

## A HIGHER-ORDER GENERALIZED SINGULAR VALUE DECOMPOSITION FOR RANK-DEFICIENT MATRICES\*

IDRIS KEMPF<sup>†</sup>, PAUL J. GOULART<sup>†</sup>, AND STEPHEN R. DUNCAN<sup>†</sup>

**Abstract.** The higher-order generalized singular value decomposition (HO-GSVD) is a matrix factorization technique that extends the GSVD to  $N \geq 2$  data matrices and can be used to identify common subspaces that are shared across multiple large-scale datasets with different row dimensions. The standard HO-GSVD factors  $N$  matrices  $A_i \in \mathbb{R}^{m_i \times n}$  as  $A_i = U_i \Sigma_i V^T$  but requires that each of the matrices  $A_i$  has full column rank. We propose a modification of the HO-GSVD that extends its applicability to rank-deficient data matrices  $A_i$ . If the matrix of stacked  $A_i$  has full rank, we show that the properties of the original HO-GSVD extend to our approach. We extend the notion of common subspaces to isolated subspaces, which identify features that are unique to one  $A_i$ . We also extend our results to the higher-order cosine-sine decomposition (HO-CSD), which is closely related to the HO-GSVD. Our extension of the standard HO-GSVD allows its application to matrices with  $m_i < n$  or  $\text{rank}(A_i) < n$ , such as those encountered in bioinformatics, neuroscience, control theory, and classification problems.

**Key words.** higher-order generalized singular value decomposition, higher-order cosine-sine decomposition, diagonalization, multimodal data fusion

**MSC codes.** 65F15, 65F55

**DOI.** 10.1137/21M1443881

**1. Introduction.** The *generalized singular value decomposition* (GSVD) [27, 21] is an extension of the well-known *singular value decomposition* (SVD) to  $N = 2$  matrices. The GSVD decomposes a pair of matrices  $A_1 \in \mathbb{R}^{m_1 \times n}$  and  $A_2 \in \mathbb{R}^{m_2 \times n}$ , with  $\text{rank}([A_1^T, A_2^T]^T) = n$ , by factorizing each of the matrices as  $A_i = U_i \Sigma_i V^T$ . The matrix of right generalized singular vectors  $V \in \mathbb{R}^{n \times n}$ , with  $\det(V) \neq 0$ , is shared between the decompositions, but unlike the standard SVD is not an orthogonal matrix. The columns of the matrices  $U_i \in \mathbb{R}^{m_i \times m_i}$  are commonly referred to as *left generalized singular vectors*, each satisfying  $U_i^T U_i = I$ . The matrices  $\Sigma_i \in \mathbb{R}^{m_i \times n}$  contain the *generalized singular values*,  $\sigma_{i,1}, \dots, \sigma_{i,n}$ , on their main diagonals, with  $\sigma_{i,k} \geq 0$  and  $r_i := \text{rank}(A_i)$  nonzero  $\sigma_{i,k}$ . The generalized singular values measure the significance of the right generalized singular vectors  $v_k$  in the factorization of each  $A_i$  [23]. If  $\sigma_{1,k} = \sigma_{2,k}$ , then  $v_k$  solves the *generalized singular value problem*  $A_1^T A_1 v_k = \mu A_2^T A_2 v_k$  with  $\mu = 1$  [7, Ch. 8.7].

The *higher-order GSVD* (HO-GSVD) [23] is an extension of the GSVD to  $N \geq 3$  matrices. Given  $N$  matrices  $A_1, \dots, A_N$ , the HO-GSVD decomposes each  $A_i$  as

$$(1.1) \quad A_i = U_i \Sigma_i V^T, \quad i = 1, \dots, N,$$

where the columns of  $U_i \in \mathbb{R}^{m_i \times n}$  are referred to as *left basis vectors* and the columns of  $V \in \mathbb{R}^{n \times n}$  with  $\det(V) \neq 0$  are referred to as *right basis vectors*, and  $\Sigma_i \in \mathbb{R}^{m_i \times n} := \text{diag}(\sigma_{i,1}, \dots, \sigma_{i,n})$  contains the generalized singular values  $\sigma_{i,k} \geq 0$ . The matrix  $V$  is

\*Received by the editors September 2, 2021; accepted for publication (in revised form) by M. W. Berry February 21, 2023; published electronically July 21, 2023.

<https://doi.org/10.1137/21M1443881>

**Funding:** The research leading to these results was supported in part by the Diamond Light Source and in part by the Engineering and Physical Sciences Research Council (EPSRC) under an Industrial Cooperative Award in Science and Technology (ICASE) studentship.

<sup>†</sup>Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, UK (idris.kempf@eng.ox.ac.uk, paul.goulart@eng.ox.ac.uk, stephen.duncan@eng.ox.ac.uk).

obtained from the eigensystem  $S_\pi V = V \Sigma$ , where  $\Sigma := \text{diag}(\varsigma_1, \dots, \varsigma_n)$  and  $S_\pi$  is the arithmetic mean of all pairwise quotients  $D_{i,\pi} D_{j,\pi}^{-1}$ ,

$$(1.2) \quad S_\pi := \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N (D_{i,\pi} D_{j,\pi}^{-1} + D_{j,\pi} D_{i,\pi}^{-1}),$$

with  $D_{i,\pi}$  defined as

$$(1.3) \quad D_{i,\pi} := A_i^T A_i + \pi A^T A, \quad \pi \geq 0,$$

where  $A := [A_1^T, \dots, A_N^T]^T$ . The case  $\pi = 0$  corresponds to the original HO-GSVD framework [23].

One shortcoming of the original HO-GSVD [23] is that the arithmetic mean (1.2) is only well defined for matrices  $A_i$  that have full column rank. If  $\text{rank}(A_i) < n$  for some  $i$ , then the inverse  $(A_i^T A_i)^{-1}$  does not exist and so  $S_0$  in (1.2) is not well defined. In addition, computing (1.2) may be inaccurate when one or more of the  $A_i$  have small singular values. Provided that the matrix  $A$  of stacked  $A_i$  has full column rank, so that  $A^T A \succ 0$ , introducing the term  $\pi A^T A$  in (1.3) with parameter  $\pi > 0$  has the effect of bounding the eigenvalues of each  $D_{i,\pi}$  away from zero. For  $\pi > 0$ , the terms  $D_{i,\pi}$  are therefore guaranteed to be invertible and the HO-GSVD can be computed for  $A_i$  with arbitrary rank.

Using the factorization (1.1), the matrices  $A_i$  can be rewritten as

$$(1.4) \quad A_i = \underbrace{\sum_{k \in \mathcal{I}_N} \sigma_{i,k} u_{i,k} v_k^T}_{\text{common}} + \underbrace{\sum_{k \in \mathcal{I}_1} \sigma_{i,k} u_{i,k} v_k^T}_{\text{isolated}} + \sum_{k \in \mathcal{I}_\perp} \sigma_{i,k} u_{i,k} v_k^T,$$

where  $\mathcal{I}_N \cup \mathcal{I}_1 \cup \mathcal{I}_\perp = \{1, \dots, n\}$ ,  $\mathcal{I}_N$ , and  $\mathcal{I}_1$  and  $\mathcal{I}_\perp$  are mutually disjoint. Indices associated with  $\mathcal{I}_N$  and  $\mathcal{I}_1$  reflect “themes” shared by matrices  $A_i$ : For  $k \in \mathcal{I}_N$ ,  $\sigma_{i,k} = 1/\sqrt{N} \forall i$ , and for  $k \in \mathcal{I}_1$ ,  $\sigma_{j,k} = 1$  for some  $j$  and  $\sigma_{i,k} = 0$  for all other  $i \neq j$  (section 3). The associated right basis vectors  $v_k$  form subspaces that are referred to as *common HO-GSVD subspace* ( $k \in \mathcal{I}_N$ ) [23] or *isolated HO-GSVD subspace* ( $k \in \mathcal{I}_1$ ), which is nonempty iff one or more of the  $A_i$  is column rank deficient (Theorem 3.5). Under certain circumstances, the left basis vectors associated with the common or isolated HO-GSVD subspaces are orthogonal to the remaining left basis vectors (Theorem 3.3 and Corollary 3.6). When all  $A_i$  have full column rank, we show that  $S_\pi$  with  $\pi > 0$  and  $S_0$  both capture the common HO-GSVD subspace of  $A_1, \dots, A_N$  (Corollary 5.2).

The HO-GSVD is a technique that is of particular use in multimodal data fusion [18], which aims to identify common features across multiple data sets that describe related phenomena. Many tensor or multimatrix decompositions are obtained from extending single-matrix factorizations to multiple matrices, such as the parallel factor analysis (PARAFAC [10] or PARAFAC2 [11]), multilinear SVDs [6], multilinear principal component analysis [8], or the higher-order eigenvalue decomposition [2]. The different extensions preserve some but not all of the single-matrix factorization properties [22], such as exactness, orthogonality, or rank conditions of the factor matrices. Some tensor decompositions require that the matrices  $A_i$  share the same dimensions, e.g., a third-order tensor  $\mathcal{A} = A_1 \times A_2 \times \dots \times A_N$  requires that all matrices  $A_i$  have dimensions  $m \times n$ , which imposes constraints on the data acquisition. In contrast, the HO-GSVD is an exact matrix factorization so that  $A_i = U_i \Sigma_i V^T$  for  $i = 1, \dots, N$ , and it can accommodate  $A_i \in \mathbb{R}^{m_i \times n}$  with different row dimensions  $m_i$ , although no constraints, such as orthogonality, can be imposed on the factor matrices.

The GSVD is closely related to the (thin) *cosine-sine decomposition* (CSD) [7, Ch. 2.5.4]. In essence, the CSD states that the SVDs of  $Q_1 \in \mathbb{R}^{m_1 \times n}$  and  $Q_2 \in \mathbb{R}^{m_2 \times n}$  satisfying  $Q_1^T Q_1 + Q_2^T Q_2 = I$  share the same matrix of standard right singular vectors [28]. The GSVD can be obtained from applying a CSD to the matrices  $Q_1$  and  $Q_2$  that are obtained from the thin QR factorization of the stacked matrices  $[A_1^T, A_2^T]^T = QR$ , where  $Q$  is conformably partitioned such that  $A_i = Q_i R$ .

Analogous to the GSVD and the CSD, the HO-GSVD is closely related to the *higher-order CSD* (HO-CSD) [29]. The HO-GSVD of  $N$  matrices  $A_i$  can be obtained from the HO-CSD of  $Q_1, \dots, Q_N$  that are obtained from the thin QR factorization of the stacked matrices  $[A_1^T, \dots, A_N^T]^T$ . As in the case of the HO-GSVD, the computation of the HO-CSD proposed in [29] is limited to the case that all  $Q_i$  have full rank. In this paper, we also propose to compute the HO-CSD in a different way, which allows for the factorization of rank-deficient  $Q_i$  satisfying  $Q_1^T Q_1 + \dots + Q_N^T Q_N = I$ .

The paper is organized as follows. Section 2 presents the HO-CSD and the HO-GSVD, which are applicable to rank-deficient matrices. In section 3, we extend the notion of common HO-CSD and HO-GSVD subspaces to rank-deficient matrices. The effect of the parameter  $\pi$  is investigated in section 4, followed by relating our findings to existing methods in section 5. In section 6, we propose an algorithm for computing the HO-GSVD and the isolated subspace. The paper is concluded with an example application of the HO-GSVD in section 7.

We use standard notation throughout the paper with  $\text{range}(A)$  and  $\ker(A)$  denoting the range (or image) and kernel of a matrix  $A$ . Positive-definite and positive-semidefinite matrices are denoted by  $A \succ 0$  and  $A \succeq 0$ , respectively, and  $\mathbb{R}_+$  and  $\mathbb{R}_{++}$  denote the set of positive and strictly positive real numbers, respectively.

**2. Main results.** Given  $N$  matrices  $A_i \in \mathbb{R}^{m_i \times n}$ , let  $A \in \mathbb{R}^{m \times n}$  with  $m := \sum_{i=1}^N m_i$  denote the matrix of stacked  $A_i$  and  $QR = A$  its thin QR factorization,

$$(2.1) \quad A = \begin{bmatrix} A_1 \\ \vdots \\ A_N \end{bmatrix} = QR = \begin{bmatrix} Q_1 \\ \vdots \\ Q_N \end{bmatrix} R, \quad Q_i \in \mathbb{R}^{m_i \times n}, \quad R \in \mathbb{R}^{n \times n},$$

where it holds that

$$(2.2) \quad Q^T Q = \sum_{i=1}^N Q_i^T Q_i = I, \quad \|Q_i\|_2 \leq 1 \quad \forall i = 1, \dots, N.$$

The matrices  $A_i = Q_i R$  can individually have arbitrary rank, but throughout the paper it is assumed that

$$(2.3) \quad \text{rank}(A) = \text{rank} \begin{pmatrix} A_1 \\ \vdots \\ A_N \end{pmatrix} = n,$$

so that  $\det(R) \neq 0$  and  $m \geq n$ . If (2.3) does not hold, the matrix  $A$  can be padded using an additional matrix  $A_{N+1}$  (see Remark 3.9 to make it full rank). The quotient terms  $D_{i,\pi}$  (2.4) of the arithmetic mean  $S_\pi$  (1.2) can be rewritten as

$$(2.4) \quad D_{i,\pi} = A_i^T A_i + \pi A^T A = R^T (Q_i^T Q_i + \pi I) R,$$

with parameter  $\pi > 0$ . Since  $A_i^T A_i \succeq 0$  and  $\pi A^T A \succ 0$ , the terms  $D_{i,\pi}$  are guaranteed to be invertible.

Most of our developments are based on the HO-CSD of  $Q_1, \dots, Q_N$ , which is obtained from the eigensystem of

$$(2.5) \quad T_\pi := \frac{1}{N} \sum_{i=1}^N (Q_i^T Q_i + \pi I)^{-1}.$$

It can be shown (Appendix A) that  $S_\pi$  and  $T_\pi$  are related by

$$(2.6) \quad R^{-T} S_\pi R^T = \frac{1}{N-1} ((1 + \pi N) T_\pi - I).$$

**THEOREM 2.1.** *Let  $T_\pi$  be defined by (2.5) and suppose that (2.2) holds. There exists an orthogonal  $Z \in \mathbb{R}^{n \times n}$  such that*

$$(2.7) \quad Z^T T_\pi Z = \text{diag}(\tau_1, \dots, \tau_n),$$

where the columns of  $Z$  are eigenvectors of  $T_\pi$  and the eigenvalues  $\tau_i$  of  $T_\pi$  satisfy

$$(2.8) \quad \tau_i \in [\tau_{\min}, \tau_{\max}] := \left[ (N^{-1} + \pi)^{-1}, \frac{N-1}{N} \pi^{-1} + \frac{1}{N} (1 + \pi)^{-1} \right].$$

To prove Theorem 2.1, we first require the following lemma.

**LEMMA 2.2.** *Let  $P = P^T \in \mathbb{R}^{n \times n}$  with  $0 \preceq P \preceq I$ . For all  $t \in \mathbb{R}^n$  and  $\pi \geq 0$ , it holds that  $t^T (\pi(1 + \pi)(\pi I + P)^{-1}) t \leq t^T ((1 + \pi)I - P) t$ . Moreover, equality holds for some  $t$  iff  $P$  has  $p \geq 1$  eigenvalues  $\lambda_1, \dots, \lambda_p \in \{0, 1\}^p$  associated with eigenvectors  $v_1, \dots, v_p$ , and  $t \in \text{span} \{v_1, \dots, v_p\}$ .*

*Proof.* The inequality  $t^T (\pi(1 + \pi)(\pi I + P)^{-1}) t \leq t^T ((1 + \pi)I - P) t$  holds for all  $t \in \mathbb{R}^n$  iff

$$(2.9) \quad (1 + \pi)I - P - \pi(1 + \pi)(\pi I + P)^{-1} \succeq 0.$$

Set  $P = V \Lambda V^T$  with  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ ,  $\lambda_i \in [0, 1]$ , so that (2.9) amounts to

$$f_i(\lambda_i) := 1 + \pi - \lambda_i - \frac{\pi(1 + \pi)}{\pi + \lambda_i} \geq 0, \quad i = 1, \dots, n.$$

Since  $f_i''(\lambda_i) = -2\pi(1 + \pi)/(\pi + \lambda_i)^3 < 0$  for  $\lambda_i \in [0, 1]$ , the function  $f_i(\lambda_i)$  is concave on  $\lambda_i \in [0, 1]$  and hence  $f_i(\lambda_i) \geq \min\{f_i(0), f_i(1)\} = \min\{0, 0\} = 0 \ \forall i = 1, \dots, n$ . Equality therefore holds iff  $\lambda_i \in \{0, 1\}$ .

For the second part of the claim, set  $t = Va$  with  $\|a\|_2 = 1$ , and pre- and post-multiply (2.9) with  $t^T$  and  $t$ , respectively, to obtain

$$(2.10) \quad \sum_{i=1}^n f_i(\lambda_i) a_i^2 \geq 0.$$

Suppose that  $t \in \text{span} \{v_1, \dots, v_p\}$ ; then  $\sum_{i=1}^n f_i(\lambda_i) a_i^2 = \sum_{i=1}^p f_i(\lambda_i) a_i^2 = 0$ . For the converse, suppose that  $t \notin \text{span} \{v_1, \dots, v_p\}$  and that equality holds in (2.10). Then there exists  $j \in \{p+1, \dots, n\}$  with  $a_j > 0$  and  $f_j(\lambda_j) > 0$ , which is a contradiction.  $\square$

*Proof of Theorem 2.1.* The existence of a matrix  $Z \in \mathbb{R}^{n \times n}$ ,  $Z^T Z = I$ , that diagonalizes  $T_\pi$  is a consequence of the symmetry in (2.5). For the lower bound in (2.8), substitute  $u = (Q_i^T Q_i + \pi I)^{\frac{1}{2}} t$  and  $v = (Q_i^T Q_i + \pi I)^{-\frac{1}{2}} t$  with  $\|t\|_2 = 1$  in the Cauchy-Schwarz inequality  $(u^T v)^2 \leq \|u\|_2^2 \|v\|_2^2$  to obtain

$$(2.11) \quad t^T (Q_i^T Q_i + \pi I)^{-1} t \geq (t^T (Q_i^T Q_i + \pi I) t)^{-1}.$$

Using (2.11) and the harmonic-mean arithmetic-mean (HM-AM) inequality [9, Thm. 16], a lower bound on  $t^T T_\pi t$  can be established as

$$(2.12a) \quad t^T T_\pi t = \frac{1}{N} \sum_{i=1}^N t^T (Q_i^T Q_i + \pi I)^{-1} t \geq \frac{1}{N} \sum_{i=1}^N \frac{1}{t^T (Q_i^T Q_i + \pi I) t}$$

$$(2.12b) \quad \geq \frac{N}{\pi N + \sum_{i=1}^N t^T (Q_i^T Q_i) t} = \tau_{\min}.$$

For the upper bound in (2.8), apply Lemma 2.2 with  $P = Q_i^T Q_i$  to each summand of  $T_\pi$ :

$$(2.13) \quad t^T T_\pi t \leq \frac{1}{N} \sum_{i=1}^N t^T \left( \frac{1}{\pi} I - \frac{1}{\pi(1+\pi)} Q_i^T Q_i \right) t = \frac{1}{\pi} - \frac{1}{N\pi(1+\pi)} = \tau_{\max}. \quad \square$$

**THEOREM 2.3.** *Let  $S_\pi$  be defined by (1.2) and suppose that (2.3) holds. There exists an invertible  $V \in \mathbb{R}^{n \times n}$  such that*

$$(2.14) \quad V^{-1} S_\pi V = \text{diag}(\varsigma_1, \dots, \varsigma_n),$$

where the columns of  $V$  are eigenvectors of  $S_\pi$  and the eigenvalues  $\varsigma_i$  satisfy

$$(2.15) \quad \varsigma_i \in [\varsigma_{\min}, \varsigma_{\max}] := \left[ 1, 1 + \frac{1}{\pi N(1+\pi)} \right].$$

*Proof.* Since (2.6) is a similarity transformation,  $V$  can be chosen as  $V := R^T Z$  with  $Z$  from Theorem 2.1, and  $V$  is invertible because  $\det(R) \neq 0$  and  $Z^T Z = I$ . The eigenvalues are  $\varsigma_i = ((1 + \pi N)\tau_i - 1)/(N - 1)$ , and the bounds (2.15) are obtained from (2.8).  $\square$

The significance of Theorems 2.1 and 2.3 is that the diagonalizable matrices  $T_\pi$  and  $S_\pi$  have eigenvalues that are both bounded away from zero and contained in finite intervals, in contrast to the original formulation [23] that requires a full rank condition and corresponds to  $\pi = 0$ . More precisely, the range of eigenvalues of  $S_\pi$  is contracted from  $[1, \infty)$  for the original formulation to  $[1, 1 + 1/(\pi N(1 + \pi))]$  in our case, which bounds the spectral condition number as  $\kappa(S_\pi) := \|S_\pi\|_2 \|S_\pi^{-1}\|_2 \leq 1 + 1/(\pi N(1 + \pi))$ .

Before examining the eigenvalues of  $S_\pi$  and  $T_\pi$  further, we state our version of the HO-CSD and HO-GSVD. The HO-CSD and HO-GSVD have already been described in [29] and [23], respectively, but our modified  $D_{i,\pi}$  from (2.4) allows us to omit the requirements that  $A_i$  and  $Q_i$  be full rank.

**DEFINITION 2.4 (HO-CSD).** *Given  $Q_1, \dots, Q_N$  satisfying (2.2) and  $N \geq 2$ , the HO-CSD of  $Q_i \in \mathbb{R}^{m_i \times n}$  is given by  $Q_i = U_i \Sigma_i Z^T$ ,  $i = 1, \dots, N$ , with  $Z$  defined as in (2.7). The matrices  $\Sigma_i \in \mathbb{R}^{n \times n}$  with  $\Sigma_i = \text{diag}(\sigma_{i,1}, \dots, \sigma_{i,n}) \succeq 0$  are obtained from*

$$B_i := Q_i Z, \quad B_i =: [b_{i,1}, \dots, b_{i,n}], \quad \sigma_{i,k} := \|b_{i,k}\|_2,$$

and  $U_i \in \mathbb{R}^{m_i \times n}$  with  $U_i := [u_{i,1}, \dots, u_{i,n}]$  from

$$u_{i,k} := \begin{cases} b_{i,k}/\sigma_{i,k} & \text{if } \sigma_{i,k} > 0, \\ u \in \mathbb{R}^{m_i} \text{ with } \|u\|_2 = 1 & \text{if } \sigma_{i,k} = 0. \end{cases}$$

The left basis vectors  $u_{i,k}$  have unit 2-norm and are, under certain circumstances, mutually orthogonal, in which case they coincide with certain left generalized singular

vectors of all pairwise standard GSVD factorizations (see section 3). Because we allow for rank-deficient  $Q_i$  ( $\text{rank}(Q_i) < \min(m_i, n)$ ) as well as  $m_i < n$ , it is possible that  $Q_i z_k = 0$  for some eigenvector  $z_k$  of  $T_\pi$ , consequently making the corresponding generalized singular value  $\sigma_{i,k} = 0$ . In these cases, the column  $u_{i,k}$  can be chosen freely or the corresponding row of  $\Sigma_i$  can be dropped. For rank-deficient  $Q_i$  ( $\text{rank}(Q_i) < \min(m_i, n)$ ), they can be chosen to be orthogonal to all other columns associated with nonzero generalized singular values, such as stated in the following lemma.

**LEMMA 2.5.** *Suppose that  $r_i := \text{rank}(Q_i) < \min(m_i, n)$ , and let the generalized singular values be ordered such that  $\sigma_{i,k} = 0$  for  $k \leq K$ , and  $\sigma_{i,j} > 0$  for  $j > K$ . There exist  $m_i - r_i \leq K$  mutually orthogonal vectors  $u_{i,1}, \dots, u_{i,(m_i-r_i)}$  such that  $u_{i,k}^T u_{i,j} = 0 \forall k \leq m_i - r_i, j > K$ .*

*Proof.* Note that  $\text{span}\{u_{i,K+1}, \dots, u_{i,n}\} = \text{range}(Q_i)$ . Since  $r_i < \min(m_i, n)$ , there exist  $m_i - r_i$  vectors  $u_{i,1}, \dots, u_{i,(m_i-r_i)}$  satisfying  $\text{span}\{u_{i,1}, \dots, u_{i,(m_i-r_i)}\} = \ker(Q_i^T)$  and  $u_{i,k}^T u_{i,j} = 0 \forall k \leq m_i - r_i, j > K$ , e.g., the last  $m_i - r_i$  columns of the matrix of standard left singular vectors of  $Q_i$ .  $\square$

**DEFINITION 2.6 (HO-GSVD).** *Given  $A_1, \dots, A_N$  satisfying (2.3) and  $N \geq 2$ , the HO-GSVD of  $A_i \in \mathbb{R}^{m_i \times n}$  is given by  $A_i = U_i \Sigma_i V^T$ , with  $V$  defined as in (2.14). The matrices  $\Sigma_i \in \mathbb{R}^{n \times n}$  with  $\Sigma_i = \text{diag}(\sigma_{i,1}, \dots, \sigma_{i,n}) \succeq 0$  are obtained from*

$$B_i := A_i V^{-T}, \quad B_i =: [b_{i,1}, \dots, b_{i,n}], \quad \sigma_{i,k} := \|b_{i,k}\|_2,$$

and  $U_i \in \mathbb{R}^{m_i \times n}$  with  $U_i := [u_{i,1}, \dots, u_{i,n}]$  from

$$u_{i,k} := \begin{cases} b_{i,k}/\sigma_{i,k} & \text{if } \sigma_{i,k} > 0, \\ u \in \mathbb{R}^{m_i} \text{ with } \|u\|_2 = 1 & \text{if } \sigma_{i,k} = 0. \end{cases}$$

In contrast to the standard SVD ( $N = 1$ ) and the GSVD ( $N = 2$ ), the left factor matrices  $U_i$  of the HO-GSVD (or HO-CSD) are not square in general. For  $m_i > n$ , it is always possible to extend  $U_i$  using the last  $m_i - n$  standard left singular vectors of  $Q_i$ , but for  $m_i < n$ , it is only possible to remove  $n - m_i$  columns of  $U_i$  if the corresponding generalized singular values are zero, which amounts to aligning  $v_1, \dots, v_n$  such that the number of nonzero  $\sigma_{i,k}$  equals  $r_i$  for each  $A_i$  (see  $B_i$  in Definitions 2.4 and 2.6). While the standard GSVD ( $N = 2$ ) finds  $n$  linearly independent vectors  $v_k$  that yield exactly  $r_i$  nonzero  $\sigma_{i,k}$  for each  $A_i$ ,  $i \in \{1, 2\}$ , straightforward geometric arguments show that it is not always possible to find  $n$  linearly independent vectors with the same property for  $N > 2$ . For example, for  $N = 3$ ,  $n = 3$ , and

$$\ker(A_1) = \text{span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}, \quad \ker(A_2) = \text{span} \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}, \quad \ker(A_3) = \text{span} \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\},$$

it is not possible to find linearly independent  $v_1, v_2$ , and  $v_3$  yielding one zero  $\sigma_{i,k}$  for each  $A_i$ .

According to Theorem 2.3, Definitions 2.4 and 2.6 are equivalent in the sense that the HO-GSVD can be obtained from setting  $V = R^T Z$ :

$$(2.16) \quad B_i = A_i V^{-T} = Q_i R R^{-1} Z = Q_i Z,$$

where the rightmost term corresponds to  $B_i$  as found in Definition 2.4. The matrix of left basis vectors  $U_i$  and the generalized singular values therefore depend only on

the column space  $Q$ . However, when the HO-GSVD and the HO-CSD are computed separately, and  $T_\pi$  and  $S_\pi$  have eigenvalues with geometric multiplicity greater than 1, it is possible that  $V \neq R^T Z$ .

*Remark 2.7.* For rank-deficient  $A_i$ , the reader may wonder why the standard formulations of  $S_\pi$  and  $T_\pi$  with  $\pi = 0$  are not adapted by substituting the pseudoinverse for the inverse in (1.2) and (2.5). The reason is that, in general,  $A_i^\dagger = (Q_i R)^\dagger \neq R^\dagger Q_i^\dagger$ , and using the pseudoinverse, the relationship (2.6) does not hold. However, relationship (2.6) is fundamental in determining the minimum and maximum eigenvalues of  $S_\pi$  that will play an important role in subsequent sections, which is why pseudoinverses are not considered further.

**3. Common and isolated subspaces.** The HO-CSD and HO-GSVD identify directions, corresponding to columns of  $Z$  and  $V$  that, in the sense of (1.4), contribute equally to the factorizations of  $Q_i$  and  $A_i$ , respectively. The directions are the right basis vectors  $v_{i,k}$  associated with generalized singular values that are identical for each  $Q_i$  and  $A_i$ , i.e.,  $\sigma_{i,k} = \sigma_{j,k}$ . These vectors form subspaces [23, 29], which are referred to as the common HO-CSD and HO-GSVD subspaces, and are defined as follows.

DEFINITION 3.1. *The common HO-CSD subspace is defined as*

$$\mathcal{T}_N \{Q_1, \dots, Q_N\} := \{z \in \mathbb{R}^n \mid T_\pi z = \tau_{\min} z\},$$

and the common HO-GSVD subspace is defined as

$$\mathcal{S}_N \{A_1, \dots, A_N\} := \{v \in \mathbb{R}^n \mid S_\pi v = \varsigma_{\min} v\},$$

where  $\tau_{\min}$  and  $\varsigma_{\min}$  are the lower bounds on the range of eigenvalues defined in Theorems 2.1 and 2.3, and  $N \geq 2$ .

By Theorem 2.3, the HO-GSVD and HO-CSD subspaces are related by

$$(3.1) \quad \mathcal{S}_N \{A_1, \dots, A_N\} = \{R^T z \in \mathbb{R}^n \mid z \in \mathcal{T}_N \{Q_1, \dots, Q_N\}\},$$

so that  $v \in \mathcal{S}_N \{A_1, \dots, A_N\}$  iff  $z = R^T v \in \mathcal{T}_N \{Q_1, \dots, Q_N\}$ . Note that for a given set of matrices  $A_1, \dots, A_N$ , the subspaces  $\mathcal{T}_N \{Q_1, \dots, Q_N\}$  and  $\mathcal{S}_N \{A_1, \dots, A_N\}$  might be empty. The following theorem characterizes situations under which the common subspaces are nonempty.

THEOREM 3.2. *The following statements are equivalent:*

- 3.2a  $\mathcal{T}_N \{Q_1, \dots, Q_N\} \neq \emptyset$ .
- 3.2b *There exists  $\hat{z} \in \mathbb{R}^n$  that is a standard right singular vector for each  $Q_i$  and associated with a standard singular value  $\hat{\sigma} = 1/\sqrt{N}$  for each  $Q_i$ .*
- 3.2c *For each  $Q_i$ , there is a left basis vector  $u_{i,k}$  satisfying  $u_{i,k}^T u_{i,p} = 0 \ \forall p \neq k$  and the corresponding generalized singular values is  $\sigma_{i,k} = 1/\sqrt{N}$  for each  $Q_i$ .*

*Proof.* The biconditional relationship  $3.2a \Leftrightarrow 3.2b$  is a consequence of Theorem 2.1. Equality holds in (2.11) iff  $t$  is an eigenvector of  $Q_i^T Q_i$  [9, Thm. 7] or consequently in (2.12a) iff  $t$  is an eigenvector of each  $Q_i^T Q_i$  for  $i = 1, \dots, N$ . Equality holds in (2.12b) iff  $t^T (Q_i^T Q_i + \pi I) t = t^T (Q_j^T Q_j + \pi I) t$  for  $i, j = 1, \dots, N$ . It follows that  $T_\pi t = \tau_{\min} t$  iff  $t$  is a standard right singular vector for each  $Q_i$  and from (2.2) that  $1 = N\hat{\sigma}^2$ , where  $\hat{\sigma} = 1/\sqrt{N}$  is the corresponding standard singular value. To show  $3.2b \Rightarrow 3.2c$ , let  $\hat{u}_{i,k}$  be the corresponding standard left singular vector; then  $Q_i z_k = \hat{\sigma} \hat{u}_{i,k}$  and from the HO-CSD,  $Q_i z_k = \sigma_{i,k} u_{i,k}$ , so the generalized singular values

satisfy  $\sigma_{i,k} = \hat{\sigma}$  since  $\|\hat{u}_{i,k}\|_2 = \|u_{i,k}\|_2 = 1$ . To show that  $u_{i,k}^T u_{i,p} = 0 \ \forall p \neq k$ , consider the following equations for  $\sigma_{i,p} \neq 0$ :

$$u_{i,k}^T u_{i,p} = \frac{b_{i,k}^T b_{i,p}}{\sigma_{i,k} \sigma_{i,p}} = \frac{z_k^T Q_i^T Q_i z_p}{\sigma_{i,k} \sigma_{i,p}} = \frac{\sigma_{i,k}}{\sigma_{i,p}} z_k^T z_p = 0,$$

where  $b_{i,k}$  denotes column  $k$  of the matrix  $B_i$  from Definition 2.4.

To show 3.2c  $\Rightarrow$  3.2b, suppose that 3.2c holds and let  $z_k$  be the corresponding right generalized singular vector. Then,  $Q_i^T Q_i z_k = Z \Sigma_i U_i U_i^T \Sigma_i Z z_k = \sigma_{i,k}^2 z_k$  since  $u_{i,k}^T u_{i,p} = 0 \ \forall p \neq k$ ; hence  $z_k$  is a shared standard right singular vector associated with a standard singular value  $\sigma_{i,k}$ .  $\square$

Note that statement 3.2c implies that the corresponding left basis vector  $u_{i,k}$  is an eigenvector for  $Q_i Q_i^T$  for each  $i$  and therefore also a *standard* left singular vector for each  $Q_i$ .

The common HO-GSVD and HO-CSD subspaces are related by (3.1), and Theorem 3.2 can be adapted for the common HO-GSVD subspace as follows.

**COROLLARY 3.3.** *The following statements are equivalent:*

3.3a  $\mathcal{S}_N\{A_1, \dots, A_N\} \neq \emptyset$ .

3.3b *For each  $A_i$ , there is a left basis vector  $u_{i,k}$  satisfying  $u_{i,k}^T u_{i,p} = 0 \ \forall p \neq k$  and the corresponding generalized singular value is  $\sigma_{i,k} = 1/\sqrt{N}$  for each  $A_i$ .*

3.3c *There exists  $v \in \mathbb{R}^n$  that is an eigenvector for each pairwise quotient  $D_{i,\pi} D_{j,\pi}^{-1}$  associated with an eigenvalue  $\lambda_{i,j} = 1$ .*

*Proof.* The biconditional relationship 3.3a  $\Leftrightarrow$  3.3b immediately follows from (3.1) and Theorem 3.2. To show 3.3b  $\Rightarrow$  3.3c, substitute the HO-GSVD in (1.3) to obtain

$$(3.2) \quad D_{i,\pi} = V \underbrace{\Sigma_i U_i^T U_i \Sigma_i}_{=: W_i} V^T + \pi A^T A = V \left( W_i + \pi \sum_{p=1}^N W_p \right) V^T,$$

so that  $D_{i,\pi} D_{j,\pi}^{-1} = V(W_i + \pi \sum_{p=1}^N W_p)(W_j + \pi \sum_{p=1}^N W_p)^{-1} V^{-1}$ . Because of 3.3b, each  $W_i$  has the block-diagonal form  $W_i = \text{diag}(\underline{W}_i, \sigma_{i,k}^2 + \pi \sum_{p=1}^N \sigma_{p,k}^2, \overline{W}_i)$ , where the scalar entry is on the  $k$ th row of  $W_i$  and  $\underline{W}_i$  and  $\overline{W}_i$  are principal submatrices of  $W_i$ . Again from 3.3b,  $\sigma_{i,k} = \sigma_{j,k}$ , so that  $D_{i,\pi} D_{j,\pi}^{-1} v = v$ . To complete the proof, we show that 3.3c  $\Rightarrow$  3.3a by right-multiplying  $S_\pi$  from (1.2) with  $v$  from 3.3c:

$$\begin{aligned} S_\pi v &= \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N (\lambda_{i,j} v + \lambda_{j,i} v) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N v \\ &= \frac{2}{N(N-1)} \sum_{i=1}^N (N-i) v = \frac{2}{N(N-1)} \left( N^2 - \frac{N^2 + N}{2} \right) v = \varsigma_{\min} v. \quad \square \end{aligned}$$

The “common features” of  $A_i$  in (1.4) can therefore be identified by the right basis vectors associated with eigenvalues of  $S_\pi$  that equal  $\varsigma_{\min}$ . In general,  $R^{-T} z$  is not an eigenvector for  $A_i^T A_i = R Q_i^T Q_i R^T$ , so that statement 3.2b cannot be adapted to the HO-GSVD subspace, and the corresponding right basis vectors associated with the common subspace are not orthogonal in general. However, the right basis vectors spanning  $\mathcal{S}_N\{A_1, \dots, A_N\}$  are eigenvectors of all pairwise quotients  $D_{i,\pi} D_{j,\pi}^{-1}$ , which is exploited in [29] to compute the common HO-GSVD subspace using the standard pairwise GSVD. In addition, one can reformulate statement 3.3c to show that there



exists a vector  $\tilde{v} = D_{j,\pi}^{-1}v = D_{i,\pi}^{-1}v$ ,  $v \in \mathcal{S}_N\{A_1, \dots, A_N\}$ , that solves the *higher-order generalized singular value problem*  $A_i^T A_i \tilde{v} = \mu A_j^T A_j \tilde{v}$  with  $\mu = 1$ .

In contrast to the common subspace, the isolated part of (1.4) that is unique to a single  $A_i$  is identified by the right basis vectors associated with eigenvalues of  $S_\pi(T_\pi)$  that equal  $\varsigma_{\max}$  ( $\tau_{\max}$ ).

DEFINITION 3.4. *The isolated HO-CSD subspace is defined as*

$$\mathcal{T}_1\{Q_1, \dots, Q_N\} := \{z \in \mathbb{R}^n \mid T_\pi z = \tau_{\max} z\},$$

and the isolated HO-GSVD subspace is defined as

$$\mathcal{S}_1\{A_1, \dots, A_N\} := \{v \in \mathbb{R}^n \mid S_\pi v = \varsigma_{\max} v\},$$

where  $\tau_{\max}$  and  $\varsigma_{\max}$  are upper bounds on the range of eigenvalues defined in Theorems 2.1 and 2.3, and  $N \geq 2$ .

THEOREM 3.5. *The following statements are equivalent:*

3.5a  $\mathcal{T}_1\{Q_1, \dots, Q_N\} \neq \emptyset$ .

3.5b *There exists  $\hat{z} \in \mathbb{R}^n$  that is a standard right singular vector for each  $Q_i$  and is associated with a standard singular value  $\hat{\sigma}_{j,k} = 1$  for one  $Q_j$  and  $\hat{\sigma}_{i,k} = 0$  for all other  $Q_i$ ,  $i \neq j$ .*

3.5c *There is a right basis vector  $z \in \mathbb{R}^n$  associated with a generalized singular value  $\sigma_{j,k} = 1$  for some  $Q_j$ , and  $\sigma_{i,k} = 0$  for all other  $Q_i$ ,  $i \neq j$ .*

*Proof.* The biconditional relationship 3.5a  $\Leftrightarrow$  3.5b is a consequence of the proof of Theorem 2.1. According to Lemma 2.2, equality is attained in (2.13) iff for each summand,  $t \in \text{span}\{v_1^i, \dots, v_p^i\}$ , where  $v_k^i$  are eigenvectors of  $Q_i^T Q_i$  associated with eigenvalues that are equal to either 0 or 1. It remains to consider (2.2). The relationship 3.5b  $\Leftrightarrow$  3.5c follows from Definition 2.4 and (2.2).  $\square$

COROLLARY 3.6. *If  $\sigma_{j,k} = 1$ , then the corresponding left basis vector  $u_{j,k}$  satisfies  $u_{j,k}^T u_{j,p} = 0 \forall p \neq k$  with  $\sigma_{j,p} \neq 0$ . If  $\text{rank}(Q_i) < \min(m_i, n)$ , the left basis vectors  $u_{i,k}$  associated with  $\sigma_{i,k} = 0$  can be chosen such that  $u_{i,k}^T u_{i,p} = 0 \forall p \neq k$  with  $\sigma_{i,p} \neq 0$ .*

*Proof.* According to Theorem 3.5, the right basis vector  $z_k$  associated with  $\sigma_{j,k} = 1$  is also a standard right singular vector of  $Q_j$ , and the proof of  $u_{j,k}^T u_{j,p} = 0 \forall p \neq k$  with  $\sigma_{j,p} \neq 0$  follows the proof of Theorem 3.5. For  $\text{rank}(Q_i) < \min(m_i, n)$ , the left basis vectors  $u_{i,k}$  associated with  $\sigma_{i,k} = 0$  can be chosen according to Lemma 2.5.  $\square$

Note that for  $m_i \geq n$ , the left basis vectors  $u_{i,k}$  associated with zero or nonzero generalized singular values can *always* be chosen to be orthogonal to the remaining left basis vectors (Lemma 2.5).

By Theorem 2.3, the isolated HO-GSVD and HO-CSD subspaces are related by

$$(3.3) \quad \mathcal{S}_1\{A_1, \dots, A_N\} = \{R^T z \in \mathbb{R}^n \mid z \in \mathcal{T}_1\{Q_1, \dots, Q_N\}\},$$

and Theorem 3.5 is reformulated for the HO-GSVD as follows.

COROLLARY 3.7. *The following statements are equivalent:*

3.7a  $\mathcal{S}_1\{A_1, \dots, A_N\} \neq \emptyset$ .

3.7b *There is a right basis vector  $v \in \mathbb{R}^n$  associated with a generalized singular value  $\sigma_{j,k} = 1$  for one  $A_j$  and  $\sigma_{i,k} = 0$  for all other  $A_i$ ,  $i \neq j$ .*

3.7c *There exist  $v \in \mathbb{R}^n$  and  $i \in \{1, \dots, N\}$  such that  $v$  is an eigenvector for each pairwise quotient  $D_{p,\pi} D_{j,\pi}^{-1}$  associated with eigenvalues  $\lambda_{i,j} = \frac{1+\pi}{\pi}$ ,  $\lambda_{j,i} = \frac{\pi}{1+\pi}$ , and  $\lambda_{p,j} = \lambda_{j,p} = 1$  for  $j = \{1, \dots, N\}$ ,  $p = \{1, \dots, N\}$ , and  $j \neq p \neq i$ .*

*Proof.* The proof follows the proof of Corollary 3.3. To show 3.7a  $\Leftrightarrow$  3.7b, use (3.3) and apply Theorem 3.5. To show 3.7b  $\Rightarrow$  3.7c, use (3.2) while considering Corollary 3.6. Finally, to show 3.7c  $\Rightarrow$  3.7a, compute  $S_\pi v$  and assume without loss of generality that  $i = 1$ :

$$\begin{aligned} S_\pi v &= \frac{1}{N(N-1)} \sum_{j=2}^N (\lambda_{1,j} + \lambda_{j,1}) v + \frac{1}{N(N-1)} \sum_{p=2}^N \sum_{j=p+1}^N (\lambda_{p,j} + \lambda_{j,p}) v \\ &= \frac{1}{N(N-1)} \sum_{j=2}^N \left( \frac{1+\pi}{\pi} + \frac{\pi}{1+\pi} \right) v + \frac{1}{N(N-1)} \sum_{p=2}^N \sum_{j=p+1}^N 2v \\ &= \frac{1}{N} \left( \frac{1+\pi}{\pi} + \frac{\pi}{1+\pi} + N-2 \right) v = \varsigma_{\max} v. \quad \square \end{aligned}$$

Note that Corollary 3.6 also applies to the left basis vectors associated with the isolated subspace of the HO-GSVD.

Statements 3.2c and 3.5c of Theorems 3.2 and 3.5 show that, in certain cases, the orthogonality of the left factor matrix, which always holds for the standard SVD and GSVD, is preserved for higher-order datasets (see also section 5). If the generalized singular values  $\sigma_{i,k}$ , the left basis vectors  $u_{i,k}$ , and the right basis vectors  $v_k$  are grouped according to whether they are associated with the common subspace ( $k \in \mathcal{I}_N$ ), the isolated subspace ( $k \in \mathcal{I}_1$ ), or neither of the subspaces ( $k \in \mathcal{I}_\perp$ ), Definition 2.6 can be refined as

$$(3.4) \quad A_i = \begin{bmatrix} U_{i,\mathcal{I}_1} & U_{i,\mathcal{I}_\perp} & U_{i,\mathcal{I}_N} \end{bmatrix} \begin{bmatrix} \Sigma_{i,\mathcal{I}_1} & & \\ & \Sigma_{i,\mathcal{I}_\perp} & \\ & & I/\sqrt{N} \end{bmatrix} \begin{bmatrix} V_{\mathcal{I}_1} & V_{\mathcal{I}_\perp} & V_{\mathcal{I}_N} \end{bmatrix}^T,$$

where  $\Sigma_{i,\mathcal{I}_1}$  contains the generalized singular values associated with  $\mathcal{S}_1\{A_1, \dots, A_N\}$  and  $\Sigma_{i,\mathcal{I}_\perp} \succ 0$ . In the notation of (3.4) and for  $\text{rank}(A_i) < \min(m_i, n)$ , the three blocks of left basis vectors are mutually orthogonal, e.g.,  $(U_{i,\mathcal{I}_N})^T U_{i,\mathcal{I}_1} = 0$ , which follows from statements 3.3b and 3.7b of Corollaries 3.3 and 3.7. Note that for the HO-GSVD, the right basis vectors are *not* orthogonal in general.

As can also be concluded from Theorems 3.2 and 3.5, the parameter  $\pi$  does not alter the common and isolated subspaces, which shows that the standard HO-GSVD formulation and the present one are equivalent.

**COROLLARY 3.8.** *The common and isolated HO-GSVD and HO-CSD subspaces are independent of the value of  $\pi$ .*

*Proof.* For the HO-CSD, the claim follows from statements 3.2b and 3.5b of Theorems 3.2 and 3.5, which are independent of the value of  $\pi$ . As a consequence of (3.1) and (3.3), the claim is also true for the HO-GSVD.  $\square$

Note that Corollary 3.8 ignores potential numerical inaccuracies, which are treated in section 6. Numerical inaccuracies can also cause rank deficiencies of the stacked matrix  $A$ , and Remark 3.9 explains how the HO-GSVD can be applied even when  $A$  does not satisfy (2.3).

**Remark 3.9.** Suppose that assumption (2.3) does *not* hold and that  $\text{rank}(A) = r < n$ . Then,  $S_\pi$  is undefined and (2.6) invalid. Let  $\text{span}\{v_1, \dots, v_{n-r}\} = \ker(A)$  be an orthogonal basis and set  $A_{N+1} := [v_1, \dots, v_{n-r}]^T$ . The HO-GSVD can be applied to the augmented dataset  $A_1, \dots, A_{N+1}$ , and at least  $n-r$  directions of the resulting isolated HO-GSVD subspace are associated with  $\ker(A)$ .

**4. The parameter  $\pi$ .** The eigenvectors of  $T_\pi$  that are in the common or isolated HO-CSD subspaces are not affected by the choice of  $\pi$ , but other (normalized) eigenvectors can be modified as  $\pi$  varies. Here, we are interested in the limits of these eigenvectors as  $\pi \rightarrow 0$  and  $\pi \rightarrow \infty$ , which can also be interpreted as  $\pi \ll \min_i \|A_i^T A_i\|_2 / \|A^T A\|_2$  and  $\pi \gg \max_i \|A_i^T A_i\|_2 / \|A^T A\|_2$  or  $\pi \ll \min_i \|Q_i^T Q_i\|_2$  and  $\pi \gg \max_i \|Q_i^T Q_i\|_2$ . Since from (2.5) it holds that  $\lim_{\pi \rightarrow \infty} S_\pi = I$  and  $\lim_{\pi \rightarrow \infty} T_\pi = 0$ , some caution is required in determining the limits of the associated eigenvectors.

Semisimple eigenvalues are expected to be associated with the common or isolated subspaces and therefore are not considered further. To examine the remaining eigenvectors associated with simple eigenvalues, we will make use of the following result in both cases.

**THEOREM 4.1** ([20, Thm. 7 & 8, Ch. 9.3, p. 130]). *Let  $M(x)$  be a differentiable square matrix-valued function of the real variable  $x$ . Suppose that  $M(0)$  has a simple eigenvalue  $m_0$ . Then for  $x$  small enough,  $M(x)$  has an eigenvalue  $m(x)$  that depends differentiably on  $x$  with  $m(0) = m_0$  and we can choose an eigenvector  $h(x)$  of  $M(x)$  pertaining to the eigenvalue  $m(x)$  that depends differentiably on  $x$ .*

**The case  $\pi \rightarrow \infty$ .**

**LEMMA 4.2** (eigenvectors of  $T_\pi$  for  $\pi \rightarrow \infty$ ). *Consider the matrix  $\tilde{T}_\infty$ ,*

$$(4.1) \quad \tilde{T}_\infty := \frac{1}{N} \sum_{i=1}^N (Q_i^T Q_i)^2,$$

*and suppose that  $\tilde{T}_\infty$  has a simple eigenvalue  $\tilde{\tau}_\infty$  associated with an eigenvector  $\tilde{z}_\infty$ . Then, there exists a differentiable  $z: \mathbb{R}_+ \mapsto \mathbb{R}^n$  such that  $z(\pi)$  is an eigenvector for  $T_\pi$  and  $\lim_{\pi \rightarrow \infty} z(\pi) = \tilde{z}_\infty$ .*

*Proof.* Use the Neumann series  $(I - M)^{-1} = \sum_{k=0}^{\infty} M^k$  with  $\|M\| < 1$  [12, Ch. 1.4] to expand each of the summands in (2.5) as

$$(Q_i^T Q_i + \pi I)^{-1} = \frac{1}{\pi} \sum_{k=0}^{\infty} \left( \frac{-1}{\pi} Q_i^T Q_i \right)^k,$$

and rewrite  $T_\pi$  as

$$T_\pi = \frac{1}{N\pi} \sum_{i=1}^N \sum_{k=0}^{\infty} \left( \frac{-1}{\pi} Q_i^T Q_i \right)^k = \frac{1}{\pi} I - \frac{1}{N\pi^2} I + \sum_{i=1}^N \frac{1}{N\pi^3} (Q_i^T Q_i)^2 + \mathcal{O}\left(\frac{1}{\pi^4}\right),$$

where  $\|\frac{1}{\pi} Q_i^T Q_i\| < 1$  for  $\pi > 1$ . Set  $\tilde{T}(\pi) = \pi^3 (T_\pi - \frac{N\pi-1}{N\pi^2} I)$ , which depends differentiably on  $\pi$  for  $\pi > 0$  and has the same eigenvectors as  $T_\pi$ . Neglecting higher-order terms  $\mathcal{O}(1/\pi^4)$ , the limit  $\lim_{\pi \rightarrow \infty} \tilde{T}(\pi) = \tilde{T}_\infty$  is obtained, where equality holds elementwise. Finally, defining

$$M(x) = \begin{cases} \tilde{T}_\infty, & x = 0, \\ \tilde{T}(1/x), & x > 0, \end{cases}$$

we see that the proof follows from Theorem 4.1.  $\square$

According to Lemma 4.2, the eigenvectors of  $T_\pi$  associated with simple eigenvalues can be chosen such that they converge to those of  $\tilde{T}_\infty$  for large  $\pi$ . Suppose that some  $Q_i$  has a “dominant” standard right singular vector  $\bar{v}$ , in the sense

that  $\bar{v}^T Q_i^T Q_i \bar{v} \gg \bar{v}^T Q_j^T Q_j \bar{v}$  for  $j \neq i$ . In this case,  $\tilde{T}_\infty$  can be rewritten as  $\tilde{T}_\infty = Q_i^T Q_i / N + \Delta$  with  $\|\Delta\|_2 \ll \|Q_i^T Q_i / N\|_2$ . According to standard perturbation theory [7, Ch. 7.2.5],  $T_\pi$  will have an eigenvector  $v = \bar{v} + \delta v$  with  $\|\delta v\|_2 \ll 1$ . By using the orthogonality property  $\sum_{i=1}^N Q_i^T Q_i = I$ , the matrix  $\tilde{T}_\infty$  defined in (4.1) can be rewritten as

$$(4.2) \quad \begin{aligned} \tilde{T}_\infty &= \frac{1}{N} \sum_{i=1}^N Q_i^T Q_i \left( I - \sum_{j \neq i} Q_j^T Q_j \right) \\ &= \frac{1}{N} I - \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^N (Q_i^T Q_i Q_j^T Q_j + Q_j^T Q_j Q_i^T Q_i), \end{aligned}$$

where the summands on the second line are referred to as *symmetrized products* or *Jordan products* of  $Q_i^T Q_i$  and  $Q_j^T Q_j$  [20, Ch. 10]. The form (4.2) shows that the eigenvectors of  $T_\pi$  will also converge to those of a “dominant” symmetrized product for large  $\pi$ .

**The case  $\pi \rightarrow 0$ .** For any rank-deficient  $Q_i$  and  $\pi = 0$ , the corresponding term  $Q_i^T Q_i + \pi I$  appearing in the definition of  $T_\pi$  in (2.5) is singular. However, by using the standard SVD  $Q_i^T Q_i = V_i \text{diag}(\sigma_{i,1}^2, \dots, \sigma_{i,r}^2, 0, \dots, 0) V_i^T$  with  $r = \text{rank}(Q_i)$ , one can show that

$$\lim_{\pi \rightarrow 0} \pi (Q_i^T Q_i + \pi I)^{-1} = V_i \underbrace{\text{diag}(0, \dots, 0)}_{r \text{ times}} \underbrace{\text{diag}(1, \dots, 1)}_{n-r \text{ times}} V_i^T,$$

where this limit is zero if  $Q_i$  is instead full rank. The following lemma provides useful information in the case where some of the  $Q_i$  are rank-deficient.

**LEMMA 4.3** (eigenvectors of  $T_\pi$  for  $\pi \rightarrow 0$ ). *Suppose that some of the  $Q_i$  are rank-deficient. Consider*

$$(4.3) \quad \tilde{T}_0 := \frac{1}{N} \sum_{i=1}^N Q_i^\dagger Q_i,$$

where  $Q_i^\dagger = \lim_{\pi \rightarrow 0} Q_i^T (\pi I + Q_i Q_i^T)^{-1}$  is the Moore–Penrose pseudoinverse of  $Q_i$  [7, P5.5.2], and suppose that  $\tilde{T}_0$  has a simple eigenvalue  $\tilde{\tau}_0$  associated with an eigenvector  $\tilde{z}_0$ . Then, there exists a differentiable  $z: \mathbb{R}_+ \mapsto \mathbb{R}^n$  such that  $z(\pi)$  is an eigenvector for  $T_\pi$  and  $\lim_{\pi \rightarrow 0} z(\pi) = \tilde{z}_0$ .

*Proof.* Use the Woodbury matrix identity [7, Ch. 2.1.4] to rewrite  $\pi T_\pi$  for  $\pi > 0$  as

$$(4.4) \quad \pi T_\pi = \frac{1}{N} \sum_{i=1}^N \left( I - Q_i^T (\pi I + Q_i Q_i^T)^{-1} Q_i \right),$$

with  $\lim_{\pi \rightarrow 0} \pi (T_\pi - \frac{1}{\pi} I) = -\tilde{T}_0$  (elementwise), where  $\tilde{T}_0$  and  $\pi (T_\pi - \frac{1}{\pi} I)$  share the same eigenspace [7, Ch. 2]. Differentiability of the matrix  $\pi T_\pi$  with respect to  $\pi$  at 0 is easily shown by substitution of the standard SVD of each  $Q_i$  into (4.4). The proof then follows from Theorem 4.1.  $\square$

Note that if *all*  $Q_i$  have full column rank, then  $Q_i^\dagger Q_i = I$  and  $\tilde{T}_0$  has no simple eigenvalues. The matrix  $Q_i^\dagger Q_i$  is the orthogonal projector onto  $\text{range}(Q_i^T)$  and  $Q_i^\dagger Q_i = I$  if  $Q_i$  has full rank. It follows that if some  $Q_j$  are rank-deficient, then the eigendecomposition of  $T_\pi$  can be chosen such that it equals the eigendecomposition of

the sum of projectors onto  $\text{range}(Q_j^T)$  (the orthogonal complement of  $\ker(Q_j)$ ), but if all  $Q_i$  have full rank, then the eigenvectors of  $\lim_{\pi \rightarrow 0} T_\pi$  are those of  $T_0$ , i.e., (2.5) with  $\pi = 0$ .

The limits for  $S_\pi$  can be obtained from pre- and post-multiplying  $\tilde{T}_0$  or  $\tilde{T}_\infty$  with  $R^T$  and  $R^{-T}$ , respectively.

Apart from rotating the eigenvectors, the choice of  $\pi$  also affects the function  $f_\pi: \mathbb{R}^n \rightarrow \mathbb{R}_{++}$ ,

$$(4.5) \quad f_\pi(v) = \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=1}^N \left( \frac{v^T(A_i^T A_i + \pi A^T A)v}{v^T(A_j^T A_j + \pi A^T A)v} + \frac{v^T(A_j^T A_j + \pi A^T A)v}{v^T(A_i^T A_i + \pi A^T A)v} \right),$$

where  $\|v\|_2 = 1$  and  $f_\pi(v) \geq 1$ . The function  $f_\pi(v)$  measures the arithmetic mean of amplifications in a particular direction  $v$  and has been shown to be related to the (HO-)GSVD [15, 5, 29]. For  $N = 2$ ,  $\pi = 0$ , and full-rank  $A_1$  and  $A_2$ , the gradient is zero for vectors that lie in the common HO-GSVD subspace [29], which can be extended to the isolated HO-GSVD subspace (Appendix C). The parameter  $\pi$  has the effect of *flattening out*  $f_\pi$  and, in particular, removing the singularities of  $f_\pi(v)$  associated with the nullspace of  $A_i$  for  $\pi > 0$  and  $\text{rank}(A) = n$ , since in that case  $v^T(A_i^T A_i + \pi A^T A)v > 0$  for  $v \neq 0$ .

**5. Comparison with standard HO-GSVD, GSVD, and SVD.** When one out of two matrices is the identity matrix, the GSVD reduces to the standard SVD [27]. The same has been shown for the full-rank HO-GSVD [23]. When  $N-1$  matrices  $A_i$  are identity matrices, then the full-rank HO-GSVD reverts to the standard SVD of  $A_j$ ,  $j \neq i$ . Here, this fact is demonstrated for our HO-GSVD as given in Definition 2.6.

**THEOREM 5.1.** *Let  $A_1$  be an arbitrary matrix and  $A_2 = \dots = A_N = I$  with  $N \geq 2$ . The HO-GSVD of  $A_1, A_2, \dots, A_N$  with  $\pi > 0$  yields the standard SVD of  $A_1$ .*

*Proof.* Substitute the standard SVD  $\hat{U}_1 \hat{\Sigma}_1 \hat{V}_1^T = A_1$  and  $A_j = I$ ,  $j = 2, \dots, N$ , in (1.3), so that

$$\hat{V}_1^T D_1 \hat{V}_1 = (1 + \pi) \hat{\Sigma}_1^T \hat{\Sigma}_1 + \pi(N-1)I, \quad \hat{V}_1^T D_j \hat{V}_1 = \pi \hat{\Sigma}_1^T \hat{\Sigma}_1 + (1 + \pi(N-1))I.$$

The summands  $D_{i,\pi} D_{j,\pi}^{-1} + D_{j,\pi} D_{i,\pi}^{-1}$  in the definition of  $S_\pi$  in (1.2) are therefore diagonalized by  $\hat{V}_1$ , and  $V = \hat{V}_1$  is an orthogonal eigenbasis for  $S_\pi$ . According to Definition 2.6, the HO-GSVD  $A_1 = U_1 \Sigma_1 V^T$  is obtained from  $B_1 = A_1 V^{-T} = A_1 \hat{V}_1 = \bar{U}_1 \bar{\Sigma}_1$ , so that  $U_1 = \hat{U}_1$  and  $\Sigma_1 = \hat{\Sigma}_1$ . If  $m_1 < n$ , at least  $n - m_1$  columns are zero, which can be dropped to obtain  $U_1$  of size  $m_1 \times m_1$ , and if  $m_1 > n$ ,  $m_1 - n$  additional columns of  $U_1$  can be found using  $\ker(A_1^T)$  (see Lemma 2.5 for rank-deficient  $A_1$ ).  $\square$

The HO-GSVD from Definition 2.6 can also be related to the GSVD. For the special case that  $N = 2$ ,  $A_1 \in \mathbb{R}^{m_1 \times n}$  with  $m_1 \geq n$  and  $\text{rank}(A_1) = n$  and an arbitrary  $A_2 \in \mathbb{R}^{m_2 \times n}$ , it can be shown that the HO-GSVD yields  $\Sigma_i$  with  $\Sigma_1^T \Sigma_1 + \Sigma_2^T \Sigma_2 = I$  and orthogonal  $U_1$  and  $U_2$ .

**THEOREM 5.2.** *For  $N = 2$  and  $\pi > 0$ , the HO-CSD from Definition 2.4 yields the standard CSD and the HO-GSVD from Definition 2.6 yields the standard GSVD.*

*Proof.* Since  $(Q_i^T Q_i + \pi I)^{-1}$  and  $Q_i^T Q_i$  with  $i = 1, 2$  and  $Q_1^T Q_1 + Q_2^T Q_2 = I$  share the same eigenspace for any  $\pi \in \mathbb{R}_{++}$  [7, Ch. 2], the eigenvectors  $z_k$  for  $T_\pi$  can be chosen such that they are right singular vectors for  $Q_1$  and  $Q_2$ . Let  $b_{i,k}$  denote the columns of  $B_i = Q_i Z$ ; then for  $j \neq k$ ,  $b_{i,k}^T b_{i,j} = z_k^T Q_i^T Q_i z_j = \hat{\sigma}_{i,j}^2 \hat{u}_{i,k}^T \hat{u}_{i,j} = 0$ , where  $\hat{\sigma}_\times$  and  $\hat{u}_\times$  denote standard singular values and left singular vectors, respectively. Hence, from  $U_i \Sigma_i = B_i$ , the columns of  $U_i$  are either zero or orthonormal. Substituting

$Q_i = U_i \Sigma_i V^T$  into  $Q_1^T Q_1 + Q_2^T Q_2 = I$  yields  $Z \Sigma_1^T \Sigma_1 Z^T + Z \Sigma_2^T \Sigma_2 Z^T = I$ , and from  $Z^T Z = I$ , it follows that  $\Sigma_1^T \Sigma_1 + \Sigma_2^T \Sigma_2 = I$ . The claim on the HO-GSVD follows from Theorem 5.2 with  $V = R^T Z$ . To obtain  $U_i$  of size  $m_i \times m_i$ , consider the closing remarks of the proof of Theorem 5.1.  $\square$

*Remark 5.3.* Lemma 5.2 shows that for  $N = 2$  the three matrices,  $T_\pi$ ,  $Q_1^T Q_1$ , and  $Q_2^T Q_2$ , share the same eigenspace, but not every eigendecomposition of  $T_\pi$  yields eigenvectors that are parallel to those of  $Q_1^T Q_1$  and  $Q_2^T Q_2$ . For example, suppose that  $\dim(\ker(Q_i)) = 1$  and that  $q_i \in \ker(Q_i)$ ,  $i = 1, 2$ , are linearly independent. From pre- and post-multiplying  $Q_1^T Q_1 + Q_2^T Q_2 = I$  with  $q_1^T$  and  $q_2$ , it holds that  $q_1^T q_2 = 0$ . It follows that  $\dim(\mathcal{T}_1) = 2$ , so that  $T_\pi$  has a semisimple eigenvalue. When the associated eigenvectors are computed using numerical software, these will not necessarily be parallel to  $q_1$  and  $q_2$ , and the HO-CSD will not necessarily yield orthonormal matrices  $U_i$ .

The HO-GSVD from Definition 2.6 can also be compared with the full-rank HO-GSVD [23]. For  $N = 2$  and full-rank matrices  $A_i$ , both HO-GSVDs have been shown to be equivalent to the GSVD. Both HO-GSVDs have also been shown to yield the SVD of  $A_j$  when  $A_i = I$  for  $i \neq j$ . For  $N > 2$ , however, the HO-GSVD from Definition 2.6 and [23] will in general *not* yield identical factorizations  $A_i = U_i \Sigma_i V^T$ , even when  $\text{rank}(A_i) = n$ . This can be seen by comparing the eigenspaces of  $T_\pi$  from (2.5) for varying  $\pi$ , where  $\pi = 0$  corresponds to the standard HO-CSD [29]. For  $N = 2$ , the eigenvectors of  $T_\pi$  are independent of the value of  $\pi$  because its eigenvectors are fixed by the orthogonality property  $Q_1^T Q_1 + Q_2^T Q_2 = I$ , while for  $N > 2$  this property is lost. From Theorem 2.3, it follows that the same holds for the HO-GSVD. However, it can be shown that in case the matrices  $A_i$  and  $Q_i$  have full rank, then the common HO-CSD and HO-GSVD subspaces will be the same for any value of  $\pi$  (Corollary 3.8) and  $N > 2$ . Moreover, it follows from Theorem 4.1 that in the full-rank case, the eigenvectors of  $T_\pi$  converge to those of the standard HO-GSVD as  $\pi \rightarrow 0$ .

**6. Computing the HO-GSVD.** The early literature on the standard GSVD ( $N = 2$ ) identified numerical issues for the case that  $A$  from (2.1) and therefore  $R$  are ill-conditioned [25, 27, 21]. This problem was resolved by basing the GSVD computation on the CSD, hereby avoiding computing the inverse of  $R$ . For the same reason, we propose to compute the HO-GSVD (1.1) using Algorithm 6.1, which in turn calls Algorithm 6.2, which computes the corresponding HO-CSD (section 6.1). An experimental MATLAB implementation is provided in [13].

Given a dataset  $A_1, \dots, A_N$  and a parameter  $\pi > 0$ , Algorithm 6.1 starts by verifying whether assumption (2.3) is satisfied.<sup>1</sup> If  $\text{rank}(A) < n$ , the dataset is padded using a matrix  $A_{N+1}$  that spans the nullspace of  $A$ , which is obtained from the standard SVD of  $A$  [7, Ch. 2.4.2]. Algorithm 6.1 proceeds by computing the thin QR factorization (2.1) on line 4 and calling the HO-CSD (section 6.1) on line 5. Finally, the shared factor matrix  $V$  is obtained on line 6 using Theorem 2.3. Without padding, the computational complexity of the HO-GSVD is dominated by the complexity of the HO-CSD, which is  $\mathcal{O}(mn^2)$  for  $m_i \geq n$ . Note that the experimental implementation [13] returns additional outputs, such as the eigenvalues of  $T_\pi$  and the indices corresponding to the isolated subspace.

<sup>1</sup>In [13], the rank is determined using  $\text{rank}(A) = |\{i \in \{1, \dots, \min(m, n)\} \mid \sigma_i > \epsilon_0\}|$ , where  $\sigma_i$  are the standard singular values of  $A$  and  $\epsilon_0 \geq 0$  is an optional parameter.

**Algorithm 6.1** HO-GSVD Computation.

---

**Input:**  $A_1, \dots, A_N \in \mathbb{R}^{m_i \times n}$ ,  $\pi \in \mathbb{R}_+$   
**Output:** Factorizations  $A_i = U_i \Sigma_i V^T$ ,  $i = 1, \dots, N$   
1: **if**  $\text{rank}(A) < n$  **then**  
2: Find  $A_{N+1}$  with  $\text{range}(A_{N+1}) = \ker(A)$  and append to  $A$   
3: **end if**  
4: Obtain  $Q_i R = A_i$ ,  $i = 1, \dots, N$ , from (2.1)  
5: Obtain  $U_i \Sigma_i Z^T = Q_i$ ,  $i = 1, \dots, N$ , from Algorithm 6.2  
6: Set  $V = R^T Z$

---

**Algorithm 6.2** HO-CSD Computation.

---

**Input:**  $Q_1, \dots, Q_N \in \mathbb{R}^{m_i \times n}$  satisfying (2.2),  $\pi, \epsilon_1, \epsilon_2 \in \mathbb{R}_+$   
**Output:** Factorizations  $Q_i = U_i \Sigma_i Z^T$ ,  $i = 1, \dots, N$   
1: Obtain  $\hat{Q}_i \hat{R}_i = [Q_i^T, \sqrt{\pi} I]^T$ ,  $i = 1, \dots, N$   $\mathcal{O}(Nn^3)$   
2: Form  $\hat{R} = [\hat{R}_1^{-1}, \dots, \hat{R}_N^{-1}]$   $\mathcal{O}(Nn^3)$   
3: Obtain  $U_{\hat{R}} \Sigma_{\hat{R}} V_{\hat{R}}^T = \hat{R}$   $\mathcal{O}(mn^2)$   
4: Set  $[z_1, \dots, z_n] = U_{\hat{R}}$  and  $\tau_k = \sigma_{\hat{R},k}^2 / N$ ,  $k = 1, \dots, n$   
5: Determine  $n_{\text{iso}}(\epsilon_1)$  using (6.6)  
6: **if**  $n_{\text{iso}}(\epsilon_1) \geq 2$  **then**  $\mathcal{O}(mnn_{\text{iso}})$   
7: Replace  $z_k$ , with  $\tilde{z}_k$ ,  $k = 1, \dots, n_{\text{iso}}(\epsilon_1)$ , from Algorithm 6.3  
8: **end if**  
9: **for**  $i = 1, \dots, N$  and  $k = 1, \dots, n$  **do**  
10: Set  $\sigma_{i,k} = \|Q_i z_k\|_2$   
11: **if**  $\sigma_{i,k} > \epsilon_2$  **then**  
12: Set  $u_{i,k} = Q_i z_k / \sigma_{i,k}$   
13: **else**  
14: Apply Lemma 2.5  
15: **end if**  
16: **end for**  $\mathcal{O}(mn^2)$

---

**6.1. Computing the HO-CSD.** According to Definition 2.4, the HO-CSD requires computing the eigenvectors and eigenvalues of  $T_\pi$ , which, according to (2.5), in turn requires accumulating  $(Q_i^T Q_i + \pi I)^{-1}$ ,  $i = 1, \dots, N$ . In order to avoid forming the products  $Q_i^T Q_i$ , each term  $Q_i^T Q_i + \pi I$  is factored as

$$(6.1) \quad Q_i^T Q_i + \pi I = \begin{bmatrix} Q_i \\ \sqrt{\pi} I \end{bmatrix}^T \begin{bmatrix} Q_i \\ \sqrt{\pi} I \end{bmatrix} =: (\hat{Q}_i \hat{R}_i)^T (\hat{Q}_i \hat{R}_i),$$

where  $\hat{Q}_i \hat{R}_i$  is the thin QR factorisation of  $[Q_i^T, \sqrt{\pi} I]^T$ . Using (6.1), equation (2.5) can be reformulated as

$$(6.2) \quad T_\pi = \frac{1}{N} \sum_{i=1}^N \hat{R}_i^{-1} \hat{R}_i^{-T} = \frac{1}{N} \underbrace{[\hat{R}_1^{-1} \dots \hat{R}_N^{-1}]}_{=: \hat{R}} \begin{bmatrix} \hat{R}_1^{-T} \\ \vdots \\ \hat{R}_N^{-T} \end{bmatrix} = \frac{1}{N} \hat{R} \hat{R}^T.$$

Substituting the thin SVD of  $\hat{R}$ ,  $\hat{R} = U_{\hat{R}} \Sigma_{\hat{R}} V_{\hat{R}}^T$ , into (6.2) yields

$$(6.3) \quad T_\pi = \frac{1}{N} (U_{\hat{R}} \Sigma_{\hat{R}} V_{\hat{R}}^T) (U_{\hat{R}} \Sigma_{\hat{R}} V_{\hat{R}}^T)^T = \frac{1}{N} U_{\hat{R}} \Sigma_{\hat{R}}^2 U_{\hat{R}}^T,$$

from which the eigenvector matrix  $Z$  and the eigenvalues  $\tau_k$  of  $T_\pi$  are obtained as

$$(6.4) \quad Z = U_{\hat{R}}, \quad \tau_k = \frac{\sigma_{\hat{R},k}^2}{N}, \quad k = 1, \dots, n,$$

where  $\sigma_{\hat{R},k}$  are the standard singular values of  $\hat{R}$ .

Equations (6.1)–(6.4) are implemented on lines 1–4 of Algorithm 6.2. To avoid the inversion of potentially ill-conditioned  $\hat{R}_i$  on line 2, the condition number  $\kappa(\hat{R}_i) = \kappa(Q_i^T Q_i + \pi I)$  can be controlled by choosing  $\pi$  as follows. Let  $\kappa_{\max} > 1$  and  $\hat{\sigma}_{i,\min}$  and  $\hat{\sigma}_{i,\max}$  denote the minimum and maximum standard singular values of  $Q_i$ , respectively; then  $\kappa(Q_i^T Q_i + \pi I) \leq \kappa_{\max} \forall i = 1, \dots, N$  if  $\pi$  is chosen such that

$$(6.5) \quad \pi \geq \min_{i \in \{1, \dots, N\}} \frac{\hat{\sigma}_{i,\max} - \kappa_{\max} \hat{\sigma}_{i,\min}}{\kappa_{\max} - 1}.$$

Considering that  $0 \leq \hat{\sigma}_{i,\min} \leq \hat{\sigma}_{i,\max} \leq 1$ , choosing  $\pi = 1 \times 10^{-3}$  guarantees that  $\kappa(\hat{R}_i) \leq 1000$ , which is used as the default value in [13]. Assuming that the eigenvalues of  $T_\pi$  on line 4 are sorted in descending order, the dimension of the isolated subspace is determined by

$$(6.6) \quad n_{\text{iso}}(\epsilon_1) := \max_{k \in \{1, \dots, n\}} k \quad \text{s.t.} \quad \tau_{\max} - \tau_k \leq \epsilon_1 (\tau_{\max} - \tau_{\min}),$$

where  $\tau_k$  is the corresponding eigenvalue of  $T_\pi$  and the scalar  $\epsilon_1 \geq 0$  is introduced to account for finite machine precision. Note the trade-off between (6.5) and (6.6): For increasing  $\pi$ , the difference  $\tau_{\max} - \tau_{\min}$  rapidly decreases, as shown in Figure 1. If the difference  $\tau_{\max} - \tau_{\min}$  is too small, numerical inaccuracies can lead to a wrong selection of directions associated with the isolated HO-CSD subspace. The same problem arises when determining the common HO-CSD subspace. If  $n_{\text{iso}}(\epsilon_1) \geq 1$ ,  $T_\pi$  has an eigenvalue that is equal to  $\tau_{\max}$  with algebraic and geometric multiplicity greater than 1, in which case the corresponding eigenvectors of  $T_\pi$  must be aligned with the shared standard right singular vectors of  $Q_1, \dots, Q_N$ . Otherwise, it is not guaranteed that the corresponding generalized singular values are either 0 or 1 (see Remark 6.1). To align the eigenvectors of  $T_\pi$ , Algorithm 6.1 makes use of Algorithm 6.3 on line 7 (section 6.2).

Finally, the matrices  $\Sigma_i$  and  $U_i$  are computed in the loop starting on line 9. If  $\sigma_{i,k} > \epsilon_2$ , where  $\epsilon_2 \geq 0$  is the zero tolerance for the generalized singular values, Algorithm 6.2 computes  $u_{i,k}$  according to Definition 2.4, but if  $\sigma_{i,k} \leq \epsilon_2$ , the algorithm substitutes a standard left singular vector of  $Q_i$  associated with  $\ker(Q_i^T)$  (see Lemma 2.5). If no such left singular vector exists, Algorithm 6.2 substitutes column  $k$  of  $Q_i$ .

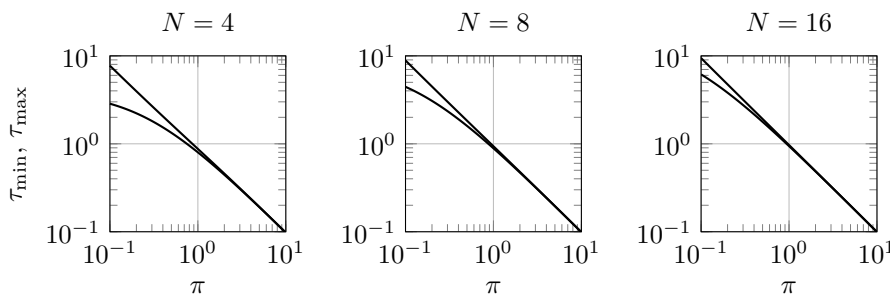


FIG. 1. Minimum and maximum eigenvalues of  $T_\pi$  as a function of  $\pi$  for different  $N$ .



An upper bound on the algorithm complexity for  $m_i \geq n$  is given by summing the shaded numbers on the right-hand side of Algorithm 6.1. The algorithm mainly uses standard routines, such as the QR decomposition or the eigendecomposition, which require roughly  $\mathcal{O}(mn^2)$  floating-point operations. If the full factorization (1.1) is not required but only the common or isolated subspace, alternative algorithms exist that compute the common HO-CSD subspace from intersecting the pairwise common HO-GSVD subspaces of  $Q_i$  and  $Q_{i+1}$  for  $i = 1, \dots, N-1$  [29, 22]. The pairwise subproblems can be solved by the standard GSVD, which exists as a built-in function in most scientific computing packages.

*Remark 6.1.* To see why the eigenvectors spanning the isolated subspace must be aligned, consider the case  $N = 3$  with  $\ker(Q_i) \neq \emptyset$ ,  $i = 1, 2, 3$ , and

$$\dim(\ker(Q_1)) = \dim(\ker(Q_2)) = 1, \quad \dim(\ker(Q_3)) = 2.$$

Furthermore, suppose that  $\mathcal{R}_1 := \text{range}(Q_1) \cap \ker(Q_2) \cap \ker(Q_3) \neq \emptyset$  and  $\mathcal{R}_2 := \text{range}(Q_2) \cap \ker(Q_1) \cap \ker(Q_3) \neq \emptyset$ , in which case  $T_\pi$  has an eigenvalue equal to  $\tau_{\max}$  with algebraic and geometric multiplicity equal to 2. The corresponding eigenvectors span  $\mathcal{R}_1 \cap \mathcal{R}_2 = \mathcal{T}_1\{Q_1, Q_2, Q_3\}$ , but are not necessarily aligned with the right standard singular vectors spanning range and kernel of  $Q_1$  and  $Q_2$ . According to Definition 2.4 and Theorem 3.5, the corresponding generalized singular values of  $Q_1$  and  $Q_2$  are (simultaneously) 0 or 1 iff the eigenvectors of  $T_\pi$  are aligned with the standard right singular vectors (see also line 11 of Algorithm 6.2).

**6.2. Computing the isolated subspace.** It follows from Definition 3.4 and Theorem 3.5 that if  $\dim(\mathcal{T}_1\{Q_1, \dots, Q_N\}) = n_{\text{iso}}(0)$ , then  $T_\pi$  has  $n_{\text{iso}}(0)$  eigenvalues equal to  $\tau_{\max}$  and each of the corresponding eigenvectors can be chosen such that it is a standard right singular vector for each  $Q_i$ . However, when  $n_{\text{iso}}(0) > 1$  the eigendecomposition of  $T_\pi$  will produce an arbitrary set of orthogonal vectors that span  $\mathcal{T}_1\{Q_1, \dots, Q_N\}$ , but that are not necessarily parallel to the shared right standard singular vectors (see Remark 6.1).

Let  $z_k$ ,  $k \leq n_{\text{iso}}(\epsilon_1)$ , be the eigenvectors obtained from (6.6) and  $Z_{\text{iso}} \in \mathbb{R}^{n \times n_{\text{iso}}(\epsilon_1)}$  contain these eigenvectors as columns. In case of infinite machine precision ( $\epsilon_1 = 0$ ), the standard singular values of  $Q_i Z_{\text{iso}}$  are either 1 or 0  $\forall i = 1, \dots, N$ , and one way to align the eigenvectors is therefore to set  $\tilde{z}_k = Z_{\text{iso}} \tilde{v}$ , where  $\tilde{v}$  is a standard right singular vector associated with a unit singular value of  $Q_i Z_{\text{iso}}$ . In case of infinite machine precision, there exist exactly  $n_{\text{iso}}(0)$  such vectors  $\tilde{z}_1, \dots, \tilde{z}_{n_{\text{iso}}(0)}$  with  $\text{span}\{\tilde{z}_1, \dots, \tilde{z}_{n_{\text{iso}}(0)}\} = \text{span}\{z_1, \dots, z_{n_{\text{iso}}(0)}\}$ . Since from Lemma 2.2 it follows that  $\|T_\pi z\|_2$  is maximized if  $z$  is parallel to the standard right singular vector of  $Q_i$  associated with the largest singular value, in the case of finite machine precision ( $\epsilon_1 > 0$ ) it appears natural to order the  $Q_i$  by magnitude of  $\|Q_i Z_{\text{iso}}\|_2$ , select the standard right singular vector  $\tilde{v}$  of  $Q_i Z_{\text{iso}}$  associated with the largest  $\|Q_i Z_{\text{iso}}\|_2$  over  $i = 1, \dots, N$ , and obtain an aligned eigenvector as  $Q_i Z_{\text{iso}} \tilde{v}$  (see [29, pp. 30–38]). Such a procedure is implemented in Algorithm 6.3, which computes a sequence of ever-thinner standard SVDs to obtain an aligned basis  $\tilde{Z}_{\text{iso}}$  from  $Z_{\text{iso}}$ , where  $\tilde{Z}_{\text{iso}}^T \tilde{Z}_{\text{iso}} = I$  and the columns of  $\tilde{Z}_{\text{iso}}$  span the same subspace as those of  $Z_{\text{iso}}$ .

The iterate  $X_k$ ,  $k = 0$ , is initialized on line 1 using  $Z_{\text{iso}}$ . In the first iteration, Algorithm 6.3 selects the index  $i$  that has the maximum amplification in the subspace spanned by the columns of  $Z_{\text{iso}}$ , i.e., by comparing  $\|Q_i Z_{\text{iso}}\|_2$ ,  $i = 1, \dots, N$ . The corresponding direction  $Z_{\text{iso}} \tilde{v}_1$  is assigned to the first column of  $\tilde{Z}_{\text{iso}}$ . Next, the algorithm selects  $n_{\text{iso}} - 1$  remaining directions that are orthogonal to  $Z_{\text{iso}} \tilde{v}_1$ . Since  $Z_{\text{iso}}$  is orthogonal and at every iteration  $\tilde{v}_1$  is orthogonal to  $\tilde{v}_2, \dots, \tilde{v}_{n_{\text{iso}}-k}$ , the resulting  $\tilde{Z}_{\text{iso}}$

**Algorithm 6.3** Isolated Subspace Computation**Input:**  $Q_1, \dots, Q_N, Z_{\text{iso}}$ **Output:** Aligned basis  $\tilde{Z}_{\text{iso}}$ 

```

1: Initialize  $X_0 := Z_{\text{iso}}$ 
2: for  $k = 0, \dots, n_{\text{iso}} - 1$  do
3:   Select  $p := \arg\max_{i=1, \dots, N} \|Q_i X_k\|_2$ 
4:   Obtain the right singular vectors  $\tilde{v}_1, \dots, \tilde{v}_{n_{\text{iso}}-k}$  of  $Q_p X_k$ 
5:   Assign  $X_k \tilde{v}_1$  to column  $k+1$  of  $\tilde{Z}_{\text{iso}}$ 
6:   Update  $X_{k+1} := X_k [\tilde{v}_2 \ \dots \ \tilde{v}_{n_{\text{iso}}-k}]$ 
7: end for

```

 $\mathcal{O}(mn n_{\text{iso}})$ 

is orthogonal too. Note that the size of  $X_k$  decreases at every iteration and that line 6 is not executed at the last iteration.

**7. Applications.** The standard (HO-)GSVD has already been applied in various fields such as bioinformatics [23, 30, 31], medicine [17], acoustics [26], and control theory [14]. On one hand, the (HO-)GSVD can be used as an exact matrix decomposition to simplify or decouple problems involving several matrices sharing the same column or row dimension, such as generalized eigenvalue or generalized total least squares problems [3]. On the other hand, the (HO-)GSVD can be used to compare  $N$  sets of measurements tabulated in matrices  $A_1, \dots, A_N$ , where matrix  $i$  represents a different organism, class, or experiment, for example. The columns of  $A_i$  usually represent a sampled coordinate, such as time or position, whereas the rows of  $A_i$  are class-specific variables that vary along the sampled coordinate.

In the form of the HO-GSVD factorization (1.4), row  $j$  of  $A_i$  is represented as a linear combination of the right basis vectors  $v_1, \dots, v_n$ , which are also weighted by the generalized singular values  $\sigma_{i,k}$ . In general, the right basis vectors are not orthogonal for  $N \geq 2$ . However, suppose  $A_1, \dots, A_N$  are such that there exists  $v \in \mathbb{R}^n$  such that  $A_i^T A_i v \neq 0$  for some  $i$  and  $A_j^T A_j v = 0$  for  $j \neq i$ , i.e.,  $v$  contributes exclusively to the rows of  $A_i$ ; then, according to Corollary 3.7,  $v$  will be an eigenvector of  $T_\pi$  associated with an eigenvalue equal to  $\tau_{\max}$ . Due to the continuity of the eigenvalues of  $T_\pi$ , it also follows that if  $\tau_k \approx \tau_{\max}$ , the corresponding right basis vector is almost exclusively used to represent the rows of  $A_i$  (see also [22, Ch. 2.3.3]). Similarly, if there exists  $\tilde{v} \in \mathbb{R}^n$  such that  $A_i^T A_i \tilde{v} = A_j^T A_j \tilde{v}$  for  $i, j = 1, \dots, N$ , then according to statement 3.3c of Corollary 3.3,  $D_{j,\pi}^{-1} \tilde{v}$  will be an eigenvector of  $T_\pi$  associated with an eigenvalue equal to  $\tau_{\min}$ . Among other cases, the condition  $A_i^T A_i \tilde{v} = A_j^T A_j \tilde{v}$  holds if  $A_1, \dots, A_N$  share a singular vector  $\tilde{v}$  associated with an identical singular value.

To examine the effect of certain right basis vectors on the rows of class  $i$ ,  $A_i$  can be reconstructed using a reduced set of right basis vectors, e.g., computing

$$(7.1) \quad A_{i,\text{iso}} := \sum_{k \in \mathcal{I}_1} \sigma_{i,k} u_{i,k} v_k^T$$

yields the reconstruction of  $A_i$  using the right basis vectors that are, in the sense of (6.6), exclusively used by class  $i$ . To see the effect of the right basis vectors associated with the common subspace,  $A_i$  can be reconstructed by summing over  $k \in \mathcal{I}_N$ .

**7.1. Numerical example.** To illustrate an example application of the HO-GSVD for rank-deficient matrices, consider the CIFAR-10 dataset, which is a collection of images used to evaluate machine learning and computer vision algorithms [16].

TABLE 1

Sample matrices extracted from the first batch of the CIFAR-10 dataset. The rows of each  $A_i \in \mathbb{R}^{m_i \times n}$  represent vectorized  $32 \times 32$  pixels large images.

	Class	$m_i$	$\text{rank}(A_i)$	$ T_i^1 $	$\dim(\mathcal{T}_1\{A_1, \dots, A_4\})$
$A_1$	Automobile	974	974	50	300
$A_2$	Cat	1016	1016	92	
$A_3$	Ship	1025	1025	101	
$A_4$	Truck	981	981	57	
$A$		3996	3072		

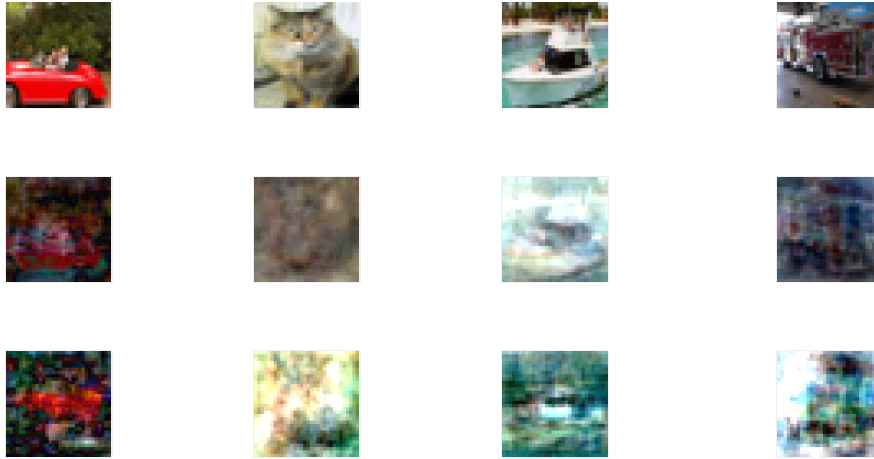


FIG. 2. First row: Example rows of  $A_1, \dots, A_4$  (left to right) reshaped into  $32 \times 32$  pixels large images. Second row: Moduli of example rows of  $A_{1,iso}, \dots, A_{4,iso}$ , where  $A_{i,iso}$  is reconstructed using right basis vectors from the isolated subspace only. Third row: Moduli of isolated right basis vectors that have the largest weight in each image.

The CIFAR-10 dataset provides 60,000  $32 \times 32$  color images of 10 different objects (classes), such as automobiles, cats, ships, and trucks. The dataset is split into 5 training batches and 1 test batch, where each batch contains 10,000 images.<sup>2</sup> To obtain rank-deficient  $A_i$  and give an example of the isolated subspace, the HO-GSVD is applied to the first training batch only and used to analyze a subset of  $N = 4$  randomly selected classes shown in Table 1. The following example can be downloaded from [13].

The images are vectorized and grouped in the matrices  $A_i \in \mathbb{R}^{m_i \times n}$ , where  $n = 32 \times 32 \times 3 = 3072$  and  $0 \leq A_i \leq 1$  (elementwise). Each  $A_i$  of the first batch of the CIFAR-10 dataset is such that  $r_i := \text{rank}(A_i) < n$ , but with the  $N = 4$  selected classes the stacked  $A \in \mathbb{R}^{m \times n}$  satisfies  $m > n$  and  $\text{rank}(A) = n$  (using  $\epsilon_0 = 10^{-14}$ ). For  $N \leq 3$  and  $\text{rank}(A) < n$ , the matrix  $A$  could be padded using an additional matrix  $A_{N+1}$  (see Remark 3.9).

The first row of Figure 2 displays row  $j_i$  for each class  $i$  as an image. The rows  $j_i$  for class  $i$  are  $j_1 = 16, j_2 = 19, j_3 = 40$ , and  $j_4 = 50$ , and have been selected to yield interpretable images for their reconstruction in the isolated subspace.

Using the HO-GSVD, the image  $j$  of class  $i$  is a row of  $A_i$  that can be represented as  $\sum_k (e_j^T u_{i,k}) \sigma_{i,k} v_k^T$ , where  $e_j$  is a standard basis vector and  $|e_j^T u_{i,k}| \leq 1$ . The

<sup>2</sup>Note that the training batches do not contain an equal number of images of each class.

columns  $v_k$  of the matrix  $V \in \mathbb{R}^{n \times n}$  with  $\det(V) \neq 0$  can be interpreted as “basis images” for the space of  $32 \times 32$  images, and class  $i$  uses  $r_i$  columns of  $V$  to represent its sample images. Note that the columns of  $V$  are not orthogonal, and some right basis vectors can therefore “cancel” each other out. The third row of Figure 2 visualizes right basis vectors  $v_{20}$ ,  $v_{82}$ ,  $v_{203}$ , and  $v_{278}$ , which are all associated with the isolated subspace (see the subsequent paragraphs).

The parameter  $\pi$  is chosen as  $\pi = 1/N = 0.25$ , which results in  $\kappa(Q_i^T Q_i + \pi I) \leq 5$ ,  $\tau_{\min} = 2$ , and  $\tau_{\max} = 3.2$ . The  $n = 3072$  eigenvalues  $\tau_k$  of  $T_\pi$  are shown in the first row of Figure 3, where  $\tau_k$  is displayed relative to  $\tau_{\min}$  and  $\tau_{\max}$  as  $(\tau_k - \tau_{\min})/(\tau_{\max} - \tau_{\min})$  sorted in descending order. It can be seen that most eigenvalues are closer to  $\tau_{\max}$  than  $\tau_{\min}$ , and that  $\tau_k \gg \tau_{\min} \quad \forall k$ , i.e., the common HO-GSVD subspace is empty. Using a tolerance of  $\epsilon_1 = 10^{-6}$ , the dimension of the isolated HO-CSD subspace is estimated as  $n_{\text{iso}}(\epsilon_1) = 300$ . The number of isolated directions per class is shown in Table 1 (column  $|\mathcal{I}_1^i|$ ).

The generalized singular values  $\Sigma_i = \text{diag}(\sigma_{i,1}, \dots, \sigma_{i,n})$  are shown in the second to fifth rows of Figure 3. For indices  $k \in \mathcal{I}_1$  that are associated with the isolated subspace, the generalized singular values are either 0 or 1. Due to numerical inaccuracies, the separation between  $\sigma_{i,k}$ ,  $k \in \mathcal{I}_1$ , and  $\sigma_{i,j}$ ,  $j \notin \mathcal{I}_1$ , is not sharp, i.e., the generalized singular values of the automobile class soar at index  $k = 301 \notin \mathcal{I}_1$  before decreasing at larger indices. Note that even though some  $\sigma_{i,k}$  equal  $1/\sqrt{N} = 0.5$ , which is the same magnitude as expected for an index  $k$  associated with the common subspace, the common subspace is empty, as can be seen from the first row of Figure 3.

From Figure 3, it becomes clear that each class  $i$  uses its own subset of isolated basis images as well as  $n - n_{\text{iso}} = 2072$  other columns of  $V$  to form its  $m_i$  samples. Class  $i$  can be reconstructed using (7.1) to obtain  $A_{i,\text{iso}}$ , which considers indices  $k \in \mathcal{I}_1^i$  only. The second row of Figure 2 shows row  $j_i$  of  $A_{i,\text{iso}}$ , where some degree of resemblance between the original and reconstructed images exists. Examples of right basis vectors are given in the third row of Figure 2, which shows  $v_{20}$ ,  $v_{82}$ ,  $v_{203}$ , and  $v_{278}$ , each of which is associated with the isolated subspace of classes  $i = 1, \dots, 4$ . The right basis vectors have been selected by determining those  $k$  that maximize  $|e_{j_i}^T u_{i,k}|$  for each image  $j_i$ , i.e., those right basis vectors have a large contribution to image  $j_i$ . As for the second row of Figure 2, it can be seen that the third row of Figure 2 resembles the original image.

To complement the numerical example, the dataset  $A$  is modified in order to artificially introduce a nonempty common subspace. According to Corollary 3.3, the common HO-GSVD subspace,  $\mathcal{S}_4\{A_1, \dots, A_4\}$ , is nonempty iff the condition  $A_i^T A_i \tilde{v} = A_j^T A_j \tilde{v}$  holds  $\forall i, j = 1, \dots, 4$  and for some  $\tilde{v}$ , which can be written out as

$$(7.2) \quad \left( \begin{bmatrix} a_{i,1}^T a_{i,1} & \dots & a_{i,1}^T a_{i,n} \\ \vdots & \ddots & \vdots \\ a_{i,n}^T a_{i,1} & \dots & a_{i,n}^T a_{i,n} \end{bmatrix} - \begin{bmatrix} a_{j,1}^T a_{j,1} & \dots & a_{j,1}^T a_{j,n} \\ \vdots & \ddots & \vdots \\ a_{j,n}^T a_{j,1} & \dots & a_{j,n}^T a_{j,n} \end{bmatrix} \right) \tilde{v} = 0,$$

where  $a_{i,k} \in \mathbb{R}^{m_i}$  denotes column  $k$  of matrix  $A_i$ . If  $\tilde{v}$  is chosen as  $[1 \ 0 \ \dots \ 0]^T$ , condition (7.2) is tantamount to requiring that  $a_{i,k}^T a_{i,1} = a_{j,k}^T a_{j,1} \quad \forall i, j = 1, \dots, 4$  and for  $k = 1, \dots, n$ , i.e., the projection of column  $k$  onto the first column of class  $i$  must equal the projection of column  $k$  onto the first column of class  $j$ . Note that condition (7.2) is *not* equivalent to inserting an identical image  $x \in \mathbb{R}^n$  in each  $A_i$ , but a simple way to satisfy (7.2) is to set  $a_{i,1} = [1 \ 0 \ \dots \ 0]^T$  and zero out the first element of  $a_{i,k}$ ,  $k = 1, \dots, n$ , for all classes  $i = 1, \dots, 4$ . This way the first image of each  $A_i$  is replaced with a black square that has one red pixel in the left corner.

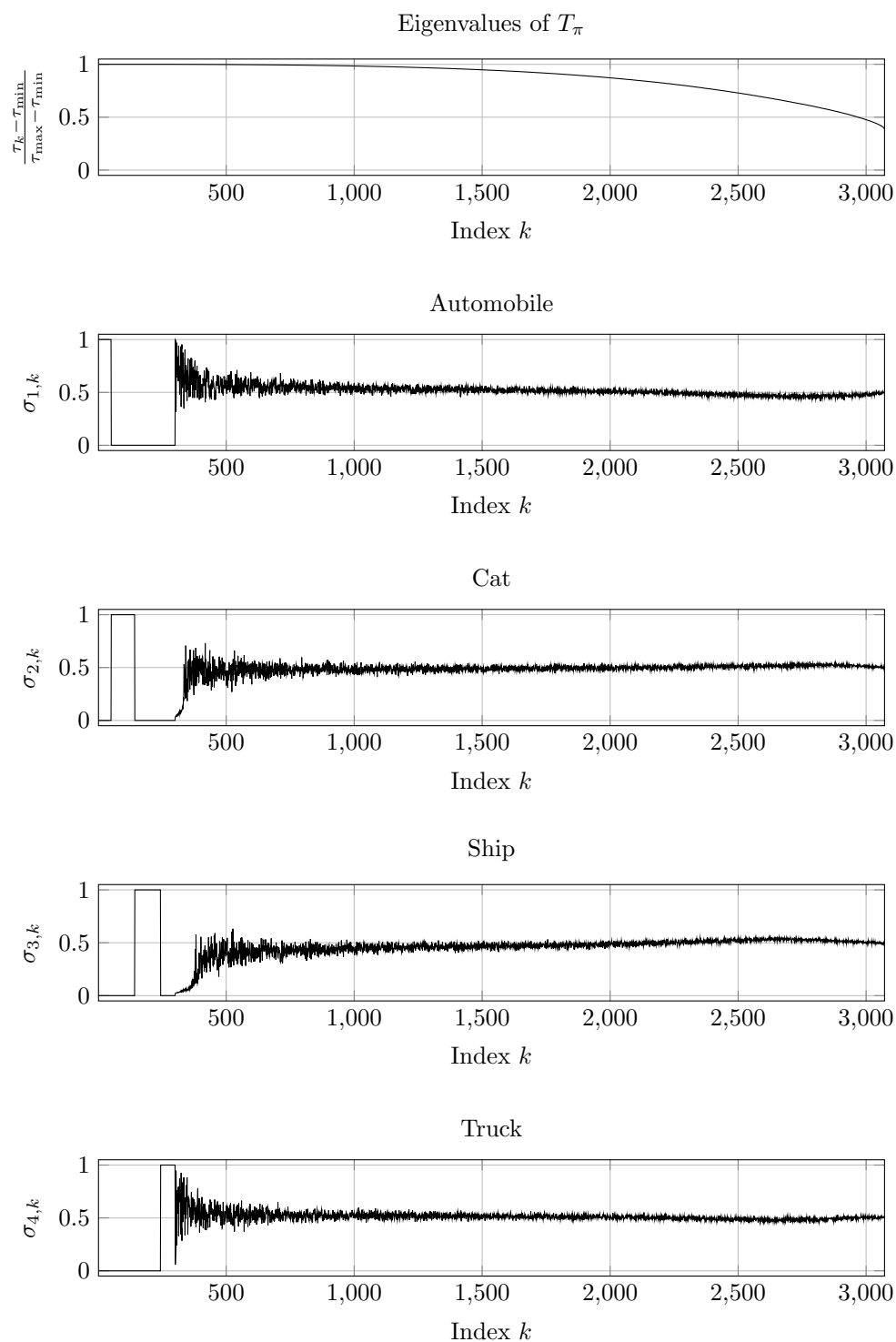


FIG. 3. Row 1: The  $n = 3072$  eigenvalues of  $T_\pi$  relative to the bounds  $\tau_{\min}$  and  $\tau_{\max}$ . Rows 2–5: Corresponding generalized singular values  $\Sigma_i = \text{diag}(\sigma_{i,1}, \dots, \sigma_{i,n})$  for classes  $i = 1, \dots, 4$ .

The eigenvalues of  $T_\pi$  and the generalized singular values for the modified dataset are shown in Figure 4. In the first row of Figure 4, it can be seen that  $\tau_k = \tau_{\min}$  for  $k = n$ , i.e., the modification successfully introduces a nonempty common subspace. The corresponding generalized singular values equal  $1/\sqrt{N} = 0.5$  for each class. By construction, the first row (image) of each  $A_i$  is orthogonal to all other rows of  $A_i$ , and therefore aligned with a shared standard right singular vector. The right basis vector associated with the common subspace,  $v_n$ , is therefore orthogonal to all other basis vectors, which is not the case in general. However, for this example it follows that  $v_n$  contributes equally to each of the matrices  $A_i$ , and for each class  $i$ ,  $v_n$  is used to represent the first image only.

**8. Conclusion.** In this paper, we have extended the standard HO-GSVD [23] to accommodate column rank deficient matrices. By adding the term  $\pi A^T A$  to each of the quotient terms  $D_{i,\pi} = A_i^T A_i + \pi A^T A$ , we shifted their eigenvalues and bounded them away from zero. This allowed us to omit the full-rank requirement on each  $A_i$ , consider the case that  $m_i < n$ , extend the HO-GSVD with the notion of isolated subspaces, and also resolve numerical issues associated with ill-conditioned  $A_i$  in the full-rank case.

The choice of adding a multiple of  $A^T A$  was motivated by the relationship between  $S_\pi$  and  $T_\pi$ , which yielded the same relationship as in [23] for  $\pi = 0$ . We bounded the eigenvalues of  $T_\pi$  and showed that the extremal eigenvalues are attained iff the corresponding eigenvectors are standard right singular vectors for each  $Q_i$  associated with a particular singular value. This led to the definition of the common and isolated HO-CSD (HO-GSVD) subspaces. In Appendix B, we also show that if the  $Q_i$  share a right singular vector  $v$  associated with a zero singular value for  $P$  matrices  $Q_i$  and with an identical singular value for the other  $N - P$  matrices  $Q_j$ , then  $T_\pi$  will have a particular eigenvalue  $\tau(P)$  associated with the eigenvector  $v$ . Future research could investigate whether a biconditional (“iff”) connection holds.

The parameter  $\pi$  was assumed to be positive, but otherwise left unspecified. The common and isolated HO-CSD and HO-GSVD subspaces are identified irrespective of the value of  $\pi$ , but other right basis vectors can be rotated for increasing values of  $\pi$ , and we have investigated the behavior of these vectors as  $\pi \rightarrow 0$  ( $\pi \ll \|A_i^T A_i\|_2 / \|A^T A\|_2$ ) and  $\pi \rightarrow \infty$  ( $\pi \gg \|A_i^T A_i\|_2 / \|A^T A\|_2$ ).

In addition, the choice of  $\pi$  also affects the condition number of  $Q_i^T Q_i + \pi I$ , which must be inverted to obtain  $T_\pi$ , as well as the range of admissible eigenvalues of  $T_\pi$ ,  $\tau_{\max} - \tau_{\min}$ . A large  $\pi$  improves the conditioning of  $Q_i^T Q_i + \pi I$  but also tightens the range of eigenvalues, which can lead to a wrong estimate of the common or isolated subspaces in the presence of numerical errors. The optimal choice of  $\pi$  remains unclear, and so future research could investigate the role of the weight  $\pi$ .

The majority of our developments were based on the HO-CSD. Using the QR factorization of  $A = [A_1^T, \dots, A_N^T]^T$ , each  $A_i$  was represented as  $A_i = Q_i R$ , and our findings were developed for  $Q_1, \dots, Q_N$ , which required  $A$  to have full column rank. For rank-deficient  $A$ , we have shown how  $A$  can be padded using an additional matrix  $A_{N+1}$  to guarantee that  $\det(R) \neq 0$ . The properties of  $A_1, \dots, A_N$  were inferred from the HO-CSD, which allowed us to avoid computing the inverse of a potentially ill-conditioned  $R$ , but a full factorization still requires one to invert the upper triangular matrices  $\hat{R}_i$  from the QR decomposition of  $[Q_i^T, \sqrt{\pi}I]^T$ . Future research could focus on finding a possibly iterative algorithm that finds the eigenvectors of  $T_\pi$  without the need for inverting the terms  $\hat{R}_i$ .

For the full-rank case, it has been shown that the common subspace can be found using a variational approach [29] and that the vectors  $v$  spanning the common

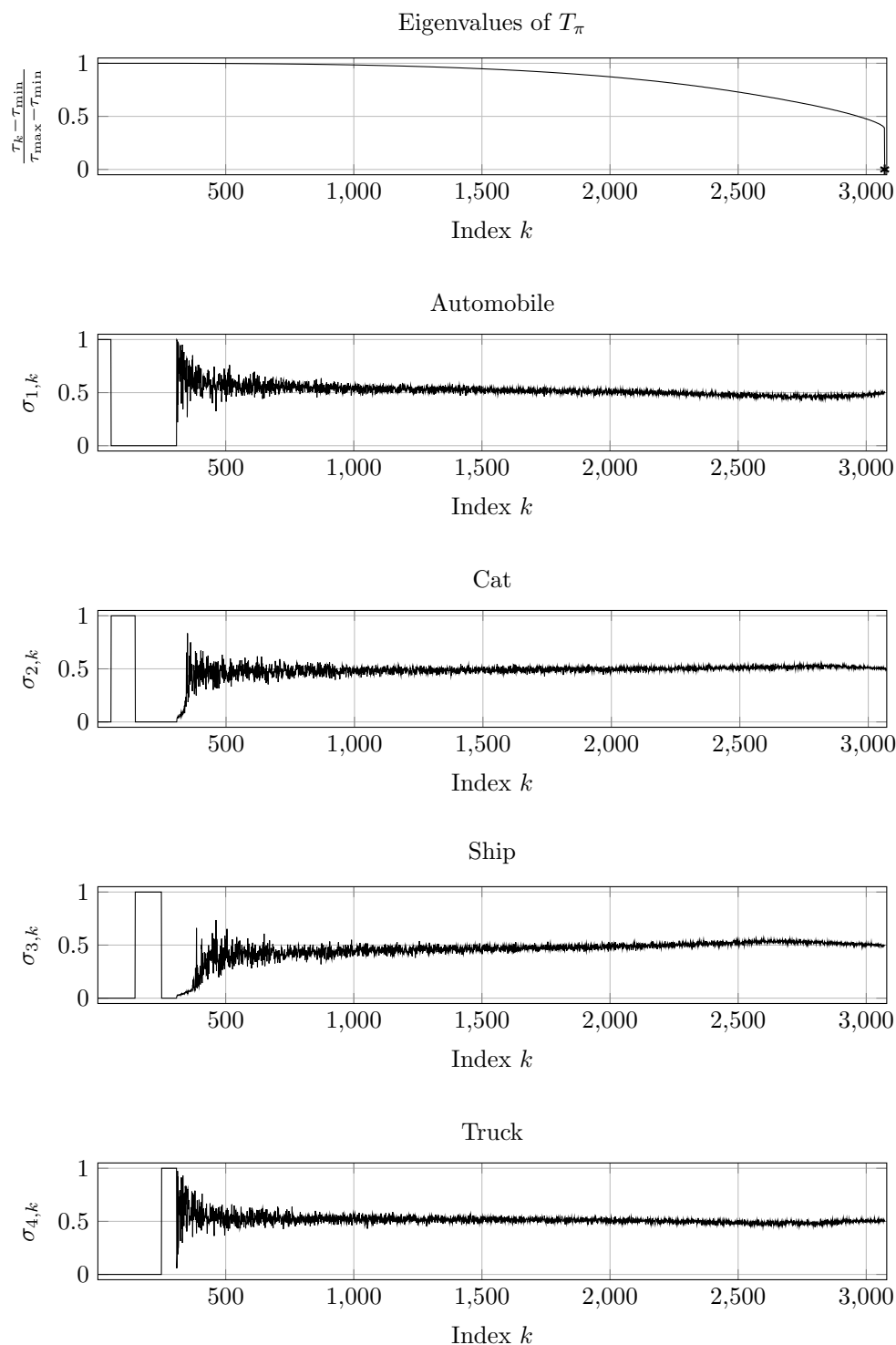


FIG. 4. Row 1: The  $n = 3072$  eigenvalues of  $T_\pi$  relative to the bounds  $\tau_{\min}$  and  $\tau_{\max}$  for the modified dataset, which has a one-dimensional common subspace associated with index  $k = 3072$  (marked by an asterisk). Rows 2–5: Corresponding generalized singular values for the modified dataset.

subspace are stationary vectors for the function  $f_\pi(v)$  in (4.5) with  $\pi = 0$ . We have shown that the same holds for  $\pi > 0$ . It remains unclear how the right basis vectors, which are not in the common or isolated subspaces, are related to  $f_\pi(v)$  and whether an eventual connection would lead to a particular choice of the parameter  $\pi$ .

An experimental MATLAB implementation was provided that replicates the proposed algorithms and makes use of standard routines, such as the QR factorization and the SVD. To accommodate large-scale problems, the QR factorization step of the HO-GSVD algorithm could be computed using sparse matrix arithmetic, but the HO-CSD computation remains a dense matrix problem. For the analysis of low-rank and large-scale datasets, sample-based algorithms have proven to be computationally efficient for tensor decompositions (see, for example, [4, 19, 1]). Recently, a sample-based algorithm has been proposed for the standard GSVD ( $N = 2$ ) [24]. Future research could focus on adapting a sample-based algorithm to the rank-deficient HO-GSVD, which could improve the computational efficiency.

**Appendix A. Relation between  $S_\pi$  and  $T_\pi$ .** Let  $D_{i,\pi} = A_i^T A_i + \pi A^T A$  and define  $K_i := Q_i^T Q_i + \pi I$ . Using (2.4), the matrix  $S_\pi$  is written as

$$\begin{aligned} S_\pi &= \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N (D_{i,\pi} D_{j,\pi}^{-1} + D_{j,\pi} D_{i,\pi}^{-1}) \\ &= \frac{1}{N(N-1)} R^T \left( \sum_{i=1}^N \sum_{j=i+1}^N K_i K_j^{-1} + K_j K_i^{-1} \right) R^{-T}, \end{aligned}$$

so that by considering  $\sum_{i=1}^N K_i = (1 + \pi N)I$

$$\begin{aligned} R^{-T} S_\pi R^T &= \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N K_i K_j^{-1} + K_j K_i^{-1} = \frac{1}{N(N-1)} \sum_{i=1}^N K_i \sum_{j=1}^N K_j^{-1} - \frac{1}{N-1} I \\ &= \frac{1}{N-1} ((1 + \pi N)T_\pi - I). \end{aligned}$$

**Appendix B. Intermediate eigenvalues of  $T_\pi$ .** If there exists a vector  $t$  with  $\|t\|_2 = 1$  in the nullspace of  $P$  matrices  $Q_j$ , but in the range of all other  $Q_i$  with index  $i \in \mathcal{R}$ , then the inequalities (2.12a)–(2.12b) can be reformulated as

$$\begin{aligned} t^T T_\pi t &= \frac{1}{N} \sum_{i=1}^N t^T (Q_i^T Q_i + \pi I)^{-1} t = \frac{P}{\pi N} + \frac{1}{N} \sum_{i \in \mathcal{R}} t^T (Q_i^T Q_i + \pi I)^{-1} t \\ \text{(B.1a)} \quad &\geq \frac{P}{\pi N} + \frac{1}{N} \sum_{i \in \mathcal{R}} \frac{1}{t^T (Q_i^T Q_i + \pi I) t} \\ \text{(B.1b)} \quad &\geq \frac{P}{\pi N} + \frac{N-P}{N} \frac{N-P}{\underbrace{\sum_{i \in \mathcal{R}} t^T (Q_i^T Q_i) t}_{=1}} = \frac{P(1 - \pi N) + \pi N^2}{\pi N(1 + \pi(N-P))}. \end{aligned}$$

The term on the right-hand side of (B.1b) corresponds to the minimum and maximum eigenvalues of  $T_\pi$  for  $P = 0$  and  $P = N - 1$ , respectively. If there exists a



shared vector  $t$  in the nullspace of  $P$  matrices  $Q_j$ , but in the range of all other  $Q_i$ , then an eigenvalue of  $T_\pi$  will be equal to the corresponding value on the right-hand side of (B.1b). Note that (B.1) does *not* prove the converse.

**Appendix C. The arithmetic mean of amplification quotients.** The HO-GSVD is related to the function  $f_\pi(v)$  (4.5), which can be simplified using the stacked QR decomposition (2.1) as

$$g_\pi(z) = \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=1}^N \left( \frac{z^T(Q_i^T Q_i + \pi I)z}{z^T(Q_j^T Q_j + \pi I)z} + \frac{z^T(Q_j^T Q_j + \pi I)z}{z^T(Q_i^T Q_i + \pi I)z} \right) \geq 1,$$

where  $z := Rv$ . The gradient  $\nabla g_\pi(z)$  of  $g_\pi(z)$  is given by

$$\begin{aligned} \nabla g_\pi(z) = & \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=1}^N \left( \frac{1}{z^T W_{j,\pi} z} \left( W_{i,\pi} z - \frac{z^T W_{i,\pi} z}{z^T W_{j,\pi} z} W_{j,\pi} z \right) \right. \\ & \left. + \frac{1}{z^T W_{i,\pi} z} \left( W_{j,\pi} z - \frac{z^T W_{j,\pi} z}{z^T W_{i,\pi} z} W_{i,\pi} z \right) \right), \end{aligned}$$

where  $W_{i,\pi} := Q_i^T Q_i + \pi I$ . To show that  $\nabla g_\pi(z) = 0$  for  $z \in \mathcal{T}_N\{Q_1, \dots, Q_N\}$  or  $z \in \mathcal{T}_1\{Q_1, \dots, Q_N\}$ , note that  $z$  must be a right singular vector for each  $Q_i$ . It follows that  $W_{i,\pi} z = (\sigma_{i,1} + \pi)z$  and

$$W_{i,\pi} z - \frac{z^T W_{i,\pi} z}{z^T W_{j,\pi} z} W_{j,\pi} z = (\sigma_{i,1} + \pi)z - \frac{\sigma_{i,1} + \pi}{\sigma_{j,1} + \pi} (\sigma_{j,1} + \pi)z = 0,$$

so that  $\nabla g_\pi(z) = 0$  if  $z \in \mathcal{T}_N\{Q_1, \dots, Q_N\}$  or  $z \in \mathcal{T}_1\{Q_1, \dots, Q_N\}$  for any value of  $\pi$ . The proof is analogous for the HO-GSVD subspaces.

#### REFERENCES

- [1] S. AHMADI-ASL, S. ABUKHOVICH, M. G. ASANTE-MENSAH, A. CICHOCKI, A. H. PHAN, T. TANAKA, AND I. OSELEDETS, *Randomized algorithms for computation of Tucker decomposition and higher order SVD (HOSVD)*, IEEE Access, 9 (2021), pp. 28684–28706.
- [2] O. ALTER AND H. GOLUB, *Reconstructing the pathways of a cellular system from genome-scale signals by using matrix and tensor computations*, Proc. Natl. Acad. Sci. USA, 102 (2005), pp. 17559–17564.
- [3] Z. BAI, *CSD, GSVD, Their Applications and Computations*, IMA Preprints Series, 1992, pp. 1–39, Retrieved from the University of Minnesota Digital Conservancy, <https://hdl.handle.net/11299/1875>.
- [4] C. BATTAGLINO, G. BALLARD, AND T. G. KOLDA, *A practical randomized CP tensor decomposition*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 876–901, <https://doi.org/10.1137/17M1112303>.
- [5] M. T. CHU, R. E. FUNDERLIC, AND G. H. GOLUB, *On a variational formulation of the generalized singular value decomposition*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 1082–1092, <https://doi.org/10.1137/S0895479895287079>.
- [6] L. DE LATHAUWER, B. DE MOOR, AND J. VANDEWALLE, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278, <https://doi.org/10.1137/S0895479896305696>.
- [7] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 4th ed., Johns Hopkins University Press, Baltimore, MD, 2013.
- [8] H. LU, K. N. PLATANOTIS, AND A. N. VENETSANOPOULOS, *A survey of multilinear subspace learning for tensor data*, Pattern Recognition, 44 (2011), pp. 1540–1551.
- [9] G. H. HARDY, J. E. LITTLEWOOD, AND G. PÓLYA, *Inequalities*, Cambridge University Press, Cambridge, UK, 1934.
- [10] R. A. HARSHMAN, *Foundations of the PARAFAC procedure: Models and conditions for an “explanatory” multimodal factor analysis*, UCLA Working Papers in Phonetics, 16 (1970), pp. 1–84.

- [11] R. A. HARSHMAN, *PARAFAC2: Mathematical and technical notes*, UCLA Working Papers in Phonetics, 22 (1972), pp. 30–44.
- [12] T. KATO, *Perturbation Theory for Linear Operators*, 2nd ed., Springer, Berlin, 1980.
- [13] I. KEMPF, *HO-GSVD*, 2022, <https://github.com/kmpape/HO-GSVD>.
- [14] I. KEMPF, S. R. DUNCAN, P. J. GOULART, AND G. REHM, *Multi-array electron beam stabilization using block-circulant transformation and generalized singular value decomposition*, in Proc. 59th IEEE Conf. Decis. Control (CDC), Jeju Island, Republic of Korea, 2020.
- [15] B. KÄGSTRÖM, *The generalized singular value decomposition and the general  $(A - \lambda B)$ -problem*, BIT, 24 (1984), pp. 568–583.
- [16] A. KRIZHEVSKY AND G. HINTON, *Learning Multiple Layers of Features from Tiny Images*, Technical report, University of Toronto, Toronto, ON, Canada, 2009.
- [17] K. VAN DEUN, A. K. SMILDE, L. THORREZ, H. A. L. KIERS, AND I. VAN MECHELEN, *Identifying common and distinctive processes underlying multiset data*, Chemom. Intell. Lab. Syst., 129 (2013), pp. 40–51.
- [18] D. LAHAT, T. ADALI, AND C. JUTTEN, *Multimodal data fusion: An overview of methods, challenges, and prospects*, Proc. IEEE, 103 (2015), pp. 1449–1477.
- [19] B. W. LARSEN AND T. G. KOLDA, *Practical leverage-based sampling for low-rank tensor decomposition*, SIAM J. Matrix Anal. Appl., 43 (2022), pp. 1488–1517, <https://doi.org/10.1137/21M1441754>.
- [20] P. D. LAX, *Linear Algebra and Its Applications*, 2nd ed., Wiley, New York, 2007.
- [21] C. C. PAIGE AND M. A. SAUNDERS, *Towards a generalized singular value decomposition*, SIAM J. Numer. Anal., 18 (1981), pp. 398–405, <https://doi.org/10.1137/0718026>.
- [22] S. P. PONNAPALLI, *Higher-Order Generalized Singular Value Decomposition: Comparative Mathematical Framework with Applications to Genomic Signal Processing*, Ph.D. thesis, University of Texas at Austin, Austin, TX, 2010.
- [23] S. P. PONNAPALLI, M. A. SAUNDERS, C. F. VAN LOAN, AND O. ALTER, *A higher-order generalized singular value decomposition for comparison of global mRNA expression from multiple organisms*, PLoS ONE, 6 (2011), pp. 1–11.
- [24] A. K. SAIBABA, J. HART, AND B. VAN BLOEMEN WAANDERS, *Randomized algorithms for generalized singular value decomposition with application to sensitivity analysis*, Numer. Linear Algebra Appl., 28 (2021), e2364.
- [25] G. W. STEWART, *On the sensitivity of the eigenvalue problem  $Ax = \lambda Bx$* , SIAM J. Numer. Anal., 9 (1972), pp. 669–686, <https://doi.org/10.1137/0709056>.
- [26] B. SUKSIRI AND M. FUKUMOTO, *An efficient framework for estimating the direction of multiple sound sources using higher-order generalized singular value decomposition*, Sensors, 19 (2019), 2977.
- [27] C. F. VAN LOAN, *Generalizing the singular value decomposition*, SIAM J. Numer. Anal., 13 (1976), pp. 76–83, <https://doi.org/10.1137/0713009>.
- [28] C. F. VAN LOAN, *Computing the CS and the generalized singular value decompositions*, Numer. Math., 49 (1985), pp. 479–491.
- [29] C. F. VAN LOAN, *Lecture 6. The Higher-Order Generalized Singular Value Decomposition*, 2015, <http://www.dm.unibo.it/~simoncin/CIME/Vanloan.Lec6.pdf>.
- [30] L. J. VAN’T VEER, ET AL., *Gene expression profiling predicts clinical outcome of breast cancer*, Nature, 415 (2002), pp. 530–536.
- [31] X. XIAO, A. MORENO-MORAL, M. ROTIVAL, L. BOTTOLO, AND E. PETRETTO, *Multi-tissue analysis of co-expression networks by higher-order generalized singular value decomposition identifies functionally coherent transcriptional modules*, PLoS Genet., 10 (2014), pp. 1–16.