SIAM J. MATRIX ANAL. APPL.
Vol. 40, No. 2, pp. 639–671

ASYMPTOTIC QUADRATIC CONVERGENCE OF THE TWO-SIDED SERIAL AND PARALLEL BLOCK-JACOBI SVD ALGORITHM*

GABRIEL OKŠA†, YUSAKU YAMAMOTO‡, MARTIN BEČKA†, AND MARIÁN VAJTERŠIC† §

Abstract. We present a proof of the global and asymptotic quadratic convergence of the serial and parallel two-sided block-Jacobi SVD algorithm with dynamic ordering. In the serial case, one pair of the off-diagonal blocks with the largest weight given as the sum of squares of Frobenius norms is annihilated. In the parallel case, using the greedy implementation of dynamic ordering and having p processors, p pairs of the off-diagonal blocks with largest weights, and disjoint block row and column indices are annihilated in each parallel iteration step. Additionally, the asymptotic quadratic convergence is also proved for the scaled iterated matrix, both in serial and parallel cases. Numerical examples confirm the developed theory.

Key words. singular value decomposition, serial and parallel two-sided SVD block-Jacobi algorithm, dynamic ordering, global convergence, asymptotic quadratic convergence

AMS subject classification. 65F30

DOI. 10.1137/18M1222727

1. Introduction.

The singular value decomposition (SVD) of a general matrix $A \in \mathbb{C}^{m\times n}$, $m \geq n$, belongs to the basic matrix factorizations and has many applications, e.g., in signal and image processing, data retrieval and reduction, system identification. It is defined as the decomposition $A = U(\Sigma, 0^T)^T V^H$, where U (of size $m \times m$) and V (of size $n \times n$) is the unitary matrix of left and right singular vectors, respectively, and Σ is the $n \times n$ diagonal matrix with real nonnegative diagonal elements $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

As is well known, the one-sided or two-sided Jacobi algorithm can be used for the eigenvalue decomposition (EVD) of a Hermitian matrix or the SVD of a general matrix. It often delivers high relative accuracy of computed eigenvalues and eigenvectors or singular triplets [6, 7, 8, 9, 19, 20].

The serial scalar two-sided Jacobi SVD algorithm was proposed by Kogbetliantz [15]. It is based on an iterative transformation of the original matrix A to the diagonal form by a series of two unitary rotations (one applied from the left side, the other from the right). In each iteration step, two off-diagonal elements are annihilated according to some ordering. For the row-cyclic ordering, the global convergence (i.e., the convergence of the off-diagonal Frobenius norm to zero) has been proven by Kogbetliantz [15]. The asymptotic quadratic convergence (AQC) has been shown by Paige [23] for well-separated (simple and multiple) singular values only, and by Charlier and Van

*Received by the editors October 24, 2018; accepted for publication (in revised form) by M. W. Berry March 25, 2019; published electronically May 23, 2019.
http://www.siam.org/journals/simax/40-2/M122272.html
Funding: The work of the first, third, and fourth authors was supported by the VEGA grant 2/0004/17. The work of the second author was partially supported by the Japanese Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific research grants 26286087, 16KT0016, and 17H02828.
†Mathematical Institute, Slovak Academy of Sciences, Bratislava, Slovak Republic (gabriel.oksa@savba.sk, martin.becka@savba.sk).
‡Department of Communication Engineering and Informatics, University of Electro-Communications, Tokyo, Japan (yusaku.yamamoto@uec.ac.jp).
§Department of Computer Sciences, University of Salzburg, Austria, (marian@cosy.sbg.ac.at).

Dooren [4] for clusters. Hari [11] proved the AQC for the case of a triangular matrix A and simple singular values with an additional assumption about the multiplicity of the minimal singular value. Considering the parallel scalar algorithm, Luk and Park [18] provided the proof of global convergence for two parallel orderings, which are equivalent to the row-cyclic ordering. Note that all of the above proofs of global convergence do not depend on the distribution of singular values.

When moving to two-sided block-Jacobi SVD algorithms, Bujanović and Drmač [3] proved the global convergence and the AQC for serial block orderings equivalent to the row-cyclic block ordering, but again for simple singular values only. Note that when using the above mentioned block orderings and starting with the upper triangular matrix, the iterated matrix remains lower triangular and upper triangular in odd and even sweeps, respectively. (For any cyclic ordering, a *sweep* is defined as some sequence of consecutive transformations, during which each off-diagonal element is annihilated exactly once.) Recently, Hari [12] proved the global convergence of general serial block-Jacobi methods for cyclic pivot strategies, which are weakly equivalent to the (block) column-cyclic strategy. His approach is very general and also includes nonunitary transformations as well as the Jacobi-type algorithms for solving the standard and generalized eigenvalue problems. However, there is no published analysis of the asymptotic convergence for the general frame of Jacobi-type algorithms.

Any cyclic ordering (serial or parallel) can be very inefficient in reducing a matrix A to its diagonal form. If the sequence of off-diagonal blocks brought to annihilation is prescribed, then the algorithm can spend many iteration steps in annihilating blocks with a relatively small Frobenius norm. Hence, many iteration steps (serial or parallel) will be needed for convergence.

The parallel *dynamic* ordering was designed by Bečka, Okša, and Vajteršic [1] for the parallel two-sided block-Jacobi SVD algorithm to achieve an optimal reduction of the off-diagonal Frobenius norm in each parallel iteration step. Essentially, it is an extension of the old Jacobi idea to the block and parallel case. Recall that in the EVD computation of a Hermitian matrix by a scalar algorithm, Jacobi [14] proposed to annihilate two off-diagonal elements with (equal) maximal modulus in each iteration step. This approach was abandoned in the second half of the 20th century, because it seemed to be very slow in comparison to EVD algorithms based on tri-diagonalization and subsequent QR iteration. For example, Golub and Van Loan argue in [10, p. 480] that finding a matrix element with maximal modulus costs $O(n^2)$ comparisons, but the subsequent matrix update requires only $O(n)$ flops because just two matrix rows and columns need to be modified. However, their argument is valid *for the first iteration step only*. In subsequent iterations, the search for an optimal off-diagonal element can also be performed in $O(n)$ comparisons—see the discussion about the computational complexity of dynamic ordering in [22, sect. 2] for the block case. The same conclusion is also valid for the serial two-sided block-Jacobi SVD algorithm.

The performance of dynamic ordering for the parallel two-sided block-Jacobi SVD algorithm was tested against the parallel cyclic CO(0) ordering in [1]. The number of parallel iteration steps needed for convergence using the parallel dynamic ordering was 2–8 times less than that for the parallel cyclic ordering (see figures in [1, p. 254]). Although dynamic ordering is required at the beginning of each parallel iteration step, its computation, using the greedy implementation of complexity $O(p^2 \log p)$ where p is the number of processors, took less than 3–5% of the total parallel execution time (see figures in [1, pp. 255–256]). Recently, the dynamic ordering was also extended to the parallel one-sided block-Jacobi SVD algorithm [2]. When the weights were computed using Level 3 basic linear algebra subroutines (BLAS), the portion of dynamic ordering was less than 1% of the total parallel execution time on modern parallel architectures—see [16] and [17] for the algorithm's performance in the EVD of Hermitian matrices.

Here we analyze the block case and prove the global convergence and the AQC of the serial and parallel two-sided block-Jacobi SVD algorithm with dynamic ordering for matrices with a general distribution of singular values (simple, multiple, clusters). The proofs are based on carefully estimated upper bounds for the change of the Frobenius norm of a nonannihilated off-diagonal block in one iteration step, and for spectral and Frobenius norms of certain submatrices of local left and right singular vectors.

This paper is the third one in a series about the AQC of the two-sided block-Jacobi algorithm. The AQC of the serial and parallel algorithm with dynamic ordering for the EVD of Hermitian matrices has been proven in [22] and [21], respectively. Although the structure of exposition is similar to the previous two papers, the SVD of a general matrix is more complex than the EVD of a Hermitian matrix. Consequently, the proofs of lemmas and theorems contain many details that are specific to the SVD. Moreover, some substantially new results are presented here:

- Introduction of the scaled iterated matrix (see (7)), and the proof of the AQC of its off-diagonal Frobenius norm for the serial and parallel algorithm (see Corollary 2.3 and (47)).
- Identification of the constant δ from Theorem 2.2 by considering the system of Sylvester equations (see Lemma 2.5).
- New upper bound (adapted from [17]) in the proof of the global convergence of the parallel algorithm (see (26)).
- New upper bound for the length of a parallel sweep W, over which the AQC of the off-diagonal Frobenius norm can be observed (see Lemma 3.2). Its value shows that the parallel algorithm is *always* more efficient than the serial one in the asymptotic regime.

The rest of the paper is organized as follows. Section 2 is devoted to the serial two-sided block-Jacobi SVD algorithm. The serial dynamic ordering is described, and the global convergence is proved. Next, the upper bound for the change of the Frobenius norm of an off-diagonal block, which was not annihilated in a given iteration step, is derived. Based on this result, the AQC is proved. The proof does not depend on the distribution of singular values of A, however, it uses an assumption about the spectral norm of a certain submatrix containing two blocks from the partition of local left and right singular vectors computed in a 2×2 block subproblem. Next, we show that this assumption holds for the case of well-separated singular values including multiple ones (subsection 2.2) as well as for clusters (subsection 2.3) under reasonable suppositions. The same structure of exposition is preserved in section 3, which deals with the parallel two-sided block-Jacobi SVD algorithm with dynamic ordering. In section 4, four numerical examples are provided (two for the serial algorithm and two for the parallel one) that illustrate the developed theory. Section 5 concludes the paper.

Regarding the notation, a set is represented by $\mathcal{P}$, and $|\mathcal{P}|$ stands for the number of its elements. Furthermore, $\|A\|_2$ and $\|A\|_F$ are the spectral and Frobenius norm of matrix A, respectively, while A^H denotes the Hermitian operator over A, i.e., its transposition and conjugation. For two integers i and j, $i \leq j$, the symbol $i : j$ denotes all integers k such that $i \leq k \leq j$.

2. Serial two-sided SVD algorithm. When dealing with the convergence analysis of any SVD algorithm, it is sufficient to consider only square matrices. If an original matrix is of size $m \times n$, $m \geq n$, one can initially compute its QR decomposition and then apply the iterative SVD algorithm to the square factor R of size n. Then, the SVD of the original matrix can be reconstructed in an obvious way.

Let us divide a square matrix A of order n into a $w \times w$ block structure with w blocks in each block row (column). Write the decomposition of n as $n = \lfloor n/w \rfloor w + r$ with $0 \leq r \leq w-1$. If $r = 0$, w divides n and nothing needs to be done. If $r > 0$, border a matrix A by adding $w-r$ zero rows to the bottom of A, $w-r$ zero columns to the right of A and $w-r$ ones to the lower part of the main diagonal of A. Hence, a bordered matrix is the direct sum $A \oplus I_{w-r}$ of size $(\lfloor n/w \rfloor + 1)w$, where I_{w-r} is the identity of order $w-r$. Note that the SVD of A can be recovered from the SVD of $A \oplus I_{w-r}$ easily. To keep the following exposition simple, we assume that w divides n.

Denote by A_{IJ} the (I,J)th block of size $\ell \times \ell$, $\ell = n/w$. Hence, there are $w(w-1)$ off-diagonal blocks in A.

Let us assume that all diagonal blocks of A were diagonalized by a series of two-sided unitary transformations at initialization step. Then, diagonal blocks remain diagonal during the whole computation.

At iteration step k of the two-sided serial block-Jacobi SVD algorithm, let us define the *weights* for off-diagonal blocks with symmetric block indices (I,J) and (J,I), $I \neq J$, by

$$w_{IJ}^{(k)} \equiv \|A_{IJ}^{(k)}\|_F^2 + \|A_{JI}^{(k)}\|_F^2. \tag{1}$$

To optimally reduce the off-diagonal Frobenius norm, a pair of off-diagonal blocks with the *maximal* weight will be annihilated (if there are more pairs of equal maximal weight, choose any one of them). Let the chosen off-diagonal blocks have block indices (X_k, Y_k) and (Y_k, X_k), i.e.,

$$w_{X_k Y_k}^{(k)} = \max_{I \neq J} w_{IJ}^{(k)}.$$

Note that, contrary to the EVD of Hermitian matrices, choosing two off-diagonal blocks with maximal weight for annihilation is *not* equivalent to choosing the off-diagonal block $A_{S_k T_k}^{(k)}$ with the *largest* Frobenius norm together with the block $A_{T_k S_k}^{(k)}$. In fact, we can easily have

$$w_{S_k T_k}^{(k)} \leq w_{X_k Y_k}^{(k)},$$

so that the off-diagonal block with the largest Frobenius norm is not annihilated.

The annihilation is performed by a two-sided unitary transformation

$$(U^{(k)})^H A^{(k)} V^{(k)} = A^{(k+1)},$$

where the $n \times n$ unitary matrices $U^{(k)}$ and $V^{(k)}$ are the matrices of local left and right singular vectors, respectively, embedded into the identity matrix I_n of order n. Four blocks of $U^{(k)}$ and $V^{(k)}$, each of order ℓ, that are different from blocks of I_n can be chosen so that

(2)

$$\begin{pmatrix} U_{X_k X_k}^{(k)} & U_{X_k Y_k}^{(k)} \\ U_{Y_k X_k}^{(k)} & U_{Y_k Y_k}^{(k)} \end{pmatrix}^H \begin{pmatrix} A_{X_k X_k}^{(k)} & A_{X_k Y_k}^{(k)} \\ A_{Y_k X_k}^{(k)} & A_{Y_k Y_k}^{(k)} \end{pmatrix} \begin{pmatrix} V_{X_k X_k}^{(k)} & V_{X_k Y_k}^{(k)} \\ V_{Y_k X_k}^{(k)} & V_{Y_k Y_k}^{(k)} \end{pmatrix} = \begin{pmatrix} A_{X_k X_k}^{(k+1)} & 0 \\ 0 & A_{Y_k Y_k}^{(k+1)} \end{pmatrix},$$

where the diagonal blocks $A^{(k+1)}_{X_kX_k}$ and $A^{(k+1)}_{Y_kY_k}$ are square, diagonal matrices of order ℓ with nonnegative diagonal elements (local singular values).

Let us define

$$\tilde{U}^{(k)} \equiv \begin{pmatrix} U^{(k)}_{X_kX_k} & U^{(k)}_{X_kY_k} \\ U^{(k)}_{Y_kX_k} & U^{(k)}_{Y_kY_k} \end{pmatrix}, \; \tilde{V}^{(k)} \equiv \begin{pmatrix} V^{(k)}_{X_kX_k} & V^{(k)}_{X_kY_k} \\ V^{(k)}_{Y_kX_k} & V^{(k)}_{Y_kY_k} \end{pmatrix}, \tag{3}$$

and

$$\tilde{A}^{(k)} \equiv \begin{pmatrix} A^{(k)}_{X_kX_k} & A^{(k)}_{X_kY_k} \\ A^{(k)}_{Y_kX_k} & A^{(k)}_{Y_kY_k} \end{pmatrix}, \; \tilde{\Sigma}^{(k)} \equiv \begin{pmatrix} A^{(k+1)}_{X_kX_k} & 0 \\ 0 & A^{(k+1)}_{Y_kY_k} \end{pmatrix}. \tag{4}$$

Because (2) is the SVD of the matrix $\tilde{A}^{(k)}$, the matricies $\tilde{U}^{(k)}$ and $\tilde{V}^{(k)}$ are the unitary matricies of left and right singular vectors of $\tilde{A}^{(k)}$, respectively.

To prove the global convergence of the serial two-sided block-Jacobi SVD algorithm, define the square of the off-diagonal Frobenius norm of $A^{(k)}$ by

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

Because the off-diagonal blocks $A^{(k)}_{X_kY_k}$ and $A^{(k)}_{Y_kX_k}$ are zero at iteration step $k+1$, we have

$$\begin{aligned} \|\mathrm{off}(A^{(k+1)})\|_F^2 &= \|\mathrm{off}(A^{(k)})\|_F^2 - \left(\|A^{(k)}_{X_kY_k}\|_F^2 + \|A^{(k)}_{Y_kX_k}\|_F^2\right) \\ &\leq \left(1 - \frac{2}{w(w-1)}\right) \|\mathrm{off}(A^{(k)})\|_F^2, \end{aligned} \tag{5}$$

where we use the bound

$$\|\mathrm{off}(A^{(k)})\|_F^2 = \sum_{I<J} \left(\|A^{(k)}_{IJ}\|_F^2 + \|A^{(k)}_{JI}\|_F^2\right) \leq \frac{w(w-1)}{2} \left(\|A^{(k)}_{X_kY_k}\|_F^2 + \|A^{(k)}_{Y_kX_k}\|_F^2\right).$$

Hence, $\|\mathrm{off}(A^{(k)})\|_F^2$ decreases at least as fast as the geometric sequence with the quotient $(W-1)/W$, $W = w(w-1)/2$, and therefore converges to zero. Note that this proof does not depend on the distribution of singular values of A.

The singular values of $\tilde{A}^{(k)}$, i.e., the elements of diagonal matrix $\tilde{\Sigma}^{(k)}$, can be computed and located in any order on the diagonal. An important variant of the local SVD has *ordered* singular values (e.g., decreasingly) on the diagonal of $\tilde{\Sigma}^{(k)}$. This can be achieved in $O(\ell^2)$ steps using a suitable permutation matrix $\Pi^{(k)}$:

$$\begin{aligned} \tilde{A}^{(k)} &= \tilde{U}^{(k)}\tilde{\Sigma}^{(k)}(\tilde{V}^{(k)})^H \\ &= \left(\tilde{U}^{(k)}(\Pi^{(k)})^H\right)\left(\Pi^{(k)}\tilde{\Sigma}^{(k)}(\Pi^{(k)})^H\right)\left(\tilde{V}^{(k)}(\Pi^{(k)})^H\right)^H. \end{aligned}$$

This variant of the SVD of a 2×2 block subproblem will be called the *local ordering of diagonal elements (LODE)*.

Now we introduce the *scaling* of the iterated matrix $A^{(k)}$, which is used in the stopping criterion of the Jacobi algorithm. Let $d^{(k)}$ be the vector of diagonal elements of $A^{(k)}$, and denote by $\mathrm{diag}(d^{(k)})$ the diagonal matrix of order n with diagonal elements from $d^{(k)}$. Let $A^{(k)}$ be given by its rows, $A^{(k)} = (r_1^{(k)}, r_2^{(k)}, \ldots, r_n^{(k)})^T$, or by its columns, $A^{(k)} = (c_1^{(k)}, c_2^{(k)}, \ldots, c_n^{(k)})$. At iteration step k, define the diagonal left and right scaling matrix $D_L^{(k)}$ and $D_R^{(k)}$, respectively, as

$$
\begin{aligned}
D_L^{(k)} &\equiv \operatorname{diag}\left(\|r_1^{(k)}\|^{1/2}, \|r_2^{(k)}\|^{1/2}, \ldots, \|r_n^{(k)}\|^{1/2}\right),\\
D_R^{(k)} &\equiv \operatorname{diag}\left(\|c_1^{(k)}\|^{1/2}, \|c_2^{(k)}\|^{1/2}, \ldots, \|c_n^{(k)}\|^{1/2}\right).
\end{aligned}
\tag{6}
$$

Assuming that no zero rows and/or columns appear in the iterated matrix through the whole Jacobi process, the *scaled* iterated matrix $A_{\text{sc}}^{(k)}$ is defined by

$$
A_{\text{sc}}^{(k)} \equiv (D_L^{(k)})^{-1} A^{(k)} (D_R^{(k)})^{-1}. \tag{7}
$$

Note that the off-diagonal Frobenius norm of the scaled iterated matrix $A_{\text{sc}}^{(k)}$ is given by

$$
\|\text{off}(A_{\text{sc}}^{(k)})\|_F = \|(D_L^{(k)})^{-1} A^{(k)} (D_R^{(k)})^{-1} - (D_L^{(k)})^{-1} \operatorname{diag}(d^{(k)}) (D_R^{(k)})^{-1}\|_F.
$$

Besides using $\|\text{off}(A_{\text{sc}}^{(k)})\|_F$ in the stopping criterion of the serial and parallel block-Jacobi SVD algorithm, its AQC will be proved (see subsections 2.1 and 3.2).

2.1. AQC. Using these preliminaries, we investigate the asymptotic convergence property of the serial two-sided block-Jacobi SVD algorithm in the general setting when no a priori assumptions are made about the distribution of singular values of A.

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

The first lemma is an obvious modification of Lemma 1 in [22]. It is devoted to the change of the Frobenius norm of a nonannihilated off-diagonal block in a given iteration step. Note that one iteration step changes only two block rows and two block columns X, Y. The lemma considers the block row and block column X. The situation for the block row and column Y is similar.

LEMMA 2.1. *Let A_{ST} be the off-diagonal block with the largest Frobenius norm. Consider the change of an off-diagonal block A_{XJ} $(J \neq X, Y)$ after application of U^H:*

$$
\hat{A}_{XJ} = U_{XX}^H A_{XJ} + U_{YX}^H A_{YJ}. \tag{8}
$$

Similarly, consider the change of an off-diagonal block A_{JX} $(J \neq X, Y)$ after application of V:

$$
\hat{A}_{JX} = A_{JX} V_{XX} + A_{JY} V_{YX}. \tag{9}
$$

Let $C = (A_{XY}, A_{YX})$. If $\left(\begin{smallmatrix} U_{YX} \\ V_{YX} \end{smallmatrix}\right)$ of size $2\ell \times \ell$ is bounded as

$$
\left\| \begin{pmatrix} U_{YX} \\ V_{YX} \end{pmatrix} \right\|_2 \leq \frac{\|C\|_F}{\delta}
$$

for some constant $\delta > 0$, then the following inequalities hold:

$$
\left| \|\hat{A}_{XJ}\|_F^2 - \|A_{XJ}\|_F^2 \right| \leq \frac{\|A_{ST}\|_F^2}{\delta^2} \|C\|_F^2 + 2\frac{\|A_{ST}\|_F}{\delta} \|C\|_F \|A_{XJ}\|_F, \tag{10}
$$

$$
\left| \|\hat{A}_{JX}\|_F^2 - \|A_{JX}\|_F^2 \right| \leq \frac{\|A_{ST}\|_F^2}{\delta^2} \|C\|_F^2 + 2\frac{\|A_{ST}\|_F}{\delta} \|C\|_F \|A_{JX}\|_F. \tag{11}
$$

Proof. Since

$$\max\{\|U_{YX}\|_2, \|V_{YX}\|_2\} \leq \left\| \begin{pmatrix} U_{YX} \\ V_{YX} \end{pmatrix} \right\|_2 \leq \frac{\|C\|_F}{\delta},$$

and both transformations in (8) and (9) are one-sided by corresponding blocks of 2×2 block unitary matrices of local left and right singular vectors, respectively, the changes of Frobenius norms can be bounded using the same technique as in the proof of Lemma 1 of [22]:

$$\begin{aligned} &\left| \|\hat{A}_{XJ}\|_F^2 - \|A_{XJ}\|_F^2 \right| \\ &\quad \leq \|U_{YX}\|_2^2 \max\left\{\|A_{XJ}\|_F^2, \|A_{YJ}\|_F^2\right\} + 2\|A_{XJ}\|_F \|U_{YX}\|_2 \|A_{YJ}\|_F \\ &\quad \leq \frac{\|A_{ST}\|_F^2}{\delta^2} \|C\|_F^2 + 2\,\frac{\|A_{ST}\|_F}{\delta} \|C\|_F \|A_{XJ}\|_F. \end{aligned}$$

A similar approach is also valid for (11). ∎

The next theorem contains our main result, namely, the proof of the AQC after $W = w(w-1)/2$ steps of the serial two-sided block-Jacobi SVD algorithm. In the following, we use the term *serial sweep* for W consecutive iteration steps. For a serial cyclic ordering, each off-diagonal matrix block is annihilated exactly once during a serial sweep. In our serial dynamic ordering, some off-diagonal blocks can be annihilated more than once, and some of them may not be annihilated at all. However, the value of W is important as can be seen in the next Theorem 2.2.

THEOREM 2.2. *Consider one sweep ($W = w(w-1)/2$ serial steps) of the block-Jacobi method. Denote the iteration steps of a sweep by $k = 0, 1, \ldots, W-1$ and the off-diagonal blocks chosen at step k for annihilation by $A^{(k)}_{X_kY_k}$ and $A^{(k)}_{Y_kX_k}$. Let*

$$C^{(k)} = \left(A^{(k)}_{X_kY_k}, A^{(k)}_{Y_kX_k}\right),$$

and let $A^{(k)}_{S_kT_k}$ be the off-diagonal block with the maximal Frobenius norm at iteration step k. If all matrices

$$\begin{pmatrix} U^{(k)}_{Y_kX_k} \\ V^{(k)}_{Y_kX_k} \end{pmatrix}$$

used at iteration steps $k = 0, 1, \ldots, W-1$ satisfy

$$\left\| \begin{pmatrix} U^{(k)}_{Y_kX_k} \\ V^{(k)}_{Y_kX_k} \end{pmatrix} \right\|_2 \leq \frac{\|C^{(k)}\|_F}{\delta}$$

for some constant $\delta > 0$, then

$$\|\mathrm{off}(A^{(W)})\|_F \leq \sqrt{2(w-2)}\,\frac{\|\mathrm{off}(A^{(0)})\|_F^2}{\delta}, \tag{12}$$

i.e., the block-Jacobi SVD algorithm converges quadratically after W serial iteration steps.

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

$$\sum_{(I,J)\in\mathcal{P}_k} \|A^{(k)}_{IJ}\|_F^2 \leq \frac{w-2}{2} \left(\frac{2\|\mathrm{off}(A^{(0)})\|_F^2 - 2\|\mathrm{off}(A^{(k)})\|_F^2}{\delta} \right)^2. \tag{13}$$

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 square of the right-hand side of (12). So it is sufficient to prove (13) instead of (12). Equation (13) will be proved by induction. When $k = 0$, it holds trivially because both sides are zero. We assume that (13) holds for some k $(0 \le k < W)$ and show that it also holds for $k+1$.

Let us choose the $2k$ off-diagonal blocks of $A^{(k)}$ that give the k smallest weights, which are computed according to (1) (some chosen weights may be equal). Denote their index set by $\mathcal{P}_k'$. It follows from the definition of weights that the set $\mathcal{P}_k'$ is symmetric, i.e., if $(I, J) \in \mathcal{P}_k'$, then $(J, I) \in \mathcal{P}_k'$, and (X_k, Y_k), $(Y_k, X_k) \notin \mathcal{P}_k'$. Note that (13) also holds for $\mathcal{P}_k'$.

To accomplish the transition from iteration step k to $k+1$, let us define the index set $\mathcal{P}_{k+1} \equiv \mathcal{P}_k' \cup \{(X_k, Y_k), (Y_k, X_k)\}$. Note that $\mathcal{P}_{k+1}$ extends the index set $\mathcal{P}_k'$ of $2k$ off-diagonal blocks with smallest weights at iteration step k by block indices of two off-diagonal blocks that will become zero at iteration step $k+1$. Hence, in some sense, $\mathcal{P}_{k+1}$ is a "bridge" between iteration steps k and $k+1$. Moreover, $\mathcal{P}_{k+1}$ is symmetric, $|\mathcal{P}_{k+1}| = 2(k+1)$ and the left-hand side of (13) can be computed for $k+1$ as

$$
\begin{aligned}
\sum_{(I,J)\in\mathcal{P}_{k+1}} \|A_{IJ}^{(k+1)}\|_F^2 &= \sum_{(I,J)\in\mathcal{P}_k'} \|A_{IJ}^{(k+1)}\|_F^2 + \|A_{X_kY_k}^{(k+1)}\|_F^2 + \|A_{Y_kX_k}^{(k+1)}\|_F^2 \\
&= \sum_{(I,J)\in\mathcal{P}_k'} \left(\|A_{IJ}^{(k)}\|_F^2 + \|A_{IJ}^{(k+1)}\|_F^2 - \|A_{IJ}^{(k)}\|_F^2\right) \\
&\le \sum_{(I,J)\in\mathcal{P}_k'} \|A_{IJ}^{(k)}\|_F^2 + \sum_{(I,J)\in\mathcal{P}_k'} \left|\|A_{IJ}^{(k+1)}\|_F^2 - \|A_{IJ}^{(k)}\|_F^2\right|.
\end{aligned}
\tag{14}
$$

To derive the second equality in (14), we used the fact that both $A_{X_kY_k}^{(k+1)}$ and $A_{Y_kX_k}^{(k+1)}$ become zero due to annihilation.

Now we show that the index set $\mathcal{P}_k'$ in the last sum of (14), which comprises all changes $|\|A_{IJ}^{(k+1)}\|_F^2 - \|A_{IJ}^{(k)}\|_F^2|$, can be reduced to a certain symmetric index subset $\mathcal{Q}_k \subseteq \mathcal{P}_k'$ defined as

$$
\begin{aligned}
\mathcal{Q}_k \equiv \{&(X_k, J), (J, X_k) | (X_k, J) \in \mathcal{P}_k', (Y_k, J) \notin \mathcal{P}_k'\} \\
&\cup \{(Y_k, J), (J, Y_k) | (Y_k, J) \in \mathcal{P}_k', (X_k, J) \notin \mathcal{P}_k'\}.
\end{aligned}
$$

Note that the index set $\mathcal{Q}_k$ has the following property: when the block index J is fixed, then only one of (X_k, J) and (Y_k, J) (or (J, X_k) and (J, Y_k)) can belong to $\mathcal{P}_k'$ together with its symmetric counterpart.

Recall that the transition from the iteration step k to $k+1$ involves a unitary update of exactly two block rows with indices X_k, Y_k from the left, and another unitary update of exactly two block columns with the same indices from the right. For a fixed block index J, the unitary update from the left combines the off-diagonal blocks $A_{X_k,J}^{(k)}$ and $A_{Y_k,J}^{(k)}$, whereas the unitary update from the right combines the off-diagonal blocks $A_{J,X_k}^{(k)}$ and $A_{J,Y_k}^{(k)}$. When $(I, J) \in \mathcal{P}_k' \backslash \mathcal{Q}_k$, either $A_{IJ}^{(k)}$ is not affected by annihilation (because it does not lie in block rows or columns X_k and Y_k), or both (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 Frobenius norms of these two blocks is not changed after annihilation (because of unitary updates). Hence, the change of $A_{IJ}^{(k)}$ only contributes to the change of $\sum_{(I,J)\in\mathcal{P}_k'} \|A_{IJ}^{(k)}\|_F^2$ when $(I, J) \in \mathcal{Q}_k$. Consequently, (14) can be written as:

$$\sum_{(I,J)\in\mathcal{P}_{k+1}} \|A_{IJ}^{(k+1)}\|_F^2 \le \sum_{(I,J)\in\mathcal{P}'_k} \|A_{IJ}^{(k)}\|_F^2 + \sum_{(I,J)\in\mathcal{Q}_k} \left| \|A_{IJ}^{(k+1)}\|_F^2 - \|A_{IJ}^{(k)}\|_F^2 \right|. \tag{15}$$

Now we evaluate the second term of (15). Let us consider the case of $I = X_k$ and $J \neq X_k, Y_k$. From the assumption $\|U_{Y_kX_k}^{(k)}\|_2 \le \|C^{(k)}\|_F/\delta$ and Lemma 1, it follows that

$$\left| \|A_{X_kJ}^{(k+1)}\|_F^2 - \|A_{X_kJ}^{(k)}\|_F^2 \right| \le \frac{\|A_{S_kT_k}^{(k)}\|_F^2}{\delta^2} \|C^{(k)}\|_F^2 + 2\frac{\|A_{S_kT_k}^{(k)}\|_F}{\delta} \|C^{(k)}\|_F \|A_{X_kJ}^{(k)}\|_F. \tag{16}$$

Other cases can be treated in a similar way (using also $\|V_{Y_kX_k}^{(k)}\|_2 \le \|C^{(k)}\|_F/\delta$). 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

$$\begin{aligned} \sum_{(I,J)\in\mathcal{Q}_k} \|A_{IJ}^{(k)}\|_F &\le \sqrt{2w-4} \sqrt{\sum_{(I,J)\in\mathcal{Q}_k} \|A_{IJ}^{(k)}\|_F^2} \\ &\le \sqrt{2w-4} \sqrt{\sum_{(I,J)\in\mathcal{P}'_k} \|A_{IJ}^{(k)}\|_F^2}, \end{aligned} \tag{17}$$

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 (16) and (17), we can evaluate the second term of (15) as

$$\begin{aligned} \sum_{(I,J)\in\mathcal{Q}_k} \left| \|A_{IJ}^{(k+1)}\|_F^2 - \|A_{IJ}^{(k)}\|_F^2 \right| &\le \frac{(2w-4)\, \|A_{S_kT_k}^{(k)}\|_F^2 \|C^{(k)}\|_F^2}{\delta^2} + \\ &+ \frac{2\sqrt{2w-4}\, \|A_{S_kT_k}^{(k)}\|_F \|C^{(k)}\|_F}{\delta} \sqrt{\sum_{(I,J)\in\mathcal{P}'_k} \|A_{IJ}^{(k)}\|_F^2}. \end{aligned}$$

Inserting this upper bound into (15) and using the estimate

$$\|A_{S_kT_k}^{(k)}\|_F \le \sqrt{\|A_{X_kY_k}^{(k)}\|_F^2 + \|A_{Y_kX_k}^{(k)}\|_F^2} = \|C^{(k)}\|_F,$$

we finally get:

$$\begin{aligned} \sum_{(I,J)\in\mathcal{P}_{k+1}} \|A_{IJ}^{(k+1)}\|_F^2 &\le \left(\frac{\sqrt{2w-4}\, \|C^{(k)}\|_F^2}{\delta} + \sqrt{\sum_{(I,J)\in\mathcal{P}'_k} \|A_{IJ}^{(k)}\|_F^2} \right)^2 \\ &\le \left(\frac{\sqrt{2w-4}\, \|C^{(k)}\|_F^2}{\delta} + \frac{\sqrt{2w-4}\, (2\|\mathrm{off}(A^{(0)})\|_F^2 - 2\|\mathrm{off}(A^{(k)})\|_F^2)}{2\delta} \right)^2 \\ &= \frac{w-2}{2} \left(\frac{2\|\mathrm{off}(A^{(0)})\|_F^2 - 2\|\mathrm{off}(A^{(k+1)})\|_F^2}{\delta} \right)^2. \end{aligned} \tag{18}$$

Here we used (13) in the second inequality, which is also valid for $\mathcal{P}'_k$ for the following reason. Since $|\mathcal{P}'_k| = |\mathcal{P}_k| = 2k$ and the symmetric index set $\mathcal{P}'_k$ contains block indices of the off-diagonal blocks of $A^{(k)}$ with the smallest weights, we have:

$$\sum_{(I,J)\in\mathcal{P}'_k} \|A_{IJ}^{(k)}\|_F^2 \le \sum_{(I,J)\in\mathcal{P}_k} \|A_{IJ}^{(k)}\|_F^2.$$

Note that because the index set $\mathcal{P}'_k$ and $\mathcal{P}_k$ was constructed at iteration step k and $k-1$, respectively, there is no simple relation between them. However, both index sets are used for summation at the *same* iteration step k.

Finally, the last equality in (18) comes from

$$2\|\mathrm{off}(A^{(k+1)})\|_F^2 = 2\|\mathrm{off}(A^{(k)})\|_F^2 - 2\|C^{(k)}\|_F^2.$$

The final upper bound in (18) shows that (13) also holds for $k+1$ and this completes the proof. ∎

From Theorem 2.2, one immediately gets the AQC of the off-diagonal Frobenius norm of the scaled iterated matrix.

COROLLARY 2.3. *Under assumptions of Theorem* 2.2 *and by setting* $\mu \equiv \delta/\sigma_1$,

$$\|\mathrm{off}(A_{\mathrm{sc}}^{(W)})\|_F \le \sqrt{2(w-2)}\,\kappa(A)\frac{\|\mathrm{off}(A_{\mathrm{sc}}^{(0)})\|_F^2}{\mu}, \tag{19}$$

where $\kappa(A) \equiv \sigma_1/\sigma_n$ *is the* 2*-norm condition number of matrix* A.

Proof. Multiply both sides of (12) by two positive constants $\|(D_L^{(W)})^{-1}\|_2$ and $\|(D_R^{(W)})^{-1}\|_2$ resulting in:

$$\begin{aligned}\|(D_L^{(W)})^{-1}\|_2\,\|\mathrm{off}(A^{(W)})\|_F\,\|(D_R^{(W)})^{-1}\|_2 \le \sqrt{2(w-2)}\|(D_L^{(W)})^{-1}\|_2 \\ \cdot\|(D_R^{(W)})^{-1}\|_2\frac{\|\mathrm{off}(A^{(0)})\|_F^2}{\delta}.\end{aligned} \tag{20}$$

Let us bound the left-hand side (LHS) of (20) from below as follows:

$$\begin{aligned}&\|(D_L^{(W)})^{-1}\|_2\,\|\mathrm{off}(A^{(W)})\|_F\,\|(D_R^{(W)})^{-1}\|_2\\ &\qquad= \|(D_L^{(W)})^{-1}\|_2\,\|A^{(W)} - \mathrm{diag}(d^{(k)})\|_F\,\|(D_R^{(W)})^{-1}\|_2\\ &\ge \|(D_L^{(W)})^{-1}\,A^{(W)}\,(D_R^{(W)})^{-1} - (D_L^{(W)})^{-1}\,\mathrm{diag}(d^{(k)})\,(D_R^{(W)})^{-1}\|_F = \|\mathrm{off}(A_{\mathrm{sc}}^{(W)})\|_F.\end{aligned}$$

Next, we bound the RHS of (20) from above. Let a matrix A have ordered singular values $\sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_n$. Since the unitary updates do not change singular values, all iterated matrices $A^{(k)}$ have the same singular values as A. The upper bounds below, which follow from the definition of the spectral norm of diagonal matrices, and well-known relations between σ_1 and σ_n on one side and Euclidean norms of rows and columns of $A^{(k)}$ on the other side, will be useful:

$$\begin{aligned}\|(D_L^{(W)})^{-1}\|_2 &= \frac{1}{\min_{1\le i\le n}\{\|r_i^{(W)}\|^{1/2}\}} \le \frac{1}{\sigma_n^{1/2}},\\ \|(D_R^{(W)})^{-1}\|_2 &= \frac{1}{\min_{1\le i\le n}\{\|c_i^{(W)}\|^{1/2}\}} \le \frac{1}{\sigma_n^{1/2}},\\ \|(D_L^{(0)})\|_2^2 &= \left(\max_{1\le i\le n}\{\|r_i^{(0)}\|^{1/2}\}\right)^2 \le \sigma_1,\\ \|(D_R^{(0)})\|_2^2 &= \left(\max_{1\le i\le n}\{\|c_i^{(0)}\|^{1/2}\}\right)^2 \le \sigma_1.\end{aligned}$$

Then

$$
\begin{aligned}
&\sqrt{2(w-2)}\, \|(D_L^{(W)})^{-1}\|_2\, \|(D_R^{(W)})^{-1}\|_2\, \frac{\|\mathrm{off}(A^{(0)})\|_F^2}{\delta} \\
&= \sqrt{2(w-2)}\, \|(D_L^{(W)})^{-1}\|_2\, \|(D_R^{(W)})^{-1}\|_2\, \frac{\|A^{(0)} - \mathrm{diag}(d^{(0)})\|_F^2}{\delta} \\
&= \sqrt{2(w-2)}\, \|(D_L^{(W)})^{-1}\|_2\, \|(D_R^{(W)})^{-1}\|_2 \\
&\quad \cdot \frac{\|D_L^{(0)}(D_L^{(0)})^{-1}A^{(0)}(D_R^{(0)})^{-1}D_R^{(0)} - D_L^{(0)}(D_L^{(0)})^{-1}\mathrm{diag}(d^{(0)})(D_R^{(0)})^{-1}D_R^{(0)}\|_F^2}{\delta} \\
&\le \sqrt{2(w-2)}\, \|(D_L^{(W)})^{-1}\|_2\, \|(D_R^{(W)})^{-1}\|_2\, \|D_L^{(0)}\|_2^2\, \|D_R^{(0)}\|_2^2\, \frac{\|\mathrm{off}(A_{\mathrm{sc}}^{(0)})\|_F^2}{\delta} \\
&\le \sqrt{2(w-2)}\, \frac{\sigma_1}{\sigma_n}\, \frac{\|\mathrm{off}(A_{\mathrm{sc}}^{(0)})\|_F^2}{(\delta/\sigma_1)}.
\end{aligned}
$$

Using μ, $\kappa(A)$, and applying the above estimates to both sides of (20), we get the AQC of the off-diagonal Frobenius norm of the scaled iterated matrix $A_{\mathrm{sc}}^{(k)}$ after W serial iteration steps. ∎

2.2. Well-separated singular values. Now we identify the constant δ for well-separated singular values. The approach is very similar to that for well-separated eigenvalues of Hermitian matrices in [22]. Let A be a square matrix of order n with q different singular values:

$$\sigma_1 = \cdots = \sigma_{s_1} > \sigma_{s_1+1} = \cdots = \sigma_{s_2} > \cdots > \sigma_{s_{q-1}+1} = \cdots = \sigma_{s_q},$$

where $n_i = s_i - s_{i-1}$, $1 \le i \le q$, is the multiplicity of σ_{s_i} (with $s_0 = 0$ and $s_q = n$). Let the *absolute gap* d_s be defined as

$$d_s \equiv \min_{i \ne j} |\sigma_{s_i} - \sigma_{s_j}|. \tag{21}$$

Writing

$$A^{(k)} = \mathrm{diag}(A^{(k)}) + \mathrm{off}(A^{(k)}), \tag{22}$$

we can make the following assumptions at some iteration step k:

A1 The off-diagonal Frobenius norm of $A^{(k)}$ is small enough:

$$\|\mathrm{off}(A^{(k)})\|_F = \sqrt{\sum_{I \ne J} \|A_{IJ}^{(k)}\|_F^2} < \frac{d_s}{4}. \tag{23}$$

A2 The main diagonal of $A^{(k)}$ is ordered (e.g., decreasingly) by suitable row and column permutations so that the diagonal elements of $A^{(k)}$ affiliated with the same multiple singular value occupy successive positions on the diagonal.

A3 The partition of $A^{(k)}$ is such that the diagonal elements affiliated with the same multiple singular value are confined to a single diagonal block.

To keep the assumption **A3** in practical computations, one needs either some a priori information about the distribution of singular values, or one should choose the maximal block size corresponding to the size of the fast internal memory of a processor to minimize the spread of multiple singular values over diagonal blocks.

Since all transformations are unitary, the singular values of A are the same as those of $A^{(k)}$. But then, according to (22), $\mathrm{off}(A^{(k)})$ is a perturbation of $\mathrm{diag}(A^{(k)})$, and it is bounded in the Frobenius norm by $d_s/4$ because of **A1**. According to the Hoffman–Wielandt theorem [13, 24], which is also valid for singular values, for each i, $1 \le i \le q$, there are exactly n_i diagonal elements of $A^{(k)}$ that lie around σ_{s_i} in the circle of radius less than $d_s/4$. Recall that according to assumption **A2** these diagonal elements occupy successive positions on the diagonal, i.e., they form *clusters* $\widehat{Cl}(k)_i$, $1 \le i \le q$. Note that two different clusters are separated *at least* by $d_s/2$ at iteration step k.

Now we show that these clusters are *stabilized*, i.e., a diagonal element that lies in the circle around σ_{s_i} can not "jump" into a circle around σ_{s_j} for $j \ne i$.

LEMMA 2.4. *Under assumptions* **A1**–**A3**, *let a cluster* $\widehat{Cl}(k)_i$, $1 \le i \le q$, *lie inside the diagonal block* $A_{tt}^{(k)}$ *for some fixed* t, $1 \le t \le w$. *Assume that the algorithm uses the LODE in each iteration step. Then, for all iteration steps* r, $r \ge k$, n_i *elements of* $\widehat{Cl}(r)_i$ *occupy successive positions on the diagonal inside the same diagonal block* $A_{tt}^{(r)}$. *Consequently, the distance between any two different clusters remains at least* $d_s/2$.

Proof. Since the Hoffman–Wielandt theorem [13] also applies for the perturbation of singular values, the proof is identical to that of Lemma 2 in [22]. ◻

The stabilization of the clusters of diagonal elements means that the diagonal elements of $\tilde{A}^{(r)}$ and $\hat{A}^{(r+1)}$ approximate the *same* singular values of A with the *same* number of corresponding diagonal elements for $r \ge k$. Moreover, due to the LODE, the diagonal elements of both $\tilde{A}^{(r)}$ and $\hat{A}^{(r+1)}$ are ordered in the same way, e.g., decreasingly.

Finally, the next lemma gives the value of constant δ.

LEMMA 2.5. *In the case of well-separated singular values (simple and/or multiple) of* A, *under assumptions* **A1**–**A3** *above and using the LODE, the constant* δ *in Theorem* 2.2 *can be set to* $\delta = \sqrt{2}d_s/4$ *where* d_s *is the absolute gap defined by* (21).

Proof. Let us analyze an iteration step $r \to r+1$, $r \ge k$. Recall that the 2×2 block subproblem from (2) has the form:

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

Applying the Hermitian operator, using the fact that both $\tilde{U}^{(r)}$ and $\tilde{V}^{(r)}$ are unitary (see (3)), and noting that the diagonal blocks $A_{XX}^{(r)}, A_{YY}^{(r)}$, $A_{XX}^{(r+1)}$, and $A_{YY}^{(r+1)}$ are diagonal and real, we get the additional relation

$$\begin{pmatrix} A_{XX}^{(r)} & A_{YX}^{(r)H} \\ A_{XY}^{(r)H} & A_{YY}^{(r)} \end{pmatrix} \begin{pmatrix} U_{XX}^{(r)} & U_{XY}^{(r)} \\ U_{YX}^{(r)} & U_{YY}^{(r)} \end{pmatrix} = \begin{pmatrix} V_{XX}^{(r)} & V_{XY}^{(r)} \\ V_{YX}^{(r)} & V_{YY}^{(r)} \end{pmatrix} \begin{pmatrix} A_{XX}^{(r+1)} & 0 \\ 0 & A_{YY}^{(r+1)} \end{pmatrix}.$$

Now take the equations for the block $(2,1)$ from both relations:

$$\begin{aligned} A_{YY}^{(r)} V_{YX}^{(r)} - U_{YX}^{(r)} A_{XX}^{(r+1)} &= -A_{YX}^{(r)} V_{XX}^{(r)}, \\ A_{YY}^{(r)} U_{YX}^{(r)} - V_{YX}^{(r)} A_{XX}^{(r+1)} &= -A_{XY}^{(r)H} U_{XX}^{(r)}. \end{aligned}$$

This system can be written as the Sylvester equation [10] for $\left(\begin{smallmatrix} U_{YX}^{(r)} \\ V_{YX}^{(r)} \end{smallmatrix}\right)$:

$$\begin{pmatrix} 0 & A_{YY}^{(r)} \\ A_{YY}^{(r)} & 0 \end{pmatrix} \begin{pmatrix} U_{YX}^{(r)} \\ V_{YX}^{(r)} \end{pmatrix} - \begin{pmatrix} U_{YX}^{(r)} \\ V_{YX}^{(r)} \end{pmatrix} A_{XX}^{(r+1)} = - \begin{pmatrix} A_{YX}^{(r)} V_{XX}^{(r)} \\ A_{XY}^{(r)H} U_{XX}^{(r)} \end{pmatrix}. \tag{24}$$

Note that the blocks $A_{YY}^{(r)}$ and $A_{XX}^{(r+1)}$ are diagonal and their eigenvalues are diagonal elements, which are all nonnegative. Hence, the spectrum of the first matrix on the LHS of (24), denoted by

$$E^{(r)} = \begin{pmatrix} 0 & A_{YY}^{(r)} \\ A_{YY}^{(r)} & 0 \end{pmatrix},$$

consists of diagonal elements of $A_{YY}^{(r)}$, where each diagonal element is present with the plus and minus sign. Recall that according to the construction of the matrix partition, the eigenvalues of $A_{YY}^{(r)}$ and $A_{XX}^{(r+1)}$ approximate *different* singular values of A. Using Lemma 2.4, the spectra of $E^{(r)}$ and $A_{XX}^{(r+1)}$ are disjoint, and the entire spectrum of $A_{XX}^{(r+1)}$ lies on the real axis either to the right of the entire spectrum of $E^{(r)}$, or between its positive and negative part. Thus the distance between the spectra of $E^{(r)}$ and $A_{XX}^{(r+1)}$ is at least $d_s/2$. Therefore, we can apply the Davis–Kahan lemma [5, Thm. 5.1] stating that the Sylvester equation (24) has the unique solution $\begin{pmatrix} U_{YX}^{(r)} \\ V_{YX}^{(r)} \end{pmatrix}$ and its spectral norm is bounded by

$$\begin{aligned} \left\| \begin{pmatrix} U_{YX}^{(r)} \\ V_{YX}^{(r)} \end{pmatrix} \right\|_2 &\leq \frac{2}{d_s} \left\| - \begin{pmatrix} A_{YX}^{(r)} V_{XX}^{(r)} \\ A_{XY}^{(r)H} U_{XX}^{(r)} \end{pmatrix} \right\|_2 = \frac{2}{d_s} \left\| \begin{pmatrix} A_{YX}^{(r)} & 0 \\ 0 & A_{XY}^{(r)H} \end{pmatrix} \begin{pmatrix} V_{XX}^{(r)} \\ U_{XX}^{(r)} \end{pmatrix} \right\|_2 \\ &\leq \frac{2}{d_s} \left\| \begin{pmatrix} A_{YX}^{(r)} & 0 \\ 0 & A_{XY}^{(r)H} \end{pmatrix} \right\|_F \left\| \begin{pmatrix} V_{XX}^{(r)} \\ U_{XX}^{(r)} \end{pmatrix} \right\|_2 \\ &= \frac{2}{d_s} \|C^{(r)}\|_F \left\| \begin{pmatrix} V_{XX}^{(r)} \\ U_{XX}^{(r)} \end{pmatrix} \right\|_2 . \end{aligned}$$

However, due to the cosine–sine decomposition [10] of unitary matrices $\tilde{U}^{(r)}$ and $\tilde{V}^{(r)}$ (see (3)),

$$\begin{aligned} \left\| \begin{pmatrix} V_{XX}^{(r)} \\ U_{XX}^{(r)} \end{pmatrix} \right\|_2^2 &= \left\| (V_{XX}^{(r)H} U_{XX}^{(r)H}) \begin{pmatrix} V_{XX}^{(r)} \\ U_{XX}^{(r)} \end{pmatrix} \right\|_2 = \|V_{XX}^{(r)H} V_{XX}^{(r)} + U_{XX}^{(r)H} U_{XX}^{(r)}\|_2 \\ &\leq \|V_{XX}^{(r)H} V_{XX}^{(r)}\|_2 + \|U_{XX}^{(r)H} U_{XX}^{(r)}\|_2 \leq 2, \end{aligned}$$

and we get

$$\left\| \begin{pmatrix} U_{YX}^{(r)} \\ V_{YX}^{(r)} \end{pmatrix} \right\|_2 \leq \frac{2\sqrt{2}}{d_s} \|C^{(r)}\|_F.$$

Hence, $\delta = \sqrt{2} d_s/4$ and the AQC proved in Theorem 2.2 is ensured. ☐

2.3. Clusters of singular values. Again, the identification of δ is very similar to the case of clusters of eigenvalues for Hermitian matrices in [22]. If A has singular values which can be divided into one or more tight clusters, the quantity d_s in (21) can be tiny. Then the assumption **A1** in subsection 2.2 becomes useless in practice because it requires that $\|\mathrm{off}(A^{(k)})\|_F$ is even smaller than d_s. For Hermitian matrices and clusters of eigenvalues, Hari [11] suggested to use another spectral gap d_c which can be much larger than d_s. His approach can be generalized to the case of clusters of singular values as follows.

Let us group the singular values of A into q sets of very close ones (clusters):

$$Cl_i = \{\sigma_{s_{i-1}+1}, \ldots, \sigma_{s_i}\}, \quad 1 \leq i \leq q,$$

where $s_0 = 0,\ s_q = n$. As above, $n_i = s_i - s_{i-1} \geq 1$ is the number of singular values inside the ith cluster Cl_i. For each cluster, define its average value,

$$c_i \equiv \frac{1}{n_i} \sum_{j=1}^{n_i} \sigma_{s_{i-1}+j},$$

and assume that c_i's are ordered decreasingly, $c_1 > c_2 > \ldots > c_q$.

Let $A = U\Sigma V^H$ be the SVD of A, and write

$$\Sigma = \Sigma_c + \Sigma_E, \quad \text{where} \quad \Sigma_c = \operatorname{diag}(c_1, \ldots, c_1, \ldots, c_q, \ldots, c_q)$$

with c_i, $1 \leq i \leq q$, repeated n_i times. Then,

$$A = A_c + E, \quad A_c = U\Sigma_c V^H, \quad E = U\Sigma_E V^H.$$

A_c has multiple singular values $c_i,\ 1 \leq i \leq q$, and $\|E\|_F$ is tiny for tight clusters. In particular, $\|E\|_2$ is not larger than the width of the largest cluster, and

$$\|E\|_F = \sqrt{\sum_{i=1}^{q} \sum_{j=1}^{n_i} |\sigma_{s_{i-1}+j} - c_i|^2}.$$

Similarly to [11], let us define $A_c^{(k)}$ and $E^{(k)}$ for $k \geq 1$ by

$$A_c^{(k+1)} \equiv (U^{(k)})^H A_c^{(k)} V^{(k)}, \quad E^{(k+1)} \equiv (U^{(k)})^H E^{(k)} V^{(k)},$$

where $A_c^{(1)} = A_c,\ E^{(1)} = E$. Then $A^{(k)} = A_c^{(k)} + E^{(k)}$ and since a two-sided unitary transformation does not change the Frobenius norm of a matrix, $\|E^{(k)}\|_F = \|E\|_F,\ k \geq 1$.

Let us define the *absolute gap for clusters* by

$$d_c \equiv \min_{i \neq j} |c_i - c_j|, \quad 1 \leq i, j \leq q. \tag{25}$$

Now we formulate asymptotic assumptions for the case of clusters of singular values at some iteration step k.

B1 $\|\operatorname{off}(A^{(k)})\|_F$ and $\|E^{(k)}\|_F = \|E\|_F$ are small quantities:

$$\|\operatorname{off}(A^{(k)})\|_F < \frac{d_c}{8}, \quad \|E^{(k)}\|_F < \frac{d_c}{8}.$$

B2 The main diagonal of $A^{(k)}$ is ordered (e.g., decreasingly) by suitable row and column permutations so that the diagonal elements of $A^{(k)}$ affiliated with the cluster of singular values $Cl_i,\ 1 \leq i \leq q$, occupy successive positions on the diagonal and can be grouped into the cluster $\widehat{Cl}(k)_i,\ 1 \leq i \leq q$.

B3 The partition of $A^{(k)}$ is such that the diagonal elements affiliated with the same cluster Cl_i of singular values are confined to a single diagonal block.

Note that the assumption $\|E^{(k)}\|_F < d_c/8$ is essentially the assumption on the tightness of clusters of A's singular values. Similarly to the case of well-separated singular values, the assumption **B3** requires either some a priori information about the distribution of singular values, or a maximal block size corresponding to the size of the fast internal memory of a processor should be used to minimize the spread of clusters over diagonal blocks.

The next lemma shows that the clusters $\widehat{Cl}(k)_i,\ 1 \leq i \leq q$, of diagonal elements of $A^{(k)}$ are stabilized.

LEMMA 2.6. *Under assumptions* **B1**–**B3**, *let a cluster* $\widehat{Cl}(k)_i$, $1 \leq i \leq q$, *lie inside the diagonal block* $A_{tt}^{(k)}$ *for some fixed* t, $1 \leq t \leq w$. *Assume that the algorithm uses the LODE in each iteration step. Then, for all iteration steps* r, $r \geq k$, n_i *elements of* $\widehat{Cl}(r)_i$ *occupy successive positions on the diagonal inside the same diagonal block* $A_{tt}^{(r)}$. *Consequently, the distance between any two different clusters remains at least* $d_c/2$.

Proof. Since the Hoffman–Wielandt theorem also applies to the perturbation of singular values, the proof is identical to that of Lemma 4 in [22]. □

Finally, the value of the constant δ can be identified.

LEMMA 2.7. *For clusters of singular values of* A, *under assumptions* **B1**–**B3** *above and using the LODE, the constant* δ *in Theorem* 2.2 *can be set to* $\delta = \sqrt{2}d_c/4$ *where* d_c *is the gap for clusters defined by* (25).

Proof. Repeating the proof of Lemma 2.5, however, now working with c_i and c_j instead of σ_{s_i} and σ_{s_j}, respectively, we get the value $d_c/2$ for the lower bound of the distance between spectra of two corresponding diagonal blocks in the Sylvester equation (24). Then, repeat the remaining estimates of Lemma 2.5. □

Remark 2.8. After the identification of δ, the constant μ in Corollary 2.3 has the value $\frac{\sqrt{2}d_s}{4\sigma_1}$ (or $\frac{\sqrt{2}d_c}{4\sigma_1}$) for well-separated singular values (or clusters); i.e., it is proportional to the absolute gap (or the absolute gap for clusters) scaled by the maximal singular value. Note that $\mu \ll 1$ for $\sigma_1 \gg d_s$ (or $\sigma_1 \gg d_c$).

3. Parallel two-sided SVD algorithm.

Let us divide a square matrix A of order n into a $w \times w$ block structure using the blocking factor $w = 2p$, $w \geq 4$, where p is the number of processors. Here we assume that w divides n (if not, use the bordering of A described at the beginning of section 2). Thus, w denotes the number of blocks in each block row (column) and each block has size $\ell \times \ell$ where $\ell = n/(2p)$.

At the beginning of a parallel iteration step k, $2p$ off-diagonal blocks of $A^{(k)}$ with block indices $(X_{k,1}, Y_{k,1})$, $(Y_{k,1}, X_{k,1})$, $\ldots$, $(X_{k,p}, Y_{k,p})$, $(Y_{k,p}, X_{k,p})$, $X_{k,i} < Y_{k,i}$ for all i, are annihilated using the greedy implementation of parallel dynamic ordering (GIPDO). The pairs of off-diagonal blocks are ordered decreasingly with respect to their weights $w_{IJ}^{(k)}$ measured by the sum of squares of their Frobenius norms (see (1)). After choosing the first pair, additional $p-1$ pairs are chosen for annihilation with a decreasing weight in a compatible way; i.e., the block indices of each new pair must be different from those of all already chosen blocks. This ensures the selection of p 2×2 block subproblems that can be solved in parallel. More details about the communication and computational complexity of GIPDO can be found in [1].

After the GIPDO is computed, p chosen pairs together with their corresponding diagonal blocks are met in p processors (one pair per processor), and p 2×2-block SVD subproblems are computed in parallel. At parallel iteration step k, the processor i, $1 \leq i \leq p$, solves the local subproblem of size $2\ell \times 2\ell$,

$$\begin{pmatrix} U_{X_{k,i}X_{k,i}}^{(k)} & U_{X_{k,i}Y_{k,i}}^{(k)} \\ U_{Y_{k,i}X_{k,i}}^{(k)} & U_{Y_{k,i}Y_{k,i}}^{(k)} \end{pmatrix}^H \begin{pmatrix} A_{X_{k,i}X_{k,i}}^{(k)} & A_{X_{k,i}Y_{k,i}}^{(k)} \\ A_{Y_{k,i}X_{k,i}}^{(k)} & A_{Y_{k,i}Y_{k,i}}^{(k)} \end{pmatrix} \begin{pmatrix} V_{X_{k,i}X_{k,i}}^{(k)} & V_{X_{k,i}Y_{k,i}}^{(k)} \\ V_{Y_{k_i}X_{k,i}}^{(k)} & V_{Y_{k,i}Y_{k,i}}^{(k)} \end{pmatrix} = \begin{pmatrix} A_{X_{k,i}X_{k,i}}^{(k+1)} & 0 \\ 0 & A_{Y_{k,i}Y_{k,i}}^{(k+1)} \end{pmatrix},$$

where the diagonal blocks $A^{(k+1)}_{X_{k,i}X_{k,i}}$ and $A^{(k+1)}_{Y_{k,i}Y_{k,i}}$ are square, diagonal matrices of order ℓ, because all diagonal blocks are diagonal after the first parallel iteration step and remain so during the whole computation.

Similarly to the serial case in (3) and (4), we define the unitary matrix of local left and right singular vectors $\tilde{U}_i^{(k)}$ and $\tilde{V}_i^{(k)}$, respectively, and the diagonal matrix $\tilde{\Sigma}_i^{(k)}$ of local singular values for each i, $1 \le i \le p$. These three matrices constitute the local SVD of matrix $\tilde{A}_i^{(k)}$.

As in the serial algorithm, local singular values in any 2×2 block subproblem can be computed and located in any order on the diagonal. If they are ordered in the same way in each processor and each parallel iteration step, we speak about the LODE. Moreover, the scaled iterated matrix $A^{(k)}_{\mathrm{sc}}$ in the parallel algorithm is defined by (7).

The proof of global convergence may be identical to that of the serial algorithm (see (5)). However, in [17] a better upper bound was derived for the case of the EVD of Hermitian matrices. This proof may also be applied to the case of SVD directly, giving the following upper bound:

$$\|\mathrm{off}(A^{(k+1)})\|_F^2 \le \left(1 - \frac{1}{2w-3}\right) \|\mathrm{off}(A^{(k)})\|_F^2. \tag{26}$$

3.1. Update of an off-diagonal block. Suppose that in a given parallel iteration step k (its index is omitted here) the off-diagonal blocks $A_{X_iY_i}$ and $A_{Y_iX_i}$, $1 \le i \le p$, were chosen by the GIPDO for annihilation. Our first step is to derive an upper bound for the change of the squared Frobenius norm of an arbitrary block that is not annihilated at parallel step k. Such a block can be written as $A_{X_iX_j}$, $A_{X_iY_j}$, $A_{Y_iX_j}$, or $A_{Y_iY_j}$, where $j \ne i$.

Let us consider the update of block rows X_i and Y_i. First, we need to evaluate the update of two off-diagonal blocks which will be combined in the subsequent update of two block columns:

$$\tilde{A}_{X_iX_j} = U^H_{X_iX_i} A_{X_iX_j} + U^H_{Y_iX_i} A_{Y_iX_j}, \tag{27}$$

$$\tilde{A}_{X_iY_j} = U^H_{X_iX_i} A_{X_iY_j} + U^H_{Y_iX_i} A_{Y_iY_j}. \tag{28}$$

Second, the update of two block columns X_j, Y_j follows from (27) and (28):

$$\begin{aligned}
\hat{A}_{X_iX_j} &= \tilde{A}_{X_iX_j} V_{X_jX_j} + \tilde{A}_{X_iY_j} V_{Y_jX_j} \\
&= U^H_{X_iX_i} A_{X_iX_j} V_{X_jX_j} + U^H_{Y_iX_i} A_{Y_iX_j} V_{X_jX_j} \\
&\quad + U^H_{X_iX_i} A_{X_iY_j} V_{Y_jX_j} + U^H_{Y_iX_i} A_{Y_iY_j} V_{Y_jX_j}.
\end{aligned}$$

In the following lemma, we bound the change of $A_{X_iX_j}$, but the same bound is applicable to the other three cases as well.

LEMMA 3.1. *Consider the change of an off-diagonal block $A_{X_iX_j}$ that was not eliminated in a given parallel iteration step k. Denote the changed block by $\hat{A}_{X_iX_j}$, and let $C_i = (A_{X_iY_i}, A_{Y_iX_i})$. Additionally, let A_{ST} be the off-diagonal block with the largest Frobenius norm, $\|A_{ST}\|_F = \max_{I \ne J} \|A_{IJ}\|_F$. If there exists a constant $\delta > 0$ such that $\|\binom{U_{Y_iX_i}}{V_{Y_iX_i}}\|_2 \le \|C_i\|_F/\delta$ for $1 \le i \le p$, then the following inequality holds:*

$$\begin{aligned}
\left| \|\hat{A}_{X_iX_j}\|_F^2 - \|A_{X_iX_j}\|_F^2 \right| \le{}& \frac{\|A_{ST}\|_F^2}{\delta^2} \left\{ (1+\sqrt{2})\|C_i\|_F^2 + (2+\sqrt{2})\|C_j\|_F^2 \right\} \\
&+ 2\,\frac{\|A_{ST}\|_F}{\delta} \left(\|C_i\|_F + \sqrt{2}\|C_j\|_F \right) \|A_{X_iX_j}\|_F.
\end{aligned} \tag{29}$$

Proof. Using the triangle inequality, we can bound the LHS of (29) as

$$\left| \|\hat{A}_{X_iX_j}\|_F^2 - \|A_{X_iX_j}\|_F^2 \right| \le \left| \|\tilde{A}_{X_iX_j}\|_F^2 - \|A_{X_iX_j}\|_F^2 \right| + \left| \|\hat{A}_{X_iX_j}\|_F^2 - \|\tilde{A}_{X_iX_j}\|_F^2 \right|. \tag{30}$$

Using (27) and (28), the first term in the RHS can be bounded using the same technique as in the proof of Lemma 1 of [22]:

$$\begin{aligned} &\left| \|\tilde{A}_{X_iX_j}\|_F^2 - \|A_{X_iX_j}\|_F^2 \right| \\ &\quad \le \|U_{Y_iX_i}\|_2^2 \max\left\{ \|A_{X_iX_j}\|_F^2, \|A_{Y_iX_j}\|_F^2 \right\} + 2\|A_{X_iX_j}\|_F \|U_{Y_iX_i}\|_2 \|A_{Y_iX_j}\|_F \\ &\quad \le \frac{\|A_{ST}\|_F^2}{\delta^2} \|C_i\|_F^2 + 2\,\frac{\|A_{ST}\|_F}{\delta} \|C_i\|_F \|A_{X_iX_j}\|_F. \end{aligned} \tag{31}$$

Similarly, the second term can be bounded as

$$\begin{aligned} &\left| \|\hat{A}_{X_iX_j}\|_F^2 - \|\tilde{A}_{X_iX_j}\|_F^2 \right| \\ &\quad \le \|V_{Y_jX_j}\|_2^2 \max\left\{ \|\tilde{A}_{X_iX_j}\|_F^2, \|\tilde{A}_{X_iY_j}\|_F^2 \right\} + 2\|\tilde{A}_{X_iX_j}\|_F \|V_{Y_jX_j}\|_2 \|\tilde{A}_{X_iY_j}\|_F. \end{aligned} \tag{32}$$

To bound the RHS, we need to evaluate $\|\tilde{A}_{X_iX_j}\|_F$ and $\|\tilde{A}_{X_iY_j}\|_F$. Using (27) and (28), we have

$$\begin{aligned} \|\tilde{A}_{X_iX_j}\|_F &\le \|A_{X_iX_j}\|_F + \|U_{Y_iX_i}\|_2 \|A_{Y_iX_j}\|_F \\ &\le \|A_{X_iX_j}\|_F + \frac{\|C_i\|_F}{\delta} \|A_{ST}\|_F. \end{aligned} \tag{33}$$

On the other hand, since the transformation is unitary, we have

$$\|\tilde{A}_{X_iX_j}\|_F^2 + \|\tilde{A}_{Y_iX_j}\|_F^2 = \|A_{X_iX_j}\|_F^2 + \|A_{Y_iX_j}\|_F^2,$$

which leads to

$$\|\tilde{A}_{X_iX_j}\|_F^2 \le \|A_{X_iX_j}\|_F^2 + \|A_{Y_iX_j}\|_F^2 \le 2\|A_{ST}\|_F^2. \tag{34}$$

Similarly,

$$\|\tilde{A}_{X_iY_j}\|_F^2 \le \|A_{X_iY_j}\|_F^2 + \|A_{Y_iY_j}\|_F^2 \le 2\|A_{ST}\|_F^2. \tag{35}$$

Putting (34) and (35) into the first term of (32) and inserting (33) and (35) into the second term of (32) gives

$$\begin{aligned} &\left| \|\hat{A}_{X_iX_j}\|_F^2 - \|\tilde{A}_{X_iX_j}\|_F^2 \right| \\ &\le \frac{\|C_j\|_F^2}{\delta^2} \cdot 2\|A_{ST}\|_F^2 + 2\left(\|A_{X_iX_j}\|_F + \frac{\|C_i\|_F}{\delta} \|A_{ST}\|_F \right) \cdot \frac{\|C_j\|_F}{\delta} \cdot \sqrt{2}\|A_{ST}\|_F \\ &= 2\,\frac{\|A_{ST}\|_F^2}{\delta^2} \left(\|C_j\|_F^2 + \sqrt{2}\,\|C_i\|_F \|C_j\|_F \right) + 2\sqrt{2}\,\frac{\|A_{ST}\|_F}{\delta} \|C_j\|_F \|A_{X_iX_j}\|_F \\ &\le \frac{\|A_{ST}\|_F^2}{\delta^2} \left\{ 2\|C_j\|_F^2 + \sqrt{2}\left(\|C_i\|_F^2 + \|C_j\|_F^2 \right) \right\} + 2\sqrt{2}\,\frac{\|A_{ST}\|_F}{\delta} \|C_j\|_F \|A_{X_iX_j}\|_F, \end{aligned} \tag{36}$$

where we used $2ab \le a^2 + b^2$ in the last inequality. Substituting (31) and (36) into (30) leads to

$$\left| \|\hat{A}_{X_iX_j}\|_F^2 - \|A_{X_iX_j}\|_F^2 \right| \le \frac{\|A_{ST}\|_F^2}{\delta^2} \left\{ (1+\sqrt{2})\|C_i\|_F^2 + (2+\sqrt{2})\|C_j\|_F^2 \right\} + 2\,\frac{\|A_{ST}\|_F}{\delta} \left(\|C_i\|_F + \sqrt{2}\|C_j\|_F \right) \|A_{X_iX_j}\|_F,$$

which completes the proof. ∎

3.2. AQC. Using Lemma 3.1, we derive a quadratic convergence bound for the parallel block Jacobi SVD algorithm with the GIPDO.

We start with the definition of two auxiliary index sets. Without loss of generality, denote the parallel iteration steps by $k = 0, 1, \ldots$, and the off-diagonal blocks chosen for annihilation at step k by $A^{(k)}_{X_{k,1}Y_{k,1}}$, $A^{(k)}_{Y_{k,1}X_{k,1}}, \ldots,$ $A^{(k)}_{X_{k,p}Y_{k,p}}$, $A^{(k)}_{Y_{k,p}X_{k,p}}$, where $A^{(k)}_{X_{k,1}Y_{k,1}}$ and $A^{(k)}_{Y_{k,1}X_{k,1}}$ are the off-diagonal blocks with the largest weight. Let $\mathcal{Q}_{k,v}$, v even, be the index set of the $v/2$ pairs of off-diagonal blocks with the smallest weights at step k (some chosen weights may be equal). Note that they are chosen in a symmetric way; i.e., if $(I,J) \in \mathcal{Q}_{k,v}$, then $(J,I) \in \mathcal{Q}_{k,v}$. Consequently, the number of elements in the set $\mathcal{Q}_{k,v}$ is $|\mathcal{Q}_{k,v}| = v$.

In addition, we define the second symmetric index set $\mathcal{P}_k$ recursively:

$$\mathcal{P}_0 \equiv \emptyset,$$

$$\mathcal{P}_{k+1} \equiv \mathcal{Q}_{k,|\mathcal{P}_k|} \cup \{(X_{k,1},Y_{k,1}),(Y_{k,1},X_{k,1}),\ldots,(X_{k,p},Y_{k,p}),(Y_{k,p},X_{k,p})\}. \tag{37}$$

Note that the set $\mathcal{P}_k$ has the same number of elements as the set $\mathcal{Q}_{k,|\mathcal{P}_k|}$.

The relation between $\mathcal{P}_{k+1}$ and $\mathcal{Q}_{k,|\mathcal{P}_k|}$ is very similar to the relation between $\mathcal{P}_{k+1}$ and $\mathcal{P}'_k$ in the serial case (see the proof of Theorem 2.2 in subsection 2.1). The index set $\mathcal{P}_{k+1}$ is a possible extension of the index set $\mathcal{Q}_{k,|\mathcal{P}_k|}$ by block indices of those off-diagonal blocks which will become zero at step $k+1$. Note, however, that in contrast to the serial case, it is not obvious how many elements the index set $\mathcal{P}_{k+1}$ will have. This is because the dynamic ordering can also choose those off-diagonal blocks for annihilation at step k, whose block indices are elements of the index set $\mathcal{Q}_{k,|\mathcal{P}_k|}$.

The next lemma gives the lower and upper bounds for the step W after which the index set $\mathcal{P}_W$ cannot be enlarged.

LEMMA 3.2. *There exists a step W, $w-1 \le W < 2w(\log w + 1)$, for which $\mathcal{P}_W$ contains the indices of all off-diagonal blocks.*

Proof. First, we show that $|\mathcal{P}_k|$ is a strictly increasing sequence and prove the existence of W. Note that $|\mathcal{P}_k|$ takes a value between 0 and $w(w-1) = 2p(2p-1)$. As will be shown below, the increase of $|\mathcal{P}_k|$ is different in each parallel step depending on the value of $|\mathcal{P}_k|$. Accordingly, we define the following $p-1$ subintervals of the interval $[0, 2p(2p-1)]$ and consider the increase of $|\mathcal{P}_k|$ in each subinterval separately:

$$I_1 \equiv [0,\, 4\cdot 3 - 2],$$
$$I_i \equiv (2i(2i-1) - 2,\, (2i+2)(2i+1) - 2] \quad i = 2, 3, \ldots, p-1.$$

Note that the intervals I_i, $2 \le i \le p-1$ are open from the left. Now consider the construction of $\mathcal{P}_{k+1}$ from $\mathcal{P}_k$. In the GIPDO, the block pair $(A^{(k)}_{X_{k,\ell}Y_{k,\ell}}, A^{(k)}_{Y_{k,\ell}X_{k,\ell}})$, $1 \le \ell \le p$, is chosen to be the pair of maximal weight under the condition that

$X_{k,\ell}, Y_{k,\ell} \notin \{X_{k,1}, Y_{k,1}, \ldots, X_{k,\ell-1}, Y_{k,\ell-1}\}$. Hence, it gives the maximal weight among $(w-(2\ell-2))(w-(2\ell-1))/2$ block pairs. In other words, among the $w(w-1)/2$ off-diagonal block pairs of $A^{(k)}$, there are at least $(w-(2\ell-2))(w-(2\ell-1))/2-1$ pairs whose weights are not larger than that of $(A^{(k)}_{X_{k,\ell}Y_{k,\ell}}, A^{(k)}_{Y_{k,\ell}X_{k,\ell}})$. Now assume that $|\mathcal{P}_k| \in I_i$. Consequently, $|\mathcal{Q}_{k,|\mathcal{P}_k|}| = |\mathcal{P}_k| \le (2i+2)(2i+1)-2$. Hence, $\mathcal{Q}_{k,|\mathcal{P}_k|}$ consists of at most $(2i+2)(2i+1)/2-1$ pairs of block indices that identify the pairs of off-diagonal blocks with the smallest weights. On the other hand, as stated above for $\ell = p-i$, there are at least

$$(w-(2(p-i)-2))(w-(2(p-i)-1))/2-1 = (2i+2)(2i+1)/2-1$$

pairs of off-diagonal blocks whose weights are not larger than the weight of the pair $(A^{(k)}_{X_{k,p-i}Y_{k,p-i}}, A^{(k)}_{Y_{k,p-i}X_{k,p-i}})$. Thus, by choosing the elements of $\mathcal{Q}_{k,|\mathcal{P}_k|}$ from them, we can ensure that the $p-i$ pairs of block indices $(X_{k,1}, Y_{k,1}), (Y_{k,1}, X_{k,1}), \ldots,$ $(X_{k,p-i}, Y_{k,p-i}), (Y_{k,p-i}, X_{k,p-i})$ do not belong to $\mathcal{Q}_{k,|\mathcal{P}_k|}$. Combining this with (37), the set $\mathcal{P}_{k+1}$ has at least $2(p-i)$ more elements than $\mathcal{P}_k$. Since the length of interval I_i is $8i+2$, it takes at most $\lfloor \frac{8i+2}{2(p-i)} \rfloor + 1$ steps for $|\mathcal{P}_k|$ to increase from the left end of I_i to its right end, where $\lfloor\cdot\rfloor$ is the floor function. Thus, the number of steps required for $|\mathcal{P}_k|$ to increase from 0 to $w(w-1)-2$ can be bounded by

$$\begin{aligned}
\sum_{i=1}^{p-1}\left(\left\lfloor \frac{8i+2}{2(p-i)} \right\rfloor + 1\right) &\le \sum_{i=1}^{p-1}\left(\frac{8i+2}{2(p-i)} + 1\right) \\
&= (4p+1)\sum_{j=1}^{p-1}\frac{1}{j} - 3(p-1) \\
&\le (4p+1)\left(1 + \int_1^{p-1}\frac{1}{x}\,dx\right) - 3(p-1) \\
&= (4p+1)(1+\log(p-1)) - 3(p-1) \\
&< 4p(1+\log p) - 1 < 2w(1+\log w) - 1.
\end{aligned}$$

Once $|\mathcal{P}_k|$ reaches $w(w-1)-2$, one more step is sufficient for it to reach $w(w-1)$. This follows from the fact that the pair of off-diagonal blocks with block indices $(X_{k,1}, Y_{k,1})$ and $(Y_{k,1}, X_{k,1})$ has the largest weight among all pairs, and therefore these block indices do not belong to $\mathcal{Q}_{k,|\mathcal{P}_k|}$ whenever $|\mathcal{Q}_{k,|\mathcal{P}_k|}| = |\mathcal{P}_k| \le w(w-1)-2$. This establishes the existence of W and, at the same time, proves its upper bound.

To prove the lower bound of W, notice that $|\mathcal{P}_k|$ can be increased by at most $w = 2p$ at each step (see (37)). Thus, it requires at least $w-1$ steps to increase $|\mathcal{P}_k|$ from 0 to $w(w-1)$. ∎

The next theorem contains the proof of the AQC of the parallel two-sided block-Jacobi SVD algorithm with the GIPDO.

THEOREM 3.3. *Consider the parallel two-sided block-Jacobi SVD algorithm with the GIPDO using the blocking factor $w = 2p$. Denote the parallel iteration steps by $k = 0, 1, \ldots$, and the off-diagonal blocks chosen for annihilation at step k by $A^{(k)}_{X_{k,1}Y_{k,1}}$, $A^{(k)}_{Y_{k,1}X_{k,1}}, \ldots, A^{(k)}_{X_{k,p}Y_{k,p}}, A^{(k)}_{Y_{k,p}X_{k,p}}$, where $A^{(k)}_{X_{k,1}Y_{k,1}}$ and $A^{(k)}_{Y_{k,1}X_{k,1}}$ are the off-diagonal blocks with the largest weight. Let $A^{(k)}_{S_kT_k}$ be the off-diagonal block with the largest*

Frobenius norm. Additionally, let $C_{k,i}^{(k)} = (A_{X_{k,i}Y_{k,i}}^{(k)}, A_{Y_{k,i}X_{k,i}}^{(k)})$, $1 \le i \le p$, *and let* W *be the integer step from Lemma* 3.2. *Furthermore, suppose that there exists a constant* $\delta > 0$ *such that*

$$\left\| \begin{pmatrix} U_{Y_{k,i}X_{k,i}}^{(k)} \\ V_{Y_{k,i}X_{k,i}}^{(k)} \end{pmatrix} \right\|_2 \le \frac{\|C_{k,i}^{(k)}\|_F}{\delta}$$

holds for all i, $1 \le i \le p$, *and* $k = 0, 1, \dots, W-1$. *Then,*

$$\|\mathrm{off}(A^{(W)})\|_F \le \sqrt{12(w-2)}\, \frac{\|\mathrm{off}(A^{(0)})\|_F^2}{\delta}, \tag{38}$$

i.e., the parallel two-sided block-Jacobi SVD algorithm with the GIPDO converges quadratically after W *parallel iteration steps.*

Proof. To prove (38), we use an alternative inequality

$$\sum_{(I,J)\in\mathcal{P}_k} \|A_{IJ}^{(k)}\|_F^2 \le \left(\frac{2+\sqrt{2}}{2}\right)^2 (w-2) \left(\frac{2\|\mathrm{off}(A^{(0)})\|_F^2 - 2\|\mathrm{off}(A^{(k)})\|_F^2}{\delta}\right)^2 \tag{39}$$

for $k = 0, 1, \dots, W$. Note that when $k = W$, the LHS becomes $\|\mathrm{off}(A^{(W)})\|_F^2$, while the RHS is smaller than the RHS of (38). So (38) follows directly from (39). We prove (39) by induction. When $k = 0$, both sides are zero, so the inequality holds trivially. We assume that (39) holds for some $k < W$ and show that it also holds for $k+1$. In the following, we omit the superscript (k) for the parallel iteration step and denote the quantities at step k and $k+1$ by symbols without and with a hat, respectively. We also write $C_{k,i}^{(k)}$, $X_{k,i}$, and $Y_{k,i}$ as C_i, X_i, and Y_i, respectively.

Let us define the index set $\mathcal{P}_k'$ by

$$\mathcal{P}_k' = \mathcal{Q}_{k,|\mathcal{P}_k|} \setminus \{(X_1, Y_1), (Y_1, X_1), \dots, (X_p, Y_p), (Y_p, X_p)\}. \tag{40}$$

Then, at parallel step $k+1$, the LHS of (39) can be evaluated as follows:

$$\begin{aligned} \sum_{(I,J)\in\mathcal{P}_{k+1}} \|\hat{A}_{IJ}\|_F^2 &= \sum_{(I,J)\in\mathcal{P}_k'} \|\hat{A}_{IJ}\|_F^2 + \sum_{i=1}^{p} (\|\hat{A}_{X_iY_i}\|_F^2 + \|\hat{A}_{Y_iX_i}\|_F^2) \\ &\le \sum_{(I,J)\in\mathcal{P}_k'} \|A_{IJ}\|_F^2 + \sum_{(I,J)\in\mathcal{P}_k'} \left| \|\hat{A}_{IJ}\|_F^2 - \|A_{IJ}\|_F^2 \right|, \end{aligned} \tag{41}$$

where we used $\|\hat{A}_{X_iY_i}\|_F^2 = \|\hat{A}_{Y_iX_i}\|_F^2 = 0$ for $1 \le i \le p$.

Let $(I, J) \in \mathcal{P}_k'$ be fixed. Since $\{X_1, Y_1, \dots, X_p, Y_p\}$ is a certain permutation of the set $\{1, 2, \dots, 2p\}$, for each I there exists exactly one index i ($1 \le i \le p$) such that $I = X_i$ or $I = Y_i$. We denote such an i by $\pi(I)$. Using the same mapping π, we can denote the index j ($1 \le j \le p$) such that $J = X_j$ or $J = Y_j$ by $\pi(J)$. Then, the off-diagonal block A_{IJ} is updated by a block rotation specified by $(X_{\pi(I)}, Y_{\pi(I)})$ from the left and by another block rotation specified by $(X_{\pi(J)}, Y_{\pi(J)})$ from the right. Note that $\pi(I) \ne \pi(J)$, because A_{IJ} is not a block chosen for annihilation. Hence, from (29) we have,

$$\begin{aligned} \left| \|\hat{A}_{IJ}\|_F^2 - \|A_{IJ}\|_F^2 \right| &\le 2\, \frac{\|A_{ST}\|_F}{\delta} (\|C_{\pi(I)}\|_F + \sqrt{2}\, \|C_{\pi(J)}\|_F) \|A_{IJ}\|_F \\ &\quad + \frac{\|A_{ST}\|_F^2}{\delta^2} \left\{ (1+\sqrt{2}) \|C_{\pi(I)}\|_F^2 + (2+\sqrt{2}) \|C_{\pi(J)}\|_F^2 \right\}. \end{aligned} \tag{42}$$

Now, consider the sum of $\|C_{\pi(I)}\|_F^2$ over $\mathcal{P}'_k$. Since $\mathcal{P}'_k \subseteq \{(I,J)\,|\,1 \le I, J \le 2p,\ \pi(I) \neq \pi(J)\}$, we can bound it by a sum over the set $\{(I,J)\,|\,1 \le I, J \le 2p,\ \pi(I) \neq \pi(J)\}$. Furthermore, there exist exactly two values of I such that $\pi(I) = i$ for each i and exactly two values of J such that $\pi(J) = j$ for each j. Hence, we can rewrite the sum over the set $\{(I,J)\,|\,1 \le I, J \le 2p,\ \pi(I) \neq \pi(J)\}$ as a sum over the set $\{(i,j)\,|\,1 \le i, j \le p,\ i \neq j\}$ multiplied by 4. Thus,

$$\begin{aligned}\sum_{(I,J)\in\mathcal{P}'_k} \|C_{\pi(I)}\|_F^2 &\le \sum_{I=1}^{2p} \sum_{\substack{J=1\\ \pi(I)\neq\pi(J)}}^{2p} \|C_{\pi(I)}\|_F^2 = 4\sum_{i=1}^{p}\sum_{\substack{j=1\\ j\neq i}}^{p} \|C_i\|_F^2 \\ &= 4(p-1)\sum_{i=1}^{p}\|C_i\|_F^2 = 2(w-2)\sum_{i=1}^{p}\|C_i\|_F^2. \end{aligned} \tag{43}$$

Similarly,

$$\sum_{(I,J)\in\mathcal{P}'_k} \|C_{\pi(J)}\|_F^2 \le 2(w-2)\sum_{i=1}^{p}\|C_i\|_F^2. \tag{44}$$

Using these results, we can bound the second term in the RHS of (41). By inserting (42) into (41), we get

$$\begin{aligned}
&\sum_{(I,J)\in\mathcal{P}'_k} \left| \|\hat{A}_{IJ}\|_F^2 - \|A_{IJ}\|_F^2 \right| \\
&\le \frac{\|A_{ST}\|_F}{\delta} \sum_{(I,J)\in\mathcal{P}'_k} (2\,\|C_{\pi(I)}\|_F + 2\sqrt{2}\,\|C_{\pi(J)}\|_F)\|A_{IJ}\|_F \\
&\quad + \frac{\|A_{ST}\|_F^2}{\delta^2} \sum_{(I,J)\in\mathcal{P}'_k} \left\{ (1+\sqrt{2})\|C_{\pi(I)}\|_F^2 + (2+\sqrt{2})\|C_{\pi(J)}\|_F^2 \right\} \\
&\le \frac{\|A_{ST}\|_F}{\delta} \left(2\sqrt{\sum_{(I,J)\in\mathcal{P}'_k} \|C_{\pi(I)}\|_F^2} + 2\sqrt{2}\sqrt{\sum_{(I,J)\in\mathcal{P}'_k} \|C_{\pi(J)}\|_F^2} \right) \sqrt{\sum_{(I,J)\in\mathcal{P}'_k} \|A_{IJ}\|_F^2} \\
&\quad + \frac{\|A_{ST}\|_F^2}{\delta^2} \left\{ (1+\sqrt{2}) \sum_{(I,J)\in\mathcal{P}'_k} \|C_{\pi(I)}\|_F^2 + (2+\sqrt{2}) \sum_{(I,J)\in\mathcal{P}'_k} \|C_{\pi(J)}\|_F^2 \right\} \\
&\le \frac{1}{\delta} \sqrt{\sum_{i=1}^{p}\|C_i\|_F^2}\,(4+2\sqrt{2})\,\sqrt{w-2}\sqrt{\sum_{i=1}^{p}\|C_i\|_F^2}\sqrt{\sum_{(I,J)\in\mathcal{P}'_k} \|A_{IJ}\|_F^2} \\
&\quad + \frac{1}{\delta^2}\left(\sum_{i=1}^{p}\|C_i\|_F^2\right)(6+4\sqrt{2})\,(w-2)\left(\sum_{i=1}^{p}\|C_i\|_F^2\right) \\
&= \frac{2(2+\sqrt{2})\sqrt{w-2}}{\delta}\left(\sum_{i=1}^{p}\|C_i\|_F^2\right)\sqrt{\sum_{(I,J)\in\mathcal{P}'_k} \|A_{IJ}\|_F^2} \\
&\quad + \frac{(2+\sqrt{2})^2(w-2)}{\delta^2}\left(\sum_{i=1}^{p}\|C_i\|_F^2\right)^2,
\end{aligned} \tag{45}$$

where we used the Cauchy–Schwarz inequality in the second inequality. In the third inequality, we used the upper bound

$$\|A_{ST}\|_F \leq \sqrt{\|A_{ST}\|_F^2 + \|A_{TS}\|_F^2} \leq \sqrt{\|A_{X_1Y_1}\|_F^2 + \|A_{Y_1X_1}\|_F^2}$$
$$\leq \sqrt{\sum_{i=1}^{p}(\|A_{X_iY_i}\|_F^2 + \|A_{Y_iX_i}\|_F^2)} = \sqrt{\sum_{i=1}^{p}\|C_i\|_F^2}$$

to bound the first factor of the first and second term, and (43) and (44) to bound the sums over $\mathcal{P}'_k$. Finally, inserting (45) into (41) gives

$$\begin{aligned}
&\sum_{(I,J)\in\mathcal{P}_{k+1}} \|\hat{A}_{IJ}\|_F^2 \\
&\quad\leq \sum_{(I,J)\in\mathcal{P}'_k} \|A_{IJ}\|_F^2 + \frac{2(2+\sqrt{2})\sqrt{w-2}}{\delta}\left(\sum_{i=1}^{p}\|C_i\|_F^2\right)\sqrt{\sum_{(I,J)\in\mathcal{P}'_k}\|A_{IJ}\|_F^2} \\
&\qquad + \frac{(2+\sqrt{2})^2(w-2)}{\delta^2}\left(\sum_{i=1}^{p}\|C_i\|_F^2\right)^2 \\
&\quad= \left\{\sqrt{\sum_{(I,J)\in\mathcal{P}'_k}\|A_{IJ}\|_F^2} + \frac{(2+\sqrt{2})\sqrt{w-2}}{2\delta}\cdot 2\sum_{i=1}^{p}\|C_i\|_F^2\right\}^2 \\
&\quad\leq \left(\frac{2+\sqrt{2}}{2}\right)^2 (w-2)\left(\frac{2\|\text{off}(A^{(0)})\|_F^2 - 2\|\text{off}(A^{(k)})\|_F^2 + 2\sum_{i=1}^{p}\|C_i\|_F^2}{\delta}\right)^2 \\
&\quad= \left(\frac{2+\sqrt{2}}{2}\right)^2 (w-2)\left(\frac{2\|\text{off}(A^{(0)})\|_F^2 - 2\|\text{off}(A^{(k+1)})\|_F^2}{\delta}\right)^2.
\end{aligned} \tag{46}$$

To derive the third inequality in (46), let us recall basic properties of the following index sets: $\mathcal{P}'_k \subseteq \mathcal{Q}_{k,|\mathcal{P}_k|}$, $\mathcal{Q}_{k,|\mathcal{P}_k|}$ contains block indices of $|\mathcal{P}_k|/2$ pairs of off-diagonal blocks that provide the smallest weights at iteration step k, and $\mathcal{Q}_{k,|\mathcal{P}_k|}$ has the same number of elements as $\mathcal{P}_k$. Then the induction assumption (39) gives the following upper bound (note that the three sums below over three index sets involve off-diagonal matrix blocks at the *same* iteration step k):

$$\begin{aligned}
\sqrt{\sum_{(I,J)\in\mathcal{P}'_k}\|A_{IJ}\|_F^2} &\leq \sqrt{\sum_{(I,J)\in\mathcal{Q}_{k,|\mathcal{P}_k|}}\|A_{IJ}\|_F^2} \\
&\leq \sqrt{\sum_{(I,J)\in\mathcal{P}_k}\|A_{IJ}\|_F^2} \\
&\leq \frac{2+\sqrt{2}}{2}\sqrt{w-2}\cdot\frac{2\|\text{off}(A^{(0)})\|_F^2 - 2\|\text{off}(A^{(k)})\|_F^2}{\delta},
\end{aligned}$$

which is used in the third inequality of (46).

The last equality in (46) follows from

$$2\|\text{off}(A^{(k+1)})\|_F^2 = 2\|\text{off}(A^{(k)})\|_F^2 - 2\sum_{i=1}^{p}\|C_i\|_F^2.$$

Hence, (46) shows that the induction assumption also holds for $k+1$ and the proof is complete. ∎

Remark 3.4. Similarly to the parallel EVD algorithm in [21], the parallel SVD algorithm can have several periods W of the AQC that are bounded by Lemma 3.2. The same arguments as in [21, Remark 1] show that the most probable period will be the shortest one, i.e., $W = w - 1$. Moreover, the upper bound from Lemma 3.2 is less than the value $W = \frac{w(w-1)}{2}$ in the serial SVD algorithm. This means that the parallel SVD algorithm is *always more efficient* than the serial one.

Starting from (38) and using the same approach as in Corollary 2.3, we get the AQC of the off-diagonal Frobenius norm of the scaled iterated matrix in the parallel two-sided SVD algorithm with the GIPDO after W parallel iteration steps:

$$\|\text{off}(A_{\text{sc}}^{(W)})\|_F \leq \sqrt{12(w-2)}\,\kappa(A)\frac{\|\text{off}(A_{\text{sc}}^{(0)})\|_F^2}{\mu}. \tag{47}$$

The identification of the constant δ is the same as for the serial algorithm. Namely, under assumptions **A1**–**A3** (or **B1**–**B3**) for well-separated singular values (or clusters) made in section 2, the proofs of the lemmas regarding the stabilization of singular values may be repeated without any change for each processor i, $1 \leq i \leq p$, and each parallel iteration step r, $r \geq k$, where k is the number of the first parallel iteration step in which the assumption **A1** (or **B1**) is met. Hence, the value of δ, which ensures the AQC, can be set to $\delta = \sqrt{2}d_s/4$ (or $\delta = \sqrt{2}d_c/4$).

4. Numerical examples.

4.1. Serial algorithm. Two numerical examples are provided which illustrate the AQC of the serial two-sided block-Jacobi SVD algorithm for well-separated singular values (SVs) and for clusters. All experiments were computed using the machine precision $\varepsilon_M \approx 2.2 \times 10^{-16}$.

The software was written in FORTRAN 90 under the operating system Linux openSUSE, release 13.1, and using the libraries LAPACK, v.3.7.1, and BLAS, v.3.7.1. The software was compiled using `gfortran`, v.4.8.1, with the optimization `-O`. It was used on a personal computer with the processor Intel Core i7-870 @ 2.93 GHz, and 8 MB cache memory.

Each example contains a combination of well-separated SVs, multiple SVs, and clusters, so that only the threshold $d_c/8$ is used to define the interval of possible quadratic convergence (see assumption **B1**).

Test matrices were constructed in a unified way. In all experiments, the matrix size and the blocking factor were $n = 1024$ and $w = 16$, respectively, so that the blocks were of size $\ell = n/w = 64$, and one sweep consisted of $W = w(w-1)/2 = 120$ iteration steps.

The test matrix was obtained by choosing its SVs that were different in the two examples below. We always started with well-separated, simple SVs defined by

$$\sigma_i = 1 + 10^{-2}(n - i + 1), \quad 1 \leq i \leq n, \tag{48}$$

which form a descending sequence. These SVs are called *nominal* in the following. After (a) possible modification(s) (see examples below), the SVs were collected in the diagonal matrix D.

Then, the random test matrix A was created using the LAPACK procedure `DLAGGE` with the vector $(19, 1, 1958, 5)$ for the input "seed" field ISEED of the generator of normally distributed values with probability density function $N(0, 1)$.

Since the SVs were in the interval $[1.01, 11.24]$, the test matrix was well-conditioned with the 2-norm condition number $\kappa(A) = 10.93$. Additionally, an ill-conditioned test matrix was constructed by changing the smallest nominal SV $\sigma_{1024} = 1.01$ to 10^{-7}, so that the 2-norm condition number changed to $\kappa(A) \approx 1.12 \times 10^8$.

Afterwards, the test matrix A entered the Jacobi iteration process. In each iteration step k, the pair of off-diagonal blocks of $A^{(k)}$ with maximal weight was zeroed using the LAPACK procedure `DGESVJ`, which is an implementation of the scalar one-sided Jacobi SVD algorithm. This procedure sorts the local SVs in descending order, so that it already performs the LODE. Next, the corresponding two block rows and columns were orthogonally updated by matrix multiplication.

When the decreasing off-diagonal Frobenius norm crossed the threshold $d_c/8$ at some iteration k_1, $A^{(k_1)}$ was re-arranged by a suitable permutation of its rows and columns so that its diagonal was sorted into a descending sequence (see assumptions **A2** or **B2**). This global ordering was only performed once during the whole computation.

In both experiments, the following stopping criterion based on $\|\mathrm{off}(A_{\mathrm{sc}}^{(k)})\|_F$ was used:

$$\|\mathrm{off}(A_{\mathrm{sc}}^{(k)})\|_F \leq n\varepsilon_M \quad \text{or} \quad \left| \|\mathrm{off}(A_{\mathrm{sc}}^{(k)})\|_F - \|\mathrm{off}(A_{\mathrm{sc}}^{(k-1)})\|_F \right| \leq 5\varepsilon_M. \tag{49}$$

Its first part depends on the matrix order n and on the machine precision ε_M. The second part does not depend on n and reflects the difficulties in the algorithm's convergence near the very end of the iteration process due to the fact that the SVD procedure `DGESVJ` does not zero the off-diagonal blocks exactly. Because of the finite arithmetic and the stopping criterion inside the procedure `DGESVJ`, the off-diagonal blocks will have a nonzero Frobenius norm after their annihilation. Consequently, the value of $\|\mathrm{off}(A_{\mathrm{sc}}^{(k)})\|_F$ will not decrease under a certain threshold, and the computation has to be stopped when the decrease of $\|\mathrm{off}(A_{\mathrm{sc}}^{(k)})\|_F$ is of order $O(\varepsilon_M)$ in two subsequent iteration steps. This stopping criterion may not be optimal from the point of view of maximal attainable accuracy, but it is sufficient for analyzing the AQC in numerical examples below.

Each example below is represented by three figures. The first figure contains the threshold value $d_c/8$ and the convergence curve of $\|\mathrm{off}(A^{(k)})\|_F$, the second one the threshold value $d_c/8$ and the convergence curve of $\|\mathrm{off}(A_{\mathrm{sc}}^{(k)})\|_F$, and the third one contains the convergence curve of

$$\left\| \begin{pmatrix} U_{Y_k X_k}^{(k)} \\ V_{Y_k X_k}^{(k)} \end{pmatrix} \right\|_2 .$$

Moreover, the first figure in each experiment also contains a suitable multiple of the upper bound (UB1) given as the RHS term in (12), because its exact value was a large overestimate. The UB1 was computed for all iterations r, $r \geq k_1 + W$, where k_1 was the first iteration step in which the assumption **B1** was met with respect to the $\|\mathrm{off}(A^{(k)})\|_F$. Note that the iteration step k_1 is depicted as the vertical line in the second figure in each experiment, which enables to analyze when the $\|\mathrm{off}(A_{\mathrm{sc}}^{(k)})\|_F$ crossed the threshold as compared to $\|\mathrm{off}(A^{(k)})\|_F$. In all graphs, the horizontal axis is calibrated in serial sweeps.

Example 4.1. Using nominal SVs, a multiple SV with multiplicity 25 was created by taking σ_{13} and letting $\sigma_{14:37} = \sigma_{13}$. Note that the multiple SV lay in one diagonal block.

Moreover, one small cluster with 10 members was defined in the following way. First, σ_{513} was taken and let $\sigma_{514:522} = \sigma_{513}$. Using the LAPACK procedure `DLARNV`, a random, normally distributed vector x of length 9 was constructed with the ISEED field $(8, 8, 2018, 13)$. Then, the above defined multiple SV was perturbed:

$$\sigma_{513+i} \leftarrow \left(1 + 10^{-6}x(i)\right)\sigma_{513+i}, \quad 1 \leq i \leq 9.$$

Hence, the cluster was confined to one diagonal block and the resulting test matrix was well-conditioned.

The results are depicted in Figures 1, 2, and 3. Figure 1 shows that the AQC of $\|\mathrm{off}(A^{(k)})\|_F$ is present after crossing the threshold $d_c/8$ because the UB1 curve in the last sweep is parallel to the convergence curve. The AQC is also present for the scaled iterated matrix (see Figure 2) although the upper bound from (19) is not depicted. Note that the curve of $\|\mathrm{off}(A^{(k)}_{\mathrm{sc}})\|_F$ crossed the threshold $d_c/8$ *before* that of $\|\mathrm{off}(A^{(k)})\|_F$. However, the interval of AQC for $\|\mathrm{off}(A^{(k)}_{\mathrm{sc}})\|_F$ is the same as for $\|\mathrm{off}(A^{(k)})\|_F$, i.e., from k_1 onwards. It is also interesting to note that the computation stopped with $\|\mathrm{off}(A^{(k)}_{\mathrm{sc}})\|_F \approx 1.20 \times 10^{-13} < n\varepsilon_M$, and $\|\mathrm{off}(A^{(k)})\|_F$ decreased to the value of 8.96×10^{-13}. In Figure 3, the curve of

$$\left\| \begin{pmatrix} U^{(k)}_{Y_k X_k} \\ V^{(k)}_{Y_k X_k} \end{pmatrix} \right\|_2$$

followed the upper bound

$$\left\| \begin{pmatrix} U^{(k)}_{Y_k X_k} \\ V^{(k)}_{Y_k X_k} \end{pmatrix} \right\|_2 \leq \frac{\|C^{(k)}\|_F}{\delta} \leq \frac{2\sqrt{2}}{d_c} \|\mathrm{off}(A^{(k)})\|_F \tag{50}$$

in the interval of AQC.

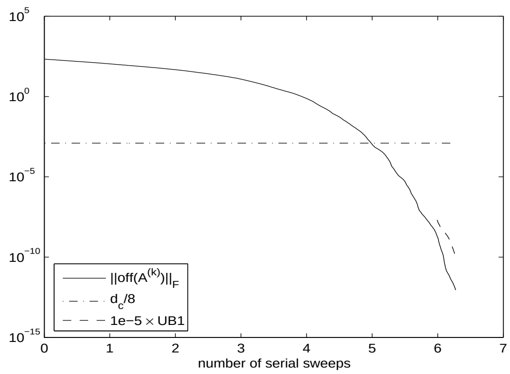

FIG. 1. *Singular value* σ_{13} *of multiplicity* 25 *and a small cluster* $\sigma_{513:522}$*;* $d_c = 10^{-2}$*,* $\kappa(A) =$ 10.93*. The off-diagonal Frobenius norm of the iterated matrix.*

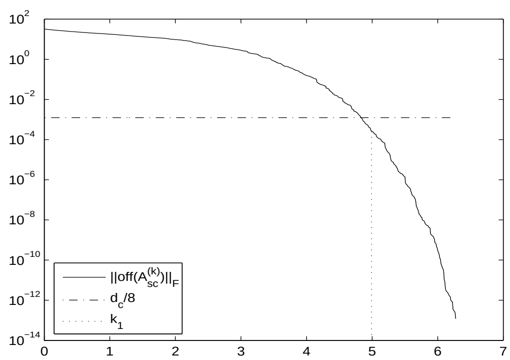

FIG. 2. *As in Figure* 1*, the off-diagonal Frobenius norm of the scaled iterated matrix.*

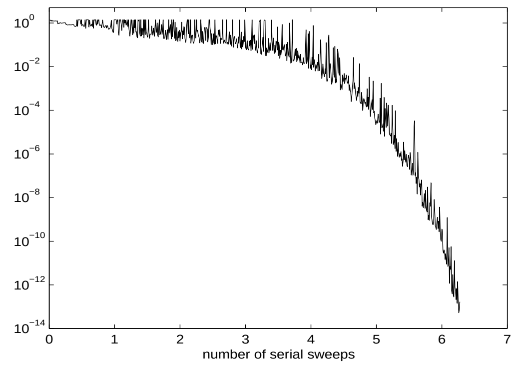

FIG. 3. *As in Figure* 1*, the graph of* $\|\binom{U^{(k)}_{Y_kX_k}}{V^{(k)}_{Y_kX_k}}\|_2$.

Example 4.2. For the ill-conditioned matrix, one multiple SV with multiplicity 65 was defined by taking σ_{580} and letting $\sigma_{581:644} = \sigma_{580}$, so that σ_{580} occupied 2 diagonal blocks. Moreover, two large clusters comprising the perturbed SVs $\sigma_{20:174}$ (155 members in 3 diagonal blocks) and $\sigma_{780:1004}$ (225 members in 4 diagonal blocks) were constructed using the procedure from Example 4.1. The ISEED field for the first cluster was $(19, 1, 1958, 31)$, whereas that for the second one was $(14, 4, 1958, 13)$.

Numerical results are summarized in Figures 4, 5, and 6. Since both assumptions **A3** and **B3** were violated, the AQC is disturbed and delayed (compare Figures 1 and 4). The algorithm needed about one serial sweep more for convergence as compared with Example 4.1. Moreover, the convergence curve of $\|\mathrm{off}(A^{(k)}_{\mathrm{sc}})\|_F$ is of the "staircase" form (see Figure 5), and the algorithm stopped with $\|\mathrm{off}(A^{(k)}_{\mathrm{sc}})\|_F \approx 1.57\times 10^{-11}$, about 2 orders of magnitude before reaching $n\varepsilon_M$. Similarly to Example 4.1, the curve of $\|\mathrm{off}(A^{(k)}_{\mathrm{sc}})\|_F$ crossed the threshold $d_c/8$ *before* that of $\|\mathrm{off}(A^{(k)})\|_F$. Perhaps the best evidence of the disturbed AQC comes from Figure 6, where the graph of

$$\left\|\begin{pmatrix} U^{(k)}_{Y_kX_k} \\ V^{(k)}_{Y_kX_k}\end{pmatrix}\right\|_2$$

did not follow the upper bound given by (50). In contrast, the maximal value of $\sqrt{2}$ can also be observed after crossing the threshold $d_c/8$.

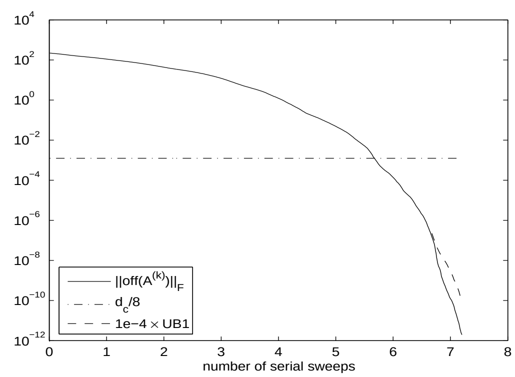

FIG. 4. *Singular value* σ_{580} *of multiplicity* 65 *and two large clusters* $\sigma_{20:174}$ *and* $\sigma_{780:1004}$*;* $d_c = 10^{-2}$*,* $\kappa(A) = 1.12 \times 10^8$*. The off-diagonal Frobenius norm of the iterated matrix.*

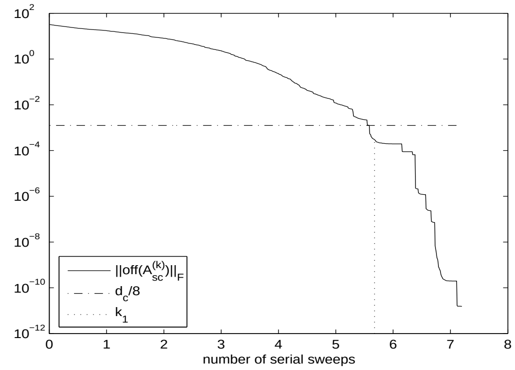

FIG. 5. *As in Figure* 4*, the off-diagonal Frobenius norm of the scaled iterated matrix.*

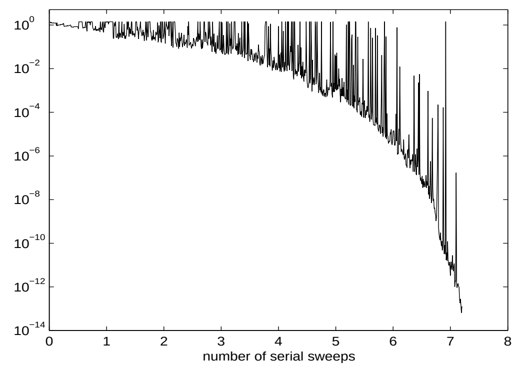

FIG. 6. *As in Figure* 1*, the graph of* $\|\begin{pmatrix} U^{(k)}_{Y_kX_k} \\ V^{(k)}_{Y_kX_k} \end{pmatrix}\|_2$.

A more detailed analysis w.r.t. the block indices of off-diagonal blocks chosen by dynamic ordering at each step gives some insight into the behavior of the "staircase" convergence curve in Figure 5 after crossing the threshold $d_c/8$, i.e., after the stabilization of diagonal elements. Recall that the multiple SV was located in diagonal blocks with block indices 10–11, the first cluster occupied diagonal blocks with block indices 1–3, and the second cluster lay inside diagonal blocks with block indices 13–16. Each serial iteration step means a zeroing of two off-diagonal blocks with maximal weight and the transfer of their Frobenius norm onto involved diagonal blocks. Since an orthogonal update preserves the Frobenius norm of the whole matrix, this means that the Frobenius norm of two diagonal blocks has increased. When the chosen block indices did not belong to a set of block indices defining a large cluster and/or multiplicity, corresponding diagonal elements were not changed, and the scaling of a large portion of off-diagonal elements used the same values as previously, hence the "stairs" arose. However, when at least one of two block indices chosen by dynamic ordering belonged to an index set of a large cluster and/or multiplicity, a significant number of involved diagonal elements could increase their value, and the subsequent scaling led to a substantial decrease of $\|\mathrm{off}(A^{(k)}_{\mathrm{sc}})\|_F$. Exactly this behavior can be observed in each decrease (approximately of one order of magnitude) of $\|\mathrm{off}(A^{(k)}_{\mathrm{sc}})\|_F$ in Figure 5. For example, at serial iteration step 767 (serial sweep 6.4), the dynamic ordering chose the block indices $(12, 15)$, whereby the second block index belonged to the set of block indices describing the second large cluster. After this step, the value of $\|\mathrm{off}(A_{\mathrm{sc}}^{(k)})\|_F$ suddenly decreased one order of magnitude from 10^{-5} to 10^{-6}, whereas it was of order 10^{-5} for about 28 steps before.

4.2. Parallel algorithm. Since only the decrease of the off-diagonal Frobenius norm during iterations is important for the verification of theory developed in section 3, the simulation software was written in FORTRAN 95 which serially computes p 2×2 block subproblems according to the GIPDO applied at the beginning of each parallel iteration step. Other details w.r.t. the implementation are identical to the serial case (see subsection 4.1).

More information about the parallel implementation of the two-sided block-Jacobi SVD algorithm with dynamic ordering can be found in [1]. This implementation uses the LAPACK procedure `DGESVD` for the SVD of each 2×2 block subproblem. To be in accordance with the theory developed in this paper, `DGESVD` should be replaced by `DGESVJ`, which executes the LODE.

In the two experiments below, $p = 32$ was the number of processors used. The matrix size and the blocking factor were $n = 4096$ and $w = 2p = 64$, respectively, so that $\ell = n/w = 64$ and one sweep consisted of $W = w - 1 = 63$ parallel iteration steps (this is the shortest period of the AQC according to Theorem 3.3).

Each of the two examples below started with nominal (i.e., well-separated, simple) SVs

$$\sigma_i = 0.1 + 10^{-2}(n - i), \quad 1 \leq i \leq n, \tag{51}$$

which were ordered decreasingly. Each example also contained different multiple SVs as well as clusters that were constructed similarly to those in subsection 4.1.

When the nominal SVs were used, the test matrix was well-conditioned with the condition number $\kappa(A) = 410.5$. Additionally, the ill-conditioned test matrix was constructed by setting $\sigma_n = 10^{-7}$, so that the 2-norm condition number changed to $\kappa(A) \approx 4.11 \times 10^8$.

Then, the random test matrix A was constructed using the LAPACK procedure `DLAGGE` with the vector $(19, 1, 1958, 5)$ for the field ISEED of the generator of normally distributed values with probability density function $N(0, 1)$.

Afterwards, the test matrix A entered the parallel block-Jacobi iteration process simulated by serial software. In the first parallel iteration step, all diagonal blocks of A were diagonalized and remained diagonal during the whole computation.

In each parallel iteration step k, $w = 2p$ off-diagonal blocks of $A^{(k)}$ that were chosen by the GIPDO were serially annihilated using the LAPACK procedure `DGESVJ`. Next, the iterated matrix $A^{(k)}$ was orthogonally updated by matrix-matrix multiplications.

When the decreasing off-diagonal Frobenius norm crossed the threshold $d_c/8$ at some iteration k_1, $A^{(k_1)}$ was re-arranged by a suitable permutation of its rows and columns so that its diagonal was sorted into a descending sequence (see assumptions **A2** or **B2**). It should be stressed that this global ordering was only performed once during the whole computation.

The graphical presentation of results has the same form as in subsection 4.1. In all figures, the horizontal axis is calibrated in parallel sweeps with $W = w - 1$, which is the shortest period given by Theorem 3.3.

Example 4.3. The well-conditioned matrix A was used starting with nominal SVs. Next, the multiplicity of σ_5 was changed to 15 by setting $\sigma_{6:19} = \sigma_5$, and the multiplicity of σ_{2180} was changed to 10 by setting $\sigma_{2181:2189} = \sigma_{2180}$. Note that each multiple SV lay in one diagonal block. Additionally, two small clusters $\sigma_{30:42}$ (13 members, ISEED = $(18, 1, 2017, 17)$) and $\sigma_{2850:2859}$ (10 members, ISEED = $(14, 4, 1958, 35)$) were created using the approach from Example 4.1 above. Consequently, each cluster was also confined to one diagonal block.

Numerical results are depicted in Figures 7, 8, and 9. They confirm the presence of the AQC. The convergence curves of $\|\mathrm{off}(A^{(k)})\|_F$ and $\|\mathrm{off}(A_{\mathrm{sc}}^{(k)})\|_F$ in Figure 7 and Figure 8, respectively, are very smooth. The curve of $\|\mathrm{off}(A_{\mathrm{sc}}^{(k)})\|_F$ crossed the threshold $d_c/8$ *before* that of $\|\mathrm{off}(A^{(k)})\|_F$, and the convergence stopped with $\|\mathrm{off}(A_{\mathrm{sc}}^{(k)})\|_F \approx 8.19 \times 10^{-13} < n\varepsilon_M$. In Figure 9, the graph of

$$\max_{1\le i\le p} \left\| \begin{pmatrix} U_{Y_{k,i}X_{k,i}}^{(k)} \\ V_{Y_{k,i}X_{k,i}}^{(k)} \end{pmatrix} \right\|_2$$

in the AQC region followed the upper bound

$$\max_{1\le i\le p} \left\| \begin{pmatrix} U_{Y_{k,i}X_{k,i}}^{(k)} \\ V_{Y_{k,i}X_{k,i}}^{(k)} \end{pmatrix} \right\|_2 \le \max_{1\le i\le p} \frac{\|C_{k,i}^{(k)}\|_F}{\delta} \le \frac{2\sqrt{2}}{d_c} \|\mathrm{off}(A^{(k)})\|_F. \tag{52}$$

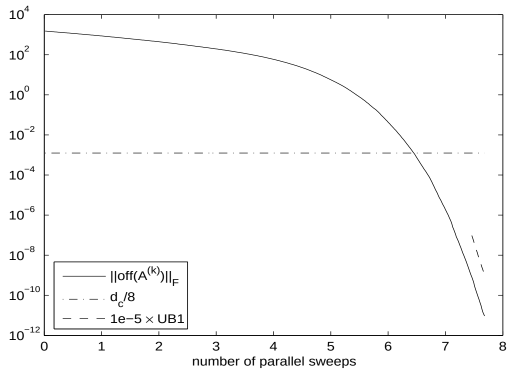

FIG. 7. *Two multiple singular values σ_5 and σ_{2180} of multiplicity* 15 *and* 10*, respectively, and two small clusters $\sigma_{30:42}$ and $\sigma_{2850:2859}$; $d_c = 10^{-2}$, $\kappa(A) = 411$. The off-diagonal Frobenius norm of the iterated matrix.*

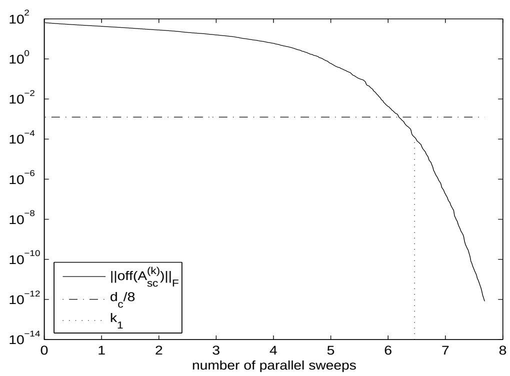

FIG. 8. *As in Figure* 7*, the off-diagonal Frobenius norm of the scaled iterated matrix.*

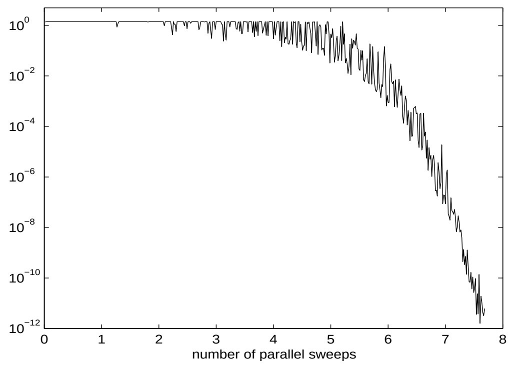

FIG. 9. *As in Figure 7, the graph of* $\max_{1\le i\le p} \|\begin{pmatrix} U^{(k)}_{Y_{k,i}} X_{k,i} \\ V^{(k)}_{Y_{k,i}} X_{k,i}\end{pmatrix}\|_2$.

Example 4.4. For the ill-conditioned matrix, the multiplicity of σ_1 was changed to 75 by setting $\sigma_{2:74} = \sigma_1$, so that the largest SV lay in 2 diagonal blocks. Additionally, two large clusters $\sigma_{125:179}$ (55 members in 2 diagonal blocks, ISEED $=$ $(29, 6, 2017, 15)$) and $\sigma_{3705:3960}$ (256 members in 5 diagonal blocks with the ISEED field $(14, 4, 1958, 37)$) were created using the approach from Example 4.1 above.

Figures 10, 11, and 12 contain numerical results. The presence of large multiplicities and large clusters of SVs causes disturbance and delay of the AQC. The algorithm needed about one parallel sweep more for convergence as compared to Example 4.3. The curve of $\|\mathrm{off}(A^{(k)}_{\mathrm{sc}})\|_F$ crossed the threshold $d_c/8$ *after* that of $\|\mathrm{off}(A^{(k)})\|_F$ (see Figure 11) and the convergence of the off-diagonal Frobenius norm of scaled iterated matrix was "staircase"-like again. The computation stopped with $\|\mathrm{off}(A^{(k)}_{\mathrm{sc}})\|_F \approx 3.39 \times 10^{-10}$, which is almost 3 orders of magnitude more than in Example 4.3. Similarly to the serial computation, the graph of

$$\max_{1\le i\le p} \left\| \begin{pmatrix} U^{(k)}_{Y_{k,i}} X_{k,i} \\ V^{(k)}_{Y_{k,i}} X_{k,i} \end{pmatrix} \right\|_2$$

did not follow the upper bound (52), and the large values observed after crossing the threshold $d_c/8$ are the best indication of the disturbed/delayed AQC.

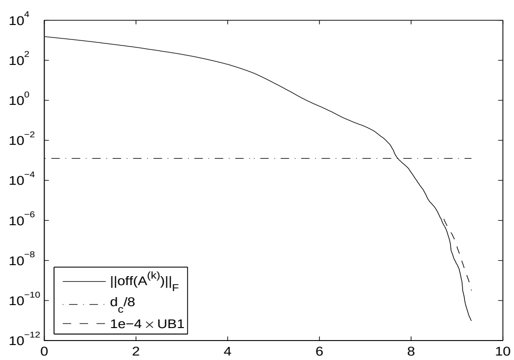

FIG. 10. *Two large clusters* $\sigma_{125:179}$ *and* $\sigma_{3705:3960}$*;* $d_c = 10^{-2}$*,* $\kappa(A) = 4.11 \times 10^8$*. The off-diagonal Frobenius norm of the iterated matrix.*

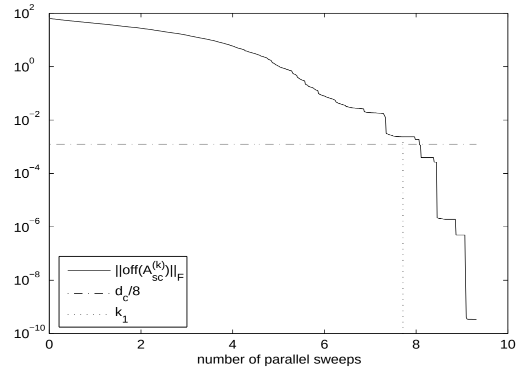

FIG. 11. *As in Figure* 10*, the off-diagonal Frobenius norm of the scaled iterated matrix.*

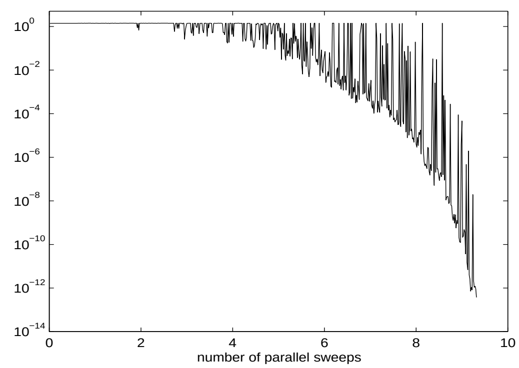

FIG. 12. *As in Figure* 10*, the graph of* $\max_{1\le i\le p} \|\begin{pmatrix} U^{(k)}_{Y_{k,i}} x_{k,i} \\ V^{(k)}_{Y_{k,i}} x_{k,i}\end{pmatrix}\|_2$.

The qualitative explanation of "stairs" in Figure 11 is similar to the analysis of Example 4.2. However, in contrast to serial experiments, *all* block rows and columns of the iterated matrix were updated in each parallel iteration step and $2p$ off-diagonal blocks with the largest weights were zeroed. Their Frobenius norms were transferred to corresponding diagonal blocks. When the block indices describing large clusters and/or multiplicities occupied *last* positions in the weight-ordered list provided by the GIPDO, then, after the stabilization of diagonal elements, the change of a large portion of diagonal elements was relatively small. Consequently, the subsequent scaling of a large portion of off-diagonal elements was computed using approximately the same values as before, and the "stairs" arose. However, when the block indices describing large clusters and/or multiplicities occupied *first* positions in the GIPDO's list, then a large portion of nearly equal diagonal elements could increase their values. The subsequent scaling led to the division of a large subset of off-diagonal elements by values larger than before, and a sudden decrease of $\|\mathrm{off}(A_{\mathrm{sc}}^{(k)})\|_F$ occurred. In our example, the multiple SV lay in diagonal blocks with block indices 1–2, the first cluster was described by block indices 2–3, and the second large cluster was located in diagonal blocks with block indices 58–62. It turns out that almost all significant decreases of $\|\mathrm{off}(A_{\mathrm{sc}}^{(k)})\|_F$ in Figure 11 took place when all block indices 58–62 (or at least three of them) were placed at the top of the ordered list of weights provided by the GIPDO. For example, at parallel iteration step 533 (parallel sweep 8.5), the top 5 places in the GIPDO's list were occupied by the following pairs of block indices: $(30, 60)$, $(56, 62)$, $(49, 58)$, $(53, 59)$, and $(44, 61)$. Note that the second block index in each pair was from the complete set of block indices describing the second large cluster. After this parallel iteration step, $\|\text{off}(A_{\text{sc}}^{(k)})\|_F$ suddenly decreased by two orders of magnitude from 10^{-4} to 10^{-6}. Similarly, at parallel iteration step 572 (parallel sweep 9.1), $\|\text{off}(A_{\text{sc}}^{(k)})\|_F$ suddenly decreased by two orders of magnitude from 10^{-7} to 10^{-9}. In this case, the four largest weights belonged to off-diagonal blocks described by index pairs $(16, 59)$, $(39, 58)$, $(21, 62)$, and $(23, 61)$, i.e., the second index in each pair was from the set of block indices describing the second large cluster again.

To summarize, in both serial and parallel two-sided block-Jacobi SVD algorithms, the AQC is present when the assumptions **A1**–**A3** and **B1**–**B3** are met. When the multiplicities and clusters of SVs are spread over more diagonal blocks, one can observe a disturbed and delayed AQC. Moreover, the convergence curve of the off-diagonal Frobenius norm of the scaled iterated matrix has a "staircase" form, which can cause an early termination of the iteration process due to the very small decrease of $\|\text{off}(A_{\text{sc}}^{(k)})\|_F$. This, in turn, can lead to SVs (or whole singular triplets) computed with less accuracy than required.

5. Conclusions. To our best knowledge, Theorems 2.2 and 3.3 together with the subsequent identification of constant δ provide first proofs of the global and asymptotic quadratic convergence of the serial and parallel two-sided block-Jacobi SVD algorithm with the GIPDO for a general distribution of singular values. The used serial and parallel dynamic ordering does not belong to the class of cyclic and quasi-cyclic orderings and is (usually) far more efficient in reducing the matrix off-diagonal Frobenius norm than any widely used cyclic ordering. Thus, in general, the cost of its computation at the beginning of each serial or parallel iteration step is acceptable due to the substantial decrease of the number of iterations needed for convergence as compared to any cyclic ordering.

It seems that the upper bounds (19) and (47) in the proof of the AQC of the off-diagonal Frobenius norm of the scaled iterated matrix for the serial and parallel algorithm, respectively, which contain the 2-norm condition number $\kappa(A)$, are large overestimates. The improvement of these upper bounds will be a topic of further research. However, the approach applied in the proof of Corollary 2.3 is general and can be used to prove the AQC of the off-diagonal Frobenius norm of the scaled iterated matrix in the case of serial [22] and parallel [21] block-Jacobi EVD algorithms for Hermitian matrices with dynamic ordering.

Acknowledgments. We would like to thank both anonymous referees for their valuable comments and suggestions that improved the paper's quality.

REFERENCES

[1] M. Bečka, G. Okša, and M. Vajteršic, *Dynamic ordering for a parallel block-Jacobi SVD algorithm*, Parallel Comp., 28 (2002), pp. 243–262, https://doi.org/10.1016/S0167-8191(01)00138-7.

[2] M. Bečka, G. Okša, and M. Vajteršic, *New dynamic orderings for the parallel one-sided block-Jacobi SVD algorithm*, Parallel Process. Lett., 25 (2015), 1550003–1–1550003–19, https://doi.org/10.1142/S0129626415500036.

[3] Z. Bujanović and Z. Drmač, *A contribution to the theory and practice of the block Kogbetliantz method for computing the SVD*, BIT Numer. Math., 52 (2012), pp. 827–849, https://doi.org/10.1007/s10543-012-0388-y.

[4] J. P. Charlier and P. V. Dooren, *On Kogbetliantz's SVD algorithm in the presence of clusters*, Linear Algebra Appl., 95 (1987), pp. 135–160, https://doi.org/10.1016/0024-3795(87)90031-0.

[5] C. Davis and W. Kahan, *The rotation of eigenvectors by a perturbation:* III, SIAM J. Numer. Anal., 7 (1970), pp. 1–46, https://doi.org/10.1137/0707001.

[6] J. W. Demmel and K. Veselić, *Jacobi's method is more accurate than QR*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 1204–1245, https://doi.org/10.1137/0613074.

[7] F. Dopico, P. Koev, and J. M. Molera, *Implicit standard Jacobi gives high relative accuracy*, Numer. Math., 113 (2009), pp. 519–553, https://doi.org/10.1007/s00211-009-0240-8.

[8] Z. Drmač and K. Veselić, *New fast and accurate Jacobi SVD algorithm:* I, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 1322–1342, https://doi.org/10.1137/050639193.

[9] Z. Drmač and K. Veselić, *New fast and accurate Jacobi SVD algorithm:* II, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 1343–1362, https://doi.org/10.1137/05063920X.

[10] G. Golub and C. F. V. Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 2013.

[11] V. Hari, *On sharp quadratic convergence bounds for the serial Jacobi methods*, Numer. Math., 60 (1991), pp. 375–406, https://doi.org/10.1007/BF01385728.

[12] V. Hari, *Convergence to diagonal form of block Jacobi-type methods*, Numer. Math., 129 (2015), pp. 449–481, https://doi.org/10.1007/s00211-014-0647-8.

[13] A. J. Hoffman and H. W. Wielandt, *The variation of the spectrum of a normal matrix*, Duke Math. J., 20 (1953), pp. 37–39, https://doi.org/10.1215/S0012-7094-53-02004-3.

[14] C. G. J. Jacobi, *Über ein leichtes Verfahren die in der Theorie der Säcularstörungen vorkommenden Gleichungen numerisch aufzulösen.*, J. Reine Angew. Math., 30 (1846), pp. 51–94, http://eudml.org/doc/147275.

[15] E. G. Kogbetliantz, *Solution of linear equations by diagonalization of coefficient matrix*, Quart. Appl. Math., 13 (1955), pp. 123–132, http://www.jstor.org/stable/43634196.

[16] S. Kudo, Y. Yamamoto, M. Bečka, and M.Vajteršic, *Performance analysis and optimization of the parallel one-sided block Jacobi SVD algorithm with dynamic ordering and variable blocking*, Concurrency Comp. Pract. Exper., 29 (2017), pp. 1–24, https://doi.org/10.1002/cpe.4059.

[17] S. Kudo, K. Yasuda, and Y. Yamamoto, *Performance of the parallel block Jacobi method with dynamic ordering for the symmetric eigenvalue problem*, JSIAM Letters, 10 (2018), pp. 41–44, https://doi.org/10.14495/jsiaml.10.41.

[18] F. T. Luk and H. Park, *A proof of convergence for two parallel Jacobi SVD algorithms*, IEEE Trans. Comp., 38 (1989), pp. 806–811, https://doi.org/10.1109/12.24289.

[19] W. F. Mascarenhas, *A note on Jacobi being more accurate than QR*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 215–218, https://doi.org/10.1137/S089547989222792X.

[20] R. Mathias, *Accurate eigensystem computations by Jacobi methods*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 977–1003, https://doi.org/10.1137/S089547989324820X.

[21] G. Okša, Y. Yamamoto, M. Bečka, and M. Vajteršic, *Asymptotic quadratic convergence of the parallel block-Jacobi EVD algorithm with dynamic ordering for Hermitian matrices*, BIT Numer. Math., 58 (2018), pp. 1099–1123, https://doi.org/10.1007/s10543-018-0711-3.

[22] G. Okša, Y. Yamamoto, and M. Vajteršic, *Asymptotic quadratic convergence of the serial block-Jacobi EVD algorithm for Hermitian matrices*, Numer. Math., 136 (2017), pp. 1071–1095, https://doi.org/10.1007/s00211-016-0863-5.

[23] C. C. Paige, *On the quadratic convergence of Kogbetliantz's algorithm for computing the singular value decomposition*, Linear Algebra Appl., 77 (1986), pp. 301–313, https://doi.org/10.1016/0024-3795(86)90173-4.

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