Available online at http://scik.org
J. Math. Comput. Sci. 3 (2013), No. 1, 332-351
ISSN: 1927-5307

USING NEWTON-KANTOROVICH METHOD TO COMPUTE QR AND $(L+I)U$- FACTORIZATIONS

MARIO PAUL AHUES[1,*], MUZAFAR FATTAH HAMA[2]

[1]Department of Mathematics, University of Jean Monnet, 42023 SAINT-ETIENNE, France

[2]School of Science - Department of Mathematics, University of Sulaimani, Sulaimani, Iraq

Abstract. In this paper, we apply Newton-Kantorovich method to compute iteratively a QR or an $(\mathsf{L}+\mathsf{I})\mathsf{U}$ factorization of a quasi-upper-triangular nonsingular matrix with no zero diagonal entries.

Keywords: $(\mathsf{L}+\mathsf{I})\mathsf{U}$ factorization; Newton method; QR factorization.

2000 AMS Subject Classification: 65F15; 65F30; 15A23

1. Introduction

The purpose of this paper is to use Newton like methods to approximate factors of a given matrix. The factorizations presented in this work are QR and $(\mathsf{L}+\mathsf{I})\mathsf{U}$ ones, where Q is a unitary matrix, R and U upper triangular matrices, L a strictly lower triangular matrix and I denotes the identity matrix. Among Newton like methods, we consider here the exact classical version extended to a more general context in [4] and [2]. The above mentioned factorizations are important tools in numerical algorithms covering a wide spectrum of problems in applied mathematics. In particular, in the domain of eigenvalue approximations performed through the iterative computation of an upper triangular

*Corresponding author

Received December 1, 2012

similar matrix, we find QR Francis and LR Rutishauser methods. The algorithms of the present paper are used in this particular context in [1].

2. The General Framework

Let $\mathbb{C}^{n\times n}$ (resp. $\mathbb{R}^{n\times n}$) be the complex (resp. real) algebra of square matrices of order n with complex (resp. real) entries. The identity matrix will be denoted by I and the null matrix by O.

Theorem 2.1.QR *factorization*

For every nonsingular matrix $\mathsf{Z} \in \mathbb{F}^{n\times n}$ *there exists a unitary matrix* Q *and an upper triangular matrix* R *such that*

$$\mathsf{Z} = \mathsf{QR}.$$

Proof. See [3].

Theorem 2.2.$(\mathsf{L}+\mathsf{I})\mathsf{U}$ *factorization*

For every nonsingular matrix $\mathsf{Z} \in \mathbb{F}^{n\times n}$ *there exists a permutation matrix* P, *a strictly lower triangular matrix* L *and nonsingular upper triangular matrix* U *such that*

$$\mathsf{Z} = (\mathsf{L}+\mathsf{I})\mathsf{UP}.$$

Proof. See [3].

The purpose of this article is to show that these forms can be approximated using Newton-Kantorovich method.

We recall the basic aspects of Newton-Kantorovich method for nonlinear equations: Let $\mathbb{B}_1$ and $\mathbb{B}_2$ be real isomorphic finite-dimensional normed linear spaces, $\mathcal{O}$ an open set of $\mathbb{B}_1$. For $i, j \in \{1,2\}$, $\mathrm{BL}(\mathbb{B}_i, \mathbb{B}_j)$ denotes the algebra of all bounded linear operators with domain $\mathbb{B}_i$ and values in $\mathbb{B}_j$, and the open subset of $\mathrm{BL}(\mathbb{B}_i, \mathbb{B}_j)$ of isomorphisms from $\mathbb{B}_i$ onto $\mathbb{B}_j$ is denoted by $\mathrm{IS}(\mathbb{B}_i, \mathbb{B}_j)$. All the norms involved in these structures are denoted by the single symbol $\|.\|$. Let $\mathcal{F} : \mathcal{O} \to \mathbb{B}_2$ be a Fréchet differentiable operator. The problem to be solved by iterations is

$$\text{Find } \varphi_\infty \in \mathcal{O} \text{ such that } \mathcal{F}(\varphi_\infty) = 0. \tag{1}$$

Let $(B_k)_{k\geq 0}$ be a sequence in $\mathrm{IS}(\mathbb{B}_1, \mathbb{B}_2)$. The so-called Newton type iterations read as

$$\varphi_0 \in \mathcal{O}, \quad \varphi_{k+1} := \varphi_k - B_k^{-1}\mathcal{F}(\varphi_k). \tag{2}$$

The Newton-Kantorovich method corresponds to the choice

$$B_k := \mathcal{F}'(\varphi_k) \text{ for all } k \geq 0. \tag{3}$$

Let $\mathcal{O}_r(\varphi)$ denote the open ball of $\mathbb{B}_1$ centered at φ with radius $r > 0$.

Theorem 2.3. A posteriori convergence of (2) with (3)

Suppose that $\mathcal{O}$, $\mathcal{F}$, $\varphi_0 \in \mathcal{O}$, $c_0 > 0$, $\ell > 0$ and $m_0 > 0$ satisfy

(2.3.1) $\mathcal{F}'(\varphi_0) \in \mathrm{IS}(\mathbb{B}_1, \mathbb{B}_2)$, $\|\mathcal{F}'(\varphi_0)^{-1}\| \leq m_0$, and $\|\mathcal{F}'(\varphi_0)^{-1}\mathcal{F}(\varphi_0)\| \leq c_0$,

(2.3.2) $\mathcal{D}_0 := \{\varphi \in \mathbb{B}_1 : \|\varphi - \varphi_0\| \leq 2c_0\}$ is included in $\mathcal{O}$,

(2.3.3) ℓ is a Lipschitz constant for $\mathcal{F}'$ on $\mathcal{D}_0$,

(2.3.4) $h_0 := m_0 \ell c_0 < 1/2$.

Then, $\mathcal{F}$ has a unique zero $\varphi_\infty \in \mathcal{D}_0$, and for all $k \geq 0$,

$$\|\varphi_{k+1} - \varphi_\infty\| \leq \frac{m_0 \ell}{1 - 2h_0}\|\varphi_k - \varphi_\infty\|^2.$$

Proof. See [4]. Some improvements on classical error bounds for Newton's method in a more general abstract framework are given in [2].

Ker will denote the kernel (or null space) of a linear operator, and Ran its range (or image space). Let $\mathbb{F}$ denote either $\mathbb{R}$ or $\mathbb{C}$. We introduce the following linear operators:

$$\mathcal{U}_\mathbb{F} : \mathbb{F}^{n\times n} \to \mathbb{F}^{n\times n}, \quad \mathcal{U}_\mathbb{F}(\mathsf{M})(i,j) := \begin{cases} \mathsf{M}(i,j) & \text{if } i \leq j, \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathcal{L}_\mathbb{F} : \mathbb{F}^{n\times n} \to \mathbb{F}^{n\times n}, \quad \mathcal{L}_\mathbb{F}(\mathsf{M})(i,j) := \begin{cases} \mathsf{M}(i,j) & \text{if } j \leq i, \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathcal{D}_\mathbb{F} : \mathbb{F}^{n\times n} \to \mathbb{F}^{n\times n}, \quad \mathcal{D}_\mathbb{F}(\mathsf{M})(i,j) := \begin{cases} \mathsf{M}(i,j) & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases}$$

for all $\mathsf{M} \in \mathbb{F}^{n\times n}$, where $\mathcal{U}$ stands for upper, $\mathcal{L}$ for lower and $\mathcal{D}$ for diagonal. With these notations, for instance, $\mathrm{Ran}(\mathcal{U}_{\mathbb{F}})$ is the space of all upper triangular matrices with coefficients in $\mathbb{F}$ and $\mathrm{Ker}(\mathcal{U}_{\mathbb{F}})$ is the space of all strictly lower triangular matrices with entries in $\mathbb{F}$.

For topological purposes, we shall consider the following inner product:

$$\langle (\mathsf{A}_1, \mathsf{A}_2, \ldots, \mathsf{A}_m), (\widehat{\mathsf{A}}_1, \widehat{\mathsf{A}}_2, \ldots, \widehat{\mathsf{A}}_m)\rangle := \sum_{i=1}^{m} \mathrm{tr}(\widehat{\mathsf{A}}_i^* \mathsf{A}_i) \tag{4}$$

for any matrices A_i, $\widehat{\mathsf{A}}_i$ in $\mathbb{F}^{p\times q}$ and any integers $m \geq 1$, $p \geq 1$ and $q \geq 1$. The corresponding induced norm will denoted by

$$\|(\mathsf{A}_1, \mathsf{A}_2, \ldots, \mathsf{A}_m)\| := \left[\sum_{i=1}^{m} \mathrm{tr}(\mathsf{A}_i^* \mathsf{A}_i)\right]^{1/2}.$$

We suppose that the matrix Z is invertible, with no zero diagonal entry, and such that $\mathsf{P} = \mathsf{I}$ in Theorem 2.2.

3. The QR factorization

3.1. Defining the nonlinear operator $\mathcal{F}$

Let A, B in $\mathbb{R}^{n\times n}$ be the real part and the imaginary part of Z respectively:

$$\mathsf{A} := \Re \mathsf{Z}, \quad \mathsf{B} := \Im \mathsf{Z}.$$

We use indifferently the notations

$$\mathsf{Z} = \mathsf{A} + \mathrm{i}\,\mathsf{B} \in \mathbb{C}^{n\times n} \quad \text{or} \quad \mathsf{Z} = (\mathsf{A}, \mathsf{B}) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}.$$

Following Theorem 2.1, there exist U_∞, V_∞ in $\mathbb{R}^{n\times n}$ and X_∞, Y_∞ in $\mathrm{Ran}(\mathcal{U}_{\mathbb{R}})$, such that

$$\mathsf{Q}_\infty := \mathsf{U}_\infty + \mathrm{i}\,\mathsf{V}_\infty, \quad \mathsf{R}_\infty := \mathsf{X}_\infty + \mathrm{i}\,\mathsf{Y}_\infty$$

satisfy

$$\mathsf{Q}_\infty \mathsf{Q}_\infty^* - \mathsf{I} \;=\; \mathsf{O}, \tag{5}$$

and

$$\mathsf{Q}_\infty \mathsf{R}_\infty - \mathsf{Z} \;=\; \mathsf{O}, \tag{6}$$

Note that if Q is such that $\mathsf{QR} = \mathsf{Z}$, and $\mathsf{QQ}^* = \mathsf{I}$, and if the diagonal entries of Q are not real then there exists a unitary diagonal matrix D such that $\mathsf{Q}_\infty := \mathsf{QD}$ has a real diagonal entries and $\mathsf{R}_\infty := \mathsf{D}^*\mathsf{R}$ is still upper triangular.

In order to help the reader to fix and clarify notations, we refer to the following table:

(7)

Role	Symbol
Unitary matrix	$\mathsf{Q} = \mathsf{U} + \mathrm{i}\,\mathsf{V}$
Increment of a unitary matrix	$\mathsf{E} = \mathsf{H} + \mathrm{i}\,\mathsf{K}$
Upper triangular matrix	$\mathsf{R} = \mathsf{X} + \mathrm{i}\,\mathsf{Y}$
Increment of an upper triangular matrix	$\mathsf{F} = \mathsf{R} + \mathrm{i}\,\mathsf{S}$

Some of these symbols may carry subscripts or upperscripts like in R_∞, X_0, Q_k, $\widehat{\mathsf{V}}$ or $\widetilde{\mathsf{H}}$.

We consider the spaces

$$\mathbb{B}_1 := \mathbb{R}^{n\times n}\times \mathrm{Ker}(\mathcal{D}_\mathbb{R})\times \mathrm{Ran}(\mathcal{U}_\mathbb{R})\times \mathrm{Ran}(\mathcal{U}_\mathbb{R}), \quad \mathbb{B}_2 := \mathbb{R}^{n\times n}\times \mathbb{R}^{n\times n}\times \mathbb{R}^{n\times n},$$

Equations (5) and (6) are equivalent to the following system:

$$\mathsf{U}_\infty \mathsf{X}_\infty - \mathsf{V}_\infty \mathsf{Y}_\infty - \mathsf{A} \;=\; \mathsf{O}, \tag{8}$$

$$\mathsf{U}_\infty \mathsf{Y}_\infty + \mathsf{V}_\infty \mathsf{X}_\infty - \mathsf{B} \;=\; \mathsf{O}, \tag{9}$$

$$\mathcal{U}_\mathbb{R}(\mathsf{U}_\infty \mathsf{U}_\infty^\top + \mathsf{V}_\infty \mathsf{V}_\infty^\top - \mathsf{I}) \;=\; \mathsf{O}, \tag{10}$$

$$\mathcal{U}_\mathbb{R}(\mathsf{V}_\infty \mathsf{U}_\infty^\top - \mathsf{U}_\infty \mathsf{V}_\infty^\top) \;=\; \mathsf{O}, \tag{11}$$

$$\mathcal{D}_\mathbb{R}(\mathsf{V}_\infty) \;=\; \mathsf{O}. \tag{12}$$

Equations (8) and (9) hold in $\mathbb{R}^{n\times n}$ and equations (10) and (11) hold in $\mathrm{Ran}(\mathcal{U}_\mathbb{R})$. We remark that, for all M, N in $\mathbb{R}^{n\times n}$,

$$\mathcal{U}_\mathbb{R}(\mathsf{MN}^\top - \mathsf{NM}^\top) \in \mathrm{Ker}(\mathcal{D}_\mathbb{R})$$

since $\mathcal{D}_\mathbb{R}(\mathsf{MN}^\top - \mathsf{NM}^\top) = \mathsf{O}$.

Let $\mathcal{F} : \mathbb{B}_1 \to \mathbb{B}_2$ be the nonlinear operator defined by

$$(13)\mathcal{F}[\mathsf{U},\mathsf{V},\mathsf{X},\mathsf{Y}] := [\mathsf{UX}-\mathsf{VY}-\mathsf{A}, \mathsf{UY}+\mathsf{VX}-\mathsf{B}, \mathcal{U}_{\mathbb{R}}(\mathsf{UU}^\top+\mathsf{VV}^\top-\mathsf{I})+\mathcal{L}_{\mathbb{R}}(\mathsf{VU}^\top-\mathsf{UV}^\top)].$$

The problem of finding a QR factorization of Z reduces to

$$(14)\qquad \text{Find } [\mathsf{U}_\infty,\mathsf{V}_\infty,\mathsf{X}_\infty,\mathsf{Y}_\infty] \in \mathbb{B}_1 \text{ such that } \mathcal{F}(\mathsf{U}_\infty,\mathsf{V}_\infty,\mathsf{X}_\infty,\mathsf{Y}_\infty) = [\mathsf{O},\mathsf{O},\mathsf{O}].$$

3.2. A Lipschitz Constant and the Invertibility of the Fréchet Derivative

The Fréchet derivative of $\mathcal{F}$ at $[\mathsf{Q},\mathsf{R}]$ is given by

$$\mathcal{F}'(\mathsf{Q},\mathsf{R})(\mathsf{E},\mathsf{F}) = \big[\mathsf{E}(\mathsf{R}+\mathsf{QF}), \mathcal{U}_{\mathbb{R}}\big(\mathsf{HU}^\top+\mathsf{UH}^\top+\mathsf{KV}^\top+\mathsf{VK}^\top\big)+\mathcal{L}_{\mathbb{R}}\big(\mathsf{KU}^\top+\mathsf{VH}^\top-\mathsf{HV}^\top-\mathsf{UK}^\top\big)\big]$$

Hence, for $[\mathsf{Q},\mathsf{R}]$, $[\widehat{\mathsf{Q}},\widehat{\mathsf{R}}]$, $[\mathsf{E},\mathsf{F}] \in \mathbb{B}_1$,

$$\begin{aligned}(\mathcal{F}'(\mathsf{Q},\mathsf{R})-\mathcal{F}'(\widehat{\mathsf{Q}},\widehat{\mathsf{R}}))(\mathsf{E},\mathsf{F}) = &\big[\mathsf{E}(\mathsf{R}-\widehat{\mathsf{R}})+(\mathsf{Q}-\widehat{\mathsf{Q}})\mathsf{F},\\ &\mathcal{U}_{\mathbb{R}}\big(\mathsf{H}(\mathsf{U}-\widehat{\mathsf{U}})^\top+(\mathsf{U}-\widehat{\mathsf{U}})\mathsf{H}^\top+\mathsf{K}(\mathsf{V}-\widehat{\mathsf{V}})^\top+(\mathsf{V}-\widehat{\mathsf{V}})\mathsf{K}^\top\big)+\\ &\mathcal{L}_{\mathbb{R}}\big(\mathsf{K}(\mathsf{U}-\widehat{\mathsf{U}})^\top+(\mathsf{V}-\widehat{\mathsf{V}})\mathsf{H}^\top-\mathsf{H}(\mathsf{V}-\widehat{\mathsf{V}})^\top-(\mathsf{U}-\widehat{\mathsf{U}})\mathsf{K}^\top\big)\big],\end{aligned}$$

and

$$\begin{aligned}\|(\mathcal{F}'(\mathsf{Q},\mathsf{R})-\mathcal{F}'(\widehat{\mathsf{Q}},\widehat{\mathsf{R}}))(\mathsf{E},\mathsf{F})\| &\le \sqrt{(\|\mathsf{E}\|\|\mathsf{R}-\widehat{\mathsf{R}}\|+\|\mathsf{F}\|\|\mathsf{Q}-\widehat{\mathsf{Q}}\|)^2+4((\|\mathsf{H}\|,\|\mathsf{K}\|).(\|\mathsf{U}-\widehat{\mathsf{U}}\|,\|\mathsf{V}-\widehat{\mathsf{V}}\|))^2}\\ &\le \sqrt{\|(\mathsf{E},\mathsf{F})\|^2\|(\mathsf{Q}-\widehat{\mathsf{Q}},\mathsf{R}-\widehat{\mathsf{R}})\|^2+4\|(\mathsf{H},\mathsf{K})\|^2\|\mathsf{Q}-\widehat{\mathsf{Q}}\|^2}\\ &\le \sqrt{5}\|(\mathsf{E},\mathsf{F})\|^2\|(\mathsf{Q}-\widehat{\mathsf{Q}},\mathsf{R}-\widehat{\mathsf{R}})\|.\end{aligned}$$

Thus we may set

$$(15)\qquad \ell := \sqrt{5}.$$

To determine a sufficient condition for the Fréchet derivative $\mathcal{F}'(\mathsf{U},\mathsf{V},\mathsf{X},\mathsf{Y})$ to be nonsingular, we study the kernel of $\mathcal{F}'(\mathsf{U},\mathsf{V},\mathsf{X},\mathsf{Y})$, where $[\mathsf{U},\mathsf{V},\mathsf{X},\mathsf{Y}]$ may be either φ_0 or φ_∞. The equation

$$(16)\qquad \mathcal{F}'(\mathsf{U},\mathsf{V},\mathsf{X},\mathsf{Y})[\mathsf{H},\mathsf{K},\mathsf{S},\mathsf{T}] = [\mathsf{O},\mathsf{O},\mathsf{O}]$$

translates into the following system:

$$\mathsf{HX} + \mathsf{US} - \mathsf{KY} - \mathsf{VR} = \mathsf{O}, \tag{17}$$

$$\mathsf{HY} + \mathsf{UR} + \mathsf{KX} + \mathsf{VS} = \mathsf{O}, \tag{18}$$

$$\mathsf{HU}^\top + \mathsf{UH}^\top + \mathsf{KV}^\top + \mathsf{VK}^\top = \mathsf{O}, \tag{19}$$

$$\mathsf{KU}^\top + \mathsf{VH}^\top - \mathsf{HV}^\top - \mathsf{UK}^\top = \mathsf{O}, \tag{20}$$

$$\mathcal{D}_{\mathbb{R}}(\mathsf{K}) = \mathsf{O}. \tag{21}$$

Theorem 3.1. *If Z and $\mathcal{D}_{\mathbb{R}}(\mathsf{U}_\infty)$ are invertible, then $\mathcal{F}'(\varphi_\infty)$ is invertible.*

Proof. Multiplying (18) by i and adding the result to (17) we get with the notations of table (7),

$$\mathsf{F} + \mathsf{GR}_\infty = \mathsf{O}. \tag{22}$$

Define

$$\mathsf{G} := \mathsf{Q}_\infty^* \mathsf{E}.$$

Multiplying (20) by i and adding the result to (19) we get

$$\mathsf{G}^* = -\mathsf{G}, \tag{23}$$

because F and R_∞ are upper triangular. The elements of the strict lower triangular part of G satisfy:

For $j < i \in [\![2, n]\!]$,

$$\sum_{k=1}^{j} \mathsf{G}(i,k)\mathsf{R}_\infty(k,j) = 0,$$

and since the diagonal entries of R_∞ are nonzero, $\mathsf{G}(i,j) = 0$ for $j < i \in [\![2, n]\!]$. Since $\mathsf{G}^* = -\mathsf{G}$, $\mathsf{G}(j,i) = 0 = \mathsf{G}(i,j)$ for $j < i \in [\![2, n]\!]$, and $\Re\mathsf{G}(l,l) = 0$ for $l \in [\![1, n]\!]$.

Hence $\mathsf{G} = \mathrm{i}\,\mathsf{D}$, where $\mathsf{D} \in \mathrm{Ran}(\mathcal{D}_{\mathbb{R}})$. So

$$\mathsf{E} = \mathrm{i}\,\mathsf{Q}_\infty \mathsf{D}, \quad \mathsf{H} = -\mathsf{V}_\infty \mathsf{D}, \quad \mathsf{K} = \mathsf{U}_\infty \mathsf{D}.$$

Since $\mathcal{D}_{\mathbb{R}}(\mathsf{K}) = \mathsf{O}$, $\mathsf{U}_\infty(j,j)\mathsf{D}(j,j) = 0$ for all $j \in [\![1, n]\!]$, and since the diagonal entries of U_∞ are nonzero, $\mathsf{D} = \mathsf{O}$. But $\mathsf{D} = \mathsf{O}$ implies $\mathsf{E} = \mathsf{G} = \mathsf{F} = \mathsf{O}$. This proves that $\mathcal{F}'(\varphi_\infty)$ is invertible. This completes the proof.

Remark 3.2. The proof of theorem 3.1 shows that the condition "for all $i \in [\![1,n]\!]$, $\mathsf{U}_\infty(i,i) \neq 0$ and $\mathsf{V}_\infty(i,i) = 0$" can be relaxed to "for all $i \in [\![1,n]\!]$, $\mathsf{Q}_\infty(i,i) \neq 0$" i.e. "for all $i \in [\![1,n]\!]$, $\mathsf{Q}_\infty(i,i) \neq 0$ if and only if $\mathsf{V}_\infty(i,i) = 0$."

3.3. Finding Constants m_0 and c_0

Suppose that $\mathsf{Q}_0 = \mathsf{I}$, and $\mathsf{R}_0 = \mathcal{U}_{\mathbb{F}}(\mathsf{Z})$ are the initial points, so for given matrices $\mathsf{N} \in \mathbb{C}^{n\times n}$ and $\mathsf{J} \in \mathbb{R}^{n\times n}$, we are led to solve

$$\mathsf{F} + \mathsf{E}\mathsf{R}_0 = \mathsf{N}, \tag{24}$$

$$\mathsf{E} + \mathsf{E}^* = \mathsf{M}, \tag{25}$$

$$\mathcal{D}_{\mathbb{R}}(\mathsf{K}) = \mathsf{O}. \tag{26}$$

where $\mathsf{M} := [\mathcal{U}_{\mathbb{R}}(\mathsf{J}) + \mathcal{L}_{\mathbb{R}}(\mathsf{J}^\top) - \mathcal{D}_{\mathbb{R}}(\mathsf{J})] + \mathrm{i}\,[\mathcal{L}_{\mathbb{R}}(\mathsf{J}) - \mathcal{U}_{\mathbb{R}}(\mathsf{J}^\top)]$. Remark that $\|\mathsf{M}\| \leq \sqrt{2}\|\mathsf{J}\|$. Because F and R_0 are upper triangular, equation (24) is equivalent to

$$\sum_{k=1}^{j} \mathsf{E}(i,k)\mathsf{R}_0(k,j) = \mathsf{N}(i,j), \tag{27}$$

where $j < i \in [\![2,n]\!]$. This implies that, for $j \in [\![1,n-1]\!]$,

$$\mathsf{E}(i,j) = \frac{1}{\mathsf{R}_0(j,j)}\left(\mathsf{N}(i,j) - \sum_{k=1}^{j-1} \mathsf{E}(i,k)\mathsf{R}_0(k,j)\right) \tag{28}$$

So for $j \in [\![1,n-1]\!]$,

$$\|\mathsf{E}_{j+1,1}(:,j)\|_2^2 \leq (n-j)\mathsf{W}(j)^2\|\mathsf{N}\|,$$

where

$$\mathsf{W}(j) := \frac{1}{|\mathsf{R}_0(j,j)|} + \sum_{k=1}^{j-1} \mathsf{W}(k)|\mathsf{R}_0(k,j)|,$$

and

$$\mathsf{E}_{p,q} := \begin{bmatrix} \mathsf{E}(p,q) & \mathsf{E}(p,q+1) & \cdots & \mathsf{E}(p,n) \\ \mathsf{E}(p+1,q) & \mathsf{E}(p+1,q+1) & \cdots & \mathsf{E}(p+1,n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{E}(n,q) & \mathsf{E}(n,q+1) & \cdots & \mathsf{E}(n,n) \end{bmatrix} \in \mathbb{C}^{(n-p)\times(n-q)},$$

$$\mathsf{N}_j := \begin{bmatrix} \mathsf{N}(j+1,j) \\ \mathsf{N}(j+2,j) \\ \vdots \\ \mathsf{N}(n,j) \end{bmatrix}, \quad \mathsf{M}_j := \begin{bmatrix} \mathsf{M}(j+1,j) \\ \mathsf{M}(j+2,j) \\ \vdots \\ \mathsf{M}(n,j) \end{bmatrix} \in \mathbb{C}^{(n-j)\times 1},$$

and equations (25), (26) imply

$$\mathsf{E}_{1,j+1}(j,:) = \overline{\mathsf{M}_j} - \overline{\mathsf{E}_{j+1,1}(:,j)}, \tag{29}$$

and for all $i \in [\![1,n]\!]$,

$$\mathsf{E}(i,i) = \frac{1}{2}\mathsf{M}(i,i).$$

So

$$\sum_{i=1}^{n} |\mathsf{E}(i,i)|^2 \le \frac{1}{4}\|\mathsf{M}\|^2 \le \frac{1}{2}\|J\|^2.$$

Also for all $j \in [\![1,n-1]\!]$,

$$\begin{aligned} \|\mathsf{E}_{1,j+1}(j,:)\|_2 &= \|\mathsf{M}_j\|_2 + \|\mathsf{E}_{j+1,1}(:,j)\|_2 \\ &\le \left(\sqrt{2} + \sqrt{(n-j)}\mathsf{W}(j)\right)\|(\mathsf{N},\mathsf{J})\| \end{aligned}$$

Now

$$\|\mathsf{E}\|^2 = \sum_{j=1}^{n-1}(\|\mathsf{E}_{j+1,1}(:,j)\|_2^2 + \|\mathsf{E}_{1,j+1}(j,:)\|_2^2) + \sum_{i=1}^{n}|\mathsf{E}(i,i)|^2,$$

So

$$\|\mathsf{E}\| \le \nu\|(\mathsf{N},\mathsf{J})\|,$$

where

$$\nu^2 := \frac{1}{2} + \sum_{j=1}^{n-1}\left[(n-j)\mathsf{W}(j)^2 + \left(\sqrt{2} + \sqrt{(n-j)}\mathsf{W}(j)\right)^2\right].$$

From equation (24),

$$\|\mathsf{F}\| \leq \|\mathsf{N}\| + \|\mathsf{R}_0\|\|\mathsf{E}\| \leq (1+\|\mathsf{R}_0\|\nu)\|(\mathsf{N},\mathsf{J})\|.$$

Thus

$$\|\mathcal{F}'(\varphi_0)^{-1}(\mathsf{N},\mathsf{J})\| \leq \sqrt{\|\mathsf{E}\|^2+\|\mathsf{F}\|^2} \leq \sqrt{\nu^2+(1+\|\mathsf{R}_0\|\nu)^2}\|(\mathsf{N},\mathsf{J})\|.$$

We can set

$$m_0 := \sqrt{\nu^2+(1+\|\mathsf{R}_0\|\nu)^2}.$$

To produce c_0 we just estimate

$$\|\mathcal{F}'(\varphi_0)^{-1}\mathcal{F}(\varphi_0)\| \leq m_0\|\mathcal{F}(\varphi_0)\| \leq m_0\|\mathsf{Z}-\mathsf{R}_0\| =: c_0.$$

Following Theorem 2.3,

$$\|\mathsf{Z}-\mathcal{U}_{\mathbb{F}}(\mathsf{Z})\| < \frac{1}{2\sqrt{5}m_0^2}$$

is a sufficient condition for convergence.

The above computations can be simplied if Z is a quasi-diagonal matrix and if we take $\mathsf{R}_0 := \mathcal{D}_{\mathbb{F}}(\mathsf{Z})$.

3.4. Performing Iterations

In order to simplify notations we will write

$$\begin{aligned} \varphi_k &:= (\widetilde{\mathsf{Q}},\widetilde{\mathsf{R}}), \text{ the current iterate,} \\ \varphi_{k+1} &:= (\mathsf{Q},\mathsf{R}), \text{ the next iterate,} \end{aligned}$$

This means that the equations of the following subsection is to be solved for (Q,R).

3.4.1. Newton-Kantorovich

The method defined by (2) and (3) amounts to solve for (Q, R) at each step k,

$$\begin{aligned} \mathsf{R} + \mathsf{G}\widetilde{\mathsf{R}} &= \widetilde{\mathsf{R}} + \widetilde{\mathsf{Q}}^{-1}\mathsf{Z}, \\ \mathsf{G} + \mathsf{G}^* &= \widetilde{\mathsf{Q}}^{-1}\widetilde{\mathsf{Q}}^{-*} + \mathsf{I}, \\ \mathcal{D}_{\mathbb{R}}(\Im\widetilde{\mathsf{Q}}\mathsf{G}) &= \mathsf{O}, \end{aligned} \tag{30}$$

where $\mathsf{G} := \widetilde{\mathsf{Q}}^{-1}\mathsf{Q}$.

3.5. Numerical Experiments

The following examples have been done with Matlab 6.5. For a given matrix W we introduce the mesure of *Departure from Unity*:

$$\mathrm{DU}(\mathsf{W}) := \frac{\|\mathsf{W}^*\mathsf{W} - \mathsf{I}\|}{\|\mathsf{W}\|^2}$$

and for a couple of matrices (W, Λ) the mesure of *Relative Residual*:

$$\mathrm{RelRes}(\mathsf{W}, \Lambda) := \frac{\|\mathsf{W}\Lambda - \mathsf{Z}\|}{\|\mathsf{W}\|\|\Lambda\|}.$$

Example 3.3.

Data:

$$\mathsf{Z} := \begin{bmatrix} 2.0 & 1.0 & 0.0 & 0.0 \\ 0.5 & 2.0 & -0.5 & 0.0 \\ 0.0 & 1.0 & 2.0 & 0.0 \\ -0.5 & 0.0 & 0.5 & 2.0 \end{bmatrix}.$$

Starting point:

$$\mathsf{Q}_0 := \mathsf{I}, \quad \mathsf{R}_0 := \mathcal{U}_{\mathbb{F}}(\mathsf{Z}).$$

Convergence table:

Example 3.3	Newton-Kantorovich	
Iteration	$\mathrm{DU}(\mathsf{Q}_k)$	$\mathrm{RelRes}(\mathsf{Q}_k, \mathsf{R}_k)$
0	$0.00E+00$	$0.16E-00$
1	$0.13E-00$	$0.88E-01$
2	$0.19E-01$	$0.10E-01$
3	$0.42E-04$	$0.33E-03$
4	$0.23E-06$	$0.25E-06$
5	$0.14E-12$	$0.16E-12$
6	$0.50E-17$	$0.92E-17$

Example 3.4.

Data:

$$\mathsf{Z} := \begin{bmatrix} 1.0 & 0.5 & 0.0 \\ 0.5 & 2.0 & 0.5 \\ 0.0 & 0.5 & 3.0 \end{bmatrix}.$$

Starting point:

$$\mathsf{Q}_0 := \mathsf{I}, \quad \mathsf{R}_0 := \mathcal{D}_{\mathbb{F}}(\mathsf{Z}).$$

Convergence table:

Example 3.4	Newton-Kantorovich	
Iteration	$\mathrm{DU}(\mathsf{Q}_k)$	$\mathrm{RelRes}(\mathsf{Q}_k, \mathsf{R}_k)$
0	$0.00E+00$	$0.15E-00$
1	$0.12E-00$	$0.13E-00$
2	$0.13E-01$	$0.11E-01$
3	$0.18E-03$	$0.11E-03$
4	$0.37E-07$	$0.25E-07$
5	$0.13E-14$	$0.54E-15$

Example 3.5.

Data:

$$\mathsf{Z} := \text{gallery('prolate',5)} := \begin{bmatrix} 0.5000 & 0.3183 & 0.0000 & -0.1061 & -0.0000 \\ 0.3183 & 0.5000 & 0.3183 & 0.0000 & -0.1061 \\ 0.0000 & 0.3183 & 0.5000 & 0.3183 & 0.0000 \\ -0.1061 & 0.0000 & 0.3183 & 0.5000 & 0.3183 \\ -0.0000 & -0.1061 & 0.0000 & 0.3183 & 0.5000 \end{bmatrix}.$$

Starting point:

$$\mathsf{Q}_0 := \mathsf{I}, \quad \mathsf{R}_0 := \mathsf{I}.$$

Convergence table:

Example 3.5	Newton-Kantorovich	
Iteration	$\mathrm{DU}(\mathsf{Q}_k)$	$\mathrm{RelRes}(\mathsf{Q}_k, \mathsf{R}_k)$
0	$0.00E+00$	$0.29E-00$
1	$0.93E-01$	$0.19E-00$
2	$0.16E-00$	$0.18E-00$
3	$0.32E-00$	$0.29E-00$
4	$0.17E-00$	$0.11E-00$
5	$0.61E-00$	$0.32E-00$
6	$0.40E-00$	$0.15E-00$
7	$0.15E-00$	$0.38E-01$
8	$0.25E-01$	$0.77E-02$
9	$0.89E-03$	$0.27E-03$
10	$0.62E-05$	$0.20E-05$
11	$0.19E-09$	$0.48E-10$
12	$0.13E-15$	$0.10E-15$

4. The $(\mathsf{L}+\mathsf{I})\mathsf{U}$ factorization

4.1. Defining the nonlinear operator $\mathcal{F}$

Following Theorem 2.2, there exist L_∞ in $\mathrm{Ker}(\mathcal{U}_\mathbb{C})$ and U_∞ in $\mathrm{Ran}(\mathcal{U}_\mathbb{C})$, such that

$$(\mathsf{L}_\infty + \mathsf{I})\mathsf{U}_\infty - \mathsf{Z} \;\; = \;\; \mathsf{O}, \tag{31}$$

Note that if L is such that $\mathsf{LU} = \mathsf{Z}$ and if the diagonal entries of L are not one then there exists an inversible diagonal matrix D such that $\mathsf{L}_\infty := \mathsf{LD}$ has unit diagonal entries and $\mathsf{U}_\infty := \mathsf{D}^{-1}\mathsf{U}$ is still upper triangular.

In order to help the reader to fix and clarify notations, we refer to the following table:

(32)

Role	Symbol
Strictly lower triangular matrix	L
Increment of a strictly lower traingular matrix	E
Upper triangular matrix	U
Increment of an upper triangular matrix	F

Some of these symbols may carry subscripts or upperscripts like in U_∞, E_0, L_k, $\widehat{\mathsf{U}}$ or $\widetilde{\mathsf{F}}$.

We consider the spaces

$$\mathbb{B}_1 := \mathrm{Ker}(\mathcal{U}_\mathbb{C}) \times \mathrm{Ran}(\mathcal{U}_\mathbb{C}), \quad \mathbb{B}_2 := \mathbb{C}^{n\times n},$$

Equation (31) is equivalent to:

$$\mathsf{L}_\infty \mathsf{U}_\infty + \mathsf{U}_\infty - \mathsf{Z}_\infty \;\; = \;\; \mathsf{O}. \tag{33}$$

Let $\mathcal{F} : \mathbb{B}_1 \to \mathbb{B}_2$ be the nonlinear operator defined by

$$\mathcal{F}[\mathsf{L}, \mathsf{U}] \;\; := \;\; \mathsf{LU} + \mathsf{U} - \mathsf{Z}. \tag{34}$$

The problem of finding a $(\mathsf{L} + \mathsf{I})\mathsf{U}$ factorization of Z reduces to

$$\text{Find } [\mathsf{L}_\infty, \mathsf{U}_\infty] \in \mathbb{B}_1 \text{ such that } \mathcal{F}(\mathsf{L}_\infty, \mathsf{U}_\infty) = \mathsf{O}. \tag{35}$$

4.2. A Lipschitz Constant and the Invertibility of the Fréchet Derivative

The Fréchet derivative of $\mathcal{F}$ at $[\mathsf{L},\mathsf{U}]$ is given by

$$\mathcal{F}'(\mathsf{L},\mathsf{U})(\mathsf{E},\mathsf{F}) = \mathsf{EU} + \mathsf{LF}.$$

Hence, for $[\mathsf{L},\mathsf{U}]$, $[\widehat{\mathsf{L}},\widehat{\mathsf{U}}]$, $[\mathsf{E},\mathsf{F}] \in \mathbb{B}_1$,

$$(\mathcal{F}'(\mathsf{L},\mathsf{U}) - \mathcal{F}'(\widehat{\mathsf{L}},\widehat{\mathsf{U}}))(\mathsf{E},\mathsf{F}) = \mathsf{E}(\mathsf{U}-\widehat{\mathsf{U}}) + (\mathsf{L}-\widehat{\mathsf{L}})\mathsf{F}$$

and

$$\|(\mathcal{F}'(\mathsf{L},\mathsf{U}) - \mathcal{F}'(\widehat{\mathsf{L}},\widehat{\mathsf{U}})(\mathsf{E},\mathsf{F})\| \le \|\mathsf{E}\|\|\mathsf{U}-\widehat{\mathsf{U}}\| + \|\mathsf{F}\|\|\mathsf{L}-\widehat{\mathsf{L}}\| \le \|(\mathsf{E},\mathsf{F})\|\|(\mathsf{L}-\widehat{\mathsf{L}},\mathsf{U}-\widehat{\mathsf{U}})\|.$$

So we may set

$$\ell := 1. \tag{36}$$

To determine a sufficient condition for the Fréchet derivative $\mathcal{F}'(\mathsf{L},\mathsf{U})$ to be nonsingular, we study the kernel of $\mathcal{F}'(\mathsf{L},\mathsf{U})$, where $[\mathsf{L},\mathsf{U}]$ may be either φ_0 or φ_∞. The equation

$$\mathcal{F}'(\mathsf{L},\mathsf{U})[\mathsf{E},\mathsf{F}] = \mathsf{O} \tag{37}$$

translates into the following equation:

$$\mathsf{EU} + \mathsf{LF} + \mathsf{F} \;\; = \;\; \mathsf{O}. \tag{38}$$

Theorem 4.1. *If Z is invertible, then $\mathcal{F}'(\varphi_\infty)$ is invertible.*

Proof. From (38) we get

$$\mathsf{GU}_\infty + \mathsf{F} \;\; = \;\; \mathsf{O}, \tag{39}$$

where

$$\mathsf{G} := (\mathsf{L}_\infty + \mathsf{I})^{-1}\mathsf{E}.$$

It is clear G is a strictly lower triangular matrix as E is. And, since F and U_∞ are upper triangular, $\mathsf{G} = \mathsf{0}$. Consequently, $\mathsf{E} = \mathsf{0}$ and $\mathsf{F} = \mathsf{0}$. This completes the proof.

4.3. Finding Constants m_0 and c_0

Suppose that $\mathsf{L}_0 = \mathsf{O}$, and $\mathsf{U}_0 = \mathcal{U}_{\mathbb{F}}(\mathsf{Z})$ are the initial points. For given matrices $\mathsf{N} \in \mathbb{C}^{n\times n}$ we are led to solve

$$\mathsf{F} + \mathsf{E}\mathsf{U}_0 \;=\; \mathsf{N}. \tag{40}$$

Because F and U_0 are upper triangular matrices, equation (40) is equivalent to

$$\sum_{k=1}^{j} \mathsf{E}(i,k)\mathsf{U}_0(k,j) = \mathsf{N}(i,j), \tag{41}$$

where $j < i \in [\![2,n]\!]$. This implies that, for $j \in [\![1,n-1]\!]$,

$$\mathsf{E}(i,j) = \frac{1}{\mathsf{U}_0(j,j)}\left(\mathsf{N}(i,j) - \sum_{k=1}^{j-1} \mathsf{E}(i,k)\mathsf{U}_0(k,j)\right). \tag{42}$$

So, for $j \in [\![1,n-1]\!]$,

$$\|\mathsf{E}_{j+1,1}(:,j)\|_2^2 \le (n-j)\mathsf{W}(j)^2\|\mathsf{N}\|,$$

where

$$\mathsf{W}(j) := \frac{1}{|\mathsf{U}_0(j,j)|} + \sum_{k=1}^{j-1}\mathsf{W}(k)|\mathsf{U}_0(k,j)|.$$

Now

$$\|\mathsf{E}\|^2 = \sum_{j=1}^{n-1}(\|\mathsf{E}_{j+1,1}(:,j)\|_2^2.$$

So

$$\|\mathsf{E}\| \le \nu\|\mathsf{N}\|,$$

where

$$\nu^2 := \sum_{j=1}^{n-1}(n-j)\mathsf{W}(j)^2.$$

From equation (40),

$$\|\mathsf{F}\| \le \|\mathsf{N}\| + \|\mathsf{U}_0\|\|\mathsf{E}\| \le (1+\|\mathsf{U}_0\|\nu)\|\mathsf{N}\|.$$

Thus

$$\|\mathcal{F}'(\varphi_0)^{-1}(\mathsf{N})\| \le \sqrt{\|\mathsf{E}\|^2+\|\mathsf{F}\|^2} \le \sqrt{\nu^2+(1+\|\mathsf{U}_0\|\nu)^2}\|\mathsf{N}\|.$$

We can set

$$m_0 := \sqrt{\nu^2+(1+\|\mathsf{U}_0\|\nu)^2}.$$

To produce c_0 we just estimate

$$\|\mathcal{F}'(\varphi_0)^{-1}\mathcal{F}(\varphi_0)\| \leq m_0\|\mathcal{F}(\varphi_0)\| \leq m_0\|\mathsf{Z} - \mathsf{U}_0\| =: c_0.$$

Following Theorem 2.3,

$$\|\mathsf{Z} - \mathcal{U}_{\mathbb{F}}(\mathsf{Z})\| < \frac{1}{2\sqrt{5}m_0^2}$$

is a sufficient condition for convergence.

As before, some simplifications are possible if Z is a quasi-diagonal matrix and if we take $\mathsf{U}_0 := \mathcal{D}_{\mathbb{F}}(\mathsf{Z})$.

5. Performing Iterations

In order to simplify notations we will write

$$\begin{aligned} \varphi_k &:= (\widetilde{\mathsf{L}}, \widetilde{\mathsf{U}}), \text{ the current iterate,} \\ \varphi_{k+1} &:= (\mathsf{L}, \mathsf{U}), \text{ the next iterate,} \end{aligned}$$

This means that the equations of the following subsection is to be solved for (L, U).

5.1. Newton-Kantorovich

The method defined by (2) and (3) amounts to solve for (L, U) at each step k,

$$\mathsf{G}\widetilde{\mathsf{U}} + \mathsf{U} = (\widetilde{\mathsf{L}} + \mathsf{I})^{-1}(\widetilde{\mathsf{L}}\widetilde{\mathsf{U}} + \mathsf{Z}), \tag{43}$$

where $\mathsf{G} := (\widetilde{\mathsf{L}} + \mathsf{I})^{-1}\mathsf{L}$.

5.2. Numerical Experiments

The following examples have been done with Matlab 6.5. For a couple of matrices (W, Λ) we introduce the mesure of *Relative Residual*:

$$\text{RelRes}(\mathsf{W}, \Lambda) := \frac{\|\mathsf{W}\Lambda - \mathsf{Z}\|}{\|\mathsf{W}\|\|\Lambda\|}.$$

Example 5.1.

Data:

$$\mathsf{Z} := \text{gallery('frank',6)} := \begin{bmatrix} 6 & 5 & 4 & 3 & 2 & 1 \\ 5 & 5 & 4 & 3 & 2 & 1 \\ & 4 & 4 & 3 & 2 & 1 \\ & & 3 & 3 & 2 & 1 \\ & & & 2 & 2 & 1 \\ & & & & 1 & 1 \end{bmatrix}.$$

Starting point:

$$\mathsf{L}_0 := 0, \quad \mathsf{U}_0 := \mathcal{U}_{\mathbb{F}}(\mathsf{Z}).$$

Convergence table:

Example 5.1	Newton-Kantorovich
Iteration	$\text{RelRes}(\mathsf{L_k}+\mathsf{I},\mathsf{U}_k)$
0	$0.42E+01$
1	$0.21E+01$
2	$0.78E-14$

Example 5.2.

Data:

$$\mathsf{Z} := \text{gallery('smok',4)} := \begin{bmatrix} 0.00+1.00i & 1.00 & & \\ 0 & -1.00+0.00i & 1.00 & \\ 0 & 0 & -0.00-1.00i & 1.00 \\ 1.00 & 0 & 0 & 1.00 \end{bmatrix}.$$

Starting point:

$$\mathsf{L}_0 := 0, \quad \mathsf{U}_0 := \mathcal{D}_{\mathbb{F}}(\mathsf{Z}).$$

Convergence table:

Example 5.2	Newton-Kantorovich
Iteration	$\text{RelRes}(\mathsf{L_k}+\mathsf{I},\mathsf{U}_k)$
0	$0.20E+01$
1	$0.12E+01$
2	$0.86E-31$
3	$0.10E+01$

Example 5.3.

Data:

$$\mathsf{Z} := \text{gallery('moler',5)} := \begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 2 & 0 & 0 & 0 \\ -1 & 0 & 3 & 1 & 1 \\ -1 & 0 & 1 & 4 & 2 \\ -1 & 0 & 1 & 2 & 5 \end{bmatrix}.$$

Starting point:

$$\mathsf{L}_0 := \mathsf{O}, \quad \mathsf{U}_0 := \mathsf{I}.$$

Convergence table:

Example 5.3	Newton-Kantorovich
Iteration	$\text{RelRes}(\mathsf{L_k}+\mathsf{I},\mathsf{U}_k)$
0	$0.71E+01$
1	$0.24E+02$
2	$0.75E+01$
3	$0.31E+01$
4	$0.13E+01$
5	$0.29E-14$

6. Complexity and Final Comments

In terms of flops (elementary operations are addition and multiplication) in real arithmetic, each iteration has a cost of the order of n^3. Details are shown in the following table :

	NK
QR	$43n^3 + 20n^2 - 37n + 14$
(L+I)U	$34n^3 + 2n^2 - 9n$

Newton type iterations show to be an efficient scheme to compute in a few flops the classical QR and (L+I)U factorizations when applied to a data matrix which is already almost upper triangular. The convergence hypotheses include the invertibility of both the data and its diagonal part. An application of these strategies is given in [1], where both factorizations are used for spectral computation purposes.

References

[1] M. Ahues, M. Hama and A. Largillier, Newton-Kantorovich method for Refining Schur and Gauss triangular forms, Inter. J. of pure and applied mathematics, Vol.65 (2010), No.2.

[2] M. Ahues, Newton Methods with Hölder Derivative, Numer. Func. Anal. and Optimiz. Vol. 25, 5&6, (2004) 1-17.

[3] G. H. Golub and C. F. Van Loan, Matrix Computations, The Johns Hopkins University Press, Baltimore, MD, 3rd edition, (1996).

[4] L.V. Kantorovich and V.I. Krylov, Approximate Methods of Higher Analysis, Wiley (Interscience), New York, (1955).