UNIVERSITÄT SALZBURG

Asymptotic Quadratic Convergence of the Serial Block-Jacobi EVD Algorithm for Hermitian Matrices

Yusaku Yamamoto Gabriel Okša Marián Vajteršic

Technical Report 2015-03 April 2015

Department of Computer Sciences

Jakob-Haringer-Straße 2
5020 Salzburg
Austria
www.cosy.sbg.ac.at

Technical Report Series

Asymptotic Quadratic Convergence of the Serial Block-Jacobi EVD Algorithm for Hermitian Matrices

Yusaku Yamamoto* Gabriel Okša† Marián Vajteršic‡

March 27, 2015

Abstract. *This report is devoted to the proof of the local (asymptotic) quadratic convergence of the serial block-Jacobi EVD algorithm for Hermitian matrices with multiple eigenvalues. At each iteration step, two off-diagonal blocks with the largest Frobenius norm are annihilated which is an extension of the 'classical' Jacobi approach to the block case.*

1 Introduction

In this report, we consider asymptotic convergence property of the 'classical' block Jacobi method for the symmetric eigenvalue problem. Let A be an $n \times n$ symmetric matrix and L be the block size which divides n. We divide A into blocks of size $L \times L$ and denote the (I, J)th block by A_{IJ}. We also denote the number of blocks in each direction by w $(= n/L)$ and the number of off-diagonal blocks in the upper triangular part by W $(= w(w-1)/n)$. Let $A^{(0)} = A$. In the kth step of the block Jacobi method, we choose the off-diagonal block $A^{(k)}_{X_k Y_k}$ with the largest Frobenius norm (F-norm) and consider the following $2L \times 2L$ pivot submatrix:

$$\tilde{A}^{(k)} = \begin{bmatrix} A^{(k)}_{X_k X_k} & A^{(k)}_{X_k Y_k} \\ A^{(k)}_{Y_k X_k} & A^{(k)}_{Y_k Y_k} \end{bmatrix}. \quad (1)$$

Now we consider the $2L \times 2L$ orthogonal matrix

$$\tilde{P}^{(k)} = \begin{bmatrix} P^{(k)}_{X_k X_k} & P^{(k)}_{X_k Y_k} \\ P^{(k)}_{Y_k X_k} & P^{(k)}_{Y_k Y_k} \end{bmatrix} \quad (2)$$

*Department of Communication Engineering and Informatics, University of Electro-Communications, Tokyo, Japan, email: yusaku.yamamoto@uec.ac.jp

†Mathematical Institute, Slovak Academy of Sciences, Bratislava, Slovak Republic, email: gabriel.oksa@savba.sk

‡Department of Computer Sciences, University of Salzburg, Austria, email: marian@cosy.sbg.ac.at

that diagonalizes $\tilde{A}^{(k)}$, embed it in the $n \times n$ identity matrix I_n to construct an $n \times n$ orthogonal matrix $P^{(k)}$, and transform the matrix $A^{(k)}$ into $A^{(k+1)}$ by the following orthogonal transformation:

$$A^{(k+1)} = \left(P^{(k)}\right)^\top A^{(k)} P^{(k)}. \tag{3}$$

Then the diagonal blocks $A^{(k+1)}_{X_k X_k}$ and $A^{(k+1)}_{Y_k Y_k}$ become diagonal, and off-diagonal blocks $A^{(k+1)}_{X_k Y_k}$ and $A^{(k+1)}_{Y_k X_k}$ become zero. This transformation reduces the sum of squares of the off-diagonal elements of $A^{(k)}$, as in the 'classical' point Jacobi method. Thus it can be shown that $A^{(k)}$ converges to a block diagonal matrix with block size L as $k \to \infty$ [9]. In this report, we make the following assumption.

A1 There exists an integer $k_0 > 0$ such that before the start of the k_0th step, at least one off-diagonal block for each row has been chosen as the pivot off-diagonal block.

This assumption ensures that all diagonal blocks of $A^{(k)}$ are diagonal for $k \geq k_0$. In the following, we shift the iteration count and assume that **A1** holds with $k_0 = 0$. Now, we introduce the square of the off-diagonal norm of $A^{(k)}$ by

$$\|\mathrm{off}(A^{(k)})\|_F^2 = \sum_{I \neq J} \|A^{(k)}_{IJ}\|_F^2. \tag{4}$$

Then it holds immediately that

$$\|\mathrm{off}(A^{(k+1)})\|_F^2 = \|\mathrm{off}(A^{(k)})\|_F^2 - 2\|A^{(k)}_{X_k Y_k}\|_F^2. \tag{5}$$

Thus $\|\mathrm{off}(A^{(k)})\|_F^2$ is a monotonically decreasing sequence and converges to zero [9].

Using these preliminaries, we investigate asymptotic convergence property of the block Jacobi method. When A has only simple eigenvalues, it has been shown in [9] that the method is ultimately quadratically convergent. However, the coefficient given there leaves much room for improvement. It is also to be noted that asymptotic convergence rate of the block cyclic Jacobi method is shown to be quadratic if the eigenvalues are simple [2]. Note that the argument there assumes that the order of elimination is either column or row cyclic, so it cannot be readily applied to the case of block Jacobi method.

In this report, we first deal with the case of simple eigenvalues. Our proof is a generalization of the quadratic convergence proof of the 'classical' point Jacobi method by Scönhage [6][7][8] to the block case and gives much better bound than the one given in [9]. We next consider the case of multiple eigenvalues, although with some restrictions. To be more precise, note that the diagonal elements of $A^{(k)}$ tend to some eigenvalues of A. Then we make the following assumption.

A2 For each eigenvalue λ_i with multiplicity greater than 1, all the diagonal elements approximating λ_i lie in the same diagonal block of $A^{(k)}$.

This assumption will be considered more precisely in Section 3. Although it is somewhat restrictive, it seems difficult to prove quadratic convergence of the block Jacobi method, or at least the method in its original form, in the presence of the multiple eigenvalues, when these are placed across the blocks. In Section 4, we will present an example to show what occurs if the assumption **A2** is violated. Note that even if $A^{(k)}$ does not satisfy **A2**, it is sometimes possible to make it satisfy **A2** by permuting its rows and columns.

In the following, we sometimes drop the superscript (k) when there is no reason of misunderstanding. In that case, we use quantities with hat (like $\hat{A}$) to denote them at the $(k+1)$th step.

2 The case of simple eigenvalues

In this section, we assume that A has simple eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. We define the minimum gap between different eigenvalues by $d = \min_{i \neq j} |\lambda_i - \lambda_j|$. In this case, it can be shown that the block Jacobi method is ultimately quadratically convergent. Our proof closely follows the proof of quadratic convergence of the 'classical' point Jacobi method given by Schönhage [6][8], with some lemmas generalized for the block case.

The first lemma, which is a generalization of the lemma given in [8] as Exercise 8.6, is one of the keys in our proof.

Lemma 1 Consider the change of an off-diagonal block A_{XJ} $(J \neq X, Y)$ after elimination of A_{XY}:

$$\hat{A}_{XJ} = P_{XX}^{\top} A_{XJ} + P_{YX}^{\top} A_{YJ}. \tag{6}$$

If P_{YX} is bounded as $\|P_{YX}\|_2 \leq \|A_{XY}\|_F / \delta$ for some $\delta > 0$, then the following inequality holds:

$$\left| \|\hat{A}_{XJ}\|_F^2 - \|A_{XJ}\|_F^2 \right| \leq \frac{\|A_{XY}\|_F^4}{\delta^2} + 2\frac{\|A_{XY}\|_F^2}{\delta} \|A_{XJ}\|_F. \tag{7}$$

Proof. Since $P_{XX}^{\top} P_{XX} + P_{YX}^{\top} P_{YX} = I_L$, we have $\|P_{XX}\|_2^2 + \|P_{YX}\|_2^2 \geq 1$. Thus it follows that $0 \leq 1 - \|P_{XX}\|_2^2 \leq \|P_{YX}\|_2^2$. Now, from the definition of $\hat{A}_{XJ}$, we have

$$\begin{aligned}
&\left| \|\hat{A}_{XJ}\|_F^2 - \|A_{XJ}\|_F^2 \right| \\
&\quad \leq \left| \|P_{XX}^{\top} A_{XJ} + P_{YX}^{\top} A_{YJ}\|_F^2 - \|A_{XJ}\|_F^2 \right| \\
&\quad \leq \left| (\|P_{XX}\|_2 \|A_{XJ}\|_F + \|P_{YX}\|_2 \|A_{YJ}\|_F)^2 - \|A_{XJ}\|_F^2 \right| \\
&\quad \leq \left| (1 - \|P_{XX}\|_2^2) \|A_{XJ}\|_F^2 - \|P_{YX}\|_2^2 \|A_{YJ}\|_F^2 \right| + 2\|P_{XX}\|_2 \|P_{YX}\|_2 \|A_{XJ}\|_F \|A_{YJ}\|_F \\
&\quad \leq \max((1 - \|P_{XX}\|_2^2) \|A_{XJ}\|_F^2, \|P_{YX}\|_2^2 \|A_{YJ}\|_F^2) + 2\|P_{YX}\|_2 \|A_{XJ}\|_F \|A_{YJ}\|_F \\
&\quad \leq \|P_{YX}\|_2^2 \|A_{XY}\|_F^2 + 2\|P_{YX}\|_2 \|A_{XJ}\|_F \|A_{XY}\|_F \\
&\quad \leq \frac{\|A_{XY}\|_F^4}{\delta^2} + 2\frac{\|A_{XY}\|_F^2}{\delta} \|A_{XJ}\|_F,
\end{aligned} \tag{8}$$

where we used $|a - b| \leq \max(a, b)$ for two nonnegative numbers a and b and $\|P_{XX}\|_2 \leq 1$ in the fourth inequality. In the fifth inequality, we used $1 - \|P_{XX}\|_2^2 \leq \|P_{YX}\|_2^2$ and the fact that A_{XY} is the off-diagonal block with the largest F-norm. (Q.E.D.)

Henceafter, we can apply the logic of the proof for the 'classical' point Jacobi method to the block case in a straightforward manner, except for the evaluation of δ, which will be done in Lemma 3.

Lemma 2 Consider one sweep ($W = w(w-1)/2$ eliminations) of the block Jacobi method. Without loss of generality, we rename the step number as $k = 0, 1, \ldots, W-1$ and denote the off-diagonal block chosen at step k as $A_{X_k Y_k}$. If there exists $\delta > 0$ and all the orthogonal matrices used at steps $k = 0, 1, \ldots, W-1$ satisfy $\|P^{(k)}_{Y_k X_k}\|_2 \leq \|A^{(k)}_{X_k Y_k}\|_F / \delta$, then

$$\|\mathrm{off}(A^{(W)})\|_F^2 \leq \frac{w-2}{2} \left(\frac{\|\mathrm{off}(A^{(0)})\|_F^2}{\delta} \right)^2, \tag{9}$$

that is, the method converges quadratically after every sweep.

Proof. We show that for each $k = 0, 1, \ldots, W$, there exists an index set $\mathcal{P}_k = \{(I,J) | I \neq J\}$ such that $|\mathcal{P}_k| = 2k$ and

$$\sum_{(I,J)\in\mathcal{P}_k} \|A^{(k)}_{IJ}\|_F^2 \leq \frac{w-2}{2} \left(\frac{\|\mathrm{off}(A^{(0)})\|_F^2 - \|\mathrm{off}(A^{(k)})\|_F^2}{\delta} \right)^2. \tag{10}$$

Note that when $k = W$, the left-hand side becomes $\|\mathrm{off}(A^{(W)})\|_F^2$, and the right-hand side is smaller than the right-hand side of Eq. (9). So it is sufficient to prove Eq. (10) instead of Eq. (9). We show Eq. (10) by induction. When $k = 0$, it holds trivially because the both sides are zero. We assume that Eq. (10) holds for some k ($0 \leq k < W$) and show that it also holds for $k+1$.

Let us choose the $2k$ smallest off-diagonal blocks of $A^{(k)}$ and denote their index sets by $\mathcal{P}'_k$. Then Eq. (10) holds also for $\mathcal{P}'_k$. Note that we can choose $\mathcal{P}'_k$ so that it is symmetric, i.e., if $(I,J) \in \mathcal{P}'_k$ then $(J,I) \in \mathcal{P}'_k$, and $(X_k, Y_k) \notin \mathcal{P}'_k$. Now, let $\mathcal{P}_{k+1} = \mathcal{P}'_k \cup \{(X_k, Y_k), (Y_k, X_k)\}$. Then $|\mathcal{P}_{k+1}| = 2(k+1)$ and the left-hand side of Eq. (10) for $k+1$ can be computed as

$$\begin{aligned} \sum_{(I,J)\in\mathcal{P}_{k+1}} \|A^{(k+1)}_{IJ}\|_F^2 &= \sum_{(I,J)\in\mathcal{P}'_k} \|A^{(k+1)}_{IJ}\|_F^2 + \|A^{(k+1)}_{X_k Y_k}\|_F^2 + \|A^{(k+1)}_{Y_k X_k}\|_F^2 \\ &= \sum_{(I,J)\in\mathcal{P}'_k} \|A^{(k)}_{IJ}\|_F^2 + \sum_{(I,J)\in\mathcal{Q}_k} \left[\|A^{(k+1)}_{IJ}\|_F^2 - \|A^{(k)}_{IJ}\|_F^2 \right], \end{aligned} \tag{11}$$

where the index set $\mathcal{Q}_k$ is defined as

$$\mathcal{Q}_k = \{(X_k, J), (J, X_k) | (X_k, J) \in \mathcal{P}'_k, (Y_k, J) \notin \mathcal{P}'_k\} \cup \{(Y_k, J), (J, Y_k) | (Y_k, J) \in \mathcal{P}'_k, (X_k, J) \notin \mathcal{P}'_k\}. \tag{12}$$

To derive the second equality, we used the fact that both $A^{(k+1)}_{X_k Y_k}$ and $A^{(k+1)}_{Y_k X_k}$ become zero due to elimination. We also used the fact that when $(I,J) \in \mathcal{P}'_k \backslash \mathcal{Q}_k$, either $A^{(k)}_{IJ}$ is not affected by the elimination, or both of (X_k, J) and (Y_k, J) (or (J, X_k) and J, Y_k) belong to $\mathcal{P}'_k$ and therefore the sum of squares of the F-norms of these two blocks is invariant after elimination. Hence, the change of $A^{(k)}_{IJ}$ contributes to the change of $\sum_{(I,J)\in\mathcal{P}'_k} \|A^{(k)}_{IJ}\|_F^2$ only when $(I,J) \in \mathcal{Q}_k$.

Now we evaluate the second term of Eq. (11). Let us consider the case of $I = X_k$ and $J \neq X_k, Y_k$. From the assumption $\|P^{(k)}_{Y_k X_k}\|_2 \leq \|A^{(k)}_{X_k Y_k}\|_F/\delta$ and Lemma 1, we have

$$\left| \|A^{(k+1)}_{X_k J}\|_F^2 - \|A^{(k)}_{X_k J}\|_F^2 \right| \leq \frac{\|A^{(k)}_{X_k Y_k}\|_F^4}{\delta^2} + 2\frac{\|A^{(k)}_{X_k Y_k}\|_F^2}{\delta}\|A^{(k)}_{X_k J}\|_F. \tag{13}$$

Other cases can be treated in the same way. Noting that $|\mathcal{Q}_k| < 2w-4$ (since only one of (X_k, J) and (Y_k, J) (or (J, X_k) and (J, Y_k)) can belong to $\mathcal{Q}_k$), we have

$$\sum_{(I,J)\in\mathcal{Q}_k} \|A^{(k)}_{IJ}\|_F \leq \sqrt{2w-4}\sqrt{\sum_{(I,J)\in\mathcal{Q}_k} \|A^{(k)}_{IJ}\|_F^2} \leq \sqrt{2w-4}\sqrt{\sum_{(I,J)\in\mathcal{P}'_k} \|A^{(k)}_{IJ}\|_F^2}, \tag{14}$$

where we used the Cauchy-Schwarz inequality in the first inequality and $\mathcal{Q}_k \subseteq \mathcal{P}'_k$ in the second inequality. By combining Eqs. (13) and (14), we can evaluate the second term of Eq. (11) as

$$\sum_{(I,J)\in\mathcal{Q}_k} \left[\|A^{(k+1)}_{IJ}\|_F^2 - \|A^{(k)}_{IJ}\|_F^2\right] \leq \frac{(2w-4)\,\|A^{(k)}_{X_k Y_k}\|_F^4}{\delta^2} + \frac{2\sqrt{2w-4}\,\|A^{(k)}_{X_k Y_k}\|_F^2}{\delta}\sqrt{\sum_{(I,J)\in\mathcal{P}'_k} \|A^{(k)}_{IJ}\|_F^2}. \tag{15}$$

Inserting this into Eq. (11) and using the relationship $\|\mathrm{off}(A^{(k)})\|_F^2 - 2\|A^{(k)}_{X_k Y_k}\|_F^2 = \|\mathrm{off}(A^{(k+1)})\|_F^2$ leads to

$$\begin{aligned}
\sum_{(I,J)\in\mathcal{P}_{k+1}} \|A^{(k+1)}_{IJ}\|_F^2 &\leq \left(\frac{\sqrt{2w-4}\,\|A^{(k)}_{X_k Y_k}\|_F^2}{\delta} + \sqrt{\sum_{(I,J)\in\mathcal{P}'_k} \|A^{(k)}_{IJ}\|_F^2}\right)^2 \\
&\leq \left(\frac{\sqrt{2w-4}\,\|A^{(k)}_{X_k Y_k}\|_F^2}{\delta} + \frac{\sqrt{2w-4}\,[\|\mathrm{off}(A^{(0)})\|_F^2 - \|\mathrm{off}(A^{(k)})\|_F^2]}{2\delta}\right)^2 \\
&= \frac{w-2}{2}\left(\frac{\|\mathrm{off}(A^{(0)})\|_F^2 - \|\mathrm{off}(A^{(k+1)})\|_F^2}{\delta}\right)^2.
\end{aligned} \tag{16}$$

Here we used Eq. (10) for $\mathcal{P}'_k$ in the second inequality. This shows that Eq. (10) holds also for $k+1$ and thus completes the proof. (Q.E.D.)

The final task is to show that the $\delta > 0$ used in Lemmas 1 and 2 actually exists and evaluate its value.

Lemma 3 Assume that $d - 4\|\mathrm{off}(A^{(k)})\|_F > 0$ holds for some k. Then the $2L \times 2L$ orthogonal matrix $\tilde{P}$ produced in this step satisfies $\|P_{YX}\|_2 \leq \|A_{XY}\|_F/(d - 2\|\mathrm{off}(A^{(k)})\|_F)$.

Proof. ¿From the Wielandt-Hoffman theorem [4, 5] applied to A, there exists a permutation $\sigma \in S_n$ such that $|a_{ii} - \lambda_{\sigma(i)}| \leq \|\mathrm{off}(A^{(k)})\|_F$ for $i = 1, \ldots, n$. We denote the eigenvalue paired with $\tilde{a}_{ii}$ by $\lambda_{\iota(i)}$, where $\iota(i)$ is an injective mapping from $\{1, \ldots, 2L\}$ to $\{1, \ldots, n\}$. Now, let the eigenvalues of the $2L \times 2L$ pivot submatrix $\tilde{A}$ be $\tilde{\mu}_i$ ($i = 1, 2, \ldots, 2L$) and apply the Wielandt-Hoffman theorem to $\tilde{A}$. Then, by renumbering $\{\tilde{\mu}_i\}$, we have $|\tilde{a}_{ii} - \tilde{\mu}_i| \leq \|\mathrm{off}(A^{(k)})\|_F$ for $i = 1, \ldots, 2L$. By combining these results, we have

$$|\tilde{\mu}_i - \lambda_{\iota(i)}| \leq |\tilde{a}_{ii} - \tilde{\mu}_i| + |\tilde{a}_{ii} - \lambda_{\iota(i)}| \leq 2\|\mathrm{off}(A^{(k)})\|_F, \quad i = 1, \ldots, 2L. \tag{17}$$

Thus $\tilde{\mu}_i \in \bar{C}_{\iota(i)}$, where $\bar{C}_l$ is the interval with center λ_l and width $4\|\mathrm{off}(A^{(k)})\|_F$. Since $\bar{C}_1, \ldots, \bar{C}_n$ are disjoint, $\tilde{\mu}_i \notin \bar{C}_l$ if $l \neq \iota(i)$.

Now we consider an $n \times n$ matrix $(\bigoplus_{I \neq X,Y} A_{II}) \oplus \tilde{A}$. A can be regarded as a perturbation of this matrix and the F-norm of the perturbation is smaller than or equal to $\|\mathrm{off}(A^{(k)})\|_F$. Hence, by applying the Wielandt-Hoffman theorem again, we know that $\tilde{\mu}_i \in C_l$ for some $l \in \{1, \ldots, n\}$, where C_l is the interval with center λ_l and width $2\|\mathrm{off}(A^{(k)})\|_F$. But because $C_l \subset \bar{C}_l$ and $\tilde{\mu}_i \notin \bar{C}_l$ if $l \neq \iota(i)$, $\tilde{\mu}_i \notin C_l$ if $l \neq \iota(i)$. Thus $\tilde{\mu}_i$ must be contained in $C_{\iota(i)}$ and we obtain a tighter bound than Eq. (17):

$$|\tilde{\mu}_i - \lambda_{\iota(i)}| \leq \|\mathrm{off}(A^{(k)})\|_F. \tag{18}$$

By combining this with $|\tilde{a}_{ii} - \lambda_{\iota(i)}| \leq \|\mathrm{off}(A^{(k)})\|_F$, we have

$$|\tilde{a}_{ii} - \tilde{\mu}_j| \geq |\lambda_{\iota(i)} - \lambda_{\iota(j)}| - |\tilde{a}_{ii} - \lambda_{\iota(i)}| - |\tilde{\mu}_i - \lambda_{\iota(i)}| \geq d - 2\|\mathrm{off}(A^{(k)})\|_F > 0 \quad \text{for} \quad i \neq j. \tag{19}$$

Now, let the eigenvector of $\tilde{A}$ belonging to $\tilde{\mu}_j$ be $\tilde{\mathbf{p}}_j$. Also, denote the jth column of I_{2L} by $\tilde{\mathbf{e}}_j$. Then, the angle θ_j between $\tilde{\mathbf{p}}_j$ and $\tilde{\mathbf{e}}_j$ can be bounded by the $\sin\theta$ theorem [1, 5] as

$$|\sin\theta_j| \leq \frac{\|\tilde{A}\tilde{\mathbf{e}}_j - (\tilde{\mathbf{e}}_j^\top \tilde{A}\tilde{\mathbf{e}}_j)\tilde{\mathbf{e}}_j\|}{\min_{i\neq j} |(\tilde{\mathbf{e}}_j^\top \tilde{A}\tilde{\mathbf{e}}_j) - \tilde{\mu}_i|} = \frac{\sqrt{\sum_{i\neq j} \tilde{a}_{ij}^2}}{\min_{i\neq j} |\tilde{a}_{jj} - \tilde{\mu}_i|} \leq \frac{1}{d - 2\|\mathrm{off}(A^{(k)})\|_F} \sqrt{\sum_{i\neq j} \tilde{a}_{ij}^2} \quad (1 \leq j \leq L). \tag{20}$$

On the other hand, for each j $(1 \leq j \leq L)$,

$$\sqrt{\sum_{i\neq j} \tilde{p}_{ij}} = \sqrt{1 - \tilde{p}_{jj}^2} = \sqrt{1 - (\tilde{\mathbf{p}}_j \cdot \tilde{\mathbf{e}}_j)^2} = |\sin\theta_j|. \tag{21}$$

¿From Eqs. (20) and (21), we have

$$\begin{aligned} \|P_{YX}\|_2 \leq \|P_{YX}\|_F &= \sqrt{\sum_{j=1}^{L} \sum_{i=L+1}^{2L} \tilde{p}_{ij}^2} \leq \sqrt{\sum_{j=1}^{L} \sum_{i\neq j} \tilde{p}_{ij}^2} = \sqrt{\sum_{j=1}^{L} \sin^2\theta_j} \\ &\leq \frac{1}{d - 2\|\mathrm{off}(A^{(k)})\|_F} \sqrt{\sum_{j=1}^{L} \sum_{i\neq j} \tilde{a}_{ij}^2} = \frac{\|A_{XY}\|_F}{d - 2\|\mathrm{off}(A^{(k)})\|_F}, \end{aligned} \tag{22}$$

where we used the fact that the diagonal blocks of $\tilde{A}$ are diagonal in the last equality. (Q.E.D.)

By combining Lemma 2 and 3, we finally obtain a theorem on quadratic convergence of the block Jacobi method in the case of simple eigenvalues.

Theorem 1 Assume that A has simple eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and let $d = \min_{i\neq j} |\lambda_i - \lambda_j|$. Also, assume that $d - 4\|\mathrm{off}(A^{(k_0)})\|_F > 0$ holds for some k_0. Then, for $k \geq k_0$,

$$\|\mathrm{off}(A^{(k+W)})\|_F^2 \leq \frac{w-2}{2} \left(\frac{\|\mathrm{off}(A^{(k)})\|_F^2}{d - 2\|\mathrm{off}(A^{(k)})\|_F} \right)^2, \tag{23}$$

that is, the block Jacobi method converges quadratically after every sweep.

Proof. Since the off-diagonal norm is monotonically decreasing, we have $d-4\|\mathrm{off}(A^{(k)})\|_F > 0$ for $k \geq k_0$. Thus, if we let $\delta = d - 2\|\mathrm{off}(A^{(k)})\|_F$, all the orthogonal matrices used at steps $k = k_0, k_0+1, \ldots$ satisfy $\|P^{(k)}_{Y_k X_k}\|_2 \leq \|A^{(k)}_{XY}\|_F/\delta$ from Lemma 3. Now we can apply Lemma 2 with $\delta = d - 2\|\mathrm{off}(A^{(k)})\|_F$ and get the desired result.

3 The case of multiple eigenvalues

Next we consider the case of multiple eigenvalues. Assume that there are q different eigenvalues:

$$\lambda_1 = \cdots = \lambda_{s_1} > \lambda_{s_1+1} = \cdots = \lambda_{s_2} > \cdots > \lambda_{s_{q-1}+1} = \cdots \lambda_n, \tag{24}$$

where $n_l = s_l - s_{l-1}$, $1 \leq l \leq q$ is the algebraic multiplicity of λ_{s_l} (defining $s_0 = 0$ and $s_q = n$). Let $d = \min_{i \neq j} |\lambda_{s_i} - \lambda_{s_j}|$ and assume that $d - 4\|\mathrm{off}(A^{(k)})\|_F > 0$ holds for some k. Also, let C_l ($1 \leq l \leq q$) be an interval with center λ_{s_l} and width $2\|\mathrm{off}(A^{(k)})\|_F$. These q intervals are mutually disjoint. Moreover, from the Wielandt-Hoffman theorem, C_l contains exactly n_l diagonal elements of $A^{(k)}$. Now, we reformulate the assumption **A2** from the Introduction as follows:

A2' The set of diagonal elements belonging to the same interval lie in the same diagonal block of $A^{(k)}$.

Hence, if $a^{(k)}_{ii}$ and $a^{(k)}_{jj}$ belong to different blocks, $|a^{(k)}_{ii} - a^{(k)}_{jj}| \geq d - 2\|\mathrm{off}(A^{(k)})\|_F$. Now we can prove the following lemma on the bound of $\|P_{YX}\|_F$.

Lemma 4 Under the assumptions of $d - 4\|\mathrm{off}(A^{(k)})\|_F > 0$ and **A2'**, the $2L \times 2L$ orthogonal matrix $\tilde{P}$ produced at step k satisfies $\|P_{YX}\|_F \leq \|A_{XY}\|_F/(d - 2\|\mathrm{off}(A^{(k)})\|_F)$.

Proof. Let us consider the $2L \times 2L$ pivot submatrix $\tilde{A}$ of $A^{(k)}$ and denote its diagonal elements and eigenvalues by $\tilde{a}_{ii}$ and $\tilde{\mu}_i$ ($i = 1, 2, \ldots, 2L$), respectively. By applying the Wielandt-Hoffman theorem to $\tilde{A}$, we know that there exists a permutation $\sigma \in S_{2L}$ such that $|\tilde{a}_{ii} - \tilde{\mu}_{\sigma(i)}| \leq \|\mathrm{off}(A^{(k)})\|_F$ for $i = 1, 2, \ldots, 2L$. Thus we renumber the eigenvalues so that $|\tilde{a}_{ii} - \tilde{\mu}_i| \leq \|\mathrm{off}(A^{(k)})\|_F$ holds for $i = 1, 2, \ldots, 2L$. If $\tilde{a}_{ii}$ belongs to an interval C_l, then $\tilde{\mu}_i$ belongs to a larger interval $\bar{C}_l$, which has the same center as C_l but has width $4\|\mathrm{off}(A^{(k)})\|_F$. Actually, we can show that $\tilde{\mu}_i$ belongs to C_k. To see this, note that the distance from $\tilde{\mu}_i$ to the center of another interval, which is an eigenvalue of $A^{(k)}$, is larger than $d-2\|\mathrm{off}(A^{(k)})\|_F > 2\|\mathrm{off}(A^{(k)})\|_F$. But by the Wieland-Hoffman theorem applied to an $n \times n$ matrix $(\bigoplus_{I \neq X,Y} A_{II}) \oplus \tilde{A}$ (see the proof of Lemma 3), the distance from $\tilde{\mu}_i$ to the nearest eigenvalue of $A^{(k)}$ is smaller than or equal to $\|\mathrm{off}(A^{(k)})\|_F$. This is possible only if $\tilde{\mu}_i$ is contained in C_l. From assumption **A2'**, the set of C_l's to which $\tilde{\mu}_1, \ldots, \tilde{\mu}_L$ belong and the set of C_l's to which $\tilde{a}_{L+1,L+1}, \ldots, \tilde{a}_{2L,2L}$ belong are disjoint. Hence we have the following inequality.

$$|\tilde{\mu}_j - \tilde{a}_{i+L,i+L}| \geq d - 2\|\mathrm{off}(A^{(k)})\|_F, \quad i, j = 1, \ldots, L. \tag{25}$$

We can use this inequality to bound $\|P_{YX}\|_F$ in the case of multiple eigenvalues. To this end, we could use the $\sin\Theta$ theorem for invariant subspaces [1], but in this report, we take a more elementary approach based on the Sylvester equation. We can write the eigendecomposition of $\tilde{A}$ in the following form.

$$\begin{bmatrix} A_{XX} & A_{XY} \\ A_{YX} & A_{YY} \end{bmatrix} \begin{bmatrix} P_{XX} & P_{XY} \\ P_{YX} & P_{YY} \end{bmatrix} = \begin{bmatrix} P_{XX} & P_{XY} \\ P_{YX} & P_{YY} \end{bmatrix} \begin{bmatrix} M_{XX} & O \\ O & M_{YY} \end{bmatrix}, \tag{26}$$

where $M_{XX} = \mathrm{diag}(\tilde{\mu}_1, \ldots, \tilde{\mu}_L)$ and $M_{YY} = \mathrm{diag}(\tilde{\mu}_{L+1}, \ldots, \tilde{\mu}_{2L})$. Note that A_{XX} and A_{YY} are also diagonal. By focusing on the $(2,1)$th block, we have the following Sylvester equation for P_{YX}:

$$P_{YX} M_{XX} - A_{YY} P_{YX} = A_{YX} P_{XX}. \tag{27}$$

This can easily be solved because both M_{XX} and A_{YY} are diagonal. By writing $A_{YX}P_{XX} = C = (c_{ij})$, we have

$$\tilde{p}_{i+L,j}\tilde{\mu}_j - \tilde{a}_{i+L,i+L}\tilde{p}_{i+L,j} = c_{ij}, \quad i, j = 1, \ldots, L. \tag{28}$$

By Eq. (25) the coefficient of $\tilde{p}_{i+L,j}$ is nonzero and the solution can be written explicitly as

$$\tilde{p}_{i+L,j} = \frac{c_{ij}}{\tilde{\mu}_j - \tilde{a}_{i+L,i+L}}. \tag{29}$$

¿From this expression, the bound on $\|P_{YX}\|_F$ is immediate:

$$\begin{aligned} \|P_{YX}\|_F^2 = \sum_{i=1}^{L}\sum_{j=1}^{L} p_{i+L,j}^2 &= \sum_{i=1}^{L}\sum_{j=1}^{L} \left(\frac{c_{ij}}{\tilde{\mu}_j - \tilde{a}_{i+L,i+L}} \right)^2 \\ &\leq \frac{1}{(d - 2\|\mathrm{off}(A^{(k)})\|_F)^2} \sum_{i=1}^{L}\sum_{j=1}^{L} c_{ij}^2 \\ &= \frac{\|C\|_F^2}{(d - 2\|\mathrm{off}(A^{(k)})\|_F)^2} \leq \frac{\|P_{XX}\|_2 \|A_{YX}\|_F^2}{(d - 2\|\mathrm{off}(A^{(k)})\|_F)^2} \\ &\leq \frac{\|A_{YX}\|_F^2}{(d - 2\|\mathrm{off}(A^{(k)})\|_F)^2}, \end{aligned} \tag{30}$$

where we used Eq. (25) in the second line and $\|P_{XX}\|_2 \leq 1$ in the last inequality. (Q.E.D.)

By combining this result with Lemma 2, we obtain a theorem of quadratic convergence for the case of multiple eigenvalues.

Theorem 2 Assume that the eigenvalues are given as Eq. (24) and let $d = \min_{i \neq j} |\lambda_{s_i} - \lambda_{s_j}|$. Also, assume that $d - 4\|\mathrm{off}(A^{(k_0)})\|_F > 0$ holds for some k_0. Assume further that **A2'** holds. Then, for $k \geq k_0$,

$$\|\mathrm{off}(A^{(k+W)})\|_F^2 \leq \frac{w-2}{2} \left(\frac{\|\mathrm{off}(A^{(k)})\|_F^2}{d - 2\|\mathrm{off}(A^{(k)})\|_F} \right)^2, \tag{31}$$

that is, the block Jacobi method converges quadratically after every sweep.

Proof. The proof is the same as that of Theorem 1, except that Lemma 4 is used instead of Lemma 3. Note that Lemma 2 makes no assumption on the eigenvalue distribution of A, so it can be used even in the presence of multiple eigenvalues. Note also that although Lemma 4 gives a bound on $\|P_{YX}\|_F$, the same bound holds also for $\|P_{YX}\|_2$ because $\|P_{YX}\|_2 \le \|P_{YX}\|_F$. (Q.E.D.)

In concluding this section, we make a note on the difference between point and block Jacobi methods in the presence of multiple eigenvalues. In the case of point Jacobi method, if $\|\mathrm{off}(A^{(k)})\|_F^2 = O(\epsilon)$, it can be shown that the off-diagonal element $a_{ij}^{(k)}$ corresponding to the two diagonal elements $a_{ii}^{(k)}$ and $a_{jj}^{(k)}$ approximating the same eigenvalue of A is $O(\epsilon^2)$ [5]. So we do not need to eliminate $a_{ij}^{(k)}$ to guarantee quadratic convergence.

On the other hand, in the block case, there are multiple diagonal elements in each diagonal block. So it can occur that the diagonal elements $a_{ii}^{(k)}$ and $a_{jj}^{(k)}$ contained in diagonal blocks $A_{11}^{(k)}$ and $A_{22}^{(k)}$, respectively, approximate the same eigenvalue, say λ_1, while a_{kk} contained in block $A_{11}^{(k)}$ approximates a different eigenvalue, say, λ_2. In that case, a_{jk} is generally not small, so the block A_{12} needs to be eliminated. But $a_{ii}^{(k)}$ and $a_{jj}^{(k)}$ can become arbitrarily close, and this can give rise to a large element in P_{YX}. Assumption **A2'** gives a sufficient condition for such a situation not to occur.

4 Example of a situation in which $\|P_{YX}\|_2$ is not small

We give an example which illustrates what can occur if the assumption **A2'** does not hold. In that case, we can show that $\|P_{YX}\|_2$ can be large even if $\|A_{XY}\|_F$ is small.

Consider an 8×8 symmetric matrix

$$A = \left[\begin{array}{cc|cc|cc|cc}
6 & & & & & & & \\
 & 4 & -s^2 & & & & -\frac{sc}{\sqrt{2}} & \frac{sc}{\sqrt{2}} \\
\hline
 & -s^2 & 4 & & & & -\frac{sc}{\sqrt{2}} & -\frac{sc}{\sqrt{2}} \\
 & & & 2 & & & & \\
\hline
 & & & & 1 & & & \\
 & -\frac{sc}{\sqrt{2}} & -\frac{sc}{\sqrt{2}} & & & 3+s^2 & & \\
\hline
 & \frac{sc}{\sqrt{2}} & -\frac{sc}{\sqrt{2}} & & & & 5-s^2 & \\
 & & & & & & & 7
\end{array}\right], \tag{32}$$

where $0 < s \ll 1$ and $c = \sqrt{1-s^2}$. All the off-diagonal elements of this matrix are $O(s)$. This matrix has the eigenvalues

$$\lambda_1 = 6, \quad \lambda_2 = \lambda_3 = 4, \quad \lambda_4 = 2, \quad \lambda_5 = 1, \quad \lambda_6 = 3, \quad \lambda_7 = 5, \quad \lambda_8 = 7, \tag{33}$$

so it has *multiple* (not clustered) eigenvalues $\lambda_2 = \lambda_3 = 4$. The corresponding eigenvectors are

$$\mathbf{p}_1 = \mathbf{e}_1, \quad \mathbf{p}_2 = \left[\begin{array}{c} 0 \\ \frac{c}{\sqrt{2}} \\ \hline \frac{c}{\sqrt{2}} \\ 0 \\ \hline 0 \\ -s \\ \hline 0 \\ 0 \end{array}\right], \quad \mathbf{p}_3 = \left[\begin{array}{c} 0 \\ -\frac{c}{\sqrt{2}} \\ \hline \frac{c}{\sqrt{2}} \\ 0 \\ \hline 0 \\ 0 \\ \hline s \\ 0 \end{array}\right], \quad \mathbf{p}_4 = \mathbf{e}_4,$$

$$\mathbf{p}_5 = \mathbf{e}_5, \quad \mathbf{p}_6 = \left[\begin{array}{c} 0 \\ \frac{s}{\sqrt{2}} \\ \hline \frac{s}{\sqrt{2}} \\ 0 \\ \hline 0 \\ c \\ \hline 0 \\ 0 \end{array}\right], \quad \mathbf{p}_7 = \left[\begin{array}{c} 0 \\ \frac{s}{\sqrt{2}} \\ \hline -\frac{s}{\sqrt{2}} \\ 0 \\ \hline 0 \\ 0 \\ \hline c \\ 0 \end{array}\right], \quad \mathbf{p}_8 = \mathbf{e}_8. \tag{34}$$

Now consider applying the block Jacobi method with block size $L = 2$ to this matrix and adopt the following top-left $2L \times 2L$ submatrix $\tilde{A}$ as the pivot submatrix.

$$\tilde{A} = \left[\begin{array}{cc|cc} 6 & & & \\ & 4 & -s^2 & \\ \hline & -s^2 & 4 & \\ & & & 2 \end{array}\right]. \tag{35}$$

$\tilde{A}$ has small off-diagonal blocks, that is, $\|A_{XY}\|_F = s^2$. The eigenvalues of $\tilde{A}$ are

$$\tilde{\mu}_1 = 6, \quad \tilde{\mu}_2 = 4 + s^2, \quad \tilde{\mu}_3 = 4 - s^2, \quad \tilde{\mu}_4 = 2, \tag{36}$$

so it has simple eigenvalues. But in this case, the diagonal elements $\tilde{a}_{22}$ and $\tilde{a}_{33}$, which approximate the multiple eigenvalues of A, $\lambda_2 = \lambda_3 = 4$, are distributed across two blocks, violating assumption **A2'**. The eigenvector matrix of $\tilde{A}$ is

$$\tilde{P} = \left[\begin{array}{cc|cc} 1 & & & \\ & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \\ \hline & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \\ & & & 1 \end{array}\right], \tag{37}$$

which is unique up to permutation of the columns and scaling by ± 1 of the columns. In this case, $\|P_{YX}\|_2 = \frac{1}{\sqrt{2}}$ and is not small.

This example shows that when A has multiple eigenvalues, without assumption **A2'**, $\|P_{YX}\|_2$ can be large even if $\|A_{XY}\|_F$ is small and the logic of quadratic convergence given in Section 3 breaks down. This suggests that some modification of the algorithm is necessary in order to guarantee quadratic convergence when the assumption **A2'** is not valid.

5 Unified treatment of simple and multiple eigenvalues

Let A be a square Hermitian matrix of order n with q different eigenvalues:

$$\lambda_1 = \cdots = \lambda_{s_1} > \lambda_{s_1+1} = \cdots \lambda_{s_2} > \cdots > \lambda_{s_{q-1}+1} = \cdots = \lambda_{s_q},$$

where $n_i = s_i - s_{i-1}$, $1 \leq i \leq q$ is the algebraic multiplicity of λ_{s_i} (defining $s_0 = 0$ and $s_q = n$). Let the *spectral gap* d be defined as

$$d \equiv \min_{i \neq j} |\lambda_{s_i} - \lambda_{s_j}|. \tag{38}$$

Notice that $d > 0$ iff $q > 1$, i.e. A has at least two different eigenvalues. If $q = 1$ then $d = 0$.

We partition A into square blocks of size $L \times L$ where $L = n/(2p)$, for some p, is the width of one block column (row), $w = 2p$ is the number of blocks per one block column (row), and $W = w(w-1)/2$ is the number of the off-diagonal blocks in the upper triangular part of A. Notice that $L \geq 1$, $w \geq 2$ and $W \geq 1$.

In the kth step of the serial 'classical' block Jacobi method, two blocks $A_{XY}^{(k)}$ and $A_{YX}^{(k)}$ with the largest Frobenius norm are eliminated by an orthogonal transformation:

$$A^{(k+1)} = (P^{(k)})^T A^{(k)} P^{(k)},$$

where the $n \times n$ orthogonal matrix $P^{(k)}$ is the 2×2 block rotation of size $2L \times 2L$ embedded into the identity of size n (see [9]). Four blocks of $P^{(k)}$—namely $P_{XX}^{(k)}$, $P_{XY}^{(k)}$, $P_{YX}^{(k)}$ and $P_{YY}^{(k)}$ that are different from blocks of I_n—are constructed so that

$$\begin{pmatrix} P_{XX}^{(k)} & P_{XY}^{(k)} \\ P_{YX}^{(k)} & P_{YY}^{(k)} \end{pmatrix}^T \begin{pmatrix} A_{XX}^{(k)} & A_{XY}^{(k)} \\ A_{YX}^{(k)} & A_{YY}^{(k)} \end{pmatrix} \begin{pmatrix} P_{XX}^{(k)} & P_{XY}^{(k)} \\ P_{YX}^{(k)} & P_{YY}^{(k)} \end{pmatrix} = \begin{pmatrix} A_{XX}^{(k+1)} & 0 \\ 0 & A_{YY}^{(k+1)} \end{pmatrix}. \tag{39}$$

Note that the diagonal blocks $A_{XX}^{(k)}$, $A_{YY}^{(k)}$, $A_{XX}^{(k+1)}$ and $A_{YY}^{(k+1)}$ are square, diagonal matrices of order L.

Let us define:

$$\tilde{A}^{(k)} \equiv \begin{pmatrix} A_{XX}^{(k)} & A_{XY}^{(k)} \\ A_{YX}^{(k)} & A_{YY}^{(k)} \end{pmatrix}, \quad \tilde{P}^{(k)} \equiv \begin{pmatrix} P_{XX}^{(k)} & P_{XY}^{(k)} \\ P_{YX}^{(k)} & P_{YY}^{(k)} \end{pmatrix}. \tag{40}$$

Since $\tilde{A}^{(k)}$ is symmetric for all k, the matrix $\tilde{P}^{(k)}$ is orthogonal and its columns are eigenvectors of $\tilde{A}^{(k)}$.

The global convergence of the method was proved in [9]. It is based on the fact that the off-diagonal Frobenius norm of $A^{(k)}$ converges to zero so that the eigenvalues of A appear on the diagonal. This proof does *not* depend on the distribution of eigenvalues.

Hence, writing

$$A^{(k)} = \text{diag}(A^{(k)}) + \text{off}(A^{(k)}), \tag{41}$$

we can make following assumptions at the iteration step k:

A3 The off-diagonal Frobenius norm is small:

$$\|\mathrm{off}(A^{(k)})\|_F \equiv \sqrt{\sum_{I \neq J} \|A_{IJ}^{(k)}\|_F^2} < \frac{d}{4}. \tag{42}$$

A4 The diagonal elements of $A^{(k)}$ affiliated with the same multiple eigenvalue occupy successive positions on the diagonal.

When $d = 0$, Eq. (42) gives $\|\mathrm{off}(A^{(k)})\|_F = 0$ and, consequently, $A^{(k)} = \lambda_1 I$, so that there is nothing to prove. Therefore we assume $q > 1$.

Since all transformations are orthogonal similarities, the eigenvalues of A are the same as those of $A^{(k)}$. But then, according to Eq. (41), $\mathrm{off}(A^{(k)})$ is a perturbation of $\mathrm{diag}(A^{(k)})$, and it is bounded in F-norm by $d/4$. According to the Wielandt-Hoffman theorem, for each i, $1 \leq i \leq q$, there are exactly n_i diagonal elements of $A^{(k)}$ that lie around λ_{s_i} in the circle of radius less than $d/4$. Recall that according to the assumption **A4** these diagonal elements occupy successive positions on the diagonal, i.e. they form *clusters* $Cl_i^{(k)}$, $1 \leq i \leq q$. Note that two different clusters are separated *at least* by $d/2$.

Now we show that these clusters are *stabilized*, i.e., at iteration step k, a diagonal element that lies in the circle around λ_{s_i} can not 'jump' into a circle around λ_{s_j} for $j \neq i$. Write a matrix with blocks of A on the left-hand side of Eq. (39) in the following way:

$$\tilde{A}^{(k)} = \begin{pmatrix} A_{XX}^{(k)} & A_{XY}^{(k)} \\ A_{YX}^{(k)} & A_{YY}^{(k)} \end{pmatrix} = \begin{pmatrix} A_{XX}^{(k)} & 0 \\ 0 & A_{YY}^{(k)} \end{pmatrix} + \begin{pmatrix} 0 & A_{XY}^{(k)} \\ A_{YX}^{(k)} & 0 \end{pmatrix} = B^{(k)} + E^{(k)}, \tag{43}$$

which is the sum of a diagonal part and an off-diagonal part (perturbation). The F-norm of perturbation is

$$\|E^{(k)}\|_F = \sqrt{\|A_{XY}^{(k)}\|_F^2 + \|A_{YX}^{(k)}\|_F^2} < \frac{d}{4}.$$

Recall that the matrix on the right-hand side of Eq. (39) (i.e., the result of step k) is diagonal and its diagonal elements are eigenvalues of the matrix on the left-hand side. At the same time, they approximate the eigenvalues of the whole A. Assume, just for a moment, that the diagonal elements of both A-matrices are ordered similarly—e.g., decreasingly. Take an arbitrary index j, pick up the diagonal elements $a_{jj}^{(k+1)}$ and choose its corresponding diagonal element $a_{jj}^{(k)}$. Let $a_{jj}^{(k)}$ belong to the cluster $Cl_i^{(k)}$ around some λ_{s_i}, so that $|a_{jj}^{(k)} - \lambda_{s_i}| < d/4$. Then, according to the Wielandt-Hoffman theorem for a 2×2 block matrix in (43),

$$|a_{jj}^{(k+1)} - a_{jj}^{(k)}| < \frac{d}{4}.$$

How far is $a_{jj}^{(k+1)}$ from λ_{s_i}?

$$|a_{jj}^{(k+1)} - \lambda_{s_i}| \leq |a_{jj}^{(k+1)} - a_{jj}^{(k)}| + |a_{jj}^{(k)} - \lambda_{s_i}| < \frac{d}{4} + \frac{d}{4} = \frac{d}{2}.$$

This means that during one iteration step a diagonal element which belongs to the cluster around λ_{s_i} can *not* move into another cluster because the distance between any two clusters is at least $d/2$.

In fact, the upper bound for a cluster's diameter *shrinks* as the iterations proceed. We know that the off-diagonal norm of $A^{(k)}$ goes to zero as $k \to \infty$. Take two iteration steps, k and $k+m$. Then there exists α_m, $0 \le \alpha_m < 1$, such that

$$\|\mathrm{off}(A^{(k+m)})\|_F = \alpha_m \, \|\mathrm{off}(A^{(k)})\|_F < \alpha_m \, \frac{d}{4}.$$

But according to the Wielandt-Hoffman theorem for $A^{(k+m)}$, for each i, $1 \le i \le q$, there are exactly n_i diagonal elements of $A^{(k+m)}$ that lie around λ_{s_i} in the circle of radius less than $\alpha_m \, d/4$.

In summary, we can safely assume that, from the iteration step k onwards, all q clusters Cl_i around λ_{s_i}, $1 \le i \le q$, are stabilized, their elements occupy consecutive positions on the diagonal, each cluster contains exactly n_i approximations to λ_{s_i} and the distance between two different clusters is at least $d/2$. Hence, we can add the third assumption to **A3** and **A4**:

A5 For each i, $1 \le i \le q$, all n_i diagonal elements that approximate the same eigenvalue λ_{s_i} remain in the same diagonal block of $A^{(s)}$ for $s \ge k$. Moreover, they are ordered non-increasingly w.r.t. the non-increasing ordering of λ_{s_i}'s.

In other words, the matrix partition does *not* divide any cluster of diagonal elements, which approximate the same eigenvalue, into two or more diagonal blocks. At the same time, the diagonal starts with approximations to λ_{s_1} (in any order) and ends with those to λ_{s_q}.

5.1 Upper bound for $\|P_{YX}^{(k)}\|_2$

We analyze one iteration step $k \to k+1$, but for simplicity the iteration index is omitted. Our task is to derive the upper bound for the spectral norm $\|P_{YX}\|_2$.

Recall that our 2×2 subproblem with the auxiliary partitioning has the form:

$$\begin{pmatrix} A_{XX}^{(k)} & A_{XY}^{(k)} \\ A_{YX}^{(k)} & A_{YY}^{(k)} \end{pmatrix} \begin{pmatrix} P_{XX} & P_{XY} \\ P_{YX} & P_{YY} \end{pmatrix} = \begin{pmatrix} P_{XX} & P_{XY} \\ P_{YX} & P_{YY} \end{pmatrix} \begin{pmatrix} A_{XX}^{(k+1)} & 0 \\ 0 & A_{YY}^{(k+1)} \end{pmatrix}.$$

We added the iteration indices to A-blocks to stress their meaning. The equation for the block $(2,1)$ reads:

$$A_{YY}^{(k)} P_{YX} - P_{YX} A_{XX}^{(k+1)} = -A_{YX}^{(k)} P_{XX}, \tag{44}$$

which is the Sylvester equation for P_{YX} [3]. Notice that the blocks $A_{YY}^{(k)}$ and $A_{XX}^{(k+1)}$ are diagonal and their eigenvalues are diagonal elements. Recall that according to the construction of the matrix partition the eigenvalues of $A_{YY}^{(k)}$ and $A_{XX}^{(k+1)}$ approximate *different* eigenvalues of A. Take an arbitrary diagonal element $a_{uu}^{(k)}$ from $A_{YY}^{(k)}$ and $a_{vv}^{(k+1)}$ from $A_{XX}^{(k+1)}$. Let $a_{uu}^{(k)}$ belong to the cluster around some λ_{s_i} and $a_{vv}^{(k+1)}$ belong to the cluster around some λ_{s_j}, $i \ne j$. Then:

$$\begin{aligned} |a_{uu}^{(k)} - a_{vv}^{(k+1)}| &= |(\lambda_{s_i} - \lambda_{s_j}) - (\lambda_{s_i} - a_{uu}^{(k)}) - (a_{vv}^{(k+1)} - \lambda_{s_j})| \\ &\ge |\lambda_{s_i} - \lambda_{s_j}| - |\lambda_{s_i} - a_{uu}^{(k)}| - |a_{vv}^{(k+1)} - \lambda_{s_j}| \\ &> d - \frac{d}{4} - \frac{d}{4} = \frac{d}{2}. \end{aligned}$$

Consequently,

$$\min_{u,v} |a_{uu}^{(k)} - a_{vv}^{(k+1)}| > \frac{d}{2},$$

so that the spectra of $A_{YY}^{(k)}$ and $A_{XX}^{(k+1)}$ are disjoint and their distance is bigger than $d/2$ and the entire spectrum of $A_{YY}^{(k)}$ lies, on the real axis, to the left of the entire spectrum of $A_{XX}^{(k+1)}$. Therefore, we can apply the Davis-Kahan lemma [1] stating that the Sylvester equation (44) has the unique solution P_{YX} and its spectral norm is bounded by

$$\|P_{YX}\|_2 = \|P_{XY}\|_2 \leq \frac{2}{d} \| - A_{YX}^{(k)} P_{XX}\|_2 \leq \frac{2}{d} \|A_{YX}^{(k)}\|_F \, \|P_{XX}\|_2 \leq \frac{2}{d} \|A_{YX}^{(k)}\|_F,$$

where we used the Cosine-Sine decomposition of $\tilde{P}$ [3] with $\|P_{XX}\|_2 \leq 1$. This bound ensures the asymptotic quadratic convergence proved in Lemma 2 above. The constant is $\delta = d/2$ and it does not depend on the off-diagonal norm of $A^{(k)}$—especially, it does not depend on the iteration index k. Also notice that we need not discuss separately the case of simple and multiple eigenvalues.

6 Conclusions

To our best knowledge, this is the first proof of the asymptotic quadratic convergence of the *block* Jacobi method for the EVD of a Hermitian matrix A. It covers the case of simple as well as multiple eigenvalues. There remains an interesting question of how to estimate δ for the case of clustered eigenvalues, when the spectral gap d can be tiny and useless for practical computations, because the off-diagonal Frobenius norm of $A^{(k)}$ should be even less than $d/4$.

References

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

[2] Z. Drmač, *Global convergence proof for cyclic Jacobi methods with block rotations*, SIAM J. Matrix Anal. Appl., 31 (2010) 1329-1350.

[3] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins UP, Baltimore, 2013.

[4] A. J. Hoffman and H. W. Wielandt, *The variation of the spectrum of a normal matrix*, Duke Math. J., 20 (1953) 37-39.

[5] B. N. Parlett, *The Symmetric Eigenvalue Problem*, SIAM, Philadelphia, 1987.

[6] A. Schönhage, *Zur Konvergenz des Jacobi-Verfahrens*, Num. Math. 3 (1961) 374-380.

[7] A. Schönhage, *Zur quadratischen Konvergenz des Jacobi-Verfahrens*, Num. Math. 6 (1964) 410-412.

[8] M. Sugihara and K. Murota, *Mathematics of Numerical Linear Algebra*, Iwanami-Shoten, 2009 (in Japanese).

[9] Y. Yamamoto, Z. Lan, S. Kudo, *Convergence analysis of the parallel classical block Jacobi method for the symmetric eigenvalue problem*, JSIAM Letters 6 (2014) 57-60.