Spectrum Cartography via Coupled Block-Term Tensor Decomposition

Guoyong Zhang, Xiao Fu, Member, IEEE, Jun Wang, Member, IEEE
Xi-Le Zhao, Member, IEEE, Mingyi Hong, Member, IEEE

Abstract—Spectrum cartography aims at estimating power propagation patterns over a geographical region across multiple frequency bands (i.e., a radio map)—from limited samples taken sparsely over the region. Classic cartography methods are mostly concerned with recovering the aggregate radio frequency (RF) information while ignoring the constituents of the radio map—but fine-grained emitter-level RF information is of great interest. In addition, many existing cartography methods work explicitly or implicitly assume random spatial sampling schemes that may be difficult to implement, due to legal/privacy/security issues. The theoretical aspects (e.g., identifiability of the radio map) of many existing methods are also unclear. In this work, we propose a joint radio map recovery and disaggregation method that is based on coupled block-term tensor decomposition. Our method guarantees identifiability of the individual radio map of each emitter (thereby that of the aggregate radio map as well), under realistic conditions. The identifiability result holds under a large variety of geographical sampling patterns, including a number of pragmatic systematic sampling strategies. We also propose effective optimization algorithms to carry out the formulated radio map disaggregation problems. Extensive simulations are employed to showcase the effectiveness of the proposed approach.

Index Terms—coupled tensor decomposition, tensor completion, block term decomposition, radio map, slab sampling, fiber sampling, spectrum cartography

I. INTRODUCTION

Improving spectrum efficiency relies on accurate, fine-grained, and agile radio frequency (RF) awareness. Spectrum sensing is considered the first step towards RF awareness. Spectrum sensing is commonly posed as a detection problem, which determines if a frequency band is used or not [1]; some works also consider it as a power spectral density (PSD) estimation problem that recovers a wideband PSD using a sub-Nyquist sampling rate [2], [3]. Spectrum cartography [4] has a more ambitious goal: It aims to construct the received signal power over some bands of interest and a geographical area—i.e., a ‘radio map’ [5].

As radio maps can provide valuable spatial and spectral information, it can be utilized to enhance the performance of many classic wireless communications/networking tasks, e.g., opportunistic access, interference management, networking planning and spectrum surveillance; see a recent survey in [5]. Radio maps also find applications in a number of emerging topics such as indoor positioning [6] and optimal relay deployment for unmanned aerial vehicle (UAV)-based networks [7], [8].

Many early approaches for radio map estimation focus on single-band radio map completion from sparsely sampled measurements over a geographical region. This problem can be understood as an image inpainting problem, and spatial smoothness of the power prorgation patterns is the main prior information exploited for handling the task. Based on this perspective, a number of approaches were proposed, e.g., the nonparametric methods (also called model-free methods) based on kernel regression, leveraging on different kernels such as Kriging [9], thin plate splines (TPS) [10], and Gaussian radial basis functions (RBF) [4]. Some parametric methods based on the knowledge of power propagation models were also proposed, e.g., the sparse representation-based works as in [11]–[14]. Estimating multi-band radio maps has also been considered in the literature [12], [15], [16]. The problem is much harder than the single-band case, since multi-band radio map is a third-order tensor—and tensor completion is a hard task. In the literature, a multi-band radio map is often modeled as a superposition of emitter PSDs scaled by their respective spatial loss fields (SLFs). Some prior knowledge of the SLFs and the PSDs is assumed to be known in [12] and [15], [16], respectively, to simplify the tensor completion task.

Both spectrum sensing and cartography have made considerable progresses since the early 2000s. However, RF awareness-enabling techniques are still far from satisfactory. First, most of the aforementioned techniques only recover aggregate PSD or SLF that are normally contributed by multiple transmitters. Simultaneously estimating PSD and SLF disaggregated to the emitter level has not been addressed. Second, most existing cartography techniques put emphases on the algorithmic aspects, e.g., interpolation and kernel design. However, it has been unclear if in theory the spectral and spatial information is recoverable, given limited samples over space and/or frequency. Third, most cartography methods exhibit the best performance under random spatial sampling (which is perhaps reminiscent of the effectiveness of matrix/tensor completion under randomly missed entries [17], [18]), while random sampling may not always be realistic, because sensing in some
areas may be subject to security/privacy/legal constraints.

Contributions. In this work, we investigate the theoretical and practical aspects of radio map recovery and disaggregation under different measurement acquisition strategies. We propose an analytical framework based on coupled tensor decomposition for radio map disaggregation from partial observations. Our framework is based on modeling the radio maps as tensors following the multilinear rank-$(L_i, L_r, 1)$ block-term decomposition (BTD) model (which will be referred to as the LL1 decomposition model for simplicity) [19]. Our formulations can be understood as joint formulations can be understood as joint formulations can be understood as joint formulations can be understood as joint. Note that our framework is based on modeling the radio maps as tensors with shared latent factors, wherein each factor consists of measurements taken from the whole radio map. We derive identifiability guarantees for the emitter PSDs and SLFs under this framework. Our identifiability results cover a variety of realistic measurement collection schemes, including systematic (non-random) sensor deployment strategies. Note that guaranteed PSD and SLF recovery under systematic sensing schemes is desired in practice, since in many cases random sensor deployment may be hard. On the algorithm side, we propose a block coordinate descent (BCD) algorithmic framework to handle the formulated joint PSD and SLF recovery problems. Extensive simulations corroborate our analysis and show the usefulness of the proposed approach under realistic scenarios with challenging fading and shadowing effects.

We should mention that there are a number of works, such as [16] and [20], [21], which also estimate emitter-level information. However, the work in [16] assumes that the PSD of each emitter is known, which may not be the case in practice. In addition, the approaches in [20], [21] by some of the authors of this work are only concerned with the emitter PSDs, without considering the SLF recovery problem.

A conference version of this work was presented at Asilomar 2019 [22]. The journal version additionally includes more sampling patterns and their associated recovery guarantees, detailed recoverability proofs, algorithm design, and comprehensive numerical study.

Notation. We largely follow the established convention in signal processing. In particular, $x$, $X$, and $\bar{X}$ denote a scalar, a vector, a matrix, and a tensor, respectively. $X(i, j, k)$ (or $x_{i,j,k}$, $\bar{X}_{i,j,k}$) denotes the $(i, j, k)$th element of a third-order tensor. $c_{k,r}$ (or $C(k,r)$) denotes the $(k, r)$th element of a matrix. $c_k$ (or $c(k)$) denotes $k$th element of a vector. The superscripts $^\top$ and $^\dagger$ denote the transpose and the pseudoinverse operations of a matrix, respectively. We use the MATLAB notations to represent a submatrix of a given matrix or a subtensor of a given tensor; e.g., $C(:, r)$ denotes $r$-th column of $C$, and $\bar{X}(:, :, k)$ denotes $k$-th frontal slab of $\bar{X}$. The cardinality of set $K$ is denoted by $|K|$. $\{K\}$ denotes the set $\{1, \ldots, |K|\}$. The outer product ‘$\odot$’ between a matrix $S$ and a vector $c$ is defined as $\bar{X} = S \odot c$ with $(i,j,k)$th element $\bar{X}(i,j,k) = S(i,j)c(k)$. The $L_1$-norm of a vector and matrix are defined as $\|x\|_1 = \sum |x_i|$ and $\|X\|_1 = \sum_{i,j} |x_{ij}|$, respectively. We use $\|x\|_F = \sqrt{\sum x_i^2}$ to denote the Frobenius norms of $X$ and $\bar{X}$, respectively.

II. Problem Statement

We consider a scenario where $R$ emitters exist within a 2-D geographical area of interest. Each emitter transmits across a certain frequency bands, and the frequency bands of different emitters could overlap. To be more precise, assuming that the signals from different emitters (or, sources) are uncorrelated, the received aggregated power spectral density (PSD) at spatial coordinates $(i, j)$ and frequency $f_k$ is represented as [15], [16]:

$$x_{i,j,k} \approx \sum_{r=1}^{R} S_r(i,j)c_{k,r},$$

where $S_r \in \mathbb{R}^{I \times J}$ denotes the (discretized) SLF of source $r$ over the geographical region of interest, $c_{k,r}$ denotes the PSD of source $r$ measured at frequency $f_k$, and $R$ the number of sources. The signal model in (1) means that the received power across frequencies at each geographical point in space is a superposition of the PSD of different sources—scaled by their respective SLFs. Eq. (1) is a valid model if the bandwidth of interest is relatively small compared to the carrier frequency [16], [20], [21] (e.g., if the band of interest spans 20MHz at a carrier frequency within 2-5 GHz—and the bandwidth can be even larger when the carrier frequency is at the range of millimeter wavelengths).

It is natural to represent the received space-space-frequency signal using tensor notations. Let us define a tensor $X \in \mathbb{R}^{I \times J \times K}$ and a matrix $C \in \mathbb{R}^{K \times R}$ such that $X(i,j,k) = x_{i,j,k}$ and $C(k,r) = c_{k,r}$. Then, if noise is absent, the signal model can be expressed as

$$X = \sum_{r=1}^{R} S_r \odot c_r,$$

where $c_r$ denotes the $r$th column of $C$ and $\odot$ denotes the outer product. We refer to the tensor $X$ as a radio map, since it reveals the RF environment across both the spatial and the spectral domains. One remark is that RF environment can change quickly in practice, and thus one may also add a temporal dimension to make the above signal model a fourth-order tensor. Nonetheless, if $X$ is measured as a geographical power spectral density, it is expected to reflect the averaged prominent RF situation over a relatively long period—which is normally informative enough. Hence, in this work, we focus on the spatial-spectral case.

To put into context, our problem is to estimate $S_r$ for $r = 1, \ldots, R$ and $C$ given

$$X(i,j,k), \quad (i, j, k) \in \Omega,$$

and $\Omega$ is the index set of observed entries of $X$. Note that normally $|\Omega| \ll IJF$—i.e., the RF tensor is heavily undersampled.

Note that if one only aims at recovering $X$, the problem is essentially a tensor completion problem under a certain observation model. Many approaches for this problem appeared in the literature [18], [23]–[29], mainly for image inpainting, denoising, and other vision problems. Nevertheless, many of these tensor completion approaches (especially those who admit theoretical supports [27]–[29]) work under an
assumption that the missing/observed values happen uniformly at random over the tensor. Translating this to our problem, it would require that many sensors are randomly deployed all over the geographical region of interest and that the sensors are randomly sensing different frequency bands. In some cases, it is hard to fulfill these requirements, especially when sensing is conducted by some secondary (or unlicensed) systems. Sensor deployment has many restrictions in practice (e.g., due to privacy/legal/security issues), and can hardly be random. This suggests systematic sampling strategies may be more appropriate. More importantly, tensor completion can only recover the ambient data $X$, but not the constituents $S_r$ and $C$—which represent the fine-grained information. In this work, we will offer solutions for radio map disaggregation, in particular, under systematic sampling schemes.

III. PRELIMINARIES: BLOCK DECOMPOSITION INTO MULTILINEAR RANK-$(L_r, L_r, 1)$ TERMS

To better understand our approach, we first introduce some relevant notions and terminologies in tensor analytics. A tensor $X$ is a multidimensional array. Unlike matrices whose rank is mathematically well-defined, the definition of tensor rank is nonsingular [30], [31]. Some popular tensor decomposition models such as the canonical polyadic decomposition (CPD) (previously known as parallel factor analysis (PARAFAC) [32] and the Tucker decomposition [33] have already triggered a large number of applications in sensing and communications [21], [34]–[37].

A. Decomposition in Multilinear Rank-$(L_r, L_r, 1)$ Terms

Beyond the CPD and Tucker models, yet another very useful tensor model is the so-called block term decomposition (BTD) model [19], [38], [39]. BTD subsumes many decomposition models as its special cases and thus is quite general. In particular, we are interested in a special kind of BTD, which is sometimes referred to as tensor decomposition in multilinear rank-$(L_r, L_r, 1)$ terms, or, simply the LL1 tensor decomposition model [40]. To be specific, under the LL1 model, a third-order tensor can be written as [19], [40]

$$X = \sum_{r=1}^{R} (A_r, B_r^\top) \circ c_r,$$

(3)

where $A_r \in \mathbb{R}^{I \times L_r}$ and $B_r \in \mathbb{R}^{J \times L_r}$ for $r = 1, \ldots, R$, $C = [c_1, \ldots, c_R] \in \mathbb{R}^{K \times R}$, and ‘$\circ$’ denotes the outer product as before. In this model, the tensor is a sum of outer products of a rank-$L_r$ matrix (we have rank$(A_r, B_r^\top) = L_r$ if $L_r \leq \min\{I, J\}$ and $A_r$ and $B_r$ have full column rank) and a vector. The block term $(A_r, B_r^\top) \circ c_r$ is a tensor, which is said to have a multilinear rank of $(L_r, L_r, 1)$ [19], [38], [39]. CPD is a special case of this model whose block terms have a multilinear rank of $(1, 1, 1)$; see Fig. 1 for the difference between the popular CPD model and the LL1 model. Also see the relationship between the BTD and Tucker decomposition models in [38].

![Illustration of tensor decomposition models. Top: CPD. Bottom: (L, L, 1)-BTD.](image)

B. Mode-$n$ Unfolding

A third-order tensor $X \in \mathbb{R}^{I \times J \times K}$ admits three modes: columns $X(:, :, k)$, rows $X(:, j, k)$, fibers $X(i, j, :)$—see Fig. 2. There are also three types of slabs associated with a third-order tensor, namely, the horizontal slabs $X(i, :, :)$, the vertical slabs $X(:, j, :)$, and the frontal slabs $X(:, :, k)$, respectively—see Fig. 3.

A tensor can be represented in a matrix form by unfolding (or matricization). For example the mode-1 unfolding of $X$, denoted as $X_1$, can be obtained by stacking all the vectorized horizontal slabs $X(i, :, :)$ as follows:

$$X_1 = [\text{vec}(X(1, :, :)), \ldots, \text{vec}(X(I, :, :))],$$

where vec$(\cdot)$ is the vectorization operator. The mode-2 and mode-3 unfoldings (denoted as $X_2$ and $X_3$) are defined in a similar way as follows:

$$X_2 = [\text{vec}(X(:, 1, :)), \ldots, \text{vec}(X(:, J, :))],$$

$$X_3 = [\text{vec}(X(:, :, 1)), \ldots, \text{vec}(X(:, :, K))].$$

To represent mode-$n$ product using the latent factors $A_r$, $B_r$ and $c_r$, we first introduce two definitions. The first one is the Khatri-Rao product (column-wise Kronecker), defined as

$$A_r \otimes B_r = [A_r(:, 1) \otimes A_r(:, 1), \ldots, A_r(:, L_r) \otimes A_r(:, L_r)] \in \mathbb{R}^{I \times L_r},$$

where $\otimes$ denotes the Kronecker product. The second one is partition-wise Khatri-Rao product between two partitioned matrices. Denote $A = [A_1, \ldots, A_R] \in \mathbb{R}^{I \times \sum_{r=1}^{R} L_r}$, $B = [B_1, \ldots, B_R] \in \mathbb{R}^{J \times \sum_{r=1}^{R} L_r}$ and $C = [c_1, \ldots, c_R] \in \mathbb{R}^{K \times R}$. The partition-wise Khatri-Rao product is defined as [19]:

$$C \odot p A = [c_1 \odot A_1, \ldots, c_R \odot A_R] \in \mathbb{R}^{IK \times \sum_{r=1}^{R} L_r}.$$
where \( S(:, r) = \text{vec}(A_r B_r^\top) \). For more details, see [19], [38], [39]. These representations will be handy in our theoretical analysis and algorithm design.

\[ C. \text{ Mode Product and BTD Representation} \]

Consider \( X \in \mathbb{R}^{I \times J \times K} \) and \( P_1 \in \mathbb{R}^{M \times I}, P_2 \in \mathbb{R}^{N \times J}, P_3 \in \mathbb{R}^{T \times K} \). The mode-\( n \) product \( G_n = X \times_n P_n \) for \( n = 1, 2, 3 \) such that \( G_1(m, j, k) = \sum_{i=1}^I X(i, j, k)P_1(m, i) \), \( G_2(i, n, k) = \sum_{j=1}^J X(i, j, k)P_2(n, j) \), and \( G_3(i, j, t) = \sum_{k=1}^K X(i, j, k)P_3(t, k) \). A key observation is that, if \( X \) follows the LL1 decomposition model as in (3), we have

\[ X \times_1 P_1 \times_2 P_2 \times_3 P_3 = \sum_{r=1}^R (P_1 A_r (P_2 B_r^\top)) \circ (P_3 c_r). \]

\[ D. \text{ Essential Uniqueness} \]

The LL1 model has a very nice property—\( \{A_r, B_r^\top\}_{r=1}^R \) and \( \{c_r\}_{r=1}^R \) are identifiable up to permutation and scaling ambiguities, under some mild conditions. In the model (3), permutation and scaling ambiguities mean that one can arbitrarily permute the terms \( \{(A_r, B_r^\top) \circ c_r\}_{r=1}^R \), and scale and counter-scale \( A_r, B_r^\top \), and \( c_r \) as long as their outer product remains unchanged. Note that \( A_r B_r^\top = A_r F_r B_r (F_r^\top) \top \) holds for any non-singular matrix \( F_r \in \mathbb{R}^{L_r \times L_r} \). Hence, \( A_r \) and \( B_r \) cannot be identified from the LL1 decomposition of \( X \) in general. In this paper, we will use the following theorem:

\[ \text{Theorem 1 [19]} \]

Let \( X = \sum_{r=1}^R (A_r, B_r^\top) \circ c_r \) and assume \( L_r = L \) for \( r = 1, \cdots, R \). Assume that \( \{A, B, C\} \) are drawn from any joint absolutely continuous distributions. Suppose the following condition holds:

\[ K \geq R \text{ and } \min \left( \left\lfloor \frac{I}{L} \right\rfloor, R \right) + \min \left( \left\lfloor \frac{J}{L} \right\rfloor, R \right) \geq R + 2. \]

Denote \( S = [\text{vec}(S_1), \cdots, \text{vec}(S_R)] \in \mathbb{R}^{I_j \times R} \) where \( S_r = A_r B_r^\top \). Then, with probability one, for any \( \tilde{S} \in \mathbb{R}^{I_j \times R} \) and \( \tilde{C} \in \mathbb{R}^{K \times R} \) satisfying \( X_3 = \tilde{S} \tilde{C}^\top \), we have

\[ \tilde{C} = \text{SIA}, \tilde{S} = \text{SIA}^{-1}, \]

where \( \Pi \) is a permutation matrix and \( \Lambda \) is a nonsingular diagonal matrix; i.e., the multilinear rank-\((L_r, L_r, 1)\) decomposition of \( X \) is essentially unique almost surely.

Many more conditions for the essential uniqueness of BTD were proposed in [19], [40]—e.g., the following one that considers the case where \( K \geq R \) is relaxed:

\[ \text{Theorem 2 [19]} \]

Let \( X = \sum_{r=1}^R (A_r B_r^\top) \circ c_r \) and assume \( L_r = L \) for \( r = 1, \cdots, R \). Assume \( \{A, B, C\} \) are drawn from certain joint absolutely continuous distributions. If \( \lfloor I \rfloor \geq 2R^2 + 2R + 1 \), then, the multilinear rank-\((L_r, L_r, 1)\) decomposition of \( X \) is essentially unique almost surely.

\[ \text{IV. PROPOSED APPROACH} \]

In this section, we propose a framework for radio map estimation and disaggregation using down-sampled data. Our method provably works with both systematic and random sensor deployment strategies. For the former one, the sensing patterns are under control of the system designers. Another salient feature of our approach is that it guarantees identifiability of the SLFs and power spectra for all the emitters in the region of interest—i.e., it allows emitter-level radio map reconstruction—under some conditions.

\[ \text{A. Signal Model} \]

Our idea is to connect the radio map signal model in (2) with the LL1 model in (3), and then use a judiciously designed decomposition criterion to recover the latent factors of the heavily down-sampled tensor—thereby achieving the goal of disaggregation.

Our approach is motivated by the following key postulate: the SLF of a source \( r \), denoted by \( S_r \in \mathbb{R}^{I_j \times J} \), is approximately low-rank for each individual source. This assumption is reasonable, because of the continuity and smoothness of power propagation over space [41]–[44]. A simulated example can be found in Fig. 4, where the SLF of a source over a 100m \( \times 100m \) region is simulated under realistic shadowing effect following the log-normal shadowing model with a standard deviation of 4 (which means a relatively strong shadowing effect) [45, Chapter 4]. We apply the singular value decomposition (SVD) to the SLF and plot the ratio \( \tau_k = \sigma_k / \sum_{k=1}^{100} \sigma_k \), where \( \sigma_k \) denotes the \( k \)th singular value of \( S_r \) in the figure. From Fig. 4, one can see \( \tau_5 = 0.90293 \), which means that the top-5 rank-one components contain 90% energy of \( S_r \). This observation suggests that the SLFs, even under shadowing effects, may be well-approximated by low-rank matrices.
Under the low-rank postulate, one can express $S_r$ by

$$S_r = A_r B_r^\top,$$

where $A_r \in \mathbb{R}^{I \times L_r}$, $B_r \in \mathbb{R}^{J \times L_r}$ and $L_r \ll \min\{I, J\}$. Consequently, we have

$$X = \sum_{r=1}^{R} S_r \circ c_r = \sum_{r=1}^{R} (A_r B_r^\top) \circ c_r;$$

i.e., the complete RF tensor follows a multilinear rank $(L_r, L_r, 1)$ BTD model. Based on this model, we will propose a number of sensing and $(S_r, c_r)$-estimation schemes in the following subsections.

B. The Moving Sensor Case: Sensing Consecutive Grids

If $X$ is available, then estimating $S_r$ and $c_r$ can be done by directly applying LL1 algorithms (e.g., that in [39]) to the tensor. In practice, however, the complete $X$ is not available. What we observe is a substantially undersampled version of $X$. One possible route is to first apply tensor completion techniques to estimate the complete $X$, and then seek the LL1 decomposition of the estimated $X$. However, random sampling schemes that may ensure provable completion of $X$ may not be possible in spectrum cartography—since sensor may not be allowed acquire samples in many regions because of legal/security/privacy issues.

To handle the above challenges, consider that one employs a moving sampling device (e.g., a UAV or a regular/self-driving car) to sample an area of interest. We assume that the spatial area is finely discretized into $I \times J$ grids, and all the grids on the route of the sampling device can be observed over a certain frequencies.

To illustrate our idea, we use a two-sensor scenario as an example—but the approach can be readily generalized to multiple sensor cases. One possible sampling scheme is illustrated in Fig. 5. In this example, two sensors are employed—they are responsible for horizontal route and the vertical route in Fig. 5, respectively. In addition, the two sensors cover different (but overlapping) frequency bands.

Assume that the moving sensors collect power spectral measurements at every $(i, j)$ grid on their routes over the assigned frequency bands. After the moving sensors completing the designed tour, two sets of samples of the RF tensor are collected. To be precise, let $S_1 \subset \{1, \cdots, I\}$ and $S_2 \subset \{1, \cdots, J\}$ contain the observed indices along the two spatial dimensions by the two sensors, respectively. Similarly, let $S_3, S_4 \subset \{1, \cdots, K\}$ be the observed indices along the frequency dimensions by the two sensors, respectively. Thus, two subtensors $X^{(1)} = X(S_1, \cdot, S_3)$ and $X^{(2)} = X(:, S_2, S_4)$, with $|S_1| = M_1$, $|S_2| = N_2$, $|S_3| = K_1$ and $|S_4| = K_2$ which satisfy $S_3 \cup S_4 = [K] = \{1, \ldots, K\}$. Define row-selection matrices $P = I_I(S_1, \cdot)$, which consists of the rows indexed by $S_1$ of an $I \times I$ identity matrix. Similarly, let $Q = I_J(S_2, \cdot)$, $R_1 = I_K(S_3, \cdot)$ and $R_2 = I_K(S_4, \cdot)$. The sampled measurements by sensor 1 and sensor 2 can be represented as follows:

$$X^{(1)} = X \times_1 P \times_3 R_1 = X(S_1, \cdot, S_3),$$

$$X^{(2)} = X \times_2 Q \times_3 R_2 = X(:, S_2, S_4).$$

C. Coupled Tensor Decomposition-Based Formulation

To proceed, recall our postulate that $S_r = A_r B_r^\top$ where $A_r \in \mathbb{R}^{I \times L_r}$ and $B_r \in \mathbb{R}^{J \times L_r}$. Hence, our sampled data can be written as

$$X^{(1)} = \sum_{r=1}^{R} ((P A_r) B_r^\top) \circ (R_1 c_r),$$

$$X^{(2)} = \sum_{r=1}^{R} (A_r (B_r^\top Q^\top)) \circ (R_2 c_r).$$

It is not hard to notice that $X^{(1)}$ and $X^{(2)}$ are two tensors that still follow the LL1 model with $R$ rank-$(L_r, L_r, 1)$ block terms. More importantly, the latent factors of the original tensor, i.e., the matrices $A_r$‘s, $B_r$’s and $C$, are captured in the two observations. Intuitively, if both $X^{(1)}$ and $X^{(2)}$ admit unique LL1 decompositions, then, one may estimate $A_r$ from $X^{(2)}$ and estimate $B_r$ from $X^{(1)}$, respectively, and then assemble the information to get $S_r$.

The idea is seemingly appealing, but has several significant hurdles. First, recall that the LL1 decomposition does not give exact $A_r$, $B_r$, but only the range spaces of them [39]. To be specific, by decomposing $X^{(1)}$, one will have $\hat{B}_r = B_r F_r$ if the BTD problem is optimally solved, where $F_r \in \mathbb{R}^{L_r \times L_r}$ is nonsingular. Similar, from $X^{(2)}$, one will have $\hat{A}_r = A_r G_r$.

In general, $G_r \neq F_r^{-1}$, and thus $\hat{S}_r \neq \hat{A}_r B_r^\top$, which means that the reconstruction fails. Second, if one of $X^{(1)}$ or $X^{(2)}$ does not admit a unique LL1 decomposition, then individually decomposing the two tensors will not lead to sensible results.
To handle these challenges, we propose the following coupled factorization criterion:

$$\text{minimize}_{\{A_r, B_r, C\}} \left\| \mathbf{X}^{(1)} - \sum_{r=1}^{R} ((PA_r)B_r^T) \circ (R_1 C_r) \right\|^2_F \text{ (8)}$$

$$+ \left\| \mathbf{X}^{(2)} - \sum_{r=1}^{R} (A_r(B_r^TQ^T)) \circ (R_2 C_r) \right\|^2_F \text{.}$$

The idea is to use shared parametrization to enforce that the two 'marginalized' tensors $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ share the same $A_r$, $B_r$, and $C$. We show that the following theorem holds:

**Theorem 3** Assume that (6) and (7) hold, and that $A_r, B_r$ for all $r$ and $C$ are drawn from any joint absolutely continuous distributions. Also assume that $|S| \geq R$, where $S = S_1 \cap S_2$, and that $L_1 = L$ for all $r$. Suppose that one of the following conditions holds

(1) $M \geq 2L$, $J \geq LR$, and $\min(|M/L|, R) + \min(|J/L|, R) \geq R + 2$;

(2) $N \geq 2L$, $I \geq LR$, and $\min(|N/L|, R) + \min(|I/L|, R) \geq R + 2$.

Then, solving Problem (8) identifies $\{S_r, C_r\}$ for $r = 1, \ldots, R$ up to scaling and permutation ambiguities, almost surely.

The proof is relegated to Appendix A in the supplementary materials. The proof is reminiscent of coupled tensor decomposition under the CPD model [46], [47]—whereas the proof of Theorem 3 is more challenging since $A_r$ and $B_r$ are not identifiable (as opposed to the case under the CPD model). Theorem 3 says that if the SLFs have low rank and the number of emitters are not too large, then, the proposed sampling strategy and the formulation guarantee recovering $S_r$ and $C_r$ for $r = 1, \ldots, R$, given $M$ and $N$ (i.e., the numbers of the partially sampled horizontal and vertical slabs, respectively) exceeding a certain threshold.

**Remark 1** We would like to mention some interesting implications of Theorem 3. First, the sensors are only required to collect data from a certain number of slabs, but it does not matter which slabs should be used. In particular, the sampled slabs do not necessarily need to be equi-spaced. Second, the sensing pattern is completely regular, without involving randomness. These properties entail the system designers flexibility to circumvent practical issues.

**Remark 2** Theorem 3 addresses identifiability of the two-sensor case that we used as an illustrative example. In practice, many more sampling patterns involving more sensors can be shown to have similar results. Nevertheless, the goal of this work is offer proof-of-concept, rather than exhausting all feasible scenarios.

**D. The Static Sensor Case: Sensing Scattered Grids**

Moving sensors have the advantage of high mobility and flexibility. However, moving sensors are not always available. In this subsection, we consider scenarios where sensors are deployed in the geographical area in advance and sense at their respective locations simultaneously. Such a setting is more conventional [4], [10], [11], [48].

While in theory the sampling patterns discussed in the previous subsection (cf. Fig. 5) can also be applied using static sensors, some times one may also hope to place the sensors over the region in a scattered way (rather than placing the sensors in consecutive grids like in Fig. 5). This kind of sensor deployment is related to a concept called fiber sampling [47], [49]. A fiber of the tensor $\mathbf{X}$ refers to $\mathbf{X}(i, j)$; see Fig. 2. In our context, a fiber is a power spectrum of the received signal measured at the geographical location $(i, j)$. We assume that the sensor at location $(i, j)$ collects the spectral information of bands indexed by $\mathcal{K} \subseteq \{1, \ldots, K\}$. Then, the collected data is denoted as $\mathbf{X}(i, j, \mathcal{K})$—which can be the whole fiber or a part of the fiber $\mathbf{X}(i, j, \cdot)$.

1) **Systematic Sensor Deployment:** To proceed, let us consider a scenario where the locations of sensors are under control—i.e., system designers can design the deployment strategy of the sensors. Suppose that we have $D$ groups of sensors. We assume that group $d$ collects data from bands indexed by $\mathcal{K}^{(d)}$. In addition, assume that the locations of group $d$ sensors are indexed by all combinations possible $(i_d, j_d)$ such that $i_d \in \mathcal{I}^{(d)}$ and $j_d \in \mathcal{J}^{(d)}$. This way, the sampled data from group $d$ forms a subtensor $\mathbf{X}^{(d)} = \mathbf{X}(\mathcal{I}^{(d)}, \mathcal{J}^{(d)}, \mathcal{K}^{(d)})$. Since $\mathbf{X}$ admits a rank-$\{L_r, L_r, 1\}$ BTD, the sampled subtensor $\mathbf{X}^{(d)}$ can be written as

$$\mathbf{X}^{(d)} = \sum_{r=1}^{R} (A_r(\mathcal{I}^{(d)}), :B_r(\mathcal{J}^{(d)}, :)^T) \circ c_r(\mathcal{K}^{(d)}). \text{ (9)}$$

To estimate $C$ and $S_r$, we propose the following criterion:

$$\text{minimize}_{\{A_r, B_r, C\}} \sum_{d=1}^{D} \left\| \mathbf{X}^{(d)} - \sum_{r=1}^{R} S_r(\mathcal{I}^{(d)}, \mathcal{J}^{(d)}) \circ c_r(\mathcal{K}^{(d)}) \right\|^2_F \text{,} \text{ (10)}$$

where we have $S_r(\mathcal{I}^{(d)}, \mathcal{J}^{(d)}) = A_r(\mathcal{I}^{(d)}, :)B_r(\mathcal{J}^{(d)}, :)^T)$. In terms of identifiability of the SLFs and PSDs, we show the following theorem:

**Theorem 4** Assume that (9) holds and $L_r = L$ for all $r$. Also assume that $A_r, B_r$ for all $r$ and $C$ are drawn from certain joint absolutely continuous distributions. Suppose that index sets $\mathcal{I}^{(d)}, \mathcal{J}^{(d)}, \mathcal{K}^{(d)}$ satisfy

$$\text{\cup}_{d=1}^{D} \mathcal{I}^{(d)} = [I], \text{ \cup}_{d=1}^{D} \mathcal{J}^{(d)} = [J], \text{ \cup}_{d=1}^{D} \mathcal{K}^{(d)} = [K], \text{ (11)}$$

$$|\mathcal{I}^{(d)}| \geq L, |\mathcal{J}^{(d)}| \geq L, |\mathcal{K}^{(d)}| \geq R,$$

and that

$$\min \left( \left\| \frac{\mathcal{I}^{(d)}}{L} \right\|_R, \min \left( \left\| \frac{\mathcal{J}^{(d)}}{L} \right\|_R \right) \right) \geq R + 2. \text{ (12)}$$

Assume that there exists a permutation of set $[D]$, denoted as $\{i_1, \ldots, i_D\}$, satisfying $\max(|\mathcal{I}^{(i_1)} \cap \mathcal{I}^{(i_2)}|, |\mathcal{J}^{(i_3)} \cap \mathcal{J}^{(i_4)}|) \geq L$ and $|\mathcal{K}^{(i_d)} \cap \mathcal{K}^{(i_{d+1})}| \geq 2$ for $d = 1, \ldots, D - 1$. Then, $\{S_r, c_r\}_{r=1}^{R}$ can be identified via solving Problem (10) up to scaling and permutation ambiguities almost surely.
Fig. 6. An example of $D = 3$ subtensors that satisfy the conditions of Theorem 4—i.e., $i_1 = 1, i_2 = 3, i_3 = 2$ and $|T^{(1)}|, |T^{(2)}| \geq L$, $|T^{(i_2)}| \geq L, |K^{(i_2)}| \geq 2, |K^{(i_2)}| \geq 2$. (a) The overlapped index sets of the first and second dimensions of $X^{(1)}, X^{(2)}, X^{(3)}$. (b) The overlapped index sets of the third dimension of $X^{(1)}, X^{(2)}, X^{(3)}$. (c) A ‘scattered’ version by permuting the spatial sampling pattern in (a).

The proof is relegated to Appendix B in the supplementary materials. Fig. 6 shows an example satisfying the condition in Theorem 4 where $D = 3$ groups of sensors are deployed. In particular, Fig. 6 (c) shows a deployment pattern where the sensors are scattered rather than being placed in consecutive grids.

One caveat is that Theorem 4 requires all $X^{(d)}$’s being identifiable LL1 tensors. However, the following theorem shows that, if there is one group of sensors that collect samples from all the frequency bands, and the corresponding subtensor is identifiable, the whole tensor is identifiable:

**Theorem 5** Assume that (9) holds. Also assume that $A_r, B_r$ for all $r$ and $C$ are drawn from certain joint absolutely continuous distributions. Suppose that the index sets $T^{(d)}, J^{(d)}, K^{(d)}$ satisfy $\sqcup_{d=1}^D T^{(d)} = [I], \sqcup_{d=1}^D J^{(d)} = [J]$, and $|T^{(d)}| \geq L, |J^{(d)}| \geq L$ for $d = 1, \ldots, D$. Assume that there exists $d_0 \in [D]$ that satisfies $K^{(d_0)} = [K]$. In addition, assume that $K \geq R$ and

$$\min \left( \left\lceil \frac{|J^{(d_0)}|}{L} \right\rceil, R \right) + \min \left( \left\lceil \frac{|T^{(d_0)}|}{L} \right\rceil, R \right) \geq R + 2,$$

and that there exists a permutation of set $[D]$, denoted as $\{i_1, \ldots, i_D\}$, satisfying $\max(|T^{(i_d)} \cap T^{(i_{d+1})}|, |J^{(i_d)} \cap J^{(i_{d+1})}|) \geq L$ for $d = 1, \ldots, D - 1$. Then, $\{S_r, c_r\}_{r=1}^R$ can be identified via solving Problem (10) up to scaling and permutation ambiguities almost surely.

The proof is relegated to Appendix C in the supplementary materials. One can see that compared to Theorem 4, the above theorem only needs one of $X^{(d)}$, namely $X^{(d_0)}$, to be identifiable, and the remain subtensors do not need to be identifiable and are only required to overlap at least $L$ indices in the first or second dimensions. However, sensors in group $d_0$ are required to sample all the frequency bands of interest under Theorem 5, while none of the groups under Theorem 4 are required to do so—but they need to cover larger space in order to recover the same tensor $\check{X}$. Apparently, these two theorems present an interesting spectral–spatial sampling trade-off.

**Remark 3** The work in [49] considered a similar setting as that in Theorems 4-5 and showed identifiability of a coupled LL1 decomposition criterion. Notably, the results in [49] can also be applied to our case. The key difference is that the work in [49] assumes full observation on the third dimension of the tensor, which means that every sensor observes all the frequency bands of interest—which is not assumed in Theorems 4-5. The work in [47] also considered fiber sampling and tensor recovery, but under the CPD model instead of the LL1 model.

2) Random Sensor Deployment: Many existing radio map recovery methods considered random sampling over the space. As a side contribution, we also show that under random sampling radio map disaggregation is possible. Let us consider a formulation when randomly deployed sensors are used:

$$\text{minimize } A, B, C \quad \| W \ast \left( X - \sum_{r=1}^R (A_r B_r^T) \circ c_r \right) \|_F^2, \quad (12)$$

where $\ast$ denotes the Hadamard product (also known as entrywise product) and $W$ is a tensor such that $W(i, j, k) = 1$ if $(i, j, k) \in \Omega$, and $W(i, j, k) = 0$ otherwise. In terms of identifiability, it can be seen that

**Proposition 1** Under the exact LL1 model of $X$, assume that, w.l.o.g., $I \leq J, J > (LR + 1)(L + 1)$ and $LR \leq \frac{1}{\epsilon}$. Also assume that $\{A_r, B_r\}$, and $C$ are drawn from any joint continuous distribution. Suppose that noise is absent. Given $0 < \epsilon \leq 1$, if $q$ entries per column of the frontal slabs $X(:, :, k), k = 1, \ldots, K$ are observed uniformly at random, where $q \geq \max \{12 \log (\frac{1}{\epsilon}) + 2LR\}$, then solving Problem (12) recovers $\{S_r, c_r\}$ for $r = 1, \ldots, R$ with a probability at least $1 - \epsilon$.

The proof is presented in Appendix D in the supplementary materials and is straightforward. In a nutshell, if the matrix rank of each slab $X(:, :, k)$ is low enough, then the slabs are identifiable—via matrix completion. Then, the SLFs and PSDs are identifiable by classic LL1 model identifiability results. Following this insight, and with careful derivations, the above theorem can be shown.

We would like to remark that Proposition 1 merely states that the SLFs and PSDs of interest can be identified under random sensor deployment. However, it does not demonstrate advantages in theory for using the LL1 model-based formulation as in (12), as opposed to performing slab-by-slab matrix completion and then factoring the completed tensor into block terms. Investigating these aspects is quite intriguing, but may go beyond the scope of this work. We defer this to a potential study.
future work. Nonetheless, in practice, using the formulation in (12) may be beneficial for a number of reasons. For example, explicitly using prior knowledge (the underlying LL1 model in this case) is known to be helpful in fending against noise. Also using an all-at-once optimization criterion for estimating the PSDs and SLFs is less possible to suffer from error propagation, compared to slab-by-slab matrix completion and then applying LL1 decomposition, as we will see shortly in the simulations.

V. ALGORITHMS AND PRACTICAL IMPLEMENTATION

In this section, we propose algorithms for handling the formulated factorization criteria in the previous section.

A. Algorithm for Solving (8)

Let us denote the objective function of (8) as \( f(A, B, C) \). In practice, instead of trying to minimize \( f(A, B, C) \), we use the following regularized version

\[
\min_{\{ A, B, C \}} \underbrace{\frac{f(A, B, C) + \lambda_1 \| A \|_F^2 + \lambda_2 \| B \|_F^2 + \lambda_3 \| C \|_F^2}{\text{Loss}(A, B, C)}}_{\text{Regularized Loss}}
\]

(13)

where the regularization terms are added in order to balance the scaling of the latent factors.

We propose to employ a block coordinate descent (BCD) scheme for handling Problem (13). To be specific, \( A, B, C \) are updated as follows:

\[
\begin{align*}
A^{t+1} &\leftarrow \arg \min A \text{ Loss}(A, B^t, C^t), \\
B^{t+1} &\leftarrow \arg \min B \text{ Loss}(A^{t+1}, B, C^t), \\
C^{t+1} &\leftarrow \arg \min C \text{ Loss}(A^{t+1}, B^{t+1}, C),
\end{align*}
\]

(14)

where superscript \( t \) is the iteration index. Note that all the subproblems above are unconstrained least squares problems and thus can be readily solved; see details in Appendix E in the supplementary materials.

As a standard exact BCD algorithm, the algorithm in (14) enjoys convergence properties as discussed in [50], [51]. Specifically, every limit point of the produced solution sequence is a stationary point of (13), when \( \lambda_n > 0 \) for \( n = 1, 2, 3 \)—since the subproblems are always strictly convex.

B. Algorithm for Solving Problems (12)

The basic idea for handling (12) is again BCD as in (14). Nevertheless, since the sampling pattern is quite different, the subproblems are solved in different ways. To be specific, defining \( Y = W \ast \tilde{X} \), let \( Y_n \) and \( W_n \) be mode-\( n \) unfoldings of \( Y \) and \( W \), respectively. The update rules for the regularized version of Problem (12) are as follows:

\[
\begin{align*}
A &\leftarrow \arg \min A \| Y_1 - W_1 \ast (C \circ_B A) A^T \|_F^2 + \lambda_1 \| A \|_F^2, \\
B &\leftarrow \arg \min B \| Y_2 - W_2 \ast (C \circ_p A) B^T \|_F^2 + \lambda_2 \| B \|_F^2, \\
C &\leftarrow \arg \min C \| Y_3 - W_3 \ast \hat{S} C^T \|_F^2 + \lambda_3 \| C \|_F^2,
\end{align*}
\]

where \( \hat{S} = [\text{vec}(A_1 B_1^T), \cdots, \text{vec}(A_K B_K^T)] \). Each subproblem boils down to a set of unconstrained least squares problems—see details in Appendix F in the supplementary materials.

Remark 4 An algorithm for handling Problem (10) can be handled by the same algorithm, with a slightly different definition of \( W \)—to take into consideration that group \( d \) and group \( d' \) are overlapped in space. To be specific, if \((i, j, k)\) is sensed by \( P \) groups simultaneously, then we have \( W(i, j, k) = \sqrt{P} \). Under this definition of \( W \), Problem (10) can be re-expressed in the form of (12), thereby using the same algorithm.

C. Performance Enhancement

Coupled tensor decomposition is nonconvex and hard to compute, even without noise. In practice, noise can make the estimation for the SLFs and PSDs even harder. In this subsection, we propose a series pragmatic engineering tricks, e.g., suitable initialization and post-processing, to help enhance performance of the propose approach.

1) Estimating \( \hat{S} \): Our identifiability theorems assume that \( S_r \)'s are exact low-rank matrices. However, in practice, the SLFs are approximately low-rank. Consequently, we notice that the estimation accuracy of the algorithm-output \( C \) is usually (much) better than that of \( \hat{S} = \hat{A}, \hat{B}^T \). It is of interest to exploit this empirical observation.

To this end, note that in (4c), \( X_1(\ell,:) = S(\ell,:)C^T \), where \( S = [\text{vec}(S_1), \cdots, \text{vec}(S_R)] \). If \( C \) is accurately estimated and \( C \) has full column rank, then, \( \hat{S}(\ell,:) = X_3(\ell,:)(C^T)^\dagger \) is a fairly accurate estimate for \( S(\ell,:) \). Note that \( S(\ell,r) = S_r(i,j) \) for \( \ell = (i-1)J + j \)—that is, there are a lot of elements of \( S_r \) can be accurate estimated. The remaining task is to utilize such information to interpolate the other elements. This, as a 2-D interpolation problem, is very well-studied in the literature [52], [53]. Since \( S_r \) is an SLF, it is smooth over space and easy to interpolate. In this work, we adopt the widely used spline interpolators, the thin-plate splines (TPS) [54]. After estimating \( \hat{S} \), the estimate of \( \tilde{X} \) can be obtained from its unfolding \( \tilde{X}_3 = \hat{S}C^T \).

2) Initialization: There are many ways for offering initializations to the BCD algorithms. For example, to initialize the algorithm for handling (8), one can decompose the ‘larger’ (and thus is more likely to admit a unique LL1 decomposition) subtensor using LL1 and extract partial information of \( A_r, B_r \) and \( C \). Then, the full latent factors can be estimated following the procedures of initialization techniques for coupled CPD as in [46], [47]. We omit the details due to space limitations.

VI. SIMULATIONS

In the simulations, the bands of interest are divided into \( K \) frequency bins, where \( K = 64 \) or 128. We assume that there are \( R \) emitters randomly deployed in a region with a size of \( 100 \times 100 \text{ m}^2 \). The region is discretized into \( 100 \times 100 \) grids. The PSDs of the emitters are generated as combinations of three randomly scaled squared sinc functions. This simulates a transmitter that occupies at most three carrier frequencies. For each of the emitters, the three carrier frequencies are randomly picked (cf. Fig. 7). To be more specific, the PSD of transmitter \( r \), i.e., \( C_r \), is generated by \( C_r(k) = \sum_{i=1}^{3} p_i^r \alpha_i^r \sin^2 \left( \frac{2\pi f_i^r}{w_{m_i}} \right) \).
where $p_i$ $(i = 1, 2, 3)$ follows a binomial distribution with equal probability; $a_i^r$ follows a uniform distribution from 0.5 to 2; $f_i^r \in \{1, \ldots, K\}$ is the center of the $i$-th sinc function $\text{sinc}^2(\cdot)$; and $w_i^r$ follows a uniform distribution between 2 and 4.

Each SLF is generated following a path-loss model and the spatial correlated log-normal shadowing model [45]. The SLF of emitter $r$ at location $y = [i, j] \in \mathbb{R}^2$ is generated by $S_r(i, j) = \|y - z_r\|_{\infty}^{-\eta_r} 10^{-2u_r(r)/10}$, where $z_r \in \mathbb{R}^2$ denotes the location of emitter $r$; $\eta_r$ is the path-loss coefficient of transmitter $r$. The correlated log-normal shadowing component $z_r(y)$ is generated as a zero-mean Gaussian random variable with autocorrelation

$$E\{z_r(y)z_r(y')\} = \sigma_r^2 \exp(-\|y - y'\|_{\infty}/X_c) \quad (15)$$

between locations $y$ and $y'$, where $X_c$ is the decorrelation distance. In this section, we take $\sigma_r = 4$, $X_c = 30$ for all $r$ unless specified. For a typical outdoor environment, $X_c$ varies from 50 to 100 [45]. We use $X_c = 30$ to test the algorithms under more critical situations. The parameter $\eta_r$ follows a uniform distribution between 2 and 3.

To evaluate the performance of estimated $\hat{C}$, $\hat{S}$ and $\widehat{X}$, we first fix the permutation ambiguity by solving the following matching criterion:

$$\hat{\pi} = \arg \min_{\pi \in \Phi} \sum_{r=1}^R \|c_r/\|c_r\|_1 - \|\hat{c}_{\pi_r}/\|\hat{c}_{\pi_r}\|_1\|_1$$

where $\Phi$ is the set of all permutations of $\{1, \ldots, R\}$ and $\pi = [\pi_1, \ldots, \pi_R]$, $\hat{\pi} = [\hat{\pi}_1, \ldots, \hat{\pi}_R]$; and $c_r$ and $\hat{c}_{\pi_r}$ are the true PSD of transmitter $r$ and its estimate, respectively. In the sequel, we assume that the permutation ambiguities of $\hat{C}$, $\hat{S}$ have been removed.

To evaluate the estimation performance of $\hat{C}$ and $\hat{S}$, we employ the normalized absolute error (NAE) for the estimated $C$:

$$\text{NAE}_C = \frac{1}{R} \sum_{r=1}^R \|c_r/\|c_r\|_1 - \|\hat{c}_{\pi_r}/\|\hat{c}_{\pi_r}\|_1\|_1,$$

the NAE of $S_r$ (denoted as NAE$_S$) is defined in an identical way. NAE of the estimated radio map is used for evaluating the radio map recovery performance, which is defined as follows:

$$\text{NAE}_X = \frac{\sum_{k=1}^K \|X(:, :, k) - \hat{X}(:, :, k)\|_1}{\sum_{k=1}^K \|X(:, :, k)\|_1}.$$

The proposed algorithms are terminated when the relative change of the loss function is less than $10^{-3}$ or when the number of iterations exceeds 100. We run 100 Monte Carlo trials and take the median of the NAEs.

To make these metrics clearly understandable, we offer a number of references in Figs. 7-9. For example, in Fig. 7, we visualize the ground-truth PSD and the estimated PSD under NAE$_C \approx 0.01, 0.1, 0.3$. When NAE$_C \approx 0.01$, the estimated PSD is almost identical the ground-truth PSD. When NAE$_C \approx 0.3$, the estimated PSD is much worse but still readable.
A. Scenario 1: Moving Sensors and Slab Sampling

In this subsection, we consider the case where we observe two subtensors \( \mathbf{X}^{(1)} = \mathbf{X}(\mathbf{S}_1, \cdots) \) and \( \mathbf{X}^{(2)} = \mathbf{X}(\mathbf{S}_2, \cdots) \), with \(|\mathbf{S}_1| = M, |\mathbf{S}_2| = N\). We use the formulation in Eq. (8) as the disaggregation criterion. The regularization parameters \( \lambda_n \) is set to \( \lambda_n = 10^{-2} \) for \( n = 1, 2, 3 \).

\[
\begin{array}{c}
\text{True } \mathbf{e}_1 \\
\text{Estimate } \mathbf{e}_1, \text{NAE} = 0.00517 \\
\text{True } \mathbf{e}_2 \\
\text{Estimate } \mathbf{e}_2, \text{NAE} = 0.00007
\end{array}
\]

Fig. 10. Ground-truth (top) and estimated (bottom) PSDs of two sources.

\[
\begin{array}{c}
\text{Aggregated SLF at the 5-th freq bin} \\
\text{True } \mathbf{S}_1 \\
\text{True } \mathbf{S}_2
\end{array}
\]

Fig. 11. Ground truth: (a1) The aggregated SLF at the 5-th frequency bin; (a2)-(a3) The ground-truth SLFs. (b1) \( M = 5 \) rows and \( N = 6 \) columns are sampled from \( \mathbf{X} \) (the unsampled entries marked in dark blue and the color of sampled entries remains unchanged); (b2) The estimated SLF of source 1; (b3) The estimated SLF of source 2.

In Figs. 10-11, we visualize the results of an \( R = 2 \) and \( K = 64 \) case. We set \( M = 5, N = 6, \) and assume that equi-spaced frontal and horizontal slabs are sampled—i.e., \( \mathbf{X}^{(1)} = \mathbf{X}(\mathbf{S}_1, \cdots) \) and \( \mathbf{X}^{(2)} = \mathbf{X}(\mathbf{S}_2, \cdots) \) are sampled (where \( \mathbf{S}_1 = \{17, 33, 49, 65, 81\}, \mathbf{S}_2 = \{15, 29, 43, 57, 71, 85\} \)). We set \( L = 4 \) and assume that \( R \) is known. The ground-truth PSDs and estimated PSDs (after removing permutations and rescaling the maximal amplitude of every PSD to 1) of the 2 sources are plotted in Fig. 10. The estimated PSDs are shown in Fig. 11 (a1-b3). One can see that both the PSDs and SLFs are reasonably recovered, under such a challenging scenario.

We also evaluate NAE\(_C\), NAE\(_S\) and NAE\(_X\) versus \( L \) under different \( M, R \). The results are presented in Table I under \( K = 128 \). One can see that the performance of the proposed approach deteriorates as \( R \) increases. This is reasonable, since many more parameters have to be estimated under a larger \( R \)—whereas the number of measurements remain the same.

Next, we compare our algorithm with some baseline methods. In [16], the authors proposed a semi-parametric regression (SR) to disaggregate the radio map and estimate the SLFs of each transmitters from multiple measurements per sensor. There, the PSD of each transmitter is assumed known. The SLFs of the emitters can be estimated under this assumption using a regression based approach. For SR, we only NAE\(_S\) and NAE\(_X\) as its performance metrics (since SR assumes that \( C \) is known). Note that SR also works with quantized measurements, which we do not consider in this work. Another baseline method is the TPS interpolation technique proposed in [10]. TPS is designed for single-band cartography. Hence, we perform TPS on each of the frontal slabs of the radio map tensor, i.e., \( \hat{\mathbf{X}}(\cdot, \cdot, k) \) for \( k = 1, \ldots, K \). We then apply the LL1 decomposition on the TPS-estimated \( \hat{\mathbf{X}} \) to disaggregate the PSDs and SLFs and observe the results. We also implement the group lasso splines (GLS) method in [15] as another baseline. GLS involves a basis selection step for representing the PSDs using an over-complete dictionary; we simplify their problem setting by giving GLS the real basis of the simulated PSDs, i.e., \( \text{sinc}^2\left(\frac{k-f_i}{B}\right) \) for \( f_i \in \mathcal{B} \), where \( \mathcal{B} \) is an index set that includes the true locations of the occupied frequency bins by emitter \( k \). This way, GLS deals with an easier task. GLS also does not estimate the emitter-level information and we apply LL1 decomposition on \( \hat{\mathbf{X}} \) as in the TPS case.

Table II shows the results under \( L = 4, R = 3, N = 6, K = 128 \) and various \( M \)'s. In general, the performance improves for all the algorithms when \( M \) increases, since more samples are available. SR exhibits the best performance in terms of running time—since it does not need to estimate \( C \). The NAE\(_S\) and NAE\(_X\) of the proposed algorithm approximate those of SR. The NAE\(_S\)'s of TPS and GLS are worse than that of the proposed coupled tensor approach, perhaps because they did not exploit the signal model information or the low-rank structure of the SLFs in (2).

Tables III-IV show the performance of the algorithms under different shadowing effect. This is of interest, since if the
shadowing effect is severe, the low-rank assumption of $S_r$ may be grossly violated. In particular, we test the algorithms under different $X_c$’s and $\sigma_r$’s in (15). Note that a larger $X_c$ means milder shadowing effect while a larger $\sigma_r$ corresponds to more severe shadowing; see Fig. 12. For outdoor environments, $X_c$ is typically between 50 and 100 [45]. Nonetheless, we test the algorithms under 10 to 100, to check if the algorithms can work under serious shadowing. We set $\sigma = \sigma_r$ for all $r$. One can see in both tables, the proposed approach outperform the baselines under most cases, even if the shadowing effect is heavy (e.g., $X_c = 10, \sigma = 8$)—which shows robustness against severe shadowing.

| Algorithm | Measure | $M$ |
|-----------|---------|-----|
| Proposed  | NAE$_C$ | 0.0381 |
|           | NAE$_S$ | 0.4717 |
|           | NAE$_X$ | 0.1589 |
|           | Running time(s) | 1.1084 |
| SR        | NAE$_C$ | 0.4580 |
|           | NAE$_S$ | 0.3795 |
|           | NAE$_X$ | 0.5995 |
|           | Running time(s) | 0.0283 |
| TPS       | NAE$_C$ | 0.5922 |
|           | NAE$_S$ | 0.9671 |
|           | NAE$_X$ | 0.5741 |
|           | Running time(s) | 0.0549 |
| GLS       | NAE$_C$ | 0.5915 |
|           | NAE$_S$ | 0.3518 |
|           | NAE$_X$ | 0.0549 |
|           | Running time(s) | 0.0549 |

Fig. 12. Shadowing effects of an SLF under various parameter settings.

| Algorithm | Measure | $X_c$ |
|-----------|---------|-----|
| Proposed  | NAE$_C$ | 0.0181 |
|           | NAE$_S$ | 0.5724 |
|           | NAE$_X$ | 0.4736 |
| SR        | NAE$_C$ | 0.5156 |
|           | NAE$_S$ | 0.4601 |
|           | NAE$_X$ | 0.3692 |
| TPS       | NAE$_C$ | 0.3692 |
|           | NAE$_S$ | 0.5724 |
|           | NAE$_X$ | 0.3692 |
| GLS       | NAE$_C$ | 0.0452 |
|           | NAE$_S$ | 0.0094 |
|           | NAE$_X$ | 0.0452 |

B. Static Sensors and Fiber Sampling

In this subsection, we consider the fiber sampling pattern of the scattered sensor deployment case. We define the sampling ratio $\rho = \frac{K}{IK}$. First, we give the results of NAE versus different sampling ratio, under $K = 128, R = 3, L = 4$. In random fiber sampling, $\rho$ varies from 0.10 to 0.25 with interval 0.05.

| Algorithm | Measure | $X_c$ |
|-----------|---------|-----|
| Proposed  | NAE$_C$ | 0.0074 |
|           | NAE$_S$ | 0.2979 |
|           | NAE$_X$ | 0.1232 |
| SR        | NAE$_C$ | 0.2742 |
|           | NAE$_S$ | 0.0305 |
|           | NAE$_X$ | 0.0129 |
| TPS       | NAE$_C$ | 0.4143 |
|           | NAE$_S$ | 0.4717 |
|           | NAE$_X$ | 0.3966 |

We visualize the estimated SLFs in Fig. 13. Here, $R = 2$ transmitters and $K = 64$ frequency bands are considered. The ground-truth SLFs are the same as those in Fig. 11. The fiber sampling pattern in (a1) is similar to the one that we saw in Fig. 6. One can see that the SLFs are well recovered. A random fiber sampling pattern is considered in Fig. 11 (b1), where approximately 10% of the fibers are sampled. One can see that our algorithm also works for random sampling, which supports our claim in Proposition 1.
C. Performance under Noise

We also consider a case where the measurements are noisy, i.e., \( \hat{X} = X + N \) where \( X \) follows the model in (2) and the entries of \( N \) follow the zero-mean i.i.d. Gaussian distribution. The signal-to-noise ratio (SNR) is defined as \( 10 \log_{10}(||X||^2_2/||N||^2_2) \).

Table VI and Table VII show the performance of the algorithms against various SNRs, under the slab sampling and fiber sampling schemes, respectively. The situations are similar as before: The proposed approach works better relative to the baselines, in terms of estimating \( S \), \( C \) and \( X \) under all SNRs. This may be because the proposed approach admits identifiability supports and explicitly utilizes the underlying signal model—which is often essential for combating noise.


equation

VII. Conclusion

In this work we proposed a novel coupled block-term tensor decomposition framework to tackle the radio map disaggregation problem. Unlike most of the existing cartography methods that are heuristic-driven, the proposed framework admits recoverability guarantees of each emitter’s radio map. In addition, the framework provably works under a number of systematic and random sampling schemes, and thus allows system designers to handle situations where sensor deployment is subject to various restrictions or regulations. Extensive simulations under heavy-shadowing scenarios show that the proposed method is promising for enhancing RF awareness of the sensing systems.

REFERENCES

[1] T. Yucek and H. Arslan, “A survey of spectrum sensing algorithms for cognitive radio applications,” IEEE Commun. Surveys Tuts., vol. 11, no. 1, pp. 116–130, 2009.
[2] B. Hamdaoui, B. Khalif, and M. Guizani, “Compressed wideband spectrum sensing: Concept, challenges, and enablers,” IEEE Commun. Mag., vol. 56, no. 4, pp. 136–141, 2018.
[3] Z. Tian and G. B. Giannakis, “Compressed sensing for wideband cognitive radios,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 2007, pp. 1357–1360.
[4] M. Hamid and B. Beferull-Lozano, “Non-parametric spectrum cartography using adaptive radial basis functions,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 2017, pp. 3599–3603.
[5] S. Bi, J. Lyu, Z. Ding, and R. Zhang, “Engineering radio maps for wireless resource management,” IEEE Trans. Wireless Commun., 2019.
[6] H. Zou, M. Jin, H. Jiang, L. Xie, and C. J. Spanos, “Winips: Wifi-based non-intrusive indoor positioning system with online radio map construction and adaptation,” IEEE Trans. Wireless Commun., vol. 16, no. 12, pp. 8118–8130, 2017.
[7] J. Chen, U. Yatamilli, and D. Gesbert, “Learning radio maps for unaware wireless networks: A segmented regression approach,” in Proc. IEEE Int. Conf. Commun. (ICC), 2017, pp. 1–6.
[8] J. Chen and D. Gesbert, “Optimal positioning of flying relays for wireless networks: A los map approach,” in Proc. IEEE Int. Conf. Commun. (ICC), 2017, pp. 1–6.
[9] G. Boccardi, G. Hernandez-Penalosa, and B. Beferull-Lozano, “Wireless sensor network for spectrum cartography based on kriging interpolation,” in Proc. IEEE PIMRC, 2012, pp. 1565–1570.
[10] S. Ureten, A. Yongaçoglu, and E. Petriu, “A comparison of interference cartography generation techniques in cognitive radio networks,” in Proc. IEEE Int. Conf. Commun. (ICC), 2012, pp. 1879–1883.
[11] B. A. Jayawickrama, E. Dutkiewicz, I. Oppermann, G. Fang, and J. Ding, “Improved performance of spectrum cartography based on compressive sensing in cognitive radio networks,” in Proc. IEEE Int. Conf. Commun. (ICC), 2013, pp. 5657–5661.
[12] J. A. Bazerque and G. B. Giannakis, “Distributed spectrum sensing for cognitive radio networks by exploiting sparsity,” IEEE Trans. Signal Process., vol. 58, no. 3, pp. 1847–1862, 2009.
[13] B. A. Jayawickrama, E. Dutkiewicz, I. Oppermann, and M. Meck, “Iteratively reweighted compressive sensing based algorithm for spectrum cartography in cognitive radio networks,” in Proc. IEEE Wireless Commun. Netw. Conf. (WCNC), 2014, pp. 719–724.
Supplementary Materials of “Spectrum Cartography via Coupled Block-term Tensor Decomposition”

G. Zhang, X. Fu, J. Wang, X.-L. Zhao, and M. Hong

APPENDIX A

PROOF OF THEOREM 3

In the noiseless case where (6) and (7) hold, we note that the optimal solutions to Problem (8) should make the two terms zero, when the noise is absent. Let \( (\hat{A}, \hat{B}, \hat{C}) \) denote any optimal solution of Problem (8) and \( (A, B, C) \) denote the ground-truth. We also define \( S_r = A_r B_r^\top \) and \( \hat{S}_r = \hat{A}_r \hat{B}_r^\top \). We aim to prove that \( \{S_r, \hat{S}_r\}_{r=1}^R \) is essentially the ground-truth \( \{S_r, c_r\}_{r=1}^R \) up to trivial ambiguities.

We show the case where condition (1) holds—the case under condition (2) shares exactly the same proof by the role of \( \{S_r, c_r\}_{r=1}^R \) and the row selection matrix \( \hat{D}_{\text{Diag}} \).

Therefore, \( C(S_r, :,:) = C(S_r, :,:) \Pi \Lambda \).

Accordingly, we have \( \hat{C}_r(S_r) = \lambda_r c_r(S_r) \), where \( \Lambda = \text{Diag}(\lambda_1, \cdots, \lambda_R) \) and \( \pi_r \in \{1, \cdots, R\} \) satisfies \( \Pi(\pi_r, r) = 1 \) for \( r = 1, \cdots, R \). Note that

\[
\begin{align*}
X_r^{(1)} &= \{\text{vec}(PA_rB_r^\top), \cdots, \text{vec}(PA_RB_R^\top)\}C(S_r, :)^T \\
&= \{\text{vec}(\hat{P}\hat{A}_r\hat{B}_r^\top), \cdots, \text{vec}(\hat{P}\hat{A}_R\hat{B}_R^\top)\}\hat{C}(S_r, :)^T.
\end{align*}
\]

Plugging \( \hat{C}(S_r, :) = C(S_r, :) \Pi \Lambda \) into (16), we have

\[
\hat{P}\hat{A}_r\hat{B}_r^\top = \lambda_r^{-1}PA_rB_r^\top.
\]

Here \( A_r, \hat{A}_r, B_r, \hat{B}_r \) are all full column-rank matrices almost surely (since they are drawn from certain joint continuous distributions), and the row selection matrix \( P \) is a full row-rank matrix. Hence, \( PA_r \in \mathbb{R}^{M \times L} \) is a submatrix of \( A_r \), and \( M \geq 2L \), which means that \( PA_r \) is a full column-rank matrix and so is \( \hat{P}\hat{A}_r \). Therefore, there exists a full rank matrix \( F_r \) satisfying \( \hat{B}_r = B_r F_r \).

Note that \( S = S_1 \cap S_4 \) and \( \hat{C}(S_r, :) = C(S_r, :) \Pi \Lambda \). Hence, we have

\[
\hat{C}(S_r, :) = C(S_r, :) \Pi \Lambda.
\]

Consider the subtensor \( X_r^{(3)} = X(:, S_2, S) \), where we have

\[
X_r^{(3)} = M_Q C(S_r, :)^T = \hat{M}_Q \hat{C}(S_r, :)^T,
\]

where

\[
M_Q = [\text{vec}(A_1 B_1^T Q^T), \cdots, \text{vec}(A_R B_R^T Q^T)]
\]

\[
\hat{M}_Q = [\text{vec}(\hat{A}_1 \hat{B}_1^T Q^T), \cdots, \text{vec}(\hat{A}_R \hat{B}_R^T Q^T)].
\]

Plugging \( \hat{C}(S_r, :) = C(S_r, :) \Pi \Lambda \) into (17), we have

\[
\hat{A}_r \hat{B}_r^T Q^T = \lambda_r^{-1} A_r B_r^T Q^T
\]

since \( C(S_r, :) \) is full column rank. Here \( Q \) is a row-selection matrix corresponding to \( S \). Plug \( \hat{B}_r = B_r F_r \) into this equation, we have

\[
\hat{A}_r F_r^T B_r^T Q^T = \lambda_r^{-1} A_r B_r^T Q^T.
\]

Since \( B_r^T Q \) is a full row rank matrix, we can infer \( \hat{A}_r = \lambda_r^{-1} A_r F_r^T \). Therefore,

\[
\hat{S}_r = \hat{A}_r \hat{B}_r^T = \lambda_r^{-1} A_r F_r^T (B_r F_r)^T
\]

\[
= \lambda_r^{-1} A_r B_r^T \hat{F}_r = \lambda_r^{-1} S_{\pi_r}.
\]

that is, we have proven that \( \{\hat{S}_r\}_{r=1}^R \) is a scaling and permutation version of ground-truth \( \{S_r\}_{r=1}^R \).

Next, we aim to prove that \( \hat{C} = \text{CPIA} \). In particular, if \( S_3 = S_4 = S = [K], \hat{C} = \text{CPIA} \) holds since the LL1 decomposition of \( X^{(1)} \) is essentially unique.

If \( S_3, S_4 \neq [K] \) while \( S_3 \cup S_4 = [K] \), note that

\[
\hat{A}_r \hat{B}_r^T Q^T = \hat{S}_r Q^T = \lambda_r^{-1} S_{\pi_r} Q^T,
\]

we have

\[
\hat{M}_Q = [\text{vec}(\hat{A}_1 \hat{B}_1^T Q^T), \cdots, \text{vec}(\hat{A}_R \hat{B}_R^T Q^T)]
\]

\[
= [\lambda_1^{-1} \text{vec}(A_{\pi_1} B_{\pi_1}^T Q^T), \cdots, \lambda_R^{-1} \text{vec}(A_{\pi_R} B_{\pi_R}^T Q^T)]
\]

\[
= M_Q \Pi \Lambda^{-1}.
\]

Consider the subtensor \( X_r^{(2)} \), we have

\[
\hat{X}_r^{(2)} = M_Q C(S_1, :)^T = \hat{M}_Q \hat{C}(S_4, :)^T.
\]

Plugging \( \hat{M}_Q = M_Q \Pi \Lambda^{-1} \) into (18), we obtain

\[
\hat{C}(S_4, :) = C(S_4, :) \Pi \Lambda.
\]

Combining with the fact that \( \hat{C}(S_3, :) = C(S_3, :) \Pi \Lambda \) and \( S_3 \cup S_4 = [K] \), we have \( \hat{C} = \text{CPIA} \).

APPENDIX B

PROOF OF THEOREM 4

In the noiseless case, we note that the optimal solutions to Problem (10) should make the two terms zero. Let \( (\hat{A}, \hat{B}, \hat{C}) \) denote any optimal solution of Problem (10) and \( (A, B, C) \) denote the ground-truth. Note that

\[
X_r^{(d)} = \sum_{r=1}^R (A_r (\tau^{(d)}), B_r (\tau^{(d)})) \circ C(\kappa^{(d)}, :)
\]

\[
= \sum_{r=1}^R (\hat{A}_r (\tau^{(d)}), \hat{B}_r (\tau^{(d)})) \circ \hat{C}(\kappa^{(d)}, :).
\]

By Theorem 1, one can see that every \( X_r^{(d)} \) admits a unique BTD, which means that \( \hat{C}(\kappa^{(d)}, :) \) is a column permuted and scaled version of \( C(\kappa^{(d)}, :) \). Therefore, there exists a permutation matrix \( \Pi^{(d)} \in \mathbb{R}^{R \times R} \) and a nonsingular diagonal matrix \( \Lambda^{(d)} \in \mathbb{R}^{R \times R} \) such that

\[
\hat{C}(\kappa^{(d)}, :) = C(\kappa^{(d)}, :) \Pi^{(d)} \Lambda^{(d)}.
\]

For any two coupled tensors \( X^{(i_2)} \) and \( X^{(i_2+1)} \), we have

\[
\hat{C}(\kappa^{(i_2)}, :) = C(\kappa^{(i_2)}, :) \Pi^{(i_2)} \Lambda^{(i_2)}
\]

\[
\hat{C}(\kappa^{(i_2+1)}, :) = C(\kappa^{(i_2+1)}, :) \Pi^{(i_2+1)} \Lambda^{(i_2+1)}.
\]
We aim to prove that $\Pi^{(i_d)} = \Pi^{(i_{d+1})}$ and $\Lambda^{(i_d)} = \Lambda^{(i_{d+1})}$ hold for $d = 1, \cdots, D - 1$. Define $R^{(i_d)} \subseteq K^{(i_d)} \cap K^{(i_{d+1})}$. Considering the common part $\hat{C}(R^{(i_d)}; \cdot)$, we have

$\hat{C}(R^{(i_d)}; :) = \hat{C}(K^{(i_d)} \cap R^{(i_d)}; :) = C(R^{(i_d)}; :)\Pi^{(i_d)}\Lambda^{(i_d)}$, $\hat{C}(R^{(i_d)}; :) = \hat{C}(K^{(i_{d+1})} \cap R^{(i_d)}; :) = C(R^{(i_d)}; :)\Pi^{(i_{d+1})}\Lambda^{(i_{d+1})}$.

Define $D^{(i_d)} = \Pi^{(i_d)}\Lambda^{(i_d)}$ and $d_r^{(i_d)} = D^{(i_d)}(:, r)$. We have

$C(R^{(i_d)}; :)D^{(i_d)} = C(R^{(i_d)}; :)D^{(i_{d+1})}$. (19)

Consider the $r$-th column of (19). We have

$C(R^{(i_d)}; :) (d_r^{(i_d)} - d_r^{(i_{d+1})}) = 0$. 

Note that $d_r^{(i_d)} \in R^R$ has only one non-zero element, then $d_r^{(i_d)} - d_r^{(i_{d+1})}$ has at most two non-zeros elements. Note that $|R^{(i_d)}| \geq 2$, which means that any two columns of $C(R^{(i_d)}; :)\Pi^{(i_d)}\Lambda^{(i_d)}$ are not proportional almost surely. Therefore, $d_r^{(i_d)} = d_r^{(i_{d+1})}$ holds for all $r = 1, \cdots, R$, and thus

$\Pi^{(i_d)}\Lambda^{(i_d)} = \Pi^{(i_{d+1})}\Lambda^{(i_{d+1})}$ holds. Therefore,

$\Pi^{(i_d)} = \Pi^{(i_{d+1})}$, $\Lambda^{(i_d)} = \Lambda^{(i_{d+1})}$ for all $d = 1, \cdots, D - 1$. Denote $\Pi = \Pi^{(i_1)}$ and $\Lambda = \Lambda^{(i_1)}$. Concatenating all $\hat{C}(K^{(i_1)}; :) = C(K^{(d_1)}; :)\Pi^{(i_1)}\Lambda^{(i_1)}$ for $d = 1, \cdots, D$, we have

$\hat{C} = \hat{C}(K^{(i_1)} \cup \cdots \cup K^{(i_D)}; :) = C\Pi\Lambda$. 

Next, we aim to prove that $\hat{S} = S\Pi\Lambda^{-1}$ where

$S = [\text{vec}(S_1), \cdots, \text{vec}(S_R)]$, $\hat{S} = [\text{vec}(\hat{S}_1), \cdots, \text{vec}(\hat{S}_R)]$, and $S_r = A_rB^\top_r$ and $\hat{S}_r = \hat{A}_r\hat{B}_r^\top$. Note that

$\widehat{X}_r^{(d)} = [\cdots, \text{vec}(S_r(I^{(d)} \cap J^{(d)})), \cdots]C(K^{(d)}; :)\Pi^{(i_1)}\Lambda^{(i_1)}$ 

and $\hat{C}(K^{(d)}; :) = C(K^{(d)}; :)\Pi^{(i_1)}\Lambda^{(i_1)}$. Therefore, we have

$\widehat{S}_r^{(d)} = S_r^{(d)}\Pi^{(i_1)}\Lambda^{(i_1)}$. (20)

where $\Lambda = \text{Diag}(\lambda_1, \cdots, \lambda_R)$ and there exists $\pi_r \in \{1, \cdots, R\}$ such that $\Pi(\pi_r, r) = 1$. Reshaping the $r$-th column of (20) to matrix, we have

$\hat{S}_r(I^{(d)} \cap J^{(d)}) = \lambda_r^{-1}S_{\pi_r}(I^{(d)} \cap J^{(d)})$. (21)

Plugging $S_r = A_rB^\top_r$ and $\hat{S}_r = \hat{A}_r\hat{B}_r^\top$ into (21), we have

$\hat{A}_r(I^{(d)}; :)\hat{B}_r(J^{(d)}; :) = \lambda_r^{-1}A_{\pi_r}(I^{(d)}; :)B_{\pi_r}(J^{(d)}; :)$.

Since both $\hat{A}_r(I^{(d)}; :)$ and $\hat{B}_r(J^{(d)}; :)$ are with full column rank almost surely for all $r$ due to $|I^{(d)}| \geq L$ and $|J^{(d)}| \geq L$, there exists a nonsingular matrix $F_r^{(d)}$ such that

$\hat{A}_r(I^{(d)}; :) = A_{\pi_r}(I^{(d)}; :)F_r^{(d)}$, 

$\hat{B}_r(J^{(d)}; :) = \lambda_r^{-1}B_{\pi_r}(J^{(d)}; :)(F_r^{(d)})^{-\top}$. 

Define $\mathcal{P}^{(i_d)} = \mathcal{I}^{(i_d)} \cap \mathcal{I}^{(i_{d+1})}$ and $\mathcal{Q}^{(i_d)} = \mathcal{J}^{(i_d)} \cap \mathcal{J}^{(i_{d+1})}$.

Note that $\max(|\mathcal{P}^{(i_d)}|, |\mathcal{Q}^{(i_d)}|) \geq L$. If we assume $|\mathcal{P}^{(i_d)}| \geq L$ holds for $d \in \{1, \cdots, D - 1\}$. Consider the common row indices of the following:

$\hat{A}_r(I^{(i_d)}; :) = A_{\pi_r}(I^{(i_d)}; :)F_r^{(i_d)}$, $\hat{A}_r(I^{(i_{d+1})}; :) = A_{\pi_r}(I^{(i_{d+1})}; :)F_r^{(i_{d+1})}$.

Then, it follows that

$A_{\pi_r}(\mathcal{P}^{(i_d)}; :)\PiF_r^{(i_{d+1})} - \PiF_r^{(i_{d+1})} = 0$.

Since $|\mathcal{P}^{(i_d)}| \geq L$, $A_{\pi_r}(\mathcal{P}^{(i_d)}; :)$ is a full rank matrix and we can conclude that $F_r^{(i_d)} = F_r^{(i_{d+1})}$ for $d \in \{1, \cdots, D - 1\}$. Similarly, if we assume that $|\mathcal{Q}^{(i_d)}| \geq L$ holds for $d \in \{1, \cdots, D - 1\}$, we can also conclude that $F_r^{(i_d)} = F_r^{(i_{d+1})}$ for $d \in \{1, \cdots, D - 1\}$. Therefore, $F_r^{(i_d)} = F_r^{(i_{d+1})}$ holds for $d = 1, \cdots, D - 1$.

Denote $F_r = F_r^{(i_1)}$. Concatenating all $\hat{A}_r(I^{(i_d)}; :) = A_{\pi_r}(I^{(i_d)}; :)F_r$,

we have

$\hat{A}_r(I^{(i_d)}; :) = A_{\pi_r}(I^{(i_d)}; :)F_r$.

Similarly, concatenating all $\hat{B}_r(J^{(d)}; :) = \lambda_r^{-1}B_{\pi_r}(J^{(d)}; :)F_r^{-\top}$, we also have

$\hat{B}_r = \lambda_r^{-1}B_{\pi_r}F_r^{-\top}$.

Therefore, we have

$\hat{S}_r = \hat{A}_r\hat{B}_r^\top = \lambda_r^{-1}A_{\pi_r}B_{\pi_r}^\top$, and

$\hat{S} = [\text{vec}(\hat{S}_1), \cdots, \text{vec}(\hat{S}_R)] = [\lambda_1^{-1}\text{vec}(S_1), \cdots, \lambda_R^{-1}\text{vec}(S_R)] = S\Pi\Lambda^{-1}$. 

**APPENDIX C**

**PROOF OF THEOREM 5**

This proof is similar to that of Theorem 4. First, we note that the optimal solutions to Problem (10) should make the two terms zero. Let $(\hat{A}, \hat{B}, \hat{C})$ denote any optimal solution of Problem (10) and $(A, B, C)$ denote the ground-truth. Note that

$X^{(d_o)} = \sum_{r=1}^{R}(A_r(I^{(d_o)}; :)B_r(J^{(d_o)}; :)\top)\circ C(K^{(d_o)}; :)$

$= \sum_{r=1}^{R}(\hat{A}_r(I^{(d_o)}; :)\hat{B}_r(J^{(d_o)}; :)\top)\circ \hat{C}(K^{(d_o)}; :)$

Note that $K^{(d_o)} = [K], K \geq R$ and

$\min\left(\frac{|I^{(d_o)}|}{L}, R\right) + \min\left(\frac{|J^{(d_o)}|}{L}, R\right) \geq R + 2$. 


By Theorem 1, there exists a permutation matrix $\Pi \in \mathbb{R}^{R \times R}$ and a nonsingular diagonal matrix $\Lambda \in \mathbb{R}^{R \times R}$ such that

$$\hat{C} = \Sigma \Pi \Lambda.$$

The proof of $\hat{S} = \Sigma \Pi \Lambda^{-1}$ is totally same as that of Theorem 4.

**APPENDIX D**

PROOF OF PROPOSITION 1

The objective function (12) can be rewritten as

$$
\sum_{k=1}^{K} \|W(:,:,k) * (X(:,:,k) - ADiag(C(:,:,k))B^\top)\|_F^2. \tag{22}
$$

Since $I \geq LR, J \geq LR$, the low-rank matrix factorization $ADiag(C(:,:,k))B^\top$ can be written as $ADiag(C(:,:,k))B^\top = U_k V_k^\top$, $U_k \in \mathbb{R}^{I \times LR}, V_k \in \mathbb{R}^{J \times LR}$. Then, the optimization problem (12) can be rewritten as

$$
\min_{A,B,C} \sum_{k=1}^{K} \|W(:,:,k) * (X(:,:,k) - U_k V_k^\top)\|_F^2
$$

subject to $U_k V_k^\top = ADiag(C(:,:,k))B^\top, k = 1, \cdots, K.$

In the noiseless case, the optimal solution should make $\|W(:,:,k) * (X(:,:,k) - U_k V_k^\top)\|_F^2$ to be zero for $k = 1, \cdots, K$. This is equivalent to solving $K$ matrix completion problems, each for one slab $X(:,:,k)$.

In [55], the authors derived sufficient conditions for rank-$r$ matrix completion, which is stated as follows:

**Lemma 1 [55]** Consider a low-rank matrix $X \in \mathbb{R}^{m \times n}$ with rank $r$, where $m \leq n$ and $r \leq m/6$. Let $0 < \epsilon \leq 1$ be given. Each column of $X$ is sampled uniformly at random, with at least $l$ entries observed. Also assume that $l$ satisfies

$$l \geq \max \left\{ 12 \log \left( \frac{d}{\epsilon} + 1 \right), 2r \right\}.$$

Then with probability at least $1 - \epsilon$, the incomplete matrix is uniquely completable if $n \geq (r + 1)(m - r)$.

Note that $X(:,:,k)$ is a rank-$LR$ matrix almost surely if $\{A_r, B_r\}$ and $C$ are drawn from any joint continuous distribution and noise is absent. Applying the above rank-$r$ matrix completion theorem to rank-$LR$ matrix $X(:,:,k)$ completion, one can conclude that $X(:,:,k)$ can be be recovered under the conditions specified in Fact 1. Once $X$ is recovered, the remaining problem is a rank-$(L_r, L_r, 1)$ BTD factorization problem. Using Theorem 1, we can reach the conclusion of Fact 1.

**APPENDIX E**

**UPDATE A, B, C FOR SOLVING (8)**

**A. Solution of (14a)**

To simplify the notation, let us rename different parts in (14a) as follows:

$$
\begin{align*}
\frac{P^\top P A}{H_1} (R_1 C \odot_p B)^\top (R_1 C \odot_p B) + \\
I_{J} A \left[ (R_2 C \odot_p QB)^\top (R_2 C \odot_p QB) + \lambda_1 I_{LR} \right] + \\
= P^\top X_1^{(1)} (R_1 C \odot_p B) + X_1^{(2)} (R_2 C \odot_p QB).
\end{align*}
$$

This simplifies the above equation as

$$H_1 A H_2 + H_3 A H_4 = H_5.$$

To find $A$ from the Sylvester equation $H_1 A H_2 + H_3 A H_4 = H_5$, we propose to employ the extended Bartels-Stewart method [56].

**B. Update B, C for Solving (8)**

The factor $B$ can be updated via solving the following equation:

$$Q^\top QB [(R_2 C \odot_p A)^\top (R_2 C \odot_p A)] + B \left[ (R_1 C \odot_p PA)^\top (R_1 C \odot_p PA) + \lambda_2 I_{LR} \right] = [X_2^1 (R_1 C \odot_p PA) + Q^\top X_2^{(2)} (R_2 C \odot_p A)].$$

Similarly, $C$ can be updated via solving the following equation:

$$R_1^T R_1 C M_1^T M_1 + R_2^T R_2 C M_2^T M_2 + \lambda_3 C = R_1^T X_3^{(1)} M_1 + R_2^T X_3^{(2)} M_2,$$

where

$$M_1 = [(B_1 \odot PA_1)_{1L}, \cdots, (B_R \odot PA_R)_{1L}] \in \mathbb{R}^{L_1 \times LR},$$

$$M_2 = [(QB_1 \odot A_1)_{1L}, \cdots, (QB_R \odot A_R)_{1L}] \in \mathbb{R}^{L_1 \times LR}.$$ 

**APPENDIX F**

**UPDATE A, B, C FOR SOLVING (12)**

The $i$th row of $A$ can be updated by

$$A(i,:) = Y_1(i,:) (C \odot_p B)$$

$$\left[ (C \odot_p B)^\top \text{Diag}(W_1(i,:)) (C \odot_p B) + \lambda_1 I_{LR} \right]^{-1}.$$ 

The $j$th row of $B$ can be updated by

$$B(j,:) = Y_2(j,:) (C \odot_p A)$$

$$\left[ (C \odot_p A)^\top \text{Diag}(W_2(j,:)) (C \odot_p A) + \lambda_2 I_{LR} \right]^{-1}.$$ 

The $k$th row of $C$ can be updated by

$$C(k,:) = Y_3(k,:) M [M^\top \text{Diag}(W_3(k,:)) M + \lambda_3 I_R]^{-1},$$

where

$$M = [(B_1 \odot A_1)_{1L}, \cdots, (B_R \odot A_R)_{1L}].$$