In addition, tensor methods usually will not experience any special difficulty when $F'(x_*)$ is singular or ill-conditioned, while methods based on (1.3) must be modified in this situation.

Systems of nonlinear equations with $F'(x_*)$ singular or ill-conditioned occur in a number of important practical situations. For example, conservation laws in stiff systems of ordinary differential equations sometimes cause the Jacobian of the associated system of nonlinear equations to be very nearly singular for all $x$. In curve tracing problems it also is not uncommon to generate systems of nonlinear equations with nearly singular Jacobians. In unconstrained minimization problems arising from data fitting, the Hessian matrix $\nabla^2 f(x_*)$ usually is singular if the problem is over-parameterized, and often $\nabla^2 f(x_*)$ is ill-conditioned because the data fitting model is far more sensitive to some parameters than to others. In all these cases, it is important to notice that the (near) rank deficiency in the derivative matrix usually is small. This is the case in which our methods are intended to improve upon standard methods. Our methods are not intended for problems where the rank of $F'(x_*)$ is small in comparison to its dimension, although sometimes they are effective in this case in practice.

The other well-known disadvantage of Newton's method is that it may not converge to any root $x_*$ if it is started too far from any root. The main remedies used in practice are augmenting (1.4) by line search or trust region algorithms,
see for example Fletcher [1980], Gill, Murray, and Wright [1981], or Dennis and Schnabel [1983]. Our new methods use the same strategies when the new local step is unsatisfactory.

It is important to consider the costs associated with solving systems of nonlinear equation by standard methods. These can be divided into three types: the arithmetic operations required by the algorithm (excluding function and derivative evaluations), the storage required, and the evaluations of the nonlinear function $F(x)$ and the Jacobian $F'(x)$, if it is provided. For algorithms that use an analytic or finite difference Jacobian, the dominant arithmetic cost is one matrix factorization per iteration, requiring $n³/3$ (for LU) or $2n³/3$ (for QR) additions and multiplications per iteration. At least $n²$ storage is required, for the Jacobian, and some algorithms store a second $n×n$ matrix. Finally, $F(x)$ must be evaluated at least once per iteration; in addition, either $F''(x)$ is evaluated once per iteration, or it is approximated by finite differences, requiring up to $n$ additional evaluations of $F(x)$ per iteration. In many practical problems, the evaluations of $F(x)$ and $F'(x)$ are expensive and dominate the other costs. The main efficiency goal of our new method is to decrease the number of function and derivative evaluations required to solve systems of nonlinear equations; however, no substantial increase in the arithmetic cost per iteration, or in the storage requirements, will be permitted.

Our new methods are based on expanding the linear model (1.3) of $F(x)$ around $x₀$ to the quadratic model

$$M_F(x₀+d) = F(x₀) + F'(x₀)d + \frac{1}{2}T_c dd$$

where $T_c ∈ R^{n×n×n}$. The three dimensional object $T_c$ often is referred to as a tensor, hence we call (1.5) a tensor model, and methods based upon (1.5) tensor methods. We define the notation $T_c dd$ used in (1.5) before proceeding.
Definition 1.1. Let $T \in R^{n \times n \times n}$. Then $T$ is composed of $n$ horizontal faces $H_i \in R^{n \times n}$, $i = 1, \ldots, n$, where $H_i[j,k] = T[i,j,k]$. For $u, w \in R^n$, $Tuw \in R^n$ with

$$Tuw[i] = u^T H_i w = \sum_{j=1}^{n} \sum_{k=1}^{n} T[i,j,k]u[j]w[k].$$

Note that $M_T(x_o + d)$ is simply the $n$-vector of $n$ quadratic models of the component functions of $F(x)$,

$$(M_T(x_o + d))[i] = f_i + g_i^T d + \frac{1}{2} d^T H_i d, \quad i = 1, \ldots, n$$

where $f_i = F(x_o)[i]$, $g_i^T$ = row $i$ of $F'(x_o)$, and $H_i$ is the Hessian matrix of the $i^{th}$ component function of $F(x)$, or an approximation to it.

The obvious choice of $T_c$ in (1.5) is the matrix $F''(x_o)$ of second partial derivatives of $F$ at $x_o$; this makes (1.5) the first three terms of the Taylor series expansion of $F$ around $x_o$. Several serious disadvantages, however, make (1.5) with $T_c = F''(x_o)$ unacceptable for algorithmic use. They include:

(1) The $n^3$ second partial derivatives of $F$ at $x_o$ would have to be computed at each iteration.

(2) The model would take more than $n^3/2$ locations to store.

(3) To find a root of the model, at each iteration one would have to solve a system of $n$ quadratic equations in $n$ unknowns.

(4) The model might not have a real root.

To use a model of form (1.5) and avoid these disadvantages, our tensor method uses a very restricted form of $T_c$. In particular, our tensor model requires no additional derivative or function information; the additional costs of forming and solving our tensor model are small compared to the $O(n^3)$ arithmetic cost per iteration of standard methods; and the additional storage required for our tensor model is small compared to the $n^2$ storage required for the Jacobian. The key contribution of this paper is showing how one may construct a useful quadratic model that satisfies these criteria. Our tensor model
still does not always have a real root, and we will address this issue.

The remainder of the paper is organized as follows. In section 2, we discuss briefly the specialization of our tensor method to one nonlinear equation in one unknown. Of course, when \( n = 1 \), many of the disadvantages stated above for a second order Taylor series model are irrelevant, and indeed, various methods for solving a single nonlinear equation are based on using quadratic models. The material in section 2 is included only to motivate our multi-variable method.

The heart of the paper, our techniques for forming and solving the tensor model for systems of nonlinear equations, is contained in sections 3 and 4, respectively. The full tensor algorithm is presented in section 5 and various implementation considerations are discussed. In section 6 we present test results of our tensor method on the problems of Moré, Garbow, and Hillstrom [1981], and on modifications of these problems constructed so that \( F'(x_k) \) is singular. We compare our tensor algorithm to an algorithm that uses the standard linear model (1.3) but is identical in virtually all other respects. We comment briefly on extensions of our tensor methods in section 7.

Notice that we have denoted members of a sequence of \( n \)-vectors \( x \) by \( \{x_k\} \) where each \( x_k \in \mathbb{R}^n \), and components of a vector \( v \in \mathbb{R}^n \) by \( v[i] \in \mathbb{R} \). The convention of using non-numerical subscripts for replications and bracket notation for scalar components is continued throughout the paper. In section 4, integer subscripts are used to denote portions of vectors or matrices that are themselves vectors or matrices; for example, the portions \( \tilde{d}_1 \in \mathbb{R}^{n-p} \) and \( \tilde{d}_2 \in \mathbb{R}^p \) of the vector \( \tilde{d} \in \mathbb{R}^n \), and the portions \( \tilde{J}_1 \in \mathbb{R}^{n \times (n-p)} \) and \( \tilde{J}_2 \in \mathbb{R}^{n \times p} \) of the matrix \( \tilde{J} \in \mathbb{R}^{n \times n} \), are defined in step 2 of Algorithm 4.1.
2. The tensor method for one equation in one unknown

In this section, we discuss briefly the restriction of our tensor method to the case \( n=1 \). The use of a quadratic model for solving one nonlinear equation in one unknown is well known and a part of some software libraries; an early reference is Muller [1956]. We make no attempt to compare the one variable version of our tensor algorithm to similar algorithms for solving a single nonlinear equation. Rather, the material in this section is included solely to motivate some features of the tensor algorithm for systems of equations that follows.

The quadratic model (1.5) restricted to \( n=1 \) is

\[
m_T(x_c + d) = f(x_c) + f'(x_c)d + \frac{1}{2}f''(x_c)d^2
\]

where all quantities now are scalars. We said in section 1 that we would not use second derivatives. Then an obvious way to select \( t_c \) is to emulate the secant method by asking the model (2.1) to interpolate the value of \( f(x) \) at the previous iterate \( x_- \). This means

\[
f(x_-) = f(x_c) + f'(x_c)s_- + \frac{1}{2}f''(x_c)s_-^2
\]

where we define

\[
s_- = x_- - x_c,
\]

the step from \( x_c \) to \( x_- \). The second derivative approximation \( t_c \) is determined uniquely by (2.2).

The roots of (2.1) are found by solving one quadratic equation in one unknown. Usually, (2.1) will have either no real roots or two real roots. If there are two real roots, then a reasonable way to choose between them is to let \( x_+ \) be the root that is closer to \( x_c \). This is written in a numerically stable way as

\[
x_+ = x_c - \frac{2f(x_c)}{f'(x_c) + \text{sign}(f'(x_c)) \sqrt{(f'(x_c))^2 - 2t_c f(x_c)}}
\]

(2.3)
If (2.1) has no real roots \(((f'(x_c))^2 < 2t_c f(x_c))\), there are two obvious alternatives: try the Newton step
\[
x_N^t = x_c - f(x_c) / f'(x_c),
\]
(2.4)
or try setting \(x_t\) to the critical point of (2.1).
\[
x_{\min}^t = x_c - f'(x_c) / t_c,
\]
(2.5)
where the absolute value of the quadratic model is smallest. The latter strategy definitely is advantageous close to a root \(x_\star\) where \(f'(x_\star) = 0 \neq f''(x_\star)\). In this case, the iterates produced by (2.2, 2.5) converge to \(x_\star\) with q-order \((1 + \sqrt{5})/2 \approx 1.61\), as this is just a variant of the secant method for minimizing or maximizing \(f(x)\). The step (2.2, 2.3) may or may not be defined in this case; if it is defined it is also quite satisfactory as the iterates it produces will converge to \(x_\star\) with q-order \((1 + \sqrt{3})/2 \approx 1.37\). (See e.g. Traub [1984] for very similar proofs.) On the other hand, we have already stated that the iterates produced by (2.2, 2.4) are q-linearly convergent to \(x_\star\) with constant converging to \(\frac{1}{2}\) in this case.

When \(f'(x_\star) \neq 0\), the quadratic model (2.1, 2.2) will have real roots for \(x_c\) sufficiently close to \(x_\star\). In general, however, (2.1) may not have real roots, and aside from the above-mentioned case, neither \(x_N^t\) or \(x_{\min}^t\) consistently is the better choice. In Frank [1982], we implemented a method for solving one nonlinear equation in one unknown based on (2.2 - 2.5) and a line search. We found that when the quadratic model had no real root, a good strategy for choosing between the steps to \(x_N^t\) and \(x_{\min}^t\) as the initial step in the line search was to choose the longer step. This is equivalent to choosing \(x_{\min}^t\) if and only if \(|m_T(x_{\min}^t)| < \frac{1}{2} |f(x_c)|\). This strategy always prefers \(x_{\min}^t\) over \(x_N^t\) sufficiently close to a root \(x_\star\) where \(f'(x_\star) = 0\), guaranteeing fast local convergence in this case.

One case where this strategy is not desirable is when the algorithm cannot locate a root of \(f(x)\) and must converge to a stationary point of \(f(x)\). In this
case the step to $z^\text{min}_i$ is shorter but considerably more desirable than the Newton step. The algorithm in Frank [1962] contains a simple strategy to recognize this situation and uses the step to $z^\text{min}_i$ instead of the Newton step in this case.

Several aspects of the algorithm for one nonlinear equation carry over to systems of equations. There we again use the extra freedom in the quadratic model to interpolate function values at past iterates. The multi-variable quadratic model again may have multiple real roots or no real roots. In the former case we again hope to find the root closest to $z_i$. In the latter case, we again use either the step to the minimizer (in the $l_2$ norm) of the quadratic model or the Newton step for our line search direction, choosing between the two steps by criteria similar to those described above.

3. Forming the tensor model

We now show how we select the tensor term $T_i \in \mathbb{R}^{n \times n \times n}$ in the model

$$M_T(x_i + d) = F(x_i) + F'(x_i)d + \frac{1}{2}T_i dd.$$  \hspace{1cm} (3.1)

Our choice of $T_i$ will cause the second order term $T_i dd$ in (3.1) to have a simple, useful form.

We have already stated that $T_i$ will not contain actual second derivative information. Another way we can use the second order term in (3.1) is to ask the model to interpolate additional values of the function $F(x)$ or the Jacobian $F'(x)$ that have already been computed by the algorithm. In our method, we will select some set of $p$ not necessarily consecutive past iterates $x_{-1}, \ldots, x_{-p}$, and require the model (3.1) to interpolate the function values $F(x_{-k})$ at these points. That is, the model should satisfy

$$F(x_{-k}) = F(x_i) + F'(x_i)s_k + \frac{1}{2}T_i s_k s_k, \quad k = 1, \ldots, p$$ \hspace{1cm} (3.2a)
where
\[ s_k = x_{-k} - x_c, \quad k = 1, \ldots, p. \quad (3.2b) \]

First we describe how the past points \( x_{-1}, \ldots, x_{-p} \) are selected. Then we show how we choose \( T_c \) to satisfy (3.2). Alternative ways to select \( T_c \) are mentioned briefly in section 7.

For the equations (3.2) to always be consistent, it is clear that the set of directions \( \{s_k\} \) from \( x_c \) to the selected past points \( x_{-k} \) must be linearly independent. In fact, our computational experience with other algorithms that interpolate information from past iterates has shown that the directions \( \{s_k\} \) should be strongly linear independent, in the sense that each direction \( s_k \) should make an angle of at least 0 degrees with the linear subspace spanned by the other directions; values of 0 between 20 and 45 degrees have proven appropriate in practice. At each iteration, therefore, we choose the past points \( \{x_{-k}\} \) that we include in (3.2) by the following procedure. We consider the past iterates in order starting with the most recent. We always select the most recent iterate, and then select each preceding past iterate if the step from it to \( x_c \) makes an angle of at least 0 degrees with the subspace spanned by the steps to the already selected more recent iterates. This procedure is implemented easily using a modified Gram-Schmidt algorithm.

We also set an upper bound \( p \) on the number of past function values interpolated by the model at each iteration. Since the set \( \{s_k\} \) must be linearly independent, clearly \( p \leq n \), but we enforce a much smaller bound,
\[ p \leq \sqrt{n}. \quad (3.3) \]

This bound also was motivated by our computational experience with other algorithms that interpolate information from past iterates; we found that using more than about \( \sqrt{n} \) interpolation conditions rarely was beneficial (see e.g. Stordahl [1980]). The bound (3.3) also is crucial to the efficiency in storage and
arithmetic operations of our tensor method. For example, since we also only consider a maximum of $\sqrt{n}$ past iterates in the above mentioned Gram-Schmidt algorithm, it requires about $n^2$ additions and multiplications.

Now we discuss how we choose $T_c$ to satisfy (3.2). It is convenient to rewrite (3.2) as

$$T_c s_k s_k = z_k, \quad k = 1, \ldots, p$$  \hspace{1cm} (3.4a)

where

$$z_k \in \mathbb{R}^n, \quad z_k = 2 \left( F(x_{-k}) - F(x_c) - F'(x_c)s_k \right).$$  \hspace{1cm} (3.4b)

This is a set of $np \leq n^{1.5}$ linear equations in the $n^3$ unknowns $T_c[i,j,k]$, $1 \leq i,j,k \leq n$. (Actually there are $(n^3+n^2)/2$ unknowns since each horizontal face of $T_c$ must be symmetric; this symmetry is provided automatically by the following derivation.) Since (3.4) is underdetermined, we follow the standard and successful practice in secant methods for nonlinear equations and optimization (see e.g. Dennis and Schnabel [1979]) and choose the $T_c$ that satisfies

$$\min_{T_c \in \mathbb{R}^{n \times n \times n}} \|T_c\|_F$$

subject to $T_c s_k s_k = z_k, \quad k = 1, \ldots, p$

where $\|T_c\|_F$, the Frobenius norm of $T_c$, is defined by

$$\|T_c\|_F^2 = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n (T_c[i,j,k])^2.$$

The solution to (3.5) is given by Theorem 3.2. First we define a rank one tensor.

**Definition 3.1.** Let $u, v, w \in \mathbb{R}^n$. The tensor $T \in \mathbb{R}^{n \times n \times n}$ for which $T[i,j,k] = u[i] \cdot v[j] \cdot w[k], \ 1 \leq i,j,k \leq n$ is called a rank one tensor and denoted $T = u v w$.

Note that the $i^{th}$ horizontal face of the rank one tensor $u v w$ is the rank one matrix $u[i](v w^T)$. Theorem 3.2 shows that the solution to (3.5) is the sum of $p$ rank one tensors whose horizontal faces are symmetric.
**Theorem 3.2.** Let \( p \leq n \), let \( s_k \in \mathbb{R}^n \), \( k = 1, \ldots, p \) with \( \{s_k\} \) linearly independent, and let \( z_k \in \mathbb{R}^n \), \( k = 1, \ldots, p \). Define \( M \in \mathbb{R}^{p \times p} \) by \( M[i,j] = (s_i^T s_j)^2 \), \( 1 \leq i, j \leq p \), \( Z \in \mathbb{R}^{n \times p} \) by column \( k \) of \( Z = z_k \), \( k = 1, \ldots, p \). Then \( M \) is positive definite, and the solution to (3.5) is

\[
T_c = \sum_{k=1}^{p} \alpha_k s_k^T s_k
\]

where \( \alpha_k \) is the \( k^{th} \) column of \( A \in \mathbb{R}^{n \times p} \), \( A \) defined by \( A = Z \ast M^{-1} \).

**Proof:** Problem (3.5) is equivalent to solving independently the \( n \) separate problems over the \( n \) horizontal faces \( H_i \) of \( T_c \)

\[
\text{minimize} \quad \|H_i\|_F \quad \text{subject to} \quad s_k^T H_i s_k = z_k[i], \quad k = 1, \ldots, p.
\]

because the constraints and the objective function of (3.5) may be decomposed into these \( n \) separate constraints and objective functions each involving one of \( H_i \)'s, and the optimal value of any one \( H_i \) clearly does not affect the optimal value of any other \( H_j \). Problem (3.7) simply is an underdetermined system of \( p \) linear equations in \( n^2 \) unknowns. To express it in standard form, let

\[
h_i \in \mathbb{R}^{n^2},
\]

\[
h_i = H_i[1,1], H_i[1,2], \ldots, H_i[1,n], H_i[2,1], \ldots, H_i[2,n], \ldots, H_i[n,1], \ldots, H_i[n,n],
\]

\[
\bar{S} \in \mathbb{R}^{p \times n^2}, \quad \text{row} \ k \ \text{of} \ \bar{S} = s_k[1] \ast s_k, \ s_k[2] \ast s_k, \ldots, s_k[n] \ast s_k,
\]

and \( \bar{z}_i = \text{row} \ i \) of \( Z \), that is,

\[
\bar{z}_i \in \mathbb{R}^p, \quad \bar{z}_i[k] = z_k[i], \quad 1 \leq i \leq n, \quad 1 \leq k \leq p.
\]

Then (3.7) is equivalent to

\[
\text{minimize} \quad \|h_i\|_2 \quad \text{subject to} \quad \bar{S} h_i = \bar{z}_i
\]

and \( \bar{S} \) has full row rank because \( \{s_k\} \) are linearly independent. Therefore the solution to (3.9) is

\[
h_i = \bar{S}^T (\bar{S} \bar{S}^T)^{-1} \bar{z}_i.
\]

It is straightforward to verify that \( \bar{S} \bar{S}^T = M \), which also shows that \( M \) is positive
definite. Since by definition row \( i \) of \( A = M^{-1} \ast (\text{row } i \text{ of } Z) \),

\[
h_i = \bar{S}^T \bar{a}_i,
\]

where \( \bar{a}_i = \text{row } i \text{ of } A \), that is,

\[
\bar{a}_i \in \mathbb{R}^p, \quad \bar{a}_i[k] = a_k[i], \quad 1 \leq i \leq n, \quad 1 \leq k \leq p.
\]

Now note that if the transformation (3.8) that transforms \( H_i \in \mathbb{R}^{n \times n} \) to \( \bar{h}_i \in \mathbb{R}^{n^2} \) is applied to \( s_k s_k^T \in \mathbb{R}^{n \times n} \), the result is row \( k \) of \( \bar{S} \). Therefore, since (3.10) is equivalent to

\[
h_i = \sum_{k=1}^{p} \bar{a}_i[k] \ast \text{row } k \text{ of } \bar{S} = \sum_{k=1}^{p} a_k[i] \ast \text{row } k \text{ of } \bar{S},
\]

transforming (3.11) back to the \( n \times n \) matrix \( H_i \) yields

\[
H_i = \sum_{k=1}^{p} a_k[i] s_k s_k^T.
\]

Finally, combining the \( n \) horizontal faces \( H_i \) given by (3.12) to reform \( T_c \) gives (3.6).

Substituting (3.6) into the tensor model (3.1) gives

\[
M_T(x_c + d) = F(x_c) + \bar{F}(x_c) d + \frac{1}{2} \sum_{k=1}^{p} a_k (d^T s_k)^2.
\]

The simple form of the second order term in (3.13) is the key to being able to efficiently form, store, and solve the tensor model. The additional storage required by (3.13) is \( 2p \) n-vectors, for \( \{a_k\} \) and \( \{s_k\} \). In addition, the \( 2p \) n-vectors \( \{x_{-k}\} \) and \( \{\bar{F}(x_{-k})\} \) must be stored. Thus the total extra storage required for our tensor model is \( 4n^{1.5} \) since \( p \leq \sqrt{n} \). The reader can verify that the entire process described above for forming \( T_c \) requires \( n^2p + O(np^2) \) multiplications and additions. The leading term comes from calculating the \( p \) matrix-vector products \( F'(x_c)s_k, k = 1, \ldots, p \); the cost of solving \( A = Z^\ast M^{-1} \) is \( O(np^2) \). Since \( p \leq \sqrt{n} \), the leading term in the cost of forming the tensor model is at most \( n^{2.5} \) multiplications and additions per iteration, small compared to the at least \( n^3/3 \) multiplications and additions per iteration required by stan-
standard methods that use analytic derivatives. In the next section, we will see that the extra cost to solve the tensor model also is at most $O(n^{2.3})$.

4. Solving the tensor model

In this section we give an efficient algorithm for finding a root of the tensor model derived in section 3, that is,

$$\text{find } d \in \mathbb{R}^n \text{ such that } (4.1)$$

$$M_T(x_c + d) = F(x_c) + F'(x_c)d + \frac{1}{2} \sum_{k=1}^{q} a_k (d^T s_k)^2 = 0.$$ 

We show that the solution of (4.1) can be reduced, in $O(n^2 p)$ operations, to the solution of a system of $q$ quadratic equations in $p$ unknowns, plus the solution of a system of $n - p$ linear equations in $n - p$ unknowns. Here $q \geq p$, with $q = p$ whenever $F'(x_c)$ is nonsingular. In addition, if $F'(x_c)$ is singular but has rank at least $n - p$, we show that $q$ still usually equals $p$ and the aforementioned system of linear equations still is well conditioned. We also show that our algorithm efficiently solves the generalization of (4.1).

$$\text{minimize } \|M_T(x_c + d)\|_2.$$ 

That is, our algorithm will find a minimizer of the tensor model when the model has no real root.

Let us define $S \in \mathbb{R}^{n \times p}$ by column $k$ of $S = s_k$.

The basic idea of the algorithm is that since (4.1) is linear on the $n - p$ dimensional subspace $\{d \mid S^T d = 0\}$, (4.1) really only should be quadratic in $p$ variables and linear in the other $n - p$. This is accomplished in steps 1 and 2 of Algorithm 4.1 by making a linear transformation of the variable space; an orthogonal transformation is used mainly because it already will be available
from the Gram-Schmidt process used to select \( \{ s_k \} \). Then a linear transformation of the equations, steps 3 and 4 of Algorithm 4.1, is used to eliminate the \( n - p \) transformed linear variables from \( p \) of the equations. The result usually is a system of \( p \) quadratic equations in \( p \) unknowns, (4.5b), that is solved in step 5 of Algorithm 4.1, and a system of \( n - p \) equations (4.5a) that are linear in the remaining \( n - p \) unknowns and can be used to compute these unknowns once the system of quadratics is solved. The exceptional case \( q > p \) arises when this system of linear equations would be singular. In this case, the number of linear equations is decreased by this rank deficiency and the number of quadratic equations is increased correspondingly, but the number of variables in each system is unaffected. Of course in practice, the mathematical notion of rank deficiency is replaced by the numerical notion of adequately small condition number.

Algorithm 4.1 gives the method we use for solving (4.2). Theorem 4.2 verifies that it solves this problem, and gives several other important properties of the algorithm. After the proof of Theorem 4.2 we discuss the efficiency and numerical stability of Algorithm 4.1. We also mention several alternative methods for solving problem (4.2).

We introduce the notation that given \( v \in \mathbb{R}^m \), \( \{ v \}^2 \) denotes the vector \( w \in \mathbb{R}^m \) for which \( w[i] = v[i]^2 \), \( i = 1, \cdots, m \). This allows us to denote the second order term of our tensor model below by \( \frac{1}{2} A \{ S^T d \}^2 \).

**Algorithm 4.1.** Let \( p \leq n \), \( F \in \mathbb{R}^n \), \( J \in \mathbb{R}^{n \times n} \), \( A, S \in \mathbb{R}^{n \times p} \), \( S \) having full column rank.

Comment: Steps 1-2 transform the system of \( n \) equations in \( n \) unknowns

\[
F + Jd + \frac{1}{2} A \{ S^T d \}^2 = 0 \quad (4.3)
\]

to the system of \( n \) equations in the \( n \) unknowns \( \hat{d}_t \in \mathbb{R}^{n-p} \) and \( \hat{d}_q \in \mathbb{R}^p \).
\[ F + \hat{J}_1 \tilde{d}_1 + \hat{J}_2 \tilde{d}_2 + \frac{1}{2} A \{ \hat{S}_2 \tilde{d}_2 \}^2 = 0. \] (4.4)

1. Find an orthogonal \( Q \in R^{n \times n} \) such that \( Q^T S = \hat{S} \), where

\[
\hat{S} \in R^{n \times p} = \begin{bmatrix}
0 & n-p \\
p & 0 \\
\end{bmatrix}
\]

and \( \hat{S}_2 \) has the triangular shape shown.

2. Calculate \( \tilde{J} = J^* Q \) and let \( \tilde{J}_1 \in R^{n \times (n-p)} \) and \( \tilde{J}_2 \in R^{n \times p} \) denote the first \( n-p \) and last \( p \) columns of \( \tilde{J} \), respectively.

Also define \( \tilde{d} = Q^T d \), and let \( \tilde{d}_1 \in R^{n-p} \) and \( \tilde{d}_2 \in R^p \) denote the first \( n-p \) and last \( p \) components of \( \tilde{d} \), respectively.

Comment: Steps 3-4 transform the system of equations (4.4) to

\[
\begin{bmatrix}
n-q \\
q \\
\end{bmatrix}
\begin{bmatrix}
\tilde{F}_1 \\
\tilde{F}_2 \\
\end{bmatrix}
+ \begin{bmatrix}
\tilde{J}_1 \\
\tilde{J}_2 \\
\end{bmatrix}
\begin{bmatrix}
\tilde{d}_1 \\
\tilde{d}_2 \\
\end{bmatrix}
+ \frac{1}{2} \begin{bmatrix}
\tilde{A}_1 \\
\tilde{A}_2 \\
\end{bmatrix} \{ \tilde{S}_2 \tilde{d}_2 \}^T = 0,
\]

that is, to the system of \( n-q \) equations in \( n \) unknowns

\[
\tilde{F}_1 + \tilde{J}_1 \tilde{d}_1 + \tilde{J}_2 \tilde{d}_2 + \frac{1}{2} \tilde{A}_1 \{ \tilde{S}_2 \tilde{d}_2 \}^T = 0 \] (4.5a)

and the system of \( q \) equations in \( p \) unknowns

\[
\tilde{F}_2 + \tilde{J}_3 \tilde{d}_2 + \frac{1}{2} \tilde{A}_2 \{ \tilde{S}_2 \tilde{d}_2 \}^T = 0. \] (4.5b)

3. Find an orthogonal \( \hat{Q} \in R^{n \times n} \) and a permutation matrix \( P \in R^{(n-p) \times (n-p)} \) such that

\[
\hat{Q} \tilde{J}_1 P = \tilde{J}_1 \begin{bmatrix}
n-p \\
q \\
\end{bmatrix}
\begin{bmatrix}
0 & n-q \\
0 & q \\
\end{bmatrix}
\]

where \( q \geq p \) and \( \tilde{J}_1 \) is upper triangular with a non-zero diagonal. Define
\[ \tilde{d}_1 = P^T \tilde{d}_1, \quad \tilde{d}_1 \in \mathbb{R}^{n-p}. \]

4. Calculate

\[
\begin{bmatrix}
\tilde{Q} \\
\tilde{J}_2 \\
\tilde{J}_3
\end{bmatrix}
= \begin{bmatrix}
P \\
J_2 \\
J_3
\end{bmatrix}
\begin{bmatrix}
n-q \\
q
\end{bmatrix}
\]

Similarly calculate \( \tilde{A} = \tilde{Q} A \), and let \( \tilde{A}_1 \in \mathbb{R}^{p \times n-q} \) and \( \tilde{A}_2 \in \mathbb{R}^{p \times q} \) denote the first \( n-q \) and the last \( q \) rows of \( \tilde{A} \), respectively; also calculate \( \tilde{F} = \tilde{Q} F \), and let \( \tilde{F}_1 \in \mathbb{R}^{n-q} \) and \( \tilde{F}_2 \in \mathbb{R}^{q} \) denote the first \( n-q \) and last \( q \) components of \( \tilde{F} \), respectively.

5. (Solve (4.5b) in the least squares sense.) Solve

\[
\begin{aligned}
\text{minimize} & \quad \| \tilde{F}_2 + \tilde{J}_3 \tilde{d}_2 + \frac{1}{2} \tilde{A}_2 \{ \tilde{S}_2 \tilde{d}_2 \}^T \|_2. \\
\text{subject to} & \quad \tilde{J}_1 \tilde{d}_1 = -\tilde{F}_1 - \tilde{A}_1 \{ \tilde{S}_2 \tilde{d}_2 \}^T.
\end{aligned}
\]

6. (Backsolve (4.5a) for \( \tilde{d}_1 \).) Find a \( \tilde{d}_1 \) that solves

\[
\tilde{J}_1 \tilde{d}_1 = -\tilde{F}_1 - \tilde{A}_1 \{ \tilde{S}_2 \tilde{d}_2 \}^T.
\]

7. Calculate \( \tilde{d}_1 = P \tilde{d}_1, \quad d = Q \tilde{d} \).

Theorem 4.2 shows that Algorithm 4.1 finds a root or minimizer of the tensor model, and gives some properties of some matrices used in Algorithm 4.1 whose relation to the numerical stability of the algorithm is discussed later in this section. Recall that for any \( W \in \mathbb{R}^{n \times m} \), the rank of \( W \) is the dimension of the linear subspace spanned by the rows or columns of \( W \), and the nullity of \( W \), the dimension of the linear subspace \( \{ y \in \mathbb{R}^m \mid Wy = 0 \} \), is \( m - \text{rank}(W) \).
Theorem 4.2. Algorithm 4.1 calculates a solution to

$$\min_{d \in \mathbb{R}^n} ||F + Jd + A\{S^T d\}^2||_2.$$  \hspace{1cm} (4.8)

Furthermore, define $J_S \in \mathbb{R}^{(n+p) \times n} = \begin{bmatrix} J \\ S^T \end{bmatrix}$. Then

$$q = p + (n - \text{rank}(J_S)).$$  \hspace{1cm} (4.9)

If rank($J_S$) = $n$, then $q = p$ and

$$K(J_1) \preceq K(J_S),$$  \hspace{1cm} (4.10)

where $K(W)$ denotes the $l_2$ condition number of $W$. Also

$$\text{rank}(\tilde{J}_3) = p - (\text{rank}(J_S) - \text{rank}(J)).$$  \hspace{1cm} (4.11)

Proof: Substituting $QQ^T d$ for $d$ in (4.3) and using the definitions in steps 1 and 2 transforms (4.3) to (4.4), and clearly the transformation does not affect the smallest $l_2$ norm value of the system of equations. Next, it is straightforward to verify that premultiplying (4.4) by $\tilde{Q}$, substituting $P \tilde{P}^T d_1$ for $\tilde{d}_1$ in (4.4), and using the definitions in steps 3 and 4 yields (4.5). The minimum $l_2$ norm values of these two systems of equations are equivalent because premultiplying a vector by an orthogonal matrix doesn't alter its $l_2$ norm. Finally, since $\tilde{J}_1$ has full row rank, given any $\tilde{d}_2$, a $\tilde{d}_1$ may be found that solves (4.5a). Therefore, the minimum value of (4.6) is the minimum $l_2$ norm value of (4.5), and by the above, of the original problem (4.8), and the $l_2$ minimizer of (4.5) is found by (4.6-4.7).

Step 7 reverses the transformations in variables made in steps 2 and 3 to obtain the minimizer of (4.8) from the minimizer of (4.5).

Now let $Q_1 \in \mathbb{R}^{n \times (n-p)}$ and $Q_2 \in \mathbb{R}^{n \times p}$ denote the first $n-p$ and last $p$ columns of the $Q$ used in step 1 of Algorithm 4.1. Then since $S^T Q = \tilde{S}^T$, we have $S^T Q_1 = 0$ and $S^T Q_2 = \tilde{S}_2^T$. Similarly from step 2 of Algorithm 4.1, $JQ_1 = \tilde{J}_1$, $JQ_2 = \tilde{J}_2$. Thus
and since $Q$ and $S_2^T$ are nonsingular,

$$\text{nullity}(J_S) = \text{nullity}(J_S Q) = \text{nullity}(\tilde{J}_1).$$  \hspace{1cm} (4.12)

From step 3, $\text{nullity}(\tilde{J}_1) = q - p$, and by definition, $\text{nullity}(J_S) = n - \text{rank}(J_S)$, so (4.12) implies (4.9). Next, if we use the notation that for $v \in \mathbb{R}^n$, $v_1$ denotes the first $n - p$ components of $v$ and $v_2$ the last $p$ components, and let $\| \cdot \|$ denote the $l_2$ vector norm, then

$$K(J_S) = K(J_S Q) = \max_{v \in \mathbb{R}^n} \frac{\|J_S Qv\| / \|v\|}{\min_{w \in \mathbb{R}^n} \|J_S Qw\| / \|w\|} \geq \frac{\max_{v \in \mathbb{R}^n, v_2 = 0} \|J_S Qv\| / \|v\|}{\min_{w \in \mathbb{R}^n, w_2 = 0} \|J_S Qw\| / \|w\|} = \frac{\max_{v_1 \in \mathbb{R}^{n-p}} \|\tilde{J}_1 v_1\| / \|v_1\|}{\min_{w_1 \in \mathbb{R}^{n-p}} \|\tilde{J}_1 w_1\| / \|w_1\|} = K(\tilde{J}_1) = K(\tilde{J}_1)$$

with the last equality a direct consequence of step 3 of Algorithm 4.1. Also, since

$$\hat{Q}JQ\hat{P} = \begin{bmatrix} \tilde{J}_1 & \tilde{J}_2 \\ 0 & \tilde{J}_3 \end{bmatrix}$$

where $\hat{P} \in \mathbb{R}^{n \times n}$ is the permutation matrix with $P$ as the upper $n - p \times n - p$ submatrix and the identity otherwise, and since $\tilde{J}_1$ has full row rank, we have

$$\text{nullity}(J) = \text{nullity}(\hat{Q}JQ\hat{P}) = \text{nullity}(\tilde{J}_1) + \text{nullity}(\tilde{J}_3).$$ \hspace{1cm} (4.13)

Substituting again $\text{nullity}(\tilde{J}_1) = q - p$ as well as the definitions $\text{nullity}(J) = n - \text{rank}(J)$ and $\text{nullity}(\tilde{J}_3) = p - \text{rank}(\tilde{J}_3)$ into (4.13) yields (4.11).

The first virtue of Algorithm 4.1 is its efficiency. Let us examine the operation counts for multiplications (and divisions); the counts for additions and subtractions are very similar. While Algorithm 4.1 is valid for any $p \leq n$, here we
reinstate the bound $p \leq \sqrt{n}$ from section 3. Then the dominant cost in Algorithm 4.1 is the QR factorization of $\hat{J}_1$ which requires about $2n^3/3 - n^2p + O(n^2)$ multiplications. The next largest cost is the $2n^3p + O(n^2)$ multiplications for the matrix multiplication $J^*Q$ in step 2. The reader can verify that all other portions of steps 1-4 and 6-7 require at most $O(n^2)$ multiplications. The remaining cost is the solution of the nonlinear least squares problem in step 5. While (4.6) must be solved by an iterative algorithm, the point is that the total cost of solving this problem is negligible. In the usual case $q = p$, each iteration of the nonlinear least squares algorithm requires $O(p^3)$ multiplications, and it is reasonable to expect at most a small multiple of $p$ iterations to suffice. (We use the bound $8p$ in our implementation.) Thus solving (4.6) usually can be expected to cost $O(p^4) \leq O(n^2)$ multiplications. If $q > p$, this cost could rise to $O(qp^3) \leq O(n^{2.5})$. However in our practical experience, $q = p$ almost all the time and $q$ is hardly greater than $p$ otherwise; some summary statistics are given at the end of Section 6. Thus it is reasonable to expect that the nonlinear least squares problem in Algorithm 4.1 requires at most $O(n^2)$ multiplications.

So the total cost of Algorithm 4.1 is about $2n^3/3 + n^2p$ multiplications, at most $n^{2.5}$ multiplications more than the QR factorization of an $n \times n$ matrix. For small $n$, these numbers are inconsequential; for moderately large $n$, the $2n^3/3$ dominates and is the same cost as a standard method for nonlinear equations would have if it used the QR factorization. While some standard algorithms for nonlinear equations do use a QR factorization (see e.g. section 6.5 of Dennis and Schnabel [1983]), others use a PLU factorization and require only $n^3/3$ multiplications per iteration if $F'(z_c)$ is nonsingular. Algorithm 4.1 also could be modified to use $n^3/3 + O(n^2p)$ multiplications per iteration when $J$ is nonsingular, by using a PLU factorization of $\hat{J}_1$ at step 3. If the resultant $p \times p$ system of quadratics (4.5b) had a solution, no further modification would be necessary. If
not, the minimum norm solution to this new system of quadratics no longer would correspond to the solution of (4.8), and steps 5-6 would have to be combined into the nonlinear least squares solution of the full system (4.5). This still could be accomplished in $O(n^2 p)$ operations when $J$ is nonsingular by using an algorithm that takes advantage of the special structure of this problem. We prefer the more expensive QR-based algorithm, however, because it is simpler and more stable numerically, especially when the tensor model has no real root.

The other virtue of Algorithm 4.1 is its numerical stability, even when the Jacobian $J$ is singular or ill-conditioned. This is reflected in the conditioning of the system of linear equations, (4.7), that is solved by Algorithm 4.1. Equations (4.9) and (4.10) of Theorem 4.2 show that if the $(n+p) \times n$ matrix $J_S$ formed by adding the $p$ rows of $S^T$ beneath $J$ has safely linearly independent columns, then this linear system will be square and at least as well conditioned as $J_S$. In practice, this means it is likely that the system of quadratics will be $p \times p$ and the linear system will be square and well conditioned whenever $J$ has at most $p$ zero or small singular values. Notice that (4.10) also implies that the conditioning of the linear system solved by the tensor algorithm always will be as good or better than the conditioning of the linear system solved by Newton's method.

Of course, if $J$ is singular, this singularity does not magically disappear in Algorithm 4.1. Equation (4.11) shows that if $J$ is rank deficient but $J_S$ has full rank (the likely situation when $\text{rank}(J) \geq n-p$), then $\tilde{J}_g$ will have the same rank deficiency as $J$. However the whole point of the tensor algorithm when $J$ is singular is that the singular submatrix $\tilde{J}_g$ is used in the system of quadratic equations (4.5b), which also contain a portion of the second order information in the tensor model. This system is not necessarily ill-conditioned even if $\tilde{J}_g$ is; for example, one quadratic equation in one unknown with no linear term is not ill-conditioned. Furthermore, this system of quadratics has maximal dimension
\( \sqrt{n} \), and in our practical experience even this small upper bound is not approached as \( n \) becomes large. Thus when the Jacobian is singular, the effect of the tensor algorithm usually is to isolate this rank deficiency as the linear term in a small system of quadratic equations that contains useful second order information and may be solved accurately without great cost.

If the tensor algorithm converges to a root \( x_* \) where \( F'(x_*) \) is singular, the Jacobians of the nonlinear least squares problems (4.6) also will become nearly singular because \( \tilde{J}_0 \) will be nearly singular and the solution \( \tilde{d}_2 \) will be small. However the cost of solving these small nearly singular nonlinear least squares problems still will be insignificant. So on nonlinear equations problems where \( F'(x_*) \) is singular, it appears the effect of the tensor algorithm is to transfer the slow convergence of Newton's method to a series of small quadratic subproblems where slow convergence does not hurt the overall efficiency of the algorithm. More importantly, no additional function evaluations are used in solving these small subproblems.

A minor deficiency of Algorithm 4.1 is that it relies upon the QRP decomposition to determine rank when \( \tilde{J}_1 \) does not have full rank \( (q > p) \); a singular value decomposition would be preferable in this case. We have already mentioned that this case is very unusual in practice. Note that when \( q > p \), Algorithm 4.1 easily is modified to find the smallest \( d \), in the \( l_2 \) norm, that minimizes the \( l_2 \) norm of the tensor model; all that is required is to find the smallest \( \tilde{d}_1 \) that solves the system of linear equations (4.7) that is underdetermined in this case.

It is interesting to consider when our tensor model has a complex root. Even though this information is not of practical value to our tensor algorithm, it turns out to give a succinct explanation of how the Jacobian \( J \), past directions \( S \), and second order information \( A \) used in Algorithm 4.1 are interrelated. Clearly a necessary condition for (4.3) to have any root is that \( F' \) be in the linear
subspace spanned by the columns of $J$ and $A$; it is easy to show by the techniques of proof of Theorem 4.2 that this is equivalent to $\tilde{F}_2$ being in the linear subspace spanned by the columns of $\tilde{J}_3$ and $\tilde{A}_3$. A sufficient condition for (4.3) to have a complex root, a direct consequence of a result of Garcia and Li [1980], is given in Theorem 4.3.

**Theorem 4.3.** Let all the notation be the same as in Algorithm 4.1, and let $J_{AS} \in \mathbb{R}^{(n+p) \times (n+p)}$,

$$J_{AS} = \begin{bmatrix} J & A \\ S_p & 0 \end{bmatrix}$$

If $J_{AS}$ is nonsingular, (4.3) has $2^p$ (not necessarily distinct) roots in $n$ dimensional complex space.

**Proof:** From the proof of Theorem 4.2, it is clear that (4.3) has a complex root if and only if (4.5b) has a complex root. It follows easily from Theorem 3.4 of Garcia and Li [1980] that (4.5b) has a complex root if $\tilde{A}_2$ and $\tilde{S}_2^T$ are nonsingular. We show that this is implied by $J_{AS}$ nonsingular.

Let $Q_+ \in \mathbb{R}^{(n+p) \times (n+p)}$ be the orthogonal matrix with $Q$ as the upper left $n \times n$ submatrix and the identity elsewhere, and let $\hat{Q}_+ \in \mathbb{R}^{(n+p) \times (n+p)}$ be the corresponding augmentation of $\hat{Q}$. Then it is straightforward to verify from Algorithm 4.1 that

$$\hat{Q}_+ J_{AS} Q_+ = \begin{bmatrix} \tilde{J}_1 & \tilde{J}_2 & \tilde{A}_1 \\ 0 & \tilde{J}_3 & \tilde{A}_2 \\ 0 & \tilde{S}_2^T & 0 \end{bmatrix} = \hat{J}_{AS}.$$

If $J_{AS}$ is nonsingular then $\hat{J}_{AS}$ is nonsingular which implies that $\tilde{J}_1$ is square (i.e. $q = p$) and nonsingular and $\tilde{S}_2^T$ is nonsingular. Since $\hat{J}_{AS}$ and $\tilde{J}_1$ are nonsingular,
the bottom right $2p \times 2p$ submatrix of $\tilde{J}_{AS}$ composed of $\tilde{J}_3$, $\tilde{A}_2$, and $\tilde{S}_2^T$ must be nonsingular, which implies that $\tilde{A}_2$ is nonsingular because $\tilde{S}_2^T$ is nonsingular.

The only case in which we can guarantee that our tensor model has a real root is when the tensor algorithm is sufficiently close to a root $z_*$ where $F'(z_*)$ is nonsingular. Here one can show that the second order term in the tensor model becomes small enough that the tensor model must have a real root.

Finally, we mention other approaches for minimizing the $l_2$ norm of the tensor model (4.1). In Frank [1982], we discuss a method that first tries to find a root of the tensor model and if it is unsuccessful finds a minimizer. It is related to the version of Algorithm 4.1 using the PLU factorization that we discussed above. Another possibility is to use Newton's method, with a line search on the $l_2$ norm of the tensor model, to solve (4.2). If the Jacobian $J$ is nonsingular, then by using the Sherman-Morrison-Woodbury formula each iteration would cost only $O(p^3)$ operations after a start-up cost of the factorization of $J$ plus $O(n^2p)$. We prefer Algorithm 4.1 because of its greater robustness and numerical stability.

5. Implementation of the tensor method

In the previous two sections we presented the main new features of our nonlinear equations method, namely, how we form our quadratic ("tensor") model of the nonlinear function, and how we calculate a root or minimizer of this model efficiently and stably. Algorithm 5.1 outlines the complete algorithm we have implemented to test these ideas. The remainder of this section clarifies various aspects of this algorithm and its computer implementation. Our test results are presented in section 6.
Algorithm 5.1. An Iteration of the Tensor Method

given $x_\circ, F(x_\circ)$

1. Calculate $F'(x_\circ)$ and decide whether to stop. If not:
2. Select the past points to use in the tensor model from among the $\sqrt{n}$ most recent past points.
3. Calculate the second order term of the tensor model, $T_\circ$, so that the tensor model interpolates $F(x)$ at all the points selected in step 2.
4. Find the root of the tensor model, or its minimizer (in the $l_2$ norm) if it has no real root.
5. Select $x_+ = x_\circ - \lambda_\circ d_\circ$, where $d_\circ$ either is the step calculated in step 4 or the Newton step, using a line search to choose $\lambda_\circ$.
6. Set $x_\circ \leftarrow x_+$, $F(x_\circ) \leftarrow F(x_+)$, go to step 1.

Several standard sections of our implementation of Algorithm 5.1 use algorithms from Appendix 1 of Dennis and Schnabel [1983]. The Jacobian is approximated by their finite difference algorithm A5.4.1. The stopping criteria are their algorithm A7.2.3; the algorithm stops if the relative size of $(x_+ - x_\circ)$ is less than $10^{-6}$ or the relative size of $(J^T F)$ is less than $10^{-5}$.

Step 2 was described completely in section 3; we set the angle $\Theta$ mentioned there to 45 degrees. Step 3 calculates the smallest $T_\circ$, in the Frobenius norm, for which the tensor model satisfies the interpolation conditions, using the procedure of Theorem 3.2. Before performing these calculations, the steps to the past points $\{s_k\}$ all are normalized to have $l_2$ norm one. This assures that the linear systems $A = Z^*M^{-1}$ solved in step 3 are well conditioned. The normalization does not alter the tensor algorithm in exact arithmetic; each $a_k$ simply is multiplied by $\|s_k\|_2^2$ and $T_\circ$ is unchanged.
Step 4 of Algorithm 5.1, the solution of the tensor model, is accomplished by Algorithm 4.1. The QR decomposition of $S$ in step 1 of Algorithm 4.1 is available from the modified Gram-Schmidt algorithm used to select the past points. The QRP decomposition of $\tilde{J}_1$ is the standard QR with column pivoting (see e.g. Dongarra et. al. [1979]); a column is considered zero if its $l_1$ norm is less than $10 \sqrt{\text{machineps}} \|F'(x_0)\|_1$. The subproblem (4.6), finding the minimum $l_2$ norm value of a system of $q$ quadratic equations in $p$ unknowns, is solved using a standard algorithm and analytic derivatives since they are easily available. The algorithm terminates when a root or minimizer of the system of equations is found, or after $8p$ iterations. When $p=q=1$, the problem is solved analytically. The one difficulty with solving this system of quadratics is that it is likely to have multiple roots or minimizers. Among these, it seems sensible to try to find the root closest to the Newton step. We attempt this by making the starting guess for the small system of quadratics the component of the Newton step in this subspace,

$$\hat{d}_2 = -\tilde{J}_3^{-1} \tilde{F}_2,$$

or the linear least squares solution to this system if $q>p$. When $\tilde{J}_3$ isn't well conditioned, a different procedure is used.

Step 5 of Algorithm 5.1 usually consists of a line search in the direction to the root or minimizer of the tensor model obtained in step 4. We use the line search algorithm A6.3.1 from Dennis and Schnabel, requiring sufficient decrease on $\|F(x)\|_2$. In some cases, a line search in the Newton direction is used instead. (The Newton direction is obtained in $O(n^2)$ operations as a byproduct of the algorithm for solving the tensor model.) These cases are when Algorithm 4.1 finds a root of the tensor model that isn’t a descent direction for $\|F(x)\|_2$, a very rare occurrence in practice but not precluded in theory; when Algorithm 4.1 fails to find a root or minimizer of the tensor model; and sometimes when Algorithm 4.1
finds a minimizer of the tensor model that is not a root. The complete strategy for determining the search direction is given in Algorithm 5.2 below. The criterion for choosing between the Newton direction and a minimizer of the tensor model is an extension of our strategy for one dimensional problems discussed in section 2.

**Algorithm 5.2.** Step Selection, Step 5 of Algorithm 5.1

Let \( J_c \) = approximation to \( F'(x_c) \),

\[
d_T = \text{root or minimizer of the tensor model},
\]

\[
d_N = \text{Newton step } -J_c^{-1}F(x_c) \text{ if } J_c \text{ is sufficiently well conditioned},
\]

Levenberg-Marquardt step \(-J_c^TJ_c + \epsilon I)^{-1}J_c^TF(x_c)\) otherwise, where \( \epsilon = \sqrt{n} \times \text{machines} \cdot \|J_c^TJ_c\|_1 \) (see Dennis and Schnabel [1983])

IF (no root or minimizer of the tensor model was found) OR ((minimizer of tensor model that is not a root was found) AND \( \|M_T(x_c + d_T)\|_2 > \frac{\varepsilon}{\|F(x_c)\|_2} \)) THEN

\[ x_+ \leftarrow x_c + \lambda_c d_N, \quad \lambda_c \in (0,1] \text{ selected by line search} \]

ELSE

\[ x_+ \leftarrow x_c + d_T \]

IF \( x_+ \) is not acceptable THEN

IF \( d_T \) is a sufficient descent direction

THEN \( x_+ \leftarrow x_c + \lambda_c d_T, \quad \lambda_c \in (0,1] \text{ selected by line search} \)

ELSE \( x_+ \leftarrow x_c + \lambda_c d_N, \quad \lambda_c \in (0,1] \text{ selected by line search} \)
6. Test Results

We tested our tensor algorithm on a variety of nonsingular and singular problems. We compared it to an algorithm that is identical except that the second order term $T_e$ always is zero. That is, the comparison algorithm is finite difference Newton's method with a line search, except that the Newton step $-J_e^{-1}F(x_e)$ is modified to the approximation to the pseudo-inverse step $-(J_e^TJ_e + \epsilon I)^{-1}J_e^TF(x_e)$ given in Algorithm 5.2 when $J_e = F'(x_e)$ is singular or sufficiently ill-conditioned. In this section we summarize our test results. Problem by problem data is provided in the appendix.

All our computation was performed on the DEC VAX 11/780 of the University of Colorado Department of Computer Science, using double precision arithmetic.

First we tested our algorithms on the set of nonlinear equations problems in Moré, Garbow, and Hillstrom [1981]. All these problems have $F'(x_*)$ nonsingular, with the exception of Powell’s Singular Function where $n = 4$ and $\text{rank}(F'(x_*)) = 2$. Our results are given in Table A.1 in the appendix, and summarized in Tables 6.1 and 6.2 below.

Several comments about the test set are necessary. The 100$x_0$ case is excluded for three problems, Brown Almost Linear, Rosenbrock, and Wood, because the objective function overflowed at the starting point or during the first iteration. (The largest real number on the VAX is about $10^{38}$.) We ran many of the variable dimension problems for $n = 10, 20, 30, 40, 50$, and include here the $n = 30$ results, which are representative. The discrete boundary problem is very easy in all cases, we include $n = 10$ because it is slightly harder than $n = 30$. The Chebyquad problems also had overflow problems when the dimension and starting point were large, so we include the cases $n = 7$ and $9$ from $x_0$ and $n = 4$ from $10x_0$, which seemed representative. Our Variable Dimension problem is a slight alteration of the problem in Moré, Garbow, and Hillstrom, which has $n = 2$
equations in \( n \) unknowns; we simply eliminate their \( n-1^{st} \) and \( n^{th} \) equations. (The first \( n \) equations in the standard problem are linear.) Our Wood function is the gradient of the standard Wood function for minimization.

The three columns in Tables 6.1 - 6.5 labelled "Average ratio, Tensor method / Standard method," contain the averages of the ratios for each problem of the number of iterations, Jacobian evaluations, and function evaluations (excluding those used for finite difference Jacobians), respectively, used by the tensor method versus the standard method. For example, if the test set contained two problems for which the tensor method required 3 and 5 iterations, respectively, and the standard method 3 and 10 iterations, respectively, then the average iteration ratio would be \( \frac{1}{2} \left( \frac{3}{3} + \frac{5}{10} \right) = 0.75 \). Thus each problem is given equal weight. Since some of the problems are very easy, a second line in each table contains the same statistics using only those problems where the slower method required at least 10 iterations. The three columns labelled "Tensor better -- Standard better -- Tie" are based on a composite consideration of iterations and function evaluations; there were no problems where one method used more function evaluations but fewer iterations. For both algorithms, the number of Jacobian evaluations per problem always is one more than the number of iterations.

The most striking aspect of the test results on nonsingular problems is that the tensor method virtually never is less efficient than the standard method, and almost always is more efficient. In fact, on problems requiring ten or more iterations of the standard method, it is always more efficient. The gains in efficiency are considerable: an average of 21-23% improvement (depending on which measure is used) if all test problems, including some very easy problems where no gains are likely, are considered, and an average of 38-39% improvement on the harder problems. This combination of consistency with reasonable
Table 6.1 -- Summary for Problems with $F'(x_\ast)$ Nonsingular

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All problems</td>
<td>28</td>
<td>0.770 0.781 0.793</td>
<td>21 1 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harder problems only</td>
<td>14</td>
<td>0.612 0.636 0.656</td>
<td>14 0 0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Additional problems solved by standard method only: 2
by tensor method only: 1

Table 6.2 -- Summary for Powell's Singular Function

| Stopping Tolerance $10^{-6}$ | 3 | 0.442 0.469 0.510 | 3 | 0 | 0 |
| Stopping Tolerance $10^{-8}$ | 3 | 0.343 0.385 0.403 | 3 | 0 | 0 |

Table 6.3 -- Summary for First Singular Test Set with Rank $(F'(x_\ast)) = n - 1$

| All Problems | 17 | 0.576 0.609 0.603 | 15 | 0 | 2 |
| Harder Problems Only | 9 | 0.392 0.429 0.434 | 9 | 0 | 0 |

Table 6.4 -- Summary for Singular Test Set with Rank $(F'(x_\ast)) = n - 2$

| All Problems | 13 | 0.631 0.664 0.729 | 11 | 2 | 0 |
| Harder Problems Only | 7 | 0.499 0.535 0.542 | 7 | 0 | 0 |

Additional problems solved by standard method only: 1
by tensor method only: 5

Table 6.5 -- Summary for Second Singular Test Set with Rank $(F'(x_\ast)) = n - 1$

| All Problems | 16 | 0.801 0.806 0.849 | 11 | 2 | 3 |
| Harder Problems Only | 11 | 0.701 0.711 0.767 | 10 | 1 | 0 |

Additional problems solved by standard method only: 1
by tensor method only: 5

* Problems where slower method required at least 10 iterations
improvement in efficiency indicates that tensor methods may be preferable to standard methods for solving nonsingular systems of nonlinear equations.

Three nonsingular problems were solved by one method and not the other. (We discount the Watson function because the standard method never really found roots and the two methods always went to different regions.) The standard method solved the $10x_0$ Biggs Exp8 problem in 119 iterations while the tensor method didn't solve it in 150; the tensor method solved the Chebyquad $n=9$ problem in 33 iterations while the standard method didn't solve it in 150; the standard method solved the $10x_0$ Wood problem in 60 iterations while the tensor method didn't solve it in 150. This last problem illustrates an occasional difficulty when testing on pathological functions: the tensor method made better progress than the standard method during the first twelve iterations, but reached a point from which neither the tensor method nor the standard method could make reasonable progress. Overall, we noticed no large difference in the success rates of the standard and tensor methods, although the tensor method did have appreciably more successes on two of the three singular test sets.

Table 6.1 also excludes four problems (Box 3D from $10x_0$ and $100x_0$, and Watson from $x_0$ and $100x_0$) where the two methods converged to different solutions. The tensor method required fewer iterations and function evaluations than the standard method in two of these cases, and the same number in the other two.

On Powell's Singular Function, Table 6.2 shows that the tensor method was 49-56% less expensive, on the average, than the standard method. The stopping tolerance we use, that the relative size of $F'(x)^TF(x)$ must be less than $10^{-5}$, is a fairly loose stopping criterion that we believe is typical of tolerances used in practice. Table 6.2 shows that if this stopping tolerance is tightened to $10^{-8}$ (about the best one can achieve on a VAX using finite difference Jacobians), the
average cost of the tensor method on Powell's Singular Function is 60-66\% less than the corresponding cost of the standard method. Presumably this is a reflection of faster local convergence by the tensor method on singular problems, an issue we comment upon later. All our subsequent tests on singular problems use the looser stopping tolerance, $10^{-6}$; the improvements by the tensor method over the standard method generally would be more dramatic with a tighter tolerance.

The only other singular nonlinear equations test problems in the literature that we are aware of are the small family proposed by Griewank [1980a]. These four problems all have $n=3$, either a one or two dimensional nullspace for $F'(x^*)$, and either a first or second order singularity. Our results on these problems are given in Table A2 in the appendix, but they don't appear very meaningful, because the standard method failed on 6 of the 12 runs and once converged to a different root than the tensor method. The tensor method was successful in all cases and always was more efficient than the standard method.

We created singular test problems by modifying the nonsingular test problems of Moré, Garbow, and Hillstrom in two different ways. The first is to create problems of the form

$$\tilde{F}(x) = F(x) - F'(x^*) A(A^TA)^{-1}A^T (x-x^*)$$

(6.1)

where $F(x)$ is the standard nonsingular test function, $x^*$ is its root, and $A \in \mathbb{R}^{n \times k}$ has full column rank with $1 \leq k \leq n$. Clearly $\tilde{F}(x^*) = 0$ and

$$\tilde{F}'(x^*) = F'(x^*) [I - A(A^TA)^{-1}A^T]$$

has rank $n-k$. A disadvantage of this problem class is that $\tilde{F}(x)$ may have roots that are not roots of $F(x)$. There is likely to be a manifold of singular Jacobians of $\tilde{F}(x)$ in a neighborhood of $x^*$, but this is not guaranteed. A manifold of singularities is considered desirable because it makes the problems harder and
because we believe it is reflective of most practical singular problems.

We used (6.1) to create two sets of singular problems, with \( \hat{F}'(x) \) having rank \( n-1 \) and \( n-2 \) respectively, by using

\[
A \in \mathbb{R}^{n \times 1}, \quad A^T = (1, 1, \ldots, 1)
\]

and

\[
A \in \mathbb{R}^{n \times 2}, \quad A^T = \begin{bmatrix}
1 & 1 & 1 & 1 & \cdots & 1 \\
1 & -1 & 1 & -1 & \cdots & 1
\end{bmatrix}
\]

respectively. We tested our method on the singular versions of all the nonsingular problems except Watson's function, which we excluded because it was quite expensive to run and the two methods never converged to the same root. Our results are given in Tables A3 and A4 in the appendix and summarized in Tables 6.3 and 6.4.

The improvement by the tensor method over the standard method on the problems with \( \text{rank}(F'(x_*) = n-1 \) is substantial: an average of 40-43\% improvement on all problems and 57-61\% on the harder problems. We speculate this is due in part to the tensor method achieving superlinear convergence in this case, and comment further on this at the end of this section. In 9 cases the two methods converged to different roots; in 6 more cases they converged to the same root but not the singular root \( x_* \). These problems are excluded from the summary statistics in Table 6.3; they point out a deficiency of the test set.

On the test set with \( \text{rank}(F'(x_*) = n-2 \), the improvement by the tensor method over the standard method was 27-37\% on all problems and 46-50\% on the harder problems. These are still substantial gains but not as large as when \( \text{rank}(F'(x_*) = n-1 \). We speculate in section 7 that our tensor method is not necessarily superlinearly convergent in this case, and mention some modifications that might make it superlinearly convergent when the rank of \( F''(x_*) \) is less than \( n-1 \). The tensor method also solved 5 of the rank \( n-2 \) prob-
lems that the standard method didn't, the standard method solved one that the
tensor method didn't.

Our second method for generating singular test functions from standard
nonsingular problems has the desirable property that \( x_0 \) is a root of the new
singular function \( \tilde{F}(x) \) if and only if it is a root of the original nonsingular
function \( F(x) \). This class of functions is described in Theorem 6.1. The functions we
generated using this method turned out to be less useful test problems than the
singular functions already described, for reasons we will discuss. However they
may be a useful class of singular problems for future testing.

**Theorem 6.1.** Let \( F(x): \mathbb{R}^n \rightarrow \mathbb{R}^n \) and \( x_0 \in \mathbb{R}^n \) with \( F(x_0) = 0 \) and \( F'(x_0) \) nonsingular. Define \( D_{F_2}(x) \in \mathbb{R}^{n \times n} \) by

\[
D_{F_2}(x) = \text{diag}\{f_1(x)^2, \ldots, f_n(x)^2\}
\]

where \( f_i(x) \) denotes the \( i \)th component function of \( F(x) \). Let \( A \in \mathbb{R}^{n \times k} \), \( 1 \leq k \leq n \),
have full column rank, and let \( u \in \mathbb{R}^n \) have the property with \( A \) that \( (Au)[i] = 0 \)
only if row \( i \) of \( A \) = 0. Define \( \tilde{F}(x): \mathbb{R}^n \rightarrow \mathbb{R}^n \) by

\[
\tilde{F}(x) = [I - A(A^T A)^{-1} A^T] F(x) + \frac{1}{2} D_{F_2}(x) Au.
\]

(6.2)

Then \( \tilde{F}(x) = 0 \) if and only if \( F(x) = 0 \), and \( \tilde{F}'(x_0) \) has rank \( n-k \).

**Proof:** It is obvious that \( \tilde{F}(x) = 0 \) if \( F(x) = 0 \). Now suppose \( \tilde{F}(x) = 0 \), and con-
sider first the case when \( A \) has no nonzero rows. Then by definition \( u = Au \) has
no nonzero components, and

\[
0 = u^T A^T \tilde{F}(x) = \frac{1}{2} u^T A^T D_{F_2}(x) Au = \frac{1}{2} \sum_{i=1}^n (f_i(x)u[i])^2
\]

which implies that \( F(x) = 0 \). Now consider the case when \( A \) has some zero rows,
and assume without loss of generality that \( A^T = [B^T \mid 0] \) where \( B \in \mathbb{R}^{m \times k} \), \( m \geq k \),
has full column rank. Let \( G(x): \mathbb{R}^n \rightarrow \mathbb{R}^m \) and \( H(x): \mathbb{R}^n \rightarrow \mathbb{R}^{n-m} \) denote the first \( m \).
and last \(n-m\) components of \(F(x)\) respectively, and similarly let \(\tilde{G}(x)\) and \(\tilde{H}(x)\) denote the first \(m\) and last \(n-m\) components of \(\tilde{F}(x)\). Then

\[
\tilde{G}(x) = [I - B(B^TB)^{-1}B^T] G(x) + \frac{1}{2} D_{G_2}(x) Bu
\]

and

\[
\tilde{H}(x) = H(x)
\]

where

\[
D_{G_2}(x) = \text{diag}\{f_1(x)^2, \ldots, f_m(x)^2\}.
\]

It follows by the same argument as we used above that \(G(\bar{x}) = 0\) if \(\tilde{G}(\bar{x}) = 0\), and since \(\tilde{H}(\bar{x})\) and \(H(\bar{x})\) are identical, \(F(\bar{x}) = 0\) if \(\tilde{F}(\bar{x}) = 0\).

Finally, it is straightforward to verify that

\[
\tilde{F}'(x) = ([I - A(A^TA)^{-1}A^T] + D_F(x) D_w) F'(x)
\]

where

\[
D_F(x) = \text{diag}\{f_1(x), \ldots, f_n(x)\}, \quad D_w = \text{diag}\{w[1], \ldots, w[n]\}.
\]

Thus

\[
\tilde{F}'(x_\star) = [I - A(A^TA)^{-1}A^T] F'(x_\star)
\]

has rank \(n-k\).

\(\tilde{F}'(x)\) given by (6.3) almost always has a manifold of singularities around \(x_\star\).

For example, if \(A\) is the first \(k\) columns of the identity matrix, then \(f_i(x) = u[i] f_i(x)^2, i = 1, \ldots, k\), so \(\tilde{F}'(x)\) is singular whenever any \(f_i(x) = 0, i = 1, \ldots, k\).

More generally, it follows from (6.3) that \(\tilde{F}'(x)\) is singular whenever \(F'(x)\) has \((n+1-k)\) zero components; this usually implies a manifold of singularities around \(x_\star\) whenever \(k \geq 2\). Finally, it is easy to show that \(\tilde{F}'(x)\) is singular whenever
\[ z[i] + f_i(z) w[i] (z[i] + (Ay)[i]) = 0, \quad i=1,...,m \]  
where \( z \in \mathbb{R}^n \) is any vector in the null space of \( A^T \) and \( y \) is any vector in \( \mathbb{R}^k \); for small \( z \), (6.4) is likely to have solutions.

We used (6.2) to generate two sets of singular test problems with \( \text{rank}(F'(x_*)) = n-1 \), by setting \( k=1, u=1, \) and \( A = (1,1,\ldots,1) \) and \( A = (1,0,0,\ldots,0,1) \) respectively. Neither was a very illuminating test set. The problems with \( A = (1,1,\ldots,1) \) were too hard for either algorithm, each solved only 30-40% of the problems. In addition, there were numerous overflows due to the squares of the original component functions appearing in the new problems, and the small exponent range of the VAX. The standard method solved 9 problems, overflowed on 7, and failed on 19, the tensor method solved 11, overflowed on 6, and failed on 18. On the few problems solved by both methods, the tensor method was always at least as efficient as the standard method, with improvements ranging from 0 to 62%. The problems with \( A = (1,0,0,\ldots,0,1) \) were easier although there still was a considerable number of overflows and failures. The results are given in Table A5 and summarized in Table 6.5. The standard method solved 20 problems, overflowed on 5, and failed on 9, the tensor method solved 24, overflowed on 4, and failed on 6. The average improvement by the tensor method was 15-20%, 21-30% on the problems that required at least ten iterations. These improvements are lower than on the first set of nonsingular problems but do not reflect the higher success rate of the tensor method.

Taken together, our test results seem to indicate that the tensor method is consistently more efficient than standard methods in solving problems where \( F'(x_*) \) has a small rank deficiency. We speculate that when \( F'(x_*) \) has rank \( n-1 \), the tensor method is superlinearly convergent in most cases. To check whether this is a reasonable possibility, we examined the sequence of ratios 

\[ \frac{||x_k - x_*||}{||x_{k-1} - x_*||} \]

produced by the standard and tensor methods on problems with \( \text{rank}(F'(x_*)) = \)
\( n - 1 \). Ratios for a typical problem are given in Table 6.6. In almost all cases, the standard method exhibited local linear convergence with constant very nearly 0.5, as the analysis of various authors mentioned in Section 1 would predict. The local convergence of the tensor method clearly is faster; the final ratio of about 0.01 is typical and might be smaller if analytic Jacobians or tighter stopping tolerances were used. Whether this is superlinear convergence remains to be determined.

---

### Table 6.6 -- Speed of convergence on a typical problem with rank \( \hat{F}'(x^*) = n - 1 \)

(Broyden Banded, \( n = 30 \), as modified by (6.1), started from \( 10x_0 \))

<table>
<thead>
<tr>
<th>Iteration ((k))</th>
<th>( |x_k - x^<em>| / |x_{k-1} - x^</em>| )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 1 )</td>
<td>( 0.638 )</td>
</tr>
<tr>
<td>( 2 )</td>
<td>( 0.511 )</td>
</tr>
<tr>
<td>( 3 )</td>
<td>( 0.502 )</td>
</tr>
<tr>
<td>( 4 )</td>
<td>( 0.426 )</td>
</tr>
<tr>
<td>( 5 )</td>
<td>( 0.330 )</td>
</tr>
<tr>
<td>( 6 )</td>
<td>( 0.204 )</td>
</tr>
<tr>
<td>( 7 )</td>
<td>( 0.0916 )</td>
</tr>
<tr>
<td>( 8 )</td>
<td>( 0.0106 )</td>
</tr>
<tr>
<td>( 9 )</td>
<td></td>
</tr>
<tr>
<td>( 10 )</td>
<td></td>
</tr>
<tr>
<td>( 11 )</td>
<td></td>
</tr>
<tr>
<td>( 12 )</td>
<td></td>
</tr>
<tr>
<td>( 13 )</td>
<td></td>
</tr>
<tr>
<td>( 14 )</td>
<td></td>
</tr>
<tr>
<td>( 15 )</td>
<td></td>
</tr>
<tr>
<td>( 16 )</td>
<td></td>
</tr>
<tr>
<td>( 17 )</td>
<td></td>
</tr>
</tbody>
</table>
It is important to comment that on the test problems where either method had difficulty, it appeared to us that a trust region method that biased short steps towards the steepest descent direction often would have helped. We used a line search algorithm in our tests because we did not want to introduce the unresolved questions about trust region strategies for nonlinear equations into our comparison of the standard and tensor models. It will be important to investigate whether the computational comparison between methods using these two models is similar in a trust region setting.

Finally, we make some comments on details of the tensor algorithm we observed in our testing. The linear independence requirement usually limited the number of past points interpolated at each iteration to a smaller number than the upper bound \( \sqrt{n} \). For example on the 100x0 Broyden banded problem where \( n = 30 \), the algorithm used one past point in 83\% of the iterations and two past points in the remaining 17\%, although it could have used up to six past points; similarly on the x0 Trigonometric problem where \( n \) also is 30, it used one, two, and three past points on 20\%, 60\%, and 20\% of the iterations, respectively. Thus the tensor method seems to obtain surprisingly large improvement from a comparatively small amount of additional information. We tested the algorithm on the nonsingular problems with the linear independence angle reduced to 22.5° (from 45°); there was some fluctuation in the results on individual problems but no overall improvement or deterioration in efficiency, and the number of past points interpolated at each iteration increased somewhat but not dramatically. From past experience, a very small angle, say less than 10°, would give inferior results. The system of linear equations that is solved as part of solving the system of quadratics at each iteration was square and reasonably well conditioned (i.e. \( q = p \)) almost all of the time; \( q \) was greater than \( p \) at about 71\% of the iterations on the singular and nonsingular Biggs functions, and at about 3\% of the iterations on all the other test problems. While the step to the
root of the tensor model is not guaranteed to be in a descent direction, on the
nonsingular problems this only occurred on 5 of the 44 problems, and there only
about 25% of the time, mostly when the method got stuck in one place.

7. Future research on tensor models

The computational results in section 6 indicate that tensor methods may be
preferable to standard methods for solving general systems of nonlinear equa-
tions where analytic or finite difference Jacobians are available, and that they
may have a substantial advantage on problems where the Jacobian at the solu-
tion has a small rank deficiency. To firmly establish such a conclusion, addi-
tional testing is required, including tests comparing trust region versions of
standard and tensor methods for nonlinear equations. Our inclination is to use
dogleg-like methods in these trust region tests.

It would be very helpful to obtain local convergence results for our tensor
algorithm applied to singular problems. Hopefully, the algorithm can be shown
to converge faster than linearly to a root \( x \) where \( F'(x) \) has rank \( n-1 \) and
\( F''(x) \) obeys appropriate conditions. Related results of this type recently have
been obtained by Griewank [1983]. Griewank shows that an algorithm that also
bases its iteration on a quadratic model with a simple second order term is
locally 2-step q-superlinearly convergent in the above case. His algorithm, how-
ever, forms the second order term in the quadratic model using information
about the singularity in \( F'(x) \) that would not be available to general purpose
nonlinear equations solvers.

We believe that the tensor method presented in this paper may not always
achieve faster than linear convergence on problems where the rank of \( F'(x) \) is
n - 2 or less. To justify this remark, suppose $z_1, \ldots, z_m$ is a basis for the null space $Z$ of $F'(x^*)$, where $m > 1$. For a method based on a quadratic model to achieve fast convergence, it seems necessary that the second order term in the model be a good approximation to $F''(x^*)$ acting on $Z$. This seems to require $(m^2 + m) / 2$ $n$-vectors of information, to characterize $F''(x^*)z_i z_j$, $1 \leq i \leq j \leq m$.

Our method, however, may not contain this much information even if the past points are in the desired directions; for example, if all the past directions were in $Z$, our method would interpolate at most $m$ function values. Thus our method does not seem to interpolate enough information to always achieve fast convergence on problems where the dimension of the null space of $F'(x^*)$ is greater than one. This speculation is not well supported by our computational results, however; our tensor method seems to perform almost as well on problems where $F'(x^*)$ has rank $n - 2$ as where $F'(x^*)$ has rank $n - 1$.

There are several ways to incorporate more information into our tensor model and eliminate the objection raised in the previous paragraph. One is to interpolate values of $F(x_0 + s_k)$ at a set of points for which the steps $\{s_k\}$ may not meet the linear independence criterion of section 3, requiring instead that the matrix $S \in R^{m \times p}$ used in Theorem 3.1 meets this criterion. It is easy to show that this procedure would allow choosing up to $(m^2 + m) / 2$ directions from an $m$ dimensional subspace while leaving the calculation of the second order term $T_c$ well conditioned. A second alternative is to choose $T_c$ using information from Jacobians at past iterations. We intend to investigate these alternatives.

The methods proposed in this paper can be adapted easily to remain efficient on large, sparse systems of nonlinear equations. In particular, the main additional computational costs of our method are Jacobian-vector products, and presumably these can be performed efficiently when the Jacobian is sparse. Two modifications required would be to select the maximum number of past points
small enough that the cost of solving \( p \times p \) linear and quadratic subproblems remains acceptably small, and to use an efficient sparse factorization in the algorithm for solving the tensor model.

The methods discussed in this paper also could be applied with very little modification to nonlinear least squares problems. Nonlinear least squares algorithms virtually always use analytic or finite difference Jacobians so the requirements of the tensor methods presented in this paper are no restriction in this case. The augmentation of the linear model by a second order term would lead to natural extensions of Gauss-Newton or Levenberg-Marquardt methods, and tensor methods might require fewer iterations and function evaluations than these methods, especially on problems where the Jacobian at the solution is rank deficient. It is not clear how tensor methods for nonlinear least squares would perform on large residual problems, and whether there is any reason to prefer them to quasi-Newton methods like those of Dennis, Gay, and Welsch [1981] in this case.

We are currently developing extensions of our tensor methods to secant methods for nonlinear equations, and to unconstrained minimization. Neither extension is straightforward. In secant methods for nonlinear equations, analytic or finite difference Jacobians are not available, but it is possible to interpolate all the function values \( F(x_k + s_k) \) used in section 3 with a linear model (see e.g. Gay and Schnabel [1978]). To create a useful second order term it is necessary to interpolate function values in (nearly) dependent directions. The primary difficulty in extending tensor methods to unconstrained minimization is that for problems where the Hessian matrix at the minimizer is singular, an approximation not to the third but to the fourth derivative matrix is necessary to speed convergence. This is because the projection of the third derivative onto the null space of the Hessian must be zero at such a minimizer. In addition, all derivative approximations for minimization must be symmetric. Our
solutions to these difficulties will be reported in future papers.

Acknowledgements

We thank Richard Byrd and Andreas Griewank for many helpful discussions related to this paper.
B. References


P. Frank [1982], "A second-order local model for solution of systems of nonlinear equations", M.S. Thesis, Department of Computer Science, University of Colorado at Boulder.


A. O. Griewank [1983], "On solving nonlinear equations with simple singularities or nearly singular solutions", preprint, Department of Mathematics, Southern Methodist University.


Appendix

Tables A1 - A5 contain the results of the tests described in Section 6. The problems in Table A1 are from Moré, Garbow, and Hillstrom [1981]. The problems in Tables A3 - A5 are singular modifications, described in Section 6, of the problems in Moré, Garbow, and Hillstrom. The starting points used for all these problems are the ones suggested by Moré, Garbow, and Hillstrom; the third column of each table designates whether the starting point is \( x_0 \), \( 10x_0 \), or \( 100x_0 \), where \( x_0 \) is the point listed in Moré, Garbow, and Hillstrom. The problems in Table A2 are taken from Griewank [1980a].

The two columns in each table labelled \("\frac{1}{2}\|F(x_u)\|_2^2"\) give half the sum of squares of the component functions values at the final iterate for the standard method and tensor method, respectively, using abbreviated notation: e.g., \( .43-12 \) means \( 0.43 \times 10^{-12} \). If the method failed on a problem, this column instead contains one of the following alphabetic codes:

\[ \begin{align*}
O & \quad \text{method overflowed} \\
D & \quad \text{divergence detected (5 consecutive very long steps)} \\
F & \quad \text{method failed to find a root in 150 iterations} \\
S & \quad \text{method stalled at non-root}
\end{align*} \]

The rightmost column in each table, labelled "same \( x_u \) ?", contains a Y (yes) if the two methods converged to the same point on this problem, a N (no) otherwise. Only problems that converged to the same root are included in the statistics in Tables 6.1 - 6.5.

In Tables A3 - A5, the two columns labelled "n.s. \( x_u \) ?" (short for "nonsingular \( x_u \)?"), contain a Y (yes) if the method converged to the same root as the corresponding nonsingular problem in Table A1, a N (no) otherwise. Only problems were both methods converged to the same root and the same root as in the nonsingular case are included in the statistics in Tables 6.3 - 6.5.
Table A1 -- Results on More', Garbow, and Hillstrom Test Set

<table>
<thead>
<tr>
<th>Function</th>
<th>n</th>
<th>$x_0$</th>
<th>Standard Method</th>
<th>Tensor Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Fcn vals</td>
<td>Itns</td>
<td>$||F(x)|^2$</td>
</tr>
<tr>
<td>Biggs</td>
<td>6</td>
<td>1 209 104</td>
<td>.97-18</td>
<td>157 70</td>
</tr>
<tr>
<td>Exp6</td>
<td>10</td>
<td>228 119</td>
<td>.22-5</td>
<td>200 150</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>306 150</td>
<td>F</td>
<td>395 150</td>
</tr>
<tr>
<td>Box 3D</td>
<td>3</td>
<td>1 5 4</td>
<td>.10-16</td>
<td>4 3</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>20 14</td>
<td>.56-19</td>
<td>12 8</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>4 3</td>
<td>.16-8</td>
<td>4 3</td>
</tr>
<tr>
<td>Brown Alm.</td>
<td>10</td>
<td>1 15 11</td>
<td>.57-12</td>
<td>10 7</td>
</tr>
<tr>
<td>Linear</td>
<td>10</td>
<td>11 9</td>
<td>.25-15</td>
<td>10 7</td>
</tr>
<tr>
<td>Broyden</td>
<td>30</td>
<td>1 6 5</td>
<td>.12-15</td>
<td>5 4</td>
</tr>
<tr>
<td>Banded</td>
<td>10</td>
<td>12 11</td>
<td>.16-22</td>
<td>9 8</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>17 16</td>
<td>.95-14</td>
<td>14 13</td>
</tr>
<tr>
<td>Broyden</td>
<td>30</td>
<td>1 5 4</td>
<td>.56-18</td>
<td>5 4</td>
</tr>
<tr>
<td>Tridiag.</td>
<td>10</td>
<td>8 7</td>
<td>.47-13</td>
<td>6 5</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>11 10</td>
<td>.64-10</td>
<td>6 5</td>
</tr>
<tr>
<td>Chebyquad</td>
<td>7</td>
<td>1 11 7</td>
<td>.38-16</td>
<td>8 6</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>332 150</td>
<td>F</td>
<td>63 33</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>10 51 35</td>
<td>.90-20</td>
<td>44 24</td>
</tr>
<tr>
<td>Discrete</td>
<td>10</td>
<td>1 3 2</td>
<td>.48-15</td>
<td>3 2</td>
</tr>
<tr>
<td>Boundary</td>
<td>10</td>
<td>4 3</td>
<td>.15-17</td>
<td>4 3</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>9 8</td>
<td>.10-13</td>
<td>8 7</td>
</tr>
<tr>
<td>Discrete</td>
<td>30</td>
<td>1 3 2</td>
<td>.18-12</td>
<td>3 2</td>
</tr>
<tr>
<td>Integral</td>
<td>10</td>
<td>4 3</td>
<td>.49-15</td>
<td>4 3</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>9 8</td>
<td>.40-11</td>
<td>9 8</td>
</tr>
<tr>
<td>Helical</td>
<td>3</td>
<td>1 12 9</td>
<td>.17-20</td>
<td>11 8</td>
</tr>
<tr>
<td>Valley</td>
<td>10</td>
<td>17 13</td>
<td>.23-21</td>
<td>16 11</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>21 16</td>
<td>.22-14</td>
<td>16 11</td>
</tr>
<tr>
<td>Powell</td>
<td>4</td>
<td>1 9 8</td>
<td>.19-7</td>
<td>4 3</td>
</tr>
<tr>
<td>Singular</td>
<td>10</td>
<td>13 12</td>
<td>.43-8</td>
<td>6 5</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>16 15</td>
<td>.10-7</td>
<td>10 8</td>
</tr>
<tr>
<td>Rosenbrock</td>
<td>2</td>
<td>1 27 14</td>
<td>.10-7</td>
<td>12 7</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>5 3</td>
<td>.0</td>
<td>6 4</td>
</tr>
<tr>
<td>Trigonometric</td>
<td>30</td>
<td>1 31 16</td>
<td>.19-12</td>
<td>21 11</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>328 150</td>
<td>F</td>
<td>145 64</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>334 150</td>
<td>F</td>
<td>132 62</td>
</tr>
<tr>
<td>Variable</td>
<td>10</td>
<td>1 14 13</td>
<td>.24-9</td>
<td>6 5</td>
</tr>
<tr>
<td>Dimension</td>
<td>10</td>
<td>15 14</td>
<td>.10-8</td>
<td>5 4</td>
</tr>
<tr>
<td>(Altered)</td>
<td>100</td>
<td>19 16</td>
<td>.10-8</td>
<td>14 11</td>
</tr>
<tr>
<td>Watson</td>
<td>31</td>
<td>5 4</td>
<td>.10-6</td>
<td>5 4</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>41 25</td>
<td>.25-6</td>
<td>378 150</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>45 28</td>
<td>.35-6</td>
<td>27 17</td>
</tr>
<tr>
<td>Wood Gradient</td>
<td>4</td>
<td>1 24 17</td>
<td>.23-28</td>
<td>11 9</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>105 61</td>
<td>.30-24</td>
<td>366 150</td>
</tr>
</tbody>
</table>
| Function            | $n$ | $x_0$ | Standard Method | Tensor Method | same $x_*$?
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fen vals</td>
<td>Itns</td>
<td>$|F(x_*)|_2^2$</td>
</tr>
<tr>
<td>Dimension of</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td>8</td>
<td>.2-10</td>
</tr>
<tr>
<td>Nullspace = 1</td>
<td>10</td>
<td>15</td>
<td>14</td>
<td>.1-10</td>
<td>7</td>
</tr>
<tr>
<td>Order of</td>
<td>100</td>
<td>21</td>
<td>19</td>
<td>.1-10</td>
<td>18</td>
</tr>
<tr>
<td>Singularity = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimension of</td>
<td>3</td>
<td>1</td>
<td>11</td>
<td>9</td>
<td>.3-10</td>
</tr>
<tr>
<td>Nullspace = 1</td>
<td>10</td>
<td>332</td>
<td>150</td>
<td>F</td>
<td>11</td>
</tr>
<tr>
<td>Order of</td>
<td>100</td>
<td>297</td>
<td>150</td>
<td>F</td>
<td>17</td>
</tr>
<tr>
<td>Singularity = 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimension of</td>
<td>3</td>
<td>1</td>
<td>11</td>
<td>9</td>
<td>.3-10</td>
</tr>
<tr>
<td>Nullspace = 2</td>
<td>10</td>
<td>332</td>
<td>150</td>
<td>F</td>
<td>11</td>
</tr>
<tr>
<td>Order of</td>
<td>100</td>
<td>297</td>
<td>150</td>
<td>F</td>
<td>17</td>
</tr>
<tr>
<td>Singularity = 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table A3 -- Results on First Singular Test Set with Rank \((F'(x_*) = n-1)\)

<table>
<thead>
<tr>
<th>Function</th>
<th>n</th>
<th>(x_0)</th>
<th>Fcn</th>
<th>Itns</th>
<th>(%|F'(x_*)|_2^2)</th>
<th>n.s.</th>
<th>Fcn</th>
<th>Itns</th>
<th>(%|F'(x_*)|_2^2)</th>
<th>n.s.</th>
<th>same</th>
<th>x_*?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biggs</td>
<td>6</td>
<td>1</td>
<td>74</td>
<td>52</td>
<td>.12-10</td>
<td>N</td>
<td>411</td>
<td>150</td>
<td>F</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Exp6</td>
<td>10</td>
<td>33</td>
<td>19</td>
<td>.66-17</td>
<td>N</td>
<td>69</td>
<td>38</td>
<td>.14-6</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>6</td>
<td>5</td>
<td>D</td>
<td>N</td>
<td>305</td>
<td>150</td>
<td>F</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box 3D</td>
<td>3</td>
<td>1</td>
<td>10</td>
<td>9</td>
<td>.96-13</td>
<td>Y</td>
<td>6</td>
<td>5</td>
<td>.57-15</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td>OF</td>
<td></td>
<td></td>
<td>11</td>
<td>10</td>
<td>.42-7</td>
<td>N</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>4</td>
<td>3</td>
<td>.46-15</td>
<td>N</td>
<td>17</td>
<td>.46-15</td>
<td>N</td>
<td>OF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brown Alm.</td>
<td>10</td>
<td>7</td>
<td>6</td>
<td>.33-7</td>
<td>Y</td>
<td>5</td>
<td>4</td>
<td>.41-7</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>10</td>
<td>22</td>
<td>21</td>
<td>.51-7</td>
<td>Y</td>
<td>9</td>
<td>7</td>
<td>.39-9</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Broyden</td>
<td>30</td>
<td>1</td>
<td>12</td>
<td>11</td>
<td>.88-11</td>
<td>Y</td>
<td>6</td>
<td>5</td>
<td>.65-14</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Banded</td>
<td>10</td>
<td>16</td>
<td>17</td>
<td>.11-10</td>
<td>Y</td>
<td>10</td>
<td>9</td>
<td>.54-15</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>24</td>
<td>23</td>
<td>.37-11</td>
<td>Y</td>
<td>14</td>
<td>13</td>
<td>.48-11</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Broyden</td>
<td>30</td>
<td>1</td>
<td>9</td>
<td>8</td>
<td>.15-9</td>
<td>Y</td>
<td>6</td>
<td>5</td>
<td>.80-12</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Tridiag.</td>
<td>10</td>
<td>14</td>
<td>13</td>
<td>.10-9</td>
<td>Y</td>
<td>6</td>
<td>5</td>
<td>.15-13</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>17</td>
<td>16</td>
<td>.32-9</td>
<td>Y</td>
<td>5</td>
<td>4</td>
<td>.16-12</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Chebyquad</td>
<td>7</td>
<td>1</td>
<td>11</td>
<td>7</td>
<td>.38-16</td>
<td>Y</td>
<td>8</td>
<td>6</td>
<td>.98-14</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>1</td>
<td>66</td>
<td>33</td>
<td>.29-13</td>
<td>N</td>
<td>294</td>
<td>103</td>
<td>.60-13</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>10</td>
<td>200</td>
<td>16</td>
<td>.90-16</td>
<td>N</td>
<td>78</td>
<td>42</td>
<td>.19-17</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Discrete</td>
<td>10</td>
<td>3</td>
<td>2</td>
<td>.16-9</td>
<td>Y</td>
<td>3</td>
<td>2</td>
<td>.14-9</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Boundary</td>
<td>10</td>
<td>5</td>
<td>4</td>
<td>.56-10</td>
<td>Y</td>
<td>5</td>
<td>4</td>
<td>.20-13</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>9</td>
<td>6</td>
<td>.43-10</td>
<td>N</td>
<td>7</td>
<td>6</td>
<td>.10-13</td>
<td>N</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrete</td>
<td>30</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>.22-9</td>
<td>Y</td>
<td>4</td>
<td>3</td>
<td>.20-11</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Integral</td>
<td>10</td>
<td>6</td>
<td>5</td>
<td>.82-9</td>
<td>Y</td>
<td>5</td>
<td>4</td>
<td>.68-12</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>10</td>
<td>9</td>
<td>.65-16</td>
<td>N</td>
<td>9</td>
<td>8</td>
<td>.32-23</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Hecial</td>
<td>3</td>
<td>1</td>
<td>8</td>
<td>7</td>
<td>.17-16</td>
<td>N</td>
<td>7</td>
<td>6</td>
<td>.64-14</td>
<td>N</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Valley</td>
<td>10</td>
<td>7</td>
<td>6</td>
<td>.38-14</td>
<td>N</td>
<td>7</td>
<td>6</td>
<td>.27-13</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>7</td>
<td>6</td>
<td>.10-14</td>
<td>N</td>
<td>7</td>
<td>6</td>
<td>.70-22</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Rosenbrock</td>
<td>2</td>
<td>1</td>
<td>15</td>
<td>14</td>
<td>.16-13</td>
<td>Y</td>
<td>4</td>
<td>3</td>
<td>.47-14</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>17</td>
<td>16</td>
<td>.77-13</td>
<td>Y</td>
<td>5</td>
<td>4</td>
<td>.22-20</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Trigonometric</td>
<td>30</td>
<td>1</td>
<td>19</td>
<td>11</td>
<td>.10-10</td>
<td>N</td>
<td>24</td>
<td>13</td>
<td>.82-10</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>10</td>
<td>1</td>
<td>14</td>
<td>13</td>
<td>.24-9</td>
<td>Y</td>
<td>10</td>
<td>9</td>
<td>.46-15</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Dimension</td>
<td>10</td>
<td>15</td>
<td>14</td>
<td>.63-12</td>
<td>N</td>
<td>12</td>
<td>11</td>
<td>.16-13</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>(Altered)</td>
<td>100</td>
<td>18</td>
<td>17</td>
<td>.14-15</td>
<td>N</td>
<td>11</td>
<td>9</td>
<td>.18-15</td>
<td>N</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wood</td>
<td>4</td>
<td>1</td>
<td>21</td>
<td>20</td>
<td>.22-15</td>
<td>Y</td>
<td>12</td>
<td>11</td>
<td>.41-18</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Gradient</td>
<td>10</td>
<td>26</td>
<td>25</td>
<td>.94-15</td>
<td>Y</td>
<td>15</td>
<td>14</td>
<td>.15-20</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>
Table A4 — Results on Singular Test Set with Rank $\langle F'(x_\ast) \rangle = n-2$

<table>
<thead>
<tr>
<th>Function</th>
<th>$n$</th>
<th>$x_0$</th>
<th>Standard Method</th>
<th>Tensor Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fcn</td>
<td>ltns</td>
<td>$|F'(x\ast)|_2^2$</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>vals</td>
<td></td>
<td>$x\ast$?</td>
<td></td>
</tr>
<tr>
<td>Biggs</td>
<td>6</td>
<td>1</td>
<td>27</td>
<td>17</td>
</tr>
<tr>
<td>Exp6</td>
<td>10</td>
<td>306</td>
<td>150</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>6</td>
<td>5</td>
<td>D</td>
</tr>
<tr>
<td>Box 3D</td>
<td>3</td>
<td>1</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>11</td>
<td>10</td>
<td>.1-12</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>6</td>
<td>5</td>
<td>D</td>
</tr>
<tr>
<td>Brown Alm.</td>
<td>10</td>
<td>1</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Linear</td>
<td>10</td>
<td>19</td>
<td>16</td>
<td>.2-15</td>
</tr>
<tr>
<td>Broyden</td>
<td>30</td>
<td>1</td>
<td>191</td>
<td>99</td>
</tr>
<tr>
<td>Banded</td>
<td>10</td>
<td>304</td>
<td>150</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>25</td>
<td>21</td>
<td>.9-19</td>
</tr>
<tr>
<td>Broyden</td>
<td>30</td>
<td>1</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Tridiag.</td>
<td>10</td>
<td>298</td>
<td>150</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>19</td>
<td>18</td>
<td>.3-11</td>
</tr>
<tr>
<td>Chebyquad</td>
<td>7</td>
<td>1</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>335</td>
<td>150</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>301</td>
<td>150</td>
<td>F</td>
</tr>
<tr>
<td>Discrete Boundary</td>
<td>10</td>
<td>1</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>.3-10</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>12</td>
<td>11</td>
<td>.1-13</td>
</tr>
<tr>
<td>Discrete Integral</td>
<td>30</td>
<td>1</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>11</td>
<td>10</td>
<td>.6-10</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>9</td>
<td>8</td>
<td>.6-18</td>
</tr>
<tr>
<td>Helical Valley</td>
<td>3</td>
<td>1</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>15</td>
<td>14</td>
<td>.6-13</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>15</td>
<td>14</td>
<td>.6-13</td>
</tr>
<tr>
<td>Rosenbrock</td>
<td>2</td>
<td>1</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>13</td>
<td>12</td>
<td>.5-8</td>
</tr>
<tr>
<td>Trigonometric</td>
<td>30</td>
<td>1</td>
<td>316</td>
<td>150</td>
</tr>
<tr>
<td>Variable</td>
<td>10</td>
<td>1</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Dimension</td>
<td>10</td>
<td>16</td>
<td>15</td>
<td>.6-10</td>
</tr>
<tr>
<td>(Altered)</td>
<td>100</td>
<td>20</td>
<td>19</td>
<td>.4-10</td>
</tr>
<tr>
<td>Wood Gradient</td>
<td>4</td>
<td>1</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>27</td>
<td>26</td>
<td>.2-16</td>
</tr>
</tbody>
</table>
Table A5 -- Results on Second Singular Test Set with Rank \( F'(x_0) = n - 1 \)

<table>
<thead>
<tr>
<th>Function</th>
<th>( n )</th>
<th>( x_0 )</th>
<th>( n )</th>
<th>Hints</th>
<th>( |F(x_0)|_2^2 )</th>
<th>n.s. ( x_* )</th>
<th>( n )</th>
<th>Hints</th>
<th>( |F(x_0)|_2^2 )</th>
<th>n.s. ( x_* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biggs</td>
<td>6</td>
<td>1</td>
<td>10</td>
<td>292</td>
<td>OF</td>
<td>-</td>
<td>351</td>
<td>150</td>
<td>F</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>304</td>
<td>F</td>
<td>-</td>
<td>303</td>
<td>150</td>
<td>F</td>
<td>-</td>
</tr>
<tr>
<td>Exp6</td>
<td>3</td>
<td>1</td>
<td>10</td>
<td>13</td>
<td>OF</td>
<td>-</td>
<td>10</td>
<td>9</td>
<td>2-10</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>303</td>
<td>F</td>
<td>-</td>
<td>15</td>
<td>14</td>
<td>3-12</td>
<td>Y</td>
</tr>
<tr>
<td>Box 3D</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>40</td>
<td>4</td>
<td>39</td>
<td>44</td>
<td>33</td>
<td>2-10</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>302</td>
<td>OF</td>
<td>-</td>
<td>28</td>
<td>16</td>
<td>1-15</td>
<td>Y</td>
</tr>
<tr>
<td>Brown Alm.</td>
<td>30</td>
<td>1</td>
<td>10</td>
<td>12</td>
<td>11</td>
<td>.4-9</td>
<td>10</td>
<td>9</td>
<td>.1-15</td>
<td>Y</td>
</tr>
<tr>
<td>Linear</td>
<td></td>
<td></td>
<td>100</td>
<td>28</td>
<td>21</td>
<td>.6-9</td>
<td>12</td>
<td>11</td>
<td>.4-18</td>
<td>Y</td>
</tr>
<tr>
<td>Broyden</td>
<td>7</td>
<td>1</td>
<td>10</td>
<td>4</td>
<td>3</td>
<td>.7-9</td>
<td>4</td>
<td>3</td>
<td>.2-10</td>
<td>Y</td>
</tr>
<tr>
<td>Banded</td>
<td>9</td>
<td>1</td>
<td>100</td>
<td>15</td>
<td>4</td>
<td>.8-9</td>
<td>5</td>
<td>4</td>
<td>.4-11</td>
<td>Y</td>
</tr>
<tr>
<td>Tridiag.</td>
<td>10</td>
<td>1</td>
<td>100</td>
<td>15</td>
<td>2</td>
<td>.6-8</td>
<td>22</td>
<td>15</td>
<td>.1-15</td>
<td>Y</td>
</tr>
<tr>
<td>Chebyquad</td>
<td>30</td>
<td>1</td>
<td>10</td>
<td>12</td>
<td>2</td>
<td>.5-10</td>
<td>3</td>
<td>2</td>
<td>.5-10</td>
<td>Y</td>
</tr>
<tr>
<td>Discrete</td>
<td>9</td>
<td>1</td>
<td>100</td>
<td>15</td>
<td>1</td>
<td>.3-7</td>
<td>6</td>
<td>5</td>
<td>.4-15</td>
<td>Y</td>
</tr>
<tr>
<td>Integral</td>
<td>10</td>
<td>1</td>
<td>100</td>
<td>15</td>
<td>14</td>
<td>.1-17</td>
<td>12</td>
<td>11</td>
<td>.4-14</td>
<td>Y</td>
</tr>
<tr>
<td>Helical</td>
<td>3</td>
<td>1</td>
<td>10</td>
<td>20</td>
<td>18</td>
<td>.7-13</td>
<td>15</td>
<td>11</td>
<td>.3-13</td>
<td>Y</td>
</tr>
<tr>
<td>Valley</td>
<td>10</td>
<td>1</td>
<td>100</td>
<td>22</td>
<td>21</td>
<td>.1-12</td>
<td>21</td>
<td>16</td>
<td>.4-17</td>
<td>Y</td>
</tr>
<tr>
<td>Rosenbrock</td>
<td>2</td>
<td>1</td>
<td>30</td>
<td>15</td>
<td>2</td>
<td>.1-9</td>
<td>23</td>
<td>12</td>
<td>.3-10</td>
<td>N</td>
</tr>
<tr>
<td>Trigonometric</td>
<td>10</td>
<td>1</td>
<td>100</td>
<td>32</td>
<td>36</td>
<td>.6-10</td>
<td>17</td>
<td>16</td>
<td>.1-10</td>
<td>Y</td>
</tr>
<tr>
<td>Dimension</td>
<td>10</td>
<td>1</td>
<td>100</td>
<td>32</td>
<td>36</td>
<td>.4-10</td>
<td>17</td>
<td>16</td>
<td>.1-10</td>
<td>Y</td>
</tr>
<tr>
<td>(Altered)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>19</td>
<td>.2-9</td>
<td>Y</td>
</tr>
<tr>
<td>Wood</td>
<td>4</td>
<td>1</td>
<td>100</td>
<td>40</td>
<td>38</td>
<td>.2-12</td>
<td>329</td>
<td>150</td>
<td>F</td>
<td>-</td>
</tr>
<tr>
<td>Gradient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>452</td>
<td>150</td>
<td>F</td>
<td>-</td>
</tr>
</tbody>
</table>
Tensor Methods for Nonlinear Equations

Robert B. Schnabel, Paul D. Frank

University of Colorado, Computer Science
Campus Box 430, Boulder, CO 80309

U. S. Army Research Office
Post Office Box 12211
Research Triangle Park, NC 27709

Approved for public release; distribution unlimited.

NA

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Nonlinear equations, tensor, singular Jacobian

A new class of methods for solving systems of nonlinear equations, called tensor methods, is introduced. Tensor methods are general purpose methods intended especially for problems where the Jacobian matrix at the solution is singular or ill-conditioned. They base each iteration on a quadratic model of the nonlinear function, the standard linear model augmented by a simple second order term. The second order term is selected so that the model interpolates function values from several previous iterations, as well as the current function value and Jacobian. The tensor method requires
April 20, 1982

U. S. Army Research Office
Box 12211
Research Triangle Park
North Carolina  27709

Attention: Robin Haley Sorrell

Dear Ms. Sorrell:

We are sending to you by United Parcel Service 60 copies of Technical Report #CU-CS-243-82 by Robert B. Schnabel and Paul D. Frank. The report is entitled "Tensor Methods for Nonlinear Equations". This work was covered under grant #DAAG 29-81-K0108.

Sincerely,

Harriet Ortiz
Staff Assistant