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 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 formulated 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 propagation 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 progress 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 (i.e., unmixed) to the emitter level has not been addressed. Second, most existing cartography techniques put emphasis on the algorithmic aspects, e.g, interpolation and kernel design. However, it has been unclear if in theory the spectral and spatial

information are 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-$(L_r, L_r, 1)$ block-term decomposition* (BTD) model (which will be referred as the LL1 decomposition model for simplicity) [19]. Our formulations can be understood as joint LL1 decomposition of tensors with shared latent factors, wherein each tensor 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 algorithmic side, we propose a *block coordinate descent* (BCD) 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 is 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] do not consider 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, $\boldsymbol{x}$, $\boldsymbol{X}$, and $\underline{\boldsymbol{X}}$ denote a scalar, a vector, a matrix, and a tensor, respectively. $\underline{\boldsymbol{X}}(i,j,k)$ (or $x_{i,j,k}$, $(\underline{\boldsymbol{X}})_{i,j,k}$) denotes the (i,j,k)th element of a third-order tensor. $c_{k,r}$ (or $\boldsymbol{C}(k,r)$) denotes the (k,r)th element of a matrix. c_k (or $\boldsymbol{c}(k)$) denotes kth element of a vector. The superscripts $^\top$ and † 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 subtensor of a given tensor; e.g., $\boldsymbol{C}(:,r)$ denotes r-th column of $\boldsymbol{C}$, and $\underline{\boldsymbol{X}}(:,:,k)$ denotes k-th frontal slab of $\underline{\boldsymbol{X}}$. The function $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. The cardinality of set $\mathcal{K}$ is denoted by $|\mathcal{K}|$. $[K]$ denotes the set $\{1, \cdots, K\}$. The outer product '$\circ$' between a matrix $\boldsymbol{S}$ and a vector $\boldsymbol{c}$ is defined as $\underline{\boldsymbol{X}} = \boldsymbol{S} \circ \boldsymbol{c}$ with

(i,j,k)th element $\underline{\boldsymbol{X}}(i,j,k) = \boldsymbol{S}(i,j)\boldsymbol{c}(k)$. The L_1-norm of a vector and matrix are defined as $\|\boldsymbol{x}\|_1 = \sum_i |x_i|$ and $\|\boldsymbol{X}\|_1 = \sum_{ij} |x_{ij}|$, respectively. We use $\|\boldsymbol{X}\|_F = \sqrt{\sum_{ij} x_{ij}^2}$ and $\|\underline{\boldsymbol{X}}\|_F = \sqrt{\sum_{ijk} x_{ijk}^2}$ to denote the Frobenius norms of $\boldsymbol{X}$ and $\underline{\boldsymbol{X}}$, respectively. The *Khatri-Rao product* (column-wise Kronecker) of two matrices $\boldsymbol{X} \in \mathbb{R}^{I \times K}$ and $\boldsymbol{Y} \in \mathbb{R}^{J \times K}$ is defined as $\boldsymbol{X} \odot \boldsymbol{Y} = [\boldsymbol{x}_1 \otimes \boldsymbol{y}_1, \cdots, \boldsymbol{x}_K \otimes \boldsymbol{y}_K]$, where $\otimes$ denotes the Kronecker product.

II. PROBLEM STATEMENT

We consider a scenario where R emitters exist within a 2-D geographical area of interest. Each emitter transmits across 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]:

$$x_{i,j,k} \approx \sum_{r=1}^{R} \boldsymbol{S}_r(i,j)c_{k,r}, \tag{1}$$

where $\boldsymbol{S}_r \in \mathbb{R}^{I \times J}$ 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 $\underline{\boldsymbol{X}} \in \mathbb{R}^{I \times J \times K}$ and a matrix $\boldsymbol{C} \in \mathbb{R}^{K \times R}$ such that $\underline{\boldsymbol{X}}(i,j,k) = x_{i,j,k}$ and $\boldsymbol{C}(k,r) = c_{k,r}$. Then, if noise is absent, the signal model can be expressed as:

$$\underline{\boldsymbol{X}} = \sum_{r=1}^{R} \boldsymbol{S}_r \circ \boldsymbol{c}_r, \tag{2}$$

where $\boldsymbol{c}_r$ denotes the rth column of $\boldsymbol{C}$ and $\circ$ denotes the outer product. We refer to the tensor $\underline{\boldsymbol{X}}$ as a *radio map*, since it reveals the RF environment across both the spatial and the spectral domains. One remark is that the 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 $\underline{\boldsymbol{X}}$ is measured as 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 $\boldsymbol{S}_r$ for $r = 1, \ldots, R$ and $\boldsymbol{C}$ given

$$\underline{\boldsymbol{X}}(i,j,k), \quad (i,j,k) \in \boldsymbol{\Omega},$$

and $\boldsymbol{\Omega}$ is the index set of observed entries of $\underline{\boldsymbol{X}}$. Note that normally $|\boldsymbol{\Omega}| \ll IJK$—i.e., the RF tensor is heavily undersampled. In this work, we refer to the problem of estimating the individual $\boldsymbol{c}_r$ and $\boldsymbol{S}_r$ as RF tensor *disaggregation* from limited observations—as opposed to approaches that only recover the manifest/aggregated information $\underline{\boldsymbol{X}}$. Obviously, if the former can be properly done, the latter is trivial to recover.

Note that if one only aims at recovering $\underline{\boldsymbol{X}}$, 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 $\underline{\boldsymbol{X}}$, but not the constituents $\boldsymbol{S}_r$ and $\boldsymbol{C}$—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 $\underline{\boldsymbol{X}}$ 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 decomposition model [40]. To be specific, under the LL1 model, a third-order tensor can be written as [19], [40]:

$$\underline{\boldsymbol{X}} = \sum_{r=1}^{R} (\boldsymbol{A}_r \boldsymbol{B}_r^\top) \circ \boldsymbol{c}_r, \tag{3}$$

Fig. 1. Illustration of tensor decomposition models. Top: CPD. Bottom: $(L, L, 1)$-BTD.

where $\boldsymbol{A}_r \in \mathbb{R}^{I \times L_r}$ and $\boldsymbol{B}_r \in \mathbb{R}^{J \times L_r}$ for $r = 1, \ldots, R$, $\boldsymbol{C} = [\boldsymbol{c}_1, \cdots, \boldsymbol{c}_R] \in \mathbb{R}^{K \times R}$, 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 $\mathrm{rank}(\boldsymbol{A}_r \boldsymbol{B}_r^\top) = L_r$ if $L_r \leq \min\{I, J\}$ and $\boldsymbol{A}_r$ and $\boldsymbol{B}_r$ have full column rank) and a vector. The block term $(\boldsymbol{A}_r \boldsymbol{B}_r^\top) \circ \boldsymbol{c}_r$ 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].

B. Mode-n Unfolding

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

A tensor can be represented in a matrix form by unfolding (or matricizing) the tensor. For example the mode-1 unfolding of $\underline{\boldsymbol{X}}$, denoted as $\boldsymbol{X}_1$, can be obtained by stacking all the vectorized horizontal slabs $\underline{\boldsymbol{X}}(i, :, :)$ as follows:

$$\boldsymbol{X}_1 = [\mathrm{vec}(\underline{\boldsymbol{X}}(1, :, :)), \cdots, \mathrm{vec}(\underline{\boldsymbol{X}}(I, :, :))]$$
$$= (\boldsymbol{C} \odot_p \boldsymbol{B}) \boldsymbol{A}^\top,$$

where $\mathrm{vec}(\cdot)$ is the vectorization operator, and the notation $\odot_p$ stands for the so-called partition-wise Khatri-Rao product, i.e., $\boldsymbol{C} \odot_p \boldsymbol{A} = [\boldsymbol{c}_1 \otimes \boldsymbol{A}_1, \cdots, \boldsymbol{c}_R \otimes \boldsymbol{A}_R] \in \mathbb{R}^{IK \times \sum_{r=1}^{R} L_r}$, in which $\boldsymbol{A} = [\boldsymbol{A}_1, \cdots, \boldsymbol{A}_R] \in \mathbb{R}^{I \times \sum_{r=1}^{R} L_r}$. The mode-2 and mode-3 unfoldings (denoted as $\boldsymbol{X}_2$ and $\boldsymbol{X}_3$) are defined in a similar way as follows:

$$\boldsymbol{X}_2 = [\mathrm{vec}(\underline{\boldsymbol{X}}(:, 1, :)), \cdots, \mathrm{vec}(\underline{\boldsymbol{X}}(:, J, :))]$$
$$= (\boldsymbol{C} \odot_p \boldsymbol{A}) \boldsymbol{B}^\top, \tag{4a}$$
$$\boldsymbol{X}_3 = [\mathrm{vec}(\underline{\boldsymbol{X}}(:, :, 1)), \cdots, \mathrm{vec}(\underline{\boldsymbol{X}}(:, :, K))]$$
$$= [(\boldsymbol{B}_1 \odot \boldsymbol{A}_1) \mathbf{1}_{L_1}, \cdots, (\boldsymbol{B}_R \odot \boldsymbol{A}_R) \mathbf{1}_{L_R}] \boldsymbol{C}^\top,$$
$$= [\mathrm{vec}(\boldsymbol{A}_1 \boldsymbol{B}_1^\top), \cdots, \mathrm{vec}(\boldsymbol{A}_R \boldsymbol{B}_R^\top)] \boldsymbol{C}^\top. \tag{4b}$$

In particular, note that one can re-write

$$\boldsymbol{X}_3 = \boldsymbol{S} \boldsymbol{C}^\top, \tag{5}$$
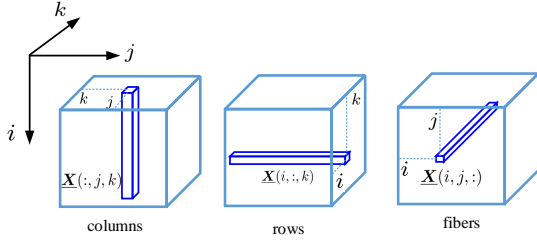
Fig. 2. Illustration of three modes, i.e., column $\underline{\boldsymbol{X}}(:,j,k)$, row $\underline{\boldsymbol{X}}(i,:,k)$ and fiber $\underline{\boldsymbol{X}}(i,j,:)$, of a third-order tensor.
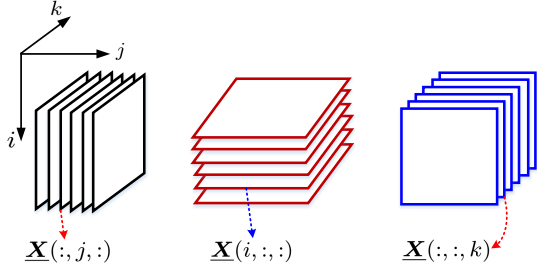
Fig. 3. Illustration of the vertical slabs $\underline{\boldsymbol{X}}(:,j,:)$ (left), the horizontal slabs $\underline{\boldsymbol{X}}(i,:,:)$ (middle), and the frontal slabs $\underline{\boldsymbol{X}}(:,:,k)$ (right) of a third-order tensor.

where $\boldsymbol{S}(:,r) = \mathrm{vec}(\boldsymbol{A}_r\boldsymbol{B}_r^\top)$. For more details, see [19], [30], [38], [39] (in particular, our unfolding strategy follows the paradigm in [30]). These representations will be handy in our theoretical analysis and algorithm design.

C. Mode Product and BTD Representation

Consider $\underline{\boldsymbol{X}} \in \mathbb{R}^{I \times J \times K}$ and $\boldsymbol{P}_1 \in \mathbb{R}^{M \times I}, \boldsymbol{P}_2 \in \mathbb{R}^{N \times J}, \boldsymbol{P}_3 \in \mathbb{R}^{T \times K}$. The mode-$n$ product $\underline{\boldsymbol{G}}_n = \underline{\boldsymbol{X}} \times_n \boldsymbol{P}_n$ for $n = 1, 2, 3$ such that $\underline{\boldsymbol{G}}_1(m,j,k) = \sum_{i=1}^{I} \underline{\boldsymbol{X}}(i,j,k)\boldsymbol{P}_1(m,i)$, $\underline{\boldsymbol{G}}_2(i,n,k) = \sum_{j=1}^{J} \underline{\boldsymbol{X}}(i,j,k)\boldsymbol{P}_2(n,j)$, and $\underline{\boldsymbol{G}}_3(i,j,t) = \sum_{k=1}^{K} \underline{\boldsymbol{X}}(i,j,k)\boldsymbol{P}_3(t,k)$. A key observation is that, if $\underline{\boldsymbol{X}}$ follows the LL1 decomposition model as in (3), we have

$$\underline{\boldsymbol{X}} \times_1 \boldsymbol{P}_1 \times_2 \boldsymbol{P}_2 \times_3 \boldsymbol{P}_3 = \sum_{r=1}^{R} (\boldsymbol{P}_1\boldsymbol{A}_r(\boldsymbol{P}_2\boldsymbol{B}_r)^\top) \circ (\boldsymbol{P}_3\boldsymbol{c}_r).$$

D. Essential Uniqueness

The LL1 model has a very nice property—$\{\boldsymbol{A}_r\boldsymbol{B}_r^\top\}_{r=1}^R$ and $\{\boldsymbol{c}_r\}_{r=1}^R$ 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 $\{(\boldsymbol{A}_r\boldsymbol{B}_r^\top) \circ \boldsymbol{c}_r\}_{r=1}^R$, and scale and counter-scale $\boldsymbol{A}_r\boldsymbol{B}_r^\top$ and $\boldsymbol{c}_r$ as long as their outer product remains uncharged. Note that $\boldsymbol{A}_r\boldsymbol{B}_r^\top = \boldsymbol{A}_r\boldsymbol{F}_r(\boldsymbol{B}_r\boldsymbol{F}_r^{-\top})^\top$ holds for any non-singular matrix $\boldsymbol{F}_r \in \mathbb{R}^{L_r \times L_r}$. Hence, $\boldsymbol{A}_r$ and $\boldsymbol{B}_r$ cannot be identified from the LL1 decomposition of $\underline{\boldsymbol{X}}$ in general. In this paper, we will use the following theorem:

Theorem 1 *[19] Let $\underline{\boldsymbol{X}} = \sum_{r=1}^{R}(\boldsymbol{A}_r\boldsymbol{B}_r^\top) \circ \boldsymbol{c}_r$ and assume $L_r = L$ for $r = 1, \cdots, R$. Assume that $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ 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 $\boldsymbol{S} = [\mathrm{vec}(\boldsymbol{S}_1), \ldots, \mathrm{vec}(\boldsymbol{S}_R)] \in \mathbb{R}^{IJ \times R}$ where $\boldsymbol{S}_r = \boldsymbol{A}_r\boldsymbol{B}_r^\top$. Then, with probability one, for any $\widetilde{\boldsymbol{S}} \in \mathbb{R}^{IJ \times R}$ and $\widetilde{\boldsymbol{C}} \in \mathbb{R}^{K \times R}$ satisfying $\boldsymbol{X}_3 = \widetilde{\boldsymbol{S}}\widetilde{\boldsymbol{C}}^\top$, we have

$$\widetilde{\boldsymbol{C}} = \boldsymbol{C}\boldsymbol{\Pi}\boldsymbol{\Lambda}, \; \widetilde{\boldsymbol{S}} = \boldsymbol{S}\boldsymbol{\Pi}\boldsymbol{\Lambda}^{-1},$$

where $\boldsymbol{\Pi}$ is a permutation matrix and $\boldsymbol{\Lambda}$ is a nonsingular diagonal matrix; i.e., the multilinear rank-$(L_r, L_r, 1)$ decomposition of $\underline{\boldsymbol{X}}$ 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:

Theorem 2 *[19] Let $\underline{\boldsymbol{X}} = \sum_{r=1}^{R}(\boldsymbol{A}_r\boldsymbol{B}_r^\top) \circ \boldsymbol{c}_r$ and assume $L_r = L$ for $r = 1, \cdots, R$. Assume $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ are drawn from certain joint absolutely continuous distributions. If $IJ \geq L^2R$ and*

$$\min\left(\left\lfloor \frac{I}{L} \right\rfloor, R\right) + \min\left(\left\lfloor \frac{J}{L} \right\rfloor, R\right) + \min(K, R) \geq 2R + 2,$$

then, the multilinear rank-$(L_r, L_r, 1)$ decomposition of $\underline{\boldsymbol{X}}$ is essentially unique almost surely.

For notational simplicity, we will by default use $L_1 = \ldots = L_R = L$ in the sequel.

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.

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 $\boldsymbol{S}_r \in \mathbb{R}^{I \times J}$, 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 100 m × 100 m 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)
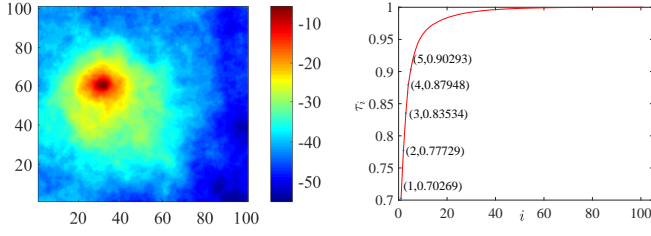
Fig. 4. Left: SLF of a source over a 100 m² region (discretized as a 101×101 matrix), with shadowing effect. Right: The plot of τ_i versus i—showing that the SLF is approximately low-rank. The first 5 i-th index and its proportion τ_i is marked in the figure, namely (i, τ_i) for $i = 1, 2, 3, 4, 5$.
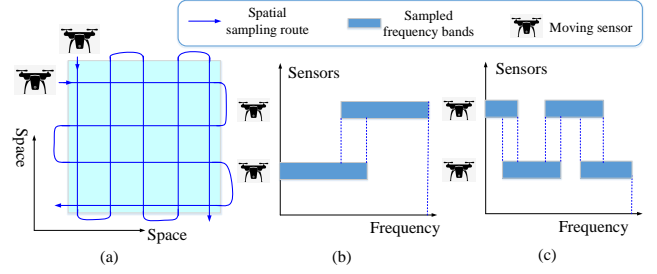

Fig. 5. Using moving sensors to sample the RF tensor. (a) An example of the sampling routes using two moving sensors; (b) an example of sampled frequency bands by two sensors; (c) another example of frequency band allocation to two sensors.

to the SLF and plot the ratio $\tau_i = \sum_{k=1}^{i} \mu_k / \sum_{k=1}^{100} \mu_k$, where μ_k denotes the kth singular value of $\boldsymbol{S}_r$ 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 $\boldsymbol{S}_r$. This observation suggests that the SLFs, even under shadowing effects, may be well-approximated by low-rank matrices. Using low-rank matrices to approximate spatial smoothness appeared in the literature, e.g., for channel gain field estimation [41] and fMRI processing [46], but has not bee considered for SLFs, to our best knowledge.

Under the low-rank postulate, one can express $\boldsymbol{S}_r$ by:

$$\boldsymbol{S}_r = \boldsymbol{A}_r \boldsymbol{B}_r^\top,$$

where $\boldsymbol{A}_r \in \mathbb{R}^{I \times L}$, $\boldsymbol{B}_r \in \mathbb{R}^{J \times L}$ and $L \ll \min\{I, J\}$. Consequently, we have

$$\underline{\boldsymbol{X}} = \sum_{r=1}^{R} \boldsymbol{S}_r \circ \boldsymbol{c}_r = \sum_{r=1}^{R} \left(\boldsymbol{A}_r \boldsymbol{B}_r^\top \right) \circ \boldsymbol{c}_r;$$

i.e., the complete RF tensor follows a multilinear rank-$(L, L, 1)$ BTD model. Based on this model, we will propose a number of sensing and $(\boldsymbol{S}_r, \boldsymbol{c}_r)$-estimation schemes in the following subsections.

B. The Moving Sensor Case: Sensing Consecutive Grids

If $\underline{\boldsymbol{X}}$ is available, then estimating $\boldsymbol{S}_r$ and $\boldsymbol{c}_r$ can be done by directly applying LL1 algorithms (e.g., that in [39]) to the tensor. In practice, however, the complete $\underline{\boldsymbol{X}}$ is not available. What we observe is a substantially undersampled version of $\underline{\boldsymbol{X}}$. One possible route is to first apply tensor completion techniques to estimate the complete $\underline{\boldsymbol{X}}$, and then seek the LL1 decomposition of the estimated $\underline{\boldsymbol{X}}$. However, random sampling schemes that may ensure provable completion of $\underline{\boldsymbol{X}}$ 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 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 $\mathcal{S}_1 \subset \{1, \cdots, I\}$ and $\mathcal{S}_2 \subset \{1, \cdots, J\}$ contain the observed indices along the two spatial dimensions by the two sensors, respectively. Similarly, let $\mathcal{S}_3, \mathcal{S}_4 \subset \{1, \cdots, K\}$ be the observed indices along the frequency dimensions by the two sensors, respectively. Then, two subtensors $\underline{\boldsymbol{X}}^{(1)} = \underline{\boldsymbol{X}}(\mathcal{S}_1, :, \mathcal{S}_3)$ and $\underline{\boldsymbol{X}}^{(2)} = \underline{\boldsymbol{X}}(:, \mathcal{S}_2, \mathcal{S}_4)$, with $|\mathcal{S}_1| = M$, $|\mathcal{S}_2| = N$, $|\mathcal{S}_3| = K_1$ and $|\mathcal{S}_4| = K_2$ which satisfy $\mathcal{S}_3 \cup \mathcal{S}_4 = [K] = \{1, \ldots, K\}$. Define row-selection matrix $\boldsymbol{P} = \boldsymbol{I}_I(\mathcal{S}_1, :)$, which consists of the rows indexed by $\mathcal{S}_1$ of an $I \times I$ identity matrix. Similarly, let $\boldsymbol{Q} = \boldsymbol{I}_J(\mathcal{S}_2, :)$, $\boldsymbol{R}_1 = \boldsymbol{I}_K(\mathcal{S}_3, :)$ and $\boldsymbol{R}_2 = \boldsymbol{I}_K(\mathcal{S}_4, :)$. The sampled measurements by sensor 1 and sensor 2 can be represented as follows:

$$\underline{\boldsymbol{X}}^{(1)} = \underline{\boldsymbol{X}} \times_1 \boldsymbol{P} \times_3 \boldsymbol{R}_1 = \underline{\boldsymbol{X}}(\mathcal{S}_1, :, \mathcal{S}_3), \quad (6a)$$

$$\underline{\boldsymbol{X}}^{(2)} = \underline{\boldsymbol{X}} \times_2 \boldsymbol{Q} \times_3 \boldsymbol{R}_2 = \underline{\boldsymbol{X}}(:, \mathcal{S}_2, \mathcal{S}_4). \quad (6b)$$

C. Coupled Tensor Decomposition-Based Formulation

To proceed, recall our postulate that $\boldsymbol{S}_r = \boldsymbol{A}_r \boldsymbol{B}_r^\top$ where $\boldsymbol{A}_r \in \mathbb{R}^{I \times L}$ and $\boldsymbol{B}_r \in \mathbb{R}^{J \times L}$. Hence, our sampled data can be written as:

$$\underline{\boldsymbol{X}}^{(1)} = \sum_{r=1}^{R} ((\boldsymbol{P} \boldsymbol{A}_r) \boldsymbol{B}_r^\top) \circ (\boldsymbol{R}_1 \boldsymbol{c}_r), \quad (7a)$$

$$\underline{\boldsymbol{X}}^{(2)} = \sum_{r=1}^{R} (\boldsymbol{A}_r (\boldsymbol{B}_r^\top \boldsymbol{Q}^\top)) \circ (\boldsymbol{R}_2 \boldsymbol{c}_r). \quad (7b)$$

It is not hard to notice that $\underline{\boldsymbol{X}}^{(1)}$ and $\underline{\boldsymbol{X}}^{(2)}$ are two tensors that still follow the LL1 model. More importantly, the latent factors of the original tensor, i.e., the matrices $\boldsymbol{A}_r$, $\boldsymbol{B}_r$ for $r = 1, \ldots, R$ and $\boldsymbol{C}$, are captured in the two observations. Intuitively, if both $\underline{\boldsymbol{X}}^{(1)}$ and $\underline{\boldsymbol{X}}^{(2)}$ admit unique LL1 decompositions, then, one may estimate $\boldsymbol{A}_r$ from $\underline{\boldsymbol{X}}^{(2)}$ and estimate $\boldsymbol{B}_r$ from $\underline{\boldsymbol{X}}^{(1)}$ respectively, and then 'assemble' the information to get $\boldsymbol{S}_r$.

The idea is seemingly appealing, but has several significant hurdles. First, recall that the LL1 decomposition does not give exact $\boldsymbol{A}_r$, $\boldsymbol{B}_r$, but only the range spaces of them [39]. To be specific, by decomposing $\underline{\boldsymbol{X}}^{(1)}$, one will have $\widehat{\boldsymbol{B}}_r = \boldsymbol{B}_r \boldsymbol{F}_r$ if the BTD problem is optimally solved, where $\boldsymbol{F}_r \in \mathbb{R}^{L \times L}$ is nonsingular. Similar, from $\underline{\boldsymbol{X}}^{(2)}$, one will have $\widehat{\boldsymbol{A}}_r = \boldsymbol{A}_r \boldsymbol{G}_r$. In general, $\boldsymbol{G}_r \neq \boldsymbol{F}_r^{-\top}$, and thus $\boldsymbol{S}_r \neq \widehat{\boldsymbol{A}}_r \widehat{\boldsymbol{B}}_r^\top$, which means that the reconstruction fails. Second, if one of $\underline{\boldsymbol{X}}^{(1)}$ or $\underline{\boldsymbol{X}}^{(2)}$ 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:

$$\underset{\{\boldsymbol{A}_r, \boldsymbol{B}_r\}, \boldsymbol{C}}{\text{minimize}} \left\| \underline{\boldsymbol{X}}^{(1)} - \sum_{r=1}^{R} ((\boldsymbol{P}\boldsymbol{A}_r)\boldsymbol{B}_r^\top) \circ (\boldsymbol{R}_1 \boldsymbol{c}_r) \right\|_F^2 \quad (8)$$
$$+ \left\| \underline{\boldsymbol{X}}^{(2)} - \sum_{r=1}^{R} (\boldsymbol{A}_r (\boldsymbol{B}_r^\top \boldsymbol{Q}^\top)) \circ (\boldsymbol{R}_2 \boldsymbol{c}_r) \right\|_F^2.$$

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

Theorem 3 *Assume that (6) and (7) hold, and that $\boldsymbol{A}_r$, $\boldsymbol{B}_r$ for all r and $\boldsymbol{C}$ are drawn from any joint absolutely continuous distributions. Also assume that $|\mathcal{S}| \geq R$, where $\mathcal{S} = \mathcal{S}_3 \cap \mathcal{S}_4$. Suppose that one of the following conditions holds*

(1) $M \geq 2L$, $J \geq LR$, and $\min(\lfloor M/L \rfloor, R) + \min(\lfloor J/L \rfloor, R) \geq R + 2$;
(2) $N \geq 2L$, $I \geq LR$, and $\min(\lfloor N/L \rfloor, R) + \min(\lfloor I/L \rfloor, R) \geq R + 2$.

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

The proof is relegated to Appendix A. The proof is reminiscent of coupled tensor decomposition under the CPD model [47], [48]—whereas the proof of Theorem 3 is more challenging since $\boldsymbol{A}_r$ and $\boldsymbol{B}_r$ 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 is not too large, then, the proposed sampling strategy and the formulation guarantee recovering $\boldsymbol{S}_r$ and $\boldsymbol{c}_r$ 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], [49].

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* [48], [50]. A fiber of the tensor $\underline{\boldsymbol{X}}$ refers to $\underline{\boldsymbol{X}}(i, j, :)$; 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 $\underline{\boldsymbol{X}}(i, j, \mathcal{K})$—which can be the whole fiber or a part of the fiber $\underline{\boldsymbol{X}}(i, j, :)$.

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 $\underline{\boldsymbol{X}}^{(d)} = \underline{\boldsymbol{X}}(\mathcal{I}^{(d)}, \mathcal{J}^{(d)}, \mathcal{K}^{(d)})$. Since $\underline{\boldsymbol{X}}$ admits a rank-$(L, L, 1)$ BTD, the sampled subtensor $\underline{\boldsymbol{X}}^{(d)}$ can be written as:

$$\underline{\boldsymbol{X}}^{(d)} = \sum_{r=1}^{R} (\boldsymbol{A}_r(\mathcal{I}^{(d)}, :) \boldsymbol{B}_r(\mathcal{J}^{(d)}, :)^\top) \circ \boldsymbol{c}_r(\mathcal{K}^{(d)}). \quad (9)$$

To estimate $\boldsymbol{C}$ and $\boldsymbol{S}_r$, we propose the following criterion:

$$\underset{\{\boldsymbol{A}_r, \boldsymbol{B}_r\}, \boldsymbol{C}}{\text{minimize}} \sum_{d=1}^{D} \left\| \underline{\boldsymbol{X}}^{(d)} - \sum_{r=1}^{R} \boldsymbol{S}_r(\mathcal{I}^{(d)}, \mathcal{J}^{(d)}) \circ \boldsymbol{c}_r(\mathcal{K}^{(d)}) \right\|_F^2, \quad (10)$$

where we have $\boldsymbol{S}_r(\mathcal{I}^{(d)}, \mathcal{J}^{(d)}) = \boldsymbol{A}_r(\mathcal{I}^{(d)}, :) \boldsymbol{B}_r(\mathcal{J}^{(d)}, :)^\top$. In terms of identifiability of the SLFs and PSDs, we show the following theorem:

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

$$\cup_{d=1}^{D} \mathcal{I}^{(d)} = [I], \quad \cup_{d=1}^{D} \mathcal{J}^{(d)} = [J], \quad \cup_{d=1}^{D} \mathcal{K}^{(d)} = [K], \quad (11)$$

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

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

Assume that there exists a permutation of set $[D]$, denoted as $\{i_1, \cdots, i_D\}$, satisfying $\max(|\mathcal{I}^{(i_d)} \cap \mathcal{I}^{(i_{d+1})}|, |\mathcal{J}^{(i_d)} \cap \mathcal{J}^{(i_{d+1})}|) \geq L$ and $|\mathcal{K}^{(i_d)} \cap \mathcal{K}^{(i_{d+1})}| \geq 2$ for $d = 1, \cdots, D-1$. 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 B. 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 $\underline{X}^{(d)}$ for $d = 1, \ldots, D$ 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 $\mathcal{I}^{(d)}, \mathcal{J}^{(d)}, \mathcal{K}^{(d)}$ satisfy $\cup_{d=1}^{D}\mathcal{I}^{(d)} = [I], \cup_{d=1}^{D}\mathcal{J}^{(d)} = [J]$. and $|\mathcal{I}^{(d)}| \geq L, |\mathcal{J}^{(d)}| \geq L$ for $d = 1, \cdots, D$. Assume that there exists $d_0 \in [D]$ that satisfies $\mathcal{K}^{(d_0)} = [K]$. In addition, assume that $K \geq R$ and*

$$\min\left(\left\lfloor \frac{|\mathcal{I}^{(d_0)}|}{L} \right\rfloor, R\right) + \min\left(\left\lfloor \frac{|\mathcal{J}^{(d_0)}|}{L} \right\rfloor, R\right) \geq R + 2,$$

and that there exists a permutation of set $[D]$, denoted as $\{i_1, \cdots, i_D\}$, satisfying $\max(|\mathcal{I}^{(i_d)} \cap \mathcal{I}^{(i_{d+1})}|, |\mathcal{J}^{(i_d)} \cap \mathcal{J}^{(i_{d+1})}|) \geq L$ for $d = 1, \cdots, D - 1$. 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. One can see that compared to Theorem 4, the above theorem only needs one of $\underline{X}^{(d)}$, namely $\underline{X}^{(d_0)}$, to be identifiable, and the remaining subtensors do not need to be identifiable and are only required to overlap at least L indices in the first or second dimension. 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 $\underline{X}$. Apparently, these two theorems present an interesting spectral-spatial sampling trade-off.

Remark 3 The work in [50] considered a similar setting as that in Theorems 4-5 and showed identifiability of a coupled LL1 decomposition criterion. Notably, the results in [50] can also be applied to our case. The key difference is that the work in [50] 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 [48] also considered fiber sampling
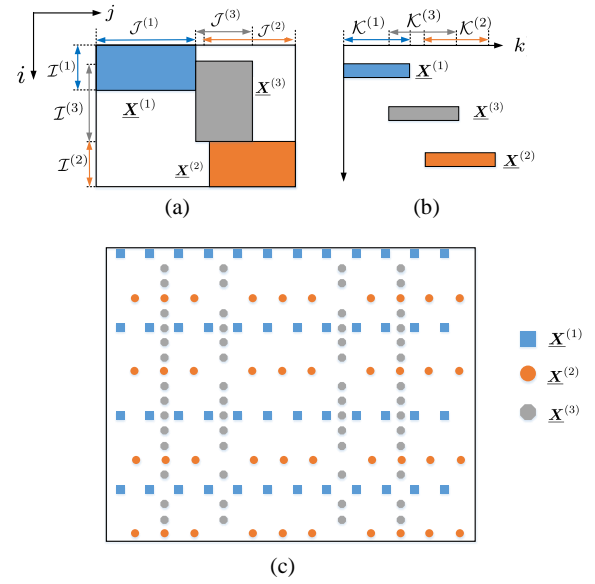

Fig. 6. An example of $D = 3$ subtensors that satisfy the conditions of Theorem 4—$i_1 = 1, i_2 = 3, i_3 = 2$ and $|\mathcal{I}^{(i_1)} \cap \mathcal{I}^{(i_2)}| \geq L, |\mathcal{J}^{(i_2)} \cap \mathcal{J}^{(i_3)}| \geq L, |\mathcal{K}^{(i_1)} \cap \mathcal{K}^{(i_2)}| \geq 2, |\mathcal{K}^{(i_2)} \cap \mathcal{K}^{(i_3)}| \geq 2$. (a) The overlapped index sets of the first and second dimensions of $\underline{X}^{(1)}, \underline{X}^{(2)}, \underline{X}^{(3)}$. (b) The overlapped index sets of the third dimension of $\underline{X}^{(1)}, \underline{X}^{(2)}, \underline{X}^{(3)}$. (c) A 'scattered' version by permuting the spatial sampling pattern in (a).

and tensor recovery, but under the CPD model instead of the LL1 model.

On a higher level, Theorems 1-5 in this work can be understood as coupled LL1 decomposition for downsampled tensor recovery with regular/systematical sampling strategies—which share insights with tensor recovery from coupled CPD models [47], [51]. The new challenge for establishing recoverability under the the coupled LL1 framework lies in the non-identifiability of A_r and B_r, which needs extra care to handle in the recoverability proofs.

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:

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

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

Proposition 1 *Under the exact LL1 model of $\underline{X}$, assume that, w.l.o.g., $I \leq J, J > (LR + 1)(I - LR)$ and $LR \leq \frac{I}{6}$. Also assume that $\{A_r, B_r\}_r$ 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 $\underline{X}(:, :, k), k = 1, \cdots, K$ are observed uniformly at random,*

where $q \geq \max\left\{12\log\left(\frac{I}{\epsilon}+1\right), 2LR\right\}$, *then solving Problem* (12) *recovers* $\{\boldsymbol{S}_r, \boldsymbol{c}_r\}$ *for* $r = 1\ldots,R$ *with a probability at least* $1-\epsilon$.

The proof is presented in Appendix D in the supplementary material and is straightforward. In a nutshell, if the matrix rank of each slab $\underline{\boldsymbol{X}}(:,:,k)$ 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 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 $f(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$. In practice, instead of trying to minimize $f(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$, we use the following regularized version

$$\underset{\{\boldsymbol{A}_r, \boldsymbol{B}_r\}, \boldsymbol{C}}{\text{minimize}} \ \underbrace{f(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}) + \lambda_1 \|\boldsymbol{A}\|_F^2 + \lambda_2 \|\boldsymbol{B}\|_F^2 + \lambda_3 \|\boldsymbol{C}\|_F^2}_{\text{Loss}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})}.$$
(13)

The regularization terms are added to alleviate the so-called scaling/counter-scaling effect. Specifically, $\alpha(\boldsymbol{A}_r\boldsymbol{B}_r^\top) \circ (1/\alpha)\boldsymbol{C}(:,r)$ may happen without changing the model fitting error, where α could be arbitrarily large, thereby affecting numerical stability. Adding regularization can prevent such pathological cases from happening; see discussions in [52].

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

$$\boldsymbol{A}^{t+1} \leftarrow \arg\min \text{Loss}(\boldsymbol{A}, \boldsymbol{B}^t, \boldsymbol{C}^t) \quad (14a)$$
$$\boldsymbol{B}^{t+1} \leftarrow \arg\min \text{Loss}(\boldsymbol{A}^{t+1}, \boldsymbol{B}, \boldsymbol{C}^t) \quad (14b)$$
$$\boldsymbol{C}^{t+1} \leftarrow \arg\min \text{Loss}(\boldsymbol{A}^{t+1}, \boldsymbol{B}^{t+1}, \boldsymbol{C}), \quad (14c)$$

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 material.

As a standard exact BCD algorithm, the algorithm in (14) enjoys convergence properties as discussed in [53], [54]. 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 $\underline{\boldsymbol{Y}} = \underline{\boldsymbol{W}} * \underline{\boldsymbol{X}}$, let $\boldsymbol{Y}_n$ and $\boldsymbol{W}_n$ be mode-n unfoldings of $\underline{\boldsymbol{Y}}$ and $\underline{\boldsymbol{W}}$, respectively. The update rules for the regularized version of Problem (12) are as follows:

$$\boldsymbol{A} \leftarrow \arg\min_{\boldsymbol{A}} \|\boldsymbol{Y}_1 - \boldsymbol{W}_1 * (\boldsymbol{C} \odot_p \boldsymbol{B})\boldsymbol{A}^\top\|_F^2 + \lambda_1 \|\boldsymbol{A}\|_F^2$$
$$\boldsymbol{B} \leftarrow \arg\min_{\boldsymbol{B}} \|\boldsymbol{Y}_2 - \boldsymbol{W}_2 * (\boldsymbol{C} \odot_p \boldsymbol{A})\boldsymbol{B}^\top\|_F^2 + \lambda_2 \|\boldsymbol{B}\|_F^2$$
$$\boldsymbol{C} \leftarrow \arg\min_{\boldsymbol{C}} \|\boldsymbol{Y}_3 - \boldsymbol{W}_3 * \boldsymbol{S}\boldsymbol{C}^\top\|_F^2 + \lambda_3 \|\boldsymbol{C}\|_F^2,$$

where $\boldsymbol{S} = [\text{vec}(\boldsymbol{A}_1\boldsymbol{B}_1^\top), \cdots, \text{vec}(\boldsymbol{A}_R\boldsymbol{B}_R^\top)]$. Each subproblem boils down to a set of unconstrained least squares problems—see details in Appendix F in the supplementary material.

Remark 4 Problem (10) can be handled by the same algorithm, with a slightly different definition of $\underline{\boldsymbol{W}}$—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 $\underline{\boldsymbol{W}}(i, j, k) = \sqrt{P}$. Under this definition of $\underline{\boldsymbol{W}}$, Problem (10) can be re-formulated in the form of (12), thereby the same algorithm can be used.

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 $\boldsymbol{S}$: Our identifiability theorems assume that $\boldsymbol{S}_r$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 $\boldsymbol{C}$ is usually (much) better than that of $\widehat{\boldsymbol{S}}_r = \widehat{\boldsymbol{A}}_r \widehat{\boldsymbol{B}}_r^\top$. It is of interest to exploit this empirical observation.

To this end, note that in (4b), $\boldsymbol{X}_3(\ell, :) = \boldsymbol{S}(\ell, :)\boldsymbol{C}^\top$, where $\boldsymbol{S} = [\text{vec}(\boldsymbol{S}_1), \cdots, \text{vec}(\boldsymbol{S}_R)]$. If $\boldsymbol{C}$ is accurately estimated and $\boldsymbol{C}$ has full column rank, then, $\widehat{\boldsymbol{S}}(\ell, :) = \boldsymbol{X}_3(\ell, :)(\widehat{\boldsymbol{C}^\top})^\dagger$ is a fairly accurate estimate for $\boldsymbol{S}(\ell, :)$. Note that $\boldsymbol{S}(\ell, r) = \boldsymbol{S}_r(i, j)$ for $\ell = (i-1)I + j$—that is, there are a lot of elements of $\boldsymbol{S}_r$ 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 [55], [56]. Since $\boldsymbol{S}_r$ 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) [57]. After estimating $\widehat{S}$, the estimate of $\widehat{X}$ can be obtained from its unfolding $\widehat{X}_3 = \widehat{S}\widehat{C}^\top$

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 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 [47], [48]. We omit the details due to space limitations.

Remark 5 Another possible performance-enhancement strategy is to incorporate prior information on c_r and S_r, e.g., nonnegativity (per their physical meaning) and spatial/spectral smoothness. In terms of algorithm design, adding nonnegativity constraints to the formulated problem (13) is readily within reach, which leads to solving a series of nonnegativity-constrained least squares problems for updating the latent factors. Using structural information may help fend against heavy noise or shadowing.

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 100×100 m^2. The region is discretized into 101×101 grids ($\{0, 1, \cdots, 100\} \times \{0, 1, \cdots, 100\}$), i.e., $I = J = 101$. 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 $c_r(k) = \sum_{i=1}^3 p_i^r a_i^r \text{sinc}^2\left(\frac{k-f_i^r}{w_i^r}\right)$, where p_i^r $(i = 1, 2, 3)$ follows a binomial distribution with equal probability; a_i^r follows a uniform distribution from 0.5 to 2; $f_i^r \in \{1, \cdots, K\}$ is the center of the i-th square 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 $y = [i, j]^\top \in \mathbb{R}^2$ is generated by $S_r(i, j) = \|y - z_r\|_2^{-\eta_r} 10^{z_r(y)/10}$, where $z_r \in \mathbb{R}^2$ denotes the location of emitter r; η_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

$$\mathbb{E}\{z_r(y)z_r(y')\} = \sigma_r^2 \exp(-\|y - y'\|_2/X_c) \quad (15)$$

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

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

$$\widehat{\pi} = \arg\min_{\pi \in \Phi} \sum_{r=1}^R \left\| \frac{c_r}{\|c_r\|_1} - \frac{\widehat{c}_{\pi_r}}{\|\widehat{c}_{\pi_r}\|_1} \right\|_1$$

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

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

$$\text{NAE}_C = \frac{1}{R}\sum_{r=1}^R \left\| \frac{c_r}{\|c_r\|_1} - \frac{\widehat{c}_r}{\|\widehat{c}_r\|_1} \right\|_1;$$

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:

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

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. For the proposed method, we set L for our method via empirical tuning—since we observe that the results are not sensitive to L. However, one may also employ existing L estimation methods for LL1 in the literature; see, e.g., [58], [59]. The parameter R is the number of emitters and is assumed to be known or has been estimated.

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 $\text{NAE}_C \approx 0.01, 0.1, 0.3$. When $\text{NAE}_C \approx 0.01$, the estimated PSD is almost identical the ground-truth PSD. When $\text{NAE}_C \approx 0.3$, 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 $\underline{X}^{(1)} = \underline{X}(\mathcal{S}_1, :, :)$ and $\underline{X}^{(2)} = \underline{X}(:, \mathcal{S}_2, :)$, with $|\mathcal{S}_1| = M, |\mathcal{S}_2| = N$. We use the formulation in Eq. (8) as the disaggregation criterion. The regularization parameters λ_n is set to $\lambda_n = 10^{-2}$ for $n = 1, 2, 3$.

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., $\underline{X}^{(1)} = \underline{X}(\mathcal{S}_1, :, :)$ and $\underline{X}^{(2)} = \underline{X}(:, \mathcal{S}_2, :)$ are sampled (where $\mathcal{S}_1 = \{17, 33, 49, 65, 81\}, \mathcal{S}_2 = \{15, 29, 43, 57, 71, 85\}$). 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 SLFs are shown in Fig. 11 (a1-b3). One can see that both the PSDs and SLFs are reasonably recovered, under such a challenging scenario.
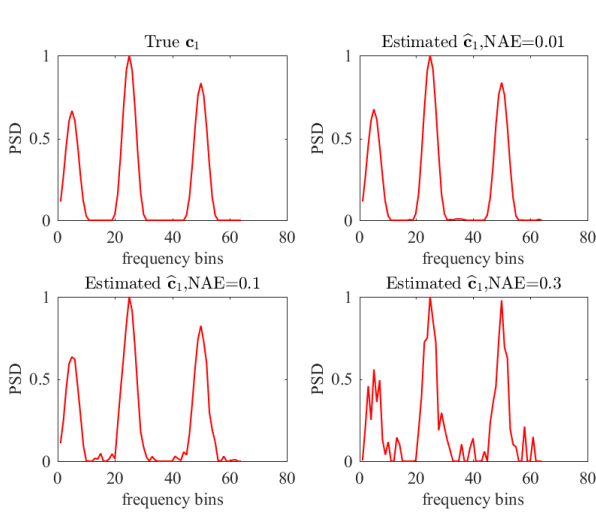
Fig. 7. Illustration of the ground-truth c_1 and the estimated $\widehat{c}_1$ corresponding to different NAEs.
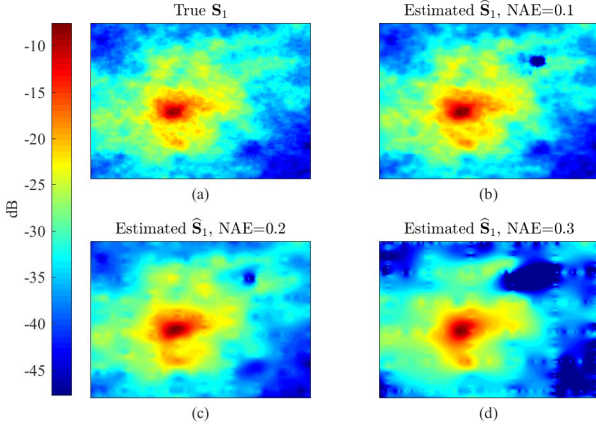

Fig. 8. Illustration of the ground-truth S_1 and the estimated $\widehat{S}_1$ corresponding to different NAEs.
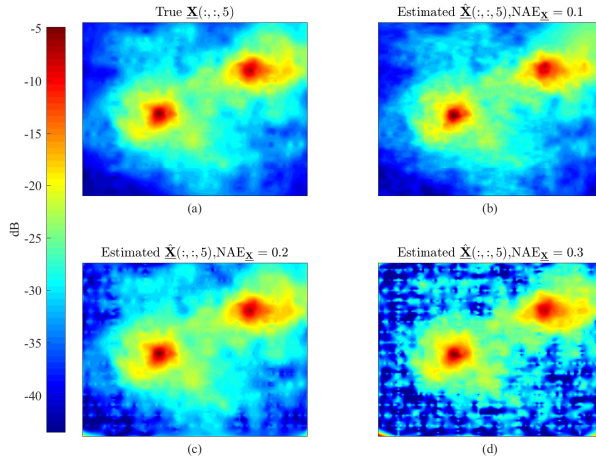

Fig. 9. Illustration of the ground-truth $\underline{X}(:,:,5)$ and the estimated $\widehat{\underline{X}}(:,:,5)$ corresponding to different NAEs.
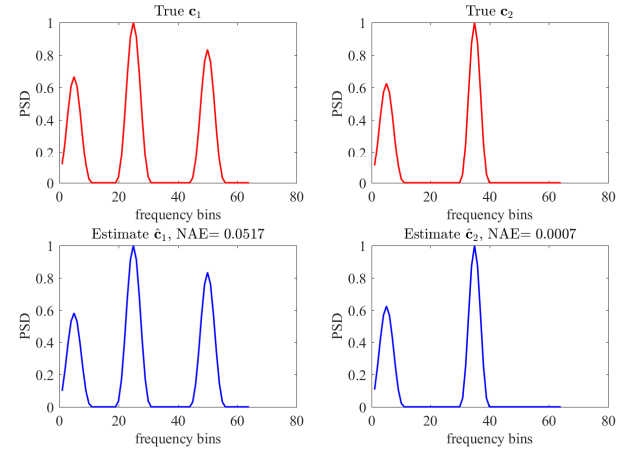

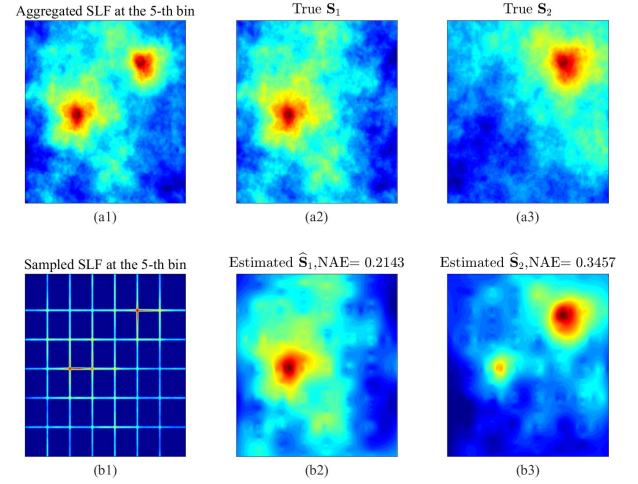Fig. 10. Ground-truth (top) and estimated (bottom) PSDs of two sources.

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 $\underline{X}$ (the unsampled entries was marked in dark blue and the color of sampled entries remains uncharged.); (b2) The estimated SLF of source 1; (b3) The estimated SLF of source 2.

We also evaluate NAE_C, NAE_S and $\text{NAE}_{\underline{X}}$ versus L under different M, N, 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. One can also see that the choice of L does affect the recovery accuracy to a certain extent, but not heavily. The chosen L controls the approximation quality for the SLFs. The tradeoff is that a small L may result in a coarse approximation for the SLFs, but a large L makes computation harder.

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 use NAE_S and $\text{NAE}_{\underline{X}}$ as its performance metrics (since SR assumes that C is known). Note that SR also works

TABLE I
NAEs UNDER DIFFERENT L, R.

NAE_C	$L=2$	$L=3$	$L=4$	$L=5$
$R=2, M=15, N=6$	0.0038	0.0009	0.0023	0.0010
$R=3, M=15, N=6$	0.0067	0.0065	0.0038	0.0056
$R=5, M=15, N=6$	0.0219	0.0198	0.0270	0.0214
$R=8, M=15, N=6$	0.0731	0.0856	0.0620	0.0923
NAE_S	$L=2$	$L=3$	$L=4$	$L=5$
$R=2, M=15, N=6$	0.2335	0.2317	0.2393	0.2456
$R=3, M=15, N=6$	0.2433	0.2515	0.2358	0.2143
$R=5, M=15, N=6$	0.2436	0.2331	0.2528	0.2539
$R=8, M=15, N=6$	0.2656	0.2786	0.2611	0.2664
$\text{NAE}_{\boldsymbol{X}}$	$L=2$	$L=3$	$L=4$	$L=5$
$R=2, M=15, N=6$	0.2087	0.2114	0.2169	0.2253
$R=3, M=15, N=6$	0.2039	0.2154	0.2025	0.2016
$R=5, M=15, N=6$	0.2118	0.2084	0.2285	0.2228
$R=8, M=15, N=6$	0.2291	0.2442	0.2265	0.2361

TABLE II
PERFORMANCE UNDER $L=4, R=3, N=6$.

Algorithm	Measure	M			
		5	10	15	20
Proposed	NAE_C	0.0381	0.0088	0.0055	0.0051
	NAE_S	0.4717	0.3420	0.2474	0.1905
	$\text{NAE}_{\boldsymbol{X}}$	0.3889	0.3017	0.2194	0.1693
	Running time(s)	1.1084	1.1716	1.4226	1.6292
SR	NAE_S	0.4580	0.3208	0.2281	0.1884
	$\text{NAE}_{\boldsymbol{X}}$	0.3795	0.2824	0.2083	0.1760
	Running time(s)	0.2883	0.4581	0.6992	0.9900
TPS	NAE_C	0.5995	0.0246	0.0117	0.0110
	NAE_S	0.9922	0.5430	0.3753	0.3249
	$\text{NAE}_{\boldsymbol{X}}$	0.6671	0.4487	0.3226	0.2492
	Running time(s)	13.3200	21.3532	31.2258	43.0450
GLS	NAE_C	0.5741	0.0216	0.0427	0.0251
	NAE_S	0.9915	0.5518	0.4487	0.4402
	$\text{NAE}_{\boldsymbol{X}}$	0.6549	0.4746	0.3196	0.2793
	Running time(s)	32.1612	59.2353	113.6869	195.5264

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

TABLE III
PERFORMANCE W.R.T DIFFERENT X_c UNDER
$M=15, N=6, K=128, R=3, \sigma=4$.

Algorithm	Measure	X_c				
		10	30	50	70	100
Proposed	NAE_C	0.0074	0.0056	0.0044	0.0038	0.0026
	NAE_S	0.2979	0.2201	0.2063	0.1826	0.1855
	$\text{NAE}_{\boldsymbol{X}}$	0.2732	0.1951	0.1841	0.1686	0.1703
SR	NAE_S	0.305	0.209	0.1944	0.196	0.1667
	$\text{NAE}_{\boldsymbol{X}}$	0.2742	0.191	0.1761	0.1701	0.1465
TPS	NAE_C	0.0129	0.0051	0.0116	0.0064	0.0074
	NAE_S	0.443	0.3582	0.4145	0.3246	0.356
	$\text{NAE}_{\boldsymbol{X}}$	0.4057	0.3035	0.3116	0.294	0.2989
GLS	NAE_C	0.0143	0.0092	0.0075	0.0092	0.0052
	NAE_S	0.4717	0.3648	0.3653	0.3875	0.3450
	$\text{NAE}_{\boldsymbol{X}}$	0.3936	0.3045	0.3251	0.3189	0.2891

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 $\boldsymbol{S}_r$ may be grossly violated. In particular, we test the algorithms under different X_cs and σ in (15). Note that a larger X_c means milder shadowing effect while a larger σ 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. 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.

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., $\underline{\boldsymbol{X}}(:,:,k)$ for $k = 1, \ldots, K$. We then apply the LL1 decomposition on the TPS-estimated $\widehat{\underline{\boldsymbol{X}}}$ 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., $\text{sinc}^2\left(\frac{k-f_i^r}{w_i^r}\right)$ 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 $\widehat{\underline{\boldsymbol{X}}}$ as in the TPS case.

Table II shows the results under $L = 4, R = 3, N = 6, K = 128$ and various Ms. 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 $\boldsymbol{C}$. The NAE_S and $\text{NAE}_{\boldsymbol{X}}$ of the proposed algorithm approximate those of SR. The NAE_Ss 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
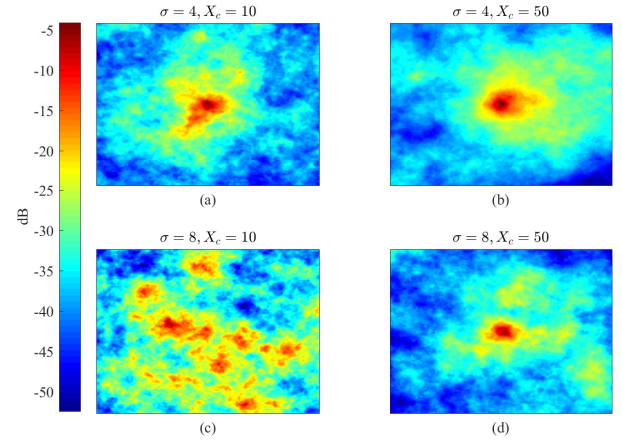
TABLE IV
PERFORMANCE W.R.T DIFFERENT X_c UNDER
$M=15, N=6, K=128, R=3, \sigma=8$.

Algorithm	Measure	X_c				
		10	30	50	70	100
Proposed	NAE_C	0.0181	0.0109	0.0076	0.0094	0.009
	NAE_S	0.5328	0.3482	0.2768	0.2543	0.2354
	$\text{NAE}_{\boldsymbol{X}}$	0.4736	0.3143	0.2544	0.2282	0.2132
SR	NAE_S	0.5156	0.332	0.2717	0.2543	0.2092
	$\text{NAE}_{\boldsymbol{X}}$	0.4601	0.3036	0.2471	0.2292	0.197
TPS	NAE_C	0.3692	0.3633	0.0325	0.0202	0.0071
	NAE_S	0.8264	0.6841	0.4503	0.397	0.385
	$\text{NAE}_{\boldsymbol{X}}$	0.6117	0.379	0.354	0.3399	0.3411
GLS	NAE_C	0.0452	0.0407	0.0575	0.0155	0.0123
	NAE_S	0.6642	0.5023	0.4213	0.4081	0.3892
	$\text{NAE}_{\boldsymbol{X}}$	0.5758	0.4080	0.3331	0.3115	0.3165

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 $\rho = \frac{\text{\# of sampled entries}}{IJK}$. First, we give the results of NAEs versus different sampling ratio, under $K = 128, R = 3, L = 4$. In random fiber sampling, ρ varies from 0.10 to 0.25 with interval 0.05.
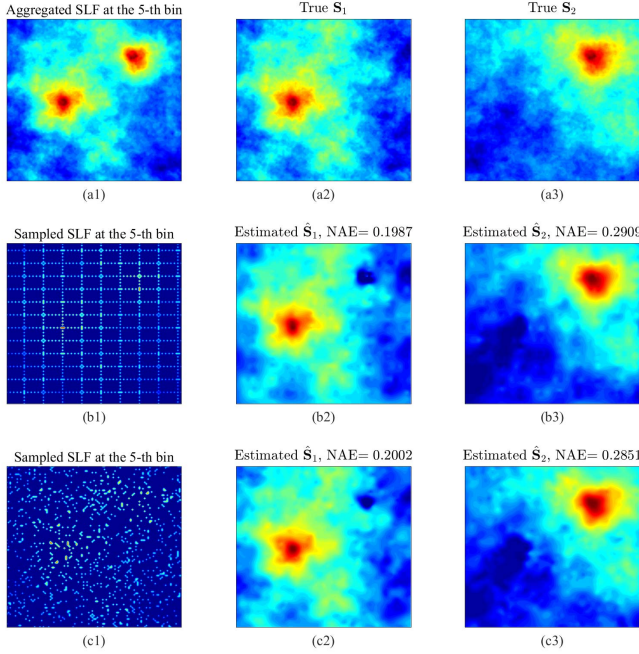

Fig. 13. Performance under regular and random sampling with $\rho = 0.10$. (a1) The ground-truth aggregated SLF at the 5-th frequency bin; (a2) - (a3) The ground-truth SLFs. (b1) regular sampling pattern similar as that in Fig. 6. (b2)-(b3): estimated SLFs by the proposed approach. (c1) random sampling pattern. (c2)-(c3) by the proposed approach.

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.

Table V shows the NAEs performance under different sampling ratios. The fibers are sampled uniformly at random. One can see that the estimation accuracy of the proposed approach is very promising, whose $\boldsymbol{S}$-estimation and $\boldsymbol{X}$-estimation accuracy approach the performance of SR—but SR assumes that $\boldsymbol{C}$ is known, while the proposed approach is a completely blind method. In addition, our method for estimating $\boldsymbol{S}$ and $\boldsymbol{X}$ is more accurate compared to TPS and GLS, which is similar to what we have observed in the slab sampling case.

C. Performance under Noise

We also consider a case where the measurements are noisy, i.e., $\underline{\tilde{\boldsymbol{X}}} = \underline{\boldsymbol{X}} + \underline{\boldsymbol{N}}$ where $\underline{\boldsymbol{X}}$ follows the model in (2)

and the entries of $\underline{\boldsymbol{N}}$ follow the zero-mean i.i.d. Gaussian distribution. The signal-to-noise ratio (SNR) is defined as $10 \log_{10}(\|\underline{\boldsymbol{X}}\|_F^2 / \|\underline{\boldsymbol{N}}\|_F^2)$. We would like to remark that the Gaussian noise is added to test robustness of the algorithms under model mismatch. In practice, the noise may not be Gaussian since the data is nonnegative.

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 fairly well in terms of estimating $\boldsymbol{S}_r$, $\boldsymbol{c}$ and $\underline{\boldsymbol{X}}$, 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.

TABLE V
PERFORMANCE UNDER $L = 4, R = 3, K = 128$.

Algorithm	Measure	ρ			
		0.1	0.15	0.2	0.25
Proposed	NAE_C	0.0176	0.0154	0.0198	0.0188
	NAE_S	0.2152	0.1721	0.1583	0.1329
	NAE_X	0.1983	0.1629	0.1400	0.1190
	Running time(s)	93.5609	92.3041	93.0684	93.6724
SR	NAE_S	0.1846	0.1537	0.1345	0.1231
	NAE_X	0.1769	0.1440	0.1267	0.1158
	Running time(s)	0.3684	0.5481	0.7880	1.1011
TPS	NAE_C	0.0120	0.0153	0.0091	0.0086
	NAE_S	0.4388	0.3688	0.3191	0.2685
	NAE_X	0.4063	0.3301	0.2786	0.2424
	Running time(s)	14.9322	30.1523	48.6299	54.0290
GLS	NAE_C	0.0306	0.0087	0.0063	0.0076
	NAE_S	0.3452	0.2558	0.2287	0.1980
	NAE_X	0.3061	0.2328	0.2059	0.1764
	Running time(s)	21.8072	49.2587	111.7062	203.3006

TABLE VI
PERFORMANCE OF SLAB SAMPLING UNDER
$L = 4, R = 3, K = 128, M = 15, N = 6$ VERSUS DIFFERENT SNR.

Algorithm	Measure	SNR(dB)			
		0	10	20	30
Proposed	NAE_C	0.1532	0.0543	0.0193	0.0098
	NAE_S	0.5441	0.3494	0.2797	0.2595
	NAE_X	0.6034	0.3431	0.2557	0.2282
SR	NAE_S	0.6742	0.3912	0.2711	0.2450
	NAE_X	0.9681	0.4288	0.2654	0.2204
TPS	NAE_C	0.1993	0.0641	0.0259	0.0236
	NAE_S	0.8855	0.5371	0.4450	0.4319
	NAE_X	10.2674	3.4584	1.2773	0.6349
GLS	NAE_C	0.0271	0.0119	0.0126	0.0137
	NAE_S	0.8381	0.4849	0.4176	0.3450
	NAE_X	1.0426	0.5175	0.3909	0.3109

TABLE VII
PERFORMANCE OF FIBER SAMPLING UNDER
$L = 4, R = 3, K = 128, p = 0.15$ VERSUS DIFFERENT SNR.

Algorithm	Measure	SNR(dB)			
		0	10	20	30
Proposed	NAE_C	0.1779	0.0753	0.0232	0.0170
	NAE_S	0.7947	0.4178	0.2538	0.2097
	NAE_X	1.1255	0.4803	0.2500	0.2007
SR	NAE_S	0.7305	0.3856	0.2590	0.2128
	NAE_X	1.1170	0.4540	0.2619	0.2053
TPS	NAE_C	0.1911	0.0820	0.0343	0.0161
	NAE_S	0.9055	0.4974	0.3382	0.3009
	NAE_X	13.9170	4.3835	1.6014	0.6653
GLS	NAE_C	0.0211	0.0111	0.0068	0.0077
	NAE_S	0.8396	0.4830	0.3312	0.2885
	NAE_X	1.1745	0.5140	0.3259	0.2564

D. Real-data Experiment

In this section, we test our algorithm on the real data collected in [60]. There, the measurements are taken in a 14×34 m² indoor area over 9 frequencies spanning from 2.412GHz to 2.462GHz. The complete RF tensor is heavily downsampled [see Fig. 14 (left)]. To be specific, the complete RF tensor has a size of $14 \times 34 \times 9$. Three horizontal slabs are observed. Three vertical slabs are observed except for $\underline{X}(6,1,:)$ and $\underline{X}(7,1,:)$. We apply our slab sampling-based methods (where $\underline{X}(6,1,:)$ and $\underline{X}(7,1,:)$ are assigned with the mean of their adjacent elements). Two subtensors $\underline{X}^{(1)} = \underline{X}(:,1:3,:)$ and $\underline{X}^{(2)} = \underline{X}([4,8,9],:,:)$ are thus formed. Note that we do not have the ground truth of the complete tensor. We also do not know the number of emitters in the region. Hence, we follow [15] to offer qualitative evaluations; i.e., we plot the recovered RF tensor within the observed region and unobserved region to visually compare with the original data. To run our algorithm, we set $R = 9$ and $L = 4$.

Our recovered $\underline{X}(:,:,k)$ seem to be visually smoother relative to the method in [15] (see Fig. 8 in [15]). This might be because the LL1 model explicitly enforces spatial smoothness (via enforcing $\boldsymbol{S}_r$ to have low rank).
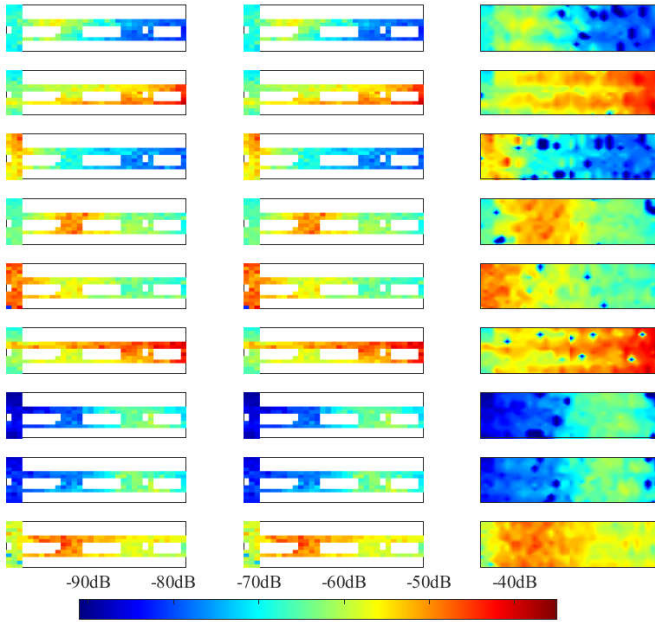
Fig. 14. Recovered $\underline{X}(:,:,k)$ using real data ($R = 9, L = 4$) by the proposed method. (Left) Original data; (center) Estimated data on observed locations; (right) Estimated data of all locations.

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.

In terms of future work, one potential direction is to study the frequency-selective fading scenario, under which the LL1 model does not hold. This may require more complex tensor models, e.g., those in [46], [61], [62]. In addition, another direction may be large-scale, online coupled LL1 decomposition algorithm design incorporating prior information—which was shown effective in other tensor models in noisy situations [63]. Last but not least, one may take temporal variations into consideration, to accommodate fast changing scenarios.

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 $(\widehat{A}, \widehat{B}, \widehat{C})$ denote any optimal solution of Problem (8) and (A, B, C) denote the ground-truth. We also define $\boldsymbol{S}_r = \boldsymbol{A}_r \boldsymbol{B}_r^\top$ and $\widehat{\boldsymbol{S}}_r = \widehat{\boldsymbol{A}}_r \widehat{\boldsymbol{B}}_r^\top$. We aim to prove that $\{\widehat{\boldsymbol{S}}_r, \widehat{\boldsymbol{c}}_r\}_{r=1}^R$ is essentially the ground-truth $\{\boldsymbol{S}_r, \boldsymbol{c}_r\}_{r=1}^R$ 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 symmetry of M and N. Note that $\underline{X}^{(1)} \in \mathbb{R}^{M \times J \times K_1}$ and that $\boldsymbol{P}\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}(\mathcal{S}_3,:)$ are latent factors of the decomposition of $\underline{X}^{(1)}$ in multilinear rank-$(L,L,1)$ terms. Since $R \le |\mathcal{S}_3 \cap \mathcal{S}_4| \le K_1, R + 2 \le \min(\lfloor M/L \rfloor, R) + \min(\lfloor J/L \rfloor, R)$, by Theorem 1, $\boldsymbol{P}\boldsymbol{A}_r\boldsymbol{B}_r^\top$ and $\boldsymbol{c}_r(\mathcal{S}_3)$ for $r = 1 \ldots, R$ can be identified from $\underline{X}^{(1)}$ up to scaling and permutation ambiguities. Therefore, $\boldsymbol{C}(\mathcal{S}_3,:)$ can be identified up to scaling and permutation ambiguities, i.e, $\widehat{\boldsymbol{C}}(\mathcal{S}_3,:) = \boldsymbol{C}(\mathcal{S}_3,:)\boldsymbol{\Pi}\boldsymbol{\Lambda}$.

Accordingly, we have $\widehat{\boldsymbol{c}}_r(\mathcal{S}_3) = \lambda_r \boldsymbol{c}_{\pi_r}(\mathcal{S}_3)$, where $\boldsymbol{\Lambda} = \text{Diag}(\lambda_1, \cdots, \lambda_R)$ and $\pi_r \in \{1, \cdots, R\}$ satisfies $\boldsymbol{\Pi}(\pi_r, r) = 1$ for $r = 1, \cdots, R$. Note that

$$
\begin{aligned}
\underline{X}_3^{(1)} &= [\text{vec}(\boldsymbol{P}\boldsymbol{A}_1\boldsymbol{B}_1^\top), \cdots, \text{vec}(\boldsymbol{P}\boldsymbol{A}_R\boldsymbol{B}_R^\top)]\boldsymbol{C}(\mathcal{S}_3,:)^\top \\
&= [\text{vec}(\boldsymbol{P}\widehat{\boldsymbol{A}}_1\widehat{\boldsymbol{B}}_1^\top), \cdots, \text{vec}(\boldsymbol{P}\widehat{\boldsymbol{A}}_R\widehat{\boldsymbol{B}}_R^\top)]\widehat{\boldsymbol{C}}(\mathcal{S}_3,:)^\top.
\end{aligned}
\tag{16}
$$

Plugging $\widehat{\boldsymbol{C}}(\mathcal{S}_3,:) = \boldsymbol{C}(\mathcal{S}_3,:)\boldsymbol{\Pi}\boldsymbol{\Lambda}$ into (16), we have $\boldsymbol{P}\widehat{\boldsymbol{A}}_r\widehat{\boldsymbol{B}}_r^\top = \lambda_r^{-1}\boldsymbol{P}\boldsymbol{A}_{\pi_r}\boldsymbol{B}_{\pi_r}^\top$. Here $\boldsymbol{A}_r, \widehat{\boldsymbol{A}}_r, \boldsymbol{B}_r, \widehat{\boldsymbol{B}}_r$ are all full column-rank matrices almost surely (since they are drawn from certain joint continuous distributions), and the row selection matrix $\boldsymbol{P}$ is a full row-rank matrix. Hence, $\boldsymbol{P}\boldsymbol{A}_r \in \mathbb{R}^{M \times L}$ is a submatrix of $\boldsymbol{A}_r$ and $M \ge 2L$, which means that $\boldsymbol{P}\boldsymbol{A}_r$ is a full column-rank matrix and so is $\boldsymbol{P}\widehat{\boldsymbol{A}}_r$. Therefore, there exists a full rank matrix $\boldsymbol{F}_r$ satisfying $\widehat{\boldsymbol{B}}_r = \boldsymbol{B}_{\pi_r}\boldsymbol{F}_r$.

Note that $\mathcal{S} = \mathcal{S}_3 \cap \mathcal{S}_4$ and $\widehat{\boldsymbol{C}}(\mathcal{S}_3,:) = \boldsymbol{C}(\mathcal{S}_3,:)\boldsymbol{\Pi}\boldsymbol{\Lambda}$. Hence, we have $\widehat{\boldsymbol{C}}(\mathcal{S},:) = \boldsymbol{C}(\mathcal{S},:)\boldsymbol{\Pi}\boldsymbol{\Lambda}$. Consider the subtensor $\underline{X}^{(3)} = \underline{X}(:,\mathcal{S}_2,\mathcal{S})$, where we have

$$
\underline{X}_3^{(3)} = \boldsymbol{M}_Q\boldsymbol{C}(\mathcal{S},:)^\top = \widehat{\boldsymbol{M}}_Q\widehat{\boldsymbol{C}}(\mathcal{S},:)^\top,
\tag{17}
$$

where

$$
\begin{aligned}
\boldsymbol{M}_Q &= [\text{vec}(\boldsymbol{A}_1\boldsymbol{B}_1^\top\boldsymbol{Q}^\top), \cdots, \text{vec}(\boldsymbol{A}_R\boldsymbol{B}_R^\top\boldsymbol{Q}^\top)] \\
\widehat{\boldsymbol{M}}_Q &= [\text{vec}(\widehat{\boldsymbol{A}}_1\widehat{\boldsymbol{B}}_1^\top\boldsymbol{Q}^\top), \cdots, \text{vec}(\widehat{\boldsymbol{A}}_R\widehat{\boldsymbol{B}}_R^\top\boldsymbol{Q}^\top)].
\end{aligned}
$$

Plugging $\widehat{C}(\mathcal{S},:) = C(\mathcal{S},:)\Pi\Lambda$ into (17), we have $\widehat{A}_r\widehat{B}_r^\top Q^\top = \lambda_r^{-1} A_{\pi_r} B_{\pi_r}^\top Q^\top$ since $C(\mathcal{S},:)$ is full column rank. Here Q is a row-selection matrix corresponding to $\mathcal{S}$. Plug $\widehat{B}_r = B_{\pi_r} F_r$ into this equation, we have $\widehat{A}_r F_r^\top B_{\pi_r}^\top Q^\top = \lambda_r^{-1} A_{\pi_r} B_{\pi_r}^\top Q^\top$. Since $B_{\pi_r}^\top Q^\top$ is a full row rank matrix, we can infer $\widehat{A}_r = \lambda_r^{-1} A_{\pi_r} F_r^{-\top}$. Therefore, $\widehat{S}_r = \widehat{A}_r\widehat{B}_r^\top = \lambda_r^{-1} A_{\pi_r} F_r^{-\top}(B_{\pi_r} F_r)^\top = \lambda_r^{-1} A_{\pi_r} B_{\pi_r}^\top = \lambda_r^{-1} S_{\pi_r}$; that is, we have proven that $\{\widehat{S}_r\}_{r=1}^R$ is a scaling and permutation version of ground-truth $\{S_r\}_{r=1}^R$.

Next, we aim to prove that $\widehat{C} = C\Pi\Lambda$. In particular, if $\mathcal{S}_3 = \mathcal{S}_4 = \mathcal{S} = [K]$, $\widehat{C} = C\Pi\Lambda$ holds since the LL1 decomposition of $\underline{X}^{(1)}$ is essentially unique.

If $\mathcal{S}_3, \mathcal{S}_4 \neq [K]$ while $\mathcal{S}_3 \cup \mathcal{S}_4 = [K]$, noting that $\widehat{A}_r\widehat{B}_r^\top Q^\top = \widehat{S}_r Q^\top = \lambda_r^{-1} S_{\pi_r} Q^\top$, we have

$$
\begin{aligned}
\widehat{M}_Q &= [\text{vec}(\widehat{A}_1\widehat{B}_1^\top Q^\top), \cdots, \text{vec}(\widehat{A}_R\widehat{B}_R^\top Q^\top) \\
&= [\lambda_1^{-1}\text{vec}(A_{\pi_1} B_{\pi_1}^\top Q^\top), \cdots, \lambda_R^{-1}\text{vec}(A_{\pi_R} B_{\pi_R}^\top Q^\top)] \\
&= M_Q\Pi\Lambda^{-1}.
\end{aligned}
$$

Consider the subtensor $\underline{X}^{(2)}$, we have

$$
\underline{X}_3^{(2)} = M_Q C(\mathcal{S}_4,:)^\top = \widehat{M}_Q \widehat{C}(\mathcal{S}_4,:)^\top. \tag{18}
$$

Plugging $\widehat{M}_Q = M_Q\Pi\Lambda^{-1}$ into (18), we obtain $\widehat{C}(\mathcal{S}_4,:) = C(\mathcal{S}_4,:)\Pi\Lambda$. Combining with the fact that $\widehat{C}(\mathcal{S}_3,:) = C(\mathcal{S}_3,:)\Pi\Lambda$ and $\mathcal{S}_3 \cup \mathcal{S}_4 = [K]$, we have $\widehat{C} = C\Pi\Lambda$.

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 $(\widehat{A}, \widehat{B}, \widehat{C})$ denote any optimal solution of Problem (10) and (A, B, C) denote the ground-truth. Note that

$$
\begin{aligned}
\underline{X}^{(d)} &= \sum_{r=1}^R (A_r(\mathcal{I}^{(d)},:)B_r(\mathcal{J}^{(d)},:)^\top) \circ C(\mathcal{K}^{(d)},:) \\
&= \sum_{r=1}^R (\widehat{A}_r(\mathcal{I}^{(d)},:)\widehat{B}_r(\mathcal{J}^{(d)},:)^\top) \circ \widehat{C}(\mathcal{K}^{(d)},:).
\end{aligned}
$$

By Theorem 1, one can see that every $\underline{X}^{(d)}$ admits a unique BTD, which means that $\widehat{C}(\mathcal{K}^{(d)},:)$ is a column permutated and scaled version of $C(\mathcal{K}^{(d)},:)$. 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

$$
\widehat{C}(\mathcal{K}^{(d)},:) = C(\mathcal{K}^{(d)},:)\Pi^{(d)}\Lambda^{(d)}.
$$

For any two coupled tensors $\underline{X}^{(i_d)}$ and $\underline{X}^{(i_{d+1})}$, we have

$$
\begin{aligned}
\widehat{C}(\mathcal{K}^{(i_d)},:) &= C(\mathcal{K}^{(i_d)},:)\Pi^{(i_d)}\Lambda^{(i_d)}, \\
\widehat{C}(\mathcal{K}^{(i_{d+1})},:) &= C(\mathcal{K}^{(i_{d+1})},:)\Pi^{(i_{d+1})}\Lambda^{(i_{d+1})}.
\end{aligned}
$$

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 $\mathcal{R}^{(i_d)} \doteq \mathcal{K}^{(i_d)} \cap \mathcal{K}^{(i_{d+1})}$. Considering the common part $\widehat{C}(\mathcal{R}^{(i_d)},:)$, we have

$$
\begin{aligned}
\widehat{C}(\mathcal{R}^{(i_d)},:) &= \widehat{C}(\mathcal{K}^{(i_d)} \cap \mathcal{R}^{(i_d)},:) = C(\mathcal{R}^{(i_d)},:)\Pi^{(i_d)}\Lambda^{(i_d)}, \\
\widehat{C}(\mathcal{R}^{(i_d)},:) &= \widehat{C}(\mathcal{K}^{(i_{d+1})} \cap \mathcal{R}^{(i_d)},:) = C(\mathcal{R}^{(i_d)},:)\Pi^{(i_{d+1})}\Lambda^{(i_{d+1})}.
\end{aligned}
$$

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

$$
C(\mathcal{R}^{(i_d)},:)D^{(i_d)} = C(\mathcal{R}^{(i_d)},:)D^{(i_{d+1})}. \tag{19}
$$

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

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

Note that $d_r^{(i_d)} \in \mathbb{R}^R$ 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 $|\mathcal{R}^{(i_d)}| \geq 2$, which means that any two columns of $C(\mathcal{R}^{(i_d)},:)$ 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 $\widehat{C}(\mathcal{K}^{(i_d)},:) = C(\mathcal{K}^{(d)},:)\Pi\Lambda$ for $d = 1, \cdots, D$, we have

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

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

$$
\begin{aligned}
S &= [\text{vec}(S_1), \cdots, \text{vec}(S_R)], \\
\widehat{S} &= [\text{vec}(\widehat{S}_1), \cdots, \text{vec}(\widehat{S}_R)].
\end{aligned}
$$

and $S_r = A_r B_r^\top$ and $\widehat{S}_r = \widehat{A}_r\widehat{B}_r^\top$. Note that

$$
\begin{aligned}
\underline{X}_3^{(d)} &= \underbrace{[\cdots, \text{vec}(S_r(\mathcal{I}^{(d)}, \mathcal{J}^{(d)})), \cdots]}_{S^{(d)}} C(\mathcal{K}^{(d)},:)^\top \\
&= \underbrace{[\cdots, \text{vec}(\widehat{S}_r(\mathcal{I}^{(d)}, \mathcal{J}^{(d)})), \cdots]}_{\widehat{S}^{(d)}} \widehat{C}(\mathcal{K}^{(d)},:)^\top,
\end{aligned}
$$

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

$$
\widehat{S}^{(d)} = S^{(d)}\Pi\Lambda^{-1}, \tag{20}
$$

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

$$
\widehat{S}_r(\mathcal{I}^{(d)}, \mathcal{J}^{(d)}) = \lambda_r^{-1} S_{\pi_r}(\mathcal{I}^{(d)}, \mathcal{J}^{(d)}). \tag{21}
$$

Plugging $S_r = A_r B_r^\top$ and $\widehat{S}_r = \widehat{A}_r\widehat{B}_r^\top$ into (21), we have

$$
\widehat{A}_r(\mathcal{I}^{(d)},:)\widehat{B}_r(\mathcal{J}^{(d)},:)^\top = \lambda_r^{-1} A_{\pi_r}(\mathcal{I}^{(d)},:)B_{\pi_r}(\mathcal{J}^{(d)},:)^\top.
$$

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

$$
\begin{aligned}
\widehat{A}_r(\mathcal{I}^{(d)},:) &= A_{\pi_r}(\mathcal{I}^{(d)},:)F_r^{(d)}, \\
\widehat{B}_r(\mathcal{J}^{(d)},:) &= \lambda_r^{-1} B_{\pi_r}(\mathcal{J}^{(d)},:)(F_r^{(d)})^{-\top}.
\end{aligned}
$$

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:

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

We have

$$\widehat{A}_r(\mathcal{P}^{(i_d)}, :) = \widehat{A}_r(\mathcal{I}^{(i_d)} \cap \mathcal{P}^{(i_d)}, :) = A_{\pi_r}(\mathcal{P}^{(i_d)}, :)F_r^{(i_d)},$$
$$\widehat{A}_r(\mathcal{P}^{(i_d)}, :) = \widehat{A}_r(\mathcal{I}^{(i_{d+1})} \cap \mathcal{P}^{(i_d)}, :) = A_{\pi_r}(\mathcal{P}^{(i_d)}, :)F_r^{(i_{d+1})}.$$

Then, it follows that

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

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 $\widehat{A}_r(\mathcal{I}^{(i_d)}, :) = A_{\pi_r}(\mathcal{I}^{(i_d)}, :)F_r$, we have

$$\widehat{A}_r = \widehat{A}_r(\mathcal{I}^{(i_1)} \cup \cdots \cup \mathcal{I}^{(i_D)}, :) = A_{\pi_r}F_r.$$

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

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

Therefore, we have

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

and

$$\begin{aligned}\widehat{S} &= [\text{vec}(\widehat{S}_1), \cdots, \text{vec}(\widehat{S}_R)] \\ &= [\lambda_1^{-1}\text{vec}(S_{\pi_1}), \cdots, \lambda_R^{-1}\text{vec}(S_{\pi_R})] \\ &= S\Pi\Lambda^{-1}.\end{aligned}$$

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 $(\widehat{A}, \widehat{B}, \widehat{C})$ denote any optimal solution of Problem (10) and (A, B, C) denote the ground-truth. Note that

$$\begin{aligned}\underline{X}^{(d_0)} &= \sum_{r=1}^{R}(A_r(\mathcal{I}^{(d_0)}, :)B_r(\mathcal{J}^{(d_0)}, :)^\top) \circ C(\mathcal{K}^{(d_0)}, :) \\ &= \sum_{r=1}^{R}(\widehat{A}_r(\mathcal{I}^{(d_0)}, :)\widehat{B}_r(\mathcal{J}^{(d_0)}, :)^\top) \circ \widehat{C}(\mathcal{K}^{(d_0)}, :).\end{aligned}$$

Note that $\mathcal{K}^{(d_0)} = [K], K \geq R$ and

$$\min\left(\left\lfloor\frac{|\mathcal{I}^{(d_0)}|}{L}\right\rfloor, R\right) + \min\left(\left\lfloor\frac{|\mathcal{J}^{(d_0)}|}{L}\right\rfloor, R\right) \geq R + 2.$$

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

$$\widehat{C} = C\Pi\Lambda.$$

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

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. Khalfi, 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 Comm.*, vol. 26, no. 2, pp. 133–141, 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. Yatnalli, and D. Gesbert, "Learning radio maps for uav-aided 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. Boccolini, G. Hernandez-Penaloza, and B. Beferull-Lozano, "Wireless sensor network for spectrum cartography based on kriging interpolation," in *Proc. IEEE PIMRC*, 2012, pp. 1565–1570.

[10] S. Üreten, A. Yongaçoğlu, 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. Mueck, "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.

[14] D.-H. Huang, S.-H. Wu, W.-R. Wu, and P.-H. Wang, "Cooperative radio source positioning and power map reconstruction: A sparse bayesian learning approach," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2318–2332, 2014.

[15] J. A. Bazerque, G. Mateos, and G. B. Giannakis, "Group-lasso on splines for spectrum cartography," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4648–4663, 2011.

[16] D. Romero, S.-J. Kim, G. B. Giannakis, and R. López-Valcarce, "Learning power spectrum maps from quantized power measurements," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2547–2560, 2017.

[17] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, p. 717, 2009.

[18] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, 2013.

[19] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms—Part II: Definitions and uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1033–1066, 2008.

[20] X. Fu, N. D. Sidiropoulos, J. H. Tranter, and W.-K. Ma, "A factor analysis framework for power spectra separation and multiple emitter localization," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6581–6594, 2015.

[21] X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Power spectra separation via structured matrix factorization." *IEEE Trans. Signal Process.*, vol. 64, no. 17, pp. 4592–4605, 2016.

[22] G. Zhang, X. Fu, J. Wang, and M. Hong, "Coupled block-term tensor decomposition based blind spectrum cartography," in *Proc Asilomar 2019,* to appear, 2019. [Online]. Available: http://people.oregonstate.edu/~fuxia/zhang2019asilomar.pdf

[23] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," in *CVPR 2011*, 2011, pp. 3449–3456.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2020.2993530, IEEE Transactions on Signal Processing

16

[24] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imaging Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.

[25] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2117–2130, 2013.

[26] D. Goldfarb and Z. Qin, "Robust low-rank tensor recovery: Models and algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 35, no. 1, pp. 225–253, 2014.

[27] Z. Zhang and S. Aeron, "Exact tensor completion using t-svd." *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1511–1526, 2017.

[28] M. Yuan and C.-H. Zhang, "On tensor completion via nuclear norm minimization," *Foundations of Computational Mathematics*, vol. 16, no. 4, pp. 1031–1068, Aug 2016. [Online]. Available: https://doi.org/10.1007/s10208-015-9269-5

[29] C. Mu, B. Huang, J. Wright, and D. Goldfarb, "Square deal: Lower bounds and improved relaxations for tensor recovery," in *Proc. 31th Int. Conf. Mach. Learn. (ICML)*, 2014, pp. 73–81.

[30] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, 2017.

[31] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[32] R. A. Harshman and M. E. Lundy, "Parafac: Parallel factor analysis," *Computational Statistics & Data Analysis*, vol. 18, no. 1, pp. 39–72, 1994.

[33] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.

[34] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2377–2388, 2000.

[35] N. D. Sidiropoulos and X. Liu, "Identifiability results for blind beamforming in incoherent multipath with small delay spread," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 228–236, 2001.

[36] Y. Rong, S. A. Vorobyov, A. B. Gershman, and N. D. Sidiropoulos, "Blind spatial signature estimation via time-varying user power loading and parallel factor analysis," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1697–1710, 2005.

[37] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, "Robust iterative fitting of multilinear models," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2678–2689, 2005.

[38] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms—Part I: Lemmas for partitioned matrices," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1022–1032, 2008.

[39] L. De Lathauwer and D. Nion, "Decompositions of a higher-order tensor in block terms—Part III: Alternating least squares algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1067–1083, 2008.

[40] L. De Lathauwer, "Blind separation of exponential polynomials and the decomposition of a tensor in rank-(l_r,l_r,1) terms," *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 4, pp. 1451–1474, 2011.

[41] D. Lee, S.-J. Kim, and G. B. Giannakis, "Channel gain cartography for cognitive radios leveraging low rank and sparsity," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5953–5966, 2017.

[42] P. N. Alevizos, X. Fu, N. D. Sidiropoulos, Y. Yang, and A. Bletsas, "Limited feedback channel estimation in massive mimo with non-uniform directional dictionaries," *IEEE Trans. Signal Process.*, vol. 66, no. 19, pp. 5127–5141, 2018.

[43] C. Qian, X. Fu, N. D. Sidiropoulos, and Y. Yang, "Tensor-based channel estimation for dual-polarized massive mimo systems," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6390–6403, 2018.

[44] D. Dupleich, S. Häfner, C. Schneider, R. Müller, R. Thomä, J. Luo, N. Iqbal, E. Schulz, X. Lu, and G. Wang, "Double-directional and dual-polarimetric indoor measurements at 70 GHz," in *Proc. IEEE 26th Ann. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Aug./Sep. 2015, pp. 2234–2238.

[45] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.

[46] C. Chatzichristos, E. Kofidis, M. Morante, and S. Theodoridis, "Blind fMRI source unmixing via higher-order tensor decompositions," *Journal of Neuroscience Methods*, vol. 315, pp. 17–47, 2019.

[47] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6503–6517, 2018.

[48] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and M. Akçakaya, "Tensor completion from regular sub-nyquist samples," *IEEE Trans. Signal Process.* to appear. *arXiv preprint arXiv:1903.00435*, 2019.

[49] Y. Teganya, L. M. Lopez-Ramos, D. Romero, and B. Beferull-Lozano, "Localization-free power cartography," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2018, pp. 3549–3553.

[50] M. Sørensen and L. De Lathauwer, "Fiber sampling approach to canonical polyadic decomposition and tensor completion," *ESAT-STADIUS, KU Leuven, Belgium, Tech. Rep*, pp. 15–151, 2017.

[51] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and M. Akçakaya, "Tensor completion from regular sub-nyquist samples," *IEEE Trans. Signal Process.*, vol. 68, pp. 1–16, 2020.

[52] X. Fu, K. Huang, W. Ma, N. D. Sidiropoulos, and R. Bro, "Joint tensor factorization and outlying slab suppression with applications," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6315–6328, Dec 2015.

[53] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.

[54] D. P. Bertsekas, "Nonlinear programming," *Journal of the Operational Research Society*, vol. 48, no. 3, pp. 334–334, 1997.

[55] J. Li and A. D. Heap, "A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors," *Ecological Informatics*, vol. 6, no. 3-4, pp. 228–241, 2011.

[56] M. N. Ikechukwu, E. Ebinne, U. Idorenyin, and N. I. Raphael, "Accuracy assessment and comparative analysis of idw, spline and kriging in spatial interpolation of landform (topography): An experimental study," *Journal of Geographic Information System*, vol. 9, no. 03, p. 354, 2017.

[57] G. Wahba, *Spline models for observational data*. Siam, 1990, vol. 59.

[58] V. Zarzoso, "Parameter estimation in block term decomposition for noninvasive atrial fibrillation analysis," in *Proc. IEEE CAMSAP*, 2017, pp. 1–5.

[59] X. Han, L. Albera, A. Kachenoura, H. Shu, and L. Senhadji, "Block term decomposition with rank estimation using group sparsity," in *Proc. IEEE CAMSAP*, 2017, pp. 1–5.

[60] T. King, S. Kopf, T. Haenselmann, C. Lubberger, and W. Effelsberg, "CRAWDAD dataset mannheim/compass (v. 2008-04-11)," Downloaded from https://crawdad.org/mannheim/compass/20080411, Apr. 2008.

[61] X. Guo, S. Miron, D. Brie, and A. Stegeman, "Uni-mode and partial uniqueness conditions for candecomp/parafac of three-way arrays with linearly dependent loadings," *SIAM Journal on Matrix Analysis and Applications*, vol. 33, no. 1, pp. 111–129, 2012.

[62] A. Stegeman and A. L. De Almeida, "Uniqueness conditions for constrained three-way factor decompositions with linearly dependent loadings," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1469–1490, 2010.

[63] M. Joneidi and N. Rahnavard, "Primary user localization and online radio cartography via structured tensor decomposition," *arXiv preprint arXiv:1905.04284*, 2019.

[64] D. L. Pimentel-Alarcón, N. Boston, and R. D. Nowak, "A characterization of deterministic sampling patterns for low-rank matrix completion," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 623–636, 2016.

[65] J. D. Gardiner, A. J. Laub, J. J. Amato, and C. B. Moler, "Solution of the sylvester matrix equation axb t+ cxd t= e," *ACM Transactions on Mathematical Software (TOMS)*, vol. 18, no. 2, pp. 223–231, 1992.
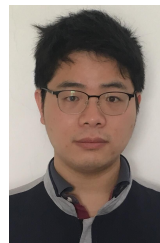
Guoyong Zhang received the B.S. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2014, where he is currently working toward the Ph.D. degree with the National Key Laboratory of Science and Technology on Communications. His currents research interests include cognitive radio and signal processing for wireless communications.
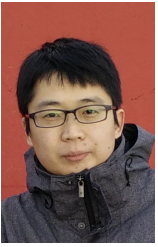
Xiao Fu (S'12-M'15) is an Assistant Professor in the School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, Oregon, United States. He received his Ph.D. degree in Electronic Engineering from The Chinese University of Hong Kong (CUHK), Hong Kong, in 2014. He was a Postdoctoral Associate in the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, United States, from 2014 to 2017. His research interests include the broad area of signal processing and machine learning. He received a Best Student Paper Award at ICASSP 2014. Two of his co-authored papers received the Best Student Awards at IEEE CAMSAP 2015 and IEEE MLSP 2019, respectively. He also received the Outstanding Postdoctoral Scholar Award from the University of Minnesota in 2016. He serves as the treasurer of IEEE Signal Processing Society Oregon Chapter. He also serves as a member of the EURASIP Technical Area Committee in Signal Processing for Multisensor Systems for the term of 2020-2022.

Jun Wang (S'03–M'09) received the B.S. degree in communication engineering and the M.S. and Ph.D. degrees in communication and information systems from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1997, 2000, and 2009, respectively. Since 2000, he has been with the National Key Laboratory of Science and Technology on Communications, UESTC, where he is currently a Professor. His research interests include signal processing for wireless communications, cognitive radio, machine learning for wireless communications. In these areas, he has published over 70 journal and conference papers, and involved over 10 major research programs.

Xi-Le Zhao received the M.S. and Ph.D. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2009 and 2012. He is currently a Professor with the School of Mathematical Sciences, UESTC. His research interests include image processing, computer vision, and machine learning. His website is https://zhaoxile.github.io/.
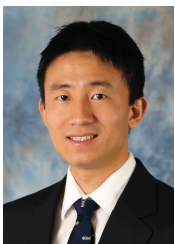
Mingyi Hong received his Ph.D. degree from the University of Virginia, Charlottesville, in 2011. He is an assistant professor in the Department of Electrical and Computer Engineering at the University of Minnesota, Minneapolis. He serves on the IEEE Signal Processing for Communications and Networking and Machine Learning for Signal Processing Technical Committees. His research interests include optimization theory and applications in signal processing and machine learning. He is a Member of the IEEE. Mingyi Hong was supported, in part, by the National Science Foundation (grant CIF-1910385) and Army Research Office (grant 73202-CS).

Supplementary Material of "Spectrum Cartography via Coupled Block-term Tensor Decomposition"

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

APPENDIX D
PROOF OF PROPOSITION 1

The objective function (12) can be rewritten as:

$$\sum_{k=1}^{K} \|\underline{W}(:,:,k) * (\underline{X}(:,:,k) - A\tilde{C}_k B^\top)\|_F^2, \quad (22)$$

where $\tilde{C}_k = \text{BlockDiag}(c_{k,1}\mathbf{1}_L, \cdots, c_{k,R}\mathbf{1}_L)$ and BlockDiag$(\cdot)$ denotes the block diagonal matrix with the augment matrices on its diagonal blocks. Since $I \geq LR, J \geq LR$, the low-rank matrix factorization $A\tilde{C}_k B^\top$ can be written as $A\tilde{C}_k B^\top = U_k V_k^\top$, $U_k \in \mathbb{R}^{I \times LR}, V_k \in \mathbb{R}^{J \times LR}$. Then, the optimization problem (12) can be rewritten as:

$$\min_{A,B,C} \sum_{k=1}^{K} \|\underline{W}(:,:,k) * (\underline{X}(:,:,k) - U_k V_k^\top)\|_F^2$$
$$\text{subject to } U_k V_k^\top = A\tilde{C}_k B^\top, k = 1, \cdots, K.$$

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

In [64], the authors derived sufficient conditions for rank-r matrix completion, which is stated as follows:

Lemma 1 *[64] 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 $\underline{X}(:,:,k)$ is a rank-LR matrix almost surely if $\{A_r, B_r\}_r$ 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 $\underline{X}(:,:,k)$ completion, one can conclude that $\underline{X}(:,:,k)$ can be be recovered under the conditions specified in Fact 1. Once $\underline{X}$ is recovered, the remaining problem is a rank-$(L, L, 1)$ 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)

The optimality condition of the subproblem w.r.t. A when we use BCD to solve Problem (8) is a system of linear equations. To see this, let us first simplify the notations and

rename different parts in the optimality condition of (14a) as follows:

$$\underbrace{P^\top P}_{H_1} A \underbrace{(R_1 C \odot_p B)^\top (R_1 C \odot_p B)}_{H_2} +$$
$$\underbrace{I_I}_{H_3} A \underbrace{\left[(R_2 C \odot_p QB)^\top (R_2 C \odot_p QB) + \lambda_1 I_{LR}\right]}_{H_4} +$$
$$= \underbrace{P^\top X_1^{(1)\top}(R_1 C \odot_p B) + X_1^{(2)\top}(R_2 C \odot_p QB)}_{H_5}.$$

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 [65].

B. Update B, C for Solving (8)

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

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

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

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

where

$$M_1 = [(B_1 \odot PA_1)\mathbf{1}_L, \cdots, (B_R \odot PA_R)\mathbf{1}_L] \in \mathbb{R}^{I_1 J \times R},$$
$$M_2 = [(QB_1 \odot A_1)\mathbf{1}_L, \cdots, (QB_R \odot A_R)\mathbf{1}_L] \in \mathbb{R}^{I J_1 \times R}.$$

APPENDIX F
UPDATE A, B, C FOR SOLVING (12)

The ith row of A can be updated by:

$$A(i,:) = Y_1(:,i)^\top (C \odot_p B)$$
$$\left[(C \odot_p B)^\top \text{Diag}(W_1(:,i))(C \odot_p B) + \lambda_1 I_{LR}\right]^{-1}.$$

The jth row of B can be updated by:

$$B(j,:) = Y_2(:,j)^\top (C \odot_p A)$$
$$\left[(C \odot_p A)^\top \text{Diag}(W_2(:,j))(C \odot_p A) + \lambda_2 I_{LR}\right]^{-1}.$$

The kth row of C can be updated by:

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

where

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