A GLOBAL CONVERGENCE PROOF FOR CYCLIC JACOBI METHODS WITH BLOCK ROTATIONS

ZLATKO DRMAČ

ABSTRACT. This paper introduces a globally convergent block (column– and row–) cyclic Jacobi method for diagonalization of Hermitian matrices and for computation of the singular value decomposition of general matrices. It is shown that a block rotation (generalization of the Jacobi's 2×2 rotation) must be computed and implemented in a particular way to guarantee global convergence. This solves a long standing open problem of convergence of block cyclic Jacobi methods. The proof includes the convergence of the eigenspaces in the general case of multiple eigenvalues.

1. INTRODUCTION AND PRELIMINARIES

State of the art accurate methods for computing the singular value decomposition (SVD) and symmetric (Hermitian) spectral decomposition are based on the classical Jacobi algorithm [21]. The algorithm starts with a Hermitian $\mathsf{H} = \mathsf{H}^{(1)} \in \mathbb{C}^{n \times n}$ and then it generates a sequence of unitary congruences, $\mathsf{H}^{(k+1)} = (\mathsf{U}^{(k)})^* \mathsf{H}^{(k)} \mathsf{U}^{(k)}$, where $\mathsf{U}^{(k)}$ is a plane rotation, i.e. $\mathsf{U}^{(k)}$ differs from the identity only at the cleverly chosen positions (i_k, i_k), (i_k, j_k), (j_k, i_k), (j_k, j_k), where $\begin{pmatrix} \mathsf{U}^{(k)}_{i_k, i_k} & \mathsf{U}^{(k)}_{i_k, j_k} \\ \mathsf{U}^{(k)}_{j_k, i_k} & \mathsf{U}^{(k)}_{j_k, j_k} \end{pmatrix} = \begin{pmatrix} \cos \phi_k & e^{i\psi_k} \sin \phi_k \\ -e^{-i\psi_k} \sin \phi_k & \cos \phi_k \end{pmatrix}$. The angles ϕ_k, ψ_k are determined to annihilate the (i_k, j_k) and (j_k, i_k) positions in $\mathsf{H}^{(k)}$,

$$(1.1) \qquad \begin{pmatrix} \cos \phi_k & -e^{i\psi_k} \sin \phi_k \\ e^{-i\psi_k} \sin \phi_k & \cos \phi_k \end{pmatrix} \begin{pmatrix} \mathsf{H}^{(k)}_{i_k i_k} & \mathsf{H}^{(k)}_{i_k j_k} \\ \mathsf{H}^{(k)}_{j_k i_k} & \mathsf{H}^{(k)}_{j_k j_k} \end{pmatrix} \begin{pmatrix} \cos \phi_k & e^{i\psi_k} \sin \phi_k \\ -e^{-i\psi_k} \sin \phi_k & \cos \phi_k \end{pmatrix} = \begin{pmatrix} \mathsf{H}^{(k+1)}_{i_k i_k} & 0 \\ 0 & \mathsf{H}^{(k+1)}_{j_k j_k} \end{pmatrix}.$$

With proper choices of the rotation angles, and under certain pivot strategies $k \mapsto (i_k, j_k)$, the matrices $\mathsf{H}^{(k)}$ converge to a diagonal matrix with eigenvalues of H along the diagonal. By implicit diagonalization of $\mathsf{H} = \mathsf{A}^* \mathsf{A}$, the algorithm can compute the SVD of a general matrix A, see [20].

An important advantage of the Jacobi algorithm is that it is more accurate than any other method that first tridiagonalizes (or bidiagonalizes in the case of SVD computation) the matrix [8], and that it can be used for highly accurate eigenvalue and singular value computation of special classes of matrices [7], [6]. Recent implementation of the Jacobi SVD algorithm [13], [14] shows that the method has the potential to achieve efficiency comparable to the fast bidiagonalization–based methods. Nontrivial modifications of the algorithm include the use of QR iterations as a preconditioner, and a specially tailored version of the implicit Jacobi algorithm for structured triangular matrices.

Future improvements of the method for large scale dense full Hermitian eigenvalue and SVD computations require *(i)* further study of the numerical properties and of the convergence of Jacobi–type iterative processes in order to find better preconditioning and better pivot strategies; *(ii)* development of parallel pivot strategies that map well to modern high performance parallel machines;

(iii) new design of the kernel routine – plane rotation has to be replaced with a block transformation that diagonalizes pivot submatrices of sizes larger than 2×2.

In this report, we give a theoretical framework for *(iii)* and prove global convergence for a class of Jacobi algorithms with block transformations. For a partition $\nu = (n_1, n_2, \ldots, n_m)$ $(m \geq 2, n_i > 0,$ $\sum_{i=1}^{m} n_i = n)$ we introduce a block partition $\mathsf{H} = (\mathsf{H}_{[ij]})_{i,j=1}^{m}$, where $\mathsf{H}_{[ij]}$ is $n_i \times n_j$. One step of the block–Jacobi method consists of choosing a pivot pair (i_k, j_k), and generalizing (1.1) to the doagonalization of $(n_{i_k} + n_{j_k}) \times (n_{i_k} + n_{j_k})$ pivot submatrix:

$$(1.2) \quad \begin{pmatrix} \mathsf{H}_{[i_k i_k]}^{(k+1)} & \mathsf{H}_{[i_k j_k]}^{(k+1)} \\ \mathsf{H}_{[j_k i_k]}^{(k+1)} & \mathsf{H}_{[j_k j_k]}^{(k+1)} \end{pmatrix} = \begin{pmatrix} (\mathsf{U}_{[i_k i_k]}^{(k)})^* & (\mathsf{U}_{[j_k i_k]}^{(k)})^* \\ (\mathsf{U}_{[i_k j_k]}^{(k)})^* & (\mathsf{U}_{[j_k j_k]}^{(k)})^* \end{pmatrix} \underbrace{\begin{pmatrix} \mathsf{H}_{[i_k i_k]}^{(k)} & \mathsf{H}_{[i_k j_k]}^{(k)} \\ \mathsf{H}_{[j_k i_k]}^{(k)} & \mathsf{H}_{[j_k j_k]}^{(k)} \end{pmatrix}}_{\mathcal{H}_{i_k j_k}^{(k)}} \underbrace{\begin{pmatrix} \mathsf{U}_{[i_k i_k]}^{(k)} & \mathsf{U}_{[i_k j_k]}^{(k)} \\ \mathsf{U}_{[j_k i_k]}^{(k)} & \mathsf{U}_{[j_k j_k]}^{(k)} \end{pmatrix}}_{\mathcal{U}_{i_k j_k}^{(k)}},$$

$$(1.3) \quad \mathsf{H}^{(k+1)} = \mathsf{U}^{(k)^*} \mathsf{H}^{(k)} \mathsf{U}^{(k)}, \quad \mathsf{U}^{(k)} = \begin{pmatrix} \mathsf{I} & 0 & 0 & 0 & 0 \\ 0 & \mathsf{U}_{[i_k i_k]}^{(k)} & 0 & \mathsf{U}_{[i_k j_k]}^{(k)} & 0 \\ 0 & 0 & \mathsf{I} & 0 & 0 \\ 0 & \mathsf{U}_{[j_k i_k]}^{(k)} & 0 & \mathsf{U}_{[j_k j_k]}^{(k)} & 0 \\ 0 & 0 & 0 & 0 & \mathsf{I} \end{pmatrix}.$$

We write the relation (1.2) as $\mathcal{H}_{i_k j_k}^{(k+1)} = (\mathcal{U}_{i_k j_k}^{(k)})^* \mathcal{H}_{i_k j_k}^{(k)} \mathcal{U}_{i_k j_k}^{(k)}$, where $\mathcal{H}_{i_k j_k}^{(k+1)}$ is diagonal, and the diagonalization procedure (1.2) is arbitrary. If the steps (1.2, 1.3) are repeated by taking the pivot strategy $k \mapsto (i_k, j_k)$ to be periodic mapping with one full period of length $\mathfrak{s} = (m-1)m/2$ (also called a *sweep*) given as

$$(1.4) \quad (1,2), (1,3), (2,3), (1,4), (2,4), (3,4), \ldots, (1,m), (2,m), \ldots, (m-1,m),$$

then we have a *block column–cyclic Jacobi method*. In a similar way, one defines a *block row–cyclic Jacobi method*, with pivot sequence $(1,2), (1,3), (1,4), \ldots, (1,m), (2,3), (2,4), \ldots, (2,m), \ldots,$ $(m-1,m)$.

In an SVD computation by one sided Jacobi algorithm, the initial matrix $\mathsf{A} \in \mathbb{C}^{r \times n}$ is partitioned in block columns as $\mathsf{A} = (\mathsf{A}_{[1]}, \ldots, \mathsf{A}_{[m]})$. In the k–th step, the matrix $\mathcal{H}_{i_k j_k}^{(k)}$ is computed as $\mathcal{H}_{i_k j_k}^{(k)} = (\mathsf{A}_{[i_k]}^{(k)}, \mathsf{A}_{[j_k]}^{(k)})^* (\mathsf{A}_{[i_k]}^{(k)}, \mathsf{A}_{[j_k]}^{(k)})$, and the transformation (1.3) is executed implicitly as

$$(1.5) \quad \mathsf{A}^{(k+1)} = \mathsf{A}^{(k)} \mathsf{U}^{(k)}.$$

Block transformations (1.3), (1.5) generate different diagonalization processes with more potential for fast convergence. Further, they are more suitable for hierarchical memory architectures because of higher *flop* to memory reference ratio, and are preferable as kernel routines in both serial and parallel environments. Thus, the update of the matrix $\mathsf{H}^{(k)}$ ($\mathsf{A}^{(k)}$) to obtain $\mathsf{H}^{(k+1)}$ ($\mathsf{A}^{(k+1)}$) can be efficiently executed by machine optimized BLAS 3 operations [10], [9].

Here we make an important distinction between the proper block methods, as described above, and the block–oriented methods. In the block–oriented methods, one uses the block partition and the pivoting (1.4), and inside each block $\mathcal{H}_{i_k j_k}^{(k)}$ one applies a sequence of 2×2 Jacobi rotations (e.g. by scanning through $\mathcal{H}_{i_k j_k}^{(k)}$ in a row–wise fashion) and accumulate them in the matrix $\mathcal{U}_{i_k j_k}^{(k)}$. In this way, several 2×2 rotations are applied as a block transformation. This is actually just a modification of the nested loops that control the sequence of 2×2 rotations, with a delayed application of individual rotations to the rest of the current iterate. The convergence of such a method is obtained by proving its equivalence (in the sense of equivalence relation between pivot strategies [17]) to some other convergent strategy with 2×2 elementary transformations (1.1), see [1], [19].

For proper block methods with the pivoting (1.4), we are not aware of any convergence result in the literature. And, in fact, we show that the convergence cannot be guaranteed unless the block transformations are implemented as described in this report. We define a new class of unitary

block–transformations and provide a novel technique to study the convergence to diagonal matrix in a framework of the polyhedral set of all diagonals of the matrices from the adjoint orbit of the given Hermitian H. Our proof of global convergence covers the general case of multiple eigenvalues, including the convergence of eigenvectors and eigenspaces, and it is to our best knowledge the first result of this kind in the theory of Jacobi methods. The convergence of the eigenspaces is obtained in a Grassmann manifold setting, by using perturbation theory, in particular the $\sin\Theta$ theorem.

2. The off–norm converges to zero

Convergence analysis of a block method driven by a pivot strategy is in many aspects similar to the classical 2×2 rotation based methods. To measure the distance to diagonal matrices, we use the so called off–norm: for $\mathsf{X} \in \mathbb{C}^{n \times n}$ the off–norm is defined as $\Omega(\mathsf{X}) = \sqrt{\sum_{i \neq j} |\mathsf{X}_{ij}|^2}$. So, our first goal is to show that $\lim_{k \to \infty} \Omega(\mathsf{H}^{(k)}) = 0$.

As in the 2×2 case, the diagonalization requirement in (1.2) can be relaxed to mere off–norm reduction by a constant factor ρ, independent of k, i.e.

$$(2.1) \qquad \Omega(\mathcal{H}_{i_k j_k}^{(k+1)}) \leq \rho\, \Omega(\mathcal{H}_{i_k j_k}^{(k)}), \;\; \rho \in [0, 1),$$

where the choice $\rho = 0$ means exact diagonalization. At each step we have

$$(2.2) \qquad \Omega^2(\mathsf{H}^{(k+1)}) \leq \Omega^2(\mathsf{H}^{(k)}) - (1 - \rho)\Omega^2(\mathcal{H}_{i_k j_k}^{(k)}) \leq \Omega^2(\mathsf{H}^{(k)}).$$

Another possibility is to use a threshold strategy: for a decreasing sequence of positive threshold values $\tau_1 > \tau_2 > \cdots$ with $\lim_{t \to \infty} \tau_t = 0$, the transformation in the k–th step is executed only if $\Omega(\mathcal{H}_{i_k j_k}^{(k)}) \geq \tau_1$, else it is skipped. After skipping one full cycle $((m-1)m/2)$ consecutive steps), the threshold is lowered to τ_2 etc. From (2.2), it follows that reduction below a given threshold is always obtained in a finite number of steps. (Cf. [27].)

The following technical lemma is common for all pivot strategies:

Lemma 2.1. *With any pivoting strategy and with transformations (1.2) that satisfy (2.1), the sequence $(\Omega(\mathsf{H}^{(k)}))_{k=1}^{\infty}$ is monotonically decreasing with the limit $\omega = \lim_{k \to \infty} \Omega(\mathsf{H}^{(k)})$. Further, the series of the squared off–norms of the pivot submatrices is convergent, $\sum_{k=1}^{\infty} \Omega^2(\mathcal{H}_{i_k j_k}^{(k)}) < \infty$. Hence, the off–diagonal entries of the pivot submatrices are always converging toward zero, $\lim_{k \to \infty} \Omega(\mathcal{H}_{i_k j_k}^{(k)}) = 0$.*

Proof: The proof follows from (2.2) and

$$\sum_{p=1}^{k} \Omega^2(\mathcal{H}_{i_p j_p}^{(p)}) \leq \frac{1}{1 - \rho}(\Omega^2(\mathsf{H}) - \Omega^2(\mathsf{H}^{(k+1)})). \qquad \boxplus$$

In this paper we will be mainly interested in (2.1) with $\rho = 0$, and for the sake of brevity we will just give hints how to proceed for $\rho \in [0, 1)$. The pivoting will be the block column–cyclic (1.4).

2.1. Uniformly bounded cosines: UBC transformations.

It is well–known that a necessary condition for the convergence of cyclic Jacobi methods is the existence of a strictly positive uniform lower bound for the cosines of the Jacobi angles [16]. How this translates to the block methods can be seen from the cosine–sine decomposition (CSD) of the matrix $\mathcal{U}_{i_k j_k}^{(k)}$ in (1.2), e.g. in the case

$n_{i_k} \geq n_{j_k}$:

$$(2.3) \qquad \begin{pmatrix} \mathsf{U}^{(k)}_{[i_k i_k]} & \mathsf{U}^{(k)}_{[i_k j_k]} \\ \mathsf{U}^{(k)}_{[j_k i_k]} & \mathsf{U}^{(k)}_{[j_k j_k]} \end{pmatrix} = \begin{pmatrix} \mathsf{Z}^{(k)}_1 & 0 \\ 0 & \mathsf{Z}^{(k)}_2 \end{pmatrix} \begin{pmatrix} \mathsf{C}^{(k)} & \mathsf{S}^{(k)} \\ -\mathsf{S}^{(k)*} & \cos\Phi^{(k)} \end{pmatrix} \begin{pmatrix} \mathsf{W}^{(k)}_1 & 0 \\ 0 & \mathsf{W}^{(k)}_2 \end{pmatrix}$$

$$(2.4) \qquad \mathsf{C}^{(k)} = \begin{pmatrix} \cos\Phi^{(k)} & 0 \\ 0 & \mathsf{I}_{n_{i_k}-n_{j_k}} \end{pmatrix}, \ \mathsf{S}^{(k)} = \begin{pmatrix} \sin\Phi^{(k)} \\ 0 \end{pmatrix},$$

where the zero block in $\mathsf{S}^{(k)}$ is $(n_{i_k} - n_{j_k}) \times n_{j_k}$, $\cos\Phi^{(k)} = \mathrm{diag}(\cos\phi^{(k)}_p)^{n_{j_k}}_{p=1}$, $\cos\phi^{(k)}_p \geq 0$, and $\sin\Phi^{(k)} = \mathrm{diag}(\sin\phi^{(k)}_p)^{n_{j_k}}_{p=1}$. From the factored representation (2.3) we see that n_{j_k} non–overlapping rotations $\begin{pmatrix} \cos\phi^{(k)}_p & \sin\phi^{(k)}_p \\ -\sin\phi^{(k)}_p & \cos\phi^{(k)}_p \end{pmatrix}$ simultaneously combine the columns from the i_k–th and the j_k–th block. Using an analogy with the 2×2 rotations we can infer that a lower bound for all $\cos\phi^{(k)}_p$ will be a necessary condition for the convergence.

The following lemma is the key result that allows us to analyze the convergence of $\Omega(\mathsf{H}^{(k)})$ towards zero in complete analogy to the classical proof due to Forsythe and Henrici [16] for the Jacobi method with 2×2 rotations.

Lemma 2.2. *Let* U *be an arbitrary* $m \times m$ *unitary matrix. Then for any partition* $m = b + \ell$, $b, \ell \in \{1, \ldots, m-1\}$ *there exists a permutation matrix* Π *such that in the block–matrix*

$$(2.5) \qquad \hat{\mathsf{U}} \equiv \mathsf{U}\Pi = \begin{pmatrix} \hat{\mathsf{U}}_{[11]} & \hat{\mathsf{U}}_{[12]} \\ \hat{\mathsf{U}}_{[21]} & \hat{\mathsf{U}}_{[22]} \end{pmatrix}, \ \hat{\mathsf{U}}_{[11]} \in \mathbb{C}^{b \times b},$$

both diagonal blocks are non–singular with

$$(2.6) \qquad 1 \geq \sigma_{\min}(\hat{\mathsf{U}}_{[11]}) = \sigma_{\min}(\hat{\mathsf{U}}_{[22]}) \geq f(b, \ell) > 0,$$

where $f(b, \ell)$ *depends solely on* b *and* ℓ.

Proof: Let U be partitioned as

$$(2.7) \qquad \mathsf{U} = \begin{pmatrix} \mathsf{U}_{[11]} & \mathsf{U}_{[12]} \\ \mathsf{U}_{[21]} & \mathsf{U}_{[22]} \end{pmatrix}, \ \mathsf{U}_{[11]} \in \mathbb{C}^{b \times b},$$

and consider the Businger–Golub column pivoted QR factorization [2] of the first block–row $\mathsf{U}_{[1:]} = \begin{pmatrix} \mathsf{U}_{[11]} & \mathsf{U}_{[12]} \end{pmatrix}$,

$$\mathsf{U}_{[1:]}\Pi = \mathsf{QR}, \ \mathsf{Q} \ b \times b \ \text{unitary}, \ \mathsf{R} \ b \times m \ \text{upper trapezoidal}.$$

We claim that the permutation Π satisfies (2.5,2.6).
If $\hat{\mathsf{U}} = \mathsf{U}\Pi$ is partitioned as in (2.5), and if we partition R as $\mathsf{R} = \begin{pmatrix} \mathsf{T} & \mathsf{K} \end{pmatrix}$ with $b \times b$ upper triangular T, then $\hat{\mathsf{U}}_{[11]} = \mathsf{QT}$, and the problem reduces to finding a lower bound for the minimal singular value of T. As a result of column pivoting, the matrix T has special diagonal dominance structure:

$$(2.8) \qquad |\mathsf{T}_{ii}|^2 \ \geq \ \sum_{j=i}^{k} |\mathsf{T}_{jk}|^2, \ i \leq k \leq b,$$

$$(2.9) \qquad |T_{bb}| \ = \ \max_{j=b:m} |\mathsf{R}_{bj}|.$$

On the other hand, since $\hat{\mathsf{U}}_{[1:]} \equiv \begin{pmatrix} \hat{\mathsf{U}}_{[11]} & \hat{\mathsf{U}}_{[12]} \end{pmatrix} = \mathsf{QR}$ and since $\hat{\mathsf{U}}_{[1:]}\hat{\mathsf{U}}^*_{[1:]} = \mathsf{RR}^* = \mathsf{I}_b$, it must hold that

$$(2.10) \qquad |\mathsf{T}_{bb}| \geq \frac{1}{\sqrt{\ell+1}}.$$

If we set $\mathsf{D} = \mathrm{diag}(\mathsf{T}_{ii})_{i=1}^{b}$, $\check{\mathsf{T}} = \mathsf{D}^{-1}\mathsf{T}$, then

$$\sigma_{\min}(\mathsf{T}) = \min_{\mathsf{x}\neq 0} \frac{\|\mathsf{T}\mathsf{x}\|_2}{\|\mathsf{x}\|_2} = \min_{\mathsf{y}\neq 0} \frac{\|\mathsf{D}\mathsf{y}\|_2}{\|\check{\mathsf{T}}^{-1}\mathsf{y}\|_2} \geq \frac{\sigma_{\min}(\mathsf{D})}{\|\check{\mathsf{T}}^{-1}\|_2} \geq \frac{1}{\sqrt{\ell+1}} \frac{1}{\|\check{\mathsf{T}}^{-1}\|_2}.$$

It remains to estimate the norm of $\check{\mathsf{T}}^{-1} = \mathsf{T}^{-1}\mathsf{D}$. Using [22], we know that, for $i = 2, \ldots, b$,

$$|\mathsf{T}^{-1}\mathsf{e}_i| \leq \frac{1}{|\mathsf{T}_{ii}|} \begin{pmatrix} 2^{i-2}, & 2^{i-3}, & \ldots & 4, & 2, & 1, & 1, & 0, & \ldots & 0 \end{pmatrix}^T,$$

where e_i is the i–th column of the identity, and the absolute value and the inequality between vectors are understood element–wise. For $i = 1$ we trivially have $\mathsf{T}^{-1}\mathsf{e}_1 = \frac{1}{\mathsf{T}_{11}}\mathsf{e}_1$, and $\check{\mathsf{T}}^{-1}\mathsf{e}_1 = \mathsf{e}_1$. For $i = 2, \ldots, b$ we use the relations

$$\check{\mathsf{T}}^{-1}\mathsf{e}_i = \mathsf{T}^{-1}\mathsf{D}\mathsf{e}_i = \mathsf{T}_{ii}\mathsf{T}^{-1}\mathsf{e}_i$$

to conclude $|\check{\mathsf{T}}^{-1}\mathsf{e}_i| \leq \begin{pmatrix} 2^{i-2}, & 2^{i-3}, & \ldots & 4, & 2, & 1, & 1, & 0, & \ldots & 0 \end{pmatrix}^T$, and thus

$$\|\check{\mathsf{T}}^{-1}\|_2 \leq \|\check{\mathsf{T}}^{-1}\|_F \leq g(b), \quad \text{where} \quad g(b) = \sqrt{b + \sum_{i=2}^{b} \sum_{j=0}^{i-2} 4^j}.$$

Finally, $\sigma_{\min}(\hat{\mathsf{U}}_{[11]}) = \sigma_{\min}(\mathsf{T}) \geq \dfrac{1}{g(b)\sqrt{\ell+1}} \equiv f(b, \ell)$, as claimed. $\boxplus$

Corollary 2.1. *Let a block partition of H be given by $\nu = (n_1, n_2, \ldots, n_m)$. Then in any pivot strategy, one can choose the block–orthogonal transformations (1.2) so that*

$$\inf_{k \geq 1} \sigma_{\min}(\mathsf{U}_{[i_k i_k]}^{(k)}) \geq \beta_\nu = \min_{i < j} f(n_i, n_j) > 0,$$

where $f(\cdot, \cdot)$ is from Lemma 2.2. If an implementation of the Jacobi methods allows changing the partition, then the lower bound β_ν can be replaced by the minimum over all (finitely many) partitions of n.

Definition 2.2. *A class of unitary transformations with given 2×2 block partition is called a UBC (Uniformly Bounded Cosines) transformation, if the singular values of the diagonal blocks can be bounded from below by a function of the dimension.*

Remark 2.1. It is possible that more sophisticated (and more expensive in practical computation) pivoting would give better lower bound for $\sigma_{\min}(\hat{\mathsf{U}}_{[11]})$, but for our purposes of proving the convergence the uniform bound from Lemma 2.2 will suffice.

Lemma 2.3. *Let the unitary matrix U be partitioned as in (2.7) and let*

$$(2.11) \qquad \begin{pmatrix} \mathsf{U}_{[11]} & \mathsf{U}_{[12]} \\ \mathsf{U}_{[21]} & \mathsf{U}_{[22]} \end{pmatrix} \begin{pmatrix} \mathsf{X} \\ \mathsf{Y} \end{pmatrix} = \begin{pmatrix} \tilde{\mathsf{X}} \\ \tilde{\mathsf{Y}} \end{pmatrix}$$

where X, Y are matrices of appropriate dimensions and $\|\mathsf{X}\|_F \leq \varepsilon$, $\|\tilde{\mathsf{Y}}\|_F \leq \varepsilon$. If the diagonal blocks $\mathsf{U}_{[11]}$, $\mathsf{U}_{[22]}$ are nonsingular, then

$$(2.12) \qquad \max\{\|\tilde{\mathsf{X}}\|_F, \|\mathsf{Y}\|_F\} \leq 2\varepsilon \|\mathsf{U}_{[11]}^{-1}\|_2.$$

Proof: We note that

$$\begin{pmatrix} \tilde{\mathsf{X}} \\ \mathsf{Y} \end{pmatrix} = \begin{pmatrix} \mathsf{I} & -\mathsf{U}_{[12]} \\ 0 & \mathsf{U}_{[22]} \end{pmatrix}^{-1} \begin{pmatrix} \mathsf{U}_{[11]} & 0 \\ -\mathsf{U}_{[21]} & \mathsf{I} \end{pmatrix} \begin{pmatrix} \mathsf{X} \\ \tilde{\mathsf{Y}} \end{pmatrix} = \begin{pmatrix} \mathsf{U}_{[11]} - \mathsf{U}_{[12]}\mathsf{U}_{[22]}^{-1}\mathsf{U}_{[21]} & \mathsf{U}_{[12]}\mathsf{U}_{[22]}^{-1} \\ -\mathsf{U}_{[22]}^{-1}\mathsf{U}_{[21]} & \mathsf{U}_{[22]}^{-1} \end{pmatrix} \begin{pmatrix} \mathsf{X} \\ \tilde{\mathsf{Y}} \end{pmatrix},$$

where in fact $\mathsf{U}_{[11]} - \mathsf{U}_{[12]}\mathsf{U}_{[22]}^{-1}\mathsf{U}_{[21]} = \mathsf{U}_{[11]}^{-*}$. $\boxplus$

2.2. **Block column cyclic off–norm reduction.** If one is familiar with the classical convergence proof for the cyclic Jacobi method due to Forsythe and Henrici [16], then one can use the results from §2.1 and complete the proof that the off–norm in the block cyclic case converges to zero. However, for the sake of completeness, we include the detailed proof, following the ideas from [16].

Let us first introduce some auxiliary notation. We use $S_{ij}^{(k)}$ to denote the off norm of the submatrix $H_{[i:j,i:j]}^{(k)}$ of $H^{(k)}$ composed of the blocks $H_{[pq]}^{(k)}$, $i \leq p, q \leq j$. Since in one cycle the pivot strategy $k \mapsto \mathfrak{p}(k) = (i_k, j_k)$ is bijective, it is sometimes convenient to use its inverse $\mathfrak{K}$, $(i,j) \mapsto k_{ij} = \mathfrak{K}(i,j)$, to match a pivot position with its iteration number.

Lemma 2.4. *Consider the column–cyclic pivoting (1.4), with the block–partition ν. Let $k_0 \in \mathbb{N}$ be chosen so that for all indices k, $k \geq k_0$ implies $\Omega(\mathcal{H}_{i_k,j_k}^{(k)}) < \varepsilon$. (See Lemma 2.1.) If for some $k_{ij} \geq k_0$, $S_{ij}^{(k_{ij})} < \varepsilon$ and $S_{i+1,j+1}^{(k_{i+1,j+1})} < \varepsilon$, then $S_{i,j+1}^{(k_{i,j+1})} < \alpha_{j-i}\varepsilon$, where α_{j-i} depends only on the difference $j - i$ and the partition ν, and can be uniformly bounded by a function of n.*

Proof: Note that in the column–cyclic strategy the indices of the transformations under consideration satisfy $k_{ij} < k_{i,j+1} < k_{i+1,j+1} = k_{i,j+1} + 1$. The task is to estimate the changes of $S_{ij}^{(k)}$ for transformation indices $k_{ij} \leq k \leq k_{i,j+1}$. We first note that

$$(2.13) \qquad \varepsilon > S_{ij}^{(k_{ij})} \geq S_{ij}^{(k_{i+1,j})} \geq \cdots \geq S_{ij}^{(k_{j-1,j})} \geq S_{ij}^{(k_{1,j+1})}$$

because the transformations with indices from k_{ij} to $k_{j-1,j}$ can be taken as the block Jacobi transformations of $H_{[i:j,i:j]}^{(k_{ij})}$, thus off–norm reducing. Further, since the transformations with indices $k_{1,j+1}, \ldots, k_{i-1,j+1}$ do not change $H_{[i:j,i:j]}^{(k_{1,j+1})}$, it holds that

$$(2.14) \qquad S_{ij}^{(k_{1,j+1})} = S_{ij}^{(k_{2,j+1})} = \ldots = S_{ij}^{(k_{i-1,j+1})} = S_{ij}^{(k_{i,j+1})}.$$

Now consider the moment immediately before the transformation at the pivot position $(i, j+1)$. Since $\Omega(\mathcal{H}_{i,j+1}^{(k_{i,j+1})}) < \varepsilon$ (as pivot submatrix, and because $k_{i,j+1} > k_0$), it also holds that $\|H_{[i,j+1]}^{(k_{i,j+1})}\|_F < \varepsilon$. After this transformation at $(i, j + 1)$ we have, using the assumption $S_{i+1,j+1}^{(k_{i+1,j+1})} < \varepsilon$,

$$(2.15) \qquad \|H_{[j+1,i+1]}^{(k_{i+1,j+1})}\|_F < \varepsilon, \;\; \|H_{[j+1,i+2]}^{(k_{i+1,j+1})}\|_F < \varepsilon, \; \ldots, \; \|H_{[j+1,j]}^{(k_{i+1,j+1})}\|_F < \varepsilon.$$

On the other hand, the blocks $H_{[j+1,q]}^{(k_{i+1,j+1})}$, $q = i + 1, \ldots, j$, are computed as

$$(2.16) \qquad \begin{pmatrix} H_{[iq]}^{(k_{i+1,j+1})} \\ H_{[j+1,q]}^{(k_{i+1,j+1})} \end{pmatrix} = \begin{pmatrix} U_{[ii]}^{(k_{i,j+1})} & U_{[i,j+1]}^{(k_{i,j+1})} \\ U_{[j+1,i]}^{(k_{i,j+1})} & U_{[j+1,j+1]}^{(k_{i,j+1})} \end{pmatrix} \begin{pmatrix} H_{[iq]}^{(k_{i,j+1})} \\ H_{[j+1,q]}^{(k_{i,j+1})} \end{pmatrix}, \;\; q = i + 1, \ldots, j.$$

From (2.13) and (2.14) we conclude that $\max_{q=i+1:j} \|H_{[iq]}^{(k_{i,j+1})}\|_F < \varepsilon$, and then, using (2.14), (2.15) and Lemma 2.3, we obtain $\max_{q=i+1:j} \|H_{[j+1,q]}^{(k_{i,j+1})}\|_F \leq 2\varepsilon/\beta_\nu$, where β_ν is from Corollary 2.1. Hence, putting it all together, we obtain

$$\begin{aligned} (S_{i,j+1}^{(k_{i,j+1})})^2 &\leq (S_{i,j}^{(k_{i,j+1})})^2 + \Omega(\mathcal{H}_{i,j+1}^{(k_{i,j+1})})^2 + 2 \sum_{q=i+1}^{j} \|H_{[j+1,q]}^{(k_{i,j+1})}\|_F^2 \\ &\leq (2 + 8(j-i)/\beta_\nu^2)\varepsilon^2 \equiv \alpha_{j-i}^2 \varepsilon^2. \qquad\qquad \boxplus \end{aligned}$$

Theorem 2.3. *For any Hermitian $H \in \mathbb{C}^{n \times n}$ with a block partition $\nu = (n_1, \ldots, n_m)$, the block Jacobi method with column–cyclic ordering generates a sequence $(H^{(k)})_{k=1}^{\infty}$ with $\lim_{k \to \infty} \Omega(H^{(k)}) = 0$.*

Proof: Let $\varepsilon_+ > 0$ be arbitrary. Define $\varepsilon = \varepsilon_+ / \prod_{p=1}^{m-2} \alpha_p$, where the α_p's are as in Lemma 2.4. Using Lemma 2.1 we can determine an index k_0 such that $\Omega(\mathcal{H}_{i_k j_k}^{(k)}) < \varepsilon$, for all $k \geq k_0$. This means that after the index k_0 the off norm of each pivot submatrix is ε small. For instance, in the case of a 6×6 partition, and without loss of generality assuming $k_0 = 1$, we have $\Omega(\mathcal{H}_{12}^{(1)}) = S_{12}^{(1)} < \varepsilon$, $\Omega(\mathcal{H}_{23}^{(3)}) = S_{23}^{(3)} < \varepsilon$, $\Omega(\mathcal{H}_{34}^{(6)}) = S_{34}^{(6)} < \varepsilon$, $\Omega(\mathcal{H}_{45}^{(10)}) = S_{45}^{(10)} < \varepsilon$, $\Omega(\mathcal{H}_{56}^{(15)}) = S_{56}^{(15)} < \varepsilon$. These inequalities can be taken in pairs as input to Lemma 2.4 (see Figure 1) to conclude that $S_{13}^{(2)} < \alpha_1 \varepsilon$, $S_{24}^{(5)} < \alpha_1 \varepsilon$, $S_{35}^{(9)} < \alpha_1 \varepsilon$, $S_{46}^{(14)} < \alpha_1 \varepsilon$. The newly obtained inequalities can be fed back into Lemma 2.4 to obtain $S_{14}^{(4)} < \alpha_2 \alpha_1 \varepsilon$, $S_{25}^{(8)} < \alpha_2 \alpha_1 \varepsilon$, $S_{36}^{(13)} < \alpha_2 \alpha_1 \varepsilon$. After another two applications of Lemma 2.4 in this manner we arrive at $\Omega(\mathsf{H}^{(11)}) \equiv S_{16}^{(11)} < \alpha_4 \alpha_3 \alpha_2 \alpha_1 \varepsilon < \varepsilon_+$. Clearly, the same reasoning applies inductively to any number of blocks and the proof is completed.   ⊞
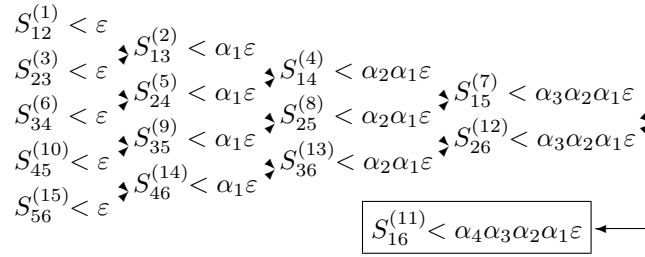
$$
\begin{array}{l}
S_{12}^{(1)} < \varepsilon \\
S_{23}^{(3)} < \varepsilon \quad S_{13}^{(2)} < \alpha_1 \varepsilon \\
\qquad\quad S_{24}^{(5)} < \alpha_1 \varepsilon \quad S_{14}^{(4)} < \alpha_2 \alpha_1 \varepsilon \\
S_{34}^{(6)} < \varepsilon \qquad\qquad\qquad S_{15}^{(7)} < \alpha_3 \alpha_2 \alpha_1 \varepsilon \\
\qquad\quad S_{35}^{(9)} < \alpha_1 \varepsilon \quad S_{25}^{(8)} < \alpha_2 \alpha_1 \varepsilon \\
S_{45}^{(10)} < \varepsilon \qquad\qquad\qquad S_{26}^{(12)} < \alpha_3 \alpha_2 \alpha_1 \varepsilon \\
\qquad\quad S_{46}^{(14)} < \alpha_1 \varepsilon \quad S_{36}^{(13)} < \alpha_2 \alpha_1 \varepsilon \\
S_{56}^{(15)} < \varepsilon \\
\qquad\qquad\qquad\qquad\boxed{S_{16}^{(11)} < \alpha_4 \alpha_3 \alpha_2 \alpha_1 \varepsilon}
\end{array}
$$

FIGURE 1.   Illustration of the proof for a 6×6 block partition and column–cyclic ordering.

Remark 2.2. The preceding convergence proof applies mutatis mutandis to the block row–cyclic pivoting. It covers the one–sided Jacobi SVD method, and we plan to adapt it for a block Kogbetliantz–type SVD computation [27] by using [16] and the new UBC transformations. Further, notice that equivalence of pivot strategies is the result of pure combinatorial relations on the set of all pivot pairs, together with the associativity of the matrix multiplication. It does not depend on the block sizes in the 2×2 block partitioned transformations. Hence, our convergence result holds true for any other equivalent pivoting, in particular including parallel strategies [25],[26].

3. CONVERGENCE TO DIAGONAL FORM

Now that we have that $(\Omega(\mathsf{H}^{(k)}))_{k=1}^{\infty}$ converges to zero, it remains to prove that the matrices $\mathsf{H}^{(k)}$ converge to a fixed diagonal matrix as $k \to \infty$. This is a nontrivial task, especially in the case of multiple eigenvalues. And, finally, we have to analyze the convergence of the infinite product $\mathsf{U}^{(1)} \mathsf{U}^{(2)} \cdots \mathsf{U}^{(k)}$, $k \to \infty$.

In this section, the assumed block column cyclic strategy can be replaced with an arbitrary block pivoting that guarantees $\lim_{k \to \infty} \Omega(\mathsf{H}^{(k)}) = 0$.

3.1. **Preliminaries.** An immediate consequence of Theorem 2.3 is the following

Corollary 3.1. *In the optimal matching distance, the diagonal entries of the sequence* $(\mathsf{H}^{(k)})_{k=1}^{\infty}$ *converge to the eigenvalues of* H. *More precisely, if* $\lambda_1, \ldots, \lambda_n$ *are the eigenvalues of* H *in an arbitrary order, then*

$$
\lim_{k \to \infty} \min_{p \in \mathbb{S}_n} \max_{i=1:n} |\mathsf{H}_{p(i)p(i)}^{(k)} - \lambda_i| = 0.
$$

For the convergence to a fixed diagonal matrix, we must prove that the diagonal entries of $\mathsf{H}^{(k)}$ cannot change their affiliation to the eigenvalues. A technical difficulty is that at step k and pivot position (i_k, j_k), $n_{i_k} + n_{j_k}$ diagonal elements of $\mathsf{H}^{(k)}$ are affected in a way that cannot be expressed by simple formulas as in the classical 2×2 case, and that UBC pivoting may introduce permutations that could preclude convergence to fixed diagonal matrix. However, there is a simple and elegant setting to treat the convergence of the diagonal entries of the matrices $\mathsf{H}^{(k)}$.

Proposition 3.1. *Let* $\mathsf{h}^{(k)} = (\mathsf{H}_{11}^{(k)}, \mathsf{H}_{22}^{(k)}, \ldots, \mathsf{H}_{nn}^{(k)})^T$, $k = 1, 2, \ldots$, *and let* $\vec{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)^T$ *be the vector of the eigenvalues of* H *in some order. Then for any implementation of the Jacobi algorithm,* $\mathsf{h}^{(k)} = (\mathsf{Q} \circ \overline{\mathsf{Q}})\vec{\lambda}$, *where* Q *is unitary,* $\circ$ *denotes the Hadamard matrix product, and* $\overline{\mathsf{Q}}$ *denotes entry–wise complex conjugation. Hence, all diagonals* $\mathsf{h}^{(k)}$ *can be considered as points in the convex polyhedral set* $\mathbf{P}(\mathsf{H})$ *whose extreme points are all permutations of the vector* $\vec{\lambda}$.

Proof: Let $\mathcal{O}(\mathsf{H}) = \{\mathsf{Q}^*\mathsf{H}\mathsf{Q}, \quad \mathsf{Q} \in \mathbb{U}(n)\}$ be the adjoint orbit of H. For any $\mathsf{M} \in \mathcal{O}(\mathsf{H})$ it holds that $\mathsf{M} = \mathsf{Q}\Lambda\mathsf{Q}^*$ with some unitary Q and $\Lambda = \mathrm{diag}(\lambda_i)_{i=1}^n$, and we directly compute that $\mathsf{m} = (\mathsf{M}_{11}, \mathsf{M}_{22}, \ldots, \mathsf{M}_{nn})^T$ can be expressed as $\mathsf{m} = (\mathsf{Q} \circ \overline{\mathsf{Q}})\vec{\lambda}$. Since all iterates $\mathsf{H}^{(k)}$ belong to $\mathcal{O}(\mathsf{H})$, the proof is completed by calling the Schur–Horn theorem, or by the Birkhoff representation theorem for doubly–stochastic matrices.

$\boxplus$

It follows from Proposition 3.1 that the convergence toward diagonal form can be analyzed by observing how $\mathsf{h}^{(k)}$ changes in a vicinity of an extreme point of $\mathbf{P}(\mathsf{H})$.

Another key issue will be the structure of the transformation matrices as $\mathsf{H}^{(k)}$, $k \to \infty$, become almost diagonal. We will use the following entry–wise form of the $\sin\Theta$ theorem [4]:

Theorem 3.2. *Let* A *be Hermitian matrix with spectral decomposition* $\mathsf{A} = \mathsf{V}\mathsf{D}_\alpha\mathsf{V}^*$, $\mathsf{D}_\alpha = \mathrm{diag}(\alpha_i)_{i=1}^n$, *and let* $\tilde{\mathsf{A}} = \mathsf{A} + \delta\mathsf{A}$ *be a Hermitian perturbation of* A, *with spectral decomposition* $\tilde{\mathsf{A}} = \tilde{\mathsf{V}}\mathsf{D}_{\tilde{\alpha}}\tilde{\mathsf{V}}^*$, $\mathsf{D}_{\tilde{\alpha}} = \mathrm{diag}(\tilde{\alpha}_j)_{j=1}^n$. *Let* $\tilde{\alpha}_j$, $j \in \mathcal{J}_j \subseteq \{1, \ldots, n\}$ *be some eigenvalues of* $\tilde{\mathsf{A}}$. *Let* $\mathcal{I}_j$ *be the indices of the corresponding eigenvalues* α_i *of* A *(in the sense that all* α_i, $i \in \mathcal{I}_j^c = \{1, \ldots, n\} \setminus \mathcal{I}_j$ *are well separated from all* $\tilde{\alpha}_j$, $j \in \mathcal{J}_j$). *If* $\Theta = \mathsf{V}^*\tilde{\mathsf{V}}$, *then*

$$\sqrt{\sum_{(i,j) \in \mathcal{I}_j^c \times \mathcal{J}_j} |\Theta_{ij}|^2} \leq \frac{\|\delta\mathsf{A}\|_F}{\min\limits_{(i,j) \in \mathcal{I}_j^c \times \mathcal{J}_j} |\tilde{\alpha}_j - \alpha_i|}.$$

Corollary 3.3. *Let* A *be almost diagonal matrix with spectral decomposition* $\mathsf{A} = \mathsf{V}\mathsf{D}_\alpha\mathsf{V}^*$. *Let* α_j, $j \in \mathcal{J}_j \subseteq \{1, \ldots, n\}$ *be some eigenvalues of* A. *Let* $\mathcal{I}_j$ *be the indices of the corresponding diagonal elements* A_{ii} *of* A, *in the sense that all* α_i, $i \in \mathcal{I}_j^c = \{1, \ldots, n\} \setminus \mathcal{I}_j$ *are well separated from all* α_j, $j \in \mathcal{J}_j$. *Then*

$$\sqrt{\sum_{(i,j) \in \mathcal{I}_j^c \times \mathcal{J}_j} |\mathsf{V}_{ij}|^2} \leq \frac{\Omega(\mathsf{A})}{\min\limits_{(i,j) \in \mathcal{I}_j^c \times \mathcal{J}_j} |\alpha_j - \mathsf{A}_{ii}|}.$$

Proof: Take A as perturbation of its diagonal part, and apply Theorem 3.2. $\boxplus$

3.2. **Simple eigenvalues.** As expected, the case of simple eigenvalues is easier to handle. Let λ_1, $\lambda_2, \ldots, \lambda_n$ be the n simple eigenvalues of H and let $\gamma = \min_{i \neq j} |\lambda_i - \lambda_j|$.

Lemma 3.1. *The block–transformation (1.2) can be implemented so that, for every index k greater or equal than a sufficiently large $k_0 \in \mathbb{N}$, the unitary matrix $\mathcal{U}_{i_k j_k}^{(k)}$ is a small perturbation of the identity.*

Proof: Let k_0 be such that $\varepsilon \equiv \Omega(\mathsf{H}^{(k_0)}) < \gamma/3$. Then the n diagonal entries of all $\mathsf{H}^{(k)}$, $k \geq k_0$, are always inside the n mutually disjoint circles of radii $\gamma/3$ around the eigenvalues of H, and each circle contains exactly one diagonal entry. As a consequence, all pivot submatrices $\mathcal{H}^{(k)}_{i_k j_k}$, $k \geq k_0$, have only simple eigenvalues. Now, consider the diagonalization of the pivot submatrix $\mathcal{H}^{(k)}_{i_k j_k}$, $\mathcal{H}^{(k+1)}_{i_k j_k} = (\mathcal{U}^{(k)}_{i_k j_k})^* \mathcal{H}^{(k)}_{i_k j_k} \mathcal{U}^{(k)}_{i_k j_k}$. If we think of $\mathcal{H}^{(k)}_{i_k j_k}$ as being a perturbation of its own diagonal part, then by Corollary 3.3 each column of $\mathcal{U}^{(k)}_{i_k j_k}$ contains exactly one element $O(\varepsilon^2)$ close to one (all remaining entries in that column are $O(\varepsilon)$ small). Since no two such big elements can be in the same row, the matrix $\mathcal{U}^{(k)}_{i_k j_k}$ is actually an $O(\varepsilon)$ perturbation of a permutation matrix. If $\mathcal{U}^{(k)}_{i_k j_k}$ is post–processed by a Businger–Golub permutation as in Lemma 2.2, then in the block partition

$$\mathcal{U}^{(k)}_{i_k j_k} = \begin{pmatrix} \mathsf{U}^{(k)}_{[i_k i_k]} & \mathsf{U}^{(k)}_{[i_k j_k]} \\ \mathsf{U}^{(k)}_{[j_k i_k]} & \mathsf{U}^{(k)}_{[j_k j_k]} \end{pmatrix}, \quad \mathsf{U}^{(k)}_{[i_k i_k]} \in \mathbb{C}^{n_{i_k} \times n_{i_k}}, \quad \mathsf{U}^{(k)}_{[j_k j_k]} \in \mathbb{C}^{n_{j_k} \times n_{j_k}}$$

all $O(1)$ entries are in the diagonal blocks $\mathsf{U}^{(k)}_{[i_k i_k]}$ and $\mathsf{U}^{(k)}_{[j_k j_k]}$. By two separate permutations $P_1 \in \mathbb{S}_{n_{i_k}}$, $P_2 \in \mathbb{S}_{n_{j_k}}$ of the first n_{i_k} and the last n_{j_k} columns of $\mathcal{U}^{(k)}_{i_k j_k}$, respectively, we can place those $O(1)$ entries to the diagonal of $\mathcal{U}^{(k)}_{i_k j_k}(P_1 \oplus P_2)$. By an additional postmultiplication with a unitary diagonal matrix $\Upsilon^{(k)}$, we can make all those big diagonal entries real and positive. This makes $\hat{\mathcal{U}}^{(k)}_{i_k j_k} \equiv \mathcal{U}^{(k)}_{i_k j_k}(P_1 \oplus P_2)\Upsilon^{(k)}$ a perturbation of the identity. More precisely, for each column index j

$$(3.1) \qquad \sqrt{\sum_{i \neq j} |(\hat{\mathcal{U}}^{(k)}_{i_k j_k})_{ij}|^2} \leq \frac{\Omega(\mathcal{H}^{(k)}_{i_k j_k})}{\gamma/3}, \quad \left| (\hat{\mathcal{U}}^{(k)}_{i_k j_k})_{jj} - 1 \right| \leq \frac{\Omega(\mathcal{H}^{(k)}_{i_k j_k})^2}{\gamma^2/9}.$$

By adding all off–diagonal entries in a divide–and–conquer fashion, we obtain

$$(3.2) \qquad \Omega(\hat{\mathcal{U}}^{(k)}_{i_k j_k}) \leq \sqrt{2 \log_2(n_{i_k} + n_{j_k})} \frac{\Omega(\mathcal{H}^{(k)}_{i_k j_k})}{\gamma/3}.$$

Finally, note that the permutations P_1 and P_2 do not change the UBC property of the transformation (see Lemma 2.2 and Corollary 2.1). $\boxplus$

Theorem 3.4. *Let H be an $n \times n$ Hermitian matrix with simple eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. If the block column–cyclic or block row–cyclic Jacobi algorithm is implemented using UBC transformations as in Lemma 3.1, then with some permutation $p \in \mathbb{S}_n$*

$$\lim_{k \to \infty} \mathsf{H}^{(k)} = \Lambda = \mathrm{diag}(\lambda_{p(1)}, \ldots, \lambda_{p(n)}),$$

where the reduction of the off–norm is quadratic per full sweep. The accumulated product of the block Jacobi transformations converges toward the corresponding eigenvector matrix,

$$\lim_{k \to \infty} (\mathsf{U}^{(1)} \mathsf{U}^{(2)} \cdots \mathsf{U}^{(k)}) = \mathsf{U}, \quad \text{where } \mathsf{U}^* \mathsf{U} = \mathsf{I}, \ \mathsf{H} \mathsf{U} = \mathsf{U} \Lambda.$$

Proof: For start, let k_0 be such that $\Omega(\mathsf{H}^{(k_0)}) < \gamma/3$. Later on, we may require larger values of k_0 to achieve smaller $\Omega(\mathsf{H}^{(k_0)})$. If we set $\mathsf{h}^{(k)} = \mathrm{diag}(\mathsf{H}^{(k)}_{ii})^n_{i=1}$, then by Corollary 3.1 there exists a permutation p such that the vector $\vec{\lambda}_p = (\lambda_{p(1)}, \lambda_{p(2)}, \ldots, \lambda_{p(n)})$ satisfies

$$(3.3) \qquad \|\vec{\lambda}_p - \mathsf{h}^{(k)}\|_2 \leq \Omega(\mathsf{H}^{(k)}) < \frac{\gamma}{3}.$$

Since for any other permutation $q \neq p$ it necessarily holds $\|\vec{\lambda}_q - \vec{\lambda}_p\|_\infty \geq \gamma$, we have that $\|\vec{\lambda}_q - \mathsf{h}^{(k)}\|_\infty > 2\gamma/3$. Hence, $\mathsf{h}^{(k)}$ is affiliated with the extreme point $\vec{\lambda}_p \in \mathbf{P}(\mathsf{H})$ and it remains to prove the same affiliation for $\mathsf{h}^{(k+1)}$, if k is sufficiently large.

For all $k \geq k_0$ and k_0 sufficiently large, each $\mathsf{U}^{(k)}$ can be written as $\mathsf{U}^{(k)} = \mathsf{I} + \mathsf{E}^{(k)}$, where by Lemma 3.1 (relation (3.2)) $\Omega(\mathsf{E}^{(k)}) \leq \sqrt{2\log_2 n}\,\Omega(\mathcal{H}^{(k)}_{i_k j_k})/(\gamma/3)$. Then

$$\mathsf{h}^{(k)} = (\mathsf{U}^{(k)} \circ \overline{\mathsf{U}}^{(k)})\mathsf{h}^{(k+1)}, \quad \mathsf{U}^{(k)} \circ \overline{\mathsf{U}}^{(k)} = \mathsf{I} + 2\mathrm{diag}(\Re(\mathsf{E}^{(k)}_{jj}))^n_{j=1} + \mathsf{E}^{(k)} \circ \overline{\mathsf{E}}^{(k)}$$

where $\Re(\cdot)$ denotes the real part of a complex number. Since $\mathsf{U}^{(k)}$ is unitary, we have

$$\text{for all } j, \;\; 2\Re(\mathsf{E}^{(k)}_{jj}) + |\mathsf{E}^{(k)}_{jj}|^2 = -\sum_{i \neq j} |\mathsf{E}^{(k)}_{ij}|^2$$

which proves that $\|\mathsf{U}^{(k)} \circ \overline{\mathsf{U}}^{(k)} - \mathsf{I}\|_2$ is quadratically small: $\|\mathsf{U}^{(k)} \circ \overline{\mathsf{U}}^{(k)} - \mathsf{I}\|_2 \leq 2\Omega^2(\mathsf{E}^{(k)})$. Using (3.2), we conclude

$$\|\mathsf{h}^{(k+1)} - \mathsf{h}^{(k)}\|_2 \;\; \leq \;\; \|\mathsf{U}^{(k)} \circ \overline{\mathsf{U}}^{(k)} - \mathsf{I}\|_2 \|\mathsf{H}\|_F \leq 36\log_2 n \frac{\Omega^2(\mathsf{H}^{(k)})}{\gamma^2}\|\mathsf{H}\|_F < \frac{\gamma}{3}$$

(3.4) $\qquad$ provided that $\qquad \dfrac{\Omega(\mathsf{H}^{(k)})}{\|\mathsf{H}\|_F} < \dfrac{1}{\sqrt{108\log_2 n}}\left(\dfrac{\gamma}{\|\mathsf{H}\|_F}\right)^{\frac{3}{2}}.$

The condition (3.4) will be satisfied for all k greater or equal than a sufficiently large k_0. Hence, for all $k \geq k_0$ we have

(3.5) $\qquad\qquad \|\vec{\lambda}_p - \mathsf{h}^{(k)}\|_2 \leq \Omega(\mathsf{H}^{(k)}) \Longrightarrow \|\vec{\lambda}_p - \mathsf{h}^{(k+1)}\|_2 \leq \Omega(\mathsf{H}^{(k+1)}).$

Thus, $\lim_{k \to \infty} \mathsf{H}^{(k)} = \mathrm{diag}(\lambda_{p(1)}, \ldots, \lambda_{p(n)})$.

The quadratic reduction of $\Omega(\mathsf{H}^{(k)})$ can be demonstrated using the Wilkinson's proof [29]. The change of the off–diagonal blocks in a block row–cyclic pivoting follows almost verbatim that of [29] – the sines and the cosines of the Jacobi angles just need to be replaced with the corresponding elements of the CS decomposition (2.3). The diagonal blocks at the end of one full cycle are actually diagonal. Hence, with the sweep length $\mathfrak{s} = m(m-1)/2$, for every $k \equiv k_{12} \geq k_0$

(3.6) $\qquad \Omega(\mathsf{H}^{(k+\mathfrak{s})}) \leq \zeta(n,\gamma)\Omega^2(\mathsf{H}^{(k)}), \;\; \zeta(n,\gamma) = \dfrac{\sqrt{2\max_{i,j}\log_2(n_i+n_j)}}{\gamma/3} \leq \dfrac{\sqrt{2\log_2 n}}{\gamma/3},$

i.e. the off–norm is reduced at quadratic rate per sweep. Note that on entry to the quadratic bound (3.6) k can correspond to any pivot position, $k = k_{12}$ is chosen for simplicity and easier comparison with [29].

Recall that for sufficiently large k, by Lemma 3.1, $\mathsf{U}^{(k)} = \mathsf{I} + \mathsf{E}^{(k)}$, where $\|\mathsf{E}^{(k)}\|_F \leq \zeta(n,\gamma)\Omega(\mathcal{H}^{(k)}_{i_k j_k})$. Using a technique by Causey and Henrici [3], we prove that

(3.7) $$\sum_{k=1}^{\infty} \|\mathsf{E}^{(k)}\|_F < \infty,$$

which is a sufficient condition for the convergence of $\mathsf{U}^{(1)}\mathsf{U}^{(2)} \cdots \mathsf{U}^{(k)}$ as $k \to \infty$. For the convergence of the series (3.7), note that for $k_0 \to \infty$

$$\begin{aligned}
\sum_{k=k_0}^{\infty} \|\mathsf{E}^{(k)}\|_F \;\; &\leq \;\; \sum_{k=k_0}^{\infty} \zeta(n,\gamma)\Omega(\mathsf{H}^{(k)}) \leq \sum_{k=k_0}^{k_0+\mathfrak{s}-1}\sum_{\ell=0}^{\infty} \zeta(n,\gamma)\Omega(\mathsf{H}^{(k+\ell\mathfrak{s})}) \\
&\leq \;\; \sum_{k=k_0}^{k_0+\mathfrak{s}-1}\sum_{\ell=0}^{\infty}\left(\zeta(n,\gamma)\Omega(\mathsf{H}^{(k)})\right)^{2^\ell} \leq \mathfrak{s}\frac{\zeta(n,\gamma)\Omega(\mathsf{H}^{(k_0)})}{1 - \zeta(n,\gamma)\Omega(\mathsf{H}^{(k_0)})} \to 0.
\end{aligned}$$

Thus, $\mathsf{U} = \lim_{k\to\infty} \mathsf{U}^{(1)}\mathsf{U}^{(2)}\cdots\mathsf{U}^{(k)}$ exists and its columns are the eigenvectors of H, since at the limit $\Lambda = \lim_{k\to\infty} \mathsf{H}^{(k+1)} = \lim_{k\to\infty} \left(\mathsf{U}^{(1)}\mathsf{U}^{(2)}\cdots\mathsf{U}^{(k)}\right)^{*} \mathsf{H}\mathsf{U}^{(1)}\mathsf{U}^{(2)}\cdots\mathsf{U}^{(k)} = \mathsf{U}^{*}\mathsf{H}\mathsf{U}$. The limit product U is necessarily unitary.   ⊞

3.3. **Multiple eigenvalues.** We now turn to the general case of multiple eigenvalues. Let $\lambda_1^{\bullet} > \lambda_2^{\bullet} > \cdots > \lambda_s^{\bullet}$ be all different eigenvalues of H with multiplicities $\mu_1, \mu_2, \ldots, \mu_s$, respectively, and with the spectral gap $\gamma = \min_{i\neq j} |\lambda_i^{\bullet} - \lambda_j^{\bullet}|$. For a sufficiently large k, precisely μ_j diagonal entries of $\mathsf{H}^{(k)}$ will be affiliated to each eigenvalue $\lambda_j^{\bullet}$, and each diagonal entry is affiliated to exactly one eigenvalue. For this kind of distribution of the diagonal entries of $\mathsf{H}^{(k)}$ it suffices to have $\Omega(\mathsf{H}^{(k)}) < \gamma/3$. All n eigenvalues (counted with multiplicities, and denoted by $\lambda_1, \lambda_2, \ldots, \lambda_n$) can be arranged in some order in a vector $\vec{\lambda}_p$, so that $\|\vec{\lambda}_p - \mathsf{h}^{(k)}\|_2 \leq \Omega(\mathsf{H}^{(k)}) < \gamma/3$. The task is to establish (3.5).

Theorem 3.5. *Let H be an $n \times n$ Hermitian matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, counted with multiplicities. The transformations (1.2,1.3) can be implemented to guarantee the convergence of the block column–cyclic Jacobi algorithm: with some $p \in \mathbb{S}_n$*

$$\lim_{k\to\infty} \mathsf{H}^{(k)} = \Lambda = \begin{pmatrix} \lambda_{p(1)} & & \\ & \ddots & \\ & & \lambda_{p(n)} \end{pmatrix}.$$

Proof: Since $\lim_{k\to\infty} \Omega(\mathsf{H}^{(k)}) = 0$, we can assume that k is taken large enough to have $\Omega(\mathsf{H}^{(k)})$ sufficiently small relative to the gaps in the spectrum. Then we consider the changes of the affected diagonal elements, and we show that certain, in a sense consistent, implementation of the block–transformation (1.2) ensures that the matrices $\mathsf{H}^{(k)}$ converge to a fixed diagonal matrix as $k \to \infty$.

It suffices to consider a k–th step and to show that for sufficiently large k the diagonal elements do not change affiliation in the transformation $\mathsf{H}^{(k+1)} = \mathsf{U}^{(k)*}\mathsf{H}^{(k)}\mathsf{U}^{(k)}$.[1] Before going to the most general case, let us note that two possible situations are quite simple:

• If all $n_{i_k} + n_{j_k}$ diagonal entries of the pivot submatrix $\mathcal{H}_{i_k j_k}^{(k)}$ are affiliated to different eigenvalues of H, then, for sufficiently large k, $\mathcal{H}_{i_k j_k}^{(k)}$ has only simple and well separated eigenvalues and we can repeat the arguments from §3.2 to conclude that there will be no change of affiliation of the affected diagonal entries in the k–th step. In fact, this completes the proof if H has only simple eigenvalues.

• If all $n_{i_k} + n_{j_k}$ diagonal entries of $\mathcal{H}_{i_k j_k}^{(k)}$ are affiliated with the same eigenvalue of H, then we cannot extract any useful structure from $\mathcal{U}_{i_k j_k}^{(k)}$ – it does not have to be close to the identity. However, the change of the diagonal can be represented as

$$(3.8) \qquad \mathsf{h}^{(k+1)} = (\mathsf{U}^{(k)*} \circ \mathsf{U}^{(k)\mathrm{T}})\mathsf{h}^{(k)} + \mathsf{e}^{(k)},$$

where $\mathsf{U}^{(k)*} \circ \mathsf{U}^{(k)\mathrm{T}}$ is uni–stochastic and $\|\mathsf{e}^{(k)}\|_2 \leq \Omega(\mathsf{H}^{(k)}) < \gamma/3$. Note that (3.8) expresses the changes of the diagonal elements as convex combinations with small additive perturbations. We immediately conclude that $\|\vec{\lambda}_p - \mathsf{h}^{(k+1)}\|_2 \leq \Omega(\mathsf{H}^{(k+1)})$, and this conclusion remains valid for any reordering of the columns of $\mathcal{U}_{i_k j_k}^{(k)}$.

For the general case we need more elaborated analysis, and we have to go down to the scalar entries of the transformation matrices. To simplify the notation, the transformation matrix $\mathcal{U}_{i_k j_k}^{(k)}$, which is assumed to be in the UBC form, will be denoted by $\hat{\mathsf{U}}^{(k)}$, the vectors of the affected diagonal entries (of $\mathcal{H}_{i_k j_k}^{(k)}$ and $\mathcal{H}_{i_k j_k}^{(k+1)}$) will be $\hat{\mathsf{h}}^{(k)}$ and $\hat{\mathsf{h}}^{(k+1)}$. Thus, $\hat{\mathsf{h}}^{(k)}, \hat{\mathsf{h}}^{(k+1)} \in \mathbb{R}^{n_{i_k}+n_{j_k}}$, and

$$(3.9) \qquad \hat{\mathsf{h}}^{(k+1)} = ((\hat{\mathsf{U}}^{(k)})^{*} \circ (\hat{\mathsf{U}}^{(k)})^{\mathrm{T}})\hat{\mathsf{h}}^{(k)} + \hat{\mathsf{e}}^{(k)}.$$

---

[1]Note that the same analysis applies to any strategy that guarantees $\lim_{k\to\infty} \Omega(\mathsf{H}^{(k)}) = 0$.

In the general case, the diagonal entries of $\mathcal{H}^{(k)}_{i_k j_k}$ can be partitioned into p clusters, affiliated with p different eigenvalues of H. Let $\mathcal{I}_1, \ldots, \mathcal{I}_p$ denote the sets of the corresponding indices of the diagonal entries (also, of the elements of $\hat{\mathsf{h}}^{(k)}$) in the clusters. The eigenvalues of $\mathcal{H}^{(k+1)}_{i_k j_k}$ (that is, the elements of $\hat{\mathsf{h}}^{(k+1)}$) must also be grouped around the same p eigenvalues of H. Let $\mathcal{J}_1, \ldots, \mathcal{J}_p$ denote the corresponding indices for the p clusters. We first conclude that $|\mathcal{I}_k| = |\mathcal{J}_k| \equiv d_k$, $k = 1, \ldots, p$. Using Corollary 3.3 we have

$$(3.10) \qquad \sqrt{\sum_{(i,q) \in \mathcal{I}^c_j \times \mathcal{J}_j} \left| \hat{\mathsf{U}}^{(k)}_{iq} \right|^2} \leq \frac{\Omega(\mathcal{H}^{(k)}_{i_k j_k})}{\gamma/3}.$$

This implies

$$(3.11) \qquad 1 \geq \sum_{i \in \mathcal{I}_j} \left| \hat{\mathsf{U}}^{(k)}_{iq} \right|^2 > 1 - \frac{\Omega(\mathcal{H}^{(k)}_{i_k j_k})^2}{\gamma^2/9}, \quad 1 \geq \sum_{q \in \mathcal{J}_j} \left| \hat{\mathsf{U}}^{(k)}_{iq} \right|^2 > 1 - \frac{\Omega(\mathcal{H}^{(k)}_{i_k j_k})^2}{\gamma^2/9}.$$

Now we note the following consequence of UBC pivoting: For each cluster $\mathcal{J}_\ell$ (representing also the corresponding column indices in $\hat{\mathsf{U}}^{(k)}$) there are in general $f_\ell \leq d_\ell$ corresponding diagonal elements from the block $\mathsf{H}^{(k)}_{[i_k i_k]}$. This means that in the corresponding submatrix $\hat{\mathsf{U}}^{(k)}(\mathcal{I}_\ell, \mathcal{J}_\ell)$ only its $f_\ell \times d_\ell$ submatrix is in the first n_{i_k} rows of $\hat{\mathsf{U}}^{(k)}$. Column pivoting will take only a $f_\ell \times f_\ell$ block out of this $f_\ell \times d_\ell$ matrix inside the $(1,1)$ block, and the rest will be moved into the $(1,2)$ block. In the second block row, the elements are moved accordingly. The UBC property does not imply any structure that can be noticed by a visual inspection, see the second plot in Figure 2. However, there is a hidden special block–structure that can be revealed by certain consistent ordering of the diagonal entries involved in the transformation.

We can separately permute the first n_{i_k} and the last n_{j_k} columns of $\hat{\mathsf{U}}^{(k)}$, where the permutations Π_1, Π_2 are determined by separately sorting the first n_{i_k} and the last n_{j_k} entries of $\hat{\mathsf{h}}^{(k+1)}$. Set $\Pi = \Pi_1 \oplus \Pi_2$. (Recall that global permutation is not allowed because of the UBC property, and that these separate permutations do not change the uniform bound from Lemma 2.2.) In the same way, separately sort the two parts of $\hat{\mathsf{h}}^{(k)}$, using permutation matrices P_1, P_2, and then set $P = P_1 \oplus P_2$.

Consistent order of $\hat{\mathsf{h}}^{(k+1)}$ is introduced as follows. Both $\hat{\mathsf{h}}^{(k)}$ and $\hat{\mathsf{h}}^{(k+1)}$ are quasi–sorted by P and Π, respectively. Now note that

$$\check{\mathsf{h}}^{(k+1)} - \hat{\mathsf{h}}^{(k)} = P \left((\Pi^T (\hat{\mathsf{U}}^{(k)})^* P) \circ (\Pi^T (\hat{\mathsf{U}}^{(k)})^{\mathrm{T}} P) \right) (P^T \hat{\mathsf{h}}^{(k)}) - \hat{\mathsf{h}}^{(k)} + P\Pi^T \hat{\mathsf{e}}^{(k)}$$

$$(3.12) \qquad\qquad = \left((\check{\mathsf{U}}^{(k)})^* \circ (\check{\mathsf{U}}^{(k)})^T \right) \hat{\mathsf{h}}^{(k)} - \hat{\mathsf{h}}^{(k)} + P\Pi^T \hat{\mathsf{e}}^{(k)},$$

where $\check{\mathsf{U}}^{(k)} = \hat{\mathsf{U}}^{(k)} \Pi P^T$, and $\check{\mathsf{h}}^{(k+1)}$ is the diagonal of $(\check{\mathsf{U}}^{(k)})^* \mathcal{H}^{(k)}_{i_k j_k} \check{\mathsf{U}}^{(k)}$. We decide to use $\check{\mathsf{U}}^{(k)}$ as a *consistently ordered* block transformation matrix. Then, the $n \times n$ unitary $\mathsf{U}^{(k)}$ is defined by placing the blocks of $\check{\mathsf{U}}^{(k)}$ to the pivot positions, see (1.3).

Consistent ordering is noting else but ensuring that $\mathcal{I}_j = \mathcal{J}_j$ for all $j = 1, \ldots, p$. Let $\check{\mathsf{U}}^{(k)}_{\mathcal{I}_j, \mathcal{J}_j}$ denote the submatrix of $\check{\mathsf{U}}^{(k)}$, obtained by taking the entries with row indices in $\mathcal{I}_j$ and column indices from $\mathcal{J}_j$, where the sets $\mathcal{I}_j, \mathcal{J}_j$ are always taken as ordered. Also, e.g. $\hat{\mathsf{h}}^{(k)}_{\mathcal{I}_j}$ denotes the subvector of $\hat{\mathsf{h}}^{(k)}$, with entries indexed by $\mathcal{I}_j$.

To compute $\check{\mathsf{h}}^{(k+1)}$ by (3.12), we first note that

$$(((\check{\mathsf{U}}^{(k)})^* \circ (\check{\mathsf{U}}^{(k)})^T)(\hat{\mathsf{h}}^{(k)}))_{\mathcal{I}_j} = ((\check{\mathsf{U}}^{(k)}_{\mathcal{I}_j \mathcal{I}_j})^* \circ (\check{\mathsf{U}}^{(k)}_{\mathcal{I}_j \mathcal{I}_j})^T) \hat{\mathsf{h}}^{(k)}_{\mathcal{I}_j} + ((\check{\mathsf{U}}^{(k)}_{\mathcal{I}^c_j \mathcal{I}_j})^* \circ (\check{\mathsf{U}}^{(k)}_{\mathcal{I}^c_j \mathcal{I}_j})^T) \hat{\mathsf{h}}^{(k)}_{\mathcal{I}^c_j}$$

$$(3.13) \qquad\qquad \equiv (\check{\mathsf{S}}^{(k)}_{\mathcal{I}_j \mathcal{I}_j} + \check{\mathsf{E}}^{(k)}_{\mathcal{I}_j \mathcal{I}_j}) \hat{\mathsf{h}}^{(k)}_{\mathcal{I}_j} + \check{\mathsf{F}}^{(k)}_{\mathcal{I}_j \mathcal{I}^c_j} \hat{\mathsf{h}}^{(k)}_{\mathcal{I}^c_j},$$

where $\check{\mathsf{S}}^{(k)}_{\mathcal{I}_j \mathcal{I}_j}$ is stochastic, and $\check{\mathsf{E}}^{(k)}_{\mathcal{I}_j \mathcal{I}_j} \to 0$, $\check{\mathsf{F}}^{(k)}_{\mathcal{I}_j \mathcal{I}_j^c} \to 0$, as $k \to \infty$. (Recall that from (3.11) it follows that $\check{\mathsf{U}}^{(k)}_{\mathcal{I}_j, \mathcal{J}_j}$ has columns of Euclidean lengths ε^2 close to one.) So,

$$(3.14) \qquad \check{\mathsf{h}}^{(k+1)}_{\mathcal{I}_j} = \check{\mathsf{S}}^{(k)}_{\mathcal{I}_j \mathcal{I}_j} \hat{\mathsf{h}}^{(k)}_{\mathcal{I}_j} + \underbrace{\check{\mathsf{E}}^{(k)}_{\mathcal{I}_j \mathcal{I}_j} \hat{\mathsf{h}}^{(k)}_{\mathcal{I}_j}}_{\to 0} + \underbrace{\check{\mathsf{F}}^{(k)}_{\mathcal{I}_j \mathcal{I}_j^c} \hat{\mathsf{h}}^{(k)}_{\mathcal{I}_j^c}}_{\to 0} + \underbrace{(P\Pi^T \hat{\mathsf{e}}^{(k)})_{\mathcal{I}_j}}_{\to 0}.$$

Relation (3.14) nicely shows that there is no change of affiliation: For k large enough, a cluster of diagonal entries converging to a multiple eigenvalue changes in one step by convex combinations of entries in the previous one ($\check{\mathsf{h}}^{(k+1)}_{\mathcal{I}_j} \approx \check{\mathsf{S}}^{(k)}_{\mathcal{I}_j \mathcal{I}_j} \hat{\mathsf{h}}^{(k)}_{\mathcal{I}_j}$), plus perturbations that converge to zero. Since for k large enough this additive perturbation is not large enough to bypass the gap of width $\gamma/3$ between the neighborhoods of different eigenvalues, we can conclude that for all k greater than some $k_0 \in \mathbb{N}$,

$$\|\mathsf{h}^{(k)} - \vec{\lambda}_p\|_2 \leq \Omega(\mathsf{H}^{(k)}) \implies \|\mathsf{h}^{(k+1)} - \vec{\lambda}_p\|_2 \leq \Omega(\mathsf{H}^{(k+1)}). \qquad \boxplus$$

Remark 3.1. The change of the diagonal entries can be better seen from

$$(3.15) \qquad \|\check{\mathsf{h}}^{(k+1)} - \hat{\mathsf{h}}^{(k)}\|_2 \leq \| \left((\Pi^T (\hat{\mathsf{U}}^{(k)})^* P) \circ (\Pi^T (\hat{\mathsf{U}}^{(k)})^{\mathrm{T}} P) \right) (P^T \hat{\mathsf{h}}^{(k)}) - P^T \hat{\mathsf{h}}^{(k)}\|_2 + \|\hat{\mathsf{e}}^{(k)}\|_2$$

where we notice an auxiliary matrix $\check{\mathsf{U}}^{(k)} = P^T \hat{\mathsf{U}}^{(k)} \Pi$, that actually reveals how $\check{\mathsf{U}}^{(k)}$ changes $\hat{\mathsf{h}}^{(k)}$. An example of this is given in Figure 2:
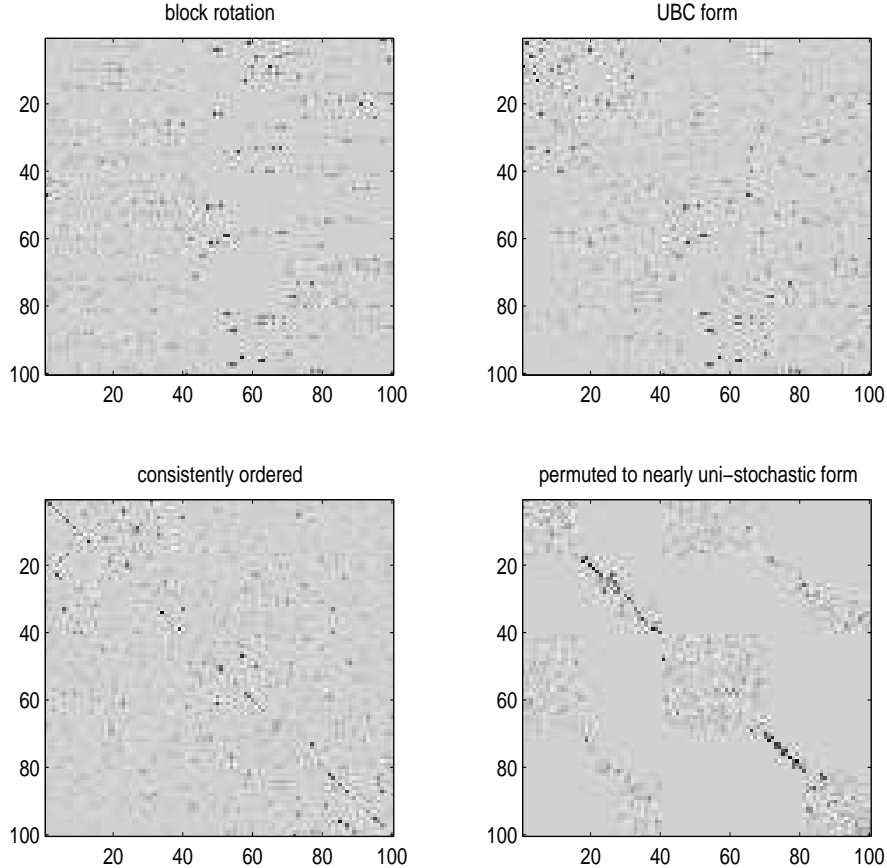

FIGURE 2.

Figure 2 illustrates the transformation of a 100×100 block, with $n_{i_k} = 40$, $n_{j_k} = 60$. The first plot shows the transformation $\hat{\mathsf{U}}^{(k)}$ before the UBC permutation. The effect of the UBC permutation is shown in the second plot. The third plot shows the consistently ordered matrix $\check{\mathsf{U}}^{(k)} = \hat{\mathsf{U}}^{(k)} \Pi P^T$. The last plot shows the auxiliary matrix $P^T \hat{\mathsf{U}}^{(k)} \Pi$ that clarifies the transformation of $\hat{\mathsf{h}}^{(k)}$. In this example, $\Omega(\mathcal{H}_{i_k j_k}^{(k)}) < 6 \cdot 10^{-7}$, $\|\check{\mathsf{h}}^{(k+1)} - \hat{\mathsf{h}}^{(k)}\|_2 < 2 \cdot 10^{-13}$, and the multiplicities of the eigenvalues (the d_k's, $k = 1, \ldots, 5$) were 46, 31, 11, 7, 5.

Remark 3.2. Clearly, sorting the diagonals of $\mathsf{H}^{(k_0)}$ when $\Omega(\mathsf{H}^{(k_0)}) < \gamma/3$ (say) will ensure that for all $k \geq k_0$ the diagonal entries that converge to the same multiple eigenvalue will occupy successive positions on the diagonal. This is important for generating quadratically or even cubically convergent modifications of the block cyclic strategy. For that, many issues, e.g. adapting the block partition $\nu = (n_1, \ldots, n_m)$ to the distribution of the eigenvalues, have to be addressed.

3.3.1. *Convergence of eigenspaces.* It remains to settle the convergence of the eigenvectors in the case of multiple eigenvalues. Of course, for a multiple eigenvalue λ_i, the eigenspace $\mathrm{Ker}(\mathsf{H} - \lambda_i \mathsf{I})$ is the right target. So, what can be said in that case about the convergence of the infinite product $\mathsf{U}^{(1)} \mathsf{U}^{(2)} \cdots \mathsf{U}^{(k)} \cdots$ of the block transformations? To our knowledge, this issue is left unanswered even in the case of the classical 2×2 Jacobi rotations.

Theorem 3.6. *Let H be an $n \times n$ Hermitian with a spectral decomposition $\mathsf{H} = \mathsf{U} \Lambda \mathsf{U}^*$ where the diagonal matrix Λ contains the eigenvalues of H in an arbitrary order and the columns of the unitary matrix U are the corresponding eigenvectors. Consider the sequence $\mathsf{H}^{(k+1)} = \mathsf{U}^{(k)*} \mathsf{H}^{(k)} \mathsf{U}^{(k)}$ computed by a convergent diagonalization process, e.g. as in Theorem 3.5. Let for $k = 1, 2, \ldots$ $\mathsf{U}^{(1:k)} = \mathsf{U}^{(1)} \mathsf{U}^{(2)} \cdots \mathsf{U}^{(k)}$, so that after k steps the computed matrix is $\mathsf{H}^{(k+1)} = (\mathsf{U}^{(1:k)})^* \mathsf{H} \mathsf{U}^{(1:k)}$. Let $\lambda_1^\bullet, \ldots, \lambda_s^\bullet$ be all different and in general multiple eigenvalues of H. Then there exists a partition $\{1, \ldots, n\} = \bigcup_{i=1}^s \mathcal{D}_i$ such that for all $i = 1, \ldots, s$, $|\mathcal{D}_i| = \mu_i$, and*

$$(3.16) \qquad \lim_{k \to \infty} \| \sin \Theta(\mathsf{U}_{:,\mathcal{D}_i}^{(1:k)}, \mathrm{Ker}(\mathsf{H} - \lambda_i^\bullet \mathsf{I})) \|_F = 0.$$

Proof: We can choose such a $k_0 \in \mathbb{N}$ that for all $k \geq k_0$ $\Omega(\mathsf{H}^{(k)})$ is as small as we like, and no change of affiliation of diagonal entries takes place in the transformation of $\mathsf{H}^{(k)}$. Hence, there is a well–defined partition $\{1, \ldots, n\} = \bigcup_{i=1}^s \mathcal{D}_i$, independent of k, such that for all $k \geq k_0$ the diagonal entries of $\mathsf{H}_{\mathcal{D}_i \mathcal{D}_i}^{(k)}$ belong to the interval $[\lambda_i^\bullet - \Omega(\mathsf{H}^{(k)}), \lambda_i^\bullet + \Omega(\mathsf{H}^{(k)})]$.

Let $\mathsf{H}^{(k+1)} = \mathsf{V}^{(k+1)} \Lambda (\mathsf{V}^{(k+1)})^*$ be a spectral decomposition of $\mathsf{H}^{(k+1)}$, and let $\mathcal{E}_i$ denote the column indices in U and in $\mathsf{V}^{(k+1)}$ that correspond to the eigenvalue $\lambda_i^\bullet$ ($\mathsf{H} \mathsf{U}_{:,\mathcal{E}_i} = \lambda_i^\bullet \mathsf{U}_{:,\mathcal{E}_i}$, $\mathsf{H}^{(k+1)} \mathsf{V}_{:,\mathcal{E}_i}^{(k+1)} = \lambda_i^\bullet \mathsf{V}_{:,\mathcal{E}_i}^{(k+1)}$, $|\mathcal{E}_i| = \mu_i =$ the multiplicity of $\lambda_i^\bullet$). We note that for all i, both $\mathsf{U}_{:,\mathcal{E}_i}$ and $\mathsf{V}_{:,\mathcal{E}_i}^{(k+1)}$ are determined up to a postmultiplication by arbitrary $\mu_i \times \mu_i$ unitary matrix. However, all quantities involved in our estimates are invariant under orthogonal changes of bases in eigenspaces, so we can simplify notation and simply ignore that vagueness in expression.[2] So, we can write $\mathsf{U}^{(1:k)*} \mathsf{U} = \mathsf{V}^{(k+1)}$ and, by Corollary 3.3,

$$(3.17) \qquad \| \sin \Theta(\mathsf{U}_{:,\mathcal{D}_i}^{(1:k)}, \mathrm{Ker}(\mathsf{H} - \lambda_i^\bullet \mathsf{I})) \|_F = \| (\mathsf{U}_{:,\mathcal{D}_i^c}^{(1:k)})^* \mathsf{U}_{:,\mathcal{E}_i} \|_F = \| \mathsf{V}_{\mathcal{D}_i^c, \mathcal{E}_i}^{(k+1)} \|_F \leq \frac{\Omega(\mathsf{H}^{(k+1)})}{\gamma/3}.$$

Thus, affiliation of diagonal entries of the iterates $\mathsf{H}^{(k)}$ to the eigenvalues is followed by the affiliation of the columns of the accumulated unitary transformations (more precisely their spans) to the corresponding eigenspaces of H.   ⊞

---

[2]See for instance the second equality in relation (3.17), and Remark 3.3.

Remark 3.3. Let the eigenvalues of H be ordered as $\lambda_1^\bullet > \cdots > \lambda_s^\bullet$, and $\Lambda = \oplus_{i=1}^s \lambda_i^\bullet \mathsf{I}_{\mu_i}$. Then $\mathcal{E}_1 = \{1, \ldots, \mu_1\}$, $\mathcal{E}_i = \{j \;:\; 1 + \sum_{p=1}^{i-1} \mu_p \leq j \leq \sum_{p=1}^i \mu_p\}$, $2 \leq i \leq s$. Let $\mathrm{G}(\mu_i, n)$ denote the set of μ_i–dimensional subspaces of $\mathbb{C}^n$ ($\mathbb{R}^n$ for real symmetric H), a complex (real in the symmetric case) Grassmann manifold (with the corresponding analytical structure). A point $\mathfrak{X}$ in $\mathrm{G}(\mu_i, n) \equiv \mathbf{U}(n)/(\mathbf{U}(\mu_i) \times \mathbf{U}(n - \mu_i))$ is represented as equivalence class $[\mathsf{X}] = \{\mathsf{XQ} \;:\; \mathsf{Q} \in \mathbf{U}(\mu_i)\}$, where the columns of X span $\mathfrak{X}$, and $\mathsf{X}^*\mathsf{X} = \mathsf{I}_{\mu_i}$. See [15] for more detailed theory. In particular, $[\mathsf{U}_{:,\mathcal{E}_i}] \in \mathrm{G}(\mu_i, n)$. If the diagonals of $\mathsf{H}^{(k_0)}$ are sorted as in Remark 3.2, then $\mathcal{D}_i = \mathcal{E}_i$, and the convergence to individual eigenspaces ($\mathsf{U}_{:,\mathcal{E}_i}^{(1:k)} \to \mathsf{U}_{:,\mathcal{E}_i}$) can be separately studied in the framework of $\mathrm{G}(\mu_i, n)$, and (3.16) states the convergence in the projection F–norm[3] (chordal) distance $d(\mathfrak{X}, \mathfrak{Y}) = \|\sin\Theta(\mathfrak{X}, \mathfrak{Y})\|_F$. Further, the eigenvector matrix U induces an orthogonal decomposition of $\mathbb{C}^n$ as $\mathbb{C}^n = \oplus_{i=1}^s \mathrm{Range}(\mathsf{U}_{:,\mathcal{E}_i})$. If we consider the set of all such decompositions, represented as quotient space $\mathfrak{F}(\mu_1, \ldots, \mu_s) = \mathbf{U}(n)/(\mathbf{U}(\mu_1) \times \cdots \times \mathbf{U}(\mu_s))$, then the eigenvector matrix U can be considered as a point in $\mathfrak{F}(\mu_1, \ldots, \mu_s)$, and the convergence $\mathsf{U}^{(1:k)} \to \mathsf{U}$ can be formulated in $\mathfrak{F}(\mu_1, \ldots, \mu_s)$ – a generalized flag manifold setting.

Remark 3.4. In practical computation, the iterations are stopped at some finite index $k = \overline{k}$ with $\Omega(\mathsf{H}^{(\overline{k})})/\|\mathsf{H}\|_F \leq \varepsilon$. After inspecting the diagonal entries of $\mathsf{H}^{(\overline{k})}$, we can identify clusters and isolated entries and make the corresponding partition of the columns of $\mathsf{U}^{(\overline{k}-1)}$. However, there is no way to tell whether an ε tight cluster of diagonal entries approximates one multiple eigenvalue or a cluster of simple and/or multiple pathologically close eigenvalues whose multiplicities sum up to the number of diagonals in the cluster.[4] It is clear that (3.17) can be applied to spectral subspaces as well: instead of a single $\mathcal{E}_i$ we can take a union of $\mathcal{E}_i$'s corresponding to pathologically close eigenvalues and then obtain a useful estimate for the corresponding spectral subspace. In the case of positive definite H, the stopping criterion is stronger, $\max_{i \neq j} |(\mathsf{H}^{(\overline{k})})_{ij}|/\sqrt{(\mathsf{H}^{(\overline{k})})_{ii}(\mathsf{H}^{(\overline{k})})_{jj}} \leq \varepsilon$, so one can use relative perturbation theory and make better estimates, but the problem is in principle the same.

4. Accuracy issues in finite precision computation

Our previous analysis was focused to theoretical aspects of the global convergence of cyclic block Jacobi methods. That is necessarily the first step on a long and rough way to an efficient and numerically reliable mathematical software. Many important implementation issues remain for a future research, such as computing the transformation $\hat{\mathsf{U}}^{(k)}$ accurately, which is very important for the overall performance, in particular in the case of the one–sided transformations (1.5). Note that in a k–th step of the one–sided Jacobi SVD algorithm we actually use $\hat{\mathsf{U}}^{(k)}$ (notation from §3.3) to compute the SVD of the pivot submatrix $(\mathsf{A}_{[i_k]}^{(k)}, \mathsf{A}_{[j_k]}^{(k)})$ as

$$(4.1) \qquad \left(\mathsf{A}_{[i_k]}^{(k+1)}, \mathsf{A}_{[j_k]}^{(k+1)}\right) = \left(\mathsf{A}_{[i_k]}^{(k)}, \mathsf{A}_{[j_k]}^{(k)}\right) \hat{\mathsf{U}}^{(k)},$$

where the columns of $(\mathsf{A}_{[i_k]}^{(k+1)}, \mathsf{A}_{[j_k]}^{(k+1)})$ are mutually orthogonal. However, if $\hat{\mathsf{U}}^{(k)}$ is computed only approximately as $\tilde{\mathsf{U}}^{(k)} = \hat{\mathsf{U}}^{(k)} + \delta\hat{\mathsf{U}}^{(k)}$, then the numerical orthogonality of the transformed columns depends on the structure of $\delta\hat{\mathsf{U}}^{(k)}$. To illustrate, consider the transformation (4.1) in infinite precision, but with $\tilde{\mathsf{U}}^{(k)}$ instead of $\hat{\mathsf{U}}^{(k)}$:

$$(4.2) \qquad \left(\mathsf{A}_{[i_k]}^{(k)}, \mathsf{A}_{[j_k]}^{(k)}\right) \tilde{\mathsf{U}}^{(k)} = \left(\mathsf{A}_{[i_k]}^{(k+1)}, \mathsf{A}_{[j_k]}^{(k+1)}\right) + \left(\mathsf{A}_{[i_k]}^{(k)}, \mathsf{A}_{[j_k]}^{(k)}\right) \delta\hat{\mathsf{U}}^{(k)}.$$

---

[3]Cf. [15].

[4]Unfortunately, we cannot "let ε go to zero".

If we set $A^{(k)} = B^{(k)}D^{(k)}$, where $B^{(k)}$ has columns of unit Euclidean length and $D^{(k)}$ is diagonal, then (4.2) reads

$$\left(A^{(k)}_{[i_k]}, A^{(k)}_{[j_k]}\right)\tilde{U}^{(k)} = \left(\left(B^{(k+1)}_{[i_k]}, B^{(k+1)}_{[j_k]}\right) + \left(B^{(k)}_{[i_k]}, B^{(k)}_{[j_k]}\right)\delta F^{(k)}\right)\begin{pmatrix} D^{(k+1)}_{[i_k i_k]} & 0 \\ 0 & D^{(k+1)}_{[j_k j_k]} \end{pmatrix},$$

$$\delta F^{(k)} = \begin{pmatrix} D^{(k)}_{[i_k i_k]} & 0 \\ 0 & D^{(k)}_{[j_k j_k]} \end{pmatrix}\delta\hat{U}^{(k)}\begin{pmatrix} (D^{(k+1)}_{[i_k i_k]})^{-1} & 0 \\ 0 & (D^{(k+1)}_{[j_k j_k]})^{-1} \end{pmatrix}.$$

Unless $\delta\hat{U}^{(k)}$ is properly graded, the best bound for $\delta F^{(k)}$ is

$$\|\delta F^{(k)}\|_2 \leq \|\delta\hat{U}^{(k)}\|_2\kappa_2((A^{(k)}_{[i_k]}, A^{(k)}_{[j_k]})) \leq \|\delta\hat{U}^{(k)}\|_2\kappa_2(A^{(k)}) = \|\delta\hat{U}^{(k)}\|_2\kappa_2(A),$$

where $\kappa_2(\cdot)$ is the spectral condition number. (To understand proper grading, define $F^{(k)}$ by replacing $\delta\hat{U}^{(k)}$ with $\hat{U}^{(k)}$ in the definition of $\delta F^{(k)}$. It is easily seen that $\|F^{(k)}\|_2 \leq 1/\sigma_{\min}((B^{(k)}_{[i_k]}, B^{(k)}_{[j_k]})) \leq \kappa_2(B^{(k)}) \to 1$ $(k \to \infty)$.) Thus, if $\tilde{U}^{(k)}$ is not good enough, the numerical orthogonality of the columns of $A^{(k)}$ (used as stopping criterion in a floating–point computation with roundoff ε) might stagnate at the level of $\varepsilon\kappa_2(A)$, where the problematic pivot pairs are those containing the smallest and the largest singular values. In such cases a cleanup sweep of ordinary 2×2 rotations might help. A particularly important issue for reliable mathematical software is how to implement the transformation in case of underflows due to high condition number of A. Our starting point for this will be [11].

Other issues include choosing the parameter ρ (or monotonically decreasing sequence of thresholds) in (2.1) to achieve optimal reduction of the off–norm, introduction of sorting of the diagonal entries and adapting the block partition to match the eigenvalue distribution in order to achieve higher order of convergence, block quasi–cycling, fast implementation of block transformations as proposed by Hari [18], or finding a convergent parallel strategy and tuning it for a particular parallel architecture.

It is desirable that the high accuracy of the Jacobi–type methods remains preserved under all modifications aimed at high run time performance. And, in fact, if the method is used in a particular way, the only condition for preserving high relative accuracy is that a numerically orthogonal (unitary) transformation $\tilde{Q}$ ($\tilde{Q}^*\tilde{Q} \approx I$) is applied to a vector x in floating–point arithmetic with round–off ε as

(4.3) $computed(\tilde{Q}x) = \hat{Q}(x + \delta x),\ \ \hat{Q}^*\hat{Q} = I,\ \ \|\hat{Q} - \tilde{Q}\|_2 \leq K_1\varepsilon,\ \ \|\delta x\|_2 \leq K_2\varepsilon\|x\|_2,$

where K_1, K_2 are moderate factors that depend on the dimension and implementation details.

Theorem 4.1. *Consider the following block–versions of two algorithms for computing the SVD and the spectral decomposition of Hermitian positive definite matrices:*

A1 *(See* [13], [14].*) SVD of an $r \times n$ matrix A, $\mathrm{rank}(A) = n$:*[5]

(1) $AP = Q\begin{pmatrix} R \\ 0 \end{pmatrix}$ *(QR factorization with optional column pivoting, such as the Businger–Golub pivoting* [2] *implemented as in* [12]*);*

(2) $R^*W = \check{V}\Sigma$ *(one sided Jacobi SVD algorithm with one–sided application of the block transformations (1.5), accumulated in W), under an arbitrary convergent serial or parallel pivot strategy)*

(3) *Output:* $A = U\begin{pmatrix} \Sigma \\ 0 \end{pmatrix}V^*$, *where* $U = Q\begin{pmatrix} W & 0 \\ 0 & I \end{pmatrix}$, $V = P\check{V}$.

A2 *(See* [28].*) Spectral decomposition of a Hermitian positive definite $H \in \mathbb{C}^{n \times n}$:*

---

[5]For the sake of simplicity, we give only a simple version of the algorithm. Similar conclusion holds for the general case of the new preconditioned Jacobi SVD algorithm in [13],[14].

(1) $\mathsf{P}^T\mathsf{H}\mathsf{P} = \mathsf{L}\mathsf{L}^*$ *(Cholesky factorization with optional pivoting)*
(2) $\mathsf{L}\mathsf{W} = \breve{\mathsf{U}}\Sigma$ *one sided Jacobi SVD algorithm with one–sided application of the block transformations 1.5, under an arbitrary serial or parallel convergent pivot strategy, and without computation of the matrix W)*
(3) *Output:* $\mathsf{H} = \mathsf{U}\Lambda\mathsf{U}^*$, *where $\Lambda = \Sigma^2$, $\mathsf{U} = \mathsf{P}\breve{\mathsf{U}}$.*

In an IEEE floating point arithmetic with round–off unit ε, and with an implementation of orthogonal transformations satisfying (4.3), the backward errors $\delta\mathsf{A}$ and $\delta\mathsf{H} = (\delta\mathsf{H})^*$ in the algorithms A1 and A2 are, respectively, bounded by

$$(4.4) \qquad \|\delta\mathsf{A}(:,i)\|_2 \;\; \leq \;\; f(r,n;\varepsilon)\varepsilon\|\mathsf{A}(:,i)\|_2, \;\; i = 1,\ldots,n;$$

$$(4.5) \qquad |\delta\mathsf{H}_{ij}| \;\; \leq \;\; g(n;\varepsilon)\varepsilon\sqrt{\mathsf{H}_{ii}\mathsf{H}_{jj}}, \;\; 1 \leq i,j \leq n.$$

Proof: To simplify the notation, assume that the matrices are already permuted, so that $\mathsf{P} = \mathsf{I}$. The computed upper triangular factor $\tilde{R}$ satisfies $\mathsf{A} + \Delta\mathsf{A} = \hat{\mathsf{Q}}\left(\begin{smallmatrix}\tilde{\mathsf{R}}\\0\end{smallmatrix}\right)$, where for all i, $\|\Delta\mathsf{A}_{:,i}\|_2 \leq q(r,n)\varepsilon\|\mathsf{A}_{:,i}\|_2$, and $\hat{\mathsf{Q}}$ is unitary and close to the actually computed numerically unitary matrix $\tilde{\mathsf{Q}}$. The factor $q(r,n)$ is a moderate polynomial in r,n and it depends on the details of the algorithm. In the second step, after $\mathfrak{s}$ sweeps of the block cyclic Jacobi method, the computed approximation $\tilde{\mathsf{V}}\tilde{\Sigma}$ of $\breve{\mathsf{V}}\Sigma$ satisfies $\tilde{\mathsf{V}}\tilde{\Sigma} = (\tilde{\mathsf{R}} + \delta\tilde{\mathsf{R}})^*\hat{\mathsf{W}}$, where $\tilde{\mathsf{V}}$ is numerically unitary, $\|\delta\tilde{\mathsf{R}}_{:,i}\|_2 \leq h(n)\varepsilon\|\tilde{\mathsf{R}}_{:i}\|_2$ for $i = 1,\ldots,n$, and $\hat{\mathsf{W}}$ is unitary and close to the actually computed numerically unitary $\tilde{\mathsf{W}}$. The latter follows from repeated applications of (4.3), where for a block partition $h = (n_1,\ldots,n_m)$, $h(n) = \mathfrak{s}O(m)K_2$. Due to the preconditioning effect of the first step [13], [14], the sweep counter $\mathfrak{s}$ is small. Since

$$\mathsf{A} + \underbrace{\Delta\mathsf{A} + \hat{\mathsf{Q}}\begin{pmatrix}\hat{\mathsf{W}} & 0\\0 & \mathsf{I}\end{pmatrix}\begin{pmatrix}\delta\tilde{\mathsf{R}}\\0\end{pmatrix}}_{\delta\mathsf{A}} = \hat{\mathsf{Q}}\begin{pmatrix}\hat{\mathsf{W}} & 0\\0 & \mathsf{I}\end{pmatrix}\begin{pmatrix}\tilde{\Sigma}\\0\end{pmatrix}\tilde{\mathsf{V}}^* \approx \tilde{\mathsf{Q}}\begin{pmatrix}\tilde{\mathsf{W}} & 0\\0 & \mathsf{I}\end{pmatrix}\begin{pmatrix}\tilde{\Sigma}\\0\end{pmatrix}\tilde{\mathsf{V}}^*,$$

the claim follows with $f(r,n;\varepsilon) = (q(r,n) + h(n)(1 + q(r,n)\varepsilon))$.

For the Hermitian case, note that the computed Cholesky factor $\tilde{\mathsf{L}} \approx \mathsf{L}$ satisfies $\tilde{\mathsf{L}}\tilde{\mathsf{L}}^* = \mathsf{H} + \Delta\mathsf{H}$ with $|(\Delta\mathsf{H})_{ij}| \leq c(n)\varepsilon\sqrt{\mathsf{H}_{ii}\mathsf{H}_{jj}}$ for all i,j. The factor $c(n)$ is $O(n)$, and if the factorization fails to compute $\tilde{\mathsf{L}}$, the input matrix H is not numerically definite, see [5], [8]. In the second step, the computed $\tilde{\mathsf{U}}\tilde{\Sigma}$ satisfies $\tilde{\mathsf{U}}\tilde{\Sigma} = (\tilde{\mathsf{L}} + \delta\tilde{\mathsf{L}})\hat{\mathsf{W}}$, where $\tilde{\mathsf{U}}$ is numerically unitary, $\hat{\mathsf{W}}$ is unitary and $\|\delta\tilde{\mathsf{L}}_{i,:}\|_2 \leq h(n)\varepsilon\|\tilde{\mathsf{L}}_{i,:}\|_2$ for all i. Let $\tilde{\Lambda} = \tilde{\Sigma}^2$, $\Delta_L\mathsf{H} = \delta\tilde{\mathsf{L}}\tilde{\mathsf{L}}^* + \tilde{\mathsf{L}}\delta\tilde{\mathsf{L}}^* + \delta\tilde{\mathsf{L}}\delta\tilde{\mathsf{L}}^*$, and $\delta\mathsf{H} = \Delta\mathsf{H} + \Delta_L\mathsf{H}$. Then

$$\tilde{\mathsf{U}}\tilde{\Lambda}\tilde{\mathsf{U}}^* = \mathsf{H} + \delta\mathsf{H}, \;\; \max_{i,j}\frac{|\delta\mathsf{H}_{ij}|}{\sqrt{\mathsf{H}_{ii}\mathsf{H}_{jj}}} \leq (c(n) + (2h(n) + h^2(n)\varepsilon)(1 + c(n)\varepsilon))\varepsilon.$$

$$\boxplus$$

Since we have structured form of the backward errors (scaling invariance), state of the art perturbation theory [23], [24] applies, thus guaranteeing accuracy superior to any other method that first tridiagonalizes H or bidiagonalizes A ([8]). This means that we have contrived a well defined framework for future development of a class of block Jacobi methods, and without trading accuracy for speed. It remains to follow through along these lines.

5. Acknowledgments

References

[1] P. Arbenz and M. Oettli, *Block implementations of the symmetric QR and Jacobi algorithms*, Research report 178, Department Informatik, ETH Zürich, June 1992.

[2] P. A. Businger and G. H. Golub, *Linear least squares solutions by Householder transformations*, Numer. Math. **7** (1965), 269–276.

[3] R. L Causey and P. Henrici, *Convergence of approximate eigenvectors in Jacobi methods*, Numerische Mathematik **2** (1960), 67–78.

[4] C. Davis and W. M. Kahan, *The rotation of eigenvectors by perturbation III*, SIAM J. Num. Anal. **7** (1970), 1–46.

[5] J. Demmel, *On floating point errors in Cholesky*, LAPACK Working Note 14, Computer Science Department, University of Tennessee, October 1989.

[6] ———, *Accurate singular value decompositions of structured matrices*, SIAM J. Matrix Anal. Appl. **21** (1999), no. 2, 562–580.

[7] J. Demmel, M. Gu, S. Eisenstat, I. Slapničar, K. Veselić, and Z. Drmač, *Computing the singular value decomposition with high relative accuracy*, Lin. Alg. Appl. **299** (1999), 21–80.

[8] J. Demmel and K. Veselić, *Jacobi's method is more accurate than QR*, SIAM J. Matrix Anal. Appl. **13** (1992), no. 4, 1204–1245.

[9] J. J. Dongarra, J. J. Du Croz, I. Duff, and S. Hammarling, *Algorithm 679: A set of Level 3 Basic Linear Algebra Subprograms: Model implementation and test programs*, ACM Trans. Math. Soft. (1990), 18–28.

[10] ———, *A set of Level 3 Basic Linear Algebra Subprograms*, ACM Trans. Math. Soft. (1990), 1–17.

[11] Z. Drmač, *Implementation of Jacobi rotations for accurate singular value computation in floating point arithmetic*, SIAM J. Sci. Comp. **18** (1997), 1200–1222.

[12] Z. Drmač and Z. Bujanović, *On the failure of rank revealing QR factorization software – a case study*, ACM Transactions on Mathematical Software (2007), ??–??, LAPACK Working Note 176.

[13] Z. Drmač and K. Veselić, *New fast and accurate Jacobi SVD algorithm: I.*, SIAM J. Matrix Anal. Appl. **29** (2008), no. 4, 1322–1342.

[14] ———, *New fast and accurate Jacobi SVD algorithm: II.*, SIAM J. Matrix Anal. Appl. **29** (2008), no. 4, 1343–1362.

[15] A. Edelman, T. Arias, and S. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix. Anal. Appl. **20** (1998), no. 2, 303–353.

[16] G. E. Forsythe and P. Henrici, *The cyclic Jacobi method for computing the principal values of a complex matrix*, Trans. Amer. Math. Soc. **94** (1960), no. 1, 1–23.

[17] E. R. Hansen, *On cyclic Jacobi methods*, Journal of the SIAM **11** (1963), no. 2, 448–459.

[18] V. Hari, *Accelerating the SVD block–Jacobi method*, Computing **75** (2005), 27–53.

[19] ———, *Convergence of a block-oriented quasi-cyclic Jacobi method*, SIAM J. Matrix Anal. Appl. **29** (2007), no. 2, 349–369.

[20] M. R. Hestenes, *Inversion of matrices by biorthogonalization and related results*, J. SIAM **6** (1958), no. 1, 51–90.

[21] C. G. J. Jacobi, *Über ein leichtes Verfahren die in der Theorie der Säcularstörungen vorkommenden Gleichungen numerisch aufzulösen*, Crelle's Journal für reine und angew. Math. **30** (1846), 51–95.

[22] C. L. Lawson and R. J. Hanson, *Solving least squares problems*, Prentice–Hall Inc., Englewood Cliffs, N. J., 1974.

[23] Ren-Cang Li, *Relative perturbation theory: I. Eigenvalue and singular value variations*, SIAM J. Matrix Anal. Appl. **19** (1998), no. 4, 956–982.

[24] ———, *Relative perturbation theory: II. Eigenspace and singular subspace variations*, SIAM J. Matrix Anal. Appl. **20** (1998), no. 2, 471–492.

[25] F. T. Luk and H. Park, *On parallel Jacobi orderings*, SIAM J. Sci. Stat. Comp. **10** (1987), no. 1, 18–26.

[26] ———, *A proof of convergence for two parallel Jacobi SVD algorithms*, IEEE Trans. Comput. **38** (1989), 806–811.

[27] C. Van Loan, *The block Jacobi method for computing the singular value decomposition*, Computational and Combinatorial Methods in Systems Theory (C. Byrnes and A. Lindquist, eds.), Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1986, pp. 245–255.

[28] K. Veselić and V. Hari, *A note on a one–sided Jacobi algorithm*, Numer. Math. **56** (1989), 627–633.

[29] J. H. Wilkinson, *Note of the quadratic convergence of cyclic Jacobi process*, Numer. Math. **4** (1962), 296–300.

Department of Mathematics, Virginia Polytechnic Institute and State University, 460 McBryde, Virginia Tech, Blacksburg, VA 24061-0123.

E-mail address: `zlatko@math.vt.edu`