MULTILINEAR SINGULAR VALUE DECOMPOSITION BASED COMPLETION WITH FIBERS OBSERVED IN A SINGLE MODE*

MIKAEL SØRENSEN[§†], STIJN HENDRIKX[§‡], AND LIEVEN DE LATHAUWER[‡]

Abstract. The multilinear singular value decomposition (MLSVD) extends the properties of the singular value decomposition to the tensor case and is a basic tool for working with tensors that are (approximately) of low multilinear rank. In the context of incomplete data, low multilinear rank tensor completion algorithms are usually optimization-based. In this work, we introduce an algorithm for computing the MLSVD of tensors with fibers that are observed along a single mode. This is important, since in real-life applications it often happens that fibers can be acquired significantly more easily along one mode than along other modes. In addition, this manner of observation allows one to formulate an algebraic algorithm, i.e., involving only standard linear algebra operations. The algorithm is guaranteed to exactly recover a low multilinear rank tensor in the noiseless case if some very reasonable conditions are satisfied. Part of the work consists of the derivation of algebraic algorithms for completing a matrix that is observed through a set of fully available submatrices. Experiments with synthetic data show that the algorithm comes close to algorithms from the literature in terms of accuracy while requiring far less computation time. As a possible use, we obtain a "completion" counterpart of initializing optimization-based low multilinear rank approximation algorithms with a truncated MLSVD. In a completion experiment with traffic speed data, our algorithm achieves a level of accuracy that is comparable to baseline algorithms.

Key words. tensor, multilinear singular value decomposition, structured matrix, completion, missing data.

1. Introduction.

The multilinear singular value decomposition (MLSVD) is a fundamental tensor tool for uses such as higher-order principal component analysis, orthogonal compression or signal subspace estimation. The MLSVD has been used in fields such as machine learning, signal processing and scientific computing [39, 14, 27, 30, 31, 17]. It is also often used in a preprocessing step to speed up the computation of other tensor decompositions [2].

In tensor completion, the goal is to fill in the missing entries of an incomplete tensor, assuming that the tensor is (approximately) low rank. Datasets can be incomplete for various reasons. For example, a matrix containing movie ratings by different people is incomplete since one cannot have watched all movies [4]. Alternatively, a very large dataset might be considered to be incomplete because using all entries might be too computationally expensive [46]. In both cases, filling in the missing entries is very useful. In the former case, it allows one to make predictions and in the

*Submitted to the editors DATE.

This work is an extension of the conference paper [21]. The present paper includes the complete proof of the uniqueness conditions for computing the MLSVD for a fiber-wise observed tensor. It also includes two extra variants of the algebraic matrix completion algorithm. The algorithm is evaluated more thoroughly in additional experiments with synthetic data and an experiment with real traffic speed data.

Funding: Research supported by: (1) Flemish Government: This research received funding from the Flemish Government (AI Research Program). Lieven De Lathauwer and Stijn Hendrikx are affiliated to Leuven.AI - KU Leuven institute for AI, B-3000, Leuven, Belgium. (2) KU Leuven Internal Funds: iBOF/23/064, C14/22/096, IDN/19/014.

§Authors sharing first authorship.

†At the time of the research, Mikael Sørensen was with Dynamical Systems, Signal Processing and Data Analytics (STADIUS), Department of Electrical Engineering (ESAT), KU Leuven, Leuven, Belgium and Group Science, Engineering and Technology, KU Leuven Kulak, Kortrijk, Belgium. (mr.mikael.sorensen@gmail.com).

‡Dynamical Systems, Signal Processing and Data Analytics (STADIUS), Department of Electrical Engineering (ESAT), KU Leuven, Leuven, Belgium. Group Science, Engineering and Technology, KU Leuven Kulak, Kortrijk, Belgium (Stijn.Hendrikx@kuleuven.be. Lieven.DeLathauwer@kuleuven.be).

latter case it offers a compact model at a feasible cost.

In literature, there are two main approaches to tensor completion. The first one minimizes the rank of the completed tensor while keeping the error between the completed tensor and the incomplete data below some tolerance. Since minimizing the rank of a matrix or tensor is a difficult and non-convex problem, convex relaxations of the rank are used instead [18, 40, 10]. In both the matrix and tensor case, the complete dataset can be exactly recovered with high probability by minimizing a relaxed problem under some assumptions regarding incoherence [9, 22, 34]. The second approach works the opposite way, namely by minimizing the error between the incomplete data and a tensor (decomposition) with a fixed rank [1, 29, 47]. This approach also has theoretical guarantees in the matrix case, namely that any local minimum of the non-convex cost function is also a global minimum with high probability [19]. However, to the best of the authors' knowledge, there are no such theoretical guarantees for the second approach in the tensor case.

The algorithm introduced in this work uses the second tensor completion approach and approximates an incomplete tensor with a tensor of fixed multilinear rank. Algorithms of this type are usually optimization-based. Some of these algorithms use Riemannian optimization [29, 26]. Alternatively, there are algorithms that optimize over the variables of a Tucker decomposition, of which the MLSVD is a specific case with additional constraints [25, 16]. For a more general overview of tensor completion algorithms, we refer the reader to [43]. Properties such as existence, uniqueness and recoverability in the absence of noise are studied using algebraic geometry in [6].

An important aspect of optimization algorithms is the selection of the starting point or initialization. In the case of complete tensors, a starting point may be obtained in an algebraic manner, that is, using only standard linear algebra operations. For example, the canonical polyadic decomposition (CPD) can be computed using a generalized eigenvalue decomposition and the basic MLSVD algorithm consists of a few matrix singular value decompositions (SVDs). Such algebraic algorithms are not available for general incomplete tensors. A new point of view was developed in [44] for a specific type of incomplete tensors, in which the observed part of the tensor consists of entire vectors or fibers along a single mode (as opposed to consisting of entries that are spread randomly throughout the tensor). It turned out that in such cases, the CPD can again be computed in an algebraic manner. The low multilinear rank counterpart, i.e., an algebraic algorithm for computing the MLSVD of a fiber-wise observed tensor and the corresponding uniqueness conditions are introduced in the present paper. In the noiseless case, the algorithm exactly recovers the original low multilinear rank tensor if these conditions are satisfied.

As mentioned above, we assume that fibers are observed along a single mode. This is an important difference with the popular technique of cross approximation (also known as skeleton approximation and CUR decomposition), in the sense that the latter obtains the decomposition from a subset of fibers observed across *all* modes [8, 33, 7, 20]. Observing fibers along one mode only makes the completion task mathematically more challenging.

In real-life applications, it often happens that fibers can be acquired significantly more easily along one mode than on other modes. By way of example, think of the acquisition of a chemical parameter as a function of time and for different concentrations of reagents. Obtaining samples along the temporal mode may be much easier and cheaper than varying the concentrations, as the latter may require conducting new experiments and human intervention. Along some modes fibers may not even exist. A popular example of the latter are collections of fluorescence excitation-emission

matrices (EEMs) stacked in a third-order tensor for different samples [42]. Emission happens at longer wavelengths than excitation, so that EEMs are triangular matrices by nature, while Rayleigh and Raman scatter happen along straight lines that traverse the EEM landscape. On the other hand, fibers along the sample mode are good candidates to be fully observed. The traffic speed data [11] used in our final experiment form another example of real-life data exhibiting a single-mode fiber-wise observation pattern.

To compute the factor matrices (except the factor matrix of the mode along which the fibers are observed), we derive algorithms for computing a column-wise orthonormal basis for the column space of an incomplete matrix. Additionally, we derive conditions under which this column space can be uniquely retrieved. The completion algorithms and corresponding uniqueness conditions introduced in this work are not specific for computing an MLSVD and can already be used for matrix completion as such.

The main contributions of this paper can be summarized as follows:

- We present algorithms for matrix completion that require only standard numerical linear algebra. Algorithm 4.1 will be of the "noise subspace" type, while Algorithm 4.2 will be of the "signal subspace" type. That is, for the user's convenience, we provide matrix completion algorithms for both types of approaches that are common in array processing. In numerical linear algebra, the two types correspond to working via the estimate of the nullspace of a matrix or via an estimate of its column/row space. The two can be traded in for each other, but one may be cheaper to obtain than the other, depending on the size and rank of the matrix at hand. We will make a comparison in the Discussion at the end of subsection 4.2. We also provide a fast version of the "signal subspace" approach (which on the other hand squares the condition number and is hence numerically less accurate for ill-conditioned matrices).
- The matrix completion algorithms are guaranteed to find the solution under mild conditions in the noise-free case. "Guaranteed" implies that there exist no solutions other than the one found by the algorithm, i.e. the solution is unique. The conditions take the form of easy-to-check algebraic conditions on the rank (or equivalently, on the dimension of the null space) of a matrix derived from fully observed submatrices of the matrix $\mathbf{M}$ that needs to be completed, namely matrix $\boldsymbol{\Upsilon}, \boldsymbol{\Xi}_{\mathbf{U}}$ and $\boldsymbol{\nabla}$ in Theorems 3.1, 3.2 and 3.4, respectively. The conditions basically balance how much of the matrix $\mathbf{M}$ has been observed versus its richness of content (i.e. its rank), in a mild manner. The mildness can be observed from the generic counts at the end of section 3.
- We use the matrix completion algorithms as building blocks to derive an algorithm for tensor completion that exploits low multilinear rank. Our Algorithm 5.1 complements the algorithm for low-rank based tensor completion in [44]. That is, the present paper is the MLSVD counterpart of the CPD-based approach in [44]. Perhaps surprisingly, both tensor approaches require only standard numerical linear algebra and they are guaranteed to find the solution under mild conditions in the noise-free case. Both algorithms are the first of their kind.
- In our approach, the tensor needs only be observed through fibers along a single mode. This is of great interest in applications, given that not all modes may be equally convenient for fiber observation, and constitutes an important advantage over (pseudo-skeleton) techniques that require the observation of fibers in all modes. On the other hand, we do need full fibers for our lin-

ear algebraic approach, while optimization-based algorithms (which are not guaranteed to find the solution in the noise-free case) allow situations where only individual entries are observed.
- In the computation of tensor decompositions in noisy settings, it is common practice to first find an algebraic approximation of the solution, which is then refined by numerical optimization. Our paper extends this common practice to tensors that are observed only partially.
- The performance of our algorithm will be evaluated in experiments with synthetic and real-life data. The experiments will show that our algorithm achieves a level of accuracy that comes close to that achieved by optimization-based tensor completion algorithms from the literature and that it is computationally much less expensive. In combination with an optimization-based refinement stage, the optimal accuracy is obtained at a fraction of the computational cost.

In the remainder of this section, the required tensor notation is introduced. Section 2 concerns the basics of the MLSVD and introduces the fiber-wise observation pattern. Next, uniqueness conditions for matrix completion are derived in section 3. In section 4, algebraic matrix completion algorithms are introduced. The algorithm for computing the overall MLSVD of a fiber-wise observed tensor and the corresponding uniqueness conditions are introduced in section 5. Last, the performance of the algorithm is evaluated in experiments with synthetic data and real-life traffic speed data.

Notation. Vectors, matrices and tensors are denoted by lower case boldface ($\mathbf{a}$), upper case boldface ($\mathbf{A}$) and upper case calligraphic letters ($\mathcal{A}$), respectively. The dimensions of a third-order tensor are denoted by $I \times J \times K$ and N denotes the order of a tensor. For simplicity, the equations in this work are given for third-order tensors. Selecting a part of a tensor is denoted by indices and colons, in which a colon corresponds to an index that has not been fixed. For example, $\mathbf{A}_{:j:} \in \mathbb{C}^{I\times K}$ corresponds to the jth mode-2 slice of $\mathcal{A} \in \mathbb{C}^{I\times J\times K}$. A permuted and reshaped version of a tensor is denoted by, for example, $\mathbf{A}_{[2;1,3]} \in \mathbb{C}^{J\times IK}$, in which the indices correspond to the modes of the original tensor $\mathcal{A}$ and the ";" separates the modes of the resulting tensor. Within one mode of the resulting tensor, the order of the indices determines how the entries are stacked. Tensor entry a_{ijk} is located at position $(j, (i-1)I+k)$ in $\mathbf{A}_{[2;1,3]}$. Matrix $\mathbf{A}_{[2;1,3]}$ has the mode-2 vectors of $\mathcal{A}$ as its columns and is commonly referred to as the mode-2 unfolding.

The conjugate, transpose, conjugate-transpose, Frobenius norm, inverse, pseudoinverse, rank, range and kernel of a matrix $\mathbf{A}$ are denoted by $\mathbf{A}^*$, $\mathbf{A}^{\mathrm{T}}$, $\mathbf{A}^{\mathrm{H}}$, $\|\mathbf{A}\|_{\mathrm{F}}$, $\mathbf{A}^{-1}$, $\mathbf{A}^{\dagger}$, $\operatorname{rank}(\mathbf{A})$, $\operatorname{range}(\mathbf{A})$, $\ker(\mathbf{A})$, respectively. The dimension of a subspace S is denoted by $\dim(S)$. The symbols $\otimes$ and $*$ denote the Kronecker and element-wise/Hadamard product.

The mode-n product $\cdot_n$ multiplies $\mathbf{B}$ with all mode-n vectors in $\mathcal{A}$. For example for $n=2$, an entry of $\mathcal{X} \in \mathbb{C}^{I\times L\times K} = \mathcal{A} \cdot_2 \mathbf{B}$, with $\mathbf{B} \in \mathbb{C}^{L\times J}$ equals $x_{ilk} = \sum_{j=1}^{J} a_{ijk} b_{lj}$, which means that $\mathbf{X}_{[2;1,3]} = \mathbf{B}\mathbf{A}_{[2;1,3]}$. The element-wise "logical or" operation of two $(0,1)$-binary matrices $\mathbf{A}$ and $\mathbf{B}$ is denoted by $\mathbf{A} \vee \mathbf{B}$, e.g., $\left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right] \vee \left[\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right] = \left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]$. Furthermore, let $\operatorname{vec}(\mathbf{A})$ denote the vector obtained by stacking the columns of $\mathbf{A} \in \mathbb{C}^{I\times J}$ into a column vector, i.e., $\operatorname{vec}(\mathbf{A}) = [\mathbf{a}_1^{\mathrm{T}}, \ldots, \mathbf{a}_J^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{IJ}$. The $I \times I$ identity matrix is denoted by $\mathbf{I}_R \in \mathbb{C}^{I\times I}$. The all-ones vector is denoted by $\mathbf{1}_R = [1, \ldots, 1]^{\mathrm{T}} \in \mathbb{C}^R$. Let $\mathbf{e}_i^{(I)} \in \{0,1\}^I$ denote the I-dimensional unit vector with a one at position i and zeros elsewhere. A set is denoted by a lower case Greek letter. The cardinality

of a set α and the complement of a set are respectively denoted by $\mathrm{card}\,(\alpha)$ and α^{c}. The union of a sequence of sets is denoted by $\cup_{\ell\in\{1,\ldots,L\}}\alpha_\ell$. The matrix that selects the rows of a $(J\times K)$ matrix with indices in the set $\alpha_\ell=\{j_1,\ldots,j_{J_\ell}\}\subseteq\{1,\ldots,J\}$ through left multiplication is

$$\mathbf{S}_{\mathrm{row}}^{(\ell)}\in\{0,1\}^{J_\ell\times J}=\begin{bmatrix}\mathbf{e}_{j_1}^{(J)} & \ldots & \mathbf{e}_{j_{J_\ell}}^{(J)}\end{bmatrix}^{\mathrm{T}}.$$

The matrix that selects the columns of a $(J\times K)$ matrix with indices in the set $\beta_\ell=\{k_1,\ldots,k_{K_\ell}\}\subseteq\{1,\ldots,K\}$ through right multiplication is

$$\mathbf{S}_{\mathrm{col}}^{(\ell)}\in\{0,1\}^{K\times K_\ell}=\begin{bmatrix}\mathbf{e}_{k_1}^{(K)} & \ldots & \mathbf{e}_{k_{K_\ell}}^{(K)}\end{bmatrix}.$$

For example, selecting the submatrix of $\mathbf{A}$ at the intersection of the row indices $\alpha_1=\{1,3\}$ and column indices $\beta_1=\{2,4\}$ corresponds to

$$\mathbf{S}_{\mathrm{row}}^{(1)}\mathbf{A}\mathbf{S}_{\mathrm{col}}^{(1)}=\begin{bmatrix}1&0&0&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}1&2&3&4\\5&6&7&8\\9&10&11&12\\13&14&15&16\end{bmatrix}\begin{bmatrix}0&0\\1&0\\0&0\\0&1\end{bmatrix}=\begin{bmatrix}2&4\\10&12\end{bmatrix}.$$

2. Multilinear singular value decomposition. The MLSVD is a special case of the Tucker decomposition, which factorizes a third-order tensor as

$$\mathcal{X}=\mathcal{G}\cdot_1\mathbf{A}\cdot_2\mathbf{B}\cdot_3\mathbf{C}, \tag{2.1}$$

with $\mathbf{A}\in\mathbb{C}^{I\times R_1}$, $\mathbf{B}\in\mathbb{C}^{J\times R_2}$, $\mathbf{C}\in\mathbb{C}^{K\times R_3}$, $\mathcal{G}\in\mathbb{C}^{R_1\times R_2\times R_3}$. The minimal tuple (R_1,R_2,R_3) is the multilinear rank of $\mathcal{X}$. The additional constraints that the MLSVD imposes on top of the Tucker decomposition are that the factor matrices $\mathbf{A},\mathbf{B}$ and $\mathbf{C}$ are column-wise orthonormal, like $\mathbf{U}$ and $\mathbf{V}$ in the matrix SVD $\mathbf{X}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}$, and that the core tensor $\mathcal{G}$ is "all-orthogonal" and "ordered". The mode-n slices of an "all-orthogonal" tensor are mutually orthogonal for $n=1,\ldots,N$. The "all-orthogonality" of the core can be seen as a generalization of $\mathbf{\Sigma}$ in the matrix SVD being diagonal, since both the rows and columns of a diagonal matrix are mutually orthogonal. Along each mode n of the core, the mode-n slices are ordered such that their Frobenius norms, which correspond to the multilinear singular values, decrease monotonically. Therefore, a low multilinear rank approximation of $\mathcal{X}$ can be obtained by truncating the core. In real-life applications, the multilinear singular values of the data tensor often exhibit a decaying trend, meaning that the tensor can be well approximated by a low multilinear rank model. Unlike in the matrix case, this approximation is not necessarily optimal in the least squares sense. However, the approximation error is bounded by the multilinear singular values that were omitted by truncating and such a low multilinear rank approximation might be sufficient in practice.

The columns of the factor matrices $\mathbf{A},\mathbf{B}$ and $\mathbf{C}$ form bases for the subspaces spanned by the mode-1, 2 and 3 vectors of $\mathcal{X}$ respectively. A straightforward way to compute these factor matrices then is to obtain them from the dominant left singular vectors of their corresponding matrix unfolding. This is the standard MLSVD algorithm and is described in Algorithm 2.1, in which $\mathbf{U}^{(1)},\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$ correspond to $\mathbf{A},\mathbf{B}$ and $\mathbf{C}$ respectively. To reduce the computational cost, the tensor can be compressed by a factor matrix as soon as the latter has been computed [45].

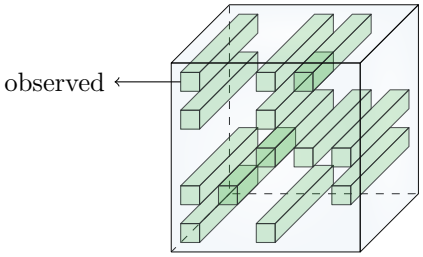
In this work, we develop an algorithm to compute the MLSVD of an incomplete tensor which consists of fibers observed along the last mode. The observed fibers are

Algorithm 2.1 The basic MLSVD method.

Input: Tensor $\mathcal{X}$
Output: Factor matrices $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{U}^{(3)}$, core $\mathcal{G}$
for $n = 1, 2, 3$ **do**
  Compute truncated SVD for mode n: $\mathbf{X}_{[n;\mathcal{I}]} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}$ where $\mathcal{I} = \{1, 2, 3\} \setminus \{n\}$
  $R_n \leftarrow \operatorname{rank}\left(\mathbf{X}_{[n;\mathcal{I}]}\right)$, $\mathbf{U}^{(n)} \leftarrow \begin{bmatrix}\mathbf{u}_1 & \dots & \mathbf{u}_{R_n}\end{bmatrix}$
end for
Compute core $\mathcal{G} = \mathcal{X} \cdot_1 \mathbf{U}^{(1)^{\mathrm{H}}} \cdot_2 \mathbf{U}^{(2)^{\mathrm{H}}} \cdot_3 \mathbf{U}^{(3)^{\mathrm{H}}}$

indicated by a binary mask $\mathbf{W} \in \{0, 1\}^{I \times J}$, meaning that $\mathbf{x}_{ij:}$ has been observed if $w_{ij} = 1$. For example, a mask and the corresponding fiber-wise observed tensor could look like in Figure 2.1. Since complete mode-3 vectors are observed, computing the third factor matrix is as straightforward as in the complete tensor case. Namely, it can be obtained as the dominant left singular vectors of a compact version of the mode-3 unfolding of $\mathcal{X}$, namely $\mathbf{X}_{[3;2,1]}\mathbf{S}_{\text{fib}}^{\mathrm{T}}$, in which all unobserved fibers have been dropped. Right multiplication with $\mathbf{S}_{\text{fib}}^{\mathrm{T}}$ selects the columns with indices in the set $\{i \in \{1, \dots, IJ\} | w_i = 1\}$, with $\mathbf{w} = \operatorname{vec}(\mathbf{W})$.

Due to the pattern of observed fibers, this straightforward way of computing the factor matrices no longer works for modes 1 and 2. For example, the mode-2 unfolding of the tensor in Figure 2.1 exhibits the pattern shown in Figure 2.2. This pattern consists of submatrices or "grids" of the unfolding that have been completely observed. In the following section, uniqueness conditions are derived that guarantee that the column space of an incomplete matrix can be computed and algebraic algorithms for this purpose are introduced in section 4. These algorithms are used for the computation of the mode-1 and 2 factor matrices of the MLSVD of a fiber-wise observed tensor in section 5.



$$\mathbf{W} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

FIG. 2.1. *Example of a fiber-wise observed tensor and the corresponding mask.*

3. Matrix completion: algebraic conditions for uniqueness. For a rank-R matrix $\mathbf{M} \in \mathbb{C}^{J \times F} = \mathbf{B}\mathbf{F}^{\mathrm{T}}$ with $\mathbf{B} \in \mathbb{C}^{J \times R}$ and $\mathbf{F} \in \mathbb{C}^{F \times R}$, we have that $(\mathbf{B}\mathbf{Z})\left(\mathbf{Z}^{-1}\mathbf{F}^{\mathrm{T}}\right)$ is an alternative factorization for any nonsingular matrix $\mathbf{Z} \in \mathbb{C}^{R \times R}$. Therefore, $\mathbf{B}$ and $\mathbf{F}$ can only be recovered up to consistent change-of-basis transformations. In this section, we will only be concerned with the unique retrieval of the subspace range($\mathbf{B}$). In the setting of matrix completion, we will consider the subspace range($\mathbf{B}$) to be unique if every pair $(\overline{\mathbf{B}}, \overline{\mathbf{F}})$ with property $\mathbf{Y} = \mathbf{V} * (\overline{\mathbf{B}}\overline{\mathbf{F}}^{\mathrm{T}})$ also satisfies the relation range($\mathbf{B}$) = range($\overline{\mathbf{B}}$). The identifiability of the subspace range($\mathbf{B}$) is a necessary, but not sufficient, condition for the identifiability of the low-rank matrix completion of $\mathbf{M}$ [35]. The latter problem, which requires that the missing entries of $\mathbf{M}$ can be exactly recovered in the noiseless case, is therefore more general than the

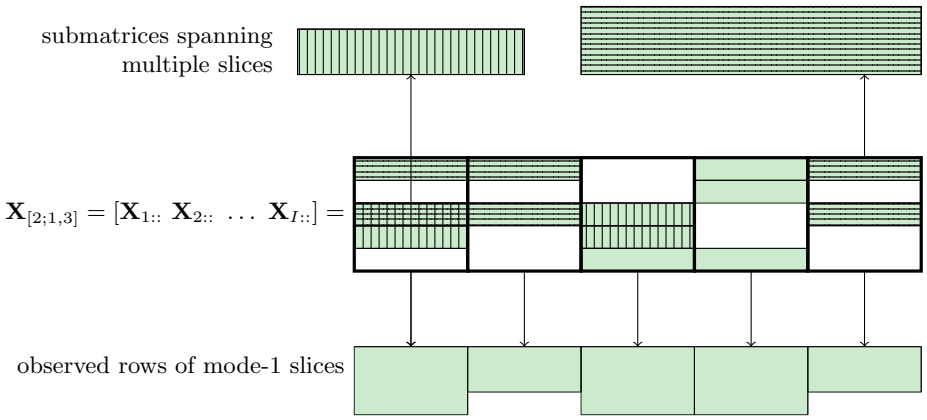


FIG. 2.2. *The mode-2 unfolding of the fiber-wise observed tensor in Figure 2.1 consists of fully observed submatrices that are potentially overlapping. There are five fully observed submatrices that are formed by the observed rows of the mode-1 slices* $\mathbf{X}_{i::}$ *for* $i = 1, \ldots, 5$. *There are also two larger fully observed submatrices that span across multiple slices, namely the submatrices marked by horizontal and vertical lines.*

former problem that is considered in this section.

For the completion of low-rank matrices, two types of identifiability conditions can be distinguished, namely statistical and deterministic conditions. Statistical conditions provide a bound on the number of randomly sampled observed entries of a low-rank matrix, such that it can be completed with a certain (high) probability [10]. On the other hand, deterministic conditions guarantee that a low-rank matrix can be completed for a specific mask, which is the pattern of observed entries. The conditions introduced in this work will be of the deterministic type. In literature, there are graph-based deterministic conditions that guarantee that the matrix can be completed if a bipartite graph, which has the mask as its biadjacency matrix, satisfies some constraints. In [28], the constraint that guarantees that a rank-R matrix can be completed is that this graph must be "R-closable". Similar graph-based matrix completion identifiability conditions can be found in [5, 13]. Furthermore, there are deterministic conditions that characterize which observation patterns are completable [36] or that follow from the connection between matrix completion and rigidity theory [41]. The former conditions are extended to low multilinear rank completion in [3].

Deterministic conditions in literature generally rely on assumptions on the low-rank matrix $\mathbf{M}$, such as incoherence or the assumption that the matrices $\mathbf{B}$ and $\mathbf{F}$ are generic[1], so that identifiability only depends on the observation pattern. In our work, the identifiability conditions do not rely on such assumptions and therefore depend on both the observation pattern and the low-rank matrix $\mathbf{M}$ itself. The approach for checking whether our conditions are satisfied is also different from the literature, as it involves checking the rank of a certain contrived matrix. More specifically, we will check the dimension of the null space of the matrix $\boldsymbol{\Upsilon}$ in (3.6) in Theorem 3.1, the dimension of the null space of the matrix $\boldsymbol{\Xi}_{\mathbf{U}}$ in (3.11) in Theorem 3.2 and the rank of the matrix $\boldsymbol{\nabla}$ in (3.13) in Theorem 3.4. These conditions basically balance how much of the matrix has been observed versus its "richness of content" (i.e. its rank). A more restrictive, but easier-to-verify condition is derived from Theorem 3.4

[1]The entries of a generic matrix have been drawn from an absolutely continuous joint probability distribution.

as Corollary 3.5. For the latter, we also provide a generic variant as (3.22) that is very easy to check given the mask, rank and matrix dimensions. The uniqueness conditions are very mild and easy to satisfy in practice. Moreover, under the conditions presented in this section, factor matrices of the MLSVD of a fiber-wise observed tensor can be computed using the algorithms in section 4. However, the conditions can also be used for matrix completion problems as such.

3.1. Problem definition. Consider the factorization of the incomplete version, in which missing entries have been replaced by zeros, of a low-rank matrix $\mathbf{M} \in \mathbb{C}^{J\times F}$:

$$\mathbf{V} * \mathbf{M} = \mathbf{V} *(\mathbf{B}\mathbf{F}^{\mathrm{T}}) \in \mathbb{C}^{J\times F}, \tag{3.1}$$

where $\mathbf{B} \in \mathbb{C}^{J\times R}$ and $\mathbf{F} \in \mathbb{C}^{F\times R}$ are rank-R matrices with $R \leq \min(J, F)$ and $\mathbf{V} \in \{0, 1\}^{J\times F}$ is a binary mask with property $v_{jk} = 1$ when entry m_{jk} is observed and $v_{jk} = 0$ otherwise. In this section, we think of $\mathbf{M}$ as a general incomplete matrix. It does not necessarily correspond to a matrix unfolding of a fiber-wise observed tensor. Whereas the mask $\mathbf{W}$ from the previous section indicates observed mode-3 fibers in a tensor $\mathcal{X}$, the mask $\mathbf{V}$ indicates observed entries in a matrix $\mathbf{M}$. The mask $\mathbf{V}$ is expressed as

$$\mathbf{V} = \bigvee_{\ell=1}^{L} \mathbf{V}^{(\ell)} = \mathbf{V}^{(1)} \vee \cdots \vee \mathbf{V}^{(L)} \qquad \text{with } v_{ij}^{(\ell)} = \begin{cases} 1 & \text{if } i \in \alpha_\ell \text{ and } j \in \beta_\ell \\ 0 & \text{otherwise} \end{cases}, \tag{3.2}$$

with the sets $\alpha_\ell \subseteq \{1, \ldots, J\}$ (with $\operatorname{card}(\alpha_\ell) = J_\ell$) and $\beta_\ell \subseteq \{1, \ldots, F\}$ (with $\operatorname{card}(\beta_\ell) = F_\ell$) for $\ell = 1, \ldots, L$.

Due to the elementwise multiplication in (3.1), each $\mathbf{V}^{(\ell)}$ selects a submatrix of $\mathbf{M}$, namely the fully observed submatrix at the intersection of the rows with index in α_ℓ and the columns with index in β_ℓ. For example, for $\alpha_1 = \{2, 3, 5\}$ and $\beta_1 = \{1, 3\}$ we have that

$$\mathbf{V}^{(1)} * \mathbf{M} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \\ 21 & 22 & 23 & 24 & 25 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 6 & 0 & 8 & 0 & 0 \\ 11 & 0 & 13 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 21 & 0 & 23 & 0 & 0 \end{bmatrix},$$

for which the corresponding fully observed submatrix of $\mathbf{M}$ is

$$\mathbf{M}_{\text{obs}}^{(1)} = \begin{bmatrix} 6 & 8 \\ 11 & 13 \\ 21 & 23 \end{bmatrix}.$$

These fully observed submatrices of $\mathbf{M}$ can be expressed as

$$\mathbf{M}_{\text{obs}}^{(\ell)} = \mathbf{S}_{\text{row}}^{(\ell)} \mathbf{M} \mathbf{S}_{\text{col}}^{(\ell)} = \mathbf{B}_{\text{obs}}^{(\ell)} \mathbf{F}_{\text{obs}}^{(\ell)^{\mathrm{T}}} \in \mathbb{C}^{J_\ell \times F_\ell} \quad \text{for } \ell = 1, \ldots, L, \tag{3.3}$$

where $\mathbf{B}_{\text{obs}}^{(\ell)} = \mathbf{S}_{\text{row}}^{(\ell)} \mathbf{B} \in \mathbb{C}^{J_\ell \times R}$ and $\mathbf{F}_{\text{obs}}^{(\ell)} = \mathbf{S}_{\text{col}}^{(\ell)^{\mathrm{T}}} \mathbf{F} \in \mathbb{C}^{F_\ell \times R}$.

The expression for the mask in (3.2) is not restrictive, as any $\mathbf{V}$ can be written in this form. However, at this point we make the working assumption that $\mathbf{B}_{\text{obs}}^{(\ell)}$ and $\mathbf{F}_{\text{obs}}^{(\ell)}$ have full column rank for $\ell = 1, \ldots, L$, i.e., each fully observed submatrix $\mathbf{M}_{\text{obs}}^{(\ell)}$ has

rank R. Note that the working assumption implies that $\min(J_\ell, F_\ell) \geq R$, $1 \leq \ell \leq L$. In the specific case where $\mathbf{M}$ corresponds to a matrix unfolding of $\mathcal{X}$, $\mathbf{V}$ is related to $\mathbf{W}$ and will be structured similarly to the example in Figure 2.2. The mask $\mathbf{V}$ can also represent more general patterns that allow overlap between the columns of the fully observed submatrices.

3.2. Uniqueness conditions. We will now develop a condition that enables us to obtain the overall range($\mathbf{B}$) from the submatrix relations (3.3). Combination of the working assumption that the fully observed submatrices $\mathbf{M}_{\text{obs}}^{(\ell)}$ of $\mathbf{M}$ have rank R with (3.3) yields that range $\left(\mathbf{M}_{\text{obs}}^{(\ell)}\right) = \text{range}\left(\mathbf{B}_{\text{obs}}^{(\ell)}\right)$. As a result, the matrices $\mathbf{M}_{\text{obs}}^{(\ell)}$ and $\mathbf{B}_{\text{obs}}^{(\ell)}$ are equal up to basis transformation matrices $\mathbf{\Gamma}^{(\ell)}$:

$$\mathbf{M}_{\text{obs}}^{(\ell)}\mathbf{\Gamma}^{(\ell)} = \mathbf{S}_{\text{row}}^{(\ell)}\mathbf{B} = \mathbf{B}_{\text{obs}}^{(\ell)} \quad \text{for } \ell = 1, \ldots, L. \tag{3.4}$$

This is a set of linear systems in the a priori unknown matrices $\mathbf{\Gamma}^{(\ell)} \in \mathbb{C}^{F_\ell \times R}$ for $\ell = 1, \ldots, L$ and $\mathbf{B}$, all of which have full column rank. Note that a solution to (3.4) can only be unique up to a change-of-basis transformational ambiguity. To be more specific, two solutions $[{\mathbf{\Gamma}^{(1)}}^{\mathrm{T}}, \ldots, {\mathbf{\Gamma}^{(L)}}^{\mathrm{T}}, \mathbf{B}^{\mathrm{T}}]^{\mathrm{T}}$ and $[{\mathbf{\Delta}^{(1)}}^{\mathrm{T}}, \ldots, {\mathbf{\Delta}^{(L)}}^{\mathrm{T}}, \widehat{\mathbf{B}}^{\mathrm{T}}]^{\mathrm{T}}$ to (3.4) are equivalent if there exists a nonsingular matrix $\mathbf{N} \in \mathbb{C}^{R \times R}$ such that

$$\begin{bmatrix} {\mathbf{\Gamma}^{(1)}}^{\mathrm{T}} & \ldots & {\mathbf{\Gamma}^{(L)}}^{\mathrm{T}} & \mathbf{B}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} {\mathbf{\Delta}^{(1)}}^{\mathrm{T}} & \ldots & {\mathbf{\Delta}^{(L)}}^{\mathrm{T}} & \widehat{\mathbf{B}}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \mathbf{N}.$$

The linear system in (3.4) can be written in matrix form as

$$\begin{bmatrix} \mathbf{M}_{\text{obs}}^{(1)} & & & -\mathbf{S}_{\text{row}}^{(1)} \\ & \ddots & & \vdots \\ & & \mathbf{M}_{\text{obs}}^{(L)} & -\mathbf{S}_{\text{row}}^{(L)} \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma}^{(1)} \\ \vdots \\ \mathbf{\Gamma}^{(L)} \\ \mathbf{B} \end{bmatrix} = \mathbf{0}. \tag{3.5}$$

If the solution to this system of linear equations is unique up to the intrinsic change-of-basis transformation, then range($\mathbf{B}$) can be computed by solving it. However, more efficient algorithms can be developed by eliminating variables, which will be explained in subsection 4.1 in the context of algebraic matrix completion algorithms.

Note that, by the possibility of completion, the first matrix in (3.5) must have a nullspace of dimension at least R as the matrices $\mathbf{\Gamma}^{(\ell)}$ for $\ell = 1, \ldots, L$ and $\mathbf{B}$ all have R columns. We can establish uniqueness by guaranteeing that the nullspace does not have a dimension larger than R. This is done in Theorem 3.1 below, which makes use of a variant of the matrix in (3.5), namely matrix

$$\mathbf{\Upsilon} := \begin{bmatrix} \mathbf{U}_{\text{obs}}^{(1)} & & & -\mathbf{S}_{\text{row}}^{(1)} \\ & \ddots & & \vdots \\ & & \mathbf{U}_{\text{obs}}^{(L)} & -\mathbf{S}_{\text{row}}^{(L)} \end{bmatrix}, \tag{3.6}$$

where the columns of $\mathbf{U}_{\text{obs}}^{(\ell)} \in \mathbb{C}^{J_\ell \times R}$ form a basis for range($\mathbf{M}_{\text{obs}}^{(\ell)}$).

THEOREM 3.1. *Let $\mathbf{M}_{obs}^{(\ell)} \in \mathbb{C}^{J_\ell \times F_\ell}$ for $\ell = 1, \ldots, L$ be fully observed submatrices of rank R of a matrix $\mathbf{M}$ of rank R. If*

$$dim(ker(\mathbf{\Upsilon})) = R \qquad \text{with } \mathbf{\Upsilon} \text{ as in (3.6)}, \tag{3.7}$$

then the overall column space range $(\mathbf{M})$ *is uniquely determined and can be explicitly obtained from solving the set of linear equations* (3.4).

Proof. Given (3.3) and that $\mathbf{M}_{\text{obs}}^{(\ell)}$ for $\ell = 1, \ldots, L$ has rank R, $\mathbf{F}_{\text{obs}}^{(\ell)}$ and $\mathbf{B}_{\text{obs}}^{(\ell)}$ have full column rank. In turn, we can from (3.3) obtain a matrix $\mathbf{U}_{\text{obs}}^{(\ell)}$ with property $\text{range}(\mathbf{U}_{\text{obs}}^{(\ell)}) = \text{range}(\mathbf{M}_{\text{obs}}^{(\ell)}) = \text{range}(\mathbf{B}_{\text{obs}}^{(\ell)})$. This also means that there exists a nonsingular matrix $\mathbf{G}_{\text{obs}}^{(\ell)} \in \mathbb{C}^{R\times R}$ such that $\mathbf{B}_{\text{obs}}^{(\ell)} = \mathbf{U}_{\text{obs}}^{(\ell)}\mathbf{G}_{\text{obs}}^{(\ell)}$. Inserting the latter factorization into (3.3) yields $\mathbf{M}_{\text{obs}}^{(\ell)} = \mathbf{B}_{\text{obs}}^{(\ell)}{\mathbf{F}_{\text{obs}}^{(\ell)}}^{\text{T}} = \mathbf{U}_{\text{obs}}^{(\ell)}\mathbf{G}_{\text{obs}}^{(\ell)}{\mathbf{F}_{\text{obs}}^{(\ell)}}^{\text{T}}$. Inserting the latter factorization of $\mathbf{M}_{\text{obs}}^{(\ell)}$ into (3.4) in turn yields $\mathbf{M}_{\text{obs}}^{(\ell)}\mathbf{\Gamma}^{(\ell)} = \mathbf{U}_{\text{obs}}^{(\ell)}\mathbf{G}_{\text{obs}}^{(\ell)}{\mathbf{F}_{\text{obs}}^{(\ell)}}^{\text{T}}\mathbf{\Gamma}^{(\ell)} = \mathbf{U}_{\text{obs}}^{(\ell)}\mathbf{V}_{\text{obs}}^{(\ell)} = \mathbf{B}_{\text{obs}}^{(\ell)}$, where $\mathbf{V}_{\text{obs}}^{(\ell)} := \mathbf{G}_{\text{obs}}^{(\ell)}{\mathbf{F}_{\text{obs}}^{(\ell)}}^{\text{T}}\mathbf{\Gamma}^{(\ell)} \in \mathbb{C}^{R\times R}$ is nonsingular. From (3.5), we now obtain the relation $\mathbf{\Upsilon}[{\mathbf{V}_{\text{obs}}^{(1)}}^{\text{T}}, \ldots, {\mathbf{V}_{\text{obs}}^{(L)}}^{\text{T}}, \mathbf{B}^{\text{T}}]^{\text{T}} = \mathbf{0}$. Since $\dim(\ker(\mathbf{\Upsilon})) = R$, we can conclude that $\text{range}\,(\mathbf{B}) = \text{range}\,(\mathbf{M})$ is unique and can be explicitly obtained from solving the set of linear equations (3.4). ∎

In Theorem 3.1 we have assumed that $\dim(\ker(\mathbf{\Upsilon})) = R$. We will now develop a condition on the factor matrices $\mathbf{B}$ and $\mathbf{F}$ under which $\dim(\ker(\mathbf{\Upsilon})) = R$ is guaranteed, implying in particular the uniqueness of $\text{range}(\mathbf{B})$. This condition will be obtained by first eliminating $\mathbf{B}$ from (3.5). To achieve this, a linear system similar to (3.5) is formulated which utilizes the overlap between the available rows of different submatrices $\mathbf{M}_{\text{obs}}^{(\ell)}$. The set of indices of the overlapping rows that are available in both $\mathbf{M}_{\text{obs}}^{(\ell_1)}$ and $\mathbf{M}_{\text{obs}}^{(\ell_2)}$ is $\alpha_{\ell_1,\ell_2} = \alpha_{\ell_1} \cap \alpha_{\ell_2}$ with $J_{\ell_1,\ell_2} = \text{card}\,(\alpha_{\ell_1,\ell_2})$. The row selection matrix that selects these overlapping rows from $\mathbf{M}$ can be constructed in two ways, namely

$$\mathbf{S}_{\text{row}}^{(\ell_1,\ell_2)} = \mathbf{S}_{\text{row}}^{(\ell_1|\ell_2)}\mathbf{S}_{\text{row}}^{(\ell_2)} \text{ or } \mathbf{S}_{\text{row}}^{(\ell_1,\ell_2)} = \mathbf{S}_{\text{row}}^{(\ell_2,\ell_1)} = \mathbf{S}_{\text{row}}^{(\ell_2|\ell_1)}\mathbf{S}_{\text{row}}^{(\ell_1)}. \tag{3.8}$$

The "conditional" row selection matrix $\mathbf{S}_{\text{row}}^{(\ell_2|\ell_1)}$, omits the rows of $\mathbf{M}_{\text{obs}}^{(\ell_1)}$ that are not available in $\mathbf{M}_{\text{obs}}^{(\ell_2)}$. Namely, it selects the rows with indices in the set

$$\alpha_{\ell_2|\ell_1} = \{i \in \{1, \ldots, \text{card}\,(\alpha_{\ell_1})\} | \alpha_{\ell_1}(i) \in \alpha_{\ell_2}\},$$

in which $\alpha_{\ell_1}(i)$ is the ith element listed for the set α_{ℓ_1}. Combination of the two expressions in (3.8) with (3.4) yields

$$\mathbf{S}_{\text{row}}^{(\ell_2|\ell_1)}\mathbf{M}_{\text{obs}}^{(\ell_1)}\mathbf{\Gamma}^{(\ell_1)} = \mathbf{S}_{\text{row}}^{(\ell_1|\ell_2)}\mathbf{M}_{\text{obs}}^{(\ell_2)}\mathbf{\Gamma}^{(\ell_2)} = \mathbf{S}_{\text{row}}^{(\ell_1,\ell_2)}\mathbf{B} \quad \text{for } 1 \le \ell_1 < \ell_2 \le L,$$

so that

$$\mathbf{S}_{\text{row}}^{(\ell_2|\ell_1)}\mathbf{M}_{\text{obs}}^{(\ell_1)}\mathbf{\Gamma}^{(\ell_1)} - \mathbf{S}_{\text{row}}^{(\ell_1|\ell_2)}\mathbf{M}_{\text{obs}}^{(\ell_2)}\mathbf{\Gamma}^{(\ell_2)} = \mathbf{0} \quad \text{for } 1 \le \ell_1 < \ell_2 \le L. \tag{3.9}$$

In words this means that the matrices, obtained by omitting the rows of $\mathbf{M}_{\text{obs}}^{(\ell_1)}$ that are not in $\mathbf{M}_{\text{obs}}^{(\ell_2)}$ and by omitting the rows of $\mathbf{M}_{\text{obs}}^{(\ell_2)}$ that are not in $\mathbf{M}_{\text{obs}}^{(\ell_1)}$ respectively, are equal up to a change-of-basis transformation. Equation (3.9) can be written in matrix form, similar to (3.5), as

$$\mathbf{\Xi}\,\mathbf{\Gamma} = \mathbf{0}, \tag{3.10}$$

with

$$\Xi = \begin{bmatrix} \mathbf{S}_{\text{row}}^{(2|1)}\mathbf{M}_{\text{obs}}^{(1)} & -\mathbf{S}_{\text{row}}^{(1|2)}\mathbf{M}_{\text{obs}}^{(2)} & \mathbf{0} & & \\ \mathbf{S}_{\text{row}}^{(3|1)}\mathbf{M}_{\text{obs}}^{(1)} & \mathbf{0} & -\mathbf{S}_{\text{row}}^{(1|3)}\mathbf{M}_{\text{obs}}^{(3)} & & \\ \vdots & & & \ddots & \\ & & & \mathbf{S}_{\text{row}}^{(L|L-1)}\mathbf{M}_{\text{obs}}^{(L-1)} & -\mathbf{S}_{\text{row}}^{(L-1|L)}\mathbf{M}_{\text{obs}}^{(L)} \end{bmatrix}$$

and

$$\mathbf{\Gamma} = \begin{bmatrix} \mathbf{\Gamma}^{(1)} \\ \vdots \\ \mathbf{\Gamma}^{(L)} \end{bmatrix}.$$

From the homogeneous matrix equation (3.10) we obtain an alternative to Theorem 3.1. More precisely, by replacing $\mathbf{\Upsilon}$ in condition (3.7) with

(3.11)
$$\mathbf{\Xi}_{\mathbf{U}} = \begin{bmatrix} \mathbf{S}_{\text{row}}^{(2|1)}\mathbf{U}_{\text{obs}}^{(1)} & -\mathbf{S}_{\text{row}}^{(1|2)}\mathbf{U}_{\text{obs}}^{(2)} & \mathbf{0} & & \\ \mathbf{S}_{\text{row}}^{(3|1)}\mathbf{U}_{\text{obs}}^{(1)} & \mathbf{0} & -\mathbf{S}_{\text{row}}^{(1|3)}\mathbf{U}_{\text{obs}}^{(3)} & & \\ \vdots & & & \ddots & \\ & & & \mathbf{S}_{\text{row}}^{(L|L-1)}\mathbf{U}_{\text{obs}}^{(L-1)} & -\mathbf{S}_{\text{row}}^{(L-1|L)}\mathbf{U}_{\text{obs}}^{(L)} \end{bmatrix}$$

where the columns of $\mathbf{U}_{\text{obs}}^{(\ell)} \in \mathbb{C}^{J_\ell \times R}$ form a basis for range$(\mathbf{M}_{\text{obs}}^{(\ell)})$, we obtain the result stated in Theorem 3.2.

THEOREM 3.2. *Let $\mathbf{M}_{obs}^{(\ell)} \in \mathbb{C}^{J_\ell \times F_\ell}$ for $\ell = 1, \ldots, L$ be fully observed submatrices of rank R of a matrix $\mathbf{M}$ of rank R. If*

$$dim(ker(\ \mathbf{\Xi}_{\mathbf{U}})) = R \qquad \text{with } \mathbf{\Xi}_{\mathbf{U}} \text{ as in (3.11)},$$

then the overall column space range$(\mathbf{M})$ is uniquely determined and can be explicitly obtained from solving the set of linear equations (3.4).

Proof. The proof is analogous to the proof of Theorem 3.1. ∎

If the uniqueness condition in Theorem 3.2 is satisfied, then this implies that the uniqueness condition in Theorem 3.1 must also be satisfied. This follows from the fact that the matrix equations (3.9) used in Theorem 3.2 are built from (a subset of) the matrix equations (3.4) used in Theorem 3.1. That is, if the solution to the former set of equations is unique up to an intrinsic change-of-basis transformation ambiguity, then the same holds for the solution to the latter set of equations.

Based on Theorem 3.2 and relation (3.9) we will derive a third uniqueness condition that is equivalent to the condition stated in Theorem 3.2, but formulated only in terms of the factor matrices $\mathbf{B}$ and $\mathbf{F}$. Relation (3.9) can also be expressed as

$$\mathbf{S}_{\text{row}}^{(\ell_2|\ell_1)}\mathbf{S}_{\text{row}}^{(\ell_1)}\mathbf{B}{\mathbf{F}_{\text{obs}}^{(\ell_1)}}^{\mathrm{T}}\mathbf{\Gamma}^{(\ell_1)} - \mathbf{S}_{\text{row}}^{(\ell_1|\ell_2)}\mathbf{S}_{\text{row}}^{(\ell_2)}\mathbf{B}{\mathbf{F}_{\text{obs}}^{(\ell_2)}}^{\mathrm{T}}\mathbf{\Gamma}^{(\ell_2)}\mathbf{0},\ 1 \le \ell_1 < \ell_2 \le L,$$

Because of (3.8), this becomes

$$\mathbf{B}_{\text{obs}}^{(\ell_1,\ell_2)}{\mathbf{F}_{\text{obs}}^{(\ell_1)}}^{\mathrm{T}}\mathbf{\Gamma}^{(\ell_1)} - \mathbf{B}_{\text{obs}}^{(\ell_1,\ell_2)}{\mathbf{F}_{\text{obs}}^{(\ell_2)}}^{\mathrm{T}}\mathbf{\Gamma}^{(\ell_2)} = \mathbf{0},\ 1 \le \ell_1 < \ell_2 \le L, \tag{3.12}$$

in which $\mathbf{B}_{\text{obs}}^{(\ell_1,\ell_2)} = \mathbf{S}_{\text{row}}^{(\ell_1,\ell_2)}\mathbf{B} \in \mathbb{C}^{\text{card}\left(\alpha_{\ell_1,\ell_2}\right)\times R}$ collects the rows of $\mathbf{B}$ that are observed in both $\mathbf{M}_{\text{obs}}^{(\ell_1)}$ and $\mathbf{M}_{\text{obs}}^{(\ell_2)}$. Based on relations (3.10) and (3.12), Theorem 3.4 below

presents a sufficient condition that ensures the identifiability of the subspace range($\mathbf{B}$). First, additional notation will be defined. Theorem 3.4 will make use of the matrix

$$\boldsymbol{\nabla} := \begin{bmatrix} \boldsymbol{\nabla}^{(1)} \\ \vdots \\ \boldsymbol{\nabla}^{(L-1)} \end{bmatrix} \in \mathbb{C}^{(\sum_{1\le\ell_1<\ell_2\le L} J_{\ell_1,\ell_2})\times(L-1)R} \tag{3.13}$$

with the submatrices $\boldsymbol{\nabla}^{(\ell_1)} \in \mathbb{C}^{(\sum_{\ell_2=\ell_1+1}^{L} J_{\ell_1,\ell_2})\times(L-1)R}$

$$\boldsymbol{\nabla}^{(\ell_1)} = \begin{bmatrix} \mathbf{0}_{J_{\ell_1,\ell_1+1}\times(\ell_1-1)R} & \mathbf{B}_{\text{obs}}^{(\ell_1,\ell_1+1)} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0}_{J_{\ell_1,\ell_1+2}\times(\ell_1-1)R} & \mathbf{B}_{\text{obs}}^{(\ell_1,\ell_1+2)} & \mathbf{B}_{\text{obs}}^{(\ell_1,\ell_1+2)} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{J_{\ell_1,L}\times(\ell_1-1)R} & \mathbf{B}_{\text{obs}}^{(\ell_1,L)} & \mathbf{B}_{\text{obs}}^{(\ell_1,L)} & \cdots & \cdots & \mathbf{B}_{\text{obs}}^{(\ell_1,L)} \end{bmatrix} \tag{3.14}$$

for $1 \le \ell_1 < L-1$. The matrix $\boldsymbol{\nabla}$ is block upper triangular. To make this more clear, let us consider an example with $L = 4$:

$$\boldsymbol{\nabla} = \begin{bmatrix} \boldsymbol{\nabla}^{(1)} \\ \boldsymbol{\nabla}^{(2)} \\ \boldsymbol{\nabla}^{(3)} \end{bmatrix} = \left[\begin{array}{c|c|c} \mathbf{B}_{\text{obs}}^{(1,2)} & \mathbf{0} & \mathbf{0} \\ \mathbf{B}_{\text{obs}}^{(1,3)} & \mathbf{B}_{\text{obs}}^{(1,3)} & \mathbf{0} \\ \mathbf{B}_{\text{obs}}^{(1,4)} & \mathbf{B}_{\text{obs}}^{(1,4)} & \mathbf{B}_{\text{obs}}^{(1,4)} \\ \hline \mathbf{0} & \mathbf{B}_{\text{obs}}^{(2,3)} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{\text{obs}}^{(2,4)} & \mathbf{B}_{\text{obs}}^{(2,4)} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{B}_{\text{obs}}^{(3,4)} \end{array}\right]. \tag{3.15}$$

Additionally, Theorem 3.4 makes use of Lemma 3.3 below, which in turn makes use of the matrices $\boldsymbol{\Theta}^{(\ell_1)} \in \mathbb{C}^{(\sum_{\ell_2=\ell_1+1}^{L} J_{\ell_1,\ell_2})\times LR}$

$$\boldsymbol{\Theta}^{(\ell_1)} = \begin{bmatrix} \mathbf{0}_{J_{\ell_1,\ell_1+1}\times(\ell_1-1)R} & \mathbf{B}_{\text{obs}}^{(\ell_1,\ell_1+1)} & -\mathbf{B}_{\text{obs}}^{(\ell_1,\ell_1+1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0}_{J_{\ell_1,\ell_1+2}\times(\ell_1-1)R} & \mathbf{B}_{\text{obs}}^{(\ell_1,\ell_1+2)} & \mathbf{0} & -\mathbf{B}_{\text{obs}}^{(\ell_1,\ell_1+2)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{J_{\ell_1,L}\times(\ell_1-1)R} & \mathbf{B}_{\text{obs}}^{(\ell_1,L)} & \mathbf{0} & \mathbf{0} & \cdots & -\mathbf{B}_{\text{obs}}^{(\ell_1,L)} \end{bmatrix}$$

for $1 \le \ell_1 < L-1$,

$$\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\Theta}^{(1)} \\ \vdots \\ \boldsymbol{\Theta}^{(L-1)} \end{bmatrix} \in \mathbb{C}^{(\sum_{1\le\ell_1<\ell_2\le L} J_{\ell_1,\ell_2})\times LR} \tag{3.16}$$

and

$$\mathbf{L} = \begin{bmatrix} \mathbf{I}_R & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{I}_R & \mathbf{I}_R & \cdots & \vdots \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{I}_R & \mathbf{I}_R & \cdots & \mathbf{I}_R \end{bmatrix} \in \mathbb{C}^{LR\times LR}. \tag{3.17}$$

To illustrate the structure of $\boldsymbol{\Theta}$, let us again consider an example with $L = 4$:

$$\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\Theta}^{(1)} \\ \boldsymbol{\Theta}^{(2)} \\ \boldsymbol{\Theta}^{(3)} \end{bmatrix} = \left[\begin{array}{cccc} \mathbf{B}_{\text{obs}}^{(1,2)} & -\mathbf{B}_{\text{obs}}^{(1,2)} & \mathbf{0} & \mathbf{0} \\ \mathbf{B}_{\text{obs}}^{(1,3)} & \mathbf{0} & -\mathbf{B}_{\text{obs}}^{(1,3)} & \mathbf{0} \\ \mathbf{B}_{\text{obs}}^{(1,4)} & \mathbf{0} & \mathbf{0} & -\mathbf{B}_{\text{obs}}^{(1,4)} \\ \hline \mathbf{0} & \mathbf{B}_{\text{obs}}^{(2,3)} & -\mathbf{B}_{\text{obs}}^{(2,3)} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{\text{obs}}^{(2,4)} & \mathbf{0} & -\mathbf{B}_{\text{obs}}^{(2,4)} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{B}_{\text{obs}}^{(3,4)} & -\mathbf{B}_{\text{obs}}^{(3,4)} \end{array}\right].$$

The proof of Theorem 3.4 is constructive. It guarantees that $\boldsymbol{\Gamma}^{(1)}, \ldots, \boldsymbol{\Gamma}^{(L)}$ can be obtained via the system of linear of equations (3.9)/(3.10). To guarantee that not only $\boldsymbol{\Gamma}^{(1)}, \ldots, \boldsymbol{\Gamma}^{(L)}$ but also the full subspace range$(\mathbf{B})$ can be obtained, each row of $\mathbf{M}$ must be observed in at least one of the submatrices $\mathbf{M}_{\text{obs}}^{(l)}$ for $\ell = 1, \ldots, L$[2]. Specifically, the necessary condition

$$\cup_{\ell=1}^{L} \alpha_\ell = \{1, \ldots, J\}$$

must hold. Finally, Theorem 3.4 makes use of the following lemma.

LEMMA 3.3. *Consider the matrices* $\boldsymbol{\nabla}$, $\boldsymbol{\Theta}$ *and* $\mathbf{L}$ *given by* (3.13), (3.16), *and* (3.17), *respectively. Then* $\boldsymbol{\Theta}\mathbf{L} = [\mathbf{0}_{(\sum_{1\le\ell_1<\ell_2\le L} J_{\ell_1,\ell_2})\times R}, -\boldsymbol{\nabla}]$.

Proof. The lemma follows with a simple technical derivation from equations (3.13), (3.16) and (3.17). Interested readers can find a formal proof based on induction in the Supplementary Materials. ∎

THEOREM 3.4. *Let* $\mathbf{M}_{obs}^{(\ell)} \in \mathbb{C}^{J_\ell \times F_\ell}$ *for* $\ell = 1, \ldots, L$ *be fully observed submatrices of rank* R *of a matrix* $\mathbf{M}$ *of rank* R. *If*

(3.18)
$$\begin{cases} \boldsymbol{\nabla} \text{ has full column rank} & \text{with } \boldsymbol{\nabla} \text{ as in (3.13)}, \\ \text{Every row of } \mathbf{M} \text{ is (partially) observed in at least one submatrix } \mathbf{M}_{obs}^{(\ell)}: \\ \quad \cup_{\ell=1}^{L} \alpha_\ell = \{1, \ldots, J\}, \end{cases}$$

then the overall column space range$(\mathbf{M})$ *is uniquely determined and can be explicitly obtained from solving the set of linear equations* (3.4).

Proof. Given (3.3) and (3.8), we have that $\mathbf{B}_{\text{obs}}^{(\ell_1,\ell_2)} \mathbf{F}_{\text{obs}}^{(\ell_1)^{\mathrm{T}}} = \mathbf{S}_{\text{row}}^{(\ell_2|\ell_1)} \mathbf{B}_{\text{obs}}^{(\ell_1)} \mathbf{F}_{\text{obs}}^{(\ell_1)^{\mathrm{T}}} = \mathbf{S}_{\text{row}}^{(\ell_2|\ell_1)} \mathbf{M}_{\text{obs}}^{(\ell_1)}$ and similarly that $\mathbf{B}_{\text{obs}}^{(\ell_1,\ell_2)} \mathbf{F}_{\text{obs}}^{(\ell_2)^{\mathrm{T}}} = \mathbf{S}_{\text{row}}^{(\ell_1|\ell_2)} \mathbf{M}_{\text{obs}}^{(\ell_2)}$. As a result, (3.10) can be expressed as

$$\boldsymbol{\Xi}\boldsymbol{\Gamma} = \mathbf{0} \Leftrightarrow \boldsymbol{\Theta}\boldsymbol{\Lambda}\boldsymbol{\Gamma} = \mathbf{0}, \tag{3.19}$$

where $\boldsymbol{\Theta} \in \mathbb{C}^{(\sum_{1\le\ell_1<\ell_2\le L} J_{\ell_1,\ell_2})\times LR}$ is defined as indicated in (3.16) and

$$\boldsymbol{\Lambda} = \begin{bmatrix} \mathbf{F}_{\text{obs}}^{(1)^{\mathrm{T}}} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{F}_{\text{obs}}^{(L)^{\mathrm{T}}} \end{bmatrix}.$$

[2]This means that every row of $\mathbf{B}$ needs to appear in (3.4), because rows that do not appear cannot be retrieved. Note also that $\dim(\ker(\boldsymbol{\Upsilon})) = R$, as in Theorem 3.1, implies that this basic condition is satisfied. Indeed, if $\cup_{\ell=1}^{L} \alpha_\ell \neq \{1, \ldots, J\}$, then $\boldsymbol{\Upsilon}$ will contain all-zero column vectors, implying that $\dim(\ker(\boldsymbol{\Upsilon})) > R$.

Next, we will argue that the conditions in (3.18) imply that $\mathbf{\Gamma}^{(1)}, \ldots, \mathbf{\Gamma}^{(L)}$ can be obtained via the system of linear of equations (3.9)/(3.10), which further implies that range($\mathbf{B}$) can be uniquely obtained from (3.4)/(3.5) under the richness-of-sampling assumption $\cup_{\ell=1}^{L} \alpha_\ell = \{1, \ldots, J\}$. First, $\mathbf{\Gamma}^{(\ell)}$ has full column rank since (3.4) holds and $\mathbf{B}_{\text{obs}}^{(\ell)}$ and $\mathbf{F}_{\text{obs}}^{(\ell)}$ have full column rank since $\mathbf{M}_{\text{obs}}^{(\ell)}$ has rank R. Second, under the same assumption, range $\left(\mathbf{M}_{\text{obs}}^{(\ell)^{\mathrm{T}}}\right)$ = range $\left(\mathbf{F}_{\text{obs}}^{(\ell)}\right)$ follows from (3.3), which implies that a basis for the latter can be obtained from the former. Let the columns of $\mathbf{V}^{(\ell)} \in \mathbb{C}^{F_\ell \times R}$ form a basis for range($\mathbf{F}_{\text{obs}}^{(\ell)}$). Third, note that in (3.9) we can w.l.o.g. choose $\mathbf{\Gamma}^{(\ell)}$ such that range($\mathbf{\Gamma}^{(\ell)}$) $\subseteq$ range($\mathbf{F}_{\text{obs}}^{(\ell)}$), as components of the columns of $\mathbf{\Gamma}^{(\ell)}$ that are orthogonal to range($\mathbf{F}_{\text{obs}}^{(\ell)}$) disappear after multiplication by $\mathbf{F}_{\text{obs}}^{(\ell)^{\mathrm{T}}}$ in (3.9). Since range($\mathbf{F}_{\text{obs}}^{(\ell)}$) = range($\mathbf{V}^{(\ell)}$), there exists a nonsingular matrix $\mathbf{G}^{(\ell)} \in \mathbb{C}^{R \times R}$ such that $\mathbf{\Gamma}^{(\ell)} = \mathbf{V}^{(\ell)}\mathbf{G}^{(\ell)}$. This means that $\mathbf{\Gamma}$ in (3.19) admits the factorization

$$\mathbf{\Gamma} = \mathbf{\Upsilon}\mathbf{G}, \tag{3.20}$$

in which

$$\mathbf{\Upsilon} = \begin{bmatrix} \mathbf{V}^{(1)} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{V}^{(L)} \end{bmatrix} \in \mathbb{C}^{(\sum_{\ell=1}^{L} F_\ell) \times LR} \quad \text{and} \quad \mathbf{G} = \begin{bmatrix} \mathbf{G}^{(1)} \\ \vdots \\ \mathbf{G}^{(L)} \end{bmatrix} \in \mathbb{C}^{LR \times R}.$$

Using factorization (3.20), relation (3.19) can be expressed as

$$\mathbf{\Theta}\mathbf{\Lambda}\mathbf{\Gamma} = \mathbf{\Theta}\mathbf{\Lambda}\mathbf{\Upsilon}\mathbf{G} = \mathbf{0}.$$

Since $\mathbf{\Lambda}\mathbf{\Upsilon} \in \mathbb{C}^{LR \times LR}$ is a nonsingular block-diagonal matrix with nonsingular matrices $\mathbf{F}_{\text{obs}}^{(\ell)^{\mathrm{T}}}\mathbf{V}^{(\ell)} \in \mathbb{C}^{R \times R}$ for $\ell = 1, \ldots, L$ on its block-diagonal, we know that ker($\mathbf{\Theta}$) = ker($\mathbf{\Theta}\mathbf{\Lambda}\mathbf{\Upsilon}$). This means that if dim(ker($\mathbf{\Theta}$)) = R, then matrix $\mathbf{G}$, and in turn also $\mathbf{\Gamma}$, can be found from ker($\mathbf{\Theta}$) up to an irrelevant change-of-basis transformation. Due to relation (3.4)/(3.5) and the condition $\cup_{\ell=1}^{L} \alpha_\ell = \{1, \ldots, J\}$, this implies the uniqueness of range ($\mathbf{B}$) = range ($\mathbf{M}$).

Finally, we will explain that condition (3.18) indeed implies that dim(ker($\mathbf{\Theta}$)) = R. From Lemma 3.3 we know that

$$\mathbf{\Theta}\mathbf{L} = [\mathbf{0}_{(\sum_{\ell_2=\ell_1+1}^{L} J_{\ell_1,\ell_2}) \times R}, -\mathbf{\nabla}],$$

where $\mathbf{L} \in \mathbb{C}^{LR \times LR}$ is a lower block-triangular matrix defined in accordance with (3.17). The full column rank condition on $\mathbf{\nabla}$ in (3.18) and the fact that $\mathbf{L}$ is nonsingular imply that dim(ker($\mathbf{\Theta}$)) = R. ∎

The number of rows of $\mathbf{\nabla}$, assuming that the number of overlapping rows between any combination of two fully observed submatrices $\mathbf{M}_{\text{obs}}^{(\ell_1)}$ and $\mathbf{M}_{\text{obs}}^{(\ell_2)}$ is cJ with $0 < c \leq 1$, is equal to $\frac{L(L-1)}{2}cJ$. Therefore, this matrix can become very large as L increases, making it computationally expensive to check whether it has full column rank and, in turn, making it impractical to verify the conditions in Theorem 3.4. The following corollary presents an easier way to guarantee that $\mathbf{\nabla}$ has full column rank, but the downside is that the resulting uniqueness conditions are more restrictive. This is achieved by checking whether the $L-1$ blocks on the block diagonal of $\mathbf{\nabla}$ have full

column rank, which in turn guarantees that $\boldsymbol{\nabla}$ itself has full column rank. So checking the rank of a large matrix has been replaced by checking the rank of multiple smaller matrices. These blocks on the block diagonal are denoted by

$$\boldsymbol{\nabla}^{(\ell_1 \to L)} = \begin{bmatrix} \mathbf{B}_{\text{obs}}^{(\ell_1,\ell_{1+1})} \\ \mathbf{B}_{\text{obs}}^{(\ell_1,\ell_{1+2})} \\ \vdots \\ \mathbf{B}_{\text{obs}}^{(\ell_1,L)} \end{bmatrix} \in \mathbb{C}^{(\sum_{\ell_2=\ell_1+1}^{L} J_{\ell_1,\ell_2}) \times R} \quad \text{for } \ell_1 = 1, \ldots, L-1, \tag{3.21}$$

in which, as stated above, $\mathbf{B}_{\text{obs}}^{(\ell_1,\ell_2)}$ collects the rows of $\mathbf{B}$ that are observed in both $\mathbf{M}_{\text{obs}}^{(\ell_1)}$ and $\mathbf{M}_{\text{obs}}^{(\ell_2)}$. In the example in (3.15), these blocks are

$$\begin{bmatrix} \mathbf{B}_{\text{obs}}^{(1,2)} \\ \mathbf{B}_{\text{obs}}^{(1,3)} \\ \mathbf{B}_{\text{obs}}^{(1,4)} \end{bmatrix}, \begin{bmatrix} \mathbf{B}_{\text{obs}}^{(2,3)} \\ \mathbf{B}_{\text{obs}}^{(2,4)} \end{bmatrix} \text{ and } \mathbf{B}_{\text{obs}}^{(3,4)}.$$

COROLLARY 3.5. *Let* $\mathbf{M}_{obs}^{(\ell)} =\in \mathbb{C}^{J_\ell \times F_\ell}$ *for* $\ell = 1, \ldots, L$ *be fully observed submatrices of rank* R *of a matrix* $\mathbf{M}$ *of rank* R. *If*

$$\begin{cases} \boldsymbol{\nabla}^{(\ell_1 \to L)} \text{ as in (3.21) has full column rank for } \ell_1 = 1, \ldots, L-1, \\ \text{Every row of } \mathbf{M} \text{ is (partially) observed in at least one submatrix } \mathbf{M}_{obs}^{(\ell)}: \\ \quad \cup_{\ell=1}^{L} \alpha_\ell = \{1, \ldots, J\}, \end{cases}$$

then the overall column space range $(\mathbf{M})$ *is uniquely determined and can be explicitly obtained by solving the set of linear equations* (3.4).

Proof. Observe that $\boldsymbol{\nabla}^{(\ell_1 \to L)}$ corresponds to the leftmost nonzero block column of $\boldsymbol{\nabla}^{(\ell_1)}$ given in (3.14). Since we assume that $\boldsymbol{\nabla}^{(\ell_1 \to L)}$ has full column rank for $\ell_1 = 1, \ldots, L-1$, the block-triangular structure of $\boldsymbol{\nabla}$ (see example (3.15)) implies that the overall matrix $\boldsymbol{\nabla}$ has full column rank. The result follows now directly from Theorem 3.4. ∎

The most restrictive part of the first set of conditions in Corollary 3.5 is that $\boldsymbol{\nabla}^{(L-1 \to L)} = \mathbf{B}_{\text{obs}}^{(L-1,L)}$ must have full column rank. As the order of the fully observed submatrices $\mathbf{M}_{\text{obs}}^{(\ell)}$ for $\ell = 1, \ldots, L$ does not matter, they can be reordered such that $\mathbf{B}_{\text{obs}}^{(L-1,L)}$ does have full column rank. The rank constraint on $\boldsymbol{\nabla}^{(\ell_1 \to L)}$ gradually relaxes as ℓ_1 decreases, because the number of vertically stacked matrices $\mathbf{B}_{\text{obs}}^{(\ell_1,\ell_2)}$ for $\ell_2 = \ell_1 + 1, \ldots, L$ in $\boldsymbol{\nabla}^{(\ell_1 \to L)}$ increases. Therefore, none of the individual submatrices $\mathbf{B}_{\text{obs}}^{(\ell_1,\ell_2)}$ for $\ell_2 = \ell_1 + 1, \ldots, L$ of $\boldsymbol{\nabla}^{(\ell_1 \to L)}$ need to have full column rank, only the overall matrix does.

Next, we present the generic variant of the uniqueness conditions in Corollary 3.5. A generic matrix $\mathbf{A} \in \mathbb{C}^{I \times J}$ has full column rank if $I \geq J$. As a result, the conditions in Corollary 3.5 are generically satisfied if the observed submatrices $\mathbf{M}_{\text{obs}}^{(\ell)}$ for $\ell = 1, \ldots, L$ have at least R rows and columns and

$$\begin{cases} \boldsymbol{\nabla}^{(\ell_1 \to L)} \text{ as in (3.21) has at least } R \text{ distinct rows for } \ell_1 = 1, \ldots, L-1, \\ \text{Every row of } \mathbf{M} \text{ is (partially) observed in at least one submatrix } \mathbf{M}_{\text{obs}}^{(\ell)}: \\ \quad \cup_{\ell=1}^{L} \alpha_\ell = \{1, \ldots, J\}. \end{cases} \tag{3.22}$$

Keeping in mind that the matrix $\mathbf{M}$ that needs to be completed has itself rank R, the above conditions are very mild indeed.

In words, we limit ourselves to fully observed submatrices of $\mathbf{M}$ that have the same rank R as $\mathbf{M}$, which is generically true for all submatrices of size at least $R \times R$. As an obvious condition, in the overall collection of these submatrices every row of $\mathbf{M}$ should be partially observed at least once (second condition in (3.22)). Let J_{ℓ_1,ℓ_2} be the number of rows that are partially observed in both $\mathbf{M}_{\text{obs}}^{(\ell_1)}$ and $\mathbf{M}_{\text{obs}}^{(\ell_2)}$, for every pair of submatrices. Then there should be an ordering of submatrices such that $\sum_{\ell_2=\ell_1+1}^{L} J_{\ell_1,\ell_2}$ is at least the rank R of $\mathbf{M}$, for every value of ℓ_1 (first condition in (3.22)). Together these conditions guarantee with probability 1 that the result is unique and can be computed using only linear algebra.

4. Matrix completion: algebraic algorithms. In this section, three algorithms are introduced for the computation of the column space of the incomplete matrix $\mathbf{M}$ from the fully observed submatrices $\mathbf{M}_{\text{obs}}^{(\ell)}$ for $\ell = 1, \ldots, L$. These algorithms can be used to compute a column-wise orthonormal basis for range $(\mathbf{M}) =$ range $(\mathbf{B})$ using the set of equations (3.4) if condition (3.18) in Theorem 3.4 is satisfied.

4.1. An SVD-based kernel subspace method. *Derivation.* If the equations (3.4) are inexact, which is the case if noise is present, they can be solved in the least squares sense by minimizing the cost function

$$\sum_{\ell=1}^{L} \left|\left| \mathbf{M}_{\text{obs}}^{(\ell)} \mathbf{\Gamma}^{(\ell)} - \mathbf{B}_{\text{obs}}^{(\ell)} \right|\right|_{\text{F}}^2 . \tag{4.1}$$

Condition (3.18) states that $\mathbf{B}_{\text{obs}}^{(\ell)}$ and $\mathbf{F}_{\text{obs}}^{(\ell)}$ have full column rank, which in turn implies that $\mathbf{M}_{\text{obs}}^{(\ell)}$ has rank R for $\ell = 1, \ldots, L$. This allows the elimination of the variables $\mathbf{\Gamma}^{(\ell)}$ by substituting $\mathbf{\Gamma}^{(\ell)} = \mathbf{M}_{\text{obs}}^{(\ell)^\dagger} \mathbf{B}_{\text{obs}}^{(\ell)}$ into (4.1), which results in the optimization problem

$$\min_{\mathbf{B}^{\text{H}}\mathbf{B}=\mathbf{I}_R} \sum_{\ell=1}^{L} \left|\left| \left(\mathbf{M}_{\text{obs}}^{(\ell)} \mathbf{M}_{\text{obs}}^{(\ell)^\dagger} - \mathbf{I}_{J_\ell} \right) \mathbf{S}_{\text{row}}^{(\ell)} \mathbf{B} \right|\right|_{\text{F}}^2 . \tag{4.2}$$

The matrix $\left(\mathbf{M}_{\text{obs}}^{(\ell)} \mathbf{M}_{\text{obs}}^{(\ell)^\dagger} - \mathbf{I}_{J_\ell} \right)$ is an orthogonal projector onto the orthogonal complement of range $\left(\mathbf{M}_{\text{obs}}^{(\ell)} \right)$. A column-wise orthonormal basis $\mathbf{Q}^{(\ell)}$ for this complement can be obtained from the SVD $\mathbf{M}_{\text{obs}}^{(\ell)} = \mathbf{U}^{(\ell)} \mathbf{\Sigma}^{(\ell)} \mathbf{V}^{(\ell)^{\text{H}}}$, namely

$$\mathbf{Q}^{(\ell)} = \begin{bmatrix} \mathbf{u}_{R+1}^{(\ell)} & \ldots & \mathbf{u}_{J_\ell}^{(\ell)} \end{bmatrix} \quad \text{for } \ell = 1, \ldots, L.$$

The orthogonal projector in (4.2) can then be replaced by $\mathbf{Q}^{(\ell)} \mathbf{Q}^{(\ell)^{\text{H}}}$, which gives

$$\min_{\mathbf{B}^{\text{H}}\mathbf{B}=\mathbf{I}_R} \sum_{\ell=1}^{L} \left|\left| \mathbf{Q}^{(\ell)} \mathbf{Q}^{(\ell)^{\text{H}}} \mathbf{S}_{\text{row}}^{(\ell)} \mathbf{B} \right|\right|_{\text{F}}^2 .$$

Here, the multiplication from the left by $\mathbf{Q}^{(\ell)}$ can be omitted because $\mathbf{Q}^{(\ell)}$ is column-wise orthonormal:

$$\min_{\mathbf{B}^{\text{H}}\mathbf{B}=\mathbf{I}_R} \sum_{\ell=1}^{L} \left|\left| \mathbf{Q}^{(\ell)^{\text{H}}} \mathbf{S}_{\text{row}}^{(\ell)} \mathbf{B} \right|\right|_{\text{F}}^2 .$$

This minimization problem can further be rewritten as

$$\min_{\mathbf{B}^{\mathrm{H}}\mathbf{B}=\mathbf{I}_R} ||\mathbf{B}^{\mathrm{H}}\mathbf{Q}||_{\mathrm{F}}^2 \quad \text{with } \mathbf{Q} = \begin{bmatrix} \mathbf{S}_{\mathrm{row}}^{(1)^{\mathrm{T}}}\mathbf{Q}^{(1)} & \dots & \mathbf{S}_{\mathrm{row}}^{(L)^{\mathrm{T}}}\mathbf{Q}^{(L)} \end{bmatrix}.$$

This shows that range $(\mathbf{B})$ is generated by the left singular vectors of $\mathbf{Q}$ that correspond to the R smallest singular values. The overall approach for the computation of a column-wise orthonormal basis for range $(\mathbf{B})$ is described in Algorithm 4.1.

Algorithm 4.1 SVD-based kernel subspace method for computing range $(\mathbf{M})$.

Input: Fully observed submatrices $\mathbf{M}_{\mathrm{obs}}^{(\ell)}$ for $\ell = 1, \dots, L$
Output: Orthonormal basis for range $(\mathbf{M})$
for $\ell = 1, \dots, L$ **do**
  Compute SVD $\mathbf{M}_{\mathrm{obs}}^{(\ell)} = \mathbf{U}^{(\ell)}\mathbf{\Sigma}^{(\ell)}\mathbf{V}^{(\ell)^{\mathrm{H}}}$
  Set $\mathbf{Q}^{(\ell)} \coloneqq \begin{bmatrix} \mathbf{u}_{R+1}^{(\ell)} & \dots & \mathbf{u}_{J_\ell}^{(\ell)} \end{bmatrix}$
end for
Compute SVD $\mathbf{Q} = \begin{bmatrix} \mathbf{S}_{\mathrm{row}}^{(1)^{\mathrm{T}}}\mathbf{Q}^{(1)} & \dots & \mathbf{S}_{\mathrm{row}}^{(L)^{\mathrm{T}}}\mathbf{Q}^{(L)} \end{bmatrix} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}$
Set $\mathbf{B} \coloneqq \begin{bmatrix} \mathbf{u}_{J-R+1} & \dots & \mathbf{u}_J \end{bmatrix}$ % range $(\mathbf{B})$ = range $(\mathbf{M})$

Discussion. Algorithm 4.1 is similar to Algorithm 1 in [37]. The latter algorithm was developed in the setting in which the observed entries can be adaptively selected. It selects entries such that for each column of the matrix, there is a fully observed submatrix with $R+1$ rows and at least R columns. As a result, the dimension of the left nullspace of each submatrix is equal to one, meaning that the equivalent of the matrix $\mathbf{Q}^{(\ell)}$ in that algorithm only consists of one column. In contrast, in Algorithm 4.1 the dimension of the nullspace of the observed rows of the submatrix $\mathbf{M}_{\mathrm{obs}}^{(\ell)}$, with J_ℓ observed rows, equals $J_\ell - R$ and $\mathbf{Q}^{(\ell)}$ has $J_\ell - R$ columns. This is an extension of the algorithm in [24] for the specific case $R = 3$, which was developed for the structure-from-motion problem in computer vision. In the next section, we introduce the principal subspace counterpart of this kernel subspace approach.

4.2. An SVD-based principal subspace method. *Derivation.* In the previous section, a column-wise orthonormal basis for range $(\mathbf{B})$ was computed by minimizing the projection of $\mathbf{B}$ onto the orthogonal complement of range $\left(\mathbf{M}_{\mathrm{obs}}^{(\ell)}\right)$ for $\ell = 1, \dots, L$ simultaneously. In contrast, here we compute range $(\mathbf{B})$ by maximizing the projection onto range $\left(\mathbf{M}_{\mathrm{obs}}^{(\ell)}\right)$ for $\ell = 1, \dots, L$. The optimization problem in (4.2) can be rewritten as

$$\min_{\mathbf{B}^{\mathrm{H}}\mathbf{B}=\mathbf{I}_R} \sum_{\ell=1}^{L} \left|\left| \left(\mathbf{M}_{\mathrm{obs}}^{(\ell)}\mathbf{M}_{\mathrm{obs}}^{(\ell)^{\dagger}}\mathbf{S}_{\mathrm{row}}^{(\ell)} - \mathbf{S}_{\mathrm{row}}^{(\ell)}\right)\mathbf{B} \right|\right|_{\mathrm{F}}^2. \tag{4.3}$$

Each term in the sum can be multiplied from the left by $\mathbf{S}_{\mathrm{row}}^{(\ell)^{\mathrm{T}}}$ since this only adds zero rows and therefore does not change the Frobenius norm of that term. For example, $\mathbf{S}_{\mathrm{row}}^{(\ell)^{\mathrm{T}}}\mathbf{M}_{\mathrm{obs}}^{(\ell)}$ yields a matrix with J rows, in which the rows with indices in α_ℓ are the rows of $\mathbf{M}_{\mathrm{obs}}^{(\ell)}$ and the other rows are zero. After multiplication with $\mathbf{S}_{\mathrm{row}}^{(\ell)^{\mathrm{T}}}$, the

optimization problem (4.3) becomes

$$\min_{\mathbf{B}^{\mathrm{H}}\mathbf{B}=\mathbf{I}_R} \sum_{\ell=1}^{L} \left|\left| \left(\mathbf{S}_{\mathrm{row}}^{(\ell)^{\mathrm{T}}} \mathbf{M}_{\mathrm{obs}}^{(\ell)} \mathbf{M}_{\mathrm{obs}}^{(\ell)^{\dagger}} \mathbf{S}_{\mathrm{row}}^{(\ell)} - \mathbf{S}_{\mathrm{row}}^{(\ell)^{\mathrm{T}}} \mathbf{S}_{\mathrm{row}}^{(\ell)} \right) \mathbf{B} \right|\right|_{\mathrm{F}}^2,$$

or equivalently,

$$\min_{\mathbf{B}^{\mathrm{H}}\mathbf{B}=\mathbf{I}_R} \sum_{\ell=1}^{L} \left|\left| \mathbf{Z}^{(\ell)} \mathbf{B} \right|\right|_{\mathrm{F}}^2, \tag{4.4}$$

with $\mathbf{Z}^{(\ell)} = \mathbf{S}_{\mathrm{row}}^{(\ell)^{\mathrm{T}}} \mathbf{M}_{\mathrm{obs}}^{(\ell)} \mathbf{M}_{\mathrm{obs}}^{(\ell)^{\dagger}} \mathbf{S}_{\mathrm{row}}^{(\ell)} - \mathbf{S}_{\mathrm{row}}^{(\ell)^{\mathrm{T}}} \mathbf{S}_{\mathrm{row}}^{(\ell)}$. The matrix $\mathbf{Z}^{(\ell)}$ is an orthogonal projector, since the relations $\mathbf{Z}^{(\ell)^{\mathrm{H}}} = \mathbf{Z}^{(\ell)}$ and $\mathbf{Z}^{(\ell)}\mathbf{Z}^{(\ell)} = \mathbf{Z}^{(\ell)}$ hold, onto the orthogonal complement of

$$\operatorname{range}\left(\begin{bmatrix} \mathbf{S}_{\mathrm{row}}^{(\ell)^{\mathrm{T}}} \mathbf{M}_{\mathrm{obs}}^{(\ell)} & \mathbf{S}_{\mathrm{row}}^{(\ell^{\mathrm{c}})^{\mathrm{T}}} \end{bmatrix} \right). \tag{4.5}$$

The matrix $\mathbf{S}_{\mathrm{row}}^{(\ell^{\mathrm{c}})^{\mathrm{T}}}$ contains unit columns with ones at the rows with indices that are in α_ℓ^{c}, that is, at the rows that have not been observed in $\mathbf{M}_{\mathrm{obs}}^{(\ell)}$. A matrix $\mathbf{P}^{(\ell)}$ with orthonormal columns that span the subspace in (4.5) can be computed from the SVD $\mathbf{M}_{\mathrm{obs}}^{(\ell)} = \mathbf{U}^{(\ell)}\mathbf{\Sigma}^{(\ell)}\mathbf{V}^{(\ell)^{\mathrm{H}}}$, namely

$$\mathbf{P}^{(\ell)} = \begin{bmatrix} \mathbf{S}_{\mathrm{row}}^{(\ell)^{\mathrm{T}}} \begin{bmatrix} \mathbf{u}_1^{(\ell)} \dots \mathbf{u}_R^{(\ell)} \end{bmatrix} & \mathbf{S}_{\mathrm{row}}^{(\ell^{\mathrm{c}})^{\mathrm{T}}} \end{bmatrix} \in \mathbb{C}^{J \times (R+J-J_\ell)}.$$

The leftmost part of $\mathbf{P}^{(\ell)}$, without $\mathbf{S}_{\mathrm{row}}^{(\ell)^{\mathrm{T}}}$, spans a part of range $(\mathbf{B})$, namely $\operatorname{range}\left(\begin{bmatrix} \mathbf{u}_1^{(\ell)} \dots \mathbf{u}_R^{(\ell)} \end{bmatrix} \right) = \operatorname{range}\left(\mathbf{S}_{\mathrm{row}}^{(\ell)} \mathbf{B} \right)$. The rightmost part $\mathbf{S}_{\mathrm{row}}^{(\ell^{\mathrm{c}})^{\mathrm{T}}}$ then adds unit columns for the rows that have not been observed in $\mathbf{M}_{\mathrm{obs}}^{(\ell)}$, which ensures that $\operatorname{range}(\mathbf{B}) \subseteq \operatorname{range}\left(\mathbf{P}^{(\ell)}\right)$. According to the orthogonal decomposition theorem, the minimization problem in (4.4) is equivalent to the following maximization problem:

$$\max_{\mathbf{B}^{\mathrm{H}}\mathbf{B}=\mathbf{I}_R} \sum_{\ell=1}^{L} \left|\left| \mathbf{P}^{(\ell)} \mathbf{P}^{(\ell)^{\mathrm{H}}} \mathbf{B} \right|\right|_{\mathrm{F}}^2.$$

Here, the multiplication from the left by $\mathbf{P}^{(\ell)}$ can be omitted because $\mathbf{P}^{(\ell)}$ is columnwise orthonormal:

$$\max_{\mathbf{B}^{\mathrm{H}}\mathbf{B}=\mathbf{I}_R} \sum_{\ell=1}^{L} \left|\left| \mathbf{P}^{(\ell)^{\mathrm{H}}} \mathbf{B} \right|\right|_{\mathrm{F}}^2.$$

The maximization problem can further be rewritten as

$$\max_{\mathbf{B}^{\mathrm{H}}\mathbf{B}=\mathbf{I}_R} ||\mathbf{B}^{\mathrm{H}}\mathbf{P}||_{\mathrm{F}}^2 \quad \text{with } \mathbf{P} = \begin{bmatrix} \mathbf{P}^{(1)} & \dots & \mathbf{P}^{(L)} \end{bmatrix}.$$

This shows that range $(\mathbf{B})$ is generated by the left singular vectors of $\mathbf{P}$ that correspond to the R largest singular values. Given that range $(\mathbf{B})$ is a subspace of range $\left(\mathbf{P}^{(\ell)}\right)$ for $\ell = 1, \dots, L$, computing range $(\mathbf{B})$ as the dominant column space of $\mathbf{P}$ corresponds to finding the intersection of the subspaces range $\left(\mathbf{P}^{(\ell)}\right)$ for $\ell = 1, \dots, L$.

The overall computation of a column-wise orthonormal basis for range $(\mathbf{B})$ is described in Algorithm 4.2.

Algorithm 4.2 SVD-based principal subspace method for computing range $(\mathbf{M})$.

Input: Fully observed submatrices $\mathbf{M}_{\text{obs}}^{(\ell)}$ for $\ell = 1, \ldots, L$
Output: Orthonormal basis for range $(\mathbf{M})$
for $\ell = 1, \ldots, L$ **do**
  Compute SVD $\mathbf{M}_{\text{obs}}^{(\ell)} = \mathbf{U}^{(\ell)}\mathbf{\Sigma}^{(\ell)}\mathbf{V}^{(\ell)\text{H}}$
  Set $\mathbf{P}^{(\ell)} \coloneqq \begin{bmatrix} \mathbf{S}_{\text{row}}^{(\ell)\text{T}} \begin{bmatrix} \mathbf{u}_1^{(\ell)} \ldots \mathbf{u}_R^{(\ell)} \end{bmatrix} & \mathbf{S}_{\text{row}}^{(\ell^{\text{c}})\text{T}} \end{bmatrix}$
end for
Compute SVD $\mathbf{P} = \begin{bmatrix} \mathbf{P}^{(1)} & \ldots & \mathbf{P}^{(L)} \end{bmatrix} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\text{H}}$
Set $\mathbf{B} \coloneqq \begin{bmatrix} \mathbf{u}_1 & \ldots & \mathbf{u}_R \end{bmatrix}$ % range $(\mathbf{B})$ = range $(\mathbf{M})$

Discussion. The computational complexity of Algorithm 4.1 and Algorithm 4.2 is dominated by the SVD of the matrix $\mathbf{Q}$ and the matrix $\mathbf{P}$, respectively. Matrix $\mathbf{P}$ is of size $J \times \left(L(R+J) - \sum_{\ell=1}^{L} J_\ell\right)$ and matrix $\mathbf{Q}$ is of size $J \times \left(\left(\sum_{\ell=1}^{L} J_\ell\right) - LR\right)$. This means that Algorithm 4.2 is less computationally expensive than Algorithm 4.1 if $\frac{1}{L}\sum_{\ell=1}^{L} J_\ell > \frac{J}{2} + R$ for $\ell = 1, \ldots, L$, that is, if the number of observed rows in the fully observed submatrices is relatively large. On the other hand, if $J_\ell \ll J$ for $\ell = 1, \ldots, L$, i.e., if the number of rows in the fully observed submatrices $\mathbf{M}_{\text{obs}}^{(\ell)}$ of $\mathbf{M}$ is relatively small, then $\mathbf{P}$ contains $\mathcal{O}(LJ)$ columns and the computational cost for finding the dominant column space of $\mathbf{P}$ increases. Since we have that range $(\mathbf{P})$ = range $(\mathbf{P}\mathbf{P}^{\text{T}})$, finding the dominant column space of range $(\mathbf{P}\mathbf{P}^{\text{T}})$ instead provides an alternative approach with a lower cost. This is because matrix $\mathbf{P}_{\text{tot}} = \mathbf{P}\mathbf{P}^{\text{T}}$ contains only J columns and can be computed without constructing $\mathbf{P}$ first. This alternative approach to computing a column-wise orthonormal basis for range $(\mathbf{B})$ is described in Algorithm 4.3. The tradeoff for the reduction in computational cost compared to Algorithm 4.2 is that the condition number of $\mathbf{P}_{\text{tot}}$ is the square of the condition number of $\mathbf{P}$.

Algorithm 4.3 SVD-based large-scale principal subspace method for computing range $(\mathbf{M})$.

Input: Fully observed submatrices $\mathbf{M}_{\text{obs}}^{(\ell)}$ for $\ell = 1, \ldots, L$
Output: Orthonormal basis for range $(\mathbf{M})$
$\mathbf{P}_{\text{tot}} = \mathbf{0}$
for $\ell = 1, \ldots, L$ **do**
  Compute SVD $\mathbf{M}_{\text{obs}}^{(\ell)} = \mathbf{U}^{(\ell)}\mathbf{\Sigma}^{(\ell)}\mathbf{V}^{(\ell)\text{H}}$
  Keep the dominant left singular vectors $\mathbf{U}^{(\ell)} \leftarrow \begin{bmatrix} \mathbf{u}_1^{(\ell)} & \ldots & \mathbf{u}_R^{(\ell)} \end{bmatrix}$
  Update $\mathbf{P}_{\text{tot}} \leftarrow \mathbf{P}_{\text{tot}} + \mathbf{S}_{\text{row}}^{(\ell)\text{T}}\mathbf{U}^{(\ell)}\mathbf{U}^{(\ell)\text{H}}\mathbf{S}_{\text{row}}^{(\ell)} + \mathbf{S}_{\text{row}}^{(\ell^{\text{c}})\text{T}}\mathbf{S}_{\text{row}}^{(\ell^{\text{c}})}$
end for
Compute SVD $\mathbf{P}_{\text{tot}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\text{H}}$
Set $\mathbf{B} \coloneqq \begin{bmatrix} \mathbf{u}_1 & \ldots & \mathbf{u}_R \end{bmatrix}$ % range $(\mathbf{B})$ = range $(\mathbf{M})$

5. MLSVD of a fiber-wise observed tensor. In this section, we introduce the algorithm for computing an MLSVD from a fiber-wise observed tensor and the cor-

responding uniqueness conditions. The first and second factor matrix of the MLSVD are computed using the matrix completion algorithms from the previous section.

5.1. Algorithm. As mentioned in section 2, $\mathbf{C}$ can be obtained from the dominant left singular vectors of $\mathbf{X}_{[3;2,1]}\mathbf{S}_{\text{fib}}^{\text{T}}$. To reduce the computational complexity of the next steps of the algorithm, the mode-3 fibers can be compressed using $\mathbf{C}$. In this way, the third dimension of the compressed tensor[3] $\mathcal{Y} = \mathcal{X} \cdot_3 \mathbf{C}^{\text{H}}$ is reduced from K to R_3. Note that the product $\mathcal{Y} = \mathcal{X} \cdot_3 \mathbf{C}^{\text{H}}$ does not need to be computed, since a matricized version of $\mathcal{Y}$ can be obtained from the SVD of $\mathbf{X}_{[3;2,1]}\mathbf{S}_{\text{fib}}^{\text{T}}$, which has already been computed for obtaining $\mathbf{C}$.

The remaining factor matrices $\mathbf{A}$ and $\mathbf{B}$ can be computed by applying either Algorithm 4.1, Algorithm 4.2 or Algorithm 4.3 to the incomplete matrix unfoldings $\mathbf{Y}_{[1;2,3]}$ and $\mathbf{Y}_{[2;1,3]}$ respectively.

Last, the core can be computed from a linear system of equations. Given that the tensor $\mathcal{X}$ admits the MLSVD (2.1), the observed part of its mode-3 unfolding is equal to

$$\mathbf{S}_{\text{fib}}\mathbf{X}_{[3;2,1]}^{\text{T}} = \mathbf{S}_{\text{fib}}(\mathbf{B} \otimes \mathbf{A})\mathbf{G}_{[3;2,1]}\mathbf{C}^{\text{T}}.$$

After the initial compression of the mode-3 vectors with $\mathbf{C}$, this becomes

$$\mathbf{S}_{\text{fib}}\mathbf{Y}_{[3;2,1]}^{\text{T}} = \mathbf{S}_{\text{fib}}(\mathbf{B} \otimes \mathbf{A})\mathbf{G}_{[3;2,1]}. \tag{5.1}$$

This is a linear system that can be solved for $\mathbf{G}_{[3;2,1]}$. The overall algorithm for computing the MLSVD of $\mathcal{X}$ is described in Algorithm 5.1. It can easily be generalized to a tensor of order N, see the Supplementary Materials.

Algorithm 5.1 (`mlsvd_fw`) MLSVD of a fiber-wise observed tensor.

Input: Observed part of mode-3 unfolding of tensor $\mathcal{X}$, i.e., $\mathbf{S}_{\text{fib}}\mathbf{X}_{[3;2,1]}^{\text{T}}$
Output: Factor matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and core $\mathcal{G}$
Compute SVD for third mode: $\mathbf{X}_{[3;2,1]}\mathbf{S}_{\text{fib}}^{\text{T}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\text{H}}$
Set $\mathbf{C} \coloneqq \begin{bmatrix}\mathbf{u}_1 & \dots & \mathbf{u}_{R_3}\end{bmatrix}$
Compress observed mode-3 vectors by setting $\mathbf{S}_{\text{fib}}\mathbf{Y}_{[3;2,1]}^{\text{T}} \coloneqq \mathbf{\Sigma}\mathbf{V}^{\text{H}}$
Compute $\mathbf{A}$ from the incomplete unfolding $\mathbf{Y}_{[1;2,3]}$ using Algorithm 4.1, 4.2 or 4.3
Compute $\mathbf{B}$ from the incomplete unfolding $\mathbf{Y}_{[2;1,3]}$ using Algorithm 4.1, 4.2 or 4.3
Solve $\mathbf{S}_{\text{fib}}\mathbf{Y}_{[3;2,1]}^{\text{T}} = \mathbf{S}_{\text{fib}}(\mathbf{B} \otimes \mathbf{A})\mathbf{G}_{[3;2,1]}$ for core $\mathbf{G}_{[3;2,1]}$

5.2. Uniqueness conditions. In order to be able to uniquely retrieve range $(\mathbf{C})$, matrix $\mathbf{X}_{[3;2,1]}\mathbf{S}_{\text{fib}}^{\text{T}}$ containing the observed mode-3 vectors must have rank R_3. The core tensor can be retrieved if the solution to the linear system in (5.1) is unique, which is the case if rank $(\mathbf{S}_{\text{fib}}(\mathbf{B} \otimes \mathbf{A})) = R_1R_2$. The conditions in Theorem 3.4 can be used to guarantee the unique retrieval of range $(\mathbf{A})$ and range $(\mathbf{B})$ from the incomplete matrix unfoldings $\mathbf{X}_{[1;2,3]}$ and $\mathbf{X}_{[2;1,3]}$, respectively. The first two sets of conditions in (3.18) can be satisfied by only selecting fully observed submatrices of $\mathbf{X}_{[1;2,3]}$ $(\mathbf{X}_{[2;1,3]})$ that are of rank R_1 (R_2). In order to differentiate between modes 1 and 2, matrix $\mathbf{\nabla}^{(\mathbf{A})}$ denotes the equivalent of the matrix in (3.13) that is constructed for $\mathbf{A}$ and matrix $\mathbf{\nabla}^{(\mathbf{B})}$ denotes its counterpart for $\mathbf{B}$. For the same reason, the sets

[3]Since $\mathcal{X}$ is fiber-wise observed, the mode-3 product is only computed for the observed fibers of $\mathcal{X}$.

α_ℓ are replaced by $\alpha^{(1)}_{\ell_1}$ and $\alpha^{(2)}_{\ell_2}$, in which the superscript corresponds to the mode of the unfolding. The overall conditions that guarantee that the factor matrices of the MLSVD of a fiber-wise observed tensor can be uniquely retrieved, up to orthogonal transformation, are listed in Theorem 5.1.

THEOREM 5.1. *Consider a fiber-wise observed tensor $\mathcal{X} \in \mathbb{C}^{I\times J\times K}$ of multilinear rank (R_1, R_2, R_3), in which the observed mode-3 vectors are indicated by the ones in $\mathbf{W} \in \{0,1\}^{I\times J}$. Denote the total number of observed mode-3 vectors as F. Given a set of L_1 fully observed submatrices of $\mathbf{X}_{[1;3,2]}$ of rank R_1 and a set of L_2 fully observed submatrices of $\mathbf{X}_{[2;3,1]}$ of rank R_2, the factor matrices $\mathbf{A} \in \mathbb{C}^{I\times R_1}$, $\mathbf{B} \in \mathbb{C}^{J\times R_2}$ and $\mathbf{C} \in \mathbb{C}^{K\times R_3}$ and the core $\mathcal{G} \in \mathbb{C}^{R_1\times R_2\times R_3}$ of the MLSVD of $\mathcal{X}$ can be retrieved if*

$$
\begin{cases}
\operatorname{rank}\left(\mathbf{X}_{[3;2,1]}\mathbf{S}_{\text{fib}}^{\mathrm{T}}\right) = R_3 & [\text{factor } \mathbf{C}] \\
\boldsymbol{\nabla}^{(\mathbf{A})} \text{ has full column rank} & [\text{factor } \mathbf{A}] \\
\cup_{\ell_1\in\{1,\ldots,L_1\}}\alpha^{(1)}_{\ell_1} = \{1,\ldots,I\} & [\text{factor } \mathbf{A}] \\
\boldsymbol{\nabla}^{(\mathbf{B})} \text{ has full column rank} & [\text{factor } \mathbf{B}] \\
\cup_{\ell_2\in\{1,\ldots,L_2\}}\alpha^{(2)}_{\ell_2} = \{1,\ldots,J\} & [\text{factor } \mathbf{B}] \\
\operatorname{rank}\left(\mathbf{S}_{\text{fib}}(\mathbf{B}\otimes\mathbf{A})\right) = R_1R_2 & [\text{core } \mathcal{G}].
\end{cases} \tag{5.2}
$$

Here, $\mathbf{X}_{[3;2,1]}\mathbf{S}_{\text{fib}}^{\mathrm{T}} \in \mathbb{C}^{I_3\times F}$ holds the F observed mode-3 vectors of $\mathcal{X}$ and $\mathbf{S}_{\text{fib}}(\mathbf{B}\otimes\mathbf{A}) \in \mathbb{C}^{F\times R_2R_1}$ holds the corresponding rows of $\mathbf{B}\otimes\mathbf{A}$. The matrices $\boldsymbol{\nabla}^{(\mathbf{A})}$ and $\boldsymbol{\nabla}^{(\mathbf{B})}$ are defined in analogy with (3.13).

In the generic case, the first and last conditions of Theorem 5.1 are satisfied if $F \geq R_3$ and $F \geq R_1R_2$ respectively, in which $F = \sum_{i,j} w_{ij}$ equals the total number of observed mode-3 vectors. Since R_3 is upper bounded by R_1R_2, which is one of the properties of multilinear rank, $F \geq R_1R_2$ implies $F \geq R_3$. A generic variant of Corollary 3.5 can be used to guarantee that $\boldsymbol{\nabla}^{(\mathbf{A})}$ and $\boldsymbol{\nabla}^{(\mathbf{B})}$ have full column rank except in a set of measure zero. The third and fifth conditions in Theorem 5.1 remain the same in the generic case.

6. Numerical experiments. The experiments in the following two subsections make use of synthetic tensors with a given multilinear rank. Such a tensor is generated by sampling the entries of the core tensor and factor matrices from the standard normal distribution. The level of additive Gaussian noise $\mathcal{N}$ added to this tensor $\mathcal{X}$ is quantified by the signal-to-noise ratio (SNR), which is defined as $20\log_{10}\frac{||\mathcal{X}||}{||\mathcal{N}||}$. Since this work considers fiber-wise observed incomplete tensors, only a fraction of the mode-3 fibers of the tensor are available to the algorithm. The available mode-3 fibers are indicated by the mask $\mathbf{W}$. The sampling ratio is the ratio of the number of available entries in a tensor to the total number of entries.

6.1. Matrix completion experiment. In this subsection, the accuracy and computation time of the matrix algorithms, introduced in section 4, are compared on synthetic data. In order to obtain a matrix with an observation pattern similar to Figure 2.2, the matrix to be completed in this experiment is the mode-2 unfolding of a fiber-wise observed tensor. Tensors of size $200 \times 200 \times 200$ with multilinear rank $(10, 10, 10)$ are generated for varying sampling ratios and an SNR of 20 dB, yielding incomplete matrices $\mathbf{M}$ of size 200×4000. The accuracy of the estimated column space of the incomplete matrix is quantified as $\angle\left(\mathbf{B}, \hat{\mathbf{B}}\right)$, which is the maximal canonical angle between the true and estimated subspaces, range $(\mathbf{B})$ and range $\left(\hat{\mathbf{B}}\right)$, respectively.

Figure 6.1 shows the median of the accuracy and computation time, aggregated over 20 repetitions. The three algorithms reach the same level of accuracy. This is because the achievable precision is determined by the noise level and sampling ratio. (Comparatively, numerical differences that have to do with the condition number, as discussed in subsection 4.2, are much smaller.) The computation time[4] of Algorithm 4.1 increases as the sampling ratio increases, whereas the opposite holds for Algorithm 4.2. This illustrates that it is advantageous to use the former when the number of observed rows in the fully observed submatrices is lower, which corresponds to a lower sampling ratio, and to use the latter when more observed rows are available. These results also validate the discussion in subsection 4.2 on computational complexity. The computation time of Algorithm 4.3 slightly increases as the sampling ratio increases. This is because the size of the fully observed submatrices also increases, of which the SVDs are computed; in Algorithm 4.3 these SVDs account for more of the overall cost than in Algorithm 4.1 and Algorithm 4.2. As this experiment confirms that Algorithm 4.3 yields a good level of accuracy at low cost, it will be used to compute the factor matrices $\mathbf{A}$ and $\mathbf{B}$ of the MLSVDs in the following experiments.

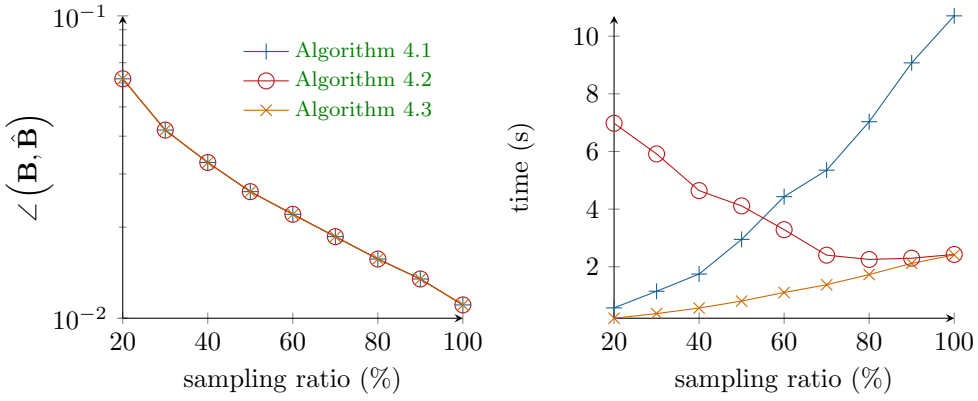


FIG. 6.1. *The three matrix completion algorithms achieve the same level of accuracy. As discussed in subsection 4.2, the computation time of Algorithm 4.1 increases as the sampling ratio increases, whereas the computation time of Algorithm 4.2 decreases. The computation time of Algorithm 4.3 is the lowest at any sampling ratio and slightly increases as the sampling ratio increases, while achieving the same level of accuracy as the other algorithms.*

6.2. Synthetic tensor completion experiments. In this subsection, the approximation error is expressed as the ratio of the Frobenius norm of the residual to the Frobenius norm of the tensor

$$E_{\text{Frob}} = \frac{||\mathcal{X} - \mathcal{G} \cdot_1 \mathbf{A} \cdot_2 \mathbf{B} \cdot_3 \mathbf{C}||_{\text{F}}}{||\mathcal{X}||_{\text{F}}}.$$

Note that this error quantifies how closely the true, noiseless and complete tensor $\mathcal{X}$ is approximated instead of the noisy tensor $\mathcal{X} + \mathcal{N}$. The results reported in figures

[4]The computation of the SVD of matrices $\mathbf{Q}$ and $\mathbf{P}$ in Algorithms 4.1 and 4.2, respectively, can be sped up by using a sparse solver. However, in this basic comparison of the "building block" Algorithms 4.1, 4.2 and 4.3, the standard dense solver in Matlab is used such that the computation time reflects the computational complexity of the algorithms themselves instead of solver specifics.

and tables are obtained by generating multiple random tensors (for each value of the parameter that is varied in that experiment) and then taking the median of the results for each tensor.

In the experiments with synthetic data, the performance of Algorithm 5.1 is compared to that of the following baseline algorithms:

- `geomCG` is a variant of nonlinear conjugated gradients that uses Riemannian optimization techniques on the manifold of tensors of fixed multilinear rank [29]. In the experiments, this optimization algorithm starts from a random starting point, unless stated otherwise.
- `tucker_opt` updates the variables of a low multilinear rank tensor in blocks, that is, the factor matrices and core are updated at each iteration in an alternating fashion. The update for each block of variables is performed using the nonlinear conjugate gradient method [16].
- `HaLRTC` relaxes the problem of minimizing the multilinear rank to minimizing the sum of the nuclear norms of the mode-n unfoldings of the tensor, namely $\sum_{n=1}^{N} ||\mathbf{X}_{[n]}||_*$. The resulting convex problem is solved using the alternating direction method of multipliers [32].

Note that `geomCG` and `tucker_opt` compute a low multilinear rank approximation (LMLRA) instead of an MLSVD. This means that no constraints are imposed on the core tensor $\mathcal{G}$ and factor matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$.

6.2.1. Accuracy and computation time. Figure 6.2 compares the accuracy and computation time obtained by the algorithms on $100 \times 100 \times 100$ tensors of multilinear rank $(8, 8, 8)$ with varying levels of noise. The mask $\mathbf{W}$ has a block diagonal structure with 9 blocks of size 20×20, in which each block has an overlap of size 10×10 with the next block. This overlap must at least be greater than the multilinear rank along the corresponding mode to satisfy the second and fourth conditions in (5.2)[5]. This mask results in a sampling ratio of 28%. The median results over 20 repetitions are shown[6]. At higher values of the SNR, the accuracy obtained by `tucker_opt` flattens off. Algorithm `geomCG` achieves the highest level of accuracy, while `mlsvd_fw` is a bit less accurate. For instance, at 40 dB SNR, `mlsvd_fw` achieves about 2.5 digits of accuracy, while `geomCG` achieves 3 digits of accuracy. On the other hand, `mlsvd_fw` requires much less computation time.

Figure 6.3 concerns $100 \times 100 \times 100$ tensors of multilinear rank $(10, 10, 10)$ at varying sampling ratios and an SNR of 20 dB. The positions of the ones in the mask $\mathbf{W}$ are randomly selected. To compute $\mathbf{A}$ (resp. $\mathbf{B}$), fully observed submatrices of $\mathbf{X}_{[1;2,3]}$ (resp. $\mathbf{X}_{[2;1,3]}$) are now selected using a simple heuristic. Namely, we use the observed parts of the mode-2 (resp. 1) slices (Figure 2.2 bottom) and also the largest fully observed submatrix of each pair of mode-2 (resp. 1) slices (Figure 2.2 top). This figure shows that `HaLRTC` performs badly at lower sampling ratios for fiber-wise observed tensors. Furthermore, `mlsvd_fw` performs badly at the low sampling ratio of 10% because the uniqueness conditions in (5.2) are not yet satisfied. This is

[5]For this mask, only subsequent fully observed submatrices ($\mathbf{M}_{\text{obs}}^{(\ell)}$ and $\mathbf{M}_{\text{obs}}^{(\ell+1)}$) have rows in common. As a result, among all the submatrices $\mathbf{B}_{\text{obs}}^{(\ell_1,\ell_2)}$ in $\boldsymbol{\nabla}^{(\mathbf{B})}$, only $\mathbf{B}_{\text{obs}}^{(\ell_1,\ell_1+1)}$ for $1 \leq \ell_1 < L$ have a nonzero number of rows. The size of the overlap guarantees that each $\mathbf{B}_{\text{obs}}^{(\ell_1,\ell_1+1)}$ has full column rank, which in turn guarantees that $\boldsymbol{\nabla}^{(\mathbf{B})}$ has full column rank. The reasoning for $\boldsymbol{\nabla}^{(\mathbf{A})}$ is analogous.

[6]This observation pattern does not suit `HaLRTC` well, as it achieves better accuracy at the same sampling ratio if the observed entries are spread randomly throughout the tensor. For fairness, this is shown in Figure 3.1 in the Supplementary Materials.

because there are not enough fully observed submatrices of the mode-1 and mode-2 unfoldings of the tensor with the required rank to guarantee that $\nabla^{(\mathbf{A})}$ and $\nabla^{(\mathbf{B})}$ have full column rank. The optimization-based algorithms `geomCG` and `tucker_opt` achieve the highest level of accuracy. The gap between the accuracy obtained by `mlsvd_fw` and the optimization-based algorithms becomes very small at sampling ratios of 40% and higher. Because the selected fully observed submatrices are at most two slices wide, it might be possible to get the accuracy even closer to that of the optimization-based methods by selecting submatrices that span across more than two slices.

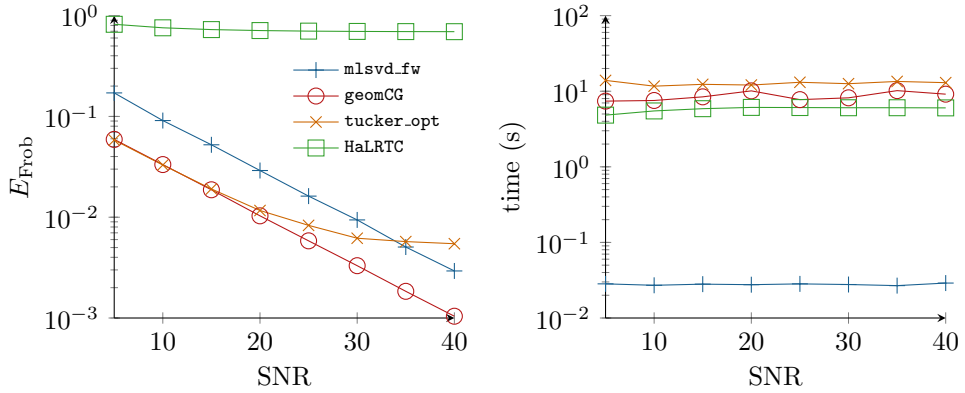


FIG. 6.2. *(left) Accuracy for the fiber-wise observed tensor with a block diagonal mask* $\mathbf{W}$ *in the first experiment. The algorithm* `mlsvd_fw`*, as it is algebraic, obtains less accurate results than the optimization-based* `geomCG` *and* `tucker_opt` *algorithms; however, the difference in accuracy is limited. (right) On the other hand,* `mlsvd_fw` *is two orders of magnitude faster.*

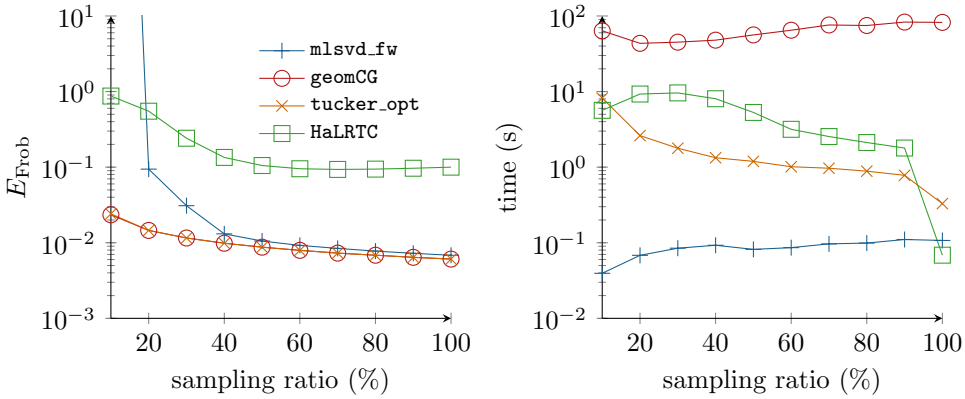


FIG. 6.3. *(left) Accuracy for the fiber-wise observed tensor with a random mask* $\mathbf{W}$ *in the second experiment. At a sampling ratio of* 10% *the working conditions of* `mlsvd_fw` *are not yet satisfied. For higher sampling ratios, the accuracy obtained by* `mlsvd_fw` *is close to that obtained by the optimization-based algorithms. (right) Meanwhile,* `mlsvd_fw` *is much less computationally expensive.*

6.2.2. Initialization of optimization algorithms. A standard approach to computing an LMLRA of a complete tensor is to first truncate the MLSVD, which can be computed algebraically, and further improve it using, for example, higher-order

orthogonal iteration [15] or an optimization algorithm on manifolds [23, 38]. Overall, this reduces the number of iterations required until convergence and it also decreases the risk of convergence to a local, non-global optimum.

Our proposed algorithm offers a similar algebraic initialization strategy in the case in which an LMLRA is computed from a fiber-wise observed tensor. In a new experiment, an LMLRA of multilinear rank $\hat{R}$ is computed of a $50 \times 50 \times 50$ tensor of multilinear rank $R = (5, 5, 5)$ with a sampling ratio of 30% and the positions of the observed fibers are selected at random. An LMLRA is computed with multilinear rank $(\hat{R}_1, \hat{R}_2, \hat{R}_3)$, with $\hat{R}_n$ chosen equal to $R_n + 5, R_n + 1, R_n, R_n - 1$ for $n = 1, 2, 3$. The LMLRA is computed using `geomCG`. The purpose of the experiment is to compare 100 different random initializations of `geomCG` with one run of `geomCG` initialized using our algorithm. For the selection of the fully observed submatrices of $\mathbf{X}_{[1;2,3]}$ (resp. $\mathbf{X}_{[2;1,3]}$) in the computation of $\mathbf{A}$ (resp. $\mathbf{B}$), we took the simplest possible approach and used only the matrices of the type in Figure 2.2 (bottom).

Figure 6.4 compares the two initialization strategies in terms of accuracy and the required number of iterations. In case the multilinear rank is overestimated, random initialization yields much less accurate results than algebraic initialization. The accuracy worsens as the overestimation of the multilinear rank increases, indicating that overfitting occurs since E_{Frob} is the error with respect to the noiseless and complete tensor $\mathcal{X}$. In case the multilinear rank is correctly estimated or underestimated, both the random and algebraic initialization strategies yield approximately equally accurate results, except for a few outliers in the former case. However, the algebraic initialization still proves useful, as it reduces the number of iterations required to reach a solution with this level of accuracy significantly and avoids outliers.

6.2.3. Scalability. In this experiment, we compare the scalability of `mlsvd_fw`, `geomCG` and `tucker_opt`. It was performed on a laptop with an AMD Ryzen 7 PRO 3700U CPU and 30GB RAM. A random tensor of dimensions $I \times I \times I$ with multilinear rank $(5, 5, 5)$ and a sampling ratio of 20% is generated for several values of I. First, a solution is computed using `mlsvd_fw` using the slices approach. The accuracy of this solution is used to set the tolerance of the residual stopping criterion of the optimization algorithms to ensure a fair comparison. In other words, the optimization algorithms stop once they achieve the same level of accuracy as `mlsvd_fw`.

Figure 6.5 shows the total time required for each algorithm for increasing I. In this experiment, the computation time of each algorithm scales similarly as I increases. However, the required computation time for `mlsvd_fw` is respectively about two and three orders of magnitude less than for `geomCG` and `tucker_opt`. Another way to interpret this figure is that the difference between the `mlsvd_fw` curve and the curve of an optimization algorithm is the computation time that can be saved by initializing using `mlsvd_fw`. This shows that `mlsvd_fw` can significantly speed up the computation of an LMLRA for a fiber-wise observed tensor.

6.3. Completion of traffic speed data. Traffic speed data is spatiotemporal and can either be collected by stationary sensors, located at different road segments, or it can be floating car data, that is, data collected by vehicles. However, the resulting datasets have missing elements in both cases. In the former case, missing data might be caused by a sensor failure that persists for a period of multiple days until it is fixed. In the latter case, it could be that no car that collects data passes through a road segment during some intervals. These missing elements are often first filled in before using the data for other purposes, such as modeling traffic, estimating travel times and navigation.

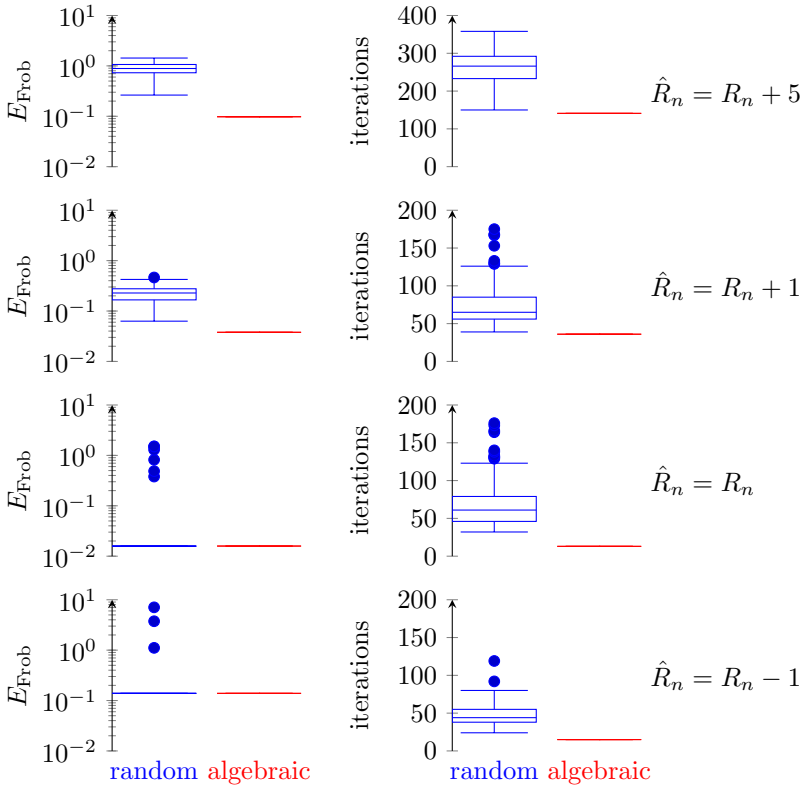


FIG. 6.4. *Comparison of random and algebraic initialization of* `geomCG` *in terms of accuracy (left) and number of iterations (right). In the case where the multilinear rank is overestimated, the optimization-based algorithm has difficulty finding good solutions, starting from* 100 *different random initializations. In the case where it is correctly estimated or underestimated, both initialization strategies yield equally good results in terms of accuracy. However, the algebraic initialization is still useful, since it significantly reduces the number of iterations required to reach this level of accuracy.*

The Guangzhou traffic dataset, which is used in this experiment, was collected from a navigation application and consists of the aggregated speed in 10-minute intervals (144 intervals in each day) from the period from 1 August 2016 to 30 September 2016 (61 days) on 214 road segments [11]. The dataset thus corresponds to a third-order tensor of dimensions $214 \times 61 \times 144$. Data that is missing for some road segments during one or more days correspond to completely missing mode-3 fibers. The data does indeed exhibit such a fiber-wise observation pattern since 1.28% out of the 1.29% missing data corresponds to completely missing fibers. The remaining 0.01% of missing entries are concentrated in three fibers that are partially missing and will be treated as completely missing for convenience.

In order to evaluate the capability of our algorithm to fill in the missing fibers, the available fibers are split into a training and a test set. Here, the sampling ratio refers to the ratio between the elements in the training set and the available elements in the entire dataset. The approximation error on the test set is quantified using the

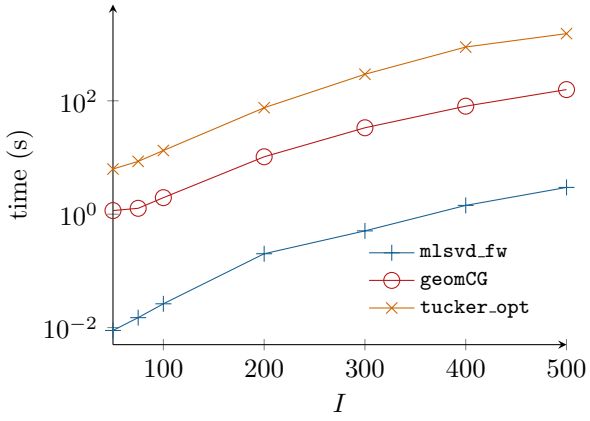


FIG. 6.5. *The computation time required by* `geomCG` *and* `tucker_opt` *to reach the same level of accuracy as* `mlsvd_fw` *is respectively about two and three orders of magnitude larger than the computation time of* `mlsvd_fw` *for the tensor dimensions I that were tested in subsection 6.2.3.*

relative Frobenius norm of the residual and the relative absolute error, namely

$$E_{\text{Frob},\alpha} = \frac{\sqrt{\sum_{i\in\alpha} r_i^2}}{\sqrt{\sum_{i\in\alpha} x_i^2}} \qquad \text{and} \qquad E_{\text{MAE},\alpha} = \frac{\sum_{i\in\alpha} |r_i|}{\sum_{i\in\alpha} |x_i|},$$

in which $\mathbf{r} = \text{vec}\left(\mathcal{X} - \mathcal{G} \cdot_1 \mathbf{A} \cdot_2 \mathbf{B} \cdot_3 \mathbf{C}\right)$ and α contains the indices of the elements of the vectorized traffic data tensor $\mathbf{x} = \text{vec}\left(\mathcal{X}\right)$ that belong to the test set. These error metrics are computed for the results obtained by `mlsvd_fw`, using only the matrices as in Figure 2.2 (bottom), and also for the results obtained after further refinement with `geomCG`. We then compare the achieved accuracy with two related tensor-based algorithms, namely `HaLRTC` and `BGCP`. In `BGCP` the data is modeled as a low-rank tensor, i.e. a CPD, subject to additive independent Gaussian noise [12]. The factor matrices and the noise level are assumed to follow certain prior distributions, which allows the factor matrices to be estimated using Bayesian inference through Gibbs sampling. A multilinear rank of $(20, 7, 20)$ was selected for `mlsvd_fw` using validation and hyperparameters for `HaLRTC` were also chosen using validation. For `BGCP`, the rank was set to 20 as this provided the best results in the case of fiber-wise missing data [12].

For each sampling ratio, 10 repetitions are performed in which the data is randomly split into different training and test sets. Table 6.1 shows the median errors of these repetitions for the different algorithms. In the current experiment, the further optimization-based refinement of the solution obtained by `mlsvd_fw` yields only a small improvement in accuracy. As a matter of fact, the accuracy achieved by `mlsvd_fw` is already comparable to `BGCP`, which was specifically designed for completing traffic data. After a little refinement, `mlsvd_fw` yields the most accurate results in most cases.

7. Conclusion. In this work, we derive an algebraic algorithm for computing the MLSVD of a fiber-wise observed tensor and conditions that guarantee that the algorithm can uniquely retrieve the factor matrices and core tensor. The algorithm is similar in spirit to the basic MLSVD algorithm for fully observed tensors, but with

Sampling ratio (%)	50	60	70	80	90
			$E_{\mathrm{MAE},\alpha}$		
`mlsvd_fw`	0.0756	0.0744	0.0737	0.0731	0.0721
`mlsvd_fw + geomCG`	**0.0730**	**0.0717**	**0.0710**	0.0705	0.0699
`BGCP`	0.0743	0.0723	0.0712	**0.0704**	**0.0697**
`HaLRTC`	0.0823	0.0794	0.0770	0.0753	0.0734
			$E_{\mathrm{Frob},\alpha}$		
`mlsvd_fw`	0.1160	0.1122	0.1102	0.1094	0.1080
`mlsvd_fw + geomCG`	**0.1099**	**0.1068**	**0.1060**	0.1051	0.1044
`BGCP`	0.1122	0.1078	0.1062	**0.1045**	**0.1036**
`HaLRTC`	0.1121	0.1095	0.1073	0.1057	0.1041

TABLE 6.1

Unlike in the experiments with synthetic data, filling in the missing fibers of this traffic speed dataset yields a far smaller difference in accuracy between `mlsvd_fw` and `geomCG`. Additionally, the accuracy achieved by `mlsvd_fw` is comparable to that of the baseline algorithms.

some of the steps consisting of a structured matrix completion problem. Our work included the development of algorithms for computing a column-wise orthonormal basis for the column space of incomplete matrices. Conditions for the unique retrieval of this column space are provided.

The performance of our algorithm is compared to related algorithms in experiments with synthetic and real-life data. The observations are analogous to the performance of the MLSVD of fully observed tensors. The accuracy of the algebraic procedure is often close to the accuracy achieved by optimization-based LMLRA algorithms. However, the algebraic procedure requires much less computation time. As for fully observed tensors, this results in a good initialization strategy for optimization-based algorithms. This strategy is especially useful if the multilinear rank of the underlying tensor is overestimated, as an optimization algorithm may struggle to find accurate solutions in this case. Additionally, the algorithm was applied to an incomplete traffic speed dataset. In this application, the algorithm achieved a level of accuracy similar to or better than that achieved by relevant algorithms from the literature.

REFERENCES

[1] E. ACAR, D. M. DUNLAVY, T. G. KOLDA, AND M. MORUP, *Scalable tensor factorizations for incomplete data*, Chemometrics and Intelligent Laboratory Systems, 106 (2011), pp. 41–56, https://doi.org/10.1016/j.chemolab.2010.08.004.

[2] C. A. ANDERSSON AND R. BRO, *Improving the speed of multi-way algorithms: Part I. Tucker3*, Chemometrics and Intelligent Laboratory Systems, 42 (1998), pp. 93–103, https://doi.org/10.1016/s0169-7439(98)00010-0.

[3] M. ASHRAPHIJUO, V. AGGARWAL, AND X. WANG, *Deterministic and probabilistic conditions for finite completability of low-Tucker-rank tensor*, IEEE Transactions on Information Theory, 65 (2019), pp. 5380–5400, https://doi.org/10.1109/TIT.2019.2919568.

[4] J. BENNETT, S. LANNING, ET AL., *The Netflix Prize*, in Proceedings of KDD cup and workshop, vol. 2007, New York, 2007, p. 35.

[5] S. BHOJANAPALLI AND P. JAIN, *Universal matrix completion*, in Proceedings of the 31st International Conference on Machine Learning, E. P. Xing and T. Jebara, eds., vol. 32 of Proceedings of Machine Learning Research, Bejing, China, 22–24 Jun 2014, PMLR, pp. 1881–1889.

[6] P. Breiding, F. Gesmundo, M. Michałek, and N. Vannieuwenhoven, *Algebraic compressed sensing*, Applied and Computational Harmonic Analysis, 65 (2023), pp. 374–406, https://doi.org/10.1016/j.acha.2023.03.006.
[7] H. Cai, K. Hamm, L. Huang, and D. Needell, *Mode-wise tensor decompositions: Multi-dimensional generalizations of cur decompositions*, Journal of Machine Learning Research, 22 (2021), pp. 1–36, http://jmlr.org/papers/v22/21-0287.html.
[8] C. F. Caiafa and A. Cichocki, *Generalizing the column–row matrix decomposition to multi-way arrays*, Linear Algebra and its Applications, 433 (2010), pp. 557–573.
[9] E. Candès and B. Recht, *Exact matrix completion via convex optimization*, Communications of the ACM, 55 (2012), p. 111–119, https://doi.org/10.1145/2184319.2184343.
[10] E. J. Candes and Y. Plan, *Matrix completion with noise*, Proceedings of the IEEE, 98 (2010), pp. 925–936, https://doi.org/10.1109/jproc.2009.2035722.
[11] X. Chen, Y. Chen, and Z. He, *Urban traffic speed dataset of Guangzhou, China*, Mar. 2018, https://doi.org/10.5281/zenodo.1205229.
[12] X. Chen, Z. He, and L. Sun, *A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation*, Transportation Research Part C: Emerging Technologies, 98 (2019), pp. 73–84, https://doi.org/10.1016/j.trc.2018.11.003.
[13] D. Cheng, N. Ruchansky, and Y. Liu, *Matrix completability analysis via graph k-connectivity*, in Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics, A. Storkey and F. Perez-Cruz, eds., vol. 84 of Proceedings of Machine Learning Research, PMLR, 09–11 Apr 2018, pp. 395–403.
[14] A. Cichocki, D. P. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. F. Caiafa, and A.-H. Phan, *Tensor decompositions for signal processing applications: From two-way to multiway component analysis*, IEEE Signal Processing Magazine, 32 (2015), pp. 145–163, https://doi.org/10.1109/msp.2013.2297439.
[15] L. De Lathauwer, B. De Moor, and J. Vandewalle, *On the best rank-1 and rank-$(R_1, R_2, \cdots, R_N)$ approximation of higher-order tensors*, SIAM Journal on Matrix Analysis and Applications, 21 (2000), pp. 1324–1342, https://doi.org/10.1137/S0895479898346995.
[16] M. Filipović and A. Jukić, *Tucker factorization with missing data with application to low-n-rank tensor completion*, Multidimensional Systems and Signal Processing, 26 (2015), pp. 677–692, https://doi.org/10.1007/s11045-013-0269-9.
[17] A. Franc, *Tensor ranks for the pedestrian for dimension reduction and disentangling interactions*, 2022, https://doi.org/10.48550/ARXIV.2201.07473, https://arxiv.org/abs/2201.07473, https://arxiv.org/abs/2201.07473.
[18] S. Gandy, B. Recht, and I. Yamada, *Tensor completion and low-n-rank tensor recovery via convex optimization*, Inverse Problems, 27 (2011), p. 025010 (19pp), https://doi.org/10.1088/0266-5611/27/2/025010.
[19] R. Ge, J. D. Lee, and T. Ma, *Matrix completion has no spurious local minimum*, in Advances in Neural Information Processing Systems, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, eds., vol. 29, Curran Associates, Inc., 2016.
[20] S. Goreinov, E. Tyrtyshnikov, and N. Zamarashkin, *A theory of pseudoskeleton approximations*, Linear Algebra and its Applications, 261 (1997), pp. 1–21, https://doi.org/ https://doi.org/10.1016/S0024-3795(96)00301-1, https://www.sciencedirect.com/science/article/pii/S0024379596003011.
[21] S. Hendrikx, M. Sørensen, and L. De Lathauwer, *Multilinear singular value decomposition of a tensor with fibers observed along one mode*, in 2023 IEEE Statistical Signal Processing Workshop (SSP), IEEE, 2023, pp. 566–570, https://doi.org/10.1109/SSP53291.2023.10207967.
[22] B. Huang, C. Mu, D. Goldfarb, and J. Wright, *Provable low-rank tensor recovery*, Optimization-Online, 4252 (2014), pp. 455–500.
[23] M. Ishteva, P.-A. Absil, S. Van Huffel, and L. De Lathauwer, *Best low multilinear rank approximation of higher-order tensors, based on the Riemannian trust-region scheme*, SIAM Journal on Matrix Analysis and Applications, 32 (2011), pp. 115–135, https://doi.org/10.1137/090764827.
[24] D. W. Jacobs, *Linear fitting with missing data for structure-from-motion*, Computer Vision and Image Understanding, 82 (2001), pp. 57–81, https://doi.org/10.1006/cviu.2001.0906.
[25] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver, *Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering*, in Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys '10, Barcelona, Spain, 2010, Association for Computing Machinery, p. 79–86, https://doi.org/10.1145/1864708.1864727.

[26] H. Kasai and B. Mishra, *Low-rank tensor completion: A Riemannian manifold preconditioning approach*, in Proceedings of The 33rd International Conference on Machine Learning, New York, New York, USA, 2016, PMLR, pp. 1012–1021.
[27] V. Khoromskaia and B. N. Khoromskij, *Tensor Numerical Methods in Quantum Chemistry*, De Gruyter, Berlin, Boston, 2018, https://doi.org/10.1515/9783110365832.
[28] F. J. Király and R. Tomioka, *A combinatorial algebraic approach for the identifiability of low-rank matrix completion*, in Proceedings of the 29th International Conference on Machine Learning, ICML, Edinburgh, Scotland, UK, 2012, icml.cc / Omnipress.
[29] D. Kressner, M. Steinlechner, and B. Vandereycken, *Low-rank tensor completion by Riemannian optimization*, BIT Numerical Mathematics, 54 (2013), pp. 447–468, https://doi.org/10.1007/s10543-013-0455-z.
[30] P. M. Kroonenberg, *Applied Multiway Data Analysis*, John Wiley & Sons, 2008.
[31] L.-H. Lim, *Tensors in computations*, Acta Numerica, 30 (2021), p. 555–764, https://doi.org/10.1017/S0962492921000076.
[32] J. Liu, P. Musialski, P. Wonka, and J. Ye, *Tensor completion for estimating missing values in visual data*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35 (2013), pp. 208–220, https://doi.org/10.1109/TPAMI.2012.39.
[33] M. W. Mahoney, M. Maggioni, and P. Drineas, *Tensor-cur decompositions for tensor-based data*, in Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '06, New York, NY, USA, 2006, Association for Computing Machinery, p. 327–336, https://doi.org/10.1145/1150402.1150440, https://doi.org/10.1145/1150402.1150440.
[34] C. Mu, B. Huang, J. Wright, and D. Goldfarb, *Square deal: Lower bounds and improved relaxations for tensor recovery*, in Proceedings of the 31st International Conference on Machine Learning, E. P. Xing and T. Jebara, eds., vol. 32 of Proceedings of Machine Learning Research, Bejing, China, 2014, PMLR, pp. 73–81.
[35] D. L. Pimentel-Alarcón, N. Boston, and R. D. Nowak, *Deterministic conditions for subspace identifiability from incomplete sampling*, in 2015 IEEE International Symposium on Information Theory (ISIT), 2015, pp. 2191–2195, https://doi.org/10.1109/ISIT.2015.7282844.
[36] D. L. Pimentel-Alarcón, N. Boston, and R. D. Nowak, *A characterization of deterministic sampling patterns for low-rank matrix completion*, IEEE Journal of Selected Topics in Signal Processing, 10 (2016), pp. 623–636, https://doi.org/10.1109/JSTSP.2016.2537145.
[37] D. L. Pimentel-Alarcón and R. D. Nowak, *Adaptive strategy for restricted-sampling noisy low-rank matrix completion*, in 2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), IEEE, 2015, pp. 429–432, https://doi.org/10.1109/CAMSAP.2015.7383828.
[38] B. Savas and L.-H. Lim, *Quasi-Newton methods on Grassmannians and multilinear approximations of tensors*, SIAM Journal on Scientific Computing, 32 (2010), pp. 3352–3393, https://doi.org/10.1137/090763172.
[39] N. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, *Tensor decomposition for signal processing and machine learning*, IEEE Transactions on Signal Processing, 65 (2017), pp. 3551–3582, https://doi.org/10.1109/tsp.2017.2690524.
[40] M. Signoretto, Q. Tran Dinh, L. De Lathauwer, and J. A. Suykens, *Learning with tensors: A framework based on convex optimization and spectral regularization*, Machine Learning, 94 (2014), pp. 303–351, https://doi.org/10.1007/s10994-013-5366-3.
[41] A. Singer and M. Cucuringu, *Uniqueness of low-rank matrix completion by rigidity theory*, SIAM Journal on Matrix Analysis and Applications, 31 (2010), pp. 1621–1641, https://doi.org/10.1137/090750688.
[42] A. K. Smilde, R. Bro, and P. Geladi, *Multi-way analysis: applications in the chemical sciences*, John Wiley & Sons, 2005.
[43] Q. Song, H. Ge, J. Caverlee, and X. Hu, *Tensor completion algorithms in big data analytics*, ACM Trans. Knowl. Discov. Data, 13 (2019), https://doi.org/10.1145/3278607.
[44] M. Sørensen and L. De Lathauwer, *Fiber sampling approach to canonical polyadic decomposition and application to tensor completion*, SIAM Journal on Matrix Analysis and Applications, 40 (2019), pp. 888–917, https://doi.org/10.1137/17m1140790.
[45] N. Vannieuwenhoven, R. Vandebril, and K. Meerbergen, *A new truncation strategy for the higher-order singular value decomposition*, SIAM Journal on Scientific Computing, 34 (2012), pp. A1027–A1052.
[46] N. Vervliet and L. De Lathauwer, *A randomized block sampling approach to canonical polyadic decomposition of large-scale tensors*, IEEE Journal on Selected Topics in Signal Processing, 10 (2016), pp. 284–295, https://doi.org/10.1109/jstsp.2015.2503260.

[47] N. Vervliet, O. Debals, L. Sorber, and L. De Lathauwer, *Breaking the curse of dimensionality using decompositions of incomplete tensors: Tensor-based scientific computing in big data analysis*, IEEE Signal Processing Magazine, 31 (2014), pp. 71–79, https://doi.org/10.1109/msp.2014.2329429.