Local Convergence Analysis of Tensor Methods for Nonlinear Equations

Dan Feng, Paul D. Frank, Robert B. Schnabel

CU-CS-591-92 April 1992

University of Colorado at Boulder
DEPARTMENT OF COMPUTER SCIENCE

LOCAL CONVERGENCE ANALYSIS OF TENSOR METHODS FOR NONLINEAR EQUATIONS

by

Dan Feng[1] , *Paul D. Frank*[2] , *Robert B. Schnabel*[3]

[1]Department of Computer Science, Campus Box 430, University of Colorado, Boulder, Colorado 80309-0430.

[2]Boeing Computer Services P.O. Box 24346 Seattle, WA 98124-0346

[3]Department of Computer Science, Campus Box 430, University of Colorado, Boulder, Colorado 80309-0430.

Abstract

Tensor methods for nonlinear equations base each iteration upon a standard linear model, augmented by a low rank quadratic term that is selected in such a way that the model is efficient to form, store, and solve. These methods have been shown to be very efficient and robust computationally, especially on problems where the Jacobian matrix at the root has a small rank deficiency. This paper analyzes the local convergence properties of two versions of tensor methods, on problems where the Jacobian matrix at the root has a null space of rank one. Both methods augment the standard linear model by a rank one quadratic term. We show under mild conditions that the sequence of iterates generated by the tensor method based upon an "ideal" tensor model converges locally and two-step Q-superlinearly to the solution with Q-order $\frac{3}{2}$, and that the sequence of iterates generated by the tensor method based upon a practical tensor model converges locally and three-step Q-superlinearly to the solution with Q-order $\frac{3}{2}$. In the same situation, it is known that standard methods converge linearly with constant converging to $\frac{1}{2}$. Hence, tensor methods have theoretical advantages over standard methods. Our analysis also confirms that tensor methods converge at least quadratically on problems where the Jacobian matrix at the root is nonsingular.

1 Introduction

Efficient numerical methods for solving the nonlinear equations problem

$$\textit{given } F : \Re^n \rightarrow \Re^n, \textit{ find } x^* \in \Re^n \textit{ such that } F(x^*) = 0, \tag{1.1}$$

have long been sought by scientists and engineers. Standard methods widely used in practice for solving (1.1) are iterative methods that base each iteration upon a linear model at current point x^c,

$$M(x^c + d) = F(x^c) + J_c d, \tag{1.2}$$

where $d \in \Re^n, J_c \in \Re^{n \times n}$. These methods can be put into two categories, those where J_c is the current Jacobian matrix or a finite difference approximation to it (Newton's methods), and those where J_c is a secant (quasi-Newton) approximation to the Jacobian. When the Jacobian is available, the linear model (1.2) becomes

$$M(x^c + d) = F(x^c) + F'(x^c)d. \tag{1.3}$$

This paper analyzes tensor methods that are extensions to the first category, those that use analytic (or finite difference) Jacobians.

The standard method for the nonlinear equations, Newton's method, is defined when $F'(x^c)$ is nonsingular, and consists of setting the next iteration x^+ to the root of (1.3),

$$x^+ = x^c - F'(x^c)^{-1}F(x^c). \tag{1.4}$$

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

$$\|x^{k+1} - x^*\| \leq c\|x^k - x^*\|^2$$

if $\|x^0 - x^*\| \leq \delta$.

Newton's method is not quickly locally convergent, however, if $F'(x^*)$ is singular. This situation is analyzed and acceleration techniques are suggested by many authors, including Reddien [15], Decker and Kelley [4],[5],[6], Decker, Keller and Kelley [3], Kelley and Suresh [13], Griewank and Osborne [12], and Griewank [11]. In summary, their papers show that when the Jacobian at the solution has a null space of dimension one, then from good starting points, Newton's method is locally Q-linearly convergent with constant converging to $\frac{1}{2}$. The acceleration techniques presented in these papers depend upon apriori knowledge that the problem is singular.

Tensor methods are intended to be efficient both for nonsingular problems and for problems with low rank deficiency. These methods augment the standard linear model by a low rank second order term, in a way that requires no additional function or Jacobian evaluations per iteration, and hardly more arithmetic per iteration or total storage, than Newton's method. The second order term supplies higher order information in recent step directions; when the Jacobian is (nearly) singular, this usually results in supplying information in directions where the Jacobian lacks information. Tensor methods were introduced by Schnabel and Frank [17], and shown to be considerably more efficient and robust than standard methods on both singular and nonsingular systems of nonlinear equations. They have since been extended to unconstrained optimization [16] and nonlinear least squares [2] with similar computational success.

A local convergence analysis of two versions of tensor methods for nonlinear equations is presented in this paper. This analysis is an extension of the unpublished analysis of tensor methods for nonlinear equations in Frank [9]. However, the convergence results presented here are much stronger, and the versions of tensor algorithms that are analyze are much closer to those that are implemented in practice. Furthermore, new techniques are developed in our new analysis, which make the proofs clearer and more succinct.

The convergence results in this paper are for algorithms using a rank one second order term. These are the simplest instances of tensor models and the only group having a closed-form solution. Since only a rank one second order term is used, a local convergence advantage over Newton's method is expected only on singular problems where $F'(x)$ has a null space of dimension one at the solution. Thus, only the one-dimensional null space case is considered.

The two tensor methods that we analyze here are an "ideal" tensor method and a practical tensor method. The ideal method is motived by the fact that the closer a singular value of $F'(x)$ is to zero, the less information $F'(x)$ provides in the direction spanned by the corresponding singular vector. If $F'(x)$ has a zero singular value at the solution x^*, then as x approaches x^*, $F'(x)$ approaches singularity, and the singular vector corresponding to the least singular value of $F'(x)$ approaches a null vector of $F'(x^*)$. From this observation, we see that an excellent choice for a second order term would be to supply second derivative information in the direction corresponding to the least singular value of $F'(x)$ at that iterate. We call the tensor method that does this an ideal tensor method. The ideal method is in fact readily implementable, but it might be impractical for some problems because it would require the computation of the smallest singular vector of the Jacobian, and one additional function evaluation (to obtain a finite difference approximation to the desired second derivative information), at each iteration. Nevertheless, the most important feature of the ideal method is that it is quite easy to analyze, and demonstrates the basic convergence behavior of tensor methods clearly. Furthermore, the techniques developed in analyzing the ideal method shed considerable light on the analysis of the practical method, which is our final target. The practical tensor method analyzed here is very close to that implemented in practice, and in fact the analysis provided here is guiding our current implementation.

The remainder of this paper is organized as follows. Tensor methods for nonlinear equations are reviewed in section 2. Some notation and assumptions needed in the analysis of both tensor methods are introduced in section 3. In section 4, an algorithm for an ideal tensor method is presented and analyzed. We show that when the Jacobian at the solution has a null space of rank one, then under mild conditions, a tensor method based upon the "ideal" tensor model produces a sequence of iterates that is locally two-step convergent to the solution with Q-order $\frac{3}{2}$. In section 5 we extend this result to show that a tensor method based on a practical tensor model produces a sequence of iterates that is locally three-step convergent with Q-order $\frac{3}{2}$, under similar conditions. Finally, in section 6 we make some brief comments about the relation of these results to the practical performance of tensor methods, and about possible extensions of these results.

2 Tensor methods for nonlinear equations

Tensor methods for nonlinear equations were introduced by Schnabel and Frank [17]. The tensor model used in these methods is a quadratic model of $F(x)$ formed by adding a second order term to (1.3), giving

$$M_T(x^c + d) = F(x^c) + J_c d + \frac{1}{2} T_c dd, \tag{2.1}$$

where $T_c \in \Re^{n \times n \times n}$ is intended to supply second order information about $F(x)$ around x^c. The second derivative of $F(x)$ at x^c, $F''(x^c) \in \Re^{n \times n \times n}$ is an obvious choice for T_c in (2.1). However this choice for T_c has several serious disadvantages that preclude its use in practice. These include the computation of $n^3/2$ second partial derivatives of $F(x)$ and a storage requirement of at least $n^3/2$ real numbers for $F''(x^c)$. Furthermore, to utilize the model (2.1) with $T_c = F''(x^c)$, at each iteration one would have to solve a system of n quadratic equations in n unknowns, which is expensive and might not have a root.

The difficulties associated with the use of $T_c = F''(x)$ in (2.1) are overcome in tensor methods by choosing T_c to have a restricted low rank form. This can be considered as an extension to second order objects of the low rank update methods used to approximate Jacobian or Hessian matrices in secant (quasi-Newton) methods. One difference is that for reasons of efficiency in arithmetic cost and storage, at each iteration the zero tensor is updated rather than the tensor from the previous iteration.

Formation of the tensor model in Schnabel and Frank [17] is based upon the interpolation of information from past iterates, and requires no additional function or derivative evaluations. This is done by selecting some set of independent past iterates $x^{-1}, \cdots, x^{-p}$ and requiring the model (2.1) to interpolate the function values $F(x^{-k})$ at these points. That is, the model is required to satisfy

$$F(x^{-k}) = F(x^c) + F'(x^c)s_k + \frac{1}{2}T_c s_k s_k, \quad k = 1, \cdots, p,$$

where

$$s_k = x^{-k} - x^c, \quad k = 1, \cdots, p.$$

The directions $\{s_k\}$ are required to be strongly linearly independent, which usually results in p being 1 or 2, although an upper bound of $p \leq \sqrt{n}$ is permitted. Then T_c is chosen to satisfy

$$\begin{array}{l} \min_{T_c \in \Re^{n\times n\times n}} \|T_c\|_F \\ \textit{subject to } T_c s_k s_k = z_k, k = 1, \cdots, p, \end{array} \tag{2.2}$$

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

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

and $z_k \in \Re^n$ is defined as

$$z_k = 2(F(x^{-k}) - F(x^c) - F'(x^c)s_k).$$

The solution of (2.2) is given by

$$T_c = \sum_{k=1}^{p} a_k s_k s_k, \tag{2.3}$$

where a_k is the kth column of $A \in \Re^{n\times p}$, A is defined by $A = ZM^{-1}$, $M \in \Re^{p\times p}$ is defined by $M[i,j] = (s_i^T s_j)^2$, $1 \leq i, j \leq p$, and $Z \in \Re^{n\times p}$ by column k of $Z = z_k$, $k = 1, \cdots, p$.

Substituting (2.3) into the tensor model (2.1) gives

$$M_T(x^c + d) = F(x^c) + F'(x^c)d + \tfrac{1}{2}\sum_{k=1}^{p} a_k(s_k^T d)^2. \tag{2.4}$$

The additional storage required by T_c is $2p$ n-vectors, for $\{a_k\}$ and $\{s_k\}$. In addition, the $2p$ n-vectors $\{x^{-k}\}$ and $\{F(x^{-k})\}$ must be stored. Thus the total extra storage required for the tensor model is at most $4n^{1.5}$ since $p \leq \sqrt{n}$, which is small compared to the n^2 storage required for the Jacobian. The entire process for forming T_c requires $n^2 p + O(np^2)$ multiplications and additions. The leading term comes from calculating the p matrix-vector products $F'(x^c)s_k$, $k = 1, \cdots, p$; the cost of solving $A = ZM^{-1}$ 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, but it is usually only a small multiple of n^2 arithmetic operations since p is usually 1 or 2. This cost also is small compared to the at least $n^3/3$ multiplications and additions per iteration required for the matrix factorizations by standard methods that use analytic or finite difference derivatives.

Solution of the tensor model with the special form of T_c given by (2.3) also can be performed efficiently in terms of algorithmic operations. The goal is to find a root of the tensor model (2.4), that is,

$$\textit{find } d \in \Re^n \textit{ such that}$$
$$M_T(x^c + d) = F(x^c) + F'(x^c)d + \tfrac{1}{2}\sum_{k=1}^{p} a_k (s_k^T d)^2 = 0. \tag{2.5}$$

Since (2.5) may not have a root, it is generalized to solving

$$\min_{d \in \Re^n} \|M_T(x^c + d)\|_2. \tag{2.6}$$

Schnabel and Frank [17] show that the solution of (2.6) can be reduced, in $O(n^2 p)$ operations, to the least squares solution of a system of q quadratic equations in p unknowns, plus the solution of a system of $n - q$ linear equations and $n - p$ unknowns. (Usually, $q = p$; the exceptional case $q > p$ arises when the system of $n - p$ linear equations and $n - p$ unknowns would be singular and generally only occurs when $rank(F'(x^c)) < n - p$.) This reduction is carried out by performing orthogonal transformations of both the variable and equation spaces in a way that isolates the quadratic terms into only p equations. The details of this process are not important to this paper, because here we deal solely with models where $p = 1$, in which case the tensor model can be solved much more simply and in closed form. In addition, the convergence analyses of sections 4 and 5 do not depend upon any particular solution techniques. For these reasons we do not discuss the solution algorithm further in this paper; for details, see [17]. The total cost of solving the tensor model is about $2n^3/3 + n^2 p + O(n^2)$ multiplications and additions, at most $n^2 p \leq n^{2.5}$ multiplications more than the QR factorization of an $n \times n$ matrix. The process generally is numerically stable even if $F'(x^c)$ is singular but has rank $\geq n - p$. If $F'(x^c)$ is nonsingular, the Newton step can be obtained very cheaply as a by-product of the tensor model solution process.

In practice, computational results in [17] show the tensor method is more efficient than an analogous standard method based upon Newton's method on both nonsingular and singular problems, with a particularly large advantage on singular problems. In tests on a standard set of nonsingular test problems, the tensor method is almost always more efficient than the standard method and is never significantly less efficient. The average improvement by the tensor method is $21 - 23\%$, in terms of iterations or function evaluations, on all test problems, and $36-39\%$ on the harder problems where one method requires at least ten iterations. The tensor method is considerably more efficient than the standard method on problems with $rank(F'(x^*)) = n-1$; the average improvement is $40-43\%$ on all problems and $57-61\%$ on the harder problems. The advantage of the tensor method over the standard method on problems with $rank(F'(x^*)) = n-2$ is not as great as for the $rank(F'(x^*)) = n-1$ case but is still considerable, an average of $27-37\%$ improvement on all problems and $57-65\%$ on the harder problems. More recent computational experiments in [1], including experiments on much larger problems, show similar advantages for tensor methods.

3 Notation and assumptions

This section introduces some notation and assumptions that are used throughout the analysis in the remainder of this paper.

We say $f(x) = \Theta(g(x))$, if both $f(x) = O(g(x))$ and $g(x) = O(f(x))$, which is a natural extension to the big-O notation. We also denote $F'(x)$ by $J(x)$ and usually abbreviate $J(x^c)$, $J(x^*)$ as J_c , J_* respectively. Similarly, we often abbreviate $F(x^c)$, $F(x^*)$, $F''(x^c)$, and $F''(x^*)$ as F_c , F_* , F''_c , and F''_* respectively. The notation $\|\cdot\|$ denotes the Euclidean vector norm.

Next, let $F'(x^c) = U_c D_c V_c^T$ be the singular value decomposition of $F'(x)$ at x^c, where $U_c = [u_1^c, u_2^c, \cdots, u_n^c]$, $V_c = [v_1^c, v_2^c, \cdots, v_n^c]$, and $D_c = diag[\sigma_1^c, \sigma_2^c, \cdots, \sigma_n^c]$, with $\sigma_1^c \geq \sigma_2^c \geq \cdots \geq \sigma_n^c \geq 0$ the singular values of $F'(x^c)$ and $\{u_i^c\}$, $\{v_i^c\}$ the corresponding left and right singular vectors. We observe that v_n^c and u_n^c are the right and left singular vectors of $F'(x)$ corresponding to the smallest singular value. Let N^c denote the space spanned by v_n^c and X^c its orthogonal complement, and let P_{N^c}, P_{X^c} be Euclidean projection matrices onto N^c and X^c, respectively. It is easy to see that $P_{N^c} = v_n^c v_n^{cT}$ and $P_{X^c} = I - P_{N^c}$. Furthermore, to simplify notation, we define for any vector $x \in \Re^n$,

$$\begin{aligned} x_{N^c} \in \Re^n &\equiv P_{N^c} x, \\ x_{X^c} \in \Re^n &\equiv P_{X^c} x. \end{aligned}$$

Similarly, let $F'(x^*) = UDV^T$. Let $\sigma_1^* \geq \sigma_2^* \geq \cdots \geq \sigma_{n-1}^* > \sigma_n^* = 0$ be the singular values of $F'(x^*)$. We denote the null spaces of $F'(x^*)$ by N and the orthogonal complement of N by X, and define P_N, P_X as the Euclidean projection matrices onto N and X, respectively. Let v and u be the right and left singular vector of $F'(x^*)$ corresponding the zero singular value, then $P_N = vv^T$ and $P_X = I - P_N$. Again, to simplify notation, we define for any vector $x \in \Re^n$,

$$\begin{aligned} x_N \in \Re^n &\equiv P_N x, \\ x_X \in \Re^n &\equiv P_X x. \end{aligned}$$

The sequence of iterates produced by our algorithms is invariant to the translations in the variable space. Thus no generality is lost by making the assumption that $x^* = 0$, and this assumption is made throughout the paper.

We assume that v_n^c and u_n^c are so chosen that $\|v_n^c - v\| = O(\|x^c\|)$ and $\|u_n^c - u\| = O(\|x^c\|)$, whenever x^c is sufficiently close to x^*. This assumption is valid from the theorems about continuity of eigenvectors in Ortega [14] and Stewart [18], as long as $F'(x)$ is continuous near x^*.

We define the cone $W(\rho, \theta)$ as

$$W(\rho,\theta) = \{x \in \Re^n \mid 0 < \|x\| \leq \rho, \|x_X\| \leq \theta \|x_N\|\}.$$

This is simply a cone rooted at x^* around the null space of $F'(x)$.

We now give Assumption 3.0, a group of assumptions that will be invoked for the remainder of this paper in every lemma and theorem involving $F(x)$. These assumptions basically state that near x^*, the second order term supplies useful information in the null space direction of $F'(x^*)$, where $F'(x^*)$ lacks information.

Assumption 3.0 *Let $F : \Re^n \to \Re^n$ have two Lipschitz continuous derivatives. Let $F(x^*) = 0, F'(x^*)$ be singular with only one zero singular value, and let u and v be the left and right singular vectors of $F'(x^*)$ corresponding to the zero singular vector. Then we assume*

$$u^T F''(x^*)vv \neq 0 \tag{3.1}$$

where $F''(x^*) \in \Re^{n \times n \times n}$.

Assumption 3.0 is satisfied by most problems with $rank(F'(x^*)) = n - 1$, and has been assumed in most papers that analyze the behavior of Newton's method on singular problems. When $n = 1$, Assumption 3.0 is equivalent to $f''(x^*) \neq 0$.

There are several consequences of Assumption 3.0 that we will use in Sections 4 and 5. First, Decker and Kelley [4] show that these assumptions imply that the singular manifold of $F'(x)$ is bounded away from N for x near x^*. That is, for θ and $\|x\|$ small enough, $F'(x)$ is nonsingular for all $x \in W(\rho, \theta)$.

Second, it is clear from (3.1) that there exist $\rho, \theta > 0$, dependent upon the value of $u^T F''(x^*)vv$, such that, if $x \in W(\rho, \theta)$, then

$$|u^T F''(x^*)xv| = \Theta(\|x\|) \tag{3.2}$$

since

$$|u^T F''(x^*)xv| \geq \|x\||cos\theta||u^T F''(x^*)vv| - |u^T F''(x^*)P_X x \cdot v|,$$

$\|P_X x\| = |sin\theta|\|x\|$, and $F''(x^*)$ is bounded. Third,

$$u_n^{cT} J_* v_n^c = O(\|x^c\|^2) \tag{3.3}$$

for x^c sufficiently close to x^*, since

$$u_n^{cT} J_* v_n^c = (u_n^c - u)^T J_* (v_n^c - v).$$

Similarly,

$$\|u_n^{c\,T} J_*\| = O(\|x^c\|). \tag{3.4}$$

Fourth, it is immediate from (3.1) that for x^c sufficiently close to x^*,

$$u_n^{c\,T} F''(x^*) v_n^c v_n^c \quad = \quad \Theta(1) \tag{3.5}$$

and

$$u_n^{c\,T} F''(x^c) v_n^c v_n^c \quad = \quad \Theta(1). \tag{3.6}$$

Finally, for $x^c \in W(\rho, \theta)$,

$$\sigma_n^c = \Theta(\|x^c\|) \tag{3.7}$$

since, using (3.3) and (3.2)

$$\begin{aligned} \sigma_n^c &= u_n^{c\,T} J_c v_n^c \\ &= u_n^{c\,T}(J_c - J_*) v_n^c + u_n^{c\,T} J_* v_n^c \\ &= u_n^{c\,T} F''(x^*) x^c v_n^c + O(\|x^c\|^2) \\ &= u^T F''(x^*) x^c v + O(\|x^c\|^2) \\ &= \Theta(\|x^c\|). \end{aligned} \tag{3.8}$$

These facts will be used throughout in the remainder of this paper.

4 Local convergence analysis of an ideal tensor method

4.1 Introduction

Suppose we know the right singular vector v_n^c corresponding to the least singular value of $F'(x^c)$ where x^c is the current iterate and $\|v_n^c\| = 1$. Then an excellent tensor model around x^c, if one is to utilize just a rank-one second order term, is

$$M_{T_i}(x^c + d) = F(x^c) + F'(x^c)d + \frac{1}{2} a_c (v_n^{cT} d)^2, \tag{4.1}$$

where $a_c = F''(x^c) v_n^c v_n^c$, because it contains the correct second order information in the direction v_n^c where the Jacobian contains the least information. The main convergence result in this section is for Algorithm 4.0, which uses (4.1) as a local model of $F(x)$.

We note that we could approximate $F''(x^c) v_n^c v_n^c$ to sufficient accuracy by finite differences, using one additional evaluation of $F(x)$, that the convergence results of this section would be unchanged. Therefore the "ideal" method analyzed in this section is indeed computationally implementable, although in comparison to Newton's method or the practical tensor method analyzed in the next section, it would require one additional evaluation of $F(x)$, and the calculation of v_n^c, at each iteration. For simplicity, in this section we will assume that $F''(x^c) v_n^c v_n^c$ is known exactly, rather than approximated by finite differences.

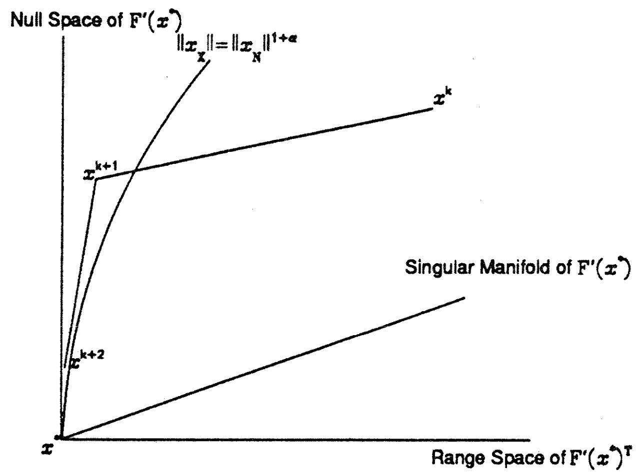

Figure 1: Illustration of Basic Idea of Proof for the Ideal Model

Algorithm 4.0

IF (4.1) has real roots THEN

$d \leftarrow d_R$ where d_R solves $M_{T_i}(x^c + d) = 0$

ELSE $d \leftarrow d_M$ where d_M minimizes $\|M_{T_i}(x^c + d)\|$ □

We will show that, given Assumption 3.0, the sequence of iterates generated by Algorithm 4.0 is locally two-step convergent to x^* with Q-order $\frac{3}{2}$. Figure 1 illustrates the basic idea of the local convergence proof.

Assume that x^k is sufficiently close to x^*, and consider the funnel about the null space $\|x_X\| \leq \Theta(\|x_N\|^{1+\alpha})$ for some $0 < \alpha \leq 1$. If x^k is outside this funnel, it is most likely that x^{k+1} will be inside the funnel, because the error in the subspace X will be reduced quadratically, while the tensor term generally does not provide enough information in this situation to reduce the error in direction N more than linearly. (This is because the error is not near the null space, so that that a second derivative cross term between the null space and its orthogonal complement would be needed for fast convergence.) Now if x^{k+1} is inside the funnel, the step to x^{k+2} will be a fast step, because the tensor model has good information in the null space direction N and the error is basically in this direction. Conversely, if x^{k+1} is outside the funnel, then the step to x^{k+1} must already have been a fast step, since the error reduction in direction N must have (nearly) matched the error reduction in subspace X. The proof implicitly combines these ideas to show that the combination of any two consecutive steps has Q-order at least $\frac{3}{2}$.

4.2 Analysis of an ideal tensor step

We analyze the ideal tensor step by examining separately, in Lemmas 4.1 and 4.2 below, the step to the minimizer of the model 4.1 when it has no root, and the step to the root of 4.1 when there is one. The lemmas show that in either case, the new error in the subspace X, $\|x_X^+\|$, is $O(\|x^c\|^2)$. The lemmas also can be viewed as showing that if $\|x_X^c\|$ is sufficiently small, then the step to the minimizer of the tensor model reduces the overall error quadratically, while a step to a root of the tensor model reduces the error by order $\frac{3}{2}$. This interpretation agrees exactly with the convergence analysis of the solution of a single nonlinear equation in one unknown using a quadratic model if $f'(x^*) = 0$ and $f''(x^*) \neq 0$. As is seen below, the multidimensional result actually is somewhat more complex.

Lemmas 4.1 and 4.2 consider the behavior of the ideal tensor method in the spaces N^c and X^c. Then lemma 4.3 easily transfers the results in subspaces N^c and X^c into the analogous results in the subspaces N and X, the null space of $F'(x^*)$ and its orthogonal complement. This is easier than dealing with the subspaces N and X directly in lemmas 4.1 and 4.2.

Lemma 4.1 *Let Assumption 3.0 hold. If the tensor model has no real roots then for x^c sufficiently close to x^*, the step $d = (x^+ - x^c)$ minimizing $\|M_{T_i}(x^c + d)\|$ gives*

$$\begin{aligned} \|x_{N^c}^+\| &= O(\|x_{X^c}^c\|) + O(\|x^c\|^2) \\ \|x_{X^c}^+\| &= O(\|x^c\|^2). \end{aligned}$$

Proof. Note that $I = \sum_{i=1}^n u_i^c u_i^{cT}$ and $J_c = \sum_{i=1}^n \sigma_i^c u_i^c v_i^{cT}$. From the orthogonality of u_i^c, $i = 1, \cdots, n$, we have

$$\begin{aligned} & M_{T_i}(x^c + d) \\ = \; & (\sum_{i=1}^n u_i^c u_i^{cT})(F_c + \sum_{i=1}^n \sigma_i^c u_i^c v_i^{cT} d + \tfrac{1}{2} a_c (v_n^{cT} d)^2) \\ = \; & \sum_{i=1}^n (\sigma_i^c v_i^{cT} d + u_i^{cT} F_c + \tfrac{1}{2} u_i^{cT} a_c (v_n^{cT} d)^2) u_i^c. \end{aligned} \tag{4.2}$$

Thus minimizing $\|M_{T_i}(x^c + d)\|$ is equivalent to minimizing independently the n separate problems, i.e., minimizing $|\sigma_i^c v_i^{cT} d + u_i^{cT} F_c + \frac{1}{2} u_i^{cT} a_c (v_n^{cT} d)^2|$, for $i = 1, \cdots, n$. For each $1 \leq i \leq n-1$, since $\sigma_i^c > 0$, d can be selected to make $|\sigma_i^c v_i^{cT} d + u_i^{cT} F_c + \frac{1}{2} u_i^{cT} a_c (v_n^{cT} d)^2|$ zero for any given $v_n^{cT} d$. Hence the minimum of $|\sigma_i^c v_i^{cT} d + u_i^{cT} F_c + \frac{1}{2} u_i^{cT} a_c (v_n^{cT} d)^2|$ has to be zero for each $1 \leq i \leq n-1$. Then $M_{T_i}(x^c + d)$ having no real roots implies that $\sigma_n^c v_n^{cT} d + u_n^{cT} F_c + \frac{1}{2} u_n^{cT} a_c (v_n^{cT} d)^2$ has no root, and is minimized at $v_n^{cT} d = -\frac{\sigma_n^c}{u_n^{cT} a_c}$. From Taylor series expansion and (3.3),

$$\begin{aligned} \sigma_n^c &= u_n^{cT} J_c v_n^c \\ &= u_n^{cT} J_* v_n^c + u_n^{cT} F_*'' x^c v_n^c + O(\|x^c\|^2) \\ &= u_n^{cT} F_*'' x^c v_n^c + O(\|x^c\|^2). \end{aligned} \tag{4.3}$$

Also

$$u_n^{c\,T} F_*'' x^c v_n^c \;=\; u_n^{c\,T} F_*'' v_n^c (v_n^{c\,T} x^c) v_n^c + u_n^{c\,T} F_*'' x_{X^c}^c v_n^c \tag{4.4}$$
$$= (v_n^{c\,T} x^c) u_n^{c\,T} F_c'' v_n^c v_n^c + O(\|x^c\|^2) + O(\|x_{X^c}^c\|). \tag{4.5}$$

Combining (4.3) and (4.5) and using $a_c = F''(x^c) v_n^c v_n^c$, and (3.6)

$$\begin{aligned} v_n^{c\,T} d &= -\frac{\sigma_n^c}{u_n^{c\,T} a_c} \\ &= -\frac{(u_n^{c\,T} a_c)(v_n^{c\,T} x^c)}{u_n^{c\,T} a_c} + \frac{O(\|x_{X^c}^c\|) + O(\|x^c\|^2)}{\Theta(1)} \\ &= -v_n^{c\,T} x^c + O(\|x_{X^c}^c\|) + O(\|x^c\|^2) \end{aligned} \tag{4.6}$$

Hence,

$$\begin{aligned} v_n^{c\,T} x^+ &= v_n^{c\,T} (x^c + d) \\ &= O(\|x_{X^c}^c\|) + O(\|x^c\|^2), \end{aligned}$$

which implies $\|x_{N^c}^+\| = O(\|x_{X^c}^c\|) + O(\|x^c\|^2)$.

Next we show that $\|x_{X^c}^+\| = O(\|x^c\|^2)$. Since (4.6) implies $v_n^{c\,T} d = O(\|x^c\|)$, from (4.2) we have that for each $1 \le i \le n-1$, $v_i^{cT} d = -\frac{u_i^{cT} F_c}{\sigma_i^c} + O(\|x^c\|^2)$. Using $F_c = -J_c x^c + V$, where $\|V\| = O(\|x^c\|^2)$, and $\sigma_i^c > \frac{1}{2}\sigma_i^* > 0$,

$$\begin{aligned} v_i^{cT} d &= -\frac{u_i^{cT} F_c}{\sigma_i^c} \\ &= -\frac{u_i^{cT} J_c x^c}{\sigma_i^c} + O(\|x^c\|^2) \\ &= -v_i^{cT} x^c + O(\|x^c\|^2). \end{aligned}$$

Therefore for each $1 \le i \le n-1$,

$$v_i^{cT} x^+ = v_i^{cT} (x^c + d) = O(\|x^c\|^2).$$

Hence $\|P_{X^c} x^+\| = \|\sum_{i=1}^{n-1} (v_i^c v_i^{cT}) x^+\| = O(\|x^c\|^2)$, which completes the proof.

Lemma 4.2 *Let Assumption 3.0 hold. If the tensor model has real roots then for x^c sufficiently close to x^*, the step $d = (x^+ - x^c)$ solving $M_{T_i}(x^c + d) = 0$ gives*

$$\begin{aligned} \|x_{N^c}^+\| &= O(\sqrt{\|x^c\| \|x_{X^c}^c\|}) + O(\|x^c\|^{\frac{3}{2}}) \\ \|x_{X^c}^+\| &= O(\|x^c\|^2). \end{aligned}$$

Proof. Using the same arguments as in the proof of Lemma 4.1, we have

$$\begin{aligned} & M_{T_i}(x^c + d) \\ = & \sum_{i=1}^{n}(\sigma_i^c v_i^{cT} d + u_i^{cT} F_c + \tfrac{1}{2} u_i^{cT} a_c (v_n^{cT} d)^2) u_i^c. \end{aligned}$$

From the orthogonality of u_i^c, $i = 1, \cdots, n$, solving $M_{T_i}(x^c + d) = 0$ is equivalent to solving independently the n separate problems, $\sigma_i^c(v_i^{cT} d) + u_i^{cT} F_c + \frac{1}{2} u_i^{cT} a_c (v_n^{cT} d)^2 = 0$, for $i = 1, \cdots, n$.

Solving

$$u_n^{cT} F_c + \sigma_n^c (v_n^{cT} d) + \tfrac{1}{2}(u_n^{cT} a_c)(v_n^{cT} d)^2 = 0$$

gives

$$v_n^{cT} d = \frac{-\sigma_n^c \pm \sqrt{\sigma_n^{c\,2} - 2(u_n^{cT} F_c)(u_n^{cT} a_c)}}{u_n^{cT} a_c}. \tag{4.7}$$

From (4.3)

$$(\sigma_n^c)^2 = (u_n^{cT} F_*'' x^c v_n^c)^2 + O(\|x^c\|^3). \tag{4.8}$$

Using (4.4) and defining $\bar{a}_c = F_*'' v_n^c v_n^c$,

$$u_n^{cT} F_*'' x^c v_n^c = (v_n^{cT} x^c)(u_n^{cT} \bar{a}_c) + u_n^{cT} F_*'' x_{X^c}^c v_n^c. \tag{4.9}$$

Thus

$$(\sigma_n^c)^2 = (v_n^{cT} x^c)^2 (u_n^{cT} \bar{a}_c)^2 + 2(v_n^{cT} x^c)(u_n^{cT} \bar{a}_c) u_n^{cT} F_*'' x_{X^c}^c v_n^c + O(\|x_{X^c}^c\|^2) + O(\|x^c\|^3) \tag{4.10}$$

Also, from Taylor series expansion, (3.3), (3.4), and $\bar{a}_c = F_*'' v_n^c v_n^c$,

$$\begin{aligned} u_n^{cT} F_c &= u_n^{cT} J_* x^c + \tfrac{1}{2} u_n^{cT} F_*'' x^c x^c + O(\|x^c\|^3) \\ &= u_n^{cT} J_* x_{X^c}^c + (u_n^{cT} J_* v_n^c)(v_n^{cT} x^c) + \tfrac{1}{2}(u_n^{cT} F_*'' v_n^c v_n^c)(v_n^{cT} x^c)^2 \\ &\quad + (u_n^{cT} F_*'' x_{X^c}^c v_n^c)(v_n^{cT} x^c) + \tfrac{1}{2} u_n^{cT} F_*'' x_{X^c}^c x_{X^c}^c \\ &= O(\|x^c\| \|x_{X^c}^c\|) + O(\|x^c\|^3) + \tfrac{1}{2}(u_n^{cT} \bar{a}_c)(v_n^{cT} x^c)^2 \\ &\quad + (u_n^{cT} F_*'' x_{X^c}^c v_n^c)(v_n^{cT} x^c) + O(\|x_{X^c}^c\|^2) \end{aligned} \tag{4.11}$$

Combining (4.10) and (4.11) and using $(u_n^{cT} \bar{a}_c) = \Theta(1)$ from (3.5), $\|\bar{a}_c - a_c\| = O(\|x^c\|)$ from the continuity of $F''(x)$, and $u_n^{cT} F_c = O(\|x^c\|^2)$ from (4.11),

$$\begin{aligned} (\sigma_n^c)^2 - 2(u_n^{cT} F_c)(u_n^{cT} a_c) &= (\sigma_n^c)^2 - 2(u_n^{cT} F_c)(u_n^{cT} \bar{a}_c) + O(\|x^c\|^3) \\ &= O(\|x^c\| \|x_{X^c}^c\|) + O(\|x^c\|^3), \end{aligned}$$

which gives

$$\frac{\sqrt{\sigma_n^{c\,2} - 2u_n^{c\,T} F_c u_n^{c\,T} a_c}}{u_n^{c\,T} a_c} = O(\sqrt{\|x^c\| \|x^c_{X^c}\|}) + O(\|x^c\|^{\frac{3}{2}}). \tag{4.12}$$

Thus, using (4.6) for $\frac{\sigma_n^c}{u_n^{c\,T} a_c}$ and $u_n^{c\,T} a_c = \Theta(1)$ from (3.6) in (4.7),

$$v_n^{c\,T} x^+ = v_n^{c\,T}(x^c + d) = O(\sqrt{\|x^c\| \|x^c_{X^c}\|}) + O(\|x^c\|^{\frac{3}{2}}).$$

The proof of $\|x^+_{X^c}\| = O(\|x^c\|^2)$ is identical as in the proof of Lemma 4.1, since $v_n^{c\,T} d = O(\|x^c\|)$. The proof is complete. □

A subtle point about Lemma 4.2 is that, even though our computational method always chooses the *closer* of the two real roots, Lemma 4.2 is true for *either* real root. This is because the roots of the model, if they exist, are very close together near a singular root.

The following lemma allows us to transfer the above results in terms of the subspaces N^c and X^c into the analogous results in terms of the subspaces N and X.

Lemma 4.3 *If x and x^c are close enough to x^* then*

$$\begin{aligned} \|x_{N^c}\| &= \|x_N\| + O(\|x\| \|x^c\|) \\ \|x_{X^c}\| &= \|x_X\| + O(\|x\| \|x^c\|). \end{aligned}$$

Proof. Since

$$\begin{aligned} \|P_{X^c} - P_X\| &= \|P_{N^c} - P_N\| = \|v_n^c v_n^{c\,T} - vv^T\| \\ &= \|\tfrac{1}{2}(v_n^c + v)(v_n^c - v)^T + \tfrac{1}{2}(v_n^c - v)(v_n^c + v)^T\| \\ &\leq \|v_n^c + v\| \|v_n^c - v\| = O(\|x^c\|), \end{aligned}$$

then

$$\begin{aligned} \|x_{X^c}\| &= \|x_X + (P_{X^c} - P_X)x\| \\ &= \|x_X\| + O(\|x^c\| \|x\|), \end{aligned}$$

and

$$\begin{aligned} \|x^c_N\| &= \|x_N + (P_{N^c} - P_N)x\| \\ &= \|x_N\| + O(\|x^c\| \|x\|). \end{aligned}$$

The proof is complete. □

By applying lemma 4.3, one can easily get the results stated in lemmas 4.1 and 4.2, but with x^c_X, x^+_X, and x^+_N replacing $x^c_{X_c}$, $x^+_{X_c}$, and $x^+_{X_c}$, respectively.

4.3 Local convergence of Algorithm 4.0

In this section the main convergence result for Algorithm 4.0 is proved. Theorem 4.4 shows that the sequence of iterates generated by Algorithm 4.0 converges to x^* from close starting points with a two-step Q-order $\frac{3}{2}$.

Theorem 4.4 *Let Assumption 3.0 hold and $\{x^k\}$ be the sequence of iterates produced by Algorithm 4.0. There exist constants K_1, K_2 such that if $\|x^0\| \leq K_1$, then the sequence $\{x^k\}$ converges to x^* and*

$$\|x^{k+2}\| \quad \leq \quad K_2\|x^k\|^{\frac{3}{2}} \tag{4.13}$$

for $k = 0, 1, 2, \cdots$.

Proof. For any iteration k, by Lemmas 4.1, 4.2 and 4.3, at least

$$\|x_N^{k+1}\| \quad = \quad O(\sqrt{\|x^k\|\|x_X^k\|}) + O(\|x^k\|^{\frac{3}{2}}) \tag{4.14}$$

$$\|x_X^{k+1}\| \quad = \quad O(\|x^k\|^2). \tag{4.15}$$

Combining (4.14) and (4.15) gives

$$\|x^{k+1}\| = O(\|x^k\|). \tag{4.16}$$

Again applying Lemmas 4.1, 4.2 and 4.3 and using (4.15) and (4.16)

$$\begin{aligned}\|x_N^{k+2}\| &= O(\sqrt{\|x^{k+1}\|\|x_X^{k+1}\|}) + O(\|x^{k+1}\|^{\frac{3}{2}}) \\ &= O(\sqrt{\|x^k\|\|x^k\|^2}) + O(\|x^k\|^{\frac{3}{2}}) \\ &= O(\|x^k\|^{\frac{3}{2}}),\end{aligned} \tag{4.17}$$

and

$$\|x_X^{k+2}\| = O(\|x^{k+1}\|^2) = O(\|x^c\|^2). \tag{4.18}$$

Using (4.17) and (4.18)

$$\|x^{k+2}\| = \|x_X^{k+2}\| + \|x_N^{k+2}\| = O(\|x^k\|^{\frac{3}{2}}).$$

The proof is complete. □

Note that Lemma 4.1 remains true even if the ideal tensor model has real roots; the proof and the results are unaffected. Using this fact, it is easy to use the techniques of the proof of Theorem 4.4 to show that if the ideal tensor method always selected the step to the minimizer, whether there was a root or not, then under Assumption 3.0, the method would be two-step quadratically convergent to x^*. It is also easy to see from the proof of Lemma 4.2 that if $F'(x^*)$ is nonsingular, then the tensor model always has real roots near x^*, and that a tensor method that always takes a step to the nearer real root is at least (one-step) quadratically convergent.

5 Local convergence analysis of a practical tensor method

5.1 Introduction

Since each iteration of the ideal tensor method requires the calculation of the singular vector v_n^c corresponding to the least singular value of $F'(x^c)$, and requires an extra evaluation of $F(x)$ to approximate $F''(x^c)v_n^c v_n^c$, Algorithm 4.0 is not as practical as one would like. In this section we analyze a practical tensor algorithm, Algorithm 5.0, that corresponds very closely to the tensor methods we have implemented and tested. Like Algorithm 4.0, it uses a rank-one second order term in the tensor model. However it uses the difference s between x^-, the previous iterate, and x^c, the current iterate, to approximate the null vector of $F'(x^*)$, and it uses the values of $F(x^-)$, $F(x^c)$, and $F'(x^c)$ to approximate the value of $F''(x^c)ss$. That is, the tensor model is

$$M_{T_r}(x^c + d) = F(x^c) + F'(x^c)d + \frac{1}{2}a(s^T d)^2, \tag{5.1}$$

where

$$s = x^- - x^c \tag{5.2}$$

$$a = \frac{2(F(x^-) - F(x^c) - J(x^c)s)}{(s^T s)^2}. \tag{5.3}$$

In this section we will establish a local convergence result for Algorithm 5.0, which uses (5.1) as a local model for $F(x)$. We denote the l_2 norm condition number of a matrix J by $\mathcal{K}(J)$.

Algorithm 5.0

The constants $C_1, C_2 > 0, \frac{1}{3} > \epsilon > 0$, are given.

Let

$$d_N = \begin{cases} -J_c^{-1}F_c & \text{if } \|F_c\| \le \frac{C_1}{\mathcal{K}(J_c)^{1+\epsilon}}, \\ -(J_c^T J_c + D)^{-1} J_c^T F_c & \text{otherwise, where } D = C_2\|F_c\| I. \end{cases}$$

Let

$$d_T = \begin{cases} d_R & \text{if } d_R \text{ solves } M_{T_r}(x^c + d) = 0, \\ d_M & \text{if } d_M \text{ minimizes } \|M_{T_r}(x^c + d)\|. \end{cases}$$

IF $\phi_1\|d_N\| < \|d_T\| < \phi_2\|d_N\|$, for $\phi_1 < \frac{1}{3}$ and $\phi_2 > 3$ being constants THEN

$d \leftarrow d_N$

ELSE

$d \leftarrow d_T$ □

There is one significant, high-level difference between Algorithms 4.0 and 5.0 : Algorithm 5.0 provides the option of using the Newton step at each iteration. This is necessary because, since the tensor model may not contain second order information in a direction near the null vector of $F'(x^*)$, the tensor step is not guaranteed to be good when $F'(x^c)$ is nearly singular. The condition for choosing between the tensor step and the Newton step will be seen to guarantee that the tensor step is selected if both x^c and x^- are inside a funnel around the null space of $F'(x^*)$, but will allow either the tensor step or the Newton step to be selected otherwise. It will also guarantee that whenever the tensor step is taken, it has the desirable properties mentioned in the discussion of Figure 2 below.

The perturbation that Algorithm 5.0 makes to the Newton step when $\mathcal{K}(J)$ is too large is a typical type of modifications used in standard methods (see e.g. [7]). The particular choice of when to perturb the Newton step used in Algorithm 5.0 guarantees that the unperturbed Newton step always is selected if x^c is close to the null space of $F'(x^*)$, and the amount of the perturbation guarantees that the linear model step step causes the next iterate to be close to the null space otherwise. We note that our convergence results could be proven using more general conditions on the perturbation matrix D, but we will just consider the simple choice in Algorithm 5.0 because it is what we would use in practice.

Algorithm 5.0 is very close to the tensor method we have previously implemented, as long as a rank-one tensor always is selected. The only differences are the specific rules for choosing between the tensor and the Newton steps, and for deciding when to perturb the Newton step; we have used slightly different rules that have very similar effects. Guided by this analysis, we expect to use the rules from Algorithm 5.0 in the future. The only modification we would make would be to approximate $\mathcal{K}(J)$ by well known efficient methods (see e.g. [8]) rather than calculating it exactly. A reasonable value for the constants C_1 and C_2 would be one.

In the remainder of this section we will show that, given Assumption 3.0, the sequence of iterates generated by Algorithm 5.0 is locally three-step convergent to x^* with Q-order $\frac{3}{2}$. Figure 2 illustrates the basic idea of the local convergence proof. Assume that x^k is sufficiently close to x^*, and consider again the funnel about the null space $\|x_X\| \leq \Theta(\|x_N\|^{1+\alpha})$ for some $0 < \alpha \leq 1$. If x^k is outside this funnel, either a Newton step or a tensor step can be taken from x^k to x^{k+1}. In either case the step either puts x^{k+1} inside the funnel (the likely case), or it has to be a fast step because either step reduces the error in the subspace X quadratically, so that the overall error has to be reduced significantly in order to stay outside the funnel. Now suppose x^{k+1} is inside the funnel. Again, either a Newton step or a tensor step can be taken to x^{k+2}. Furthermore, either x^{k+2} is inside the funnel, in which case the step selection rules guarantee that the step is long enough that it is nearly parallel to the null vector of $F'(x^*)$ (this is the likely case), or the step must be a fast step. Finally, suppose both x^{k+1} and x^{k+2} are inside the funnel and the step from x^{k+1} to x^{k+2} is near the null vector. Then the tensor model at x^{k+2} has good second order information in the null space direction. In this case Algorithm 5.0 selects a tensor model based step to x^{k+3} and the step is fast. The proof again implicitly combines these ideas to obtain desired convergence result.

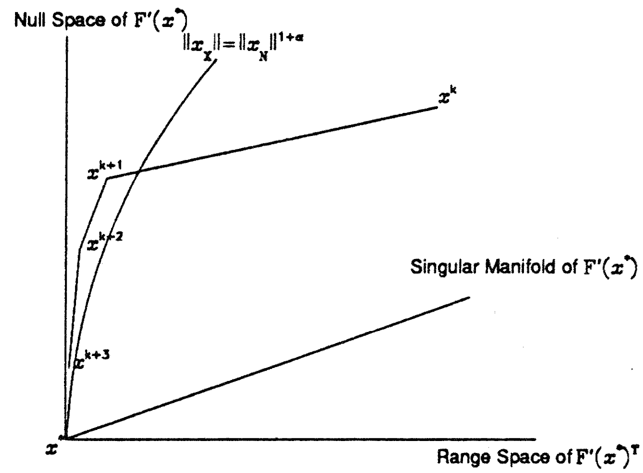

Figure 2: Illustration of Basic Idea of Proof for the Practical Model

Note that the practical method obtains a three step result, versus the two-step result for the ideal method. This is because once an iterate (x^{k+1} in Figures 1 and 2) is in the funnel, the ideal method immediately takes a fast step, whereas at least in theory, the practical method may require an additional step to make the previous step direction nearly parallel to the null vector before it takes a fast step.

For the sake of convenience, we rewrite the tensor model (5.1) in an equivalent form. From Taylor expansion,

$$F(x^-) \;=\; F(x^c + s) = F_c + J_c s + \frac{1}{2}F_c''ss + \Delta_1$$

where $||\Delta_1|| = O(||s||^3)$. Substituting this into (5.3),

$$a_c' \;=\; a||s||^2 = F_c''pp + \Delta_2 \tag{5.4}$$

where $p = \frac{s}{||s||}$ and $||\Delta_2|| = O(||s||)$. (5.1) can be rewritten as

$$M_{T_r}(x^c + d) \;=\; F_c + F_c'd + \frac{1}{2}a_c'(p^T d)^2, \tag{5.5}$$

which has a form similar to (4.1).

Note that when the angle γ between p and v_n^c is $O(||x^c||^\alpha)$,

$$\begin{aligned} ||p_{X^c}||^2 &= ||(I - v_n^c v_n^{cT})p||^2 \\ &= 1 - (p^T v_n^c)^2 = 1 - cos^2\gamma = sin^2\gamma = O(||x^c||^{2\alpha}), \end{aligned} \tag{5.6}$$

or $\|p_{X^c}\| = O(\|x^c\|^\alpha)$. Furthermore, if we define

$$\varphi = \begin{cases} 1 & \text{if } \|x^c\| > \|x^-\|, \\ \varphi \geq 1 & \text{for which } \|x^c\| = \|x^-\|^\varphi \text{ otherwise,} \end{cases} \tag{5.7}$$

so that $\|s\| = O(\|x^c\|^{\frac{1}{\varphi}})$, then using also $p = (p^T v_n^c) v_n^c + p_{X^c}$, (5.4) and (5.6) we have

$$\begin{aligned} \|a_c' - a_c\| &= \|F_c'' pp + \Delta_2 - F_c'' v_n^c v_n^c\| \\ &\leq \|F_c'' pp - F_c'' v_n^c v_n^c\| + O(\|s\|) \\ &= \|((p^T v_n^c)^2 - 1) F_c'' v_n^c v_n^c + 2 p^T v_n^c F_c'' p_{X^c} v_n^c + F_c'' p_{X^c} p_{X^c}\| + O(\|x^c\|^{\frac{1}{\varphi}}) \\ &\leq O(|sin^2 \gamma|) + O(\|p_{X^c}\|) + O(\|p_{X^c}\|^2) + O(\|x^c\|^{\frac{1}{\varphi}}) \\ &= O(\|x^c\|^{min(\alpha, \frac{1}{\varphi})}). \end{aligned} \tag{5.8}$$

It is also easy to verify that $u_n^{c\,T} a_c' = \Theta(1)$ from (5.8) and $u_n^{c\,T} a_c = \Theta(1)$. These facts will be used in the proofs in this section.

The remainder of this section is organized as follows. Subsection 5.2 analyzes the Newton step defined in Algorithm 5.0. Subsection 5.3 analyzes the practical tensor step. Finally, the overall convergence theorem for Algorithm 5.0 is stated and proved in subsection 5.4.

5.2 Analysis of a Newton step

The following lemma analyzes the linear model step d_N defined in Algorithm 5.0. This step is either the Newton step if $\|F_c\| \leq \frac{C_1}{\mathcal{K}(J_c)^{1+\epsilon}}$ (which will be shown to imply that $\|x_{X^c}^c\| \leq O(\|x^c\|^{1+\epsilon})$), or the perturbed Newton step otherwise. The lemma states that in either case, the linear model step reduces the error in the subspace X^c quadratically. In addition, when the unperturbed Newton step is selected it asymptotically reduces the error in the null space direction by $\frac{1}{2}$, otherwise the perturbed Newton step keeps the error in the null space in the same order of magnitude.

Lemma 5.1 *Let Assumption 3.0 hold and let $x^+ = x^c + d_N$, where d_N is defined by Algorithm 5.0. At the current iteration x^c, if $\|F_c\| \leq \frac{C_1}{\mathcal{K}(J_c)^{1+\epsilon}}$, then*

$$\|x_{N^c}^+\| = \frac{1}{2}\|x_{N^c}^c\| + O(\|x^c\|^{1+\epsilon}) \tag{5.9}$$

$$\|x_{X^c}^+\| = O(\|x^c\|^2), \tag{5.10}$$

otherwise,

$$\|x_{N^c}^+\| = O(\|x^c\|) \tag{5.11}$$

$$\|x_{X^c}^+\| = O(\|x^c\|^2). \tag{5.12}$$

Proof. From Taylor series expansion, $J_c = \sum_{i=1}^{n} \sigma_i^c u_i^c v_i^{cT}$, and $\sigma_n^c = O(\|x^c\|)$,

$$\begin{aligned} \|F_c\| &= \|F_* + J_c x^c\| + O(\|x^c\|^2) \\ &= \sqrt{\sum_{i=1}^{n} [\sigma_i^c (v_i^{cT} x^c)]^2 + O(\|x^c\|^2)} \\ &\geq \sigma_{n-1}^c \sqrt{\sum_{i=1}^{n-1} (v_i^{cT} x^c)^2 + O(\|x^c\|^2)} \\ &= \sigma_{n-1}^c \|x_{X^c}^c\| + O(\|x^c\|^2). \end{aligned} \tag{5.13}$$

Clearly,

$$\frac{1}{\mathcal{K}(J_c)} = \frac{\sigma_n^c}{\sigma_1^c}. \tag{5.14}$$

First, $\|F_c\| \leq \frac{C_1}{\mathcal{K}(J_c)^{1+\epsilon}}$ implies

$$\sigma_{n-1}^c \|x_{X^c}^c\| + O(\|x^c\|^2) \leq \left(\frac{\sigma_n^c}{\sigma_1^c}\right)^{1+\epsilon}$$

or

$$\|x_{X^c}^c\| \leq O(\|x^c\|^{1+\epsilon})$$

since $\sigma_n^c = O(\|x^c\|)$ and σ_{n-1}^c and σ_1^c are bounded below. This implies x^c is inside the funnel $\|x_{X^c}^c\| = \Theta(\|x^c\|^{1+\epsilon})$, hence it is in $W(\rho, \theta)$. Therefore, σ_n^c satisfies $\sigma_n^c = \Theta(\|x^c\|)$ from (3.8).

Note also that if $\|x_{X^c}^c\| < \|x^c\|^{1+\bar{\epsilon}}$ for any $1 > \bar{\epsilon} > \epsilon$, then $\|F_c\| \leq \frac{C_1}{\mathcal{K}(J_c)^{1+\epsilon}}$, since in this case $\sigma_n^c = \Theta(\|x^c\|)$ from (3.8) so that

$$\begin{aligned} \|F_c\| &\leq \sigma_1^c \|x_{X^c}^c\| + O(\|x^c\|^2) \\ &\leq 2\sigma_1^c \|x^c\|^{1+\bar{\epsilon}} \\ &< \frac{C_1}{(\sigma_1^c)^{1+\epsilon}} (\sigma_n^c)^{1+\epsilon} \\ &= \frac{C_1}{\mathcal{K}(J_c)^{1+\epsilon}} \end{aligned}$$

for x^c sufficiently small. This fact will be used in the final portion of this proof and in the proof of Theorem 5.5.

By the definition of the Newton step, $v_n^{cT} d = -v_n^{cT} J_c^{-1} F_c = -\frac{u_n^{cT} F_c}{\sigma_n^c}$. From Taylor series expansion,

$$\begin{aligned} u_n^{cT} F_c &= u_n^{cT} (F_* + J_c x^c - \tfrac{1}{2} F_c'' x^c x^c) + O(\|x^c\|^3) \\ &= \sigma_n^c (v_n^{cT} x^c) - \tfrac{1}{2} u_n^{cT} F_c'' x^c x^c + O(\|x^c\|^3). \end{aligned} \tag{5.15}$$

From the continuity of F'',

$$
\begin{aligned}
u_n^{cT} F_c'' x^c x^c &= u_n^{cT} F_*'' x^c x^c + O(\|x^c\|^3) \\
&= (v_n^{cT} x^c)^2 u_n^{cT} F_*'' v_n^c v_n^c + 2(v_n^{cT} x^c) u_n^{cT} F_*'' x_{X^c}^c v_n^c \\
&\quad + u_n^{cT} F_*'' x_{X^c}^c x_{X^c}^c + O(\|x^c\|^3) \\
&= (v_n^{cT} x^c)^2 u_n^{cT} F_*'' v_n^c v_n^c + O(\|x^c\|^{2+\epsilon}) \qquad (5.16)
\end{aligned}
$$

since $\|x_{X^c}^c\| = O(\|x^c\|^{1+\epsilon})$. Furthermore,

$$
\begin{aligned}
\sigma_n^c &= u_n^{cT} J_c v_n^c \\
&= u_n^{cT} J_* v_n^c + u_n^{cT} F_*'' x^c v_n^c + O(\|x^c\|^2) \\
&= (v_n^{cT} x^c) u_n^{cT} F_*'' v_n^c v_n^c + u_n^{cT} F_*'' x_{X^c}^c v_n^c + O(\|x^c\|^2) \\
&= (v_n^{cT} x^c) u_n^{cT} F_*'' v_n^c v_n^c + O(\|x^c\|^{1+\epsilon}). \qquad (5.17)
\end{aligned}
$$

Hence from (5.15), (5.16) and (5.17), and using $v_n^{cT} x^c = \Theta(\|x^c\|)$ and (3.6),

$$
\begin{aligned}
v_n^{cT} d &= -\frac{u_n^{cT} F_c}{\sigma_n^c} \\
&= -v_n^{cT} x^c + \frac{\frac{1}{2} u_n^{cT} F_c'' x^c x^c + O(\|x^c\|^3)}{\sigma_n^c} \\
&= -v_n^{cT} x^c + \frac{\frac{1}{2}(v_n^{cT} x^c)^2 u_n^{cT} F_*'' v_n^c v_n^c + O(\|x^c\|^{2+\epsilon})}{(v_n^{cT} x^c) u_n^{cT} F_*'' v_n^c v_n^c + O(\|x^c\|^{1+\epsilon})} \\
&= -v_n^{cT} x^c + \tfrac{1}{2}(v_n^{cT} x^c) + O(\|x^c\|^{1+\epsilon}),
\end{aligned}
$$

which implies

$$
v_n^{cT} x^+ = \tfrac{1}{2}(v_n^{cT} x^c) + O(\|x^c\|^{1+\epsilon})
$$

or

$$
\|x_{N^c}^+\| = \tfrac{1}{2}\|x_{N^c}^c\| + O(\|x^c\|^{1+\epsilon}).
$$

Also, by the definition of the Newton step, $v_i^{cT} d = -v_i^{cT} J_c^{-1} F_c = -\frac{u_i^{cT} F_c}{\sigma_i^c}$, $i = 1, \cdots, n-1$. For each $1 \le i \le n-1$, from Taylor series expansion, and $\sigma_i^c = \Theta(1)$,

$$
\begin{aligned}
v_i^{cT} d &= -\frac{u_i^{cT} F_c}{\sigma_i^c} \\
&= -\frac{u_i^{cT} J_c x^c}{\sigma_i^c} + O(\|x^c\|^2) \\
&= -v_i^{cT} x^c + O(\|x^c\|^2),
\end{aligned}
$$

which implies

$$\begin{aligned}
\|x^+_{X^c}\| &= \|(\sum_{i=1}^{n-1} v_i^c v_i^{cT})(x^c + d)\| \\
&= \|(\sum_{i=1}^{n-1} v_i^c)(v_i^{cT} x^c + v_i^{cT} d)\| \\
&= O(\|x^c\|^2).
\end{aligned}$$

Second, when $\|F_c\| > \frac{C_1}{\mathcal{K}(J_c)^{1+\epsilon}}$,

$$\begin{aligned}
d &= -(J_c^T J_c + C_2\|F_c\| I)^{-1} J_c^T F_c \\
&= -\sum_{i=1}^{n} \frac{\sigma_i^c (u_i^{cT} F_c)}{(\sigma_i^c)^2 + C_2\|F_c\|} v_i^c.
\end{aligned}$$

Then for $1 \leq i \leq n-1$,

$$v_i^{cT} d = -(u_i^{cT} F_c) \frac{\sigma_i^c}{(\sigma_i^c)^2 + C_2\|F_c\|},$$

and using $u_i^{cT} F_c = \sigma_i^c (v_i^{cT} x^c) + O(\|x^c\|^2)$ as was shown in the preceding part of this proof, $\sigma_i^c = \Theta(1)$, and $\|F_c\| = O(\|x^c\|)$, we have

$$\begin{aligned}
v_i^{cT} d &= -(\sigma_i^c (v_i^{cT} x^c) + O(\|x^c\|^2))(\frac{1}{\sigma_i^c} + O(\|x^c\|)) \\
&= -v_i^{cT} x^c + O(\|x^c\|^2).
\end{aligned}$$

As above, this establishes $\|x^+_{X^c}\| = O(\|x^c\|^2)$. Finally, it was shown earlier in this proof that in this case $\|x^c_{X^c}\| \geq \|x^c\|^{1+\bar{\epsilon}}$ for any $1 > \bar{\epsilon} > \epsilon$, so that $\|x^c_{X^c}\| \geq \|x^c\|^{1+2\epsilon}$. Thus from (5.13),

$$\|F_c\| \geq \sigma_{n-1}^c \|x^c\|^{1+2\epsilon}. \qquad (5.18)$$

Now using Taylor series expansion, $\sigma_n^c = O(\|x^c\|)$ and (5.18),

$$\begin{aligned}
v_n^{cT} d &= -\frac{\sigma_n^c}{\sigma_n^{c\,2} + C_2\|F_c\|}(\sigma_n^c (v_n^{cT} x^c) - \tfrac{1}{2} u_n^{cT} F_*'' x^c x^c + O(\|x^c\|^3)) \\
&= O(\|x^c\|^{2-2\epsilon}),
\end{aligned}$$

which implies

$$\begin{aligned}
v_n^{cT} x^+ &= v_n^{cT} x^c + v_n^{cT} d \\
&= v_n^{cT} x^c + O(\|x^c\|^{2-2\epsilon}) \\
&= O(\|x^c\|).
\end{aligned}$$

Hence $\|x^+_{N^c}\| = O(\|x^c\|)$. The proof is complete. □

Using Lemma 4.3, we can replace N^c and X^c in Lemma 5.1 by N and X, respectively.

5.3 Analysis of a practical tensor step

When the current iterate is inside a funnel around the null space of $F'(x^*)$ and the previous step is along the null space, the tensor model (5.1) has good second order information, which can be expected to compensate for the lack of useful first order derivative information in the null space due to the singularity of the $F'(x^*)$. In this situation, we can show that a tensor step will reduce the error in the null space direction N substantially while reducing the error in the orthogonal subspace X quadratically. This is shown in Lemmas 5.2 and 5.3 below, for the cases when the tensor model does not and does have a root, respectively. Finally Lemma 5.4 shows that in any situation where Algorithm 5.0 selects the tensor step, this step reduces the error in the subspace X quadratically and at least keeps the total error at the same order of magnitude.

First we consider the case when tensor model has no real roots at x^c. The key term $min(\alpha, \delta, \frac{1}{\varphi})$ in the statements of both Lemmas 5.2 and 5.3 captures the fact that either the previous step was fast ($\varphi > 1$ and/or $\delta < 1$) or else the current step will be fast because α will be significantly greater than one. The way these parts come together to prove the three-step convergence result will be seen in the proof of Theorem 5.5.

Lemma 5.2 *Let Assumption 3.0 hold and let φ be defined by (5.7). Assume that the tensor model (5.1) has no real roots at x^c, the current iterate. If the angle between p and v_n^c is $O(\|x^c\|^\alpha)$ and $\|x^c_{X^c}\| = O(\|x^c\|^{1+\delta})$ for $\alpha, \delta > 0$, then for x^c sufficiently close to x^*, the step $d = (x^+ - x^c)$ minimizing (5.1) in the least squares sense satisfies*

$$\begin{aligned} \|x^+_{N^c}\| &= O(\|x^c\|^{1+min(\alpha,\delta,\frac{1}{\varphi})}) \\ \|x^+_{X^c}\| &= O(\|x^c\|^2). \end{aligned}$$

Proof. First we show $p^T d = O(\|x^c\|)$. If the tensor model has no real roots, then the step d is the minimizer of $\|M_{T_r}(x^c + d)\|$. Taking the gradient of $\frac{1}{2}M_{T_r}(x^c + d)^T M_{T_r}(x^c + d)$, we find that at a minimizer of $\|M_{T_r}(x^c + d)\|$,

$$(F'(x^c) + (p^T d)a'_c p^T)^T M_{T_r}(x^c + d) = 0,$$

which implies that the matrix $F'(x^c) + (p^T d)a'_c p^T$ must be singular since $M_{T_r}(x^c + d) \neq 0$. From the Sherman-Morrison-Woodbury formula and the nonsingularity of $F'(x^c)$ for $x^c \in W(\rho, \theta)$, we conclude that $F'(x^c) + (p^T d)a'_c p^T$ is singular if and only if $p^T d = \frac{-1}{p^T F'(x^c)^{-1} a'_c}$. However,

$$\begin{aligned} \frac{-1}{p^T J_c^{-1} a'_c} &= \frac{-1}{p^T(\sum_{i=1}^n \frac{1}{\sigma_i^c} v_i^c u_i^{cT})a'_c} \\ &= O(\frac{\sigma_n^c}{u_n^{cT} a'_c}) \\ &= O(\|x^c\|), \end{aligned}$$

since $p^T v_i^c = O(\|x^c\|^\alpha)$ for $i = 1, \cdots, n-1$ from (5.6), $\sigma_i^c = \Theta(1)$ for $i = 1, \cdots, n-1$, $\sigma_n^c = \Theta(\|x^c\|)$, and $u_n^{cT} a'_c = \Theta(1)$. Hence

$$p^T d = O(\|x^c\|). \tag{5.19}$$

From the same reasoning as in the proof of Lemma 4.1, one can easily obtain

$$\begin{aligned} & M_{T_r}(x^c + d) \\ = & \sum_{i=1}^{n}(\sigma_i^c v_i^{cT} d + u_i^{cT} F_c + \tfrac{1}{2} u_i^{cT} a_c'(p^T d)^2) u_i^c. \end{aligned}$$

Then minimizing $\|M_{T_r}(x^c + d)\|$ is equivalent to minimizing independently the n separate quantities,

$$\sigma_i^c v_i^{cT} d + u_i^{cT} F_c + \tfrac{1}{2} u_i^{cT} a_c'(p^T d)^2 \tag{5.20}$$

for $i = 1, \cdots, n$.

First we show that (5.20) has a real root for each $i = 1, \cdots, n-1$. We know that the step d must satisfy

$$\|M_{T_r}(x^c + d)\| \leq \|M_{T_r}(x^c)\| = \|F_c\| = O(\|x^c\|^{1+\delta}). \tag{5.21}$$

Observe that from the Taylor series expansion,

$$\begin{aligned} u_i^{cT} F_c &= -u_i^{cT} J_c x^c + O(\|x^c\|^2) \\ &= -\sigma_i^c v_i^{cT} x^c + O(\|x^c\|^2) \\ &= O(\|x^c\|^{1+\delta}), \end{aligned}$$

for $1 \leq i \leq n-1$, since $\|x^c_{X^c}\| = O(\|x^c\|^{1+\delta})$. Then from (5.20), using $\sigma_i^c = \Theta(1)$ for $1 \leq i \leq n-1$, $u_n^{cT} a_c' = \Theta(1)$, and (5.19), (5.21) implies that

$$v_i^{cT} d = O(\|x^c\|^{1+\delta}), \quad i = 1, \cdots, n-1. \tag{5.22}$$

since any larger value would violate (5.21). Since

$$p^T d = \sum_{i=1}^{n-1}(p^T v_i^c)(v_i^{cT} d) + (p^T v_n^c)(v_n^{cT} d), \tag{5.23}$$

from (5.19), (5.22) and $p^T v_i^c = O(\|x^c\|^{\alpha})$ and $p^T v_n^c = \Theta(1)$ from (5.6), we have

$$v_n^{cT} d = O(\|x^c\|). \tag{5.24}$$

For each $i = 1, \cdots, n-1$, using (5.23), (5.24), (5.6) and the bounds on σ_i^c, $u_i^{cT} F_c$, and $u_i^{cT} a_c'$ mentioned above, (5.20) becomes

$$\begin{aligned} & \sigma_i^c v_i^{cT} d + u_i^{cT} F_c + \tfrac{1}{2} u_i^{cT} a_c'((p^T v_i^c)(v_i^{cT} d) + O(\|x^c\|))^2 \\ = & (\sigma_i^c + O(\|x^c\|^{1+\alpha})) v_i^{cT} d + u_i^{cT} F_c + O(\|x^c\|^2) + O(\|x^c\|^{2\alpha})(v_i^{cT} d)^2, \end{aligned}$$

and it is immediate that the discriminant of this quadratic in ${v_i^c}^T d$ is greater than zero. Hence (5.20) has a real root for each $i = 1, \cdots, n-1$. By this fact and (5.22), for $i = 1, \cdots, n-1$ the root d to (5.20) must satisfy

$$\begin{aligned} & (\sigma_i^c + O(\|x^c\|^{1+\alpha})){v_i^c}^T d + {u_i^c}^T F_c + O(\|x^c\|^2) + O(\|x^c\|^{2\alpha})({v_i^c}^T d)^2 \\ = \ & \sigma_i^c {v_i^c}^T d + {u_i^c}^T F_c + O(\|x^c\|^2) = 0, \end{aligned}$$

or

$$ {v_i^c}^T d = -\frac{{u_i^c}^T F_c}{\sigma_i^c} + O(\|x^c\|^2) \tag{5.25}$$

since σ_i^c is bounded below. From the same analysis as in Lemma 5.1 for $\|x_{X^c}^+\|$, we can obtain $\|x_{X^c}^+\| = O(\|x^c\|^2)$ from (5.25).

As a consequence of above analysis, $M_{T_r}(x^c + d)$ has no real roots is equivalent to

$$ \sigma_n^c {v_n^c}^T d + {u_n^c}^T F_c + \tfrac{1}{2} {u_n^c}^T a_c' (p^T d)^2 \tag{5.26}$$

has no real root. From (5.6) and (5.22),

$$ p^T d \ = \ O(\|x^c\|^{1+\alpha}) + ({v_n^c}^T p)({v_n^c}^T d). $$

Hence (5.26) is equivalent to

$$ (\sigma_n^c + O(\|x^c\|^{1+\alpha})){v_n^c}^T d + {u_n^c}^T F_c + O(\|x^c\|^{2(1+\alpha)}) + \tfrac{1}{2}({u_n^c}^T a_c')({v_n^c}^T p)^2({v_n^c}^T d)^2. $$

Thus the minimizer of (5.26) obeys

$$ (\sigma_n^c + O(\|x^c\|^{1+\alpha})) + ({u_n^c}^T a_c')({v_n^c}^T p)^2({v_n^c}^T d) = 0 $$

or

$$\begin{aligned} {v_n^c}^T d &= -\frac{\sigma_n^c + O(\|x^c\|^{1+\alpha})}{{u_n^c}^T a_c' ({v_n^c}^T p)^2} \\ &= -\frac{\sigma_n^c + O(\|x^c\|^{1+\alpha})}{{u_n^c}^T a_c + O(\|x^c\|^{min(\alpha, \frac{1}{\varphi})})} \\ &= -\frac{\sigma_n^c}{{u_n^c}^T a_c} + O(\|x\|^{1+min(\alpha,\frac{1}{\varphi})}) \\ &= -{v_n^c}^T x^c + O(\|x_{X^c}^c\|) + O(\|x\|^{1+min(\alpha,\frac{1}{\varphi})}) \\ &= -{v_n^c}^T x^c + O(\|x^c\|^{1+\delta}) + O(\|x\|^{1+min(\alpha,\frac{1}{\varphi})}) \\ &= -{v_n^c}^T x^c + O(\|x^c\|^{1+min(\alpha,\delta,\frac{1}{\varphi})}), \end{aligned}$$

from $\frac{\sigma_n^c}{{u_n^c}^T a_c} = -{v_n^c}^T x^c + O(\|x_{X^c}^c\|) + O(\|x^c\|^2)$ in the proof of Lemma 4.1 and $\|x_{X^c}^c\| = O(\|x^c\|^{1+\delta})$, which gives

$$ {v_n^c}^T (x^c + d) \ = \ O(\|x^c\|^{1+min(\alpha,\delta,\frac{1}{\varphi})}). $$

Hence, $\|x^+_{N^c}\| = O(\|x^c\|^{1+min(\alpha,\delta,\frac{1}{\varphi})})$. The proof is complete. □

Using Lemma 4.3, we can change the results in Lemma 5.2 to $\|x^+_X\| = O(\|x^c\|^2)$ and $\|x^+_N\| = O(\|x^c\|^{1+min(\alpha,\delta,\frac{1}{\varphi})})$.

Secondly we consider the case when tensor model has real roots at x^c.

Lemma 5.3 *Let Assumption 3.0 hold and let φ be defined by (5.7). Assume that the tensor model (5.1) has real roots at x^c, the current iterate. If the angle between p and v^c_n is $O(\|x^c\|^\alpha)$ and $\|x^c_{X^c}\| = O(\|x^c\|^{1+\delta})$ for $\alpha, \delta > 0$, then for x^c sufficiently close to x^*, the step $d = (x^+ - x^c)$ solving model (5.1) satisfies*

$$\begin{aligned} \|x^+_{N^c}\| &= O(\|x^c\|^{1+\frac{1}{2}min(\alpha,\delta,\frac{1}{\varphi})}) \\ \|x^+_{X^c}\| &= O(\|x^c\|^2). \end{aligned}$$

Proof. First we show $p^T d = O(\|x^c\|)$. Since $M_{T_r}(x^c + d) = 0$ and J_c is nonsingular,

$$F_c + J_c d + \tfrac{1}{2}a'_c(p^T d)^2 = 0$$

implies

$$p^T J_c^{-1} F_c + p^T d + \tfrac{1}{2}p^T J_c^{-1} a'_c (p^T d)^2 = 0,$$

which in turn implies

$$p^T d = \frac{-1 \pm \sqrt{1 - 2p^T J_c^{-1} a'_c p^T J_c^{-1} F_c}}{p^T J_c^{-1} a'_c}.$$

However, using $\sigma^c_n = \Theta(\|x^c\|)$, $\sigma^c_i = \Theta(1)$ for $i = 1, \cdots, n-1$, $p^T v^c_n = \Theta(1)$ and $u^{c\,T}_n a'_c = \Theta(1)$,

$$\begin{aligned} p^T J_c^{-1} F_c &= \sum_{i=1}^{n} \frac{(p^T v^c_i)(u^{c\,T}_i F_c)}{\sigma^c_i} \\ &= \sum_{i=1}^{n} \frac{(p^T v^c_i)(u^{c\,T}_i F_* + u^{c\,T}_i J_c x^c + O(\|x^c\|^2))}{\sigma^c_i} \\ &= \sum_{i=1}^{n-1} (p^T v^c_i)(v^{c\,T}_i x^c) + O(\|x^c\|^2) + (p^T v^c_n)(v^{c\,T}_n x^c) + O(\|x^c\|) \\ &= O(\|x^c\|), \end{aligned}$$

and

$$\begin{aligned} p^T J_c^{-1} a'_c &= \sum_{i=1}^{n} \frac{(p^T v^c_i)(u^{c\,T}_i a'_c)}{\sigma^c_i} \\ &= \left(\sum_{i=1}^{n-1} \frac{(p^T v^c_i)(u^{c\,T}_i a'_c)}{\sigma^c_i}\right) + \frac{(p^T v^c_n)(u^{c\,T}_n a'_c)}{\sigma^c_n} \\ &= \Theta(\|x^c\|^{-1}). \end{aligned}$$

Thus

$$\begin{aligned} p^T d &= \frac{-1 \pm \sqrt{1 - 2O(\|x^c\|)\Theta(\|x^c\|^{-1})}}{\Theta(\|x^c\|^{-1})} \\ &= O(\|x^c\|). \end{aligned}$$

Using the same arguments as stated in the proof of Lemma 5.2,

$$\begin{aligned} & M_{T_r}(x^c + d) \\ = & \sum_{i=1}^{n} (\sigma_i^c v_i^{cT} d + u_i^{cT} F_c + \tfrac{1}{2} u_i^{cT} a_c' (p^T d)^2) u_i^c. \end{aligned}$$

Then solving $M_{T_r}(x^c + d) = 0$ is equivalent to solving independently the n separate equations,

$$\begin{aligned} & \sigma_i^c v_i^{cT} d + u_i^{cT} F_c + \tfrac{1}{2} u_i^{cT} a_c' (p^T d)^2 \\ = & \sigma_i^c v_i^{cT} d + u_i^{cT} F_c + O(\|x^c\|^2) = 0 \end{aligned}$$

for $i = 1, \cdots, n$. By the identical reasoning as in the proof of Lemma 5.2 for $\|x_{X^c}^+\|$, we have $\|x_{X^c}^+\| = O(\|x^c\|^2)$ immediately.

On the other hand, for $i = n$

$$\sigma_n^c v_n^{cT} d + u_n^{cT} F_c + \tfrac{1}{2} u_n^{cT} a_c' (p^T d)^2 = 0.$$

From the same arguments as in the proof of Lemma 5.2

$$\begin{aligned} & \sigma_n^c v_n^{cT} d + u_n^{cT} F_c + \tfrac{1}{2} u_n^{cT} a_c' (p^T d)^2 \\ = & u_n^{cT} F_c + O(\|x^c\|^{2(1+\alpha)}) + (\sigma_n^c + O(\|x^c\|^{1+\alpha})) v_n^{cT} d \\ & + \tfrac{1}{2} u_n^{cT} a_c' (v_n^{cT} p)^2 (v_n^{cT} d)^2, \end{aligned}$$

which is solved with

$$\begin{aligned} v_n^{cT} d &= \frac{-\sigma_n^c + O(\|x^c\|^{1+\alpha})}{(v_n^{cT} p)^2 u_n^{cT} a_c'} \\ & \quad \frac{\pm\sqrt{(\sigma_n^c + O(\|x^c\|^{1+\alpha}))^2 - 2(u_n^{cT} F_c + O(\|x^c\|^{2(1+\alpha)}))(v_n^{cT} p)^2 (u_n^{cT} a_c')}}{(v_n^{cT} p)^2 u_n^{cT} a_c'} \\ &= \frac{-\sigma_n^c + O(\|x^c\|^{1+\alpha})}{u_n^{cT} a_c + O(\|x^c\|^{min(\alpha, \frac{1}{\varphi})})} \\ & \quad \pm \frac{\sqrt{\sigma_n^{c2} - 2(u_n^{cT} F_c)(u_n^{cT} a_c) + O(\|x^c\|^{2+min(\alpha, \frac{1}{\varphi})})}}{u_n^{cT} a_c + O(\|x^c\|^{min(\alpha, \frac{1}{\varphi})})}, \end{aligned}$$

from (5.6), (5.8) and ${u_n^c}^T F_c = O(\|x^c\|^2)$ from (4.11). Now using (4.12), ${u_n^c}^T a_c = \Theta(1)$, and $\|x^c_{X^c}\| = O(\|x^c\|^{1+\delta})$,

$$\begin{aligned}
{v_n^c}^T d &= -\frac{\sigma_n^c}{{u_n^c}^T a_c} + O(\|x^c\|^{min(1+\alpha, 1+\frac{1}{\varphi})}) \\
&\quad \pm \frac{\sqrt{O(\|x^c\|\|x^c_{X^c}\|) + O(\|x^c\|^3) + O(\|x^c\|^{2+min(\alpha,\frac{1}{\varphi})})}}{{u_n^c}^T a_c + O(\|x^c\|^{min(\alpha,\frac{1}{\varphi})})} \\
&= -\frac{\sigma_n^c}{{u_n^c}^T a_c} + O(\|x^c\|^{1+min(\alpha,\frac{1}{\varphi})}) + O(\|x^c\|^{1+\frac{1}{2}min(\alpha,\delta,\frac{1}{\varphi})}) \\
&= -\frac{\sigma_n^c}{{u_n^c}^T a_c} + O(\|x^c\|^{1+\frac{1}{2}min(\alpha,\delta,\frac{1}{\varphi})})
\end{aligned}$$

Hence, from the analysis of $\frac{\sigma_n^c}{{u_n^c}^T a_c}$ in the proof of Lemma 4.1, one can easily obtain

$${v_n^c}^T x^+ = O(\|x^c\|^{1+\frac{1}{2}min(\alpha,\delta,\frac{1}{\varphi})}).$$

The proof is complete. □

Using Lemma 4.3, we can change the results in Lemma 5.3 to $\|x^+_X\| = O(\|x^c\|^2)$ and $\|x^+_N\| = O(\|x^c\|^{1+\frac{1}{2}min(\alpha,\delta,\frac{1}{\varphi})})$.

Note that, as with the analysis of the ideal tensor model, Lemma 5.3 is true no matter which of the two real roots of the tensor model is selected.

Now we consider the case when the tensor step is selected by Algorithm 5.0 even though the current and/or previous iterates may not be close to the null space of $F'(x^*)$.

Lemma 5.4 *Let Assumption 3.0 hold. If Algorithm 5.0 selects the tensor step d_T, then $x^+ = x^c + d_T$ satisfies*

$$\begin{aligned}
\|x^+_{X^c}\| &= O(\|x^c\|^2) \\
\|x^+\| &= O(\|x^c\|).
\end{aligned}$$

Proof. Using the same reasoning as in the proofs of Lemmas 5.2 and 5.3, the step d_T must be a root or minimizer of

$$\sigma_i^c {v_i^c}^T d + {u_i^c}^T F_c + \tfrac{1}{2} {u_i^c}^T a_c' (p^T d)^2 \tag{5.27}$$

for $i = 1, \cdots, n$. From Lemma 5.1, the linear model step d_N satisfies $d_N = O(\|x^c\|)$. Therefore if the tensor model step d_T is selected, it follows from the selection rule in Algorithm 5.0 that $d_T = O(\|x^c\|)$. This shows that $\|x^+\| = O(\|x^c\|)$. Also $d_T = O(\|x^c\|)$ implies that $p^T d_T = O(\|x^c\|)$, and ${v_i^c}^T d_T = O(\|x^c\|)$ for $i = 1, \cdots, n$. Thus from the identical reasoning as in the proof of Lemma 5.2, (5.27) must have real roots for $i = 1, \cdots, n-1$, and

$${v_i^c}^T d_T = -\frac{{u_i^c}^T F_c}{\sigma_i^c} + O(\|x^c\|^2).$$

Therefore using the same analysis as in Lemma 5.1 for $\|x^+_{X^c}\|$, we obtain $\|x^+_{X^c}\| = O(\|x^c\|^2)$, which completes the proof. $\square$

Using Lemma 4.3, we can change the result in Lemma 5.4 to $\|x^+_X\| = O(\|x^c\|^2)$.

We note that, using similar techniques to those used in the proof of Lemma 5.4, it is easy to show that if $F'(x^*)$ is *nonsingular* then the tensor step reduces the error quadratically, i.e. it has at least the convergence rate of the Newton step. A sketch of the proof is as follows. From $\|F(x^c)\| = O(\|x^c\|)$ and the fact that the singular values of $F'(x^c)$ are bounded below, it is easily shown that the tensor model (5.1) must have a root, and that the closer root satisfies $d_T = O(\|x^c\|)$. Then using the same analysis as in Lemma 5.1, it is immediate that the closer root obeys $\|x^+\| = O(\|x^c\|^2)$.

5.4 Local convergence of Algorithm 5.0

Theorem 5.5 gives the main convergence result for the practical tensor method. It shows that given Assumption 3.0, the sequence of iterates produced by Algorithm 5.0 converges to x^* from close starting points with a three-step Q-order $\frac{3}{2}$.

Theorem 5.5 *Let Assumption 3.0 hold and $\{x^k\}$ be the sequence of iterates generated by Algorithm 5.0. There exist constants K_1, K_2 such that if $\|x^0\| \le K_1$ then sequence $\{x^k\}$ converges to x^* and*

$$\|x^{k+3}\| \le K_2\|x^k\|^{\frac{3}{2}} \tag{5.28}$$

for $k = 0, 1, 2, \cdots$.

Proof. Select $\bar{\epsilon}$ to be any number in $(\epsilon, \frac{1}{3})$, where $\epsilon < \frac{1}{3}$ is given in Algorithm 5.0. If $\|x^{k+1}_X\| \ge \|x^{k+1}\|^{1+\bar{\epsilon}}$, then since Lemmas 5.1-5.4 show that any step taken by Algorithm 5.0 satisfies $\|x^{k+1}_X\| = O(\|x^k\|^2)$,

$$\|x^{k+1}\| \le \|x^{k+1}_X\|^{\frac{1}{1+\bar{\epsilon}}} = O(\|x^k\|^{\frac{2}{1+\bar{\epsilon}}}) = O(\|x^k\|^{\frac{3}{2}}).$$

Hence at least (5.28) holds. Similarly if $\|x^{k+2}_X\| \ge \|x^{k+2}\|^{1+\bar{\epsilon}}$, then since also $\|x^{k+1}\| = O(\|x^k\|)$ from Lemmas 5.1-5.3,

$$\|x^{k+2}\| \le \|x^{k+2}_X\|^{\frac{2}{1+\bar{\epsilon}}} = O(\|x^{k+1}\|^{\frac{3}{2}}) = O(\|x^k\|^{\frac{3}{2}})$$

and at least (5.28) holds. Otherwise, let $\|x^{k+1}_X\| = \|x^{k+1}\|^{1+\delta_1}$ and $\|x^{k+2}_X\| = \|x^{k+2}\|^{1+\delta_2}$, for $\delta_1, \delta_2 > \bar{\epsilon} > \epsilon$.

Since Lemmas 5.1-5.3 show that any step taken by Algorithm 5.0 satisfies $\|x^{k+1}_X\| = O(\|x^k\|^2)$, then from $\|x^{k+1}_X\| = \|x^k\|^{1+\delta_1}$,

$$O(\|x^k\|^2) = \|x^{k+1}\|^{1+\delta_1},$$

which along with $\|x^{k+1}\| = O(\|x^k\|)$ implies

$$\|x^{k+1}\| = O(\|x^k\|^{max(\frac{2}{1+\delta_1},1)}). \tag{5.29}$$

Likewise

$$\|x^{k+2}\| = O(\|x^{k+1}\|^{max(\frac{2}{1+\delta_2},1)}). \tag{5.30}$$

We will denote the angle between v_n^{k+2} and $x^{k+1} - x^{k+2}$ by γ, and let φ be the rate of error reduction from x^{k+1} to x^{k+2}, i.e.

$$\varphi = \begin{cases} 1 & \text{if } \|x^{k+2}\| > \|x^{k+1}\|, \\ \varphi \geq 1 & \text{for which } \|x^{k+2}\| = \|x^{k+1}\|^{\varphi} \text{ otherwise.} \end{cases} \tag{5.31}$$

We will consider two cases: when $\|x^{k+1}\|^{1+\delta_1} \geq \|x^{k+2}\|^{1+\delta_2}$, and when $\|x^{k+1}\|^{1+\delta_1} < \|x^{k+2}\|^{1+\delta_2}$. In either case, since $\|x_X^{k+1}\| = O(\|x^{k+1}\|^{1+\delta_1}) = O(\|x^{k+1}\|^{1+\bar{\epsilon}})$, the analysis in the proof of Lemma 5.1 shows that the linear model step from x^{k+1}, d_N^{k+1}, is the Newton step and $\|d_N^{k+1}\| = \frac{1}{2}\|x^{k+1}\| + O(\|x^{k+1}\|^{1+\epsilon})$. Thus by Algorithm 5.0, if the tensor step is selected at iteration $k+1$, $\|d_T^{k+1}\| \in [\phi_1, \phi_2]\|d_N^{k+1}\|$ which implies $\|d_T^{k+1}\| = \Theta(\|x^{k+1}\|)$. Therefore in any case

$$\|x^{k+1} - x^{k+2}\| = \Theta(\|x^{k+1}\|). \tag{5.32}$$

Now consider the first case, when $\|x^{k+1}\|^{1+\delta_1} \geq \|x^{k+2}\|^{1+\delta_2}$. Then

$$\|x^{k+1}\|^{1+\delta_1} \geq \|x^{k+1}\|^{(1+\delta_2)\varphi},$$

which implies

$$1 + \delta_1 \leq (1+\delta_2)\varphi$$

or

$$\delta_2 \geq \frac{1+\delta_1-\varphi}{\varphi}. \tag{5.33}$$

Furthermore, from Lemma 4.3, $\|x^{k+1}\| = O(\|x^{k+2}\|^{\frac{1}{\varphi}})$ by the definition (5.31), and (5.32)

$$\begin{aligned} sin\gamma_1 &= \frac{\|(x^{k+1} - x^{k+2})_{X^{k+2}}\|}{\|x^{k+1} - x^{k+2}\|} \\ &\leq \frac{\|x_X^{k+1}\| + \|x_X^{k+2}\| + O(\|x^{k+1}\|\|x^{k+2}\|)}{\Theta(\|x^{k+1}\|)} \\ &= \frac{\|x^{k+1}\|^{1+\delta_1} + \|x^{k+2}\|^{1+\delta_2} + O(\|x^{k+1}\|\|x^{k+2}\|)}{\Theta(\|x^{k+1}\|)} \\ &\leq \frac{2\|x^{k+1}\|^{1+\delta_1} + O(\|x^{k+1}\|\|x^{k+2}\|)}{\Theta(\|x^{k+1}\|)} \\ &= O(\|x^{k+1}\|^{\delta_1}) + O(\|x^{k+2}\|) \\ &= O(\|x^{k+2}\|^{min(\frac{\delta_1}{\varphi},1)}) \end{aligned}$$

Hence $\gamma = O(\|x^{k+2}\|^{min(\frac{\delta_1}{\varphi},1)})$. Thus by Lemmas 5.2 and 5.3, the step from x^{k+2}, d_T^{k+2}, satisfies

$$\|d_T^{k+2}\| = \|x^{k+2}\| + O(\|x^{k+2}\|^{1+\frac{1}{2}min(\frac{1}{\varphi},\frac{\delta_1}{\varphi},\delta_2)}).$$

Also, the linear model step, d_N^{k+2}, is the Newton step because $\|x_X^{k+2}\| = O(\|x^{k+2}\|^{1+\delta_2}) = O(\|x^{k+2}\|^{1+\bar{\epsilon}})$, and by Lemma 5.1, $\|d_N^{k+2}\| = \frac{1}{2}\|x^{k+2}\| + O(\|x^{k+2}\|^{1+\epsilon})$. Therefore for any $\phi_1 < \frac{1}{3}$ and $\phi_2 > 3$,

$$\phi_1\|d_N^{k+2}\| < \|d_T^{k+2}\| < \phi_2\|d_N^{k+2}\|$$

so that the step from x^{k+2} to x^{k+3} is the tensor step, and from Lemmas 5.2 and 5.3

$$\|x^{k+3}\| \le O(\|x^{k+2}\|^{1+\frac{1}{2}min(\delta_2,min(\frac{\delta_1}{\varphi},1),\frac{1}{\varphi})}). \tag{5.34}$$

From (5.29), (5.31), (5.34), and recalling $\varphi > 1$, the three-step rate of error reduction from x^k to x^{k+3} is at least

$$\begin{aligned}
& max(\frac{2}{1+\delta_1},1)\cdot\varphi\cdot(1+\tfrac{1}{2}min(\delta_2,min(\frac{\delta_1}{\varphi},1),\frac{1}{\varphi})) \\
\ge\ & max(\frac{2}{1+\delta_1},1)\cdot\varphi\cdot(1+\tfrac{1}{2}min(\frac{1+\delta_1-\varphi}{\varphi},\frac{\delta_1}{\varphi},\frac{1}{\varphi})) \\
=\ & max(\frac{2}{1+\delta_1},1)\cdot(\varphi+\tfrac{1}{2}min(1+\delta_1-\varphi,\delta_1,1)) \\
=\ & max(\frac{2}{1+\delta_1},1)\cdot\tfrac{1}{2}min(1+\delta_1+\varphi,\delta_1+2\varphi,1+2\varphi) \\
=\ & \begin{cases} \ge \frac{3}{2} & \text{if } \delta_1 \ge 1, \\ \ge \frac{2}{1+\delta_1}\cdot\frac{2+\delta_1}{2} \ge \frac{3}{2} & \text{if } \delta_1 < 1. \end{cases}
\end{aligned}$$

Second, consider the case when $\|x^{k+1}\|^{1+\delta_1} < \|x^{k+2}\|^{1+\delta_2}$. Then from Lemma 4.3, (5.32) and $\|x^{k+1}\| = \Theta(\|x^{k+2}\|^{\frac{1}{\varphi}})$ from (5.31) and (5.32),

$$\begin{aligned}
sin\gamma_1 &= \frac{\|(x^{k+1}-x^{k+2})_{X^{k+2}}\|}{\|x^{k+1}-x^{k+2}\|} \\
&\le \frac{\|x_X^{k+1}\| + \|x_X^{k+2}\| + O(\|x^{k+1}\|\|x^{k+2}\|)}{\Theta(\|x^{k+1}\|)} \\
&= \frac{\|x^{k+1}\|^{1+\delta_1} + \|x^{k+2}\|^{1+\delta_2} + O(\|x^{k+1}\|\|x^{k+2}\|)}{\Theta(\|x^{k+1}\|)} \\
&\le \frac{2\|x^{k+2}\|^{1+\delta_2} + O(\|x^{k+1}\|\|x^{k+2}\|)}{\Theta(\|x^{k+1}\|)} \\
&= \frac{2\|x^{k+2}\|^{1+\delta_2}}{\Theta(\|x^{k+2}\|^{\varphi})} + O(\|x^{k+2}\|) \\
&= O(\|x^{k+2}\|^{min(1+\delta_2-\frac{1}{\varphi},1)}),
\end{aligned}$$

Hence $\gamma_1 = O(\|x^{k+2}\|^{min(1+\delta_2-\frac{1}{\varphi},1)})$.

Now by the identical reasoning as was used in the first case, the step from x^{k+2} to x^{k+3} is the tensor step and obeys

$$\|x^{k+3}\| \leq O(\|x^{k+2}\|^{1+\frac{1}{2}min(\delta_2,1+\delta_2-\frac{1}{\varphi},\frac{1}{\varphi})}). \tag{5.35}$$

From (5.29), (5.30) and (5.35), the three-step rate of error reduction from x^k to x^{k+3} is at least

$$max(\frac{2}{1+\delta_1},1)\cdot\varphi\cdot(1+\tfrac{1}{2}min(\delta_2,1+\delta_2-\frac{1}{\varphi},\frac{1}{\varphi})). \tag{5.36}$$

Also since (5.30) implies

$$\varphi \geq max(\frac{2}{1+\delta_2},1), \tag{5.37}$$

we have

$$(1+\delta_2)\varphi \geq 2. \tag{5.38}$$

Therefore from(5.36), (5.38), and $\varphi \geq 1$, the three-step rate of convergence is at least

$$\begin{aligned} & 1\cdot\varphi\cdot(1+\tfrac{1}{2}min(\delta_2,1+\delta_2-\frac{1}{\varphi},\frac{1}{\varphi})) \\ = & \tfrac{1}{2}min(\varphi\delta_2+2\varphi,\varphi\delta_2+3\varphi-1,2\varphi+1) \\ = & \tfrac{1}{2}min(\varphi+\varphi(1+\delta_2),\varphi(1+\delta_2)+2\varphi-1,2\varphi+1) \\ \geq & \frac{3}{2}. \end{aligned}$$

Combining the above results obtained completes the proof. □

Note that it is straightforward to combine the proof of $p^T d = O(\|x^c\|)$ in Lemma 5.3 with the remainder of the proof of Lemma 5.2 to show that Lemma 5.2 remains true even if the tensor model has real roots. Using this fact, one can use the techniques of the proof of Theorem 5.5 to show that if the tensor method always selects the step to the minimizer, whether the model has a root or not, then under Assumption 3.0, the method is three-step convergent to x^* with rate $2-\epsilon$, for any fixed $\epsilon > 0$.

6 Discussion and conclusions

Theorem 5.5 shows that a practical tensor method possesses a significantly faster local convergence rate than Newton's method on an important class of singular problems. While the theorem indicates that the tensor method possesses a three-step convergence behavior, in our computational experience, tensor methods generally appear to exhibit one-step local superlinear convergence on problems where $rank(F'(x^*)) = n-1$. A proof of such a result would entail showing that once the tensor method's iterates enter a funnel close to the null space, they remain in this funnel. In this case a one-step superlinear result would be immediate from Lemmas 5.2 and 5.3. In fact analysis shows that such behavior is likely, but may not occur if certain error terms in the null space component of the tensor step cancel each other. This cancellation could cause the iterates to leave the funnel close to the null space. The method would then require three steps to recover fast convergence. Thus a one step superlinear result appears to require extra assumptions to ensure that the null space error reduction is always subquadratic. However, since a one step superlinear result is prevented only by the possibility of cancellation in the null space error terms, one would expect one step superlinear convergence to occur often in practice, as has been observed.

An obvious theoretical question related to this paper is whether the results of this paper can be extended to problems where $rank(F'(x^*)) < n-1$. In practice, for example, tensor methods have been observed to exhibit considerably faster local convergence than Newton's method on problems where $rank(F'(x^*)) = n-2$. We do not expect, however, that one can prove a faster than linear convergence result for the tensor methods of Schnabel and Frank [16] when $rank(F'(x^*)) < n-1$, even if one uses a higher rank second order term in the tensor model. The reason is that the current tensor model does not provide enough information to approximate all the components of $F''(x^*)$ that appear to be needed for such a result; it would appear necessary to either use past Jacobian values or previous iterates in linearly dependent directions to obtain the necessary second order information. We doubt whether this effort is warranted in practice.

An interesting computational issue related to this paper is how the ideal tensor method analyzed in Section 4 would perform in practice. In particular, would it be superior to the practical tensor method in terms of the number of iterations required to solve given problems, and would this advantage be sufficient to outweigh the extra costs associated with each iteration? We hope to investigate these questions computationally in the future.

Finally, an intriguing question is whether the computational and theoretical advantages of tensor methods can be extended to nonlinearly constrained optimization problems where the Jacobian matrix of the constraints is rank deficient at the solution. We are currently investigating this issue.

References

[1] A. Bouaricha, Solving large sparse systems of nonlinear equations and nonlinear least squares problems using tensor methods on sequential and parallel computers, Ph.D. dissertation, Department of Computer Science, University of Colorado (1992).

[2] A. Bouaricha and R. B. Schnabel, A software package for tensor methods for nonlinear equations and nonlinear least squares, in preparation.

[3] D. W. Decker, H. B. Keller and C. T. Kelley, Convergence rate for Newton's method at singular points, *SIAM J. Numer. Anal.*, 20 (1983), pp. 296-314.

[4] D. W. Decker and C. T. Kelley, Newton's method at singular points I, *SIAM J. Numer. Anal.*, 17 (1980), pp. 66-70.

[5] D. W. Decker and C. T. Kelley, Newton's method at singular points II, *SIAM J. Numer. Anal.*, 17 (1980), pp. 465-471.

[6] D. W. Decker and C. T. Kelley, Convergence acceleration for Newton's method at singular points, *SIAM J. Numer. Anal.*, 19 (1981), pp. 219-229.

[7] J. E. Dennis Jr. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, New Jersey, (1983).

[8] J. J. Dongarra, J. R. Bunch, C. B. Moler, and G. W. Stewart, *LINPACK Users Guide*, SIAM Publications, Philadelphia, (1979).

[9] P. D. Frank, Tensor methods for solving systems of nonlinear equations, Ph.D. dissertation, Department of Computer Science, University of Colorado (1982).

[10] A. Griewank, Analysis and modification of Newton's method at singularities, Ph.D. dissertation, Australian National University (1980).

[11] A. Griewank, On solving nonlinear equations with simple singularities or nearly singular solutions, *SIAM Rev.*, 27 (1985), pp. 537-563.

[12] A. Griewank and M. R. Osborne, Analysis of Newton's method at irregular singularities, *SIAM J. Numer. Anal.*, 1983 (20), pp. 747-773.

[13] C. T. Kelley and R. Suresh, A new acceleration method for Newton's method at singular points, *SIAM J. Numer. Anal.*, 20 (1983), pp. 1001-1009.

[14] J. M. Ortega, *Numerical Analysis*, Academic Press, New York.

[15] G. W. Reddien, On Newton's method for singular problems, *SIAM J. Numer. Anal.*, 15 (1978), pp. 993-996.

[16] R. B. Schnabel and T. Chow, Tensor methods for unconstrained optimization using second derivatives, *SIAM J. Opt.*, 1 (1991), pp. 293-315.

[17] R. B. Schnabel and P. D. Frank, Tensor methods for nonlinear equations, *SIAM J. Numer. Anal.*, 21 (1984), pp. 815-843.

[18] G. W. Stewart, Error and perturbation bounds for subspaces associated with certain eigenvalue problems, *SIAM Rev.*, Vol. 15, No. 4 (1973), pp. 727-764.