Pp. 1447-1457 in *Progress in Analysis*, Vol. 2 (Heinrich G W Begehr. Robert P Gilbert and Man Wah Wong, Editors), World Scientific, Singapore, 2003, ISBN 981-238-967-9

AN INVERSE-FREE DIRECTIONAL NEWTON METHOD FOR SOLVING SYSTEMS OF NONLINEAR EQUATIONS

YURI LEVIN AND ADI BEN-ISRAEL

Abstract. A directional Newton method is proposed for solving systems of m equations in n unknowns. The method does not use the inverse, or generalized inverse, of the Jacobian, and applies to systems of arbitrary m, n. Quadratic convergence is established under typical assumptions (first derivative "not too small", second derivative "not too large"). The method is stable under singularities in the Jacobian.

1. Introduction

Consider a system of m equations in n unknowns:

$$\mathbf{f}(\mathbf{x}) \;=\; \mathbf{0} \;,\; \text{ or } \quad f_i(x_1, x_2, \ldots, x_n) \;=\; 0 \;,\; i \in \overline{1, m} \;. \tag{1}$$

If $m = n$ the Newton method for solving (1) uses the iterations, see e.g. [14],[10],

$$\mathbf{x}^{k+1} \;:=\; \mathbf{x}^k - J_{\mathbf{f}}(\mathbf{x}^k)^{-1}\, \mathbf{f}(\mathbf{x}^k) \;,\; k = 0, 1, \ldots \;, \tag{2}$$

where the Jacobian matrix

$$J_{\mathbf{f}}(\mathbf{x}) \;:=\; \left(\frac{\partial f_i}{\partial x_j}\right)$$

is assumed nonsingular. If $J_{\mathbf{f}}(\mathbf{x})$ is singular, or if $m \neq n$, a suitable generalized inverse of $J_{\mathbf{f}}(\mathbf{x})$ can be used in (2) e.g. ([1],[4]) without losing quadratic convergence, e.g. [9]. In particular, the Moore-Penrose inverse in (2) gives the method

$$\mathbf{x}^{k+1} \;:=\; \mathbf{x}^k - J_{\mathbf{f}}(\mathbf{x}^k)^{\dagger}\, \mathbf{f}(\mathbf{x}^k) \;,\; k = 0, 1, \ldots \;, \tag{3}$$

see e.g. [8] where quadratic convergence was established under typical assumptions. This method is applicable to least squares problems, because every limit point $\mathbf{x}^{\infty}$ of (3) is a stationary point of the sum of squares $\sum f_i^2(\mathbf{x})$,

$$\nabla \sum f_i^2(\mathbf{x}^{\infty}) \;=\; 0 \;.$$

Both methods (2) and (3) require matrix inversions. We present here a Newton method for solving systems of equations that does not use any matrix inversion. This requires two steps

---

Date: November 5, 2000.

1991 *Mathematics Subject Classification.* Primary 65H05, 65H10; Secondary 49M15.

Key words and phrases. Directional Newton method, Inverse-free method, Single equations, Systems of equations.

Step 1: Transform (1) to a single equation in n unknowns

$$F(\mathbf{x}) = 0 \ , \quad \text{or} \quad F(x_1, x_2, \ldots, x_n) = 0 \ , \tag{4}$$

such that (1) and (4) have the same solutions.

Step 2: Solve (4) by a **directional Newton method**, see e.g [7],

$$\mathbf{x}^{k+1} := \mathbf{x}^k - \frac{F(\mathbf{x}^k)}{\|\nabla F(\mathbf{x}^k)\|^2} \nabla F(\mathbf{x}^k) \ , \quad k = 0, 1, \ldots \tag{5}$$

In Step 1 a natural transformation is

$$F(\mathbf{x}) := \sum_{i=1}^{m} f_i^2(\mathbf{x}) \ = \ 0 \ , \tag{6}$$

and several authors (e.g. [2, Vol II, p. 165], [12, p. 362]) suggested solving (6) by minimizing the sum of squares $\sum f_i^2(\mathbf{x})$ using a suitable method, such as the steepest descent method. However, if one solves (6) using a directional Newton method, quadratic convergence is lost because the gradient of F

$$\nabla F(\mathbf{x}) = 2 \sum_{i=1,\ldots,m} f_i(\mathbf{x}) \, \nabla f_i(\mathbf{x})$$

approaches the zero vector as the values $f_i(\mathbf{x})$ tend to zero. To prevent this, we propose an alternative transformation

$$F(\mathbf{x}) := \sum_{i=1}^{m} \left(\sqrt{f_i^2(\mathbf{x}) + \theta_i^2} - \theta_i \right) \ = \ 0 \ , \ \text{where } \theta_i \geq 0, \ i \in \overline{1, m} \ , \tag{7}$$

with gradient (existing if all $f_i(\mathbf{x}) \neq 0$)

$$\nabla F(\mathbf{x}) = \sum_{i=1}^{m} \left(\frac{f_i\,(\mathbf{x})}{\sqrt{f_i^2(\mathbf{x}) + \theta_i^2}} \nabla f_i\,(\mathbf{x}) \right) \ , \tag{8}$$

whose behavior, as $f_i(\mathbf{x}) \to 0$, can be controlled by adjusting the parameters θ_i.
In [8] we established the quadratic convergence of The directional Newton method (5), and the more general directional method

$$\mathbf{x}^{k+1} \ := \ \mathbf{x}^k - \frac{F(\mathbf{x}^k)}{\nabla F(\mathbf{x}^k) \cdot \mathbf{d}^k} \, \mathbf{d}^k \ , \quad k = 0, 1, \ldots \ , \tag{9}$$

under typical assumptions on the function F around the initial point $\mathbf{x}_0$, and the successive directions $\{\mathbf{d}^k\}$. However, the convergence proofs in [8] are not applicable to the special case of F given by (7).
The main results of this paper are Theorems 1 and 2. Theorem 1 gives conditions for the quadratic convergence of the directional Newton method (5), conditions that are natural in the special case of $F(\mathbf{x})$ given by (7). Theorem 2 then establishes the quadratic convergence of (5) when applied to the equivalent equation $F(\mathbf{x}) = 0$.

We call the method $\{(7),(5)\}$ an *inverse–free Newton method.* This method is adapted to least squares solutions and optimization problems in §§ 4–5. The method is illustrated by numerical examples in § 6.
Since the inverse–free Newton method does not require inversion, it is well suited for dealing with singularities in the Jacobian, along the iterations $\{\mathbf{x}^k\}$ or in their limits $\mathbf{x}^\infty$.

2. Convergence of the Directional Newton Method (5)

In this section we give a new proof of the convergence of the directional Newton method (5) that apply naturally to F of (7). The main tool is the majorizing sequence, due to Kantorovich and Akilov [6], see also [10, Chapter 12.4].

Definition 1. A sequence $\left\{y^k\right\}$, $y^k \geq 0$, $y^k \in \mathbb{R}$ for which $\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| \leq y^{k+1} - y^k$, $k = 0, 1, ...$ is called a **majorizing sequence** for $\left\{\mathbf{x}^k\right\}$.

Note that any majorizing sequence is necessarily monotonically increasing.
The following two lemmas are used below.

Lemma 1. If $F : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable in an open set $\mathcal{S}$ then for any $\mathbf{x}, \mathbf{y} \in \mathcal{S}$,

$$\|\nabla F(\mathbf{y}) - \nabla F(\mathbf{x})\| \leq \|\mathbf{y} - \mathbf{x}\| \sup_{0 \leq t \leq 1} \|F''(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))\| \ .$$

Lemma 2. If $\left\{y^k\right\}$, $y^k \geq 0$, $y^k \in \mathbb{R}$ is a majorizing sequence for $\left\{\mathbf{x}^k\right\}$, $\mathbf{x}^k \in \mathbb{R}$ and $\lim_{k\to\infty} y^k = y^* < \infty$, then there exists $\mathbf{x}^* = \lim_{k\to\infty} \mathbf{x}^k$ and $\left\|\mathbf{x}^* - \mathbf{x}^k\right\| \leq y^* - y^k$, $k = 0, 1, ...$

Lemma 1 is consequence of the Mean Value Theorem, and Lemma 2 is proved in [10, Chapter 12.4, Lemma 12.4.1].
To prove the convergence of (5), we write it as

$$\mathbf{x}^{k+1} := \mathbf{x}^k - F\left(\mathbf{x}^k\right)\mathbf{v}^k, \tag{10a}$$

$$\text{where } \mathbf{v}^k := \frac{\nabla F(\mathbf{x}^k)}{\|\nabla F(\mathbf{x}^k)\|^2}. \tag{10b}$$

Theorem 1. Let $F : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function, $\mathbf{x}^0 \in \mathbb{R}^n$, and assume that

$$\sup_{\mathbf{x}\in X_0} \|f''(\mathbf{x})\| = M, \tag{11}$$

where X_0 is defined as

$$X_0 := \left\{\mathbf{x} : \left\|\mathbf{x} - \mathbf{x}^0\right\| \leq R\right\} \ , \tag{12}$$

for R given in terms of constants M, B, C that are assumed to satisfy

$$\left|F\left(\mathbf{x}^0\right)\right| \leq C \,, \tag{13a}$$

$$\left\|\nabla F\left(\mathbf{x}^0\right)\right\| \geq \frac{1}{B} \,, \tag{13b}$$

$$MB^2C < \frac{1}{2} \,, \tag{13c}$$

$$R := \frac{1-\sqrt{1-2MB^2C}}{MB} \,. \tag{13d}$$

Then:
(a) All the points $\mathbf{x}^{k+1} := \mathbf{x}^k - F\left(\mathbf{x}^k\right)\mathbf{v}^k,\ k = 0, 1, ...$ lie in X_0.
(b) $\mathbf{x}^* = \lim_{k\to\infty} \mathbf{x}^k$ exists, $\mathbf{x}^* \in \mathbf{X}_0$, and $f\left(\mathbf{x}^*\right) = 0$.
(c) The order of covergence of the directional Newton method (5) is quadratic.

Proof. We construct a majorizing sequence for $\{\mathbf{x}^k\}$ in terms of the auxiliary function

$$\varphi\left(y\right) = \frac{M}{2}y^2 - \frac{1}{B}\,y + C \,. \tag{14}$$

By (13c), the quadratic equation $\varphi(y) = 0$ has two roots $r_1 = \frac{1-\sqrt{1-2MB^2C}}{MB} = R$ and $r_2 = \frac{1+\sqrt{1-2MB^2C}}{MB}$. Also, $\varphi'(y) = My - \frac{1}{B}$ and $\varphi''(y) = M$.
Starting from $y^0 = 0$, apply the scalar Newton iteration to the function $\varphi\left(y\right)$ to get

$$\begin{aligned} y^{k+1} &= y^k - \frac{\varphi\left(y^k\right)}{\varphi'\left(y^k\right)},\ k = 0, 1, 2, ... \\ &= y^k - \frac{\frac{M}{2}\left(y^k\right)^2 - \frac{1}{B}\,y^k + C}{M\,y^k - \frac{1}{B}},\ k = 0, 1, 2, \ldots \\ &= \frac{\frac{M}{2}\left(y^k\right)^2 - C}{My^k - \frac{1}{B}},\ k = 0, 1, 2, \ldots \end{aligned} \tag{15}$$

Next we prove that the sequences $\left\{\mathbf{x}^k\right\}$ and $\left\{y^k\right\}$, generated by (5) and (15) respectively, satisfy for $k = 0, 1, ...$

$$\left|F\left(\mathbf{x}^k\right)\right| \leq \varphi\left(y^k\right) \,, \tag{16a}$$

$$\left\|\mathbf{v}^k\right\| \leq -\frac{1}{\varphi'\left(y^k\right)} \,, \tag{16b}$$

$$\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| \leq y^{k+1} - y^k \,. \tag{16c}$$

Statement (16c) says that $\left\{y^k\right\}$ is a majorizing sequence for $\left\{\mathbf{x}^k\right\}$. The proof is by induction.

<u>Verification for $k = 0$:</u>

$$\left|F\left(\mathbf{x}^0\right)\right| \leq C = \varphi\left(y^0\right) = \varphi\left(0\right),$$

$$\left\|\mathbf{v}^0\right\| \leq B = -\frac{1}{\varphi'\left(y^0\right)} \quad ,$$

$$\begin{aligned}
\left\|\mathbf{x}^1-\mathbf{x}^0\right\| &= \left\|F\left(\mathbf{x}^0\right)\mathbf{v}^0\right\| \leq \left|F\left(\mathbf{x}^0\right)\right|\left\|\mathbf{v}^0\right\| \\
&\leq -\frac{\varphi\left(y^0\right)}{\varphi'\left(y^0\right)} = y^1 - y^0,
\end{aligned}$$

showing that equations (16a)-(16c) hold for $k=0$. Suppose (16a)-(16c) hold for $k \leq n$.
Proof of (16a) for $n+1$:

$$\begin{aligned}
\left\|\mathbf{x}^{n+1}-\mathbf{x}^0\right\| &= \left\|\sum_{k=0}^{n}\left(\mathbf{x}^{k+1}-\mathbf{x}^k\right)\right\| \leq \sum_{k=0}^{n}\left(y^{k+1}-y^k\right) \\
&= y^{n+1}-y^0 = y^{n+1} \leq R \, . \\
\therefore\ \mathbf{x}^{n+1} &\in X_0 \, . \\
\text{Let } a_n(\mathbf{x}) &:= (\mathbf{x}-\mathbf{x}^n)\left\|\nabla F(\mathbf{x}^n)\right\|^2 + F(\mathbf{x}^n)\nabla F(\mathbf{x}^n) \, . \\
\therefore\ a_n(\mathbf{x}) &= (\mathbf{x}-\mathbf{x}^n)\left\|\nabla F(\mathbf{x}^n)\right\|^2 \, . \\
\text{Let } p_n(\mathbf{x}) &:= \frac{\nabla F(\mathbf{x}^n)}{\left\|\nabla F(\mathbf{x}^n)\right\|^2}\, a_n(\mathbf{x}) \, .
\end{aligned}$$

Note that $p_n(\mathbf{x}^{n+1})=0$. On the other hand, $p_n(\mathbf{x}^{n+1})$ can be represented as

$$\begin{aligned}
p_n\left(\mathbf{x}^{n+1}\right) &= F\left(\mathbf{x}^n\right) + \nabla F\left(\mathbf{x}^n\right)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) \, . \\
\therefore\ 0 &= F\left(\mathbf{x}^n\right) + \nabla F\left(\mathbf{x}^n\right)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) \, . \\
\therefore\ F\left(\mathbf{x}^{n+1}\right) &= F\left(\mathbf{x}^{n+1}\right) - F\left(\mathbf{x}^n\right) - \nabla F\left(\mathbf{x}^n\right)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) \, .
\end{aligned}$$

So, by induction

$$\begin{aligned}
\left|F\left(\mathbf{x}^{n+1}\right)\right| &\leq \frac{M}{2}\left\|\mathbf{x}^{n+1}-\mathbf{x}^n\right\|^2 \leq \frac{M}{2}\left(y^{n+1}-y^n\right)^2 = \varphi\left(y^{n+1}\right) \, , \\
\text{since } \varphi\left(y^{n+1}\right) &= \frac{M}{2}\left(y^{n+1}-y^n\right)^2 .
\end{aligned}$$

Proof of (16b) for $n+1$:

$$\begin{aligned}
\left\|\mathbf{v}^{n+1}\right\| &= \frac{1}{\left\|\nabla F\left(\mathbf{x}^{n+1}\right)\right\|} = \frac{1}{\left\|\nabla F\left(\mathbf{x}^0\right)-\left(\nabla F\left(\mathbf{x}^0\right)-\nabla F\left(\mathbf{x}^{n+1}\right)\right)\right\|} \\
&\leq \frac{1}{\left\|\nabla F\left(\mathbf{x}^0\right)\right\| - \left\|\nabla F\left(\mathbf{x}^0\right)-\nabla F\left(\mathbf{x}^{n+1}\right)\right\|} \\
&= \frac{1}{\left\|\nabla F\left(\mathbf{x}^0\right)\right\|\left|1-\frac{\left\|\nabla F(\mathbf{x}^0)-\nabla F(\mathbf{x}^{n+1})\right\|}{\left\|\nabla F(\mathbf{x}^0)\right\|}\right|} \, , \text{ by Lemma 1 ,} \\
&\leq \frac{B}{1-MBy^{n+1}} = \frac{1}{\frac{1}{B}-My^{n+1}} = \frac{1}{-\varphi'\left(y^{n+1}\right)} \, .
\end{aligned}$$

Proof of (16c) for $n+1$:

$$\begin{aligned}\left\|\mathbf{x}^{n+2}-\mathbf{x}^{n+1}\right\| &= \left\|F\left(\mathbf{x}^{n+1}\right)\mathbf{v}^{n+1}\right\| \leq \left|F\left(\mathbf{x}^{n+1}\right)\right|\left\|\mathbf{v}^{n+1}\right\| \\ &\leq -\frac{\varphi\left(y^{n+1}\right)}{\varphi'\left(y^{n+1}\right)}\ ,\ \text{by } (16a) \text{ and } (16b)\ , \\ &= y^{n+2}-y^{n+1}\ .\end{aligned}$$

Consequently, (16a)-(16c) hold for all $n \geq 0$. Since the sequence $\{\mathbf{x}^k\}$ is majorized by the sequence $\{y^k\}$ it follows from Lemma 2 that $\lim_{k\to\infty} \mathbf{x}^k = \mathbf{x}^*$ and $\mathbf{x}^* \in X_0$.

The scalar Newton method has a quadratic rate of convergence, and $\left|y^{k+1}-y^k\right| \leq \beta\frac{\theta^{2^k}}{1-\theta^{2^k}}$, where $\theta < 1$, see [13]. Therefore $\left\|\mathbf{x}^{k+1}-\mathbf{x}^k\right\| \leq \beta\frac{\theta^{2^k}}{1-\theta^{2^k}}$, and the sequence $\{\mathbf{x}^k\}$ has at least quadratic rate of convergence. □

3. Application to the solution of a system of equations

Recall the system (1) and the equivalent equation (7). The following theorem gives conditions for the the convergence of the directional Newton method (5) when applied to (7).

Theorem 2. Let $f_i : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable functions, $i \in \overline{1,m}$, let $\mathbf{x}^0 \in \mathbb{R}^n$, and assume that

$$\|f_i''(\mathbf{x})\| \leq \frac{M}{2m}\ ,\ \mathbf{x} \in X_0\ ,\ i \in \overline{1,m}\ , \tag{17}$$

$$\|\nabla f_i(\mathbf{x})\|^2 \leq \frac{M}{2m}\theta_i\ ,\ \mathbf{x} \in X_0\ ,\ i \in \overline{1,m}\ , \tag{18}$$

$$\text{where } X_0 \text{ is defined by } X_0 := \left\{\mathbf{x} : \left\|\mathbf{x}-\mathbf{x}^0\right\| \leq R\right\}\ , \tag{19}$$

for R given in terms of constants M, B, C that are assumed to satisfy

$$\sum_{i=1}^m \left(\sqrt{f_i^2(\mathbf{x}^0)+\theta_i^2}-\theta_i\right) \leq C\ , \tag{20a}$$

$$\left\|\sum_{i=1}^m \left(\frac{f_i\left(\mathbf{x}^0\right)}{\sqrt{f_i^2(\mathbf{x}^0)+\theta_i^2}}\nabla f_i\left(\mathbf{x}^0\right)\right)\right\| \geq \frac{1}{B}\ , \tag{20b}$$

$$MB^2C < \frac{1}{2}\ , \tag{20c}$$

$$R := \frac{1-\sqrt{1-2MB^2C}}{MB}\ . \tag{20d}$$

Then the function $F(\mathbf{x}) := \sum_{i=1}^m \left(\sqrt{f_i^2(\mathbf{x})+\theta_i^2}-\theta_i\right)$, where $\theta_i \geq 0$, $i \in \overline{1,m}$, satisfies all conditions of Theorem 1 with corresponding M, B, C, R, X_0 .

Proof. We show a function $F(\mathbf{x}) := \sum_{i=1}^{m} \left(\sqrt{f_i^2(\mathbf{x}) + \theta_i^2} - \theta_i\right)$, where $\theta_i \geq 0$, $i \in \overline{1,m}$, satisfies all conditions of the Theorem 1 with corresponding M, B, C, R, X_0.

$$\left|F(\mathbf{x}^0)\right| = \sum_{i=1}^{m} \left(\sqrt{f_i^2(\mathbf{x}^0) + \theta_i^2} - \theta_i\right) \leq C, \text{ by } (20a).$$

So, (13a) holds.

$$\nabla F(\mathbf{x}) = \sum_{i=1}^{m} \left(\frac{f_i(\mathbf{x})}{\sqrt{f_i^2(\mathbf{x}) + \theta_i^2}} \nabla f_i(\mathbf{x})\right),$$

$$\because \quad \|\nabla F(\mathbf{x})\| = \left\|\sum_{i=1}^{m} \left(\frac{f_i(\mathbf{x})}{\sqrt{f_i^2(\mathbf{x}) + \theta_i^2}} \nabla f_i(\mathbf{x})\right)\right\| , \text{ and}$$

$$\left\|\nabla F(\mathbf{x}^0)\right\| = \left\|\sum_{i=1}^{m} \left(\frac{f_i(\mathbf{x}^0)}{\sqrt{f_i^2(\mathbf{x}^0) + \theta_i^2}} \nabla f_i(\mathbf{x}^0)\right)\right\| \geq \frac{1}{B}, \text{ by } (20b).$$

So, (13b) holds.

$$F''(\mathbf{x}) = \sum_{i=1}^{m} \left(\frac{f_i(\mathbf{x})}{\sqrt{f_i^2(\mathbf{x}) + \theta_i^2}} f_i''(\mathbf{x}) + \frac{\theta_i^2 \nabla f_i(\mathbf{x}) \cdot \nabla f_i(\mathbf{x})^T}{(f_i^2(\mathbf{x}) + \theta_i^2)^{\frac{3}{2}}}\right),$$

$$\because \quad \|F''(\mathbf{x})\| = \left\|\sum_{i=1}^{m} \left(\frac{f_i(\mathbf{x})}{\sqrt{f_i^2(\mathbf{x}) + \theta_i^2}} f_i''(\mathbf{x}) + \frac{\theta_i^2 \nabla f_i(\mathbf{x}) \cdot \nabla f_i(\mathbf{x})^T}{(f_i^2(\mathbf{x}) + \theta_i^2)^{\frac{3}{2}}}\right)\right\|$$

$$\leq \sum_{i=1}^{m} \left(\|f_i''(\mathbf{x})\| + \frac{\theta_i^2 \|\nabla f_i(\mathbf{x})\|^2}{(f_i^2(\mathbf{x}) + \theta_i^2)^{\frac{3}{2}}}\right) \leq \sum_{i=1}^{m} \left(\|f_i''(\mathbf{x})\| + \frac{\|\nabla f_i(\mathbf{x})\|^2}{\theta_i}\right)$$

$$\leq \sum_{i=1}^{m} \left(\frac{M}{2m} + \frac{M}{2m}\right), \text{ by (17) and (18).}$$

So, (11) holds, where $X_0 := \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}^0\| \leq R\}$, $R := \frac{1-\sqrt{1-2MB^2C}}{MB}$ and $MB^2C < \frac{1}{2}$. The function $F(\mathbf{x}) := \sum_{i=1}^{m} \left(\sqrt{f_i^2(\mathbf{x}) + \theta_i^2} - \theta_i\right)$, where $\theta_i \geq 0$, $i \in \overline{1,m}$, thus satisfies all conditions of the Theorem 1 with corresponding M, B, C, R, X_0 . □

4. Application to least squares problems

The system (1) was replaced above by an equivalent single equation $F(\mathbf{x}) = 0$, with

$$F(\mathbf{x}) := \sum_{i=1}^{m} \left(\sqrt{f_i^2(\mathbf{x}) + \theta_i^2} - \theta_i\right) = 0 \text{ , where } \theta_i \geq 0,\ i = 1, \ldots, m \text{ ,} \tag{7}$$

and then solved by the directional Newton method

$$\mathbf{x}^{k+1} := \mathbf{x}^k - \frac{F(\mathbf{x}^k)}{\|\nabla F(\mathbf{x}^k)\|^2} \nabla F(\mathbf{x}^k) \ , \quad k = 0, 1, \ldots \tag{5}$$

where

$$\nabla F(\mathbf{x}) = \sum_{i=1}^{m} \left(\frac{f_i(\mathbf{x})}{\sqrt{f_i^2(\mathbf{x}) + \theta_i^2}} \nabla f_i(\mathbf{x}) \right) \ . \tag{8}$$

In practice it is often advantageous to use the following modification of (5)

$$\mathbf{x}^{k+1} := \mathbf{x}^k - \frac{F(\mathbf{x}^k)}{\|\nabla F(\mathbf{x}^k)\|^2} \sum_{i=1}^{m} f_i(\mathbf{x}^k) \nabla f_i(\mathbf{x}^k) \ , \quad k = 0, 1, \ldots \tag{21}$$

where the last occurrence of $\nabla F(\mathbf{x})$ in (5) is replaced by

$$\sum_{i=1}^{m} f_i(\mathbf{x}) \nabla f_i(\mathbf{x}) \ , \ \text{that is the gradient of} \ \sum_{i=1}^{m} f_i(\mathbf{x})^2 \ . \tag{22}$$

Experience shows similar iteration counts for (21) and (5), giving an advantage to (21) because of less work per iteration. Another advantage of (21) stems from (22): all limit points of (21) are stationary points of the sum of squares $\sum f_i(\mathbf{x})^2$. The modified method (21) can therefore be used to find least squares solutions, like the method (3) but without matrix inversion.
In contrast, the method (5) converges only if the system (1) has a solution.

5. Application to optimization

Optimization problems often call for solving

$$\nabla f(\mathbf{x}) \ = \ 0 \ , \tag{23}$$

where f is the objective function. This is a system of nonlinear equations

$$\frac{\partial f(\mathbf{x})}{\partial x_i} = 0, \ \ i \in \overline{1,n} \ .$$

that can be solved by (5) with F and ∇F replaced by

$$\begin{aligned} F(\mathbf{x}) \ &:= \ \sum_{i=1}^{n} \left(\sqrt{\left(\frac{\partial f(\mathbf{x})}{\partial x_i}\right)^2 + \theta_i^2} \ - \ \theta_i \right) \\ \nabla F(\mathbf{x}) \ &= \ \sum_{i=1}^{n} \left(\frac{\frac{\partial f(\mathbf{x})}{\partial x_i}}{\sqrt{\left(\frac{\partial f(\mathbf{x})}{\partial x_i}\right)^2 + \theta_i^2}} \left(\nabla \frac{\partial f}{\partial x_i} \right) (\mathbf{x}) \right) \ . \end{aligned}$$

6. Numerical Examples

In this section we illustrate the inverse-free Newton method for four numerical examples. In Example 4 we also compare this method with method (3). Example 5 compares the modified inverse-free method (21) with (3) for least-squares problems.

Example 1. Consider the system ([5, p. 186])

$$\begin{aligned} x &= x^2 + y^3 + z^5 \\ y &= x^3 + y^5 + z^7 \\ z &= x^5 + y^7 + z^{11}. \end{aligned}$$

The Newton method (2) converges for $\mathbf{x}^0 = (.8, .5, .3)$ to the root

$$\mathbf{z}^1 = (.7916675708, .5443461301, .3251333166) \ ,$$

but diverges for other initial points "between" the roots $\mathbf{z}^1$ and $\mathbf{z}^2 = (0, 0, 0)$, for example, $\mathbf{x}^0 = (.4, .3, .2)$. The inverse-free method (21) converges for $\mathbf{x}^0$, giving after 7 iterations

$$(.26804377\, 10^{-7}, -.40713982\, 10^{-7}, -.53475305\, 10^{-8})$$

sufficiently close to the root $\mathbf{z}^2$.

Example 2. Consider the system

$$\begin{aligned} x^2 + y &= 0, \\ -x^2 + y &= 0. \end{aligned}$$

The Newton method (2) cannot be used with initial points on the y–axis, where the Jacobian is singular. Both the inverse free method (5) and the generalized inverse Newton method (3) converge to the solution $(0, 0)$ in one iteration from any point on the y–axis.

Example 3. Consider the system (see [3, p. 470])

$$\begin{aligned} x^3 + xy &= 0 \\ y + y^2 &= 0. \end{aligned}$$

The Jacobian of this system is singular at any point $(x, -0.5)$, and the Newton method is inapplicable there. The inverse-free methods (5) and (21) are not affected by the singularity of the Jacobian, and converge also for initial points on the line $\{(x, -0.5) : x \in \mathbb{R}\}$.

Example 4. Consider a system of 10 equations in 10 unknowns ([8, Example 2]),

$$f_k(\mathbf{x}) := \sum_{i=1}^{10} x_i^k - 10 = 0 \ , \ k \in \overline{1, 10} \ , \tag{24}$$

with solution $\mathbf{x} = (1, 1, \cdots, 1)$. Problem (24) was solved using method (3) and the inverse-free method (5), with $\mathbf{x}^0 = (2, 2, \cdots, 2)$. The Newton method (2) cannot be used with $\mathbf{x}^0$ because the Jacobians $J_f(\mathbf{x})$ are singular (have rank 1) at all successive iterations.

After 10 iterations, both methods gave the solution $(1,1,\cdots,1)$ of system (24). Table 1 shows fast decreases in the **sum of squares errors** (SSE), $\sum_k f_k(\mathbf{x})^2$ for both methods. Note, that (5) does not compute the inverse of the Jacobian, whereas (3) does.

Table 1. Comparison of sum of squares errors in Example 4 for Methods (3) and (5), with initial solution $\mathbf{x}^0 = (2,2,\cdots,2)$

	Method (3)	**Method** (5)
Iteration	**SSE**	**SSE**
0	$.139401800 \cdot 10^9$	$.139401800 \cdot 10^9$
1	$.1721211495 \cdot 10^8$	$.1461084826 \cdot 10^8$
2	$.2132634809 \cdot 10^7$	$.1490439773 \cdot 10^7$
3	263707.9109	146690.3099
4	31756.60306	13490.88384
5	3425.414715	1014.499162
6	257.4808354	39.38440501
7	6.733861299	.2195197771
8	.01109470826	.00001080291589
9	$.372021265 \cdot 10^{-7}$	$.385 \cdot 10^{-13}$
10	0	0

Example 5. Consider a system of 10 equations in 10 unknowns ([8, Example 2]),

$$f_k(\mathbf{x}) := \sum_{i=1}^{10} x_i^k - 5 = 0 , \; k \in \overline{1,10} , \tag{25}$$

which has no solution, and a least squares solution is sought. Problem (25) was solved using method (3) and the modified inverse-free method (21), with $\mathbf{x}^0 = (2,2,\cdots,2)$. Both methods ((21) after 7 iterations and (3) after 10) gave the least squares solution $(.888,.888,\cdots,.888)$ of system (25). Table 2 shows fast decreases in the **sum of squares errors** (SSE), $\sum_k f_k(\mathbf{x})^2$ for both methods.

Acknowledgement

We thank Mikhail Nediak for his valuable comments and suggestions.

References

[1] A. Ben-Israel, *A Newton-Rapshon method for the solution of sytems of equations*, J. Math. Anal. Appl. **15**(1966), 243–252

[2] I.S. Berezin and N.P. Zhidkov, *Computing Methods*, Pergamon Press, 1965

[3] D. Decker and C. Kelley, *Newton's method at singular points*, SIAM J. Numer. Anal. **17-3**(1980), 465–471

[4] P. Deuflhard and G. Heindl, *Affine invariant convergence theorems for Newton's method and extensions to related methods*, SIAM J. Numer. Anal. **16**(1979), 1–10

TABLE 2. Comparison of sum of squares errors in Example 5 for Methods (3) and (21), with initial solution $\mathbf{x}^0 = (2, 2, \cdots, 2)$

	Method (3)	**Method** (21)
Iteration	**SSE**	**SSE**
0	$.139401800 \cdot 10^9$	$.139605650 \cdot 10^9$
1	$.1726341351 \cdot 10^8$	$.145970247848 \cdot 10^8$
2	$.2149328180 \cdot 10^7$	$.148160940564 \cdot 10^7$
3	269887.1376	144861.825286
4	34116.82173	13443.8154470
5	4282.049124	1133.94896877
6	523.6762415	93.5000837323
7	80.36779417	37.1186876848
8	39.43144346	–
9	37.19011123	–
10	37.11975344	–

[5] C.-E. Fröberg, *Mumerical Mathematics: Theory and Computer Applications*, Benjamin, 1985
[6] L.V. Kantorovich and G.P. Akilov, *Functional Analysis in Normed Spaces*, Macmillan, New York, 1964
[7] Y. Levin and A. Ben-Israel, *Directional Newton methods in n variables*, Math. of Computation (to appear)
[8] Y. Levin and A. Ben-Israel, *A Newton method for systems of m equations in n unknowns*, Nonlinear Analyses (to appear)
[9] M.Z. Nashed and X. Chen, *Convergence of Newton-like methods for singular operator equations using outer inverses*, Numer. Math. **66**(1993), 235–257
[10] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, 1970
[11] A.M. Ostrowski, *Solution of Equations in Euclidean and Banach Spaces*, 3rd Edition, Academic Press, 1973
[12] A. Ralston and P. Rabinowitz, *A First Course in Numerical Analysis*, 2nd edition, McGraw-Hill, 1978
[13] J. Stoer and K. Bulirsch, *Introduction to Numerical Analysis*, Springer-Verlag, 1976
[14] J.F. Traub, *Iterative Methods for the Solution of Equations*, Prentice–Hall, 1964

YURI LEVIN, RUTCOR–RUTGERS CENTER FOR OPERATIONS RESEARCH, RUTGERS UNIVERSITY, 640 BARTHOLOMEW RD, PISCATAWAY, NJ 08854-8003, USA
E-mail address: ylevin@rutcor.rutgers.edu

ADI BEN-ISRAEL, RUTCOR–RUTGERS CENTER FOR OPERATIONS RESEARCH, RUTGERS UNIVERSITY, 640 BARTHOLOMEW RD, PISCATAWAY, NJ 08854-8003, USA
E-mail address: bisrael@rutcor.rutgers.edu