MATHEMATICS OF COMPUTATION
VOLUME 47, NUMBER 176
OCTOBER 1986, PAGES 609–623

A Shamanskii-Like Acceleration Scheme for Nonlinear Equations at Singular Roots*

By C. T. Kelley

Abstract. A variation of the Shamanskii method is used to obtain a superlinearly convergent method for a class of nonlinear equations having singular Fréchet derivative at the root. The cost of a superlinear step is one derivative evaluation and two function evaluations.

1. Introduction. Nonlinear equations of the form $F(x) = 0$ are often solved by variations of Newton's method. If F is a sufficiently smooth map from a Banach space E into itself and $x^* \in E$ is a solution to

$$F(x^*) = 0, \tag{1.1}$$

then the Newton sequence,

$$x_{n+1} = x_n - F'(x_n)^{-1}F(x_n), \tag{1.2}$$

will converge quadratically to x^* if the initial iterate x_0 is sufficiently near x^* and $F'(x^*)$ is nonsingular [10], [17], [23]. Evaluation of the derivative and solution of the linear equation for the step, $s(x_n) = -F'(x_n)^{-1}F(x_n)$, may be very expensive. An alternative approach that almost entirely avoids this cost is the chord method. Here we only compute $F'(x_0)$; the iterates are

$$x_{n+1} = x_n - F'(x_0)^{-1}F(x_n). \tag{1.3}$$

For a finite-dimensional problem, the Jacobian matrix need only be factored once. Hence the cost of each iterate is very low. The convergence rate, however, is degraded to linear [10], [17], [23]. In fact, for x_0 sufficiently near x^* and $F'(x^*)$ nonsingular, we have, for some $K_C > 0$,

$$\|x_{n+1} - x^*\| \leq K_C\|x_0 - x^*\|\,\|x_n - x^*\|. \tag{1.4}$$

We observe, directly from (1.4), that the rate of linear convergence improves as the initial guess gets better. This motivates a method that lies between the chord method and Newton's method. This method, due originally to Shamanskii [28], and analyzed in detail by others as well [1], [23], [32], proceeds as follows: Given an integer m and an initial iterate x_0, we move from x_n to x_{n+1} through an intermediate sequence

Received April 29, 1985; revised January 10, 1986.
1980 *Mathematics Subject Classification*. Primary 49D15; Secondary 65H10.
*This work supported by NSF grants DMS-8300841 and DMS-8500944.

0025-5718/86 $1.00 + $.25 per page

$\{y_{n,p}\}_{p=1}^{m}$, which is one Newton iterate followed by several chord iterates,

$$\begin{aligned} y_{n,1} &= x_n - F'(x_n)^{-1}F(x_n), \\ y_{n,p+1} &= y_{n,p} - F'(x_n)^{-1}F(y_{n,p}), \qquad 1 \leqslant p \leqslant m-1, \\ x_{n+1} &= y_{n,m}. \end{aligned} \tag{1.5}$$

Note that when $m = 1$, Newton's method results, and when $m = \infty$, we obtain the chord method. If $F'(x^*)$ is nonsingular and x_0 is sufficiently near x^*, the convergence of the Shamanskii method is m-step q-superlinear with q-order at least $m + 1$. This means that there is $K_S > 0$, such that

$$\|x_{n+1} - x^*\| \leqslant K_S \|x_n - x^*\|^{m+1}. \tag{1.6}$$

The convergence rate, (1.6), for the Shamanskii method is an easy consequence of (1.4) and the quadratic convergence for Newton's method. In fact, if x_0 is sufficiently near x^* there is $K_N > 0$, such that

$$\|y_{n,1} - x^*\| \leqslant K_N \|x_n - x^*\|^2. \tag{1.7}$$

Repeated applications of (1.4) imply that, for $1 \leqslant p \leqslant m - 1$,

$$\|y_{n,p+1} - x^*\| \leqslant K_C^p K_N \|x_n - x^*\|^{p+2}. \tag{1.8}$$

(1.8) gives (1.6) with $K_S = K_C^{m-1} K_N$.

If, instead of the Shamanskii method, one does m complete Newton steps, the overall improvement in the accuracy is

$$\|x_{n+m} - x^*\| \leqslant K_N^m \|x_n - x^*\|^{2^m}. \tag{1.9}$$

The tradeoff, therefore, is between the faster convergence of Newton's method and the reduced cost in derivative evaluations and matrix factorizations for the Shamanskii method. This situation was analyzed in [1], for several methods, including the Shamanskii method, by Brent. He found that, if one measures work in terms of function evaluations and counts a Jacobian evaluation as M function evaluations, where $M < \infty$ is the dimension of E, the Shamanskii method is always more efficient than Newton's method. Moreover, the relative performance of the Shamanskii method improves as the dimension increases if the optimal value of m is chosen

The purpose of the present paper is consideration of these issues if $F'(x^*)$ is singular. Before discussing the specific new results, we will give some known facts about Newton and Newton-like methods for problems having singular derivative at the root. We will call such problems singular, and call problems with $F'(x^*)$ invertible nonsingular.

We begin by considering a simple problem in one variable. Let $f \in C^{k+2}(R)$ satisfy

$$f(x^*) = f'(x^*) = \cdots = f^{(k)}(x^*) = 0 \neq f^{(k+1)}(x^*). \tag{1.10}$$

It is easy to show that if the initial guess, x_0, is sufficiently near, but not equal to, x^*, the Newton iterates, $\{x_n\}$, will converge linearly to x^* with an asymptotic linear ratio of $k/(k + 1)$. This means that

$$\lim_{n \to \infty} \frac{|x_{n+1} - x^*|}{|x_n - x^*|} = \frac{k}{k+1}. \tag{1.11}$$

In this situation, quadratic convergence may be restored by modifying the Newton step, [31]. The following sequence of iterates will converge quadratically to x^*, if the initial guess is in a sufficiently small deleted neighborhood of x^*:

$$x_{n+1} = x_n - (k+1)f(x_n)/f'(x_n), \qquad n \geqslant 0. \tag{1.12}$$

For Newton's method, therefore, linear convergence will be the best one can hope for, but it is possible to recover superlinear convergence by modifying the iterates in a simple way.

Methods that do not reevaluate the derivative at each step, like the chord and Shamanskii methods, do not fare as well. For the chord method, sublinear convergence should be expected, [12], [8]. Hence, the chord method is not a reasonable approach to these problems. The Shamanskii method will, as we shall see, converge linearly at an asymptotic linear rate that depends both on the order, k, of the root and the number m. This linear rate, which we denote by $r_{k,m}$, is given by the recurrence relation

$$r_{k,1} = k/(k+1), \qquad r_{k,p+1} = \left(1 - r_{k,p}^k/(k+1)\right) r_{k,p}. \tag{1.13}$$

We prove this in Section 2. We will discuss the approximate size of the linear rates, $r_{k,m}$ after the statement of Theorem 1.3.

The behavior of iterative methods applied to singular problems in more than one variable is more complex [2], [4]–[9], [12]–[16], [19]–[21], [24], [25], [29], [30]. To begin with, it is no longer the case that $F'(x)$ remains nonsingular in a deleted neighborhood of the root. Instead, $F'(x)$ is generally singular for x in a collection of codimension one smooth manifolds passing through x^*. A consequence of this is that the set of initial iterates that produce convergence is not a deleted neighborhood of x^*, but a region that avoids the set on which $F'(x)$ is singular. The construction of such regions is an important part of an analysis of convergence.

The issue of computation of rates of convergence is also made more complex for higher-dimensional problems. It is not the case that the Newton iterates must converge at a linear rate of the form $k/(k+1)$ [15], [25]. In fact, examples have been created for which the Newton iterates diverge for almost every initial guess [15]. Most work in this area has avoided this strange behavior by imposing some structure on the nature of the singularity. We do this later in this section.

For many problems [2], [4]–[7], [9], [12]–[16], [24], the rates of convergence for Newton's method are of the form $k/(k+1)$. A natural question is whether the acceleration method (1.12) will recover quadratic convergence, as it does for problems in one variable. The answer is no, because the modified iterate, as we shall see shortly, may leave the region of valid initial guesses, and there will be no guarantee that subsequent iterates will be defined. Designing acceleration schemes is complicated for this reason [6], [7], [9], [12], [16], [19]–[21], [29], [30].

In this paper we give regions of valid initial iterates and compute rates of convergence for the Shamanskii method for a class of singular problems. We also derive a new acceleration strategy, based on the Shamanskii method, that is more efficient than the existing methods. Before stating the main results, we establish some notation and discuss the nature of the singularities of interest.

We assume that $F'(x^*)$ has finite-dimensional null space N and range X with $E = N \oplus X$. Let P_N denote a projection onto N parallel to X and let $P_X = I - P_N$. We will let $\hat{F} = F'(x^*)$. Often $\hat{F}$ will be considered as a map on X alone; when this is done, $\hat{F}$ is invertible and we let $\hat{F}^{-1}$ denote that inverse. Formally, $\hat{F}^{-1}F'(x^*) = F'(x^*)\hat{F}^{-1} = P_X$.

As we said before, it is not always the case that the asymptotic linear rate of convergence for Newton's method for singular problems is of the form $k/(k+1)$. We make an assumption on the singularity to insure that this is the case. This assumption, called "regularity" in [12]–[16] and also used in [7], [21], [27], is somewhat technical. After stating the assumption and giving the theorem on Newton's method that requires that assumption, we discuss its geometric meaning and its significance in the design of acceleration methods. The assumption is:

(R) Let k be the least integer such that (we assume that $F \in C^{(k+3)}$) $P_N F^{(k+1)}(x^*)(z_1, \ldots, z_{k+1}) \neq 0$ for some $\{z_j\} \subset E$. There is $\phi_0 \in N$, such that the operator H, defined on N by $H = P_N F^{(k+1)}(x^*)(\phi_0^k, P_N \cdot)$, is invertible.

The integer k is sometimes called the order of the singularity [12]–[16], [21].

One consequence of assumption (R) is the geometric statement that the set, M, on which $F'(x)$ is singular, passes through x^* transverse to the affine set $\{x \mid x = x^* + e,\ e \in N\}$. Assumption (R) has also, in the case where $k = 1$ and $\dim(N) = 1$, been used to argue that x^* is an isolated root of F, [18]. Further discussion of the geometric consequences of (R) must await the description of the behavior of the Newton sequence.

We let P_0 denote a projection in N onto the one-dimensional subspace spanned by ϕ_0; we let $\delta P_N = P_N - P_0$. For convenience we let $\tilde{x} = x - x^*$. We let $\beta_k(x)$ denote any term that is $O(\|\tilde{x}\|^k)$. For ρ, θ, and η positive we define a cone in E, $W(\rho, \theta, \eta)$, by

$$W(\rho, \theta, \eta) = \{x \mid 0 < \|\tilde{x}\| \leqslant \rho,\ \|P_X\tilde{x}\| \leqslant \theta\|P_N\tilde{x}\|,\ \|\delta P_N\tilde{x}\| \leqslant \eta\|P_N\tilde{x}\|\}. \tag{1.14}$$

With this notation in mind, we quote as a theorem a combination of results from [2], [4]–[7], [9], [12]–[16], [24].

THEOREM 1.1. *Assume that* (R) *holds. Then there exists* $\hat{\rho}$, $\hat{\theta}$, *and* $\hat{\eta}$ *such that for all* $x \in W(\hat{\rho}, \hat{\theta}, \hat{\eta})$, $F'(x)^{-1}$ *exists and satisfies*

$$F'(x)^{-1} = \beta_{-k}(x). \tag{1.15}$$

In addition, for ρ, θ, *and* η *sufficiently small and* $x_0 \in W(\rho, \theta, \eta)$, *the Newton sequence,*

$$x_{n+1} = x_n - F'(x_n)^{-1}F(x_n), \qquad n \geqslant 0, \tag{1.16}$$

is in $W(\hat{\rho}, \hat{\theta}, \hat{\eta})$ *for all n and converges to* x^* *with rates given by*

$$\begin{aligned} P_N\tilde{x}_{n+1} = {} & \frac{k}{k+1}P_N\tilde{x}_n + P_N(\theta_n\beta_1(x_n) + \beta_2(x_n)) \\ & + P_X(\theta_n\beta_2(x_n) + \beta_3(x_n)), \end{aligned} \tag{1.17}$$

where $\theta_n = \|P_X\tilde{x}_n\|/\|P_N\tilde{x}_n\|$ *and* $\lim_{n\to\infty}\theta_n = 0$. *For* $n \geqslant 1$, $\theta_n = \beta_1(x_n)$.

The above theorem is much more than the simple statement that the Newton iterates converge to x^* with an asymptotic linear rate of $k/(k+1)$. Equation (1.15) is a restriction on the size of $F'(x)^{-1}$ for $x \in W(\hat{\rho}, \hat{\theta}, \hat{\eta})$; this keeps the Newton step from becoming so large that the iterates leave $W(\hat{\rho}, \hat{\theta}, \hat{\eta})$. Equation (1.17) describes the limiting behavior of $\|\tilde{x}_n\|$ and θ_n. This in turn leads to important information on the path taken by the Newton iterates. This path becomes tangential to the affine space $\{x \in E \mid x = x^* + e, e \in N\}$ in a better than parabolic way. We have

$$\lim_{n \to \infty} \|P_X \tilde{x}_n\| / \|P_N \tilde{x}_n\|^2 = 0. \tag{1.18}$$

Therefore, the direction of the errors, $\tilde{x}_n$, rapidly becomes dominated by N, and $\tilde{x}_n$ moves farther away in direction from the set on which $F'(x)$ is singular.

The above geometric properties play an important role in acceleration methods which recover superlinear convergence. If we use a modified Newton step, as one would for the one-variable case, (1.12), the modified iterate could well leave $W(\hat{\rho}, \hat{\theta}, \hat{\eta})$, as there is no guarantee that $\|P_N \tilde{x}_{n+1}\|$ and $\|P_X \tilde{x}_{n+1}\|$ would not be roughly the same size. The solution is to take an intermediate Newton iterate before the modified step. The properties of the direction of $\tilde{x}_n$ allow us to still use a modified step similar to (1.12) and take into account the requirement that the iterates remain in $W(\hat{\rho}, \hat{\theta}, \hat{\eta})$. This was done in [21]. To state this result, we write the Newton step as

$$s^N(x) = -F'(x)^{-1} F(x). \tag{1.19}$$

THEOREM 1.2. *Assume that* (R) *holds. Let* $W(\hat{\rho}, \hat{\theta}, \hat{\eta})$ *be as in Theorem* 1.1, *and let* $C \neq 0$ *and* $\alpha \in (0,1)$ *be given. Then, if* $x_0 \in W(\rho, \theta, \eta)$ *for* ρ, θ, *and* η *sufficiently small, and* y_{n+1} *and* x_{n+1} *are given, for* $n \geq 0$, *by*

$$y_{n+1} = x_n + s^N(x_n), \tag{1.20}$$

$$x_{n+1} = y_{n+1} + (k+1) s^N(y_{n+1}) - \left[C \|s^N(y_{n+1})\|^{\alpha} s^N(y_{n+1}) \right], \tag{1.21}$$

then the sequence $\{x_n\}$ *is defined, remains in* $W(\hat{\rho}, \hat{\theta}, \hat{\eta})$, *and converges to* x^* *q-superlinearly with q-order* $1 + \alpha$.

The algorithm given by Theorem 1.2 provides two-step superlinear convergence. The cost of a superlinear step is two derivative evaluations. It is appropriate here to discuss the geometric aspects of Theorem 1.2 in order to explain the purpose of the intermediate step, (1.20), to explain the modified step, (1.21), and to give some insight into the choice of C and α. As one can see from Theorem 1.1, a Newton step has the effect of further reducing the weight of the X component of $\tilde{x}$; after taking such a step, the result of (1.12) might well leave $W(\hat{\rho}, \hat{\theta}, \hat{\eta})$, but would not miss by much. Hence, the additional correction term in the brackets in (1.21) can force the direction of $\tilde{x}_{n+1}$ to be close to that of $\tilde{x}_n$. This is why x_{n+1} remains in $W(\hat{\rho}, \hat{\theta}, \hat{\eta})$. A similar analysis will be applied to the Shamanskii method in the next section. Now, if C is taken too small, or α too large, this correction may not be enough to force $x_{n+1} \in W(\hat{\rho}, \hat{\theta}, \hat{\eta})$. On the other hand, if C is too large, or α too small, the sequence $\{x_n\}$, while in $W(\hat{\rho}, \hat{\theta}, \hat{\eta})$, will converge more slowly to x^* than a combination of smaller C or larger α. Balancing this is somewhat problem-dependent, but we have found that $C = 1$ and $\alpha = .9$ are generally good choices.

The main result of this paper is that in (1.21), one does not need to compute the derivative, $F'(y_{n+1})$, but can use, instead of the Newton step, a Shamanskii step. This new method is more efficient, even though an extra initial Newton iterate must be taken and a smaller value of α used. Our results also describe the behavior of the Shamanskii method when not modified for superlinear convergence.

For x such that $F'(x)^{-1}$ exists and y in E, we write the Shamanskii step as

$$s^S(x, y) = -F'(x)^{-1}F(y). \tag{1.22}$$

When no confusion can result, we will suppress the x-dependence in (1.22). This will occur when x is considered as fixed and we wish to analyze the step for various values of y. We prove two results in Section 2. The first is a description of the behavior of the unmodified Shamanskii method for singular problems satisfying (R). We will denote by $S^{(m)}$ the Shamanskii method with a given derivative used m times (so $S^{(1)}$ is Newton's method, etc.).

THEOREM 1.3. *Assume that* (R) *holds and let* $W(\hat{\rho}, \hat{\theta}, \hat{\eta})$ *be as in Theorem* 1.1. *Then, for a given* m, *there exist* ρ, θ, *and* η *such that if* $x_0 \in W(\rho, \theta, \eta)$ *and* $\{x_n\}$ *are the iterates given by* $S^{(m)}$, *the* x_n*'s are in* $W(\hat{\rho}, \hat{\theta}, \hat{\eta})$ *and converge linearly to* x^*. *The linear rate is given by*

$$\lim_{n \to \infty} \|\tilde{x}_{n+1}\|/\|\tilde{x}_n\| = r_{k,m}, \tag{1.23}$$

where $r_{k,m}$ *is given by* (1.13).

We shall see in the proof that (1.8) does not hold in general, but does if $k = 1$. In general, all that one can guarantee is that the tract of the iterates is roughly parabolic. This is the only difference in the geometry that will be important later on. This fact will make the proof of the acceleration result a bit different from that in [21].

For finite-dimensional problems we may compare the result in Theorem 1.3 with the results in [1] as far as the best choice of m goes. If one views (1.13) as a differential equation for $r_{k,m}$ as a function of m, one finds that $r_{k,m}$ may be approximated as

$$r_{k,m} \simeq (km/(k+1)+1)^{-1/k} \simeq (m+1)^{-1/k}. \tag{1.24}$$

Then, if one follows [1] and defines efficiency as the ratio,

$$\text{eff} = \frac{|\log(\text{linear rate})|}{\text{work}} \simeq \frac{\log(1+m)}{k(m+M)}. \tag{1.25}$$

Here M would be the dimension of E if we measured work in terms of function evaluations. Alternatively, we can regard M as the ratio of the work required to evaluate and factor the Jacobian matrix to that required to evaluate F and solve a factored system. So efficiency is roughly maximized at an m that is independent of k and is the same as found in [1] for regular problems. While this is not exactly the case, we have found the optimal m for regular problems to be very close to the optimal m for the singular problems. For example, if $M = 10$, the optimal m for regular problems is 7, for singular problems we found the optimal m by inspection for values of k between 1 and 20. For $k = 1$ the optimal m was 7; for $2 \leqslant k \leqslant 20$

the optimal m was 6. For larger M the variation was somewhat larger. For $M = 100$, the optimal m for regular problems was 37; for the singular problems, the optimal m ranged from 39 for $k = 1$ to 35 for $k \geq 8$.

The next theorem, an acceleration result, uses only $m = 2$. Theorem 1.4 is a more efficient form of Theorem 1.2 in that the derivative at the intermediate iterate need not be computed or factored.

THEOREM 1.4. *Assume that* (R) *holds and* $W(\hat{\rho}, \hat{\theta}, \hat{\eta})$ *is as in Theorem* 1.1. *Let* x_{-1} *be given and let* $x_0 = x_{-1} + s^N(x_{-1})$. *Consider the sequence* $\{x_n\}$, *defined for* $n \geq 1$, $C \neq 0$, $\alpha > 0$, *and* $n \geq 0$ *by*

$$y_{n+1} = x_n + s^N(x_n),$$

$$x_{n+1} = y_{n+1} + \left(\frac{(k+1)^{k+1}}{k^k}\right) s^S(x_n, y_{n+1}) \tag{1.26}$$

$$- \left[C\|s^S(x_n, y_{n+1})\|^\alpha s^S(x_n, y_{n+1})\right].$$

Let $\alpha \in (0, (\sqrt{5} - 1)/2)$ *if* $k = 1$ *and let* $\alpha \in (0, \sqrt{2} - 1)$ *if* $k > 1$. *Then, for* ρ, θ, *and* η *sufficiently small and* $x_{-1} \in W(\rho, \theta, \eta)$, *we have* $x_n \in W(\hat{\rho}, \hat{\theta}, \hat{\eta})$ *for all* $n \geq 0$, *and the sequence* $\{x_n\}$ *converges q-superlinearly to* x^* *with q-order* $1 + \alpha$.

In Theorem 1.4 the choice of α is more restricted than in Theorem 1.2. Using a slight variation of the analysis in [1], we measure efficiency as the ratio of the log of the q-order of convergence of the sequence x_n to the number of derivative evaluations required to move from x_n to x_{n+1}. When $k = 1$, $C = 1$, and $\alpha = .6$ is used for the new algorithm, and $\alpha = .9$ for that of Theorem 1.2, the new algorithm is more efficient by a factor of roughly 1.46. We believe that this overrides the cost of the additional iterate, from x_{-1} to x_0 at the beginning. This is certainly the case in the examples considered in Section 3. However, if $k > 1$, the choice of α is even more restricted. The algorithm in Theorem 1.4 is still slightly more efficient. If $k > 1$ and we use $\alpha = .4$ for Theorem 1.4, and $\alpha = .9$ for Theorem 1.2, the new algorithm is only more efficient by a factor of 1.04. This small increase may not offset the cost of the extra iterate at the start. In the examples we only consider problems for which $k = 1$.

We remark that if (R) holds, k can be determined by observation of the progress of the iterates. In fact, if $x_1 = x_0 + s^N(x_0)$, then k can be computed by consideration of the number R given by

$$R = \|s^N(x_0)\|/\|s^N(x_1)\| = (k+1)/k + \theta_0\beta_1(x_0) + \beta_2(x_0). \tag{1.27}$$

For ρ and θ sufficiently small, k is the nearest integer to $(R-1)^{-1}$. For the Shamanskii iterates we define R by

$$R = \|s^N(x_0)\|/\|s^S(x_0, y_1)\| = ((k+1)/k)^{k+1} + \theta\beta_0(x_0) + \beta 1(x_0). \tag{1.28}$$

It would be more difficult to determine k from the R given in (1.28) than the one given in (1.27). If k is not known we would recommend taking two Newton iterates and using (1.27) to find it.

In Section 2 we prove Theorems 1.3 and 1.4 by analysis of the Shamanskii step. In Section 3 we present some numerical examples.

2. Analysis of the Shamanskii Step. In this section we consider the behavior of the m-step Shamanskii method, $S^{(m)}$. We consider first one cycle; then we will use that analysis inductively to prove the main results. We take as the initial iterate $x_0 = x \in W(\rho, \theta, \eta) \subseteq W(\hat{\rho}, \hat{\theta}, \hat{\eta})$, where $W(\hat{\rho}, \hat{\theta}, \hat{\eta})$ is as in Theorem 1.1. We will use the methods and notation developed in [7] that we have used in [8], [9], [21] for Newton's method and the chord method. We begin with an expansion of $F'(x)^{-1}$.

We define operators A, B, C, and D by:

$$A(x) = P_X F'(x) P_X, \qquad B(x) = P_X F'(x) P_N, \tag{2.1}$$
$$C(x) = P_N F'(x) P_X, \qquad D(x) = P_N F'(x) P_N.$$

We note that for ρ sufficiently small, A is an invertible operator on X. In fact, $A(x)^{-1} = \hat{F}^{-1} + \beta_1(x)$. We let $\tilde{D}$ be defined as a map on N by

$$\tilde{D}(x) = D(x) - C(x)A(x)^{-1}B(x). \tag{2.2}$$

Assumption (R) implies that $\tilde{D}$ is an invertible map on N and that $\tilde{D}(x)^{-1} = \beta_{-k}(x)$. For $x \in W(\hat{\rho}, \hat{\theta}, \hat{\eta})$,

$$P_N F'(x)^{-1} = P_N \tilde{D}(x)^{-1}\left[P_N - C(x)A(x)^{-1}\right], \tag{2.3}$$
$$P_X F'(x)^{-1} = A(x)^{-1}\left[P_X - B(x)P_N F'(x)^{-1}\right].$$

This is far more precise than the simple statement that $F'(x)^{-1} = \beta_{-k}(x)$ and will be used to obtain detailed information on the directions of the steps.

Proof of Theorem 1.3. We now begin the proof of Theorem 1.3. We will denote our initial iterate by x, and take m as given. We will denote the Shamanskii iterates from the starting point x by $\{y_j\}_{j=1}^m$. The first iterate, y_1, is a Newton iterate. Hence, by Theorem 1.1, we have

$$P_N \tilde{y}_1 = (k/k+1)P_N \tilde{x} + \theta\beta_1(x) + \beta_2(x) \quad \text{and} \quad P_X \tilde{y}_1 = \theta\beta_2(x) + \beta_3(x). \tag{2.4}$$

For y_j in the Shamanskii sequence, we define $\bar{\rho}_j$ and $\bar{\theta}_j$ by

$$\bar{\rho}_j = \|\tilde{y}_j\|, \qquad \bar{\theta}_j = \|P_X \tilde{y}_j\|/\|P_N \tilde{y}_j\|. \tag{2.5}$$

Then $y_j \in W(\bar{\rho}_j, \bar{\theta}_j, \bar{\eta}_j)$, for some $\bar{\eta}_j$. Our analysis will show that $\{\bar{\rho}_j\}$ is a decreasing sequence, that for ρ, θ, and η sufficiently small, the sequence $\{\bar{\eta}_j\}$ remains bounded by $\hat{\eta}$, and that $\bar{\theta}_m < \theta$.

We make one important observation about the relative sizes of $\tilde{x}$ and $\tilde{y}_j$. The sequence $\{r_{k,p}\}$, defined by (1.13), is monotone decreasing, and hence, for $p \leqslant m$, has a lower bound. This will mean that terms like $\beta_i(x)$ and $\beta_i(y_j)$ will differ by a multiplicative factor that is bounded from above by 1 and away from zero from below. We can therefore regard $\tilde{y}_j$ and $\tilde{x}$ to have the same order. This observation allows us to express all error terms in x, rather than having terms in y_j for several different j's.

We will denote the Shamanskii step $s^S(x, y)$ simply by $s^S(y)$, as x will be fixed for this part of the analysis. In order to compute this step, we require an expansion of $F(y_j)$. We will assume that y_j is related to x as follows:

$$P_N \tilde{y}_j = \bar{r}_j P_N \tilde{x} + \theta\beta_1(x) + \beta_2(x), \qquad \bar{\theta}_j < \theta, \ \bar{\rho}_j < \rho. \tag{2.6}$$

Equation (2.6) clearly holds when $j = 1$; here, $\bar{r}_1 = k/(k+1)$. We note that (2.4) implies that

$$\bar{\theta}_1 = \theta\beta_1(x) + \beta_2(x). \tag{2.7}$$

We now mention some consequences of assumption (R). To begin with, the leading term in the Taylor series expansion of $P_N F(y_j)$ about x^* is the $(k+1)$st. This means that

$$P_N F(y_j) = \frac{1}{k+1}\left[D(y_j) + C(y_j)\right]\tilde{y}_j + \beta_{k+2}(x). \tag{2.8}$$

Hence, by (2.6), and the fact that $C(x)\tilde{y}_j = \bar{\theta}_j\beta_{k+1}(x)$,

$$\begin{aligned} P_N F(y_j) &= \frac{\bar{r}_j^k}{k+1}\left[D(x) + C(x)\right]\tilde{y}_j + \beta_{k+2}(x) + \theta\beta_{k+1}(x) \\ &= \frac{\bar{r}_j^k}{k+1}D(x)\tilde{y}_j + \beta_{k+2}(x) + \theta\beta_{k+1}(x). \end{aligned} \tag{2.9}$$

As far as $P_X F$ is concerned, assumption (R) says very little explicitly. Clearly, however,

$$P_X F(y_j) = \hat{F}\tilde{y}_j + \tfrac{1}{2}P_X F''(x^*)(\tilde{y}_j, \tilde{y}_j) + \beta_3(x). \tag{2.10}$$

In order to compute $s^S(y_j)$, we use (2.3) and (2.10). The result is different depending on whether $k = 1$ or $k > 1$. We have, since $\hat{F} = A(x) + \beta_1(x)P_X$, that

$$\begin{aligned} P_X F(y_j) &= A(x)\tilde{y}_j + \frac{\bar{r}_j^k}{k+1}B(x)\tilde{y}_j + \beta_2(x) && \text{if } k > 1, \\ P_X F(y_j) &= A(x)\tilde{y}_j + \frac{\bar{r}_j}{2}B(x)\tilde{y}_j + \theta\beta_2(x) + \beta_3(x) && \text{if } k = 1. \end{aligned} \tag{2.11}$$

We then obtain

$$\begin{aligned} -P_N s^S(y_j) &= \left(\frac{\bar{r}_j^k}{k+1}\right)P_N\tilde{D}(x)^{-1}\left[D(x) - C(x)A(x)^{-1}B(x)\right]\tilde{y}_j \\ &\quad - \tilde{D}(x)^{-1}C(x)\tilde{y}_j + \theta\beta_1(x) + \beta_2(x) \\ &= \left(\frac{\bar{r}_j^k}{k+1}\right)P_N\tilde{y}_j + \theta\beta_1(x) + \beta_2(x). \end{aligned} \tag{2.12}$$

The final stage in this part of the analysis is to compute $P_X s^S(y_j)$. Again we use (2.3),

$$-P_X s^S(y_j) = A(x)^{-1}P_X F(y_j) - A(x)^{-1}B(x)P_N s^S(y_j). \tag{2.13}$$

From (2.11) and (2.12) we have that

$$\begin{aligned} A(x)^{-1}P_X F(y_j) &= P_X\tilde{y}_j + A(x)^{-1}B(x)P_N s^S(y_j) \\ &\quad + \beta_2(x) \quad \text{if } k > 1, \quad \text{and} \\ A(x)^{-1}P_X F(y_j) &= P_X\tilde{y}_j + A(x)^{-1}B(x)P_N s^S(y_j) \\ &\quad + \theta\beta_2(x) + \beta_3(x) \quad \text{if } k = 1. \end{aligned} \tag{2.14}$$

Hence,

$$\begin{aligned} -P_X s^S(y_j) &= P_X\tilde{y}_j + \beta_2(x) && \text{if } k > 1, \quad \text{and} \\ -P_X s^S(y_j) &= P_X\tilde{y}_j + \theta\beta_2(x) + \beta_3(x) && \text{if } k = 1. \end{aligned} \tag{2.15}$$

We may now compute y_{j+1},

$$P_N\tilde{y}_{j+1} = \left(1 - \frac{\bar{r}_j^k}{k+1}\right)P_N\tilde{y}_j + \theta\beta_1(x) + \beta_2(x), \qquad P_X\tilde{y}_{j+1} = \beta_2(x). \tag{2.16}$$

The dependency of the estimates in (2.11) and (2.14)–(2.15) on the value of k will be used in the proof of Theorem 1.4, but only the weaker of the estimates will be needed in the remainder of this proof.

We use (2.16) to estimate the quantities $\bar{\rho}_{j+1}$, $\bar{\theta}_{j+1}$, $\bar{\eta}_{j+1}$, and $\bar{r}_{j+1}$. We have

$$\begin{aligned}
\bar{\rho}_{j+1} &= \left(1 - \frac{\bar{r}_j^k}{k+1}\right)\bar{\rho}_j + \theta\beta_1(x) + \beta_2(x), \qquad \bar{\theta}_{j+1} = \beta_1(x), \\
\bar{\eta}_{j+1} &= \frac{\bar{\eta}_j + \theta\beta_0(x) + \beta_1(x)}{1 + \theta\beta_0(x) + \beta_1(x)}, \\
\bar{r}_{j+1} &= \bar{r}_j\left(1 - \frac{\bar{r}_j^k}{k+1}\right) + \theta\beta_0(x) + \beta_1(x).
\end{aligned} \tag{2.17}$$

Now, since $\bar{\rho}_1 = k/(k+1) + \theta\beta_0(x) + \beta_1(x)$, if ρ, θ, and η are sufficiently small, we obtain immediately from (2.17) that there is $c_1 > 0$ such that

$$r_{k,j+1} - c_1(\rho + \theta) < \bar{r}_{j+1} < k/(k+1) + c_1(\rho + \theta) < 1. \tag{2.18}$$

In (2.18), $r_{k,j+1}$ is given by (1.13). Therefore, there is c_2 such that, for all $1 \leqslant j < m$,

$$\big(r_{k,j+1} - c_2(\theta + \rho)\big)\rho < \bar{\rho}_{j+1} < \big(r_{k,j+1} + c_2(\theta + \rho)\big)\rho. \tag{2.19}$$

Hence, for θ and ρ sufficiently small, we obtain, for all $1 \leqslant j \leqslant m-1$, $\bar{\rho}_{j+1} < \bar{\rho}_j$. Hence, the intermediate Shamanskii steps are moving toward the root with a rate described by (2.18). It remains to show that y_m, the final iterate in the cycle, lies in $W(\hat{\rho}, \hat{\theta}, \hat{\eta})$, and hence a Newton step may be taken. We proceed, as in [10], [17], [23], by first looking at $\bar{\theta}_j$. By (2.16), there is c_3 such that $\bar{\theta}_j < c_3\rho$, for all $1 \leqslant j \leqslant m$. We demand that θ and ρ be small and related so that there is $r < 1$ such that,

$$c_3\rho < \hat{\theta}, \qquad \bar{\rho}_j < r\rho \quad \text{for all } 1 \leqslant j \leqslant m. \tag{2.20}$$

At this stage we have shown that our next iterate, x_1, lies in $W(\rho_1, \theta_1, \eta_1)$, where

$$\begin{aligned}
&\theta_1 < c_3\rho < \hat{\theta}, \qquad \rho_1 < r\rho, \\
&\eta_1 \leqslant \max_{1 \leqslant j \leqslant m} \bar{\eta}_j \leqslant \frac{\eta + c_4(\theta + \rho)}{1 - c_4(\theta + \rho)} \quad \text{for some } c_4 > 0.
\end{aligned} \tag{2.21}$$

Hence, for ρ, θ, and η small, $x_1 = y_m \in W(\hat{\rho}, \hat{\theta}, \hat{\eta})$. The remainder of the proof uses the above formulae inductively. If we let $\theta_n = \|P_X\tilde{x}_n\|/\|P_N\tilde{x}_n\|$ and $\rho_n = \|\tilde{x}_n\|$, and have $x_n \in W(\rho_n, \theta_n, \eta_n) \subseteq W(\hat{\rho}, \hat{\theta}, \hat{\eta})$, where, for $1 \leqslant p \leqslant n$,

$$\begin{aligned}
&\text{(a)} \quad \rho_p < r\rho_{p-1}, \\
&\text{(b)} \quad \theta_p < \hat{\theta}, \\
&\text{(c)} \quad \eta_p < \frac{\eta_{p-1} + c_4(\theta_{p-1} + \rho_{p-1})}{1 - c_4(\theta_{p-1} + \rho_{p-1})},
\end{aligned} \tag{2.22}$$

then (2.22) holds for $p = n+1$ by (2.20) and (2.21). In addition, since the sequence $\{\bar{r}_j\}$ has a positive lower bound, we have, for $n \geqslant 2$,

$$\theta_n < c_3\rho_{n-1}, \tag{2.23}$$

and, therefore, the sequence $\{\theta_n\}$ converges to zero. Therefore, the sequence η_n is bounded by $\hat{\eta}$ for ρ and η sufficiently small, [7], [14], [21], [27]. Note that we have

also shown that ρ_n and θ_n converge to zero as n tends to infinity. The convergence rate, (1.23), in Theorem 1.3 follows from (2.17). This completes the proof.

Note that, if $k > 1$, Theorem 1.3 does not include a detailed description of the error like (1.17) that leads to a result like (1.18). For $k > 1$, in fact, the X-component of $\tilde{x}_n$ is in general $\beta_2(x_{n-1})$ for $n > 1$. If $k = 1$, then (1.18) holds by (2.15).

Proof of Theorem 1.4. The proof of Theorem 1.4 is based on a simple observation. If $x_{-1} \in W(\rho, \theta, \eta)$, then $x = x_0 \in W(\rho_0, \theta_0, \eta_0)$ and $\theta_0 = \theta\beta_1(x)$ by Theorem 1.1. Hence, $P_X\tilde{x}_0 = \beta_{1+\sigma}(x)$ for any $\sigma \leqslant 1$. Hence, $P_X\tilde{y}_1 = \beta_{2+\sigma}(x)$. This extra reduction in the X-direction is the reason for the special initial iterate. The approach is to find α and an appropriate σ so that the iterates given by Theorem 1.4 remain in $W(\hat{\rho}, \hat{\theta}, \hat{\eta})$ and continue to satisfy $P_X\tilde{x} = \beta_{1+\sigma}(x)$. In order to indicate how the result depends on whether or not $k = 1$, we will indicate error terms that are only present if $k > 1$ by multiplying them by $(k - 1)$.

If we consider the Shamanskii step $s^S(x, y_1)$, we find from (2.12) and (2.15) that

$$-P_N s^S(x, y_1) = \frac{(k/k+1)^k}{k+1} P_N\tilde{y}_1 + \beta_{1+\sigma}(x) + \beta_2(x), \quad \text{and}$$

$$\begin{aligned} -P_X s^S(x, y_1) &= P_X\tilde{y}_1 + (k-1)\beta_2(x) + \beta_{2+\sigma}(x) \\ &= (k-1)\beta_2(x) + \beta_{2+\sigma}(x). \end{aligned} \tag{2.24}$$

Hence, if we define z by

$$z = y_1 + \frac{(k+1)^{k+1}}{k^k} s^S(x, y_1), \tag{2.25}$$

we have

$$\tilde{z} = P_N\beta_{1+\sigma}(x) + P_X(\beta_{2+\sigma}(x) + (k-1)\beta_2(x)). \tag{2.26}$$

When we add the correction term given in Theorem 1.4 to z to obtain x_1, we find that

$$\begin{aligned} \tilde{x}_1 &= \tilde{z} - \left[C\|s^S(x, y_1)\|^\alpha s^S(x, y_1)\right] \\ &= \lambda_{k,\alpha} C P_N\tilde{x} + P_N\beta_{1+\sigma}(x) + (k-1)\beta_2(x) + \beta_{2+\sigma}(x). \end{aligned} \tag{2.27}$$

In (2.27) the quantity $\lambda_{k,\alpha}$ is computed from (2.12),

$$\lambda_{k,\alpha} = \left(k^k/(k+1)^{k+1}\right)^{1+\alpha} \|\tilde{x}\|^\alpha. \tag{2.28}$$

Equation (2.28) implies that $\|\tilde{x}_1\| \leqslant K\|\tilde{x}\|^{1+\alpha}$ for some $K > 0$.

Hence, we may proceed with the iteration if $x_1 \in W(\hat{\rho}, \hat{\theta}, \hat{\eta})$ and $P_X\tilde{x}_1 = \beta_{1+\sigma}(x_1)$. We require that the error in the N-component of $P_N\tilde{x}_1$ in (2.27) be smaller than the main term; this requires that $\alpha < \sigma$. Now note that $P_X\tilde{x}_1 = (k-1)\beta_2(x) + \beta_{2+\sigma}(x)$. Hence, there is $c_6 > 0$ such that

$$\begin{aligned} \theta_1 &= \|P_X\tilde{x}_1\|/\|P_N\tilde{x}_1\| = \left[(k-1)\beta_2(x) + \beta_{2+\sigma}(x)\right]\beta_{-1-\alpha}(x) \\ &< c_6\left[(k-1)\|\tilde{x}\|^{1-\alpha} + \|\tilde{x}\|^{1+\sigma-\alpha}\right]. \end{aligned} \tag{2.29}$$

Hence, if ρ is sufficiently small and $\alpha < 1$, $\theta_1 < \hat{\theta}$. The condition that $P_X\tilde{x}_1 = \beta_{1+\sigma}(x_1)$, which is required for the computation of x_2, is satisfied by $\sigma = \min(1, (1/\alpha) - 1)$ if $k = 1$ and by $\sigma = (1 - \alpha)/(1 + \alpha)$ if $k > 1$. Since $\sigma > \alpha$ is also required, we must have $\alpha < (\sqrt{5} - 1)/2$ if $k = 1$ and $\alpha < \sqrt{2} - 1$ if $k > 1$. Finally, to control the η's, one proceeds as in the proof of Theorem 1.3. This completes the proof.

The remarks made after the statement of Theorem 1.2 about the choice of C and α remain valid in the context of Theorem 1.4. We found $C = 1$ and $\alpha = .6$ to work well.

3. Examples and Numerical Results. In this section we reconsider the two examples from [21]. The first of these is a three-dimensional problem that we use to illustrate the behavior of the parameter η in Theorems 1.3 and 1.4. The equation is

$$F(x_1, x_2, x_3) = \begin{pmatrix} x_1 + x_1x_2 + x_2^2 \\ x_1^2 - 2x_1 + x_2^2 \\ x_1 + x_3^2 \end{pmatrix}. \tag{3.1}$$

Here F satisfies the hypotheses of Theorems 1.1–1.4 with N being the (x_1, x_2)-plane, X the x_1-axis, and $\phi_0 = (0, 0, 1)$. We let $x_0 = (.1, .5, 1)$ and use the l_1-norm. For this example, $x^* = (0, 0, 0)$, and $k = 1$. In Table 1 we illustrate Theorem 1.3 for various

TABLE 1

m	n	ρ_n	θ_n	η_n	$\delta_{m,n}$
2	1	9.58663E–002	1.65350E–004	3.50496E–001	2.79134E–001
2	2	3.59174E–002	1.90626E–006	3.50537E–001	3.39011E–004
2	3	1.34689E–002	8.09709E–009	3.50537E–001	4.46143E–006
2	4	5.05082E–003	1.28023E–011	3.50537E–001	1.99774E–008
2	5	1.89406E–003	7.56883E–015	3.50537E–001	3.22669E–011
2	15	1.04161E–007	6.68713E–024	3.50537E–001	1.66777E–016
2	16	3.90604E–008	1.34245E–024	3.50537E–001	3.17671E–017
3	1	5.38932E–002	1.09717E–004	3.50509E–001	2.50794E–001
3	2	1.64151E–002	6.93196E–007	3.50536E–001	1.01474E–004
3	3	5.00147E–003	1.33089E–009	3.50537E–001	5.67678E–007
3	4	1.52389E–003	7.77993E–013	3.50537E–001	1.05043E–009
3	12	1.13185E–007	2.49151E–023	3.50537E–001	2.45030E–017
3	13	3.44861E–008	2.03567E–024	3.50537E–001	2.19280E–017
10	1	8.64112E–003	3.81655E–005	3.50526E–001	1.30261E–001
10	2	1.20018E–003	3.85352E–008	3.50535E–001	1.05230E–005
10	3	1.66707E–004	5.40404E–012	3.50535E–001	9.34826E–009
10	4	2.31558E–005	1.05259E–016	3.50535E–001	1.30162E–012
10	5	3.21639E–006	1.51138E–022	3.50535E–001	2.69831E–017
10	6	4.46762E–007	9.40199E–024	3.50535E–001	4.03188E–017
10	7	6.20561E–008	6.46542E–024	3.50535E–001	1.36203E–017

values of m. We tabulate the following quantities for various values of n:

$$\rho_n = \|\tilde{x}_n\|, \qquad \theta_n = \|P_X(\tilde{x}_n)\|/\|P_N(\tilde{x}_n)\|,$$
$$\delta_{m,n} = |\,\|\tilde{x}_n\|/\|\tilde{x}_{n-1}\| - r_{k,m}|, \qquad \eta_n = \|(P_N - P_0)\tilde{x}_n\|/\|P_N\tilde{x}_n\|. \tag{3.2}$$

We consider $m = 2, 3$, and 10. Our termination criterion for the Shamanskii method is that if $\|s^N(x_n)\| < 10^{-7}$, then x_{n+1} will be the final iterate. This roughly reflects the fact that for a kth order singularity, the maximum attainable accuracy will be the kth root of machine unit roundoff. For each m the table gives at least the first four and the last two iterates. Recall that each iterate requires one Jacobian evaluation.

Note that θ_n and $\delta_{m,n}$ tend to zero as n increases, as Theorem 1.3 would predict. One can also observe the linear convergence rate and the boundedness of η. Note that η does not tend to zero.

We compared the superlinearly convergent algorithms of Theorems 1.2 and 1.4. We use $C = 1$ and $\alpha = .9$ for the algorithm of Theorem 1.2 and $\alpha = .6$ for that of Theorem 1.4 for two examples for which $k = 1$. We terminated the iteration as follows. As soon as $\|s^N(x_n)\|^{1+\alpha} < 10^{-7}$ we took the following modified step and stopped the iteration. The reasoning here is that the predicted size of $\|\tilde{x}_{n+1}\|$ is less than the square root of machine unit roundoff. For the algorithm of Theorem 1.2 the iteration terminated after eight Jacobian evaluations; for the new algorithm of Theorem 1.4 termination took four Jacobian evaluations.

Our second example is a nonlinear integral equation that arises in radiative transfer [3], [22],

$$H(\mu) = \left(1 - .5\int_0^1 \frac{\mu H(\nu)}{\mu + \nu}\, d\nu\right)^{-1}. \tag{3.3}$$

In (3.3), the unknown function H is continuous on $[0, 1]$. This equation satisfies the hypotheses of the theorems in this paper with $k = 1$, $\dim(N) = 1$, and $N = \text{span}(\mu H(\mu))$, [3], [6], [19]–[22]. Moreover, if the equation is approximated with a quadrature rule that integrates constants exactly, the resulting finite-dimensional problem has similar properties. We approximated the integral with composite 20-point Gauss rules. The sup norm was used. As a termination criterion we use the same rules as for the previous example.

In Table 2 we tabulate the number of Jacobian evaluations required for termination as a function of the dimension of the problem and m. The dimension will be a multiple of 20, as mentioned above. The initial guess was the vector with all

TABLE 2

Number of Jacobian evaluations required for termination as a function of m and problem dimension.

n	$m = 1$	$m = 2$	$m = 3$	$m = 6$	$m = 11$	$m = 21$
20	23	15	12	8	6	5
40	23	15	12	8	6	5
60	23	15	12	8	6	5
80	23	15	12	8	6	5

components equal to one. Values of m ranged from 1 (Newton's method) to 21. As the table will show, increasing m offers diminishing returns as m gets large. Also note that the results do not change as the dimension of the problem increases. The linear problems for the steps were solved by obtaining a QR factorization [11] of the Jacobian, and using that to find the Newton step and each subsequent Shamanskii step.

As in the previous example, we compared the algorithms from Theorems 1.2 and 1.4. We found, for dimensions 20, 40, 60, and 80, that the algorithm of Theorem 1.2 terminated after six derivative evaluations and that of Theorem 1.4 after four.

Department of Mathematics
North Carolina State University
Raleigh, North Carolina 27695-8205

1. R. P. BRENT, "Some efficient algorithms for solving systems of nonlinear equations," *SIAM J. Numer. Anal.*, v. 10, 1973, pp. 327–344.

2. R. C. CAVANAUGH, *Difference Equations and Iterative Processes*, Thesis, Computer Science Dept., Univ. of Maryland, College Park, 1970.

3. S. CHANDRASEKHAR, *Radiative Transfer*, Dover, New York, 1960.

4. D. W. DECKER & C. T. KELLEY, "Newton's method at singular points I," *SIAM J. Numer. Anal.*, v. 17, 1980, pp. 66–70.

5. D. W. DECKER & C. T. KELLEY, "Newton's method at singular points II," *SIAM J. Numer. Anal.*, v. 17, 1980, pp. 465–471.

6. D. W. DECKER & C. T. KELLEY, "Convergence acceleration for Newton's method at singular points," *SIAM J. Numer. Anal.*, v. 19, 1982, pp. 219–229.

7. D. W. DECKER, H. B. KELLER & C. T. KELLEY, "Convergence rates for Newton's method at singular points," *SIAM J. Numer. Anal.*, v. 20, 1983, pp. 296–314.

8. D. W. DECKER & C. T. KELLEY, "Sublinear convergence of the chord method at singular points," *Numer. Math.*, v. 42, 1983, pp. 147–154.

9. D. W. DECKER & C. T. KELLEY, "Expanded convergence domains for Newton's method at nearly singular roots," *SIAM J. Sci. Statist. Comput.*, v. 6, 1985, pp. 951–966.

10. J. E. DENNIS & R. B. SCHNABEL, *Numerical Methods for Nonlinear Equations and Unconstrained Optimization*, Prentice-Hall, Englewood Cliffs, N.J., 1983.

11. G. H. GOLUB & C. F. VAN LOAN, *Matrix Computations*, Johns Hopkins Univ. Press, Baltimore, 1983.

12. A. GRIEWANK, *Analysis and Modification of Newton's Method at Singularities*, Thesis, Australian National University, Canberra, 1980.

13. A. GRIEWANK, "Starlike domains of convergence for Newton's method at singularities," *Numer. Math.*, v. 35, 1980, pp. 95–111.

14. A. GRIEWANK & M. R. OSBORNE, "Newton's method for singular problems when the dimension of the null space is > 1," *SIAM J. Numer. Anal.*, v. 18, 1981, pp. 179–189.

15. A. GRIEWANK & M. R. OSBORNE, "Analysis of Newton's method at irregular singular points," *SIAM J. Numer. Anal.*, v. 20, 1983, pp. 747–773.

16. A. GRIEWANK, "On solving nonlinear equations with simple singularities or nearly singular solutions," *SIAM Rev.*, v. 27, 1985, pp. 537–564.

17. L. V. KANTOROVICH & G. P. AKILOV, *Functional Analysis in Normed Spaces*, Pergamon Press, New York, 1964.

18. H. B. KELLER, "Geometrically isolated nonisolated solutions and their approximation," *SIAM J. Numer. Anal.*, v. 18, 1981, pp. 822–838.

19. C. T. KELLEY, "Solution of the Chandrasekhar H-equation by Newton's method," *J. Math. Phys.*, v. 21, 1980, pp. 1625–1628.

20. C. T. KELLEY, "Approximate methods for the solution of the Chandrasekhar H-equation," *J. Math. Phys.*, v. 23, 1982, pp. 2097–2100.

21. C. T. KELLEY & R. SURESH, "A new acceleration method for Newton's method at singular points," *SIAM J. Numer. Anal.*, v. 20, 1983, pp. 1001–1009.

22. T. W. MULLIKIN, "Some probability distributions for neutron transport in a half space," *J. Appl. Probab.*, v. 5, 1968, pp. 357–374.

23. J. M. ORTEGA & W. C. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

24. L. B. RALL, "Convergence of the Newton process to multiple solutions," *Numer. Math.*, v. 9, 1961, pp. 23–37.

25. L. B. RALL, *Rates of Convergence of Newton's method*, MRC Technical Summary Report no. 1224, Mathematics Research Center, Madison, Wisc., 1972.

26. G. W. REDDIEN, "On Newton's method for singular problems," *SIAM J. Numer. Anal.*, v. 15, 1978, pp. 993–996.

27. G. W. REDDIEN, "Newton's method and high order singularities," *Comput. Math. Appl.*, v. 5, 1980, pp. 79–86.

28. V. E. SHAMANSKII, "A modification of Newton's method," *Ukrain. Mat. Zh.*, v. 19, 1967, pp. 133–138.

29. R. B. SCHNABEL, *Conic Methods for Unconstrained Minimization Problems and Tensor Methods for Nonlinear Equations*, University of Colorado Technical Report, Dept. of Computer Science, University of Colorado, Boulder, Colorado, 1982.

30. R. B. SCHNABEL & P. D. FRANK, *Tensor Methods for Nonlinear Equations*, University of Colorado Report, Dept. of Computer Science, University of Colorado, Boulder, Colorado, 1983.

31. E. SCHRÖDER, "Ueber unendlich viele Algorithmen zur Auflösung der Gleichungen," *Math. Ann.*, v. 2, 1870, pp. 317–365.

32. J. F. TRAUB, *Iterative Methods for the Solution of Equations*, Prentice-Hall, Englewood Cliffs, N.J., 1964.