SIAM J. Matrix Anal. Appl.
Vol. 44, No. 1, pp. 382–407

THE JOINT BIDIAGONALIZATION METHOD FOR LARGE GSVD COMPUTATIONS IN FINITE PRECISION[*]

ZHONGXIAO JIA[†] AND HAIBO LI[†]

Abstract. The joint bidiagonalization (JBD) method has been used to compute some extreme generalized singular values and vectors of a large regular matrix pair $\{A, L\}$. We make a numerical analysis of the underlying JBD process and establish relationships between it and two mathematically equivalent Lanczos bidiagonalizations in finite precision. Based on the results of numerical analysis, we investigate the convergence of the approximate generalized singular values and vectors of $\{A, L\}$. The results show that, under some mild conditions, the semiorthogonality of Lanczos-type vectors suffices to deliver approximate generalized singular values with the same accuracy as the full orthogonality does, meaning that it is only necessary to seek for efficient semiorthogonalization strategies for the JBD process. We establish a sharp bound for the residual norm of an approximate generalized singular value and corresponding approximate right generalized singular vectors, which can reliably estimate the residual norm without explicitly computing the approximate right generalized singular vectors before the convergence occurs.

Key words. generalized singular value decomposition, joint bidiagonalization, Lanczos bidiagonalization, rounding error, orthogonality level, Ritz value, reorthogonalization, residual norm

MSC codes. 65F15, 65F20, 65F25, 15A18, 65F50, 65G50

DOI. 10.1137/22M1483608

1. Introduction.

The generalized singular value decomposition (GSVD) of two matrices A and L was introduced by Van Loan [36] and further developed by Paige and Saunders [26]. For $A \in \mathbb{R}^{m\times n}$ and $L \in \mathbb{R}^{p\times n}$, to ease the presentation, suppose that the matrix pair $\{A, L\}$ is regular, i.e., $\operatorname{rank}((A^T, L^T)^T) = n$ with $\operatorname{rank}(\cdot)$ being the rank of a matrix. Then the GSVD of $\{A, L\}$ is

$$A = P_A C_A X^{-1}, \quad L = P_L S_L X^{-1} \tag{1.1}$$

where P_A and P_L are orthogonal, X is nonsingular, and the diagonal matrices $C_A = \operatorname{diag}(c_1, c_2, \ldots, c_n) \in \mathbb{R}^{m\times n}$ and $S_L = \operatorname{diag}(s_1, s_2, \ldots, s_n) \in \mathbb{R}^{p\times n}$ with $0 \le c_i, s_i \le 1$ and $c_i^2 + s_i^2 = 1$. The case $\operatorname{rank}((A^T, L^T)^T) = r < n$ is called singular. The GSVD of the singular matrix pair $\{A, L\}$ is defined and analyzed in [26], where the matrix pair has $n - r$ arbitrary pairs $\{c, s\}$ as generalized singular values, called trivial ones. More details on the GSVD with $\operatorname{rank}((A^T, L^T)^T) = r < n$ can be found in [26]. The singular case is subtle and more complicated, as the generalized singular values are discontinuous with respect to matrix entries when the pair $\{A, L\}$ is singular. We refer the reader to [1, section 2.6.9] for a discussion on the discontinuity of the generalized eigenvalue problem of $\{A^T A, L^T L\}$, which is equivalent to the GSVD of $\{A, L\}$ [26]. In this paper, we restrict ourselves to the GSVD of the regular matrix pair $\{A, L\}$.

Specifically, label the $c_i < 1$ in nonincreasing order, and let these c_i and the corresponding s_i constitute the first diagonal blocks of C_A and S_L. Write $X = (x_1, \ldots, x_n)$,

[*] Received by the editors March 10, 2022; accepted for publication (in revised form) by M. W. Berry November 23, 2022; published electronically March 17, 2023.
https://doi.org/10.1137/22M1483608

Funding: The work of the authors was partially supported by National Natural Science Foundation of China grant 12171273.

[†] Department of Mathematical Sciences, Tsinghua University, 100084 Beijing, People‘s Republic of China (jiazx@tsinghua.edu.cn, lee12358@163.com).

$P_A = (p_1^A, \dots, p_m^A)$, and $P_L = (p_1^L, \dots, p_p^L)$. Then for $\{A, L\}$ regular, the number of such pairs $\{c_i, s_i\}$ is $q = n - q_1 - q_2$, where $q_1 = \dim(\mathcal{N}(A))$ and $q_2 = \dim(\mathcal{N}(B))$ with $\mathcal{N}(\cdot)$ the null space of a matrix. They correspond to the nonzero and finite generalized singular values c_i/s_i, $i = 1, 2, \dots, q$, called nontrivial ones. In terms of these notations, we can write the corresponding GSVD part in (1.1) in the vector form

$$\begin{cases} Ax_i = c_i p_i^A, \\ Lx_i = s_i p_i^L, \\ s_i A^T p_i^A = c_i L^T p_i^L, \end{cases} \qquad i = 1, \dots, q, \tag{1.2}$$

where the ith large generalized singular value of $\{A, L\}$ is c_i/s_i and the ith corresponding generalized singular vectors are x_i, p_i^A, and p_i^L, respectively. We call x_i the right generalized singular vector and p_i^A and p_i^L the left generalized singular vectors corresponding to c_i/s_i. We also use pair $\{c_i, s_i\}$ to denote a generalized singular value. We mention that each x_i satisfies the normalization $x_i^T(A^TA + L^TL)x_i = 1$. For more details on the GSVD of the regular matrix pair $\{A, B\}$, we refer the reader to [9].

In [37], Zha presents a joint bidiagonalization process (JBD) that jointly bidiagonalizes a large sparse or structured matrix pair $\{A, L\}$ to upper diagonal forms successively. He exploits the JBD process to compute a few extreme generalized singular values and vectors of $\{A, L\}$. Kilmer, Hansen, and Español [17] develop a variant of the JBD process that jointly reduces $\{A, L\}$ to lower and upper bidiagonal forms. Besides the computation of a few extreme GSVD components, this variant is used to solve large-scale linear discrete ill-posed problems with general-form regularization, where L is the regularization matrix [6, 7, 16, 17].

Let

$$\begin{pmatrix} A \\ L \end{pmatrix} = QR = \begin{pmatrix} Q_A \\ Q_L \end{pmatrix} R \tag{1.3}$$

be the compact QR factorization of the stacked matrix, where $Q \in \mathbb{R}^{(m+p)\times n}$ is column orthonormal and $R \in \mathbb{R}^{n\times n}$ is upper triangular and $Q_A \in \mathbb{R}^{m\times n}$ and $Q_L \in \mathbb{R}^{p\times n}$. Then the GSVD (1.1) of $\{A, L\}$ is related to the CS decomposition

$$Q_A = P_A C_A W^T, \quad Q_L = P_L S_L W^T \tag{1.4}$$

of $\{Q_A, Q_L\}$ [5, section 2.5.4], where W is orthogonal and $X = R^{-1}W$.

The k-step JBD process [17] of $\{A, L\}$ is mathematically equivalent to lower and upper Lanczos bidiagonalizations [27] of Q_A and Q_L,

$$Q_A V_k = U_{k+1} B_k, \quad Q_A^T U_{k+1} = V_k B_k^T + \alpha_{k+1} v_{k+1} e_{k+1}^T, \tag{1.5}$$

$$Q_L \widehat{V}_k = \widehat{U}_k \widehat{B}_k, \quad Q_L^T \widehat{U}_k = \widehat{V}_k \widehat{B}_k^T + \hat{\beta}_k \hat{v}_{k+1} e_k^T, \tag{1.6}$$

where e_k is the last column of the identity matrix I_k of order k,

$$B_k = \begin{pmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \beta_3 & \ddots & \\ & & \ddots & \alpha_k \\ & & & \beta_{k+1} \end{pmatrix} \in \mathbb{R}^{(k+1)\times k}, \; \widehat{B}_k = \begin{pmatrix} \hat{\alpha}_1 & \hat{\beta}_1 & & \\ & \hat{\alpha}_2 & \ddots & \\ & & \ddots & \hat{\beta}_{k-1} \\ & & & \hat{\alpha}_k \end{pmatrix} \in \mathbb{R}^{k\times k} \tag{1.7}$$

and

$$U_{k+1}=(u_1,\ldots,u_{k+1})\in\mathbb{R}^{m\times(k+1)},\quad V_k=(v_1,\ldots,v_k)\in\mathbb{R}^{n\times k} \tag{1.8}$$

and

$$\widehat{U}_k=(\hat{u}_1,\ldots,\hat{u}_k)\in\mathbb{R}^{p\times k},\quad \widehat{V}_k=(\hat{v}_1,\ldots,\hat{v}_k)\in\mathbb{R}^{n\times k} \tag{1.9}$$

with the starting vector $\hat{v}_1=v_1$ of the upper Lanczos bidiagonalization (1.6).

It is proved in [17, 37] that

$$\hat{v}_{i+1}=(-1)^i v_{i+1},\quad \hat{\alpha}_i\hat{\beta}_i=\alpha_{i+1}\beta_{i+1}. \tag{1.10}$$

It is well known from, e.g., [3], that the lower bidiagonal B_k is the Ritz–Galerkin projection of Q_A on the left subspace span(U_{k+1}) and the right subspace span(V_k), while the upper bidiagonal $\widehat{B}_k$ is the Ritz–Galerkin projection of Q_L on the left and right subspaces span$(\widehat{U}_k)$ and span(V_k), where span$(\cdot)$ denotes the subspace spanned by the columns of a matrix. Therefore, the extreme singular values of Q_A or Q_L can be approximated by those of B_k or $\widehat{B}_k$.

In finite precision arithmetic, we have numerically observed that the four sets of basis vectors computed by the JBD process, whose algorithmic implementation will be detailed in the next section, lose orthogonality gradually. This is a typical phenomenon in Lanczos-type algorithms, such as the symmetric Lanczos process [18] and the Lanczos bidiagonalization process [19]. The loss of orthogonality of Lanczos vectors leads to a delay of convergence of some extreme eigenvalues and the appearance of spurious computed Ritz values, i.e., ghost Ritz values, [21, 22, 23, 25]. To fix this deficiency, several reorthogonalization strategies have been proposed to maintain some level of orthogonality of the computed Lanczos vectors in order to avoid these consequences [29, 31, 32]. Particularly, Simon [31] proves that the semiorthogonality of Lanczos vectors suffices to guarantee that the computed Ritz values have the same accuracy as the full orthogonality does and avoid spurious computed Ritz values. These results have been extended by Larsen [19] to Lanczos bidiagonalization, based on which he proposes an efficient partial reorthogonalization strategy. Later, Simon and Zha in [33] proposed a one-sided reorthogonalization strategy on the right Lanczos vectors. Barlow [2] makes a backward error analysis of the one-sided reorthogonalization scheme and proves that Lanczos bidiagonalization of a matrix C produces Krylov subspaces generated by a nearby matrix $C+E$, where E is an error matrix depending on the orthogonality level of the computed right Lanczos vectors.

Denote the unit roundoff by ϵ. In finite precision, among many others, a central concern is whether or not the JBD process for computing U_{k+1}, V_k, and B_k is equivalent to the standard lower Lanczos bidiagonalization of Q_A with the rounding error $O(\epsilon)$ and whether or not the process for computing $\widehat{U}_k$, $\widehat{V}_k$, and $\widehat{B}_k$ is equivalent to the upper Lanczos bidiagonalization of Q_L with the rounding error $O(\epsilon)$. There has been yet no result on the finite precision behavior of the JBD process. In this paper, we will focus on it and, based on some underlying roundoff error models and results, make a numerical analysis of the JBD process. We will derive a number of properties of the JBD process. Our contributions mainly consist of the following three parts.

First, we will show that the equivalence of the JBD process and standard lower and upper Lanczos bidiagonalizations holds only conditionally in finite precision. That is, the finite precision forms resulting from the JBD process may be no longer the corresponding ones of standard Lanczos bidiagonalizations if there are no additional conditions. We will investigate what a role rounding errors play in the loss of this equivalence and in what way rounding errors are amplified.

Second, we will show that the orthogonality levels of U_{k+1}, $\widetilde{V}_k$, and $\widehat{U}_k$ are closely related and that those of $\widetilde{V}_k$ and V_k interact too. In particular, we derive an upper bound for the orthogonality level of $\widehat{U}_k$, which is shown to be controlled by not only the orthogonality levels of $\widetilde{V}_k$ and U_{k+1} but also a gradually growing quantity $\|\widehat{B}_k^{-1}\|$. The result indicates that the orthogonality level of $\widehat{U}_k$ is similar to those of U_{k+1} and $\widetilde{V}_k$, provided that $\widehat{B}_k$ is not ill conditioned. In the meantime, we prove that the orthogonality level of $\widetilde{V}_k$ is controlled by that of V_k. Therefore, when designing a reorthogonalization strategy for the JBD process, one only needs to reorthogonalize u_i and v_i. This way saves reorthogonalization cost considerably.

Third, we shall investigate the convergence of the JBD method for computing extreme GSVD components of $\{A, L\}$. We show that, under the assumptions that $\|B_k^{-1}\|$ and $\|\widehat{B}_k^{-1}\|$ are modest uniformly with k, the semiorthogonality of Lanczos-type vectors suffices to guarantee that the approximate generalized singular values have the same accuracy as the full orthogonality does. Here the semiorthogonality means that the absolute value of inner product of two unit length vectors is $O(\epsilon^{1/2})$, in contrast to the full orthogonality level $O(\epsilon)$. Therefore, the semiorthogonality of basis vectors suffices for the JBD method when computing generalized singular values accurately. In the meantime, we study the residual norm of an approximate generalized singular value and approximate right generalized singular vector, whose size is used to design a stopping tolerance for the JBD method. In finite precision, we derive an upper bound for the residual norm and show that this upper bound can replace the residual norm to design a reliable stopping criterion without explicitly computing approximate right generalized singular vectors before the convergence occurs. We only compute the approximate right generalized singular vectors by solving certain consistent least squares problems with the coefficient matrix $(A^T, L^T)^T$ at the convergence rather than doing so at each iteration.

The paper is organized as follows. In section 2, we describe the JBD process in exact arithmetic. In section 3, we make a numerical analysis of the JBD process in finite precision. We establish relationships between the JBD process and two lower and upper Lanczos bidiagonalizations and investigate interactions of orthogonality levels of the computed basis vectors. In section 4, we describe the JBD method for computing a number of extreme generalized singular values and vectors of $\{A, L\}$ and discuss the convergence and stopping criteria. In section 5, we report numerical experiments to confirm our results. Finally, we conclude the paper with some remarks and future work in section 6.

Throughout the paper, we denote by I_k the identity matrix of order k and by 0_k and $0_{k\times l}$ the k-dimensional zero vector and the $k \times l$ zero matrix, respectively. The transpose of a matrix C is denoted by C^T, and $\|\cdot\|$ is the 2-norm of a matrix.

2. The JBD process in exact arithmetic.

Algorithm 1 describes the JBD process in [17]. For A and L large, notice that the explicit computation of QR factorization (1.3) is generally impractical due to the excessive storage and/or computational cost. Thus, both Q and R are generally not available in practical computations. The JBD process avoids this difficulty in the following way: Let $\tilde{u}_i = (u_i^T, 0_p^T)^T$, $i = 1, 2, \ldots, k+1$. Algorithm 1 needs to compute $QQ^T\tilde{u}_i$ at steps 2 and 6. Since $QQ^T\tilde{u}_i$ is nothing but the orthogonal projection of $\tilde{u}_i$ onto the column space of $(A^T, L^T)^T$, we have $QQ^T\tilde{u}_i = (A^T, L^T)^T\tilde{x}_i$, where

$$\tilde{x}_i = \arg\min_{\tilde{x}\in\mathbb{R}^n} \left\| \begin{pmatrix} A \\ L \end{pmatrix} \tilde{x} - \tilde{u}_i \right\|. \tag{2.1}$$

This large-scale least squares problem can be solved by an iterative solver, e.g., the most commonly used LSQR algorithm [27].

Algorithm 1 The k-step JBD process.

```
Choose a nonzero starting vector $b \in \mathbb{R}^m$, and let $\beta_1 u_1 = b,\ \beta_1 = \|b\|$
$\alpha_1 \tilde{v}_1 = QQ^T \begin{pmatrix} u_1 \\ 0_p \end{pmatrix}$
$\hat{\alpha}_1 \hat{u}_1 = \tilde{v}_1(m+1:m+p)$
for $i = 1, 2, \ldots, k$, do
    $\beta_{i+1} u_{i+1} = \tilde{v}_i(1:m) - \alpha_i u_i$
    $\alpha_{i+1} \tilde{v}_{i+1} = QQ^T \begin{pmatrix} u_{i+1} \\ 0_p \end{pmatrix} - \beta_{i+1} \tilde{v}_i$
    $\hat{\beta}_i = (\alpha_{i+1} \beta_{i+1}) / \hat{\alpha}_i$
    $\hat{\alpha}_{i+1} \hat{u}_{i+1} = (-1)^i \tilde{v}_{i+1}(m+1:m+p) - \hat{\beta}_i \hat{u}_i$
end for
```

Suppose that $QQ^T \tilde{u}_i,\ i = 1, 2, \ldots, k+1$ are computed accurately. In exact arithmetic, the k-step JBD process produces the two bidiagonal matrices B_k, $\widehat{B}_k$ and three orthonormal matrices U_{k+1}, $\widehat{U}_k$ in (1.8)–(1.9) and

$$\widetilde{V}_k = (\tilde{v}_1, \ldots, \tilde{v}_k) \in \mathbb{R}^{(m+p) \times k}, \tag{2.2}$$

where $\tilde{v}_i = Qv_i$ with v_i the ith column of V_k in (1.8), i.e., $v_i = Q^T \tilde{v}_i$. The process can be written as

$$(I_m, 0_{m \times p}) \widetilde{V}_k = U_{k+1} B_k, \tag{2.3}$$

$$QQ^T \begin{pmatrix} U_{k+1} \\ 0_{p \times (k+1)} \end{pmatrix} = \widetilde{V}_k B_k^T + \alpha_{k+1} \tilde{v}_{k+1} e_{k+1}^T, \tag{2.4}$$

$$(0_{p \times m}, I_p) \widetilde{V}_k D_k = \widehat{U}_k \widehat{B}_k, \tag{2.5}$$

where $D_k = \operatorname{diag}(1, -1, \ldots, (-1)^{k-1}) \in \mathbb{R}^{k \times k}$ and e_{k+1} is the last column of I_{k+1}.

By the QR factorization (1.3), relations (2.3) and (2.5) are precisely

$$AZ_k = U_{k+1} B_k, \quad LZ_k = \widehat{U}_k \bar{B}_k, \tag{2.6}$$

where $Z_k = R^{-1} V_k = (z_1, \ldots, z_k)$ and $\bar{B}_k = \widehat{B}_k D_k$ and

$$B_k^T B_k + \bar{B}_k^T \bar{B}_k = I_k. \tag{2.7}$$

Therefore, the singular values of B_k and $\bar{B}_k$ can be determined by each other.

3. The JBD process in finite precision.

In the following, we do not consider the solution accuracy of the inner least squares problem (2.1) at each iteration and always assume that (2.1) is solved accurately.

Before proceeding, we define the orthogonality level of a set of vectors as follows.

DEFINITION 3.1. *For a rectangular matrix $W_k = (w_1, \dots, w_k) \in \mathbb{R}^{r\times k}$ with $\|w_j\|=1$, $j = 1, \dots, k$, we call $\xi_{ij} = |w_i^T w_j|$ the orthogonality level among w_i and w_j. The orthogonality level of $\{w_1, \dots, w_k\}$ or W_k is measured by one of*

$$\xi(W_k) = \max_{1\leq i\neq j\leq k} \xi_{ij}, \tag{3.1}$$

$$\eta(W_k) = \|I_k - W_k^T W_k\|. \tag{3.2}$$

Two measures are equivalent since $\xi(W_k) \leq \eta(W_k) \leq k\xi(W_k)$. It is also known from, e.g., [2], that

$$\|W_k\| \leq \sqrt{1+\eta(W_k)}. \tag{3.3}$$

We next state a set of basics on the behavior of the rounding errors occurring in the JBD process, which are adapted from the symmetric Lanczos process and Lanczos bidiagonalization. From now on, without confusion, we use the same notation as before to denote the computed ones in finite precision. In this case, relations (2.3)–(2.5) add rounding error terms (cf. [28, section 13.4]) and become

$$(I_m, 0_{m\times p})\widetilde{V}_k = U_{k+1}B_k + \widetilde{F}_k, \tag{3.4}$$

$$QQ^T \begin{pmatrix} U_{k+1} \\ 0_{p\times(k+1)} \end{pmatrix} = \widetilde{V}_k B_k^T + \alpha_{k+1}\tilde{v}_{k+1}e_{k+1}^T + \widetilde{G}_{k+1}, \tag{3.5}$$

$$(0_{p\times m}, I_p)\widetilde{V}_k D_k = \widehat{U}_k \widehat{B}_k + \bar{F}_k, \tag{3.6}$$

where the rounding error matrices $\widetilde{F}_k = (\tilde{f}_1, \dots, \tilde{f}_k)$, $\widetilde{G}_{k+1} = (\tilde{g}_1, \dots, \tilde{g}_{k+1})$, and $\bar{F}_k = (\bar{f}_1, \dots, \bar{f}_k)$ satisfy

$$\|\widetilde{F}_k\|, \|\widetilde{G}_{k+1}\|, \|\bar{F}_k\| = O(\epsilon). \tag{3.7}$$

Second, the following local orthogonality of u_i holds, similar to [24, 31]:

$$\beta_{i+1}|u_{i+1}^T u_i| = O(c_1(m,n)\epsilon), \tag{3.8}$$

where $c_1(m,n)$ is a modest constant depending on m and n.

3.1. Relationships between the JBD process and Lanczos bidiagonalizations in finite precision.

We first present the following results.

THEOREM 3.1. *Let $v_i = Q^T\tilde{v}_i$, $V_k = (v_1, \dots, v_k)$ and $\underline{B}_k = \begin{pmatrix} B_{k-1}^T \\ \alpha_k e_k^T \end{pmatrix} \in \mathbb{R}^{k\times k}$. Then*

$$\|\widetilde{V}_k - QV_k\| \leq \|\widetilde{G}_k \underline{B}_k^{-1}\| = O(\|\underline{B}_k^{-1}\|\epsilon) \tag{3.9}$$

with $\widetilde{G}_k$ defined in (3.5) *and*

$$\xi(\widetilde{V}_k) = \xi(V_k) + O(\|\underline{B}_k^{-1}\|\epsilon). \tag{3.10}$$

Proof. Write the matrix $C = (A^T, L^T)^T$. Then

$$QQ^T = CC^{\dagger},\ QQ^T C = C,\ QQ^T \begin{pmatrix} U_k \\ 0_{p\times k} \end{pmatrix} = CX_k,$$

where "†" denotes the Moore–Penrose inverse of a matrix and $X_k = C^{\dagger} \begin{pmatrix} U_k \\ 0_{p\times k} \end{pmatrix}$. From (3.5) and $V_k = Q^T \widetilde{V}_k$, we have $CX_k = \widetilde{V}_k \underline{B}_k + \widetilde{G}_k$, leading to $\widetilde{V}_k = CX_k \underline{B}_k^{-1} - \widetilde{G}_k \underline{B}_k^{-1}$. Therefore, we obtain

$$\begin{aligned} \widetilde{V}_k - QV_k &= \widetilde{V}_k - QQ^T \widetilde{V}_k = \widetilde{V}_k - QQ^T (CX_k \underline{B}_k^{-1} - \widetilde{G}_k \underline{B}_k^{-1}) \\ &= \widetilde{V}_k - CX_k \underline{B}_k^{-1} + QQ^T \widetilde{G}_k \underline{B}_k^{-1} = -\widetilde{G}_k \underline{B}_k^{-1} + QQ^T \widetilde{G}_k \underline{B}_k^{-1} \\ &= (QQ^T - I_{m+p}) \widetilde{G}_k \underline{B}_k^{-1}. \end{aligned} \tag{3.11}$$

Taking norms in both sides proves (3.9).

Since Q is orthonormal, it is easily justified that the orthogonality level $\xi(\widetilde{V}_k)$ is as accurate as $\xi(V_k)$ within $O(\|\underline{B}_k^{-1}\|\epsilon)$. This proves (3.10). □

Using (3.9), we can rewrite (3.4) as

$$(I_m, 0_{m\times p}) QV_k = U_{k+1} B_k + F_k, \tag{3.12}$$

where

$$F_k = \widetilde{F}_k - (I_m, 0_{m\times p})(\widetilde{V}_k - QV_k). \tag{3.13}$$

Then from (3.7) and (3.9), we have $\|F_k\| = O(\|\underline{B}_k^{-1}\|\epsilon)$. Premultiplying (3.5) by Q^T and exploiting $(I_m, 0_{m\times p})QV_k = Q_A V_k$ straightforwardly yields the lower Lanczos bidiagonalization of Q_A in finite precision resulting from the JBD process.

THEOREM 3.2. *Suppose that the inner least squares problem* (2.1) *is solved accurately. In finite precision, we have*

$$Q_A V_k = U_{k+1} B_k + F_k, \tag{3.14}$$

$$Q_A^T U_{k+1} = V_k B_k^T + \alpha_{k+1} v_{k+1} e_{k+1}^T + G_{k+1}, \tag{3.15}$$

where $G_{k+1} = Q^T \widetilde{G}_{k+1}$ *with* $\widetilde{G}_{k+1}$ *in* (3.5) *and* $\|F_k\| = O(\|\underline{B}_k^{-1}\|\epsilon)$, $\|G_{k+1}\| = O(\epsilon)$.

This theorem indicates that the error term F_k is amplified gradually as $\|\underline{B}_k^{-1}\|$ grows with k. Importantly, for Q_A strictly rectangular, i.e., $m > n$, the size of $\|\underline{B}_k^{-1}\|$ may be uncontrollably large, as shown below: From the second relation in (1.5), we obtain

$$U_k^T Q_A Q_A^T U_k = \underline{B}_k^T \underline{B}_k,$$

indicating that the eigenvalues of $\underline{B}_k^T \underline{B}_k$ are the Ritz values of the *singular* matrix $Q_A Q_A^T$ with respect to span(U_k) and lie between the largest and smallest eigenvalues of $Q_A Q_A^T$. Notice that span(U_k) is the Krylov subspace generated by

$$u_1, Q_A Q_A^T u_1, \ldots, (Q_A Q_A^T)^{k-1} u_1.$$

Then the smallest eigenvalue of $\underline{B}_k^T \underline{B}_k$ may approach the zero eigenvalue of $Q_A Q_A^T$ as k increases, so that $\|\underline{B}_k^{-1}\|$ may become uncontrollably large; on the other hand, for

Q_A flat or square, i.e., $m \leq n$, and having full row rank, however, such a phenomenon definitively cannot occur, and the smallest eigenvalue of $\underline{B}_k^T \underline{B}_k$ is bounded from below by the smallest positive one of $Q_A Q_A^T$. We refer the reader to [12] on a detailed analysis on $\underline{B}_k$ and its singular values. As a result, the JBD process for computing U_{k+1}, V_k, and B_k is conditionally equivalent to the lower Lanczos bidiagonalization of Q_A, whose rounding error term in the place of F_k is always $O(\|Q_A\|\epsilon) = O(\epsilon)$ in size.

Similarly, from (1.10) and the first relation in (1.6), we have

$$V_k^T Q_L^T Q_L V_k = \bar{B}_k^T \bar{B}_k.$$

Since $\mathrm{span}(V_k)$ is the Krylov subspace generated by $v_1, Q_L^T Q_L v_1, \ldots, (Q_L^T Q_L)^{k-1} v_1$, we can make a similar analysis to the above. Specifically, if Q_L is rectangular or square and of full column rank, then the smallest singular value of $\bar{B}_k$ converges to the smallest one of Q_L from above as k increases, and it is bounded from below by it, meaning that $\|\bar{B}_k^{-1}\|$ is controllable; if Q_L is flat, then the smallest singular value of $\bar{B}_k$ may approach zero as k increases, causing that $\|\bar{B}_k^{-1}\|$ may not be controlled and become large as k increases.

The above analysis and assertions suggest us to first check the orders of A and L and then perform the JBD process on either $\{A, L\}$ or $\{L, A\}$ when attempting to ensure that the resulting $\|\underline{B}_k^{-1}\|$ and $\|\bar{B}_k^{-1}\|$ are bounded whenever possible. As will be seen later, their boundedness is desirable for the JBD process and the JBD method for the GSVD computation in finite precision.

The following results are presented in [13, Theorem 3.1] and its proof, which will be used later.

THEOREM 3.3. *With the hypothesis of Theorem* 3.2*, in finite precision, we have*

$$B_k^T B_k + \bar{B}_k^T \bar{B}_k = I_k + E_k, \tag{3.16}$$

$$\alpha_{i+1}\beta_{i+1} = \hat{\alpha}_i \hat{\beta}_i + \gamma_i, \tag{3.17}$$

where E_k is symmetric tridiagonal with its nonzero elements being $O(c_3(m,n,p)\epsilon)$ and $c_3(m,n,p) = c_1(m,n) + c_2(p,n)$ in size and $|\gamma_i| \leq [1 + O(c_1(m,n)\epsilon)]\epsilon = O(\epsilon)$.

Remarkably, (3.16) holds independently of the orthogonality levels of $U_{k+1}, \widehat{U}_k$ and $\widetilde{V}_k$; it indicates that the squares of singular values of B_k and $\bar{B}_k$ can be determined by each other with the error $O(\epsilon)$.

For later use, we establish sharp upper bounds for $\|B_k\|$ and $\|\bar{B}_k\|$. From (3.4) and (3.6), at the ith step, we have

$$\begin{aligned}
&\tilde{v}_i(1:m) = \alpha_i u_i + \beta_{i+1} u_{i+1} + \tilde{f}_i, \\
&(-1)^{i-1} \tilde{v}_i(m+1:m+p) = \hat{\alpha}_i \hat{u}_i + \hat{\beta}_{i-1} \hat{u}_{i-1} + \bar{f}_i.
\end{aligned}$$

Thus, $\|\alpha_i u_i + \beta_{i+1} u_{i+1}\|^2 = \|\tilde{v}_i(1:m) - \tilde{f}_i\|^2$, which leads to

$$\begin{aligned}
\alpha_i^2 + \beta_{i+1}^2 &= \|\tilde{v}_i(1:m)\|^2 + \|\tilde{f}_i\|^2 - 2\tilde{f}_i^T \tilde{v}_i(1:m) - 2\alpha_i\beta_{i+1} u_{i+1}^T u_i \\
&\leq 1 + O(c_1(m,n)\epsilon),
\end{aligned} \tag{3.18}$$

where we have used (3.8). Similarly, we obtain

$$\hat{\alpha}_i^2 + \hat{\beta}_{i-1}^2 \leq 1 + O(c_2(p,n)\epsilon). \tag{3.19}$$

Therefore, we have proved the following lemma.

Lemma 3.1. *In finite precision, we have*[1]

$$\|B_k\| \leq \sqrt{2} \max_{1\leq i\leq k} (\alpha_i^2 + \beta_{i+1}^2)^{1/2} \leq \sqrt{2} + O(c_1(m,n)\epsilon), \tag{3.20}$$

$$\|\bar{B}_k\| \leq \sqrt{2} \max_{1\leq i\leq k} (\hat{\alpha}_i^2 + \hat{\beta}_{i-1}^2)^{1/2} \leq \sqrt{2} + O(c_2(p,n)\epsilon). \tag{3.21}$$

Theorem 3.4. *With the hypothesis of Theorem* 3.2, *in finite precision, we have*

$$Q_L\widehat{V}_k = \widehat{U}_k\widehat{B}_k + \widehat{F}_k, \tag{3.22}$$

$$Q_L^T\widehat{U}_k = \widehat{V}_k\widehat{B}_k^T + \hat{\beta}_k\hat{v}_{k+1}e_k^T + \widehat{G}_k, \tag{3.23}$$

where

$$\|\widehat{F}_k\| = O(\|\underline{B}_k^{-1}\|\epsilon), \tag{3.24}$$

$$\|\widehat{G}_k\| = O(c_4(m,n,p,k)\epsilon) \tag{3.25}$$

with

$$c_4(m,n,p,k) = (\|\widehat{B}_k^{-1}\| + 1)\|\underline{B}_k^{-1}\| + c_3(m,n,p)\|\widehat{B}_k^{-1}\|. \tag{3.26}$$

Proof. Recall the notation in (1.10), (2.5), and (1.6). Then exploiting Theorem 3.1, we can rewrite (3.6) as

$$(0_{p\times m}, I_p)\widehat{V}_k = \widehat{U}_k\widehat{B}_k + \widehat{F}_k,$$

where

$$\widehat{F}_k = \bar{F}_k - (0_{p\times m}, I_p)(\widetilde{V}_k - QV_k)D_k. \tag{3.27}$$

Therefore, (3.22) holds.

From (3.14) and (3.15), we obtain

$$\begin{aligned} Q_A^TQ_AV_k &= Q_A^TU_{k+1}B_k + Q_A^TF_k \\ &= (V_kB_k^T + \alpha_{k+1}v_{k+1}e_{k+1}^T + G_{k+1})B_k + Q_A^TF_k \\ &= V_kB_k^TB_k + \alpha_{k+1}\beta_{k+1}v_{k+1}e_k^T + G_{k+1}B_k + Q_A^TF_k. \end{aligned} \tag{3.28}$$

Premultiplying and postmultiplying (3.22) by Q_L^T and D_k gives

$$Q_L^TQ_LV_k = (Q_L^T\widehat{U}_k\widehat{B}_k + Q_L^T\widehat{F}_k)D_k.$$

Summing the above two equalities and exploiting (3.16) yields

$$\begin{aligned} V_k &= (Q_A^TQ_A + Q_L^TQ_L)V_k \\ &= V_kB_k^TB_k + Q_L^T\widehat{U}_k\widehat{B}_kD_k + \alpha_{k+1}\beta_{k+1}v_{k+1}e_k^T + (G_{k+1}B_k + Q_A^TF_k + Q_L^T\widehat{F}_kD_k) \\ &= V_k(I_k - D_k\widehat{B}_k^T\widehat{B}_kD_k + E_k) + Q_L^T\widehat{U}_k\widehat{B}_kD_k + \alpha_{k+1}\beta_{k+1}v_{k+1}e_k^T \\ &\quad + (G_{k+1}B_k + Q_A^TF_k + Q_L^T\widehat{F}_kD_k). \end{aligned}$$

[1] Here we use the result of an exercise from [8, Chapter 6, Problem 6.14], which gives an upper bound for the p-norm of a row/column sparse matrix.

Postmultiplying the last relation by D_k and exploiting (3.17), $\widehat{V}_k = V_k D_k$, and $D_k^2 = I_k$, we obtain

$$\begin{aligned}\widehat{V}_k \widehat{B}_k^T \widehat{B}_k &= Q_L^T \widehat{U}_k \widehat{B}_k + \alpha_{k+1}\beta_{k+1} v_{k+1} e_k^T D_k + (G_{k+1}B_k + Q_A^T F_k + Q_L^T \widehat{F}_k D_k + V_k E_k) D_k \\ &= Q_L^T \widehat{U}_k \widehat{B}_k - (\hat{\alpha}_k \hat{\beta}_k + \gamma_k)\hat{v}_{k+1} e_k^T + (G_{k+1}B_k + Q_A^T F_k + Q_L^T \widehat{F}_k D_k + V_k E_k) D_k,\end{aligned}$$

which shows that

$$Q_L^T \widehat{U}_k = \widehat{V}_k \widehat{B}_k^T + \hat{\beta}_k \hat{v}_{k+1} e_k^T - E_1 - E_2, \tag{3.29}$$

where

$$E_1 = [(G_{k+1}B_k + V_k E_k)D_k - \gamma_k \hat{v}_{k+1} e_k^T]\widehat{B}_k^{-1}, \quad E_2 = (Q_A^T F_k D_k + Q_L^T \widehat{F}_k)\widehat{B}_k^{-1}.$$

Notice that γ_k in (3.17) satisfies $|\gamma_k| = O(\epsilon)$ and that the elements of E_k in (3.16) are $O(c_3(m,n,p))$ in size. Making use of (3.3) and the upper bounds for $\|B_k\|$ in (3.20), we have

$$\|E_1\| = O(\bar{c}_1(m,n,p,k)\epsilon)$$

with $\bar{c}_1(m,n,p,k) = (\sqrt{2} + c_3(m,n,p))\|\widehat{B}_k^{-1}\|$. Using the expressions of F_k, $\widehat{F}_k$, and $\widetilde{V}_k - QV_k$ in (3.13), (3.27), and (3.11), respectively, we obtain

$$\begin{aligned}Q_A^T F_k D_k + Q_L^T \widehat{F}_k &= \begin{pmatrix} Q_A^T & Q_L^T \end{pmatrix} \begin{pmatrix} F_k D_k \\ \widehat{F}_k \end{pmatrix} \\ &= Q^T \Big[\begin{pmatrix} \widetilde{F}_k D_k \\ \bar{F}_k \end{pmatrix} - \begin{pmatrix} I_m & 0_{m\times p} \\ 0_{p\times m} & I_p \end{pmatrix} (\widetilde{V}_k - QV_k) D_k \Big] \\ &= Q^T \Big[\begin{pmatrix} \widetilde{F}_k D_k \\ \bar{F}_k \end{pmatrix} + (I_{m+p} - QQ^T)\widetilde{G}_k \underline{B}_k^{-1} D_k \Big].\end{aligned}$$

Since

$$\|\underline{B}_k^{-1} D_k \widehat{B}_k^{-1}\| = \|(\bar{B}_k \underline{B}_k)^{-1}\| \le \|\bar{B}_k^{-1}\| \|\underline{B}_k^{-1}\|$$

and $\|\bar{B}_k^{-1}\| = \|\widehat{B}_k^{-1}\|$, it holds that

$$\|E_2\| = O(\bar{c}_2(k)\epsilon)$$

with $\bar{c}_2(k) = (\|\widehat{B}_k^{-1}\| + 1)\|\underline{B}_k^{-1}\|$. Letting $\widehat{G}_k = -E_1 - E_2$ in (3.29) and substituting the estimates on $\|E_1\|$ and $\|E_2\|$ leads to the desired result. ∎

Notice that $\|\widehat{B}_k^{-1}\| \ge 1$. We see that $\|\widehat{G}_k\| = O(\|\widehat{B}_k^{-1}\| \|\underline{B}_k^{-1}\|\epsilon)$ and, particularly, $\|\widehat{G}_k\| = O((\|\widehat{B}_k^{-1}\| + \|\underline{B}_k^{-1}\|)\epsilon)$, provided that the size of either one is not large. The theorem indicates that the JBD process for computing $\widehat{U}_k$, $\widehat{V}_k$, and $\widehat{B}_k$ is conditionally equivalent to the standard upper Lanczos bidiagonalization of Q_L, whose rounding error is always $O(\|Q_L\|\epsilon) = O(\epsilon)$ in finite precision.

3.2. Loss of orthogonality of the basis vectors. Once the orthogonality is lost at one iteration, the errors will propagate to later steps, which leads to the more loss of orthogonality of subsequently computed basis vectors. As it will turn out, the orthogonality levels of U_{k+1}, $\widetilde{V}_k$, and $\widehat{U}_k$ are closely related. Below we prove how the orthogonality level of $\widehat{U}_k$ is affected by those of U_{k+1} and $\widetilde{V}_k$.

THEOREM 3.5. *With the hypothesis of Theorem* 3.2*, in finite precision, we have*

$$\eta(\widehat{U}_k) \le \|\widehat{B}_k^{-1}\|^2 \big[\eta(\widetilde{V}_k) + 2\eta(U_{k+1}) + O(c_3(m,n,p)\epsilon)\big]. \tag{3.30}$$

Proof. From (3.4) and (3.6), we have

$$\widetilde{V}_k = \begin{pmatrix} U_{k+1}B_k \\ \widehat{U}_k \bar{B}_k \end{pmatrix} + \begin{pmatrix} \widetilde{F}_k \\ \bar{F}_k D_k \end{pmatrix},$$

which shows that

$$\widetilde{V}_k^T \widetilde{V}_k = B_k^T U_{k+1}^T U_{k+1} B_k + \bar{B}_k^T \widehat{U}_k^T \widehat{U}_k \bar{B}_k + E_3, \tag{3.31}$$

where

$$E_3 = B_k^T U_{k+1}^T \widetilde{F}_k + \bar{B}_k^T \widehat{U}_k^T \bar{F}_k D_k + \widetilde{F}_k^T U_{k+1} B_k + D_k \bar{F}_k^T \widehat{U}_k \bar{B}_k + \widetilde{F}_k^T \widetilde{F}_k + D_k \bar{F}_k^T \bar{F}_k D_k.$$

From (3.16) and (3.31), we obtain

$$I_k - \widetilde{V}_k^T \widetilde{V}_k = B_k^T (I_{k+1} - U_{k+1}^T U_{k+1}) B_k + \bar{B}_k^T (I_k - \widehat{U}_k^T \widehat{U}_k) \bar{B}_k - E_k - E_3,$$

which yields

$$I_k - \widehat{U}_k^T \widehat{U}_k = \bar{B}_k^{-T} \big[(I_k - \widetilde{V}_k^T \widetilde{V}_k) - B_k^T (I_{k+1} - U_{k+1}^T U_{k+1}) B_k + E_k + E_3\big] \bar{B}_k^{-1}. \tag{3.32}$$

By (3.3), we have $\|U_{k+1}\| \le (1+\eta(U_{k+1}))^{1/2}$ and $\|\widehat{U}_k\| \le (1+\eta(\widehat{U}_k))^{1/2}$. Using the bounds for $\|B_k\|$ and $\|\widehat{B}_k\|$ in (3.20) and (3.21), respectively, by a simple calculation, we obtain

$$\|E_3\| = O(\epsilon).$$

Using the bound for $\|B_k\|$, we have

$$\begin{aligned} \|B_k^T (I_{k+1} - U_{k+1}^T U_{k+1}) B_k\| &\le \|B_k\|^2 \|I_{k+1} - U_{k+1}^T U_{k+1}\| \\ &\le 2\|I_{k+1} - U_{k+1}^T U_{k+1}\| + O(c_1(m,n)\epsilon). \end{aligned}$$

Taking norms in (3.32) proves the desired result. □

This theorem indicates that provided $\|\widehat{B}_k^{-1}\| = \|\bar{B}_k^{-1}\|$ is not large, the orthogonality of $\widehat{U}_k$ is as good as those of U_{k+1} and $\widetilde{V}_k$. Therefore, it is only necessary to perform some sort of reorthogonalization strategies to maintain desired orthogonality levels of U_{k+1} and $\widetilde{V}_k$.

The following result shows that the orthogonality levels of the long $\widetilde{V}_k \in \mathbb{R}^{(m+p)\times k}$ and the short $V_k \in \mathbb{R}^{p\times k}$ are the same within $O(\epsilon)$ under some mild condition.

THEOREM 3.6. *With definition* (3.2)*, it holds that*

$$|\eta(\widetilde{V}_k) - \eta(V_k)| = O(\|\underline{B}_k^{-1}\|^2 \epsilon^2). \tag{3.33}$$

Proof. Since $(I_{m+p} - QQ^T)^2 = I_{m+p} - QQ^T$, from (3.11), we have

$$\begin{aligned} \widetilde{V}_k^T (I_{m+p} - QQ^T) \widetilde{V}_k &= \widetilde{V}_k^T (I_{m+p} - QQ^T)(I_{m+p} - QQ^T) \widetilde{V}_k \\ &= (\widetilde{V}_k - QV_k)^T (I_{m+p} - QQ^T)(I_{m+p} - QQ^T)(\widetilde{V}_k - QV_k) \\ &= \underline{B}_k^{-T} \widetilde{G}_k^T (I_{m+p} - QQ^T) \widetilde{G}_k \underline{B}_k^{-1}. \end{aligned}$$

Thus,

$$\begin{aligned} I_k - V_k^T V_k &= I_k - \widetilde{V}_k^T QQ^T \widetilde{V}_k = I_k - \widetilde{V}_k^T \widetilde{V}_k + \widetilde{V}_k^T (I_{m+p} - QQ^T)\widetilde{V}_k \\ &= (I_k - \widetilde{V}_k^T \widetilde{V}_k) + \underline{B}_k^{-T} \widetilde{G}_k^T (I_{m+p} - QQ^T) \widetilde{G}_k \underline{B}_k^{-1}. \end{aligned}$$

Therefore, (3.33) holds. □

This theorem shows that provided $\|\underline{B}_k^{-1}\|$ is not too large, say, no more than $\epsilon^{-1/2}$, we have $|\eta(\widetilde{V}_k) - \eta(V_k)| = O(\epsilon)$. Therefore, it is only necessary to maintain the same orthogonality level of V_k in order to make $\widetilde{V}_k$ achieve a desired orthogonality level. From (3.30), we only need to maintain desired orthogonality levels of U_{k+1} and V_k in order to make $\widehat{U}_k$ have a desired orthogonality level. In summary, for the four sets of basis vectors generated by the JBD process, it generally suffices to reorthogonalize U_{k+1} and V_k.

4. The JBD method for the GSVD computation.

In this section, we describe the JBD method for computing some extreme GSVD components of $\{A, L\}$ and make an analysis on the convergence of the approximate generalized singular values in finite precision by exploiting the previous results.

4.1. The JBD method.

For ease of presentation, we do not take into account rounding errors when computing the GSVD of $\{B_k, \bar{B}_k\}$ or the SVD of B_k or $\bar{B}_k$; that is, we assume that the compact SVD of B_k is computed accurately,

$$B_k = P_k \Theta_k W_k^T, \quad \Theta_k = \operatorname{diag}(c_1^{(k)}, \dots, c_k^{(k)}), \quad 1 \geq c_1^{(k)} > \dots > c_k^{(k)} \geq 0, \tag{4.1}$$

where $P_k = (p_1^{(k)}, \dots, p_k^{(k)}) \in \mathbb{R}^{(k+1)\times k}$ is orthonormal and $W_k = (w_1^{(k)}, \dots, w_k^{(k)}) \in \mathbb{R}^{k\times k}$ is orthogonal. The SVD (4.1) can be obtained by a standard SVD algorithm since B_k is small. Then we have k approximate generalized singular values $\{c_i^{(k)}, (1 - (c_i^{(k)})^2)^{1/2}\}$, $i = 1, 2, \dots, k$, of $\{A, L\}$, and the approximate right generalized singular vectors are the $x_i^{(k)} = R^{-1} V_k w_i^{(k)}$, and the approximations to the left generalized singular vectors p_i^A are $y_i^{(k)} = U_{k+1} p_i^{(k)}$. Among these k approximations, we pick up a few of the largest and/or smallest ones as approximations to the largest and/or smallest c_i/s_i and the corresponding x_i and p_i^A.

If we also want to compute an approximation of the left generalized singular vector p_i^L, we need to compute the SVD of $\bar{B}_k$. From (2.7), it is known that $(B_k^T, \bar{B}_k^T)^T$ is column orthonormal. Therefore, the CS decomposition of the pair $\{B_k, \bar{B}_k\}$ is its GSVD, and the right singular vectors of B_k and $\bar{B}_k$ are identical. As a result, we can assume that the SVD of $\bar{B}_k$ is

$$\bar{B}_k = \bar{P}_k \Psi_k W_k^T, \quad \Psi_k = \operatorname{diag}(\bar{s}_1^{(k)}, \dots, \bar{s}_k^{(k)}), \quad 0 \leq \bar{s}_1^{(k)} < \dots < \bar{s}_k^{(k)} \leq 1, \tag{4.2}$$

where $\bar{P}_k = (\bar{p}_1^{(k)}, \dots, \bar{p}_k^{(k)}) \in \mathbb{R}^{k\times k}$ and $W_k = (w_1^{(k)}, \dots, w_k^{(k)}) \in \mathbb{R}^{k\times k}$ are orthogonal. Then $z_i^{(k)} = \widehat{U}_k \bar{p}_i^{(k)}$, $i = 1, 2, \dots, k$ are approximate left generalized singular vectors for L. The approximate generalized singular values and the corresponding approximate right generalized singular vectors are $\{(1 - (\bar{s}_i^{(k)})^2)^{1/2}, \bar{s}_i^{(k)}\}$ and $R^{-1} V_k w_i^{(k)}$, respectively.

Alternatively, we compute the GSVD of the pair $\{B_k, \bar{B}_k\}$,

$$B_k = P_k C_k W_k^T, \ \bar{B}_k = \bar{P}_k S_k W_k^T, \tag{4.3}$$

where $C_k = \operatorname{diag}(c_1^{(k)}, \dots, c_k^{(k)})$ and $S_k = \operatorname{diag}(s_1^{(k)}, \dots, s_k^{(k)})$ and P_k and $\bar{P}_k$ are as those defined in (4.1) and (4.2). The approximate generalized singular values are $\{c_i^{(k)}, s_i^{(k)}\}$

or $c_i^{(k)}/s_i^{(k)}$; the corresponding left approximate generalized singular vectors for A and L are $U_{k+1}p_i^{(k)}$ and $\widehat{U}_k\bar{p}_i^{(k)}$, respectively; and the right approximate generalized singular vectors are $x_i^{(k)} = Z_k w_i^{(k)} = R^{-1}V_k w_i^{(k)}$.

For the computation of $x_i^{(k)}$, it is shown in [37] that the explicit inversion R^{-1} can be avoided by noticing that

$$\begin{pmatrix} A \\ L \end{pmatrix} x_i^{(k)} = QRR^{-1}V_k w_i^{(k)} = \widetilde{V}_k w_i^{(k)}. \tag{4.4}$$

Then, solving the corresponding consistent linear system by an iterative solver, e.g., the LSQR algorithm, we obtain $x_i^{(k)}$.

4.2. Convergence, accuracy, and reorthogonalization. We investigate the convergence of the computed generalized singular values of the JBD method. We only focus on the approach of using the SVD of B_k. The same analysis and results hold for the approaches of using the SVD of $\bar{B}_k$ and the GSVD of $\{B_k, \bar{B}_k\}$. Since $c_i^2 + s_i^2 = 1$, in order to compute the generalized singular value $\{c_i, s_i\}$, we only need to compute c_i, which is a singular value of Q_A. Note that $c_i^{(k)}$, the singular value of B_k, is a computed Ritz value of Q_A since the k-step JBD process for computing B_k is Lanczos bidiagonalization applied to Q_A with the rounding error $O(\|\underline{B}_k^{-1}\|\epsilon)$; see Theorem 3.2. In exact arithmetic, the eigenvalues of $B_k^T B_k$ are the Ritz values of $Q_A^T Q_A$ with respect to the Krylov subspace generated by $Q_A^T b, (Q_A^T Q_A)Q_A^T b, \ldots, (Q_A^T Q_A)^{k-1}Q_A^T b$, and the k-step lower bidiagonalization of Q_A is equivalent to the symmetric Lanczos process applied to $Q_A^T Q_A$ and the starting vector $Q_A^T b/\|Q_A^T b\|$. Therefore, the convergence theory of the symmetric Lanczos method applies (cf. [28, 30]), and the singular values of B_k generally favor some of the largest and smallest ones of Q_A. More convergence results and details have been given in [10, 11, 16]. In finite precision, typical convergence features of the symmetric Lanczos method carry over to our case.

In finite precision, because of the loss of orthogonality of basis vectors, some of the singular values of B_k could be numerically multiple as the iteration number k increases, which may produce ghost approximations to some generalized singular values of $\{A, L\}$. A direct consequence is that a simple or genuine multiple generalized singular value of $\{A, L\}$ could be approximated by numerically multiple computed Ritz values, which could lead to a convergence delay of computed Ritz values. These phenomena can be avoided by using some types of reorthogonalization strategies, such as full reorthogonalization or the more efficient one-sided reorthogonalization [33].

By Theorem 3.2, the JBD process for computing B_k is the lower Lanczos bidiagonalization of Q_A with the rounding error $O(\|\underline{B}_k^{-1}\|\epsilon)$, which is comparable to $O(\epsilon)$ whenever the size of $\|\underline{B}_k^{-1}\|$ is modest. If the JBD process is implemented with one-sided reorthogonalization of v_i such that the orthogonality level of V_k achieves $O(\epsilon)$, exploiting the backward error results on the Lanczos bidiagonalization with one-sided reorthogonalization [2, Theorem 5.2 and Corollary 5.1], we can deduce that the computed B_k is the exact one generated by the Lanczos bidiagonalization of a nearby matrix $Q_A + E_k$ with $\|E_k\| = O(\|\underline{B}_k^{-1}\|\epsilon)$. Therefore, by the perturbation theory of the singular values [5, Corollary 8.6.2], the extreme singular values of Q_A can be computed with the ultimate accuracy $O(\|\underline{B}_k^{-1}\|\epsilon)$.

The following theorem is due to the authors of [13], which relaxes the requirement on the full orthogonality of basis vectors.

THEOREM 4.1. *Assume that the compact QR factorizations of U_{k+1} and V_k are $U_{k+1} = M_{k+1}R_{k+1}$ and $V_k = N_k S_k$, where the diagonals of R_{k+1} and S_k are positive,* *and let $\delta = O(\|\underline{B}_k^{-1}\|\epsilon)$. If U_{k+1} and V_k satisfy the semiorthogonality*

$$\xi(U_{k+1}),\ \xi(V_k) \leqslant \sqrt{\delta/(2k+1)}, \tag{4.5}$$

then

$$M_{k+1}^T Q_A N_k = B_k + \widetilde{E}_k, \tag{4.6}$$

where the elements of $\widetilde{E}_k$ are $O(\delta) = O(\|\underline{B}_k^{-1}\|\epsilon)$ in size.[2]

Notice that $M_{k+1}^T Q_A N_k$ is precisely the Ritz–Galerkin projection matrix of Q_A with respect to the left and right subspaces span(U_{k+1}) and span(V_k), whose singular values are the *true* Ritz values of Q_A with respect to these two subspaces, while the singular values of B_k are the *computed* Ritz values when semiorthogonality is met. Theorem 4.1 indicates that once the orthogonality levels of U_{k+1} and V_k are below $(\delta/(2k+1))^{1/2}$, the computed Ritz values are close to those true ones within $O(\epsilon)$, provided that $\|\underline{B}_k^{-1}\|$ is modest. Since the true Ritz values are never ghosts, provided that no breakdown occurs before iteration k, we avoid the appearance of ghost-computed Ritz values whenever U_{k+1} and V_k have semiorthogonality. Consequently, as long as true Ritz values are approximations to some singular values of Q_A with the accuracy $O(\|\underline{B}_k^{-1}\|\epsilon)$, the corresponding computed Ritz values have the same approximation accuracy too. In the meantime, it is easily justified that, when U_{k+1} and V_k have full orthogonality levels $O(\epsilon)$, Theorem 4.1 holds with the norm of the error matrix in the right-hand side still being $O(\|\underline{B}_k^{-1}\|\epsilon)$. Therefore, the semiorthogonality of U_{k+1} and V_k suffices for computing generalized singular values accurately. We have made a detailed investigation on the JBD process with semiorthogonalization strategy and proposed an efficient partial reorthogonalization strategy in [13].

There is a corresponding counterpart of Theorem 4.1 for $\bar{B}_k$, as stated below.

THEOREM 4.2. *Let $\hat{\delta} = O(c_4(m,n,p,k)\epsilon)$ with $c_4(m,n,p,k)$ defined by* (3.26) *and the compact QR factorizations of $\widehat{U}_k$ and V_k be $\widehat{U}_k = \widehat{M}_k\widehat{R}_k$ and $V_k = N_kS_k$, where the diagonals of $\widehat{R}_k$ and S_k are positive. If $\widehat{U}_k$ and $\widehat{V}_k$ satisfy the semiorthogonality*

$$\xi(\widehat{U}_k),\ \xi(V_k) \leqslant \sqrt{\hat{\delta}/(2k+1)},$$

then

$$\widehat{M}_k^T Q_L N_k = \bar{B}_k + \bar{E}_k,$$

where the elements of $\bar{E}_k$ are $O(\hat{\delta}) = O(\|\underline{B}_k^{-1}\|\|\bar{B}_k^{-1}\|\epsilon)$ in size.

Comparing Theorem 4.2 with Theorem 4.1, we find that Theorem 4.2 requires that the sizes of $\|\underline{B}_k^{-1}\|$ and $\|\bar{B}_k^{-1}\|$ be controllable, stronger than Theorem 4.1 does.

4.3. Residual norm and stopping criterion. Now we concentrate on designing an effective and efficient stopping criterion for the GSVD computation based

[2] In Theorem 5 of Larsen [19], the right-hand side of (4.5) is $\sqrt{\delta/k}$ instead of $\sqrt{\delta/(2k+1)}$, but Larsen does not justify it rigorously. In fact, this result is a corresponding counterpart of [31, Theorem 4]. Since the k-step Lanczos bidiagonalization of Q_A with the starting vector b is equivalent to the $(2k+1)$-step symmetric Lanczos process [3, section 7.6.1] of $\bar{C} = \begin{pmatrix} 0 & Q_A \\ Q_A^T & 0 \end{pmatrix}$ with the starting vector $\bar{b} = \begin{pmatrix} b \\ 0 \end{pmatrix}$, which holds not only in exact arithmetic but also in finite precision, the denominator in (4.5) should be $2k+1$.

on the JBD process. Still, we only assume rounding errors in the JBD process and suppose that the other computations are exact.

It is known (e.g., [36]) that the GSVD (1.1) of $\{A, L\}$ is mathematically equivalent to the generalized eigenvalue problem $s_i^2 A^T A x_i = c_i^2 L^T L x_i$. Based on this equivalence, Zha in [37] uses the residual norm

$$\|r_i^{(k)}\| = \|((s_i^{(k)})^2 A^T A - (c_i^{(k)})^2 L^T L) x_i^{(k)}\| \tag{4.7}$$

to design a stopping criterion for an approximate generalized singular value pair $\{c_i^{(k)}, s_i^{(k)}\}$ and the corresponding right vector $x_i^{(k)}$, where $(c_i^{(k)})^2 + (s_i^{(k)})^2 = 1$. Clearly, the computation of $\|r_i^{(k)}\|$ is expensive since it needs to compute $x_i^{(k)}$ explicitly by solving the large-scale problem (4.4) at each iteration k until the convergence occurs. In exact arithmetic, Zha [37] has established a sharp bound,

$$\|r_i^{(k)}\| \leq \|R\| \alpha_{k+1} \beta_{k+1} |e_k^T w_i^{(k)}|, \tag{4.8}$$

with R defined in (1.3), so that $\|R\| \alpha_{k+1} \beta_{k+1} |e_k^T w_i^{(k)}|$ can be used to design a stopping criterion if $\|R\|$ or its reasonable estimate is available. From (1.3), for $C = (A^T, L^T)^T$, we have $\|R\| = \|C\| = \sigma_{\max}(C)$, the largest singular value of C. Therefore, $\frac{\|r_i^{(k)}\|}{\|R\|}$ can be regarded as a relative residual norm of the approximate eigenvalue $(c_i^{(k)}/s_i^{(k)})^2$ and eigenvector $x_i^{(k)}$ of $s_i^2 A^T A_i = c_i^2 L^T L x_i$. In finite precision, we can obtain the following upper bound for $\|r_i^{(k)}\|$.

THEOREM 4.3. *Suppose that the inner least squares problem* (2.1) *is solved accurately and that* $x_i^{(k)} = R^{-1} V_k w_i^{(k)}$ *is the approximate right generalized singular vector by the JBD method. Then it holds that*

$$\left\| [(s_i^{(k)})^2 A^T A - (c_i^{(k)})^2 L^T L] x_i^{(k)} \right\| \leq \|R\| \left(\alpha_{k+1} \beta_{k+1} |e_k^T w_i^{(k)}| + O(\|\underline{B}_k^{-1}\| \epsilon) \right). \tag{4.9}$$

Proof. From (3.28), we have

$$Q_A^T Q_A V_k = V_k B_k^T B_k + \alpha_{k+1} \beta_{k+1} v_{k+1} e_k^T + G_{k+1} B_k + Q_A^T F_k.$$

From (4.1), we have

$$B_k^T B_k = W_k \Theta_k^2 W_k^T.$$

Notice that $(s_i^{(k)})^2 = 1 - (c_i^{(k)})^2$ and $x_i^{(k)} = R^{-1} V_k w_i^{(k)}$. Using the above two relations and (1.3), we obtain

$$\begin{aligned} \left((s_i^{(k)})^2 A^T A - (c_i^{(k)})^2 L^T L \right) x_i^{(k)} &= \left(A^T A - (c_i^{(k)})^2 (A^T A + L^T L) \right) R^{-1} V_k W_k e_i \\ &= R^T \left(Q_A^T Q_A V_k W_k - (c_i^{(k)})^2 V_k W_k \right) e_i \\ &= R^T \left(\alpha_{k+1} \beta_{k+1} v_{k+1} e_k^T w_i^{(k)} + (G_{k+1} B_k + Q_A^T F_k) w_i^{(k)} \right), \end{aligned} \tag{4.10}$$

where e_i is the ith column of I_k. From Theorem 3.2, we have

$$\|(G_{k+1} B_k + Q_A^T F_k) w_i^{(k)}\| = O(\|\underline{B}_k^{-1}\| \epsilon).$$

Therefore, by taking norms in (4.10), we prove the desired result. □

Recall the QR factorization (1.3) of C. If we perform Lanczos bidiagonalization on C several steps, then the largest Ritz value is a reasonably good lower bound for

$\sigma_1(C)$. However, we should remind the reader that a roughly good upper bound for $\|C\|$ suffices for our use here. Notice

$$\|R\|^2 = \|C\|^2 = \|C^TC\| \leq \|A^TA\| + \|L^TL\| \leq \|A\|_1\|A\|_\infty + \|L\|_1\|L\|_\infty,$$

where $\|\cdot\|_1$ and $\|\cdot\|_\infty$ denote the 1-norm and the infinity norm, which is cheap to compute when A and L are explicitly stored in a certain sparse format. We then take the square root of the above right-hand side as an estimate for $\|R\|$. Alternatively, it is simpler to use $\|C\|_1 \leq \|A\|_1 + \|L\|_1$ or $\|C\|_\infty = \max\{\|A\|_\infty, \|L\|_\infty\}$ as a replacement of $\|R\|$. Suppose that $\|\underline{B}_k^{-1}\| = O(1)$. Since $e_k^T w_i^{(k)}$ is available from the SVD of B_k or $\bar{B}_k$ or the GSVD of $\{B_k, \bar{B}_k\}$, the quantity $|\|R\|\alpha_{k+1}\beta_{k+1}|e_k^T w_i^{(k)}|$ can be used as a reliable stopping criterion, provided that the stopping tolerance for the (relative) residual norm is not required to achieve the level of ϵ. Computationally, we benefit very much from this criterion since we avoid the explicit computation of $x_i^{(k)}$ before the convergence. We will numerically confirm the reliability of the criterion.

5. Numerical experiments.

We report numerical experiments to justify the results obtained, except Theorem 3.3, which has been numerically confirmed in [13]. All the numerical experiments were performed on an Intel Core i7-7700 CPU at 3.60 GHz with a main memory of 8 GB using Matlab R2017a with the machine precision $\epsilon = 2.22 \times 10^{-16}$ under the Miscrosoft Windows 10 64-bit system. For each matrix pair $\{A, L\}$, we use $b = (1, \ldots, 1)^T \in \mathbb{R}^m$ as the starting vector of the JBD process, and each inner least squares problem (2.1) is solved accurately by computing the QR factorization (1.3) and QQ^Tw for a given w.

5.1. Examples for the JBD process in finite precision.

We choose four matrix pairs to confirm the numerical behavior of the JBD process in finite precision. We construct the first pair $\{A_c, L_s\}$ as follows: Take $n = 800$ and $C_A =$ diag(c), $S_L =$ diag(s), where $c = (\frac{3n}{2}, \frac{3n}{2} - 1, \ldots, \frac{n}{2} + 1)/2n$ and $s = (\sqrt{1-c_1^2}, \ldots, \sqrt{1-c_n^2})$. Let D be the symmetric orthogonal matrix generated by the Matlab built-in function `D=gallery('orthog',n,2)`. We then define $A = C_AD$ and $L = S_LD$. By construction, the ith generalized singular value of $\{A, L\}$ is $\{c_i, s_i\}$, the corresponding right vector x_i is the ith column of D, and the left generalized singular vectors p_i^A and p_i^L are the ith column e_i of I_n, $i = 1, \ldots, n$. The remaining three pairs use sparse matrices from [4], where

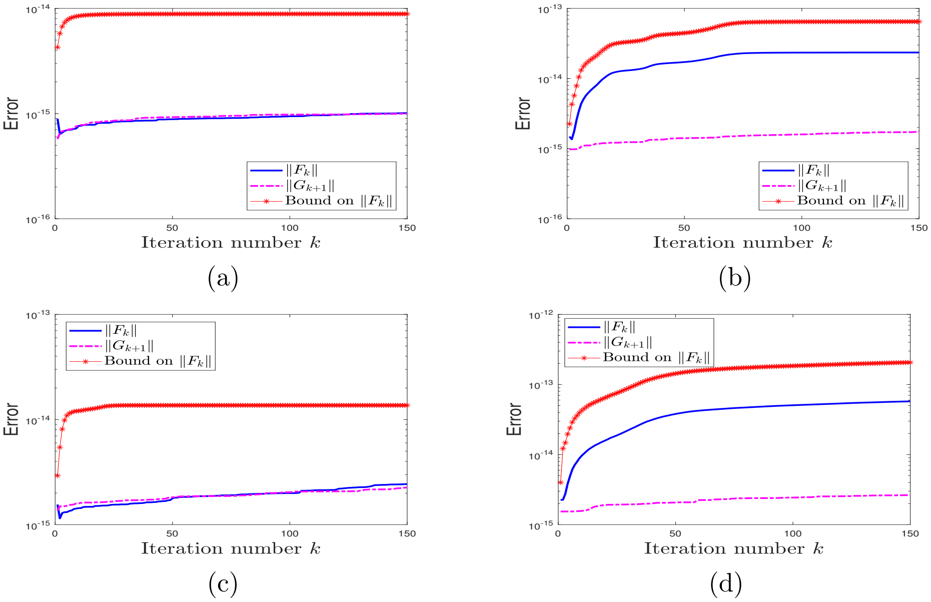
$$L = L_1 = \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(n-1)\times n} \tag{5.1}$$

with $n = 712$, which is the scaled discrete approximation of the first order derivative operator, and $L = L_n =$ diag(l) with $l = (2n, 2n-1, \ldots, n+1)/1000$, $n = 3969$. Some properties of the four test matrix pairs are described in Table 1, where $\kappa(A) = \|A\|\|A^\dagger\|$ is the condition number of A.

Figure 1 depicts the growths of $\|F_k\|$ and $\|G_{k+1}\|$ in (3.14) and (3.15) as the iteration number k increases from 1 to 150. By Theorem 3.2, we take $O(\|\underline{B}_k^{-1}\|\epsilon) = 10\|\underline{B}_k^{-1}\|\epsilon$. For the four test problems, it is seen from Figures 1(a)–1(d) that $\|F_k\|$ grows very slowly as k increases. For the four matrix pairs, $O(\|\underline{B}_k^{-1}\|\epsilon)$ is indeed a

TABLE 1
Properties of the test matrices.

A	$m \times n$	$\kappa(A)$	L	$p \times n$	$\kappa(L)$
A_c	800×800	2.99	L_s	800×800	1.46
well1850	1850×712	111.31	L_1	711×712	453.27
rdb2048	2048×2048	2026.80	dw2048	2048×2048	5301.50
c-23	3969×3969	22795.9	L_n	3969×3969	1.9995



(a)

(b)

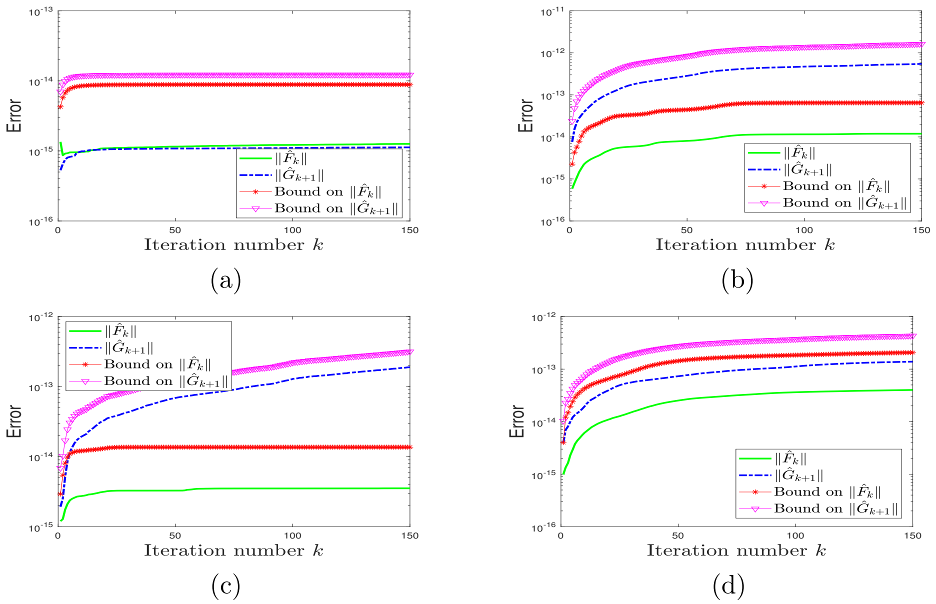
(c)

(d)

FIG 1. *Estimated error bound for* $\|F_k\|$*:* (a) $\{A_c,\ L_s\}$*;* (b) $\{$`well1850`$,\ L_1\}$*;* (c) $\{$`rdb2048`, `dw2048`$\}$*;* (d) $\{$`c-23`$,\ L_n\}$*.*

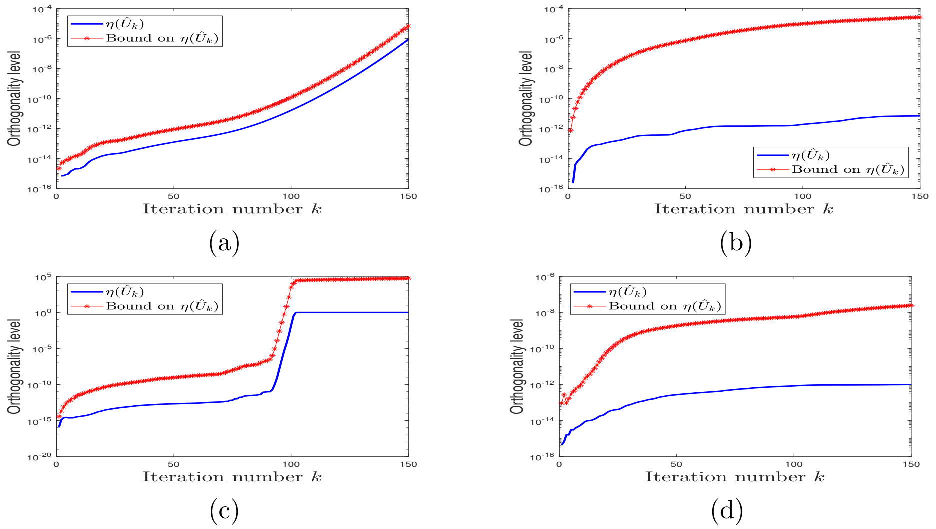
very good upper bound for $\|F_k\|$ within ten times, and the growth trends of $\|F_k\|$ and $\|\underline{B}_k^{-1}\|$ are similar. This indicates that the growth of $\|F_k\|$ is mainly affected by the growth of $\|\underline{B}_k^{-1}\|$. Since $QQ^T\tilde{u}_i$ is explicitly computed at each step in our experiments, $\|G_{k+1}\| = O(\epsilon)$ remains almost unchanged.

Figure 2 depicts the growths of $\|\widehat{F}_k\|$ and $\|\widehat{G}_k\|$ in (3.22) and (3.23). By Theorem 3.4, we take $O(\|\underline{B}_k^{-1}\|\epsilon) = 10\|\underline{B}_k^{-1}\|\epsilon$ and $O((\|\underline{B}_k^{-1}\| + \|\widehat{B}_k^{-1}\|)\epsilon) = 10(\|\underline{B}_k^{-1}\| + \|\widehat{B}_k^{-1}\|)\epsilon$, respectively. From the figures, we see that $O(\|\underline{B}_k^{-1}\|\epsilon)$ and $O((\|\underline{B}_k^{-1}\| + \|\widehat{B}_k^{-1}\|)\epsilon)$ are indeed reasonable upper bounds for $\|\widehat{F}_k\|$ and $\|\widehat{G}_k\|$, and the growths of $\|\widehat{F}_k\|$ and $\|\widehat{G}_k\|$ are critically affected by those of $\|\underline{B}_k^{-1}\|$ and $\|\underline{B}_k^{-1}\| + \|\widehat{B}_k^{-1}\|$, respectively. For the four matrix pairs, $\|\underline{B}_k^{-1}\|$ always grows slowly, but $\|\widehat{B}_k^{-1}\| = \|\bar{B}_k^{-1}\|$ grows faster for $\{$well1850, $L_1\}$ than for the other three pairs. This is because L_1 is truly flat, and the smallest singular value of $\bar{B}_k$ approaches zero as k increases, causing that $\|\bar{B}_k^{-1}\|$ is ultimately very large, as we have shown after Theorem 3.2. In contrast, the smallest singular value of $\bar{B}_k$ converges to the nonzero smallest one of L for the other three matrix pairs, and $\|\bar{B}_k^{-1}\|$ is uniformly bounded by the reciprocal of the smallest singular value of L.



(a) (b) (c) (d)

FIG 2. $\|\widehat{F}_k\|$ *and* $\|\widehat{G}_k\|$ *as well as estimated error bounds for them:* (a) $\{A_c,\ L_s\}$*;* (b) {`well1850`, L_1}*;* (c) {`rdb2048, dw2048`}*;* (d) {`c-23`, L_n}.



(a) (b) (c) (d)

FIG 3. *The orthogonality level* $\eta(\widehat{U}_k)$ *and its upper bound:* (a) $\{A_c,\ L_s\}$*;* (b) {`well1850`, L_1}*;* (c) {`rdb2048, dw2048`}*;* (d) {`c-23`, L_n}.

Figure 3 depicts the orthogonality level $\eta(\widehat{U}_k)$ as k increases from 1 to 150. The upper bound for $\eta(\widehat{U}_k)$ is (3.30), and we use 10ϵ as an estimate for $O(c_3(m,n,p)\epsilon)$. We observe that the orthogonality of $\widehat{U}_k$ is lost gradually. Particularly, for the test problem {rdb2048, dw2048}, the columns of $\widehat{U}_k$ lose orthogonality completely and become numerically linearly dependent after $k = 100$. The growth trends of $\eta(\widehat{U}_k)$ and its bound resemble, meaning that the orthogonality level of $\widehat{U}_k$ is affected not only by $\eta(U_k)$ and $\eta(\tilde{V}_k)$ but also by $\|\widehat{B}_k^{-1}\|$.

5.2. Examples for the GSVD computation. We illustrate the performance of the JBD method for computing a few extreme GSVD components of $\{A, L\}$ and justify the upper bound in (4.9). Whenever reorthogonalization is used, we mean the full reorthogonalizations of all the sets of basis vectors. For effective partial reorthogonalization of selected ones, we refer the reader to [13], where the numerical experiments have confirmed Theorems 4.1–4.2 when the basis vectors are only semiorthogonal.

Example 1. We show the convergence of the singular values of B_k. Take $m = n = p = 500$. We construct a row vector $c = (c_1, \ldots, c_n)$ with

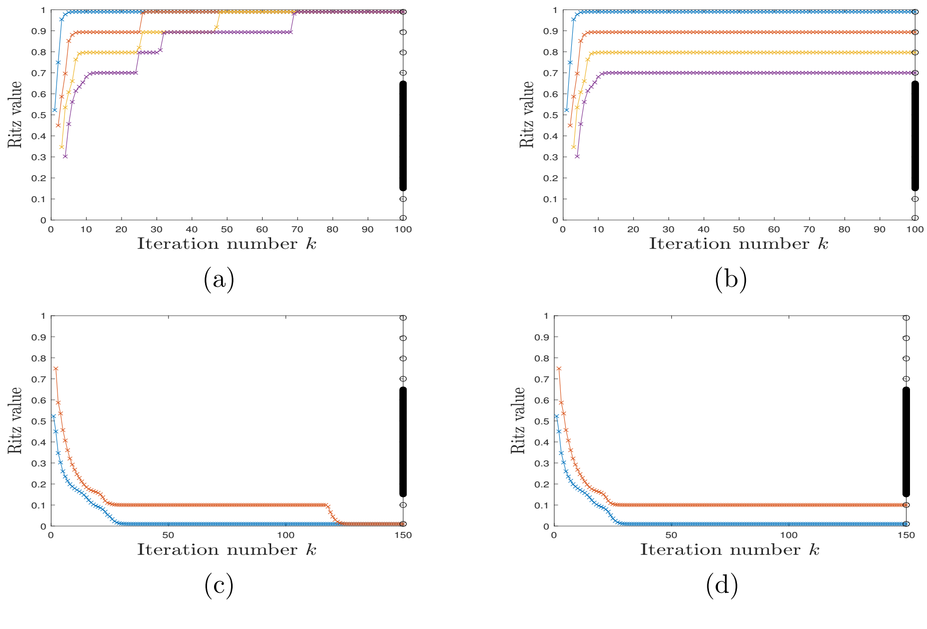
$$\begin{aligned} &l_{\max} = 4, l_{\min} = 2, c_{(1:l_{\max})} = \texttt{linspace}(0.99, 0.7, l_{\max}),\\ &c_{(n-l_{\min}+1:n)} = \texttt{linspace}(0.10, 0.01, l_{\min}),\\ &c_{(l_{\max}+1:n-l_{\min})} = \texttt{linspace}(0.65, 0.15, n - l_{\max} - l_{\min}) \end{aligned}$$

and

$$s = (\sqrt{1-c_1^2}, \ldots, \sqrt{1-c_n^2}),$$

where `linspace` is the Matlab built-in function. We then define $C_A = \operatorname{diag}(c)$, $S_L = \operatorname{diag}(s)$, and `D = gallery('orthog',n,2)` and take $A = C_A D$ and $L = S_L D$. By construction, $\kappa(A) = 6.6000$, $\kappa(L) = 7.0888$, and the ith large generalized singular value pair of $\{A, L\}$ is $\{c_i, s_i\}$.

Figure 4 depicts the convergence processes of the first four largest and two smallest Ritz values computed by the SVD of B_k, which correspond to the four largest and two smallest generalized singular values of $\{A, L\}$, respectively, where we implemented the



(a) (b) (c) (d)

FIG 4. *Convergence of Ritz values computed by the SVD of* B_k*:* (a) *the first four largest Ritz values without reorthogonalization;* (b) *the first four largest Ritz values with full reorthogonalization;* (c) *the first two smallest Ritz values without reorthogonalization;* (d) *the first two smallest Ritz values with full reorthogonalization.*

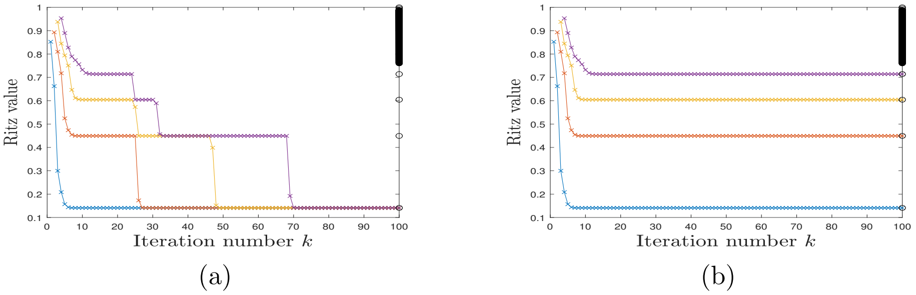


FIG 5. *Convergence of Ritz values computed by the SVD of* $\bar{B}_k$*:* (a) *the first four smallest Ritz values without reorthogonalization;* (b) *the first four smallest Ritz values with full reorthogonalization.*

JBD process without and with reorthogonalization. The right vertical line indicates the values of c_i for $i=1,\ldots,500$, and the left and right panels exhibit the convergence behavior of the JBD method without and with reorthogonalization. We observe from Figure 4(a) that each of the converged second to fourth Ritz values suddenly jumps to become a ghost and then converges to the next large singular values after several iterations. Such a phenomenon repeats several times and corresponds to the appearance of spurious copies each time. More precisely, as Figure 4(a) indicates, the four largest Ritz values gradually become numerically multiple and ultimately converge to the largest singular value of Q_A as k increases. Similarly, as Figure 4 (c) shows, the second smallest Ritz value first converges to the second smallest singular value of Q_A, then starts to converge to the first smallest one and ultimately is numerically the same as the first converged Ritz value.

However, when the JBD process is performed with full reorthogonalization, the convergence of the Ritz values changes and becomes regular, as Figures 4(b) and 4(d) indicate. In the right panels, the convergence behavior is much simpler and is in accordance with the theoretical results in exact arithmetic. It is clear that a simple generalized singular value is approximated by a single Ritz value and that no ghosts appear. We also observe that the large Ritz values converge more quickly than relatively interior Ritz values; that is, the Ritz values closer to the rightmost generalized singular values stabilize more early, which confirms the theory that the JBD method generally favors the extreme generalized singular values.

Figure 5 depicts the convergence processes of the first four smallest Ritz values computed by the SVD of $\bar{B}_k$, which corresponds to the first four largest generalized singular values of $\{A,L\}$. The right vertical line indicates the values of s_i for $i=1,\ldots,500$. From Figure 5(a), we observe the "ghost" phenomenon that some converged Ritz values suddenly jump and then converge to the next small singular values of Q_L after several iterations. The convergence phenomena are similar to Figures 4(a) and 4(c). Figure 5(b) shows the convergence of Ritz values with full reorthogonalization, from which it is clear that the JBD method converges regularly and that there are no spurious copies. Figure 5(b) demonstrates that the JBD method favors the extreme generalized singular values.

Example 2. We investigate the convergence of the approximate generalized singular values and vectors of $\{A,L\}$, which are computed by using both the SVDs of B_k and $\bar{B}_k$ and the GSVD of $\{B_k,\bar{B}_k\}$. We test two matrix pairs. The first pair is the problem in Example 1, and the second matrix pair $\{\mathsf{dw256A},\mathsf{dw256B}\}$ is an

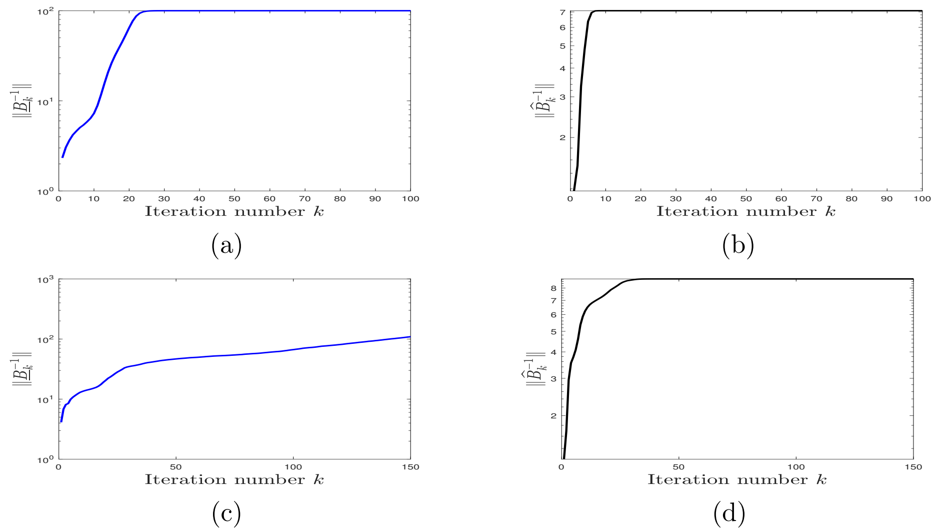


FIG 6. *Growths of* $\|\underline{B}_k^{-1}\|$ *and* $\|\widehat{B}_k^{-1}\|$*:* (a), (b) $\{A_{500},\ L_{500}\}$ *in Example* 1*;* (c), (d) $\{dw256A, 256B\}$ *in Example* 2*.*

electromagnetic problem with $m=n=p=512$ from the non-Hermitian Eigenvalue Problem Collection in the Matrix Market,[3] where $\kappa(A)=11490.4$ and $\kappa(L)=3.7328$.

We use the JBD method with full reorthogonalization to compute the largest generalized singular value and vectors. Instead of the SVD of the individual B_k or $\bar{B}_k$, we compute the SVDs of B_k and $\bar{B}_k$ simultaneously and take $\{c_1^{(k)}, \bar{s}_1^{(k)}\}$ to approximate $\{c_1, s_1\}$, where $c_1^{(k)}$ is the largest singular value of B_k and $\bar{s}_1^{(k)}$ is the smallest singular value of $\bar{B}_k$. The approximations to the right and left generalized singular vectors x_1 and p_1^A are computed from the SVD of B_k, and the approximation to the left generalized singular vectors p_1^L is computed from the SVD of $\bar{B}_k$. Alternatively, we also compute the GSVD of $\{B_k, \bar{B}_k\}$ and obtain the approximation $\{c_1^{(k)}, s_1^{(k)}\}$ to $\{c_1, s_1\}$ and the approximations $x_1^{(k)}, y_1^{(k)}, z_1^{(k)}$ to x_1, p_1^A, p_1^L.

We use the angle error

$$\sin\theta_k = |\bar{s}_1^{(k)} c_1 - s_1 c_1^{(k)}| \text{ or } |s_1^{(k)} c_1 - s_1 c_1^{(k)}|$$

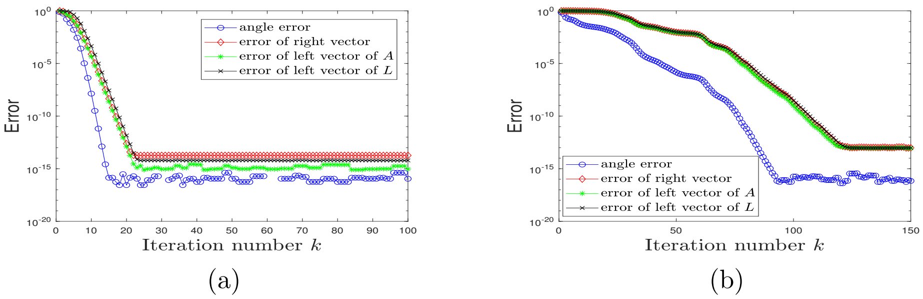
to measure the error between $\{c_1^{(k)}, \bar{s}_1^{(k)}\}$ or $\{c_1^{(k)}, s_1^{(k)}\}$ and $\{c_1, s_1\}$ [35], where θ_k denotes the angle between the vectors $(c_1, s_1)^T$ and $(c_1^{(k)}, \bar{s}_1^{(k)})^T$ or $(c_1^{(k)}, s_1^{(k)})^T$. For the corresponding generalized singular vectors, we measure the errors

$$\sin\angle(x_1, x_1^{(k)}), \quad \sin\angle(p_1^A, y_1^{(k)}), \quad \sin\angle(p_1^L, z_1^{(k)}).$$

Figures 6(a), 6(b), and 6(d) show that $\|\underline{B}_k^{-1}\|$ and $\|\widehat{B}_k^{-1}\|$ stabilize in several iterations, showing that the smallest singular values of $\underline{B}_k$ and $\widehat{B}_k$ have converged to the smallest *positive* ones of Q_A and Q_L, respectively, as commented after Theorem 3.2.

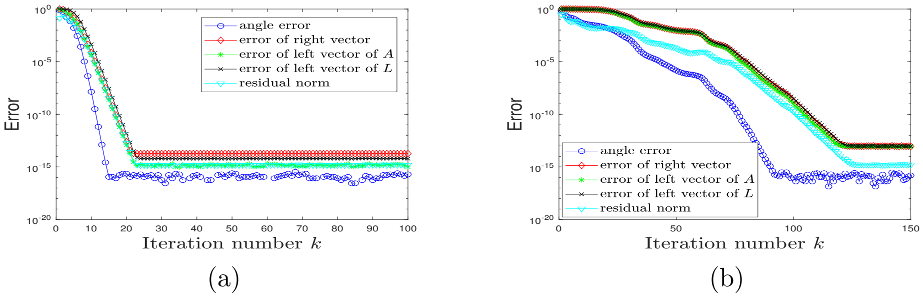
Figure 7 draws the approximation processes of the approximate generalized singular values and vectors obtained by the SVDs of B_k and $\bar{B}_k$ as k increases, while Figure 6 depicts the growths of $\|\underline{B}_k^{-1}\|$ and $\|\widehat{B}_k^{-1}\|$. We have found that $\|\underline{B}_k^{-1}\|$ and

[3] https://math.nist.gov/MatrixMarket.



(a) (b)

FIG 7. *Convergence processes of the approximate GSVD components:* (a) $\{A_{500},\ L_{500}\}$ *in Example* 1*;* (b) $\{\mathsf{dw256A}, \mathsf{dw256B}\}$ *in Example* 2.



(a) (b)

FIG 8. *Convergence processes of the approximate GSVD components based on the GSVD of* $\{B_k, \bar{B}_k\}$*:* (a) $\{A_{500},\ L_{500}\}$ *in Example* 1*;* (b) $\{\mathsf{dw256A}, \mathsf{dw256B}\}$.

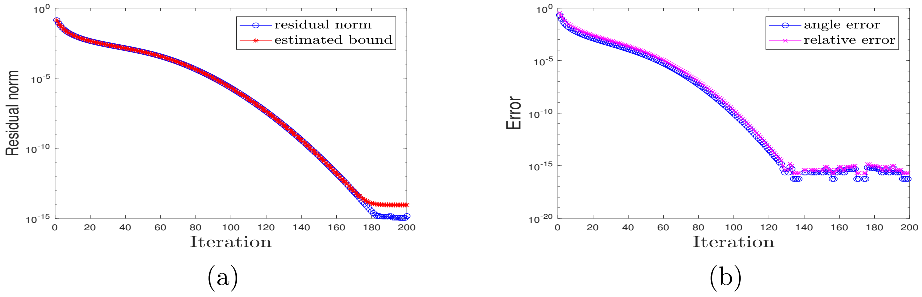
$\|\widehat{B}_k^{-1}\|$ grow quite slowly and are very modest for Example 1 when $k=1\sim 150$ and for Example 2 when $k=1\sim 150$, respectively. We have also seen from Figure 7 that the computed Ritz value converges to the the largest generalized singular value with accuracy (ϵ). This confirms Theorems 4.1–4.2 and the comments on them.

Figure 8 depicts the convergence processes of approximate GSVD components computed by the GSVD of $\{B_k, \bar{B}_k\}$. In this figure, we also draw the curves of residual norms. Clearly, the computed results are very similar to those obtained by the SVDs of B_k and $\bar{B}_k$ until the errors reach the level of ϵ. Therefore, the approximate GSVD components converge regularly, the JBD method converges fast, and all the errors achieve the level of ϵ after 20 iterations for Example 1. For the GSVD of $\{\mathsf{dw256A}, \mathsf{dw256B}\}$, the JBD method computes the largest GSVD component quite accurately, and the relative residual norm reaches 10^{-8} after 100 iterations and stabilizes at $O(\epsilon)$ after 125 iterations.

Example 3. We show the residual norm and its upper bound (4.9). The matrix pair $\{A, L\}$ is chosen to be $\{A_{800}, L_{800}\}$ in Table 1, and we use the largest singular value of B_k to compute an approximation to the largest generalized singular value. From the construction, we have $\|(A_{800}^T, L_{800}^T)^T\| = 1$, and the largest generalized singular value is $\{c_1, s_1\}$, where $c_1 = 0.75$ and $s_1 = \sqrt{1-c_1^2}$.

In Figure 9, we draw the convergence histories of the approximate largest generalized singular value and the residual norm as well as those of the approximate generalized singular values by using both the angle error and relative error.

$$|c_1^{(k)}/s_1^{(k)} - c_1/s_1|/(c_1/s_1).$$



(a) (b)

FIG 9. *Convergence history of the approximate largest generalized singular value of* $\{A_{800}, L_{800}\}$*:* (a) *residual norm and its upper bound;* (b) *angle error and relative error.*

From Figure 9(b), It is found that the approximate largest generalized singular value $c_1^{(k)}/s_1^{(k)}$ converges to c_1/s_1 with the ultimate relative error $O(\epsilon)$. This justifies the comments after Theorem 4.1. As is expected, the angle errors and relative errors resemble very much since $c_1/s_1 = O(1)$. We observe from Figure 9(a) that the residual norm and its upper bound are almost the same as k increases. The true residual norm decays until the level of ϵ, but the estimated upper bound stagnates at the level that is a little bit higher than ϵ since the upper bound for $\|r_i^{(k)}\|$ has a term $O(\|\underline{B}_k^{-1}\|\epsilon)$, which is considerably bigger than ϵ when $\|\underline{B}_k^{-1}\| > 1$ considerably. For the case that $\|\underline{B}_k^{-1}\|$ remains modest, the term $O(\|\underline{B}_k^{-1}\|\epsilon)$ plays no role in the upper bound until the bound reaches $O(\epsilon)$. Therefore, the upper bound $\|R\|\alpha_{k+1}\beta_{k+1}|e_k^T w_i^{(k)}|$ can be used as a reliable stopping criterion for the JBD algorithm. We point out that, in large matrix computations, a (relative) stopping tolerance is usually $O(\epsilon^{1/2})$. Therefore, provided that $\|\underline{B}_k^{-1}\| \leq O(\epsilon^{-1/2})$, our upper bound is a reliable estimate for $\|r_i^{(k)}\|$.

6. Conclusions and future work. We have made a numerical analysis of the JBD process on $\{A, L\}$ in finite precision and established close relationships between it and respective lower and upper Lanczos bidiagonalizations of Q_A and Q_L. The results have shown that the k-step JBD process for computing U_{k+1}, V_k, and B_k is equivalent to the lower Lanczos bidiagonalization of Q_A with the error $\delta = O(\|\underline{B}_k^{-1}\|\epsilon)$ and for computing $\widehat{U}_{k+1}$, V_k, and $\widehat{B}_k$ is equivalent to the upper Lanczos bidiagonalization of Q_L with the error $\hat{\delta} = O(\|\underline{B}_k^{-1}\|\|\widehat{B}_k^{-1}\|\epsilon)$. We have investigated the loss of orthogonality of the computed basis vectors and established a relationship between the orthogonality levels $\xi(\widetilde{V}_k)$ and $\xi(V_k)$ and an upper bound for the orthogonality level $\eta(\widehat{U}_k)$, showing that $\eta(\widehat{U}_k)$ is controlled by $\eta(U_{k+1})$, $\eta(V_k)$ and $\|\widehat{B}_k^{-1}\|$.

We have described a JBD method that computes approximate generalized singular values and vectors of $\{A, L\}$ and considered the convergence and accuracy of the approximate generalized singular values. The results have indicated that the generalized singular values of B_k and $\bar{B}_k$ are as accurate as the true Ritz values of Q_A and Q_L with respect to the given subspaces within $\mathcal{O}(\epsilon)$, provided that the basis vectors have semiorthogonality levels and $\underline{B}_k$ and $\bar{B}_k$ are not ill conditioned. Under these conditions, it is only necessary to maintain the desired semiorthogonality in order to obtain the approximate GSVD components with the same accuracy as those obtained by the JBD method with full reorthogonalization. An efficient partial reorthogonalization strategy has been proposed in [13] for this purpose.

In the meantime, we have established a compact upper bound for the residual norm $\|r_i^{(k)}\|$ of an approximate generalized singular value and approximate right generalized singular vector in finite precision and shown that it can be used as a cheap and reliable stopping criterion without explicitly computing the approximate right generalized singular vector until the convergence occurs. Finally, we have reported numerical experiments to justify the results obtained and assertions.

There remain some important issues. For instance, due to the limitation of storage, it is generally necessary to restart the JBD method. A commonly used restarting technique is the implicit restarting proposed in [34] for the eigenvalue problem and adapted to the SVD computation in [14, 15, 20]. How to adapt the implicit restart to the JBD method and develop efficient algorithms is very significant. Also, notice that the residual norm (4.7) is used to measure the convergence of the JBD method, which is the residual norm of an approximate generalized eigenpair $((c_i^{(k)}/s_i^{(k)})^2, x_i^{(k)})$ of $s_i^2 A^T A x_i = c_i^2 L^T L x_i$ and does not take into account approximate left generalized singular vectors $y_i^{(k)}$ for A and $z_i^{(k)}$ for L. Indeed, the convergence and accuracy of approximate right and two left generalized singular vectors may differ greatly for some problems, so do two approximate left ones. Therefore, a more reliable and general-purpose criterion should take into consideration the approximate generalized singular value pair and corresponding right and two left singular vectors and measure their residual norm as an approximate generalized singular component of the original GSVD of the matrix pair $\{A, L\}$. That is, it is much more proper to measure the residual norm of the approximate GSVD components $(c_i^{(k)}, s_i^{(k)}, x_i^{(k)}, y_i^{(k)}, z_i^{(k)})$, which, by the definition (1.2) of GSVD of $\{A, L\}$, is

$$\|r_{i,\text{new}}^{(k)}\| = \sqrt{\|Ax_i^{(k)} - c_i^{(k)} y_i^{(k)}\|^2 + \|Lx_i^{(k)} - s_i^{(k)} z_i^{(k)}\|^2 + \|s_i^{(k)} A^T y_i^{(k)} - c_i^{(k)} L^T z_i^{(k)}\|^2}.$$

For the JBD method, the first two terms in the square root are zeros, leading to

$$\|r_{i,\text{new}}^{(k)}\| = \|s_i^{(k)} A^T y_i^{(k)} - c_i^{(k)} L^T z_i^{(k)}\|.$$

Therefore, one can compute it directly without involving the approximate right generalized singular vector $x_i^{(k)}$. Similarly to the derivations of (4.8) and (4.9), one can establish sharp upper bounds for $\|r_{i,\text{new}}^{(k)}\|$ in exact arithmetic and in finite precision, which do not need to compute the approximate left generalized singular vector $y_i^{(k)}$ and $z_i^{(k)}$ before the occurrence of convergence, and, meanwhile, save the matrix-vector products $A^T y_i^{(k)}$ and $L^T z_i^{(k)}$. As a result, one can design an efficient general-purpose stopping criterion for the JBD method. These issues are not focuses of this paper, and we do not give details on them.

Acknowledgments. We thank the two referees and the editor Professor Berry for their comments and suggestions, which made us improve the presentation considerably.

REFERENCES

[1] Z. Bai, J. Demmel, J. Dongrarra, A. Ruhe, and H. van der Vorst , *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, SIAM, Philadelphia, 2000.

[2] J. L. Barlow, *Reorthogonalization for the Golub-Kahan-Lanczos bidiagonal reduction*, Numer. Math., 124 (2013), pp. 237–278.

[3] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.

[4] T. A. Davis and Y. Hu, *The University of Florida sparse matrix collection*, ACM Trans. Math. Softw., 38 (2011), pp. 1–25.

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

[6] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, SIAM, Philadelphia, 1998.

[7] P. C. Hansen, *Discrete Inverse Problems: Insight and Algorithms*, SIAM, Philadelphia, 2010.

[8] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002.

[9] J. Huang and Z. Jia, *Two harmonic Jacobi-Davidson methods for computing a partial generalized singular value decomposition of a large matrix pair*, J. Sci. Comput., 93 (2022), 41.

[10] Z. Jia, *The low rank approximations and Ritz values in LSQR for linear discrete ill-posed problem*, Inverse Problems, 36 (2020), 045013.

[11] Z. Jia, *Regularization properties of LSQR for linear discrete ill-posed problems in the multiple singular value case and best, near best and general low rank approximations*, Inverse Problems, 36 (2020), 085009.

[12] Z. Jia, *Regularization properties of Krylov iterative solvers CGME and LSMR for linear dicrete ill-posed problems with an application to truncated randomized SVDs*, Numer. Algorithms, 85 (2020), pp. 1281–1310.

[13] Z. Jia and H. Li, *The joint bidiagonalization process with partial reorthogonalization*, Numer. Algorithms, 88 (2021), pp. 965–992.

[14] Z. Jia and D. Niu, *An implicitly restarted refined bidiagonalization Lanczos method for computing a partial singular value decomposition*, SIAM J. Matrix Anal. Appl., 25 (2003), pp. 246–265.

[15] Z. Jia and D. Niu, *A refined harmonic Lanczos bidiagonalization method and an implicitly restarted algorithm for computing the smallest singular triplets of large matrices*, SIAM J. Sci. Comput., 32 (2010), pp. 714–744.

[16] Z. Jia and Y. Yang, *A joint bidiagonalization based algorithm for large scale general-form Tikhonov regularization*, Appl. Numer. Math., 157 (2020), pp. 159–177.

[17] M. E. Kilmer, P. C. Hansen, and M. I. Español, *A projection-based approach to general-form Tikhonov regularization*, SIAM J. Sci. Comput., 29 (2007), pp. 315–330.

[18] C. Lanczos, *An iteration method for the solution of eigenvalue problem of linear differential and integral operators*, J. Res. Natl. Bureau Standards, 45 (1950), pp. 255–282.

[19] R. M. Larsen, *Lanczos Bidiagonalization with Partial Reorthogonalization*, Technical report DAIMI PB-357, University of Aarhus, 1998, http://sun.stanford.edu/∼rmunk/PROPACK/paper.pdf.

[20] R. M. Larsen, *Combining Implicit Restarts and Partial Reorthogonalization in Lanczos Bidiagonalization*, Technical report, SCCM, Stanford University, 2001.

[21] G. Meurant and Z. Strakoš, *The Lanczos and conjugate gradient algorithms in finite precision arithmetic*, Acta Numer., 15 (2006), pp. 471–542.

[22] C. C. Paige, *The Computation of Eigenvalues and Eigenvectors of Very Large Sparse Matrices*, Ph.D. thesis, London University, 1971.

[23] C. C. Paige, *Computational variants of the Lanczos method for the eigenproblem*, J. Inst. Math. Appl., 10 (1972), pp. 373–381.

[24] C. C. Paige, *Error analysis of the Lanczos algorithm for tridiagonalizing a symmetric matrix*, J. Inst. Math. Appl., 18 (1976), pp. 341–349.

[25] C. C. Paige, *Accuracy and effectiveness of the Lanczos algorithm for the symmetric eigenproblem*, Linear Algebra Appl., 34 (1980), pp. 235–258.

[26] C. C. Paige and M. A. Saunders, *Towards a generalized singular value decomposition*, SIAM J. Numer. Anal., 18 (1981), pp. 398–405.

[27] C. C. Paige and M. A. Saunders, *LSQR: An algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Softw., 8 (1982), pp. 43–71.

[28] B. N. Parlett, *The Symmetric Eigenvalue Problem*, SIAM, Philadelphia, 1998.

[29] B. N. Parlett and D. S. Scott, *The Lanczos algorithm with selective orthogonalization*, Math. Comp., 33 (1979), pp. 217–238.

[30] Y. Saad, *On the rates of convergence of the Lanczos and the block-Lanczos methods*, SIAM J. Numer. Anal., 17 (1980), pp. 687–706.

[31] H. D. Simon, *Analysis of the symmetric Lanczos algorithm with reorthogonalization methods*, Linear Algebra Appl., 61 (1984), pp. 101–131.

[32] H. D. Simon, *The Lanczos algorithm with partial reorthogonalization*, Math. Comp., 42 (1984), pp. 115–142.

[33] H. D. Simon and H. Zha, *Low-rank matrix approximation using the Lanczos bidiagonalization process with applications*, SIAM J. Sci. Comput., 21 (2000), pp. 2257–2274.

[34] D. C. Sorensen, *Implicit application of polynomial filters in a k-step Arnoldi method*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 357–385.

[35] J.-G. Sun, *Perturbation analysis for the generalized singular value problem*, SIAM J. Numer. Anal., 20 (1983), pp. 611–625.

[36] C. F. Van Loan, *Generalizing the singular value decompositions*, SIAM J. Numer. Anal., 13 (1976), pp. 76–83.

[37] H. Zha, *Computing the generalized singular values/vectors of large sparse or structured matrix pairs*, Numer. Math., 72 (1996), pp. 391–417.