CHARACTERIZING GSVD BY SINGULAR VALUE EXPANSION OF LINEAR OPERATORS AND ITS COMPUTATION *

HAIBO LI†

Abstract. The generalized singular value decomposition (GSVD) of a matrix pair $\{A, L\}$ with $A \in \mathbb{R}^{m\times n}$ and $L \in \mathbb{R}^{p\times n}$ generalizes the singular value decomposition (SVD) of a single matrix. In this paper, we provide a new understanding of GSVD from the viewpoint of SVD, based on which we propose a new iterative method for computing nontrivial GSVD components of a large-scale matrix pair. By introducing two linear operators $\mathcal{A}$ and $\mathcal{L}$ induced by $\{A, L\}$ between two finite-dimensional Hilbert spaces and applying the theory of singular value expansion (SVE) for linear compact operators, we show that the GSVD of $\{A, L\}$ is nothing but the SVEs of $\mathcal{A}$ and $\mathcal{L}$. This result characterizes completely the structure of GSVD for any matrix pair with the same number of columns. As a direct application of this result, we generalize the standard Golub-Kahan bidiagonalization (GKB) that is a basic routine for large-scale SVD computation such that the resulting generalized GKB (gGKB) process can be used to approximate nontrivial extreme GSVD components of $\{A, L\}$, which is named the gGKB_GSVD algorithm. We use the GSVD of $\{A, L\}$ to study several basic properties of gGKB and also provide preliminary results about convergence and accuracy of gGKB_GSVD for GSVD computation. Numerical experiments are presented to demonstrate the effectiveness of this method.

Key words. GSVD, linear operator, singular value expansion, generalized Golub-Kahan bidiagonalization, Krylov subspace

MSC codes. 15A22, 47A05, 65F99

1. Introduction.

The generalized singular value decomposition (GSVD) of a matrix pair is an extension of the singular value decomposition (SVD) of a single matrix. First introduced by Van Loan [52] and further developed by many others [41, 49, 53], now the GSVD has been a standard matrix decomposition [8, 19]. The GSVD provides an important mathematical tool for analyzing relationships between two sets of variables or matrices, which is particularly useful in various applications, including signal processing [37, 48], statistics [39, 43], computational biology [1], and many others [7, 17, 22, 27, 29].

Let I_k denote the identity matrix of order k and $\mathbf{0}$ denote the zero matrix or vector with dimensions clarified by the context. For any two matrices with the same number of columns, the general-form GSVD is stated as follows [41]:

THEOREM 1.1 (GSVD). *Let $A \in \mathbb{R}^{m\times n}$ and $L \in \mathbb{R}^{p\times n}$ with* $\mathrm{rank}((A^\top, L^\top)^\top) = r$. *Then the GSVD of $\{A, L\}$ is*

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

with

$$C_A = \begin{array}{cc} \begin{pmatrix} \Sigma_A & \mathbf{0} \end{pmatrix} & m \\ \begin{matrix} r & \quad n-r \end{matrix} & \end{array}, \quad S_L = \begin{array}{cc} \begin{pmatrix} \Sigma_L & \mathbf{0} \end{pmatrix} & p \\ \begin{matrix} r & \quad n-r \end{matrix} & \end{array} \tag{1.1b}$$

and

$$\Sigma_A = \begin{array}{cc} \begin{pmatrix} I_{q_1} & & \\ & C_{q_2} & \\ & & \mathbf{0} \end{pmatrix} & \begin{matrix} q_1 \\ q_2 \\ m-q_1-q_2 \end{matrix} \\ \begin{matrix} q_1 & q_2 & q_3 \end{matrix} & \end{array}, \quad \Sigma_L = \begin{array}{cc} \begin{pmatrix} \mathbf{0} & & \\ & S_{q_2} & \\ & & I_{q_3} \end{pmatrix} & \begin{matrix} p-r+q_1 \\ q_2 \\ q_3 \end{matrix} \\ \begin{matrix} q_1 & q_2 & q_3 \end{matrix} & \end{array}, \tag{1.1c}$$

*Submitted to the editors DATE.

†School of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia. (haibo.li@unimelb.edu.au).

where $q_1+q_2+q_3=r$, and $P_A\in\mathbb{R}^{m\times m}$, $P_L\in\mathbb{R}^{p\times p}$ are orthogonal, $X\in\mathbb{R}^{n\times n}$ is invertible, and $\Sigma_A^T\Sigma_A+\Sigma_L^T\Sigma_L=I_r$. The values of q_1, q_2 and q_3 are defined internally by the matrices A and L.

If $r=n$, then $\{A,L\}$ is called a regular matrix pair. Discussions about GSVD for regular and nonregular matrix pairs can be found in [35, 49] and [38, 41, 50], respectively. Write $C_{q_2}=\mathrm{diag}(c_{q_1+1},\ldots,c_{q_1+q_2})$ with $1>c_{q_1+1}\geq\cdots\geq c_{q_1+q_2}>0$ and $S_{q_2}=\mathrm{diag}(s_{q_1+1},\ldots,s_{q_1+q_2})$ with $0<s_{q_1+1}\leq\cdots\leq s_{q_1+q_2}<1$. Let $c_1=\cdots=c_{q_1}=1,\ c_{q_1+q_2+1}=\cdots=c_r=0$ and $s_1=\cdots=s_{q_1}=0,\ s_{q_1+q_2+1}=\cdots=s_r=1$. Write $X=(x_1,\ldots,x_n)$, $P_A=(p_{A,1},\ldots,p_{A,m})$ and $P_L=(p_{L,1},\ldots,p_{L,p})$. We call the tuple $(c_i,s_i,x_i,p_{A,i},p_{L,i})$ the i-th nontrivial GSVD components, and the i-th largest generalized singular value is $\gamma_i:=c_i/s_i$ satisfying $c_i^2+s_i^2=1$, where $1\leq i\leq r$. In this paper, we consider the nontrivial GSVD components and their computations.

Despite its remarkable capabilities, computing the GSVD poses significant challenges. Early computational approaches for the GSVD were built upon adaptations of algorithms designed for the SVD; for small-scale matrices, there are several such numerical algorithms for full GSVD computation [5, 40, 53]. Recent development on stable computation of the CS decomposition (CSD) [51] provides another alternative for small-scale GSVD computation. For large and sparse problems, obtaining the full GSVD may not be feasible, yet it is often necessary to compute only a subset of GSVD components relevant to practical applications. Typically, this refers to certain extreme GSVD components, which are those with the largest or smallest corresponding generalized singular values, or interior GSVD components, which are those whose corresponding desired generalized singular values are closest to a specified target.

For large-scale GSVD computation, the first step is usually tranforming it as an equivalent generalized eigendecomposition (GED) problem [4] or CSD problem. The joint bidiagonalization (JBD) method proposed by Zha [54] can be used to compute a few extreme GSVD components, which is essentially an indirect procedure for CSD of the Q-factor of a regular $\{A,L\}$ (i.e. the Q matrix in the QR factorization). This method relies on a JBD process that iteratively reduces $\{A,L\}$ to bidiagonal forms simultaneously. Jia and Li [25,26] made a detailed analysis for the numerical behavior of the JBD method and the convergence behavior for extreme GSVD components in finite precision arithmetic. They proposed the semi-orthogonalization strategy and design a partial reorthogonalization procedure to maintain regular convergence of the computed approximate GSVD components. Subsequently, Li [33] analyzed the influence of computational errors resulting from inaccurate inner iterations in JBD on the convergence and final accuracy of computed GSVD components and proposed a modified version of the JBD method. Recently, Alvarruiz et al. [2] developed a thick restart technique for JBD to compute a partial GSVD, enabling the storage and computation cost further saved. On the other hand, the Jacobi–Davidson type algorithms [21] are capable of computing a few interior GSVD components. A representative work is the Jacobi–Davidson GSVD (JDGSVD) method proposed by Hochstenbach [20], which formulates the GSVD of $\{A,L\}$ as a GED problem of an augmented symmetric matrix pair. This method is further analyzed and developed in several subsequent work; see e.g. [23, 24, 45]. As the development of contour integration technique for eigenvalue problems of finding interior eigenvalues [44,47], recently a contour integral-based algorithm has been adopted for interior GSVD components computation [36].

Apart from regarding the GSVD as an equivalent CSD or GED, there is very little work on understanding and analyzing GSVD from other perspectives. In [10], the authors studied the GSVD using a variational formulation analogous to that of the SVD, providing a new understanding of the generalized singular vectors. Recently, by treating $(A^\top, L^\top)^\top$ (more precisely, the range space of it) as a point in the real Grassmann manifold $\mathrm{Gr}(m+p, r)$ — the manifold composed of all r-dimensional subspaces of $\mathbb{R}^{m+p}$ — the authors in [15] interpreted a modified form of GSVD as a coordinate representation of $(A^\top, L^\top)^\top$. For developing practical GSVD algorithms, however, these new perspectives on GSVD are not enough. It would be beneficial to understand the GSVD from the viewpoint of SVD so that many existing algorithms for large-scale SVD are possible to be adapted for large-scale GSVD computation. One well-known result is that $\{\gamma_i\}$ are the singular values of $AL^\dagger$ if L has full column rank [54], where "$\dagger$" is the Moor-Penrose pseudoinverse. But for noninvertible and nonsquare L, generally the GSVD of $\{A, L\}$ is not related to the SVD of $AL^\dagger$; this issue becomes much more complicated for nonregular matrix pairs.

In this paper, we provide a new understanding of GSVD from the viewpoint of SVD. This new perspective relies on the theory of singular value expansion (SVE) for linear compact operators [16, §2.2], which is essentially the SVD if the compact operator is a matrix between two Euclidean spaces. Denote by $\mathcal{R}(\cdot)$ and $\mathcal{N}(\cdot)$ the range space and null space of a matrix, respectively. By defining the positive semidefinite matrix $M = A^\top A + L^\top L$, we first investigate the structure of trivial and nontrivial GSVD components x_i, showing that those trivial $\{x_i\}$ form a basis for $\mathcal{N}(M)$ and any nontrivial x_i belongs to the coset $\bar{x}_i + \mathcal{N}(M)$, where $\bar{x}_i \in \mathcal{R}(M)$ is a nontrivial GSVD component. Then we consider the nontrivial $x_i \in \mathcal{R}(M)$ and other corresponding GSVD components. By introducing a linear operator $\mathcal{A}$ induced by $\{A, L\}$ between two finite-dimensional Hilbert spaces, where a non-Euclidean inner-product is used, we show that the SVE of $\mathcal{A}$ is just the nontrivial GSVD components of A, i.e. the first decomposition in (1.1a). Similarly, we introduce a linear operator $\mathcal{L}$ induced by $\{A, L\}$ and show that the SVE of $\mathcal{L}$ is just the nontrivial GSVD components of L. This result reveals that the nontrivial part of the GVSD of $\{A, L\}$ is nothing but the SVEs of the two linear operators induced by $\{A, L\}$. Combined with the trivial GSVD components $\{x_i\}$, it completely characterizes the structure of GSVD for any matrix pair with the same number of columns.

As a direct application of the above result, we propose a new iterative method for computing several extreme nontrivial GSVD components of $\{A, L\}$. This iterative method is a natural extension of the Golub-Kahan bidiagonalization (GKB) method for large-scale SVD computation [18], which iteratively reduces a matrix to a bidiagonal form by a Lanczos-type iterative process. There are several variants and extensions for the standard GKB of a single matrix, which are proposed to solve large-scale generalized least squares problems [3, 6], saddle point problems [14], or regularization of inverse problems [11, 12, 31, 32, 34]; most of them are named the *generalized Golub-Kahan bidiagonalization* (gGKB) while some have other different names. As a natural analogy of the standard GKB for SVD computation, we develop an operator-type GKB for linear operators $\mathcal{A}$ and $\mathcal{L}$ to approximate their SVE components, which is also named the gGKB process. We derive matrix-form recursive relations for this operator-type GKB so that it can be used in practical computations. Moreover, this approach offers a unified and general treatment for extending the standard GKB, which can be used to derive nearly all of the aforementioned gGKB processes.

Using the GSVD of $\{A, L\}$, we study several basic properties of the proposed gGKB process. Due to the correspondence of GSVD and SVE, the gGKB method can approximate well the extreme nontrivial GSVD components of $\{A, L\}$, resulting in the gGKB_GSVD algorithm. We derive a relative residual norm and its sharp upper bound for the computed nontrivial GSVD components, which is a good measure of the approximating accuracy and can be used in a stopping criterion for practical computations. Several preliminary results about the convergence and accuracy of gGKB_GSVD in exact arithmetic are provided, showing the effectiveness of this method.

The paper is organized as follows. In Section 2, we review several basic properties of the GSVD. In Section 3, we introduce two linear operators induced by $\{A, L\}$ to characterize the structure of GSVD by the SVEs of them. In Section 4 we propose the new gGKB process and study its basic properties. In Section 5, we propose the gGKB_GSVD algorithm for computing nontrivial GSVD components. Numerical experiments are presented in Section 6, and concluding remarks follow in Section 7.

2. GSVD, SVD and Golub-Kahan bidiagonalization. We review several basic properties of the GSVD of $\{A, L\}$ presented in Theorem 1.1. The nontrivial generalized singular values of $\{A, L\}$ in descending order are

$$\underbrace{\infty, \ldots, \infty}_{q_1}, \ \underbrace{c_{q_1+1}/s_{q_1+1}, \ldots, c_{q_1+q_2}/s_{q_1+q_2}}_{q_2}, \ \underbrace{0, \ldots, 0}_{q_3}. \tag{2.1}$$

We remark that the three numbers q_1, q_2, q_3 are uniquely determined by the property of $\{A, L\}$, and some of them may be zero in certain instances. The nontrivial GSVD components are linked by the vector-form relations

$$\begin{cases} Ax_i = c_i p_{A,i} \\ Lx_i = s_i p_{L,i} \\ s_i A^T p_{A,i} = c_i L^T p_{L,i} \end{cases} \tag{2.2}$$

for $i = 1, \ldots, r$. For those trivial GSVD components, it holds that

$$Ax_i = \mathbf{0}, \quad Lx_i = \mathbf{0}, \quad A^T p_{A,i} = \mathbf{0}, \quad L^T p_{L,i} = \mathbf{0} \tag{2.3}$$

for $i = r+1, \ldots, n$. The following result describes the structure of trivial and nontrivial GSVD components $\{x_i\}$.

PROPOSITION 2.1. *Let $M = A^\top A + L^\top L$ and partition X as $\begin{pmatrix} \underset{r}{X_1} & \underset{n-r}{X_2} \end{pmatrix}$. Then $\mathcal{R}(X_2) = \mathcal{N}(M)$. Moreover, let*

$$\bar{X} = (\bar{X}_1 \ \ X_2), \quad \bar{X}_1 = (\mathcal{P}_{\mathcal{R}(M)} x_1, \ldots, \mathcal{P}_{\mathcal{R}(M)} x_r). \tag{2.4}$$

Then it holds

$$A = P_A C_A \bar{X}^{-1}, \quad L = P_L S_L \bar{X}^{-1}. \tag{2.5}$$

Proof. It is clear that $\mathcal{N}(M) = \mathcal{N}(A) \cap \mathcal{N}(L)$, and using th GSVD of $\{A, L\}$, it is easy to obtain that $\mathcal{N}(A) \cap \mathcal{N}(L) = \mathcal{R}(X_2)$. Thus, we have $\mathcal{R}(X_2) = \mathcal{N}(M)$. Using the relation that $\mathcal{P}_{\mathcal{R}(M)} x_i = x_i - \mathcal{P}_{\mathcal{N}(M)} x_i$, for $1 \le i \le r$ we have

$$A\mathcal{P}_{\mathcal{R}(M)} x_i = Ax_i - A\mathcal{P}_{\mathcal{N}(M)} x_i = Ax_i, \quad L\mathcal{P}_{\mathcal{R}(M)} x_i = Lx_i - L\mathcal{P}_{\mathcal{N}(M)} x_i = Lx_i.$$

Using the above two relations, it is easy to verify (2.5). ∎

Since $\dim(\mathcal{N}(M)) = n - r = \text{rank}(\{x_i\}_{i=r+1}^n)$, it follows that $\{x_i\}_{i=r+1}^n$ forms a basis for $\mathcal{N}(M)$. Proposition 2.1 also indicates that for any x_i with $1 \le i \le r$, $\mathcal{P}_{\mathcal{R}(M)} x_i$ is also an i-th generalized singular vector of $\{A, L\}$. Therefore, any nontrivial x_i can be decomposed into two components, one being $\mathcal{P}_{\mathcal{R}(M)}x_i \in \mathcal{R}(M)$ and the other being an arbitrary vector in $\mathcal{N}(M)$, which means that any nontrivial x_i belongs to the coset $\bar{x}_i + \mathcal{N}(M)$, where $\bar{x}_i \in \mathcal{R}(M)$ is the i-th nontrivial GSVD component. In particular, we can focus on those nontrivial x_i in $\mathcal{R}(M)$, which results in the new form of GSVD (2.5). In the subsequent part, we always consider this form of GSVD by requiring

$$x_i \in \mathcal{R}(M) \quad \text{for} \quad 1 \leq i \leq r. \tag{2.6}$$

There exists a direct relationship between the SVD and GSVD for a matrix pair with a special property. If L has full column rank, it follows from (1.1) that $r = n$ and $q_1 = 0$, leading to

$$AL^{\dagger} = P_A C_A X^{-1}[(P_L S_L)X^{-1}]^{\dagger} = P_A C_A X^{-1} X (P_L S_L)^{\dagger} = P_A (C_A S_L^{\dagger}) P_L^{\top}, \tag{2.7}$$

where we have used the property that $(M_1 M_2)^{\dagger} = M_2^{\dagger} M_1^{\dagger}$ if M_1 has full column rank and M_2 has full row rank. Therefore, the SVD of $AL^{\dagger}$ is $P_A(C_A S_L^{\dagger})P_L^{\top}$ with the singular values be $\{\gamma_i\}_{i=1}^{n}$.

For the above case, one can compute the SVD of $AL^{\dagger}$ to get the GSVD components [1]. For a large-scale matrix, the GKB process is a basic routine for computing a few extreme SVD components. At each iteration, it reduces the matrix to a lower-order bidiagonal matrix and generates two Krylov subspaces. The Rayleigh-Ritz procedure is then exploited to approximate extreme SVD components using the bidiagonal matrix and Krylov subspaces [4].

In the following part of the paper, we characterize the GSVD from the viewpoint of SVD for any matrix pair with the same number of columns. Then we generalize the GKB process so that it can be used to compute extreme GSVD components.

3. Characterizing GSVD by singular value expansion of linear operators.

We first discuss linear operators between two finite-dimensional Hilbert spaces. Then we study the GSVD of $\{A, L\}$ using the singular value expansion of linear operators. Note that Subsection 3.1 is quite general without requiring $M = A^{\top}A + L^{\top}L$.

3.1. Linear operator induced by matrices.

Suppose $G \in \mathbb{R}^{m \times m}$ is symmetric positive definite. It is obviously that $\langle u, u' \rangle_G := u^{\top} G u'$ defines an inner product on $\mathbb{R}^m$ such that $(\mathbb{R}^m, \langle \cdot, \cdot \rangle_G)$ is an m-dimensional Hilbert space. On the other hand, for any symmetric positive semidefinite matrix $M \in \mathbb{R}^{n \times n}$ with $\mathrm{rank}(M) = r$, the bilinear form $\langle v, v' \rangle_M := v^{\top} M v'$ is not a well-defined inner product on $\mathbb{R}^n$ if $r < n$. In this case, we consider the inner product on the subspace $\mathcal{R}(M)$.

LEMMA 3.1. *The bilinear form $\langle v, v' \rangle_M := v^{\top} M v'$ for any $v, v' \in \mathcal{R}(M)$ is an inner product, and $(\mathcal{R}(M), \langle \cdot, \cdot \rangle_M)$ is an r-dimensional Hilbert space.*

Proof. The statement $\dim(\mathcal{R}(M)) = \mathrm{rank}(M)$ is a basic property. We need to show that $\langle \cdot, \cdot \rangle_M$ is indeed an inner product, i.e., it is a symmetric and positive bilinear form on $\mathcal{R}(M) \times \mathcal{R}(M)$. We only need to prove the positiveness. To see it, let $v \in \mathcal{R}(M)$ such that $\langle v, v \rangle_M = v^{\top} M v = 0$. It follows $v \in \mathcal{N}(M)$. Note that $\mathcal{R}(M) \cap \mathcal{N}(M) = \{\mathbf{0}\}$ since M is symmetric, which leads to $v = \mathbf{0}$. ∎

Given a matrix $A \in \mathbb{R}^{m \times n}$. Define the linear map

$$\mathcal{A} : (\mathcal{R}(M), \langle \cdot, \cdot \rangle_M) \to (\mathbb{R}^m, \langle \cdot, \cdot \rangle_G), \quad v \mapsto Av, \tag{3.1}$$

[1] For numerical computations, it is not recommended to explicitly form $AL^{\dagger}$ due to its numerical unstability, especially when L is close to column rank-deficient.

where v and Av are column vectors under the canonical bases of $\mathbb{R}^n$ and $\mathbb{R}^m$. Let $W_r \in \mathbb{R}^{n\times r}$ and $Z \in \mathbb{R}^{m\times m}$ be two matrices whose columns constitute orthonormal bases of $(\mathcal{R}(M), \langle\cdot,\cdot\rangle_M)$ and $(\mathbb{R}^m, \langle\cdot,\cdot\rangle_G)$, respectively, i.e. $W_r^\top M W_r = I_r$ and $Z^\top G Z = I_m$. Then we have the commutative diagram:

$$\begin{tikzcd}
(\mathcal{R}(M), \langle\cdot,\cdot\rangle_M) \arrow[r, "\mathcal{A}"] & (\mathbb{R}^m, \langle\cdot,\cdot\rangle_G) \\
(\mathbb{R}^r, \langle\cdot,\cdot\rangle_2) \arrow[u, "\pi_1"] \arrow[r, "{[\mathcal{A}]}"'] & (\mathbb{R}^m, \langle\cdot,\cdot\rangle_2) \arrow[u, "\pi_2"']
\end{tikzcd}, \tag{3.2}$$

where $[\mathcal{A}]$ denotes the matrix representation of $\mathcal{A}$ under bases W_r and Z, and π_1 and π_2 are two linear maps:

$$\pi_1 : (\mathbb{R}^r, \langle\cdot,\cdot\rangle_2) \to (\mathcal{R}(M), \langle\cdot,\cdot\rangle_M), \quad x \mapsto W_r x, \tag{3.3}$$

$$\pi_2 : (\mathbb{R}^m, \langle\cdot,\cdot\rangle_2) \to (\mathbb{R}^m, \langle\cdot,\cdot\rangle_G), \quad y \mapsto Zy. \tag{3.4}$$

Note that π_1 and π_2 are two isomorphism maps such that $(\mathcal{R}(M), \langle\cdot,\cdot\rangle_M) \cong (\mathbb{R}^r, \langle\cdot,\cdot\rangle_2)$ and $(\mathbb{R}^m, \langle\cdot,\cdot\rangle_G) \cong (\mathbb{R}^m, \langle\cdot,\cdot\rangle_2)$. Since $\mathcal{A}$ is a bounded linear operator, we can define its adjoint

$$\mathcal{A}^* : (\mathbb{R}^m, \langle\cdot,\cdot\rangle_G) \to (\mathcal{R}(M), \langle\cdot,\cdot\rangle_M), \quad u \mapsto \mathcal{A}^* u \tag{3.5}$$

by the relation

$$\langle \mathcal{A}v, u\rangle_G = \langle \mathcal{A}^* u, v\rangle_M \tag{3.6}$$

for any $v \in (\mathcal{R}(M), \langle\cdot,\cdot\rangle_M)$ and $u \in (\mathbb{R}^m, \langle\cdot,\cdot\rangle_G)$. We have the following corresponding commutative diagram:

$$\begin{tikzcd}
(\mathcal{R}(M), \langle\cdot,\cdot\rangle_M) & (\mathbb{R}^m, \langle\cdot,\cdot\rangle_G) \arrow[l, "\mathcal{A}^*"'] \\
(\mathbb{R}^r, \langle\cdot,\cdot\rangle_2) \arrow[u, "\pi_1"] & (\mathbb{R}^m, \langle\cdot,\cdot\rangle_2) \arrow[u, "\pi_2"'] \arrow[l, "{[\mathcal{A}^*]}"]
\end{tikzcd}, \tag{3.7}$$

where $[\mathcal{A}^*]$ is the matrix representation of $\mathcal{A}^*$ under bases W_r and Z. The following result gives the matrix-forms of $[\mathcal{A}]$ and $[\mathcal{A}^*]$.

Lemma 3.2. *The matrix representations of $\mathcal{A}$ and $\mathcal{A}^*$ under bases W_r and Z are*

$$[\mathcal{A}] = Z^{-1} A W_r, \quad [\mathcal{A}^*] = W_r^\top A^\top G Z. \tag{3.8}$$

Proof. By the commutative diagram (3.2), we have $\mathcal{A}\circ\pi_1(x) = \pi_2 \circ [\mathcal{A}](x)$ for any $x \in \mathbb{R}^r$, which is equivalent to $AW_r x = Z[\mathcal{A}]x$ Thus, we have $AW_r = Z[\mathcal{A}]$, leading to $[\mathcal{A}] = Z^{-1}AW_r$. Similarly, by the commutative diagram (3.7), we have

$$\mathcal{A}^* \circ \pi_2(y) = \pi_1 \circ [\mathcal{A}^*](y) \quad \Leftrightarrow \quad \mathcal{A}^* Z y = W_r[\mathcal{A}^*]y$$

for any $y \in \mathbb{R}^m$, which leads to

$$\mathcal{A}^* Z = W_r[\mathcal{A}^*]. \tag{3.9}$$

From the definition of $\mathcal{A}^*$, we have $\langle \mathcal{A}\circ\pi_1(x), \pi_2(y)\rangle_G = \langle \pi_1(x), \mathcal{A}^*\circ\pi_2(y)\rangle_M$ for any $x \in \mathbb{R}^r$ and $y \in \mathbb{R}^m$, which can also be written as

$$(AW_r x)^\top G Z y = (W_r x)^\top M(\mathcal{A}^* Z y) \quad \Leftrightarrow \quad x^\top W_r^\top A^\top G Z y = x^\top W_r^\top M \mathcal{A}^* Z y.$$

Thus, we have

$$W_r^\top A^\top GZ = W_r^\top M\mathcal{A}^* Z. \tag{3.10}$$

Combining (3.9) and (3.10) and using $W_r^\top MW_r = I_r$, we finally obtain

$$[\mathcal{A}^*] = W_r^\top MW_r[\mathcal{A}^*] = W_r^\top M\mathcal{A}^* Z = W_r^\top A^\top GZ.$$

The proof is completed. □

The following result will be used throughout the paper.

LEMMA 3.3. *If $\mathcal{R}(W_r) = \mathcal{R}(M)$ and $W_r^\top MW_r = I_r$, then the Moor-Penrose pseudoinverse of M can be expressed as $M^\dagger = W_rW_r^\top$.*

Proof. Let $\bar{M} = W_rW_r^\top$. We only need to verify the following four identities:

$$\begin{aligned} M\bar{M}M &= M, \quad (M\bar{M})^\top = M\bar{M}, \\ \bar{M}M\bar{M} &= \bar{M}, \quad (\bar{M}M)^\top = \bar{M}M. \end{aligned}$$

The third identity is the easiest to verify: $\bar{M}M\bar{M} = W_rW_r^\top MW_rW_r^\top = W_rW_r^\top = \bar{M}$. Suppose the compact-form eigenvalue decomposition of M is $M = P_r\Lambda_rP_r^\top$ with $\Lambda_r = \mathrm{diag}(\lambda_1, \dots, \lambda_r)$, where $\lambda_1 \geq \cdots \geq \lambda_r > 0$ and $P_r \in \mathbb{R}^{n\times r}$ with 2-orthonormal columns. Since $\mathcal{R}(W_r) = \mathcal{R}(M) = \mathcal{R}(P_r)$, there exist $D \in \mathbb{R}^{r\times r}$ such that $W_r = P_rD$. It follows that $I_r = W_r^\top MW_r = D^\top\Lambda_rD$. Therefore, it follows that $M\bar{M}M = MP_rDD^\top P_r^\top M = P_r\Lambda_rDD^\top\Lambda_rP_r^\top = P_r\Lambda_rP_r^\top = M$, since $\Lambda_rDD^\top = D^\top\Lambda_rD = I_r$. Similarly, we have

$$\begin{aligned} M\bar{M} &= P_r\Lambda_rP_r^\top P_rDD^\top P_r^\top = P_r\Lambda_rDD^\top P_r^\top = P_rP_r^\top = (M\bar{M})^\top, \\ \bar{M}M &= P_rDD^\top P_r^\top M = P_rDD^\top\Lambda_rP_r^\top = P_rP_r^\top = (\bar{M}M)^\top. \end{aligned}$$

Now all the four identities have been verified. □

3.2. Characterizing GSVD by singular value expansion. In this subsection, we consider the simpler case that $G = I_m$, since it has direct connections with the GSVD of a matrix pair. For notational simplicity, let $\mathcal{X} = (\mathcal{R}(M), \langle\cdot,\cdot\rangle_M)$ and $\mathcal{Y} = (\mathbb{R}^m, \langle\cdot,\cdot\rangle_2)$. For the linear compact operator

$$\mathcal{A}: \mathcal{X} \to \mathcal{Y}, \quad v \mapsto Av \tag{3.11}$$

between the two Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$, where v and Av are column vectors under the canonical bases of $\mathbb{R}^n$ and $\mathbb{R}^m$, it has the singular value expansion (SVE) with finite terms; see e.g. [28, §15.4]. Here we use the terminology "SVE" instead of "SVD" to distinguish it from the SVD of a matrix. The theory of SVE for $\mathcal{A}$ states that there exist positive scalars $\sigma_1 \geq \cdots \geq \sigma_d > 0$, two orthonormal systems $\{f_i\}_{i=1}^d \subseteq \mathcal{X}$ and $\{h_i\}_{i=1}^d \subseteq \mathcal{Y}$ such that

$$\mathcal{A}f_i = \sigma_ih_i, \quad \mathcal{A}^*h_i = \sigma_if_i, \tag{3.12}$$

and any $v \in \mathcal{X}$ has the expansion

$$v = v_0 + \sum_{i=1}^d \langle v, f_i\rangle_M f_i \tag{3.13}$$

with some $v_0 \in \ker(\mathcal{A})$, and

$$\mathcal{A}v = \sum_{i=1}^{d} \sigma_i \langle v, f_i\rangle_M h_i, \tag{3.14}$$

where $d = \dim(\mathrm{im}(\mathcal{A}))$. Here we use $\ker(\cdot)$ and $\mathrm{im}(\cdot)$ to denote the kernel and image of a linear operator, respectively, to distinguish them from the null space $\mathcal{N}(\cdot)$ and range space $\mathcal{R}(\cdot)$ of a matrix.

The following result provides more details about the SVE of $\mathcal{A}$.

THEOREM 3.4. *For any $A \in \mathbb{R}^{m\times n}$ and symmetric positive semidefinite matrix $M \in \mathbb{R}^{n\times n}$ with rank r, define the linear operator $\mathcal{A}$ as (3.11). Then there exist an M-orthonormal matrix $F \in \mathbb{R}^{n\times r}$, a 2-orthonormal matrix $H \in \mathbb{R}^{m\times m}$ and a diagonal matrix $\Sigma \in \mathbb{R}^{m\times r}$, such that for any $v \in \mathcal{X}$ and $u \in \mathcal{Y}$, it holds that*

$$\mathcal{A}v = H\Sigma F^\top M v, \quad \mathcal{A}^* u = F\Sigma^\top H^\top u \tag{3.15}$$

under the canonical bases of $\mathbb{R}^n$ and $\mathbb{R}^m$.

Proof. Let $\mathcal{X}_1 = \mathrm{span}\{f_i\}_{i=1}^d$. We first prove $\mathcal{X} = \ker(\mathcal{A}) \oplus \mathcal{X}_1$. Noticing (3.13), we only need to prove $\ker(\mathcal{A}) \cap \mathcal{X}_1 = \{\mathbf{0}\}$. Let $v = \sum_{j=1}^d \mu_j f_j \in \ker(\mathcal{A}) \cap \mathcal{X}_1$. By (3.14), it follows $\mathbf{0} = \mathcal{A}v = \sum_{j=1}^n \sigma_i\mu_i h_i$, leading to $\sigma_i\mu_i = 0$ for $1 \le i \le d$. Since $\sigma_i > 0$, we have $\mu_i = 0$ for $1 \le i \le d$, thereby $v = \mathbf{0}$. Then we prove $\ker(\mathcal{A}) \perp_M \mathcal{X}_1$, where $\perp_M$ is the orthogonal relation in $\mathcal{X}$. For any $v \in \ker(\mathcal{A})$ and any f_i, by (3.12) we have $f_i = \sigma_i^{-1}\mathcal{A}^* h_i$, thereby

$$\langle v, f_i\rangle_M = \langle v, \sigma_i^{-1}\mathcal{A}^* h_i\rangle_M = \langle \mathcal{A}v, \sigma_i^{-1} h_i\rangle_2 = \langle \mathbf{0}, \sigma_i^{-1} h_i\rangle_2 = 0.$$

Note $\dim(\mathcal{X}_1) = d$. Therefore, we can find $r - d$ M-orthonormal vectors in $\ker(\mathcal{A})$ that are M-orthogonal to each f_i. Denote these vectors by $\{f_{d+1}, \dots, f_r\}$. Then $\{f_i\}_{i=1}^r$ constitute a complete orthonormal basis for $\mathcal{X}$. From (3.14) we have $\mathrm{im}(\mathcal{A}) = \mathrm{span}\{h_i\}_{i=1}^d =: \mathcal{Y}_1$. Using the relation $\ker(\mathcal{A}^*) = \mathrm{im}(\mathcal{A})^\perp = \mathcal{Y}_1^\perp$, where the orthogonality is taken in $(\mathbb{R}^m, \langle\cdot,\cdot\rangle_2)$, there exist $m - d$ 2-orthonormal $\{h_{d+1}, \dots, h_m\} \subseteq \ker(\mathcal{A}^*)$ such that $\{h_i\}_{i=1}^m$ constitute a complete orthonormal basis for $\mathcal{Y}$.

Therefore, for any $v \in \mathcal{X}$, it can be written as $v = \sum_{i=1}^r \langle v, f_i\rangle_M f_i$, and thereby

$$\mathcal{A}v = \sum_{i=1}^{r} \langle v, f_i\rangle_M \mathcal{A} f_i = \sum_{i=1}^{r} \sigma_i \langle v, f_i\rangle_M h_i,$$

where we define $\sigma_{d+1} = \cdots = \sigma_r = 0$. Similarly, for any $u \in \mathcal{Y}$ with the expansion $u = \sum_{i=1}^m \langle u, h_i\rangle_2 h_i$, it holds

$$\mathcal{A}^* u = \sum_{i=1}^{m} \langle u, h_i\rangle_M \mathcal{A}^* h_i = \sum_{i=1}^{r} \sigma_i \langle u, h_i\rangle_2 f_i.$$

Let the matrices $F = (f_1, \dots, f_r)$, $H = (h_1, \dots, h_m)$ and $\Sigma = \begin{pmatrix} \Sigma_d & \\ & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m\times r}$ with $\Sigma_d = \mathrm{diag}(\sigma_1, \dots, \sigma_d)$. Then (3.15) is just the matrix-form of the above two identities. ∎

One can verify that (3.12)–(3.14) can be derived from (3.15). Therefore, the two relations in (3.15) describe completely the SVE of $\mathcal{A}$. In the following part, we use the notation

$$\mathcal{A} \sim H\Sigma F^{\top} \tag{3.16}$$

to denote the SVE of $\mathcal{A}$. From the proof of Theorem 3.4, we have the following basic relations for the four important subspaces:

$$\ker(\mathcal{A}) = \mathrm{span}\{f_i\}_{i=d+1}^{r}, \quad \mathrm{im}(\mathcal{A}) = \mathrm{span}\{h_i\}_{i=1}^{d}, \tag{3.17a}$$

$$\ker(\mathcal{A}^*) = \mathrm{span}\{h_i\}_{i=d+1}^{m}, \quad \mathrm{im}(\mathcal{A}^*) = \mathrm{span}\{f_i\}_{i=1}^{d}. \tag{3.17b}$$

From the theory of SVE for linear compact operators, if the multiplicity of σ_i is 1, then the corresponding f_i and h_i are uniquely determined (at most differing by a sign). If the multiplicity of σ_i is $m_i > 1$, then there are m_i corresponding linear independent $\{f_i\}$ and $\{h_i\}$, respectively, which are M-orthonormal and 2-orthonormal.

Based on Theorem 3.4, now we can use the SVE of $\mathcal{A}$ to characterize the GSVD of $\{A, L\}$. Remember that we consider those $x_i \in \mathcal{R}(M)$ for $1 \leq i \leq r$.

THEOREM 3.5. *Let the GSVD of $\{A, L\}$ be (1.1) and let $\mathcal{A}$ be defined as (3.11) with $M = A^{\top}A + L^{\top}L$. Partition P_A and X as*

$$P_A = \begin{array}{c} \begin{pmatrix} P_{A1} & P_{A2} & P_{A3} \end{pmatrix} m \\ \quad q_1 \quad\; q_2 \quad\; m-q_1-q_2 \end{array}, \quad X = \begin{array}{c} \begin{pmatrix} X_1 & X_2 & X_3 & X_4 \end{pmatrix} n \\ \quad q_1 \quad q_2 \quad q_3 \quad n-r \end{array} \tag{3.18}$$

and let $\widetilde{X}_1 = (X_1 \ X_2 \ X_3)$. Then the SVE of $\mathcal{A}$ has the form

$$\mathcal{A} \sim P_A \Sigma_A \widetilde{X}_1^{\top}, \tag{3.19}$$

and it holds that

$$\ker(\mathcal{A}) = \mathcal{R}(X_3), \quad \mathrm{im}(\mathcal{A}) = \mathcal{R}((P_{A1} \ P_{A2})), \tag{3.20a}$$

$$\ker(\mathcal{A}^*) = \mathcal{R}(P_{A3}), \quad \mathrm{im}(\mathcal{A}^*) = \mathcal{R}((X_1 \ X_2)). \tag{3.20b}$$

Proof. Using the GSVD of $\{A, L\}$, we have

$$A^{\top}A + L^{\top}L = X^{-\top}\left(\begin{pmatrix} \Sigma_A^{\top}\Sigma_A & \\ & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \Sigma_L^{\top}\Sigma_L & \\ & \mathbf{0} \end{pmatrix}\right)X^{-1} = X^{-\top}\begin{pmatrix} I_r & \\ & \mathbf{0} \end{pmatrix}X^{-1},$$

which leads to $\mathrm{rank}(M) = r$ and

$$\begin{pmatrix} I_r & \\ & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \widetilde{X}_1^{\top} \\ X_4^{\top} \end{pmatrix} M \begin{pmatrix} \widetilde{X}_1 & X_4 \end{pmatrix} = \begin{pmatrix} \widetilde{X}_1^{\top}M\widetilde{X}_1 & \widetilde{X}_1^{\top}MX_4 \\ X_4^{\top}M\widetilde{X}_1 & X_4^{\top}MX_4 \end{pmatrix}.$$

Therefore, we have $\widetilde{X}_1^{\top}M\widetilde{X}_1 = I_r$. Note that $\mathcal{R}(\widetilde{X}_1) \subseteq \mathcal{R}(M)$. It follows that $\widetilde{X}_1$ is an M-orthonormal basis of $(\mathcal{R}(M), \langle\cdot,\cdot\rangle_M)$, thereby we obtain from Lemma 3.3 that $M^{\dagger} = \widetilde{X}_1\widetilde{X}_1^{\top}$. Notice from (1.1) that

$$A(\widetilde{X}_1 \ X_4) = P_A(\Sigma_A \ \mathbf{0}) \Rightarrow A\widetilde{X}_1 = P_A\Sigma_A.$$

Thus, we have $AM^{\dagger} = A\widetilde{X}_1\widetilde{X}_1^{\top} = P_A\Sigma_A\widetilde{X}_1^{\top}$. For any $v \in (\mathcal{R}(M), \langle\cdot,\cdot\rangle_M)$, it holds

$$\mathcal{A}v = \mathcal{A}\mathcal{P}_{\mathcal{R}(M)}v = AM^{\dagger}Mv = P_A\Sigma_A\widetilde{X}_1^{\top}Mv.$$

Using the commutative diagram (3.7) and noticing $G = Z = I_m$ for the current case, we have for any $u \in (\mathbb{R}^m, \langle\cdot,\cdot\rangle_2)$ that

$$\mathcal{A}^* u = \pi_1 \circ [\mathcal{A}^*](u) = \widetilde{X}_1(\widetilde{X}_1^\top A^\top)u = M^\dagger A^\top u = (AM^\dagger)^\top u = \widetilde{X}_1 \Sigma_A^\top P_A^\top u, \tag{3.21}$$

where we have used $[\mathcal{A}^*] = \widetilde{X}_1^\top A^\top$ by Lemma 3.2. This proves that the SVE of $\mathcal{A}$ has the form $P_A \Sigma_A \widetilde{X}_1^\top$.

From the SVE of $\mathcal{A}$ we have $\dim(\mathrm{im}(\mathcal{A})) = q_1 + q_2$. Using the relations (3.17), it follows that $\mathrm{im}(\mathcal{A}) = \mathcal{R}((P_{A1}\ P_{A2}))$. Since P_A is a 2-orthogonal matrix, we then have $\ker(\mathcal{A}^*) = \mathcal{R}(P_{A3})$. The other two relations can also be verified easily. ∎

Corresponding to Theorem 3.5, we have the following result.

THEOREM 3.6. *Define $\mathcal{L}$ as*

$$\mathcal{L} : (\mathcal{R}(M), \langle\cdot,\cdot\rangle_M) \to (\mathbb{R}^p, \langle\cdot,\cdot\rangle_2), \quad v \mapsto Lv, \tag{3.22}$$

where v and Lv are column vectors under the canonical bases of $\mathbb{R}^n$ and $\mathbb{R}^p$. Partition P_L as

$$P_L = \begin{array}{c} \begin{pmatrix} P_{L1} & P_{L2} & P_{L3} \end{pmatrix} p. \\ \quad p-q_2-q_3 \quad\ q_2 \quad\ q_3 \end{array} \tag{3.23}$$

Then the SVE of $\mathcal{L}$ has the form

$$\mathcal{L} \sim P_L \Sigma_L \widetilde{X}_1^\top, \tag{3.24}$$

and it holds that

$$\ker(\mathcal{L}) = \mathcal{R}(X_1), \qquad \mathrm{im}(\mathcal{L}) = \mathcal{R}((P_{L2}\ P_{L3})), \tag{3.25a}$$

$$\ker(\mathcal{L}^*) = \mathcal{R}(P_{L1}), \qquad \mathrm{im}(\mathcal{L}^*) = \mathcal{R}((X_2\ X_3)). \tag{3.25b}$$

Proposition 2.1 together with Theorem 3.5 and Theorem 3.6 characterizes completely the GSVD of $\{A, L\}$ based on the SVEs of linear operators $\mathcal{A}$ and $\mathcal{L}$. Particularly, they show that the nontrivial part $\widetilde{X}_1$ is the common right SVE components of $\mathcal{A}$ and $\mathcal{L}$, while P_A and P_L are the left SVE components of $\mathcal{A}$ and $\mathcal{L}$, respectively. Moreover, the relations (3.20) and (3.25) use the image spaces and kernel spaces of $\mathcal{A}$, $\mathcal{A}^*$ and $\mathcal{L}$, $\mathcal{L}^*$ to describe the structure of each GSVD blocks and give a new explanation of the three numbers q_1, q_2 and q_3 in (1.1).

Based on the SVE characterization of GSVD, we can expect to modify those algorithms for large-scale SVD computation for large-scale GSVD computation. To compute nontrivial extreme GSVD components, we generalize the standard GKB process from the viewpoint of linear operators.

4. Generalizing the Golub-Kahan bidiagonalization. In this section, the generalization of GKB is quite general without requiring $M = A^\top A + L^\top L$, and we follow the notations and assumptions in Subsection 3.1. For the linear operator in (3.1), the iterative process of GKB can be described as follows; see [9] for discussions about GKB for linear compact operators. Choosing a nonzero vector $b \in (\mathbb{R}^m, \langle\cdot,\cdot\rangle_G)$, the basis recursive relations are

$$\begin{cases} \beta_1 u_1 = b, \\ \alpha_i v_i = \mathcal{A}^* u_i - \beta_i v_{i-1}, \\ \beta_{i+1} u_{i+1} = \mathcal{A} v_i - \alpha_i u_i, \end{cases} \tag{4.1}$$

where $u_i \in (\mathbb{R}^m, \langle\cdot,\cdot\rangle_G)$ and $v_i \in (\mathcal{R}(M), \langle\cdot,\cdot\rangle_M)$, and α_i and β_i are positive scalars such that $\|v_i\|_M = \|u_i\|_G = 1$. Note that $v_0 := \mathbf{0}$ for the initial step.

For the purpose of practical computation, we need to derive a matrix-form expression of the recursive relations. Using the isomorphisms π_1 and π_2, denote u_i and v_i by $u_i = Zy_i$ and $v_i = W_r x_i$ with $y_i \in \mathbb{R}^m$ and $x_i \in \mathbb{R}^r$ for any $i \geq 1$. Then we have

$$\begin{aligned}\mathcal{A}v_i &= \mathcal{A}\circ\pi_1(x_i) = \pi_2\circ[\mathcal{A}]x_i = ZZ^{-1}AW_r x_i = Av_i,\\ \mathcal{A}^* u_i &= \mathcal{A}^*\circ\pi_2(y_i) = \pi_1\circ[\mathcal{A}^*]y_i = W_r W_r^\top A^\top GZy_i = M^\dagger A^\top G u_i.\end{aligned}$$

Therefore, the last two recursions in (4.1) can be written in the matrix-vector forms

$$\begin{cases}\alpha_i v_i = M^\dagger A^\top G u_i - \beta_i v_{i-1}\\ \beta_{i+1}u_{i+1} = Av_i - \alpha_i u_i.\end{cases} \tag{4.2}$$

Using (4.2), the GKB of $\mathcal{A}$ can be proceeded step by step. We name the above iterative process the *generalized Golub-Kahan bidiagonalization* (gGKB). The pseudocode of gGKB is shown in Algorithm 4.1.

Remark 4.1. *If G is also positive semidefinite, define the linear operator $\mathcal{A}$: $(\mathcal{R}(M), \langle\cdot,\cdot\rangle_M) \to (\mathcal{R}(G), \langle\cdot,\cdot\rangle_G)$ be $v \mapsto Av$. In this case, a similar gGKB process can be obtained. A slight difference is that the initial vector should satisfy $b \in \mathcal{R}(G)$.*

Algorithm 4.1 Generalized Golub-Kahan bidiagonalization (gGKB)

Input: $A \in \mathbb{R}^{m\times n}$, $M \in \mathbb{R}^{n\times n}$, $G \in \mathbb{R}^{m\times m}$, $b \in \mathbb{R}^m$

1: Initialize: let $\beta_1 = \|b\|_G$, $u_1 = b/\beta_1$,
2: Compute $\bar{s} = A^\top G u_1$, $s = M^\dagger \bar{s}$,
3: $\alpha_1 = \|s\|_M$, $v_1 = s/\alpha_1$
4: **for** $i = 1, 2, \ldots, k$, **do**
5: $\quad q = Av_i - \alpha_i u_i$,
6: $\quad \beta_{i+1} = \|q\|_G$, $u_{i+1} = q/\beta_{i+1}$;
7: $\quad \bar{s} = A^\top G u_{i+1}$, $s = M^\dagger \bar{s} - \beta_{i+1} v_i$
8: $\quad \alpha_{i+1} = \|s\|_M$, $v_{i+1} = s/\alpha_{i+1}$
9: **end for**

Output: $\{\alpha_i, \beta_i\}_{i=1}^{k+1}$, $\{u_i, v_i\}_{i=1}^{k+1}$

For large-scale matrices, we can not directly compute $M^\dagger$. In this case, using the relation

$$M^\dagger \bar{s} = \underset{s\in\mathbb{R}^n}{\operatorname{argmin}} \|Ms - \bar{s}\|_2, \tag{4.3}$$

we compute $M^\dagger \bar{s}$ by iteratively solving the above least squares problems. This can be done efficiently by using the LSQR algorithm [42]. In this case, gGKB has the nested inner-outer iteration structure.

If $G = I_m$ and $M = I_n$, the gGKB becomes the standard GKB. If $G = I_m$ and M is invertible, the gGKB is equivalent to the generalizations of GKB prosed in [3, 11, 12, 31, 32] with different forms. The following result describes the basic property of gGKB, very similar to that of the standard GKB.

Proposition 4.1. *For each v_i generated by gGKB, it holds that $v_i \in \mathcal{R}(M)$. The group of vectors $\{v_i\}_{i=1}^k$ is an M-orthonormal basis of the Krylov subspace*

$$\mathcal{K}_k(M^\dagger A^\top GA, M^\dagger A^\top Gb) = \operatorname{span}\{(M^\dagger A^\top GA)^i M^\dagger A^\top Gb\}_{i=0}^{k-1}, \tag{4.4}$$

and $\{u_i\}_{i=1}^k$ *is a* G*-orthonormal basis of the Krylov subspace*

$$\mathcal{K}_k(AM^\dagger A^\top G, b) = \mathrm{span}\{(AM^\dagger A^\top G)^i b\}_{i=0}^{k-1}. \tag{4.5}$$

Proof. To get a better understanding of gGKB, we give two proofs.

The first proof. We prove $v_i \in \mathcal{R}(M)$ by mathematical induction. First note $\mathcal{R}(M^\dagger) = \mathcal{R}(M)$ for the symmetric M. For $i = 1$, we have $\alpha_1 = M^\dagger A^\top G u_1 \in \mathcal{R}(M)$. Suppose $v_i \in \mathcal{R}(M)$ for $i \geq 1$. From the recursion (4.2) we obtain

$$\alpha_{i+1} v_{i+1} = M^\dagger A^\top G u_i - \beta_{i+1} v_i \in \mathcal{R}(M),$$

leading to $v_{i+1} \in \mathcal{R}(M)$. To prove the second property, we exploit the theory about the GKB of $\mathcal{A} : (\mathcal{R}(M), \langle\cdot,\cdot\rangle_M) \to (\mathbb{R}^m, \langle\cdot,\cdot\rangle_G)$ with starting vector b, which states that $\{v_i\}_{i=1}^k$ and $\{u_i\}_{i=1}^k$ are M-orthonormal basis and G-orthonormal basis of the two Krylov subspaces

$$\mathcal{K}_k(\mathcal{A}^*\mathcal{A}, \mathcal{A}^* b) := \mathrm{span}\{(\mathcal{A}^*\mathcal{A})^i \mathcal{A}^* b\}_{i=0}^{k-1},$$
$$\mathcal{K}_k(\mathcal{A}\mathcal{A}^*, b) := \mathrm{span}\{(\mathcal{A}\mathcal{A}^*)^i b\}_{i=0}^{k-1},$$

respectively. For any $u \in \mathbb{R}^m = \pi_2(y) = Zy$, from the commutative diagram (3.7) and using Lemma 3.2 and Lemma 3.3, we obtain

$$\mathcal{A}^* u = \mathcal{A}^* \circ \pi_2(y) = \pi_1 \circ [\mathcal{A}^*](y) = W_r(W_r^\top A^\top G Z) y = M^\dagger A^\top G u.$$

Therefore, we have

$$(\mathcal{A}^*\mathcal{A})^i \mathcal{A}^* b = (M^\dagger A^\top G A)^i M^\dagger A^\top G b, \quad (\mathcal{A}\mathcal{A}^*)^i b = (AM^\dagger A^\top G)^i b.$$

The second proof. Using the coordinates of u_i and v_i under bases W_r and Z, we can write the last two relations in (4.1) as

$$\alpha_i W_r x_i = \mathcal{A}^* Z y_i - \beta_i W_r x_{i-1}, \quad \beta_{i+1} Z y_{i+1} = \mathcal{A} W_r x_i - \alpha_i Z y_i,$$

where $v_i = W_r x_i$ and $u_i = Z y_i$. Note that $Z^\top G Z = I_m$ implies $Z^{-1} = Z^\top G$. Letting $\bar{b} = Z^{-1} b$, multiplying from left the first and second relations by $W_r^\top M$ and Z^{-1}, and using (3.10), we obtain

$$\begin{cases} \beta_1 y_1 = \bar{b}_1 \\ \alpha_i x_i = W_r^\top A^\top G Z y_i - \beta_i x_{i-1} \\ \beta_{i+1} y_{i+1} = Z^\top G A W_r x_i - \alpha_i y_i. \end{cases} \tag{4.6}$$

Since $G^\top = G$, if follows that (4.6) is the standard GKB of matrix $Z^\top G A W_r$ with starting vector $\bar{b}$. Therefore, $\{x_i\}_{i=1}^k$ and $\{y_i\}_{i=1}^k$ are 2-orthonormal bases of the two Krylov subspaces

$$\mathrm{span}\{((Z^\top GAW_r)^\top Z^\top GAW_r)^i (Z^\top GAW_r)^\top \bar{b}\}_{i=0}^{k-1},$$
$$\mathrm{span}\{(Z^\top GAW_r (Z^\top GAW_r)^\top)^i \bar{b}\}_{i=0}^{k-1},$$

respectively. Note that

$$\begin{aligned} &W_r((Z^\top GAW_r)^\top Z^\top GAW_r)^i (Z^\top GAW_r)^\top \bar{b} \\ &= W_r(W_r^\top A^\top G Z Z^\top G A W_r)^i W_r^\top A^\top G Z \bar{b} = W_r(W_r^\top A^\top G A W_r)^i W_r^\top A^\top G b \\ &= (W_r W_r^\top A^T G A)^i W_r W_r^\top A^\top G b = (M^\dagger A^\top G A)^i M^\dagger A^\top G b. \end{aligned}$$

We have $v_i = W_r x_i \in \mathcal{R}(M)$ and obtain (4.4). Similarly, we can obtain (4.5). ∎

It is easy to verify that (4.6) is equivalent to (4.1). Note from Lemma 3.2 that $[\mathcal{A}^*] = [\mathcal{A}]^\top$ since $Z^{-1} = Z^\top G$. Therefore, the matrix representations of $\mathcal{A}$ and $\mathcal{A}^*$ are $Z^\top GAW_r \in \mathbb{R}^{m\times r}$ and $(Z^\top GAW_r)^\top$, respectively, which maps a coordinate vector from $\mathbb{R}^r$ to $\mathbb{R}^m$. In this sense, we can say that the recursive relation (4.6) is the coordinate representation for the gGKB of $\mathcal{A}$ under bases W_r and Z.

From the first proof of Proposition 4.1, we have $s \in \mathcal{R}(M)$. Therefore, for each $i \geq 1$, if $s = M^\dagger A^\top Gu_i - \beta_i v_{i-1} \neq \mathbf{0}$, then $\alpha_i = \|s\|_M \neq 0$. This indicates that even if M is not positive definite, the gGKB does not terminate as long as s or $q = Av_i - \alpha_i u_i$ is nonzero. Here "terminate" means that α_i or β_i equals zero at the current step. Suppose gGKB does not terminate before the k-th iteration, i.e. $\alpha_i\beta_i \neq 0$ for $1 \leq i \leq k$. Then the k-step gGKB process generates an M-orthonormal matrix $V_{k+1} = (v_1, \dots, v_{k+1}) \in \mathbb{R}^{n\times(k+1)}$ and a G-orthonormal matrix $U_{k+1} = (u_1, \dots, u_{k+1}) \in \mathbb{R}^{m\times(k+1)}$, satisfying the relations

$$\beta_1 U_{k+1} e_1 = b, \tag{4.7a}$$

$$AV_k = U_{k+1} B_k, \tag{4.7b}$$

$$M^\dagger A^\top G U_{k+1} = V_k B_k^T + \alpha_{k+1} v_{k+1} e_{k+1}^\top, \tag{4.7c}$$

where e_1 and e_{k+1} are the first and $(k+1)$-th columns of I_{k+1}, and

$$B_k = \begin{pmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \beta_3 & \ddots & \\ & & \ddots & \alpha_k \\ & & & \beta_{k+1} \end{pmatrix} \in \mathbb{R}^{(k+1)\times k} \tag{4.8}$$

has full column rank. Note that it may occurs that $\beta_{k+1} = 0$, which means gGKB terminates just at the k-th step, and in this case $v_{k+1} = \mathbf{0}$.

We emphasize that gGKB will eventually terminate at most $\min\{m, r\}$ steps, since the column rank of U_k or V_k can not exceed $\min\{m, r\}$. Using the GSVD of $\{A, L\}$, we can give a detailed description of the "terminate step" of gGKB, defined as

$$k_t = \min\{k : \alpha_{k+1}\beta_{k+1} = 0\}. \tag{4.9}$$

For any closed subspace $\mathcal{G}$ of a Hilbert space, denote by $\mathcal{P}_\mathcal{G}$ the projection operator onto $\mathcal{G}$. We have the following result.

THEOREM 4.2. *Define the linear operator $\mathcal{A}$ as (3.11) with $M = A^\top A + L^\top L$, where the GSVD of $\{A, L\}$ is as (1.1). Suppose there are l distinct positive c_i in Σ_A with l subspaces $\mathcal{G}_1, \dots, \mathcal{G}_l$ spanned by the corresponding $p_{A,i}$. Then k_t equals to the number of nonzero elements in $\{\mathcal{P}_{\mathcal{G}_1} b, \dots, \mathcal{P}_{\mathcal{G}_l} b\}$.*

The following lemma is needed to prove this theorem.

LEMMA 4.3. *For any square matrix C and a vector v, define the degree of v with respect to C as*

$$\deg_C(v) = \min\{k : \exists\ p \in \mathcal{P}_k \text{ s.t. } p(C)v = \mathbf{0}\},$$

where $\mathcal{P}_k$ is the set of all polynomials with degrees not bigger than k. Then we have

$$\deg_{AM^\dagger A^\top}(b) = \deg_{M^\dagger A^\top A}(M^\dagger A^\top b) = l. \tag{4.10}$$

Proof. First notice that $\deg_C(v)$ is nothing but the maximum rank of $\{C^i v\}_{i=0}^k$ with respect to $k \geq 0$. By Theorem 3.5, we have $AM^\dagger A^\top = A\widetilde{X}_1\widetilde{X}_1^\top A^\top$. Using the relation $A\widetilde{X}_1 = P_A\Sigma_A$, we have $AM^\dagger A^\top = P_A(\Sigma_A\Sigma_A^\top)P_A^\top$, which is the eigenvalue decomposition of $AM^\dagger A^\top$. Thus, the positive eigenvalues of $AM^\dagger A^\top$ are $1, c_{q_1+1}^2, \dots, c_{q_1+q_2}^2$ with the corresponding eigenvectors be the columns of $(P_{A1}\ P_{A2})$, and the corresponding eigenspaces are subspaces $\mathcal{G}_1, \dots, \mathcal{G}_l$. Denote the l distinct positive eigenvalues by $\lambda_1, \dots, \lambda_l$ and let G_i be those matrices with 2-orthonormal columns spanning $\mathcal{G}_i$ for $1 \leq i \leq l$. Then we can write the eigenvalue decomposition of $AM^\dagger A^\top$ as $AM^\dagger A^\top = \sum_{i=1}^l \lambda_i G_i G_i^\top$, and we have $\mathcal{P}_{\mathcal{G}_i} = G_i G_i^\top$. For each $j \geq 0$, it follows that

$$w_j := (AM^\dagger A^\top)^j b = \sum_{i=1}^l (\lambda_i G_i G_i^\top)^j b = \sum_{i=1}^l \lambda_i^j G_i G_i^\top b,$$

since G_i are mutually 2-orthogonal. Without loss of generality, suppose there are s nonzero elements in $\{\mathcal{P}_{\mathcal{G}_1} b, \dots, \mathcal{P}_{\mathcal{G}_l} b\}$ and $g_i = \mathcal{P}_{\mathcal{G}_i} b / \|\mathcal{P}_{\mathcal{G}_i} b\|_2 \neq \mathbf{0}$ for $1 \leq i \leq s$. Then $\{g_i\}_{i=1}^s$ are mutually 2-orthogonal, and

$$w_j = \sum_{i=1}^s \lambda_i^j g_i \|\mathcal{P}_{\mathcal{G}_i} b\|_2 = \sum_{i=1}^s \lambda_i^j g_i (g_i^\top b),$$

since

$$g_i^\top b = (G_i G_i^\top b)^\top b / \|\mathcal{P}_{\mathcal{G}_i} b\|_2 = \|G_i^\top b\|_2^2 / \|G_i^\top b\|_2 = \|\mathcal{P}_{\mathcal{G}_i} b\|_2.$$

Thus, the rank of $\{w_j\}_{j=0}^k$ is at most s, leading to $\deg_{AM^\dagger A^\top}(b) \leq s$. On the other hand, for $1 \leq k \leq s$, by setting $\bar{w}_j := g_j(g_j^\top b)$, we have $(w_1 \dots, w_k) = (\bar{w}_1, \dots, \bar{w}_s) T_k$, where

$$T_k = \begin{pmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{k-1} \\ 1 & \lambda_2 & \cdots & \lambda_2^{k-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & \lambda_s & \cdots & \lambda_s^{k-1} \end{pmatrix} =: \begin{array}{cc} \begin{pmatrix} T_{k1} \\ T_{k2} \end{pmatrix} & \begin{array}{l} k \\ s-k \end{array} \\ k & \end{array} .$$

Since $\lambda_i \neq \lambda_j$ for $1 \leq i \neq j \leq k$, the Vandermonde matrix T_{k1} is invertible, thereby T_k has full column rank. Therefore, the rank of $\{w_i\}_{i=1}^k$ is k for $1 \leq k \leq s$, leading to $\deg_{AM^\dagger A^\top}(b) \geq s$. This proves $\deg_{AM^\dagger A^\top}(b) = s$.

To prove $\deg_{M^\dagger A^\top A}(M^\dagger A^\top b) = s$, it is sufficient to show

$$\operatorname{rank}\left(\{(M^\dagger A^\top A)^j M^\dagger A^\top b\}_{i=0}^k\right) = \operatorname{rank}\left(\{w_j\}_{j=0}^k\right)$$

for any $k \geq 0$. Notice that

$$(M^\dagger A^\top A)^j M^\dagger A^\top b = M^\dagger A^\top (AM^\dagger A^\top)^j b = M^\dagger A^\top w_i$$

and

$$M^\dagger A^\top = \widetilde{X}_1 (A\widetilde{X}_1)^\top = \widetilde{X}_1 \Sigma_A^\top P_A^\top = (X_1\ X_2 C_{q_2})(P_{A1}\ P_{A2})^\top.$$

Let $\tilde{w}_j = (M^\dagger A^\top A)^j M^\dagger A^\top b$. It follows that $\operatorname{rank}\left(\{\tilde{w}_j\}_{j=0}^k\right) \leq \operatorname{rank}\left(\{w_j\}_{j=0}^k\right)$. To prove the inverse inequality, suppose $\{w_j\}_{j=0}^k$ are independent. We only need to show $\{\tilde{w}_j\}_{j=0}^k$ are independent. If there exist real numbers $\mu_0, \ldots, \mu_k$ such that $\sum_{j=0}^k \mu_j \tilde{w}_j = \mathbf{0}$. Then $M^\dagger A^\top W z = \mathbf{0}$, where $W = (w_0, \ldots, w_k)$ has full column rank and $z = (\mu_0, \ldots, \mu_k)^\top$. By the expression of $M^\dagger A^\top$, it follows that $Wz \in \mathcal{N}(M^\dagger A^\top) = \mathcal{R}(P_{A3})$. On the other hand, from $AM^\dagger A^\top = P_A(\Sigma_A \Sigma_A^\top) P_A^\top$ we get $Wz \in \mathcal{R}(W) \subseteq \mathcal{R}(AM^\dagger A^\top) = \mathcal{R}((P_{A1}\ P_{A2}))$. Since $\mathcal{R}((P_{A1}\ P_{A2})) \cap \mathcal{R}(P_{A3}) = \{\mathbf{0}\}$, we obtain $Wz = \mathbf{0} \Rightarrow z = \mathbf{0}$, meaning that $\{\tilde{w}_j\}_{j=0}^k$ are independent. This completes the proof. $\square$

Proof of Theorem 4.2. Suppose gGKB terminates at the k_t-th step. By Proposition 4.1, the rank of $\{u_i\}_{i=1}^{k_t}$ is k_t, implying $k_t \leq \deg_{AM^\dagger A^\top}(b) = s$ by Lemma 4.3. Then we show $k_t \geq s$. Notice from the relations (4.1) and (4.2) that

$$\alpha_1 \beta_1 v_1 = M^\dagger A^\top b,$$
$$\alpha_{i+1}\beta_{i+1} v_{i+1} = M^\dagger A^\top A v_i - (\alpha_i^2 + \beta_{i+1}^2) v_i - \alpha_i \beta_i v_{i-1}$$

for $1 \leq i \leq k_t$, where we have used

$$\begin{aligned} M^\dagger A^\top A v_i &= \alpha_i M^\dagger A^\top u_i + \beta_{i+1} M^\dagger A^\top u_{i+1} \\ &= \alpha_i(\alpha_i v_i + \beta_i v_{i+1}) + \beta_{i+1}(\alpha_{i+1} v_{i+1} + \beta_{i+1} v_i). \end{aligned}$$

Therefore, it follows that

$$v_{i+1} = \frac{1}{\alpha_{i+1}\beta_{i+1}} \left(M^\dagger A^\top A v_i - (\alpha_i^2 + \beta_{i+1}^2) v_i - \alpha_i \beta_i v_{i-1} \right)$$

for $1 \leq i < k_t$. Combining with $v_1 = \frac{1}{\alpha_1\beta_1} M^\dagger A^\top b$, the above recursion leads to

$$v_{k+1} = \sum_{i=0}^{k} \xi_i (M^\dagger A^\top A)^i M^\dagger A^\top b, \quad \xi_k = 1/\Pi_{i=1}^{k+1} \alpha_i \beta_i \neq 0$$

for $1 \leq k < k_t$. Since $\mathbf{0} = \alpha_{k_t+1}\beta_{k_t+1} v_{k_t+1}$ is a linear combination of v_{k_t} and v_{k_t-1} with nonzero coefficients, the above identity implies that $\alpha_{k_t+1}\beta_{k_t+1} v_{k_t+1}$ must be a linear combination of $\{(M^\dagger A^\top A)^i M^\dagger A^\top b\}_{i=0}^{k_t}$ with nonzero coefficients, thereby $\{(M^\dagger A^\top A)^i M^\dagger A^\top b\}_{i=0}^{k_t}$ is linear dependent. By Lemma 4.3, it follows that $k_t \geq s$. Finally, we obtain $k_t = s$. $\square$

Just as the standard GKB can be employed to approximate extreme SVD components, we will utilize gGKB to approximate nontrivial extreme GSVD components.

5. GSVD computation by generalized Golub-Kahan bidiagonalization.

We first show that gGKB can be used to approximate the SVE components. Then we use Theorem 3.5 and Theorem 3.6 to relate these approximations to the nontrivial GSVD components.

5.1. Computing nontrivial GSVD components by gGKB.

Suppose gGKB does not terminate before the k-th step. Then the compact-form SVD of B_k can be written as

$$B_k = Y_k \Theta_k H_k^\top, \quad \Theta_k = \mathrm{diag}\left(\theta_1^{(k)}, \ldots, \theta_k^{(k)}\right), \quad \theta_i^{(k)} > \cdots > \theta_k^{(k)} > 0, \tag{5.1}$$

where $Y_k = \left(y_1^{(k)}, \ldots, y_k^{(k)}\right) \in \mathbb{R}^{(k+1)\times k}$ and $H_k = \left(h_1^{(k)}, \ldots, h_k^{(k)}\right) \in \mathbb{R}^{k \times k}$ are two 2-orthonormal matrices. The approximation to the SVE triplet $(c_i, p_{A,i}, x_i)$ of $\mathcal{A}$ is defined as $\left(\bar{c}_i^{(k)}, \bar{p}_{A,i}^{(k)}, \bar{x}_i^{(k)}\right) := \left(\theta_i^{(k)}, U_{k+1} y_i^{(k)}, V_k h_i^{(k)}\right)$. To measure the quality of this approximation, we give the following result.

THEOREM 5.1. *The approximate SVE triplet for $\mathcal{A}$ satisfies*

$$\mathcal{A}\bar{x}_i^{(k)} - \bar{c}_i^{(k)}\bar{p}_{A,i}^{(k)} = 0, \tag{5.2a}$$

$$\mathcal{A}^*\bar{p}_{A,i}^{(k)} - \bar{c}_i^{(k)}\bar{x}_i^{(k)} = \alpha_{k+1}v_{k+1}e_{k+1}^\top y_i^{(k)}. \tag{5.2b}$$

Proof. Note that $\mathcal{A}v = Av$. The first relation can be verified using (4.7b):

$$\mathcal{A}\bar{v}_i^{(k)} - \bar{\sigma}_i^{(k)}\bar{u}_i^{(k)} = AV_kh_i^{(k)} - \theta_i^{(k)}U_{k+1}y_i^{(k)} = U_{k+1}\left(B_kh_i^{(k)} - \theta_i^{(k)}y_i^{(k)}\right) = 0.$$

For the second relation, using (3.21) that $\mathcal{A}^*u = M^\dagger A^\top u$, we obtain from (4.7c) that

$$\begin{aligned}
\mathcal{A}^*\bar{p}_{A,i}^{(k)} - \bar{c}_i^{(k)}\bar{x}_i^{(k)} &= M^\dagger A^\top U_{k+1}y_i^{(k)} - \theta_i^{(k)}V_kh_i^{(k)} \\
&= \left(V_kB_k^\top + \alpha_{k+1}v_{k+1}e_{k+1}^\top\right)y_i^{(k)} - \theta_i^{(k)}V_kh_i^{(k)} \\
&= V_k(B_k^\top y_i^{(k)} - \theta_i^{(k)}h_i^{(k)}) + \alpha_{k+1}v_{k+1}e_{k+1}^\top y_i^{(k)} \\
&= \alpha_{k+1}v_{k+1}e_{k+1}^\top y_i^{(k)}.
\end{aligned}$$

The proof is completed. ∎

Therefore, the triplet $\left(\bar{\sigma}_i^{(k)}, \bar{p}_{A,i}^{(k)}, \bar{x}_i^{(k)}\right)$ can be accepted as a satisfied SVE triplet at the iteration that $\left|\alpha_{k+1}v_{k+1}e_{k+1}^\top y_i^{(k)}\right|$ is sufficiently small. Using the connection between the SVE of $\mathcal{A}$ and the GSVD of $\{A, L\}$ revealed by Theorem 3.5, the tuple $\left(\bar{c}_i^{(k)}, \bar{s}_i^{(k)}, \bar{p}_{A,i}^{(k)}, \bar{x}_i^{(k)}\right) := \left(\theta_i^{(k)}, (1-(\theta_i^{(k)})^2)^{1/2}, U_{k+1}y_i^{(k)}, V_kh_i^{(k)}\right)$ can be used as a good approximation to a GSVD component. To further measure the quality of this approximation, note from (2.2) that

$$s_i^2A^\top Ax_i = c_i^2L^\top Lx_i, \quad 1 \leq i \leq r. \tag{5.3}$$

This is a well-known basic relation for GSVD, which indicates that the nontrivial generalized eigenvalues of the generalized eigenvalue problem $A^TAx = \lambda L^TLx$ are $\{\gamma_i^2\}_{i=1}^r$ and the corresponding generalized eigenvectors are $\{x_i\}_{i=1}^r$ [19, §8.7]. Now we can give the following result.

THEOREM 5.2. *The above approximate GSVD tuple for $\{A, L\}$ satisfies*

$$(\bar{s}_i^{(k)})^2A^\top A\bar{x}_i^{(k)} - (\bar{c}_i^{(k)})^2L^\top L\bar{x}_i^{(k)} = \alpha_{k+1}\beta_{k+1}Mv_{k+1}e_k^\top h_i^{(k)}. \tag{5.4}$$

Proof. First notice from (5.3) that

$$\begin{aligned}
A^\top Ax_i &= (c_i^2+s_i^2)A^\top Ax_i = c_i^2(A^\top A + L^\top L)x_i = c_i^2Mx_i, \\
L^\top Lx_i &= (c_i^2+s_i^2)L^\top Lx_i = s_i^2(A^\top A + L^\top L)x_i = s_i^2Mx_i
\end{aligned}$$

for $1 \leq i \leq r$. Since $\widetilde{X}_1 = (x_1, \ldots, x_r)$ is an M-orthonormal basis of $(\mathcal{R}(M), \langle\cdot,\cdot\rangle_M)$, it follows that $A^\top A\mathcal{R}(M) \subseteq \mathcal{R}(M)$ and $L^\top L\mathcal{R}(M) \subseteq \mathcal{R}(M)$. Therefore, we have $A^\top A\bar{x}_i^{(k)}, L^\top L\bar{x}_i^{(k)} \in \mathcal{R}(M)$ due to $\bar{x}_i^{(k)} = V_kh_i^{(k)} \in \mathcal{R}(M)$. By Theorem 5.1, we have

$$\begin{aligned}
&M^\dagger[(\bar{s}_i^{(k)})^2A^\top A\bar{x}_i^{(k)} - (\bar{c}_i^{(k)})^2L^\top L\bar{x}_i^{(k)}] = M^\dagger[A^\top A\bar{x}_i^{(k)} - (\bar{c}_i^{(k)})^2M\bar{x}_i^{(k)}] \\
&= \theta_i^{(k)}M^\dagger A^\top U_{k+1}y_i^{(k)} - (\theta_i^{(k)})^2V_kh_i^{(k)} \\
&= \theta_i^{(k)}\left(V_kB_k^\top + \alpha_{k+1}v_{k+1}e_{k+1}^\top\right)y_i^{(k)} - (\theta_i^{(k)})^2V_kh_i^{(k)} \\
&= \alpha_{k+1}\beta_{k+1}v_{k+1}e_k^\top h_i^{(k)},
\end{aligned}$$

where we have used $B_k^\top y_i^{(k)} = \theta_i^{(k)} h_i^{(k)}$ and $B_k h_i^{(k)} = \theta_i^{(k)} y_i^{(k)}$. Multiplying the above equality by M and using $\mathcal{P}_{\mathcal{R}(M)} = MM^\dagger$, we finally obtain (5.4). ◻

Combining Theorems 5.1 and 5.2, it is more proper to use the residual norm

$$\|r_i^{(k)}\|_2 := \left(\|A\bar{x}_i^{(k)} - \bar{c}_i^{(k)}\bar{p}_{A,i}^{(k)}\|_2^2 + \|(\bar{s}_i^{(k)})^2 A^\top A\bar{x}_i^{(k)} - (\bar{c}_i^{(k)})^2 L^\top L\bar{x}_i^{(k)}\|_2^2\right)^{1/2} \tag{5.5}$$

to measure the quality of the approximate GSVD components of A. Since $\|v_{k+1}\|_M = 1$, it follows from (5.4) that

$$\|r_i^{(k)}\|_2 / \|(A^\top, L^\top)^\top\|_2 \leq \alpha_{k+1}\beta_{k+1}|e_k^\top h_i^{(k)}|, \tag{5.6}$$

because $\|M\|_2^{1/2} = \|(A^\top, L^\top)^\top\|_2$. The easily computed quantity $\alpha_{k+1}\beta_{k+1}|e_k^\top h_i^{(k)}|$ is an upper bound of the scaling-invariant relative residual norm $\|r_i^{(k)}\|_2/\|(A^\top, L^\top)^\top\|_2$, which can be used in a stopping criterion.

We present the pseudocode of the gGKB-based GSVD computation (computing the GSVD components of A) in Algorithm 5.1. We remark that in order to approximate the GSVD components of L, the gGKB of $\mathcal{L}$ should be used; the spirit is the same as that for $\mathcal{A}$ and we omit it.

Algorithm 5.1 The gGKB-based GSVD computation (gGKB_GSVD)

Input: $A \in \mathbb{R}^{m\times n}$, $L \in \mathbb{R}^{p\times n}$, tol > 0

```
Initialize: choose a nonzero b ∈ ℝ^m; form M = AᵀA + LᵀL
Compute β_1, α_1, u_1, v_1 by gGKB
for i = 1, 2, ..., k, do
    Compute β_{k+1}, α_{k+1}, u_{k+1}, v_{k+1} by gGKB; form B_k, U_{k+1} and V_k
    Compute the SVD of B_k as (5.1)
    if α_{k+1} β_{k+1} |e_kᵀ h_i^{(k)}| < tol then
        Compute (c̄_i^{(k)}, s̄_i^{(k)}, p̄_{A,i}^{(k)}, x̄_i^{(k)}) := (θ_i^{(k)}, (1 − (θ_i^{(k)})²)^{1/2}, U_{k+1} y_i^{(k)}, V_k h_i^{(k)})
    end if
end for
```

Output: Approximate GSVD components $\left(\bar{c}_i^{(k)}, \bar{s}_i^{(k)}, \bar{p}_{A,i}^{(k)}, \bar{x}_i^{(k)}\right)$

5.2. Convergence and accuracy. We provide preliminary results about the convergence and accuracy of gGKB_GSVD for GSVD computation. The following result demonstrates the good property of gGKB_GSVD at the terminate step.

THEOREM 5.3. *Following the notations and assumptions of Theorem 4.2, then at the k_t-th step, the gGKB_GSVD algorithm computes exactly k_t GSVD components corresponding to the nonzero elements in $\{\mathcal{P}_{\mathcal{G}_1}b, \ldots, \mathcal{P}_{\mathcal{G}_l}b\}$.*

Proof. By Theorems 5.1 and 5.2, at the terminate step of gGKB, it computes exact the SVE components of $\mathcal{A}$, which are also the exact GSVD components of $\{A, L\}$ by Theorem 3.5. Following the notations in the proof of Lemma 4.3, we need to prove that the s vectors $\bar{p}_{A,i}^{(s)}$ belong separately to the invariant subspaces $\mathcal{G}_1, \ldots, \mathcal{G}_s$. Since $\theta_i^{(s)} > 0$ have different values and $\mathcal{G}_i$ are mutually 2-orthogonal, these $\bar{p}_{A,i}^{(s)}$ must belong to different invariant subspaces. Therefore, we only need to prove $\mathcal{P}_{\mathcal{G}}\bar{p}_{A,i}^{(s)} = \bar{p}_{A,i}^{(s)}$ for each $1 \leq i \leq s$, where $\mathcal{G} = \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_s$. From the proof of Lemma 4.3, we have

$\bar{p}_{A,i}^{(s)} \in \mathcal{K}_s(AM^{\dagger}A^{\top}, b) = \mathrm{span}\{w_i\}_{i=0}^{s-1}$, and

$$\widetilde{W}_s := (w_0, \dots, w_{s-1}) = \widetilde{G}_s \begin{pmatrix} g_1^{\top}b & & \\ & \ddots & \\ & & g_1^{\top}b \end{pmatrix} \begin{pmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{s-1} \\ 1 & \lambda_2 & \cdots & \lambda_2^{s-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & \lambda_s & \cdots & \lambda_s^{s-1} \end{pmatrix} =: \widetilde{G}_s \Lambda_s T_s,$$

where $\widetilde{G}_s = (g_1, \dots, g_s)$. Since $g_i^{\top}b \neq 0$ and T_s is nonsingular, it follows that $\mathcal{R}(\widetilde{W}_s) = \mathcal{R}(\widetilde{G}_s)$. Thus, we can write $\bar{p}_{A,i}^{(s)}$ as $\bar{p}_{A,i}^{(s)} = \widetilde{G}_s z$ with a nonzero $z \in \mathbb{R}^s$. Now we immediately obtain

$$\mathcal{P}_{\mathcal{G}} \bar{p}_{A,i}^{(s)} = \sum_{i=1}^{s} \mathcal{P}_{\mathcal{G}_i} \bar{p}_{A,i}^{(s)} = \sum_{i=1}^{s} G_i G_i^{\top} (\widetilde{G}_s z) = (\widetilde{G}_s \widetilde{G}_s^{\top}) \widetilde{G}_s z = \bar{p}_{A,i}^{(s)},$$

which is the desired result. ∎

To investigate the convergence behavior of the approximations, we give the following result to describe the convergence speed of the Ritz values $\theta_i^{(k)}$.

THEOREM 5.4. *For any $1 \leq i \leq q_1 + q_2$, let*

$$\phi_i = \arccos \frac{|c_i p_{A,i}^{\top} b|}{\|\Sigma_A^{\top} P_A^{\top} b\|_2}. \tag{5.7}$$

Then at the k-th iteration of gGKB_GSVD, it holds for $1 \leq i \leq k$ that

$$0 \leq c_i^2 - (\theta_i^{(k)})^2 \leq (c_1^2 - c_r^2) \left(\frac{\kappa_i^{(k)} \tan \phi_i}{C_{k-i}(1 + 2\gamma_i)} \right), \tag{5.8}$$

where $C_j(t)$ is the j-th Chebyshev polynomial

$$C_j(t) = \frac{1}{2} \left[(t + \sqrt{t^2 - 1})^k + (t + \sqrt{t^2 - 1})^{-k} \right], \quad |t| \geq 1$$

and

$$\gamma_i = \frac{c_i^2 - c_{i+1}^2}{c_{i+1}^2 - c_r^2}, \quad \kappa_i^{(k)} = \prod_{j=1}^{i-1} \frac{(\theta_j^{(k)})^2 - c_r^2}{(\theta_j^{(k)})^2 - c_i^2} \ (i > 1), \quad \kappa_1^{(k)} = 1 \ (i = 1).$$

Proof. Using the relations (4.6) with $G = Z = I_m$ and $W_r = \widetilde{X}_1$, it follows that the coordinate representation of gGKB is the standard GKB of $A\widetilde{X}_1$ with starting vector b. This GKB process is equivalent to the symmetric Lanczos process of $(A\widetilde{X}_1)^{\top} A\widetilde{X}_1 \in \mathbb{R}^{r \times r}$ with starting vector $(A\widetilde{X}_1)^{\top} b$, which generates 2-orthogonal vectors $\{u_i\}_{i=1}^{k}$ and the symmetric tridiagonal matrix $B_k^{\top} B_k$; see e.g. [30]. Since $A\widetilde{X}_1 = P_A \Sigma_A$, it follows that $(A\widetilde{X}_1)^{\top} A\widetilde{X}_1 = I_r (\Sigma_A^{\top} \Sigma_A) I_r^{\top}$ is the eigenvalue decomposition of $(A\widetilde{X}_1)^{\top} A\widetilde{X}_1$, and $(A\widetilde{X}_1)^{\top} b = \Sigma_A^{\top} P_A^{\top} b$. Since the i-th eigenvector of $(A\widetilde{X}_1)^{\top} A\widetilde{X}_1$ is e_i and

$$(A\widetilde{X}_1)^{\top} b = \Sigma_A^{\top} P_A^{\top} b = \begin{pmatrix} (P_{A1}^{\top} b)^{\top} & (P_{A2}^{\top} b)^{\top} & \mathbf{0} \end{pmatrix}^{\top},$$

the angle between $(A\widetilde{X}_1)^{\top} b$ and e_i for $q_1 + q_2 + 1 \leq i \leq r$ is $\pi/2$, and for $1 \leq i \leq q_1 + q_2$ the angle is expressed as (5.7). Notice that the eigenvalues of $B_k^{\top} B_k$ are $(\theta_i^{(k)})^2$. Using the convergence theory of the symmetric Lanczos process (see e.g. [46, Theorem 6.4]), we immediately obtain (5.8). ∎

Theorem 5.4 indicates that the convergence rate of $\theta_i^{(k)}$ primarily depends on two factors: the clossness between b and the corresponding vector $p_{A,i}$ and the degree of separation of c_i from others. Therefore, usually we can expect rapid convergence to the extreme and well-separated positive c_i. Note again that the approximations will not converge to the GSVD components corresponding to those zero c_i, since the angle between $(A\widetilde{X}_1)^\top b$ and e_i is $\pi/2$ for $q_1+q_2+1 \leq i \leq r$. The convergence behavior of $\bar{p}_{A,i}^{(k)}$ and $\bar{x}_i^{(k)}$ can also be described similarly based on the convergence theory of the symmetric Lanczos process, but the mathematical expressions are more complex. Interested readers can refer to [46, §6.6]

We remark that all the aforementioned results are derived for the gGKB with exact computations., i.e. we do not take into account rounding errors and computational errors arising from iteratively solving (4.3). In the presence of rounding errors, the Lanczos-type iterative process behaves very differently from that in exact arithmetic. One well-known result is that the orthogonality of u_i and v_i will gradually lost, which leads to a delay of convergence of approximations and the appearance of spurious convergent quantities [30]. Also, the inaccurate computation of $M^\dagger\bar{s}$ may affect the final accuracy of the approximations. These issues for gGKB_GSVD will be addressed in future work. We will demonstrate several of them in the subsequent numerical experiments.

6. Experimental results. We report some experimental results to demonstrate the performance of gGKB_GSVD for computing nontrivial extreme GSVD components. All the experiments are performed in MATLAB R2023b using double precision. The codes are available at https://github.com/Machealb/gsvd_iter. For the starting vector of gGKB for A and L, we use the random vector $b = $ `randn(m,1)` and $b = $ `randn(p,1)` with random seed `rng(2024)`, respectively.

Example 1. The matrix pair $\{A, L\}$ is constructed as follows. Set $m = n = p = 1000$. Let $C_A = \mathrm{diag}(\{c_i\}_{i=1}^n)$ with $c(1) = 1, c(2) = 0.95, c(3) = 0.90$, $c(4:n-6) = $ `linspace(0.88,0.12,n-6)` and $c(n-2) = 0.1, c(n-1) = 0.05, c(n) = 0.01$, and let $S_L = \mathrm{diag}(\{s_i\}_{i=1}^n)$ with $s_i = (1-c_i^2)^{1/2}$. Then let W be an orthogonal matrix by letting $W = $ `gallery('orthog',n,2)`, and $D = $ `diag(linspace(1,100,n))`. Finally let $A = C_A W^\top D$ and $L = S_L W^\top D$. By the construction, $\{A, L\}$ is a regular matrix pair, and the i-th generalized singular value of $\{A, L\}$ are c_i/s_i, where the corresponding generalized singular vectors are the i-th columns of I_n, I_n and $D^{-1}W$.

In this experiment, we use gGKB_GSVD to compute several largest and smallest generalized singular values, and show the convergence behavior of Ritz values $\theta_i^{(k)}$ where gGKB is performed with and without reorthogonalization. This is a small-scale problem, thereby we directly compute M^{-1} for computing $M^{-1}\bar{s}$ at each iteration of gGKB. Figure 6.1 shows the convergence of the first three largest and smallest Ritz values, where in the top two subfigures the right vertical lines indicate the values of c_i. There are four findings. (1) If no reorthogonalization is used for gGKB, then as the iteration proceeds, the converged Ritz values may suddenly jump up (also may jump down for converging to those smallest c_i) to become a ghost and then converge to the next larger c_i; this phenomenon leads to the appearance of spurious copies of computed c_i. (2) If gGKB is performed with full reorthogonalization, the convergence of the Ritz values remains regular, and the first three largest and smallest Ritz values converge to the first three largest and smallest c_i, respectively. (3) The final accuracy of the approximated c_i is around $\mathcal{O}(\mathbf{u})$, where $\mathbf{u} = 2^{-53} \approx 10^{-16}$ is the roundoff unit of double precision. (4) The convergence to those largest c_i is faster than the convergence to those smallest c_i; also, the convergence to c_1, c_2 and c_n, c_{n-1} is faster

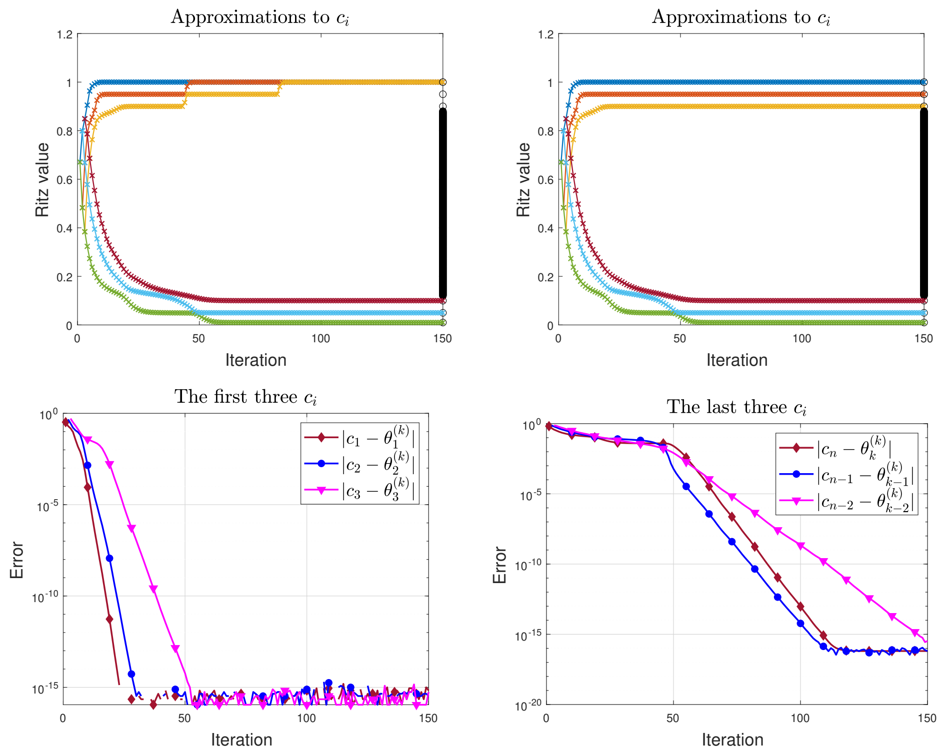

FIG. 6.1. *Convergence and accuracy of approximations to c_i. Top: convergence of Ritz values $\theta_i^{(k)}$ to largest/smallest c_i by gGKB_GSVD without reorthogonalization (left) and with full reorthogonalization (right). Bottom: error curves (full reorthogonalization).*

than that to c_3 and c_{n-2}. This is because c_1, c_2 are more well-separated from others than c_3; the same reason applies to c_n, c_{n-1} and c_{n-2}.

We also test using the gGKB of $\mathcal{L}$ to approximate several largest and smallest s_i. The convergence behavior of the Ritz values and error curves are plotted in Figure 6.2. In addition to the findings closely resembling those depicted in Figure 6.1, there are two additional insights. First, we find that the first three smallest Ritz values converge to s_2, s_3, s_4 instead of s_1, s_2, s_3. This is because $s_1 = 0$, which can not be converged upon by Ritz values, as revealed by Theorem 5.3 and Theorem 5.4. Therefore, we should use gGKB of $\mathcal{A}$ to compute the generalized singular values with value ∞. Second, we find from the bottom two subfigures that those smallest s_i can be approximated more quickly than the largest ones, due to their well-separated locations. Given that these smallest s_i correspond to those largest c_i, it is unsurprising that they exhibit similar convergence behaviors.

Example 2. The matrix A named well1850 is taken from the SuiteSparse matrix collection [13], and the matrix L is set as

$$L = \begin{pmatrix} 1.1 & -1 & & & \\ & 1.1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1.1 & -1 \end{pmatrix} \in \mathbb{R}^{(n-1)\times n}. \quad (16)$$

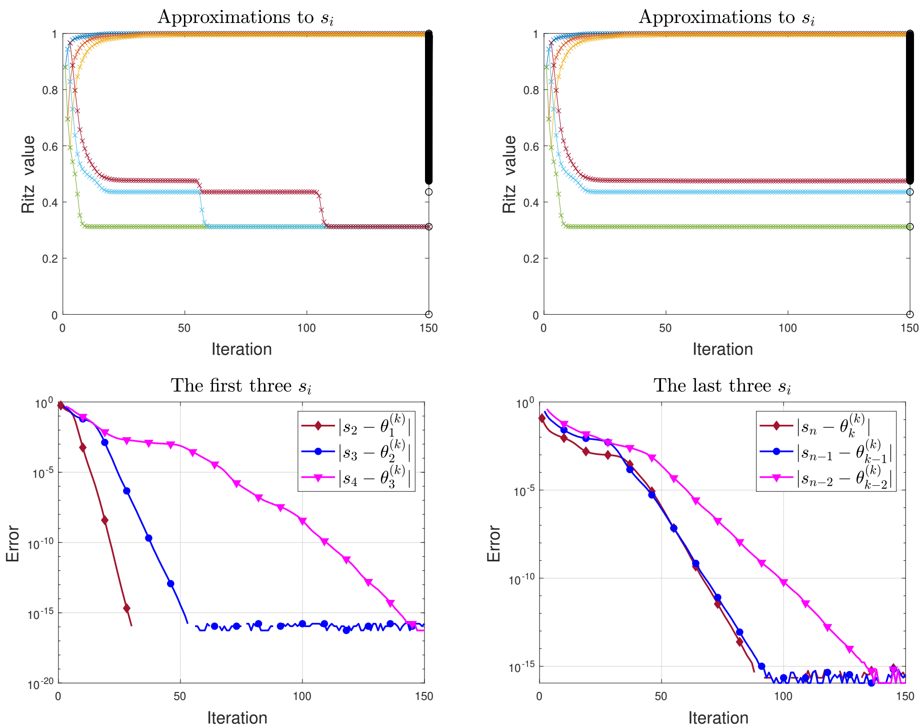

FIG. 6.2. *Convergence and accuracy of approximations to s_i. Top: convergence of Ritz values $\theta_i^{(k)}$ to largest/smallest s_i by* **gGKB_GSVD** *without reorthogonalization (left) and with full reorthogonalization (right). Bottom: error curves (full reorthogonalization).*

This is a regular matrix pair. We use the Matlab build-in function `gsvd.m` to compute the full GSVD of $\{A, L\}$ as the baseline of comparison.

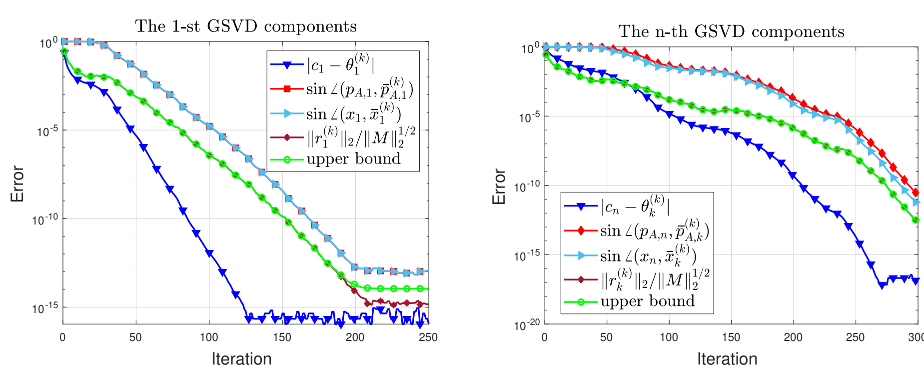

FIG. 6.3. *Error curves of the approximate GSVD components by* **gGKB_GSVD** *and relative residual norm with its upper bound. Left: approximations to the 1-st GSVD components. Right: approximations to the n-th GSVD components.*

In this experiment, we test the performance of gGKB_GSVD for computing the first and n-th GSVD components of A for a matrix pair with nonsquare matrices. We only show the results for the gGKB of $\mathcal{A}$ and omit the similar results for the gGKB of $\mathcal{L}$. Full reorthogonalization is used and M^{-1} is computed directly. The errors for the approximated generalized singular vectors are measured by $\sin\angle(x,y)$ between two vectors. We also plot the variation of the relative residual norm and its upper bound $\alpha_{k+1}\beta_{k+1}|e_k^\top h_i^{(k)}|$. Figure 6.3 shows that gGKB_GSVD can approximate very well the two group extreme GSVD components, with final accuracy around $\mathcal{O}(\mathbf{u})$. The upper bound $\alpha_{k+1}\beta_{k+1}|e_k^\top h_i^{(k)}|$ exhibits a nearly identical decreasing trend as the relative residual norm. Therefore, it is a highly suitable quantity to be employed in the stopping criterion. We also observe that the convergence to the first GSVD components is faster than the convergence to the and n-th.

Example 3. The matrix pair $\{A, L\}$ is constructed as follows. Set $m = n = p = 1000$ and set $r = 900$. Let $C_A = (\Sigma_A, \mathbf{0})$ with $C_A = \mathrm{diag}(\{c_i\}_{i=1}^r)$, where $c(1) = 0.99, c(2) = 0.98$, $c(3 : r-4) =$ `linspace(0.96,0.06,r-4)` and $c(r-1) = 0.04, c(r-1) = 0.02$, and let $S_L = (\Sigma_L, \mathbf{0})$ with $\Sigma_L = \mathrm{diag}(\{s_i\}_{i=1}^r)$ and $s_i = (1-c_i^2)^{1/2}$. Then let $W =$ `gallery('orthog',n,2)` be an orthogonal matrix and $D =$ `diag(linspace(1,10,n))`. Finally let $A = C_A W^\top D$ and $L = S_L W^\top D$. By the construction, we have $\mathrm{rank}((A^\top, L^\top)^\top) = r < n$, and the nontrivial GSVD components are c_i, s_i and the i-th columns of I_n, I_n and $D^{-1}W$ for $1 \le i \le r$. For each nontrivial x_i, we compute $\mathcal{P}_{\mathcal{R}(M)} x_i = MM^\dagger x_i$ to get the corresponding right generalized singular vector belonging to $\mathcal{R}(M)$. We use gGKB_GSVD to compute x_i that belongs to $\mathcal{R}(M)$.

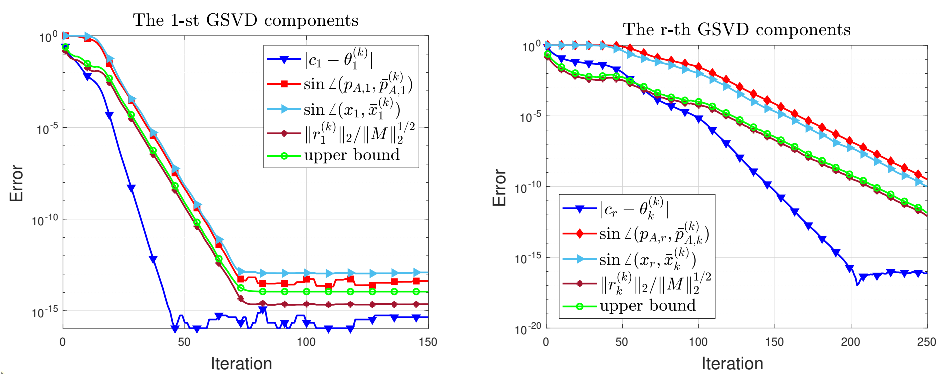

FIG. 6.4. *Error curves of the approximate GSVD components by* gGKB_GSVD *and relative residual norm with its upper bound, where* $\mathrm{rank}((A^\top, L^\top)^\top) = r < n$. *Left: approximations to the 1-st GSVD components. Right: approximations to the r-th GSVD components.*

In this experiment, we test the performance of gGKB_GSVD for computing the first and r-th GSVD components for a nonregular matrix pair. We directly compute $M^\dagger$ and use full reorthogonalization for gGKB. Figure 6.4 shows very good performance of the algorithm: (1) the two extreme GSVD components $(c_i, p_{A,i}, x_i)$ for $i = 1, r$ can be approximated with final accuracy around $\mathcal{O}(\mathbf{u})$; (2) the relative residual norm and its upper bound $\alpha_{k+1}\beta_{k+1}|e_k^\top h_i^{(k)}|$ follow nearly identical decreasing curves, with both eventually stabilizing at a level around $\mathcal{O}(\mathbf{u})$. Again, we observe that the convergence to the first GSVD components is faster than the convergence to the r-th.

Example 4. The matrix pair $\{A, L\}$ is constructed as follows. Set $m = n = p = 5000$. Let $C_A = \mathrm{diag}(\{c_i\}_{i=1}^r)$ with $c(1) = 0.99, c(2) = 0.97$, $c(3 : n-4) =$ `linspace(0.95,0.15,n-4)` and $c(n-1) = 0.1, c(n) = 0.05$. Let $S_L = \mathrm{diag}(\{s_i\}_{i=1}^n)$ with $s_i = (1-c_i^2)^{1/2}$. Then let $W =$ `gallery('orthog',n,2)` be an orthogonal matrix and $D =$ `diag(linspace(1,10,n))`. Finally let $A = C_A W^\top D$ and $L = S_L W^\top D$.

By the construction, $\{A, L\}$ is a regular matrix pair, and the i-th GSVD components are $c_i,\ s_i$ and i-th columns of I_n, I_n and $D^{-1}W$.

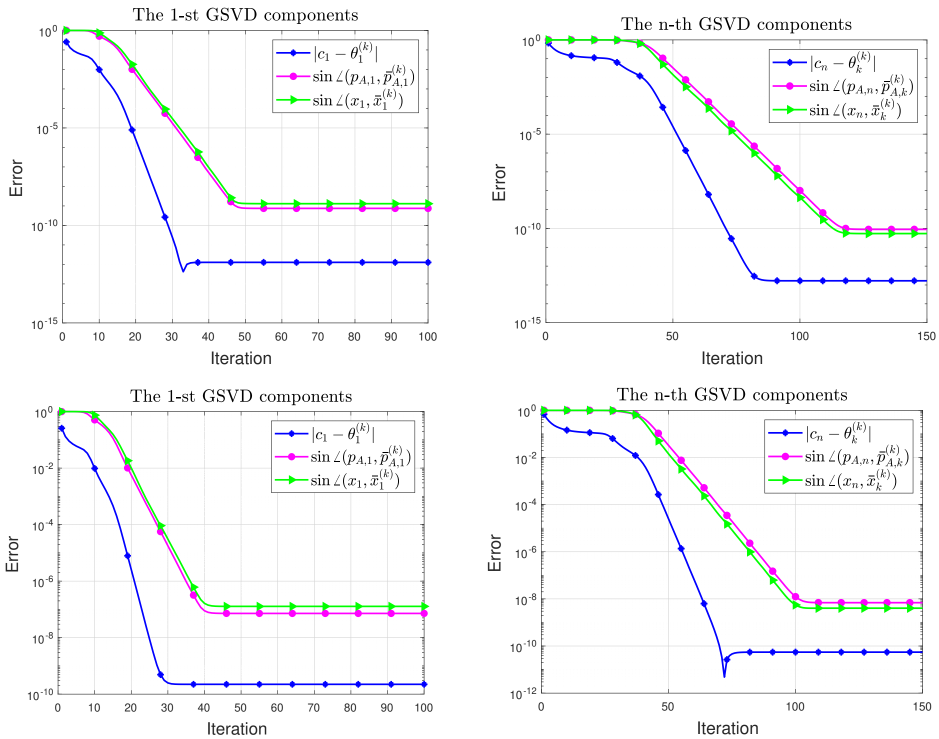

FIG. 6.5. *Accuracy of computed GSVD components by* gGKB_GSVD*, where* $s = M^{\dagger}\bar{s}$ *at each* gGKB *iteration is computed by solving* $\min_s \|Ms - \bar{s}\|_2$ *using* `lsqr.m` *with different stopping tolerance* `tol`*. Top:* `tol`$=10^{-10}$*. Top:* `tol`$=10^{-8}$*.*

We use this experiment to demonstrate the impact of inaccuracy in the computation of $M^{\dagger}\bar{s}$ on the final accurate of the approximate GSVD components. We use the Matlab build-in function `lsqr.m` to solve (4.3) iteratively with stopping tolerance `tol` $= 10^{-10}, 10^{-8}$ at each iteration of gGKB, respectively. Figure 6.5 shows the decrease of relative errors of the first and n-th approximate GSVD components with the two stopping tolerances. We observe that the computational accuracy of $M^{\dagger}\bar{s}$ significantly affects the final accuracy of both the generalized singular values and vectors. As the computational accuracy deteriorates, so does the final accuracy of the computed GSVD components. Further theoretical investigation into this issue should be conducted in future research.

7. Conclusion and outlook. Based on the theory of singular value expansion (SVE) of linear compact operators, we have provided a new understanding of the GSVD of $\{A, L\}$ with $A \in \mathbb{R}^{m\times n}$ and $L \in \mathbb{R}^{p\times n}$. By defining the positive semidefinite matrix $M = A^{\top}A + L^{\top}L$, we have shown that: (1) the trivial GSVD components $\{x_i\}$ form a basis for $\mathcal{N}(M)$ and any nontrivial x_i belongs to the coset $\bar{x}_i + \mathcal{N}(M)$, where $\bar{x}_i \in \mathcal{R}(M)$ is a nontrivial GSVD component; (2) the nontrivial GSVD components of A and L are just the SVEs of the linear operators $\mathcal{A} : (\mathcal{R}(M), \langle\cdot,\cdot\rangle_M) \to (\mathbb{R}^m, \langle\cdot,\cdot\rangle_2),\ v \mapsto Av$ and $\mathcal{L} : (\mathcal{R}(M), \langle\cdot,\cdot\rangle_M) \to (\mathbb{R}^p, \langle\cdot,\cdot\rangle_2),\ v \mapsto Lv$, respectively. As a direct application of this result, we have developed an operator-type Golub-Kahan bidiagonalization (GKB) for $\mathcal{A}$ and $\mathcal{L}$, leading to a novel generalized GKB (gGKB) process. We have used the GSVD of $\{A, L\}$ to study basic properties of gGKB and proposed the gGKB_GSVD algorithm to compute several nontrivial extreme GSVD components of large-scale matrix pairs. Preliminary results about convergence and accuracy of gGKB_GSVD for GSVD computation have been provided, and numerical experiments are presented to demonstrate the effectiveness of this method.

The idea of this paper offers potential directions for developing new algorithms for large-scale GSVD computation. Note that the SVE of $\mathcal{A}$ or $\mathcal{L}$ can be treated as a "weighted" SVD, where the weight matrix M induces a non-Euclidean inner product. Therefore, existing SVD algorithms based on Krylov subspace projection may be modified to approximate the SVE and consequently, the nontrivial GSVD components.

REFERENCES

[1] O. Alter, P. O. Brown, and D. Botstein, *Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms*, Proceedings of the National Academy of Sciences, 100 (2003), pp. 3351–3356.

[2] F. Alvarruiz, C. Campos, and J. E. Roman, *Thick-restarted joint Lanczos bidiagonalization for the GSVD*, Journal of Computational and Applied Mathematics, 440 (2024), p. 115506.

[3] M. Arioli, *Generalized Golub–Kahan bidiagonalization and stopping criteria*, SIAM Journal on Matrix Analysis and Applications, 34 (2013), pp. 571–592.

[4] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, *Templates for the solution of algebraic eigenvalue problems: a practical guide*, SIAM, 2000.

[5] Z. Bai and J. W. Demmel, *Computing the generalized singular value decomposition*, SIAM Journal on Scientific Computing, 14 (1993), pp. 1464–1486.

[6] S. J. Benbow, *Solving generalized least-squares problems with LSQR*, SIAM Journal on Matrix Analysis and Applications, 21 (1999), pp. 166–177.

[7] K. Bhuyan, S. Singh, and P. Bhuyan, *Application of generalized singular value decomposition to ionospheric tomography*, in Annales Geophysicae, vol. 22, Copernicus Publications Göttingen, Germany, 2004, pp. 3437–3444.

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

[9] N. A. Caruso and P. Novati, *Convergence analysis of LSQR for compact operator equations*, Linear Algebra Appl., 583 (2019), pp. 146–164.

[10] M. T. Chu, R. E. Funderlic, and G. H. Golub, *On a variational formulation of the generalized singular value decomposition*, SIAM Journal on Matrix Analysis and Applications, 18 (1997), pp. 1082–1092.

[11] J. Chung and A. K. Saibaba, *Generalized hybrid iterative methods for large-scale Bayesian inverse problems*, SIAM Journal on Scientific Computing, 39 (2017), pp. S24–S46.

[12] J. Chung, A. K. Saibaba, M. Brown, and E. Westman, *Efficient generalized Golub–Kahan based methods for dynamic inverse problems*, Inverse Problems, 34 (2018), p. 024005.

[13] T. A. Davis and Y. Hu, *The university of florida sparse matrix collection*, ACM Trans. Math. Software (TOMS), 38 (2011), pp. 1–25, http://www.cise.ufl.edu/research/sparse/matrices.

[14] A. Dumitrasc, C. Kruse, and U. Ruede, *Generalized Golub-Kahan bidiagonalization for nonsymmetric saddle point systems*, arXiv preprint arXiv:2310.06952, (2023).

[15] A. Edelman and Y. Wang, *The GSVD: Where are the ellipses?, matrix trigonometry, and more*, SIAM Journal on Matrix Analysis and Applications, 41 (2020), pp. 1826–1856.

[16] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, 2000.

[17] L. M. Ewerbring and F. T. Luk, *Canonical correlations and generalized SVD: applications and new algorithms*, Journal of computational and applied mathematics, 27 (1989), pp. 37–52.

[18] G. Golub and W. Kahan, *Calculating the singular values and pseudo-inverse of a matrix*, Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis, 2 (1965), pp. 205–224.

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

[20] M. Hochstenbach, *A Jacobi–Davidson type method for the generalized singular value problem*,

Linear Algebra and its Applications, 431 (2009), pp. 471–487.

[21] M. E. Hochstenbach and Y. Notay, *The Jacobi–Davidson method*, GAMM-Mitteilungen, 29 (2006), pp. 368–382.

[22] P. Howland, M. Jeon, and H. Park, *Structure preserving dimension reduction for clustered text data based on the generalized singular value decomposition*, SIAM Journal on Matrix Analysis and Applications, 25 (2003), pp. 165–179.

[23] J. Huang and Z. Jia, *On choices of formulations of computing the generalized singular value decomposition of a large matrix pair*, Numerical Algorithms, 87 (2021), pp. 689–718.

[24] J. Huang and Z. Jia, *Two harmonic Jacobi–Davidson methods for computing a partial generalized singular value decomposition of a large matrix pair*, Journal of Scientific Computing, 93 (2022), p. 41.

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

[26] Z. Jia and H. Li, *The joint bidiagonalization method for large GSVD computations in finite precision*, SIAM Journal on Matrix Analysis and Applications, 44 (2023), pp. 382–407.

[27] B. Kågström, *The generalized singular value decomposition and the general $(A-\lambda B)$-problem*, BIT Numerical Mathematics, 24 (1984), pp. 568–583.

[28] R. Kress, *Linear Integral Equations*, Springer New York, NY, 2013, https://doi.org/10.1007/978-1-4614-9593-2.

[29] S. Kuo, W. Yeih, and Y. Wu, *Applications of the generalized singular-value decomposition method on the eigenproblem using the incomplete boundary element formulation*, Journal of Sound and Vibration, 235 (2000), pp. 813–845.

[30] R. M. Larsen, *Lanczos bidiagonalization with partial reorthogonalization*, DAIMI Report Series, (1998).

[31] H. Li, *A preconditioned Krylov subspace method for linear inverse problems with general-form Tikhonov regularization*, arXiv preprint arXiv:2308.06577, (2023).

[32] H. Li, *Subspace projection regularization for large-scale Bayesian linear inverse problems*, arXiv preprint arXiv:2310.18618, (2023).

[33] H. Li, *The joint bidiagonalization of a matrix pair with inaccurate inner iterations*, SIAM Journal on Matrix Analysis and Applications, 45 (2024), pp. 232–259.

[34] H. Li, J. Feng, and F. Lu, *Scalable iterative data-adaptive RKHS regularization*, arXiv preprint arXiv:2401.00656, (2024).

[35] R.-C. Li, *Bounds on perturbations of generalized singular values and of associated subspaces*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 195–234.

[36] Y. Liu, X. Shan, and M. Shao, *A contour integral-based algorithm for computing generalized singular values*, arXiv preprint arXiv:2401.00121, (2023).

[37] K. Nakamura, K. Nakadai, and G. Ince, *Real-time super-resolution sound source localization for robots*, in 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2012, pp. 694–699.

[38] C. C. Paige, *A note on a result of Sun Ji-Guang: Sensitivity of the CS and GSV decompositions*, SIAM J. Numer. Anal., 21 (1984), pp. 186–191, https://doi.org/10.1137/0721013.

[39] C. C. Paige, *The general linear model and the generalized singular value decomposition*, Linear Algebra and its applications, 70 (1985), pp. 269–284.

[40] C. C. Paige, *Computing the generalized singular value decomposition*, SIAM Journal on Scientific and Statistical Computing, 7 (1986), pp. 1126–1146.

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

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

[43] C. H. Park and H. Park, *A relationship between linear discriminant analysis and the generalized minimum squared error solution*, SIAM Journal on Matrix Analysis and Applications, 27 (2005), pp. 474–492.

[44] P. T. Peter Tang and E. Polizzi, *FEAST as a subspace iteration eigensolver accelerated by approximate spectral projection*, SIAM Journal on Matrix Analysis and Applications, 35 (2014), pp. 354–390.

[45] A. Refahi Sheikhani and S. Kordrostami, *New iterative methods for generalized singular-value problems*, Mathematical Sciences, 11 (2017), pp. 257–265.

[46] Y. Saad, *Numerical methods for large eigenvalue problems: revised edition*, SIAM, 2011.

[47] T. Sakurai and H. Sugiura, *A projection method for generalized eigenvalue problems using numerical integration*, Journal of computational and applied mathematics, 159 (2003), pp. 119–128.

[48] J. M. Speiser and C. Van Loan, *Signal processing computations using the generalized singular*

value decomposition, in Real-Time Signal Processing VII, vol. 495, SPIE, 1984, pp. 47–57.
[49] J.-G. Sun, *Perturbation analysis for the generalized singular value problem*, SIAM J. Numer. Anal., 20 (1983), pp. 611–625.
[50] J.-g. Sun, *Perturbation theorems for generalized singular values*, J. Comput. Math., 1 (1983), pp. 233–242.
[51] B. D. Sutton, *Stable computation of the CS decomposition: Simultaneous bidiagonalization*, SIAM Journal on Matrix Analysis and Applications, 33 (2012), pp. 1–21.
[52] C. F. Van Loan, *Generalizing the singular value decomposition*, SIAM J. Numer. Anal., 13 (1976), pp. 76–83.
[53] C. F. Van Loan, *Computing the CS and the generalized singular value decompositions*, Numer. Math., 46 (1985), pp. 479–491.
[54] H. Zha, *Computing the generalized singular values/vectors of large sparse or structured matrix pairs*, Numer. Math., 72 (1996), pp. 391–417.