DECOMPOSITION OF A TENSOR INTO MULTILINEAR RANK-$(M_r, N_r, \cdot)$ TERMS*

IGNAT DOMANOV†, NICO VERVLIET‡, AND LIEVEN DE LATHAUWER†‡

Key words. multilinear algebra, third-order tensor, block term decomposition, multilinear rank

AMS subject classifications. 15A23, 15A69

1. Introduction.

1.1. Terminology and problem statement. Throughout the paper $\mathbb{F}$ denotes the field of real or complex numbers; bold-face lower case, bold-face capital case, and capital calligraphic letters refer to vectors, matrices, and tensors, respectively (e.g., $\mathbf{a}_1$, $\mathbf{B}_2$, and $\mathcal{T}$); $\mathbf{O}$ and $\mathbf{0}$ denote a zero matrix and a zero vector, respectively; $\mathbf{I}_n$ and $\mathbf{O}_{m\times n}$ denote the $n \times n$ identity matrix and the $m \times n$ zero matrix, respectively; $\mathbf{O}_n$ a is shorthand for $\mathbf{O}_{n\times n}$; $\mathrm{Null}\,(\cdot)$ denotes the null space of a matrix; "$*$" , "$^{\mathrm{T}}$", and "$^{\mathrm{H}}$" denote the conjugation, transpose, and hermitian transpose, respectively.

Let tensor $\mathcal{T} \in \mathbb{F}^{I\times J\times K}$ and let $\mathbf{T}_1, \ldots \mathbf{T}_K \in \mathbb{F}^{I\times J}$ denote the frontal slices of $\mathcal{T}$. The matrices

$$\mathbf{T}_{(1)} = [\mathbf{T}_1 \ \ldots \ \mathbf{T}_K] \in \mathbb{F}^{I\times JK},$$
$$\mathbf{T}_{(2)} = [\mathbf{T}_1^{\mathrm{T}} \ \ldots \ \mathbf{T}_K^{\mathrm{T}}] \in \mathbb{F}^{J\times IK},$$
$$\mathbf{T}_{(3)} = [\mathrm{vec}(\mathbf{T}_1) \ \ldots \mathrm{vec}(\mathbf{T}_K)]^{\mathrm{T}} \in \mathbb{F}^{K\times IJ},$$

are called the *matrix unfoldings* of $\mathcal{T}$ and the triplet of their ranks $(r_{\mathbf{T}_{(1)}}, r_{\mathbf{T}_{(2)}}, r_{\mathbf{T}_{(3)}})$ is called the *multiLinear (ML) rank* of $\mathcal{T}$.

Tensor $\mathcal{T}$ is ML rank-$(M, N, \cdot)$ if $r_{\mathbf{T}_{(1)}} = M$, $r_{\mathbf{T}_{(2)}} = N$ and $r_{\mathbf{T}_{(3)}}$ is not specified. It can easily be shown that $\mathcal{T}$ is ML rank-$(M, N, \cdot)$ if and only if its frontal slices can be simultaneously factorized as

$$\mathbf{T}_k = \mathbf{A}\mathbf{D}_k\mathbf{B}^{\mathrm{T}}, \qquad \mathbf{D}_k \in \mathbb{F}^{M\times N}, \qquad k = 1, \ldots, K, \tag{1.1}$$

where the factors $\mathbf{A} \in \mathbb{F}^{I\times M}$ and $\mathbf{B} \in \mathbb{F}^{J\times N}$ have full column rank and the matrices

$$\mathbf{D}_{(1)} := [\mathbf{D}_1 \ \ldots \ \mathbf{D}_K] \in \mathbb{F}^{M\times NK} \ \text{ and } \ \mathbf{D}_{(2)} := [\mathbf{D}_1^{\mathrm{T}} \ \ldots \ \mathbf{D}_K^{\mathrm{T}}] \in \mathbb{F}^{N\times MK}$$

have full row rank. (One can, for instance, take $\mathbf{A}$ and $\mathbf{B}$ equal to the "U" factors in the compact SVDs of $\mathbf{T}_{(1)}$ and $\mathbf{T}_{(2)}$, respectively, and set $\mathbf{D}_k = \mathbf{A}^{\mathrm{H}}\mathbf{T}_k\mathbf{B}^*$.) It is worth noting that the factors $\mathbf{A}$, $\mathbf{D}_k$, and $\mathbf{B}$ in (1.1) are not unique. Moreover, a generic

*Submitted to the editors DATE

Funding: This work was supported by the Flemish Government: (1) This research received funding from the Flemish Government (AI Research Program). Lieven De Lathauwer and Nico Vervliet are affiliated to Leuven.AI—KU Leuven institute for AI, B-3000, Leuven, Belgium. (2) This work was supported by the Fonds de la Recherche Scientifique – FNRS and the Fonds Wetenschappelijk Onderzoek – Vlaanderen under EOS Project no G0F6718N (SeLMA) (3) PhD/Postdoc grants. This work was further supported via KU Leuven Internal Funds: C16/15/059, IDN/19/014. N. Vervliet holds a research grant from the Research Foundation Flanders (FWO) (12ZM220N).

†Group Science, Engineering and Technology, KU Leuven - Kulak, E. Sabbelaan 53, 8500 Kortrijk, Belgium and (Ignat.Domanov@gmail.com, Lieven.DeLathauwer@kuleuven.be).

‡Dept. of Electrical Engineering ESAT/STADIUS, KU Leuven, Kasteelpark Arenberg 10, bus 2446, B-3001 Leuven, Belgium, Nico.Vervliet@kuleuven.be

This manuscript is for review purposes only.

ML rank-$(M, N, \cdot)$ tensor can parameterized with $(I - M)M + (J - N)N + MNK$ parameters. Indeed, if $\mathbf{X} \in \mathbb{F}^{I \times I}$ and $\mathbf{Y} \in \mathbb{F}^{J \times J}$ are nonsingular matrices, then $\mathbf{T}_1, \ldots, \mathbf{T}_K$ can also be factorized as

$$\mathbf{T}_k = (\mathbf{A}\mathbf{X}) \left(\mathbf{X}^{-1}\mathbf{D}_k\mathbf{Y}^{-\mathrm{T}}\right) (\mathbf{B}\mathbf{Y})^{\mathrm{T}} = \widehat{\mathbf{A}}\widehat{\mathbf{D}}_k\widehat{\mathbf{B}}^{\mathrm{T}}, \qquad k = 1, \ldots, K. \tag{1.2}$$

Assuming that $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{D}_k$ in (1.2) are generic we can set $\mathbf{X}$ (resp. $\mathbf{Y}$) equal to the inverse of the matrix formed by the first M (resp. N) rows of $\mathbf{A}$ (resp. $\mathbf{B}$), implying that the right-hand side of the K equations

$$\mathbf{T}_k = \widehat{\mathbf{A}}\widehat{\mathbf{D}}_k\widehat{\mathbf{B}}^{\mathrm{T}} = \begin{bmatrix} \mathbf{I}_M \\ * \end{bmatrix} [*] \begin{bmatrix} \mathbf{I}_N \\ * \end{bmatrix}^{\mathrm{T}}, \qquad k = 1, \ldots, K.$$

can be parameterized with $(I - M)M + (J - N)N + MN \cdot K$ parameters.

We will write the K factorizations in (1.1) as

$$\mathcal{T} = \mathcal{D} \bullet_1 \mathbf{A} \bullet_2 \mathbf{B},$$

where $\mathcal{D}$ denotes an $M \times N \times K$ tensor with frontal slices $\mathbf{D}_1, \ldots, \mathbf{D}_K$, and $\bullet_n$ denotes the mode-n tensor-matrix product.

In this paper we study the decomposition of $\mathcal{T}$ into a sum of ML rank-$(M_r, N_r, \cdot)$ tensors which is illustrated in Figure 1.1 and defined as:

$$\mathcal{T} = \sum_{r=1}^{R} \mathcal{D}_r \bullet_1 \mathbf{A}_r \bullet_2 \mathbf{B}_r, \qquad \mathcal{D}_r \in \mathbb{F}^{M_r \times N_r \times K}, \quad \mathbf{A}_r \in \mathbb{F}^{I \times M_r}, \quad \mathbf{B}_r \in \mathbb{F}^{J \times N_r}. \tag{1.3}$$

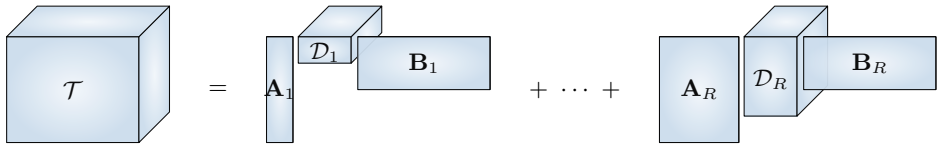


FIG. 1.1. *ML rank-$(M_r, N_r, \cdot)$ decomposition of a third-order tensor*

DEFINITION 1.1. *ML rank-$(M_r, N_r, \cdot)$ decomposition of tensor $\mathcal{T}$ is called* unique *if for any two decompositions of the form (1.3) one can be obtained from another by a permutation of summands.*

DEFINITION 1.2. *ML rank-$(M_r, N_r, \cdot)$ decomposition is called* generically unique *if*

$$\mu\{(\mathbf{A}_1, \ldots, \mathbf{A}_R, \mathbf{B}_1, \ldots, \mathbf{B}_R, \mathcal{D}_1, \ldots, \mathcal{D}_R) : \text{the decomposition of } \mathcal{T} = \sum_{r=1}^{R} \mathcal{D}_r \bullet_1 \mathbf{A}_r \bullet_2 \mathbf{B}_r \text{ is not unique}\} = 0,$$

where μ denotes a measure on $\mathbb{F}^{I\sum M_r + J\sum N_r + K\sum M_r N_r}$ absolutely continuous with respect to the Lebesgue measure.

We say that $\mathcal{T}$ is a sum of R *generic* ML rank-$(M_r, N_r, \cdot)$ terms if the entries of $\mathbf{A}_r$, $\mathbf{B}_r$, and $\mathcal{D}_r$ in (1.3) are randomly sampled from an absolutely continuous distribution. Thus, the generic uniqueness means uniqueness that holds with probability one.

Let

$$\mathbf{A} = [\mathbf{A}_1 \ \dots \ \mathbf{A}_R] \in \mathbb{F}^{I \times \sum M_r} \quad \text{and} \quad \mathbf{B} = [\mathbf{B}_1 \ \dots \ \mathbf{B}_R] \in \mathbb{F}^{J \times \sum N_r}. \tag{1.4}$$

Throughout the paper we assume that the matrices

$$\mathbf{A} \quad \text{and} \quad \mathbf{B} \quad \text{have full column rank.} \tag{1.5}$$

1.2. Previous results. Let $\mathbf{D}_{r,k} \in \mathbb{F}^{M_r \times N_r}$ denote the k-th frontal slice of the tensor $\mathcal{D}_r$ and let

$$\mathbf{D}_k = \operatorname{blockdiag}(\mathbf{D}_{1,k}, \dots, \mathbf{D}_{R,k}) \in \mathbb{F}^{\sum M_r \times \sum N_r}, \qquad k = 1, \dots, K \tag{1.6}$$

denote a block-diagonal matrix with the matrices $\mathbf{D}_{1,k}, \dots, \mathbf{D}_{R,k}$ on the diagonal. By (1.1), the decomposition (1.3) can be interpreted as the problem of (possibly nonsymmetric and nonsquare) *joint block diagonalization* (JDB) of the frontal slices of $\mathcal{T}$:

$$\mathbf{T}_k = \mathbf{A}\mathbf{D}_k\mathbf{B}^{\mathrm{T}} = \mathbf{A} \operatorname{blockdiag}(\mathbf{D}_{1,k}, \dots, \mathbf{D}_{R,k})\mathbf{B}^{\mathrm{T}}, \qquad k = 1, \dots, K, \tag{1.7}$$

where $\mathbf{A}$ and $\mathbf{B}$ are defined in (1.4).

Problem (1.7) (both uniqueness and algorithms) has been relatively well studied in the following cases:

1. JBD by congruence and $*$-congruence, i.e., the cases $\mathbf{B} = \mathbf{A}$ and $\mathbf{B} = \mathbf{A}^*$, implying the constraint $M_r = N_r$ for all r; see, for instance, [1, 2, 12] and the references therein.
2. All blocks are row vectors, i.e., $M_1 = \cdots = M_R = 1$, implying that the ML rank-$(M_r, N_r, \cdot)$ terms in (1.3) are actually ML rank-$(1, N_r, N_r)$; see, for instance, [3, 4, 16, 13].
3. All blocks are column vectors, i.e., $N_1 = \cdots = N_R = 1$ which follows from 2. by taking transpose of the identities in (1.7).

Apart from the items 1.–3. above the generic uniqueness of **??** was proved in [3] under the assumptions $M_r = N_r$ for all r and $K \geq 3$.

2. Uniqueness result and algorithm. We obtain results on uniqueness and present algorithms for the computation of ML rank-$(M_r, N_r, \cdot)$ tensors in (1.3) in the case where $M_r \neq N_r$ for some r.

2.1. Exact decomposition.

THEOREM 2.1. *Let $\mathcal{T}$ admit the ML rank-$(M_r, N_r, \cdot)$ decomposition (1.3), the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{D}_1, \dots, \mathbf{D}_K$ be defined in (1.4) and (1.6), respectively, and*

$$\mathbf{M}_{\mathcal{D}} = \begin{bmatrix} \mathbf{D}_1^T \otimes \mathbf{I}_I & -\mathbf{I}_J \otimes \mathbf{D}_1 \\ \vdots & \vdots \\ \mathbf{D}_K^T \otimes \mathbf{I}_I & -\mathbf{I}_J \otimes \mathbf{D}_K \end{bmatrix}. \tag{2.1}$$

Assume that

1. $\mathbf{A}$ *and* $\mathbf{B}$ *have full column rank and*
2. $\dim \operatorname{Null}(\mathbf{M}_{\mathcal{D}}) = R$.

Then the ML rank-$(M_r, N_r, \cdot)$ decomposition of $\mathcal{T}$ is unique and can be computed by means of coupled simultaneous eigenvalue decomposition (CS-EVD).

Proof. See section 4. ∎

Assumption 2) in Theorem 2.1 expresses the fact that the system of matrix equations

$$\mathbf{X}\mathbf{D}_k = \mathbf{D}_k\mathbf{Y}, \qquad k = 1, \dots, K \tag{2.2}$$

has exactly R linearly independent solutions $(\mathbf{X}, \mathbf{Y}) \in \mathbb{F}^{\sum M_r \times \sum M_r} \times \mathbb{F}^{\sum N_r \times \sum N_r}$. If

$$\begin{aligned} \mathbf{P}_{I,r} &= \text{blockdiag}(\mathbf{O}_{M_1}, \dots, \mathbf{O}_{M_{r-1}}, \mathbf{I}_{M_r}, \mathbf{O}_{M_{r+1}}, \dots, \mathbf{O}_{M_R}) \quad \text{and} \\ \mathbf{P}_{J,r} &= \text{blockdiag}(\mathbf{O}_{N_1}, \dots, \mathbf{O}_{N_{r-1}}, \mathbf{I}_{N_r}, \mathbf{O}_{N_{r+1}}, \dots, \mathbf{O}_{N_R}), \qquad r = 1, \dots, R, \end{aligned} \tag{2.3}$$

then $(\mathbf{X}, \mathbf{Y}) = (\mathbf{P}_{I,r}, \mathbf{P}_{J,r})$ is, obviously, a solution of (2.2), as

$$\begin{aligned} &\mathbf{P}_{I,r}\mathbf{D}_k = \mathbf{D}_k\mathbf{P}_{J,r} = \\ &\qquad \text{blockdiag}(\mathbf{O}_{M_1 \times N_1}, \dots, \mathbf{O}_{M_{r-1} \times N_{r-1}}, \mathbf{D}_{r,k}, \mathbf{O}_{M_{r+1} \times N_{r+1}}, \dots, \mathbf{O}_{M_R \times N_R}). \end{aligned}$$

Thus, assumption 2) in Theorem 2.1 means that the dimension of the subspace of solutions of (2.2) is minimal.

The algebraic procedure related to Theorem 2.1 is summarized in Algorithm 2.1. By assumption 1) in Theorem 2.1, the number of columns of $\mathbf{A}$ and $\mathbf{B}$ cannot exceed the number of rows, i.e., $\sum M_r \leq I$ and $\sum N_r \leq J$. To simplify the presentation we assume in Algorithm 2.1 that $\sum M_r = I$ and $\sum N_r = J$, implying that $\mathbf{A}$ and $\mathbf{B}$ are square nonsingular matrices and $\mathcal{T} \in \mathbb{F}^{\sum M_r \times \sum N_r \times K}$. Since, by Lemma 4.1 below, $\mathcal{T}$ has ML rank-$(\sum M_r, \sum N_r, \cdot)$, this compression step can always be performed without loss of generality, e.g., via the multilinear singular value decomposition [5]. Note that Algorithm 2.1 also recovers the dimensions of the blocks, i.e., the values M_r and N_r. Indeed, in steps 1–4 of Algorithm 2.1 we construct matrices $\mathbf{U}_{I,1}, \dots, \mathbf{U}_{I,R} \in \mathbb{F}^{I \times I}$ and $\mathbf{U}_{J,1}, \dots, \mathbf{U}_{J,R} \in \mathbb{F}^{J \times J}$ that can be simultaneously diagonalized by $\mathbf{A}$ and $\mathbf{B}^{-\mathrm{T}}$, respectively. In step 5 of Algorithm 2.1 we use the fact that for each r the spectra of $\mathbf{U}_{I,r}$ and $\mathbf{U}_{J,r}$ relate to each other as in (2.5)–(2.6). Thus, if the values $\alpha_1, \dots, \alpha_R$ are generic, then the distinct eigenvalues of the matrices $\mathbf{F} = \alpha_1\mathbf{U}_{I,1} + \cdots + \alpha_R\mathbf{U}_{I,R}$ and $\mathbf{G} = \alpha_1\mathbf{U}_{J,1} + \cdots + \alpha_R\mathbf{U}_{J,R}$ are the same and their multiplicities are $M_1, \dots, M_R$ and $N_1, \dots, N_R$ respectively. Hence, the values M_r, N_r and the corresponding blocks $\mathbf{A}_r$, $\mathbf{B}_r$ can be recovered from the EVDs of $\mathbf{F}$ and $\mathbf{G}$.

The following corollary completes the results on uniqueness and computation of the ML rank-$(1, N_r, N_r)$ decomposition in [3, 4, 16].

COROLLARY 2.2. *Let $\mathcal{T}$ admit the ML rank-$(1, N_r, N_r)$ decomposition (1.3), the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{D}_1, \dots, \mathbf{D}_K$ be defined as in (1.4) and (1.6), respectively, and let*

$$\mathbf{C}_r := [\mathbf{D}_{r,1}^T \ \dots \ \mathbf{D}_{r,K}^T]^T \in \mathbb{F}^{K \times N_r}, \qquad r = 1, \dots, R. \tag{2.7}$$

Assume that

1. *$\mathbf{A}$ and $\mathbf{B}$ have full column rank and*
2. *$r_{[\mathbf{C}_i \ \mathbf{C}_j]} \geq \max(N_i, N_j) + 1$ for all $1 \leq i < j \leq R$.*

Then the ML rank-$(1, N_r, N_r)$ decomposition of $\mathcal{T}$ is unique and can be computed by means of CS-EVD.

Proof. See section 4. ∎

The following theorem contains bounds on the tensor dimensions that guarantee that assumptions 1) and 2) in Theorem 2.1 hold for generic matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{D}_1, \dots, \mathbf{D}_K$.

Algorithm 2.1 Computation of BTD (1.3) by CS-EVD (see Theorem 2.1)

Require: tensor $\mathcal{T} \in \mathbb{F}^{I \times J \times K}$ such that $I = \sum M_r$ and $J = \sum N_r$

1: Construct the IJK-by-$(I^2 + J^2)$ matrix $\mathbf{M}_{\mathcal{T}}$ as

$$\mathbf{M}_{\mathcal{T}} = \begin{bmatrix} \mathbf{T}_1^{\mathrm{T}} \otimes \mathbf{I}_I & -\mathbf{I}_J \otimes \mathbf{T}_1 \\ \vdots & \vdots \\ \mathbf{T}_K^{\mathrm{T}} \otimes \mathbf{I}_I & -\mathbf{I}_J \otimes \mathbf{T}_K \end{bmatrix} \tag{2.4}$$

2: Compute $\mathbf{u}_r = [u_{1,r} \ \ldots \ u_{I^2+J^2,r}]^{\mathrm{T}}$, $r = 1, \ldots, R$ that form a basis of Null $(\mathbf{M}_{\mathcal{T}})$
3: For $r = 1, \ldots, R$ reshape $[u_{1,r} \ \ldots \ u_{I^2,r}]^{\mathrm{T}}$ into the $I \times I$ matrix $\mathbf{U}_{I,r}$
4: For $r = 1, \ldots, R$ reshape $[u_{I^2+1,r} \ \ldots \ u_{I^2+J^2,r}]^{\mathrm{T}}$ into the $J \times J$ matrix $\mathbf{U}_{J,r}$
5: Solve CS-EVD problem

$$\mathbf{U}_{I,r} = \mathbf{A} \, \mathrm{blockdiag}(g_{1r}\mathbf{I}_{M_1}, \ldots, g_{Rr}\mathbf{I}_{M_R})\mathbf{A}^{-1}, \tag{2.5}$$

$$\mathbf{U}_{J,r} = \mathbf{B}^{-\mathrm{T}} \, \mathrm{blockdiag}(g_{1r}\mathbf{I}_{N_1}, \ldots, g_{Rr}\mathbf{I}_{N_R})\mathbf{B}^{T}, \quad r = 1, \ldots, R \tag{2.6}$$

with respect to the matrices $\mathbf{A}$, $\mathbf{B}$ and the values $M_1, \ldots, M_R$, $N_1, \ldots, N_R$
6: Solve (1.7) with respect to $\mathbf{D}_1, \ldots, \mathbf{D}_K$
7: For $r = 1, \ldots, R$ stack $\mathbf{D}_{r,1}, \ldots, \mathbf{D}_{r,K}$ into the $M_r \times N_r \times K$ tensor $\mathcal{D}_r$
8: **return** $\mathbf{A}_r \in \mathbb{F}^{I \times M_r}$, $\mathbf{B}_r \in \mathbb{F}^{J \times N_r}$ and $\mathcal{D}_r \in \mathbb{F}^{M_r \times N_r \times K}$ such that (1.3) holds

THEOREM 2.3. *Let $\mathcal{T}$ be a sum of R generic ML rank-$(M_r, N_r, \cdot)$ terms. Assume that*

$$I \geq \sum M_r, \qquad J \geq \sum N_r, \qquad K \geq \max\left(\max\left\lceil \frac{M_r}{N_r} \right\rceil + \max\left\lceil \frac{N_r}{M_r} \right\rceil, 3\right), \tag{2.8}$$

where $\lceil x \rceil$ denotes the smallest integer not less than x. Then the decomposition of $\mathcal{T}$ into a sum of ML rank-$(M_r, N_r, \cdot)$ terms is unique and can be computed by means of CS-EVD.

Proof. See section 5. ∎

The following corollaries easily follow from Theorem 2.3.

COROLLARY 2.4. *Let $\mathcal{T}$ be a sum of $R_1 \geq 1$ generic ML rank-$(1, N_r, N_r)$ terms and $R_2 \geq 1$ generic ML rank-$(M_r, 1, M_r)$ terms. Assume that*

$$I \geq R_1 + \sum M_r, \qquad J \geq R_2 + \sum N_r, \qquad \text{and} \qquad K \geq \max M_r + \max N_r.$$

Then the decomposition of $\mathcal{T}$ into a sum of R_1 ML rank-$(1, N_r, N_r)$ terms and R_2 ML rank-$(M_r, 1, M_r)$ terms is unique and can be computed by means of CS-EVD.

COROLLARY 2.5. *Let $\mathcal{T}$ be a sum of R generic ML rank-$(1, N_r, N_r)$ terms. Assume that*

$$I \geq R, \qquad J \geq \sum N_r, \qquad \text{and} \qquad K \geq \max N_r + 1. \tag{2.9}$$

Then the decomposition of $\mathcal{T}$ into a sum of ML rank-$(1, N_r, N_r)$ terms is unique and can be computed by means of CS-EVD.

COROLLARY 2.6. *Let $\mathcal{T}$ be a sum of R generic ML rank-$(M, N, \cdot)$ terms. Assume that*

$$I \geq RM, \qquad J \geq RN, \qquad K \geq \max\left(\left\lceil \frac{M}{N} \right\rceil + 1, \left\lceil \frac{N}{M} \right\rceil + 1, 3\right).$$

Then the decomposition of $\mathcal{T}$ into a sum of ML rank-$(M, N, \cdot)$ terms is unique and can be computed by means of CS-EVD.

The bounds in Corollary 2.4, Corollary 2.5, and Corollary 2.6 are new and guarantee that a generic ML rank $(M_r, N_r, \cdot)$ decomposition can be computed by Algorithm 2.1. The bound on K in (2.9) complements the bounds for generic uniqueness of the ML rank-$(1, N, N)$ decomposition in [3, Section 4] and is more relaxed than the bound for generic uniqueness of ML rank-$(1, N_r, N_r)$ decomposition in [13, Theorem 1.7 (E)].

In the remaining part of this section, we apply Theorem 2.1 to (generic) constrained JBD problems in Table 2.1. Uniqueness of the constrained JBD problem

TABLE 2.1
Some examples of constrained JBD problems

Type of JBD problem	Constraints in (1.7)	$\mathbb{F}$
JBD by congruence	$\mathbf{B} = \mathbf{A}$	$\mathbb{R}$ or $\mathbb{C}$
symmetric JBD	$\mathbf{B} = \mathbf{A}$ and $\mathbf{D}_k = \mathbf{D}_k^{\mathrm{T}}$, $k = 1, \ldots, K$	$\mathbb{R}$ or $\mathbb{C}$
JBD by $*$-congruence	$\mathbf{B} = \mathbf{A}^*$	$\mathbb{C}$
Hermitian JBD	$\mathbf{B} = \mathbf{A}^*$ and $\mathbf{D}_k = \mathbf{D}_k^{\mathrm{H}}$, $k = 1, \ldots, K$	$\mathbb{C}$

means uniqueness of the corresponding constrained ML rank-$(N_r, N_r, 1)$ decomposition. Generic uniqueness of the constrained JBD problem can be defined in a similar way as in Definition 1.2. By way of example, we resort to the following definition of generic uniqueness of the symmetric JBD problem.

DEFINITION 2.7. *Let L_r denote a linear bijection between $\mathbb{F}^{K\frac{N_r(N_r+1)}{2}}$ and the subspace of $N_r \times N_r \times K$ tensors whose frontal slices are symmetric matrices. Let also μ be a measure on $\mathbb{F}^{I\sum N_r + K\sum \frac{N_r(N_r+1)}{2}}$ absolutely continuous with respect to the Lebesgue measure. A solution of the generic symmetric JBD problem is called unique if*

$$\mu\{(\mathbf{A}_1, \ldots, \mathbf{A}_R, \mathbf{d}_1, \ldots, \mathbf{d}_R) : \text{the constrained decomposition of } \mathcal{T} = \sum_{r=1}^{R} L_r(\mathbf{d}_r) \cdot_1 \mathbf{A}_r \cdot_2 \mathbf{A}_r \text{ is not unique}\} = 0.$$

Theorem 2.3 guarantees that if $K \geq 3$ and the diagonal blocks $\mathbf{D}_{r,k}$ in JBD problem (1.7) are square and generic, then the decomposition is unique and can be computed by means of CS-EVD. The following theorem states that the bound $K \geq 3$ remains valid for all constrained JBD problems in Table 2.1 as well and that the computation relies on S-EVD.

THEOREM 2.8. *Assume that $I \geq \sum N_r$, and $K \geq 3$. Then the solution of a generic constrained JBD problems in Table 2.1 is unique and can be computed by means of S-EVD.*

Steps 1–4 and modified step 5 of Algorithm 2.1 constitute the algebraic procedure related to Theorem 2.8. In step 5, because of the constraints $\mathbf{B} = \mathbf{A}$ and $\mathbf{B} = \mathbf{A}^*$,

the CS-EVD problem (2.5)–(2.6) turns into the S-EVD problem. It can be shown that in the case of the JBD by congruence problem the proposed reduction to S-EVD essentially coincides with the one in [1, Subsection 2.3].

Proof. See section 5. ∎

2.2. Implementation. The mathematical description of Algorithm 1 leaves room for several implementation decisions. We now discuss an algebraic version which does not depend on optimization and estimates all parameters when decomposing a (noisy) tensor $\bar{\mathcal{T}} \in \mathbb{F}^{\bar{I} \times \bar{J} \times \bar{K}}$.

First, Algorithm 1 assumes that $\mathcal{T} \in \mathbb{F}^{I \times J \times K}$ with $I := \sum_r M_r \leq \bar{I}$, $J := \sum_r N_r \leq \bar{J}$, and $K \leq \bar{K}$ with unknown M_r and N_r. As $\mathbf{A}$ and $\mathbf{B}$ are non-singular by (1.5), (I, J, K) is the multilinear rank of $\bar{\mathcal{T}}$ which can be determined via the MLSVD. In the noisy or inexact case, rank estimation strategies can be used. In this paper, we use $I = \bar{I} - R_1$ and $J = \bar{J} - R_2$ in which R_n, $n = 1, 2$, is the largest r such that the variance $\mathrm{Var}\left(\{\sigma_i^{(n)}\}_{i=1}^{r}\right) \leq \theta$ with $\sigma_i^{(n)}$ the ith smallest mode-n singular value. The threshold $\theta = 0.1$ is chosen by the user. The core tensor of the truncated MLSVD is then used as input $\mathcal{T}$ to Algorithm 2.1.

Second, the matrix $\mathbf{M}_{\mathcal{T}}$ is constructed using (2.4) and a basis for $\mathrm{Null}\,(\mathbf{M}_{\mathcal{T}})$ is computed via the SVD by selecting the right singular vectors corresponding to the R smallest singular values which are exactly zero in the noiseless, exact case, and small otherwise. In the latter case, any rank estimation strategy can be used, e.g., the variance-based strategy outlined above. Note that the smallest singular value σ_1 of $\mathrm{Null}\,(\mathbf{M}_{\mathcal{T}})$ is ignored in the computation of the variance as it is always (numerically) zero because of the trivial solution **COMMENT**: @Ignat: what can I best refer to? .

Third, the CS-EVD can be computed using algebraic or optimization-based techniques. For the experiments, an algebraic version using the generalized eigenvalue decomposition (GEVD) is implemented. Stacking $\mathbf{U}_{I,r}$, $r = 1, \ldots, R$, as the frontal slices of $\mathcal{X} \in \mathbb{F}^{I \times I \times R}$ and $\mathbf{U}_{J,r}$, $r = 1, \ldots, R$, as the frontal slices of $\mathcal{Y} \in \mathbb{F}^{J \times J \times R}$, the rank-$I$ and rank-J canonical polyadic decompositions(CPD) of $\mathcal{X} = [\![\mathbf{A}, \mathbf{A}^{-\mathrm{T}}, \mathbf{G}_1]\!]$ and $\mathcal{Y} = [\![\mathbf{B}^{-\mathrm{T}}, \mathbf{B}, \mathbf{G}_2]\!]$, resp., are computed independently using a GEVD-based CPD algorithm [11, 8] to obtain $\mathbf{A}$ and $\mathbf{B}$. Note that the order of the columns in $\mathbf{A}$ and $\mathbf{B}$ is arbitrary and does not necessarily match the partitioning in (1.4). As the algebraic algorithm does not enforce the inverse constraint, and does not take the repeated columns in $\mathbf{G}_1$ and $\mathbf{G}_2$ into account (see (2.5) and (2.6)), $\mathbf{G}_1$ and $\mathbf{G}_2$ are estimated separately from the least squares systems:

$$(\mathbf{A}^{-\mathrm{T}} \odot \mathbf{A})\mathbf{G}_1 = \mathbf{X}_{(3)}^{\mathrm{T}}, \quad \text{and} \quad (\mathbf{B} \odot \mathbf{B})\mathbf{G}_2 = \mathbf{Y}_{(3)}^{\mathrm{T}}. \tag{2.10}$$

From the experiments, we see that the accuracy of the overall decomposition into multilinear rank $(M_r, N_r, \cdot)$ terms can be improved by computing the CPDs more accurately, e.g., using the generalized eigenspace decomposition [9], or by following the GEVD by an optimization-based refinement step, e.g., the Gauss–Newton algorithm in `cpd_nls` [14, 17].

Fourth, to estimate M_r and N_r the columns of $\mathbf{G}_1$ and $\mathbf{G}_2$, resp., are jointly clustered into R groups with M_r columns from $\mathbf{G}_1$ and N_r from $\mathbf{G}_2$. To achieve this, an alternating k-means algorithm is used: k-means is run on the columns of $\mathbf{G}_1$ to obtain R cluster centers which are then used as initialization for k-means applied to the columns of $\mathbf{G}_2$. The new cluster centers are then used for a new k-means run on $\mathbf{G}_1$, and this process is repeated until the cluster centers do not change. To improve

clustering performance, a row in which all entries are identical is removed; this row is the result of the trivial solution. If M_r and N_r are provided by the user, the clustering algorithm is repeated until the desired cluster sizes have been found.

Finally, using the clustering results the columns of $\mathbf{A}$ and $\mathbf{B}$ can be ordered to match the partitioning in (1.4) and the core tensors $\mathcal{D}_r$ are computed by solving

$$\begin{bmatrix} \mathbf{B}_1 \otimes \mathbf{A}_1 & \dots & \mathbf{B}_R \otimes \mathbf{A}_R \end{bmatrix} \begin{bmatrix} \operatorname{vec}(\mathbf{D}_{1,k}) \\ \vdots \\ \operatorname{vec}(\mathbf{D}_{R,k}) \end{bmatrix} = \operatorname{vec}(\mathbf{T}_k), \quad k = 1, \dots, K, \tag{2.11}$$

in least-squares sense. This procedure is more accurate in a noisy setting than bringing $\mathbf{A}$ and $\mathbf{B}$ to the left-hand side in (1.7) as the zero-blocks are forced to be zero.

3. Simulations.

3.1. Recovering a decomposition without prior knowledge. The proposed CS-EVD algorithm can recover a decomposition into multilinear rank $(M_r, N_r, \cdot)$ terms without prior knowledge, as shown using the following small-scale experiment. A tensor $\mathcal{T} \in \mathbb{C}^{9\times 8\times 5}$ which can be written as the sum of $R = 3$ terms with multilinear ranks $(2, 3, \cdot)$, $(4, 2, \cdot)$ and $(3, 3, \cdot)$, is constructed; hence $M = \begin{bmatrix} 2, 4, 3 \end{bmatrix}$ and $N = \begin{bmatrix} 3, 2, 3 \end{bmatrix}$. The matrices $\mathbf{A} \in \mathbb{C}^{9\times 9}$ and $\mathbf{B} \in \mathbb{C}^{8\times 8}$ have entries of which the real and imaginary parts are drawn independently from a normal distribution. If the condition numbers $\operatorname{cond}(\mathbf{A})$ or $\operatorname{cond}(\mathbf{B})$ are larger than 10, a new matrix is generated. The entries of the core tensors $\mathcal{D}_r \in \mathbb{C}^{M_r\times N_r\times 5}$ are similarly drawn from a normal distribution. The tensor $\tilde{\mathcal{T}} = \mathcal{T} + \mathcal{N}$ can be perturbed by a Gaussian independently and identically distributed noise tensor $\mathcal{N}$ which is scaled such that the signal-to-noise ratio (SNR) $||\mathcal{T}|| \,/\, ||\mathcal{N}||$ has a specified value. We use the proposed algebraic version of CS-EVD which uses the GEVD and estimates all parameters, unless stated otherwise; see subsection 2.2. The performance is compared to standard optimization routines: alternating least squares algorithm for the $(M_r, N_r, \cdot)$ decomposition [6, type-2 BTD], denoted by ALS, and the Gauss–Newton implementation `btd_nls` [15, 17] of the general BTD [3], denoted by NLS. For the optimization algorithms, 25 random initializations are generated per experiment; see `btd_rnd` [17]. We generate 100 random tensors $\mathcal{T}$ and test the performance for five noise levels and the noiseless case. Only relevant experiments are shown. All code and data is available online[1].

In contrast to optimization methods, CS-EVD does not require estimates for the parameters R, M_r and N_r in the exact case. In Table 3.1, we see that these parameters can be estimated even under noisy conditions provided the signal-to-noise ratio is sufficiently high, and that if R is estimated correctly, the values M_r and N_r are also correct. For the experiments where the SNR is 20 dB, the dimension of $(\mathbf{M}_{\mathcal{T}})$ is often estimated incorrectly as the singular value gap σ_{R+1}/σ_R is at most 1.76, which is too small for the rank estimation procedure. As this rank estimation step is not the core of the algorithm, we assume R is given for the remainder of this experiment; this will only influence the success rate in the case the SNR is 20 dB.

Next, we investigate whether the correct underlying decomposition can be recovered by CS-EVD when only I and J are known. A decomposition is successfully recovered if, for two decompositions $\mathcal{T} = \sum_r \mathcal{D}_r \bullet_1 \mathbf{A}_r \bullet_2 \mathbf{B}_r$ and $\hat{\mathcal{T}} = \sum_r \hat{\mathcal{D}}_r \bullet_1 \hat{\mathbf{A}}_r \bullet_2 \hat{\mathbf{B}}_r$,

[1] https://www.tensorlabplus.net/domanov2021mndot.html

the maximal relative term error E

$$E = \max_r \frac{\left|\left|\mathcal{D}_r \cdot_1 \mathbf{A}_r \cdot_2 \mathbf{B}_r - \hat{\mathcal{D}}_r \cdot_1 \hat{\mathbf{A}}_r \cdot_2 \hat{\mathbf{B}}_r\right|\right|}{||\mathcal{D}_r \cdot_1 \mathbf{A}_r \cdot_2 \mathbf{B}_r||} \leq \tau. \tag{3.1}$$

(See Table 3.2 for the values of τ that are used.) From Table 3.2, we conclude that CS-EVD can indeed recover the decomposition as long as the SNR is sufficiently high. However, if the CS-EVD solution is provided as the initialization for NLS, the success rate improves significantly. Hence, even for low SNR, CS-EVD provides a good initialization as long as R is estimated correctly. The optimization methods recover the solution almost always, but, as can be seen in Figure 3.1, for only a third of the initializations, the algorithms converge to the correct solution[2]. Moreover, these optimization methods require an initial guess and therefore the underlying parameters M_r and N_r need to be known.

In Figure 3.2, we show that the accuracy of CS-EVD is indeed limited which causes the lower success rate. For both the CS-EVD and the random initializations, NLS converges to the same solution. Finally, we investigate whether the CS-EVD algorithm can be improved by providing additional prior knowledge, or by using the extra (internal) refinement step using `cpd_nls`; see subsection 2.2. In Figure 3.3, we see that the extra information does not improve the accuracy, but the extra refinement step does. In all cases, refining the solution of CS-EVD using ALS or NLS is beneficial.

TABLE 3.1
Starting without prior knowledge, the proposed CS-EVD algorithm can recover the correct number of terms and their sizes in all cases except when the SNR is very low.

	successful estimation (/100 experiments)				
	SNR				
estimated	20 dB	40 dB	80 dB	160 dB	no noise
R	36	100	100	100	100
M_r, N_r	36	100	100	100	100

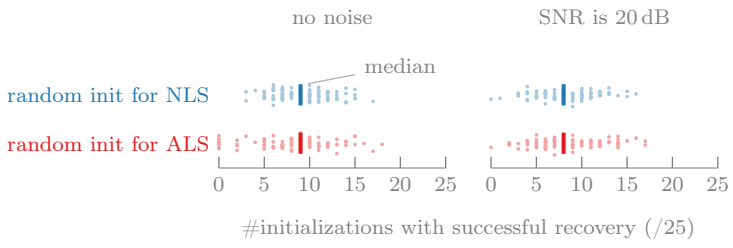


FIG. 3.1. *Only for approximately one in three random initializations (median) the correct solution is found by the optimization based algorithms. Note that for some experiments, the correct solution is found for only a few or even no random initializations. Each dot represents one out of 100 experiments and is shifted vertically and randomly to improve the visualization.*

[2]We assume that the correct solution is only slightly perturbed compared to the generated factors and core tensors.

TABLE 3.2

The proposed CS-EVD algorithm successfully recovers the decomposition, unless the tensor is highly perturbed by noise. When the solution is used as initialization for an optimization algorithm, almost all tensors are recovered if R is estimated correctly, hence accuracy is the main issue. An experiment is considered successful if the relative term error $E \leq \tau$ with $\tau = 0.25, 0.2, 0.002, 2 \cdot 10^{-7}, 10^{-13}$ for each column, resp.

		success rate (/100 experiments)				
		SNR				
	known	20 dB	40 dB	80 dB	160 dB	no noise
CS-EVD	–	9	99	100	100	100
	R	10	99	100	100	100
CS-EVD as init for NLS	–	34	100	100	100	100
	R	94	100	100	100	100
random init for NLS	M_r, N_r	99	100	100	100	100
random init for ALS	M_r, N_r	99	100	100	100	91

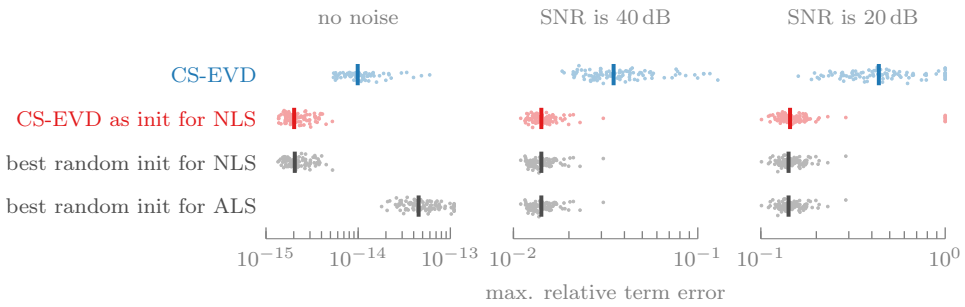


FIG. 3.2. *The accuracy of the decomposition computed using the CS-EVD method is slightly lower than when optimization-based methods such as NLS or ALS are used. However, CS-EVD provides a good initialization and uses only R as prior knowledge. (Not knowing R only changes the success ratio for the experiment with an SNR of 20 dB, as R is estimated correctly for the other cases; see Table 3.1.) Each dot represents one out of 100 experiments and is shifted vertically and randomly to improve the visualization; the bar represents the median.*

3.2. A practical, larger-scale example. In a more practical setting, the goal is often to decompose a noisy tensor $\bar{\mathcal{T}} \in \mathbb{C}^{\bar{I} \times \bar{J} \times \bar{K}}$ with $\bar{I} \geq I$, $\bar{J} \geq J$ and $\bar{K} \geq K$, into an unknown sum of simple rank-1 terms and more complex multilinear rank $(L_r, L_r, 1)$ or $(M_r, N_r, \cdot)$ terms. In this experiment, we investigate the ability of the algorithm to find the correct decomposition when changing the known parameters: I and J, R, or all values M_r and N_r. As we will show, the performance depends on the effectiveness of the rank estimation procedures, which are not at the core of the proposed algorithm.

A similar strategy as in subsection 3.1 is used to generate a noiseless tensor $\mathcal{T}$, but now with $M = [2, 4, 3, 5, 3, 1, 1]$, $N = [3, 2, 3, 2, 1, 1, 1]$ and $K = 50$; hence $I = \sum_r M_r = 19$ and $J = \sum_r N_r = 13$. Next, the tensor $\bar{\mathcal{T}} \in \mathbb{R}^{100 \times 100 \times 50}$ is constructed as $\bar{\mathcal{T}} = \mathcal{T} \cdot_1 \mathbf{U} \cdot_2 \mathbf{V}$ in which $\mathbf{U} \in \mathbb{C}^{100 \times I}$ and $\mathbf{V} \in \mathbb{C}^{100 \times J}$ are random matrices with orthonormal columns. Finally, the tensor is perturbed by noise such that the SNR is a given value. For these experiments, the results of the GEVDs in (2.5) and (2.6) are refined using a few iterations of Gauss–Newton; see subsection 2.2.

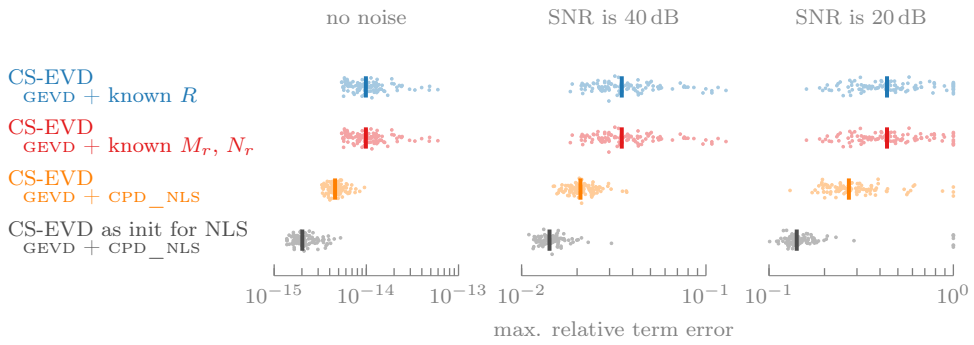


FIG. 3.3. *While the default algebraic version of the proposed CS-EVD algorithm is reasonably accurate, the accuracy can still be improved by replacing the GEVD algorithm used within. Note that knowing the correct parameters M_r and N_r, $r = 1, \ldots, R$ does not improve accuracy; it does improve the success rate as seen in Table 3.2. The accuracy can be improved by using the computed result as initialization for an optimization-based algorithm, e.g., based on NLS. Each dot represents one out of 100 experiments and is shifted vertically and randomly to improve the visualization; the bar represents the median.*

If stated explicitly, the algebraic result is further refined using ALS.

The problem studied in this section is a hard problem for optimization algorithms, as ALS and NLS do not converge to the correct solution for any random initialization for 51 and 70 experiments, resp., as can be seen in Table 3.3 for the case where the SNR is 10 dB. Moreover, if the correct solution is found, for only 1 experiment 3 initializations converged to this solution. In contrast, the CS-EVD method finds the correct solution for 83 and 98 experiments (for an SNR of 10 dB, resp. 40 dB), given no extra information.

In terms of accuracy, it is clear from Figure 3.4 that knowing more parameters does not lower the error; only the success rate improves. Again, using a refinement step results in a lower error. The improvement is limited in this case though: for the 10 dB experiment the median error reduces from 0.070 if no parameters are known to 0.053 if the result is refined.

There are multiple reasons why an experiment can be unsuccessful: the result is inaccurate, i.e., the error is higher than a particular threshold, the block size parameters M_r and N_r are estimated incorrectly, the estimated Null $(\mathbf{M}_{\mathcal{T}})$ is incorrect, or the estimated multilinear rank is wrong. From Figure 3.5, we see that if the multilinear is estimated correctly, the main reason is that dimension of the null space is estimated wrongly. The null space is not clearly defined as the singular value gap is too small: Let σ_r be the rth smallest singular value, then in almost all failed cases, $\sigma_{R+1} \leq 2\sigma_R$. If all parameters are known, and therefore the correct null space is used, the CS-EVD algorithm is still successful for smaller singular value gaps.

Finally, for reference we give indicative median timings for the 10 dB experiment[3]. While ALS and NLS require 7.30 s and 79.18 s, resp., the CS-EVD approach only requires 1.91 s which includes 0.05 s for MLSVD compression, 0.28 s for the computation of the GEVD, 1.28 s for the (internal) refinement step with `cpd_nls` and clustering, and 0.30 s for refinement of the CS-EVD solution using ALS. Hence, while the CS-EVD method does not only not require parameter estimates or an initialization, it is

[3]The experiments are performed on a laptop with an Intel Xeon E-2176M CPU, running Ubuntu 20.04.3, Matlab R2019b and Tensorlab 3.0.

also fast, even for larger tensors.

TABLE 3.3
The proposed CS-EVD algorithm recovers the correct solution for most (10 dB) or almost all (40 dB) experiments, while the optimization routines, which require prior information, have a much lower success rate. Even if the optimization method finds the correct solution, the number of initialization for which this happens is at most 3. The number of experiments that have at least k out of 100 random initializations that converge to the correct solution, as well as the total number successful experiment are shown for 100 experiments.

	SNR = 10 dB				SNR = 40 dB			
	k				k			
	1	2	3	# successful	1	2	3	# successful
CS-EVD as init for ALS	–	–	–	83	–	–	–	98
random init for ALS	49	8	1	49	42	8	1	42
random init for NLS	30	5	1	30	21	2	0	21

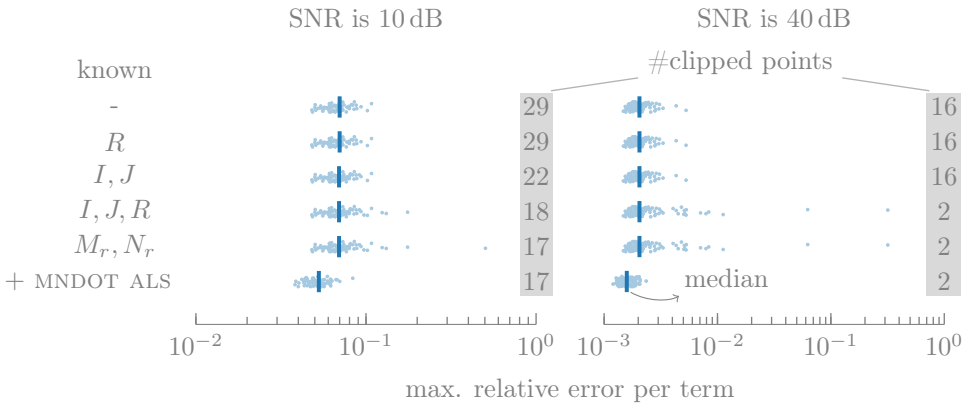


FIG. 3.4. *The knowledge of additional information does not improve the accuracy, only the success rate; see Figure 3.5. An additional refinement step using an optimization-based algorithm such as ALS can improve the accuracy slightly. Results shown for a decomposition of a noisy $100 \times 100 \times 50$ tensor into $R = 7$ multilinear rank $(M_r, N_r, \cdot)$ terms with $M = [2, 4, 3, 5, 3, 1, 1]$ and $N = [3, 2, 3, 2, 1, 1, 1]$, hence $I = 19$ and $J = 13$. Each dot represents one out of 100 experiments and is shifted vertically and randomly to improve the visualization; the same shifts are used for all plots. Relative errors larger than 1 are clipped.*

4. Proof of Theorem 2.1 and Corollary 2.2.

We need the following lemmas.

LEMMA 4.1. *Let the conditions of Theorem 2.1 hold and let $\mathbf{U}_1$ and $\mathbf{U}_2$ denote the "U" factors in the compact SVDs of $\mathbf{T}_{(1)}$ and $\mathbf{T}_{(2)}$, respectively. Then*

1. *tensor $\mathcal{T} \in \mathbb{F}^{I \times J \times K}$ is ML rank-$(\sum M_r, \sum N_r, \cdot)$;*
2. *the "compressed" matrices $\mathbf{A}^c := \mathbf{U}_1^H \mathbf{A}$ and $\mathbf{B}^c := \mathbf{U}_2^H \mathbf{B}$ are square and non-singular;*
3. *$\mathbf{A} = \mathbf{U}_1 \mathbf{A}^c$ and $\mathbf{B} = \mathbf{U}_2 \mathbf{B}^c$;*
4. *the "compressed" tensor $\mathcal{T}^c := \mathcal{T} \cdot_1 \mathbf{U}_1^H \cdot_2 \mathbf{U}_2^H \in \mathbb{F}^{\sum M_r \times \sum N_r \times K}$ is ML rank-$(\sum M_r, \sum N_r, \cdot)$;*

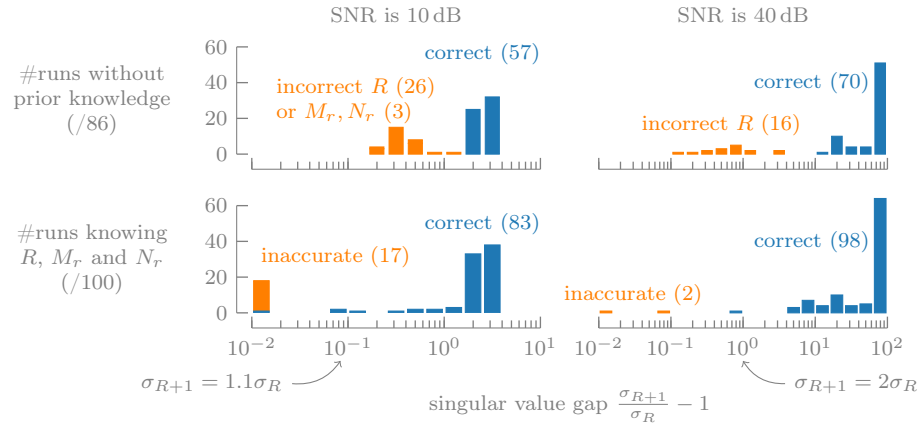


FIG. 3.5. *Most failures are due to an insufficiently large singular value gap, which causes* $\text{Null}\,(\mathbf{M}_{\mathcal{T}})$ *to be estimated incorrectly. For each experiment, the reason why the decomposition is not found is indicated. Results shown for a decomposition of a noisy* $100 \times 100 \times 50$ *tensor into* $R = 7$ *multilinear rank* $(M_r, N_r, \cdot)$ *terms with* $M = [2, 4, 3, 5, 3, 1, 1]$ *and* $N = [3, 2, 3, 2, 1, 1, 1]$. *In the top row, 14 experiments are excluded as* I *and* J *are estimated incorrectly.*

5. *tensor* $\mathcal{T}^c$ *admits the ML rank-*$(M_r, N_r, \cdot)$ *decomposition*

$$\mathcal{T}^c = \sum_{r=1}^{R} \mathcal{D}_r \bullet_1 (\mathbf{U}_1^H \mathbf{A}_r) \bullet_2 (\mathbf{U}_2^H \mathbf{B}_r),$$

where the tensors $\mathcal{D}_r \in \mathbb{F}^{M_r \times N_r \times K}$ *are the same as in decomposition* (1.3).

Proof. 1. First we show that the matrices

$$\mathbf{D}_{(1)} = [\mathbf{D}_1 \ \ldots \ \mathbf{D}_K] \quad \text{and} \quad \mathbf{D}_{(2)} = [\mathbf{D}_1^{\mathrm{T}} \ \ldots \ \mathbf{D}_K^{\mathrm{T}}] \quad \text{have full row rank.} \tag{4.1}$$

Assume that $\mathbf{f}^{\mathrm{T}}\mathbf{D}_{(1)} = \mathbf{0}$. Then $\mathbf{1}\mathbf{f}^{\mathrm{T}}\mathbf{D}_k = \mathbf{D}_k\mathbf{O}_{\sum N_r}$, where $\mathbf{1}$ denotes the $\sum M_r \times 1$ vector of ones. Hence $\begin{bmatrix} \mathbf{f} \otimes \mathbf{1} \\ \mathbf{0} \end{bmatrix} \in \text{Null}\,(\mathbf{M}_{\mathcal{D}})$. Since $\dim \text{Null}\,(\mathbf{M}_{\mathcal{D}}) = R$, it follows that the rank-1 matrix $\mathbf{1}\mathbf{f}^{\mathrm{T}}$ is a linear combination of the matrices $\mathbf{P}_{I,1}, \ldots, \mathbf{P}_{I,R}$. Hence $\mathbf{f} = \mathbf{0}$. Thus $\mathbf{D}_{(1)}$ has full row rank. In a similar way one can prove that $\mathbf{D}_{(2)}$ has also full row rank. Now, from assumption 1 in Theorem 2.1, (4.1), and the identities

$$\mathbf{T}_{(1)} = \mathbf{A}\mathbf{D}_{(1)}\,\text{blockdiag}(\mathbf{B}^{\mathrm{T}}, \ldots, \mathbf{B}^{\mathrm{T}}), \quad \mathbf{T}_{(2)} = \mathbf{B}\mathbf{D}_{(2)}\,\text{blockdiag}(\mathbf{A}^{\mathrm{T}}, \ldots, \mathbf{A}^{\mathrm{T}}) \tag{4.2}$$

it follows that

$$\text{col}(\mathbf{T}_{(1)}) = \text{col}(\mathbf{A}) \quad \text{and} \quad \text{col}(\mathbf{T}_{(2)}) = \text{col}(\mathbf{B}), \tag{4.3}$$

where $\text{col}(\cdot)$ denotes the column space of a matrix. Hence $r_{\mathbf{T}_{(1)}} = r_{\mathbf{A}} = \sum M_r$ and $r_{\mathbf{T}_{(2)}} = r_{\mathbf{B}} = \sum N_r$, implying that the tensor $\mathcal{T} \in \mathbb{F}^{I \times J \times K}$ is ML rank-$(\sum M_r, \sum N_r, \cdot)$.

2. and 3. follow from (4.3) and the fact that $\mathbf{U}_1$ and $\mathbf{U}_2$ are the "U" factors in the compact SVDs of $\mathbf{T}_{(1)}$ and $\mathbf{T}_{(2)}$, respectively.

4. From the construction of $\mathbf{T}^c_{(1)}$ and (4.2) it follows that

$$\mathbf{T}^c_{(1)} = [\mathbf{U}_1^{\mathrm{H}}\mathbf{T}_1\mathbf{U}_2^* \ \dots \ \mathbf{U}_1^{\mathrm{H}}\mathbf{T}_K\mathbf{U}_2^*] = \mathbf{U}_1^{\mathrm{H}}\mathbf{T}_{(1)}\,\mathrm{blockdiag}(\mathbf{U}_2^*,\dots,\mathbf{U}_2^*) = \mathbf{U}_1^{\mathrm{H}}\mathbf{A}\mathbf{D}_{(1)}\,\mathrm{blockdiag}((\mathbf{U}_2^{\mathrm{H}}\mathbf{B})^{\mathrm{T}},\dots,(\mathbf{U}_2^{\mathrm{H}}\mathbf{B})^{\mathrm{T}}) = \mathbf{A}^c\mathbf{D}_{(1)}\,\mathrm{blockdiag}((\mathbf{B}^c)^{\mathrm{T}},\dots,(\mathbf{B}^c)^{\mathrm{T}}). \tag{4.4}$$

Hence, by statement 2. and (4.1), $r_{\mathbf{T}^c_{(1)}} = r_{\mathbf{A}^c} = \sum M_r$. In a similar way one can prove that $r_{\mathbf{T}^c_{(2)}} = r_{\mathbf{B}^c} = \sum N_r$. Thus, $\mathcal{T}^c$ is ML rank-$(\sum M_r, \sum N_r, \cdot)$.

5. Follows from (4.4) and 2. ∎

LEMMA 4.2. *Let $I = \sum M_r$, $J = \sum N_r$, $\mathbf{A}$ and $\mathbf{B}$ be nonsingular, and let the matrices $\mathbf{M}_{\mathcal{D}}$ and $\mathbf{M}_{\mathcal{T}}$ be defined in (2.1) and (2.4), respectively. Then*

1) the matrices $\mathbf{M}_{\mathcal{D}}$ and $\mathbf{M}_{\mathcal{T}}$ are related by

$$\mathbf{M}_{\mathcal{T}} = \mathrm{blockdiag}(\mathbf{B}\otimes\mathbf{A},\dots,\mathbf{B}\otimes\mathbf{A})\cdot\mathbf{M}_{\mathcal{D}}\cdot\begin{bmatrix}\mathbf{A}^T\otimes\mathbf{A}^{-1} & \mathbf{O}\\ \mathbf{O} & \mathbf{B}^{-1}\otimes\mathbf{B}^T\end{bmatrix};$$

2) the matrices $\mathbf{X}\in\mathbb{F}^{I\times I}$ and $\mathbf{Y}\in\mathbb{F}^{J\times J}$ satisfy the equations $\mathbf{X}\mathbf{D}_k = \mathbf{D}_k\mathbf{Y}$, $k = 1,\dots,K$ if and only if $\mathbf{M}_{\mathcal{D}}\begin{bmatrix}\mathrm{vec}(\mathbf{X})\\ \mathrm{vec}(\mathbf{Y})\end{bmatrix} = \mathbf{0}$;

3) the matrices $\mathbf{X}\in\mathbb{F}^{I\times I}$ and $\mathbf{Y}\in\mathbb{F}^{J\times J}$ satisfy the equations $\mathbf{X}\mathbf{T}_k = \mathbf{T}_k\mathbf{Y}$, $k = 1,\dots,K$ if and only if $\mathbf{M}_{\mathcal{T}}\begin{bmatrix}\mathrm{vec}(\mathbf{X})\\ \mathrm{vec}(\mathbf{Y})\end{bmatrix} = \mathbf{0}$;

4) $\mathrm{Null}\,(\mathbf{M}_{\mathcal{D}}) \supseteq \mathrm{span}\left\{\begin{bmatrix}\mathrm{vec}(\mathbf{P}_{I,r})\\ \mathrm{vec}(\mathbf{P}_{J,r})\end{bmatrix} : \ r = 1,\dots,R\right\}$, where the projectors $\mathbf{P}_{I,r}$ and $\mathbf{P}_{J,r}$ are defined in (2.3);

5) $\mathrm{Null}\,(\mathbf{M}_{\mathcal{T}}) = \begin{bmatrix}\mathbf{A}^{-T}\otimes\mathbf{A} & \mathbf{O}\\ \mathbf{O} & \mathbf{B}\otimes\mathbf{B}^{-T}\end{bmatrix}\mathrm{Null}\,(\mathbf{M}_{\mathcal{D}})$;

6) $\dim\mathrm{Null}\,(\mathbf{M}_{\mathcal{T}}) = \dim\mathrm{Null}\,(\mathbf{M}_{\mathcal{D}}) \geq R$;

7) if $\dim\mathrm{Null}\,(\mathbf{M}_{\mathcal{D}}) = R$, then the matrices

$$[\mathbf{A}\mathbf{D}_1\mathbf{B}^T \ \dots \ \mathbf{A}\mathbf{D}_K\mathbf{B}^T] \in \mathbb{F}^{I\times JK} \quad \text{and} \quad [\mathbf{B}\mathbf{D}_1^T\mathbf{A}^T \ \dots \ \mathbf{B}\mathbf{D}_K^T\mathbf{A}^T] \in \mathbb{F}^{J\times IK}$$

have full row rank.

Proof. 1) follows from (1.7), (2.1), and (2.4); 2)–4) are trivial; 5) follows from 1); 6) follows from 5) and 4); 7) follows from Lemma 4.1 1. and (4.2). ∎

Now we are ready to prove Theorem 2.1.

Proof of Theorem 2.1. By Lemma 4.1, without loss of generality (WLOG), we can assume that $I = \sum M_r$ and $J = \sum N_r$.

1) First we show that decomposition (1.3) can be computed algebraically. From assumption 2) and Lemma 4.2 4), 6) we have that

$$\mathrm{Null}\,(\mathbf{M}_{\mathcal{D}}) = \mathrm{span}\left\{\begin{bmatrix}\mathrm{vec}(\mathbf{P}_{I,r})\\ \mathrm{vec}(\mathbf{P}_{J,r})\end{bmatrix} : \ r = 1,\dots,R\right\} \quad \text{and} \quad \dim\mathrm{Null}\,(\mathbf{M}_{\mathcal{T}}) = R.$$

Hence, by Lemma 4.2 5),

$$\mathrm{Null}\,(\mathbf{M}_{\mathcal{T}}) = \begin{bmatrix}\mathbf{A}^{-\mathrm{T}}\otimes\mathbf{A} & \mathbf{O}\\ \mathbf{O} & \mathbf{B}\otimes\mathbf{B}^{-\mathrm{T}}\end{bmatrix}\mathrm{span}\left\{\begin{bmatrix}\mathrm{vec}(\mathbf{P}_{I,r})\\ \mathrm{vec}(\mathbf{P}_{J,r})\end{bmatrix} : \ r = 1,\dots,R\right\}. \tag{4.5}$$

Let the columns of $\mathbf{U} \in \mathbb{F}^{(I^2+J^2)\times R}$ form a basis of $\operatorname{Null}(\mathbf{M}_{\mathcal{T}})$. We represent $\mathbf{U}$ as

$$\mathbf{U} = \begin{bmatrix} \operatorname{vec}(\mathbf{U}_{I,1}) & \dots & \operatorname{vec}(\mathbf{U}_{I,R}) \\ \operatorname{vec}(\mathbf{U}_{J,1}) & \dots & \operatorname{vec}(\mathbf{N}_{J,R}) \end{bmatrix}, \tag{4.6}$$

where $\mathbf{U}_{I,r} \in \mathbb{F}^{I\times I}$ and $\mathbf{U}_{J,r} \in \mathbb{F}^{J\times J}$, $r = 1, \dots, R$. From (4.5) and (4.6), it follows that there exists a unique nonsingular $R \times R$ matrix $\mathbf{G}$ such that

$$\begin{aligned} \begin{bmatrix} \operatorname{vec}(\mathbf{U}_{I,1}) & \dots & \operatorname{vec}(\mathbf{U}_{I,R}) \end{bmatrix} &= \left(\mathbf{A}^{-\mathrm{T}} \otimes \mathbf{A}\right) [\operatorname{vec}(\mathbf{P}_{I,1}) \ \dots \ \operatorname{vec}(\mathbf{P}_{I,R})]\mathbf{G}, \\ \begin{bmatrix} \operatorname{vec}(\mathbf{U}_{J,1}) & \dots & \operatorname{vec}(\mathbf{U}_{J,R}) \end{bmatrix} &= \left(\mathbf{B} \otimes \mathbf{B}^{-\mathrm{T}}\right) [\operatorname{vec}(\mathbf{P}_{J,1}) \ \dots \ \operatorname{vec}(\mathbf{P}_{J,R})]\mathbf{G}. \end{aligned}$$

Hence, (2.5)–(2.6) hold. Since the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{G}$ are nonsingular, it follows that the matrices $\mathbf{U}_{I,1}, \dots, \mathbf{U}_{I,R}$ are linearly independent. Thus, we can recover $\mathbf{A}$ from (2.5). Similarly, we can recover $\mathbf{B}$ from (2.6).

2) Now we show that decomposition (1.3) is unique. Assume that (1.3) holds for $\mathbf{A}_r$, $\mathbf{B}_r$, etc. replaced by $\widetilde{\mathbf{A}}_r$, $\widetilde{\mathbf{B}}_r$, etc. To prove the uniqueness it is sufficient to show that $\widetilde{\mathbf{A}}$, $\widetilde{\mathbf{B}}$ and $\mathbf{M}_{\widetilde{\mathcal{D}}}$ satisfy assumptions 1) and 2), i.e., that $\widetilde{\mathbf{A}}$ and $\widetilde{\mathbf{B}}$ are nonsingular and $\dim \operatorname{Null}\left(\mathbf{M}_{\widetilde{\mathcal{D}}}\right) = R$. By (1.7),

$$\begin{aligned} [\mathbf{A}\mathbf{D}_1\mathbf{B}^{\mathrm{T}} \ \dots \ \mathbf{A}\mathbf{D}_K\mathbf{B}^{\mathrm{T}}] =& [\mathbf{T}_1 \ \dots \ \mathbf{T}_K] = \\ & [\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}_1\widetilde{\mathbf{B}}^{\mathrm{T}} \ \dots \ \widetilde{\mathbf{A}}\widetilde{\mathbf{D}}_K\widetilde{\mathbf{B}}^{\mathrm{T}}] = \widetilde{\mathbf{A}}[\widetilde{\mathbf{D}}_1\widetilde{\mathbf{B}}^{\mathrm{T}} \ \dots \ \widetilde{\mathbf{D}}_K\widetilde{\mathbf{B}}^{\mathrm{T}}], \end{aligned} \tag{4.7}$$

$$\begin{aligned} [\mathbf{B}\mathbf{D}_1^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} \ \dots \ \mathbf{B}\mathbf{D}_K^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}] =& [\mathbf{T}_1^{\mathrm{T}} \ \dots \ \mathbf{T}_K^{\mathrm{T}}] = \\ & [\widetilde{\mathbf{B}}\widetilde{\mathbf{D}}_1^{\mathrm{T}}\widetilde{\mathbf{A}}^{\mathrm{T}} \ \dots \ \widetilde{\mathbf{B}}\widetilde{\mathbf{D}}_K^{\mathrm{T}}\widetilde{\mathbf{A}}^{\mathrm{T}}] = \widetilde{\mathbf{B}}[\widetilde{\mathbf{D}}_1^{\mathrm{T}}\widetilde{\mathbf{A}}^{\mathrm{T}} \ \dots \ \widetilde{\mathbf{D}}_K^{\mathrm{T}}\widetilde{\mathbf{A}}^{\mathrm{T}}]. \end{aligned} \tag{4.8}$$

Since, by Lemma 4.2 7), the matrices in the left-hand side of (4.7) and (4.8) have full row rank, it follows that the matrices $\widetilde{\mathbf{A}}$ and $\widetilde{\mathbf{B}}$ are nonsingular. Since, by assumption 2), $\dim \operatorname{Null}(\mathbf{M}_{\mathcal{D}}) = R$, and, by Lemma 4.2 5),

$$\begin{aligned} \operatorname{Null}(\mathbf{M}_{\mathcal{T}}) = \begin{bmatrix} \mathbf{A}^{-\mathrm{T}} \otimes \mathbf{A} & \mathbf{O} \\ \mathbf{O} & \mathbf{B} \otimes \mathbf{B}^{-\mathrm{T}} \end{bmatrix} \operatorname{Null}(\mathbf{M}_{\mathcal{D}}) = \\ \begin{bmatrix} \widetilde{\mathbf{A}}^{-\mathrm{T}} \otimes \widetilde{\mathbf{A}} & \mathbf{O} \\ \mathbf{O} & \widetilde{\mathbf{B}} \otimes \widetilde{\mathbf{B}}^{-\mathrm{T}} \end{bmatrix} \operatorname{Null}\left(\mathbf{M}_{\widetilde{\mathcal{D}}}\right), \end{aligned}$$

it follows that $\dim \operatorname{Null}\left(\mathbf{M}_{\widetilde{\mathcal{D}}}\right) = R$. ∎

Proof of Corollary 2.2. It is sufficient to prove that assumption 2 in Corollary 2.2 implies assumption 2 in Theorem 2.1. Thus, we need to show that the subspace of solution of (2.2) is spanned by the trivial solutions $(\mathbf{X}, \mathbf{Y}) = (\mathbf{P}_{I,r}, \mathbf{P}_{J,r})$, $r = 1, \dots, R$, where $\mathbf{P}_{I,r}$ and $\mathbf{P}_{J,r}$ are defined in (2.3) and $M_1 = \dots = M_R = 1$.

Let $\mathbf{X} = (x_{ij})_{i,j=1}^{R}$ and let $\mathbf{Y} = (\mathbf{Y}_{ij})_{i,j=1}^{R}$ consists of the blocks $\mathbf{Y}_{ij} \in \mathbb{F}^{N_i\times N_j}$. Since the matrices $\mathbf{D}_k$ are block-diagonal, we can rewrite (2.2) as

$$x_{ij}\mathbf{D}_{j,k} = \mathbf{D}_{i,k}\mathbf{Y}_{ij}, \qquad 1 \le i, j \le R, \qquad k = 1, \dots, K,$$

which, by (2.7), is equivalent to

$$x_{ij}\mathbf{C}_j = \mathbf{C}_i\mathbf{Y}_{ij}, \qquad 1 \le i, j \le R. \tag{4.9}$$

We need to show that

$$\mathbf{Y}_{ii} = x_{ii}\mathbf{I}_{N_i}, \qquad 1 \le i \le R, \tag{4.10}$$

$$\mathbf{Y}_{ij} = \mathbf{O}, \qquad x_{ij} = 0, \qquad 1 \le i \ne j \le R. \tag{4.11}$$

(i) Let us prove (4.10). By (4.9),

$$\mathbf{C}_i(x_{ii}\mathbf{I}_{N_i} - \mathbf{Y}_{ii}) = \mathbf{O}, \qquad 1 \le i \le R. \tag{4.12}$$

Since the $1 \times N_i \times K$ tensor $\mathcal{D}_i$ in (1.3) is ML rank-$(1, N_i, N_i)$ and the second matrix unfolding of $\mathcal{D}_i$ coincides with $\mathbf{C}_i^{\mathrm{T}}$, it follows that $\mathbf{C}_i^{\mathrm{T}}$ has full row rank or that $\mathbf{C}_i$ has full column rank. Hence (4.10) follows from (4.12).

(ii) Let us prove (4.11). It is clear that (4.9) can be rewritten as

$$\mathbf{O} = [\mathbf{C}_i \;\; \mathbf{C}_j] \begin{bmatrix} \mathbf{Y}_{ij} \\ x_{ij}\mathbf{I}_{N_j} \end{bmatrix}. \tag{4.13}$$

Assume to the contrary that $x_{ij} \neq 0$. Then from (4.13), Sylvester's rank inequality, and assumption 2 it follows that

$$0 \ge r_{[\mathbf{C}_i \;\; \mathbf{C}_j]} + r_{[\mathbf{Y}_{ij}^{\mathrm{T}} \;\; x_{ij}\mathbf{I}_{N_j}]^{\mathrm{T}}} - (N_i + N_j) \ge \max(N_i, N_j) + 1 + N_j - (N_i + N_j) \ge 1,$$

which is a contradiction. Thus, $x_{ij} = 0$. Hence, by (4.9), $\mathbf{C}_i\mathbf{Y}_{ij} = \mathbf{O}$. Since $\mathbf{C}_i$ has full column rank, it follows that $\mathbf{Y}_{ij} = \mathbf{O}$. ∎

5. Proofs of Theorem 2.3 and Theorem 2.8. We need the following two lemmas.

LEMMA 5.1. *Let*

$$K = \max\left(\left\lceil \frac{M}{N} \right\rceil + 1, \left\lceil \frac{N}{M} \right\rceil + 1, 3\right). \tag{5.1}$$

Then there exist matrices $\mathbf{E}_1, \ldots, \mathbf{E}_K \in \mathbb{F}^{M \times N}$ *such that the system of matrix equations*

$$\mathbf{X}\mathbf{E}_k = \mathbf{E}_k\mathbf{Y}, \qquad \mathbf{X} \in \mathbb{F}^{M \times M}, \;\; \mathbf{Y} \in \mathbb{F}^{N \times N} \qquad\qquad k = 1, \ldots, K \tag{5.2}$$

has only the trivial solutions $\mathbf{X} = c\mathbf{I}_M$ *and* $\mathbf{Y} = c\mathbf{I}_N$*, where* $c \in \mathbb{F}$*.*

Proof. It is clear that by taking the transpose of (5.2) we obtain an equivalent problem of the form (5.2) in which M and N are switched. Thus, w.l.o.g., we can always assume that $M \ge N$. We consider three cases: $M = N$, $N < M \le 2N$, and $2N < M$.

Case $M = N$. By assumption (5.1), $K = 3$. Let $\mathbf{a} \in \mathbb{F}^N$ be a vector with distinct entries and let $\mathbf{D} \in \mathbb{F}^{N \times N}$ be a matrix with nonzero entries. We show that if $\mathbf{E}_1 = \mathbf{I}_N$, $\mathbf{E}_2 = \operatorname{diag}(\mathbf{a})$, and $\mathbf{E}_3 = \mathbf{D}$, then (5.2) has only trivial solutions.

The equation $\mathbf{X}\mathbf{E}_1 = \mathbf{E}_1\mathbf{Y}$ implies that $\mathbf{X} = \mathbf{Y}$. Hence,

$$\mathbf{Y}\operatorname{diag}(\mathbf{a}) = \mathbf{Y}\mathbf{E}_2 = \mathbf{X}\mathbf{E}_2 = \mathbf{E}_2\mathbf{Y} = \operatorname{diag}(\mathbf{a})\mathbf{Y}.$$

Since the entries of $\mathbf{a}$ are distinct, it follows that $\mathbf{Y}$ is a diagonal matrix, $\mathbf{Y} = \operatorname{diag}(y_1, \ldots, y_N)$. Finally, since the entries of $\mathbf{E}_3 = \mathbf{D}$ are nonzero, the equation

$$\operatorname{diag}(y_1, \ldots, y_N)\mathbf{D} = \mathbf{Y}\mathbf{E}_3 = \mathbf{X}\mathbf{E}_3 = \mathbf{E}_3\mathbf{Y} = \mathbf{D}\operatorname{diag}(y_1, \ldots, y_N)$$

implies that $y_1 = \cdots = y_N =: c$. Thus, $\mathbf{X} = \mathbf{Y} = c\mathbf{I}_N$.

Case $N < M \le 2N$. By assumption (5.1), $K = 3$. Let $\mathbf{a} \in \mathbb{F}^N$ be a vector with distinct entries, $\mathbf{D} \in \mathbb{F}^{(M-N)\times(2N-M)}$ be a full row rank matrix with nonzero entries,

and let $\widehat{\mathbf{I}}_{M-N}$ denote the matrix formed by the first $M-N$ rows of $\mathbf{I}_N$. We show that if

$$\mathbf{E}_1 = \begin{bmatrix} \mathbf{I}_N \\ \mathbf{O} \end{bmatrix}, \qquad \mathbf{E}_2 = \begin{bmatrix} \mathbf{O} \\ \mathbf{D} \end{bmatrix}, \qquad \mathbf{E}_3 = \begin{bmatrix} \operatorname{diag}(\mathbf{a}) \\ \widehat{\mathbf{I}}_{M-N} \end{bmatrix},$$

then (5.2) has only trivial solutions. Let $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_3 \\ \mathbf{X}_2 & \mathbf{X}_4 \end{bmatrix}$ with $\mathbf{X}_1 \in \mathbb{F}^{N\times N}$, $\mathbf{X}_2, \mathbf{X}_3^{\mathrm{T}} \in \mathbb{F}^{(M-N)\times N}$ and $\mathbf{X}_4 \in \mathbb{F}^{(M-N)\times(M-N)}$. Then the equations $\mathbf{X}\mathbf{E}_1 = \mathbf{E}_1\mathbf{Y}$, $\mathbf{X}\mathbf{E}_2 = \mathbf{E}_2\mathbf{Y}$, and $\mathbf{X}\mathbf{E}_3 = \mathbf{E}_3\mathbf{Y}$ imply

$$\begin{aligned} \mathbf{X}_1 = \mathbf{Y} \quad &\text{and} \quad \mathbf{X}_2 = 0, \\ \mathbf{X}_3\mathbf{D} = \mathbf{O} \quad &\text{and} \quad \mathbf{X}_4\mathbf{D} = \mathbf{D}\mathbf{Y}, \end{aligned} \tag{5.3}$$

$$\mathbf{X}_1 \operatorname{diag}(\mathbf{a}) + \mathbf{X}_3\widehat{\mathbf{I}}_{M-N} = \operatorname{diag}(\mathbf{a})\mathbf{Y} \quad \text{and} \quad \mathbf{X}_2 \operatorname{diag}(\mathbf{a}) + \mathbf{X}_4\widehat{\mathbf{I}}_{M-N} = \widehat{\mathbf{I}}_{M-N}\mathbf{Y}, \tag{5.4}$$

respectively. Since $\mathbf{X}_3\mathbf{D} = \mathbf{O}$ and $\mathbf{D}$ has full row rank, it follows that $\mathbf{X}_3 = \mathbf{O}$. Hence, by (5.3) and (5.4),

$$\mathbf{Y} \operatorname{diag}(\mathbf{a}) = \mathbf{X}_1 \operatorname{diag}(\mathbf{a}) + \mathbf{X}_3\widehat{\mathbf{I}}_{M-N} = \operatorname{diag}(\mathbf{a})\mathbf{Y}.$$

Since all entries of $\mathbf{a}$ are distinct, it follows that $\mathbf{Y}$ is a diagonal matrix, $\mathbf{Y} = \operatorname{diag}(y_1, \dots, y_N)$. From the second identities in (5.3) and (5.4) it follows that $\mathbf{X}_4 = \operatorname{diag}(y_1, \dots, y_{M-N})$. Finally, since the entries of $\mathbf{D}$ are nonzero, it follows from $\mathbf{X}_4\mathbf{D} = \mathbf{D}\mathbf{Y}$ that $y_1 = \cdots = y_N =: c$. Hence, $\mathbf{Y} = c\mathbf{I}_N$ and $\mathbf{X} = \operatorname{blockdiag}(\mathbf{X}_1, \mathbf{X}_4) = c\mathbf{I}_M$.

Case $2N < M$. By assumption (5.1), $K = \left\lceil \frac{M}{N} \right\rceil + 1$. Let $r = M - (K-2)N$. Then $0 < r \le N$. We define $\mathbf{E}_1, \dots, \mathbf{E}_{K-2} \in \mathbb{F}^{M\times N}$ and $\widehat{\mathbf{E}}_{K-1} \in \mathbb{F}^{M\times r}$ as blocks in the partition $\mathbf{I}_M = [\mathbf{E}_1 \ \dots \ \mathbf{E}_{K-2} \ \widehat{\mathbf{E}}_{K-1}]$ and set $\mathbf{E}_{K-1} = [\widehat{\mathbf{E}}_{K-1} \ \mathbf{O}]$, where the zero matrix has $N-r$ columns. We also set $\mathbf{E}_k = \begin{bmatrix} \operatorname{diag}(\mathbf{a}) \\ \mathbf{D} \\ \mathbf{O} \end{bmatrix}$, where $\mathbf{a} \in \mathbb{F}^N$ is a vector with distinct entries and $\mathbf{D} \in \mathbb{F}^{N\times N}$ is a matrix with nonzero entries. We show that for such choice of the matrices $\mathbf{E}_k$ system (5.2) has only the trivial solution. From the first $K-1$ equations in (5.2) it follows that $\mathbf{X} = \operatorname{blockdiag}(\mathbf{Y}, \dots, \mathbf{Y}, \widehat{\mathbf{Y}})$, where $\widehat{\mathbf{Y}}$ denotes a matrix located at the intersection of the first r rows and r columns of $\mathbf{Y}$. The remaining equation $\mathbf{X}\mathbf{E}_K = \mathbf{E}_K\mathbf{Y}$ in (5.2) implies that $\mathbf{Y}\operatorname{diag}(\mathbf{a}) = \operatorname{diag}(\mathbf{a})\mathbf{Y}$ and $\mathbf{Y}\mathbf{D} = \mathbf{D}\mathbf{Y}$. Now, in exactly the same way as "*Case* $M = N$" one can obtain that $\mathbf{Y} = c\mathbf{I}_N$ for some $c \in \mathbb{F}$. Hence, $\mathbf{X} = c\mathbf{I}_M$. ∎

LEMMA 5.2. *Let*

$$K = \begin{cases} 2, \text{ if } \ M_1 > N_1 \ \text{ and } \ M_2 < N_2, \\ \lceil M_2/N_2 \rceil + 1, \text{ if } \ M_1 > N_1 \ \text{ and } \ M_2 \ge N_2, \\ \lceil N_1/M_1 \rceil + 1, \text{ if } \ M_1 \le N_1 \ \text{ and } \ M_2 < N_2, \\ \lceil M_2/N_2 \rceil + \lceil N_1/M_1 \rceil, \text{ if } \ M_1 \le N_1 \ \text{ and } \ M_2 \ge N_2. \end{cases} \tag{5.5}$$

Then there exist matrices $\mathbf{E}_1, \dots, \mathbf{E}_K \in \mathbb{F}^{M_2\times N_2}$ *and* $\mathbf{F}_1, \dots, \mathbf{F}_K \in \mathbb{F}^{M_1\times N_1}$ *such that the system of matrix equations*

$$\mathbf{X}\mathbf{E}_k = \mathbf{F}_k\mathbf{Y}, \qquad \mathbf{X} \in \mathbb{F}^{M_1\times M_2}, \ \mathbf{Y} \in \mathbb{F}^{N_1\times N_2} \qquad k = 1, \dots, K \tag{5.6}$$

has only the zero solution.

Proof. We consider each of the four cases in (5.5) separately.

Case $M_1 > N_1$ and $M_2 < N_2$. By (5.5), $K = 2$. By definition put

$$\mathbf{E}_1 = \mathbf{O}, \qquad \mathbf{E}_2 = [\mathbf{I}_{M_2}\ \mathbf{O}], \qquad \mathbf{F}_1 = \begin{bmatrix} \mathbf{I}_{N_1} \\ \mathbf{O} \end{bmatrix}, \qquad \mathbf{E}_2 = \mathbf{O}.$$

Then the equations $\mathbf{X}\mathbf{E}_1 = \mathbf{F}_1\mathbf{Y}$ and $\mathbf{X}\mathbf{E}_2 = \mathbf{F}_2\mathbf{Y}$ imply that $\mathbf{Y} = \mathbf{O}$ and $\mathbf{X} = \mathbf{O}$, respectively.

Case $M_1 > N_1$ and $M_2 \geq N_2$. By (5.5), $K = \lceil M_2/N_2 \rceil + 1$. Let $\mathbf{E}_1, \ldots, \mathbf{E}_{K-1}$ be matrices such that $[\mathbf{E}_1\ \ldots \mathbf{E}_{K-1}]$ has full row rank, let $\mathbf{E}_K$ be an arbitrary matrix, $\mathbf{F}_1, \ldots, \mathbf{F}_{K-1}$ be zero matrices, and $\mathbf{F}_K = \begin{bmatrix} \mathbf{I}_{N_1} \\ \mathbf{O} \end{bmatrix}$. Then from the first $K-1$ equations in (5.6) it follows that $\mathbf{X}[\mathbf{E}_1\ \ldots\ \mathbf{E}_{K-1}] = 0$. Hence $\mathbf{X} = \mathbf{O}$. Now the remaining equation $\mathbf{X}\mathbf{E}_K = \mathbf{F}_K\mathbf{Y}$ implies that $\mathbf{Y} = \mathbf{O}$.

Case $M_1 \leq N_1$ and $M_2 < N_2$. By (5.5), $K = \lceil N_1/M_1 \rceil + 1$. The problem can be reduced to the setting used in the previous case by taking the transpose of the equations in (5.6).

Case $M_1 \leq N_1$ and $M_2 \geq N_2$. By (5.5), $K = \lceil M_2/N_2 \rceil + \lceil N_1/M_1 \rceil$. Let $\mathbf{E}_1, \ldots, \mathbf{E}_{\lceil M_2/N_2 \rceil}$ be matrices such that $[\mathbf{E}_1\ \ldots\ \mathbf{E}_{\lceil M_2/N_2 \rceil}]$ has full row rank and let $\mathbf{E}_{\lceil M_2/N_2 \rceil + 1} = \mathbf{E}_K = \mathbf{O}$. Let also $\mathbf{F}_1 = \cdots = \mathbf{F}_{\lceil M_2/N_2 \rceil} = 0$ and $\mathbf{F}_{\lceil M_2/N_2 \rceil + 1}, \ldots, \mathbf{F}_K$ be matrices such that $[\mathbf{F}_{\lceil M_1/N_2 \rceil + 1}^{\mathrm{T}}\ \ldots\ \mathbf{F}_K^{\mathrm{T}}]^{\mathrm{T}}$ has full column rank. Then from the first $\lceil M_2/N_2 \rceil$ and the last $\lceil N_1/M_1 \rceil$ equations in (5.6) it follows that

$$\mathbf{X}[\mathbf{E}_1\ \ldots\ \mathbf{E}_{\lceil M_2/N_2 \rceil}] = \mathbf{O} \qquad \text{and} \qquad \mathbf{O} = [\mathbf{F}_{\lceil M_1/N_2 \rceil + 1}^{\mathrm{T}}\ \ldots\ \mathbf{F}_K^{\mathrm{T}}]^{\mathrm{T}}\mathbf{Y},$$

respectively. Hence, $\mathbf{X} = \mathbf{O}$ and $\mathbf{Y} = \mathbf{O}$. ∎

Now we are ready to prove Theorem 2.3.

Proof of Theorem 2.3. We show that assumptions 1) and 2) in Theorem 2.1 hold for generic matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{D}_1, \ldots, \mathbf{D}_K$. The first two inequalities in (2.8) express the fact that the number of columns of $\mathbf{A}$ and $\mathbf{B}$ cannot exceed the number of rows. Since $\mathbf{A}$ and $\mathbf{B}$ are generic this means that assumption 1) in Theorem 2.1 holds. To prove that assumption 2) in Theorem 2.1 also holds it is sufficient to show that the subspace of solutions of (2.2) has dimension R. In the remaining part of the proof we show that for generic matrices $\mathbf{D}_1, \ldots, \mathbf{D}_K$ the subspace of solutions of (2.2) is spanned by $(\mathbf{P}_{I,1}, \mathbf{P}_{J,1}), \ldots, (\mathbf{P}_{I,R}, \mathbf{P}_{J,R})$, where the projectors $\mathbf{P}_{I,r}$, $\mathbf{P}_{J,r}$ are defined in (2.3).

Let $\mathbf{X} = (\mathbf{X}_{ij})_{i,j=1}^{R}$ and $\mathbf{Y} = (\mathbf{Y}_{ij})_{i,j=1}^{R}$ consist of the blocks $\mathbf{X}_{ij} \in \mathbb{F}^{M_i \times M_j}$ and $\mathbf{Y}_{ij} \in \mathbb{F}^{N_i \times N_j}$, respectively. Since the matrices $\mathbf{D}_k$ are block-diagonal, we can rewrite (2.2) in terms of the blocks $\mathbf{X}_{ij}$ and $\mathbf{Y}_{ij}$ as

$$\mathbf{X}_{ij}\mathbf{D}_{j,k} = \mathbf{D}_{i,k}\mathbf{Y}_{ij}, \qquad k = 1, \ldots, K, \tag{5.7}$$

where $i, j \in \{1, \ldots, R\}$. Let the subsets

$$\Omega_{ij},\ \Omega_i \subseteq \left(\mathbb{F}^{M_1 \times N_1} \times \cdots \times \mathbb{F}^{M_R \times N_R}\right) \times \cdots \times \left(\mathbb{F}^{M_1 \times N_1} \times \cdots \times \mathbb{F}^{M_R \times N_R}\right)$$

consist of the elements $(\mathbf{D}_{1,1}, \ldots, \mathbf{D}_{R,1}, \ldots, \mathbf{D}_{1,K}, \ldots, \mathbf{D}_{R,K})$ and be defined as follows: the elements in Ω_{ij} $(i \neq j)$ are such that (5.7) has only the zero solution and the elements in Ω_i are such that (5.7) has only the trivial solution, i.e., $\mathbf{X}_{ii} = c_i\mathbf{I}_{M_i}$ and $\mathbf{Y}_{ii} = c_i\mathbf{I}_{N_i}$ for some $c_i \in \mathbb{F}$.

To prove that $(\mathbf{X}, \mathbf{Y})$ belongs to the linear subspace spanned by $(\mathbf{P}_{I,1}, \mathbf{P}_{J,1}), \ldots, (\mathbf{P}_{I,R}, \mathbf{P}_{J,R})$ we have to show that $(\cap\Omega_{ij}) \bigcap (\cap\Omega_i)$ is a set of full measure. Since the intersection of a finite number of sets of full measure is also a set of full measure it is sufficient to prove that all sets in the intersection have full measure. This can be done as follows.

1) Let $i \neq j$. W.l.o.g. we can assume that $i = 1$ and $j = 2$. Then, by (5.7),

$$\mathbf{X}_{12}\mathbf{D}_{2,k} = \mathbf{D}_{1,k}\mathbf{Y}_{12}, \qquad k = 1, \ldots, K. \tag{5.8}$$

It is clear that system (5.8) has only the zero solution if and only if the matrix

$$\mathbf{M} = \begin{bmatrix} \mathbf{D}_{2,1}^{\mathrm{T}} \otimes \mathbf{I}_{M_1} & -\mathbf{I}_{N_2} \otimes \mathbf{D}_{1,1} \\ & \vdots \\ \mathbf{D}_{2,K}^{\mathrm{T}} \otimes \mathbf{I}_{M_1} & -\mathbf{I}_{N_2} \otimes \mathbf{D}_{1,K} \end{bmatrix}$$

has full column rank. Hence, from Lemma 5.2 and assumption (2.8) it follows that there exist matrices $\mathbf{D}_{1,1}, \ldots, \mathbf{D}_{1,K}$ and $\mathbf{D}_{2,1} \ldots \mathbf{D}_{2,K}$ such that $\mathbf{M}$ has full column rank. Now we assume that $\mathbf{D}_{1,1}, \ldots, \mathbf{D}_{1,K}$ and $\mathbf{D}_{2,1} \ldots \mathbf{D}_{2,K}$ in (5.8) are generic. Then, by [7, Lemma 6.3], $\mathbf{M}$ has full column rank, implying that system (5.8) has only the zero solution. Hence, by Fubini's theorem [10, Theorem C, p.148], Ω_{12} is a set of full measure.

2) Let $i = j$. W.l.o.g. we can assume that $i = 1$. We have to prove that the system

$$\mathbf{X}_{11}\mathbf{D}_{1,k} = \mathbf{D}_{1,k}\mathbf{Y}_{11}, \qquad k = 1, \ldots, K, \tag{5.9}$$

has only the trivial solution $\mathbf{X}_{11} = c\mathbf{I}_{M_1}$ and $\mathbf{Y}_{11} = c\mathbf{I}_{N_1}$, where $c \in \mathbb{F}$. It is clear that system (5.9) has only the trivial solution if and only if the dimension of the null space of the matrix

$$\mathbf{M} = \begin{bmatrix} \mathbf{D}_{1,1}^{\mathrm{T}} \otimes \mathbf{I}_{M_1} & -\mathbf{I}_{N_1} \otimes \mathbf{D}_{1,1} \\ & \vdots \\ \mathbf{D}_{1,K}^{\mathrm{T}} \otimes \mathbf{I}_{M_1} & -\mathbf{I}_{N_1} \otimes \mathbf{D}_{1,1} \end{bmatrix}$$

is one. Hence, from Lemma 5.1 and assumption (2.8) it follows that there exist matrices $\mathbf{D}_{1,1}, \ldots, \mathbf{D}_{1,K}$ such that $M_1^2 + N_1^2 - 1$ columns of $\mathbf{M}$ are linearly independent. Now we assume that $\mathbf{D}_{1,1}, \ldots, \mathbf{D}_{1,K}$ are generic. Then, by [7, Lemma 6.3], the same $M_1^2 + N_1^2 - 1$ columns of $\mathbf{M}$ are linearly independent, implying that system (5.9) has only the trivial solution. Hence, by Fubini's theorem [10, Theorem C, p.148], Ω_{11} is a set of full measure. ∎

Proof of Theorem 2.8. We show that assumptions 1) and 2) in Theorem 2.1 hold for generic matrix $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{D}_1, \ldots, \mathbf{D}_K$, where $\mathbf{B} = \mathbf{A}$ or $\mathbf{B} = \mathbf{A}^*$ and $\mathbf{D}_1, \ldots, \mathbf{D}_K$ may be symmetric, Hermitian or unconstrained.

1) Since $\mathbf{A}$ is generic and $I \geq \sum N_r$, it follows that $\mathbf{A}$ and $\mathbf{B}$ has full column rank.

2) WLOG we assume that $K = 3$. We proceed as in the proof of Theorem 2.3. First we show that the subspace of the solutions of (2.2) has dimension R for a specific choice of $\mathbf{D}_1$, $\mathbf{D}_2$, and $\mathbf{D}_3$. Let $\mathbf{D}_1 = \mathbf{I}_{\sum N_r}$, $\mathbf{D}_2$ be a diagonal matrix with distinct real values on the main diagonal and let $\mathbf{D}_3 = \operatorname{blockdiag}(\mathbf{D}_{1,3}, \ldots, \mathbf{D}_{R,3})$, where the entries of $\mathbf{D}_{1,3}, \ldots, \mathbf{D}_{R,3}$ are nonzero. Then $\mathbf{X}\mathbf{D}_1 = \mathbf{D}_1\mathbf{Y}$ implies that $\mathbf{X} = \mathbf{Y}$. From $\mathbf{X}\mathbf{D}_2 = \mathbf{D}_2\mathbf{Y} = \mathbf{D}_2\mathbf{X}$, it follows that $\mathbf{X} = \operatorname{diag}(x_1, \ldots, x_{\sum N_r})$.

Finally, from $\operatorname{diag}(x_1, \ldots, x_{\sum N_r})\mathbf{D}_3 = \mathbf{X}\mathbf{D}_3 = \mathbf{D}_3\mathbf{Y} = \mathbf{D}_3 \operatorname{diag}(x_1, \ldots, x_{\sum N_r})$ it follows that $\mathbf{X} = \mathbf{Y} = \operatorname{blockdiag}(c_1\mathbf{I}_{N_1}, \ldots, c_R\mathbf{I}_{N_R})$ for some $c_1, \ldots, c_R \in \mathbb{F}$. Thus, assumption 2) in Theorem 2.1 holds for a specific choice of $\mathbf{D}_1$, $\mathbf{D}_2$, and $\mathbf{D}_3$. Hence some $2(\sum N_r)^2 - R$ columns of $\mathbf{M}_{\mathcal{D}}$ are linearly independent. From [7, Lemma 6.3] it follows that if $\mathbf{D}_1$, $\mathbf{D}_2$, and $\mathbf{D}_3$ are generic symmetric, Hermitian or unconstrained matrices, then the same $2(\sum N_r)^2 - R$ columns of $\mathbf{M}_{\mathcal{D}}$ are linearly independent, implying that $\dim \operatorname{Null}(\mathbf{M}_{\mathcal{D}}) = R$. ∎

6. Conclusion. COMMENT: TBD

REFERENCES

[1] Y. Cai and C. Liu, *An algebraic approach to nonorthogonal general joint block diagonalization*, SIAM J. Matrix Anal. Appl., 38(1) (2017), pp. 50–71.

[2] G. Chabriel, M. Kleinsteuber, E. Moreau, H. Shen, P. Tichavsky, and A. Yeredor, *Joint matrices decompositions and blind source separation: A survey of methods, identification, and applications*, IEEE Signal Processing Magazine, 31 (2014), pp. 34–43.

[3] L. De Lathauwer, *Decompositions of a higher-order tensor in block terms — Part II: Definitions and uniqueness*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1033–1066.

[4] L. De Lathauwer, *Blind separation of exponential polynomials and the decomposition of a tensor in rank-$(L_r, L_r, 1)$ terms*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 1451–1474.

[5] L. De Lathauwer, B. De Moor, and J. Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278.

[6] L. De Lathauwer and D. Nion, *Decompositions of a higher-order tensor in block terms — Part III: Alternating least squares algorithms*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1067–1083.

[7] I. Domanov and L. De Lathauwer, *On the uniqueness of the canonical polyadic decomposition of third-order tensors — Part II: Overall uniqueness*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 876–903.

[8] I. Domanov and L. De Lathauwer, *Canonical polyadic decomposition of third-order tensors: Reduction to generalized eigenvalue decomposition*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 636–660, https://doi.org/10.1137/130916084.

[9] E. Evert, M. Vandecappelle, and L. De Lathauwer, *A recursive eigenspace computation for the canonical polyadic decomposition*, Tech. Report 20-80, ESAT-STADIUS, KU Leuven, Leuven, Belgium, 2020.

[10] P. R. Halmos, *Measure theory*, Springer-Verlag, New-York, 1974.

[11] S. E. Leurgans, R. T. Ross, and R. B. Abel, *A decomposition for three-way arrays*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 1064–1083, https://doi.org/10.1137/0614071, http://dx.doi.org/10.1137/0614071.

[12] T. Maehara and K. Murota, *Algorithm for error-controlled simultaneous block-diagonalization of matrices*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 605–620.

[13] Y. Ming, *On partial and generic uniqueness of block term tensor decompositions*, Annali Dell'Universita'Di Ferrara, 60 (2014), pp. 465–493.

[14] L. Sorber, M. Van Barel, and L. De Lathauwer, *Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-$(L_r, L_r, 1)$ terms, and a new generalization*, SIAM J. Optim., 23 (2013), pp. 695–720.

[15] L. Sorber, M. Van Barel, and L. De Lathauwer, *Structured data fusion*, IEEE J. Sel. Topics Signal Process., 9 (2015), pp. 586–600, https://doi.org/10.1109/JSTSP.2015.2400415.

[16] M. Sørensen and L. De Lathauwer, *Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank-$(L_{r_n}, L_{r_n}, 1)$ terms—Part I: Uniqueness*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 496–522.

[17] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, *Tensorlab 3.0*, 3 2016. Available online at https://www.tensorlab.net.