Journal of Computational and Applied Mathematics 27 (1989) 17–36
North-Holland

A Jacobi-like algorithm for computing the generalized Schur form of a regular pencil

J.-P. CHARLIER and P. VAN DOOREN
Philips Research Laboratory, Av. Van Becelaere 2, Box 8, B-1170 Brussels, Belgium

Received 29 June 1988
Revised 13 October 1988

Abstract: We develop a Jacobi-like scheme for computing the generalized Schur form of a regular pencil of matrices $\lambda B - A$. The method starts with a preliminary triangularization of the matrix B and iteratively reduces A to triangular form, while maintaining B triangular. The scheme heavily relies on the technique of Stewart for computing the Schur form of an arbitrary matrix A. Just as Stewart's algorithm, this one can efficiently be implemented in parallel on a square array of processors. This explains some of its peculiarities, and at the same time yields further insight in Stewart's algorithm.

Keywords: Generalized Schur decomposition, parallel algorithm, linear algebra.

1. Introduction

The cyclic by rows version of the Jacobi algorithm for computing the eigenvalue decomposition of an $n \times n$ Hermitian matrix performs iteratively "sweeps" of unitary transformations:

$$\begin{matrix} (1,2)(1,3)(1,4)\ldots(1,n-1)(1,n) \\ (2,3)(2,4)\ldots(2,n-1)(2,n) \\ \ldots \\ (n-2,n-1)(n-2,n) \\ (n-1,n) \end{matrix} \qquad (1)$$

where (i, j) denotes a Givens rotation that only affects rows and columns i and j such that the elements (i, j) and (j, i) are annihilated. For each of these annihilations, there are two possible angles from which the smaller (or *inner*) one is chosen.

Recently, Brent, Luk and Van Loan have proposed a parallel implementation of this algorithm [1,2]. It consists in a reordering of the rotations (1) in order to execute them efficiently on a square grid of systolic processors. With such an array of $O(n \times n)$ processors, the diagonalization is then achieved in *linear*, i.e. $O(n)$, *time*. This striking result is due to the facts that

(i) several of the rotations in (1) can be performed in parallel,

(ii) successive "groups" of rotations can be pipelined on the square grid of processors.

Each of these two factors yields a speedup of the order of n. On the other hand, the convergence of the algorithm is such that, in practice, the number of sweeps is almost independent of n (see [1]).

Various extensions of this basic algorithm were soon presented for related decompositions of a matrix A or of a pair of matrices A and B. Those extensions differ mainly by the effect of appropriate unitary transformations on the 2×2 diagonal blocks of A or of the pair (A, B). They include the eigenvalue decomposition algorithm for normal matrices proposed by Goldstine and Horwitz [10]; the Schur decomposition algorithm proposed by Stewart [24], see also [4]; the singular value decomposition originally proposed by Kogbetliantz [15,16] and rederived for parallel computers by Brent et al. [1,2]; and the generalized singular value decomposition algorithm presented by Paige [19]. Other related developments are the QR-decomposition algorithm proposed by Luk [17] (which in fact is not iterative but terminates after $\frac{3}{2}n$ time steps), the product singular value decomposition proposed by Heath et al. [13] and by Fernando and Hammarling [8], and the construction of the "closest matrix" proposed by Ruhe [22].

Among all (standard or generalized) eigenvalue and singular value decompositions involving only unitary transformations, there is definitely one that is missing and prevents the picture from being complete: the *generalized Schur form* of a regular (i.e. $\det(B - A) \not\equiv 0$) pencil $\lambda B - A$ with A and B arbitrary in $\mathbb{C}^{n \times n}$ [11, p. 253]. It consists in constructing unitary matrices U and V such that

$$U^* = (\lambda B - A)V = (\lambda B_s - A_s), \tag{2}$$

where $A_s \equiv (\hat{a}_{ij})$ and $B_s \equiv (\hat{b}_{ij})$ are upper triangular. On sequential machines, (2) is typically computed by the QZ-algorithm of Moler and Stewart [18]. Here we present instead a Jacobi-like method for constructing iteratively the matrices A_s and B_s. Let us introduce it briefly. By analogy with the above-mentioned algorithms, U (resp. V) will be approached by successive application of Givens rotations $G_{ij}(\phi_k, d_k)$ (resp. $G_{ij}(\psi_k, e_k)$):

$$G_{ij}(\phi_k, d_k) = \begin{bmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & \cos\phi_k & \dots & d_k \sin\phi_k & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & -\bar{d}_k \sin\phi_k & \dots & \cos\phi_k & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{bmatrix} \begin{matrix} \\ \\ \leftarrow i \\ \\ \leftarrow j \\ \\ \\ \end{matrix} \tag{3}$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \uparrow i \qquad\qquad\quad \uparrow j$$

$(k = 1, 2, \dots)$; $\bar{d}_k$ stands for the complex conjugate of d_k, $|d_k| = 1$, and it is assumed that, at step k, a rotation in the plane (i, j) is performed. Denoting by $A_k \equiv (a_{lm}^{(k)})$ and $B_k \equiv (b_{lm}^{(k)})$ the iterates after execution of the kth step, and writing only the effect on the related diagonal blocks, we characterize the method as

$$\begin{bmatrix} a_{ii}^{(k)} & a_{ij}^{(k)} \\ 0 & a_{jj}^{(k)} \end{bmatrix} = \begin{bmatrix} \cos\phi_k & -d_k \sin\phi_k \\ \bar{d}_k \sin\phi_k & \cos\phi_k \end{bmatrix} \begin{bmatrix} a_{ii}^{(k-1)} & a_{ij}^{(k-1)} \\ a_{ji}^{(k-1)} & a_{jj}^{(k-1)} \end{bmatrix} \begin{bmatrix} \cos\psi_k & e_k \sin\psi_k \\ -\bar{e}_k \sin\psi_k & \cos\psi_k \end{bmatrix}, \tag{4}$$

$$\begin{bmatrix} b_{ii}^{(k)} & b_{ij}^{(k)} \\ 0 & b_{jj}^{(k)} \end{bmatrix} = \begin{bmatrix} \cos\phi_k & -d_k \sin\phi_k \\ \bar{d}_k \sin\phi_k & \cos\phi_k \end{bmatrix} \begin{bmatrix} b_{ii}^{(k-1)} & b_{ij}^{(k-1)} \\ b_{ji}^{(k-1)} & b_{jj}^{(k-1)} \end{bmatrix} \begin{bmatrix} \cos\psi_k & e_k \sin\psi_k \\ -\bar{e}_k \sin\psi_k & \cos\psi_k \end{bmatrix}. \tag{5}$$

An elementary 2×2 generalized Schur decomposition is thus realized at each step. It is easily seen that (4)–(5) amounts to two 2×2 Schur decompositions since the matrices

$$\begin{bmatrix} \cos \psi_k & -e_k \sin \psi_k \\ \bar{e}_k \sin \psi_k & \cos \psi_k \end{bmatrix} \begin{bmatrix} b_{ii}^{(k-1)} & b_{ij}^{(k-1)} \\ b_{ji}^{(k-1)} & b_{jj}^{(k-1)} \end{bmatrix}^{-1} \begin{bmatrix} a_{ii}^{(k-1)} & a_{ij}^{(k-1)} \\ a_{ji}^{(k-1)} & a_{jj}^{(k-1)} \end{bmatrix} \times \begin{bmatrix} \cos \psi_k & e_k \sin \psi_k \\ -\bar{e}_k \sin \psi_k & \cos \psi_k \end{bmatrix} \tag{6}$$

and

$$\begin{bmatrix} \cos \phi_k & -d_k \sin \phi_k \\ \bar{d}_k \sin \phi_k & \cos \phi_k \end{bmatrix} \begin{bmatrix} a_{ii}^{(k-1)} & a_{ij}^{(k-1)} \\ a_{ji}^{(k-1)} & a_{jj}^{(k-1)} \end{bmatrix} \begin{bmatrix} b_{ii}^{(k-1)} & b_{ij}^{(k-1)} \\ b_{ji}^{(k-1)} & b_{jj}^{(k-1)} \end{bmatrix}^{-1} \times \begin{bmatrix} \cos \phi_k & d_k \sin \phi_k \\ -\bar{d}_k \sin \phi_k & \cos \phi_k \end{bmatrix} \tag{7}$$

are now triangular. Yet none of these matrices is actually a submatrix of $B_k^{-1}A_k$ or $A_kB_k^{-1}$. In this sense this method differs from Stewart's standard Schur algorithm [24]. Nevertheless, when convergence is almost achieved, the matrices A_k and B_k are both *nearly* triangular. If moreover $j = i + 1$, then the 2×2 matrices (6) and (7) are *near* to the corresponding 2×2 blocks of $B_k^{-1}A_k$ and $A_kB_k^{-1}$, respectively. This will be further analyzed in the sequel.

In the next section, we develop preliminary results about "normal pencils", needed for a better understanding of our method. In Section 3 the method and its possible variants are explained in more detail and are related to Stewart's Schur decomposition. Global and asymptotic convergence are then analyzed in Sections 4 and 5, respectively. In Section 6 we give some test examples illustrating the convergence analysis. Finally concluding remarks include comments about the possible derivation of a real variant of the Schur algorithm.

2. Normal pencils

The standard Schur form of a matrix A is diagonal if and only if A is normal. Similarly, in the generalized situation, special forms occur when the pencil is "normal" in some sense. We investigate these forms here. For convenience, one of both matrices of the pencil is first assumed to be invertible, but it will be argued that this constraint is not crucial. *Normality* is important because it can be associated to fast asymptotic convergence of Jacobi-like methods for computing Schur decompositions (see [24] and later sections).

Theorem 2.1. *Let $\lambda B - A$ be a regular pencil with B invertible. Then there always exist unitary transformations U and V yielding a generalized Schur decomposition*

$$\lambda B_s - A_s = U^*(\lambda B - A)V \tag{8}$$

of the form

(i) $\lambda B_s - A_s = T(\lambda D_b - D_a)$, *if $B^{-1}A$ is normal,*

(ii) $\lambda B_s - A_s = (\lambda D_b - D_a)T$, *if AB^{-1} is normal,*

(iii) $\lambda B_s - A_s = \lambda D_b - D_a$, *if both $B^{-1}A$ and AB^{-1} are normal,*

where D_a and D_b are diagonal and T is unit upper triangular.

Proof. We start from *any* generalized Schur decomposition (8), which always exists. Decompose then A_s as $T_a D_a$ and B_s as $T_b D_b$ with both T_a and T_b unit upper triangular. If $B^{-1}A$ is normal, $D_b^{-1}(T_b^{-1}T_a)D_a$ is normal and upper triangular by construction. Therefore, it must also be diagonal. If D_a is non-singular, then one has $T_b^{-1}T_a = I$ and (i) follows with $T = T_a = T_b$. If D_a is singular, then $T_b^{-1}T_a \doteq T_{\text{up}}$ is the identity matrix *except* possibly for non-zero elements above the diagonal in the columns of T_{up} corresponding to zero diagonal elements in D_a. Hence $T_{\text{up}}D_a = D_a$. But then A_s could as well be decomposed as $A_s = \hat{T}_a D_a$ with $\hat{T}_a \doteq T_a T_{\text{up}}^{-1}$. Therefore $T_b^{-1}\hat{T}_a = I$ and (i) follows now with $T = \hat{T}_a = T_b$.

If AB^{-1} is normal, a similar reasoning yields (ii).

Finally, if both $B^{-1}A$ and AB^{-1} are normal, then we have simultaneously that $\lambda B_s - A_s = T_l(\lambda D_b - D_a) = (\lambda D_b - D_a)T_r$ for some unit upper triangular matrices T_l and T_r. If $\lambda D_b - D_a$ has distinct diagonal elements for *some* value of λ, then one must have $T_l = T_r = I$ since this is the only upper triangular matrix commuting with a diagonal matrix with distinct diagonal elements, and (iii) follows immediately. On the other hand, if $\lambda D_b - D_a$ has repeated diagonal elements for *all* values λ, then this must also be the case for D_b and D_a separately. We show that (8) can then be updated by performing additional row and column transformations such that T_r and T_l become both the identity matrix. For simplicity of the argument, let us suppose there is only one repeated value and the equal diagonal elements in D_a and D_b are adjacent, say,

$$\begin{aligned} D_a &= \operatorname{diag}\{x_1, \ldots, x_k, \alpha, \ldots, \alpha, x_l, \ldots, x_n\}, \\ D_b &= \operatorname{diag}\{y_1, \ldots, y_k, \beta, \ldots, \beta, y_l, \ldots, y_n\}. \end{aligned} \tag{9}$$

Then it follows that

$$T_r = T_l = \operatorname{diag}\{I_k, \hat{T}, I_{n-l}\} \tag{10}$$

with $\hat{T}$ unit upper triangular. Let now $\hat{T} = \hat{U}\hat{\Sigma}\hat{V}^*$ be the singular value decomposition of this diagonal block. Because of the special form of D_a and D_b, the factors $\hat{U}$ and $\hat{V}$ can be "absorbed" in the matrices U and V and the new diagonal blocks of D_a and D_b become respectively $\alpha\hat{\Sigma}$ and $\beta\hat{\Sigma}$. In this updated decomposition one clearly has $\hat{T}_l = \hat{T}_r = I$ and (iii) is proved. □

An annoying detail in this theorem is the condition that B must be invertible. This can be avoided as follows. Let $\lambda B - A$ be a regular pencil and pick arbitrary values s and c (with $s^2 + c^2 = 1$). Consider the new pencil

$$\lambda'(sB - cA) - (cB + sA). \tag{11}$$

The eigenvalues of this pencil and those of $\lambda B - A$ are related by

$$\lambda'_i = (c\hat{b}_{ii} + s\hat{a}_{ii})/(s\hat{b}_{ii} - c\hat{a}_{ii}), \qquad \lambda_i = \hat{a}_{ii}/\hat{b}_{ii}, \tag{12}$$

where $\hat{a}_{ii}$ and $\hat{b}_{ii}$ are the diagonal elements of the generalized Schur decomposition (2) of $\lambda B - A$. This follows easily from the fact that if $\lambda B_s - A_s = U^*(\lambda B - A)V$ is a generalized Schur decomposition for $\lambda B - A$, then $\lambda'(sB_s - cA_s) - (cB_s + sA_s) = U^*[\lambda'(sB - cA) - (cB + sA)]V$ is a generalized Schur decomposition for $\lambda'(sB - cA) - (cB + sA)$. One easily checks then that if $\lambda B - A$ is regular, so is $\lambda'(sB - cA) - (cB + sA)$. Moreover, there is always a point s/c which is *not* an eigenvalue of $\lambda B - A$, and hence $sB - cA$ is then regular. Finally, whenever the appropriate matrices are invertible, one easily checks that $B^{-1}A$ is normal iff $(sB - cA)^{-1}(cB +$

sA) is normal and that AB^{-1} is normal iff $(cB + sA)(sB - cA)^{-1}$ is normal. The transformation (11) thus preserves the decompositions of Theorem 2.1 for *all* pairs (s, c) and these exist if for *some* pair (s, c) the matrices $(sB - cA)^{-1}(cB + sA)$ and/or $(cB + sA)(sB - cA)^{-1}$ are normal. From this we are led to the following definition of what could be called a "normal pencil" $\lambda B - A$.

Definition 2.2. A regular pencil $\lambda B - A$ is said (i) *left normal*, (ii) *right normal*, and (*iii*) *normal*, iff there exist unitary transformations U and V yielding a generalized Schur decomposition

$$\lambda B_s - A_s = U^*(\lambda B - A)V \tag{13}$$

respectively of the form

(i) $\lambda B_s - A_s = T(\lambda D_b - D_a)$,
(ii) $\lambda B_s - A_s = (\lambda D_b - D_a)T$,
(iii) $\lambda B_s - A_s = \lambda D_b - D_a$,

where D_a and D_b are diagonal and T is unit upper triangular.

Since the invertibility of B is not essential in this anymore, we will suppose in the sequel of this paper that B is invertible in order to simplify all discussions.

3. Description of the method

Basically, we want to obtain the generalized Schur decomposition of a pencil $\lambda B - A$ by applying to it rotations of the type (4)–(5) in an iterative manner. Notice that one of the two matrices, say B, can be made triangular in a *finite* number of steps by a preliminary QR-decomposition. We shall see that triangularity is then automatically maintained for the iterates B_k in the method described below. Although this preprocessing is not essential, it simplifies notations and derivations, and also somewhat decreases the complexity of each iteration. Moreover, it can be executed systolically in $\frac{3}{2}n$ time steps [17]. Unless otherwise stated, we thus assume in the sequel that B is upper triangular.

Let then L_k be the strictly lower part of A_k. A method is said to be convergent when the Frobenius norm of L_k, i.e.

$$\| L_k \| \doteq \sqrt{\sum_{l>m} |a_{lm}^{(k)}|^2}, \tag{14}$$

tends to 0. Before examining this in further sections, we have to choose a particular ordering of the elementary rotations and to specify which angles are to be considered at a given step.

Remark first that ϕ_k and ψ_k are each one of the two solutions of a quadratic equation which can be derived from (6) or (7). Let us denote by ϕ_O (resp. ψ_O) the solution for which $|\sin \phi_k|$ (resp. $|\sin \psi_k|$) is the nearest to 1, and by ϕ_I (resp. ψ_I) the other solution, at step k. The rotations corresponding to ϕ_O and ψ_O will be referred to as the "outer" rotations, while those corresponding to ϕ_I will be referred to as the "inner" rotations. In both pairs (ϕ_k, ψ_k) determined according to (4)–(5), inner and outer rotations are not necessarily associated with each other. It is not difficult to derive from (6) that the product $|\tan \psi_I \cdot \tan \psi_O|$ is given by the

ratio of the off-diagonal elements in the 2×2 block before rotation:

$$|\tan \psi_I \cdot \tan \psi_O| = \left| \frac{b_{ii}^{(k-1)} a_{ji}^{(k-1)}}{b_{jj}^{(k-1)} a_{ij}^{(k-1)} - b_{ij}^{(k-1)} a_{jj}^{(k-1)}} \right|, \tag{15}$$

and, similarly from (7), that

$$|\tan \phi_I \cdot \tan \phi_O| = \left| \frac{a_{ji}^{(k-1)} b_{jj}^{(k-1)}}{a_{ij}^{(k-1)} b_{ii}^{(k-1)} - a_{ii}^{(k-1)} b_{ij}^{(k-1)}} \right|. \tag{16}$$

These relations are used later on.

We now describe the method. As it is heavily inspired by the one proposed by Stewart in the standard case ($B = I$), we limit the description to essential features, referring to Stewart's paper [24, particularly Sections 3 and 4] for further developments. The method of Stewart is based on the two following choices:

1. *Only rotations in planes* $(i, i+1)$ *are performed.* At each step, the transfer between the lower and the upper triangular parts of the matrix is limited inside the diagonal block, and is only due to the annihilation of the element in position $(i+1, i)$. Other transfers, in particular undesirable ones from the upper part to the lower part, are thus ruled out.
2. According to what precedes, only elements of the first subdiagonal are annihilated. Therefore, in order to maximize the "mixing" of the matrix at each step, *only outer rotations are considered.* This tends to ensure that a significant part of other elements of the lower diagonal part move into the first subdiagonal and be subsequently annihilated. Nevertheless, in a number of situations, the outer rotations are close or even equal to the identity matrix, and the algorithm may not converge. Attempts to basically improve this behavior have failed so far [4].

Choice 1 is maintained in the generalized case. At step k, we thus have

$$\| L_k \|^2 = \| L_{k-1} \|^2 - \left| a_{i+1,i}^{(k-1)} \right|^2. \tag{17}$$

The norm of L_k never increases for increasing values of k. Additional features are that

(i) B_k is then upper triangular as B_{k-1} was,

(ii) the product of the 2×2 diagonal blocks of A_k and B_k^{-1} is close to the corresponding block of $A_k B_k^{-1}$.

The iterative process is then divided in a number of sweeps, during each of which all the lower diagonal elements would be temporarily annihilated. As in the standard case [24], we consider here two kinds of sweeps. A *forward sweep* consists of the following sequence of rotation planes:

$$\begin{array}{l} (1,2)(2,3) \ldots (n-2, n-1)(n-1, n) \\ (1,2)(2,3) \ldots (n-2, n-1) \\ \ldots \\ (1,2)(2,3) \\ (1,2) \end{array} \tag{18}$$

and a *backward sweep* corresponds to the sequence:

$$\begin{array}{l}(n-1,\ n)(n-2,\ n-1)\ \ldots\ (2,\ 3)(1,\ 2)\\ (n-1,\ n)(n-2,\ n-1)\ \ldots\ (2,\ 3)\\ \ldots\\ (n-1,\ n)(n-2,\ n-1)\\ (n-1,\ n)\end{array} \tag{19}$$

In the standard case, if the outer rotations are distant enough from the identity matrix, the application of a forward or of a backward sweep of outer rotations essentially reduces to the inversion of the order of the lower diagonals. The parallel implementation of these orderings in the generalized case is the same as in the standard case, except that two matrices, instead of one, are mapped on the array of processors (see [24, Fig. 4.3]). In particular a *double sweep*, consisting of a forward and a backward sweep, can efficiently be pipelined on such an array.

The generalization of choice 2 is less immediate. In contrast to the standard case, two angles are to be computed at each step and, since they are not independent, it is not possible in general to retain for both the solution corresponding to the outer rotation. Also, to some extent, convergence properties depend jointly on two choices: the side (left or right) on which an outer rotation is applied, and the type of sweep (forward or backward) which is performed. Anticipating on the next sections, we make this a little more precise in the two following points:

(1) It will be shown (Section 4) that (roughly) the process converges provided that $|\sin \phi_k|$ is large enough at each step of a forward sweep and that $|\sin \psi_k|$ is large enough at each step of a backward sweep. Therefore, we choose the solution for ϕ_k (resp. ψ_k) corresponding to the outer rotation at any step of a forward (resp. backward) sweep.

(2) Let us assume first that the process has reached a stage near the convergence, i.e. that $a_{i+1,i}^{(k)}$ is close to 0, and secondly that the pencil is right normal, i.e. that $A_k B_k^{-1}$ is normal. Applying Theorem 2.1, we thus have $a_{i,i+1}^{(k)} b_{ii}^{(k)} \simeq b_{i,i+1}^{(k)} a_{ii}^{(k)}$. Hence, except possibly for special matrix patterns, the product (15) is close to 0, while (16) takes yet a finite value. Since both inner rotations are then close to the identity, the choice of the outer rotation for the transformation on the left side (angle ϕ_k) seems to be appropriate. Conversely, if the pencil was left normal, the outer rotation to the right side (angle ψ_k) would be chosen. Indeed, we shall prove (Section 5) that the convergence of the process, when applied to a right (resp. left) normal pencil, is "ultimately" quadratic through any forward (resp. backward) sweep if an outer rotation is performed on the left (resp. to the right) side at each step.

Summing up, we propose the following method (its features will be analyzed and tested in the rest of the paper):

Method. *Let $\lambda B - A$ be an arbitrary pencil with B upper triangular. Sequences of iterates A_k and B_k ($k = 1, 2, \ldots$) are generated by applying to it an alternance of forward sweeps* (18) *and backward sweeps* (19), *until the Frobenius norm* (17) *of the strictly lower part of A_k is smaller than a prescribed value or stagnates. At each step of a forward (resp. backward) sweep, an outer rotation is performed on the left side (resp. right side), from which the right side (resp. left side) rotation is deduced, such that the elementary decomposition* (4)–(5) *is obtained.*

4. Global convergence

Stewart's algorithm for computing the standard Schur decomposition of a matrix is not convergent in general. The algorithm discussed here for computing the generalized Schur decomposition of a pencil is not convergent either, to the same extent: there does not exist any neat characterization of the whole class of pencils for which this algorithm converges, and not any modification of the method is known which could guarantee convergence. Nevertheless, it is possible to derive a nontrivial *sufficient condition* for convergence. We develop it now.

Let us apply the algorithm to some pencil $\lambda B_0 - A_0$, and let us follow its evolution over, say, a forward sweep (a similar argument should hold for a backward sweep). For notational convenience, the reasoning will be illustrated for matrices of order 5. Such a formal simplification was already used by Wilkinson in proving the ultimate quadratic convergence of the standard Jacobi method for Hermitian matrices [26], and, more recently, by Fernando in the global convergence proof of a particular implementation of the Kogbetliantz method for computing the singular value decomposition of an arbitrary matrix [7]. In some respects, our result is related to Fernando's.

The pencil $\lambda B_0 - A_0$ is thus transformed by a sequence of ten Givens rotations, with the following order of transformation planes:

$$\begin{array}{l}(1,2)\,(2,3)\,(3,4)\,(4,5)\\(1,2)\,(2,3)\,(3,4)\\(1,2)\,(2,3)\\(1,2)\end{array} \tag{20}$$

At step k, corresponding to the plane $(i, i+1)$, a pair of rotation angles (ϕ_k, ψ_k) is computed in such a way that the elements $(i+1, i)$ of A_k and B_k are set or maintained to 0. Looking at the sequence $\{A_k\}$, we aim at bounding $\| L_{10} \|$ in terms of $\| L_0 \|$ (see (14)).

Let us detail the iterates A_k. The first ones are shown below. Each lower element is identified by a letter the index of which increases only when the value of the element is modified; x is generic.

$$A_0 = \begin{bmatrix} x & x & x & x & x \\ a_0 & x & x & x & x \\ b_0 & e_0 & x & x & x \\ c_0 & f_0 & h_0 & x & x \\ d_0 & g_0 & i_0 & j_0 & x \end{bmatrix} \underset{(1,2)}{\overset{k=1}{\rightarrow}} \begin{bmatrix} x & x & x & x & x \\ 0 & x & x & x & x \\ b_1 & e_1 & x & x & x \\ c_1 & f_1 & h_0 & x & x \\ d_1 & g_1 & i_0 & j_0 & x \end{bmatrix} \underset{(2,3)}{\overset{k=2}{\rightarrow}} \begin{bmatrix} x & x & x & x & x \\ a_2 & x & x & x & x \\ b_2 & 0 & x & x & x \\ c_1 & f_2 & h_1 & x & x \\ d_1 & g_2 & i_1 & j_0 & x \end{bmatrix}$$

$$\underset{(3,4)}{\overset{k=3}{\rightarrow}} \begin{bmatrix} x & x & x & x & x \\ a_2 & x & x & x & x \\ b_3 & e_3 & x & x & x \\ c_2 & f_3 & 0 & x & x \\ d_1 & g_2 & i_2 & j_1 & x \end{bmatrix} \underset{(4,5)}{\overset{k=4}{\rightarrow}} \begin{bmatrix} x & x & x & x & x \\ a_2 & x & x & x & x \\ b_3 & e_3 & x & x & x \\ c_3 & f_4 & h_3 & x & x \\ d_2 & g_3 & i_3 & 0 & x \end{bmatrix} = A_4. \tag{21}$$

The elements of the bottom row of A_4 can be written as

$$\begin{bmatrix} j_2 \\ i_3 \\ g_3 \\ d_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ i_2 & 0 & 0 \\ g_2 & f_2 & 0 \\ d_1 & c_1 & b_1 \end{bmatrix} \begin{bmatrix} \cos \phi_4 \\ \bar{d}_4 \sin \phi_4 \cdot \cos \phi_3 \\ \bar{d}_4 \sin \phi_4 \cdot \bar{d}_3 \sin \phi_3 \cdot \cos \phi_2 \end{bmatrix}, \tag{22}$$

where part of their history is made explicit from (21). The Euclidean norm of the right-hand side vector is readily seen to be equal to $(1 - \sin^2\phi_4 \cdot \sin^2\phi_3 \cdot \sin^2\phi_2)^{1/2}$. On the other hand, the Frobenius norm of the right-hand side matrix equals $\| L_4 \|$; indeed the equalities

$$\begin{aligned} &|b_1|^2 + |c_1|^2 + |d_1|^2 = |a_2|^2 + |b_3|^2 + |c_3|^2 + |d_2|^2, \\ &|g_2|^2 + |f_2|^2 = |e_3|^2 + |f_4|^2 + |g_3|^2, \qquad |i_2|^2 = |h_3|^2 + |i_3|^2 \end{aligned} \tag{23}$$

follow from the fact that, at a given step, the norm of a line is altered only if one of its elements is annihilated. Summing up, and taking into account that $\| L_k \|$ is not increasing when k increases, we have from (22) that

$$\begin{aligned} |i_3|^2 + |g_3|^2 + |d_2|^2 &\leqslant \| L_4 \|^2 (1 - \sin^2\phi_4 \cdot \sin^2\phi_3 \cdot \sin^2\phi_2) \\ &= \| L_4 \|^2 (1 - \sin^2\phi_{2,4}) = \| L_0 \|^2 (1 - \sin^2\hat{\phi}_{2,4}) \end{aligned} \tag{24}$$

for *some* angles $\hat{\phi}_{2,4}$ and $\phi_{2,4}$ satisfying

$$\sin^2\hat{\phi}_{2,4} \geqslant \sin^2\phi_{2,4} \geqslant \sin^2\phi_4 \cdot \sin^2\phi_3 \cdot \sin^2\phi_2 . \tag{25}$$

The norm of the bottom row is not modified by subsequent steps and the expression (24) thus remains valid through them.

A same argument is now applied to the strictly lower triangular part of the $(n-1) \times (n-1)$ leading principal submatrix of A_4. The next three steps are

$$\begin{bmatrix} x & x & x & x & x \\ 0 & x & x & x & x \\ b_4 & e_4 & x & x & x \\ c_4 & f_5 & h_3 & x & x \\ d_3 & g_4 & i_3 & j_2 & x \end{bmatrix} \underset{(1,2)}{\overset{k=5}{\rightarrow}} \begin{bmatrix} x & x & x & x & x \\ a_4 & x & x & x & x \\ b_5 & 0 & x & x & x \\ c_4 & f_6 & h_4 & x & x \\ d_3 & g_5 & i_4 & j_2 & x \end{bmatrix} \underset{(2,3)}{\overset{k=6}{\rightarrow}}$$

$$\underset{(3,4)}{\overset{k=7}{\rightarrow}} \begin{bmatrix} x & x & x & x & x \\ a_4 & x & x & x & x \\ b_6 & e_6 & x & x & x \\ c_5 & f_7 & 0 & x & x \\ d_3 & g_5 & i_5 & j_3 & x \end{bmatrix} = A_7 .$$

Subsequent steps do not alter the norm of the $(n-1)$th row, which can be expressed as

$$\begin{aligned} |c_5|^2 + |f_7|^2 &\leqslant \left[\|L_7\|^2 - \|L_4\|^2 \cos^2\phi_{2,4}\right](1 - \sin^2\phi_7 \cdot \sin^2\phi_6) \\ &\leqslant \|L_7\|^2 \sin^2\phi_{2,4}(1 - \sin^2\phi_7 \cdot \sin^2\phi_6) \\ &= \|L_7\|^2 \sin^2\phi_{2,4}(1 - \sin^2\phi_{6,7}) \\ &= \|L_0\|^2 \sin^2\phi_{2,4}(1 - \sin^2\hat{\phi}_{6,7}) \end{aligned} \tag{26}$$

with

$$\sin^2\hat{\phi}_{6,7} \geqslant \sin^2\phi_{6,7} \geqslant \sin^2\phi_7 \cdot \sin^2\phi_6. \tag{27}$$

Similarly, after the 9th step:

$$\overset{\substack{k=8\\ \rightarrow\\ (1,2)}}{\begin{bmatrix} x & x & x & x & x \\ 0 & x & x & x & x \\ b_7 & e_7 & x & x & x \\ c_6 & f_8 & h_5 & x & x \\ d_4 & g_6 & i_5 & f_3 & x \end{bmatrix}} \overset{\substack{k=9\\ \rightarrow\\ (2,3)}}{\begin{bmatrix} x & x & x & x & x \\ a_6 & x & x & x & x \\ b_8 & 0 & x & x & x \\ c_6 & f_9 & h_6 & x & x \\ d_4 & g_7 & i_6 & j_3 & x \end{bmatrix}} = A_9$$

we have

$$\begin{aligned} |b_8|^2 &\leqslant \left[\|L_9\|^2 - \|L_4\|^2 \cos^2\phi_{2,4} - \|L_7\|^2 \sin^2\phi_{2,4} \cdot \cos^2\phi_{6,7}\right](1 - \sin^2\phi_9) \\ &\leqslant \|L_9\|^2 \sin^2\phi_{2,4} \cdot \sin^2\phi_{6,7}(1 - \sin^2\phi_9) \\ &= \|L_9\|^2 \sin^2\phi_{2,4} \cdot \sin^2\phi_{6,7}(1 - \sin^2\phi_{9,9}) \\ &= \|L_0\|^2 \sin^2\phi_{2,4} \cdot \sin^2\phi_{6,7}(1 - \sin^2\hat{\phi}_{9,9}) \end{aligned} \tag{28}$$

with

$$\sin^2\hat{\phi}_{9,9} \geqslant \sin^2\phi_{9,9} \geqslant \sin^2\phi_9. \tag{29}$$

Finally, the last step

$$\overset{\substack{k=10\\ \rightarrow\\ (1,2)}}{\begin{bmatrix} x & x & x & x & x \\ 0 & x & x & x & x \\ b_9 & e_9 & x & x & x \\ c_7 & f_{10} & h_6 & x & x \\ d_5 & g_8 & i_6 & j_2 & x \end{bmatrix}} = A_{10}, \quad \text{yields } |a_7|^2 = 0. \tag{30}$$

Adding the contribution of each row (i.e. (24), (26), (28), and (30)), we can characterize the effect of the entire forward sweep by

$$\begin{aligned} \|L_{10}\|^2 &= \|L_0\|^2(\cos^2\hat{\phi}_{2,4} + \sin^2\phi_{2,4} \cdot \cos^2\hat{\phi}_{6,7} + \sin^2\phi_{2,4} \cdot \sin^2\phi_{6,7} \cdot \cos^2\hat{\phi}_{9,9}) \\ &\leqslant \|L_0\|^2(1 - \sin^2\hat{\phi}_{2,4} \cdot \sin^2\hat{\phi}_{6,7} \cdot \sin^2\hat{\phi}_{9,9}) \end{aligned} \tag{31}$$

or, in terms of the rotation angles

$$\|L_{10}\|^2 \leqslant \|L_0\|^2\left[1 - (\sin^2\phi_2 \cdot \sin^2\phi_3 \cdot \sin^2\phi_4)(\sin^2\phi_6 \cdot \sin^2\phi_7)(\sin^2\phi_9)\right]. \tag{32}$$

Clearly, this relation does not depend in any way on the triangularity of B. In fact, if B is full,

the same inequality can be written for the iterates B_k's, with L_k redefined accordingly. Also, generalization to pencils of arbitrary order is immediate. On the other hand, a similar bound holds for a backward sweep: simply, the angles ψ_k's are then to be considered instead of the ϕ_k's. As already mentioned, this result, as well as its proof, is formally comparable to one obtained by Fernando in another context [7]. Moreover, related bounds have been derived for various Jacobi-like diagonalization processes, but both angle sets are involved in general [12,14]. Finally, note that (32) is also valid in the standard case (Stewart's method); the distinction between forward and backward sweeps is however not relevant anymore, since both angles are identical at each step.

From (32) and the corresponding inequality for a backward sweep, we can directly infer the following *sufficient condition* for convergence to upper triangular forms:

Theorem 4.1. *Any Jacobi-like method, consisting of forward and backward sweeps of elementary rotations* (4)–(5) *between adjacent rows and columns, is convergent if*

$$\phi_k \notin [-\epsilon, \epsilon] \cup [\pi - \epsilon, \pi + \epsilon] \quad \textit{through any forward sweep}$$

and

$$\psi_k \notin [-\epsilon, \epsilon] \cup [\pi - \epsilon, \pi + \epsilon] \quad \textit{through any backward sweep},$$

where ϵ is a positive constant, independent of k.

The requirement that ϵ be independent of k is introduced here to rule out limit situations. A similar constraint was already considered by Forsythe and Henrici when studying the global convergence of the standard Jacobi method for diagonalizing Hermitian matrices [9]. Theorem 4.1 obviously applies to the method defined in Section 3 and, *a fortiori*, to Stewart's method. As already mentioned, it remains an open question to characterize the class of pencils for which global convergence occurs in the standard as well as in the generalized cases.

5. Ultimate convergence

Near the convergence ("ultimately"), most of the Jacobi-like methods (with appropriate orderings) converge quadratically. In general, this does not depend on whether they are globally convergent or not. Ultimate quadratic convergence means (roughly) that, if the norm of the matrix part which is to be annihilated is already smaller than some distance between the diagonal elements (or their limit values, e.g. the eigenvalues), then its decrease over a subsequent sweep of elementary transformations is quadratic. When proposing his method for computing the standard Schur decomposition of an arbitrary matrix, Stewart indicated that its ultimate convergence is quadratic for normal matrices having distinct eigenvalues [24, Section 5]. He gave a qualitative analysis based on continuity arguments. We show here that a similar result holds for the method of Section 2 for computing the generalized Schur decomposition of a (left or right) normal pencil (see Definition 2.2), and we develop a *quantitative* reasoning which in return applies to the standard case. The following simple lemma will be useful.

Lemma 5.1. *Let N be a normal matrix. Then, for any partition*

$$N = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \tag{33}$$

with square diagonal blocks, we have

$$\| N_{12} \| = \| N_{21} \|. \tag{34}$$

Proof. Due to $NN^* = N^*N$, one has $N_{11}N_{11}^* + N_{12}N_{12}^* = N_{11}^*N_{11} + N_{21}^*N_{21}$. The result follows from comparing traces in this relation. □

We now state the main result. The parameter S stands for the number of elementary transformations in a forward or backward sweep: $S \doteq \frac{1}{2}[n(n-1)]$.

Theorem 5.2. *Let the pencil $\lambda B - A$ have distinct eigenvalues*

$$2\delta \doteq \min_{i \neq j} |\lambda_i - \lambda_j| > 0. \tag{35}$$

Assume that B is upper triangular and non-singular. Generate the sequence $\{A_k, B_k\}$ by the method defined in Section 3. Assume also that a state has been reached when

$$(1 + \sqrt{n-1}) \| L_r \| \| B^{-1} \|_2 < \tfrac{1}{2}\delta, \tag{36}$$

where L_r denotes the strictly lower part of A_r. Then the iterates A_k produced at the subsequent steps converge to the upper triangular form according to

$$\| L_{r+S} \| \leqslant \frac{4\sqrt{S}\, \| B^{-1} \|_2}{\delta} \| L_r \|^2 \tag{37}$$

over

- *any forward sweep, if $\lambda B - A$ is right normal,*
- *any backward sweep, if $\lambda B - A$ is left normal,*
- *any (forward or backward) sweep, if $\lambda B - A$ is normal.*

Proof. The proof is inspired, while it is more complex, by those for standard eigenvalue and singular value decompositions by Jacobi methods [20,21,26]. We detail the case of a right normal pencil, assuming thus that AB^{-1} is normal. For left normal or normal pencils, the argument is quite analogous.

Consider some step k $(k \geqslant r)$ in a forward sweep, corresponding to the rotation plane $(i, j = i+1)$ and to the rotation angles (ϕ_k, ψ_k). Denote by a hat the corresponding 2×2 diagonal blocks of A_k and B_k. At the step, the 2×2 matrix

$$\hat{A}_k \hat{B}_k^{-1} \equiv \begin{bmatrix} m_{ii} & m_{ij} \\ m_{ji} & m_{jj} \end{bmatrix} \tag{38}$$

is implicitly triangularized by the outer rotation of angle ϕ_k (see (7) and the definition of the method in Section 3) or, equivalently, the matrix

$$\begin{bmatrix} m_{jj} & -m_{ji} \\ -m_{ij} & m_{ii} \end{bmatrix} \tag{39}$$

is implicitly triangularized by the inner rotation of angle $(\phi_k - \frac{1}{2}\pi)$. This angle is easily shown to verify

$$|\tan(\phi_k - \tfrac{1}{2}\pi)| \leqslant \frac{2|m_{ij}|}{|m_{ii} - m_{jj}|} \tag{40}$$

provided that

$$\frac{|m_{ij}||m_{ji}|}{|m_{ii} - m_{jj}|^2} \leqslant \frac{1}{4} \tag{41}$$

holds. This is directly derived from an analysis of the quadratic equation yielding $\tan(\phi_k - \frac{1}{2}\pi)$ or, more generally, follows from a perturbation result of Stewart [23, Theorem 4.11] applied to the matrix (39). We verify condition (41) in the following three steps.

- Denote $N_k \doteq A_k B_k^{-1} \equiv (n_{lm})$. Remark that N_k is normal and B_k is upper triangular at every step k. For any block partitioning of the type (33) of N_k, A_k, and B_k, we have $(N_k)_{21} = (A_k)_{21}(B_k)_{11}^{-1}$. This and Lemma 5.1 gives $(k \geqslant r)$

$$\begin{aligned} \|(N_k)_{12}\| = \|(N_k)_{21}\| &= \|(A_k)_{21}(B_k)_{11}^{-1}\| \leqslant \|L_k\| \, \|(B_k)_{11}^{-1}\|_2 \leqslant \|L_k\| \, \|B^{-1}\|_2 \\ &\leqslant \|L_r\| \, \|B^{-1}\|_2, \end{aligned} \tag{42}$$

  where the invariance of the 2-norm with respect to orthogonal transformations and the monotonic decrease of $\|L_k\|$ for increasing values of k are taken into account.
- It is easily seen that the difference between $\hat{A}_k \hat{B}_k^{-1}$ and the corresponding 2×2 block of $A_k B_k^{-1}$, say X, satisfies

$$\begin{bmatrix} n_{ii} & n_{ij} \\ n_{ji} & n_{jj} \end{bmatrix} = \hat{A}_k \hat{B}_k^{-1} + X, \quad \text{with } \|X\| \leqslant \|L_r\| \, \|B^{-1}\|_2. \tag{43}$$

  Note that this bound clearly requires $j = i + 1$. We thus have

$$|m_{ii} - n_{ii}|, \, |m_{jj} - n_{jj}| \leqslant \|L_r\| \, \|B^{-1}\|_2, \tag{44}$$

  and, using (42),

$$|m_{ij}|, \, |m_{ji}| \leqslant 2\|L_r\| \, \|B^{-1}\|_2. \tag{45}$$

- On the other hand, the Gershgorin circle theorem [11, p. 200] yields here

$$|\lambda_i - n_{ii}| \leqslant \sum_{m=1}^{i-1} |n_{im}| + \sum_{m=i+1}^{n} |n_{im}| \leqslant \sqrt{n-1} \left(\sum_{m=1}^{i-1} n_{im}^2 + \sum_{m=i+1}^{n} n_{im}^2 \right)^{1/2} \tag{46}$$

  for some eigenvalue λ_i. Hence

$$|\lambda_i - n_{ii}| \leqslant \sqrt{n-1} \, \|L_r\| \, \|B^{-1}\|_2 \tag{47}$$

  results from the application of Lemma 5.1 to the second sum in (46) and from the fact that the Frobenius norm of the strictly lower part of N_k is smaller than $\|L_r\| \, \|B^{-1}\|_2$. Due to (44) and (46), we thus have

$$|\lambda_i - m_{ii}| \leqslant |\lambda_i - n_{ii}| + |n_{ii} - m_{ii}| \leqslant (1 + \sqrt{n-1}) \, \|L_r\| \, \|B^{-1}\|_2. \tag{48}$$

  Same bounds hold for the distance between n_{jj} or m_{jj} and another eigenvalue λ_j. Therefore,

the assumption (36) ensures that every diagonal element is associated unequivocally to one eigenvalue. Moreover,

$$\begin{aligned} |m_{ii} - m_{jj}| &= |(m_{ii} - \lambda_i) - (\lambda_i - \lambda_j) - (\lambda_j - m_{jj})| \\ &\geqslant |\lambda_i - \lambda_j| - |\lambda_i - m_{ii}| - |\lambda_j - m_{jj}| \\ &> 2\delta - 2(1 + \sqrt{n-1})\,\|L_r\|\,\|B^{-1}\|_2 > \delta. \end{aligned} \tag{49}$$

Combining (37), (45), and (49), one easily verifies that the condition (41) is satisfied (if $n \geqslant 2$). The bound (40) is then valid. We obtain

$$|\tan(\phi_k - \tfrac{1}{2}\pi)| < \frac{4\,\|L_r\|\,\|B^{-1}\|_2}{\delta}. \tag{50}$$

By using here the estimation (32), valid over a forward sweep, in the form

$$\|L_{r+S}\|^2 \leqslant \|L_r\|^2 \left(1 - \prod_{k=r}^{r+S} \sin^2\phi_k\right) \leqslant \|L_r\|^2 \sum_{k=r}^{r+S} \cos^2\phi_k, \tag{51}$$

and noting that $|\cos\phi_k| \leqslant |\tan(\phi_k + \frac{1}{2}\pi)|$, we finally have

$$\|L_{r+S}\|^2 \leqslant \|L_r\|^2 \sum_{k=r}^{r+S} \frac{16\|L_r\|^2\|B^{-1}\|_2^2}{\delta^2} = \frac{16S\|L_r\|^4\|B^{-1}\|_2^2}{\delta^2}, \tag{52}$$

i.e. (37). □

This result can be commented in several respects:

(i) According to Theorem 5.2, the ultimate convergence of our method is quadratic during (at least) every second sweep if the pencil is right or left normal and during any sweep if the pencil is merely normal. The "type of normality" of the current pencil is thus not presumed. On the contrary, if a pencil was known to be left (or right) normal, a variant of the method could of course be devised where only backward (or forward) sweeps would be performed. These situations, as well as the behavior of the method for non-normal pencils, are illustrated in the next section. Note however that, as mentioned in Section 3, the most efficient parallel implementation is obtained by alternating forward and backward sweeps throughout the iterative process.

(ii) The main difference between the estimate (37) and other ones valid for standard Jacobi-like methods is the presence of $\|B^{-1}\|_2$ in the coefficient. Whether this factor reflects real features is also tested in the next section.

(iii) The assumption that B is upper triangular is not essential. Without it, an inequality of the type (37) could still be obtained. But the condition (36) and the coefficient of the quadratic term in (37) would then take a (much) more complicated form, due in particular to the harder derivation of analogues of (42) and (43). Since moreover triangularity leads to lower computational complexity, we do not go deeper into the full case.

(iv) Clearly, Theorem 5.2 applies to Stewart's method for computing the standard Schur decomposition [24]. It suffices to set B to I in the above statement. Furthermore a factor 2 can be saved in (37) by looking closely at the proof. Indeed, in the standard case, we have $m_{lm} = n_{lm}$ ($X \equiv 0$ in (43)). Hence, the inequality $|n_{ij}|, |n_{ji}| \leqslant \|L_r\|$ is obtained instead of (45), and (48) reduces to (47). Taking this into account, we can write the following slightly sharper result:

Theorem 5.3. *Let the normal matrix A have distinct eigenvalues*

$$2\delta \doteq \min_{i \neq j} |\lambda_i - \lambda_j| > 0. \tag{53}$$

Generate the sequence $\{A_k\}$ *by Stewart's method. Assume that a state has been reached when*

$$\sqrt{n-1}\,\|L_r\| < \tfrac{1}{2}\delta, \tag{54}$$

where L_r *denotes the strictly lower part of* A_r*. Then the iterates* A_k *produced at the subsequent steps converge to the upper triangular form according to*

$$\|L_{r+s}\| \leqslant \frac{2\sqrt{s}}{\delta}\|L_r\|^2. \tag{55}$$

Again, no distinction has to be made here between forward and backward sweeps since both angles are identical at each step. Since the Schur form of a normal matrix is diagonal, Theorem 5.3 slightly extends a result of Ruhe [21] who proved the ultimate quadratic convergence of the Jacobi method for diagonalizing a normal matrix by the "optimal" procedure of Goldstine and Horwitz [10], i.e. by minimizing

$$|a_{ij}^{(k)}|^2 + |a_{ji}^{(k)}|^2$$

at each step k. The Schur method, while not optimal in this sense, exhibits the same convergence rate; moreover, the coefficient in the bound (55) is very close to Ruhe's.

(v) Theorems 5.2 and 5.3 are valid for pencils having distinct eigenvalues. No attempt was made to generalize it for pencils having multiple eigenvalues. Nevertheless, it can be conjectured (as Stewart did) that in the latter case ultimate quadratic convergence still holds, provided that the diagonal elements associated to the multiple eigenvalues occupy adjacent positions. Indeed, under this condition, proofs have been given for a number of Jacobi-like processes (e.g. the diagonalization of Hermitian matrices [25] and the singular value decomposition of triangular matrices [3]). In particular, the above-mentioned result of Ruhe [21] was originally stated also for this situation.

6. Numerical tests

We illustrate now the analysis presented in the previous sections. In particular, we focus on properties of the method that are not retrieved in the special case $B = I$ where it boils down to Stewart's one.

As in the standard case, convergence may stagnate when outer angles tend to 0. A typical example of this is

$$\lambda \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \end{bmatrix}. \tag{56}$$

It is easily seen that all angles of inner and outer rotations are here equal to 0 and hence that the

matrix does not change anymore. Yet the pencil is *normal* (B is the identity and A is unitary) and ultimate quadratic convergence will result, provided stagnation does not occur. Clearly the recommendation of Stewart to perform a sweep of *random* rotations applies here too. But such examples are pathological and usually are not encountered.

In the numerical examples detailed below, stagnation was unlikely to occur since a random generator was used to construct them. Following Theorem 2.1, we use the respective decompositions:

$$\lambda B_{\mathrm{n}} - A_{\mathrm{n}} = U(\lambda D_b - D_a)V^*,$$
$$\lambda B_{\mathrm{rn}} - A_{\mathrm{rn}} = U(\lambda D_b - D_a)TV^*,$$
$$\lambda B_{\mathrm{ln}} - A_{\mathrm{ln}} = UT(\lambda D_b - D_a)V^*,$$
$$\lambda B - A = U\left[\lambda D_b T - D_a(T + \alpha E)\right]V^*$$

for a normal pencil, right normal pencil, left normal pencil, and arbitrary pencil. Here D_a and D_b are random diagonal matrices, T is a random unit upper triangular matrix, V is a random unitary matrix, and U is a unitary matrix chosen such that B is upper triangular (A, of course, is full in general). Finally, E is a random strictly upper triangular matrix (with zero diagonal) which makes $\lambda B - A$ non-normal for any value of $\alpha \neq 0$. Only 10×10 real matrices are considered. By construction all these pencils clearly have real eigenvalues (namely the elements of $D_a D_b^{-1}$). All tests were performed on a VAX-3200 with relative precision $\epsilon \approx 1.4\text{E}{-17}$.

We first deal with convergence rate for pencils getting closer to a normal one. We apply the method of Section 3 to pencils $\lambda B - A = U[\lambda D_b - D_a(I + \alpha E)]V^*$ for several values of α, see Table 1.

Table 1

K	$\alpha = 1$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.001$	$\alpha = 0$
0	1.33E+00	6.88E–01	6.92E–01	6.94E–01	6.94E–01
1	4.52E–01	2.99E–01	3.00E–01	3.01E–01	3.01E–01
2	1.75E–01	1.40E–01	1.09E–01	1.03E–01	1.03E–01
3	9.02E–02	3.50E–02	2.02E–02	1.45E–02	1.40E–02
4	6.41E–02	6.34E–03	5.52E–04	2.01E–04	1.80E–04
5	5.04E–02	2.29E–03	7.28E–06	2.66E–07	4.45E–08
6	4.20E–02	4.85E–04	6.38E–08	2.98E–10	–
7	3.70E–02	1.61E–04	1.18E–09	5.34E–13	–
8	3.29E–02	5.78E–05	3.08E–13	–	–
9	3.02E–02	1.54E–05	7.83E–15	–	–
10	2.81E–02	8.86E–06	–	–	–
11	2.64E–02	2.14E–06	–	–	–
12	2.53E–02	1.37E–06	–	–	–
13	2.41E–02	3.24E–07	–	–	–
14	2.34E–02	2.12E–07	–	–	–
15	2.25E–02	4.97E–08	–	–	–
16	2.19E–02	3.27E–08	–	–	–
17	2.12E–02	7.66E–09	–	–	–
18	2.08E–02	5.05E–09	–	–	–
19	2.01E–02	1.17E–09	–	–	–
20	1.98E–02	7.77E–10	–	–	–

Table 2

$\lambda B_{rn} - A_{rn}$				$\lambda B_{ln} - A_{ln}$			
K	Forward	Backward	Alternate	K	Forward	Backward	Alternate
0	6.01E-01	6.01E-01	6.01E-01	0	4.77E-01	4.77E-01	4.77E-01
1	3.03E-01	2.60E-01	3.03E-01 (f)	1	2.21E-01	2.04E-01	2.21E-01 (f)
2	1.61E-01	1.49E-01	9.68E-02 (b)	2	1.15E-01	1.00E-01	6.67E-02 (b)
3	7.41E-02	9.54E-02	2.97E-02 (f)	3	8.41E-02	7.39E-02	3.44E-02 (f)
4	3.67E-02	7.06E-02	1.88E-02 (b)	4	7.16E-02	5.62E-02	7.58E-03 (b)
5	1.36E-02	5.38E-02	4.63E-03 (f)	5	6.29E-02	3.40E-02	4.09E-03 (f)
6	3.28E-03	4.53E-02	3.41E-03 (b)	6	5.50E-02	6.11E-03	7.85E-05 (b)
7	8.42E-05	3.80E-02	3.57E-05 (f)	7	5.01E-02	1.52E-04	4.37E-05 (f)
8	6.11E-08	3.31E-02	3.77E-06 (b)	8	4.59E-02	3.59E-08	9.83E-09 (b)
9	2.50E-14	2.85E-02	4.14E-11 (f)	9	4.22E-02	5.00E-15	1.55E-09 (f)
10	–	2.52E-02	2.17E-12 (b)	10	3.88E-02	–	–
11	–	2.20E-02	–	11	3.58E-02	–	–

In this example the gap between any two eigenvalues is $2\delta \approx 0.06$ and $\|B^{-1}\|_2 \approx 8$. The parameter K denotes the index of the sweep. Values smaller than 1.E-16 are left out as an indication of completed convergence.

The behavior is very similar to that of the standard case [24]. When a pencil is more distant from a normal pencil, one observes gradual deterioration of the quadratic convergence as was also reported in [24]. The convergence with $\alpha = 1$ is linear and very slow. For examples with a larger gap, a better convergence has been observed.

The second example involves two pencils $\lambda B_{rn} - A_{rn}$ and $\lambda B_{ln} - A_{ln}$. For each of these pencils we use three different methods: one involving only *forward* sweeps, one with only *backward*

Table 3

	K	Right normal	Left normal
	0	4.88E-01	1.54E-01
(f)	1	1.95E-02	6.99E-03
(b)	2	7.26E-03	3.19E-03
(f)	3	3.59E-03	1.89E-03
(b)	4	1.18E-03	1.29E-03
(f)	5	6.29E-04	8.56E-04
(b)	6	2.87E-04	6.17E-04
(f)	7	8.48E-05	5.27E-04
(b)	8	6.31E-05	6.27E-05
(f)	9	1.60E-05	3.84E-05
(b)	10	1.58E-05	6.08E-06
(f)	11	3.45E-06	5.02E-06
(b)	12	3.44E-06	2.13E-06
(f)	13	1.76E-07	1.82E-06
(b)	14	1.60E-07	7.57E-09
(f)	15	2.37E-10	6.70E-09
(b)	16	2.31E-10	9.40E-14
(f)	17	–	4.20E-14

Table 4

	K	Normal	Right normal	Left normal
	0	5.37E-01	5.04E-01	4.11E-01
(f)	1	2.36E-01	1.73E-01	1.81E-01
(b)	2	6.92E-02	5.40E-02	2.93E-02
(f)	3	1.08E-02	7.75E-03	9.31E-03
(b)	4	2.65E-04	1.38E-03	1.78E-03
(f)	5	7.29E-08	2.96E-04	3.90E-04
(b)	6	–	2.65E-04	4.94E-07
(f)	7	–	1.42E-06	9.21E-08
(b)	8	–	1.24E-06	3.00E-15
(f)	9	–	6.18E-12	–
(b)	10	–	3.15E-13	–

sweeps, and one where forward and backward sweeps *alternate* (i.e. the method we finally recommended in Section 3). The eigenvalues of the pencils are the same as in the previous example ($2\delta \approx 0.06$) and the "inverse norms" are $\|B_{\mathrm{rn}}^{-1}\|_2 \approx 11$ and $\|B_{\mathrm{ln}}^{-1}\|_2 \approx 15$, see Table 2.

One observes that quadratic convergence is indeed only obtained for *forward* sweeps in the *right normal* case and for *backward* sweeps in the *left normal* case. Notice that the alternate method converges in approximately the same number of sweeps although quadratic convergence occurs only every other sweep (the forward and the backward sweeps of the alternate method are marked in the last column). The convergence appears to be faster in the beginning of the process, which is not explained by our analysis but ties up with Stewart's remark that one double sweep seems to perform better than two forward or two backward sweeps.

In the third example we apply our method to two pencils $\lambda B_{\mathrm{rn}} - A_{\mathrm{rn}}$ and $\lambda B_{\mathrm{ln}} - A_{\mathrm{ln}}$ with large inverse norms $\|B_{\mathrm{rn}}^{-1}\|_2 \approx 1.\mathrm{E}+05$ and $\|B_{\mathrm{ln}}^{-1}\|_2 \approx 1.\mathrm{E}+05$, in order to check the convergence results of Theorem 5.2. The large inverse norms were obtained by using a badly conditioned T matrix. The gap is still $2\delta \approx 0.06$, see Table 3.

One observes here that quadratic convergence starts only around steps 14–15 (Theorem 5.2 guarantees that it occurs after $\|L_k\| < 1.\mathrm{E}{-}07$) and that it is significantly attenuated because of the factor $4\sqrt{S}\,\|B^{-1}\|_2/\delta$ (approximately $1.\mathrm{E}+08$ here).

The final example deals with close eigenvalues (not adjacent in the final form). We generated three pencils $\lambda B_{\mathrm{n}} - A_{\mathrm{n}}$, $\lambda B_{rn} - A_{rn}$, and $\lambda B_{\mathrm{ln}} - A_{\mathrm{ln}}$, all having the same eigenvalues. The gap is $2\delta \approx 0.0008$ and the inverse norms are $\|B_{\mathrm{n}}^{-1}\|_2 \approx 50$, $\|B_{\mathrm{rn}}^{-1}\|_2 \approx 190$, and $\|B_{\mathrm{ln}}^{-1}\|_2 \approx 180$; see Table 4.

These last two examples suggest that, while the condition (36) for quadratic convergence seems to reflect practical behavior, the coefficient of the bound (37) could be overestimated. In particular, further tests and possibly a deeper theoretical analysis are needed to estimate the exact influence of $\|B^{-1}\|_2$ and δ on the convergence rate.

7. Conclusion

We have presented and analyzed a Jacobi-like method for computing the generalized Schur decomposition of a regular pencil. To some extent, this work may seem to be academic. Nevertheless, its interest is (at least) threefold:

- It fills a gap. The obtention of the generalized Schur form by a Jacobi-like method is the only classical decomposition by unitary transformations that has not been investigated yet. Such methods have benefited from a renewed attention for a few years due to their high inherent parallelism. Moreover, to be complete, it is worthwhile to mention that a generalized eigenvalue decomposition algorithm for symmetric-definite pencils, using *non-unitary* elementary congruences, has been proposed by Falk and Langemeyer [5,6] and by Zimmermann [27].
- It generalizes and completes previous results. Our method extends the one of Stewart [24] from matrices to pencils. Also quantitative results are given for global and ultimate convergence which are valid for both the standard and the generalized cases, whereas Stewart's convergence results for the standard case are only qualitative. Interestingly, the bounds we obtain here are similar to those derived for various decompositions (e.g. [7,12,14,20,21,26]).
- A class of "normal" pencils is introduced as a natural extension of normal matrices. Our method shows ultimate quadratic convergence for these pencils in precisely the same manner as Stewart's method behaves for normal matrices.

A few questions remain unanswered:

- The ultimate convergence is proved to be quadratic for pencils having distinct eigenvalues, but only conjectured to be so for multiple or clustered eigenvalues, provided they are adjacent on the diagonal.
- The influence of B^{-1} and of the gap 2δ (see Theorem 5.2) on the convergence rate is observed up to some extent in our examples, but not completely understood.
- For (pencils of) real matrices, one could reformulate the method such that only real arithmetic is used. This then involves 4×4 real orthogonal transformations as basic operations of the method. Outer rotations have to be defined appropriately. Their computation requires the solution of either a 4×4 (generalized) eigenvalue problem, or a (set of) quadratic 2×2 matrix equation(s).

References

[1] R.P. Brent, F.T. Luk and C.F. Van Loan, Computation of the singular value decomposition using mesh-connected processors, *J. VLSI Comput. Systems* **1** (1984) 242–270.

[2] R.P. Brent and F.T. Luk, The solution of singular-value and symmetric eigenvalue problems on multiprocessor arrays, *SIAM J. Sci. Statist. Comput.* **6** (1985) 69–84.

[3] J.-P. Charlier and P. van Dooren, On Kogbetliantz's SVD algorithm in the presence of clusters, *Linear Algebra Appl.* **95** (1987) 135–160.

[4] P.J. Eberlein, On the Schur decomposition of a matrix for parallel computation, *IEEE Trans. Comput.* **36** (1987) 167–174.

[5] S. Falk and P. Langemeyer, Das Jacobische Rotationsverfahren für reellsymmetrische Matrizenpaare I, *Elektron. Datenverarbeitung* (1960) 30–34.

[6] S. Falk and P. Langemeyer, Das Jacobische Rotationsverfahren für reellsymmetrische Matrizenpaare Teil 2, *Elektron. Datenverarbeitung* (1960) 35–43.

[7] K.V. Fernando, Global convergence of the cyclic Kogbetliantz method, NAG Techn. Rept. TR6/86, Numerical Algorithms Group Ltd., Oxford, 1986.

[8] K.V. Fernando and S.J. Hammarling, A generalised singular value decomposition for a product of two matrices and balanced realisation, NAG Techn. Rept. TR1/87, Numerical Algorithms Group Ltd., Oxford, 1987.

[9] G. Forsythe and P. Henrici, The cyclic Jacobi method for computing the principal values of a complex matrix, *Trans. Amer. Math. Soc.* **94** (1960) 1–23.

[10] H.H. Goldstine and L.P. Horwitz, A procedure for the diagonalization of normal matrices, *J. ACM* **6** (1959) 176–195.
[11] G.H. Golub and C.F. Van Loan, *Matrix Computations* (North Oxford Academic, Oxford, 1983).
[12] V. Hari, On the convergence of cyclic Jacobi-like processes, *Linear Algebra Appl.* **81** (1986) 105–127.
[13] M. Heath, A. Laub, C. Paige and R. Ward, Computing the singular value decomposition of a product of matrices, *SIAM J. Sci. Statist. Comput.* 7 (1986) 1147–1159.
[14] P. Henrici and K. Zimmermann, An estimate for the norm of certain cyclic Jacobi operators, *Linear Algebra Appl.* **1** (1968) 489–501.
[15] E. Kogbetliantz, Diagonalization of general complex matrices as a new method for solution of linear equations, *Proc. Intern. Congr. Math., Vol. 2* (Amsterdam, 1954) 356–357.
[16] E. Kogbetliantz, Solution of linear equations by diagonalization of coefficient matrices, *Quart. Appl. Math.* **13** (1955) 123–132.
[17] F.T. Luk, A rotation method for computing the QR-decomposition, *SIAM J. Sci. Statist. Comput.* **7** (1986) 452–459.
[18] C.B. Moler and G.W. Stewart, An algorithm for generalized matrix eigenvalue problems, *SIAM J. Numer. Anal.* **10** (1973) 241–256.
[19] C.C. Paige, Computing the generalized singular value decomposition, *SIAM J. Sci. Statist. Comput.* **7** (1986) 1126–1146.
[20] C.C. Paige and P. van Dooren, On the quadratic convergence of Kogbetliantz's algorithm for computing the singular value decomposition, *Linear Algebra Appl.* 77 (1986) 301–313.
[21] A. Ruhe, On the quadratic convergence of the Jacobi method for normal matrices, *BIT* **7** (1967) 305–313.
[22] A. Ruhe, Closest normal matrix finally found!, *BIT* **27** (1987) 585–598.
[23] G.W. Stewart, Error and perturbation bounds for subspaces associated with certain eigenvalue problems, *SIAM Rev.* **15** (1973) 727–764.
[24] G.W. Stewart, A. Jacobi-like algorithm for computing the Schur decomposition of a nonhermitian matrix, *SIAM J. Sci. Statist. Comput.* **6** (1985) 853–864.
[25] H.P.M. van Kempen, On the quadratic convergence of the special cyclic Jacobi method, *Numer. Math.* **9** (1966) 19–22.
[26] J.H. Wilkinson, Note on the quadratic convergence of the cyclic Jacobi process, *Numer. Math.* **4** (1962) 296–300.
[27] K. Zimmermann, Zur Konvergenz eines Jacobiverfahren für gewöhnliche und verallgemeinerte Eigenwertprobleme, Dissertation Nr. 4305, Eidgenössischen Technischen Hochschule Zürich, 1969.