A SUFFICIENTLY EXACT INEXACT NEWTON STEP BASED ON REUSING MATRIX INFORMATION

Anders FORSGREN*

Technical Report TRITA-MAT-2009-OS7
Department of Mathematics
Royal Institute of Technology
November 2009

Abstract

Newton's method is a classical method for solving a nonlinear equation $F(z) = 0$. We derive inexact Newton steps that lead to an inexact Newton method, applicable near a solution. The method is based on solving for a particular $F'(z_{k'})$ during p consecutive iterations $k = k', k'+1, \dots, k'+p-1$. One such p-cycle requires $2^p - 1$ solves with the matrix $F'(z_{k'})$. If matrix factorization is used, it is typically more computationally expensive to factorize than to solve, and we envisage that the proposed inexact method would be useful as the iterates converge. The inexact method is shown to be p-step convergent with Q-factor 2^p under standard assumptions where Newton's method has quadratic rate of convergence. The method is thus sufficiently exact in the sense that it mimics the convergence rate of Newton's method. It may interpreted as a way of performing iterative refinement by solving the subproblem $F(z_k) + F'(z_k)d = 0$ sufficiently exactly by a simplified Newton method. The method is contrasted to a simplified Newton method, where it is known that a cycle of $2^p - 1$ iterations gives the same type of convergence. We present some numerical results and also discuss how this method might be used in the context of interior methods for linear programming.

1. Introduction

This paper concerns solving a nonlinear equation

$$F(z) = 0, \tag{1.1}$$

where $F : I\!R^n \to I\!R^n$ is continuously differentiable and F' is Lipschitz continuous. The "standard" method for solving (1.1) is Newton's method, which for each iterate z_k finds a zero of the linearization of F at z_k. This means that iteration k consists of solving the equation

$$F(z_k) + F'(z_k)d = 0, \tag{1.2}$$

so that $d_k = -F'(z_k)^{-1}F(z_k)$, and then $z_{k+1} = z_k + d_k$. As the name sugggests, this is a classical method, see, e.g., Ortega and Rheinboldt [13] for a detailed treatment.

*Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden (andersf@kth.se). Research supported by the Swedish Research Council (VR).

Our concern is the asymptotic rate of convergence of Newton's method. Throughout, we consider a solution z^* to (1.1) at which the following assumption holds.

Assumption 1.1. The function $F : I\!R^n \to I\!R^n$ is continuously differentiable and z^* is a solution to $F(z) = 0$ at which $F'(z^*)$ is nonsingular. There is an open neighborhood $N_1(z^*)$ of z^* and a constant L such that $\|F'(u) - F'(v)\| \leq L\|u - v\|$ for $u \in N_1(z^*)$, $v \in N_1(z^*)$.

Given Assumption 1.1, it is well known that Newton's method converges locally with a Q-quadratic rate.

Theorem 1.1. (Convergence of Newton's method) *Assume that Assumption 1.1 holds. Then, there exists a positive constant C such that if $z_0 \in N_1(z^*)$ and $C\|z_0 - z^*\| \leq \beta < 1$, then the Newton iterates of (1.2) satisfy*

$$\|z_k - z^*\| \leq C\|z_{k-1} - z^*\|^2 \leq \beta^{2^k-1}\|z_0 - z^*\|, \quad k = 1, \ldots. \tag{1.3}$$

The work involved in Newton's method is to solve the linear equation of (1.2) for $k = 1, 2, \ldots$. Various ways of solving (1.2) inexactly have been suggested, starting with the work of Dembo, Eisenstat and Steihaug [5]. Many such methods are based on iterative methods for solving the linear equation (1.2), see, e.g., Dembo and Steihaug [6], Ypma [18], Cai and Keyes [3]. One such method that combines a factorization with preconditioned conjugate-gradient iterations has been proposed by Deng and Wang [7]. Our concern is the case when (1.2) is solved by a method for which it is less costly to solve with $F'(z_0)$ several times than to solve with $F'(z_k)$ for each iteration k. One such situation is when a factorization of $F'(z_0)$ is used. Then it is typically more expensive to factorize than to solve with the factors. Making use of $F'(z_0)$ at consecutive iterations has been considered by several authors, e.g., Brent [1] and Moré and Cosnard [11].

A "natural" initial approximation is to replace $F'(z_k)$ by $F'(z_0)$ in (1.2) for r steps and thus solve

$$F'(z_o)d_k = -F(z_k), \quad k = 0, \ldots, r-1. \tag{1.4}$$

For $k = 0$, this is a Newton iteration, and for $k \geq 1$, such a d_k is referred to as a *simplified Newton step* in the classical book of Ortega and Rheinboldt [13, p. 187]. The combination of Newton step and simplified Newton steps are described in Traub [16, Section 8.5] and Ortega and Rheinboldt [13, p. 315]. They show that the r steps of (1.4) give

$$\|z_r - z^*\| \leq M_r\|z_0 - z^*\|^{r+1} \tag{1.5}$$

for some constant M_r, in a neighborhood of z^*. In general, M_r depends on r.

With Newton's method, (1.3) implies that p steps would give

$$\|z_p - z^*\| \leq C\|z_{p-1} - z^*\|^2 \leq \ldots \leq C^{2^p-1}\|z_0 - z^*\|^{2^p}. \tag{1.6}$$

In particular, by letting $r = 2^p - 1$ in the scheme with simplified Newton steps of (1.4), it follows from (1.5) that

$$\|z_{2^p-1} - z^*\| \leq M_{2^p-1}\|z_0 - z^*\|^{2^p}, \tag{1.7}$$

for some constant M_{2^p-1}, which is comparable to the bound (1.6) given by Newton's method. Hence, we may view a cycle of 2^p-1 simplified Newton steps as a substitute for p Newton steps. Such a cycle has $2^p - 1$ iterations and requires $2^p - 1$ solves with $F'(z_0)$. Each iteration requires the evaluation of $F(z_k)$. Tapia et al. [15] give an approach based on simplified Newton iterations of the form (1.4) for $r = 2$ and apply it to interior methods for linear programming. They also discuss varying r but do not analyze this situation.

In a sense, we now have two extreme approaches for computing a step by solving the Newton iteration (1.2): (i) Newton's method in which d_k solves (1.2) exactly, and (ii) the simplified Newton method, in which we may view d_k of (1.4) as one step towards solving (1.2) by a simplified Newton method. We propose an alternative scheme, based on taking m_k simplified Newton steps towards solving (1.2) at iteration k.

At iteration k, we initialize with $d_{k,0} = 0$. Then, for $i = 0, \ldots, m_k - 1$, let $p_{k,i}$ and $d_{k,i+1}$ be given by

$$F'(z_0)p_{k,i} = -(F(z_k) + F'(z_k)d_{k,i}), \tag{1.8a}$$

$$d_{k,i+1} = d_{k,i} + p_{k,i}. \tag{1.8b}$$

Finally, let $d_k = d_{k,m_k}$.

Note that $d_{0,1}$ is the Newton step at iteration 0 and $d_{k,1}$ is the simplified Newton step at iteration k. In a neighborhood of the solution, it holds that $\lim_{m_k\to\infty} d_{k,m_k} = -F'(z_k)^{-1}F(z_k)$. Hence, by increasing m_k, the level of exactness of the approximation of the Newton step is increased. This is made precise in this paper. In particular, we show in Theorem 3.1 that $m_k = 2^k$, $k = 0, \ldots, p-1$, suffices for giving

$$\|z_p - z^*\| \leq M^{2^p-1}\|z_0 - z^*\|^{2^p}, \tag{1.9}$$

where M is given by (2.1). Such a cycle has p iterations and requires $2^p - 1$ solves with $F'(z_0)$. Each iteration requires the evaluation of $F(z_k)$ and $F'(z_k)$, just as Newton's method. The bound (1.9) has the same structure as the corresponding bound for Newton's method given by (1.6), but the constant C of (1.3) is weakened to M of (2.1).

Throughout the paper, the Euclidean norm is used. For a square matrix M, we denote the spectral radius of M by $\rho(M)$.

2. An expression for the Newton step

Our analysis concerns Newton iterates and approximate Newton iterates in a neighborhood of the solution. In this section we express the Newton step $-F'(z_k)^{-1}F(z_k)$ as the limit of the simplified Newton steps d_{k,m_k} of (1.8) when m_k tends to infinity. The approach of (1.8) which we suggest falls in the category of *generalized linear* *equations*, see, e.g., Ortega and Rheinboldt [13, Sections 7.4 and 10.3]. It may be viewed as based on splitting $F'(z_k)$ as $F'(z_k) = F'(z_0) + (F'(z_k) - F'(z_0))$, and then solve (1.2) inexactly by a simplified Newton method. This gives an expression in increasing powers of the matrix $I - F'(z_0)^{-1}F'(z_k)$. The following definition gives a neighborhood in which our analysis applies.

Definition 2.1. (Neighborhood of z^*) The neighborhood $N(z^*)$ is defined as all $z \in N_1(z^*)$, with $N_1(z^*)$ given by Assumption 1.1, such that in addition $\|F'(z)^{-1}\| \leq 2\|F'(z^*)^{-1}\|$ and $M\|z - z^*\| \leq \beta < 1$, where

$$M = C + 8L\|F'(z^*)^{-1}\|, \tag{2.1}$$

with C given by Theorem 1.1.

The neighborhood $N(z^*)$ of Definition 2.1 is well defined as $F'(z)$ is continuous on $N_1(z^*)$. For our analysis to apply, $F'(z_k)$ and $F'(z_0)$ must be sufficiently close in the sense that $\rho(I - F'(z_0)^{-1}F'(z_k)) < 1/2$. The following lemma shows that this is the case if $z_0 \in N(z^*)$ and $\|z_k - z^*\| \leq \|z_0 - z^*\|$.

Lemma 2.1. *If z_0 and z_k are such that $z_0 \in N(z^*)$, with $N(z^*)$ given by Definition 2.1, and $\|z_k - z^*\| \leq \|z_0 - z^*\|$, then $\rho(I - F'(z_0)^{-1}F'(z_k)) < 1/2$.*

Proof. If $z_0 \in N(z^*)$ and $\|z_k - z^*\| \leq \|z_0 - z^*\|$, it follows that $z_k \in N(z^*)$ with

$$\begin{aligned}
\rho(I - F'(z_0)^{-1}F'(z_k)) &\leq \|F'(z_0)^{-1}(F'(z_k) - F'(z_0))\| \\
&\leq \|F'(z_0)^{-1}\|\|F'(z_k) - F'(z_0)\| \\
&\leq 4\|F'(z^*)^{-1}\|\|F'(z_0) - F'(z^*)\| \\
&\leq 4L\|F'(z^*)^{-1}\|\|z_0 - z^*\| < \frac{1}{2},
\end{aligned}$$

completing the proof. ∎

The following lemma gives an expression for d_{k,m_k} and also shows that d_{k,m_k} converges to $-F(z_k)^{-1}F(z_k)$ as a geometric series. Note that $d_{0,1}$ is the Newton step at iteration 0, $d_{k,1}$ is the simplified Newton step at iteration k and $\lim_{m_k\to\infty} d_{k,m_k}$ is the Newton step at iteration k. Hence, by varying m_k, the level of exactness of the approximation of the Newton step is prescribed.

Lemma 2.2. *Assume that $\rho(I - F'(z_0)^{-1}F'(z_k)) < 1$. Let $m_k \geq 1$. For $d_{k,0} = 0$, let $p_{k,i}$ and $d_{k,i+1}$, $i = 0, 1, \ldots, m_k - 1$, be given by (1.8). Then,*

$$d_{k,m_k} = \sum_{i=0}^{m_k-1} p_{k,i}, \tag{2.2}$$

with

$$p_{k,i} = -(I - F'(z_0)^{-1}F'(z_k))^i F'(z_0)^{-1}F(z_k), \quad i = 0, 1, \ldots, m_k - 1. \tag{2.3}$$

In particular, $\lim_{m_k\to\infty} d_{k,m_k} = -F'(z_k)^{-1}F(z_k)$ *with*

$$d_{k,m_k} = -(I - (I - F'(z_0)^{-1}F'(z_k))^{m_k})F'(z_k)^{-1}F(z_k). \tag{2.4}$$

Proof. Letting $i = m_k - 1$ in (1.8), (1.8a) gives $p_{k,m_k-1} = d_{k,m_k} - d_{k,m_k-1}$, which inserted into (1.8b) yields

$$\begin{aligned} d_{k,m_k} &= -F'(z_0)^{-1}F(z_k) + (I - F'(z_0)^{-1}F'(z_k))d_{k,m_k-1} \\ &= \ldots = -\sum_{i=0}^{m_k-1} (I - F'(z_0)^{-1}F'(z_k))^i F'(z_0)^{-1}F(z_k), \end{aligned} \tag{2.5}$$

so that (2.2) holds with $p_{k,i}$ given by (2.3). As $\rho(I - F'(z_0)^{-1}F'(z_k)) < 1$, it holds that

$$\sum_{i=0}^{\infty}(I - F'(z_0)^{-1}F'(z_k))^i = (I - (I - F'(z_0)^{-1}F'(z_k)))^{-1} = F'(z_k)^{-1}F'(z_0), \tag{2.6}$$

see, e.g., Saad [14, Theorem 4.1]. Consequently, (2.5) and (2.6) imply that $d_{k,\infty}$ given by $d_{k,\infty} = \lim_{m_k\to\infty} d_{k,m_k}$ is well defined with

$$d_{k,\infty} = -\sum_{i=0}^{\infty}(I - F'(z_0)^{-1}F'(z_k))^i F'(z_0)^{-1}F(z_k) = -F'(z_k)^{-1}F(z_k). \tag{2.7}$$

Finally, a combination of (2.5) and (2.7) gives

$$d_{k,m_k} - d_{k,\infty} = -\sum_{j=m_k}^{\infty} p_{k,j} = -(I - F'(z_0)^{-1}F'(z_k))^{m_k} d_{k,\infty}, \tag{2.8}$$

so that (2.4) follows from (2.7) and (2.8), completing the proof. ∎

Note that (2.3) of Lemma 2.2 gives an alternative way of computing $p_{k,i}$, $i = 0, 1 \ldots, m_k - 1$, from

$$\begin{aligned} &F'(z_0)p_{k,0} = -F(z_k), \quad \text{and} \\ &F'(z_0)p_{k,i} = -(F'(z_k) - F'(z_0))p_{k,i-1}, \quad i = 1, 2, \ldots, m_k - 1. \end{aligned}$$

Lemma 2.2 shows that d_{k,m_k} may be interpreted in two ways: (i) as the m_kth iteration of a simplified Newton method applied to the linear equation $F'(z_k)d + F(z_k) = 0$, where $F'(z_0)$ is used instead of $F'(z_k)$, and (ii) as a linear combination of vectors formed by increasing powers of the matrix $I - F'(z_0)^{-1}F'(z_k)$, which by our assumption will be small in norm. This is the basis for the inexact Newton method which we propose.

3. A sufficiently exact inexact Newton step

Lemma 2.2 gives an expression for $-F'(z_k)^{-1}F(z_k)$ in terms of increasing powers of the matrix $I - F'(z_0)^{-1}F'(z_k)$. As we assume that $I - F'(z_0)^{-1}F'(z_k)$ is small, an inexact Newton direction is obtained by including only the leading m_k directions $p_{k,i}$, $i = 0, \ldots, m_k - 1$. Our main result is to show that it suffices to let $m_k = 2^k$ to match the quadratic rate of convergence enjoyed by Newton's method.

We first characterize the difference between d_{k,m_k} and $-F(z_k)^{-1}F(z_k)$.

Lemma 3.1. *Let* z_0 *and* z_k *be such that* $z_0 \in N(z^*)$ *of Definition 2.1 and* $\|z_k - z^*\| \leq \|z_0 - z^*\|$. *Let* $m_k \geq 1$. *For* $d_{k,0} = 0$, *let* $p_{k,i}$ *and* $d_{k,i+1}$, $i = 0, 1, \ldots, m_k - 1$, *be given by (1.8). Then,*

$$\|d_{k,m_k} + F(z_k)^{-1}F(z_k)\| \leq 2\left(2L\|F'(z^*)^{-1}\|\|z_k - z_0\|\right)^{m_k}\|z_k - z^*\|$$
$$\leq 2\left(4L\|F'(z^*)^{-1}\|\|z_0 - z^*\|\right)^{m_k}\|z_k - z^*\|. \quad (3.1)$$

Proof. Lemma 2.1 shows that $\rho(I - F'(z_0)^{-1}F'(z_k)) < 1$. Hence, Lemma 2.2 gives

$$\begin{aligned} d_{k,m_k} + F(z_k)^{-1}F(z_k) &= -(I - F'(z_0)^{-1}F'(z_k))^{m_k}F'(z_k)^{-1}F(z_k) \\ &= (-F'(z_0)^{-1}(F'(z_k) - F'(z_0)))^{m_k}(z_k - F'(z_k)^{-1}F(z_k) - z^*) \\ &- (-F'(z_0)^{-1}(F'(z_k) - F'(z_0)))^{m_k}(z_k - z^*). \end{aligned}$$

Taking into account that $z_0 \in N(z^*)$ and $z_k \in N(z^*)$, it follows that $\|F'(z_0)^{-1}\| \leq 2\|F'(z^*)^{-1}\|$, $\|F'(z_k) - F'(z_0)\| \leq L\|z_k - z_0\|$ and $C\|z_k - z^*\| \leq 1$. Hence,

$$\begin{aligned} \|d_{k,m_k} + F(z_k)^{-1}F(z_k)\| &\leq (2L\|F'(z^*)^{-1}\|\|z_k - z_0\|)^{m_k}(C\|z_k - z^*\|^2 + \|z_k - z^*\|) \\ &\leq 2(2L\|F'(z^*)^{-1}\|\|z_k - z_0\|)^{m_k}\|z_k - z^*\|, \end{aligned}$$

proving the first inequality of (3.1). As it is assumed that $\|z_k - z^*\| \leq \|z_0 - z^*\|$, it holds that $\|z_k - z_0\| \leq 2\|z_0 - z^*\|$, so that the second equality of (3.1) follows from the first one. ∎

Lemma 3.1 suggests that if $\|z_k - z^*\| \approx \|z_0 - z^*\|^{2^k}$, then letting $m_k = 2^k$, i.e., utilizing $d_{k,2^k}$ instead of $-F'(z_k) \inf F(z_k)$, would suffice to make $\|z_{k+1} - z^*\| \approx \|z_0 - z^*\|^{2^{k+1}}$, i.e., to mimic the rate of convergence of Newton's method. This is made precise in the following theorem.

Theorem 3.1. *Let* p *be a positive integer and let* z_0 *be such that* $z_0 \in N(z^*)$, *with* $N(z^*)$ *given by Definition 2.1. Let* $z_{k+1} = z_k + d_{k,2^k}$, $k = 0, \ldots, p-1$, *with* $d_{k,2^k}$ *given by (1.8) for* $m_k = 2^k$. *Then,*

$$\|z_k - z^*\| \leq M^{2^k - 1}\|z_0 - z^*\|^{2^k} \leq \beta^{2^k - 1}\|z_0 - z^*\|, \quad k = 1, \ldots, p, \quad (3.2)$$

where M *and* β *are given by Lemma 2.1. The directions that define a* p*-cycle, i.e.,* $d_{k,2^k}$, $k = 0, \ldots, p-1$, *may be computed by in total* $2^p - 1$ *solves with the matrix* $F'(z_0)$.

Proof. The proof is by induction. As the first step is a Newton step and $z_0 \in N(z^*)$, (3.2) holds for $k = 1$. Assume that (3.2) holds for $k = k'$. As $\beta < 1$, we obtain $\|z_{k'} - z_0\| \leq \|z_0 - z^*\|$ so that $z_{k'} \in N(z^*)$. Then, Lemma 2.1 shows that $\rho(I - F'(z_0)^{-1}F'(z_{k'})) < 1$. As $z_{k'} \in N(z^*) \subseteq N_1(z^*)$, Theorem 1.1 ensures that

$$\|z_{k'} - F(z_{k'})^{-1}F(z_{k'}) - z^*\| \leq C\|z_{k'} - z^*\|^2. \quad (3.3)$$

Hence, (3.3) in conjunction with Lemma 3.1 give

$$\begin{aligned}\|z_{k'} + d_{k',2^{k'}} - z^*\| &\le \|z_{k'} - F(z_{k'})^{-1}F(z_{k'}) - z^*\| + \|d_{k',2^{k'}} + F(z_{k'})^{-1}F(z_{k'})\| \\ &\le C\|z_{k'} - z^*\|^2 \\ &\quad + 2\left(4L\|F'(z^*)^{-1}\|\|z_0 - z^*\|\right)^{2^{k'}} \|z_{k'} - z^*\|. \end{aligned} \tag{3.4}$$

To simplify the notation, let $A = 4L\|F'(z^*)^{-1}\|$. Then, with M given by (2.1), $M = C + 2A$, so that $A \le M$ and $C \le M$. By our induction hypothesis, (3.2) gives

$$\|z_{k'} - z^*\| \le M^{2^{k'}-1}\|z_0 - z^*\|^{2^{k'}}. \tag{3.5}$$

Letting $z_{k'+1} = z_{k'} + d_{k',2^{k'}}$, insertion of (3.5) in (3.4) gives

$$\begin{aligned}\|z_{k'+1} - z^*\| &\le C\|z_{k'} - z^*\|^2 + 2\left(A\|z_0 - z^*\|\right)^{2^{k'}} \|z_{k'} - z^*\| \\ &\le C\left(M^{2^{k'}-1}\|z_0 - z^*\|^{2^{k'}}\right)^2 + 2A^{2^{k'}}\|z_0 - z^*\|^{2^{k'}} M^{2^{k'}-1}\|z_0 - z^*\|^{2^{k'}} \\ &\le C\left(M^{2^{k'}-1}\|z_0 - z^*\|^{2^{k'}}\right)^2 + 2A\left(M^{2^{k'}-1}\|z_0 - z^*\|^{2^{k'}}\right)^2 \\ &= (C + 2A)\left(M^{2^{k'}-1}\|z_0 - z^*\|^{2^{k'}}\right)^2 \\ &= M\left(M^{2^{k'}-1}\|z_0 - z^*\|^{2^{k'}}\right)^2 = M^{2^{k'+1}-1}\|z_0 - z^*\|^{2^{k'+1}},\end{aligned}$$

proving the first inequality of (3.2) for $k = k' + 1$. The second inequality follows from the fact that $M\|z_0 - z^*\| \le 1$, as

$$\begin{aligned}\|z_{k'+1} - z^*\| &\le M^{2^{k'+1}-1}\|z_0 - z^*\|^{2^{k'+1}} = (M\|z_0 - z^*\|)^{2^{k'+1}-1}\|z_0 - z^*\| \\ &\le \beta^{2^{k'+1}-1}\|z_0 - z^*\|,\end{aligned}$$

completing the proof of (3.2) for $k = k' + 1$.

As for the number of solves with $F'(z_0)$, (1.8) shows that $d_{k,2^k}$ can be computed by 2^k solves with $F'(z_0)$. Hence, in total we get $\sum_{k=0}^{p-1} 2^k$ solves in one p-cycle, which equals $2^p - 1$. ∎

We thus have $\|z_p - z^*\| \le M^{2^p-1}\|z_0 - z^*\|^{2^p}$ after p iterations. This means that if p steps are made, we expect $\|z_k - z^*\| \approx M\|z_{k-1} - z^*\|^2$, $k = 1, \ldots, p$, i.e., similar to Newton's method. The constant C of Newton's method has been replaced by M of (2.1), so in practice it cannot be expected that the the inexact method achieves the same constant as Newton's method in the quadratic rate of convergence.

By repeating the p-cycles, we obtain a method which is p-step convergent with Q-factor 2^p, as stated in the following corollary.

Corollary 3.1. *Let p be a positive integer and let z_0 and z_k be such that $z_0 \in N(z^*)$, with $N(z^*)$ given by Definition 2.1. Given z_0, let a p-cycle of an inexact Newton method be defined by Theorem 3.1. Let the inexact Newton method be defined by repeating these p-cycles. Then,*

$$\|z_{jp} - z^*\| \le M^{2^p-1}\|z_{(j-1)p} - z^*\|^{2^p} \le \beta^{2^p-1}\|z_{(j-1)p} - z^*\|, \quad j = 1, 2, \dots,$$

where M and β are given by Lemma 2.1.

The p-step inexact Newton method is described in Algorithm 3.1. To emphasize the fact that we make repeated use of matrix factors, the method is described with an LU factorization. This can of course be substituted for any suitable factorization.

Another comment worth making is that we have stated the algorithm for $m_k = 2^k$, which is the smallest value of m_k for which quadratic rate of convergence can be proved along the lines of the proof of Theorem 3.1. In this case, both constants A and M in the proof are of order one. A higher value of m_k gives a closer approximation of the Newton step.

Finally, for a given k, we expect $p_{k,i}$, $i = 0, 1, \dots, m_k - 1$, to be of decreasing norm, as they correspond to increasing powers of the matrix $F'(z_0)^{-1}(F'(z_k) - F'(z_0))$. Hence, as the computations are made in finite precision, one may omit computing $p_{k,i}$, $i = i' + 1, , \dots, m_k - 1$ if $p_{k,i'}$ is "sufficiently small", for example if the norm is smaller than the optimality tolerance.

$\frac{w_F}{w_S}$	p^*	$r(p^*)$
10	3	0.52
100	5	0.26
1000	8	0.16

Table 1: Optimal value of p as a function of w_F/w_S.

The point of the proposed method is for it to be applied when the work of a solve is less than the work of a factorize. If the work of a factorize is denoted by w_F and the work of a solve is denoted by w_S, the total amount of work in a p-cycle is given by $w_F + (2^p - 1)w_S$, where we assume that $w_F > w_S$. The work of p Newton iterations is $p(w_F + w_S)$. To get a feeling of what vales of p that might be of interest, we assume that the proposed method obtains the same rate of quadratic convergence as Newton's method. Given this assumption, the ratio between the two methods is given by $r(p)$, with

$$r(p) = \frac{w_F + (2^p - 1)w_S}{p(w_F + w_S)}.$$

As $1/p$ is strictly convex for $p > 0$ and $2^p/p$ is logarithmically convex for $p > 0$, we conclude that $r(p)$ is strictly convex for $p > 0$. In addition, $\lim_{p\to 0_+} r(p) = \infty$ and $\lim_{p\to\infty} r(p) = \infty$. Hence, there is a unique number $\widehat{p}$ such that $r'(\widehat{p}) =$

0. By rounding $\widehat{p}$ up and down, we find the optimal integer p^*. Assuming LU-decomposition and dense arithmetic, we have $w_F = O(n^3)$ and $w_S = O(n^2)$, i.e, the ratio $w_F/w_S = n$. We do not expect this ratio for sparse problem, but include the optimal ratio in Table 3 to get a feeling that rather small values of p^* are to be expected. Note that in general $r(p^*)$ is expected to overemphasize the efficiency of the inexact Newton method, as we do not expect the inexact Newton method to have the same constant for the expected quadratic rate of convergence as Newton's method. Also, when considering the situation when the iterates converge it might be of interest to see how many more iterations are expected, and then see if more factorizations should be carried out. For example, if $\|F(z_0)\| \approx 10^{-2}$ one might expect three iterations to obtain $\|F(z_3)\| \approx 10^{-16}$. If $F'(z_0)$ is used throughout, we would expect seven solves with the inexact Newton method. This should then be compared to the cost of factorizing.

Also note that by (1.7), the simplified Newton method, i.e, $m_k = 1$, is expected to give the same convergence behavior if one compares the number of solves. The inexact Newton method, with $m_k = 2^k$, makes $2^k - 1$ solves for k iterations, whereas the simplified Newton method, with $m_k = 1$, makes $2^k - 1$ solves for $2^k - 1$ iterations. Hence, if $k = l$ for $m_k = 2^k$, it should be compared to $k = 2^l - 1$ for $m_k = 1$, and consequently $p = p'$ for $m_k = 2^k$ should be compared to $p = 2^{p'} - 1$ for $m_k = 1$. We can think of $2^k - 1$ as the cost in number of solves with $F'(z_0)$ to obtain quadratic rate of convergence up to iteration k, either with the inexact Newton method, $m_k = 2^k$, or the simplified Newton method, $m_k = 1$.

4. Test problems and numerical results

We illustrate our results on two test problems. Both problems have a unique solution z^*, where $F'(z^*)$ is nonsingular.

Problem 4.1. This is a one-dimensional problem with $F(z) = 2 - 1/z$ for which $z^* = 1/2$. The initial point is given by $z_0 = 0.49$.

What is nice with Problem 4.1 is that Newton's method has exact quadratic rate of convergence in a neighborhood of z^* with factor 2.

Problem 4.2. This is a primal-dual nonlinear equation arising in interior methods for linear programming. Given $A \in I\!R^{m\times\bar{n}}$ of rank m, $b \in I\!R^m$, $c \in I\!R^{\bar{n}}$ and $\mu \in I\!R_+$, we let $z = (x, y, s)$ with $x \in I\!R^{\bar{n}}_{++}$, $y \in I\!R^m$ and $s \in I\!R^{\bar{n}}_{++}$, and define

$$F(z) = \begin{pmatrix} Ax - b \\ A^T y + s - c \\ XSe - \mu e \end{pmatrix} \quad \text{with} \quad F'(z) = \begin{pmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S & 0 & X \end{pmatrix},$$

where $X = \operatorname{diag}(x)$, $S = \operatorname{diag}(s)$ and $e = (1, 1, \ldots, 1)^T$. Our test data is from problem Blend (put in standard form) in the Netlib test set [12] with $\mu = 1$. Here $m = 74$ and $\bar{n} = 114$, which gives $z \in I\!R^n$ for $n = m + 2\bar{n} = 302$. The initial point is given by "`z0=zstar+0.01*rand(size(zstar))`". The optimal solution z^* has been computed by Newton's method.

Problem 4.2 is interesting since we believe that interior methods for linear programming is an area where the results of this paper would be useful as the work involved is exactly in solving the primal-dual nonlinear equations, and the cost of evaluating F and F' is negligible. This will be briefly discussed in Section 5. The initial point is chosen such that the x and s components of z are strictly positive, and that we are in the region of quadratic convergence of Newton's method.

$p = 1$ (Newton's method)				
It	#f	#s	zdiff	zratio
0			1.00e-02	
1	1	1	2.00e-04	2.00e+00
2	2	2	8.00e-08	2.00e+00
3	3	3	1.28e-14	1.99e+00

$p = 2$, $m_k = 2^k$				
It	#f	#s	zdiff	zratio
0			1.00e-02	
1	1	1	2.00e-04	2.00e+00
2	1	3	3.81e-07	9.54e+00
3	2	4	2.91e-13	2.00e+00
4	2	6	0.00e+00	0.00e+00

$p = 3$, $m_k = 2^k$				
It	#f	#s	zdiff	zratio
0			1.00e-02	
1	1	1	2.00e-04	2.00e+00
2	1	3	3.81e-07	9.54e+00
3	1	7	1.23e-12	8.45e+00
4	2	8	0.00e+00	0.00e+00

$p = 4$, $m_k = 2^k$				
It	#f	#s	zdiff	zratio
0			1.00e-02	
1	1	1	2.00e-04	2.00e+00
2	1	3	3.81e-07	9.54e+00
3	1	7	1.23e-12	8.45e+00
4	1	15	0.00e+00	0.00e+00

Table 2: Results of Algorithm 3.1 applied to Problem 4.1 with $m_k = 2^k$.

$p = 1$ (Newton's method)				
It	#f	#s	zdiff	zratio
0			1.01e-01	
1	1	1	2.03e-03	2.00e-01
2	2	2	1.19e-08	2.88e-03
3	3	3	6.01e-13	4.26e+03

$p = 2$, $m_k = 2^k$				
It	#f	#s	zdiff	zratio
0			1.01e-01	
1	1	1	2.03e-03	2.00e-01
2	1	3	1.17e-05	2.83e+00
3	2	4	3.74e-13	2.73e-03

$p = 3$, $m_k = 2^k$				
It	#f	#s	zdiff	zratio
0			1.01e-01	
1	1	1	2.03e-03	2.00e-01
2	1	3	1.17e-05	2.83e+00
3	1	7	1.78e-09	1.30e+01
4	2	8	4.68e-13	1.47e+05

$p = 4$, $m_k = 2^k$				
It	#f	#s	zdiff	zratio
0			1.01e-01	
1	1	1	2.03e-03	2.00e-01
2	1	3	1.17e-05	2.83e+00
3	1	7	1.78e-09	1.30e+01
4	1	15	4.68e-13	1.47e+05

Table 3: Results of Algorithm 3.1 applied to Problem 4.2 with $m_k = 2^k$.

The inexact Newton method of Algorithm 3.1 has been applied to Problem 4.1 and Problem 4.2 respectively. The method has been implemented in Matlab and run on an Apple Macbook Pro with Mac OS 10.5.8. Throughout, `tol` $= 10^{-12}$ has been used. The headings in the tables are as follows:

$p = 1$ (Newton's method)				
It	#f	#s	zdiff	zratio
0			1.00e-02	
1	1	1	2.00e-04	2.00e+00
2	2	2	8.00e-08	2.00e+00
3	3	3	1.28e-14	1.99e+00

$p = 3$, $m_k = 1$ (Simplified Newton)				
It	#f	#s	zdiff	zratio
0			1.00e-02	
1	1	1	2.00e-04	2.00e+00
3	1	3	3.10e-07	7.76e+00
4	2	4	1.93e-13	2.00e+00

$p = 7$, $m_k = 1$ (Simplified Newton)				
It	#f	#s	zdiff	zratio
0			1.00e-02	
1	1	1	2.00e-04	2.00e+00
3	1	3	3.10e-07	7.76e+00
7	1	7	7.63e-13	7.92e+00
8	2	8	0.00e+00	0.00e+00

$p = 15$, $m_k = 1$ (Simplified Newton)				
It	#f	#s	zdiff	zratio
0			1.00e-02	
1	1	1	2.00e-04	2.00e+00
3	1	3	3.10e-07	7.76e+00
7	1	7	7.63e-13	7.92e+00
8	1	8	3.02e-14	

Table 4: Results of Algorithm 3.1 applied to Problem 4.1 with $m_k = 1$.

$p = 1$ (Newton's method)				
It	#f	#s	zdiff	zratio
0			1.01e-01	
1	1	1	2.03e-03	2.00e-01
2	2	2	1.19e-08	2.88e-03
3	3	3	6.01e-13	4.26e+03

$p = 3$, $m_k = 1$ (Simplified Newton)				
It	#f	#s	zdiff	zratio
0			1.01e-01	
1	1	1	2.03e-03	2.00e-01
3	1	3	1.17e-05	2.83e+00
4	2	4	1.27e-13	9.29e-04

$p = 7$, $m_k = 1$ (Simplified Newton)				
It	#f	#s	zdiff	zratio
0			1.01e-01	
1	1	1	2.03e-03	2.00e-01
3	1	3	1.17e-05	2.83e+00
7	1	7	1.78e-09	1.30e+01
8	2	8	3.29e-13	1.04e+05

$p = 15$, $m_k = 1$ (Simplified Newton)				
It	#f	#s	zdiff	zratio
0			1.01e-01	
1	1	1	2.03e-03	2.00e-01
3	1	3	1.17e-05	2.83e+00
7	1	7	1.78e-09	1.30e+01
9	1	9	2.33e-11	

Table 5: Results of Algorithm 3.1 applied to Problem 4.2 with $m_k = 1$.

It	Iteration number.
#f	Total number of factorizations after iteration k.
#s	Total number of solves after iteration k.
zdiff	Difference from optimality, $\|z_k - z^*\|$.
zratio	Rate of quadratic convergence, $\|z_k - z^*\|/(\|z_{k-1} - z^*\|^2)$[1]

Results with $m_k = 2^k$ are presented in Table 2 and Table 3 for Problem 4.1 and Problem 4.2 respectively. The inexact Newton method shows quadratic rate of convergence in practice, but not with quite the same constant as Newton's method.

Finally, we have run the simplified Newton method on both problems by letting $m_k = 1$. The results for iteration $2^l - 1$ are presented, for $l = 0, 1, \ldots$, in Tables 4 and 5 respectively. As was discussed at the end of Section 3, these are the iterates where the number of solves is comparable to the number of solves of the proposed inexact Newton method. Also, the number `zratio` is modified accordingly. This means that Table 4 is comparable to Table 2 and Table 5 is comparable to Table 3. It can be seen that the behavior of the method with $m_k = 2^k$ and $m_k = 1$ are very similar in this comparison. As the convergence test is made per iteration k, the simplified Newton method is able to terminate earlier.

5. Summary and discussion

We have proposed an inexact Newton method with the aim to avoid matrix factorizations at the expense of more solves with the matrix. We envisage that such a scheme will be useful near a solution where quadratic rate of convergence is achieved by Newton's method, and one may view the final convergence as refinement of the solution. We have shown that the choice of $m_k = 2^k$ in the subproblems allows p-step convergence with Q-factor 2^p. Our scheme allows a dynamic level of accuracy in the subproblem with $m_k = 1$ as the least accurate. This corresponds to a simplified Newton method. We have also included some numerical results that demonstrate the behavior of the method.

Of particular interest are optimization methods of penalty and barrier type, where typically a perturbed set of nonlinear equations are solved, see, e.g., Dussault [8], Cores and Tapia [4] and Byrd, Curtis and Nocedal [2]. In particular, interior methods for linear programming are of high interest, since the linear algebra work is there the dominating factor in terms of computational effort, see., e.g., Mehrotra [9, 10], Tapia et al. [15], Wright [17]. Also, an advantage of the approach suggested here compared to an approach based on iterative methods for solving the Newton equation (1.2) inexatly is that the linear equations will automatically be satisfied at $z_1, z_2, \ldots, z_p$. We envisage that the techniques proposed in this paper might be useful there.

Acknowledgement

I thank Philip Gill, Jorge Moré and Margaret Wright for bibliographical assistance.

[1] For the simplified Newton method, with $m_k = 1$, `zratio` is presented for iterates with $k = 2^l - 1$, where l is an integer. Then `zratio` is given as $\|z_k - z^*\|/(\|z_{k'} - z^*\|^2)$, with $k' = 2^{l-1} - 1$.

References

[1] R. P. Brent. Some efficient algorithms for solving systems of nonlinear equations. *SIAM J. Numer. Anal.*, 10:327–344, 1973. Collection of articles dedicated to the memory of George E. Forsythe.

[2] R. H. Byrd, F. E. Curtis, and J. Nocedal. Infeasibility detection and SQP methods for nonlinear optimization. Technical report, Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA, 2008.

[3] X.-C. Cai and D. E. Keyes. Nonlinearly preconditioned inexact Newton algorithms. *SIAM J. Sci. Comput.*, 24(1):183–200 (electronic), 2002.

[4] D. Cores and R. A. Tapia. Perturbation lemma for the Newton method with application to the SQP Newton method. *J. Optim. Theory Appl.*, 97(2):271–280, 1998.

[5] R. S. Dembo, S. C. Eisenstat, and T. Steihaug. Inexact Newton methods. *SIAM J. Numer. Anal.*, 19:400–408, 1982.

[6] R. S. Dembo and T. Steihaug. Truncated-Newton algorithms for large-scale unconstrained optimization. *Math. Program.*, 26:190–212, 1983.

[7] N. Y. Deng and Z. Z. Wang. Theoretical efficiency of an inexact Newton method. *Journal of Optimization Theory and Applications*, 105(1):97–112, 2000.

[8] J. Dussault. Numerical stability and efficiency of penalty algorithms. *SIAM J. Numer. Anal.*, 32:296–317, 1995.

[9] S. Mehrotra. On finding a vertex solution using interior point methods. *Linear Algebra Appl.*, 152:233–253, 1991.

[10] S. Mehrotra. On the implementation of a primal-dual method. *SIAM J. Optim.*, 2:575–601, 1992.

[11] J. J. Moré and M. Y. Cosnard. Numerical solution of nonlinear equations. *ACM Trans. Math. Software*, 5(1):64–85, 1979.

[12] Netlib linear programming test problems. http://www.netlib.org/lp/data/.

[13] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, 1970.

[14] Y. Saad. *Iterative methods for sparse linear systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2003.

[15] R. A. Tapia, Y. Zhang, M. Saltzman, and A. Weiser. The predictor-corrector interior-point method as a perturbed composite Newton method. *SIAM J. Optim.*, 6:47–56, 1996.

[16] J. F. Traub. *Iterative methods for the solution of equations*. Prentice-Hall Series in Automatic Computation. Prentice-Hall Inc., Englewood Cliffs, N.J., 1964.

[17] S. J. Wright. *Primal-Dual Interior-Point Methods*. SIAM, Society for Industrial and Applied Mathematics, Philadelphia, 1997. ISBN 0-89871-382-X.

[18] T. J. Ypma. Local convergence of inexact Newton methods. *SIAM J. Numer. Anal.*, 21(3):583–590, 1984.

Algorithm 3.1. *The p-step inexact Newton method*

```
% [zk]=inexactNewton(z0,p,getFandFprim,tol)
% Input data:
% zk            initial value of z
% p             number of steps in loop
% getFandFprim  function that gives Fk and Fprimk given zk
% tol           optimality tolerance
% Output data:
% zk            estimate of solution z
  it=0; converged=false;
  while not(converged)
      for k=0:p-1
          [Fk,Fprimk]=feval(getFandFprim,zk);
          converged=(norm(Fk)<tol);
          if (converged) break; end
          it=it+1;
          if k==0
              F0=Fk; Fprim0=Fprimk;
              [L0,U0]=lu(Fprim0);   % Or other factorization
          end
          pk0=-L0\Fk; pk0=U0\pk0;
          dk=pk0;
          mk=2^k;     % mk=1 gives simplified Newton
          for i=1:mk-1;
              pki=-L0\((Fprimk-Fprim0)*pki); pki=U0\pki;
              dk=dk+pki;
          end
          zk=zk+dk;
      end
  end
```