DIMACS Technical Report 2000-13
May 2000

Directional Newton Methods in n Variables

by

Yuri Levin
[1] RUTCOR–Rutgers Center for Operations Research,
Rutgers University,
640 Bartholomew Rd, Piscataway, NJ 08854-8003, USA
Email address: ylevin@rutcor.rutgers.edu

Adi Ben-Israel
RUTCOR–Rutgers Center for Operations Research,
Rutgers University,
640 Bartholomew Rd, Piscataway, NJ 08854-8003, USA
Email address: bisrael@rutcor.rutgers.edu

[1]The work was supported by DIMACS

DIMACS is a partnership of Rutgers University, Princeton University, AT&T Labs-Research, Bell Labs, NEC Research Institute and Telcordia Technologies (formerly Bellcore).

DIMACS is an NSF Science and Technology Center, funded under contract STC–91–19999; and also receives support from the New Jersey Commission on Science and Technology.

ABSTRACT

Directional Newton methods for functions f of n variables are shown to converge, under typical assumptions, to a solution of $f(\mathbf{x}) = 0$. The rate of convergence is quadratic, for near-gradient directions, and directions along components of the gradient of f with maximal modulus. These methods are applied to solving systems of equations without inversion of the Jacobian matrix.

Key words and phrases. Newton Method, Single equations, Systems of equations
Mathematics Subject Classification. Primary 65H05, 65H10; Secondary 49M15

1 Introduction

Consider a single equation in n unknowns,

$$f(\mathbf{x}) = 0 \ , \quad \text{or} \quad f(x_1, x_2, \ldots, x_n) = 0 \ . \tag{1}$$

Given a point $\mathbf{x}^0$ where f is differentiable and a direction vector $\mathbf{d}$, we restrict f to the line

$$L := \{\mathbf{x}^0 + t\,\mathbf{d} : t \in \mathbb{R}\} \ ,$$

where it is a function of one variable

$$F(t) := f(\mathbf{x}^0 + t\,\mathbf{d}) \ .$$

The Newton iteration for F at $t^0 = 0$ gives the next point

$$t^1 := -\frac{F(0)}{F'(0)} \ ,$$

and the corresponding iteration for f is

$$\mathbf{x}^1 := \mathbf{x}^0 - \frac{f(\mathbf{x}^0)}{\nabla f(\mathbf{x}^0) \cdot \mathbf{d}}\,\mathbf{d} \tag{2}$$

since $F(0) = f(\mathbf{x}^0)$ and $F'(0)$ is the directional derivative

$$F'(0) = f'(\mathbf{x}^0, \mathbf{d}) = \nabla f(\mathbf{x}^0) \cdot \mathbf{d} \ .$$

Continuing in this fashion we get the iterations

$$\mathbf{x}^{k+1} := \mathbf{x}^k - \frac{f(\mathbf{x}^k)}{\nabla f(\mathbf{x}^k) \cdot \mathbf{d}^k}\,\mathbf{d}^k \ , \quad k = 0, 1, \ldots \tag{3}$$

that we call a **directional Newton method**. The geometric interpretation of (3) is: the points $\mathbf{x}^{k+1}$, for all directions $\mathbf{d}$, describe the intersection of $\mathbb{R}^n$ and the tangent (hyperplane in $\mathbb{R}^{n+1}$) of the graph of f at $(\mathbf{x}^k, f(\mathbf{x}^k))$, see Lemma 1.

For directions $\mathbf{d}^k$ sufficiently close to the gradients $\nabla f(\mathbf{x}^k)$, Theorems 1–2 establish quadratic convergence of the method (3) under typical assumptions, namely

- the gradient of f not "too small", see (10b) and (24b), and
- the second derivative (Hessian matrix) of f not "too large", see (10a).

A special case is the gradient method (29) studied in Corollary 1.

Another choice of $\mathbf{d}^k$ is the unit vector $\mathbf{e}^{m(k)}$ where $m(k)$ is the index of a component of $\nabla f(\mathbf{x}^k)$ of maximal modulus

$$\left|\nabla f(\mathbf{x}^k)[m(k)]\right| := \max_{j=1,2,\ldots,n} \left|\nabla f(\mathbf{x}^k)[j]\right| \ . \tag{4}$$

For this choice of $\mathbf{d}^k$, the directional Newton method (3) becomes

$$\mathbf{x}^{k+1} := \mathbf{x}^k - \frac{f(\mathbf{x}^k)}{\nabla f(\mathbf{x}^k)[m(k)]}\, \mathbf{e}^{m(k)} , \quad k = 0, 1, \dots \tag{5}$$

a method whose convergence is proved in Theorem 3. This method is suitable for parallel implementations.

These results are applied in § 5 to general systems of m equations in n unknowns.

Appendix A gives MAPLE programs for the methods considered here.

Notation: We use the Euclidean norm $\| \mathbf{x} \|$, and the corresponding matrix norm $\| A \|$, except in § 4 where the ∞-norm is used for vectors and matrices, denoted by $\| \mathbf{x} \|_\infty$ and $\| A \|_\infty$ respectively.

The angle between two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is denoted $\angle(\mathbf{u}, \mathbf{v})$ and given by

$$\angle(\mathbf{u}, \mathbf{v}) := \arccos \left(\frac{\mathbf{u} \cdot \mathbf{v}}{\| \mathbf{u} \| \, \| \mathbf{v} \|} \right) . \tag{6}$$

2 Geometric interpretation of the directional Newton method

For arbitrary directions $\mathbf{d}^k$, the points $\mathbf{x}^{k+1}$ defined by (3) describe a hyperplane in $\mathbb{R}^n$. It is the intersection of $\mathbb{R}^n$ and the tangent of the graph of f at the point $(\mathbf{x}^k, f(\mathbf{x}^k))$. It suffices to prove this for $k = 0$.

Lemma 1. *Assume $f(\mathbf{x}^0) \neq 0$ and $\nabla f(\mathbf{x}^0) \neq \mathbf{0}$. Let $\mathbf{d}$ be an arbitrary direction, and define $\mathbf{x}^1(\mathbf{d})$ by (2). Then the set*

$$\{\mathbf{x}^1(\mathbf{d}) : \| \mathbf{d} \| = 1, \, \mathbf{d} \in \mathbb{R}^n\}$$

is the intersection of $\mathbb{R}^n$ and the tangent hyperplane (in $\mathbb{R}^{n+1}$) of the graph of f at $(\mathbf{x}^0, f(\mathbf{x}^0))$.

Proof. Since $\nabla f(\mathbf{x}^0) \neq \mathbf{0}$ it follows that the tangent hyperplane of the graph of f at $(\mathbf{x}^0, f(\mathbf{x}^0))$ is "not horizontal". Its intersection with $\mathbb{R}^n$ is the hyperplane

$$f(\mathbf{x}^0) + \nabla f(\mathbf{x}^0) \cdot (\mathbf{x} - \mathbf{x}^0) = 0 , \tag{7}$$

which does not contain $\mathbf{x}^0$, since $f(\mathbf{x}^0) \neq 0$. Therefore, any point $\mathbf{x}$ in the above intersection is of the form

$$\mathbf{x} = \mathbf{x}^0 + t \, \mathbf{d} \tag{8}$$

where $\| \mathbf{d} \| = 1$ and $t \neq 0$. Substituting (8) in (7) we get

$$t := -\frac{f(\mathbf{x}^0)}{\nabla f(\mathbf{x}^0) \cdot \mathbf{d}}$$

proving the claim. □

The step length t is shortest if $\mathbf{d}$ is along the gradient $\nabla f(\mathbf{x}^0)$. Only directions close to the gradient are useful here. In particular, the step t is infinite if $\mathbf{d}$ is orthogonal to $\nabla f(\mathbf{x}^0)$.

3 Gradient and near gradient methods

In this section we study the convergence of the directional Newton method (3), along gradient and near gradient directions.

Theorem 1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable, let $\mathbf{x}^0$ be a point where*

$$f(\mathbf{x}^0) \neq 0 \ , \quad \nabla f(\mathbf{x}^0) \neq \mathbf{0} \ ,$$

let $\mathbf{d}^0 \in \mathbb{R}^n$ be such that $\| \mathbf{d}^0 \| = 1$ and let

$$\mathbf{h}^0 \ := \ -\frac{f(\mathbf{x}^0)}{\nabla f(\mathbf{x}^0) \cdot \mathbf{d}^0}\, \mathbf{d}^0 \ , \tag{9a}$$

$$\mathbf{x}^1 \ := \ \mathbf{x}^0 + \mathbf{h}^0 \ . \tag{9b}$$

Consider the ball

$$X_0 := \{\mathbf{x} : \| \mathbf{x} - \mathbf{x}^1 \| \leq \| \mathbf{h}^0 \|\}$$

and assume that $f \in C^2[X_0]$ and

$$\sup_{\mathbf{x} \in X_0} \| f''(\mathbf{x}) \| \ = \ M \ , \tag{10a}$$

$$| \nabla f(\mathbf{x}^0) \cdot \mathbf{d}^0 | \ \geq \ 2 \, \| \mathbf{h}^0 \| \, M \ , \tag{10b}$$

where f'' is the Hessian matrix of f and $\| f'' \|$ is its matrix norm corresponding to the Euclidean norm.

Consider the sequence $\{\mathbf{x}^i : i = 1, 2, \ldots\}$ defined recursively by

$$\mathbf{x}^{i+1} \ := \ \mathbf{x}^i + \mathbf{h}^i \ , \tag{11a}$$

$$\text{where} \quad \mathbf{h}^i \ := \ -\frac{f(\mathbf{x}^i)}{\nabla f(\mathbf{x}^i) \cdot \mathbf{d}^i}\, \mathbf{d}^i \ , \tag{11b}$$

$\mathbf{d}^i \in \mathbb{R}^n$ is such that $\| \mathbf{d}^i \| = 1$ and

$$\angle(\mathbf{d}^{i+1}, \nabla f(\mathbf{x}^{i+1})) \ \leq \ \angle(\mathbf{d}^i, \nabla f(\mathbf{x}^i)) \ , \ i = 0, 1, \ldots \ . \tag{12}$$

Then:

(a) All $\mathbf{x}^i$ lie in X_0.

(b) The sequence $\{\mathbf{x}^i\}$ converges, as $i \to \infty$, to a point $\mathbf{x}^* \in X_0$ that is a zero of f.

(c) Moreover, $\nabla f(\mathbf{x}^*) \neq \mathbf{0}$ unless $\| \mathbf{x}^* - \mathbf{x}^0 \| = \ 2 \ \| \mathbf{h}^0 \|$.

(d) Further, for $i = 1, 2, \ldots$

$$\| \mathbf{x}^{i+1} - \mathbf{x}^{i} \| \;\leq\; \frac{M}{2\,|\,\nabla f(\mathbf{x}^{i}) \cdot \mathbf{d}^{i}\,|}\;\| \mathbf{x}^{i} - \mathbf{x}^{i-1} \|^{2}\,, \tag{13a}$$

$$\| \mathbf{x}^{*} - \mathbf{x}^{i+1} \| \;\leq\; \frac{M}{2\,|\,\nabla f(\mathbf{x}^{i}) \cdot \mathbf{d}^{i}\,|}\;\| \mathbf{x}^{i} - \mathbf{x}^{i-1} \|^{2}\;. \tag{13b}$$

Note. Since $\| \mathbf{d}^{i} \| = 1$ for all i, the condition (12) is equivalent to:

$$\frac{|\,\nabla f(\mathbf{x}^{i+1}) \cdot \mathbf{d}^{i+1}\,|}{\| \nabla f(\mathbf{x}^{i+1}) \|} \;\geq\; \frac{|\,\nabla f(\mathbf{x}^{i}) \cdot \mathbf{d}^{i}\,|}{\| \nabla f(\mathbf{x}^{i}) \|}\;,\;\; i = 0, 1, \ldots\;. \tag{14}$$

Proof.
Part 1. Proof that

$$\| \nabla f(\mathbf{x}^{1}) \| \geq \frac{1}{2}\,\| \nabla f(\mathbf{x}^{0}) \|\;. \tag{15}$$

Each component of $f''(\mathbf{x})d\mathbf{x}$ is an exact 1-form, and by [1, Theorem 6.1],

$$\nabla f(\mathbf{x}) - \nabla f(\mathbf{x}^{0}) = \int_{\mathbf{x}^{0}}^{\mathbf{x}} f''(\mathbf{x})d\mathbf{x}\;.$$

Therefore, by (10a),

$$\| \nabla f(\mathbf{x}) - \nabla f(\mathbf{x}^{0}) \| \leq\; M\,\| \mathbf{x} - \mathbf{x}^{0} \| \tag{16}$$

for all $\mathbf{x} \in X_0$. In particular, for $\mathbf{x}^1$,

$$\begin{aligned}
\| \nabla f(\mathbf{x}^{1}) - \nabla f(\mathbf{x}^{0}) \| \;&\leq\; M\,\| \mathbf{x}^{1} - \mathbf{x}^{0} \| =\; M\;\| \mathbf{h}^{0} \| \\
&\leq\; \frac{1}{2}\;|\,\nabla f(\mathbf{x}^{0}) \cdot \mathbf{d}^{0}\,|\;,\;\text{by (10b)}\;, \\
&\leq\; \frac{1}{2}\,\| \nabla f(x_0) \|\;,\;\text{since}\;\; \| \mathbf{d}^{0} \| = 1\;. \qquad (17) \\
\therefore\; \| \nabla f(\mathbf{x}^{1}) \| \;&\geq\; \| \nabla f(\mathbf{x}^{0}) \| - \| \nabla f(\mathbf{x}^{1}) - \nabla f(\mathbf{x}^{0}) \|\;, \\
&\geq\; \| \nabla f(\mathbf{x}^{0}) \| - \frac{1}{2}\,\| \nabla f(\mathbf{x}^{0}) \|\;,\;\text{by (17)}\;, \\
&=\; \frac{1}{2}\,\| \nabla f(\mathbf{x}^{0}) \|\;,\;\text{proving (15)}\;.
\end{aligned}$$

Part 2. Proof that

$$|\,f(\mathbf{x}^{1})\,| \;\leq\; \frac{1}{2}\,\| \mathbf{h}^{0} \|^{2}\,M\;, \tag{18a}$$

$$|\,\nabla f(\mathbf{x}^{1}) \cdot \mathbf{d}^{1}\,| \;\geq\; 2\,\| \mathbf{h}^{1} \|\,M\;, \tag{18b}$$

$$\text{and}\;\; \| \mathbf{h}^{1} \| \;\leq\; \frac{1}{2}\,\| \mathbf{h}^{0} \|\;. \tag{18c}$$

First we prove that

$$f(\mathbf{x}^1) \;=\; \int_{\mathbf{x}^0}^{\mathbf{x}^1} (\mathbf{x}^1 - \mathbf{x}) f''(\mathbf{x}) d\mathbf{x}\;, \tag{19}$$

using integration by parts:

$$\begin{aligned}
\int_{\mathbf{x}^0}^{\mathbf{x}^1} (\mathbf{x}^1 - \mathbf{x}) \cdot f''(\mathbf{x}) d\mathbf{x} &= -(\mathbf{x}^1 - \mathbf{x}^0) \cdot \nabla f(\mathbf{x}^0) + f(\mathbf{x}^1) - f(\mathbf{x}^0) \\
&= -\mathbf{h}^0 \cdot \nabla f(\mathbf{x}^0) + f(\mathbf{x}^1) - f(\mathbf{x}^0) \\
&= \frac{f(\mathbf{x}^0)}{(\nabla f(\mathbf{x}^0) \cdot \mathbf{d}^0)} (\mathbf{d}^0 \cdot \nabla f(\mathbf{x}^0)) + f(\mathbf{x}^1) - f(\mathbf{x}^0) \\
&= f(\mathbf{x}^1)\;,\; \text{by (9a)}\;.
\end{aligned}$$

Next we prove (18a) by integrating (19) using a variable t defined by $\mathbf{x} = \mathbf{x}^0 + t\mathbf{h}^0$

$$\therefore\;\; \mathbf{x}^1 - \mathbf{x} = \mathbf{x}^1 - \mathbf{x}^0 - t\mathbf{h}^0 = \mathbf{h}^0 - t\mathbf{h}^0 = (1-t)\mathbf{h}^0\;,\;\; d\mathbf{x} = \mathbf{h}^0 dt\;.$$

Thus (19) becomes

$$f(\mathbf{x}^1) = \int_0^1 (1-t)\, \mathbf{h}^0 \cdot f''(\mathbf{x}^0 + t\,\mathbf{h}^0)\mathbf{h}^0\, dt\;.$$

Since $1 - t \geq 0$, it follows that

$$\begin{aligned}
|\, f(\mathbf{x}^1)\,| &\leq \int_0^1 (1-t) \frac{|f(\mathbf{x}^0)|^2}{|\,\nabla f(\mathbf{x}^0) \cdot \mathbf{d}^0\,|^2}\, \mathbf{d}^0 \cdot f''(\mathbf{x}^0 + t\,\mathbf{h}^0)\, \mathbf{d}^0\, dt\;, \\
&\leq \frac{M\,|f(\mathbf{x}^0)|^2}{2\,|\,\nabla f(\mathbf{x}^0) \cdot \mathbf{d}^0\,|^2}\, \|\,\mathbf{d}^0\,\|^2\;,\; \text{by (10a)}\;, \\
&= \frac{1}{2}\, \|\,\mathbf{h}^0\,\|^2\, M\;,\; \text{by (9a), proving (18a)}\;.
\end{aligned}$$

For $\mathbf{h}^1$ given by (11b) we have

$$\begin{aligned}
\| \mathbf{h}^1 \| &= \frac{| f(\mathbf{x}^1) |}{| \nabla f(\mathbf{x}^1) \cdot \mathbf{d}^1 |} \leq \frac{\frac{1}{2} \| \mathbf{h}^0 \|^2 M}{| \nabla f(\mathbf{x}^1) \cdot \mathbf{d}^1 |} , \text{ by (18a)} , \\
&\leq \frac{\frac{1}{2} \| \mathbf{h}^0 \|^2 M}{| \nabla f(\mathbf{x}^0) \cdot \mathbf{d}^0 |} \frac{\| \nabla f(\mathbf{x}^0) \|}{\| \nabla f(\mathbf{x}^1) \|} , \text{ by (14)} \\
&\leq \frac{\| \mathbf{h}^0 \|^2 M}{| \nabla f(\mathbf{x}^0) \cdot \mathbf{d}^0 |} , \text{ by (15)} . \qquad (20) \\
\therefore \frac{2M \| \mathbf{h}^1 \|}{\| \nabla f(\mathbf{x}^1) \|} &\leq \frac{2M \| \mathbf{h}^1 \|}{| \nabla f(\mathbf{x}^1) \cdot \mathbf{d}^1 |} , \text{ since } \| \mathbf{d}^1 \| = 1 , \\
&\leq \frac{2M^2 \| \mathbf{h}^0 \|^2}{| \nabla f(\mathbf{x}^1) \cdot \mathbf{d}^1 | \, | \nabla f(\mathbf{x}^0) \cdot \mathbf{d}^0 |} , \text{ by (20)} , \\
&\leq \frac{2M^2 \| \mathbf{h}^0 \|^2}{| \nabla f(\mathbf{x}^0) \cdot \mathbf{d}^0 |^2} \frac{\| \nabla f(\mathbf{x}^0) \|}{\| \nabla f(\mathbf{x}^1) \|} , \text{ by (14)} , \\
&\leq \frac{2^2 M^2 \| \mathbf{h}^0 \|^2}{| \nabla f(\mathbf{x}^0) \cdot \mathbf{d}^0 |^2} , \text{ by (15)} , \\
&= \left(\frac{2M \| \mathbf{h}^0 \|}{| \nabla f(\mathbf{x}^0) \cdot \mathbf{d}^0 |} \right)^2 \leq 1 , \text{ by (10b)} , \text{ proving (18b)} .
\end{aligned}$$

From (20) we also get

$$\frac{\| \mathbf{h}^1 \|}{\| \mathbf{h}^0 \|} \leq \frac{1}{2} \left(\frac{2 \| \mathbf{h}^0 \| M}{| \nabla f(\mathbf{x}^0) \cdot \mathbf{d}^0 |} \right) \leq \frac{1}{2} , \text{ by (10b), proving (18c)} .$$

Part 3. Proof of claims (a) and (b).
It follows from (18c) that

$$\| \mathbf{x}^2 - \mathbf{x}^1 \| = \| \mathbf{h}^1 \| \leq \frac{1}{2} \| \mathbf{h}^0 \|$$

showing that $\mathbf{x}^2 \in X_0$. Further, the ball

$$X_1 := \{ \mathbf{x} : \| \mathbf{x} - \mathbf{x}^2 \| \leq \| \mathbf{h}^1 \| \}$$

is contained in X_0. The inequality (18b) shows that the hypotheses of our theorem remain true if we replace $\mathbf{x}^0$ and $\mathbf{h}^0$ by $\mathbf{x}^1$ and $\mathbf{h}^1$, respectively. The same argument can be repeated for all $\mathbf{x}^i$ and $\mathbf{h}^i$ given by (11), $i = 0, 1, \ldots$. The following analogs of (18) therefore hold for all i,

$$| f(\mathbf{x}^i) | \leq \frac{1}{2} \| \mathbf{h}^{i-1} \|^2 \; M , \qquad (21a)$$

$$| \nabla f(\mathbf{x}^i) \cdot \mathbf{d}^i | \geq 2 \| \mathbf{h}^i \| \; M , \qquad (21b)$$

$$\text{and } \| \mathbf{h}^{i+1} \| \leq \frac{1}{2} \| \mathbf{h}^i \| , \qquad (21c)$$

showing that the nested balls

$$X_i := \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}^{i+1}\| \le \|\mathbf{h}^i\|\}$$

have radii converging to zero. Therefore there is a unique point $\mathbf{x}^*$ contained in all the balls X_i, and $\mathbf{x}^i \to \mathbf{x}^*$ since each $\mathbf{x}^i$ is the center of the ball X_{i+1}.

We prove now that $\mathbf{x}^*$ is a zero of $f(\mathbf{x})$.

$$\begin{aligned} |f(\mathbf{x}^i)| &\le \frac{1}{2}\|\mathbf{h}^{i-1}\|^2 \ M \ , \text{ by (21a)} \ , \\ &\le \left(\frac{1}{2}\right)^i \|\mathbf{h}^0\|^2 \ M \ , \text{ by (21c)} \ . \\ \therefore \quad f(\mathbf{x}^*) &= 0 \ . \end{aligned}$$

Part 4. Proof of claim (c).
We prove that $\nabla f(\mathbf{x}^*) \ne \mathbf{0}$ except if $\mathbf{x}^*$ is diametrically opposite to $\mathbf{x}^0$. For any $\mathbf{x} \in X_0$ we have, by (16),

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{x}^0)\| \le M \ \|\mathbf{x} - \mathbf{x}^0\| \le 2M \ \| \ \mathbf{h}^0\| \ .$$

If $\mathbf{x}$ is not diametrically opposite to $\mathbf{x}^0$, i.e. if $\|\mathbf{x} - \mathbf{x}^0\| < 2 \ \|\mathbf{h}^0\|$, then

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{x}^0)\| < 2M \ \|\mathbf{h}^0\| \le \|\nabla f(\mathbf{x}^0)\| \ ,$$

or

$$\|\nabla f(\mathbf{x}^0\| > \|\nabla f(\mathbf{x}^0) - \nabla f(\mathbf{x})\| \ ,$$

showing that $\nabla f(\mathbf{x}) \ne \mathbf{0}$.
Part 5. Proof of claim (d).
The inequality (13a) is equivalent to

$$\|\mathbf{h}^i\| \le \frac{M \ \|\mathbf{h}^{i-1}\|^2}{2 \ \|\nabla f(\mathbf{x}^i)\|} \ , \ i = 1, 2, \ldots \tag{22}$$

Indeed,

$$\|\mathbf{h}^i\| = \frac{|f(\mathbf{x}^i)|}{|\ \nabla f(\mathbf{x}^i) \cdot \mathbf{d}^i \ |} \le \frac{M \ \|\mathbf{h}^{i-1}\|^2}{2 \ |\ \nabla f(\mathbf{x}^i) \cdot \mathbf{d}^i \ |} \ , \text{ by (21a)} \ .$$

To prove (13b), we note

$$\|\mathbf{x}^* - \mathbf{x}^{i+1}\| \le \|\mathbf{x}^{i+1} - \mathbf{x}^i\| = \|\mathbf{h}^i\| \ ,$$

since $\mathbf{x}^{i+1}$ is the center of the ball X_i of radius $\|\mathbf{h}^i\|$, $\mathbf{x}^i$ is on the boundary of X_i and $\mathbf{x}^*$ is in X_i. Then (13b) follows from (13a). $\square$

Lemma 2. *If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable for all $\mathbf{x}$ in a convex region $X_0 \in \mathbb{R}^n$, and if there exists a constant M such that*

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \;\leq\; M\ \|\mathbf{x}-\mathbf{y}\|\ ,\ \forall\ \mathbf{x},\mathbf{y} \in X_0\ , \tag{23a}$$

then

$$|\ f(\mathbf{x}) - f(\mathbf{y}) - \nabla f(\mathbf{y})\cdot(\mathbf{x}-\mathbf{y})\ | \;\leq\; \frac{M}{2}\ \|\mathbf{x}-\mathbf{y}\|^2\ ,\ \forall\ \mathbf{x},\mathbf{y} \in X_0\ . \tag{23b}$$

Proof. Follows from [9, Lemma 5.3.1]. □

Remark. The Lipschitz bound $\mathbf{M}$ in (23a) can be used in (10a) if f is twice differentiable.

Theorem 2. *(Quadratic convergence of the Directional Newton Method)*
Let the assumptions of Theorem 1 hold, and in addition assume that there exist positive constants L, M such that

$$\|\nabla f(\mathbf{x})\ -\ \nabla f(\mathbf{y})\| \;\leq\; M\ \|\mathbf{x}\ -\ \mathbf{y}\|\ ,\ \forall\ \mathbf{x},\mathbf{y} \in X_0\ , \tag{24a}$$

$$|\ \nabla f(\mathbf{x})\cdot \mathbf{d}\ | \;\geq\; \frac{1}{L}\ ,\ \forall\ \mathbf{x} \in X_0\,,\ \mathbf{d} \in \mathbb{R}^n\ ,\ \|\mathbf{d}\| = 1\ , \tag{24b}$$

$$\text{and } q \;:=\; \frac{ML\ \|\mathbf{h}^0\|}{2} \;<\; 1\ . \tag{24c}$$

Then

$$\|\mathbf{x}^*\ -\ \mathbf{x}^i\| \leq \|\mathbf{h}^0\|\ \frac{q^{2^i-1}}{1-q^{2^i}}\ ,\ i = 1, 2, \ldots \tag{25}$$

Note: Since $0 < q < 1$, the inequalities (25) show that the directional Newton method is at least quadratically convergent.

Proof. By Theorem 1, part (a), all iterates $\mathbf{x}^i$ lie in X_0.

$$\begin{aligned} \therefore\ \|\mathbf{x}^{i+1}\ -\ \mathbf{x}^i\| \;&=\; \frac{|\ f(\mathbf{x}^i)\ |}{|\ \nabla f(\mathbf{x}^i)\cdot \mathbf{d}^i\ |} \\ &\leq\; L\ |\ f(\mathbf{x}^i)\ |\ ,\ \text{by (24b)}\ , \\ &=\; L\ |\ f(\mathbf{x}^i) - f(\mathbf{x}^{i-1}) - \nabla f(\mathbf{x}^{i-1})\cdot(\mathbf{x}^i - \mathbf{x}^{i-1})\ |\ , \end{aligned}$$

since

$$f(\mathbf{x}^{i-1}) + \nabla f(\mathbf{x}^{i-1})\cdot(\mathbf{x}^i - \mathbf{x}^{i-1}) \;=\; 0$$

follows by multiplying both sides of

$$\mathbf{x}^i - \mathbf{x}^{i-1} = -\frac{f(\mathbf{x}^{i-1})}{\nabla f(\mathbf{x}^{i-1})\cdot \mathbf{d}^{i-1}}\,\mathbf{d}^{i-1} \quad \text{with} \quad \nabla f(\mathbf{x}^{i-1})\ .$$

$$\therefore \| \mathbf{x}^{i+1} - \mathbf{x}^i \| \leq \frac{LM}{2} \| \mathbf{x}^i - \mathbf{x}^{i-1} \|^2 \ , \ \text{by (23b)} \ . \tag{26}$$

We prove by induction that

$$\| \mathbf{x}^{i+1} - \mathbf{x}^i \| \leq \| \mathbf{h}^0 \| \ q^{2^i - 1} \ , \ i = 0, 1, \ldots \ . \tag{27}$$

By definition this inequality holds for $i = 0$. Assume it is correct for $i - 1$.

$$\begin{aligned}
\therefore \| \mathbf{x}^{i+1} - \mathbf{x}^i \| &\leq \frac{LM}{2} \| \mathbf{x}^i - \mathbf{x}^{i-1} \|^2 \ , \ \text{by (26)} \ , \\
&\leq \frac{LM}{2} \| \mathbf{h}^0 \|^2 \ q^{2^i - 2} \ , \ \text{by (27) for } i - 1 \ , \\
&= \| \mathbf{h}^0 \| \ q^{2^i - 1} \ , \ \text{proving (27) for } i \ .
\end{aligned}$$

For $m > n$ we therefore have

$$\begin{aligned}
\| \mathbf{x}^m - \mathbf{x}^n \| &\leq \| \mathbf{x}^m - \mathbf{x}^{m-1} \| + \| \mathbf{x}^{m-1} - \mathbf{x}^{m-2} \| + \ldots + \| \mathbf{x}^{n+1} - \mathbf{x}^n \| \\
&\leq \| \mathbf{h}^0 \| \ q^{2^n - 1} \left(1 + q^{2^n} + (q^{2^n})^2 + \cdots + (q^{2^n})^{m-n-1}\right) \ , \ \text{by (27)} \ , \\
&\leq \| \mathbf{h}^0 \| \ q^{2^n - 1} \left(1 + q^{2^n} + (q^{2^n})^2 + \cdots \right) \ , \\
&= \frac{\| \mathbf{h}^0 \| \ q^{2^n - 1}}{1 - q^{2^n}} \ . \\
\therefore \lim_{m \to \infty} \| \mathbf{x}^m - \mathbf{x}^n \| &= \| \mathbf{x}^* - \mathbf{x}^n \| \leq \frac{\| \mathbf{h}^0 \| \ q^{2^n - 1}}{1 - q^{2^n}} \ , \ \text{proving (25)} \ .
\end{aligned}$$

$\square$

The main special case of Theorem 1 is the **gradient method** discussed next.

Corollary 1. *For $i = 0, 1, \ldots$ define the direction $\mathbf{d}^i$ and step $\mathbf{h}^i$ by*

$$\mathbf{d}^i := \frac{\nabla f(\mathbf{x}^i)}{\| \nabla f(\mathbf{x}^i \|} \ , \tag{28a}$$

$$\mathbf{h}^i := -\frac{f(\mathbf{x}^i)}{\| \nabla f(\mathbf{x}^i) \|^2} \nabla f(\mathbf{x}^i) \ , \tag{28b}$$

and let $f, \mathbf{x}^0, X_0$, M, L, q be as in Theorems 1–2. Then the iterations

$$\mathbf{x}^{i+1} := \mathbf{x}^i - \frac{f(\mathbf{x}^i)}{\| \nabla f(\mathbf{x}^i) \|^2} \nabla f(\mathbf{x}^i) \ , \ i = 0, 1, \ldots \ , \tag{29}$$

satisfy all conclusions of Theorems 1–2. $\square$

4 Newton directions along maximal modulus components of the gradient

A directional Newton method not covered by Theorem 1 is where the direction $\mathbf{d}$ in each iteration is chosen as the unit vector along the maximal absolute value $|\frac{\partial f}{\partial x_j}|$. However, the proof of Theorem 1 applies here, with small changes.

Theorem 3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable, let $\mathbf{x}^0$ be a point where*

$$f(\mathbf{x}^0) \neq 0 \ , \quad \nabla f(\mathbf{x}^0) \neq \mathbf{0} \ .$$

Let $m(0)$ be an index such that

$$|f_*'(\mathbf{x}^0)| := \left| \frac{\partial f}{\partial x_{m(0)}}(\mathbf{x}^0) \right| = \max_{j=1,\dots,n} \left| \frac{\partial f}{\partial x_j}(\mathbf{x}^0) \right|$$

and define the vectors $\mathbf{h}^0$ and $\mathbf{x}^1$ by

$$\mathbf{h}^0[k] \ := \ \begin{cases} -\dfrac{f(\mathbf{x}^0)}{|f_*'(\mathbf{x}^0)|} & , \ k = m(0) \ , \\ 0 & , \ k \neq m(0) \ , \end{cases} \tag{30a}$$

$$\mathbf{x}^1 \ := \ \mathbf{x}^0 + \mathbf{h}^0 \ . \tag{30b}$$

Consider the interval

$$X_0 \ := \ \{\mathbf{x} : \| \mathbf{x} - \mathbf{x}^1 \|_\infty \leq \| \mathbf{h}^0 \|_\infty \}$$

and assume that $f \in C^2[X_0]$, and

$$\sup_{\mathbf{x} \in X_0} \| f''(\mathbf{x}) \|_\infty \ = \ M \ , \tag{31a}$$

$$\| \nabla f(\mathbf{x}^0) \|_\infty \ \geq \ 2 \, \| \mathbf{h}^0 \|_\infty \ M \ . \tag{31b}$$

Define sequences $\{\mathbf{x}^i\}$, $\{\mathbf{h}^i\}$ recursively as follows. Let $m(i)$ be an index of the maximal modulus of $\frac{\partial f(\mathbf{x}^i)}{\partial x_j}$,

$$|f_*'(\mathbf{x}^i)| := \left| \frac{\partial f}{\partial x_{m(i)}}(\mathbf{x}^i) \right| = \max_{j=1,\dots,n} \left| \frac{\partial f}{\partial x_j}(\mathbf{x}^i) \right| \ ,$$

$$\mathbf{h}^i[k] \ := \ \begin{cases} -\dfrac{f(\mathbf{x}^i)}{|f_*'(\mathbf{x}^i)|} & , \ k = m(i) \ , \\ 0 & , \ k \neq m(i) \ , \end{cases} \tag{32a}$$

$$\mathbf{x}^{i+1} \ := \ \mathbf{x}^i + \mathbf{h}^i \ . \tag{32b}$$

Then

(a) All $\mathbf{x}^i$ *lie in* X_0.

(b) The sequence $\{\mathbf{x}^i\}$ *converges, as* $i \to \infty$*, to a point* $\mathbf{x}^* \in X_0$ *that is a zero of* f.

(c) Moreover, $\nabla f(\mathbf{x}^*) \neq \mathbf{0}$ *unless* $\|\mathbf{x}^* - \mathbf{x}^0\|_\infty = 2\;\|\mathbf{h}^0\|_\infty$.

(d) Further, for $i = 1, 2, \ldots$

$$\|\mathbf{x}^{i+1} - \mathbf{x}^i\|_\infty \;\leq\; \frac{M}{2\,\|\nabla f(\mathbf{x}^i)\|_\infty}\;\|\mathbf{x}^i - \mathbf{x}^{i-1}\|_\infty^2\,, \tag{33a}$$

$$\|\mathbf{x}^* - \mathbf{x}^{i+1}\|_\infty \;\leq\; \frac{M}{2\,\|\nabla f(\mathbf{x}^i)\|_\infty}\;\|\mathbf{x}^i - \mathbf{x}^{i-1}\|_\infty^2\;. \tag{33b}$$

Proof. We break the proof into 5 parts as in the proof of Theorem 1. Parts 1 and 3–5 of the latter proof can be used verbatim, by replacing each occurrence of $|\,\nabla f(\mathbf{x}^i)\cdot\mathbf{d}^i\,|$ by $\|\,\nabla f(\mathbf{x}^i)\,\|_\infty$, and by using the ∞–norm instead of the Euclidean norm. For example, the inequality (15) becomes

$$\|\nabla f(\mathbf{x}^1)\|_\infty \geq \frac{1}{2}\,\|\nabla f(\mathbf{x}^0)\|_\infty\;. \tag{34}$$

It remains to prove:

Part 2.

$$|\,f(\mathbf{x}^1)\,| \;\leq\; \frac{1}{2}\,\|\mathbf{h}^0\|_\infty^2\;M\,, \tag{35a}$$

$$\|\nabla f(\mathbf{x}^1)\|_\infty \;\geq\; 2\,\|\mathbf{h}^1\|_\infty\;M\,, \tag{35b}$$

$$\text{and}\;\;\|\mathbf{h}^1\|_\infty \;\leq\; \frac{1}{2}\,\|\mathbf{h}^0\|_\infty\;. \tag{35c}$$

Integrating (19) using the variable t as in Part 2 of the proof of Theorem 1 we get

$$\begin{aligned}
f(\mathbf{x}^1) &= \int_0^1 (1-t)\,\mathbf{h}^0\cdot f''(\mathbf{x}^0 + t\,\mathbf{h}^0)\mathbf{h}^0\,dt \\
&= \|\mathbf{h}^0\|_\infty^2 \int_0^1 (1-t)\frac{\partial^2 f}{\partial x_{m(0)}^2}(\mathbf{x}^0 + t\,\mathbf{h}^0)dt\,. \\
\therefore\,|\,f(\mathbf{x}^1)\,| &\leq \|\mathbf{h}^0\|_\infty^2 \int_0^1 (1-t)\left|\frac{\partial^2 f}{\partial x_{m(0)}^2}(\mathbf{x}^0 + t\mathbf{h}^0)\right| dt\,. \\
\therefore\,|\,f(\mathbf{x}^1)\,| &\leq \|\mathbf{h}^0\|_\infty^2\;M\int_0^1 (1-t)\,dt\;,\;\text{by (31a)}\;, \\
&= \frac{1}{2}\,\|\mathbf{h}^0\|_\infty^2\;M\;,\;\text{proving (35a)}\;.
\end{aligned}$$

Next prove (35b).

$$\begin{aligned}
\|\,\mathbf{h}^1\,\|_\infty &= \frac{|\,f(\mathbf{x}^1)\,|}{|\,f'_*(\mathbf{x}^1)\,|} = \frac{|\,f(\mathbf{x}^1)\,|}{\|\,\nabla f(\mathbf{x}^1\,\|_\infty} \\
&\le \frac{\frac{1}{2}\,\|\,\mathbf{h}^0\,\|_\infty^2\,M}{\frac{1}{2}\,\|\,\nabla f(\mathbf{x}^0)\,\|_\infty}\ ,\ \text{by (34) and (35a)}\ . \\
\therefore\ \frac{2\,M\ \|\,\mathbf{h}^1\,\|_\infty}{\|\,\nabla f(\mathbf{x}^1)\,\|} &\le \frac{2^2\ \|\,\mathbf{h}^0\,\|_\infty\ M^2}{\|\,\nabla f(\mathbf{x}^0)\,\|_\infty^2} = \left(\frac{2\ \|\,\mathbf{h}^0\,\|_\infty\ M}{\|\,\nabla f(\mathbf{x}^0)\,\|_\infty}\right)^2 \le 1
\end{aligned}$$

by (31b), proving (35b). The inequality (35c) then follows from (35a) and (35b). □

The convergence–rate of method (5) is quadratic under conditions analogous to Theorem 2. The proof is analogous to that of Theorem 2.

5 Systems of equations

Consider an arbitrary system of m equations in n unknowns:

$$\mathbf{f}(\mathbf{x}) = \mathbf{0}\ ,\ \text{or}\quad f_i(x_1, x_2, \ldots, x_n) = 0\ ,\ i = 1, \ldots, m\ . \tag{36}$$

If $m = n$ the usual Newton method for solving (36) uses the iterations

$$\mathbf{x}^{k+1} := \mathbf{x}^k - J_{\mathbf{f}}(\mathbf{x}^k)^{-1}\,\mathbf{f}(\mathbf{x}^k)\ ,\ k = 0, 1, \ldots\ , \tag{37}$$

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 $m \neq n$, the Moore-Penrose inverse can be used in (37),

$$\mathbf{x}^{k+1} := \mathbf{x}^k - J_{\mathbf{f}}(\mathbf{x}^k)^{\dagger}\,\mathbf{f}(\mathbf{x}^k)\ ,\ k = 0, 1, \ldots\ , \tag{38}$$

see [2]. Note that the gradient method (29) is a special case of (38), corresponding to $m = 1$.

Alternatively, several authors (e.g. [4, Vol II, p. 165], [8, p. 362]) suggest minimizing the sum of squares $\sum\limits_{i=1,\ldots,m} f_i^2(\mathbf{x})$, using a suitable method, such as the steepest descent method. Following this idea, consider the single equation,

$$F(\mathbf{x}) := \sum_{i=1,\ldots,m} f_i^2(\mathbf{x}) = 0\ , \tag{39}$$

that is equivalent to the system (36), in the sense of having the same solutions. There are now two (Newton) approaches to solving (36) or (39):

Approach 1 Apply a vector Newton method, e.g. (37) or (38), to the system (36).

Approach 2 Apply a directional Newton method, e.g. (29) or (5), to equation (39).

An advantage of Approach 2 is that it avoids the inversion of the Jacobian matrix, as in (37) or (38). See e.g. the comparison in Examples A.1 and A.4.

Another advantage is suitability for parallel implementation.

Disadvantages of Approach 2 include slower convergence. Indeed, the gradient of (39) is

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

and as the values $f_i(\mathbf{x})$ get close to zero, the condition (10b) may no longer hold. Thus quadratic convergence is lost.

Example 1. Another example of what can go wrong in Approach 2 is illustrated by the system of n equations in n unknowns,

$$\begin{aligned} f_1(\mathbf{x}) &:= 1 - x_1 = 0 \,, \\ f_i(\mathbf{x}) &:= 10\,(x_{i-1} - x_i^2) = 0 \,,\ i = 2, \dots, n \,, \end{aligned}$$

that is solved quickly by the Newton method (37), for all n that can be handled. However, the sum of squares $\sum f_i^2$ is the notorious Rosenbrock function

$$R_n(\mathbf{x}) := (1 - x_1)^2 + 100 \sum_{i=2}^{n} (x_{i-1} - x_i^2)^2 \,, \tag{40}$$

for which no directional Newton method works well, even for small n. See e.g. Example A.5. □

To salvage Approach 2 we can apply, in iteration k, the Newton method to the modified function

$$\frac{F(\mathbf{x})}{\|\,\mathbf{x} - \mathbf{x}^{k-1}\,\|^{\alpha}} \,, \text{ for suitably chosen } \alpha > 0 \,, \tag{41}$$

to get a next point $\mathbf{x}^{k+1}$, see e.g. [3]. The denominator in (41) creates a barrier at $\mathbf{x}^{k-1}$, and the resulting method may be called a **Newton barrier method**, see § A.4.

The modified function (41) is "steeper" than the original F, making it more likely to satisfy (10b). However, the modified function is less likely to satisfy the second derivative condition (10a). The choice of α will be studied elsewhere.

Example 2. (**Intersection of surfaces**) In computer graphics it is often required to compute and display the intersection $\mathcal{C} := \mathcal{S}_1 \cap \mathcal{S}_2$ of two surfaces $\mathcal{S}_1, \mathcal{S}_2$ in $\mathbb{R}^3$. The surfaces in question may be represented explicitly

$$\mathcal{S}_1 := \{(x, y, z) :\ z = f(x, y)\} \,,\ \mathcal{S}_2 := \{(x, y, z) :\ z = g(x, y)\}$$

in which case their intersection $\mathcal{C}$ is the set of points $\mathbf{r} = (x, y, z)$ satisfying

$$f(x, y) = g(x, y) \tag{42}$$

and then $z := f(x, y)$. If the surfaces are given implicitly

$$\mathcal{S}_1 := \{(x, y, z) :\ F(x, y, z) = 0\}\ ,\ \mathcal{S}_2 := \{(x, y, z) :\ G(x, y, z) = 0\}$$

then the intersection points satisfy

$$F(x, y, z) = 0\ ,\ G(x, y, z) = 0\ . \tag{43}$$

The so-called **marching methods** compute a **starting solution** $\mathbf{r}^0 = (x^0, y^0, z^0) \in \mathcal{C}$ and from it compute other intersection points sufficiently close to each other to enable displaying the curve $\mathcal{C}$, see [6].

A starting solution requires solving the single equation (42) in 2 unknowns, or the system (43) of 2 equations in 3 unknowns. Either way, the Newton method (37) cannot be used. The remaining alternatives include the method (38) using the generalized inverse of the Jacobian matrix, (see [2], [6]), or the directional Newton methods of this paper.

For any point $\mathbf{r}^k = (x^k, y^k, z^k) \in \mathcal{C}$, a next point $\mathbf{r}^{k+1}$ is found by solving (42) or (43) with an additional equation, such as

$$x = x^k + \Delta\ ,\ \Delta \text{ sufficiently small}\ ,$$

to give movement in the x direction. One can solve (in parallel) for movements in the y and z directions, then select the point closest to $\mathbf{r}^k$. $\square$

Example 3. (**Complex roots of non-analytic functions**) If $f(z)$ is analytic, its roots $z = x + i\,y$ can be approximated by the complex Newton method

$$z^{k+1} := z^k - \frac{f(z^k)}{f'(z^k)}\ ,\ k = 0, 1, \ldots \tag{44}$$

or equivalently, by solving the system of 2 equations in 2 unknowns

$$\Re f(x, y) = 0,\ \Im f(x, y) = 0\ . \tag{45}$$

If f is not analytic then (44) cannot be used, however the system (45) may still be solved by the real Newton method (37). If f is a non-analytic function of several complex variables, say $f(z_1, \overline{z_1}, z_2, \overline{z_2}, \ldots, z_m, \overline{z_m})$, then (45) is a system of 2 equations in $2\,m$ variables, and its surrogate equation $\Re^2 f + \Im^2 f = 0$ can be solved by the above directional Newton methods. $\square$

Acknowledgment

We thank Professor Arkadi Nemirovski for his help and constructive suggestions.

References

[1] W. Fleming, *Functions of Several Variables*, 2nd Edition, Springer, 1977

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

[3] A. Ben-Israel, *Newton's method with modified functions*, Contemp. Math. **204**(1997), 39–50

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

[5] C.-E. Fröberg, *Mumerical Mathematics: Theory and Computer Applications*, Benjamin, 1985

[6] G. Lukács, *The generalized inverse matrix and the surface-surface intersection problem*, pp. 167–185 in *Theory and Practice of Geometric Modeling* (W. Strasser and H.-P. Seidel, editors), Springer-Verlag, 1989

[7] A.M. Ostrowski, *Solution of Equations in Euclidean and Banach Spaces*, 3rd Edition, Academic Press, 1973

[8] A. Ralston and P. Rabinowitz, *A First Course in Numerical Analysis*, 2nd edition, McGraw-Hill, 1978

[9] J. Stoer and K. Bulirsch, *Introduction to Numerical Analysis*, Springer-Verlag, 1976

Appendix A: Maple Programs

Note. In the examples below all equations have zero RHS's, so the values of the functions give an indication of the error. Also, the functions use a vector variable `x` of unspecified dimension, making it necessary to define the dimension, say

```
> x:=array(0..3):
```

before using a function. We use the `linalg` package, so

```
> restart:with(linalg):
```

The function **SOS(x)** computes the sum of squares of the components of the vector **x**. It is used in some of the functions below, and works better than the MAPLE function `norm(x,2)`^{2} which is not differentiable if any $x_i = 0$.

```
> SOS:=proc(x)
> local n,k;
> n:=vectdim(x):
> sum(x[k]^2,k=1..n);
> end:
```

A.1 The Newton method (37)

The function **NewtonNext(f, x, x0)** computes the next Newton iterate of $\mathbf{f}(\mathbf{x})$ at $\mathbf{x}^0$

```
> NewtonNext:=proc(f,x,x0)
> local val,jac;
> val:=eval(subs(x=x0,f)):
> jac:=eval(jacobian(f,x),x=x0):jac:=inverse(jac);
> evalm(x0-jac &* val);
> end:
```

The function **Newton(f, x, x0, N)** computes N Newton iterations for $\mathbf{f}(\mathbf{x})$ and initial $\mathbf{x}^0$, and prints the last iterate, and the corresponding function value.

```
> Newton:=proc(f,x,x0,N)
> local k,sol,valf;
> sol:=array(0..N):sol[0]:=x0:valf:=eval(subs(x=x0,f)):
> lprint(ITERATION,0):print(sol[0]):
> lprint(function):print(valf):
> for k from 1 to N
>   do   sol[k]:=NewtonNext(f,x,sol[k-1]):
>        valf:=eval(subs(x=sol[k],f)):
>   od:
> lprint(ITERATION,k-1):print(sol[k-1]):
> lprint(function):print(valf):
> end:
```

Example A.1. $\mathbf{f}(\mathbf{x}) = (\exp\{1 - x_1 - x_2 - x_3\} - 1, x_1^2 - x_2, x_1^3 - x_3)$, $\mathbf{x}^0 = (1.0, 1.0, 1.2)$, 10 iterations

```
> x:=array(1..3):
> Newton([exp(1-x[1]-x[2]-x[3])-1,x[1]^2-x[2],
x[1]^3-x[3]],x,[1.,1., 1.2],10);

ITERATION   0
```

$$[1., 1., 1.2]$$

```
function
```

$$[-.8891968416, 0, -.2]$$

```
ITERATION   10
```

$$[.7019574100, .351893856, -.056449766]$$

```
function
```

$$[.002601879, .1408503495, .4023352123]$$

A.2 The directional Newton iteration (3)

The function **NewtonDirNext(f, x, x0, d)**computes the next directional Newton iterate of $f(\mathbf{x})$ at $\mathbf{x}^0$ in the direction $\mathbf{d}$

```
> NewtonDirNext:=proc(f,x,x0,d)
```

```
>  local val,gr,c;
>  val:=eval(subs(x=x0,f)):
>  gr:=eval(subs(x=x0,grad(f,x))):c:=dotprod(gr,d):
>  evalm(x0-(val/c)*d);
>  end:
```

A.3 The gradient method (29)

The function **NewtonGrad(f, x, x0, N)** computes N iterations for $f(\mathbf{x})$ starting at $\mathbf{x}^0$

```
>  NewtonGrad:=proc(f,x,x0,N)
>  local d,valf,sol,k;
>  sol:=array(0..N):sol[0]:=x0:valf:=eval(subs(x=x0,f)):
>  lprint(ITERATION,0):print(sol[0]):
>  lprint(function):print(valf):
>  for k from 1 to N
>    do  d:=eval(subs(x=sol[k-1],grad(f,x))):
>         sol[k]:=NewtonDirNext(f,x,sol[k-1],d):
>         valf:=eval(subs(x=sol[k],f)):
>    od:
>  lprint(ITERATION,k-1):print(sol[k-1]):
>  lprint(function):print(valf):
>  end:
```

Example A.2. $f(\mathbf{x}) = e^{1-x_1^2-x_2} - 1$, $\mathbf{x}^0 = (1.0, 1.2)$, 10 iterations.

```
>  x:=array(1..2):
>  NewtonGrad(exp(1-x[1]^2-x[2])-1,x,[1.,1.2],10);
```

`ITERATION   0`

$$[1., 1.2]$$

`function`

$$-.6988057881$$

`ITERATION   10`

$$[.1100943600, .9878783427]$$

`function`

$$.889\,10^{-6}$$

A.4 The Newton barrier method

The function **NewtonBarrier(f, x, x0, init, alpha, N)** computes N iterations of the directional gradient method (29) for the modified function

$$g(\mathbf{x}) := \frac{f(\mathbf{x})}{\|\,\mathbf{x} - \mathbf{x}^{k-1}\,\|^{\alpha}}$$

starting at $\mathbf{x}^0$, with $\mathbf{x}^{-1}$ =`init`.

```
>  NewtonBarrier:=proc(f,x,x0,init,alpha,N)
```

```
> local d,valf,sol,k,n,g;
> sol:=array(-1..N):sol[-1]:=init:sol[0]:=x0:
> valf:=eval(subs(x=x0,f)):
> lprint(ITERATION,0):print(x0):
> lprint(function):print(valf):
> for k from 1 to N
>   do   g:=f/((SOS(evalm(x-sol[k-2])))^(alpha/2)):
>        d:=eval(subs(x=sol[k-1],grad(g,x))):
>        sol[k]:=NewtonDirNext(g,x,sol[k-1],d):
>        valf:=eval(subs(x=sol[k],f)):
>   od:
> lprint(ITERATION,k-1):print(sol[k-1]):
> lprint(function):print(valf):
> end:
```

A.5 The method (5)

The function **Maximum(x)** finds the first component of **x** of maximal modulus, and its index

```
> Maximum:=proc(x) local i,g;
> g:=array(1..2);
> g[1]:=1; g[2]:=x[1];
> for i from 2 to vectdim(x)
>   do  if  (abs(x[i])>abs(g[2]))
>         then  g[1]:=i; g[2]:=x[i];
>        fi
>   od; [g[1],g[2]];
> end:
```

Example A.3. The third component of the vector $(1, 2, -5, -3, 5, 2)$ is the first occurrence of maximum modulus

```
> Maximum([1,2,-5,-3,5,2]);
```

$$[3, -5]$$

The function **NewtonMax(f, x, x0, N)** computes N iterations of (5) for f, starting at $\mathbf{x}^0$

```
> NewtonMax:=proc(f,x,x0,N)
> local d,valf,sol,k,gr,g,n,m;
> n:=vectdim(x0):sol:=array(0..N):sol[0]:=x0:
> valf:=eval(subs(x=x0,f)):
> lprint(ITERATION,0):print(sol[0]):
> lprint(function):print(valf):
> for k from 1 to N
>   do   gr:=eval(subs(x=sol[k-1],grad(f,x))):
>        g:=Maximum(gr):d:=array(1..n):
>        for m from 1 to n:
```

```
>           do   d[m]:=0
>           od:
>         d[g[1]]:=1:
>         sol[k]:=NewtonDirNext(f,x,sol[k-1],d):
>         valf:=eval(subs(x=sol[k],f)):
>   od:
> lprint(ITERATION,k-1):print(sol[k-1]):
> lprint(function):print(valf):
> end:
```

A.6 Systems of equations

The function **SystemNewtonGrad(f, x, x0, N)** computes N iterations of the method (29) for the sum of squares $\sum_{i=1}^{n} f_i^2(\mathbf{x})$, starting at $\mathbf{x}^0$

```
> SystemNewtonGrad:=proc(f,x,x0,N)
> local F;
> F:=SOS(f); print(F);
> NewtonGrad(F,x,x0,N);
> end:
```

The function **SystemNewtonBarrier(f,x,x0,init,alpha,N)** computes N iterations of the Newton Barrier method for the sum of squares $\sum_{i=1}^{n} f_i^2(\mathbf{x})$, starting at $\mathbf{x}^0$ with initial `init`.

```
> SystemNewtonBarrier:=proc(f,x,x0,init,alpha,N)
> local F;
> F:=SOS(f);print(F);
> NewtonBarrier(F,x,x0,init,alpha,N);
> end:
```

The function **SystemNewtonMax(f, x, x0, N)** computes N iterations of the method (5) for the sum of squares $\sum_{i=1}^{n} f_i^2(\mathbf{x})$, starting at $\mathbf{x}^0$

```
> SystemNewtonMax:=proc(f,x,x0,N)
> local F;
> F:=SOS(f); print(F);
> NewtonMax(F,x,x0,N);
> end:
```

Example A.4. Applying **SystemNewtonGrad** to the system in Example A.1, we get

```
> x:=array(1..3):
> SystemNewtonGrad([exp(1-x[1]-x[2]-x[3])-1,x[1]^2-x[2],
x[1]^3-x[3]],x,[1., 1., 1.2],10);
```

$$(\exp(1 - x_1 - x_2 - x_3) - 1)^2 + (x_1^2 - x_2)^2 + (x_1^3 - x_3)^2$$

```
ITERATION   0
```

$$[1., 1., 1.2]$$

```
function
```

$$.8306710231$$

```
ITERATION   10
```

$$[.6663989356,\ .3016884459,\ .1488239919]$$

```
function
```

$$.05409448940$$

a smaller error than the sum of squares of errors after 10 iterations of the Newton method, see Example A.1.

A.7 The Rosenbrock function

The function **Rosenbrock(n)** computes the Rosenbrock function (40)

```
>  Rosenbrock:=proc(n)
>  local i,s; global x;
>  x:=array(1..n);s:=(1-x[1])^2;
>  for i from 1 to n-1
>    do  s:=s+100*(x[i]-x[i+1]^2)^2
>    od;
>  end:
```

Example A.5. The Rosenbrock function in $\mathbb{R}^3$

```
>  Rosenbrock(3);
```

$$(1 - x_1)^2 + 100\,(x_1 - {x_2}^2)^2 + 100\,(x_2 - {x_3}^2)^2$$

Applying the Newton method (37) to the square roots of the components of the above Rosenbrock function we get quick convergence. For example, starting at $\mathbf{x}^0 = (0.6, 0.7, 0.4)$, 5 iterations suffice:

```
>  x:=array(1..3):
>  Newton([1-x[1],10*(x[1]-x[2]^2),10*(x[2]-x[3]^2)],x, [0.6,0.7,0.4],5);
```

```
ITERATION   0
```

$$[.6,\ .7,\ .4]$$

```
function
```

$$[.4,\ 1.10,\ 5.40]$$

```
ITERATION   5
```

$$[1.000000000,\ 1.000000000,\ 1.000007651]$$

```
function
```

$$[0,\ 0,\ -.00015302]$$

However, applying a directional Newton method to the Rosenbrock function gives very slow convergence, for example

```
>  NewtonGrad(Rosenbrock(3),x,[0.6,0.7,0.4],200);
```

```
ITERATION   0
```

$$[.6,\ .7,\ .4]$$

```
function
```

$$30.5300$$

```
ITERATION   200
```

$$[.9619433387,\ .9817641324,\ .9882554815]$$

```
function
```

$$.004432543367$$

Example A.6. Consider the system

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

The Newton method (37) converges for $\mathbf{x}^0 = (.8, .5, .3)$ to the root

$$\mathbf{z}^1 = (0.7916675703,\ 0.5443461301,\ 0.3251333166)\ ,$$

see [5, p. 186], but diverges for some other points $\mathbf{x}^0$ "between" the roots $\mathbf{z}^1$ and $\mathbf{z}^2 = (0, 0, 0)$, for example, $\mathbf{x}^0 = (.4, .3, .2)$,

```
>  x:=array(1..3):
>  Newton([x[1]^2-x[1]+x[2]^3+x[3]^5,x[1]^3+x[2]^5-x[2]
+x[3]^7,x[1]^5+x[2]^7+x[3]^11-x[3]], x,[0.4,0.3,0.2],5);
```

```
ITERATION   0
```

$$[.4,\ .3,\ .2]$$

```
function
```

$$[-.21268,\ -.2335572,\ -.1895412795]$$

```
ITERATION   5
```

$$[-.1521795626\,10^{149},\ -.589159899\,10^{157},\ .7450247900\,10^{29}]$$

```
function
```

$$[-.2045029311\,10^{471},\ -.7098488699\,10^{784},\ -.2463952058\,10^{1098}]$$

For this system of equations, and similarly chosen $\mathbf{x}^0$, the directional Newton methods **SystemNewtonGrad**, **SystemNewtonMax**, and **SystemNewtonBarrier** converged to one of the two roots $\mathbf{z}^1$ and $\mathbf{z}^2$.