Directional Halley and Quasi-Halley Methods in n Variables

Yuri Levin[a] * and Adi Ben-Israel[b]

[a]RUTCOR–Rutgers Center for Operations Research, Rutgers University, 640 Bartholomew Rd, Piscataway, NJ 08854-8003, USA. E-mail: ylevin@rutcor.rutgers.edu

[b]RUTCOR–Rutgers Center for Operations Research, Rutgers University, 640 Bartholomew Rd, Piscataway, NJ 08854-8003, USA. E-mail: bisrael@rutcor.rutgers.edu

A directional Halley method for functions f of n variables is shown to converge, at a cubic rate, to a solution. To avoid the second derivative needed in Halley method we propose a directional quasi-Halley method, with one more function evaluation per iteration than the directional Newton method, but with convergence rates comparable to the Halley method.

1. Introduction

When describing iterations such as

$$x^{k+1} := \Phi(x^k, d^k)\ ,\ \text{or}\ x^{k+1} := \Psi(x^k, x^{k-1})\ ,\quad k = 0, 1, \ldots \tag{1}$$

we sometimes denote the current point by x, the next point by x_+, and the previous point by x_-, so that (1) is written simply as

$$x_+ := \Phi(x, d)\ ,\ \text{or}\ x_+ := \Psi(x, x_-)\ .$$

Consider a single equation in n unknowns,

$$f(\mathbf{x}) = 0\ ,\quad \text{or}\quad f(x_1, x_2, \ldots, x_n) = 0\ . \tag{2}$$

Under standard assumptions on f and the initial approximation, the *directional Newton method*

$$\mathbf{x}_+ := \mathbf{x} - \frac{f(\mathbf{x})}{\nabla f(\mathbf{x}) \cdot \mathbf{d}}\, \mathbf{d}\ , \tag{3}$$

converges to a solution at a quadratic rate, for certain directions $\mathbf{d}$ related to the gradient $\nabla f(\mathbf{x})$, see [6, Theorems 2–3]. An important special case of (3) is when $\mathbf{d}$ is along the gradient $\nabla f(\mathbf{x})$, giving

$$\mathbf{x}_+ := \mathbf{x} - \frac{f(\mathbf{x})}{||\nabla f(\mathbf{x})||^2}\, \nabla f(\mathbf{x})\ . \tag{4}$$

*The work was supported by DIMACS

For $n = 1$, method (3) reduces to the scalar Newton method

$$x_+ := x - \frac{f(x)}{f'(x)} \tag{5}$$

with which it shares quadratic convergence. Applying (5) to the function $f(x)/\sqrt{f'(x)}$ we get

$$x_+ := x - \frac{f(x)}{f'(x) - \dfrac{f''(x)f(x)}{2f'(x)}}\ , \tag{6}$$

the (scalar) *Halley method* with cubic rate of convergence. The *quasi-Halley method* of [1] replaces the second derivative in (6) by a difference $(f'(x) - f'(x_-))/(x - x_-)$,

$$x_+ := x - \frac{f(x)}{f'(x) - \dfrac{(f'(x) - f'(x_-)f(x)}{2(x - x_-)f'(x)}} \tag{7}$$

without losing much in convergence rate, see [1, Theorem 3] where (7) was shown to have order $1 + \sqrt{2}$, and [1, Theorem 5] where the Halley step h and quasi-Halley step q were shown to satisfy $|h - q| = O(|u_-|^3)$ where u_- is the previous Newton step. This shows that sufficiently close to a solution, the Halley and quasi-Halley methods are indistinguishable, as confirmed by numerical experiments.

A Halley method for solving operator equations in Banach space was given by Safiev [7] and Yao [9]. They assume three times differentiable mappings, with bijective first derivatives. Specializing to functions $f : \mathbb{R}^n \to \mathbb{R}^n$, where the second derivative f'' is a tensor, the Halley method of [7] and [9] is of theoretical interest, but difficult to implement.

On the other hand, a Halley method for solving (2), a single equation in n unknowns, is practical. For $f : \mathbb{R}^n \to \mathbb{R}$ we define, in analogy with the directional Newton method (3), the *directional Halley Method* as:

$$\mathbf{x}_+ := \mathbf{x} - \frac{f(\mathbf{x})}{\left(\nabla f(\mathbf{x}) - \dfrac{f(\mathbf{x})}{2\nabla f(\mathbf{x}) \cdot \mathbf{d}}\, f''(\mathbf{x})\mathbf{d}\right) \cdot \mathbf{d}}\, \mathbf{d}\ , \tag{8}$$

where f'' is the Hessian matrix of f. For $n = 1$, (8) reduces to the scalar Halley method (6).

We establish cubic convergence for the method (8) for the directions $\mathbf{d}$ in two cases:

- directions $\mathbf{d}$ nearly constant throughout the iterations, see § 2, Theorem 1, and
- directions $\mathbf{d}$ along the gradient $\nabla f(\mathbf{x})$, in which case (8) becomes

$$\mathbf{x}_+ := \mathbf{x} - \frac{f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|^2 - \dfrac{f(\mathbf{x})}{2\,\|\nabla f(\mathbf{x})\|^2}\, \nabla f(\mathbf{x}) \cdot f''(\mathbf{x})\, \nabla f(\mathbf{x})}\, \nabla f(\mathbf{x})\ , \tag{9}$$

see § 3, Theorem 2.

In § 4 we propose the following *directional quasi-Halley method*

$$\mathbf{x}_+ := \mathbf{x} - \frac{f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|^2 \left(1 - f\left(\mathbf{x} - \frac{f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|^2}\nabla f(\mathbf{x})\right)/f(\mathbf{x})\right)}\nabla f(\mathbf{x}) \,, \tag{10}$$

obtained by approximating the term involving the second derivative in (9)

$$\frac{f(\mathbf{x})}{2\,\|\nabla f(\mathbf{x})\|^2}\nabla f(\mathbf{x}) \cdot f''(\mathbf{x})\,\nabla f(\mathbf{x}) \quad \text{by} \quad \frac{\|\nabla f(\mathbf{x})\|^2}{f(\mathbf{x})} f\left(\mathbf{x} - \frac{f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|^2}\nabla f(\mathbf{x})\right) \,.$$

The advantages of the directional quasi–Halley method (10) include:

- it avoids the second derivative needed in (9), requiring only one more function evaluation per iteration than the directional Newton method (4),
- if both methods (9) and (10) converge, their steps near a solution are approximately equal, see Theorem 3.

In §§ 5–6 we study three directional methods, along the direction of the gradient:

- the directional Newton method (4),
- the directional Halley method (9), and
- the directional quasi-Halley method (10).

We will drop the adjective "directional" when referring to these methods. The corresponding steps, along the gradient $\nabla f(\mathbf{x})$,

$$\text{Newton step}\,, \quad \mathbf{u} = -\frac{f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|^2}\nabla f(\mathbf{x}) \,, \tag{11a}$$

$$\text{Halley step}\,, \quad \mathbf{h} = -\frac{f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|^2 - \frac{f(\mathbf{x})}{2\,\|\nabla f(\mathbf{x})\|^2}\nabla f(\mathbf{x}) \cdot f''(\mathbf{x})\nabla f(\mathbf{x})}\nabla f(\mathbf{x}) \,, \text{ and} \tag{11b}$$

$$\text{quasi-Halley step}\,, \quad \mathbf{q} = -\frac{f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|^2 \left(1 - f\left(\mathbf{x} - \frac{f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|^2}\nabla f(\mathbf{x})\right)/f(\mathbf{x})\right)}\nabla f(\mathbf{x}) \tag{11c}$$

are compared in § 5. In numerical experiments, reported in § 6, the three methods were applied to randomly generated test problems with polynomials in several variables. In terms of average number of iterations, the Halley and quasi-Halley methods are comparable, and both are superior to the Newton method (4).

2. The directional Halley method (8) with nearly constant directions

In this section we study the convergence of the method (8), for directions $\{\mathbf{d}^i : i = 0, 1, \cdots\}$ that are nearly constant in the sense of condition (16) below. We use the following result, a consequence of the Mean Value Theorem,

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

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

The main tool for proving convergence is the majorizing sequence, due to Kantorovich and Akilov [5], see also [4, Chapter 12.4] and [9], where a majorizing sequence was used to prove cubic convergence for Halley's method in Banach space.

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

Note that any majorizing sequence is necessarily monotonically increasing.
We recall the following lemma, proved in [4, Chapter 12.4, Lemma 12.4.1].

Lemma 2. *Let $\{y^k\}$ is a majorizing sequence for $\{\mathbf{x}^k\} \subset \mathbb{R}^n$ and $\lim_{k\to\infty} y^k = y^* < \infty$. Then there exists $\mathbf{x}^* = \lim_{k\to\infty} \mathbf{x}^k$ and $\|\mathbf{x}^* - \mathbf{x}^k\| \leq y^* - y^k$, $k = 0, 1, \ldots$.* □

To prove the convergence of (8), we write that iteration as

$$\mathbf{x}^{k+1} := \mathbf{x}^k - \alpha^k f\left(\mathbf{x}^k\right) \mathbf{v}^k \text{ , where} \tag{12a}$$

$$\mathbf{v}^k := \frac{\mathbf{d}^k}{\nabla f(\mathbf{x}^k) \cdot \mathbf{d}^k} \text{ , and} \tag{12b}$$

$$\alpha^k := \frac{1}{1 - \frac{1}{2} f(\mathbf{x}^k)\, \mathbf{v}^k \cdot f''(\mathbf{x}^k) \mathbf{v}^k} . \tag{12c}$$

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

$$\sup_{\mathbf{x}\in X_0} \|f''(\mathbf{x})\| = M , \tag{13a}$$

$$\sup_{\mathbf{x}\in X_0} \|f'''(\mathbf{x})\| = N , \tag{13b}$$

where X_0 is defined as

$$X_0 := \left\{\mathbf{x} := \|\mathbf{x} - \mathbf{x}^0\| \leq (1 + q) B\right\} , \tag{14}$$

for q and B given in terms of constants L, T, C that are assumed to satisfy

$$\left|\nabla f(\mathbf{x}^0) \cdot \mathbf{d}^0\right| \geq \frac{1}{L} , \tag{15a}$$

$$|f(\mathbf{x}^0)| \leq \frac{B}{L} , \tag{15b}$$

$$T := CLB < \frac{1}{2} \text{ , where } C := \sqrt{M^2 + \frac{2}{3}\frac{N}{L(1 - \frac{1}{2}MLB)}} , \tag{15c}$$

$$q = \frac{1 - \sqrt{1 - 2T}}{1 + \sqrt{1 - 2T}} , \tag{15d}$$

$\mathbf{d}^0$ is the initial direction, and all directions satisfy

$$\angle(\mathbf{d}^i, \nabla f(\mathbf{x}^0)) \leq \angle(\mathbf{d}^0, \nabla f(\mathbf{x}^0)) \ , \ \mathbf{d}^i \in \mathbb{R}^n \ , \ \| \mathbf{d}^i \| = 1 \ , \ i = 0, 1, \dots \ . \tag{16}$$

Then:
(a) All the points $\mathbf{x}^{k+1} := \mathbf{x}^k - \alpha^k f\left(\mathbf{x}^k\right) \mathbf{v}^k$, $k = 0, 1, \cdots$ lie in X_0.
(b) $\mathbf{x}^* = \lim_{k\to\infty} \mathbf{x}^k$ exists, $\mathbf{x}^* \in X_0$, and $f\left(\mathbf{x}^*\right) = 0$.
(c) The order of convengence of the directional Halley method (8) is cubic.

Note: We use condition (16) in its equivalent form

$$\mathbf{d}^i \cdot \nabla f(\mathbf{x}^0) \geq \mathbf{d}^0 \cdot \nabla f(\mathbf{x}^0) \ , \ \mathbf{d}^i \in \mathbb{R}^n \ , \ \| \mathbf{d}^i \| = 1 \ , \ i = 0, 1, \dots \ . \tag{17}$$

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

$$\varphi\left(y\right) := \frac{C}{2} y^2 - \frac{1}{L} y + \frac{B}{L} \ . \tag{18}$$

The quadratic equation $\varphi\left(y\right) = 0$ has two roots $r_1 = (1+q)\, B$, $r_2 = \left(1 + \frac{1}{q}\right) B$, and $0 < r_1 < r_2$. Then $\varphi\left(y\right) = \frac{C}{2}(y - r_1)(y - r_2)$, $\varphi'\left(y\right) = \frac{C}{2}((y - r_1) + (y - r_2))$, and $\varphi''\left(y\right) = C$.

Starting from $y^0 = 0$, apply the scalar Halley iteration (6) to the function $\varphi(y)$ to get

$$\begin{aligned} y^{k+1} &= y^k - \frac{\varphi\left(y^k\right)}{\varphi'\left(y^k\right) - \frac{1}{2}\frac{\varphi\left(y^k\right)\varphi''\left(y^k\right)}{\varphi'\left(y^k\right)}} \qquad , \ k = 0, 1, 2, \ldots \\ &= y^k - \frac{\frac{C}{2}\left(y^k\right)^2 - \frac{1}{L} y^k + \frac{B}{L}}{C y^k - \frac{1}{L} - \frac{C}{2}\frac{\left(\frac{C}{2}\left(y^k\right)^2 - \frac{1}{L} y^k + \frac{B}{L}\right)}{C y^k - \frac{1}{L}}} \qquad , \ k = 0, 1, 2, .. \end{aligned} \tag{19}$$

obtained by substituting $\varphi'(y) = C y - \dfrac{1}{L}$, $\varphi''(y) = C$.

Next we prove that the sequences $\left\{\mathbf{x}^k\right\}$ and $\{y^k\}$, generated by (8) and (19) respectively, satisfy for $k = 0, 1, \dots$,

$$|f(\mathbf{x}^k)| \leq \varphi(y^k) \ , \tag{20a}$$

$$\|\mathbf{v}^k\| \leq -\frac{1}{\varphi'(y^k)} \ , \tag{20b}$$

$$|\alpha^k| \leq \frac{1}{1 - \frac{M}{2}\frac{\varphi\left(y^k\right)}{(\varphi'\left(y^k\right))^2}} \ , \tag{20c}$$

$$\|\mathbf{x}^{k+1} - \mathbf{x}^k\| \leq y^{k+1} - y^k \ . \tag{20d}$$

Statement (20d) says that $\{y^k\}$ is a majorizing sequence for $\{\mathbf{x}^k\}$. The proof is by induction.

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

$$
\begin{aligned}
|f\left(\mathbf{x}^0\right)| &\leq \frac{B}{L} = \varphi\left(y^0\right) = \varphi(0)\ , \\
\|\mathbf{v}^0\| \leq L &= -\frac{1}{\varphi'\left(y^0\right)}\ , \\
|\alpha^0| &= \left|\frac{1}{1-\frac{1}{2}f\left(\mathbf{x}^0\right)\mathbf{v}^0\cdot f''\left(\mathbf{x}^0\right)\mathbf{v}^0}\right| \leq \frac{1}{1-\frac{1}{2}M\frac{\varphi(y^0)}{(\varphi'(y^0))^2}}, \\
\left\|\mathbf{x}^1-\mathbf{x}^0\right\| &= \left\|\alpha^0\mathbf{v}^0 f\left(\mathbf{x}^0\right)\right\| \leq \left|\alpha^0\right|\,\left\|\mathbf{v}^0\right\|\,\left|f\left(\mathbf{x}^0\right)\right| \\
&\leq \frac{-\frac{\varphi\left(y^0\right)}{\varphi'(y^0)}}{1-\frac{1}{2}M\frac{\varphi(y^0)}{(\varphi'(y^0))^2}} \leq \frac{-\frac{\varphi\left(y^0\right)}{\varphi'(y^0)}}{1-\frac{1}{2}C\frac{\varphi(y^0)}{(\varphi'(y^0))^2}} = y^1-y^0,
\end{aligned}
$$

showing that equations (20a)-(20d) hold for $k = 0$. Suppose (20a)-(20d) hold for $k \leq n$.
<u>Proof of (20a) for $n + 1$</u>:

$$
\begin{aligned}
\left\|\mathbf{x}^{n+1}-\mathbf{x}^0\right\| &= \left\|\sum_{k=0}^{n}\left(\mathbf{x}^{k+1}-\mathbf{x}^k\right)\right\| \leq \sum_{k=0}^{n}\left(y^{k+1}-y^k\right) = y^{n+1}-y^0 = y^{n+1} \leq r_1\ . \\
\therefore\ \mathbf{x}^{n+1} &\in X_0\ . \\
\text{Let } a_n(\mathbf{x}) &:= (\mathbf{x}-\mathbf{x}^n)\left(\nabla f(\mathbf{x}^n)\cdot\mathbf{d}^n - \frac{f(\mathbf{x}^n)}{2\nabla f(\mathbf{x}^n)\cdot\mathbf{d}^n}\,\mathbf{d}^n\cdot f''(\mathbf{x}^n)\,\mathbf{d}^n\right) + f(\mathbf{x}^n)\mathbf{d}^n\ . \\
\therefore\ a_n(\mathbf{x}) &= (\mathbf{x}-\mathbf{x}^{n+1})\left(\nabla f(\mathbf{x}^n)\cdot\mathbf{d}^n - \frac{f(\mathbf{x}^n)}{2\nabla f(\mathbf{x}^n)\cdot\mathbf{d}^n}\,\mathbf{d}^n\cdot f''(\mathbf{x}^n)\,\mathbf{d}^n\right) \\
&\quad + (\mathbf{x}^{n+1}-\mathbf{x}^n)\left(\nabla f(\mathbf{x}^n)\cdot\mathbf{d}^n - \frac{f(\mathbf{x}^n)}{2\nabla f(\mathbf{x}^n)\cdot\mathbf{d}^n}\,\mathbf{d}^n\cdot f''(\mathbf{x}^n)\,\mathbf{d}^n\right) + f(\mathbf{x}^n)\mathbf{d}^n \\
&= (\mathbf{x}-\mathbf{x}^{n+1})\left(\nabla f(\mathbf{x}^n)\cdot\mathbf{d}^n - \frac{f(\mathbf{x}^n)}{2\nabla f(\mathbf{x}^n)\cdot\mathbf{d}^n}\,\mathbf{d}^n\cdot f''(\mathbf{x}^n)\,\mathbf{d}^n\right)\ . \\
\text{Let } p_n(\mathbf{x}) &:= \frac{\nabla f(\mathbf{x}^n)}{\nabla f(\mathbf{x}^n)\cdot\mathbf{d}^n}\cdot a_n(\mathbf{x}) \\
&= (\mathbf{x}-\mathbf{x}^n)\cdot\left(\nabla f\left(\mathbf{x}^n\right) - \frac{f(\mathbf{x}^n)}{2\nabla f(\mathbf{x}^n)\cdot\mathbf{d}^n}\,f''(\mathbf{x}^n)\,\mathbf{d}^n\right) + f\left(\mathbf{x}^n\right)\ .
\end{aligned}
$$

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(\mathbf{x}^{n+1}) &= f\left(\mathbf{x}^n\right) + \nabla f\left(\mathbf{x}^n\right)\cdot\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) - \frac{f(\mathbf{x}^n)}{2\nabla f(\mathbf{x}^n)\cdot\mathbf{d}^n}\,\mathbf{d}^n\cdot f''(\mathbf{x}^n)\,\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\\
&\quad +\frac{1}{2}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\left(\mathbf{x}^n\right)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) - \frac{1}{2}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\left(\mathbf{x}^n\right)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\\
&= f\left(\mathbf{x}^n\right) + \nabla f\left(\mathbf{x}^n\right)\cdot\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) + \frac{1}{2}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\left(\mathbf{x}^n\right)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\\
&\quad -\frac{1}{2}\left(\frac{f\left(\mathbf{x}^n\right)\mathbf{d}^n}{\nabla f(\mathbf{x}^n)\cdot\mathbf{d}^n} + \mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\left(\mathbf{x}^n\right)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\\
&= f\left(\mathbf{x}^n\right) + \nabla f\left(\mathbf{x}^n\right)\cdot\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) + \frac{1}{2}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\left(\mathbf{x}^n\right)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\\
&\quad -\frac{1}{4}\frac{f\left(\mathbf{x}^n\right)\mathbf{d}^n\cdot f''(\mathbf{x}^n)\,\mathbf{d}^n}{(\nabla f(\mathbf{x}^n)\cdot\mathbf{d}^n)^2}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot 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)\cdot\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) + \frac{1}{2}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\left(\mathbf{x}^n\right)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\\
&\quad -\frac{1}{4}\frac{f\left(\mathbf{x}^n\right)\mathbf{d}^n\cdot f''(\mathbf{x}^n)\,\mathbf{d}^n}{(\nabla f(\mathbf{x}^n)\cdot\mathbf{d}^n)^2}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\left(\mathbf{x}^n\right)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right).
\end{aligned}$$

Thus,

$$\begin{aligned}
f\left(\mathbf{x}^{n+1}\right) &= f\left(\mathbf{x}^{n+1}\right) - f\left(\mathbf{x}^n\right) - \nabla f\left(\mathbf{x}^n\right)\cdot\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) - \frac{1}{2}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\left(\mathbf{x}^n\right)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\\
&\quad +\frac{1}{4}\frac{f\left(\mathbf{x}^n\right)\mathbf{d}^n\cdot f''(\mathbf{x}^n)\,\mathbf{d}^n}{(\nabla f(\mathbf{x}^n)\cdot\mathbf{d}^n)^2}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\left(\mathbf{x}^n\right)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right). \qquad (21)
\end{aligned}$$

So, by induction

$$\begin{aligned}
\left|f\left(\mathbf{x}^{n+1}\right)\right| &\le \frac{N}{6}\left\|\mathbf{x}^{n+1}-\mathbf{x}^n\right\|^3 + \frac{M^2}{4}\frac{\left|f\left(\mathbf{x}^n\right)\right|}{(\nabla f(\mathbf{x}^n)\cdot\mathbf{d}^n)^2}\left\|\mathbf{x}^{n+1}-\mathbf{x}^n\right\|^2\\
&\le \left\{\frac{N}{6}\left\|\mathbf{x}^{n+1}-\mathbf{x}^n\right\| + \frac{M^2}{4}\frac{\varphi\left(y^n\right)}{\left(\varphi'\left(y^n\right)\right)^2}\right\}\left(y^{n+1}-y^n\right)^2,\ \text{by induction}\\
&= \left\{\frac{N}{6}\left|\alpha^n\right|\left\|\mathbf{v}^n\right\|\left|f\left(\mathbf{x}^n\right)\right| + \frac{M^2}{4}\frac{\varphi\left(y^n\right)}{\left(\varphi'\left(y^n\right)\right)^2}\right\}\left(y^{n+1}-y^n\right)^2\\
&\le \left\{\frac{N}{6}\frac{-\frac{\varphi(y^n)}{\varphi'(y^n)}}{1-\frac{M}{2}\frac{\varphi(y^n)}{(\varphi'(y^n))^2}} + \frac{M^2}{4}\frac{\varphi\left(y^n\right)}{\left(\varphi'\left(y^n\right)\right)^2}\right\}\left(y^{n+1}-y^n\right)^2\\
&= \left\{\frac{N}{6}\frac{-\varphi'\left(y^n\right)}{1-\frac{M}{2}\frac{\varphi(y^n)}{(\varphi'(y^n))^2}} + \frac{M^2}{4}\right\}\frac{\varphi\left(y^n\right)}{\left(\varphi'\left(y^n\right)\right)^2}\left(y^{n+1}-y^n\right)^2\ .
\end{aligned}$$

The following result is analogous to (21), and is proved the same way

$$\varphi\left(y^{n+1}\right) = \frac{C^2}{4}\frac{\varphi\left(y^n\right)}{\left(\varphi'\left(y^n\right)\right)^2}\left(y^{n+1}-y^n\right)^2. \qquad (22)$$

For any $y \in [0, r_1]$ the function $\frac{\varphi(y)}{(\varphi'(y))^2}$ is decreasing, since

$$\begin{aligned}
\left(\frac{\varphi(y)}{(\varphi'(y))^2}\right)' &= \frac{(r_2 - r_1)^2}{\frac{C}{2}(2y - r_1 - r_2)^3} < 0, \ y \in [0, r_1] , \\
\text{and } \frac{\varphi(y^n)}{(\varphi'(y^n))^2} &\leq \frac{\varphi(y^0)}{(\varphi'(y^0))^2} . \\
\therefore \ \frac{-\varphi'(y^n)}{1 - \frac{M}{2}\frac{\varphi(y^n)}{(\varphi'(y^n))^2}} &\leq \frac{-\varphi'(y^0)}{1 - \frac{M}{2}\frac{\varphi(y^0)}{(\varphi'(y^0))^2}} = \frac{1}{L(1 - \frac{M}{2}LB)} . \qquad (23)
\end{aligned}$$

Consequently,

$$\begin{aligned}
\left|f\left(\mathbf{x}^{n+1}\right)\right| &\leq \frac{1}{4}\left(M^2 + \frac{2}{3}\frac{N}{L\left(1 - \frac{MBL}{2}\right)}\right)\frac{\varphi(y^n)}{(\varphi'(y^n))^2}\left(y^{n+1} - y^n\right)^2, \ \text{by (23)} \\
&= \frac{C^2}{4}\frac{\varphi(y^n)}{(\varphi'(y^n))^2}\left(y^{n+1} - y^n\right)^2 = \varphi\left(y^{n+1}\right), \ \text{by (22)} \ .
\end{aligned}$$

Proof of (20b) for $n+1$:

$$\begin{aligned}
\left\|\mathbf{v}^{n+1}\right\| &= \frac{1}{|\nabla f(\mathbf{x}^{n+1}) \cdot \mathbf{d}^{n+1}|} = \frac{1}{|\nabla f(\mathbf{x}^0) \cdot \mathbf{d}^{n+1} - (\nabla f(\mathbf{x}^0) - \nabla f(\mathbf{x}^{n+1})) \cdot \mathbf{d}^{n+1}|} \\
&= \frac{1}{|\nabla f(\mathbf{x}^0) \cdot \mathbf{d}^{n+1}| \left|1 - \frac{(\nabla f(\mathbf{x}^0) - \nabla f(\mathbf{x}^{n+1})) \cdot \mathbf{d}^{n+1}}{\nabla f(\mathbf{x}^0) \cdot \mathbf{d}^{n+1}}\right|} \\
&\leq \frac{1}{|\nabla f(\mathbf{x}^0) \cdot \mathbf{d}^0| \left|1 - \frac{(\nabla f(\mathbf{x}^0) - \nabla f(\mathbf{x}^{n+1})) \cdot \mathbf{d}^{n+1}}{\nabla f(\mathbf{x}^0) \cdot \mathbf{d}^0}\right|} , \ \text{by (17)} , \\
&\leq \frac{L}{1 - LM\left\|\mathbf{x}^{n+1} - \mathbf{x}^0\right\|} , \ \text{by Lemma 2} , \\
&\leq \frac{L}{1 - LCy^{n+1}} = \frac{1}{\frac{1}{L} - Cy^{n+1}} = \frac{1}{-\varphi'(y^{n+1})} .
\end{aligned}$$

Proof of (20c) for $n+1$:

$$\begin{aligned}
\left|\alpha^{n+1}\right| &= \frac{1}{\left|1 - \frac{1}{2}f\left(\mathbf{x}^{n+1}\right)\mathbf{v}^{n+1} \cdot f''\left(\mathbf{x}^{n+1}\right)\mathbf{v}^{n+1}\right|} \\
&\leq \frac{1}{1 - \frac{M}{2}\frac{\varphi(y^{n+1})}{(\varphi'(y^{n+1}))^2}}, \ \text{by } (20a) \text{ and } (20b).
\end{aligned}$$

Proof of (20d) for $n+1$:

$$\begin{aligned}
\left\|\mathbf{x}^{n+2} - \mathbf{x}^{n+1}\right\| &= \left\|\alpha^{n+1}f(\mathbf{x}^{n+1})\mathbf{v}^{n+1}\right\| \leq \frac{-\frac{\varphi\left(y^{n+1}\right)}{\varphi'(y^{n+1})}}{1 - \frac{M}{2}\frac{\varphi(y^{n+1})}{(\varphi'(y^{n+1}))^2}}, \ \text{by } (20a), (20b) \text{ and } (20c). \\
&\leq \frac{-\frac{\varphi\left(y^{n+1}\right)}{\varphi'(y^{n+1})}}{1 - \frac{C}{2}\frac{\varphi(y^{n+1})}{(\varphi'(y^{n+1}))^2}} = y^{n+2} - y^{n+1}.
\end{aligned}$$

Consequently, (20a)–(20d) hold for all $n \geq 0$. Since the sequence $\left\{\mathbf{x}^k\right\}$ is majorized by the sequence $\left\{y^k\right\}$ it follows from Lemma 2 that $\lim\limits_{k\to\infty} \mathbf{x}^k = \mathbf{x}^*$ and $\mathbf{x}^* \in X_0$.

The scalar Halley method has a cubic rate of convergence, [8], and as shown in [9], $\left|y^{k+1} - y^k\right| \leq \beta\frac{\theta^{3^k}}{1-\theta^{3^k}}$, where $\theta < 1$. Therefore $\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\| \leq \beta\frac{\theta^{3^k}}{1-\theta^{3^k}}$, and the sequence $\{\mathbf{x}^k\}$ has at least cubic rate of convergence. □

Remark 1. Condition (16) says that all directions $\mathbf{d}^i$ are as close, in angle, to the initial gradient $\nabla f(\mathbf{x}^0)$ as the initial direction $\mathbf{d}^0$. Equivalently, all directions $\mathbf{d}^i$ are in a circular cone with axis $\nabla f(\mathbf{x}^0)$, generated by $\mathbf{d}^0$. In particular, (16) holds if the directions are fixed

$$\mathbf{d}^i = \mathbf{d}^0 \ , \ i = 1, 2, \cdots \ , \tag{24}$$

in which case, (8) reduces to the scalar Halley method (6) for the function $F(t) := f(\mathbf{x}^0 + t\mathbf{d}^0)$, along the line $\mathbf{L} := \{\mathbf{x}^0 + t\mathbf{d} : t \in \mathbb{R}\}$. The scalar quasi–Halley method (7) can also be used for $F(t)$ along $\mathbf{L}$.

3. A gradient–directional Halley method

To prove the convergence of (9), we write that iteration as

$$\mathbf{x}^{k+1} := \mathbf{x}^k - \alpha^k f\left(\mathbf{x}^k\right)\mathbf{v}^k \ , \text{ where} \tag{25a}$$

$$\mathbf{v}^k := \frac{\nabla f(\mathbf{x}^k)}{\|\nabla f(\mathbf{x}^k)\|^2} \ , \text{ and} \tag{25b}$$

$$\alpha^k := \frac{1}{1 - \frac{1}{2} f(\mathbf{x}^k)\, \mathbf{v}^k \cdot f''(\mathbf{x}^k)\mathbf{v}^k} \ . \tag{25c}$$

The following theorem is analogous to Theorem 1. The proof is given in Appendix A.

Theorem 2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a three times differentiable function, $\mathbf{x}^0 \in \mathbb{R}^n$, and assume that

$$\sup_{\mathbf{x}\in X_0} \|f''(\mathbf{x})\| = M \ , \tag{26a}$$

$$\sup_{\mathbf{x}\in X_0} \|f'''(\mathbf{x})\| = N \ , \tag{26b}$$

where X_0 is defined in (27d) below. Let there exist constants B, L such that

$$\left\|\nabla f(\mathbf{x}^0)\right\| \geq \frac{1}{L} \ , \tag{27a}$$

$$|f(\mathbf{x}^0)| \leq \frac{B}{L} \ , \tag{27b}$$

$$T := CLB < \frac{1}{2} \ , \text{ where } C := \sqrt{M^2 + \frac{2}{3}\frac{N}{L(1 - \frac{1}{2}MLB)}} \ , \tag{27c}$$

$$\text{and finally let } X_0 := \left\{\mathbf{x} := \left\|\mathbf{x} - \mathbf{x}^0\right\| \leq (1+q)\, B\right\} \ , \text{ where } q = \frac{1 - \sqrt{1 - 2T}}{1 + \sqrt{1 - 2T}} \ . \tag{27d}$$

Then:

(a) All the points $\mathbf{x}^{k+1} := \mathbf{x}^k - \alpha^k f\left(\mathbf{x}^k\right)\mathbf{v}^k$, $k = 0, 1, \cdots$ lie in X_0.

(b) $\mathbf{x}^* = \lim\limits_{k\to\infty} \mathbf{x}^k$ exists, $\mathbf{x}^* \in X_0$, and $f\left(\mathbf{x}^*\right) = 0$.

(c) The order of convengence of the Halley method (9) is cubic.

4. Directional Quasi-Halley Iteration

Consider the Taylor expansion of $f(x^k - \frac{f(\mathbf{x}^k)}{\|\nabla f(\mathbf{x}^k)\|^2}\nabla f(\mathbf{x}^k))$:

$$
\begin{aligned}
f(x^k - \frac{f(\mathbf{x}^k)}{\|\nabla f(\mathbf{x}^k)\|^2}\nabla f(\mathbf{x}^k)) &= f(\mathbf{x}^k) - \nabla f(\mathbf{x}^k)\cdot\frac{f(\mathbf{x}^k)}{\|\nabla f(\mathbf{x}^k)\|^2}\nabla f(\mathbf{x}^k) \\
&\quad + \frac{(f(\mathbf{x}^k))^2\,\nabla f(\mathbf{x}^k)\cdot f''(\mathbf{x}^k)\,\nabla f(\mathbf{x}^k)}{2\,\|\nabla f(\mathbf{x}^k)\|^4} + O\left(\frac{|f(\mathbf{x}^k)|^2}{||\nabla f(\mathbf{x}^k)||^2}\right) \\
&= \frac{f(\mathbf{x}^k)}{\|\nabla f(\mathbf{x}^k)\|^2}\frac{f(\mathbf{x}^k)\,\nabla f(\mathbf{x}^k)\cdot f''(\mathbf{x}^k)\,\nabla f(\mathbf{x}^k)}{2\,\|\nabla f(\mathbf{x}^k)\|^2} + O\left(||\mathbf{u}^k||^2\right)\ ,\ \text{see (11a)}\ .
\end{aligned}
$$

Multiplying by $\frac{\|\nabla f(\mathbf{x}^k)\|^2}{f(\mathbf{x}^k)}$ we get

$$
\frac{\|\nabla f(\mathbf{x}^k)\|^2}{f(\mathbf{x}^k)}\,f(x^k - \frac{f(\mathbf{x}^k)}{\|\nabla f(\mathbf{x}^k)\|^2}\nabla f(\mathbf{x}^k)) = \frac{f(\mathbf{x}^k)\,\nabla f(\mathbf{x}^k)\cdot f''(\mathbf{x}^k)\nabla f(\mathbf{x}^k)}{2\,\|\nabla f(\mathbf{x}^k)\|^2} + O(|f(\mathbf{x}^k)|)\ ,
$$

showing that

$$
\frac{f(\mathbf{x}^k)\nabla f(\mathbf{x}^k)\cdot f''(\mathbf{x}^k)\nabla f(\mathbf{x}^k)}{2\,\|\nabla f(\mathbf{x}^k)\|^2}\ \text{can be approximated by}\ \frac{\|\nabla f(\mathbf{x}^k)\|^2}{f(\mathbf{x}^k)}f(x^k - \frac{f(\mathbf{x}^k)}{\|\nabla f(\mathbf{x}^k)\|^2}\nabla f(\mathbf{x}^k))
$$

if $|f(\mathbf{x}^k)|$ is sufficiently small. Substituting this approximation to (9), we get the following iteration

$$
\mathbf{x}^{k+1} := \mathbf{x}^k - \frac{f(\mathbf{x}^k)}{\|\nabla f(\mathbf{x}^k)\|^2\left(1 - \dfrac{f\left(\mathbf{x}^k - \frac{f(\mathbf{x}^k)}{\|\nabla f(\mathbf{x}^k)\|^2}\nabla f(\mathbf{x}^k)\right)}{f(\mathbf{x}^k)}\right)}\nabla f(\mathbf{x}^k)\ ,\ k = 0, 1, \ldots \tag{28a}
$$

$$
= \mathbf{x}^k - \frac{f(\mathbf{x}^k)}{f(\mathbf{x}^k + \mathbf{u}^k) - f(\mathbf{x}^k)}\,\mathbf{u}^k\ ,\ k = 0, 1, \ldots \tag{28b}
$$

This **quasi-Halley method** does not reduce to its scalar counterpart (7) for $n = 1$.

5. Comparison of steps

We use the Taylor expansion of f, see [2, (8.14.3)],

$$
f(\mathbf{x} + \mathbf{u}) = f(\mathbf{x}) + f'(\mathbf{x})\cdot\mathbf{u} + \frac{1}{2}\,f''(\mathbf{x})\cdot\mathbf{u}^{(2)} + \frac{1}{3!}\,f'''(\xi)\cdot\mathbf{u}^{(3)}\ ,
$$

for some ξ betwee $\mathbf{x}$ and $\mathbf{x} + \mathbf{u}$, that we write as

$$
f(\mathbf{x} + \mathbf{u}) = f(\mathbf{x}) + \nabla f(\mathbf{x})\cdot\mathbf{u} + \frac{1}{2}\,\mathbf{u}\cdot f''(\mathbf{x})\mathbf{u} + \frac{1}{3!}\,f'''(\xi)\cdot\mathbf{u}^{(3)}\ . \tag{29}
$$

Two iterative methods are comparable locally if at a given point they produce comparable steps. We will compare the steps in terms of length.

We first compare the steps of the Newton and Halley methods, assuming both steps emanate from the same point $\mathbf{x}^k$, arrived at by Newton's method. This corresponds to a hypothetical situation where at an iterate $\mathbf{x}^k$ of the Newton method we have an option continuing (and making a Newton step) or switching to the Halley method (9), making a *Halley step*

To simplify the writing we denote by f_k the function f evaluated at $\mathbf{x}^k$. Similarly, ∇f_k and f''_k denote the gradient ∇f and the Hessian f'' evaluated at $\mathbf{x}^k$. The steps to be compared are the Newton step $\mathbf{u}^k$ and the Halley step $\mathbf{h}^k$, see (11a)–(11b), written as

$$\mathbf{u}^k = -\frac{f_k}{\|\nabla f_k\|^2}\nabla f_k \quad \text{and} \quad \mathbf{h}^k = -\frac{f_k}{\|\nabla f_k\|^2 - \frac{f_k \cdot \nabla f_k \cdot f''_k \cdot \nabla f_k}{2 \cdot \|\nabla f_k\|^2}}\nabla f_k .$$

The next lemma gives a condition for the Newton Halley steps to have the same sign.

Lemma 3. *The steps $\mathbf{u}^k$ and $\mathbf{h}^k$ have the same sign if and only if*

$$\|\nabla f_k\|^2 > \frac{f_k \nabla f_k \cdot f''_k \nabla f_k}{2\,\|\nabla f_k\|^2}, \tag{30}$$

in which case

$$\|\mathbf{h}^k\| \geq \|\mathbf{u}^k\| \quad \text{if } f_k \nabla f_k \cdot f''_k \nabla f_k \geq 0, \tag{31a}$$
$$\|\mathbf{h}^k\| < \|\mathbf{u}^k\| \quad \text{if } f_k \nabla f_k \cdot f''_k \nabla f_k < 0 . \tag{31b}$$

Proof. The steps $\mathbf{u}^k$ and $\mathbf{h}^k$ have the same sign if and only if

$$\begin{aligned}
\|\nabla f_k\|^2 - \frac{f_k \nabla f_k \cdot f''_k \nabla f_k}{2\,\|\nabla f_k\|^2} &\geq 0 , \\
\text{that is } \|\nabla f_k\|^2 &\geq \frac{f_k \nabla f_k \cdot f''_k \nabla f_k}{2\,\|\nabla f_k\|^2}, \\
\text{in which case } \|\nabla f_k\|^2 &\geq \left| \|\nabla f_k\|^2 - \frac{f_k \nabla f_k \cdot f''_k \nabla f_k}{2\,\|\nabla f_k\|^2} \right| \quad \text{if} \quad f_k \nabla f_k \cdot f''_k \nabla f_k \geq 0 , \\
\|\nabla f_k\|^2 &< \left| \|\nabla f_k\|^2 - \frac{f_k \nabla f_k \cdot f''_k \nabla f_k}{2\,\|\nabla f_k\|^2} \right| \quad \text{if} \quad f_k \nabla f_k \cdot f''_k \nabla f_k < 0 .
\end{aligned}$$

□

The point $\mathbf{x}^k$ where the steps $\mathbf{u}^k$ and $\mathbf{h}^k$ are compared is arrived at by the Newton method. It is therefore reasonable to assume that the following conditions hold, see [6, eq. 10a-10b]

$$\|f''_k\| \leq M , \tag{32a}$$
$$\|\nabla f_k\| \geq 2\,\|\mathbf{u}^k\|\, M, \; k = 1, 2, ... \tag{32b}$$

Theorem 3. *If conditions (32a) and (32b) hold, then the Newton step $\mathbf{u}^k$ and the Halley step $\mathbf{h}^k$ have the same sign, and are related by*

$$\frac{2}{3}\,\|\mathbf{u}^k\| \leq \|\mathbf{h}^k\| \leq \frac{4}{3}\,\|\mathbf{u}^k\| . \tag{33}$$

Proof.

$$||\nabla f_k|| \geq 2\left\|\mathbf{u}^k\right\| M = 2\frac{|f_k|}{\|\nabla f_k\|}M, \text{ by } (32b)$$

$$\therefore \|\nabla f_k\|^2 \geq 2\,|f_k|\,M \geq 2\,|f_k|\,\|f_k''\| \geq 2\,|f_k|\,\frac{\nabla f_k \cdot f_k'' \nabla f_k}{\|\nabla f_k\|^2}, \text{ by } (32a).$$

Therefore (30) holds, showing that $\mathbf{u}^k$ and $\mathbf{h}^k$ have the same sign. Then

$$\begin{aligned}\mathbf{u}^k - \mathbf{h}^k &= \frac{f_k}{\|\nabla f_k\|^2 - \frac{f_k \nabla f_k \cdot f_k'' \nabla f_k}{2\|\nabla f_k\|^2}}\nabla f_k - \frac{f_k}{\|\nabla f_k\|^2}\nabla f_k \\ &= \frac{(f_k)^2(\frac{\nabla f_k \cdot f_k'' \nabla f_k}{2\|\nabla f_k\|^2})}{\|\nabla f_k\|^2\,(\|\nabla f_k\|^2 - \frac{f_k \nabla f_k \cdot f_k'' \nabla f_k}{2\|\nabla f_k\|^2})}\nabla f_k\end{aligned}$$

$$\therefore \; ||\mathbf{u}^k - \mathbf{h}^k|| = \frac{(f_k)^2\frac{\left|\nabla f_k \cdot f_k'' \nabla f_k\right|}{2\|\nabla f_k\|^2}}{\|\nabla f_k\|\left|\|\nabla f_k\|^2 - \frac{f_k \nabla f_k \cdot f_k'' \nabla f_k}{2\|\nabla f_k\|^2}\right|}. \tag{34}$$

From (32a) and (32b) it follows that $|\nabla f_k \cdot f_k'' \nabla f_k| \leq M \|\nabla f_k\|^2$, and $\frac{|f_k|M}{2\|\nabla f_k\|^2} \leq \frac{1}{4}, \quad k = 0, 1, 2, \ldots$, which substituted in (34) gives

$$||\mathbf{u}^k - \mathbf{h}^k|| \leq \frac{(f_k)^2\frac{M}{2}}{\|\nabla f_k\|^3\,(1 - \frac{|f_k|M}{2\|\nabla f_k\|^2})} \leq \frac{(f_k)^2\frac{M}{2}}{\|\nabla f_k\|^3\,|f_k|\,M}\frac{4}{3} = \frac{1}{3}\frac{f_k}{\|\nabla f_k\|} = \frac{||\mathbf{u}^k||}{3},$$

proving (33). ☐

The Halley step $\mathbf{h}^k$ therefore lies in the interval with endpoints $\frac{2}{3}\mathbf{u}^k$ and $\frac{4}{3}\mathbf{u}^k$.
We next compare the Halley step $\mathbf{h}^k$ and the quasi-Halley step (11c)

$$\mathbf{q}^k = -\frac{f(\mathbf{x}^k)}{\|\nabla f(\mathbf{x}^k)\|^2\left(1 - \frac{f(\mathbf{x}^k + \mathbf{u}^k)}{f(\mathbf{x}^k)}\right)}\nabla f(\mathbf{x}^k)\,,\; k = 0, 1, \ldots \tag{35}$$

evaluated at the same point $\mathbf{x}^k$ where $f(\mathbf{x}^k) \neq 0$. Numerical experience shows that, close to a root to which both methods converge, the Halley step and the quasi-Halley step are very close. This is explained by the the following theorem.

Theorem 4. *Let f have continuous third derivative in the interval X_0, and let*

$$\sup_{\mathbf{x} \in X_0} \|f'''(\mathbf{x})\| = N.$$

If conditions (32a) and (32b) hold, then

$$\left\|\mathbf{h}^k - \mathbf{q}^k\right\| \leq \frac{4}{27}\frac{N}{M}\left\|\mathbf{u}^k\right\|^2, \tag{36}$$

i.e. the difference between the Halley step $\mathbf{h}^k$ and the quasi-Halley step $\mathbf{q}^k$ is of order $O(||\mathbf{u}^k||^2)$, where $\mathbf{u}^k$ is the Newton step at the current point.

Proof. Let $\mathbf{g}_k := \nabla f_k$. Then

$$
\begin{aligned}
\mathbf{q}^k - \mathbf{h}^k &= \frac{f_k \mathbf{g}_k}{\|\mathbf{g}_k\|^2}\left(\frac{1}{1-\frac{f_k\,\mathbf{g}_k\cdot f_k''\mathbf{g}_k}{2\|\mathbf{g}_k\|^4}} - \frac{1}{1-\frac{f(\mathbf{x}^k+\mathbf{u}^k)}{f_k}}\right)\\
&= -\frac{\mathbf{g}_k}{\|\mathbf{g}_k\|^2}\,\frac{f(\mathbf{x}^k+\mathbf{u}^k) - \frac{(f_k)^2\,\mathbf{g}_k\cdot f_k''\,\mathbf{g}_k}{2\|\mathbf{g}_k\|^4}}{\left(1-\frac{f_k\,\mathbf{g}_k\cdot f_k''\,\mathbf{g}_k}{2\|\mathbf{g}_k\|^4}\right)\left(1-\frac{f(\mathbf{x}^k+\mathbf{u}^k)}{f_k}\right)}\\
&= -\frac{\mathbf{g}_k}{\|\mathbf{g}_k\|^2}\,\frac{f_k + \mathbf{g}_k\cdot\mathbf{u}^k + \frac{1}{2}\,\mathbf{u}^k\cdot f_k''\,\mathbf{u}^k + \frac{1}{6}f'''(\xi)\cdot(\mathbf{u}^k)^{(3)} - \frac{(f_k)^2\,\mathbf{g}_k\cdot f_k''\,\mathbf{g}_k}{2\|\mathbf{g}_k\|^4}}{\left(1-\frac{f_k\,\mathbf{g}_k\cdot f_k''\,\mathbf{g}_k}{2\|\mathbf{g}_k\|^4}\right)\left(1-\frac{f(\mathbf{x}^k+\mathbf{u}^k)}{f_k}\right)}\ ,\\
&\qquad \text{for some } \xi \text{ between } \mathbf{x}^k+\mathbf{u}^k \text{ and } \mathbf{x}^k\ ,\\
&= -\frac{\mathbf{g}_k}{\|\mathbf{g}_k\|^2}\,\frac{\frac{1}{6}\,f'''(\xi)\cdot(\mathbf{u}^k)^{(3)}}{\left(1-\frac{f_k\,\mathbf{g}_k\cdot f_k''\,\mathbf{g}_k}{2\|\mathbf{g}_k\|^4}\right)\left(1-\frac{f(\mathbf{x}^k+\mathbf{u}^k)}{f_k}\right)}\ .
\end{aligned}
$$

Note,

$$
\begin{aligned}
\frac{\left|f(\mathbf{x}^k+\mathbf{u}^k)\right|}{|f_k|} &\le \frac{\left|\frac{1}{2}\,\mathbf{u}^k\cdot f_k''(\theta)\,\mathbf{u}^k\right|}{|f_k|}\ , \text{ for some } \theta \text{ between } \mathbf{x}^k+\mathbf{u}^k \text{ and } \mathbf{x}^k\ ,\\
&\le \frac{\frac{1}{2}\,M\,\left\|\mathbf{u}^k\right\|^2}{|f_k|} \le \frac{1}{4}\ , \text{ since } |f_k| \ge 2M\left\|\mathbf{u}^k\right\|^2\ .
\end{aligned}
$$

$$
\begin{aligned}
\therefore\ \left\|\mathbf{h}^k-\mathbf{q}^k\right\| &\le \frac{\frac{1}{6}N\left\|\mathbf{u}^k\right\|^3}{\|\mathbf{g}_k\|\left|1-\frac{f_k\,\mathbf{g}_k\cdot f_k''\mathbf{g}_k}{2\|\mathbf{g}_k\|^4}\right|\frac{3}{4}} = \frac{\frac{2}{9}N\left\|\mathbf{u}^k\right\|^3 f_k}{\|\mathbf{g}_k\|\left|f_k - \frac{1}{2}\,\mathbf{u}^k\cdot f_k''\,\mathbf{u}^k\right|} \le \frac{\frac{2}{9}N\left\|\mathbf{u}^k\right\|^4}{|f_k| - \left|\frac{1}{2}\,\mathbf{u}^k\cdot f_k''\,\mathbf{u}^k\right|}\\
&\le \frac{\frac{2}{9}\,N\left\|\mathbf{u}^k\right\|^4}{2M\left\|\mathbf{u}^k\right\|^2 - \frac{1}{2}\,M\left\|\mathbf{u}^k\right\|^2}\ , \text{ since } |f_k| \ge 2M\left\|\mathbf{u}^k\right\|^2\\
&= \frac{4}{27}\frac{N}{M}\left\|\mathbf{u}^k\right\|^2\ .
\end{aligned}
$$

□

The comparison (36) between the Halley step $\mathbf{h}^k$ and the quasi-Halley step $\mathbf{q}^k$ is in terms of the current Newton step $\mathbf{u}^k$. The same comparison in terms of the previous Newton step $\mathbf{u}^{k-1}$ (assuming the current point $\mathbf{x}^k$ is arrived at by the Newton method) gives

$$
\left\|\mathbf{h}^k-\mathbf{q}^k\right\| \le \frac{NM}{27\,\|\mathbf{g}_k\|^2}\left\|\mathbf{u}^{k-1}\right\|^4\ .
$$

Example 1. As in [6], an arbitrary system of m equations in n unknowns:

$$
\mathbf{f}(\mathbf{x}) = \mathbf{0}\ , \text{ or } \quad f_i(x_1, x_2, \ldots, x_n) = 0\ ,\ i = 1, \ldots, m \tag{37}
$$

can be replaced by an equivalent single equation

$$
\sum_{i=1}^{m} f_i^2(\mathbf{x}) = 0\ , \tag{38}
$$

which can be solved by both the Halley and the quasi-Halley methods. An example is given in the Appendix A.5.

6. Numerical experience

The Halley method (9) and the quasi-Halley method (10) were tested on polynomials in n variables, $n = 2, \cdots, 9$, and degree d, $d = 2, \cdots, 10$. The *degree* of a polynomial in n variables is the maximal degree of its terms, e.g. $degree(x^2y^3 + x^4 + y^4) = 5$. For each combination of n and d, 100 random polynomials were generated and solved by both methods, starting with the same initial point $\mathbf{x}^0 = (1, 1, \cdots, 1)$. For comparison, each such polynomial was also solved by the Newton method (4). The stopping rule in all methods was identical, the norm of the step value is less than 10^{-12}, with an upper bound of 30 iterations. The average numbers of iterations of the directional Newton, Halley and quasi-Halley methods, for 100 random polynomials with degree $d = 2, \cdots, 10$ and n variables, $n = 2, \cdots, 9$ are tabulated in Table 1.

Figures 1–2 illustrate two typical sections of Table 1.

In Figure 1 the number of variables is fixed at 5. For each degree from 2 to 10, 100 random polynomials were generated and solved by the three methods, recording the average number of iterations as a function of the degree.

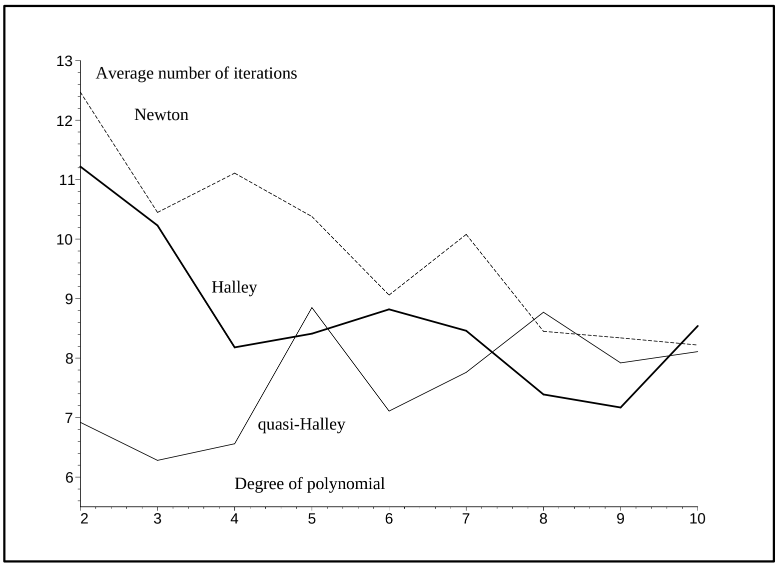

Figure 1. Comparison of the Newton, Halley and quasi-Halley methods for polynomials with 5 variables.

In Figure 2 the polynomial degree is fixed at 5. For each number of variables from 2 to 10, 100 random polynomials were generated and solved by the three methods, recording the average number of iterations as a function of the number of variables.

Number of variables	2			3			4			5		
Degree	**N**	**H**	**q-H**	**N**	**H**	**q-H**	**N**	**H**	**q-H**	**N**	**H**	**q-H**
2	12.45	11.86	7.87	10.05	10.40	7.93	10.46	11.69	6.82	12.47	11.22	6.92
3	11.60	12.25	8.88	10.75	10.98	7.75	11.05	9.66	6.64	10.45	10.23	6.28
4	13.74	11.16	8.09	13.24	10.70	8.15	10.66	10.73	7.45	11.11	8.18	6.56
5	14.30	11.96	9.65	12.38	11.77	8.79	10.16	7.80	8.02	10.38	8.41	8.85
6	12.94	10.48	9.26	11.92	9.82	9.25	10.58	9.26	8.21	9.06	8.82	7.11
7	13.25	10.22	10.11	11.50	9.16	8.00	10.21	7.36	8.06	10.08	8.46	7.76
8	9.47	10.56	8.90	12.54	9.44	9.75	11.69	8.74	7.98	8.45	7.39	8.77
9	10.86	8.99	9.09	10.33	10.17	9.09	10.00	10.23	9.61	8.34	7.17	7.92
10	11.61	10.04	8.99	10.23	10.76	9.76	9.20	7.83	9.32	8.22	8.54	8.11

Number of variables	6			7			8			9		
Degree	**N**	**H**	**q-H**	**N**	**H**	**q-H**	**N**	**H**	**q-H**	**N**	**H**	**q-H**
2	9.00	7.95	7.23	9.12	8.14	8.37	11.20	10.59	9.13	9.82	10.08	7.77
3	9.64	10.67	6.00	10.88	9.36	7.44	10.03	8.38	6.92	9.24	9.92	7.14
4	8.69	7.18	6.25	9.68	8.85	7.46	10.02	9.15	6.95	8.26	7.48	6.96
5	9.93	9.32	6.57	8.89	7.37	8.43	8.45	6.62	6.55	8.54	7.00	7.40
6	9.35	8.32	7.08	7.65	7.41	6.77	7.98	6.55	6.61	7.70	7.52	5.48
7	7.80	7.47	8.24	7.65	6.03	6.38	7.62	8.37	6.66	8.21	6.89	7.80
8	8.79	6.27	6.91	7.66	7.03	7.64	7.33	7.43	7.51	6.75	4.98	7.34
9	8.21	6.03	7.05	8.47	7.20	7.69	7.74	6.58	6.68	8.25	6.66	7.30
10	7.26	7.53	7.26	8.10	6.70	7.45	6.78	6.35	7.14	6.51	5.66	6.38

Table 1
Comparison of the directional Newton (**N**), Halley (**H**) and quasi-Halley (**q-H**) methods in terms of the average number of iterations for 100 random polynomials with the given degree and number of variables.

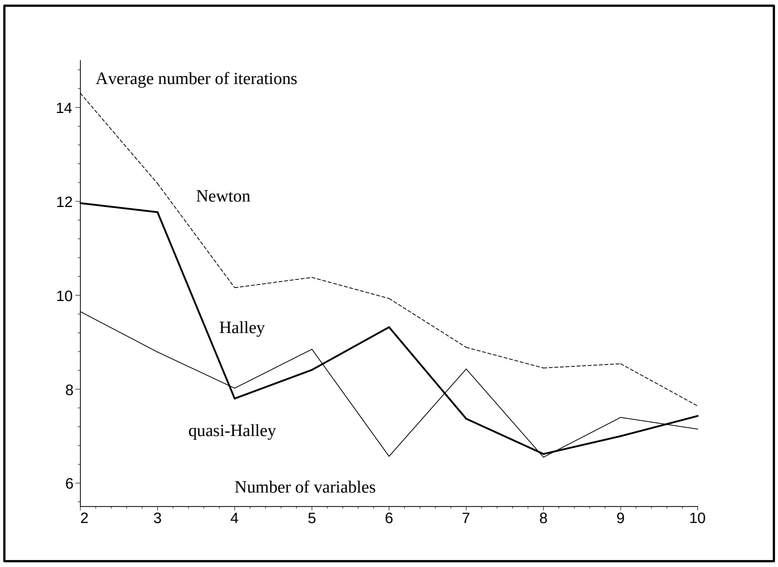

Figure 2. Comparison of the Newton, Halley and quasi-Halley methods for polynomials of degree 5.

7. Conclusions

Three methods for solving a single equation in n unknowns are discussed here:

- the Newton method (4),
- the Halley method (9), and
- the quasi-Halley method (10).

The Newton method is of order 2, see [6, Theorem 1], and requires the evaluation of $f(\mathbf{x})$, $\nabla f(\mathbf{x})$ (the *Newton data*) at each iteration. We use this method as our basis for comparison.

The Halley method is of order 3, see Theorem 1 above, but requires the Hessian $f''(\mathbf{x})$ in addition to the Newton data.

The order of the quasi-Halley method is unknown (the related scalar quasi-Halley method (7) has order $1+\sqrt{2}$, see [1]), however in practice it performs similarly to the Halley method, see the comparison of steps in Theorem 4, and the numerical experience reported in § 6. In terms of work per iteration, the quasi-Halley method requires one more function evaluation than the Newton method, $f(\mathbf{x}+\mathbf{u})$ where $\mathbf{u}$ is the next Newton step. Another advantage of the quasi-Halley method is that it is amenable for parallel implementation, since the matrix $f''(\mathbf{x})$ is avoided.

REFERENCES

1. A. Ben-Israel, Newton's method with modified functions, *Contemp. Math.* **204**(1997), 39–50
2. J. Dieudonné, *Foundations of Modern Analysis*, Academic Press, 1960
3. C.-E. Fröberg, *Numerical Mathematics: Theory and Computer Applications*, (Benjamin, 1985)

4. J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, (Academic Press, 1970)
5. L.V. Kantorovich and G.P. Akilov, *Functional Analysis in Normed Spaces*, (Macmillan, New York, 1964)
6. Y. Levin and A. Ben-Israel, Directional Newton methods in n variables, *Rutcor Research Report* **19**(1999).
7. R.A. Safiev, The methods of tangent hyperbolas, *Soviet Math. Dokl.* **4**(1963), 482-485
8. J.F. Traub, *Iterative Methods for the Solution of Equations*, (Prentice–Hall, 1964)
9. Q. Yao, On Halley iteration, *Numer. Math.* **81**(1999), 647-677

Appendix A: Proof of Theorem 2

Proof. Let the function φ and the sequence $\{y^k\}$ be given by (18) and (19) respectively, and let $\mathbf{v}^k$ and α^k be defined by (25b)–(25c). We show that $\{y^k\}$ is a majorizing sequence for the sequence $\left\{\mathbf{x}^k\right\}$ generated by (9). This is statement (A.1d), proved below for all k. Indeed, we prove for $k = 0, 1, \cdots$,

$$\begin{aligned}
|f(\mathbf{x}^k)| &\leq \varphi(y^k)\ , && \text{(A.1a)}\\
\|\mathbf{v}^k\| &\leq -\frac{1}{\varphi'(y^k)}\ , && \text{(A.1b)}\\
|\alpha^k| &\leq \frac{1}{1-\frac{M}{2}\frac{\varphi\left(y^k\right)}{(\varphi'\left(y^k\right))^2}}\ , && \text{(A.1c)}\\
\|\mathbf{x}^{k+1}-\mathbf{x}^k\| &\leq y^{k+1}-y^k\ . && \text{(A.1d)}
\end{aligned}$$

The proof is by induction.
<u>Verification for $k = 0$</u>:

$$\begin{aligned}
|f\left(\mathbf{x}^0\right)| &\leq \frac{B}{L} = \varphi\left(y^0\right) = \varphi(0)\ ,\\
\|\mathbf{v}^0\| \leq L &= -\frac{1}{\varphi'\left(y^0\right)}\ ,\\
|\alpha^0| &= \left|\frac{1}{1-\frac{1}{2}f\left(\mathbf{x}^0\right)\mathbf{v}^0\cdot f''\left(\mathbf{x}^0\right)\mathbf{v}^0}\right| \leq \frac{1}{1-\frac{1}{2}M\frac{\varphi(y^0)}{(\varphi'(y^0))^2}},\\
\left\|\mathbf{x}^1-\mathbf{x}^0\right\| &= \left\|\alpha^0\mathbf{v}^0 f\left(\mathbf{x}^0\right)\right\| \leq \left|\alpha^0\right|\left\|\mathbf{v}^0\right\|\left|f\left(\mathbf{x}^0\right)\right|\\
&\leq \frac{-\frac{\varphi\left(y^0\right)}{\varphi'(y^0)}}{1-\frac{1}{2}M\frac{\varphi(y^0)}{(\varphi'(y^0))^2}} \leq \frac{-\frac{\varphi\left(y^0\right)}{\varphi'(y^0)}}{1-\frac{1}{2}C\frac{\varphi(y^0)}{(\varphi'(y^0))^2}} = y^1-y^0,
\end{aligned}$$

showing that equations (A.1a)-(A.1d) hold for $k = 0$.
Suppose (A.1a)-(A.1d) hold for $k \leq n$.

Proof of (A.1a) for $n+1$:

$$\begin{aligned}
\left\|\mathbf{x}^{n+1}-\mathbf{x}^0\right\| &= \left\|\sum_{k=0}^{n}\left(\mathbf{x}^{k+1}-\mathbf{x}^k\right)\right\| \le \sum_{k=0}^{n}\left(y^{k+1}-y^k\right) = y^{n+1}-y^0 = y^{n+1} \le r_1 \,. \\
\therefore\ \mathbf{x}^{n+1} &\in X_0 \,. \\
\text{Let } a_n(\mathbf{x}) &:= (\mathbf{x}-\mathbf{x}^n)\,(\|\nabla f\,(\mathbf{x}^n)\|^2 - \frac{f(\mathbf{x}^n)}{2\,\|\nabla f(\mathbf{x}^n)\|^2}\,\nabla f(\mathbf{x}^n)\cdot f''(\mathbf{x}^n)\,\nabla f(\mathbf{x}^n)) + f(\mathbf{x}^n)\nabla f(\mathbf{x}^n) \,. \\
\therefore\ a_n(\mathbf{x}) &= (\mathbf{x}-\mathbf{x}^{n+1})(\|\nabla f\,(\mathbf{x}^n)\|^2 - \frac{f(\mathbf{x}^n)}{2\,\|\nabla f(\mathbf{x}^n)\|^2}\,\nabla f(\mathbf{x}^n)\cdot f''(\mathbf{x}^n)\,\nabla f(\mathbf{x}^n)) \\
&\quad +(\mathbf{x}^{n+1}-\mathbf{x}^n)(\|\nabla f\,(\mathbf{x}^n)\|^2 - \frac{f(\mathbf{x}^n)}{2\,\|\nabla f(\mathbf{x}^n)\|^2}\,\nabla f(\mathbf{x}^n)\cdot f''(\mathbf{x}^n)\,\nabla f(\mathbf{x}^n)) + f(\mathbf{x}^n)\nabla f(\mathbf{x}^n) \\
&= (\mathbf{x}-\mathbf{x}^{n+1})(\|\nabla f\,(\mathbf{x}^n)\|^2 - \frac{f(\mathbf{x}^n)}{2\,\|\nabla f(\mathbf{x}^n)\|^2}\,\nabla f(\mathbf{x}^n)\cdot f''(\mathbf{x}^n)\,\nabla f(\mathbf{x}^n)) \,. \\
\text{Let } p_n(\mathbf{x}) &:= \frac{\nabla f(\mathbf{x}^n)}{\|\nabla f(\mathbf{x}^n)\|^2}\cdot a_n(\mathbf{x}) \\
&= (\mathbf{x}-\mathbf{x}^n)\cdot\left(\nabla f\,(\mathbf{x}^n) - \frac{f(\mathbf{x}^n)}{2\,\|\nabla f(\mathbf{x}^n)\|^2}\,f''(\mathbf{x}^n)\,\nabla f(\mathbf{x}^n)\right) + f\,(\mathbf{x}^n) \,.
\end{aligned}$$

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

$$\begin{aligned}
p_n(\mathbf{x}^{n+1}) &= f\,(\mathbf{x}^n) + \nabla f\,(\mathbf{x}^n)\cdot\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) - \frac{f(\mathbf{x}^n)}{2\,\|\nabla f(\mathbf{x}^n)\|^2}\,\nabla f(\mathbf{x}^n)\cdot f''(\mathbf{x}^n)\,\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) \\
&\quad +\frac{1}{2}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\,(\mathbf{x}^n)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) - \frac{1}{2}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\,(\mathbf{x}^n)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) \\
&= f\,(\mathbf{x}^n) + \nabla f\,(\mathbf{x}^n)\cdot\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) + \frac{1}{2}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\,(\mathbf{x}^n)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) \\
&\quad -\frac{1}{2}\left(\frac{f\,(\mathbf{x}^n)\,\nabla f\,(\mathbf{x}^n)}{\|\nabla f(\mathbf{x}^n)\|^2} + \mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\,(\mathbf{x}^n)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) \\
&= f\,(\mathbf{x}^n) + \nabla f\,(\mathbf{x}^n)\cdot\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) + \frac{1}{2}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\,(\mathbf{x}^n)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) \\
&\quad -\frac{1}{4}\frac{f\,(\mathbf{x}^n)\,\nabla f\,(\mathbf{x}^n)\cdot f''(\mathbf{x}^n)\,\nabla f(\mathbf{x}^n)}{\|\nabla f(\mathbf{x}^n)\|^4}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\,(\mathbf{x}^n)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) \,. \\
\therefore\ 0 &= f\,(\mathbf{x}^n) + \nabla f\,(\mathbf{x}^n)\cdot\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) + \frac{1}{2}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\,(\mathbf{x}^n)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) \\
&\quad -\frac{1}{4}\frac{f\,(\mathbf{x}^n)\,\nabla f\,(\mathbf{x}^n)\cdot f''(\mathbf{x}^n)\,\nabla f(\mathbf{x}^n)}{\|\nabla f(\mathbf{x}^n)\|^4}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\,(\mathbf{x}^n)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) \,.
\end{aligned}$$

Thus,

$$\begin{aligned}
f\left(\mathbf{x}^{n+1}\right) &= f\left(\mathbf{x}^{n+1}\right) - f\,(\mathbf{x}^n) - \nabla f\,(\mathbf{x}^n)\cdot\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) - \frac{1}{2}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\,(\mathbf{x}^n)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) \\
&\quad +\frac{1}{4}\frac{f\,(\mathbf{x}^n)\,\nabla f\,(\mathbf{x}^n)\cdot f''(\mathbf{x}^n)\,\nabla f(\mathbf{x}^n)}{\|\nabla f(\mathbf{x}^n)\|^4}\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right)\cdot f''\,(\mathbf{x}^n)\left(\mathbf{x}^{n+1}-\mathbf{x}^n\right) \,.
\end{aligned}$$

So, by induction

$$\left|f\left(\mathbf{x}^{n+1}\right)\right| \leq \frac{N}{6}\left\|\mathbf{x}^{n+1}-\mathbf{x}^{n}\right\|^{3}+\frac{M^{2}}{4}\frac{\left|f\left(\mathbf{x}^{n}\right)\right|}{\left\|\nabla f\left(\mathbf{x}^{n}\right)\right\|^{2}}\left\|\mathbf{x}^{n+1}-\mathbf{x}^{n}\right\|^{2}$$

$$\leq\left\{\frac{N}{6}\left\|\mathbf{x}^{n+1}-\mathbf{x}^{n}\right\|+\frac{M^{2}}{4}\frac{\varphi\left(y^{n}\right)}{\left(\varphi'\left(y^{n}\right)\right)^{2}}\right\}\left(y^{n+1}-y^{n}\right)^{2}, \text{ by induction}$$

$$=\left\{\frac{N}{6}\left|\alpha^{n}\right|\left\|\mathbf{v}^{n}\right\|\left|f\left(\mathbf{x}^{n}\right)\right|+\frac{M^{2}}{4}\frac{\varphi\left(y^{n}\right)}{\left(\varphi'\left(y^{n}\right)\right)^{2}}\right\}\left(y^{n+1}-y^{n}\right)^{2}$$

$$\leq\left\{\frac{N}{6}\frac{-\frac{\varphi(y^{n})}{\varphi'(y^{n})}}{1-\frac{M}{2}\frac{\varphi(y^{n})}{(\varphi'(y^{n}))^{2}}}+\frac{M^{2}}{4}\frac{\varphi\left(y^{n}\right)}{\left(\varphi'\left(y^{n}\right)\right)^{2}}\right\}\left(y^{n+1}-y^{n}\right)^{2}$$

$$=\left\{\frac{N}{6}\frac{-\varphi'\left(y^{n}\right)}{1-\frac{M}{2}\frac{\varphi(y^{n})}{(\varphi'(y^{n}))^{2}}}+\frac{M^{2}}{4}\right\}\frac{\varphi\left(y^{n}\right)}{\left(\varphi'\left(y^{n}\right)\right)^{2}}\left(y^{n+1}-y^{n}\right)^{2}.$$

For any $y \in [0, r_1]$ the function $\frac{\varphi(y)}{(\varphi'(y))^2}$ is decreasing, since

$$\left(\frac{\varphi\left(y\right)}{\left(\varphi'\left(y\right)\right)^{2}}\right)' = \frac{\left(r_{2}-r_{1}\right)^{2}}{\frac{C}{2}\left(2y-r_{1}-r_{2}\right)^{3}}<0,\ y\in\left[0,r_{1}\right].$$

$$\text{Also, } \frac{\varphi\left(y^{n}\right)}{\left(\varphi'\left(y^{n}\right)\right)^{2}} \leq \frac{\varphi\left(y^{0}\right)}{\left(\varphi'\left(y^{0}\right)\right)^{2}}.$$

$$\therefore\ \frac{-\varphi'\left(y^{n}\right)}{1-\frac{M}{2}\frac{\varphi(y^{n})}{(\varphi'(y^{n}))^{2}}} \leq \frac{-\varphi'\left(y^{0}\right)}{1-\frac{M}{2}\frac{\varphi(y^{0})}{(\varphi'(y^{0}))^{2}}}=\frac{1}{L(1-\frac{M}{2}LB)}.$$

$$\therefore\ \left|f\left(\mathbf{x}^{n+1}\right)\right| \leq \frac{1}{4}\left(M^{2}+\frac{2}{3}\frac{N}{L\left(1-\frac{MBL}{2}\right)}\right)\frac{\varphi\left(y^{n}\right)}{\left(\varphi'\left(y^{n}\right)\right)^{2}}\left(y^{n+1}-y^{n}\right)^{2}$$

$$=\frac{C^{2}}{4}\frac{\varphi\left(y^{n}\right)}{\left(\varphi'\left(y^{n}\right)\right)^{2}}\left(y^{n+1}-y^{n}\right)^{2}=\varphi\left(y^{n+1}\right), \text{ by (22)}.$$

<u>Proof of (A.1b) for $n+1$:</u>

$$\left\|\mathbf{v}^{n+1}\right\| = \frac{1}{\left\|\nabla f\left(\mathbf{x}^{n+1}\right)\right\|}=\frac{1}{\left\|\nabla f\left(\mathbf{x}^{0}\right)-\left(\nabla f\left(\mathbf{x}^{0}\right)-\nabla f\left(\mathbf{x}^{n+1}\right)\right)\right\|}$$

$$=\frac{1}{\left\|\nabla f\left(\mathbf{x}^{0}\right)\right\|\left\|\frac{\nabla f(\mathbf{x}^{0})}{\|\nabla f(\mathbf{x}^{0})\|}-\frac{\nabla f(\mathbf{x}^{0})-\nabla f(\mathbf{x}^{n+1})}{\|\nabla f(\mathbf{x}^{0})\|}\right\|}\leq\frac{L}{1-LM\left\|\mathbf{x}^{n+1}-\mathbf{x}^{0}\right\|}, \text{ by lemma 2}$$

$$\leq \frac{L}{1-LCy^{n+1}}=\frac{1}{\frac{1}{L}-Cy^{n+1}}=\frac{1}{-\varphi'\left(y^{n+1}\right)}.$$

<u>Proof of (A.1c) for $n+1$:</u>

$$\left|\alpha^{n+1}\right| = \frac{1}{\left|1-\frac{1}{2}f\left(\mathbf{x}^{n+1}\right)\mathbf{v}^{n+1}\cdot f''\left(\mathbf{x}^{n+1}\right)\mathbf{v}^{n+1}\right|}$$

$$\leq \frac{1}{1-\frac{M}{2}\frac{\varphi(y^{n+1})}{(\varphi'(y^{n+1}))^{2}}}, \text{ by } (20a) \text{ and } (20b).$$

Proof of (A.1d) for $n+1$:

$$\left\|\mathbf{x}^{n+2}-\mathbf{x}^{n+1}\right\| = \left\|\alpha^{n+1} f(\mathbf{x}^{n+1})\mathbf{v}^{n+1}\right\| \le \frac{-\frac{\varphi\left(y^{n+1}\right)}{\varphi'(y^{n+1})}}{1-\frac{M}{2}\frac{\varphi(y^{n+1})}{(\varphi'(y^{n+1}))^2}}, \text{ by } (20a), (20b) \text{ and } (20c).$$

$$\le \frac{-\frac{\varphi\left(y^{n+1}\right)}{\varphi'(y^{n+1})}}{1-\frac{C}{2}\frac{\varphi(y^{n+1})}{(\varphi'(y^{n+1}))^2}} = y^{n+2}-y^{n+1}.$$

Consequently, (A.1a)–(A.1d) hold for all $n \ge 0$. So by Lemma 2, $\lim_{k\to\infty} \mathbf{x}^k = \mathbf{x}^*$ and $\mathbf{x}^* \in X_0$.

So, the sequence $\left\{\mathbf{x}^k\right\}$ is majorized by the sequence $\left\{\mathbf{y}^k\right\}$ generated by the scalar Halley method. The scalar Halley method has a cubic rate of convergence, [8], and $\left|y^{k+1}-y^k\right| \le \beta\frac{\theta^{3^k}}{1-\theta^{3^k}}$, where $\theta < 1$, [9]. So, $\left\|\mathbf{x}^{k+1}-\mathbf{x}^k\right\| \le \beta\frac{\theta^{3^k}}{1-\theta^{3^k}}$, and the sequence $\{\mathbf{x}^k\}$ has at least cubic rate of convergence. □

Appendix B: 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(1..3):
```

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

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

B.1. The directional Halley Method (8).

The function **HalleyDirNext(f,x,x0,d)** computes the next directional Halley iterate of $f(x)$ at x^0 in the direction d

```
> HalleyDirNext:=proc(f,x,x0,d)
> local val,gr,c,cc,hes;
> val:=eval(subs(x=x0,f)):
> gr:=eval(subs(x=x0,grad(f,x))):
> c:=dotprod(gr,d):
> hes:=eval(subs(x=x0,hessian(f,x))):
> cc:=dotprod(d,evalm(hes &* d)):
> evalm(x0-(val/(c-(cc*val)/(2*c))*d));
> end:
```

B.2. The gradient Halley Method (9).

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

```
> HalleyGrad:=proc(f,x,x0,N)
> local d,sol,valf; global k;
> k:=0; sol:=array(0..N):sol[0]:=x0:
> valf:=eval(subs(x=x0,f)):
```

```
> print(f);
> lprint(Iterate,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]:=HalleyDirNext(f,x,sol[k-1],d):
> valf:=eval(subs(x=sol[k],f)):
> if (sqrt(dotprod(sol[k]-sol[k-1],sol[k]-sol[k-1]))<eps) then
> break fi:
> od:
> lprint(Iterate,k-1):print(sol[k-1]):
> lprint(function):print(valf):
> end:
```

Example B.2. $f(x) = \exp(1 - x_1^2 - x_2) - 1$, $x^0 = (1.0, 1.2)$, 10 *iterations.*

```
> HalleyGrad(exp(1-x[1]-x[2])-1,x,[1.,1.2],3);
```

$e^{(1-x_1-x_2)} - 1$

`Iterate   0`

$$[1.,\ 1.2]$$

`Function`

$$-.6988057881$$

`Iterate   3`

$$[.4000000000,\ .6000000000]$$

`Function`

$$0$$

B.3. The gradient quasi-Halley Method (10).

The function **QuasiDirNext(f, x, x0, d)** computes the next directional quasi-Halley iterate of $f(x)$ at x^0 in the direction $\nabla f(x^0)$.

```
> QuasiDirNext:=proc(f,x,x0)
> local val,val1,x1,gr,c; val:=eval(subs(x=x0,f)):
> gr:=eval(subs(x=x0,grad(f,x))):  c:=dotprod(gr,gr):
> x1:=evalm(x0-(val/c)*gr):  val1:=eval(subs(x=x1,f)):
> if (val1<>val) then evalm(x0-val/(c*(1-val1/val))*gr)
> else evalm(x0-val/(c)*gr); fi:
> end:
```

The function **QuasiGrad(f,x,x0,N)** computes N iterations of the function $f(x)$ starting at $\mathbf{x}^0$.

```
> QuasiGrad:=proc(f,x,x0,N)
> local d,sol,valf; global k; k:=0;
> sol:=array(0..N):
> sol[0]:=x0:valf:=eval(subs(x=x0,f)):
> print(f); lprint(Iterate,0):print(sol[0]):  lprint(Function):print(valf):
> for k from 1 to N do
```

```
> sol[k]:=QuasiDirNext(f,x,sol[k-1]):
> valf:=eval(subs(x=sol[k],f)):
> if (sqrt(dotprod(sol[k]-sol[k-1],sol[k]-sol[k-1]))<eps) then break fi:
> od:
> lprint(Iterate,k-1):print(sol[k-1]):  lprint(Function):print(valf):
> end:
```

Example B.3. $f(x) = x_1^2 - x_2$, $x^0 = (2.1, 1.2)$, 4 *iterations.*

```
> QuasiGrad(x[1]^2-x[2],x,[2.1,1.2],4);
```

$$x_1{}^2 - x_2$$

```
Iterate   0
```

$$[2.1,\ 1.2]$$

```
Function
```

$$3.21$$

```
Iterate   3
```

$$[1.192944003,\ 1.423115393]$$

```
Function
```

$$.1\,10^{-8}$$

B.4. Comparison Between Directional methods: Newton, Halley and quasi-Halley.

Derby(deg, iter, N) generates N random polynomials of degree *deg* (at most), and solves each using Newton, Halley and QuasiHalley, allowing at most *iter* iterations.

```
> Derby:=proc(deg,iter,N)
> local degr,x,j,c,f,x0,init,ncount,hcount,qcount;
> global k;
> ncount:=0;hcount:=0;qcount:=0;
> x:=array(1..2);
> for degr from 2 to deg do
> k:=0; ncount:=0; hcount:=0; qcount:=0;
> for j from 1 to N do
> c:=evalf(rand()/10^12):
> f:=c+randpoly([x[1],x[2]],degree=degr);
> x0:=[1.0,1.0];
> lprint(Newton):
> NewtonGrad(f,x,x0,iter):ncount:=ncount+k;print(k);
> lprint(Halley):
> HalleyGrad(f,x,x0,iter):hcount:=hcount+k;print(k);
> lprint(Quasi-Halley):
> NewQuasiGrad(f,x,x0,iter):qcount:=qcount+k;print(k);
> c:='c';
> od:
> lprint(Average-no-iterations);
> lprint(Newton);print(evalf(ncount/N));
> lprint(Halley);print(evalf(hcount/N));
```

```
> lprint(Quasi-Halley);print(evalf(qcount/N));
> od:  end:
```

```
> Derby(2,30,100);
```

```
Average-no-iterations

Newton
                                12.450000000
Halley
                                11.860000000
Quasi-Halley
                                7.870000000
```

obtaining the first comparison in Table 1.

B.5. Systems of equations.

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:
```

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

```
> SystemHalleyGrad:=proc(f,x,x0,N)
> local n,F;
> n:=vectdim(x0);F:=SOS(f);print(f);
> HalleyGrad(F,x,x0,N);
> end:
```

The function **SystemQuasiGrad(f,x,x0,N)** computes N iterations of the directional quasi-Halley method for the sum of squares $\sum_{i=1}^{n} f_i^2(x)$, starting at x^0.

```
> SystemQuasiGrad:=proc(f,x,x0,N)
> local n,F;
> n:=vectdim(x0);F:=SOS(f);print(f);
> QuasiGrad(F,x,x0,N);
> end:
```

Example B.4. *Fröberg, p. 186, Example 1*

```
> x:=array(1..3):
> SystemHalleyGrad([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],10):
```

$$[{x_1}^2 - x_1 + {x_2}^3 + {x_3}^5,\ {x_1}^3 + {x_2}^5 - x_2 + {x_3}^7,\ {x_1}^5 + {x_2}^7 + {x_3}^{11} - x_3]$$

$$({x_1}^2 - x_1 + {x_2}^3 + {x_3}^5)^2 + ({x_1}^3 + {x_2}^5 - x_2 + {x_3}^7)^2 + ({x_1}^5 + {x_2}^7 + {x_3}^{11} - x_3)^2$$

```
Iterate   0
```

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

```
Function
```

$$.1357076447$$

```
Iterate   10
```

$$[.002243051296,\ .0002858171153,\ \text{-}.0002540074383]$$

```
Function
```

$$.5154938245\,10^{-5}$$

```
> x:=array(1..3):
> SystemQuasiGrad([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],10);
```

$$[{x_1}^2 - x_1 + {x_2}^3 + {x_3}^5,\ {x_1}^3 + {x_2}^5 - x_2 + {x_3}^7,\ {x_1}^5 + {x_2}^7 + {x_3}^{11} - x_3]$$

$$({x_1}^2 - x_1 + {x_2}^3 + {x_3}^5)^2 + ({x_1}^3 + {x_2}^5 - x_2 + {x_3}^7)^2 + ({x_1}^5 + {x_2}^7 + {x_3}^{11} - x_3)^2$$

```
Iterate   0
```

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

```
Function
```

$$.1357076447$$

```
Iterate   10
```

$$[.0001876563761,\ .4627014469\,10^{-5},\ -.3061094461\,10^{-5}]$$

```
Function
```

$$.3523247963\,10^{-7}$$