On the (non)existence of best low-rank approximations of generic IxJx2 arrays

Stegeman, Alwin

Published in:
Linear Algebra and Its Applications

DOI:
10.48550/arXiv.1309.5727

Publication date:
2022

On the (non)existence of best low-rank approximations of generic $I \times J \times 2$ arrays

Alwin Stegeman †

March 21, 2022

Abstract

Several conjectures and partial proofs have been formulated on the (non)existence of a best low-rank approximation of real-valued $I \times J \times 2$ arrays. We analyze this problem using the Generalized Schur Decomposition and prove (non)existence of a best rank-R approximation for generic $I \times J \times 2$ arrays, for all values of I, J, R. Moreover, for cases where a best rank-R approximation exists on a set of positive volume only, we provide easy-to-check necessary and sufficient conditions for the existence of a best rank-R approximation.

Keywords: tensor decomposition, low-rank approximation, Candecomp, Parafac, Generalized Schur Decomposition.

AMS subject classifications: 15A18, 15A22, 15A69, 49M27, 62H25.

Acknowledgments. This manuscript has benefited from comments of anonymous reviewers who carefully checked the proofs and suggested changes to improve their readability. Also, an anonymous reviewer has stimulated the search for a theoretical proof or rigorous numerical evidence for Conjecture 3.2. Unfortunately, neither have been found by the author.

This research was supported by the Dutch Organisation for Scientific Research (NWO), VIDI grant 452-08-001.

†A. Stegeman is with the Heijmans Institute for Psychological Research, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands, phone: ++31 50 363 6193, fax: ++31 50 363 6304, e-mail: a.w.stegeman@rug.nl, URL: http://www.gmw.rug.nl/∼stegeman.

1 Introduction

We consider the problem of finding a best low-rank approximation to a generic three-way array or order-3 tensor $\mathcal{Z} \in \mathbb{R}^{I\times J\times K}$. The rank of a three-way array $\mathcal{Y}$ is defined as the smallest number of rank-1 arrays whose sum equals $\mathcal{Y}$. A three-way array has rank 1 if it is the outer vector product of three nonzero vectors. The outer vector product $\mathbf{Y} = \mathbf{a} \circ \mathbf{b} = \mathbf{a}\,\mathbf{b}^T$ is a rank-1 matrix (or order-2 tensor) with entries $y_{ij} = a_i\, b_j$. The outer vector product $\mathcal{Y} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$ is rank-1 tensor with entries $y_{ijk} = a_i\, b_j\, c_k$. The problem of finding a best rank-R approximation to $\mathcal{Z}$ can be denoted as

$$\min_{\substack{\mathbf{a}_r\in\mathbb{R}^I,\,\mathbf{b}_r\in\mathbb{R}^J,\,\mathbf{c}_r\in\mathbb{R}^K,\\ r=1,\ldots,R}} \quad \|\mathcal{Z} - \sum_{r=1}^{R} (\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r)\|_F^2\,, \tag{1.1}$$

where $\|\cdot\|_F$ denotes the Frobenius norm (i.e., the square root of the sum-of-squares). For N-way arrays (or order-N tensors), this problem has been introduced by Hitchcock [14] [15]. The form of the rank-R approximation is known as Candecomp/Parafac [12] [4] and also as Canonical Polyadic Decomposition (CPD). It can be seen as a multi-way (or higher-order) generalization of component analysis for matrices. Applications of the CPD are found in chemometrics [32], the behavioral sciences [24], signal processing [6] [8], algebraic complexity theory [2] [3] (see [35] for a discussion), and data mining in general. An overview of applications of tensor decompositions can be found in [20] [1]. For the computation of a best low-rank approximation an iterative algorithm is used. For an overview and comparison of CPD algorithms, see [16] [44] [5]. It has been proven that determining the rank of an order-3 tensor or computing its best rank-1 approximation are NP-hard problems [13].

We denote the frontal $I \times J$ slices of $\mathcal{Z} \in \mathbb{R}^{I\times J\times K}$ as $\mathbf{Z}_k$, $k = 1,\ldots,K$. Let $\mathbf{A} = [\mathbf{a}_1 \ldots \mathbf{a}_R]$, $\mathbf{B} = [\mathbf{b}_1 \ldots \mathbf{b}_R]$, and $\mathbf{C} = [\mathbf{c}_1 \ldots, \mathbf{c}_R]$. Problem (1.1) can be written slicewise as

$$\min_{\substack{\mathbf{A}\in\mathbb{R}^{I\times R},\,\mathbf{B}\in\mathbb{R}^{J\times R},\\ \mathrm{diag}(\mathbf{C}_k)\in\mathbb{R}^R,\,k=1,\ldots,K}} \quad \sum_{k=1}^{K} \|\mathbf{Z}_k - \mathbf{A}\,\mathbf{C}_k\,\mathbf{B}^T\|_F^2\,, \tag{1.2}$$

where $\mathbf{C}_k$ is $R\times R$ diagonal with row k of $\mathbf{C}$ as diagonal, $k = 1,\ldots,K$. The set of $I \times J \times K$ arrays with rank at most R is denoted by

$$S_R(I,J,K) = \{\mathcal{Y} \in \mathbb{R}^{I\times J\times K} :\ \mathrm{rank}(\mathcal{Y}) \leq R\}\,. \tag{1.3}$$

Problem (1.1) can also be written as:

$$\min_{\mathcal{Y}\in S_R(I,J,K)} \|\mathcal{Z} - \mathcal{Y}\|_F^2\,. \tag{1.4}$$

Unfortunately, for $R \geq 2$, the problem may not have an optimal solution because the set $S_R(I, J, K)$ is not closed [9]. In such a case, trying to compute a best rank-R approximation yields a rank-R sequence converging to a boundary point $\mathcal{X}$ of $S_R(I, J, K)$ with $\text{rank}(\mathcal{X}) > R$. As a result, while running the iterative CPD algorithm, the decrease of the objective function becomes very slow, and some (groups of) columns of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ become nearly linearly dependent, while their norms increase without bound [25] [22] [9]. This phenomenon is known as "diverging CP components" or "degenerate solutions" or "diverging rank-1 terms". Needless to say, diverging rank-1 terms should be avoided if an interpretation of the rank-1 terms is needed. Note that diverging rank-1 terms are used in algebraic complexity theory to obtain a fast and arbitrarily accurate approximation to the computation of bilinear forms (see [35] for a discussion).

Nonexistence of a best rank-R approximation can be avoided by imposing constraints on the rank-1 terms in $(\mathbf{A}, \mathbf{B}, \mathbf{C})$. Imposing orthogonality constraints on (one of) the component matrices guarantees existence of a best rank-R approximation [22], and the same is true for nonnegative $\mathcal{Z}$ under the restriction of nonnegative $\mathbf{A}, \mathbf{B}, \mathbf{C}$ [26]. Also, [27] show that constraining the magnitude of the inner products between pairs of columns of $\mathbf{A}, \mathbf{B}, \mathbf{C}$ guarantees existence of a best rank-R approximation. However, imposing constraints will not be suitable for all CPD applications. As an alternative to deal with diverging rank-1 terms, methods have been developed to obtain the limit point $\mathcal{X}$ of the diverging rank-R sequence and a sparse decomposition of $\mathcal{X}$ [41] [30] [38] [39] [40].

There are very few theoretical results on the (non)existence of a best rank-R approximation for specific three-way arrays or sizes. It has been proven that $2 \times 2 \times 2$ arrays of rank 3 do not have a best rank-2 approximation [9], and conjectures on $I \times J \times 2$ arrays are formulated and partly proven in [35]. In simulation studies with random $\mathcal{Z}$, diverging rank-1 terms occur very often [33] [35] [34] [38]. Although diverging rank-1 terms may also occur due to a bad choice of starting point for the iterative algorithm [28] [36], if trying many random starting points does not help, then this is strong evidence for nonexistence of a best rank-R approximation.

In this paper, we consider (non)existence of best rank-R approximations for generic $I \times J \times 2$ arrays. The use of the term *generic* implies that the entries are randomly sampled from an $IJ2$-dimensional continuous distribution (for which sets of positive Lebesgue measure also have positive probability). Properties that hold for a generic array hold "with probability one", "almost surely", or "almost everywhere". Properties that hold on a set of positive Lebesgue measure but not almost everywhere, hold on a set of "positive volume" or "with positive probability". Using the relations between the CPD and the Generalized Schur Decomposition (GSD) formulated in [7] [41] [37], we are able to prove the conjectures formulated in [35]. Our main result concerns generic $I \times I \times 2$ arrays, which have ranks I and $I+1$ on sets of positive Lebesgue measure. It has been conjectured that generic $I \times I \times 2$ arrays of rank $I+1$ do not have a best rank-I approximation [33] [35]. So far, this has only been proven for $I = 2$ [9]. We provide a proof for $I \geq 2$. Our proofs of the (non)existence of best rank-R approximations for generic $I \times J \times 2$ arrays make use of our main result. In some cases, we prove that existence of a best rank-R approximation holds on a set of positive volume only. For such arrays we also provide easy-to-check necessary and sufficient conditions for the existence of a best rank-R approximation.

Classically, $I \times J \times 2$ arrays were classified as matrix pencils, where a matrix pencil $\mu\,\mathbf{X}_1 + \lambda\,\mathbf{X}_2$ consists of two $I \times J$ matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ and scalars μ and λ. A matrix pencil is called *regular* if both $\mathbf{X}_1$ and $\mathbf{X}_2$ are square matrices and there exist μ and λ such that $\det(\mu\mathbf{X}_1 + \lambda\mathbf{X}_2) \neq 0$. In all other cases, the pencil is called *singular*. For regular matrix pencils, equivalence results and a canonical form were established by Weierstrass [45]. The corresponding theory for singular pencils was developed by Kronecker [23]. Ja' Ja' [19] extended Kronecker's [23] equivalence results for matrix pencils to $I \times J \times 2$ arrays. As an anonymous reviewer pointed out, orbit classification results for matrix pencils such as those recently obtained for complex pencils in [29] could form a different approach to prove (non)existence of best rank-R approximations. However, we have taken a different approach via the relation between the CPD and GSD. A more detailed discussion of the relation between classical matrix pencil theory and the rank of real $I \times J \times 2$ arrays can be found in [35].

The paper is organized as follows. In section 2, we consider the relation between the CPD and GSD for $I \times J \times 2$ arrays and state the conjectures of [35]. In section 3, we formulate our main result for $I \times I \times 2$ arrays and $R = I$, and sketch its proof. The proof itself is contained in the appendix. In section 4, we prove a case that cannot be proven by using the GSD. In section 5, we extend our analysis and proof from section 3 to $I \times J \times 2$ arrays and $R \leq \min(I, J)$. Finally, section 6 contains a discussion of our findings.

We use the following notation. The notation $\mathcal{Y}$, $\mathbf{Y}$, $\mathbf{y}$, y is used for a three-way array, a matrix, a column vector, and a scalar, respectively. All arrays, matrices, vectors, and scalars are real-valued. Matrix transpose and inverse are denoted as $\mathbf{Y}^T$ and $\mathbf{Y}^{-1}$, respectively. An zero matrix of size $p \times q$ is denoted by $\mathbf{O}_{p,q}$. An zero column vector is denoted by $\mathbf{0}$. A $p \times p$ matrix $\mathbf{Y}$ is called orthonormal if $\mathbf{Y}^T\mathbf{Y} = \mathbf{Y}\mathbf{Y}^T = \mathbf{I}_p$. A $p \times q$ matrix has orthogonal columns if $\mathbf{Y}^T\mathbf{Y}$ is diagonal.

2 The CPD and GSD for $I \times J \times 2$ arrays

We begin by defining the Generalized Schur Decomposition (GSD) for $I \times J \times 2$ arrays. Analogous to (1.2), fitting a GSD to $\mathcal{Z}$ can be written slicewise as

$$\min_{\substack{\mathbf{Q}_a \in \mathbb{R}^{I\times R},\, \mathbf{Q}_b \in \mathbb{R}^{J\times R},\\ \mathbf{Q}_a^T\mathbf{Q}_a = \mathbf{Q}_b^T\mathbf{Q}_b = \mathbf{I}_R,\\ \mathbf{R}_k \in \mathbb{R}^{R\times R}\text{ upper triangular}, k=1,2.}} \sum_{k=1}^{2} \|\mathbf{Z}_k - \mathbf{Q}_a\,\mathbf{R}_k\,\mathbf{Q}_b^T\|_F^2\,. \tag{2.1}$$

Note that the GSD is only defined for $R \leq \min(I, J)$. We define the GSD solution set as

$$P_R(I, J, 2) = \{\mathcal{Y} \in \mathbb{R}^{I\times J\times 2} :\ \mathbf{Y}_k = \mathbf{Q}_a\,\mathbf{R}_k\,\mathbf{Q}_b^T\,,\ k = 1, 2\}\,. \tag{2.2}$$

It has been shown that $P_R(I, J, 2)$ is equal to the closure of $S_R(I, J, 2)$ [41] [37]. Moreover, a best fitting GSD always exists and it can be transformed to a best rank-R approximation if it exists [41]. If a best rank-R approximation does not exist, then a CPD algorithm trying to find a best rank-R approximation yields a sequence of rank-R arrays converging to an optimal solution of (2.1), and the CPD sequence features diverging components.

Showing that $\mathcal{Z}$ has no best rank-R approximation is equivalent to showing that all optimal solutions of (2.1) have rank larger than R. Let $\mathbf{G}_a$ and $\mathbf{G}_b$ be such that $\widetilde{\mathbf{Q}}_a = [\mathbf{Q}_a\ \mathbf{G}_a]$ and $\widetilde{\mathbf{Q}}_b = [\mathbf{Q}_b\ \mathbf{G}_b]$ are square and orthonormal matrices. When the slices of the GSD solution array are premultiplied by $\widetilde{\mathbf{Q}}_a^T$ and postmultiplied by $\widetilde{\mathbf{Q}}_b$, we obtain slices

$$\begin{bmatrix} \mathbf{R}_k & \mathbf{O}_{R,J-R} \\ \mathbf{O}_{I-R,R} & \mathbf{O}_{I-R,J-R} \end{bmatrix}, \qquad k = 1, 2\,, \tag{2.3}$$

where $\mathbf{O}_{p,q}$ denotes an zero $p \times q$ matrix. This implies that the rank of the GSD solution array is equal to the rank of the $R \times R \times 2$ array $\mathcal{R}$ with slices $\mathbf{R}_1$ and $\mathbf{R}_2$. To establish the rank of $\mathcal{R}$, we use the following lemma.

Lemma 2.1 *Let $\mathcal{Y} \in \mathbb{R}^{R\times R\times 2}$ with nonsingular $R \times R$ slices $\mathbf{Y}_1$ and $\mathbf{Y}_2$. The following statements hold:*

(i) If $\mathbf{Y}_2\mathbf{Y}_1^{-1}$ has R real eigenvalues and is diagonalizable, then $\mathcal{Y}$ has rank R.

(ii) If $\mathbf{Y}_2\mathbf{Y}_1^{-1}$ has R real eigenvalues but is not diagonalizable, then $\mathcal{Y}$ has at least rank $R + 1$.

(iii) If $\mathbf{Y}_2\mathbf{Y}_1^{-1}$ has at least one pair of complex eigenvalues, then $\mathcal{Y}$ has at least rank $R + 1$.

Proof. See [18, section 3]. □

Suppose that $\mathbf{R}_1$ and $\mathbf{R}_2$ are nonsingular. Since $\mathbf{R}_2\mathbf{R}_1^{-1}$ is upper triangular, it has R real eigenvalues. By Lemma 2.1, the rank of $\mathcal{R}$ is R when $\mathbf{R}_2\mathbf{R}_1^{-1}$ has R linearly independent eigenvectors. Otherwise, the rank of $\mathcal{R}$ is larger than R. In this case, the GSD solution array is the limit point of a CPD sequence featuring diverging rank-1 terms. Moreover, the diverging rank-1 terms are defined by groups of identical eigenvalues that do not have the same number of linearly independent associated eigenvectors [33] [35] [41] [42].

In this paper, we consider the conjectures of [35] on the (non)existence of best low-rank approximations for generic $I \times J \times 2$ arrays. These conjectures are given in Table 1. Note that existence of a best rank-R approximation is formulated in terms of volume, but can analogously be formulated in terms of probability. The rank values for the generic arrays are derived from the following. For a generic $I \times I \times 2$ array $\mathcal{Z}$, the matrix $\mathbf{Z}_2\mathbf{Z}_1^{-1}$ has I distinct eigenvalues. By Lemma 2.1 and [43] [33], the array satisfies either (i) and has rank I, or (iii) and has rank $I+1$; see also [43]. This is also formulated as $I \times I \times 2$ arrays having *typical rank* $\{I, I+1\}$. For generic $I \times J \times 2$ arrays with $I > J \geq 2$, the rank is given by $\min(I, 2J)$ [43]. In other words, $I \times J \times 2$ arrays with $I > J \geq 2$ have *generic rank* $\min(I, 2J)$. The notion of typical rank is used when several rank values occur on sets of positive Lebesgue measure.

In cases 1, 4, and 6 in Table 1, the value of R is larger than or equal to the rank of $\mathcal{Z}$. Hence, in these cases the best rank-R approximation of $\mathcal{Z}$ is $\mathcal{Z}$ itself. Case 2 is proven in section 3. Cases 3, 5, 8, and 9 are proven in section 5. In case 7 we have $R > J$ and cannot use the GSD to analyze the problem. This case is proven in section 4.

3 Case 2: $I \times I \times 2$ arrays of rank $I+1$ and $R = I$

We consider the GSD problem for generic $I \times I \times 2$ arrays and $R = I$. We rewrite the GSD problem (2.1) as

$$\min_{\substack{\mathbf{Q}_a \in \mathbb{R}^{I \times R},\, \mathbf{Q}_b \in \mathbb{R}^{J \times R}, \\ \mathbf{Q}_a^T \mathbf{Q}_a = \mathbf{Q}_b^T \mathbf{Q}_b = \mathbf{I}_R, \\ \mathbf{R}_k \in \mathbb{R}^{R \times R} \text{ upper triangular}, k=1,2.}} \quad \sum_{k=1}^{2} \|\mathbf{Q}_a^T \mathbf{Z}_k \mathbf{Q}_b - \mathbf{R}_k\|_F^2 . \tag{3.1}$$

For each $\mathbf{Q}_a$ and $\mathbf{Q}_b$, the optimal $\mathbf{R}_k$ are found as the upper triangular parts of $\mathbf{Q}_a^T\mathbf{Z}_k\mathbf{Q}_b$, respec-

Case	$\mathcal{Z} \in \mathbb{R}^{I\times J\times 2}$	rank($\mathcal{Z}$)	R	Best rank-R approx. exists ?
1	$I = J$	$I + 1$	$R \geq I + 1$	always
2	$I = J$	$I + 1$	$R = I$	zero volume
3	$I = J$	$I + 1$	$R < I$	positive volume
4	$I = J$	I	$R \geq I$	always
5	$I = J$	I	$R < I$	positive volume
6	$I > J$	$\min(I, 2J)$	$R \geq \min(I, 2J)$	always
7	$I > J$	$\min(I, 2J)$	$\min(I, 2J) > R > J$	almost everywhere
8	$I > J$	$\min(I, 2J)$	$R = J$	positive volume
9	$I > J$	$\min(I, 2J)$	$R < J$	positive volume

Table 1: Results (cases 1, 4, and 6) and conjectures (cases 2, 3, 5, 7, 8, and 9) of [35] on the existence of a best rank-R approximation of generic $I \times J \times 2$ arrays. Here, $I \geq J \geq 2$ and $R \geq 2$.

tively. Hence, problem (3.1) can be written as

$$\min_{\substack{\mathbf{Q}_a\in\mathbb{R}^{I\times R},\,\mathbf{Q}_b\in\mathbb{R}^{J\times R},\\ \mathbf{Q}_a^T\mathbf{Q}_a=\mathbf{Q}_b^T\mathbf{Q}_b=\mathbf{I}_R}} \quad \sum_{k=1}^{2} \|\mathbf{Q}_a^T\,\mathbf{Z}_k\,\mathbf{Q}_b\|_{LFs}^2\,, \tag{3.2}$$

where $\|\cdot\|_{LFs}$ denotes the Frobenius norm of the strictly lower triangular part. The optimal $\mathbf{Q}_a$ and $\mathbf{Q}_b$ can be obtained by iterating over Givens rotations (De Lathauwer, De Moor, and Vandewalle [7]). The optimal $\mathbf{Q}_a$ and $\mathbf{Q}_b$ are then the products of the consecutive optimal Givens rotation matrices. Each rotation affects rows and columns i and j ($i < j$) of $\mathbf{Q}_a^T\mathbf{Z}_k\mathbf{Q}_b$, $k = 1, 2$. For rotation (i, j), the corresponding Givens rotation matrices $\mathbf{U}_a$ and $\mathbf{U}_b$ are equal to $\mathbf{I}_I$ except:

$$(\mathbf{U}_a)_{ii} = (\mathbf{U}_a)_{jj} = \cos(\alpha)\,, \qquad (\mathbf{U}_a)_{ji} = -(\mathbf{U}_a)_{ij} = \sin(\alpha)\,, \tag{3.3}$$

$$(\mathbf{U}_b)_{ii} = (\mathbf{U}_b)_{jj} = \cos(\beta)\,, \qquad (\mathbf{U}_b)_{ji} = -(\mathbf{U}_b)_{ij} = \sin(\beta)\,. \tag{3.4}$$

The Jacobi-type algorithm of [7] to solve problem (3.2) iterates over all rotations (i, j), $1 \leq i < j \leq I$. In each iteration, α and β are computed that minimize $\sum_{k=1}^{2} \|\mathbf{U}_a^T\,\mathbf{Q}_a^T\,\mathbf{Z}_k\,\mathbf{Q}_b\,\mathbf{U}_b\|_{LFs}^2$, where $\mathbf{Q}_a$ and $\mathbf{Q}_b$ are the current updates. Next, $\mathbf{Q}_a$ is replaced by $\mathbf{Q}_a\mathbf{U}_a$ and $\mathbf{Q}_b$ is replaced by $\mathbf{Q}_b\mathbf{U}_b$. A necessary condition for reaching an optimal solution is that no rotation (i, j) can further decrease the objective function in (3.2). To derive the equations defining local minima for each rotation (i, j), we use the following lemma.

Lemma 3.1 *For vectors* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$ *and* $\alpha \in \mathbb{R}$, *define the rotation*

$$[\tilde{\mathbf{x}} \;\; \tilde{\mathbf{y}}] = [\mathbf{x} \;\; \mathbf{y}] \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}.$$

Let $f(\alpha) = \|\tilde{\mathbf{x}}\|^2$. *We have*

$$\frac{\partial f}{\partial \alpha} = 2\,\tilde{\mathbf{x}}^T \tilde{\mathbf{y}}\,, \qquad \frac{\partial^2 f}{\partial \alpha^2} = 2\,(\tilde{\mathbf{y}}^T \tilde{\mathbf{y}} - \tilde{\mathbf{x}}^T \tilde{\mathbf{x}})\,. \tag{3.5}$$

Moreover, if $\frac{\partial f}{\partial \alpha}(\alpha) = \frac{\partial^2 f}{\partial \alpha^2}(\alpha) = 0$ *for some* α, *then* $f(\alpha) = \mathbf{x}^T \mathbf{x}$ *is constant.*

Proof. We write $\|\tilde{\mathbf{x}}\|^2 = \sum_{i=1}^{p} (\cos(\alpha)\, x_i + \sin(\alpha)\, y_i)^2$. The first derivative in (3.5) follows from

$$\frac{\partial f}{\partial \alpha} = 2 \sum_{i=1}^{p} (\cos(\alpha)\, x_i + \sin(\alpha)\, y_i)\,(-\sin(\alpha)\, x_i + \cos(\alpha)\, y_i) = 2 \sum_{i=1}^{p} \tilde{x}_i\, \tilde{y}_i = 2\,\tilde{\mathbf{x}}^T \tilde{\mathbf{y}}\,.$$

The second derivative is obtained as

$$\begin{aligned} \frac{\partial^2 f}{\partial \alpha^2} &= 2\, \frac{\partial f}{\partial \alpha} \sum_{i=1}^{p} (\cos(\alpha)\, x_i + \sin(\alpha)\, y_i)\,(-\sin(\alpha)\, x_i + \cos(\alpha)\, y_i) \\ &= 2 \sum_{i=1}^{p} ((-\sin(\alpha)\, x_i + \cos(\alpha)\, y_i)^2 - (\cos(\alpha)\, x_i + \sin(\alpha)\, y_i)^2) \\ &= 2 \sum_{i=1}^{p} \tilde{y}_i^2 - \tilde{x}_i^2 = 2\,(\tilde{\mathbf{y}}^T \tilde{\mathbf{y}} - \tilde{\mathbf{x}}^T \tilde{\mathbf{x}})\,. \end{aligned}$$

Next, suppose the first and second derivatives are zero for some α. That is, $\tilde{\mathbf{x}}^T \tilde{\mathbf{y}} = 0$ and $\tilde{\mathbf{x}}^T \tilde{\mathbf{x}} = \tilde{\mathbf{y}}^T \tilde{\mathbf{y}}$ for some α. We write

$$\tilde{\mathbf{x}}^T \tilde{\mathbf{y}} = (\cos^2(\alpha) - \sin^2(\alpha))\, \mathbf{x}^T \mathbf{y} + \sin(\alpha)\cos(\alpha)\, (\mathbf{y}^T \mathbf{y} - \mathbf{x}^T \mathbf{x}) = 0\,, \tag{3.6}$$

$$\tilde{\mathbf{y}}^T \tilde{\mathbf{y}} - \tilde{\mathbf{x}}^T \tilde{\mathbf{x}} = (\cos^2(\alpha) - \sin^2(\alpha))\, (\mathbf{y}^T \mathbf{y} - \mathbf{x}^T \mathbf{x}) - 4\, \sin(\alpha)\cos(\alpha)\, \mathbf{x}^T \mathbf{y} = 0\,, \tag{3.7}$$

$$f(\alpha) = \tilde{\mathbf{x}}^T \tilde{\mathbf{x}} = \cos^2(\alpha)\, \mathbf{x}^T \mathbf{x} + \sin^2(\alpha)\, \mathbf{y}^T \mathbf{y} + 2\, \sin(\alpha)\cos(\alpha)\, \mathbf{x}^T \mathbf{y}\,. \tag{3.8}$$

When $\sin(\alpha) = 0$ or $\cos(\alpha) = 0$, it follows from (3.6)-(3.7) that $\mathbf{x}^T \mathbf{y} = 0$ and $\mathbf{x}^T \mathbf{x} = \mathbf{y}^T \mathbf{y}$. By (3.8), this implies the desired result $f(\alpha) = \mathbf{x}^T \mathbf{x}$. Next, suppose $\sin(\alpha)\cos(\alpha) \neq 0$. Combining (3.6)-(3.7) yields

$$\mathbf{y}^T \mathbf{y} - \mathbf{x}^T \mathbf{x} = -\left(\frac{(\cos^2(\alpha) - \sin^2(\alpha))^2}{4\, \sin^2(\alpha)\cos^2(\alpha)} \right) (\mathbf{y}^T \mathbf{y} - \mathbf{x}^T \mathbf{x})\,. \tag{3.9}$$

Since the term depending on α in (3.9) is nonpositive, it follows that $\mathbf{x}^T\mathbf{x} = \mathbf{y}^T\mathbf{y}$. Then $\mathbf{x}^T\mathbf{y} = 0$ follows from (3.6)-(3.7), and we again obtain $f(\alpha) = \mathbf{x}^T\mathbf{x}$. This completes the proof. $\square$

Let $\widetilde{\mathbf{Z}}_k = \mathbf{Q}_a^T\,\mathbf{Z}_k\,\mathbf{Q}_b$, $k = 1, 2$, and define the 2-dimensional vectors

$$\tilde{\mathbf{z}}_{(m,n)} = \begin{pmatrix} (\widetilde{\mathbf{Z}}_1)_{mn} \\ (\widetilde{\mathbf{Z}}_2)_{mn} \end{pmatrix} = \begin{pmatrix} \tilde{z}_{mn1} \\ \tilde{z}_{mn2} \end{pmatrix}, \qquad m = 1, \ldots, I\,, \quad n = 1, \ldots, I\,. \tag{3.10}$$

The vectors (3.10) are also known as mode-3 vectors of the array with slices $\widetilde{\mathbf{Z}}_1$ and $\widetilde{\mathbf{Z}}_2$. As in [7], we determine the stationary points for rotation (i, j) by setting the derivatives with respect to α and β of $\sum_{k=1}^2 \|\mathbf{U}_a^T\,\widetilde{\mathbf{Z}}_k\,\mathbf{U}_b\|_{LFs}^2$ equal to zero. When rotating rows i and j (with $i < j$) the entries (i, r) and (j, r) with $r = 1, \ldots, i - 1$ stay in the strictly lower triangular part. Their Frobenius norm is not changed. Analogously, the entries (i, r) and (j, r) with $r = j, \ldots, I$ stay in the upper triangular part, and do not affect the objective function (3.2). Hence, the rotation of rows i and j can change the objective function only via entries (i, r) and (j, r) with $r = i, \ldots, j - 1$. Let $\tilde{\mathbf{y}}_k = [\tilde{z}_{i,i,k}, \ldots, \tilde{z}_{i,j-1,k}]^T$ and $\tilde{\mathbf{x}}_k = [\tilde{z}_{j,i,k}, \ldots, \tilde{z}_{j,j-1,k}]^T$, $k = 1, 2$. Next we apply Lemma 3.1 with objective function $f(\alpha) = \|\tilde{\mathbf{x}}_1\|^2 + \|\tilde{\mathbf{x}}_2\|^2$, which yields the first-order condition $\tilde{\mathbf{y}}_1^T\tilde{\mathbf{x}}_1 + \tilde{\mathbf{y}}_2^T\tilde{\mathbf{x}}_2 = 0$. We rewrite this condition for a stationary point as

$$\sum_{r=i}^{j-1} \tilde{\mathbf{z}}_{(i,r)}^T \tilde{\mathbf{z}}_{(j,r)} = 0\,, \qquad 1 \le i < j \le I\,. \tag{3.11}$$

When rotating columns i and j (with $i < j$) we obtain analogously that the objective function can be changed only via entries (r, i) and (r, j) with $r = i + 1, \ldots, j$. Analogous to (3.11), it follows from Lemma 3.1 that a stationary point satisfies

$$\sum_{r=i+1}^{j} \tilde{\mathbf{z}}_{(r,i)}^T \tilde{\mathbf{z}}_{(r,j)} = 0\,, \qquad 1 \le i < j \le I\,. \tag{3.12}$$

Equations (3.11) and (3.12) are the first-order optimality conditions. Hence, in an optimal solution of problem (3.2) equations (3.11) and (3.12) will hold.

We obtain second-order optimality conditions from Lemma 3.1, where positive second derivatives are required for a local minimum for each rotation. Lemma 3.1 shows that a second derivative being zero at a stationary point implies a constant objective function and infinitely many optimal rotation angles. For the rotation of rows i and j with $i < j$, we define $\tilde{\mathbf{x}}_k$ and $\tilde{\mathbf{y}}_k$ as above, $k = 1, 2$. For the minimization of $f(\alpha) = \|\tilde{\mathbf{x}}_1\|^2 + \|\tilde{\mathbf{x}}_2\|^2$, the second-order condition is given by

$\|\tilde{\mathbf{y}}_1\|^2 + \|\tilde{\mathbf{y}}_2\|^2 \geq \|\tilde{\mathbf{x}}_1\|^2 + \|\tilde{\mathbf{x}}_2\|^2$; see Lemma 3.1. We rewrite this second-order condition as

$$\sum_{r=i}^{j-1} \tilde{\mathbf{z}}_{(i,r)}^T \tilde{\mathbf{z}}_{(i,r)} - \sum_{r=i}^{j-1} \tilde{\mathbf{z}}_{(j,r)}^T \tilde{\mathbf{z}}_{(j,r)} \geq 0\,, \qquad 1 \leq i < j \leq I\,. \tag{3.13}$$

Analogously, the second-order condition for the rotation of columns i and j with $i < j$ equals

$$\sum_{r=i+1}^{j} \tilde{\mathbf{z}}_{(r,j)}^T \tilde{\mathbf{z}}_{(r,j)} - \sum_{r=i+1}^{j} \tilde{\mathbf{z}}_{(r,i)}^T \tilde{\mathbf{z}}_{(r,i)} \geq 0\,, \qquad 1 \leq i < j \leq I\,. \tag{3.14}$$

We work under the following conjecture of positive second-order conditions. Numerical evidence and more theoretical context are provided in Appendix A.

Conjecture 3.2 *Let $\mathcal{Z} \in \mathbb{R}^{I\times I\times 2}$ be generic with* $\text{rank}(\mathcal{Z}) = I + 1$. *Let* $(\mathbf{Q}_a, \mathbf{Q}_b, \mathbf{R}_1, \mathbf{R}_2)$ *be an optimal solution of the GSD problem* (2.1) *with $R = I$. Then the second-order conditions* (3.13)-(3.14) *are strictly positive.* □

We use the first and second-order optimality conditions to obtain the following result.

Theorem 3.3 *Let $\mathcal{Z} \in \mathbb{R}^{I\times I\times 2}$ be generic with* $\text{rank}(\mathcal{Z}) = I + 1$. *Let* $(\mathbf{Q}_a, \mathbf{Q}_b, \mathbf{R}_1, \mathbf{R}_2)$ *be an optimal solution of the GSD problem* (2.1) *with $R = I$ and strictly positive second-order conditions* (3.13)-(3.14). *Then the rank of the $I \times I \times 2$ array $\mathcal{R}$ with slices $\mathbf{R}_1$ and $\mathbf{R}_2$ is larger than I.*

Proof. See Appendix B. □

Theorem 3.3 and Conjecture 3.2 imply that any optimal solution array of the GSD problem (2.1), which has slices $\mathbf{Q}_a\mathbf{R}_k\mathbf{Q}_b^T$, $k = 1, 2$, has rank larger than I. As explained in section 2, this is equivalent to $\mathcal{Z}$ not having a best rank-I approximation. Hence, we obtain the following.

Corollary 3.4 *Let $\mathcal{Z} \in \mathbb{R}^{I\times I\times 2}$ be generic with* $\text{rank}(\mathcal{Z}) = I + 1$. *Then $\mathcal{Z}$ does not have a best rank-I approximation.* □

Note that the formulation of Corollary 3.4 is equivalent to an $I \times I \times 2$ array of rank $I + 1$ having a best rank-I approximation at most on a set of zero volume, as is stated in case 2 of Table 1.

4 Case 7: $I \times J \times 2$ arrays with $I > J$ and $\min(I, 2J) > R > J$

Since $R > J$, we cannot use the GSD in this case. We define the set

$$W_R(I,J,2) = \{\mathcal{Y} \in \mathbb{R}^{I\times J\times 2} :\ \mathrm{rank}([\mathbf{Y}_1 \,|\, \mathbf{Y}_2]) \leq R\}\,. \tag{4.1}$$

and the problem

$$\min_{\mathcal{Y}\in W_R(I,J,2)} \|\mathcal{Z} - \mathcal{Y}\|_F^2\,. \tag{4.2}$$

Stegeman [35] shows that $W_R(I,J,2)$ is the closure of the rank-R set $S_R(I,J,2)$ when $I > J \geq 2$ and $\min(I,2J) > R > J$. In our proof below, the set $W_R(I,J,2)$ plays the role of the GSD solution set $P_R(I,J,2)$ in section 3. We have the following result.

Theorem 4.1 *Let $\mathcal{Z} \in \mathbb{R}^{I\times J\times 2}$ with $I > J \geq 2$ and $\min(I,2J) > R > J$ be generic. Then the optimal solution $\mathcal{X}$ of problem (4.2) is unique and $\mathrm{rank}(\mathcal{X}) = R$.*

Proof. Problem (4.2) is in fact a matrix problem. Namely, the closest rank-R matrix $\mathbf{Y} = [\mathbf{Y}_1 \,|\, \mathbf{Y}_2]$ to a generic $I \times 2J$ matrix $\mathbf{Z} = [\mathbf{Z}_1 \,|\, \mathbf{Z}_2]$ is asked for. It is well known that this problem is solved by the truncated singular value decomposition (SVD) of $\mathbf{Z}$ [10]. Let the SVD of $\mathbf{Z}$ be given as $\mathbf{Z} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^T$. Without loss of generality we assume $I \geq 2J$. Matrix $\mathbf{U}$ is $I \times 2J$ and columnwise orthonormal, $\mathbf{S}$ is $2J \times 2J$ diagonal and nonsingular, and $\mathbf{V}$ is $2J \times 2J$ and orthonormal. The singular values on the diagonal of $\mathbf{S}$ are assumed to be in decreasing order. Since the singular values of $\mathbf{Z}$ are distinct ($\mathbf{Z}$ is generic), matrix problem (4.2) has a unique solution $\mathbf{X} = \mathbf{U}_R\,\mathbf{S}_R\,\mathbf{V}_R^T$. Here, $\mathbf{U}_R$ and $\mathbf{V}_R$ contain the first R columns of $\mathbf{U}$ and $\mathbf{V}$, respectively, and $\mathbf{S}_R$ is $R \times R$ diagonal and contains the R largest singular values.

The optimal solution $\mathcal{X}$ of problem (4.2) has slices $\mathbf{X}_1 = \mathbf{U}_R\,\mathbf{S}_R\,\mathbf{V}_{R,1}^T$ and $\mathbf{X}_2 = \mathbf{U}_R\,\mathbf{S}_R\,\mathbf{V}_{R,2}^T$, where $\mathbf{V}_{R,1}^T$ contains columns $1,\ldots,J$ of $\mathbf{V}_R^T$, and $\mathbf{V}_{R,2}^T$ contains columns $J+1,\ldots,2J$ of $\mathbf{V}_R^T$. The rank of $\mathcal{X}$ is equal to the rank of the $R \times J \times 2$ array $\mathcal{V}_R$ with slices $\mathbf{S}_R\mathbf{V}_{R,1}^T$ and $\mathbf{S}_R\mathbf{V}_{R,2}^T$. Hence, the proof is complete if we show that $\mathrm{rank}(\mathcal{V}_R) = R$.

We have the eigendecomposition $\mathbf{Z}^T\mathbf{Z} = \mathbf{V}\,\mathbf{S}^2\,\mathbf{V}^T$, where $\mathbf{Z}^T\mathbf{Z}$ is a generic symmetric positive definite $2J\times 2J$ matrix. Since the set of all $\mathbf{Z}^T\mathbf{Z}$ has positive measure in the set of symmetric $2J\times 2J$ matrices, it has dimensionality $2J(2J+1)/2$. The set with the parameterization $\mathbf{V}\,\mathbf{S}\,\mathbf{V}^T$ must have the same dimensionality. The dimensionality of the set of all $\mathbf{S}$ equals $2J$ and the dimensionality of the set of all $2J\times 2J$ orthonormal $\mathbf{V}$ equals $2J(2J-1)/2$, with the sum being $2J(2J+1)/2$. Hence, $\mathbf{V}$ may be considered a generic $2J \times 2J$ orthonormal matrix. Analogously, $\mathbf{V}_R$ may be considered a generic $2J \times R$ columnwise orthonormal matrix. The $R \times J \times 2$ array $\mathcal{V}_R$ may be considered generic under the condition that the rows of its matrix unfolding $\mathbf{S}_R \mathbf{V}_R^T = [\mathbf{S}_R \mathbf{V}_{R,1}^T \,|\, \mathbf{S}_R \mathbf{V}_{R,2}^T]$ are orthogonal. Premultiplying the slices of $\mathcal{V}_R$ by a generic $R \times R$ matrix yields a generic $R \times J \times 2$ array, with rank equal to $\text{rank}(\mathcal{V}_R)$. Hence, $\text{rank}(\mathcal{V}_R)$ is equal to the rank of generic $R \times J \times 2$ arrays. When $2J > R > J \geq 2$, the latter rank is given by $\min(R, 2J) = R$ [43]. This completes the proof. □

Since $W_R(I, J, 2)$ is the closure of $S_R(I, J, 2)$, it follows that the optimal solution $\mathcal{X}$ in Theorem 4.1 is an optimal solution of the best rank-R approximation problem (1.4). Hence, we obtain the following.

Corollary 4.2 *Let $\mathcal{Z} \in \mathbb{R}^{I \times J \times 2}$ with $I > J \geq 2$ and $\min(I, 2J) > R > J$ be generic. Then $\mathcal{Z}$ has a best rank-R approximation.* □

Note that the formulation of Corollary 4.2 is equivalent to an $I \times J \times 2$ array with $I > J \geq 2$ and $\min(I, 2J) > R > J$ having a best rank-I approximation almost everywhere, as is stated in case 7 of Table 1.

5 Extension to $I \times J \times 2$ arrays and $R \leq \min(I, J)$

Here, we consider the GSD problem (2.1) for cases 3, 5, 8, and 9 in Table 1. Hence, we have $R < I$ or $R < J$ or both. Also, $R < \text{rank}(\mathcal{Z})$. In these cases, [35] conjectures that the set of arrays that have a best rank-R approximation, and the set of arrays that do not have a best rank-R approximation, both have positive volume. Below, we analyze this using the GSD framework. In section 5.1, we consider the GSD algorithm when $R < I$ or $R < J$ or both, which was presented in [41]. We derive equations defining a stationary point, which we use in our proofs. In section 5.2 we prove case 8, in which $R = J < I$. In section 5.3 we prove cases 3, 5, and 9, in which $R < \min(I, J)$.

5.1 The GSD algorithm when $R < I$ or $R < J$ or both

For $R = I = J$, the optimal $\mathbf{Q}_a$ and $\mathbf{Q}_b$ are found by minimizing the Frobenius norm of the strictly lower triangular parts of $\mathbf{Q}_a^T \mathbf{Z}_k \mathbf{Q}_b$, $k = 1, 2$; see (3.2). Since $\mathbf{Q}_a$ and $\mathbf{Q}_b$ are orthonormal, we have $\|\mathbf{Q}_a^T \mathbf{Z}_k \mathbf{Q}_b\|_F = \|\mathbf{Z}_k\|_F$, $k = 1, 2$. This implies that solving (3.2) is equivalent to maximizing the Frobenius norm of the upper triangular parts of $\mathbf{Q}_a^T \mathbf{Z}_k \mathbf{Q}_b$, $k = 1, 2$. Analogously, for $R < I$ or

$R < J$ or both, we maximize the upper triangular part of the first R rows and columns of $\widetilde{\mathbf{Q}}_a^T \, \mathbf{Z}_k \, \widetilde{\mathbf{Q}}_b$, $k = 1, 2$. Here, $\widetilde{\mathbf{Q}}_a$ $(I \times I)$ and $\widetilde{\mathbf{Q}}_b$ $(J \times J)$ are orthonormal, and $\mathbf{Q}_a$ and $\mathbf{Q}_b$ are taken as the first R columns from $\widetilde{\mathbf{Q}}_a$ and $\widetilde{\mathbf{Q}}_b$, respectively.

Updating $\widetilde{\mathbf{Q}}_a$ and $\widetilde{\mathbf{Q}}_b$ is done via Givens rotations, as for $R = I = J$ in section 3. We have four different kinds of Givens rotations. Rotations of rows i and j or columns i and j with $1 \leq i < j \leq R$ are the same as described in section 3. Conditions for stationary points with respect to these rotations are given by (3.11) and (3.12) for $1 \leq i < j \leq R$, where we now define $\widetilde{\mathbf{Z}}_k = \widetilde{\mathbf{Q}}_a^T \, \mathbf{Z}_k \, \widetilde{\mathbf{Q}}_b$, $k = 1, 2$. For convenience, we repeat these equations as

$$\sum_{r=i}^{j-1} \tilde{\mathbf{z}}_{(i,r)}^T \tilde{\mathbf{z}}_{(j,r)} = 0 \,, \qquad 1 \leq i < j \leq R \,, \tag{5.1}$$

$$\sum_{r=i+1}^{j} \tilde{\mathbf{z}}_{(r,i)}^T \tilde{\mathbf{z}}_{(r,j)} = 0 \,, \qquad 1 \leq i < j \leq R \,. \tag{5.2}$$

When $R < I$, we have additional rotations of rows i and j with $i > R$ or $j > R$ or both. Rotations of rows i and j with $R < i < j$ do not change the upper triangular part of the first R rows. Hence, they can be left out of consideration. Rotations of rows i and j with $1 \leq i \leq R$ and $R+1 \leq j \leq I$ change the upper triangular part of the first R rows via entries (i, r) with $r = i, \ldots, R$. Analogous to (5.1) and (5.2), this yields the following equations for stationary points:

$$\sum_{r=i}^{R} \tilde{\mathbf{z}}_{(i,r)}^T \tilde{\mathbf{z}}_{(j,r)} = 0 \,, \qquad 1 \leq i \leq R \,, \quad R+1 \leq j \leq I \,. \tag{5.3}$$

When $R < J$, we also have rotations of columns i and j with $i > R$ or $j > R$ or both. Analogous to row rotations, we only need to consider i and j with $1 \leq i \leq R$ and $R+1 \leq j \leq J$. In the upper triangular part of the first R columns only the entries (r, i) with $r = 1, \ldots, i$ are changed. This yields the following equations for stationary points:

$$\sum_{r=1}^{i} \tilde{\mathbf{z}}_{(r,i)}^T \tilde{\mathbf{z}}_{(r,j)} = 0 \,, \qquad 1 \leq i \leq R \,, \quad R+1 \leq j \leq J \,. \tag{5.4}$$

Hence, stationary points of the GSD problem (2.1) satisfy (5.1)–(5.4).

For fixed $i \in \{1, \ldots, R\}$, the GSD algorithm of [41] combines row rotations (i, j) for all $j = R+1, \ldots, I$ using a singular value decomposition. The same holds for column rotations (i, j) with fixed $i \in \{1, \ldots, R\}$ and all $j = R+1, \ldots, J$. However, the GSD algorithm can also be programmed in the way described above, i.e., solving each rotation separately. For each rotation, the optimal rotation angle α can be computed by setting the derivative in (3.5) equal to zero. After dividing by

$\cos^2(\alpha)$, this yields a second degree polynomial in $\tan(\alpha)$. Numerical experiments show that, for the same generic $\mathcal{Z}$, the two GSD algorithms yield different $\mathbf{Q}_a$, $\mathbf{Q}_b$, $\mathbf{R}_1$, and $\mathbf{R}_2$, but the GSD solution array is identical, and also the eigenvalues and number of eigenvectors of $\mathbf{R}_2\mathbf{R}_1^{-1}$ are identical.

5.2 Case 8: $I \times J \times 2$ arrays with $I > J = R$

We proceed analogous to case 7 in section 4. We define the set $W_R(I,J,2)$ as in (4.1) and consider the best approximation of $\mathcal{Z}$ from $W_R(I,J,2)$ in (4.2). We have $S_R(I,J,2) \subset W_R(I,J,2)$, see [35]. Hence, if the best approximation $\mathcal{X}$ from the set $W_R(I,J,2)$ has rank at most R, then $\mathcal{Z}$ has a best rank-R approximation. As in the proof of Theorem 4.1, the best approximation from $W_R(I,J,2)$ is unique and given by the truncated singular value decomposition (SVD) of $\mathbf{Z} = [\mathbf{Z}_1 \,|\, \mathbf{Z}_2]$, which we denote as $\mathbf{X} = \mathbf{U}_R\,\mathbf{S}_R\,\mathbf{V}_R^T$. The corresponding array $\mathcal{X}$ has slices $\mathbf{X}_1 = \mathbf{U}_R\,\mathbf{S}_R\,\mathbf{V}_{R,1}^T$ and $\mathbf{X}_2 = \mathbf{U}_R\,\mathbf{S}_R\,\mathbf{V}_{R,2}^T$, where $\mathbf{V}_R^T = [\mathbf{V}_{R,1}^T \,|\, \mathbf{V}_{R,2}^T]$. The rank of $\mathcal{X}$ is equal to the rank of the $R \times R \times 2$ array $\mathcal{V}_R$ with $R \times R$ slices $\mathbf{S}_R\mathbf{V}_{R,1}^T$ and $\mathbf{S}_R\mathbf{V}_{R,2}^T$. As stated above, $\mathrm{rank}(\mathcal{X}) = \mathrm{rank}(\mathcal{V}_R) \leq R$ implies that $\mathcal{Z}$ has a best rank-R approximation. The rank of $\mathcal{V}_R$ can be checked by making use of Lemma 2.1.

We have the following result for the case where $\mathrm{rank}(\mathcal{X}) = \mathrm{rank}(\mathcal{V}_R) > R$.

Theorem 5.1 *Let $\mathcal{Z} \in \mathbb{R}^{I\times J\times 2}$ with $I > J = R \geq 2$ be generic. Let $\mathcal{X}$ be the optimal solution of problem* (4.2). *Let $(\mathbf{Q}_a, \mathbf{Q}_b, \mathbf{R}_1, \mathbf{R}_2)$ be an optimal solution of the GSD problem* (2.1) *with strictly positive second-order conditions* (3.13)-(3.14) *for $1 \leq i < j \leq R$. If $\mathrm{rank}(\mathcal{X}) > R$, then the rank of the $R \times R \times 2$ array $\mathcal{R}$ with slices $\mathbf{R}_1$ and $\mathbf{R}_2$ is larger than R.*

Proof. See Appendix C. □

Analogous to Conjecture 3.2, we assume that the second-order conditions (3.13)-(3.14), $1 \leq i < j \leq R$, are strictly positive for generic $\mathcal{Z} \in \mathbb{R}^{I\times J\times 2}$ with $I > J = R \geq 2$. As in Appendix A, we have found no counterexamples in numerical experiments (results not reported). Analogous to Corollary 3.4 following from Theorem 3.3, we obtain the following.

Corollary 5.2 *Let $\mathcal{Z} \in \mathbb{R}^{I\times J\times 2}$ with $I > J = R \geq 2$ be generic, and let $\mathcal{X}$ be the optimal solution of problem* (4.2). *If $\mathrm{rank}(\mathcal{X}) > R$, then $\mathcal{Z}$ does not have a best rank-R approximation.* □

Corollary 5.2 implies that we now have an easy-to-check criterion to determine whether $\mathcal{Z}$ has a best rank-R approximation or not. First, compute the truncated SVD of $\mathbf{Z} = [\mathbf{Z}_1 \,|\, \mathbf{Z}_2]$ as $\mathbf{X} =$

$\mathbf{U}_R\,\mathbf{S}_R\,\mathbf{V}_R^T$. As in the proof of Theorem 4.1, the array $\mathcal{V}_R$ with slices $\mathbf{S}_R\mathbf{V}_{R,k}^T$, $k = 1, 2$, may be considered a generic $R \times R \times 2$ array. Hence, its rank is either R or $R+1$, both on sets of positive Lebesgue measure [43]. Next, compute the eigenvalues of $\mathbf{S}_R\mathbf{V}_{R,2}^T(\mathbf{S}_R\mathbf{V}_{R,1}^T)^{-1}$ (or just $\mathbf{V}_{R,2}^T(\mathbf{V}_{R,1}^T)^{-1}$), which are distinct. If all eigenvalues are real, then $\mathrm{rank}(\mathcal{X}) = \mathrm{rank}(\mathcal{V}_R) = R$ (Lemma 2.1) and $\mathcal{Z}$ has a best rank-R approximation, which can be taken equal to $\mathcal{X}$. If some eigenvalues are complex, then $\mathrm{rank}(\mathcal{X}) = \mathrm{rank}(\mathcal{V}_R) = R+1$ (Lemma 2.1) and $\mathcal{Z}$ does not have a best rank-R approximation. Since both situations occur on sets of positive Lebesgue measure, this completes the proof of case 8 of Table 1.

5.3 Cases 3, 5, 9: $I \times J \times 2$ arrays with $R < \min(I, J)$

We proceed analogous to case 8 in section 5.2, except now the situation is more complicated. We define the set

$$\widetilde{W}_R(I,J,2) = \{\mathcal{Y} \in \mathbb{R}^{I\times J\times 2} :\ \mathrm{rank}([\mathbf{Y}_1\,|\,\mathbf{Y}_2]) \leq R\,,\ \text{and rank}\left(\begin{bmatrix}\mathbf{Y}_1\\ \mathbf{Y}_2\end{bmatrix}\right) \leq R\}\,. \tag{5.5}$$

and the problem

$$\min_{\mathcal{Y}\in\widetilde{W}_R(I,J,2)} \|\mathcal{Z}-\mathcal{Y}\|_F^2\,. \tag{5.6}$$

Since $\widetilde{W}_R(I,J,2)$ is closed, problem (5.6) is guaranteed to have an optimal solution. We have $S_R(I,J,2) \subset \widetilde{W}_R(I,J,2)$, see [35]. Hence, if a best approximation $\mathcal{X}$ from the set $\widetilde{W}_R(I,J,2)$ has rank at most R, then $\mathcal{Z}$ has a best rank-R approximation. Note that problem (5.6) is equivalent to finding a best multilinear rank-$(R,R,2)$ approximation of $\mathcal{Z}$, with no transformation in the third mode. Algorithms to solve this problem can be found in [31] [17].

Next, we present an algorithm to solve problem (5.6) by using Givens rotations. We make use of this algorithm in our proof for cases 3, 5, and 9. Let $\mathcal{Y} \in \widetilde{W}_R(I,J,2)$ have the following SVDs of its matrix unfoldings:

$$[\mathbf{Y}_1\,|\,\mathbf{Y}_2] = \mathbf{U}_1\,\mathbf{S}_1\,\mathbf{V}_1^T\,, \qquad \begin{bmatrix}\mathbf{Y}_1\\ \mathbf{Y}_2\end{bmatrix} = \mathbf{V}_2\,\mathbf{S}_2\,\mathbf{U}_2^T\,, \tag{5.7}$$

with $\mathbf{U}_1$ $(I\times I)$, $\mathbf{V}_1$ $(2J\times 2J)$, $\mathbf{V}_2$ $(2I\times 2I)$, and $\mathbf{U}_2$ $(J\times J)$ orthonormal. Since both unfoldings have rank at most R, only the first R diagonal entries of $\mathbf{S}_1$ and $\mathbf{S}_2$ are nonzero. It follows that

$$\mathbf{U}_1^T\,\mathbf{Y}_k\,\mathbf{U}_2 = \begin{bmatrix}\mathbf{G}_k & \mathbf{O}_{R,J-R}\\ \mathbf{O}_{I-R,R} & \mathbf{O}_{I-R,J-R}\end{bmatrix}\,, \qquad k = 1,2\,, \tag{5.8}$$

where $\mathbf{G}_k$ is $R \times R$, $k = 1,2$. Hence, $\mathcal{Y} \in \widetilde{W}_R(I,J,2)$ satisfies $\mathbf{Y}_k = \mathbf{U}_{1,R}\,\mathbf{G}_k\,\mathbf{U}_{2,R}^T$, $k = 1,2$, where $\mathbf{U}_{1,R}$ $(I \times R)$ and $\mathbf{U}_{2,R}$ $(J \times R)$ consist of the first R columns of $\mathbf{U}_1$ and $\mathbf{U}_2$, respectively. Analogous to the GSD algorithm discussed in section 5.1, problem (5.6) can be solved by finding orthonormal $\mathbf{U}_1$ and $\mathbf{U}_2$ that maximize the Frobenius norm of the first R rows and columns of $\mathbf{U}_1^T\,\mathbf{Z}_k\,\mathbf{U}_2$, $k = 1,2$. The best approximation $\mathcal{X}$ from $\widetilde{W}_R(I,J,2)$ then has slices $\mathbf{X}_k = \mathbf{U}_{1,R}\,\mathbf{G}_k\,\mathbf{U}_{2,R}^T$, $k = 1,2$, where $\mathbf{U}_{1,R}$ and $\mathbf{U}_{2,R}$ consist of the first R columns of $\mathbf{U}_1$ and $\mathbf{U}_2$, respectively, and $\mathbf{G}_k$ is taken as the first R rows and columns of $\mathbf{U}_1^T\,\mathbf{Z}_k\,\mathbf{U}_2$, $k = 1,2$.

Finding $\mathbf{U}_1$ and $\mathbf{U}_2$ can be done by Givens rotations as follows. We write

$$\mathbf{U}_1^T\,\mathbf{Z}_k\,\mathbf{U}_2 = \begin{bmatrix} \mathbf{G}_k & \mathbf{L}_k \\ \mathbf{H}_k & \mathbf{M}_k \end{bmatrix}, \qquad k = 1,2\,. \tag{5.9}$$

To maximize $\|\mathbf{G}_1\|_F^2 + \|\mathbf{G}_2\|_F^2$, iterative Givens rotations can be used for each pair of rows (i,j), $1 \le i \le R$, $R+1 \le j \le I$, in $\left[\begin{array}{c|c} \mathbf{G}_1 & \mathbf{G}_2 \\ \mathbf{H}_1 & \mathbf{H}_2 \end{array}\right]$, and for each pair of columns (i,j), $1 \le i \le R$, $R+1 \le j \le J$, in $\left[\begin{array}{cc} \mathbf{G}_1 & \mathbf{L}_1 \\ \hline \mathbf{G}_2 & \mathbf{L}_2 \end{array}\right]$. Analogous to the derivation of first-order optimality conditions for the GSD algorithm in section 5.1, we obtain the following conditions for stationary points of problem (5.6) in terms of (5.9):

- All rows of $[\mathbf{H}_1\,|\,\mathbf{H}_2]$ are orthogonal to all rows of $[\mathbf{G}_1\,|\,\mathbf{G}_2]$.
- All columns of $\begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \end{bmatrix}$ are orthogonal to all columns of $\begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix}$.

The rank of a best approximation $\mathcal{X}$ from the set $\widetilde{W}_R(I,J,2)$ is equal to the rank of the $R \times R \times 2$ array $\mathcal{G}$ with slices $\mathbf{G}_1$ and $\mathbf{G}_2$. As stated above, $\text{rank}(\mathcal{X}) = \text{rank}(\mathcal{G}) \le R$ implies that $\mathcal{Z}$ has a best rank-R approximation. The rank of $\mathcal{G}$ can be checked by making use of Lemma 2.1.

We have the following result for the case where $\text{rank}(\mathcal{X}) = \text{rank}(\mathcal{G}) > R$.

Theorem 5.3 *Let $\mathcal{Z} \in \mathbb{R}^{I\times J\times 2}$ with $2 \le R < \min(I,J)$ be generic. Let $(\mathbf{Q}_a, \mathbf{Q}_b, \mathbf{R}_1, \mathbf{R}_2)$ be an optimal solution of the GSD problem* (2.1) *with strictly positive second-order conditions* (3.13)-(3.14) *for $1 \le i < j \le R$. If all optimal solutions $\mathcal{X}$ of problem* (5.6) *have* $\text{rank}(\mathcal{X}) > R$*, then the rank of the $R \times R \times 2$ array $\mathcal{R}$ with slices $\mathbf{R}_1$ and $\mathbf{R}_2$ is larger than R.*

Proof. See Appendix D. □

Analogous to Conjecture 3.2, we assume that the second-order conditions (3.13)-(3.14), $1 \leq i < j \leq R$, are strictly positive for generic $\mathcal{Z} \in \mathbb{R}^{I\times J\times 2}$ with $2 \leq R < \min(I, J)$. As in Appendix A, we have found no counterexamples in numerical experiments (results not reported). Analogous to Corollary 5.2 following from Theorem 5.1, we obtain the following.

Corollary 5.4 *Let $\mathcal{Z} \in \mathbb{R}^{I\times J\times 2}$ with $2 \leq R < \min(I, J)$ be generic. If all optimal solutions $\mathcal{X}$ of problem* (5.6) *have* $\text{rank}(\mathcal{X}) > R$, *then $\mathcal{Z}$ does not have a best rank-R approximation.* □

Corollary 5.4 implies that we now have an easy-to-check criterion to determine whether $\mathcal{Z}$ has a best rank-R approximation or not. First, compute a best approximation $\mathcal{X}$ from the set $\widetilde{W}_R(I, J, 2)$ by using the algorithm with Givens rotations or the algorithms in [31] [17]. A number of runs with random starting values can be executed to make sure the global maximum is obtained and the optimal solution $\mathcal{X}$ is unique. As in case 8, array $\mathcal{G}$ (corresponding to $\mathcal{X}$) may be considered a generic $R \times R \times 2$ array. Hence, $\text{rank}(\mathcal{X}) = \text{rank}(\mathcal{G})$ equals R or $R + 1$, both on sets of positive Lebesgue measure [43]. Next, compute the eigenvalues of $\mathbf{G}_2\mathbf{G}_1^{-1}$, which are distinct. If all eigenvalues are real, then $\text{rank}(\mathcal{X}) = \text{rank}(\mathcal{G}) = R$ (Lemma 2.1) and $\mathcal{Z}$ has a best rank-R approximation, which can be taken equal to $\mathcal{X}$. If some eigenvalues are complex, then $\text{rank}(\mathcal{X}) = \text{rank}(\mathcal{G}) = R + 1$ (Lemma 2.1) and $\mathcal{Z}$ does not have a best rank-R approximation. Since both situations occur on sets of positive Lebesgue measure, this completes the proof of cases 3, 5, and 9 of Table 1.

6 Discussion

Using the Generalized Schur Decomposition (GSD) and its relation to the CPD, we have proven all conjectures of [35] on the (non)existence of best rank-R approximations for generic $I \times J \times 2$ arrays. Our main result is that generic $I \times I \times 2$ arrays of rank $I+1$ do not have a best rank-I approximation. So far, this was only proven for $I = 2$ [9], which was the only result on (non)existence of low-rank approximations for generic three-way arrays in the literature.

In cases 3, 5, 8, and 9 of Table 1, existence of a best rank-R approximation holds on a set of positive volume only. For such arrays we have obtained easy-to-check necessary and sufficient conditions for the existence of a best rank-R approximation. In case 8, it suffices to solve problem (4.2) by computing a truncated SVD and computing the eigenvalues of a corresponding matrix. In cases 3, 5, and 9, problem (5.6) needs to be solved, and the eigenvalues of a matrix corresponding to the optimal solution of (5.6) need to be computed. To the author's knowledge, this is the first time such conditions are formulated.

In our proofs, we have made use of the fact that the GSD solution set $P_R(I, J, 2)$ is the closure of the set $S_R(I, J, 2)$ of arrays with rank at most R [37]. Unfortunately, this result does not generalize to $I \times J \times K$ arrays with $K \geq 3$ and the Simultaneous Generalized Schur Decomposition [7], nor do we know any other closed form description of the closure of the rank-R set $S_R(I, J, K)$. Hence, at present the results for $I \times J \times 2$ arrays in this paper do not seem to be generalizable to $I \times J \times K$ arrays.

Appendix A: numerical evidence for Conjecture 3.2

Evidence for Conjecture 3.2 is provided by running the GSD algorithm for random $I \times I \times 2$ arrays $\mathcal{Z}$ of rank $I + 1$. The GSD algorithm is terminated when the relative decrease in error sum of squares drops below 10^{-9}. For each array, we record the maximum absolute value (over all pairs (i, j)) of the first-order conditions (3.11) and (3.12), and the minimum value (over all pairs (i, j)) of the second-order conditions (3.13) and (3.14). According to Conjecture 3.2, the latter should be strictly positive for an optimal GSD solution. To evaluate whether the GSD algorithm has terminated in a local minimum, we also compute the Hessian matrix of second order derivatives. Let the orthonormal $\mathbf{Q}_a$ and $\mathbf{Q}_b$ be parameterized by $I(I-1)/2$ Gives rotations each. We consider the Hessian matrix of the problem (3.2), where the variables are the $I(I-1)$ rotation angles. The first-order conditions are then equal to (3.11) and (3.12), and the second-order conditions (3.13) and (3.14) form the diagonal of the Hessian matrix. The Hessian matrix of the GSD problem was not considered in [7].

Note that the GSD algorithm decreases the objective value (3.2) with every Givens rotation, unless the corresponding second-order condition equals zero (then the objective value remains unchanged; see Lemma 3.1). Also, the GSD algorithm is expected to converge to a stationary point, i.e., with zero first-order conditions (3.11) and (3.12). Indeed, if a first-order condition is nonzero, then a Givens rotation exists that will decrease the objective value. Of course, since the GSD algorithm is terminated after a finite number of iterations, the first-order conditions (3.11) and (3.12) will not be exactly zero after convergence.

In [7], the GSD algorithm is initialized by taking $\mathbf{Q}_a$ and $\mathbf{Q}_b$ from the generalized real Schur decomposition [11, theorem 7.7.2] computed via QZ iteration, and an extensive simulation study yields no cases of suboptimal GSD solutions. Here, we also consider initialization with random orthonormal $\mathbf{Q}_a$ and $\mathbf{Q}_b$ and check whether the GSD algorithm terminates in suboptimal solutions. A GSD solution is suboptimal when a better GSD solution is found for different initial values in the GSD algorithm. For each $I \in \{2,3,4,5,6,7\}$ we generate 10 arrays $\mathcal{Z}$ and run the GSD algorithm 100 times with random initial values, and 1 time with the QZ initial values. In Table 2 (top half) the results are presented. All runs result in first-order conditions close to zero, with largest absolute value below 10^{-4}. All runs result in strictly positive second-order conditions, with the smallest value being 0.14 for $I = 3$. It is clear that Conjecture 3.2 is not violated in these simulations. All runs result in a positive definite Hessian matrix, indicating a local minimum, although for some runs the Hessian has a very small positive eigenvalue. Hence, we do not find any saddle points.

I	suboptimal %	max(abs(1st))	min(2nd)	min(eig(Hessian))
$I = 2$	0	$7 \cdot 10^{-16}$	0.15	0.0215
$I = 3$	3.9	$5 \cdot 10^{-5}$	0.14	0.0082
$I = 4$	6.4	$1 \cdot 10^{-4}$	0.77	0.0028
$I = 5$	15.1	$8 \cdot 10^{-5}$	0.59	0.0010
$I = 6$	29.5	$8 \cdot 10^{-5}$	0.59	0.0017
$I = 7$	25.2	$9 \cdot 10^{-5}$	0.86	0.0005
$I = 2$	-	$9 \cdot 10^{-16}$	0.01	0.0024
$I = 3$	-	$4 \cdot 10^{-16}$	0.01	0.0009
$I = 4$	-	$9 \cdot 10^{-5}$	0.08	0.0002
$I = 5$	-	$1 \cdot 10^{-4}$	0.29	$9 \cdot 10^{-5}$
$I = 6$	-	$1 \cdot 10^{-4}$	0.28	$7 \cdot 10^{-6}$
$I = 7$	-	$1 \cdot 10^{-4}$	0.36	$1 \cdot 10^{-5}$

Table 2: Percentage of suboptimal solutions, maximum absolute value of first-order conditions (3.11)-(3.12), minimum value of second-order conditions (3.13)-(3.14), and smallest eigenvalue of the Hessian for runs of the GSD algorithm for random $I \times I \times 2$ arrays $\mathcal{Z}$ of rank $I + 1$. Top half: 10 arrays per I; for each array 101 runs are executed, 100 with random starting values and 1 with QZ starting values; results for optimal runs are reported. Bottom half: 1000 arrays per I; for each array 1 run with QZ starting values is executed.

Quite some runs result in a suboptimal local minimum, but the QZ initialized runs result in a suboptimal solution for only 3 out of the 60 arrays. In all following simulations, we use only one QZ initialized run for each array. Next, we generate 1000 arrays $\mathcal{Z}$ for each $I \in \{2,3,4,5,6,7\}$ and again check the first- and second-order conditions, and the eigenvalues of the Hessian matrix. The results are reported in Table 2 (bottom half). Again the first-order conditions are close to zero, with largest absolute value below 10^{-4}. The second-order conditions are all positive, but the minimal values for $I = 2,3,4$ are rather small. However, they are still 10^{14} times larger than the largest first-order condition for $I = 2,3$, and 10^3 times larger for $I = 4$. Therefore, we do not consider this a violation of Conjecture 3.2. The smallest eigenvalues of the Hessian matrix are positive but small. Again we do not encounter any saddle points.

For larger values of I the symbolic computation of the Hessian matrix takes a lot of time (for $I = 7$ it takes 6 hours on a regular PC). Hence, we omit the computation of the Hessian in the final simulations for $I = 2,\ldots,25$. For each value of I, we generate 10 arrays $\mathcal{Z}$ and run the GSD algorithm with QZ initial values. The results are depicted in Figure 1 below. As can be seen, for all arrays the largest absolute value of the first-order conditions is below 10^{-4}, and the smallest value of the second-order conditions is strictly positive (with the smallest values being 0.24 for $I = 2$ and $I = 3$). Hence, also here Conjecture 3.2 is not violated.

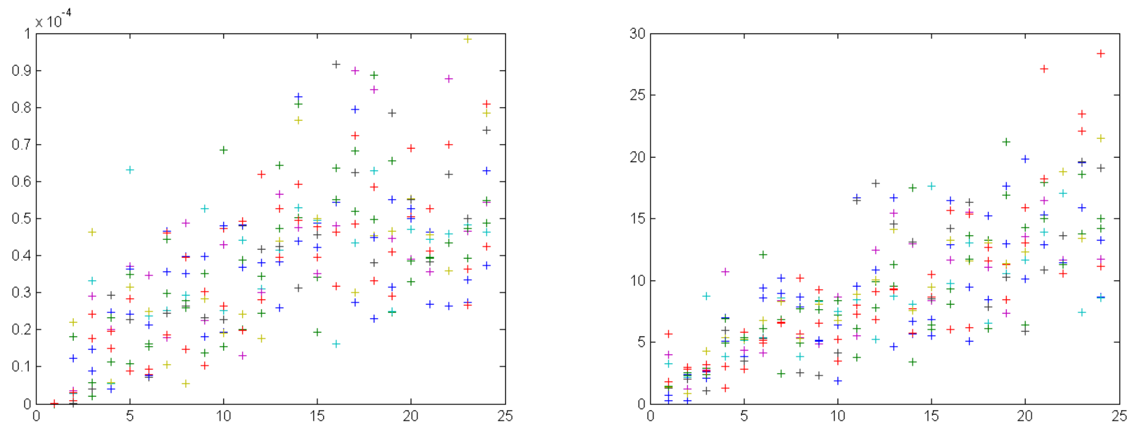

Figure 1: Maximum absolute value of first-order conditions (3.11)-(3.12) (left), and minimum value of second-order conditions (3.13)-(3.14) (right), after running the GSD algorithm for 10 random $I \times I \times 2$ arrays $\mathcal{Z}$ of rank $I+1$, for $I = 2,\ldots,25$.

Note that Conjecture 3.2 does not imply that zero second-order conditions do not occur for $I \times I \times 2$ arrays $\mathcal{Z}$ of rank $I+1$. The conjecture only states that this does not occur for *generic* $I \times I \times 2$ arrays $\mathcal{Z}$ of rank $I+1$. Below are two examples with zero second-order conditions for $I = 3$. Let

$$[\widetilde{\mathbf{Z}}_1 \,|\, \widetilde{\mathbf{Z}}_2] = \left[\begin{array}{ccc|ccc} 0 & 1 & 2 & 0 & 2 & -2 \\ 0 & 2 & 3 & 0 & 3 & -4 \\ 0.1 & 0 & 2 & 0.2 & 0 & -1 \end{array}\right] . \tag{A.1}$$

The matrix $\widetilde{\mathbf{Z}}_2\widetilde{\mathbf{Z}}_1^{-1}$ has complex eigenvalues, which implies rank$(\mathcal{Z}) = 4$ by Lemma 2.1. It can be verified that (A.1) satisfies the first-order conditions (3.11)-(3.12) and the second-order conditions (3.13)-(3.14). However, the second-order condition (3.13) is zero for $(i,j) = (1,2)$. The Hessian corresponding to (A.1) has eigenvalues -0.1, 0, 10.0, 22.1, 36.1, and 63.7. Hence, this is an example of a saddle point.

A second example is obtained by setting the (3,3) entries of $\widetilde{\mathbf{Z}}_1$ and $\widetilde{\mathbf{Z}}_2$ in (A.1) to zero. Let

$$[\widetilde{\mathbf{Z}}_1 \,|\, \widetilde{\mathbf{Z}}_2] = \left[\begin{array}{ccc|ccc} 0 & 1 & 2 & 0 & 2 & -2 \\ 0 & 2 & 3 & 0 & 3 & -4 \\ 0.1 & 0 & 0 & 0.2 & 0 & 0 \end{array}\right] . \tag{A.2}$$

Again $\widetilde{\mathbf{Z}}_2\widetilde{\mathbf{Z}}_1^{-1}$ has complex eigenvalues, and the first- and second-order conditions are satisfied. The second-order condition (3.13) is zero for $(i,j) = (1,2)$, and the second-order condition (3.14) is zero for $(i,j) = (2,3)$. The corresponding Hessian has eigenvalues 0, 0, 0.01, 20.9, 36.0, and 54.9. Hence, this is not a saddle point. Both examples (A.1) and (A.2) are suboptimal GSD solutions and can be improved by using the GSD algorithm with random starting values.

An alternative way to verify the strict positivity of the second-order conditions is by considering *all stationary points* of the GSD problem (3.2), and using a dimensionality argument. Consider the set Q_I of all $\mathbf{Q}_a$ and $\mathbf{Q}_b$ satisfying the first-order optimality conditions (3.11) and (3.12). We rewrite the latter explicitly in terms of $\mathbf{Q}_a$ and $\mathbf{Q}_b$ and obtain

$$\begin{aligned} Q_I = \{\mathbf{Q}_a, \mathbf{Q}_b \in \mathbb{R}^{I\times I} \quad : \quad & \mathbf{Q}_a^T\mathbf{Q}_a = \mathbf{Q}_b^T\mathbf{Q}_b = \mathbf{I}_I\,, \\ & \sum_{r=i}^{j-1}\sum_{k=1}^{2}(\mathbf{q}_{a,i}^T\mathbf{Z}_k\mathbf{q}_{b,r})(\mathbf{q}_{a,j}^T\mathbf{Z}_k\mathbf{q}_{b,r}) = 0\,, \quad 1 \le i < j \le I\,, \\ & \sum_{r=i+1}^{j}\sum_{k=1}^{2}(\mathbf{q}_{a,r}^T\mathbf{Z}_k\mathbf{q}_{b,i})(\mathbf{q}_{a,r}^T\mathbf{Z}_k\mathbf{q}_{b,j}) = 0\,, \quad 1 \le i < j \le I\,\}\,, \end{aligned} \tag{A.3}$$

where $\mathbf{q}_{a,r}$ and $\mathbf{q}_{b,r}$ denote the rth columns of $\mathbf{Q}_a$ and $\mathbf{Q}_b$, respectively. The number of equations in $\mathbf{Q}_a^T\mathbf{Q}_a = \mathbf{Q}_b^T\mathbf{Q}_b = \mathbf{I}_I$ equals $2\,I(I+1)/2 = I(I+1)$. The first-order optimality conditions form $2\,I(I-1)/2 = I(I-1)$ equations. Hence, the total number of equations in (A.3) equals $I(I+1) + I(I-1) = 2\,I^2$, which is equal to the number of unknowns in $\mathbf{Q}_a$ and $\mathbf{Q}_b$. If the dimensionality of the set Q_I equals zero, then it contains at most a finite number of $\mathbf{Q}_a$ and $\mathbf{Q}_b$. Since the GSD problem (3.2) always has an optimal solution, the set Q_I is not empty. Moreover, due to permutation ambiguities in the GSD, the set Q_I always contains more than one pair of $\mathbf{Q}_a$ and $\mathbf{Q}_b$ [7] [41]. Lemma 3.1 shows that a second-order condition being zero at a stationary point implies infinitely many optimal $\mathbf{Q}_a$ or $\mathbf{Q}_b$ (a one-dimensional set, in fact). Hence, this would contradict the dimensionality of the set Q_I being zero. With the use of a computer algebra package, the number of (real) solutions to the equations defining Q_I could be computed, for some random $I \times I \times 2$ array $\mathcal{Z}$ of rank $I+1$. When this number is finite, the second-order conditions are strictly positive. However, although this approach was numerically feasible for $I = 2$, for $I = 3$ no solution to the equations was found after 6 hours on a regular PC using Mathematica 9. Note that for $I = 3$ we have 18 4th degree poynomial equations in 18 variables. Another approach to verify that $\dim(Q_I) = 0$ is by using an algebraic geometry program to compute the dimension of the ideal generated by the polynomial equations in Q_I, with the rational numbers as base field for the coefficients. Hence, here $\mathbf{Z}_1$ and $\mathbf{Z}_2$ should be random and rational, which approximates the real case. We have used the Macaulay2 package and for $I = 2$ the dimensionality of the ideal was indeed zero. However, for $I = 3$ computing the dimensionality of the ideal did not produce an answer after several hours on a regular PC. Since the approach via $\dim(Q_I) = 0$ was not successful, we verified the strict positivity of the second-order conditions using the GSD algorithm instead.

In [7] the GSD algorithm is used as an algorithm for simultaneous matrix diagonalization. When the stationary points in Q_I could be computed easily, then a new fast algorithm for simultaneous matrix diagonalization would have been obtained. Moreover, all local and global minima would be known. This would be a huge improvement with respect to all other exisiting algorithms for simultaneous matrix diagonalization. Therefore, we do not expect that computing all solutions to the equations in Q_I is numerically feasible for I not very small.

Appendix B: proof of Theorem 3.3

First, we show that we may assume without loss of generality that the optimal $\mathbf{R}_1$ and $\mathbf{R}_2$ are nonsingular when the second-order conditions (3.13)-(3.14) are strictly positive. The optimal $\mathbf{R}_k$ equals the upper triangular part of $\widetilde{\mathbf{Z}}_k = \mathbf{Q}_a^T\,\mathbf{Z}_k\,\mathbf{Q}_b$, $k = 1, 2$. A singular $\mathbf{R}_k$ implies that it has one or more diagonal entries equal to zero. The second-order optimality conditions (3.13)-(3.14) for rotations $(i, i+1)$ are $\tilde{\mathbf{z}}_{(i,i)}^T \tilde{\mathbf{z}}_{(i,i)} \geq \tilde{\mathbf{z}}_{(i+1,i)}^T \tilde{\mathbf{z}}_{(i+1,i)}$, $i = 1, \ldots, I-1$, and $\tilde{\mathbf{z}}_{(i,i)}^T \tilde{\mathbf{z}}_{(i,i)} \geq \tilde{\mathbf{z}}_{(i,i-1)}^T \tilde{\mathbf{z}}_{(i,i-1)}$, $i = 2, \ldots, I$. Hence, positive second-order conditions imply that $\mathbf{R}_1$ and $\mathbf{R}_2$ do not both have a zero on position (i,i), $i = 1, \ldots, I$. In that case, a nonsingular slicemix $\mathbf{S}$ (2×2) exists, such that the mixed slices $\widehat{\mathbf{R}}_k = s_{k1}\,\mathbf{R}_1 + s_{k2}\,\mathbf{R}_2$, $k = 1, 2$, are nonsingular. The matrix $\mathbf{S}$ may be taken orthonormal. Then the GSD solution $(\mathbf{Q}_a, \mathbf{Q}_b, \widehat{\mathbf{R}}_1, \widehat{\mathbf{R}}_2)$ is an optimal solution of the GSD problem for array $\widehat{\mathcal{Z}}$ with mixed slices $\widehat{\mathbf{Z}}_k = s_{k1}\,\mathbf{Z}_1 + s_{k2}\,\mathbf{Z}_2$, $k = 1, 2$. The optimal $\widehat{\mathbf{R}}_k$, $k = 1, 2$, are nonsingular, $\text{rank}(\mathcal{Z}) = \text{rank}(\widehat{\mathcal{Z}})$, and $\text{rank}(\mathcal{R}) = \text{rank}(\widehat{\mathcal{R}})$. Hence, a proof of $\text{rank}(\widehat{\mathcal{R}}) > I$ implies $\text{rank}(\mathcal{R}) > I$. Therefore, positive second-order conditions imply that we may assume without loss of generality that the optimal $\mathbf{R}_1$ and $\mathbf{R}_2$ are nonsingular.

As rank criterion for $\mathcal{R}$ we use Lemma 2.1 (ii). We show that

(Bi) $\mathbf{R}_2\mathbf{R}_1^{-1}$ has some identical eigenvalues,

(Bii) $\mathbf{R}_2\mathbf{R}_1^{-1}$ does not have I linearly independent eigenvectors.

First, we prove (Bi) by contradiction. Suppose the eigenvalues of $\mathbf{R}_2\mathbf{R}_1^{-1}$ are distinct. Then $\text{rank}(\mathcal{R}) = I$ by Lemma 2.1 (i). The eigenvalues of $\mathbf{R}_2\mathbf{R}_1^{-1}$ are equal to the diagonal entries of $\widetilde{\mathbf{Z}}_2$ divided by those of $\widetilde{\mathbf{Z}}_1$. Statement (Bi) not holding is equivalent to none of the vectors $\tilde{\mathbf{z}}_{(i,i)}$, $i = 1, \ldots, I$, being proportional. The nonsingularity of $\mathbf{R}_1$ and $\mathbf{R}_2$ implies that vectors $\tilde{\mathbf{z}}_{(i,i)}$ do not contain zeros, $i = 1, \ldots, I$. Consider the optimality conditions (3.11) and (3.12) for rotations $(i, i+1)$, which are: $\tilde{\mathbf{z}}_{(i,i)}^T \tilde{\mathbf{z}}_{(i+1,i)} = \tilde{\mathbf{z}}_{(i+1,i)}^T \tilde{\mathbf{z}}_{(i+1,i+1)} = 0$. Since $\tilde{\mathbf{z}}_{(i,i)}$ and $\tilde{\mathbf{z}}_{(i+1,i+1)}$ are not proportional and not zero, this implies $\tilde{\mathbf{z}}_{(i+1,i)} = \mathbf{0}$ for $i = 1, \ldots, I-1$. Here, $\mathbf{0}$ denotes the zero vector in $\mathbb{R}^2$.

Using this result, we consider (3.11) and (3.12) for rotations $(i, i+2)$. These equations now become (the sum terms vanish): $\tilde{\mathbf{z}}_{(i,i)}^T \tilde{\mathbf{z}}_{(i+2,i)} = \tilde{\mathbf{z}}_{(i+2,i)}^T \mathbf{z}_{(i+2,i+2)} = 0$. Since $\tilde{\mathbf{z}}_{(i,i)}$ and $\tilde{\mathbf{z}}_{(i+2,i+2)}$ are not proportional and not zero, this implies $\tilde{\mathbf{z}}_{(i+2,i)} = \mathbf{0}$ for $i = 1, \ldots, I-2$. By consecutively considering rotations $(i, i+q)$ in this way, it is clear that we obtain $\tilde{\mathbf{z}}_{(j,i)} = \mathbf{0}$ for $1 \leq i < j \leq I$. This implies that $\widetilde{\mathbf{Z}}_k = \mathbf{Q}_a^T\,\mathbf{Z}_k\,\mathbf{Q}_b$, $k = 1, 2$, are upper triangular. Moreover, $\widetilde{\mathbf{Z}}_k = \mathbf{R}_k$, $k = 1, 2$, and $\text{rank}(\mathcal{Z}) = \text{rank}(\mathcal{R}) = I$, which contradicts the assumption of $\text{rank}(\mathcal{Z}) = I + 1$ in Theorem 3.3. Hence, we have proven that (Bi) holds.

For the proof of (Bii) we first reorder the eigenvalues of $\mathbf{R}_2\mathbf{R}_1^{-1}$ such that identical eigenvalues appear in contiguous groups. The reordering can be done within the GSD; see Kressner [21]. The proof of (Bii) is by contradiction. We suppose that $\mathbf{R}_2\mathbf{R}_1^{-1}$ has I linearly independent eigenvectors, which implies rank($\mathcal{R}$) $= I$ by Lemma 2.1 (i). Hence, for eigenvalue λ of $\mathbf{R}_2\mathbf{R}_1^{-1}$ with multiplicity d there are d linearly independent eigenvectors. In other words, $\text{rank}(\mathbf{R}_2\mathbf{R}_1^{-1} - \lambda\,\mathbf{I}_I) = I - d$, which is equivalent to $\text{rank}(\mathbf{R}_2 - \lambda\,\mathbf{R}_1) = I - d$. The diagonal of $\mathbf{R}_2 - \lambda\,\mathbf{R}_1$ consists of $I - d$ nonzero entries and d zeros which form the diagonal of a $d \times d$ strictly upper triangular block $\mathbf{V}$. We consider $\mathbf{R}_2 - \lambda\,\mathbf{R}_1$ as a block upper triangular matrix. Its rank is at least equal to the sum of the ranks of its diagonal blocks. These blocks are $I - d$ nonzero scalars and the block $\mathbf{V}$. Hence, the lower bound for the rank of $\mathbf{R}_2 - \lambda\,\mathbf{R}_1$ includes the value $I - d$ only if $\text{rank}(\mathbf{V}) = 0$. Conversely, row and column operations can be used to show that $\text{rank}(\mathbf{V}) = 0$ implies $\text{rank}(\mathbf{R}_2 - \lambda\,\mathbf{R}_1) = I - d$. Rank($\mathbf{V}$) $= 0$ is equivalent to all vectors $\tilde{\mathbf{z}}_{(m,n)}$ contained in the upper triangular part of the $d \times d$ block (with $m \leq n$) being either proportional or zero. Proposition B.1 below shows that the vectors $\tilde{\mathbf{z}}_{(m,n)}$ with $m > n$ in the block are zero. To summarize, let the corresponding $d \times d$ diagonal block of $\widetilde{\mathbf{Z}}_k = \mathbf{Q}_a^T\,\mathbf{Z}_k\,\mathbf{Q}_b$ be denoted by $\mathbf{W}_k$, $k = 1, 2$. Then $\mathbf{W}_1$ and $\mathbf{W}_2$ are upper triangular and $\mathbf{W}_2 - \lambda\,\mathbf{W}_1 = \mathbf{O}_{d,d}$, where $\mathbf{O}_{d,d}$ denotes the $d \times d$ zero matrix.

The last part of the proof of (Bii) is similar to the proof of (Bi): we show that $\widetilde{\mathbf{Z}}_k$, $k = 1, 2$, are upper triangular, which implies rank($\mathcal{Z}$) $=$ rank($\mathcal{R}$) $= I$. The contradiction with rank($\mathcal{Z}$) $= I + 1$ then implies that $\mathbf{R}_2\mathbf{R}_1^{-1}$ does not have I linearly independent eigenvectors. The diagonal of $\widetilde{\mathbf{Z}}_k$ consists of blocks $\mathbf{W}_k^{(1)}, \ldots, \mathbf{W}_k^{(L)}$ corresponding to L distinct eigenvalues of $\mathbf{R}_2\mathbf{R}_1^{-1}$. Let blocks $\mathbf{W}_k^{(l)}$, $k = 1, 2$, have size $d_l \times d_l$, $l = 1, \ldots, L$. A 1×1 block corresponds to a unique eigenvalue, and a $d_l \times d_l$ block with $d_l \geq 2$ corresponds to an eigenvalue with multiplicity d_l. From Proposition B.1 we know that the $d_l \times d_l$ blocks $\mathbf{W}_k^{(l)}$, $k = 1, 2$, are upper triangular. Consider the optimality conditions (3.11) and (3.12) for rotations $(i, i+1)$ such that entry $(i, i+1)$ is not part of any block $\mathbf{W}_k^{(l)}$. These are: $\tilde{\mathbf{z}}_{(i,i)}^T\tilde{\mathbf{z}}_{(i+1,i)} = \tilde{\mathbf{z}}_{(i+1,i)}^T\tilde{\mathbf{z}}_{(i+1,i+1)} = 0$. Since $\tilde{\mathbf{z}}_{(i,i)}$ and $\tilde{\mathbf{z}}_{(i+1,i+1)}$ are not proportional (they are not in a block $\mathbf{W}_k^{(l)}$) and not zero, this implies $\tilde{\mathbf{z}}_{(i+1,i)} = \mathbf{0}$. Note that $\tilde{\mathbf{z}}_{(i+1,i)} = \mathbf{0}$ for all $(i + 1, i)$ in block $\mathbf{W}_k^{(l)}$ by Proposition B.1.

As in the proof of (Bi), next we consider (3.11) and (3.12) for rotations $(i, i + 2)$ such that entry $(i, i + 2)$ is not in a block $\mathbf{W}_k^{(l)}$. These equations now become (the sum terms vanish): $\tilde{\mathbf{z}}_{(i,i)}^T\tilde{\mathbf{z}}_{(i+2,i)} = \tilde{\mathbf{z}}_{(i+2,i)}^T\mathbf{z}_{(i+2,i+2)} = 0$. Since $\tilde{\mathbf{z}}_{(i,i)}$ and $\tilde{\mathbf{z}}_{(i+2,i+2)}$ are not proportional and not zero, this implies $\tilde{\mathbf{z}}_{(i+2,i)} = \mathbf{0}$. Proceeding in the same way, we obtain $\tilde{\mathbf{z}}_{(j,i)} = \mathbf{0}$ for $1 \leq i < j \leq I$, which implies that $\widetilde{\mathbf{Z}}_k$, $k = 1, 2$, are upper triangular. This completes the proof of (Bii).

It remains to state and prove Proposition B.1.

Proposition B.1 *Let* $\mathcal{Z} \in \mathbb{R}^{I\times I\times 2}$ *be generic with* $\text{rank}(\mathcal{Z}) = I + 1$. *Let* $(\mathbf{Q}_a, \mathbf{Q}_b, \mathbf{R}_1, \mathbf{R}_2)$ *be an optimal solution of the GSD problem* (2.1)*, with nonsingular* $\mathbf{R}_1$ *and* $\mathbf{R}_2$. *For* $d \geq 2$*, let* $\mathbf{W}_k$ *be a* $d \times d$ *diagonal block of* $\mathbf{Q}_a^T\, \mathbf{Z}_k\, \mathbf{Q}_b$, $k = 1, 2$*, such that the upper triangular part of* $\mathbf{W}_2 - \lambda\, \mathbf{W}_1$ *is zero for some* $\lambda \neq 0$. *Then* $\mathbf{W}_1$ *and* $\mathbf{W}_2$ *are upper triangular.*

Proof. The proof is by induction on d. Let $\mathbf{W}_k^{(h)}$ be the $d \times d$ submatrix of $\widetilde{\mathbf{Z}}_k = \mathbf{Q}_a^T\, \mathbf{Z}_k\, \mathbf{Q}_b$ consisting of rows $h, \ldots, h+d-1$ and columns $h, \ldots, h+d-1$. Recall the definition of the vectors $\tilde{\mathbf{z}}_{(m,n)}$ in (3.10). Note that since $\mathbf{R}_k$ are nonsingular, $k = 1, 2$, the vectors $\tilde{\mathbf{z}}_{(i,i)}$ do not contain zeros, $i = 1, \ldots, I$.

First, we consider $d = 2$. We write the $2 \times 2 \times 2$ array $\mathcal{W}^{(h)}$ with unfolding $[\mathbf{W}_1^{(h)}\ \mathbf{W}_2^{(h)}]$ in terms of its 3-mode vectors $\tilde{\mathbf{z}}_{(m,n)}$ as

$$\begin{bmatrix} \tilde{\mathbf{z}}_{(h,h)} & \tilde{\mathbf{z}}_{(h,h+1)} \\ \tilde{\mathbf{z}}_{(h+1,h)} & \tilde{\mathbf{z}}_{(h+1,h+1)} \end{bmatrix}, \tag{B.1}$$

where $\tilde{\mathbf{z}}_{(h,h)}$ and $\tilde{\mathbf{z}}_{(h+1,h+1)}$ are proportional, and $\tilde{\mathbf{z}}_{(h,h+1)}$ is either zero or proportional to $\tilde{\mathbf{z}}_{(h,h)}$. The proof is by contradiction. Suppose $\tilde{\mathbf{z}}_{(h+1,h)} \neq \mathbf{0}$. By the proportionality of $\tilde{\mathbf{z}}_{(h,h+1)}$ and $\tilde{\mathbf{z}}_{(h,h)}$, an orthonormal rotation of the rows of $\mathbf{W}_k^{(h)}$, $k = 1, 2$, exists that makes $\tilde{\mathbf{z}}_{(h,h+1)}$ zero. This rotation is not needed when $\tilde{\mathbf{z}}_{(h,h+1)} = \mathbf{0}$. Next, swapping rows and columns yields upper triangular blocks $\mathbf{W}_k^{(h)}$, $k = 1, 2$. This implies a better GSD solution has been found, which is a contradiction. Hence, it follows that $\tilde{\mathbf{z}}_{(h+1,h)} = \mathbf{0}$. Note that the transformations used do not affect the GSD objective function outside of the blocks $\mathbf{W}_k^{(h)}$, $k = 1, 2$. This completes the proof for $d = 2$.

Next, we consider $d = 3$. We write the $3 \times 3 \times 2$ array $\mathcal{W}^{(h)}$ with unfolding $[\mathbf{W}_1^{(h)}\ \mathbf{W}_2^{(h)}]$ as

$$\begin{bmatrix} \tilde{\mathbf{z}}_{(h,h)} & \tilde{\mathbf{z}}_{(h,h+1)} & \tilde{\mathbf{z}}_{(h,h+2)} \\ \tilde{\mathbf{z}}_{(h+1,h)} & \tilde{\mathbf{z}}_{(h+1,h+1)} & \tilde{\mathbf{z}}_{(h+1,h+2)} \\ \tilde{\mathbf{z}}_{(h+2,h)} & \tilde{\mathbf{z}}_{(h+2,h+1)} & \tilde{\mathbf{z}}_{(h+2,h+2)} \end{bmatrix}, \tag{B.2}$$

where $\tilde{\mathbf{z}}_{(h+i,h+i)}$, $i = 0, 1, 2$, are proportional, and $\tilde{\mathbf{z}}_{(h+i,h+j)}$ with $0 \leq i < j \leq 2$ are either zero or proportional to $\tilde{\mathbf{z}}_{(h,h)}$. The proof for $d = 2$ applies to the subblock consisting of the first two rows and columns, and to the subblock consisting of the last two rows and columns. This implies $\tilde{\mathbf{z}}_{(h+1,h)} = \tilde{\mathbf{z}}_{(h+2,h+1)} = \mathbf{0}$. The proof is by contradiction. Suppose $\tilde{\mathbf{z}}_{(h+2,h)} \neq \mathbf{0}$. Let 'Row$(i,j)$' to denote an orthonormal rotation of rows i and j of $\mathbf{W}_k^{(h)}$, $k = 1, 2$. Next, we apply the following sequence of orthonormal row rotations:

$$\begin{bmatrix} \tilde{\mathbf{z}}_{(h,h)} & \tilde{\mathbf{z}}_{(h,h+1)} & \tilde{\mathbf{z}}_{(h,h+2)} \\ \mathbf{0} & \tilde{\mathbf{z}}_{(h+1,h+1)} & \tilde{\mathbf{z}}_{(h+1,h+2)} \\ \tilde{\mathbf{z}}_{(h+2,h)} & \mathbf{0} & \tilde{\mathbf{z}}_{(h+2,h+2)} \end{bmatrix} \overset{\text{Row}(h,h+1)}{\longrightarrow} \begin{bmatrix} \bar{\mathbf{z}}_{(h,h)} & \mathbf{0} & \bar{\mathbf{z}}_{(h,h+2)} \\ \bar{\mathbf{z}}_{(h+1,h)} & \bar{\mathbf{z}}_{(h+1,h+1)} & \bar{\mathbf{z}}_{(h+1,h+2)} \\ \tilde{\mathbf{z}}_{(h+2,h)} & \mathbf{0} & \tilde{\mathbf{z}}_{(h+2,h+2)} \end{bmatrix}$$

$$\overset{\text{Row}(h,h+2)}{\longrightarrow} \begin{bmatrix} \bar{\bar{\mathbf{z}}}_{(h,h)} & \mathbf{0} & \mathbf{0} \\ \bar{\mathbf{z}}_{(h+1,h)} & \bar{\mathbf{z}}_{(h+1,h+1)} & \bar{\mathbf{z}}_{(h+1,h+2)} \\ \bar{\mathbf{z}}_{(h+2,h)} & \mathbf{0} & \bar{\mathbf{z}}_{(h+2,h+2)} \end{bmatrix}.$$

Hence, first we rotate rows h and $h+1$ such that $\tilde{\mathbf{z}}_{(h,h+1)}$ becomes zero (when $\tilde{\mathbf{z}}_{(h,h+1)}$ is not zero already). Note that $\bar{\mathbf{z}}_{(h,h+2)}$ and $\tilde{\mathbf{z}}_{(h+2,h+2)}$ are proportional or $\bar{\mathbf{z}}_{(h,h+2)} = \mathbf{0}$. Then we rotate rows h and $h+2$ to make $\bar{\mathbf{z}}_{(h,h+2)} = \mathbf{0}$. After this, swapping columns $h+1$ and $h+2$ and swapping rows $h+1$ and $h+2$ makes the blocks lower triangular. Then reversing the order of the rows and reversing the order of the columns makes the blocks upper triangular. This yields a better GSD solution, which is a contradiction. Hence, we obtain that $\tilde{\mathbf{z}}_{(h+2,h)} = \mathbf{0}$. Note that none of the transformations used affects the GSD objective function outside of the blocks $\mathbf{W}_k^{(h)}$, $k = 1, 2$. This completes the proof for $d = 3$.

Next, we assume the result holds for d and prove it for $d+1$. By the induction hypothesis, the $(d+1) \times (d+1) \times 2$ array $\mathcal{W}^{(h)}$ is of the form

$$\begin{bmatrix} \tilde{\mathbf{z}}_{(h,h)} & \tilde{\mathbf{z}}_{(h,h+1)} & \cdots & \cdots & \tilde{\mathbf{z}}_{(h,h+d)} \\ \mathbf{0} & \tilde{\mathbf{z}}_{(h+1,h+1)} & & & \vdots \\ \vdots & \ddots & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & \ddots & \ddots & \vdots \\ \tilde{\mathbf{z}}_{(h+d,h)} & \mathbf{0} & \cdots & \mathbf{0} & \tilde{\mathbf{z}}_{(h+d,h+d)} \end{bmatrix}. \tag{B.3}$$

The proof is by contradiction. Suppose $\tilde{\mathbf{z}}_{(h+d,h)} \neq \mathbf{0}$. Next, we apply consecutive rotations of rows h and $h+i$ of $\mathbf{W}_k^{(h)}$, $k = 1, 2$, to make the current $\tilde{\mathbf{z}}_{(h,h+i)}$ zero (if it is not zero already), for $i = 1, \ldots, d$. This yields the form

$$\begin{bmatrix} \bar{\mathbf{z}}_{(h,h)} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \bar{\mathbf{z}}_{(h+1,h)} & \bar{\mathbf{z}}_{(h+1,h+1)} & \cdots & \cdots & \bar{\mathbf{z}}_{(h+1,h+d)} \\ \vdots & \mathbf{0} & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \bar{\mathbf{z}}_{(h+d,h)} & \mathbf{0} & \cdots & \mathbf{0} & \bar{\mathbf{z}}_{(h+d,h+d)} \end{bmatrix}. \tag{B.4}$$

Reversing the order of columns $h+1,\ldots,h+d$ and reversing the order of rows $h+1,\ldots,h+d$ then makes the blocks lower triangular. Next, reversing the order of the rows and reversing the order of the columns makes the blocks upper triangular. This yields a better GSD solution, which is a contradiction. Hence, we obtain $\tilde{\mathbf{z}}_{(h+d,h)} = \mathbf{0}$. Again, note that none of the transformations used affects the GSD objective function outside of the blocks $\mathbf{W}_k^{(h)}$, $k=1,2$. This completes the proof of Proposition B.1. □

Appendix C: proof of Theorem 5.1

The structure of the proof is analogous to the proof of Theorem 3.3. As explained in Appendix B, the strictly positive second-order optimality conditions (3.13)-(3.14) for $1 \leq i < j \leq R$ imply that we may assume without loss of generality that $\mathbf{R}_1$ and $\mathbf{R}_2$ of an optimal GSD solution are nonsingular. As rank criterion for $\mathcal{R}$ we use again Lemma 2.1 (ii). We show that

(Ci) $\mathbf{R}_2\mathbf{R}_1^{-1}$ has some identical eigenvalues,

(Cii) $\mathbf{R}_2\mathbf{R}_1^{-1}$ does not have R linearly independent eigenvectors.

We write

$$\widetilde{\mathbf{Z}}_k = \widetilde{\mathbf{Q}}_a^T\,\mathbf{Z}_k\,\mathbf{Q}_b = \begin{bmatrix} \widetilde{\mathbf{G}}_k \\ \widetilde{\mathbf{H}}_k \end{bmatrix}, \qquad k=1,2\,, \tag{C.1}$$

where $\widetilde{\mathbf{G}}_k$ is $R\times R$, and $\widetilde{\mathbf{H}}_k$ is $(I-R)\times R$, $k=1,2$. The GSD algorithm finds $\widetilde{\mathbf{Q}}_a$ and $\mathbf{Q}_b$ such that the Frobenius norm of the upper triangular parts of $\widetilde{\mathbf{G}}_k$, $k=1,2$, is maximized. The optimal $\mathbf{R}_k$ is taken as the upper triangular part of $\widetilde{\mathbf{G}}_k$, $k=1,2$. Note that $\widetilde{\mathbf{Q}}_b = \mathbf{Q}_b$ since $R=J$.

First, we prove (Ci) by contradiction. Suppose (Ci) does not hold, i.e., all eigenvalues of $\mathbf{R}_2\mathbf{R}_1^{-1}$ are distinct, which implies $\text{rank}(\mathcal{R}) = R$ by Lemma 2.1 (i). As in the proof of Theorem 3.3, optimality conditions (5.1)–(5.2) then imply that $\widetilde{\mathbf{G}}_k$ are upper triangular, $k=1,2$. We write $\widetilde{\mathbf{G}}_k = \mathbf{R}_k$, $k=1,2$. The SVD of $[\mathbf{Z}_1\,|\,\mathbf{Z}_2]$ is given as $\mathbf{U}\,\mathbf{S}\,\mathbf{V}^T$, with $\mathbf{U}^T\mathbf{U} = \mathbf{I}_I$ and $\mathbf{V}^T\mathbf{V} = \mathbf{I}_{2J}$. The best approximation $\mathcal{X}$ from the set $W_R(I,J,2)$ is given by the truncated SVD and has slices $[\mathbf{X}_1\,|\,\mathbf{X}_2] = \mathbf{U}_R\,\mathbf{S}_R\,\mathbf{V}_R^T = \mathbf{U}_R\,[\mathbf{S}_R\mathbf{V}_{R,1}^T\,|\,\mathbf{S}_R\mathbf{V}_{R,2}^T]$. The rank of $\mathcal{X}$ is equal to the rank of the $R\times R\times 2$ array $\mathcal{V}_R$ with $R\times R$ slices $\mathbf{S}_R\mathbf{V}_{R,1}^T$ and $\mathbf{S}_R\mathbf{V}_{R,2}^T$. In Theorem 5.1 it is assumed that $\text{rank}(\mathcal{X}) = \text{rank}(\mathcal{V}_R) > R$.

Note that $\widetilde{\mathbf{Z}}_k = \widetilde{\mathbf{Q}}_a^T \mathbf{Z}_k \mathbf{Q}_b = (\mathbf{U}^T\widetilde{\mathbf{Q}}_a)^T (\mathbf{U}^T\mathbf{Z}_k) \mathbf{Q}_b$, $k = 1, 2$. We write

$$[\widetilde{\mathbf{Z}}_1 \,|\, \widetilde{\mathbf{Z}}_2] = (\mathbf{U}^T\widetilde{\mathbf{Q}}_a)^T (\mathbf{S}\mathbf{V}^T) \begin{bmatrix} \mathbf{Q}_b & \mathbf{O}_{J,J} \\ \mathbf{O}_{J,J} & \mathbf{Q}_b \end{bmatrix} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \widetilde{\mathbf{H}}_1 & \widetilde{\mathbf{H}}_2 \end{bmatrix}. \tag{C.2}$$

The optimality conditions (5.3), together with $\mathbf{R}_1$ and $\mathbf{R}_2$ being upper triangular, yield that all rows of $[\widetilde{\mathbf{H}}_1\ \widetilde{\mathbf{H}}_2]$ are orthognal to all rows of $[\mathbf{R}_1\ \mathbf{R}_2]$. Postmultiplying (C.2) by its transpose yields

$$(\mathbf{U}^T\widetilde{\mathbf{Q}}_a)^T \mathbf{S}\mathbf{S}^T (\mathbf{U}^T\widetilde{\mathbf{Q}}_a) = \begin{bmatrix} \mathbf{P}_1 & \mathbf{O}_{R,I-R} \\ \mathbf{O}_{I-R,R} & \mathbf{P}_2 \end{bmatrix}, \tag{C.3}$$

where $\mathbf{P}_1$ is $R \times R$ and symmetric and nonsingular (since $[\mathbf{R}_1\ \mathbf{R}_2]$ has full row rank due to nonsingularity of $\mathbf{R}_k$), and $\mathbf{P}_2$ is $(I-R) \times (I-R)$ and symmetric. Matrix $\mathbf{S}$ is $I \times 2J$ and contains the $\min(I, 2J)$ nonzero singular values on its diagonal. Note that all singular values are nonzero and distinct since $[\mathbf{Z}_1 \,|\, \mathbf{Z}_2]$ is generic.

First, suppose $I \leq 2J$. Then $\mathbf{S}\mathbf{S}^T$ is $I \times I$ diagonal and nonsingular. Since the diagonal entries of $\mathbf{S}\mathbf{S}^T$ are positive and distinct, and $\mathbf{U}^T\widetilde{\mathbf{Q}}_a$ is orthonormal, equation (C.3) is the eigendecomposition of an $I \times I$ symmetric matrix that is nonsingular. In fact, (C.3) can be obtained as the superposition of the eigendecompositions of $\mathbf{P}_1$ and $\mathbf{P}_2$ (both nonsingular). This implies that

$$\mathbf{U}^T\widetilde{\mathbf{Q}}_a = \begin{bmatrix} \widetilde{\mathbf{Q}}_a^{(1)} & \mathbf{O}_{R,I-R} \\ \mathbf{O}_{I-R,R} & \widetilde{\mathbf{Q}}_a^{(2)} \end{bmatrix}, \tag{C.4}$$

where $\widetilde{\mathbf{Q}}_a^{(1)}$ $(R \times R)$ and $\widetilde{\mathbf{Q}}_a^{(2)}$ $((I-R) \times (I-R))$ are orthonormal. From (C.2) we then obtain

$$\begin{aligned} \mathbf{S}\,\mathbf{V}^T &= \begin{bmatrix} \widetilde{\mathbf{Q}}_a^{(1)} & \mathbf{O}_{R,I-R} \\ \mathbf{O}_{I-R,R} & \widetilde{\mathbf{Q}}_a^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \widetilde{\mathbf{H}}_1 & \widetilde{\mathbf{H}}_2 \end{bmatrix} \begin{bmatrix} \mathbf{Q}_b^T & \mathbf{O}_{J,J} \\ \mathbf{O}_{J,J} & \mathbf{Q}_b^T \end{bmatrix} \\ &= \begin{bmatrix} \widetilde{\mathbf{Q}}_a^{(1)}\mathbf{R}_1\mathbf{Q}_b^T & \widetilde{\mathbf{Q}}_a^{(1)}\mathbf{R}_2\mathbf{Q}_b^T \\ \widetilde{\mathbf{Q}}_a^{(2)}\widetilde{\mathbf{H}}_1\mathbf{Q}_b^T & \widetilde{\mathbf{Q}}_a^{(2)}\widetilde{\mathbf{H}}_2\mathbf{Q}_b^T \end{bmatrix}. \end{aligned} \tag{C.5}$$

This implies that array $\mathcal{V}_R$ has slices $\mathbf{S}_R\mathbf{V}_{R,k}^T = \widetilde{\mathbf{Q}}_a^{(1)}\,\mathbf{R}_k\,\mathbf{Q}_b^T$, $k = 1, 2$. Hence, $\text{rank}(\mathcal{V}_R) = \text{rank}(\mathcal{R}) = R$, which contradicts $\text{rank}(\mathcal{X}) = \text{rank}(\mathcal{V}_R) > R$. Hence, $\widetilde{\mathbf{G}}_k$, $k = 1, 2$, cannot be upper triangular and (Ci) must hold.

Next, suppose $I > 2J$. Then $\mathbf{S}\mathbf{S}^T$ is $I \times I$ diagonal with the first $2J$ diagonal entries positive and distinct and the last $I - 2J$ diagonal entries zero. In (C.3), matrix $\mathbf{P}_1$ is nonsingular and $\mathbf{P}_2$ has rank $J = R$. As above, it follows that

$$(\mathbf{U}^T\widetilde{\mathbf{Q}}_a)^T = \begin{bmatrix} (\widetilde{\mathbf{Q}}_a^{(1)})^T & \mathbf{O}_{R,J} & \mathbf{N}_1 \\ \mathbf{O}_{I-R,R} & (\widehat{\mathbf{Q}}_a^{(2)})^T & \mathbf{N}_2 \end{bmatrix}, \tag{C.6}$$

where $(\widehat{\mathbf{Q}}_a^{(2)})^T$ is $(I-R)\times J$ and contains the eigenvectors of $\mathbf{P}_2$, and $\mathbf{N}_1$ and $\mathbf{N}_2$ have $I-2J$ columns. Since the matrix $\mathbf{U}^T\widetilde{\mathbf{Q}}_a$ is orthonormal, and $\widetilde{\mathbf{Q}}_a^{(1)}$ is orthonormal, it follows that $\mathbf{N}_1$ is zero. Hence, $\mathbf{U}^T\widetilde{\mathbf{Q}}_a$ is of the same form as in (C.4), and the remaining part of the proof is as above. This completes the proof of (Ci).

Finally, we prove (Cii) by contradiction. Suppose $\mathbf{R}_2\mathbf{R}_1^{-1}$ has R linearly independent eigenvectors. Hence, rank($\mathcal{R}$) $= R$ by Lemma 2.1 (i). As in the proof of Theorem 3.3, the optimality conditions (5.1)–(5.2) then imply that $\widetilde{\mathbf{G}}_k$, $k=1,2$, are upper triangular. This yields the contradiction rank($\mathcal{X}$) $=$ rank($\mathcal{V}_R$) $=$ rank($\mathcal{R}$) $= R$ as shown in the proof of (Ci) above. Hence, (Cii) must hold. This completes the proof of Theorem 5.1.

Appendix D: proof of Theorem 5.3

The structure of the proof is analogous to the proof of Theorem 5.1. As before, the strictly positive second-order conditions imply that we may assume without loss of generality that $\mathbf{R}_1$ and $\mathbf{R}_2$ of an optimal GSD solution are nonsingular. As rank criterion for $\mathcal{R}$ we use again Lemma 2.1 (ii). We show that

(Di) $\mathbf{R}_2\mathbf{R}_1^{-1}$ has some identical eigenvalues,

(Dii) $\mathbf{R}_2\mathbf{R}_1^{-1}$ does not have R linearly independent eigenvectors.

We write

$$\widetilde{\mathbf{Z}}_k = \widetilde{\mathbf{Q}}_a^T\,\mathbf{Z}_k\,\widetilde{\mathbf{Q}}_b = \begin{bmatrix} \widetilde{\mathbf{G}}_k & \widetilde{\mathbf{L}}_k \\ \widetilde{\mathbf{H}}_k & \widetilde{\mathbf{M}}_k \end{bmatrix}, \qquad k=1,2\,. \tag{D.1}$$

where $\widetilde{\mathbf{G}}_k$ is $R\times R$, $\widetilde{\mathbf{H}}_k$ is $(I-R)\times R$, $\widetilde{\mathbf{L}}_k$ is $R\times(J-R)$, and $\widetilde{\mathbf{M}}_k$ is $(I-R)\times(J-R)$, $k=1,2$. The GSD algorithm finds $\widetilde{\mathbf{Q}}_a$ and $\widetilde{\mathbf{Q}}_b$ such that the Frobenius norm of the upper triangular parts of $\widetilde{\mathbf{G}}_k$, $k=1,2$, is maximized. The optimal $\mathbf{R}_k$ is taken as the upper triangular part of $\widetilde{\mathbf{G}}_k$, $k=1,2$. The optimal $\mathbf{Q}_a$ and $\mathbf{Q}_b$ of the GSD solution consist of the first R columns of $\widetilde{\mathbf{Q}}_a$ and $\widetilde{\mathbf{Q}}_b$, respectively.

First, we prove (Di) by contradiction. Suppose all eigenvalues of $\mathbf{R}_2\mathbf{R}_1^{-1}$ are distinct, which implies rank($\mathcal{R}$) $= R$ by Lemma 2.1 (i). As in the proof of Theorem 3.3, optimality conditions (5.1)–(5.2) then imply that $\widetilde{\mathbf{G}}_k$ are upper triangular, $k=1,2$. We write $\widetilde{\mathbf{G}}_k=\mathbf{R}_k$, $k=1,2$. The optimality conditions (5.3)–(5.4), together with the upper triangularity of $\mathbf{R}_1$ and $\mathbf{R}_2$, yield that:

- All rows of $[\widetilde{\mathbf{H}}_1\,|\,\widetilde{\mathbf{H}}_2]$ are orthogonal to all rows of $[\mathbf{R}_1\,|\,\mathbf{R}_2]$.

- All columns of $\begin{bmatrix} \widetilde{\mathbf{L}}_1 \\ \widetilde{\mathbf{L}}_2 \end{bmatrix}$ are orthogonal to all columns of $\begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{bmatrix}$.

Note that these are also the conditions for a stationary point of problem (5.6). We have

$$\widetilde{\mathbf{Q}}_a^T \, [\mathbf{Z}_1\mathbf{Q}_b |\, \mathbf{Z}_2\mathbf{Q}_b] = \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \widetilde{\mathbf{H}}_1 & \widetilde{\mathbf{H}}_2 \end{bmatrix} . \tag{D.2}$$

Let the SVD of $[\mathbf{Z}_1\mathbf{Q}_b |\, \mathbf{Z}_2\mathbf{Q}_b]$ $(I \times 2R)$ be given by $\mathbf{V}_1\, \mathbf{S}_1\, \mathbf{W}_1^T$, with $\mathbf{V}_1^T\mathbf{V}_1 = \mathbf{I}_I$ and $\mathbf{W}_1^T\mathbf{W}_1 = \mathbf{I}_{2R}$. Postmultiplying (D.2) by its transpose yields

$$(\mathbf{V}_1^T\widetilde{\mathbf{Q}}_a)^T\, \mathbf{S}_1\mathbf{S}_1^T\, (\mathbf{V}_1^T\widetilde{\mathbf{Q}}_a) = \begin{bmatrix} \mathbf{P}_1 & \mathbf{O}_{R,I-R} \\ \mathbf{O}_{I-R,R} & \mathbf{P}_2 \end{bmatrix} , \tag{D.3}$$

where $\mathbf{P}_1$ is $R \times R$ and symmetric and nonsingular (since $[\mathbf{R}_1\ \mathbf{R}_2]$ has full row rank due to nonsingularity of $\mathbf{R}_k$), and $\mathbf{P}_2$ is $(I-R) \times (I-R)$ and symmetric. Matrix $\mathbf{S}_1$ is $I \times 2R$ and contains the $\min(I, 2R)$ nonzero singular values on its diagonal. Note that all singular values are nonzero and distinct since $[\mathbf{Z}_1 \,|\, \mathbf{Z}_2]$ is generic and $\mathbf{Q}_b$ has full column rank R.

Suppose $I \leq 2R$. Then $\mathbf{S}_1\mathbf{S}_1^T$ is $I \times I$ diagonal and nonsingular. Since the diagonal entries of $\mathbf{S}_1\mathbf{S}_1^T$ are positive and distinct, and $\mathbf{V}_1^T\widetilde{\mathbf{Q}}_a$ is orthonormal, equation (D.3) is the eigendecomposition of an $I \times I$ symmetric matrix that is nonsingular. In fact, (D.3) can be obtained as the superposition of the eigendecompositions of $\mathbf{P}_1$ and $\mathbf{P}_2$ (both nonsingular). This implies that

$$\mathbf{V}_1^T\widetilde{\mathbf{Q}}_a = \begin{bmatrix} \widetilde{\mathbf{Q}}_a^{(1)} & \mathbf{O}_{R,I-R} \\ \mathbf{O}_{I-R,R} & \widetilde{\mathbf{Q}}_a^{(2)} \end{bmatrix} , \tag{D.4}$$

where $\widetilde{\mathbf{Q}}_a^{(1)}$ $(R \times R)$ and $\widetilde{\mathbf{Q}}_a^{(2)}$ $((I-R) \times (I-R))$ are orthonormal. Hence, $\widetilde{\mathbf{Q}}_a$ is such that, for given $\widetilde{\mathbf{Q}}_b$, the Frobenius norm of the first R rows of $\widetilde{\mathbf{Q}}_a^T\, [\mathbf{Z}_1\mathbf{Q}_b |\, \mathbf{Z}_2\mathbf{Q}_b]$ is maximal.

As in the proof of Theorem 5.1 (see (C.6)), when $I > 2R$ it also follows that $\mathbf{V}_1^T\widetilde{\mathbf{Q}}_a$ is of the form (D.4).

Next, we consider $\widetilde{\mathbf{Q}}_b$ for given $\widetilde{\mathbf{Q}}_a$. We have

$$\begin{bmatrix} \mathbf{Q}_a^T\mathbf{Z}_1 \\ \mathbf{Q}_a^T\mathbf{Z}_2 \end{bmatrix} \widetilde{\mathbf{Q}}_b = \begin{bmatrix} \mathbf{R}_1 & \widetilde{\mathbf{L}}_1 \\ \mathbf{R}_2 & \widetilde{\mathbf{L}}_2 \end{bmatrix} . \tag{D.5}$$

Let the SVD of $\begin{bmatrix} \mathbf{Q}_a^T\mathbf{Z}_1 \\ \mathbf{Q}_a^T\mathbf{Z}_2 \end{bmatrix}$ $(2R \times J)$ be given by $\mathbf{W}_2\, \mathbf{S}_2\, \mathbf{V}_2^T$, with $\mathbf{V}_2^T\mathbf{V}_2 = \mathbf{I}_J$ and $\mathbf{W}_2^T\mathbf{W}_2 = \mathbf{I}_{2R}$. Premultiplying (D.5) by its transpose yields

$$(\mathbf{V}_2^T\widetilde{\mathbf{Q}}_b)^T\, \mathbf{S}_2^T\mathbf{S}_2\, (\mathbf{V}_2^T\widetilde{\mathbf{Q}}_b) = \begin{bmatrix} \mathbf{K}_1 & \mathbf{O}_{R,J-R} \\ \mathbf{O}_{J-R,R} & \mathbf{K}_2 \end{bmatrix} , \tag{D.6}$$

where $\mathbf{K}_1$ is $R \times R$ and symmetric and nonsingular, and $\mathbf{K}_2$ is $(J-R) \times (J-R)$ and symmetric. Matrix $\mathbf{S}_2$ is $2R \times J$ and contains the $\min(J, 2R)$ nonzero singular values on its diagonal.

Suppose $J \leq 2R$. Then $\mathbf{S}_2^T \mathbf{S}_2$ is $J \times J$ diagonal and nonsingular. Since the diagonal entries of $\mathbf{S}_2^T \mathbf{S}_2$ are positive and distinct, and $\mathbf{V}_2^T \widetilde{\mathbf{Q}}_b$ is orthonormal, equation (D.6) is the eigendecomposition of a $J \times J$ symmetric matrix that is nonsingular. In fact, (D.6) can be obtained as the superposition of the eigendecompositions of $\mathbf{K}_1$ and $\mathbf{K}_2$ (both nonsingular). This implies that

$$\mathbf{V}_2^T \widetilde{\mathbf{Q}}_b = \begin{bmatrix} \widetilde{\mathbf{Q}}_b^{(1)} & \mathbf{O}_{R,J-R} \\ \mathbf{O}_{J-R,R} & \widetilde{\mathbf{Q}}_b^{(2)} \end{bmatrix}, \tag{D.7}$$

where $\widetilde{\mathbf{Q}}_b^{(1)}$ $(R \times R)$ and $\widetilde{\mathbf{Q}}_b^{(2)}$ $((J-R) \times (J-R))$ are orthonormal. Hence, $\widetilde{\mathbf{Q}}_b$ is such that, for given $\widetilde{\mathbf{Q}}_a$, the Frobenius norm of the first R columns of $\begin{bmatrix} \mathbf{Q}_a^T \mathbf{Z}_1 \\ \mathbf{Q}_a^T \mathbf{Z}_2 \end{bmatrix} \widetilde{\mathbf{Q}}_b$ is maximal. As above, when $J > 2R$ it also follows that $\mathbf{V}_2^T \widetilde{\mathbf{Q}}_b$ is of the form (D.7).

From the above (also see the first-order optimality conditions of problem (5.6) in section 5.3), it follows that $\widetilde{\mathbf{Q}}_a$ and $\widetilde{\mathbf{Q}}_b$ are such that the Frobenius norm of $\widetilde{\mathbf{G}}_k = \mathbf{R}_k$, $k = 1, 2$ is maximal in (D.1). Therefore, we have obtained an optimal solution $\mathcal{X}$ of problem (5.6) with slices $\mathbf{X}_k = \mathbf{Q}_a \, \mathbf{R}_k \, \mathbf{Q}_b^T$, $k = 1, 2$, and $\text{rank}(\mathcal{X}) = \text{rank}(\mathcal{R}) = R$. This contradicts the assumption in Theorem 5.3 that all optimal solutions $\mathcal{X}$ of problem (5.6) have rank larger than R. Hence, (Di) must hold.

Finally, we prove (Dii) by contradiction. Suppose $\mathbf{R}_2 \mathbf{R}_1^{-1}$ has R linearly independent eigenvectors. Hence, $\text{rank}(\mathcal{R}) = R$ by Lemma 2.1 (i). As in the proof of Theorem 3.3, the optimality conditions (5.1)–(5.2) then imply that $\widetilde{\mathbf{G}}_k$, $k = 1, 2$, are upper triangular. As in the proof of (Di) above, we obtain an optimal solution $\mathcal{X}$ of problem (5.6) with $\text{rank}(\mathcal{X}) = \text{rank}(\mathcal{R}) = R$, which is a contradiction. Hence, (Dii) must hold. This completes the proof of Theorem 5.3.

References

[1] Acar, E., & Yener, B. (2009). Unsupervised multiway data analysis: a literature survey. *IEEE Transactions on Knowledge and Data Engineering*, **21**, 1–15.

[2] Bini, D., Capovani, M., Romani, F., & Lotti, G. (1979). $O(n^{2.7799})$ complexity for $n \times n$ approximate matrix multiplication. *Information Processing Letters*, **8**, 234–235.

[3] Bini, D., Lotti, G., & Romani, F. (1980). Approximate solutions for the bilinear form computational problem. *SIAM Journal on Computing*, **9**, 692–697.

[4] Carroll, J.D., & Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition. *Psychometrika*, **35**, 283–319.

[5] Comon, P., Luciani, X., & de Almeida, A.L.F. (2009). Tensor decompositions, alternating least squares and other tales. *Journal of Chemometrics*, **23**, 393–405.

[6] Comon, P., & De Lathauwer, L. (2010). Algebraic identification of under-determined mixtures, pp.325–366 in: *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, P. Comon and C. Jutten (Eds.), Academic Press.

[7] De Lathauwer, L., De Moor, B., & Vandewalle, J. (2004). Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition. *SIAM Journal on Matrix Analysis and Applications*, **26**, 295–327.

[8] De Lathauwer, L. (2010). Algebraic methods after prewhitening, pp.155–178 in: *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, P. Comon and C. Jutten (Eds.), Academic Press.

[9] De Silva, V., & Lim, L.-H. (2008). Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, **30**, 1084–1127.

[10] Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, **1**, 211–218.

[11] Golub, G.H., & Van Loan, C.F. (1996). *Matrix Computations*, 3rd edition, Johns Hopkins University Press, Baltimore.

[12] Harshman, R.A. (1970). Foundations of the Parafac procedure: models and conditions for an "explanatory" multimodal factor analysis. *UCLA Working Papers in Phonetics*, **16**, 1–84.

[13] Hillar, C.J., & Lim, L.-H. (2013). Most tensor problems are NP-hard. *Journal of the ACM*, **60**, Article 45.

[14] Hitchcock, F.L. (1927). The expression of a tensor or a polyadic as a sum of products, *Journal of Mathematics and Physics*, **6**, 164–189.

[15] Hitchcock, F.L. (1927). Multiple invariants and generalized rank of a p-way matrix or tensor, *Journal of Mathematics and Physics*, **7**, 39–70.

[16] Hopke, P.K., Paatero, P., Jia, H., Ross, R.T., & Harshman, R.A. (1998). Three-way (Parafac) factor analysis: examination and comparison of alternative computational methods as applied to ill-conditioned data. *Chemometrics and Intelligent Laboratory Systems*, **43**, 25–42.

[17] Ishteva, M., Absil, P.-A., Van Huffel, S., & De Lathauwer, L. (2011). Best low multilinear rank approximation of higher-order tensors, based on the Riemannian trust-region scheme. *SIAM Journal on Matrix Analysis and Applications*, **32**, 115–135.

[18] Ja' Ja', J.J. (1979). Optimal evaluation of pairs of bilinear forms. *SIAM Journal on Computing*, **8**, 443–462.

[19] Ja' Ja', J.J. (1979). An addendum to Kronecker's theory of pencils. *SIAM Journal on Applied Mathematics*, **37**, 700–712.

[20] Kolda, T.G., & Bader, B.W. (2009). Tensor decompositions and applications. *SIAM Review*, **51**, 455–500.

[21] Kressner, D. (2006). Block algorithms for reordering standard and generalized Schur forms. *ACM Transactions on Mathematical Software*, **32**, 521–532.

[22] Krijnen, W.P., Dijkstra, T.K., & Stegeman, A. (2008). On the non-existence of optimal solutions and the occurrence of "degeneracy" in the Candecomp/Parafac model. *Psychometrika*, **73**, 431–439.

[23] Kronecker, L. (1890). Algebraische reduction der schaaren bilinearer formen. *Sitzungberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*, 763–776.

[24] Kroonenberg, P.M. (2008). *Applied Multiway Data Analysis*, Wiley Series in Probability and Statistics.

[25] Kruskal, J.B., Harshman, R.A., & Lundy, M.E. (1989). How 3-MFA data can cause degenerate Parafac solutions, among other relationships, pp. 115–121 in: *Multiway Data Analysis*, R. Coppi and S. Bolasco (Eds.), North-Holland.

[26] Lim, L.-H., & Comon, P. (2009). Nonnegative approximations of nonnegative tensors. *Journal of Chemometrics*, **23**, 432–441.

[27] Lim, L.-H., & Comon, P. (2010). Multiarray signal processing: tensor decomposition meets compressed sensing. *Comptes-Rendus de l'Académie des Sciences, Mécanique*, **338**, 311–320.

[28] Paatero, P. (2000). Construction and analysis of degenerate Parafac models. *Journal of Chemometrics*, **14**, 285–299.

[29] Pervouchine, D.D. (2004). Hierarchy of closures of matrix pencils. *Journal of Lie Theory*, **14**, 443–479.

[30] Rocci, R., & Giordani, P. (2010). A weak degeneracy revealing decomposition for the Candecomp/Parafac model. *Journal of Chemometrics*, **24**, 57–66.

[31] Savas, B., & Lim, L.-H. (2010). Quasi-Newton methods on Grassmannians and multilinear approximation of tensors. *SIAM Journal on Scientific Computing*, **32**, 3352–3393.

[32] Smilde, A., Bro, R., & Geladi, P. (2004). *Multi-way Analysis: Applications in the Chemical Sciences*. Chichester: Wiley.

[33] Stegeman, A. (2006). Degeneracy in Candecomp/Parafac explained for $p \times p \times 2$ arrays of rank $p+1$ or higher. *Psychometrika*, **71**, 483–501.

[34] Stegeman, A. (2007). Degeneracy in Candecomp/Parafac explained for several three-sliced arrays with a two-valued typical rank. *Psychometrika*, **72**, 601–619.

[35] Stegeman, A. (2008). Low-rank approximation of generic $p \times q \times 2$ arrays and diverging components in the Candecomp/Parafac model. *SIAM Journal on Matrix Analysis and Applications*, **30**, 988–1007.

[36] Stegeman, A. (2009). Using the Simultaneous Generalized Schur Decomposition as a Candecomp/Parafac algorithm for ill-conditioned data. *Journal of Chemometrics*, **23**, 385–392.

[37] Stegeman, A. (2010). The Generalized Schur Decomposition and the rank-R set of real $I \times J \times 2$ arrays. Technical Report, available online as `arXiv:1011.3432`

[38] Stegeman, A. (2012). Candecomp/Parafac - from diverging components to a decomposition in block terms. *SIAM Journal on Matrix Analysis and Applications*, **33**, 291–316.

[39] Stegeman, A. (2013). A three-way Jordan canonical form as limit of low-rank tensor approximations. *SIAM Journal on Matrix Analysis and Applications*, **34**, 624–650.

[40] Stegeman, A. (2014). Finding the limit of diverging components in three-way Candecomp/Parafac - a demonstration of its practical merits. *Computational Statistics and Data Analysis*, **75**, 203-216.

[41] Stegeman, A., & De Lathauwer, L. (2009). A method to avoid diverging components in the Candecomp/Parafac model for generic $I \times J \times 2$ arrays. *SIAM Journal on Matrix Analysis and Applications*, **30**, 1614–1638.

[42] Stegeman, A., & De Lathauwer, L. (2011). Are diverging CP components always nearly proportional? Technical Report, available online as `arXiv:1110.1988`

[43] Ten Berge, J.M.F., & Kiers, H.A.L. (1999). Simplicity of core arrays in three-way principal component analysis and the typical rank of $p \times q \times 2$ arrays. *Linear Algebra and its Applications*, **294**, 169–179.

[44] Tomasi, G., & Bro, R. (2006). A Comparison of algorithms for fitting the Parafac model. *Computational Statistics & Data Analysis*, **50**, 1700-1734.

[45] Weierstrass, K. (1867). Zur theorie der bilinearen und quadratischen formen. *Monatsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*, 310–338.