**Electronic Transactions on Numerical Analysis.**
**Volume 26, pp. 82-102, 2007.**

**ISSN 1068-9613.**

**ETNA**
**Kent State University**
**etna@mcs.kent.edu**

# COMPUTING QUATERNIONIC ROOTS BY NEWTON'S METHOD*

DRAHOSLAVA JANOVSKÁ† AND GERHARD OPFER‡

**Abstract.** Newton's method for finding zeros is formally adapted to finding roots of Hamilton's quaternions. Since a derivative in the sense of complex analysis does not exist for quaternion valued functions we compare the resulting formulas with the more classical formulas obtained by using the Jacobian matrix and the Gâteaux derivative. The latter case includes also the so-called damped Newton form. We investigate the convergence behavior and show that under one simple condition all cases introduced, produce the same iteration sequence and have thus the same convergence behavior, namely that of locally quadratic convergence. By introducing an analogue of Taylor's formula for $x^n,\ n \in \mathbb{Z}$, we can show the local, quadratic convergence independently of the general theory. It will also be shown that the application of damping proves to be very useful. By applying Newton iterations backwards we detect all points for which the iteration (after a finite number of steps) must terminate. These points form a nice pattern. There are explicit formulas for roots of quaternions and also numerical examples.

**Key words.** Roots of quaternions, Newton's method applied to finding roots of quaternions.

**AMS subject classifications.** 11R52, 12E15, 30G35, 65D15

**1. Introduction.** The newer literature on quaternions is in many cases concerned with algebraic problems. Let us mention in this context the survey paper by Zhang [15]. Here, for the first time we try to apply an analytic tool, namely Newton's method, to finding roots of quaternions, numerically. Let $g : \mathbb{R}^n \to \mathbb{R}^n$ be a given mapping with continuous partial derivatives. Then, the *classical Newton form* for finding solutions of $g(x) = 0$ is given by

$$g(x) + g'(x)\eta = 0, \quad x_{\text{new}} := x + \eta, \tag{1.1}$$

where $g'$ stands for the matrix of partial derivatives of $g$, which is also called *Jacobian matrix*. The equation (1.1) has to be regarded as a linear system for $\eta$ with known $x$. The further steps consist of iteratively solving this system with $x_{\text{new}}$.

In this paper we want to treat a special problem $g(x) = 0$ with $g : \mathbb{H} \to \mathbb{H}$, where $\mathbb{H}$ denotes the (skew) field of quaternions. We use the letter $\mathbb{H}$ in honor of William Rowan $\mathbb{H}$amilton (1805 – 1865), the inventor of quaternions. In this setting we will try also other forms of derivatives of $g$ than the matrix of partial derivatives.

For illustration in this introduction, we use the simple equation $g(x) := x^2 - a$ with $a, x \in \mathbb{H}$. If we follow the real or complex case for defining derivatives, we have two possibilities because of the non commutativity of the multiplication in $\mathbb{H}$, namely

$$g'(x) := \lim_{h\to 0}\{(g(x+h) - g(x))h^{-1}\} = \lim_{h\to 0}((x+h)^2 - x^2)h^{-1} = x + \lim_{h\to 0} hxh^{-1},$$
$$g'(x) := \lim_{h\to 0}\{h^{-1}(g(x+h) - g(x))\} = \lim_{h\to 0} h^{-1}((x+h)^2 - x^2) = x + \lim_{h\to 0} h^{-1}xh.$$

If we put $y_h := hxh^{-1}$ for any $h \neq 0$ then from later considerations we know that $|y_h| = |x|$ and $(y_h)_1 = x_1$. Thus, $y_h$ fills the surface of a three dimensional ball and there is no unique limit. In other words, the above requirement for differentiability is too strong. One can even show that only the quaternion valued functions $g(z) := az + b,\ g(z) := za + b$, $a, b \in \mathbb{H}$, respectively, are differentiable with respect to the two given definitions, Sudbery [13, Theorem 1].

*Received April 18, 2006. Accepted for publication October 18, 2006. Recommended by L. Reichel.
† Institute of Chemical Technology, Prague, Department of Mathematics, Technická 5, 166 28 Prague 6, Czech Republic (janovskd@vscht.cz).
‡University of Hamburg, Faculty for Mathematics, Informatics, and Natural Sciences [MIN], Bundestraße 55, 20146 Hamburg, Germany (opfer@math.uni-hamburg.de).

**ETNA**
Kent State University
etna@mcs.kent.edu



In approximation theory and optimization a much weaker form of derivative is employed very successfully. It is the *one sided directional derivative of* $g : \mathbb{H} \to \mathbb{H}$ *in direction* $h$ or *one sided Gâteaux*[1] derivative of $g$ *in direction* $h$ (for short only *Gâteaux derivative*) which for $x, h \in \mathbb{H}$ is defined as follows:

$$g'(x,h) := \lim_{\substack{\alpha\to 0\\ \alpha>0}} \frac{g(x+\alpha h)-g(x)}{\alpha} = \lim_{\substack{\alpha\to 0\\ \alpha>0}} \frac{(x+\alpha h)^2 - x^2}{\alpha} = xh + hx. \tag{1.2}$$

Let $h \in \mathbb{R}\backslash\{0\}$, then $g'(x,h) = 2hx$ and from (1.1) replacing $g'(x)$ with $g'(x,h)$ we obtain the *damped Newton form*

$$x_{\text{new}} := N(x) := x + \frac{1}{2h}\left(x^{-1}a - x\right)$$

if $h > 1$. For $h = 1$ we obtain the *common Newton form* for square roots.

If we work with partial derivatives, the equation $g(x) := x^2 - a$ implies

$$g'(x) := 2\begin{pmatrix} x_1 & -x_2 & -x_3 & -x_4 \\ x_2 & x_1 & 0 & 0 \\ x_3 & 0 & x_1 & 0 \\ x_4 & 0 & 0 & x_1 \end{pmatrix}. \tag{1.3}$$

Matrices of this form are known as *arrow matrices*. They belong to a class of *sparse matrices* for which many interesting quantities can be computed explicitly, Reid [11], Walter, Lederbaum, and Schirmer [14], and Arbenz and Golub [1] for eigenvalue computations. The special cases $a, x \in \mathbb{R}$ and $a, x \in \mathbb{C}$ reduce immediately to the common Newton form

$$x_{\text{new}} := N(x) := \frac{1}{2}\left(x + \frac{a}{x}\right).$$

The treatment of analytic problems in $\mathbb{H}$ goes back to Fueter [5]. A more recent overview including new results is given by Sudbery [13]. However, Gâteaux derivatives do not occur in this article.

We start with some information on explicit formulas for roots of quaternions. Then we adjust the common Newton formula for the $n$-th root of a real (positive) or complex number to the case of quaternions. Because of the non commutativity of the multiplication we obtain two slightly different formulas. We will see that under a simple condition both formulas produce the same sequence. We see by examples that in this case the convergence is fast and we also see from various examples that in case the formulas produce different sequences, the convergence is slow or even not existing. Later we apply the Gâteaux derivative and the Jacobian matrix of the partial derivatives to formula (1.1) and show that under the same condition the same formulas can be derived which proves that the convergence is locally quadratic. The Gâteaux derivative gives also rise to the damped Newton form which turns out to be very successful and superior to the ordinary Newton technique.

## 2. Roots of quaternions.

We start by describing a method for finding the solutions of

$$g(x) := x^n - a = 0, \quad a \in \mathbb{H}\backslash\mathbb{R}, \quad n \in \mathbb{N}, \quad n \geq 2, \tag{2.1}$$

explicitly. The solutions of $g(x) = 0$ will be called *roots of* $a$. We need some preparations. If $a = (a_1, a_2, a_3, a_4) \in \mathbb{H}$ we will also use the notation

$$a = a_1 + a_2\,\mathbf{i} + a_3\,\mathbf{j} + a_4\,\mathbf{k},$$

[1] René Gâteaux, French mathematician (Vitry 1889 – [Verdun?] 1914)

**ETNA**
Kent State University
etna@mcs.kent.edu



where $\mathbf{i}, \mathbf{j}, \mathbf{k}$ stand for the units $(0,1,0,0)$, $(0,0,1,0)$, $(0,0,0,1)$, respectively.

DEFINITION 2.1. *Two quaternions $a, b$ are called* equivalent, *denoted by $a \sim b$, if there is $h \in \mathbb{H}\backslash\{0\}$ such that $a = h^{-1}bh$ (or $ha = bh$). The set of all quaternions equivalent to $a$ is denoted by $[a]$. Let $a := (a_1, a_2, a_3, a_4) \in \mathbb{H}\backslash\mathbb{R}$. We call $a_v := (0, a_2, a_3, a_4)$ the* vector part *of $a$. By assumption $a_v \neq 0$. The complex number*

$$\tilde{a} := (a_1, \sqrt{a_2^2 + a_3^2 + a_4^2}, 0, 0) =: a_1 + |a_v|\,\mathbf{i} \tag{2.2}$$

*has the property that it is equivalent to $a$ (cf. (2.3)) and it is the only equivalent complex number with positive imaginary part. We shall call this number $\tilde{a}$* the complex equivalent *of $a$.*

Because of

$$a = h^{-1}bh = \left(\frac{h}{|h|}\right)^{-1} b \frac{h}{|h|}$$

there is no loss of generality if we assume that $|h|^2 = 1$. Since $a \in \mathbb{R}$ commutes with all elements in $\mathbb{H}$ we have $[a] = \{a\}$. In other words, for real numbers $a$ the equivalence class $[a]$ consists only of the single element $a$. Let $c \in \mathbb{C}$, then $c$ and the complex conjugate $\overline{c}$ belong to the same class $[c]$ because of $\overline{c} = (\mathbf{j})^{-1} c\, \mathbf{j}$.

LEMMA 2.2. *The above notion of equivalence defines an equivalence relation. And we have $a \sim b$ if and only if*

$$\Re a = \Re b, \quad |a| = |b|. \tag{2.3}$$

*Proof.* Let $ha = bh$ for some $h \neq 0$. Then, the general rule $|xy| = |x||y|$ yields $|a| = |b|$. Let us put $a = h^{-1}bh$ and apply another general rule $\Re(xy) = \Re(yx)$. Then $\Re a = \Re(h^{-1}bh) = \Re((h^{-1}b)h) = \Re(hh^{-1}b) = \Re b$. It remains to show that (2.3) implies the existence of an $h \neq 0$ such that $ha = bh$. Let $a \in \mathbb{R}$. Then (2.3) implies $a = b$ and hence, $h = 1$. Otherwise, (2.3) is equivalent to a real, linear, homogeneous $4 \times 4$ system. It can be shown, that the rank of the corresponding matrix is two. □

There are situations where there are infinitely many roots.

THEOREM 2.3. *Let $g$ be defined as in (2.1) but with real $a$. If there exists a complex root $x$ of $a$ which is not real, then there will be infinitely many quaternionic roots of $a$.*

*Proof.* Let $x := x_1 + x_2\mathbf{i}$ be a root of $a$ with $x_2 \neq 0$. We have $g(x) := x^n - a = 0$. Let $h \in \mathbb{H}\backslash\{0\}$. We multiply the last equation from the left by $h^{-1}$ and from the right by $h$ and obtain

$$h^{-1}g(x)h = h^{-1}x^n h - h^{-1}ah = (h^{-1}xh)^n - a = 0 \tag{2.4}$$

since real numbers commute with quaternions. Therefore, $g(h^{-1}xh) = 0$ or, in other words, the whole equivalence class $[x]$ of $x$ consists of roots. □

COROLLARY 2.4. *Let $a \neq 0$ be real. For $n \geq 3$ there are always infinitely many roots of $a$. For $n = 2$ there are infinitely many roots if $a < 0$.*

The finding of roots of quaternions is based on the following lemma.

LEMMA 2.5. *Let $a \in \mathbb{H}\backslash\mathbb{R}$ and let $\tilde{a}$ be the corresponding complex equivalent of $a$ where $\tilde{a} = h^{-1}ah$ for some $h \neq 0$ such that $\Im\tilde{a} > 0$. Then, $x$ will be a root of $a$ if and only if $\hat{x} := h^{-1}xh$ is a root of $\tilde{a}$.*

*Proof.* (i) Let $x$ be a root of $a$. By applying (2.4) we obtain $\hat{x}^n - \tilde{a} = 0$. (ii) Let $\hat{x}$ be a root of $\tilde{a}$. I.e. we have $\hat{x}^n - \tilde{a} = 0$. Multiplying from the left by $h$ and from the right by $h^{-1}$ gives the desired result. □

This lemma yields the following steps for solving (2.1) for $a := (a_1, a_2, a_3, a_4) \notin \mathbb{R}$.

**ETNA**
Kent State University
etna@mcs.kent.edu



(i) Compute $\tilde{a} := (a_1, \sqrt{a_2^2 + a_3^2 + a_4^2}, 0, 0) = a_1 + |a_v|\mathbf{i} \in \mathbb{C}$.
(ii) Let $\hat{x}_k \in \mathbb{C}$ be the roots of $\tilde{a} \in \mathbb{C}$: $\hat{x}_k = |a|^{1/n} \exp(\mathbf{i}\,\frac{\alpha+2k\pi}{n})$, $k = 0, 1, \ldots, n-1$, $\cos\alpha = \frac{a_1}{|a|}$, $\alpha \in [0, \pi[$.
(iii) Find $h \in \mathbb{H}$ such that $\tilde{a} := h^{-1}ah \in \mathbb{C}$.
(iv) Then, the sought after roots are $x_k = h\hat{x}_k h^{-1}$.

The equivalence $a \sim \tilde{a}$, expressed in (iii) may be regarded as a *linear mapping*

$$\mathbf{H}\,a = \tilde{a}, \text{ where } \mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & \tilde{\mathbf{H}} \end{pmatrix} \in \mathbb{R}^{4\times 4} \tag{2.5}$$

and $\tilde{\mathbf{H}}$ is a $(3\times 3)$ *Householder matrix*

$$\tilde{\mathbf{H}} := \mathbf{I} - \frac{2}{v^{\mathrm{T}}v} vv^{\mathrm{T}}, \quad v := \begin{pmatrix} a_2 - |a_v| \\ a_3 \\ a_4 \end{pmatrix} \text{ with}$$

$$\tilde{\mathbf{H}} \begin{pmatrix} a_2 \\ a_3 \\ a_4 \end{pmatrix} = \begin{pmatrix} |a_v| \\ 0 \\ 0 \end{pmatrix}.$$

Now, in (iv) we need the inverse mapping $\mathbf{H}^{-1} = \mathbf{H}$, thus, the roots are

$$x_k := \mathbf{H} \begin{pmatrix} \Re\hat{x}_k \\ \Im\hat{x}_k \\ 0 \\ 0 \end{pmatrix} = |a|^{1/n} \begin{pmatrix} \cos\frac{\alpha+2k\pi}{n} \\ \frac{a_2}{|a_v|}\sin\frac{\alpha+2k\pi}{n} \\ \frac{a_3}{|a_v|}\sin\frac{\alpha+2k\pi}{n} \\ \frac{a_4}{|a_v|}\sin\frac{\alpha+2k\pi}{n} \end{pmatrix}, \quad k = 0, 1, \ldots, n-1. \tag{2.6}$$

The right hand side of (2.6) was already given by Kuba [10]. However, the above derivation using Householder transformations is new. It allows a very easy proof of the following lemma.

LEMMA 2.6. *Let $n \geq 2$ and $a \in \mathbb{H}\backslash\mathbb{R}$ be given and let $x_k,\ k = 0, 1, 2, \ldots, n-1$, be the roots of a according to (2.6). Then (i) $|x_k| = |a|^{1/n}$ for all $k = 0, 1, 2, \ldots, n-1$, and (ii) the real $(4\times n)$ matrix $\mathbf{X} := (\,x_0 \quad x_1 \quad \cdots \quad x_{n-1}\,)$ of all roots has rank two.*

*Proof*. (i) The matrix $\mathbf{H}$ is orthogonal and thus, does not change norms: $|x_k| = |\Re\hat{x}_k + \Im\hat{x}_k\mathbf{i}| = |\,\hat{x}_k\,| = |a|^{1/n}$. (ii) The matrix $\mathbf{H}$ is non singular and thus, does not change the dimension of the image space. □

COROLLARY 2.7. *Under the same assumptions as in the previous lemma all roots $x_k$ of $a$ are located on a (two dimensional) circle on the surface of the four dimensional ball with radius $|a|^{1/n}$.*

Let $x \in \mathbb{H}$ be a root of $a \in \mathbb{H}\backslash\mathbb{R}$ and let $\tilde{x}, \tilde{a}$ be the complex equivalents of $x, a$, respectively. The Lemma 2.5 does not state that $\tilde{x}$ is a root of $\tilde{a}$. Nevertheless, it is half way true. For any real number $y$ we define $\lfloor y \rfloor$ as the largest integer not exceeding $y$. For a complex number $z$, the quantity $\overline{z}$ is defined as the complex conjugate of $z$.

LEMMA 2.8. *Let $a \in \mathbb{H}\backslash\mathbb{R}$ be given and let $x_k$ be the roots of $a$ in the ordering $k = 0, 1, \ldots, n-1$ given in (2.6). Let $\tilde{a}$ be the complex equivalent of $a$ and $\tilde{x}_k$ be the complex equivalents of $x_k,\ k = 0, 1, \ldots, n-1$. Then, $\tilde{x}_k$ is a root of $\tilde{a}$ for $k = 0, 1, \ldots, \lfloor (n-1)/2 \rfloor$ and $\overline{\tilde{x}_k}$ is a root of $\tilde{a}$ for the remaining $k$.*

*Proof*. We only show the essential part: If $x$ is a root of $a$, then either $\tilde{x}$ or $\overline{\tilde{x}}$ is a root of $\tilde{a}$. Let $\tilde{a} = h^{-1}ah$ and $\tilde{x} = \tilde{h}^{-1}x\tilde{h}$. By applying (2.4) we have $(h^{-1}xh)^n - \tilde{a} = 0$. Since $h^{-1}xh$ and $\tilde{h}^{-1}x\tilde{h}$ are both complex, they differ by Lemma 2.2 at most in the sign of the imaginary part and the statement is proved. □

Let us illustrate this lemma by a little example.

EXAMPLE 2.9. Let $n = 2$. The two roots of $a := (-4, 40, 30, -20)$ are $x_0 :=$

ETNA
Kent State University
etna@mcs.kent.edu



$(5,-4,3,-2)$, $x_1 = -x_0$ and $\tilde{a} = -4+10\sqrt{29}\,\mathbf{i}, \tilde{x}_0 = 5+\sqrt{29}\,\mathbf{i}$, $\tilde{x}_1 = -5+\sqrt{29}\,\mathbf{i}$. We have $\tilde{x}_0^2 = \tilde{a}$ and $(\overline{\tilde{x}}_1)^2 = \tilde{a}$.

If we use numerical methods for finding roots of $a \in \mathbb{H}$ we will find only one of the quaternionic roots, say $r$. Let $\tilde{a}, \tilde{r}$ be the complex equivalents of $a, r$, respectively. Then, according to Lemma 2.8, $\tilde{r}$ or $\overline{\tilde{r}}$ is a complex root of $\tilde{a}$. We define

$$\hat{r} := \begin{cases} \tilde{r} & \text{if } \tilde{r}^n = \tilde{a}, \\ \overline{\tilde{r}} & \text{otherwise.} \end{cases}$$

All further roots $\hat{r}_k$ of $\tilde{a}$ follow the equation

$$\hat{r}_k = \hat{r}\exp\frac{2k\pi}{n}\mathbf{i}, \quad k = 1,2,\ldots,n-1. \tag{2.7}$$

It should be observed that the factor $\exp\frac{2k\pi}{n}\mathbf{i}$ apart from $n$ does not contain any information about the root $\hat{r}$. In order to find all quaternionic roots we only need to apply (2.6) again. We put $\hat{r} := u + v\mathbf{i}$ and $\sigma_k := \frac{2k\pi}{n}$ and obtain the other roots by

$$r_k := \mathbf{H}\begin{pmatrix} \Re\hat{r}_k \\ \Im\hat{r}_k \\ 0 \\ 0 \end{pmatrix} = \mathbf{H}\begin{pmatrix} u\cos\sigma_k - v\sin\sigma_k \\ v\cos\sigma_k + u\sin\sigma_k \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} s_k \\ \rho_2\, t_k \\ \rho_3\, t_k \\ \rho_4\, t_k \end{pmatrix}, \tag{2.8}$$

where $r =: (\rho_1, \rho_2, \rho_3, \rho_4)$, $\quad |r_v| := \sqrt{(\rho_2)^2 + (\rho_3)^2 + (\rho_4)^2}$, and where

$$s_k := u\cos\sigma_k - v\sin\sigma_k, \quad t_k := \frac{\operatorname{sign} v}{|r_v|}(v\cos\sigma_k + u\sin\sigma_k), \quad k = 1,2,\ldots,n-1.$$

EXAMPLE 2.10. Let $n = 3$ and $a = (-86, 52, -78, 104)$. Then, $r = (1,-2,3,-4)$ is one of the quaternionic roots and the corresponding complex equivalents are $\tilde{a} = -86 + 26\sqrt{29}\mathbf{i}$, $\tilde{r} = 1 + \sqrt{29}\mathbf{i}$. We have $\hat{r} = 1 - \sqrt{29}\mathbf{i}$, $|r_v| = \sqrt{29}$, $u = 1$, $v = -\sqrt{29}, \sigma_1 = 2\pi/3, \sigma_2 = 4\pi/3, s_1 = -0.5(1+\sqrt{87}) = -5.1637, s_2 = 0.5(\sqrt{87}-1) = 4.1637$, $t_1 = -0.5(1 + \frac{\sqrt{87}}{29}) = -0.6608$, $t_2 = 0.5(\frac{\sqrt{87}}{29} - 1) = -0.3392$. Then the two other quaternionic roots are $r_1 := (-5.1637, 0.6784, -1.0175, 1.3567)$, $r_2 := (4.1637, 1.3216, -1.9825, 2.6433)$.

**3. Newton iterations for roots of quaternions.** Newton iterations for finding the $n$-th root of a positive number $a$ is commonly defined by the repeated application of

$$x_{\text{new}} := N(x) := \frac{1}{n}\left((n-1)x + \frac{a}{x^{n-1}}\right). \tag{3.1}$$

What happens if $a$ is a quaternion? There are the two following analogues of Newton's formula (3.1):

$$x_{\text{new}} := N_1(x) := \frac{1}{n}\left((n-1)x + x^{1-n}a\right), \tag{3.2}$$

$$y_{\text{new}} := N_2(y) := \frac{1}{n}\left((n-1)y + ay^{1-n}\right). \tag{3.3}$$

Both formulas have to be started with some value $x_0 \neq 0, y_0 \neq 0$, respectively. The quantities $x_0, y_0$ will be called *initial guesses* for $N_1, N_2$, respectively. In the first place we do not know what formula to use. But there is the following important information.

LEMMA 3.1. *Let the initial guess $x_0 \in \mathbb{H}\backslash\{0\}$ be the same for both formulas (3.2) and (3.3). (i) The formulas $N_1$ and $N_2$ generate the same sequences $x_0, x_1, x_2 \ldots$ if $x_0$ and $\overline{a}$*

**ETNA**
Kent State University
etna@mcs.kent.edu



*commute and in this case $x_j$ and $\overline{a}$ commute for all $j \geq 0$. (ii) Let $n = 2$. Then $x_j = y_j$ for all $j \geq 0$ implies that $x_j$ and $\overline{a}$ commute for all $j \geq 0$.*

*Proof*. Let $N_1$ produce the sequence $x_0, x_1, x_2 \ldots$ and $N_2$ the sequence $x_0, y_1, y_2 \ldots$

(i) Assume that $x_0$ and $\overline{a}$ commute. Using formulas (3.2) and (3.3), we obtain

$$x_{j+1} - y_{j+1} = \frac{1}{n}\left(x_j^{1-n}a - ay_j^{1-n} + (n-1)(x_j - y_j)\right), \tag{3.4}$$

$$x_{j+1}\overline{a} - \overline{a}y_{j+1} = \frac{1}{n}\left((n-1)(x_j\overline{a} - \overline{a}y_j) + |a|^2(x_j^{1-n} - y_j^{1-n})\right). \tag{3.5}$$

We first show the following implication:

$$\text{(a)}\quad x\overline{a} - \overline{a}x = 0 \quad \Rightarrow \quad \text{(b)}\quad x^{1-n}a - ax^{1-n} = 0 \quad \text{for any } x \in \mathbb{H}\backslash\{0\}. \tag{3.6}$$

For $a = 0$ this implication is true. Let $a \neq 0$. Then (a) implies $x^k\overline{a} = \overline{a}x^k$ for all $k \in \mathbb{N}$ and hence, $\overline{a}^{-1}x^{-k} = x^{-k}\overline{a}^{-1}$. Since $\overline{a}^{-1} = \frac{a}{|a|^2}$ (b) follows. We shall prove by induction that

$$x_j - y_j = 0, \quad x_j\overline{a} - \overline{a}y_j = 0 \text{ for all } j \geq 0. \tag{3.7}$$

By assumption, (3.7) is valid for $j = 0$. Assume that it is valid for any positive $j$. Then by (3.4) and by (3.6), we have $x_{j+1} - y_{j+1} = 0$. And (3.5) implies $x_{j+1}\overline{a} - \overline{a}y_{j+1} = 0$. Thus, (3.7) is valid for all $j \in \mathbb{N}$.

(ii) Let $x_j = y_j$ for all $j \geq 0$. Then, (3.4), (3.5) reduce to

$$x_j^{1-n}a - ax_j^{1-n} = 0, \tag{3.8}$$

$$x_{j+1}\overline{a} - \overline{a}x_{j+1} = \frac{n-1}{n}\left(x_j\overline{a} - \overline{a}x_j\right). \tag{3.9}$$

For $n = 2$ equation (3.8) reads $x_j^{-1}a = ax_j^{-1}$ which implies $a^{-1}x_j = x_ja^{-1}$. Since $a^{-1} = \frac{\overline{a}}{|a|^2}$ it follows that $\overline{a}x_j = x_j\overline{a}$ and hence by (3.9), we have $\overline{a}x_{j+1} = x_{j+1}\overline{a}$. ☐

It should be noted that part (i) is already mentioned by Smith [12, Theorem 3.1], though in a matrix setting. In the above lemma it was assumed that $x_0$ and $\overline{a}$ commute. However, it is an easy exercise to see that this is equivalent to the commutation of $x_0$ and $a$. Only in our context it was a little more convenient to assume that $x_0$ and $\overline{a}$ commute.

Let $n \in \mathbb{N}$ be arbitrary. Then $x_j = y_j$ for all $j \geq 0$ implies (3.8). However, for $n \geq 3$ the implication (3.6) is not an equivalence. Take $n = 3$ and $x := \mathbf{i}$, then (b) of (3.6) is valid, but not necessarily (a) of (3.6).

In the next example we show, that for $n \geq 3$ the necessary condition (3.8) for $x_1 = y_1$ does not imply $x_2 = y_2$.

EXAMPLE 3.2. Let $n = 3$ and $x_0 = \mathbf{i}$. Then (3.8) is valid for $j = 0$ and all $a \in \mathbb{H}$ and as a consequence $x_1 = y_1 = \frac{1}{3}(2\mathbf{i} - a)$. However, $x_0\overline{a} - \overline{a}x_0 \neq 0$ and $x_2 \neq y_2$ for some $a$.

In Lemma 3.1 we have shown that the commutation of $a$ and $x_0$ implies the commutation of $a$ and $x_j$ for all $j \geq 0$. If $x_j$, $j \geq 0$, are the members of any sequence of approximation for an $n$-th root of $a \in \mathbb{H}$, then the property that $a$ and $x_j$ commute is intrinsic to the problem.

LEMMA 3.3. *For a given $a \in \mathbb{H}$ let $x$ be a solution of $g(x) := x^n - a = 0$, $n \in \mathbb{N}$. Then $a$ and $x$ commute.*

*Proof*. Multiply $g(x) := x^n - a = 0$ from either side by $x$ and subtract the resulting equations. Then $ax = xa$. ☐

Lemma 3.1 does not exclude the case that $x_j = 0$ for some $j > 0$. This means that both sequences stop at the same stage. However, we will show that this cannot happen if $x_{j-1}$ is already close to or far away from one of the roots of $a$. We introduce the *residual* $r_j$ of $x_j$ by

$$r_j := a - x_j^n.$$

**ETNA**
**Kent State University**
**etna@mcs.kent.edu**



It is a computable quantity.

LEMMA 3.4. *Let us consider the two values $x_{j-1}, x_j,\ j \geq 1$, generated by $N_1$ defined in (3.2) under the only assumption that $x_{j-1} \neq 0$. Let the residual $r_{j-1}$ have the property that*

$$|r_{j-1}| \leq |a| \text{ or } |r_{j-1}| > 2|a|. \tag{3.10}$$

*Then $x_j \neq 0$ and consequently, $x_{j+1}$ is well defined.*

*Proof*. It is clear from (3.2) that $x_j := N_1(x_{j-1}) = 0$ can happen if and only if $(n-1)x_{j-1}^n + a = 0$ or $x_{j-1}^n = -\frac{1}{n-1}\, a$. Then, in this case $r_{j-1} := a - x_{j-1}^n = a + \frac{1}{n-1}\, a = \frac{n}{n-1}\, a$, which contradicts our assumption. □

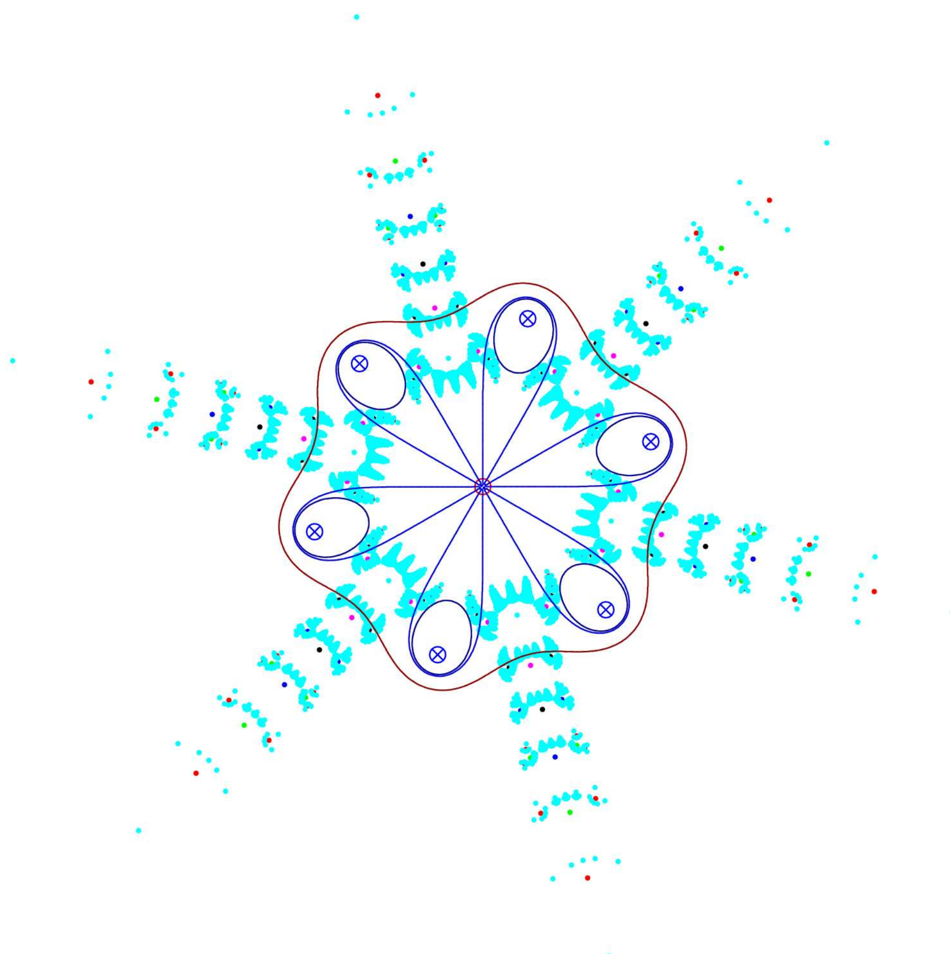

FIG. 3.1. *Exceptional points $E_n(a)$ for $n = 6$ and roots of $a = \mathbf{i}$ marked $\otimes$.*

Let $N_1$ be given by (3.2). It is easy and also interesting to find all *exceptional points*

$$E_n(a) := \{x : N_1(x) = 0, x \neq 0\} \cup \{0\}$$

for which the Newton iteration will terminate. For this purpose we write the Newton iteration backwards, i. e. we switch $x_{j+1}, x_j$ and obtain the equation

$$p(x_{j+1}) := (n-1)x_{j+1}^n - n x_{j+1}^{n-1} x_j + a = 0, \quad j = 0, 1, \ldots, \quad x_0 = 0. \tag{3.11}$$

In a first step, starting with $x_0 = 0$ we obtain $n$ solutions $x_1$ of $p(x_1) = 0$, repeat with all $n$ solutions $x_1$, obtain $n^2$ solutions $x_2$ etc. In this way, we generate $e_d := 1 + n + n^2 + \cdots + n^d = (n^{d+1} - 1)/(n-1)$ points of $E_n(a)$ if we stop after $d$ cycles. Since $x_0 = 0$ we can apply the techniques from Section 2 reducing equation (3.11) for all $j \geq 0$ to an equation with complex coefficients with the consequence that all solutions are complex as well and $E_n(a) \subset \mathbb{C}$. For $n = 2$ the set $E_n(a)$ is located on a straight line passing through the origin and having slope $\alpha = \arctan(\Im x_1 / \Re x_1)$ where $x_1 := (-a)^{1/2}$. For $n > 2$ the set $E_n(a)$ is rotational invariant under rotations of $2\pi/n$ and shows typical self-similarity. The sets $E_n(a)$ and $E_n(b)$ differ only by scaling and rotation. Or in other words, the qualitative look of $E_n(a)$ is independent of $a$. Since the exceptional points are apart from rotation the same in each of the $n$ sectors there are $(e_d - 1)/n = e_{d-1} = (n^d - 1)/(n-1)$ points in each sector. An example with $d = 7$ cycles, $n = 6$, and $a := \mathbf{i}$ is shown in Figure 3.1. It contains 335 923 points. We have also included the three *level curves*

$$\mathrm{l}_c := \{z \in \mathbb{C} : |z^n - a| = c|a|\} \text{ for } c = 0.9, 1, 2.$$

## 4. Inclusion properties.

Newton iterations can be written in the form

$$N_1(x) := \frac{n-1}{n} x + \frac{1}{n} x^{1-n} a. \tag{4.1}$$

Thus, $N_1(x)$ is a convex combination of $x$ and $x^{1-n}a$. Let $a := (a_1, a_2, a_3, a_4)$, $b := (b_1, b_2, b_3, b_4)$ be two arbitrary quaternions. With the help of the (closed, non empty) intervals

$$I_j := [\min(a_j, b_j), \max(a_j, b_j)], \; j = 1, 2, 3, 4,$$

we define the *segment*

$$\Big(a, b\Big) := (I_1, I_2, I_3, I_4).$$

LEMMA 4.1. *Let $x_0, x_1, \ldots$ be the sequence generated by $N_1$ for a given $a \in \mathbb{H}$. Then, for all $j \geq 0$ we have (componentwise)*

$$x_{j+1} \in \Big(x_j, x_j^{1-n} a\Big). \tag{4.2}$$

*Proof.* Follows immediately from (4.1). □

TABLE 4.1
*Inclusion property for some selected values $x_j, x_j^{-2}a$.*

| $\sqrt[3]{a}$ | = | 1 | −2 | 3 | −4 |
|---|---|---|---|---|---|
| $x_3$ | = | −2.2416 | −1.9163 | 2.8744 | −3.8326 |
| $x_4$ | = | 0.2017 | −1.4054 | 2.1081 | −2.8108 |
| $x_3^{-2}a$ | = | 5.0882 | −0.3837 | 0.5755 | −0.7673 |
| $x_4$ | = | 0.2017 | −1.4054 | 2.1081 | −2.8108 |
| $x_5$ | = | 1.7739 | −2.2159 | 3.3238 | −4.4318 |
| $x_4^{-2}a$ | = | 4.9184 | −3.8368 | 5.7552 | −7.6737 |

EXAMPLE 4.2. Use Example 2.10 again: $n = 3, a := (-86, 52, -78, 104)$ with $x_0 := \overline{a}/8$. We obtain (monotonicity is missing) the above numbers (in Table 4.1) and a graphical

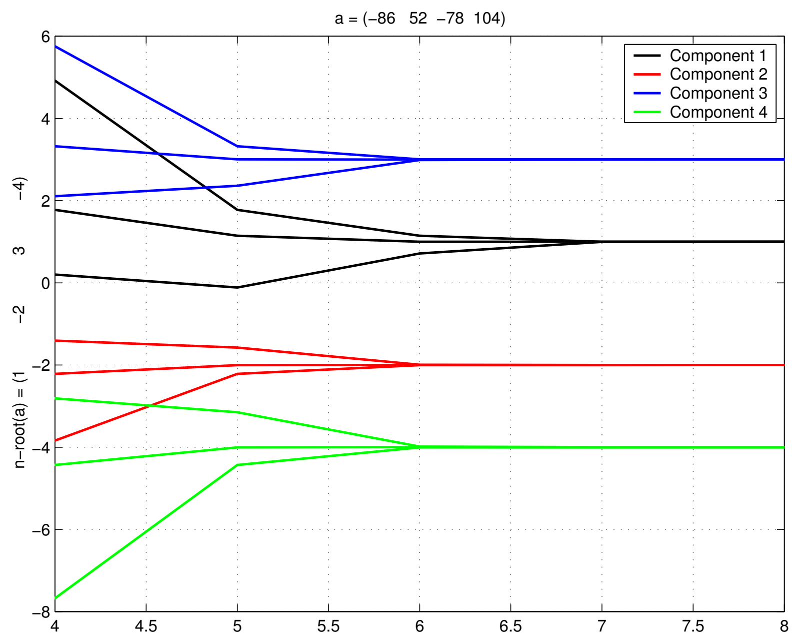


FIG. 4.1. *Inclusion property of Newton iterations from step 4 to step 8.*

representation in Figure 4.1. We also see that the inclusion is very quickly so precise that the three curves cannot be distinguished by inspection of the graph.

As we see from the table the inclusion $\sqrt[n]{a} \in \left(x_j, x_j^{1-n}a\right)$ which is valid for real roots is not true in general.

**5. Numerical behavior of Newton iterations.** There are three cases:

(i) The iterates converge quickly (quadratically).
(ii) The iterates converge slowly (linearly).
(iii) The iterates do not converge.

**Case i.)** We choose an arbitrary $a$ and select the initial guess $x_0$ so that $\overline{a}$ and $x_0$ commute ($\Rightarrow N_1 = N_2$). We observe fast (quadratic) convergence. In the Figures 7.1, 7.2, left side, p. 95, we see 16 examples for $n = 3$ and for $n = 7$, showing the absolute value of the residuals. In all examples the convergence is eventually quadratic.

**Case ii.)** We choose $a$ and $x_0$ randomly and independently. Ten examples are exhibited in Figure 5.1 where the horizontal axis represents the number of iterations and where the vertical axis represents the exponent of the absolute value of the residuals with respect to base ten. In all cases the convergence is slow (linear).

**Case iii.)** We look at the following special example.

EXAMPLE 5.1. Let $a := (0,0,1,0)$, $h := \sqrt{3+2\sqrt{2}}$, $\beta^4 := (5\sqrt{2}-7)\frac{h}{8}$, $\alpha := h\beta$ and $n = 4$. Then, ($\alpha \approx 0.9239$, $\beta \approx 0.3827$)

$$\sqrt[4]{a} \in \{(\alpha,0,\beta,0), (-\alpha,0,-\beta,0), (-\beta,0,\alpha,0), (\beta,0,-\alpha,0)\}.$$

If we start both iterations for this case with $x_0 = (0,0,0,1)$, we have $x_0\overline{a} \neq \overline{a}x_0$ and we obtain different iterates. And even worse, if we continue the computation (see Figure 5.2, showing the absolute value of the residual), we observe that the first and third component of

ETNA
Kent State University
etna@mcs.kent.edu



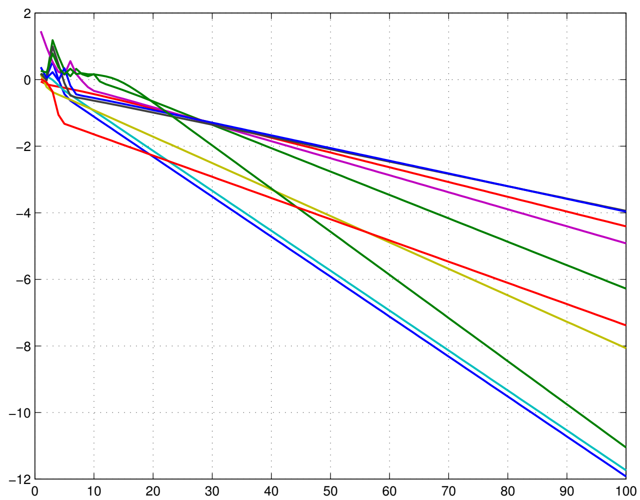


FIG. 5.1. *Fourth root of quaternion $a$, $a$ and initial guess $x_0$ random.*

all iterates will remain zero. Thus, convergence is impossible. Observe, that those elements which commute with $a$ have the form $x = (x_1, 0, x_3, 0)$.

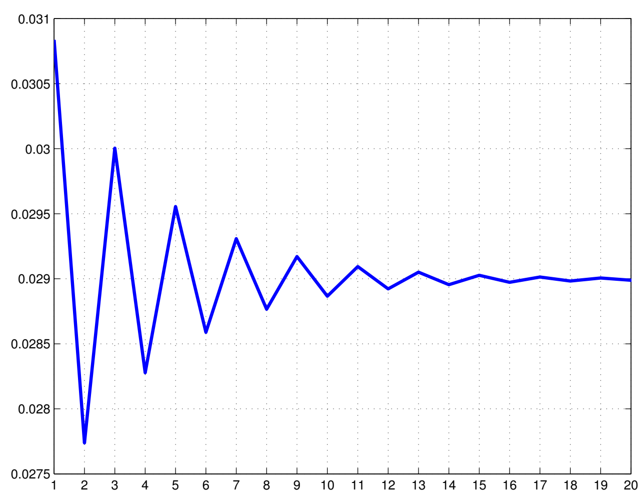


FIG. 5.2. *Fourth root of quaternion $a = (0, 0, 1, 0)$, with initial guess $x_0 = (0, 0, 0, 1)$.*

**6. Convergence of Newton iterations.** According to our previous investigations, the two Newton iterations defined in (3.2), (3.3) may converge slowly or may not converge in case the initial guess $x_0$ and the given $a$ do not commute. Therefore, we assume throughout this section that $a$ and $x_0$ commute. We already mentioned that equivalently, $x_0$ and $\overline{a}$ commute. Then, according to Lemma 3.1 the two formulas produce the same sequence. Therefore, we only use formula (3.2). We want to show that in this case the convergence is fast. The details will be specified later.

Let $g$ be defined by $g(x) := x^n - a$ where $a, x \in \mathbb{H}$ and $a \neq 0$. We will compare the iteration generated by formula (3.2) with the classical Newton iteration which is defined by the linear $(4 \times 4)$ system

$$g(x_j) + g'(x_j)\eta_j = 0, \quad x_{j+1} := x_j + \eta_j, \quad j = 0, 1, \ldots, \tag{6.1}$$

where $g'$ is the already mentioned $(4 \times 4)$ Jacobian matrix whose columns are the partial derivatives of $g$ with respect to the four components of $x = (x_1, x_2, x_3, x_4)^{\mathrm{T}}$. The equation (6.1) is a linear system for the unknown $\eta_j$ where $x_j$ is known. Here and in the sequel of this section, it is reasonable to assume that $x_j, \eta_j$ have the form of column vectors. An explicit formula for $g'$ for $n = 2$ was already given in the Introduction, formula (1.3). For the general case, we will develop a recursive and an explicit formula for $g'$. Let us denote by $g^{(j)}$ the column vector of the partial derivative of $g$ with respect to the variable $x_j,\ j = 1, 2, 3, 4$. Then $g' = (g^{(1)}, g^{(2)}, g^{(3)}, g^{(4)})$. We will use the formulas

$$(x^2)^{(j)} := (xx)^{(j)} = xx^{(j)} + x^{(j)}x, \quad j = 1, 2, 3, 4, \tag{6.2}$$

$$(x^n)^{(j)} := (xx^{n-1})^{(j)} = x(x^{n-1})^{(j)} + x^{(j)}x^{n-1}, \quad j = 1, 2, 3, 4, \quad n \geq 3. \tag{6.3}$$

Since $x = x_1 + x_2\mathbf{i} + x_3\mathbf{j} + x_4\mathbf{k}$ we have $x^{(1)} = 1, x^{(2)} = \mathbf{i}, x^{(3)} = \mathbf{j}, x^{(4)} = \mathbf{k}$. For $n = 2$ we have therefore

$$g'(x) = (x + x, x\mathbf{i} + \mathbf{i}x, x\mathbf{j} + \mathbf{j}x, x\mathbf{k} + \mathbf{k}x) = x\mathbf{E} + \mathbf{E}x,$$

where

$$\mathbf{E} := (\,1, \quad \mathbf{i}, \quad \mathbf{j}, \quad \mathbf{k}\,),$$

and the multiplications $x\mathbf{E}, \mathbf{E}x$ are not matrix multiplications but simply componentwise multiplications with the (quaternionic) constant $x$. If $\mathbf{E}$ is considered a matrix, then it is the identity matrix. For a general $n \geq 3$ we obtain from (6.3)

$$g'(x) = x\Big((x^{n-1})^{(1)}, (x^{n-1})^{(2)}, (x^{n-1})^{(3)}, (x^{n-1})^{(4)}\Big) + \mathbf{E}x^{n-1}.$$

In order for the multiplication with $x$ to be correct, each column $(x^{n-1})^{(j)}$, $j = 1, 2, 3, 4$, has to be understood as a quaternion.

Let us write instead of $g'$ a little more accurately $g_n'$ if the Jacobian matrix is derived from $g_n(x) := x^n - a$. Then the formulas (6.2), (6.3) read

$$g_2'(x) = x\mathbf{E} + \mathbf{E}x, \quad g_n'(x) = xg_{n-1}' + \mathbf{E}x^{n-1}, \quad n \geq 3. \tag{6.4}$$

From these formulas it is easy to derive the following explicit formula

$$g_{n+1}'(x) = \sum_{j=0}^{n} x^{n-j}\mathbf{E}x^j, \quad n \geq 0, \tag{6.5}$$

where we also allow $g_1' := \mathbf{E}$. In particular, we have $g_n'(0) = \mathbf{0}$ for $n \geq 2$. Since we have already computed $g_2'$ in (1.3) we can compute $g_3'$ quite easily by using (6.4):

$$g_3'(x) = xg_2'(x) + \mathbf{E}x^2 = \tag{6.6}$$

$$\begin{pmatrix} 3(x_1^2 - x_2^2 - x_3^2 - x_4^2) & -6x_1x_2 & -6x_1x_3 & -6x_1x_4 \\ 6x_1x_2 & 3x_1^2 - 3x_2^2 - x_3^2 - x_4^2 & -2x_2x_3 & -2x_2x_4 \\ 6x_1x_3 & -2x_2x_3 & 3x_1^2 - x_2^2 - 3x_3^2 - x_4^2 & -2x_3x_4 \\ 6x_1x_4 & -2x_2x_4 & -2x_3x_4 & 3x_1^2 - x_2^2 - x_3^2 - 3x_4^2 \end{pmatrix}.$$

ETNA
Kent State University
etna@mcs.kent.edu



This expression is quite complicated. However, we do not need any explicit formula like (6.6) for numerical purposes, because we can create the needed values by evaluating (6.4), or (6.5) directly.

We shall show below that, roughly, the classical Newton iterates governed by (6.1) are identical with the iterates produced by (3.2) or (3.3). However, there is a difference in the break down behavior. We have already seen (proof of Lemma 3.4) that the iteration defined by (3.2) can break down if and only if $N_1(x) = 0$, which would imply that the Jacobian matrix $g_n'(x)$ is the zero matrix. Thus, the classical Newton iteration will also break down. However, there is the possibility that $g_n'$ is not the zero matrix but nevertheless singular, implying that the classical Newton iteration breaks down, whereas the other iteration still works. It is best to present an example for this case.

EXAMPLE 6.1. Let $n = 4,\ a = x_0 = (0, 0, 1, 0)$. Then (cf. (6.5))

$$g_4'(x_0) = \begin{pmatrix} 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \\ -4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and the classical Newton iteration cannot be continued. However, $x_1 := N_1(x_0) = (-1/4, 0, 3/4, 0)$ and the following values converge quickly to $(-\beta, 0, \alpha, 0)$. Compare to Example 5.1. A remedy would be to start the classical Newton iteration with $x_1$.

The connection between the two iterations (3.2) and (6.1) is established in the following theorem.

THEOREM 6.2. *Let $g_n$ be defined by $g_n(x) := x^n - a$ for $x, a \in \mathbb{H},\ a \neq 0$ and $n \geq 2$. Let the initial guess $x_0 \neq 0$ commute with $\overline{a}$ and let $x_0$ be the same for both iterations (3.2), (6.1). Then, both iterations produce the same sequences, provided the Jacobian matrix $g_n'$ is not singular.*

*Proof*. We prove that

$$\eta_0 := \frac{1}{n}\Big(x_0^{1-n}a - x_0\Big) \tag{6.7}$$

solves (6.1) for $j = 0$. This is sufficient because of $x_1 = x_0 + \eta_0 = x_0 + \frac{1}{n}\Big(x_0^{1-n}a - x_0\Big) = \frac{1}{n}\Big((n-1)x_0 + x_0^{1-n}a\Big) =: N_1(x_0)$. If we use formula (6.5) we have to show that

$$x_0^n - a + \frac{1}{n}\Big[\sum_{j=0}^{n-1} x_0^{n-1-j}\mathbf{E}x_0^j\Big]\Big(x_0^{1-n}a - x_0\Big) = 0.$$

Inside the square brackets are matrices. Vectors are in round or in no parentheses. The former equation is equivalent to

$$n(x_0^n - a) + \Big[\sum_{j=0}^{n-1} x_0^{n-1-j}\mathbf{E}x_0^j\Big]x_0^{1-n}a - \Big[\sum_{j=0}^{n-1} x_0^{n-1-j}\mathbf{E}x_0^j\Big]x_0 = 0.$$

Thus, it suffices to show that

$$\Big[\sum_{j=0}^{n-1} x_0^{n-1-j}\mathbf{E}x_0^j\Big]x_0 = nx_0^n, \quad \Big[\sum_{j=0}^{n-1} x_0^{n-1-j}\mathbf{E}x_0^j\Big]x_0^{1-n}a = na.$$

ETNA
Kent State University
etna@mcs.kent.edu



The first equation is a special case of the second equation, put $a = x_0^n$. It is therefore sufficient to show the validity of the second equation. We prove the second equation by induction. We shall use that $a$ and $x_0$ commute with the consequence that $a$ and $x_0^k$ also commute for all $k \in \mathbb{Z}$. See (3.6). For $n = 1$ the equation is true. Suppose it is true as it stands. Then

$$\Big[\sum_{j=0}^{n} x_0^{n-j}\mathbf{E}x_0^j\Big]x_0^{-n}a = \Big[\sum_{j=0}^{n-1} x_0^{n-j}\mathbf{E}x_0^j + \mathbf{E}x_0^n\Big]x_0^{-n}a$$

$$= \underbrace{x_0\,\underbrace{\Big[\sum_{j=0}^{n-1} x_0^{n-1-j}\mathbf{E}x_0^j\Big]x_0^{1-n}a}_{=na}\,x_0^{-1}}_{=na} + \underbrace{\Big[\mathbf{E}x_0^n\Big]x_0^{-n}a}_{=a} = (n+1)a.$$

Thus, we have shown, that $\eta_0$ solves (6.1) for $j = 0$. This will even be true, if $g_n'$ is singular. □

By this theorem we have shown, that the iteration defined by (3.2) coincides with the classical Newton iteration via the Jacobian matrix $g'$ of the partial derivatives. Therefore, all known features are valid: The iteration converges locally and quadratically to one of the roots. The iteration generated by (3.2) has the advantage that, numerically, the case $N_1(x) = 0$ is practically impossible (cf. Proof of Lemma 3.4) since this requires, that the components of $x$ are irrational numbers which, however, have in general no representation in a computer.

In the last section (no. 9) we shall give an independent proof for the local, quadratic convergence of Newton's method for finding roots by showing that an analogue of Taylor's theorem can be applied to $N_1$ or $N_2$.

**7. The Gâteaux derivative and the damped Newton iteration.** The Gâteaux derivative of a mapping $g : \mathbb{H} \to \mathbb{H}$ was already defined in (1.2). Let $g_n(x) := x^n - a$ for $x, a \in \mathbb{H}$, then

$$g_n'(x,h) = \sum_{j=0}^{n-1} x^{n-1-j}hx^j.$$

For real $h$ this specializes to $g_n'(x,h) = nhx^{n-1}$ and if we introduce this expression into the classical Newton form (1.1) (replacing $g'(x)$ with $g_n'(x,h)$) we obtain

$$x_{\text{new}} := N(x) := x + \frac{1}{nh}\big(x^{1-n}a - x\big)$$

which coincides with $N_1$ defined in (3.2) if $h = 1$, otherwise it can be regarded as a damped Newton form with *damping factor* $\lambda := 1/h$. Damping is normally used in the beginning of the iteration. It enlarges (sometimes) the basin of attraction. In order to apply damping we write

$$x_{\text{new}}(\lambda) := N(x,\lambda) := x + \lambda\frac{1}{n}\big(x^{1-n}a - x\big) \tag{7.1}$$

and carry out the following test

$$|g_n(x_{\text{new}}(\lambda))| < |g_n(x)|, \quad \lambda := 1, \frac{1}{2}, \frac{1}{4}, \ldots$$

The first (largest) $\lambda$ which passes this test will be used to define $x_{\text{new}}(\lambda)$ for the next step. This strategy proved to be very useful in all examples we used.

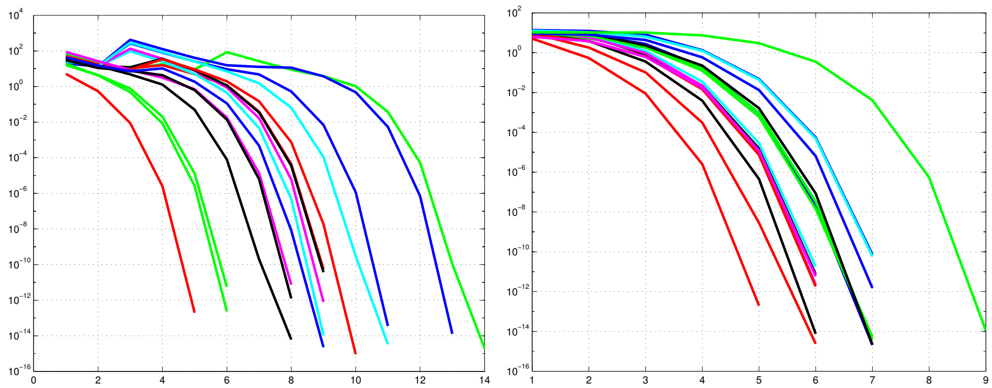

FIG. 7.1. *Newton without and with damping, applied to the computation of third roots.*

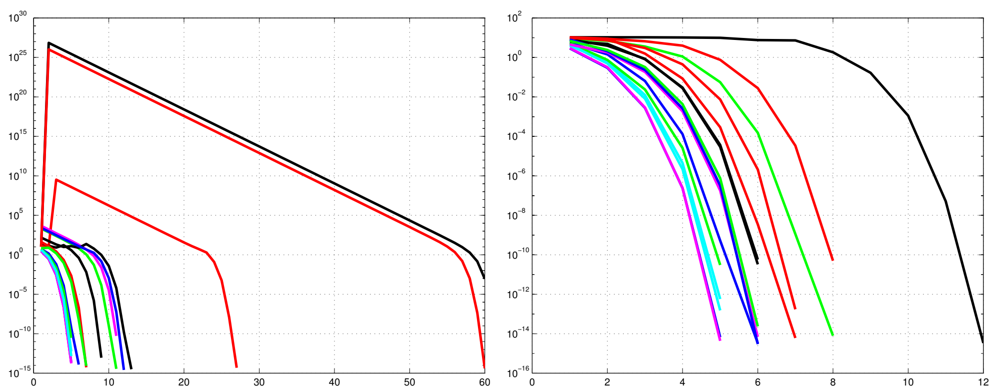

FIG. 7.2. *Newton without and with damping, applied to the computation of seventh roots.*

As expected, the damping is used only in the beginning of the iteration, with the consequence that the convergence order is not changed, and, in addition, only few damping steps were applied. We show the effect in Figures 7.1 and 7.2, where 16 cases are exhibited each for $n = 3$ and $n = 7$. The initial data are identical for the undamped and damped case. In the case of $n = 2$ the undamped and damped case look alike.

We also compared the number of calls of $N$ (defined in (7.1)) for the damped Newton iteration and for $N_1$ (defined in (3.2)) for the undamped Newton iteration. For $n = 2$ and $n = 3$ these numbers are similar, but from $n = 5$ on there is a clear difference. We made 1000 tests for $n = 3, 5$, and for $n = 7$. For $n = 5$ the number of calls with damping is about 22% smaller than that without damping. For $n = 7$ those figure is 25%.

**8. The Schur decomposition of quaternions.** We start with a definition.

DEFINITION 8.1. *Let $a_1, a_2, a_3, a_4$ be any four real numbers. We form the two complex numbers $\alpha := a_1 + a_2\mathbf{i}$, $\beta := a_3 + a_4\mathbf{i}$ and the following two matrices:*

$$\mathbf{A} := \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix}, \quad \mathbf{B} := \begin{pmatrix} a_1 & -a_2 & -a_3 & -a_4 \\ a_2 & a_1 & -a_4 & a_3 \\ a_3 & a_4 & a_1 & -a_2 \\ a_4 & -a_3 & a_2 & a_1 \end{pmatrix}. \tag{8.1}$$

*The matrix* $\mathbf{A}$ *will be called* complex q-matrix*, the matrix* $\mathbf{B}$ *will be called* real q-matrix.

Both types of matrices are isomorphic to quaternions $a := (a_1, a_2, a_3, a_4)$ with respect to matrix multiplication. We have $|a| = ||\mathbf{A}|| = ||\mathbf{B}||$ with the consequence that the conditions of $\mathbf{A}$ and $\mathbf{B}$ are equal to one. Further, $\mathbf{A}\mathbf{A}^* = |a|^2\,\mathbf{I}$, $\mathbf{B}\mathbf{B}^{\mathrm{T}} = |a|^2\,\mathbf{I}$. The eigenvalues of $\mathbf{A}$ and $\mathbf{B}$ are the same, only in $\mathbf{B}$ all eigenvalues appear twice. The two eigenvalues of $\mathbf{A}$ are $\sigma_\pm := a_1 \pm \sqrt{a_2^2 + a_3^2 + a_4^2}\,\mathbf{i}$. They are distinct if $a \notin \mathbb{R}$.

In Björck and Hammarling [2] the authors develop methods to finding the square root of a matrix. In more recent papers these methods are extended to the computation of $n$-th roots of matrices, Smith [12], Higham [6], Iannazzo [7]. For finding a root of a matrix $\mathbf{C}$ the authors use the Schur decomposition of $\mathbf{C}$. If $\mathbf{C}$ is any complex square matrix, then the (complex) Schur decomposition which always exists has the form

$$\mathbf{S} = \mathbf{U}^*\mathbf{C}\mathbf{U},$$

where $\mathbf{S}$ is upper triangular, thus, having the eigenvalues of $\mathbf{C}$ on its diagonal, and $\mathbf{U}$ is unitary (i.e. $\mathbf{U}^*\mathbf{U} = \mathbf{I}$). If one knows an $n$-th root $\mathbf{Y}$ of $\mathbf{S}$, then $\mathbf{C} = \mathbf{U}\mathbf{S}\mathbf{U}^* = \mathbf{U}\mathbf{Y}^n\mathbf{U}^* = (\mathbf{U}\mathbf{Y}\mathbf{U}^*)^n =: \mathbf{X}^n$. Thus, $\mathbf{X}$ is an $n$-th root of $\mathbf{C}$.

An application to quaternions results in the question: Can $\mathbf{A}$ or $\mathbf{B}$ have a Schur decomposition, in terms of q-matrices? If we pose this problem for complex q-matrices we have to ask whether a decomposition of the following form is possible:

$$\begin{pmatrix} \sigma_+ & 0 \\ 0 & \sigma_- \end{pmatrix} =: \begin{pmatrix} \sigma & 0 \\ 0 & \overline{\sigma} \end{pmatrix} = \begin{pmatrix} \overline{u} & -v \\ \overline{v} & u \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix} \begin{pmatrix} u & v \\ -\overline{v} & \overline{u} \end{pmatrix}, \tag{8.2}$$

where $\alpha, \beta$ are arbitrary, given complex numbers and $\sigma, u, v$ are wanted complex numbers such that $|u|^2 + |v|^2 = 1$. If we rewrite this equation with quaternions, it reads

$$\sigma = \overline{\mathbf{u}}a\mathbf{u}, \quad |\mathbf{u}| = 1, \tag{8.3}$$

where $\mathbf{u}$ is the quaternion defining the q-matrix $\mathbf{U}$, i. e. $\mathbf{u} := (\Re(u), \Im(u), \Re(v), \Im(v))$. Since $|\mathbf{u}|^2 = |u|^2 + |v|^2 = 1$ we have $\mathbf{u}^{-1} = \overline{\mathbf{u}}$. Thus, equation (8.3) defines an equivalence between $\sigma$ and $a$. Our former Lemma 2.2 confirms that $\sigma$ and $a$ are indeed equivalent. This may be summarized as follows.

THEOREM 8.2. *Let $a$ be a quaternion and $\sigma$ the complex representative of $a$. Then (8.3) is the Schur decomposition of $a$.*

*Proof.* Rewrite (8.3) in form of complex q-matrices. ☐

In terms of quaternions, the application of the Schur decomposition leads to the explicit determination of the roots as already described in Section 2.

Because of the isomorphy between complex and real q-matrices, corresponding results for real q-matrices can be directly copied from the case of complex q-matrices and are deleted here.

In order to find $\mathbf{u}$, equation (8.3) may be regarded as a linear, homogeneous, real system of four equations in the four components of $\mathbf{u}$. In a former paper, [8], we have already solved a similar system. It has the form

$$\mathbf{D}\mathbf{u} = \mathbf{0}, \quad \mathbf{D} := \begin{pmatrix} 0 & -a_2 + |a_v| & -a_3 & -a_4 \\ a_2 - |a_v| & 0 & -a_4 & a_3 \\ a_3 & a_4 & 0 & -a_2 - |a_v| \\ a_4 & -a_3 & a_2 + |a_v| & 0 \end{pmatrix}.$$

The matrix has rank two for $a \in \mathbb{H}\backslash\mathbb{R}$. We find two independent solutions as follows:

$$\mathbf{u}_1 := (|a_v| + a_2, |a_v| + a_2, a_3 - a_4, a_3 + a_4), \quad \mathbf{u}_2 := (a_3 - a_4, a_3 + a_4, |a_v| - a_2, |a_v| - a_2),$$

provided $a_3$ or $a_4$ is not vanishing. In case $a_3 = a_4 = 0$ and $a_2 > 0$, $\mathbf{u}_1 := (1,0,0,0)$, $\mathbf{u}_2 := (0,1,0,0)$ are independent solutions. In case $a_3 = a_4 = 0$ and $a_2 < 0$, $\mathbf{u}_1 := (0,0,1,0)$, $\mathbf{u}_2 := (0,0,0,1)$ are independent solutions. The general solution of (8.3) and of (8.2) as well is, therefore,

$$\mathbf{u} := \frac{\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2}{|\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2|}, \quad \alpha_1, \alpha_2 \in \mathbb{R}, \quad |\alpha_1| + |\alpha_2| > 0. \tag{8.4}$$

We could choose $\alpha_1, \alpha_2$ such that one of the four components of $\mathbf{u}$ is vanishing, which would simplify the resulting matrix $\mathbf{U}$ slightly. E. g. $\alpha_1 := -a_3 - a_4, \alpha_2 := |a_v| + a_2$ would make the second component of $\mathbf{u}$ vanish and the corresponding complex $\mathbf{U}$ would have a real diagonal (provided $|a_3| + |a_4| > 0$). But we would like to point out that the considerations of this section are of theoretical nature and not used in our numerical computations. The Householder transformation, developed from (2.5) to (2.6) is to our taste much neater and does not need the explicit knowledge of $\mathbf{u}$.

In view of the isomorphic representations (8.1) of quaternions in matrix forms, it is of course tempting to use matrix algorithms for treating quaternions. As far as only elementary arithmetic operations are used, there will be no problem. But there is already a difference in the amount of arithmetic work. To invert a quaternion, 11 (real) flops are needed. To invert a corresponding complex $(2 \times 2)$ matrix requires 300 flops and to invert a real $(4 \times 4)$ matrix requires 350 flops (`matlab` counts). Since in general matrix operations do not know about the underlying quaternionic structure, problems of ignoring the matrix structure can be avoided by simply using quaternion arithmetic. This is supported in two papers by Dongarra, Gabriel, Koelling, and Wilkinson, [3], [4]. There is a very simple example, see the present authors [9], of computing eigenvalues of a quaternion valued $(2 \times 2)$ matrix where an application of an eigenvalue algorithm to the corresponding complex $(4 \times 4)$ matrix gives bad results. The matrix structure is ignored and the precision is reduced significantly. Another example: If one computes the `matlab` Schur decomposition of $\mathbf{A}$, $\mathbf{B}$ the resulting unitary matrices $\mathbf{U}$ do not belong into the class of q-matrices.

**9. Taylor for $x^n$ in the quaternionic case.** The question is whether there are some possibilities to extend Taylor's theorem also to quaternionic valued functions, though derivatives in the strong (complex) sense do not exists. We will only treat the question for simple functions $f$ defined by

$$f(x) := x^n, \quad n \in \mathbb{Z},\ x \in \mathbb{H},$$

and we will replace derivatives of $f$ by the derivatives we know from the real and complex case, namely

$$f'(x) := nx^{n-1}, \quad f''(x) := n(n-1)x^{n-2}, \quad n \in \mathbb{Z},\ x \in \mathbb{H}, \tag{9.1}$$

and we will call these functions, $f', f''$ *derivatives*. We shall show that a Taylor formula of the form

$$f(x) = f(x_0) + f'(\xi)(x - x_0), \tag{9.2}$$

is possible which reads in our special case

$$x^n = x_0^n + n\xi^{n-1}(x - x_0), \tag{9.3}$$

which leads for $n \neq 0$ to

$$\xi^{n-1} = \frac{1}{n}(x^n - x_0^n)(x - x_0)^{-1}. \tag{9.4}$$

ETNA
Kent State University
etna@mcs.kent.edu



That means we can find $n-1$ values of $\xi$ such that formula (9.2) is valid. However, this is quite trivial. What we want to know is some information on the location of $\xi$ in relation to $x$ and $x_0$. If we do not make special assumptions on $x$ and $x_0$ we are not able to make forecasts about $\xi$. But if we assume that $x, x_0$ commute then the situation changes. For commuting $x, x_0$ we have the formula

$$f'(\xi) = (x^n - x_0^n)(x - x_0)^{-1} = \sum_{j=0}^{n-1} x^j x_0^{n-j-1}, \quad n \geq 1. \tag{9.5}$$

The same formula for negative $m$ reads

$$f'(\xi) = (x^m - x_0^m)(x - x_0)^{-1} = -\sum_{j=0}^{-m-1} x^{-j-1} x_0^{j+m}, \quad m \leq -1. \tag{9.6}$$

These formulas are also valid for $m = n = 0$, but they are trivial in this case. If we go one step further with Taylor's formula we obtain

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(\eta)}{2}(x - x_0)^2. \tag{9.7}$$

If we put $f(x) := x^n$ then for $\eta$ we obtain (for $n \neq 0, n-1 \neq 0$) the formula

$$\eta^{n-2} = \frac{2}{n(n-1)}\Big((x^n - x_0^n)(x - x_0)^{-2} - n x_0^{n-1}(x - x_0)^{-1}\Big). \tag{9.8}$$

With the help of (9.4), (9.5), and (9.6) we obtain

$$\begin{aligned} \frac{f''(\eta)}{2} &= (x^n - x_0^n)(x - x_0)^{-2} - n x_0^{n-1}(x - x_0)^{-1} \\ &= \sum_{j=1}^{n-1} (n-j) x^{j-1} x_0^{n-j-1}, \quad n \geq 1, \end{aligned} \tag{9.9}$$

$$\begin{aligned} \frac{f''(\eta)}{2} &= (x^m - x_0^m)(x - x_0)^{-2} - m x_0^{m-1}(x - x_0)^{-1} \\ &= \sum_{j=0}^{-m-1} (-m-j) x^{-j-1} x_0^{m+j-1}, \quad m \leq -1. \end{aligned} \tag{9.10}$$

If we express $\xi^{n-1}$ defined in (9.4) either by (9.5) or by (9.6) and $\eta^{n-2}$ defined in (9.8) either by (9.9) or by (9.10), then $\xi^{n-1}, \eta^{n-2}$ have one common feature. They all represent convex combinations. Therefore, we have the following inclusion properties:

$$\xi^{n-1} \in \Big(\min_{j=0,1,\ldots,n-1} x^j x_0^{n-j-1}, \max_{j=0,1,\ldots,n-1} x^j x_0^{n-j-1}\Big), \ n \geq 1, \tag{9.11}$$

$$\xi^{m-1} \in \Big(\min_{j=0,1,\ldots,-m-1} x^{-j-1} x_0^{m+j}, \max_{j=0,1,\ldots,-m-1} x^{-j-1} x_0^{m+j}\Big), \ m \leq -1, \tag{9.12}$$

$$\eta^{n-2} \in \Big(\min_{j=1,2,\ldots,n-1} x^{j-1} x_0^{n-j-1}, \max_{j=1,2,\ldots,n-1} x^{j-1} x_0^{n-j-1}\Big), \ n \geq 2, \tag{9.13}$$

$$\eta^{m-2} \in \Big(\min_{j=0,1,\ldots,-m-1} x^{-j-1} x_0^{m+j-1}, \max_{j=0,1,\ldots,-m-1} x^{-j-1} x_0^{m+j-1}\Big), \ m \leq -1, \tag{9.14}$$

where in all cases the minima and maxima have to be applied componentwise. More exactly, one could also say that these values are all contained in the convex hull of the given points. The situation is particularly simple in the cases where $n$ is small:

**ETNA**
Kent State University
etna@mcs.kent.edu



$$\xi = \frac{1}{2}(x + x_0), \quad n = 2,$$
$$\xi^2 = \frac{1}{3}(x^2 + xx_0 + x_0^2), \quad n = 3,$$
$$\xi^3 = \frac{1}{4}(x^3 + x^2x_0 + xx_0^2 + x_0^3), \quad n = 4,$$
$$\xi^{-2} = x^{-1}x_0^{-1}, \quad m = -1,$$
$$\xi^{-3} = \frac{1}{2}(x^{-2}x_0^{-1} + x^{-1}x_0^{-2}), \quad m = -2$$
$$\xi^{-4} = \frac{1}{3}(x^{-3}x_0^{-1} + x^{-2}x_0^{-2} + x^{-1}x_0^{-3}), \quad m = -3$$
$$\eta = \frac{1}{3}(2x_0 + x), \quad n = 3,$$
$$\eta^2 = \frac{1}{6}(3x_0^2 + 2xx_0 + x^2), \quad n = 4,$$
$$\eta^3 = \frac{1}{10}(4x_0^3 + 3x_0^2x + 2x_0x^2 + x^3), \quad n = 5,$$
$$\eta^{-3} = x^{-1}x_0^{-2}, \quad m = -1,$$
$$\eta^{-4} = \frac{1}{3}(x^{-2}x_0^{-2} + 2x^{-1}x_0^{-3}), \quad m = -2,$$
$$\eta^{-5} = \frac{1}{6}(x^{-3}x_0^{-2} + 2x^{-2}x_0^{-3} + 3x^{-1}x_0^{-4}), \quad m = -3.$$

We summarize our results so far.

THEOREM 9.1. *(Taylor form 1) Let $f : \mathbb{H} \to \mathbb{H}$ be defined by $f(x) := x^n, n \in \mathbb{Z}$, and define $f', f''$ according to (9.1). Assume that $x, x_0 \in \mathbb{H}$ commute. Then there is an element $\xi \in \mathbb{H}$ and an element $\eta \in \mathbb{H}$ such that*

$$f(x) = f(x_0) + f'(\xi)(x - x_0),$$
$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(\eta)}{2}(x - x_0)^2,$$

*where for $\xi, \eta$ we have the inclusions given in (9.11) to (9.14).*

We are mainly interested in the case where

$$x - x_0 =: \varepsilon$$

is small. The commutation of $x, x_0$ implies that also $\varepsilon$ commutes with $x$ and with $x_0$ because

$$\varepsilon x = (x - x_0)x = x^2 - x_0x = x^2 - xx_0 = x\varepsilon,$$
$$\varepsilon x_0 = (x - x_0)x_0 = xx_0 - x_0^2 = x_0x - x_0^2 = x_0\varepsilon.$$

Since the commutation of $x, x_0$ also implies the commutation of $x^j, x_0^k$ for arbitrary $j, k \in \mathbb{Z}$, this applies also for the two commuting pairs $\varepsilon, x$; $\varepsilon, x_0$. Thus, the binomial formula for $x^j = (x_0 + \varepsilon)^j$ is valid in the ordinary sense.

THEOREM 9.2. *(Taylor form 2) Let $f : \mathbb{H} \to \mathbb{H}$ be defined by $f(x) := x^n, n \in \mathbb{Z}$, and define $f', f''$ according to (9.1). Assume that $x, x_0 \in \mathbb{H}$ commute. Then with $\varepsilon := x - x_0$ we have*

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + O(\varepsilon^2), \tag{9.15}$$
$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + O(\varepsilon^3), \tag{9.16}$$

**ETNA**
Kent State University
etna@mcs.kent.edu



*where $O(h)$ is an abbreviation for an expression with the property*

$$\lim_{|h|\to 0} O(h)h^{-1} = \text{const}\,.$$

*Proof.* (i) Let $n \geq 1$. [a] From (9.2) and (9.5) by letting $x := x_0 + \varepsilon$ we obtain

$$\begin{aligned}
f(x) &= f(x_0) + \sum_{j=0}^{n-1} (x_0 + \varepsilon)^j x_0^{n-j-1}\varepsilon \\
&= f(x_0) + \sum_{j=0}^{n-1} \Big( \sum_{k=0}^{j} \binom{j}{k} x_0^{j-k}\varepsilon^k \Big) x_0^{n-j-1}\varepsilon \\
&= f(x_0) + \sum_{j=0}^{n-1}\sum_{k=0}^{j} \binom{j}{k} x_0^{n-k-1}\varepsilon^{k+1} \\
&= f(x_0) + \sum_{j=0}^{n-1} \Big( x_0^{n-1}\varepsilon + j x_0^{n-2}\varepsilon^2 + \cdots \Big) \\
&= f(x_0) + f'(x_0)(x - x_0) + \sum_{j=1}^{n-1} \Big( j x_0^{n-2}\varepsilon^2 + \cdots \Big) \\
&= f(x_0) + f'(x_0)(x - x_0) + O(\varepsilon^2).
\end{aligned}$$

[b] From (9.7) and (9.9) by letting $x = x_0 + \varepsilon$ we obtain

$$\begin{aligned}
f(x) &= f(x_0) + f'(x_0)(x - x_0) + \sum_{j=1}^{n-1} (n-j)(x_0 + \varepsilon)^{j-1} x_0^{n-j-1}\varepsilon^2 \\
&= f(x_0) + f'(x_0)(x - x_0) + \sum_{j=1}^{n-1} (n-j) \Big( \sum_{k=0}^{j-1} \binom{j-1}{k} x_0^{j-1-k}\varepsilon^k \Big) x_0^{n-j-1}\varepsilon^2 \\
&= f(x_0) + f'(x_0)(x - x_0) + \sum_{j=1}^{n-1} (n-j) \sum_{k=0}^{j-1} \binom{j-1}{k} x_0^{n-2-k}\varepsilon^{k+2} \\
&= f(x_0) + f'(x_0)(x - x_0) + \sum_{j=1}^{n-1} (n-j) \Big( x_0^{n-2}\varepsilon^2 + (j-1)x_0^{n-2}\varepsilon^3 + \cdots \Big) \\
&= f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \sum_{j=2}^{n-1} (n-j) \Big( (j-1)x_0^{n-2}\varepsilon^3 + \cdots \Big) \\
&= f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + O(\varepsilon^3).
\end{aligned}$$

(ii) Now, let $m \leq -1$ and define $e$ by $x = x_0 + e x_0$. Then, $\varepsilon := x - x_0 = e x_0$. Assume that $e, \varepsilon$ are small. [a] We use (9.2) and (9.6) and obtain

$$\begin{aligned}
f(x) &= f(x_0) - \sum_{j=0}^{-m-1} (x_0 + e x_0)^{-j-1} x_0^{j+m} e x_0 \\
&= f(x_0) - \sum_{j=0}^{-m-1} x_0^{-j-1}(1+e)^{-j-1} x_0^{j+m+1} e
\end{aligned}$$

$$= f(x_0) - x_0^m \sum_{j=0}^{-m-1} (1+e)^{-j-1} e = f(x_0) - x_0^m \sum_{j=0}^{-m-1} (1 - e + e^2 - e^3 \cdots)^{j+1} e$$
$$= f(x_0) - x_0^{m-1}(x - x_0) \sum_{j=0}^{-m-1} (1 - e + e^2 - e^3 \cdots)^{j+1}$$
$$= f(x_0) - x_0^{m-1}(x - x_0)\Big( - m - c_1 e + c_2 e^2 - c_3 e^3 \cdots \Big)$$
$$= f(x_0) + f'(x_0)(x - x_0) + c_1 x_0^m e^2 + \cdots = f(x_0) + f'(x_0)(x - x_0) + O(\varepsilon^2),$$

where $c_1, c_2, c_3, \ldots$ are positive constants (e.g. $c_1 = \frac{-m(-m+1)}{2}$).

[b] We use (9.7) and (9.10) and obtain

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \sum_{j=0}^{-m-1} (-m-j)(x_0 + e x_0)^{-j-1} x_0^{m+j-1} x_0^2 e^2$$
$$= f(x_0) + f'(x_0)(x - x_0) + \sum_{j=0}^{-m-1} (-m-j) x_0^{-j-1} (1+e)^{-j-1} x_0^{m+j+1} e^2$$
$$= f(x_0) + f'(x_0)(x - x_0) + x_0^m \sum_{j=0}^{-m-1} (-m-j)(1+e)^{-j-1} e^2$$
$$= f(x_0) + f'(x_0)(x - x_0) + x_0^m \sum_{j=0}^{-m-1} (-m-j)(1 - e + e^2 - \cdots)^{j+1} e^2$$
$$= f(x_0) + f'(x_0)(x - x_0) + x_0^m \sum_{j=0}^{-m-1} (-m-j)(1 - c_1^{(j)} e + c_2^{(j)} e^2 - \cdots) e^2$$
$$= f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + O(\varepsilon^3),$$

where the constants $c_1^{(j)}, c_2^{(j)}, \ldots$ could be computed by a recursion formula. □

Some generalizations are possible. If we multiply the formulas given in Theorem 9.1, and Theorem 9.2 from the left by any constant $a \in \mathbb{H}$ and take into account the fact that $aO(h) = O(h)$ then we see that we can apply these theorems also to $f(x) := ax^n$, $n \in \mathbb{Z}$, where the derivatives of $f$ are defined as usual. If $f, g$ are two functions for which the two theorems are valid, then these theorems are also valid for the sum $f + g$ because of $O(h) + O(h) = O(h)$. Since Newton's formula for computing the root is a sum of this type we have the following result.

COROLLARY 9.3. *Let $a, x \in \mathbb{H}$ and let $r$ be one of the possible solutions of $r^n = a$ for $n \geq 2$ and assume that $r$ is commuting with $x$. Define*

$$N(x) := \frac{1}{n}\Big((n-1)x + ax^{1-n}\Big).$$

*Then $a$ is also commuting with $x$ and*

$$N(x) = r + \frac{n-1}{2} r^{-1}(x - r)^2 + O((x - r)^3). \tag{9.17}$$

*Proof.* Since $r$ and $x$ commute we have $xr = rx$ implying $r = x^{-1}rx$ and $r^n = (x^{-1}rx)^n = x^{-1}r^n x$. Since $r^n = a$ the elements $a$ and $x$ commute. Formula (9.17) is the second Taylor formula of Theorem 9.2. □

This corollary proves the local, quadratic convergence of Newton's method for computing quaternionic roots without relying on any global theory.

COROLLARY 9.4. *Let $m \leq 0 \leq n$ and let $\Pi_{m,n}$ be the set of all polynomials of the form*

$$p(z) := \sum_{j=m}^{n} a_j z^j, \quad a_j \in \mathbb{H}.$$

*Define the first derivative $p'$ and the second derivative $p''$ of $p$ as in the complex case. Let $x, x_0 \in \mathbb{H}$ be commuting elements. Then for $p \in \Pi_{m,n}$ we have*

$$p(x) = p(x_0) + p'(x_0)(x - x_0) + O((x - x_0)^2);$$
$$p(x) = p(x_0) + p'(x_0)(x - x_0) + \frac{p''(x_0)}{2}(x - x_0)^2 + O((x - x_0)^3).$$

**Acknowledgment.** The authors acknowledge with pleasure the support of the Grant Agency of the Czech Republic (grant No. 201/06/0356). The work is a part of the research project MSM 6046137306 financed by MSMT, Ministry of Education, Youth and Sports, Czech Republic. The authors also thank Professor Ron B. Guenther, Oregon State University, Corvallis, Oregon, USA, for valuable advice.

REFERENCES

[1] P. ARBENZ AND G. H. GOLUB, *QR-like algorithms for symmetric arrow matrices*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 655–658.
[2] Å. BJÖRCK AND S. HAMMARLING, *A Schur method for the square root of a matrix*, Linear Algebra Appl., 52/53 (1983), pp. 127–140.
[3] J. J. DONGARRA, J. R. GABRIEL, D. D. KOELLING, AND J. H. WILKINSON, *Solving the secular equation including spin orbit coupling for systems with inversion and time reversal symmetry*, J. Comput. Phys., 54 (1984), pp. 278–288.
[4] J. J. DONGARRA, J. R. GABRIEL, D. D. KOELLING, AND J. H. WILKINSON, *The eigenvalue problem for hermitian matrices with time reversal symmetry*, Linear Algebra Appl., 60 (1884), pp. 27–42.
[5] R. FUETER, *Die Funktionentheorie der Differentialgleichungen $\Delta u = 0$ und $\Delta\Delta u = 0$ mit vier reellen Variablen*, Comment. Math. Helv., 7 (1935), pp. 307–330.
[6] N. HIGHAM, *Convergence and stability of iterations for matrix functions*, 21st Biennial Conference on Numerical Analysis, Dundee, 2005.
[7] B. IANNAZZO, *On the Newton method for the matrix pth root*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 503–523.
[8] D. JANOVSKÁ AND G. OPFER, *Givens' transformation applied to quaternion valued vectors*, BIT, 43 (2003), Suppl., pp. 991–1002.
[9] D. JANOVSKÁ AND G. OPFER, *Fast Givens transformation for quaternionic valued matrices applied to Hessenberg reductions*, Electron. Trans. Numer. Anal., 20 (2005), pp. 1–26. http://etna.math.kent.edu/vol.20.2005/pp1-26.dir/pp1-26.html.
[10] G. KUBA, *Wurzelziehen aus Quaternionen*, Mitt. Math. Ges. Hamburg, 23/1 (2004), pp. 81–94 (in German: Finding roots of quaternions).
[11] J. K. REID, *Solution of linear systems of equations: direct methods*, in Sparse Matrix Techniques, V. A. Barker, ed., Lecture Notes in Math., 572, Springer, Berlin, 1977, 109.
[12] M. I. SMITH, *A Schur algorithm for computing matrix pth roots*, SIAM J. Matrix Anal. Appl., 24 (2003), pp. 971–989.
[13] A. SUDBERY, *Quaternionic analysis*, Math. Proc. Camb. Phil. Soc., 85 (1979), pp. 199–225.
[14] O. WALTER, L. S. LEDERBAUM, AND J. SCHIRMER, *The eigenvalue problem for 'arrow' matrices*, J. Math. Phys., 25 (1984), pp. 729–737.
[15] F. ZHANG, *Quaternions and matrices of quaternions*, Linear Algebra Appl., 251 (1997), pp. 21–57.