Power Spectra Separation via Structured Matrix Factorization

Xiao Fu, *Member, IEEE*, Nicholas D. Sidiropoulos, *Fellow, IEEE*, and Wing-Kin Ma, *Senior Member, IEEE*

***Abstract*—Power spectra separation aims at extracting the individual power spectra of multiple emitters from the received mixtures. Traditional spectrum sensing for dynamic spectrum sharing is mostly concerned with detecting or estimating the aggregate spectrum. Spectra separation can be considered as a further step towards full awareness of the radio frequency (RF) environment, which may enable judicious routing, scheduling and beamforming with more effective interference avoidance. In other applications such as geoscience, astronomy, and chemometrics, separating the spectra of the objects/analytes from the sensed mixtures is also of great interest. Our prior work tackled this problem from a tensor decomposition point of view, but this requires delicate and careful receiver setups, and the algorithms are computationally heavy and difficult to decentralize. In this work, we propose to solve the power spectra separation problem using a *structured* matrix factorization model, where the columns of one of the two factor matrices live in the unit simplex. The salient features of this new framework are that 1) the receivers can be far simpler in terms of hardware, 2) an algebraically very simple algorithm can be employed for the centralized case, 3) and that effective decentralized algorithms can be devised under this framework. Numerical simulations and a laboratory experiment using real software-defined radios are presented to demonstrate the effectiveness of the proposed algorithms.**

***Index Terms*—Spectrum estimation, spectra separation, nonnegative matrix factorization, sparse optimization, cognitive radio.**

I. Introduction

SPECTRUM sensing plays an essential role in many applications. In wireless communications, spectrum sensing enables effective interference avoidance and features high-efficiency transmit design. In radio astronomy, spectrum sensing and estimation are employed to observe celestial targets of interest. The techniques of estimating spectrum are also instrumental in geoscience and analytical chemistry, where the spectra of certain objects/analytes are of interest.

There is rich literature on spectrum sensing, particularly in the context of cognitive radio. The approaches range from per-bin detection-based methods [2], to sub-Nyquist wideband sensing [3], cooperative sensing [4], [5], power spectrum-based compressive sensing [6], and distributed power spectrum sensing from coarsely quantized measurements [7]–[10]. Power spectrum sensing (as opposed to Fourier spectrum sensing) makes much sense for wireless communication applications like dynamic spectrum sharing, when the objective is not to demodulate the signal of interest, but rather to estimate spectral occupancy. Another very important application of power spectrum sensing is radio astronomy, where the power spectrum of some celestial target is of interest and is usually sensed by radio telescopes.

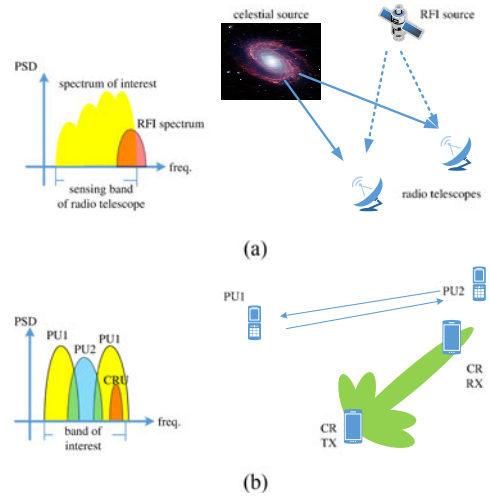


Fig. 1. Motivating examples of spectra separation. (a) Spectra separation of celestial target and radio frequency interference. (b) Cognitive radio user (CRU) transmitting at the band of primary user (PU1) with beamforming towards the direction of PU2 will not interfere either of the PUs.

Despite recent interest in power spectrum sensing, the blind power spectra separation problem had not been considered until our preliminary work in [11], [12]. The power spectra separation problem comes from a practical concern: In realistic scenarios where multiple emitter co-exist in the same band, the received signals at the receivers are mixtures. Apparently, extracting the individual power spectrum of each emitter from the received mixtures is of greater interest than merely sensing the aggregate power spectrum. In fact, knowing the individual emitted spectra enables a number of applications. For example, intelligent beamforming and judicious routing for wireless communication systems will be possible, if the spectral occupancy and the location of each active wireless user are known to the incoming users. Spectra separation is also much relevant to radio astronomy since it can help eliminate radio frequency interferences (RFIs). Some motivating examples are illustrated in Fig. 1. On the other hand, power spectra separation is by nature a very challenging signal processing problem, since the emitters

Manuscript received January 06, 2016; revised April 05, 2016; accepted April 14, 2016. Date of publication April 29, 2016; date of current version July 22, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Pascal Larzabal. This work was supported in part by ARO STIR W911NF-15-1-0384 and in part by the Hong Kong Research Grants Council (RGC) General Research Fund (GRF), under Project CUHK 14205414. This work was presented in part at the IEEE Sensor Array and Multichannel Workshop, A Corunã, Spain, 2014.

X. Fu and N. D. Sidiropoulos are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: xfu@umn.edu; nikos@ece.umn.edu).

W.-K. Ma is with the Department of Electronic Engineering, Chinese University of Hong Kong, Shatin, N.T., Hong Kong (e-mail: wkma@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2016.2560142

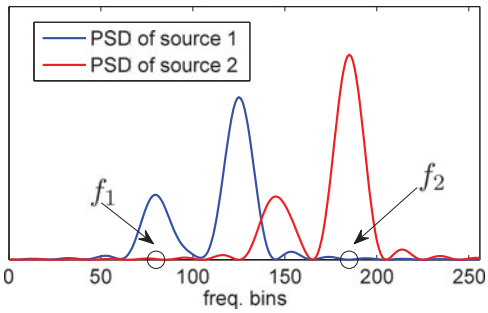


Fig. 2. An illustration with two power spectra that satisfy (A1).

(e.g., wireless transmitters and celestial sources) are in general not cooperating with the receivers. To tackle the spectra separation problem under such a 'blind' setting, theoretical aspects such as identifiability and practical concerns like efficient and decentralized algorithms need to be carefully considered.

Our prior work [11], [12] made the first attempt towards this direction. There, receivers that carry aligned sensor arrays were employed to perform cooperative power spectra separation. By a careful formulation, the power spectra separation problem was posed as a *parallel factor analysis* (PARAFAC) problem [13] in the temporal correlation domain. Identifiability of the power spectra was guaranteed under this framework, and tensor decomposition algorithms can be applied. There are, however, some difficulties left. First, the problem setups in [11], [12] are somewhat restricted: All the receivers are required to have a multiple-antenna array, and these arrays have to be aligned to satisfy some geometrical specifications. Second, while the PARAFAC *formulation* guarantees the identifiability of the power spectra under mild conditions, the PARAFAC *decomposition problem* itself is hard to solve. Third, the formulated PARAFAC problem is difficult to decentralize, while decentralized algorithms are indeed desired in many cases, e.g., when long-distance communication between receivers is prohibitive, no fusion center is available, or only limited data sharing is allowed.

Contributions: Our endeavor in this work is driven by the above concerns. The first contribution of this paper lies in formulating the spectra separation problem as a structured matrix factorization (SMF) problem where the columns of the data matrix lie in a convex hull. The new formulation only involves asynchronous single-antenna receivers, and no geometrical alignment among the receivers is required. Under this formulation, different scenarios and solutions are considered. First, by making use of a local sparsity property in the frequency domain, which is naturally satisfied when the spectra of emitters are centered at different carriers, an algebraically simple algorithm is employed to solve the spectra separation problem in closed form. We also show that the proposed SMF framework can deal with the critical situation where the local sparsity condition is violated; the technique is based on the *volume-minimization* (VolMin) criterion [14]. Second, to deal with the case where the number of emitters exceeds that of the receivers, an alternative formulation is proposed, where the closed-form solution and VolMin can still be applied—the price paid is that receiver synchronization is required.

The second major contribution of this work is decentralized algorithms for power spectra separation. This line of development starts from a sparse optimization-based interpretation to the SMF problem under local sparsity [15]–[18]. The next step is to adopt a convex relaxation approach and decentralize the resulting problem by a careful design of an *alternating direction method of multipliers* (ADMM) algorithm [19]. Also, an iteratively reweighted strategy which can be easily incorporated into the decentralized algorithm is proposed to enhance the performance in practice. Simulations and a laboratory experiment are presented to showcase the effectiveness of the proposed algorithms.

A conference version of this work appeared in [1], which discussed the local sparsity-based centralized algorithms. In this journal version, we additionally include the discussion on cases where local sparsity does not hold, the sparse optimization-based formulation and its identifiability, the decentralized algorithms, extensive simulations, and a real experiment using software-defined radios.

Notation: ^{T}, ^{H} and * denote transpose, Hermitian transpose and conjugate, respectively; † and $^{-1}$ denotes the Moore–Penrose pseudoinverse and matrix inverse, respectively; $\otimes$ and $\odot$ represent the Kronecker product and the Khatri-Rao product, respectively; $\mathrm{rank}(\cdot)$ is the matrix rank; $\boldsymbol{X}_{m,:}([\boldsymbol{X}]_{m,:})$ and $\boldsymbol{X}_{:,n}([\boldsymbol{X}]_{:,n})$ represent the m-th row and the n-th column of matrix $\boldsymbol{X}$, respectively; $\| \boldsymbol{x} \|_p$ and $\| \boldsymbol{X} \|_F$ denote the vector p-(quasi-)norm $(p > 0)$ and the matrix Frobenius norm, respectively; $\| \boldsymbol{X} \|_{\mathrm{row}-0}$ denotes a row-wise zero-norm which counts the number of non-zero rows of $\boldsymbol{X}$;$\| \boldsymbol{X} \|_{q,p}^p = \sum_{i=1}^m \| \boldsymbol{X}_{m,:} \|_q^p$ for $q \geq 1$ and $0 < p \leq 1$ represents the ℓ_q/ℓ_p mixed (quasi-)norm; $\mathrm{vec}(\boldsymbol{X})$ denotes the operator that concatenates the columns of $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$ such that $\mathrm{vec}(\boldsymbol{X}) = [\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_n^T]^T$; $\mathrm{unvec}(\boldsymbol{x})$ denotes the inverse operator of $\mathrm{vec}(\cdot)$; $\mathrm{Diag}(x_1, \ldots, x_n)$ denotes a diagonal matrix with $x_1, \ldots, x_n$ as its diagonal elements; $\boldsymbol{e}_k$ denotes the kth column of the identity matrix; $\boldsymbol{P}_{\boldsymbol{X}}^{\perp}$ denotes projector to the orthogonal complement of the range space of $\boldsymbol{X}$; $(\boldsymbol{x})_+ = \max(\boldsymbol{x}, \boldsymbol{0})$ denotes thresholding operator; $\mathrm{Tr}(\boldsymbol{X})$ denotes the trace of $\boldsymbol{X}$, $\mathrm{conv}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ denotes the convex hull of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$.

II. Matrix Factorization-Based Power Spectra Separation

A. Signal Model

We consider the following received signal model at each sensor:

$$y_n(t) = \sum_{k=1}^{K} a_{n,k} x_k(t) + v_n(t), \quad t = 0, 1, \ldots, \qquad (1)$$

where $x_k(t) \in \mathbb{C}$ denotes the transmitted signal of source k, which is assumed to be wide-sense stationary (WSS), $a_{n,k} \in \mathbb{C}$ is the channel response from source k to sensor n, $v_n(t)$ denotes the noise at sensor n, which is assumed to be i.i.d. zero-mean circularly symmetric Gaussian with variance σ_n^2, and N and K denote the number of sensors and sources, respectively. Assume that the sources are mutually uncorrelated, and that K is known

or has been previously acquired. Our objective is to estimate the power spectra of the sources.

In this work, we focus on the model in (1) and assume no multipath effects. Two specific scenarios where the model in (1) is valid are i) when airborne sensors are employed for reconnaissance, search and rescue, or target detection and spectral signature identification; and ii) when the target sources are celestial objects and the sensors are radio telescopes. More generally, if the source signals are narrowband relative to the carrier frequency (e.g., bandwidth smaller than 20 MHz on a 2.4 GHz carrier), frequency-selective multipath effects are usually negligible. For wideband sources, one can employ long-term correlation averaging, which under certain conditions can mitigate the effects of different frequency-selective multipath from sensor to sensor—see Appendix A of [7] for detailed conditions and derivations. Finally, subband processing can be used to ensure that multipath effects are tolerable within each subband. Each subband can be separately analyzed, and the results can be combined at the end.

B. Proposed Approach

To formulate the problem, we start by computing the auto-correlations of the received signal at each sensor; i.e., sensor n locally computes

$$c_n(\ell) = \mathrm{E}\left\{y_n(t)\, y_n^*(t-\ell)\right\}, \quad n = 1, \ldots, N,$$

where $\ell \in \mathbb{Z}$ denotes the index of the time lag. From the signal model in (1), and under the assumption that the sources are uncorrelated, it can be readily shown that

$$\begin{aligned} c_n(\ell) &= \sum_{k=1}^{K} a_{n,k} a_{n,k}^* \mathrm{E}\left\{x_k(t) x_k^*(t-\ell)\right\} \\ &\quad + \mathrm{E}\left\{v_n(t)\, v_n^*(t-\ell)\right\} \\ &= \sum_{k=1}^{K} |a_{n,k}|^2 r_k(\ell) + \sigma_n^2 \delta(\ell), \end{aligned} \tag{2}$$

where $r_k(\ell) = \mathrm{E}\{x_k(t) x_k^*(t-\ell)\}$ represents the temporal auto-correlation of source k at time lag ℓ, $\mathrm{E}\{v_n(t) v_n^*(t-\ell)\} = \sigma_n^2 \delta(\ell)$ is obtained by assuming that the noise is i.i.d. Gaussian distributed, and $\delta(\ell)$ denotes the Kronecker delta function. Taking the Discrete Fourier Transform (DFT) of $c_n(\ell)$ for $n = 1, \ldots, N$, we obtain

$$C_n(\omega) = \sum_{\ell=-\infty}^{\infty} c_n(\ell)\, e^{-j\omega\ell} = \sum_{k=1}^{K} |a_{n,k}|^2 S_k(\omega) + \sigma_n^2,$$

where $S_k(\omega)$ denotes the power spectral density (PSD) of source k at frequency ω. We discretize $C_n(\omega)$ by letting $G_n(f) = C_n(\frac{2\pi(f-1)}{F}) = \sum_{k=1}^{K} |a_{n,k}|^2 S_k(\frac{2\pi(f-1)}{F}) + \sigma_n^2$ for $f = 1, \ldots, F$, and construct a matrix $\boldsymbol{G} \in \mathbb{R}^{N \times F}$ such that $\boldsymbol{G}_{n,f} = G_n(f)$. In matrix form, we have

$$\boldsymbol{G} = \boldsymbol{B}\boldsymbol{S} + \boldsymbol{\eta}\mathbf{1}^T, \tag{3}$$

where $\boldsymbol{B} \in \mathbb{R}^{N \times K}$ is such that $\boldsymbol{B}_{n,k} = |a_{n,k}|^2$, $\boldsymbol{S} \in \mathbb{R}^{K \times F}$, $\boldsymbol{S}_{k,:} = [S_k(0), \ldots, S_k(\frac{2\pi(F-1)}{F})]$ is the discretized power spectrum of source k, $\boldsymbol{\eta} = [\sigma_1^2, \ldots, \sigma_N^2]^T$, and $\mathbf{1}$ denotes an all-one vector with proper length.

There are many ways to estimate σ_n^2 for $n = 1, \ldots, N$ [12], [13]. After obtaining estimates of the noise variances, i.e., $\widehat{\boldsymbol{\eta}} = [\widehat{\sigma_1}^2, \ldots, \widehat{\sigma_N}^2]^T$, we can remove $\boldsymbol{\eta}$ from $\boldsymbol{G}$ in (3) via subtraction. Let $\widetilde{\boldsymbol{G}} := \boldsymbol{G} - \widehat{\boldsymbol{\eta}}\mathbf{1}^T$. Then, the following noise-free model can be (approximately) obtained,

$$\widetilde{\boldsymbol{G}} = \boldsymbol{B}\boldsymbol{S}. \tag{4}$$

Since both $\boldsymbol{B}$ and $\boldsymbol{S}$ are nonnegative, estimating $\boldsymbol{S}$ from $\widetilde{\boldsymbol{G}}$ can be treated as a *nonnegative matrix factorization* (NMF) problem, which recently found diverse applications in machine learning and signal processing. Hence, one possible way of estimating $\boldsymbol{S}$ is to apply some existing algorithms that tackle the following optimization criterion for NMF:

$$\boldsymbol{S} = \arg \min_{\boldsymbol{B} \geq \mathbf{0}, \boldsymbol{S} \geq \mathbf{0}} \left\| \widetilde{\boldsymbol{G}} - \boldsymbol{B}\boldsymbol{S} \right\|_F^2. \tag{5}$$

However, as mentioned in [12], [20], the difficulty is that NMF does not necessarily lead to a unique solution, which means that the identifiability of $\boldsymbol{S}$ is not guaranteed; plus, the optimization problem in (5) is hard to solve. To ensure uniqueness, it is sometimes more convenient to explore some other structures of one or both of $\boldsymbol{B}$ and $\boldsymbol{S}$ [14], [21], [22], rather than relying on nonnegativity alone. To explain, we first assume that the matrix $\boldsymbol{B}$ has full column rank, i.e., $\mathrm{rank}(\boldsymbol{B}) = K$. We also consider the following condition:

(A1) For each $k \in \{1, \ldots, K\}$, there exists a frequency index $f_k \in \{1, 2, \ldots, F\}$ such that $S_k(\frac{2\pi(f_k-1)}{F}) > 0$ and $S_{j \neq k}(\frac{2\pi(f_k-1)}{F}) = 0$.

(A1) means that each source dominates at least one particular frequency, which is common in wireless communication—different carrier frequencies are allocated to different transmitters to avoid mutual interference. Notice that (A1) is identical to the so-called *separability*, *local dominance*, and *pure-pixel assumption* in NMF [22], blind source separation (nBSS)[1] [21] and hyperspectral unmixing [15], respectively. In this work, we refer to the matrix factorization task under (A1) as *structured matrix factorization* (SMF), to distinguish from general NMF. Given that (A1) holds, the task of matrix factorization amounts to finding the index set $\Lambda = \{f_1, \ldots, f_K\}$, since the factor $\boldsymbol{B}$ (with column permutation and scaling ambiguities) can be simply 'read out' from the corresponding columns of $\widetilde{\boldsymbol{G}}$, i.e., $\widehat{\boldsymbol{B}} = \widetilde{\boldsymbol{G}}_{:,\Lambda}$; and $\boldsymbol{S}$ can then be estimated up to row permutation and scaling ambiguities by solving

$$\widehat{\boldsymbol{S}} = \arg \min_{\boldsymbol{S} \geq \mathbf{0}} \| \widetilde{\boldsymbol{G}} - \widehat{\boldsymbol{B}}\boldsymbol{S} \|_F^2. \tag{6}$$

We should mention that, to make the formulated factorization model work, we do not necessarily need (A1). Approaches that can deal with the situation where (A1) is violated will also be briefly introduced later—see Remarks 2 and 4.

Many algorithms have been proposed to identify Λ; see, e.g., [21], for a review. In this paper, we propose to employ the

[1]The subtle difference between nBSS and NMF is that nBSS only requires $\boldsymbol{S}$ to be nonnegative, while imposing no nonnegativity constraint on $\boldsymbol{B}$.

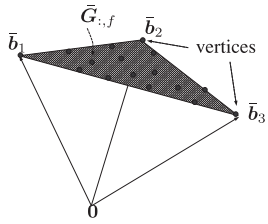


Fig. 3. Geometry of $\bar{\boldsymbol{G}} = \bar{\boldsymbol{B}}\bar{\boldsymbol{S}}$. The shaded region is $\mathrm{conv}\{\bar{\boldsymbol{b}}_1, \bar{\boldsymbol{b}}_2, \bar{\boldsymbol{b}}_3\}$.

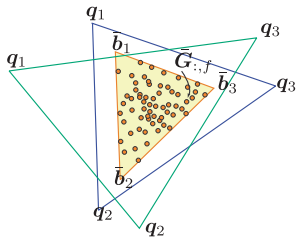


Fig. 4. Intuition of VolMin. The illustration is on the hyperplane where $\mathrm{conv}\{\bar{\boldsymbol{b}}_1, \bar{\boldsymbol{b}}_2, \bar{\boldsymbol{b}}_3\}$ lies in.

so-called *successive projection algorithm* (SPA) [22], [23], which consists of closed-form solutions for identifying $f_1, \ldots, f_K$. To describe the process, let us first normalize the columns of $\widetilde{\boldsymbol{G}}$; i.e., we let $\bar{\boldsymbol{G}}_{:,f} = \widetilde{\boldsymbol{G}}_{:,f} / \| \widetilde{\boldsymbol{G}}_{:,f} \|_1$ for $f \in \mathcal{F}$, where $\mathcal{F} = \{f | \ \| \widetilde{\boldsymbol{G}}_{:,f} \|_1 > 0, f \in \{1, \ldots, F\}\}$, and let $\bar{\boldsymbol{G}}_{:,f} = \widetilde{\boldsymbol{G}}_{:,f}$ otherwise. By this normalization, we see

$$\bar{\boldsymbol{G}}_{:,f} = \sum_{k=1}^{K} \bar{\boldsymbol{B}}_{:,k} \bar{\boldsymbol{S}}_{k,f}, \tag{7}$$

where $\bar{\boldsymbol{B}}_{:,k} = \boldsymbol{B}_{:,k} / \| \boldsymbol{B}_{:,k} \|_1$ and $\bar{\boldsymbol{S}}_{k,f} = \boldsymbol{S}_{k,f} \| \boldsymbol{B}_{:,k} \|_1 / \| \sum_{k=1}^{K} \boldsymbol{S}_{k,f} \boldsymbol{B}_{:,k} \|_1$. Then, we have a signal model in the following compact form:

$$\bar{\boldsymbol{G}} = \bar{\boldsymbol{B}}\bar{\boldsymbol{S}}, \tag{8}$$

in which, by invoking the nonnegativity of $\bar{\boldsymbol{B}}$, one can easily show that [22]

$$\mathbf{1}^T \bar{\boldsymbol{S}} = \mathbf{1}, \bar{\boldsymbol{S}} \geq \mathbf{0}. \tag{9}$$

Under (9) and $\mathrm{rank}(\boldsymbol{B}) = K$, all the columns of $\bar{\boldsymbol{G}}_{:,f}$ live in the convex hull $\mathrm{conv}\{\bar{\boldsymbol{b}}_1, \ldots, \bar{\boldsymbol{b}}_K\}$, and the columns corresponding to $f_1, \ldots, f_K$ in $(\mathsf{A1})$ are the vertices of this convex hull—the problem can also be viewed as a vertex-index picking problem; see Fig. 3 for an illustration. SPA picks out $f_1, \ldots, f_K$ using the following simple steps:

$$\hat{f}_k = \arg \max_{f \in \{1, \ldots, F\}} \left\| \boldsymbol{P}^{\perp}_{\widehat{\boldsymbol{B}}_{1:k-1}} \bar{\boldsymbol{G}}_{:,f} \right\|_2^2, \qquad k = 1, \ldots, K, \tag{10}$$

where $\widehat{\boldsymbol{B}}_{1:k-1} = [\bar{\boldsymbol{G}}_{:,\hat{f}_1}, \ldots, \bar{\boldsymbol{G}}_{:,\hat{f}_{k-1}}]$ and $\widehat{\boldsymbol{B}}_{1:0} = \boldsymbol{I}$; the readers are referred to [21], [22] for the derivation.

Remark 1: The proposed SPA-based solution to power spectra separation has several attractive features. First, the formulation in (4) does not require synchronization between the sensors, which spares much effort and saves considerable communication overhead. The reason is that $c_n(\ell)$(cf. (2)) is autocorrelation of a stationary signal which is only related to the *time lag*, i.e., ℓ, rather than any specific time shift. Second, SPA itself is very efficient. This desired feature makes SPA particularly suitable to be used in real-time applications such as wireless communications. Also, it has been proved that SPA is robust to modeling errors and noise [22], which means that even when $(\mathsf{A1})$ is violated to some extent, e.g., when source k does not exactly occupy f_k by itself but merely dominates the others at f_k, the proposed approach can still work.

Remark 2: In more challenging cases where $(\mathsf{A1})$ is grossly violated, it is still possible to recover $\bar{\boldsymbol{B}}$ from $\bar{\boldsymbol{G}}$(and thus $\boldsymbol{S}$) by leveraging the volume-minimization (VolMin)-based matrix factorization criterion. VolMin finds $\bar{\boldsymbol{B}}$ via finding the minimum-volume convex hull that encloses all the columns of $\bar{\boldsymbol{G}}$. The VolMin criterion is as follows:

$$\begin{aligned} (\bar{\boldsymbol{B}}, \bar{\boldsymbol{S}}) = \ & \arg \min_{\boldsymbol{Q} \in \mathbb{R}^{M \times K}, \boldsymbol{\Theta} \in \mathbb{R}^{K \times F}} \det \left(\boldsymbol{Q}^T \boldsymbol{Q} \right) \\ \text{s.t. } \bar{\boldsymbol{G}} = \ & \boldsymbol{Q}\boldsymbol{\Theta}, \qquad \mathbf{1}^T \boldsymbol{\Theta} = \mathbf{1}^T, \boldsymbol{\Theta} \geq \mathbf{0}, \end{aligned} \tag{11}$$

where $\boldsymbol{Q} = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_K]$, the cost function measures the volume of the sought convex hull, and the constraints mean that the columns of $\bar{\boldsymbol{G}}$ are enclosed by the convex hull $\mathrm{conv}\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_K\}$. The intuition of VolMin is illustrated in Fig. 4 using a $K = 3$ case, where we see that the enclosing convex hull with minimum volume coincides with $\mathrm{conv}\{\bar{\boldsymbol{b}}_1, \bar{\boldsymbol{b}}_2, \bar{\boldsymbol{b}}_3\}$. Recently, we have shown that (11) is a sound criterion that guarantees identifiability of $\bar{\boldsymbol{B}}$ without $(\mathsf{A1})$, provided that the $\bar{\boldsymbol{G}}_{:,f}$'s are sufficiently spread in $\mathrm{conv}\{\bar{\boldsymbol{b}}_1, \ldots, \bar{\boldsymbol{b}}_K\}$[14]. Nevertheless, VolMin algorithms (e.g., those in [24], [25]) in general do not have closed-form solution as SPA does, so there is a trade-off between robustness to modeling assumptions and implementation simplicity.

Remark 3: In many applications, estimating the locations of the sources is also of interest. This is seriously compounded by superposition, so spectra separation is instrumental for localization as well. Specifically, under $(\mathsf{A1})$, if f_k has been identified, the sensors can localize source k within a narrow bandwith centered at frequency f_k, where interference from other sources is weak, using various received signal strength (RSS)-based algorithms, e.g., [26] or [27]–[29].

C. The Under-Determined Case

Applying SPA requires that $\mathrm{rank}(\boldsymbol{B}) = K$, which implicitly assumes that the number of sensors is larger than or equal to that of sources. In practice, however, there are cases where such an assumption does not hold. Under such circumstances, there exist other algorithms that can potentially identify Λ, e.g., the successive nonnegative projection algorithm (SNPA) [30] which is an advanced version of SPA with more complicated updates. Nevertheless, we can still apply the simple plain SPA to resolve the power spectra of the sources, if the sensors are *synchronized*. To explain, we construct $\boldsymbol{y}(t) = [y_1(t), \ldots, y_N(t)]^T$ by collecting the received signals of the sensors at time t. Let

$$\boldsymbol{D}(\ell) = \mathrm{E}\left\{ \boldsymbol{y}(t) \boldsymbol{y}^H (t - \ell) \right\}, \tag{12}$$

denote the temporal cross-correlation matrix of the received signals at time lag ℓ. Using $\boldsymbol{y}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{v}(t)$, where $\boldsymbol{x}(t) = [x_1(t), \ldots, x_K(t)]^T$ and $\boldsymbol{v}(t) = [v_1(t), \ldots, v_N(t)]^T$, and the assumption that the noise is spatially and temporally white, we have

$$\boldsymbol{D}(\ell) = \boldsymbol{A}\boldsymbol{R}(\ell)\boldsymbol{A}^H + \Xi\delta(\ell), \tag{13}$$

where $\delta(\ell)$ denotes the Kronecker delta, $\boldsymbol{R}(\ell) = \mathrm{Diag}(\boldsymbol{r}(\ell))$, $\boldsymbol{r}(\ell) = [r_1(\ell), \ldots, r_K(\ell)]^T$, $\Xi = \mathrm{Diag}(\boldsymbol{\eta})$, and $\boldsymbol{\eta}$ is defined as before.

The signal model in (13) has been extensively used in the context of blind source separation (BSS) [31] *for estimating the mixing system* $\boldsymbol{A}$ via the joint diagonalization technique, with the ultimate goal of separating the sources, e.g., by (pseudo-)inverting $\boldsymbol{A}$ to estimate $\boldsymbol{x}(t)$ as $\widehat{\boldsymbol{x}}(t) = \boldsymbol{A}^{\dagger}\boldsymbol{y}(t)$. Our interest here lies in sensing the spectra of the sources instead. To this end, we define

$$\boldsymbol{z}(\ell) = \mathrm{vec}(\boldsymbol{D}(\ell)) = (\boldsymbol{A}^* \odot \boldsymbol{A})\boldsymbol{r}(\ell) + \delta(\ell)\boldsymbol{g},$$

where $\boldsymbol{g} = (\boldsymbol{I} \odot \boldsymbol{I})\boldsymbol{\eta}$, and $\boldsymbol{A}^* \odot \boldsymbol{A} = [\boldsymbol{a}_1^* \otimes \boldsymbol{a}_1, \ldots, \boldsymbol{a}_K^* \otimes \boldsymbol{a}_K]$. It can be seen that the Fourier Transform of $z_i(\ell)$, i.e., the ith element of $\boldsymbol{z}(\ell)$, is

$$Z_i(\omega) = \sum_{\ell=-\infty}^{\infty} z_i(\ell)\, e^{-jw\ell} = \sum_{k=1}^{K} [\boldsymbol{A}^* \odot \boldsymbol{A}]_{i,k} S_k(\omega) + g_i,$$

where $i = 1, \ldots, N^2$, and $S_k(\omega) = \sum_{\ell=-\infty}^{\infty} r_k(\ell)e^{-jw\ell}$ is the power spectrum of source k at frequency ω. Hence, by letting $W_i(f) = Z_i(\frac{2\pi(f-1)}{F})$ for $f = 1, \ldots, F$ and $\boldsymbol{W}_{i,f} = W_i(f)$, we come up with

$$\boldsymbol{W} = (\boldsymbol{A}^* \odot \boldsymbol{A})\boldsymbol{S} + \boldsymbol{g}\boldsymbol{1}^T. \tag{14}$$

The noise term $\boldsymbol{g}$ can be estimated and canceled. Consequently, we obtain the following signal model,

$$\widetilde{\boldsymbol{W}} = (\boldsymbol{A}^* \odot \boldsymbol{A})\boldsymbol{S}. \tag{15}$$

We still wish to apply SPA to the signal model in (15) to resolve the power spectra of the sources. The difficulty is that $\boldsymbol{A}^* \odot \boldsymbol{A}$ is not nonnegative and the normalization step in the last section is no longer applicable. Here, a slightly different normalization can be employed, following the insight of the method in [14]. To see the procedure, let us denote

$$\mathrm{unvec}\left(\widetilde{\boldsymbol{W}}_{:,f}\right) = \boldsymbol{A}\mathrm{Diag}(\boldsymbol{S}_{:,f})\boldsymbol{A}^H.$$

Then, for f such that $\mathrm{Tr}(\mathrm{unvec}(\widetilde{\boldsymbol{W}}_{:,f})) > 0$, we construct

$$\bar{\boldsymbol{W}}_{:,f} = \frac{\widetilde{\boldsymbol{W}}_{:,f}}{\mathrm{Tr}\left(\mathrm{unvec}\left(\widetilde{\boldsymbol{W}}_{:,f}\right)\right)} = \boldsymbol{H}\check{\boldsymbol{S}}_{:,f}, \tag{16}$$

where $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]$, $\boldsymbol{h}_k = \frac{\boldsymbol{a}_k^* \otimes \boldsymbol{a}_k}{\|\boldsymbol{a}_k\|_2^2}$, and $\check{\boldsymbol{S}}_{k,f} = \frac{\|\boldsymbol{a}_k\|_2^2 \boldsymbol{S}_{k,f}}{\sum_{k=1}^{K} \boldsymbol{S}_{k,f}\|\boldsymbol{a}_k\|_2^2}$. We see that, by this construction, $\boldsymbol{1}^T\check{\boldsymbol{S}} = \boldsymbol{1}^T$ and $\check{\boldsymbol{S}} \geq \boldsymbol{0}$ are satisfied. To apply SPA, we now only require that $\mathrm{rank}(\boldsymbol{H}) = K$, which can be satisfied under mild conditions. For example, if the entries of $\boldsymbol{A}$ are drawn from some continuous distribution, $\mathrm{rank}(\boldsymbol{H}) = K$ is fulfilled almost surely if $K \leq N^2$[32]—this implies the ability of resolving K power spectra using only $\mathcal{O}(\sqrt{K})$ sensors.

Algorithm 1: Centralized Power Spectra Separation.

input : $x_n(t)$; K.
if $K \leq N$ **then**
  construct $\bar{\boldsymbol{G}}$ following (2)–(8);
  let $\bar{\boldsymbol{Y}} = \bar{\boldsymbol{G}}$ and $\boldsymbol{Y} = \hat{\boldsymbol{G}}$;
else
  synchronize the sensors;
  construct $\boldsymbol{W}$ following (12)–(16);
  let $\bar{\boldsymbol{Y}} = \bar{\boldsymbol{W}}$ and $\boldsymbol{Y} = \tilde{\boldsymbol{W}}$;
end
let $\hat{\boldsymbol{Y}}_{1:0} = \boldsymbol{I}$;
for $k = 1 : K$ **do**
  $\hat{f}_k = \arg\max_{f \in \mathcal{F}} \left\| \boldsymbol{P}^{\perp}_{\hat{\boldsymbol{Y}}_{1:k-1}} \bar{\boldsymbol{Y}}_{:,f} \right\|_2^2$;
  $\hat{\boldsymbol{Y}}_{1:k} = [\bar{\boldsymbol{Y}}_{:,\hat{f}_1}, \ldots, \bar{\boldsymbol{Y}}_{:,\hat{f}_k}]$;
  $\hat{\Lambda} = [\hat{f}_1, \ldots, \hat{f}_k]$;
end
solve $\hat{\boldsymbol{S}} = \arg\min_{\boldsymbol{S} \geq \boldsymbol{0}} \|\boldsymbol{Y} - \boldsymbol{Y}_{:,\Lambda}\boldsymbol{S}\|_F^2$;
output: $\hat{\boldsymbol{S}}$; $\hat{\Lambda}$.

Remark 4: When (A1) is violated, the VolMin approach discussed in Remark 2 can be employed in place of SPA to (16) to recover $\boldsymbol{S}$. Moreover, using the matrices $\{\mathrm{unvec}(\widetilde{\boldsymbol{W}}_{:,f})\}_{f=1}^{F}$, a low-rank tensor factorization-based formulation can also be considered:

$$\min_{\boldsymbol{A},\boldsymbol{S}} \sum_{f=1}^{F} \left\| \mathrm{unvec}\left(\widetilde{\boldsymbol{W}}_{:,f}\right) - \boldsymbol{A}\mathrm{Diag}(\boldsymbol{S}_{:,f})\boldsymbol{A}^H \right\|_F^2. \tag{17}$$

Problem (17) is known as the least squares optimization criterion for *parallel factor analysis* (PARAFAC) with an additional symmetry constraint. Under the framework of PARAFAC, the factors (i.e., $\boldsymbol{A}$ and $\boldsymbol{S}$) are also identifiable when $N < K$[13]. Another merit of using (17) is that the identifiability of $\boldsymbol{S}$ holds under even milder conditions compared to that ensure the identifiability of VolMin. However, the trade-off is that solving Problem (17) is difficult, which is even more cumbersome than dealing with VolMin in practice.

We summarize the SPA algorithm on the signal models in both (8) and (16) in Algorithm 1. It should be noted that for the under-determined case, an extra synchronization stage is required for applying SPA, which is a price to pay for dealing with this challenging scenario. Also notice that in this section, we implicitly assumed that there is a fusion center that collects and processes the data received by the sensors. Hence, we call Algorithm 1 the centralized algorithm. In the next section, decentralized algorithms will be proposed.

III. Decentralized Power Spectra Separation

In many cases, decentralized algorithms are desired. For example, sometimes there is no fusion center. Moreover, long-distance communication is sometimes expensive or prohibitive and thus the sensors can only exchange information with their neighbors. Decentralized algorithms are also needed for enhancing network robustness against sensor failure, and for security

considerations such as data sharing restrictions. In this section, we propose decentralized power spectra separation algorithms that only require each sensor to communicate with its one-hop neighbors.

A. A Sparse Optimization Criterion

Our idea is to utilize the signal model in (8) again. Notice that $\widetilde{\boldsymbol{G}}_{n,:}$ is received and stored at sensor n, and $\| \widetilde{\boldsymbol{G}}_{:,f} \|_1 = \sum_{n=1}^{N} \widetilde{\boldsymbol{G}}_{n,f}$ can be easily obtained at each sensor by some existing distributed aggregation algorithms, e.g., the gossip algorithm [33] and the first order algorithm for unconstrained consensus [34]. Hence, the normalized data $\bar{\boldsymbol{G}}_{n,:}$ can be obtained at each sensor without any centralized process. Beginning from this point, we consider a recently proposed formulation in [15], [16], [35] for estimating Λ:

$$\begin{aligned} \min_{\boldsymbol{C}\in\mathbb{R}^{F\times F}} & \| \boldsymbol{C} \|_{\mathrm{row}-0} \\ \text{s.t. } & \bar{\boldsymbol{G}} = \bar{\boldsymbol{G}}\boldsymbol{C} \\ & \mathbf{1}^T\boldsymbol{C} = \mathbf{1}^T, \boldsymbol{C} \geq \mathbf{0}. \end{aligned} \tag{18}$$

The intuition behind the above formulation is to use a row-sparse $\boldsymbol{C}$ to select some columns of $\bar{\boldsymbol{G}}$ as a basis which should enclose all the data points within its convex hull. To see the reason why Problem (18) helps identify Λ, let us consider an illustrative case where $\Lambda = \{1, 2, \ldots, K\}$. Given that $(\mathsf{A1})$ holds and that $\mathrm{rank}(\boldsymbol{B}) = K$, Problem (18) has the optimal solution $\boldsymbol{C}^\star = [\bar{\boldsymbol{S}}^T, \mathbf{0}^T]^T$ [15]. Notice that the constraints in Problem (18) are all satisfied by $\boldsymbol{C}^\star$ and it is impossible to find another feasible $\boldsymbol{C}$ with fewer nonzero rows, which validate the optimality of $\boldsymbol{C}^\star$. Such a structure of the optimal solution leads to identification of Λ, which can be found by simply inspecting the non-zero rows of $\boldsymbol{C}^\star$. In more general cases where $\Lambda = \{f_1, \ldots, f_K\}$, we have [15]

$$\boldsymbol{C}^\star = \boldsymbol{\Pi}^T \begin{bmatrix} \bar{\boldsymbol{S}} \\ \mathbf{0} \end{bmatrix}, \tag{19}$$

where $\boldsymbol{\Pi}$ is a permutation matrix such that $\bar{\boldsymbol{S}}\boldsymbol{\Pi} = [\boldsymbol{I}, \bar{\boldsymbol{S}}']$, and Λ can also be known by inspecting the nonzero rows of $\boldsymbol{C}^\star$.

To see how we decentralize the problem, we consider the mixed-norm relaxation (see [16]), where $\| \boldsymbol{C} \|_{\mathrm{row}-0}$ is replaced by $\| \boldsymbol{C} \|_{q,1}$ for $q > 1$:

$$\begin{aligned} \min_{\boldsymbol{C}\in\mathbb{R}^{F\times F}} & \| \boldsymbol{C} \|_{q,1} \\ \text{s.t. } & \bar{\boldsymbol{G}} = \bar{\boldsymbol{G}}\boldsymbol{C} \\ & \mathbf{1}^T\boldsymbol{C} = \mathbf{1}^T, \boldsymbol{C} \geq \mathbf{0}. \end{aligned} \tag{20}$$

Curiously, although such a relaxation technique is widely adopted in many applications, particularly in remote sensing [36], [37], there has been no formal proof regarding its identifiability of Λ, to our best knowledge.[2] Here, as a side contribution, we fill this gap by providing the following theorem:

Theorem 1: Let $\boldsymbol{C}_{\mathrm{opt}}$ denote the optimal solution to Problem (20). Assume that $\bar{\boldsymbol{G}}$ has no repeated columns and $\mathrm{rank}(\boldsymbol{B}) = K$. Then, under the model in (8) and $(\mathsf{A1})$, $\boldsymbol{C}_{\mathrm{opt}} = \boldsymbol{C}^\star$ holds for $q > 1$, where $\boldsymbol{C}^\star$ is as defined in (19).

The proof of Theorem 1 can be found in Appendix A, which verifies the soundness of employing Problem (20) as an optimization surrogate.

B. ADMM-Based Decentralization

In practice, to deal with noise and modeling error, and to derive a decentralized algorithm, we are interested in dealing with the following LASSO-type variant of Problem (20) [16]:

$$\begin{aligned} \min_{\boldsymbol{C}\in\mathbb{R}^{F\times F}} & \lambda \| \boldsymbol{C} \|_{q,1} + \frac{1}{2}\big\|\bar{\boldsymbol{G}} - \bar{\boldsymbol{G}}\boldsymbol{C}\big\|_F^2 \\ \text{s.t. } & \mathbf{1}^T\boldsymbol{C} = \mathbf{1}^T, \boldsymbol{C} \geq \mathbf{0}. \end{aligned} \tag{21}$$

By relaxing $\bar{\boldsymbol{G}} = \bar{\boldsymbol{G}}\boldsymbol{C}$, the above problem is considered more robust to noise and modeling errors; also see some theoretical evidence in [18] for the $q = \infty$ case. We wish that every sensor finds a local solution of $\boldsymbol{C}$ and achieves 'local consensus' with its one-hop neighbors, so that the overall consensus can be eventually achieved. To this end, let us first denote $\mathcal{N}_n$ as the index set of the one-hop neighbors of sensor n. Then, taking into consideration that $\bar{\boldsymbol{G}}_{n,:}$ is the (normalized) data at sensor n, we come up with the following formulation:

$$\begin{aligned} \min_{\{\boldsymbol{\Gamma}_r^n\},\{\boldsymbol{C}^n\}} & \frac{1}{2}\sum_{n=1}^{N}\big\|\bar{\boldsymbol{G}}_{n,:} - \bar{\boldsymbol{G}}_{n,:}\boldsymbol{C}^n\big\|_2^2 + \sum_{n=1}^{N}\frac{\lambda}{N|\mathcal{N}_n|}\| \boldsymbol{\Gamma}_r^n \|_{q,1} \\ \text{s.t. } & \mathbf{1}^T\boldsymbol{C}^n = \mathbf{1}^T, \\ & \boldsymbol{C}^n = \boldsymbol{\Gamma}_r^n, \boldsymbol{C}^r = \boldsymbol{\Gamma}_r^n, r \in \mathcal{N}_n \\ & \boldsymbol{\Gamma}_r^n \geq \mathbf{0}. \end{aligned} \tag{22}$$

Instead of optimizing with respect to a global variable $\boldsymbol{C}$, Problem (22) aims at finding $\boldsymbol{C}^n$ at sensor n, and a local consensus variable $\boldsymbol{\Gamma}_r^n$ with its neighbors (sensor r for $r \in \mathcal{N}_n$). As pointed out in [19], [38], if there always exists a single or multi-hop path that links any two sensors, such a reformulated local consensus problem in (22) and the original problem in (21) are equivalent. To solve Problem (22), we propose the following *alternating direction method of multipliers* (ADMM)-based algorithm [19], whose updates can be seen in

$$\begin{aligned} \boldsymbol{\Gamma}_r^n := \ & \arg\min_{\boldsymbol{\Gamma}_r^n \geq \mathbf{0}} \frac{\lambda}{N|\mathcal{N}_n|}\| \boldsymbol{\Gamma}_r^n \|_{q,1} \\ & + \frac{\rho}{2}\|\boldsymbol{C}^n - \boldsymbol{\Gamma}_r^n + \boldsymbol{U}_r^n\|_F^2 \\ & + \frac{\rho}{2}\|\boldsymbol{C}^r - \boldsymbol{\Gamma}_r^n + \boldsymbol{V}_n^r\|_F^2 \end{aligned} \tag{23a}$$

$$\begin{aligned} \boldsymbol{C}^n := \ & \arg\min_{\mathbf{1}^T\boldsymbol{C}^n = \mathbf{1}^T} \frac{1}{2}\big\|\bar{\boldsymbol{G}}_{n,:} - \bar{\boldsymbol{G}}_{n,:}\boldsymbol{C}^n\big\|_2^2 \\ & + \frac{\rho}{2}\sum_{r\in\mathcal{N}_n}\|\boldsymbol{C}^n - \boldsymbol{\Gamma}_r^n + \boldsymbol{U}_r^n\|_F^2 \\ & + \frac{\rho}{2}\sum_{r\in\mathcal{N}_n}\|\boldsymbol{C}^n - \boldsymbol{\Gamma}_n^r + \boldsymbol{V}_r^n\|_F^2 \end{aligned} \tag{23b}$$

$$\boldsymbol{U}_r^n := \boldsymbol{C}^n - \boldsymbol{\Gamma}_r^n + \boldsymbol{U}_r^n, \tag{23c}$$

$$\boldsymbol{V}_n^r := \boldsymbol{C}^r - \boldsymbol{\Gamma}_r^n + \boldsymbol{V}_n^r. \tag{23d}$$

[2] In [16], a statement was made regarding the identifiability, but the proof was not provided.

There, $\boldsymbol{U}_r^n$ is the dual variable associated with the equality constraint $\boldsymbol{C}^n = \boldsymbol{\Gamma}_r^n$, and $\boldsymbol{V}_n^r$ with $\boldsymbol{C}^r = \boldsymbol{\Gamma}_r^n$. In the process described in (23), $\boldsymbol{\Gamma}_r^n$, $\boldsymbol{C}^n$ and $\boldsymbol{U}_r^n$ are calculated at sensor n, and $\boldsymbol{V}_n^r$ is calculated at neighbors of sensor n. Each sensor needs to gather information from its neighbors for computing (23a) and (23b). Then, the calculated $\boldsymbol{\Gamma}_r^n$ is passed to its neighbors to compute (23d). Therefore, the algorithm can be carried out in a sensor network where only one-hop communication is available.

Notice that the subproblems (23a) and (23b) admit simple solutions. Subproblem (23b) is an equality-constrained quadratic program, and the optimal $\boldsymbol{C}^n$ can be obtained by solving the following optimality conditions of Problem (23b):

$$\begin{bmatrix} 2|\mathcal{N}_n|\rho\boldsymbol{I} + \left(\bar{\boldsymbol{G}}_{n,:}\right)^T\bar{\boldsymbol{G}}_{n,:}, & \boldsymbol{1} \\ \boldsymbol{1}^T, & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{C}^n \\ \boldsymbol{\nu} \end{bmatrix} = \begin{bmatrix} \left(\bar{\boldsymbol{G}}_{n,:}\right)^T\bar{\boldsymbol{G}}_{n,:} + \rho\left(\boldsymbol{\Gamma}_r^n - \boldsymbol{U}_r^n + \boldsymbol{\Gamma}_n^r - \boldsymbol{V}_r^n\right) \\ \boldsymbol{1}^T \end{bmatrix}. \quad (24)$$

Problem (23a) is a constrained proximal operator. To solve it, we only need to consider the problems w.r.t. each row of $\boldsymbol{\Gamma}_r^n$, i.e., for $f \in \{1, \ldots, F\}$,

$$\min_{\boldsymbol{\Gamma}_{f,:}^n \geq \boldsymbol{0}} \frac{\rho}{2}\left\| \frac{\boldsymbol{C}_{f,:}^n + [\boldsymbol{U}_r^n]_{f,:} + \boldsymbol{C}_{f,:}^r + [\boldsymbol{V}_n^r]_{f,:}}{2} - [\boldsymbol{\Gamma}_r^n]_{f,:}\right\|_2^2 + \frac{\lambda}{2N|\mathcal{N}_n|}\left\|[\boldsymbol{\Gamma}_r^n]_{f,:}\right\|_q. \quad (25)$$

Problem (25) can be converted into the following form [39]:

$$\min_{[\boldsymbol{\Gamma}_r^n]_{f,:}} \frac{1}{2}\left\| \left(\frac{\boldsymbol{C}_{f,:}^n + [\boldsymbol{U}_r^n]_{f,:} + \boldsymbol{C}_{f,:}^r + [\boldsymbol{V}_n^r]_{f,:}}{2} \right)_+ - [\boldsymbol{\Gamma}_r^n]_{f,:}\right\|_2^2 + \frac{\lambda}{2\rho N|\mathcal{N}_n|}\left\|[\boldsymbol{\Gamma}_r^n]_{f,:}\right\|_q. \quad (26)$$

For the $q = 2$ case, we have

$$[\boldsymbol{\Gamma}_r^n]_{f,:} := \begin{cases} \boldsymbol{0}, & \left\|\boldsymbol{\beta}_f^n\right\|_2 < \frac{\lambda}{2N|\mathcal{N}_n|\rho} \\ \left(1 - \frac{\frac{\lambda}{2N|\mathcal{N}_n|\rho}}{\left\|\boldsymbol{\beta}_f^n\right\|_2}\right)\boldsymbol{\beta}_f^n, & \text{otherwise} \end{cases} \quad (27)$$

where $\boldsymbol{\beta}_f^n = \left(\frac{\boldsymbol{C}_{f,:}^n + [\boldsymbol{U}_r^n]_{f,:} + \boldsymbol{C}_{f,:}^r + [\boldsymbol{V}_n^r]_{f,:}}{2}\right)_+$; for $q = \infty$, the proximal operator can also be solved via computing projection onto an ℓ_1-norm ball [40], [41], whose complexity order can be as small as $\mathcal{O}(F)$ [42].

Following the proposed ADMM algorithm, if there always exists a single or multi-hop path that links any two sensors, the solution of the decentralized algorithm converges to the solution of Problem (21) [19], [38]. After obtaining $\boldsymbol{C}^n$ at each sensor, an estimated index set, i.e., $\widehat{\Lambda}^n$, is known at each sensor. The next step is to estimate the power spectra in a decentralized manner by solving the nonnegative least squares problem (6). Problem (6) can be solved following the same decentralizing strategy as we described for Problem (22), so that a local estimate $\widehat{\boldsymbol{S}}^n$ can be obtained at sensor n; the detailed algorithm is relegated to Appendix B.

C. Reweighting for Performance Enhancement

To enhance the row-sparsity, a commonly used trick is to solve a series of convex relaxation problems with a reweighting scheme. Specifically, in our context, instead of solving Problem (22) once, we repeatedly solve its weighted version with the objective function

$$\frac{1}{2}\sum_{n=1}^N \left\|\bar{\boldsymbol{G}}_{n,:} - \bar{\boldsymbol{G}}_{n,:}\boldsymbol{C}^n\right\|_F^2 + \sum_{n=1}^N \frac{\lambda}{N|\mathcal{N}_n|}\left\|\boldsymbol{\Sigma}^n\boldsymbol{\Gamma}_r^n\right\|_{q,1} \quad (28)$$

for several times, where $\boldsymbol{\Sigma}^n \in \mathbb{R}^{F\times F}$ is a diagonal matrix. Notice that such a change does not affect the ADMM updates, except that (26) is replaced by the following:

$$\min_{[\boldsymbol{\Gamma}_r^n]_{f,:}} \frac{1}{2}\left\| \left(\frac{\boldsymbol{C}_{f,:}^n + [\boldsymbol{U}_r^n]_{f,:} + \boldsymbol{C}_{f,:}^r + [\boldsymbol{V}_n^r]_{f,:}}{2} \right)_+ - [\boldsymbol{\Gamma}_r^n]_{f,:}\right\|_2^2 + \frac{\lambda}{2N|\mathcal{N}_n|\rho}\left\|\boldsymbol{\Sigma}_{f,f}^n[\boldsymbol{\Gamma}_r^n]_{f,:}\right\|_q \quad (29)$$

Solving the above reweighted version is equally easy to solving Problem (26); for example, when $q = 2$, the only change is to replace λ in (27) by $\lambda\boldsymbol{\Sigma}_{f,f}^n$. A good way of choosing $\boldsymbol{\Sigma}^n$ is to set $\boldsymbol{\Sigma}_{f,f}^n = p\left\|\boldsymbol{C}_{f,:}^n\right\|_q^{p-1}$ for $f = 1, \ldots, F$, where $0 < p < 1$. Under such a choice of $\boldsymbol{\Sigma}^n$, the iteratively reweighted algorithm can be interpreted as a successive convex approximation algorithm for solving Problem (21) with $\|\boldsymbol{C}\|_{q,1}$ replaced by $\|\boldsymbol{C}\|_{q,p}^p$, and the latter is provably a better approximation to $\|\boldsymbol{C}\|_{\text{row}-0}$[18]. In our experience, only three to five iterations of reweighting gives much better solution than solving Problem (22) only once.

Remark 5: A side benefit of using formulation in (18) and its mixed-norm optimization surrogates is that the number of sources, i.e., K, is not necessarily needed. In other words, K can be inferred by the number of rows with nonzero (or, practically, significant) norms. A practical way for sensor n to estimate K is to set up a threshold α, and count the number of rows that admit norms larger than α; i.e., one can let the number of sources estimated at sensor n, denoted by $\hat{K}^n$, to be $\hat{K}^n = |\widehat{\Lambda}^n|$ where $\widehat{\Lambda}^n = \{f|\ \|\boldsymbol{C}_{f,:}^n\|_q > \alpha\}$.

To conclude this section, we summarize the decentralized algorithm in Algorithm 2, where one can clearly see the message exchange protocol.

IV. Numerical Results

In this section, we provide numerical results to validate the proposed algorithms. In the simulations, the band of interest consists of B subchannels ($B > K$), which are indexed by $\{1, \ldots, B\}$. Subchannel i for $i = 1, \ldots, K$ is assigned to emitter i so that (A1) is fulfilled. In addition, each transmitter randomly picks J other subchannels from $K+1$ to B. In each occupied subchannel, the corresponding source signal is generated by filtering i.i.d. zero-mean circularly symmetric Gaussian signals by a sinc-shape filter with a random scaling. The channel matrix $\boldsymbol{A}$ is also randomly generated following the zero-mean and unit-variance circularly symmetric Gaussian distribution. For the noise at each sensor, we also generate it

Algorithm 2: Decentralized Power Spectra Separation.

input : $\bar{\boldsymbol{G}}$; $q \in \{2, \infty\}$; K (optional); $p \in (0, 1]$; MaxRwIt.
if $p = 1$ **then**
  MaxRwIt $= 1$;
end
RwIt $= 1$;
let $\boldsymbol{\Gamma}_r^n = \boldsymbol{I}$, $\boldsymbol{C}^n = \boldsymbol{I}$, $\boldsymbol{U}_r^n = \boldsymbol{0}$ and $\boldsymbol{V}_n^r = \boldsymbol{0}$ $\forall n, r \in \mathcal{N}_n$;
while RwIt $\leq$ MaxRwIt **do**
  repeat
    if $p = 1$ **then**
      sensor n updates $\boldsymbol{\Gamma}_r^n$ according to (26);
    else
      sensor n updates $\boldsymbol{\Gamma}_r^n$ according to (29);
    end
    sensor n broadcasts $\boldsymbol{\Gamma}_r^n$ to sensor $r \in \mathcal{N}_n$;
    sensor n updates $\boldsymbol{C}^n$ according to (23b);
    sensor n updates $\boldsymbol{U}_r^n$ according to (23c);
    sensor $r \in \mathcal{N}_n$ updates $\boldsymbol{V}_n^r$ according to (23d);
    sensor $r \in \mathcal{N}_n$ sends $\boldsymbol{V}_n^r$ and $\boldsymbol{C}^r$ to sensor n;
  until *some stopping criterion is satisfied*;
  sensor n computes $\boldsymbol{\Sigma}_{f,f}^n := p\|\boldsymbol{C}_{f,:}^n\|_q^{p-1}$;
  RwIt := RwIt $+ 1$;
end
if K *is given* **then**
  sensor n sorts the frequencies such that $\|\boldsymbol{C}_{f_1,:}^n\|_q \geq \ldots, \geq \|\boldsymbol{C}_{f_F,:}^n\|_q$;
  sensor n finds $\hat{\Lambda}^n = \{f_1, \ldots, f_K\}$;
else
  let $\hat{\Lambda}^n = \{f \mid \|\boldsymbol{C}_{f,:}^n\|_q > \alpha\}$;
end
estimate $\{\hat{\boldsymbol{S}}^n\}$ following the decentralized algorithm in Appendix B;
output: $\{\hat{\boldsymbol{S}}^n\}$; $\{\hat{\Lambda}^n\}$.

following $v_n(t) \sim \mathcal{CN}(0, \sigma^2)$. The signal-to-noise ratio is defined as $\text{SNR} = \text{E}\{\|\boldsymbol{A}\boldsymbol{x}(t)\|_2^2\}/N\sigma^2$, In all the simulations, unless specified, $B = 8$ subbands and $T = 10^6$ samples are used, and the band of interest is discretized into $F = 64$ frequency bins. The noise variance σ_n^2 for $n = 1, \ldots, N$ is estimated and canceled following the same way that was proposed in [12]. All the simulations are carried out using Matlab codes on a desktop computer with i7 2.7 GHz CPU and 16 GB RAM.

A. Centralized Cases

Fig. 5 shows an illustrative example, where the results of applying SPA to the formulation in (8) and the Khatri-Rao structured formulation in (16) are presented; the results of the NMF algorithm, namely, accelerated hierarchical alternating least squares (HALS) [43], that aims at solving (5) are also presented. We assume that the sensors are synchronized so that SPA can be legitimately applied to (16). The shapes of the power spectral densities (PSDs) are fixed for better visual illustration, but the channels and noise are randomly generated at each of the 100 trials. In this simulation, we set $\text{SNR} = 10$ dB, $(N, K) = (5, 2)$ and $J = 4$. We see that applying SPA to both (8) and (16) can successfully separate the individual power spectra. Particularly, using the Khatri-Rao structured formulation in (16) gives visually better results in this case, which we believe is because the structure of $\boldsymbol{A} \odot \boldsymbol{A}^*$ brings more degrees of freedom—$\boldsymbol{A} \odot \boldsymbol{A}^*$ is virtually an $N^2 \times K$ matrix while $\boldsymbol{B}$ is an $N \times K$ matrix. In Fig. 5, we also notice that HALS fails in many trials even when the underlying spectra are separable via SPA, i.e., when (A1) is satisfied. This verifies that NMF without considering additional structure is not promising in this application.

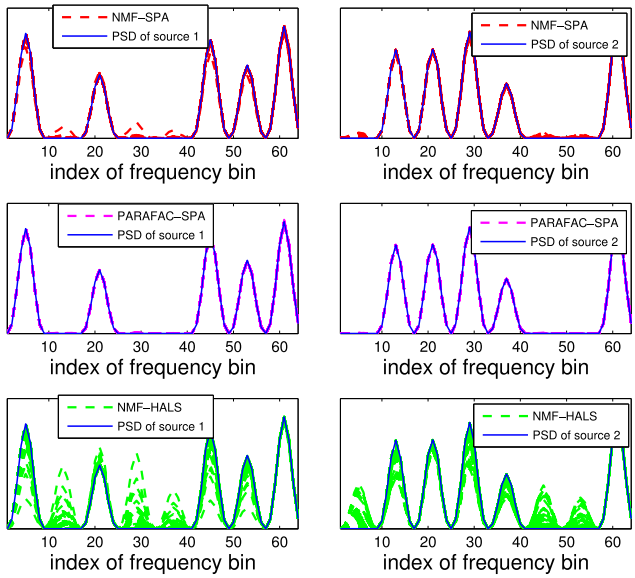


Fig. 5. The separated power spectra of two source by the algorithms. $\text{SNR} = 10$ dB; $(N, K) = (5, 2)$.

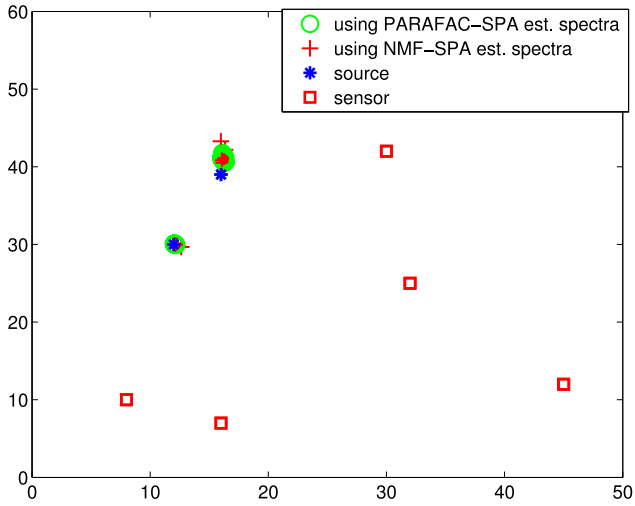


Fig. 6. The source localization result. $\text{SNR} = 10$ dB; $(N, K) = (5, 2)$.

Fig. 6 follows Remark 3, where we present the results of source localization using a RSS-based localization algorithm at the dominant frequencies identified by SPA. The results of 100 trials are shown in this figure. We assume a free-space propagation model within a 60 m $\times$ 50 m region, and the localization algorithm that we employ here is a suboptimal yet simple *least squares* (LS)-based algorithm [27]. Notice that more sophisticated (but computationally more expensive) algorithms such as the semidefinite relaxation (SDR)-based approach [29] can also be adopted to improve the performance for more critical situations, e.g., lower SNR cases. One can see that the estimated locations of the sources are reasonably accurate in this scenario,

TABLE I
THE MSES (DB) AND RUNTIMES (SEC.) OF THE ALGORITHMS UNDER VARIOUS SNRS

Algorithm	measure	SNR(dB)			
		-10	0	10	20
SPA to (8) w/ sync.	MSE	-7.881	-21.605	-22.5294	-23.7287
	time	**0.00067**	**0.00066**	**0.00067**	**0.00067**
SPA to (16) w/ sync.	MSE	-10.3443	**-33.4507**	**-33.3283**	**-33.1508**
	time	0.0015	0.0015	0.0015	0.0015
HALS w/ sync.	MSE	**-10.6787**	-13.6543	-14.0214	-14.1874
	time	0.0177	0.0179	0.018	0.018
SPA to (8) w/o sync.	MSE	-8.0924	**-21.4171**	**-22.6332**	**-22.4693**
	time	**0.00078**	**0.00068**	**0.00068**	**0.00068**
SPA to (16) w/o sync.	MSE	-3.3157	-2.9023	-2.9946	-2.9129
	time	0.0016	0.0015	0.0015	0.0015
HALS w/o sync.	MSE	**-11.6295**	-14.4095	-13.4925	-14.8746
	time	0.0177	0.0177	0.0177	0.0177

TABLE II
THE MSES (DB) AND RUNTIMES (SEC.) OF THE ALGORITHMS UNDER VARIOUS F'S

Algorithm	measure	F				
		32	64	128	256	512
SPA to (8) w./ sync	MSE	**-15.049**	-37.7977	-35.6596	-30.393	-30.8694
	time	**0.0006**	**0.0006**	**0.0007**	**0.0008**	**0.0009**
SPA to (16) w./sync	MSE	-10.5176	**-43.4463**	**-41.498**	**-38.9308**	**-36.3401**
	time	0.0009	0.001	0.0013	0.0017	0.0023
HALS w./ sync.	MSE	-9.03	-9.9069	-9.3839	-11.021	-10.3319
	time	0.028	0.0306	0.0357	0.0417	0.0834
SPA to (8)w./o sync	MSE	**-16.005**	**-38.8587**	**-36.3758**	**-34.1453**	**-30.2041**
	time	**0.001**	**0.0007**	**0.0007**	**0.0009**	**0.0012**
SPA to (16) w./o sync	MSE	-0.7127	-0.3739	-0.3695	-0.3254	-0.4126
	time	0.0011	0.0009	0.0014	0.002	0.0022
HALS w./o. sync.	MSE	-9.9439	-10.4277	-9.8991	-10.3322	-9.7375
	time	0.0276	0.0293	0.0354	0.0418	0.0853

TABLE III
THE MSES (DB) AND RUNTIMES (SEC.) OF THE ALGORITHMS UNDER VARIOUS T'S

Algorithm	measure	T				
		5×10^4	10^5	5×10^5	10^6	5×10^6
SPA to (8) w./ sync	MSE	-25.5018	-27.9365	-34.301	-35.9448	-41.8932
	time	**0.0006**	**0.0007**	**0.0006**	**0.0005**	**0.0005**
SPA to (16) w./sync	MSE	**-28.6534**	**-31.6864**	**-38.1962**	**-40.7278**	**-45.6481**
	time	0.0009	0.001	0.001	0.0007	0.0007
HALS w./ sync.	MSE	-9.9907	-9.9616	-10.0713	-9.8781	-10.5409
	time	0.0303	0.0331	0.0301	0.0252	0.0239
SPA to (8) w./o sync	MSE	**-24.7178**	**-27.9727**	**-35.4133**	**-37.0959**	**-42.098**
	time	**0.0006**	**0.0005**	**0.0006**	**0.0007**	**0.0007**
SPA to (16) w./o sync	MSE	-0.1425	-0.3737	-0.2699	-0.1916	-0.3243
	time	0.0007	0.0007	0.0009	0.001	0.001
HALS w./o. sync.	MSE	-9.7433	-9.7758	-9.7586	-9.8624	-10.3678
	time	0.0249	0.0245	0.0287	0.0296	0.0301

and using the dominant frequencies given by SPA to (16) slightly outperforms using that of SPA to (8), which is consistent with the last example.

In Table I, the mean squared errors (MSEs) of the estimated power spectra of a Monte Carlo simulation is presented; the MSE is defined by

$$\mathrm{MSE} = \min_{\boldsymbol{\pi}\in\Pi} \frac{1}{K}\sum_{k=1}^{K}\left\| \frac{\boldsymbol{S}_{k,:}}{\|\boldsymbol{S}_{k,:}\|_2} - \frac{\widehat{\boldsymbol{S}}_{\pi_k,:}}{\|\widehat{\boldsymbol{S}}_{\pi_k,:}\|_2}\right\|_2^2,$$

where Π is the set of all permutations of $\{1,2,\ldots,K\}$; and $\mathbf{S}_{k,:}$ and $\widehat{\mathbf{S}}_{k,:}$ are the true power spectrum of source k and the corresponding estimate, respectively. The corresponding runtimes are also presented. Here, we set $(N,K)=(5,3)$ and $J=3$,

TABLE IV
THE MSES (DB) AND RUNTIMES (SEC.) OF THE ALGORITHMS UNDER VARIOUS K'S

Algorithm	measure	K			
		2	4	6	8
SPA to (8) w./ sync	MSE	-36.2619	-36.7749	-35.4727	-31.8923
	time	**0.0005**	**0.0006**	**0.0007**	**0.0009**
SPA to (16) w./sync	MSE	**-43.2344**	**-43.0395**	**-42.9522**	**-42.2817**
	time	0.0007	0.0011	0.0024	0.0041
HALS w./ sync.	MSE	-10.9741	-9.3406	-7.4169	-7.091
	time	0.0187	0.0208	0.0254	0.0287
SPA to (8) w./o sync	MSE	**-39.0452**	**-37.1393**	**-34.4917**	**-31.9785**
	time	**0.0006**	**0.0005**	**0.0007**	**0.00091**
SPA to (16) w./o sync	MSE	-0.6718	-0.1136	-0.3186	-0.4186
	time	0.0016	0.0011	0.0022	0.0039
HALS w./o. sync.	MSE	-12.9286	-9.1927	-7.132	-6.7413
	time	0.0199	0.0204	0.0253	0.0288

TABLE V
THE MSES (DB) AND RUNTIMES (SEC.) OF THE ALGORITHMS OF AN UNDER-DETERMINED CASE

Algorithm	measure	SNR(dB)			
		-10	0	10	20
SPA to (16)	MSE	-6.9942	-20.0584	-21.9437	-21.9767
	time	**0.0011**	**0.001**	**0.001**	**0.001**
TALS (random init.)	MSE	-8.7057	-9.9954	-9.6203	-9.8454
	time	1.086	1.0648	1.1076	1.0951
TALS (SPA init.)	MSE	**-11.6296**	**-27.8282**	**-32.7283**	**-33.6526**
	time	0.6575	0.4381	0.396	0.3927

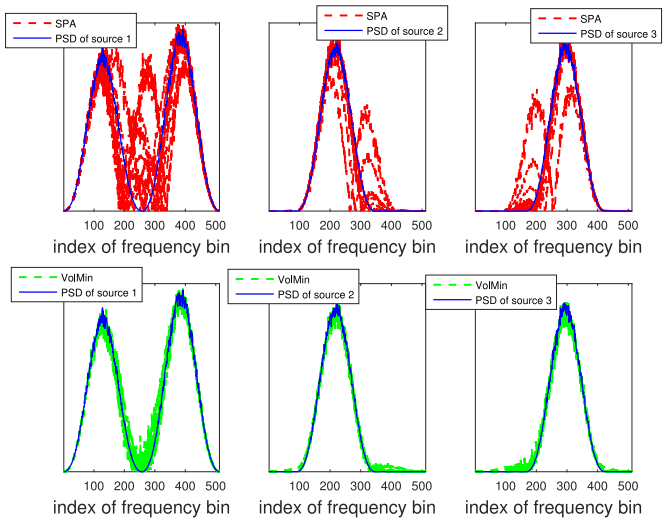


Fig. 7. A case where (A1) does not hold.

and the results are obtained by averaging 100 trials. In this simulation, we test the performance of the algorithms with both synchronous and asynchronous sensors. We see that, with synchronization, SPA applied to (16) yields the most favorable MSE performance, and SPA applied to (8) also gives reasonable estimates. However, if the sensors are not synchronized, SPA applied to (16) fails as expected, while the performance of SPA applied to (8) is not affected. In terms of runtime, we see that SPA applied to (8) outperforms the other two algorithms.

Table V presents the MSE and runtime performance of the algorithms with $(N,K)=(3,6)$ and synchronous sensors; the other settings follow the previous simulation. Notice that

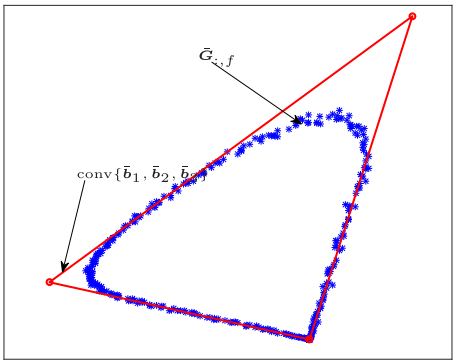


Fig. 8. The geometry of the simulated case in Fig. 7.

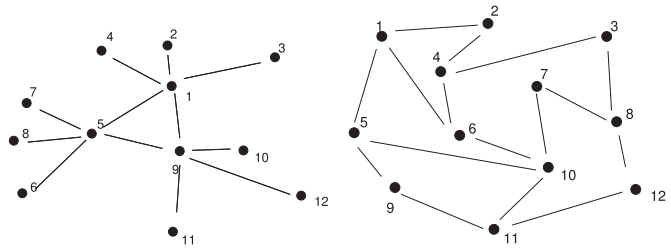


Fig. 9. Topologies of the two simulated sensor networks. The dots represent the sensors, and the lines represent the communication links.

under such circumstances, SPA cannot be applied to (8). We use the PARAFAC algorithm (cf. Remark 4), namely, trilinear alternating least squares (TALS) [13], as benchmark. TALS is stopped when the number of iterations reaches 1000 or the relative change of its cost function is smaller than 10^{-9}. We see that SPA applied to (16) gives good estimation accuracy of the desired power spectra, and the speed is quite appealing, which is more than 300 times faster than the algorithms under comparison. TALS using random initializations cannot give reasonable estimates of the power spectra, since it cannot converge to a good solution within 1000 iterations. However, using the results of SPA as initialization, we see TALS gives excellent estimation results. This suggests that the usage of SPA can be twofold in practice: one can either apply SPA for real-time applications that put emphasis on computational efficiency, or can use it to initialize a PARAFAC algorithm for better accuracy.

In Tables II and III, we test how the performance scales with F and T, respectively. We fix $\mathrm{SNR} = 10$ dB in both simulations, and the other settings are unchanged. We see that applying SPA to (8) yields reasonable results under all situations, but the other algorithms are either not working (NMF-HALS) or sensitive to synchronization (SPA applied to (16)). From Table II, we see that the algorithms perform better for $F = 64$ and $F = 128$ under the current settings—to obtain the same accuracy under finer spectral resolution, one may need more samples. This is also reflected in Table III, where we see that a larger number of samples leads to better performance.

In Table IV, we observe the performance of the algorithms versus the number of sources. We set up $B = 20$ subchannels, assign 3 to each source, and vary K from 2 to 8. The ith subchannel is assigned to source i as before. We let $N = K + 3$, $T = 10^6$, $F = 64$, and $\mathrm{SNR} = 10$ dB. We see that although SPA is a greedy algorithm that may suffer from error accumulation and performs better for smaller K, it gives very satisfactory spectra estimation accuracy even when $K = 8$.

In Figs. 7–8, we show an example where local dominance does not hold. The spectra of the three sources are shown in Fig. 7 with blue solid lines. Fig. 8 shows the columns of $\bar{\boldsymbol{G}}$ projected onto $\mathrm{conv}\{\bar{\boldsymbol{a}}_1, \bar{\boldsymbol{a}}_2, \bar{\boldsymbol{a}}_3\}$. Clearly, two vertices are not touched and (A1) is significantly violated. Under this setup, we run SPA and VolMin for 10 different trials, where the sources and the channels are randomly generated at each trial. We see that SPA does not succeed in recovering the spectra under this situation, but VolMin works well. The results here echo Remark 2 that, in challenging cases, VolMin can still provide reasonable solution by better exploiting the same signal model. The cost is that the average runtime of the VolMin algorithm (0.0435 sec.) that we employ here (i.e., the algorithm in [24]) is around 20 times slower than that of SPA (0.0023 sec).

B. Decentralized Cases

In this subsection, we use simulations to verify the effectiveness of the proposed decentralized algorithm. We test the proposed algorithm with and without reweighting, respectively. Throughout this section, we set the maximal iterations of reweighting (i.e., MaxRwIt in Algorithm 2) to be 5 and let $p = 0.05$. For the regularization parameter λ, we fix it for the algorithm with and without reweighting to be 10^{-3} and 0.1, respectively. For the $p = 1$ case (no reweighting), the step size of ADMM is set to be $\rho = 10^{-3}$. ADMM is stopped if the number of iterations reaches 10^5 or $\| \boldsymbol{\Gamma}_r^1 - \boldsymbol{C}^1 \|_F \leq \eta$ where $\eta = 10^{-7}$ for some $r \in \mathcal{N}_n$. For the $p < 1$ case, we adopt a more aggressive stopping criterion: for the first four reweighted subproblems, we set $\eta = 10^{-2}$, and for the last subproblem, we set $\eta = 10^{-4}$. As we will show, this stopping criterion significantly decreases the total number of iterations, while not affecting the performance when combined with the reweighting strategy. We assume that K is known and pick up $\widehat{\Lambda}^n$ by identifying the rows of $\boldsymbol{C}^1$ with the K largest norms.

We test the algorithms under the sensor networks depicted in Fig. 9. The histograms of MSEs of the estimated power spectra obtained at sensor 1 from 100 trials are presented Fig. 10. The results of the centralized algorithm, i.e., SPA applied to (8), are also presented as baseline. In both networks, we set $\mathrm{SNR} = 10$ and let $q = 2$ and $(N, K) = (12, 3)$. We see that the proposed algorithm without reweighting gives reasonable MSE performance for most of the trials, but there exist several trials where the algorithm fails. On the other hand, we see that reweighting indeed improves the performance—at all trials, the iteratively reweigted algorithm succeeds.

Fig. 11 may shed some light on the reason why the reweighted algorithm works better. Here, we present the $\{\| \boldsymbol{C}_{f,:}^n \|_2\}$ obtained at one instance with the network in Fig. 9. We take the obtained solution at sensor 1 for observation. We see that the proposed algorithm without reweighting works reasonably well—the rows with the 3 largest 2-norms correspond to the desired

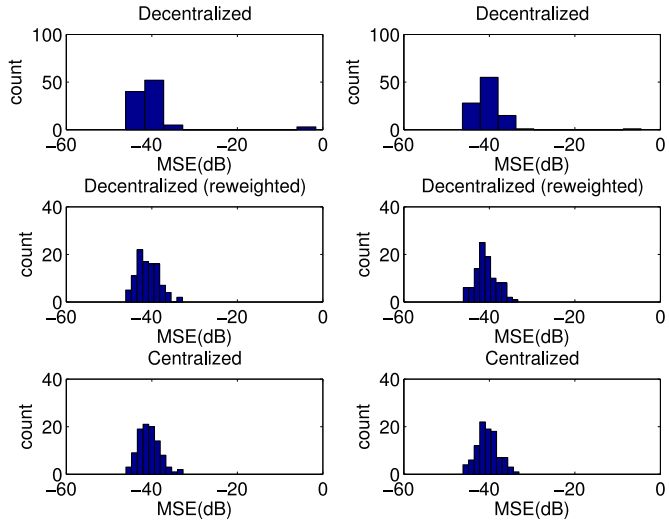


Fig. 10. The left and right columns are the MSE histograms of the algorithms with the left and right networks in Fig. 9, respectively.

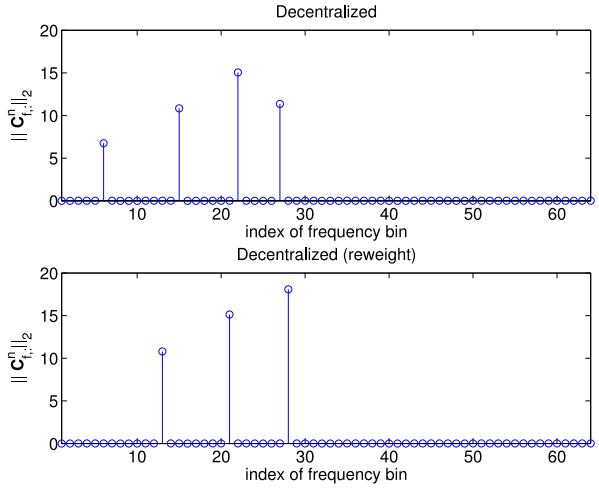


Fig. 11. The values of $\{\| \boldsymbol{C}^1_{f,:}\|_2\}$ yielded by the proposed algorithm with and without reweighting, respectively, at one randomly picked instance; the network is as in Fig. 9; $\mathrm{SNR} = 10$ dB.

TABLE VI
THE ACTIVE ROWS OF THE OBTAINED $\boldsymbol{C}^n$ FOR $n = 1$ BY THE PROPOSED ALGORITHMS AND THE TOTAL ITERATIONS NEEDED; $\mathrm{SNR} = 10$ DB

Algorithm	measure	network 1 (Fig. 9, left)	network 2 (Fig. 9, right)
w/o reweight.	iterations	7020	7754
	$\mathrm{prob}(\hat{K} = K)$	29%	36%
w/ reweight.	iterations	512	510
	$\mathrm{prob}(\hat{K} = K)$	97%	93%

dominant frequencies. However, there is another frequency that is 'active', which could confuse the frequency picking stage sometimes. But for the reweighted version, we see the solution is very clean, only three active rows of $\boldsymbol{C}^n$ exist. The reason is that using $p < 1$ is able to yield sparser solutions. Table VI confirms this observation. For both networks under test, the reweighted algorithm gives smaller numbers of active rows, $\hat{K}$, (i.e., rows with 2-norms larger than $\alpha = 0.1$) of $\boldsymbol{C}^1$, which are

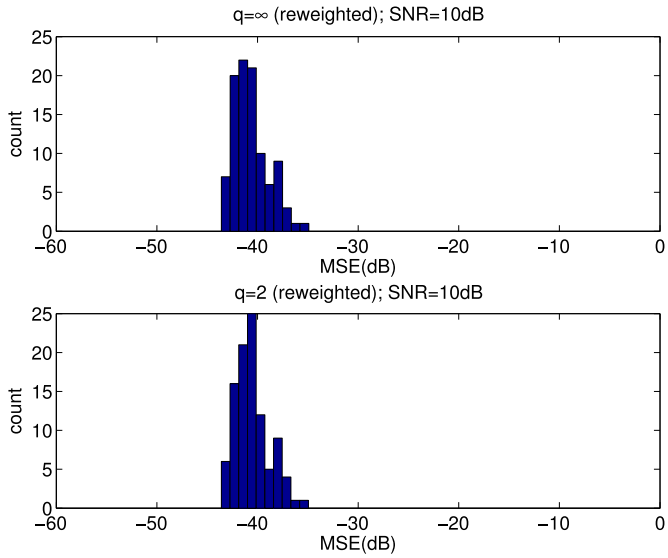


Fig. 12. MSE histograms of the decentralized algorithm using $q = \infty$ and $q = 2$, respectively.

with high probability to be K. This result also supports our claim in Remark 5 that K is not necessarily needed as an input if we adopt the sparse optimization-based criterion.

Another improvement of using the iteratively reweighted algorithm lies in the number of iterations. In general, ADMM can converge slowly [19]. But when combining with an iterative reweighting procedure, one may terminate the reweighted subproblem early and go on to the next subproblem—it is not necessary to solve each subproblem to a fine-accuracy level. By doing so, we see a big saving in the iterations: in Table VI, we see that solving the problem without reweighting costs more than 7000 iterations, but it requires only around 500 iterations in total for solving 5 reweighted subproblems.

We also test different q's under the network depicted on the right of Fig. 9. In Fig. 12, we see that using both $q = 2$ and $q = \infty$ yield very similar results. However, in terms of optimization complexity, $q = 2$ is preferable since it admits closed-form update of $\boldsymbol{\Gamma}^n_r$.

C. Experiment Using Software-Defined Radios

In this subsection, we present a laboratory experiment using real software-defined radios. The experiment was implemented in the communications laboratory of the Department of Electrical and Computer Engineering at the University of Minnesota, and we used the Universal Software Radio Peripheral (USRP) radios by Ettus Research as both transmitters and receivers. Two sources and two sensors were used in the experiment, and the sensors were not synchronized. The sources were placed approximately 3–4 m away from the sensors, and at 2–3m from each other. All radios were communicating at a carrier frequency of 2.5 GHz, with the sources transmitting random BPSK signals with a bandwidth of 100 KHz. We discretized the 300 KHz bandwidth into 1024 frequency bins. Bins 411–614 were allocated to source 1 and bins 205–410 and 615–819 to source 2.

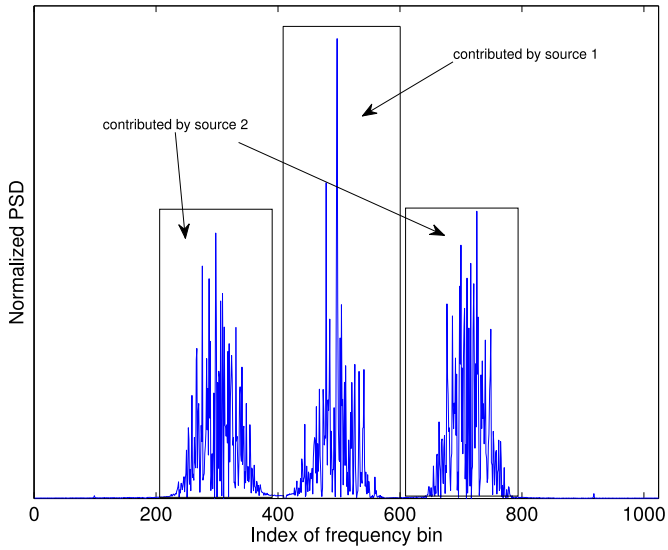


Fig. 13. The measured mixed power spectra at sensor 2.

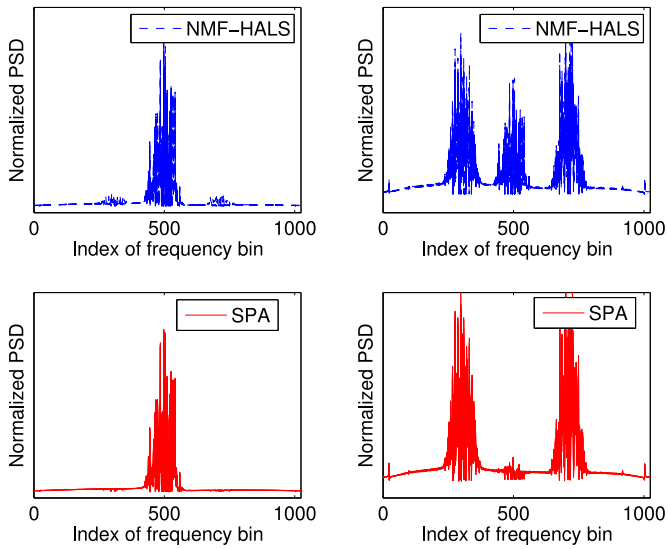


Fig. 14. The separated power spectra by SPA and HALS.

The received signals are mixtures; see Fig. 13 for an illustration of the power spectrum of $y_2(t)$, i.e., the actual signal that is received at sensor 2. Fig. 14 illustrates the separated spectra that are obtained by applying SPA on (8) and HALS, respectively. We plot the separated spectra from 10 different measured datasets under the same setups, and we see that SPA works consistently well. The results are quite encouraging: We see that for all the 10 trials, SPA successfully separated the two spectra even in the indoor laboratory environment, where modeling errors are inevitable. We also notice that the NMF algorithm, i.e., HALS, failed at this experiment, which is consistent with our simulations.

V. Conclusion

In this work, we revisited the power spectra separation problem. Unlike our prior work [11], [12] which made use of baseline-aligned multiple-antenna sensors, we reformulated the problem using simple single-antenna sensors without any geometrical specification on their deployment. A structured matrix factorization-based framework was established, which guarantees the identifiability of the power spectra under a realistic local dominance assumption in the spectral domain. Efficient centralized and effective decentralized algorithms were proposed under this framework. Both simulations and a real experiment using prototype software-defined radios demonstrated the effectiveness of the proposed algorithms.

Appendix A
Proof of Theorem 1

For notational convenience, we assume that, without loss of generality, $\mathbf{\Pi} = \boldsymbol{I}$, or, equivalently, $\Lambda = \{1, \ldots, K\}$. Hence, we see that $\boldsymbol{C}^\star$ can be expressed as $\boldsymbol{C}^\star = \left[\begin{smallmatrix} \bar{\boldsymbol{S}} \\ \boldsymbol{0} \end{smallmatrix}\right]$. For any feasible solution, we have $\bar{\boldsymbol{G}} = \bar{\boldsymbol{G}}\boldsymbol{C}$. Hence,

$$\bar{\boldsymbol{G}}\boldsymbol{C}^\star = \bar{\boldsymbol{G}}\boldsymbol{C} \Rightarrow \bar{\boldsymbol{B}}\bar{\boldsymbol{S}}\boldsymbol{C}^\star = \bar{\boldsymbol{B}}\bar{\boldsymbol{S}}\boldsymbol{C}.$$

Recall that $\boldsymbol{B}$ has no null space. Hence, we now have

$$\bar{\boldsymbol{S}}\boldsymbol{C}^\star = \bar{\boldsymbol{S}}\boldsymbol{C} \Rightarrow \boldsymbol{C}^\star_{\Lambda,:} = \bar{\boldsymbol{C}}^\star = \bar{\boldsymbol{S}} = \bar{\boldsymbol{S}}\boldsymbol{C}.$$

Following the above, we see that $\bar{\boldsymbol{C}}^\star_{k,:} = \sum_{f=1}^{F} \bar{\boldsymbol{S}}_{k,f}\boldsymbol{C}_{f,:}$. Hence, for $k = 1, \ldots, K$, we have

$$\| \bar{\boldsymbol{C}}^\star_{k,:} \|_q \leq \sum_{f=1,\ldots,F} \bar{\boldsymbol{S}}_{k,f} \| \boldsymbol{C}_{f,:} \|_q \tag{30}$$

by the triangle inequality. Notice that, the equality holds if and only if

$$\bar{\boldsymbol{C}}^\star_{k,:} = \boldsymbol{C}_{k,:}, \quad \text{and} \quad \boldsymbol{C}_{f,:} = \boldsymbol{0}, \quad \forall f > K,$$

since $\bar{\boldsymbol{S}}_{:,1:K} = \boldsymbol{I}$ and there exists no other unit vectors in $\bar{\boldsymbol{S}}$ by the assumption that $\bar{\boldsymbol{G}}$ has no repeated columns. Now, summing up across k we have

$$\sum_{k=1}^{K} \| \bar{\boldsymbol{C}}^\star_{k,:} \|_q \leq \sum_{k=1}^{K}\sum_{f=1}^{F} \bar{\boldsymbol{S}}_{k,f} \| \boldsymbol{C}_{f,:} \|_q.$$

By the fact that $\sum_{k=1}^{K} \bar{\boldsymbol{S}}_{k,f} = 1$, we see that

$$\sum_{k=1}^{K} \| \bar{\boldsymbol{C}}^\star_{k,:} \|_q = \| \boldsymbol{C}^\star \| \leq \sum_{f=1}^{F} \| \boldsymbol{C}_{f,:} \|_q = \| \boldsymbol{C} \|_{q,1}.$$

Clearly, the lower bound is achieved if and only if the lower bound in (30) is attained for all $k = 1, \ldots, K$. That means,

$$\boldsymbol{C}^\star = \boldsymbol{C}_{\text{opt}},$$

which completes the proof.

APPENDIX B
DECENTRALIZED NONNEGATIVE LEAST SQUARES

We can recast the above into the following form:

$$\begin{aligned}\min_{\{\boldsymbol{\Upsilon}_r^n\},\{\boldsymbol{S}^n\}} \quad & \frac{1}{2}\sum_{n=1}^{N}\left\|\boldsymbol{G}_{n,:}-\widehat{\boldsymbol{B}}_{n,:}\boldsymbol{S}^n\right\|_F^2\\ & \text{s.t. } \boldsymbol{S}^n=\boldsymbol{\Upsilon}_r^n, \boldsymbol{S}^r=\boldsymbol{\Upsilon}_r^n, r\in\mathcal{N}_n,\\ & \boldsymbol{\Upsilon}_r^n\geq\boldsymbol{0},\end{aligned}$$

where $\widehat{\boldsymbol{B}}_{n,:}=\boldsymbol{G}_{n,\Lambda^n}$, which can be obtained locally at sensor n. The ADMM updates are presented in (31),

$$\begin{aligned}\boldsymbol{\Upsilon}_r^n :=\ & \arg\min_{\boldsymbol{\Upsilon}_r^n\geq\boldsymbol{0}}\frac{\rho}{2}\left\|\boldsymbol{S}^n-\boldsymbol{Y}_r^n+\widetilde{\boldsymbol{U}}_r^n\right\|_F^2\\ & +\frac{\rho}{2}\left\|\boldsymbol{S}^r-\boldsymbol{Y}_r^n+\widetilde{\boldsymbol{V}}_n^r\right\|_F^2 \qquad (31a)\\ \boldsymbol{S}^n :=\ & \arg\min_{\boldsymbol{S}^n}\frac{1}{2}\|\boldsymbol{G}_{n,:}-\boldsymbol{B}_{n,:}\boldsymbol{S}^n\|_F^2\\ & +\frac{\rho}{2}\sum_{r\in\mathcal{N}_n}\left\|\boldsymbol{S}^n-\boldsymbol{\Upsilon}_r^n+\widetilde{\boldsymbol{U}}_r^n\right\|_F^2\\ & +\frac{\rho}{2}\sum_{r\in\mathcal{N}_n}\left\|\boldsymbol{S}^n-\boldsymbol{\Upsilon}_n^r+\widetilde{\boldsymbol{V}}_r^n\right\|_F^2 \qquad (31b)\\ \widetilde{\boldsymbol{U}}_r^n :=\ & \boldsymbol{S}^n-\boldsymbol{\Upsilon}_r^n+\widetilde{\boldsymbol{U}}_r^n, \qquad (31c)\\ \widetilde{\boldsymbol{V}}_n^r :=\ & \boldsymbol{S}^r-\boldsymbol{\Upsilon}_r^n+\widetilde{\boldsymbol{V}}_n^r, \qquad (31d)\end{aligned}$$

where $\widetilde{\boldsymbol{U}}_r^n$ and $\widetilde{\boldsymbol{V}}_n^r$ are the dual variables associated with $\boldsymbol{S}^n=\boldsymbol{\Upsilon}_r^n$ and $\boldsymbol{S}^r=\boldsymbol{\Upsilon}_r^n$, respectively. Notice that Problems (31a)–(31b) both admit closed-form solutions. Specifically, we have

$$\boldsymbol{\Upsilon}_r^n:=\left(\frac{\boldsymbol{S}^n+\widetilde{\boldsymbol{U}}_r^n+\boldsymbol{S}^r+\widetilde{\boldsymbol{V}}_n^r}{2}\right)_+,$$

and

$$\begin{aligned}\boldsymbol{S}^n :=\ & \left(2|\mathcal{N}_n|\rho\boldsymbol{I}+(\boldsymbol{B}_{n,:})^T\boldsymbol{B}_{n,:}\right)^{-1}\\ & \times\left((\boldsymbol{B}_{n,:})^T\boldsymbol{G}_{n,:}+\rho\left(\boldsymbol{\Upsilon}_r^n-\widetilde{\boldsymbol{U}}_r^n+\boldsymbol{\Upsilon}_n^r-\widetilde{\boldsymbol{V}}_r^n\right)\right).\end{aligned}$$

REFERENCES

[1] X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Tensor-based power spectra separation and emitter localization for cognitive radio," in *Proc. IEEE SAM*, 2014.

[2] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio: State-of-the-art and recent advances," *IEEE Signal Process. Mag.*, vol. 29, no. 3, pp. 101–116, May 2012.

[3] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. ICASSP*, 2007, vol. 4, pp. 1357–1360.

[4] Z. Fanzi, C. Li, and Z. Tian, "Distributed compressive spectrum sensing in cooperative multihop cognitive networks," *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 1, pp. 37–48, Feb. 2011.

[5] 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, Mar. 2010.

[6] D. D. Arianandа and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4775–4789, Sep. 2012.

[7] O. Mehanna and N. D. Sidiropoulos, "Frugal sensing: Wideband power spectrum sensing from few bits," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2693–2703, May 2013.

[8] O. Mehanna and N. D. Sidiropoulos, "Channel tracking and transmit beamforming with frugal feedback," *IEEE Trans. Signal Process.*, vol. 62, no. 24, pp. 6402–6413, Dec. 2014.

[9] O. Mehanna and N. D. Sidiropoulos, "Maximum likelihood passive and active sensing of wideband power spectra from few bits," *IEEE Trans. Signal Process.*, vol. 63, no. 6, pp. 1391–1403, Mar. 2015.

[10] A. Konar, N. D. Sidiropoulos, and O. Mehanna, "Parametric frugal sensing of power spectra for moving average models," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1073–1085, Mar. 2015.

[11] X. Fu, N. D. Sidiropoulos, W.-K. Ma, and J. H. Tranter, "Blind spectra separation and direction finding for cognitive radio using temporal correlation-domain ESPRIT," in *Proc. IEEE ICASSP*, 2014.

[12] X. Fu, N. D. Sidiropoulos, J. H. Tranter, and W.-K. Ma, "A factor analysis framework for power spectra separation and emitter localization," *IEEE Trans. Signal Process.*, 2015.

[13] 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, May 2005.

[14] X. Fu, W.-K. Ma, K. Huang, and N. D. Sidiropoulos, "Blind separation of quasi-stationary sources: Exploiting convex geometry in covariance domain," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2306–2320, May 2015.

[15] X. Fu, W.-K. Ma, T.-H. Chan, and J. M. Bioucas-Dias, "Self-dictionary sparse regression for hyperspectral unmixing: Greedy pursuit and pure pixel search are related," *IEEE J. Sel. Top. Signal Process.*, 2015.

[16] E. Elhamifar, G. Sapiro, and R. Vidal, "See all by looking at a few: Sparse modeling for finding representative objects," in *Proc. IEEE CVPR*, 2012, pp. 1600–1607.

[17] X. Fu, W.-K. Ma, T.-H. Chan, J.-M. Bioucas-Dias, and M.-D. Iordache, "Greedy algorithms for pure pixels identification in hyperspectral unmixing: A multiple-measurement vector viewpoint," in *Proc. EUSIPCO*, 2013.

[18] X. Fu and W.-K. Ma, "Robustness analysis of structured matrix factorization via self-dictionary mixed-norm optimization," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 60–64, 2016.

[19] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, pp. 1–122, 2011.

[20] K. Huang, N. D. Sidiropoulos, and A. Swami, "Non-negative matrix factorization revisited: New uniqueness results and algorithms," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 211–224, Jan. 2014.

[21] W.-K. Ma *et al.*, "A signal processing perspective on hyperspectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, Jan. 2014.

[22] N. Gillis and S. Vavasis, "Fast and robust recursive algorithms for separable nonnegative matrix factorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 698–714, Apr. 2014.

[23] U. Araújo *et al.*, "The successive projections algorithm for variable selection in spectroscopic multicomponent analysis," *Chemometrics Intell. Lab. Syst.*, vol. 57, no. 2, pp. 65–73, 2001.

[24] J. M. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. IEEE WHISPERS*, 2009, pp. 1–4.

[25] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4418–4432, Nov. 2009.

[26] H. C. So and L. Lin, "Linear least squares approach for accurate received signal strength based source localization," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 4035–4040, Aug. 2011.

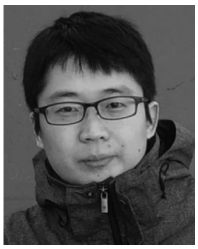
[27] P. Stoica and J. Li, "Lecture notes—Source localization from range-difference measurements," *IEEE Signal Process. Mag.*, vol. 23, no. 6, pp. 63–66, Nov. 2006.

[28] D. Li and Y. H. Hu, "Least square solutions of energy based acoustic source localization problems," in *Proc. ICPPW*, 2004, pp. 443–446.

[29] A. Beck, P. Stoica, and J. Li, "Exact and approximate solutions of source localization problems," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1770–1778, May 2008.

[30] N. Gillis, "Successive nonnegative projection algorithm for robust nonnegative blind source separation," *SIAM J. Imaging Sci.*, vol. 7, no. 2, pp. 1420–1450, 2014.

[31] A. Belouchrani, K. A. Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.
[32] L. D. Lathauwer and J. Castaing, "Blind identification of underdetermined mixtures by simultaneous matrix diagonalization," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1096–1105, Mar. 2008.
[33] A. Dimakis, S. Kar, J. Moura, M. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proc. IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.
[34] W. Shi, Q. Ling, G. Wu, and W. Yin, "Extra: An exact first-order algorithm for decentralized consensus optimization," *SIAM J. Optim.*, vol. 25, no. 2, pp. 944–966, 2015.
[35] E. Esser, M. Moller, S. Osher, G. Sapiro, and J. Xin, "A convex model for nonnegative matrix factorization and dimensionality reduction on physical space," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3239–3252, Jul. 2012.
[36] R. Ammanouil, A. Ferrari, C. Richard, and D. Mary, "Blind and fully constrained unmixing of hyperspectral images," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5510–5518, Dec. 2014.
[37] Q. Qu, N. Nasrabadi, and T. Tran, "Subspace vertex pursuit: A fast and robust near-separable nonnegative matrix factorization method for hyperspectral unmixing," *IEEE J. Select. Top. Signal Process.*, 2015.
[38] 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, 2010.
[39] R. Tandon and S. Sra, "Sparse nonnegative matrix approximation: New formulations and algorithms," *Rapport Technique*, vol. 193, pp. 38–42, 2010.
[40] E. Elhamifar, G. Sapiro, and S. S. Sastry, "Dissimilarity-based sparse subset selection," 2014. [Online]. Available: http://arxiv.org/abs/1407.6810
[41] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 123–231, 2013.
[42] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projections onto the ℓ_1-ball for learning in high dimensions," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 272–279.
[43] N. Gillis and F. Glineur, "Accelerated multiplicative updates and hierarchical als algorithms for nonnegative matrix factorization," *Neural Comput.*, vol. 24, no. 4, pp. 1085–1105, 2012.



Xiao Fu (S'12–M'15) received the B.Eng. and M.Eng. degrees in communication and information engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2005 and 2010, respectively, and the Ph.D. degree in electronic engineering from the Chinese University of Hong Kong (CUHK), Shatin, N.T., Hong Kong, in 2014.

From 2005 to 2006, he was an Assistant Engineer at China Telecom Co. Ltd., Shenzhen, China. He is currently a Postdoctoral Associate at the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA. His research interests include signal processing and machine learning, with a recent emphasis on factor analysis and its applications.

Dr. Fu received the Best Student Paper Award at IEEE ICASSP 2014, and was a finalist of the Best Student Paper Competition at IEEE SAM 2014. He also coauthored a paper that received the Best Student Paper Award at IEEE CAMSAP 2015.

Nicholas D. Sidiropoulos (F'09) received the Diploma in electrical engineering from the Aristotelian University of Thessaloniki, Greece, in 1988 and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland—College Park, College Park, MD, USA, in 1990 and 1992, respectively.

He served as Assistant Professor at the University of Virginia (1997–1999); Associate Professor at the University of Minnesota—Minneapolis (2000–2002); Professor at the Technical University of Crete, Greece (2002–2011); and Professor at the University of Minnesota—Minneapolis (2011–). His current research focuses primarily on signal and tensor analytics, with applications in cognitive radio, big data, and preference measurement.

Dr. Sidiropoulos received the NSF/CAREER award (1998), the IEEE Signal Processing Society (SPS) Best Paper Award (2001, 2007, 2011), and the IEEE SPS Meritorious Service Award (2010). He has served as IEEE SPS Distinguished Lecturer (2008–2009), and Chair of the IEEE Signal Processing for Communications and Networking Technical Committee (2007–2008). He received the Distinguished Alumni Award of the Department of Electrical and Computer Engineering, University of Maryland—College Park in 2013, and was elected EURASIP Fellow in 2014.

Wing-Kin Ma (M'01–SM'11) received the B.Eng. degree in electrical and electronic engineering from the University of Portsmouth, Portsmouth, U.K., in 1995 and the M.Phil. and Ph.D. degrees in electronic engineering from The Chinese University of Hong Kong (CUHK), Shatin, N.T., Hong Kong, in 1997 and 2001, respectively.

He is currently an Associate Professor with the Department of Electronic Engineering, CUHK. From 2005 to 2007, he was also an Assistant Professor with the Institute of Communications Engineering, National Tsing Hua University, Taiwan. Prior to becoming a faculty member, he held various research positions with McMaster University, Canada; CUHK; and the University of Melbourne, Australia. His research interests are in signal processing, communications and optimization, with a recent emphasis on MIMO transceiver designs and interference management, blind separation and structured matrix factorization, and hyperspectral unmixing in remote sensing.

Dr. Ma currently serves as the Senior Area Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the Associate Editor of *Signal Processing*. Also, he has previously served as an Associate Editor and Guest Editor of several journals, which include the IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE SIGNAL PROCESSING LETTERS, the IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS, and the IEEE SIGNAL PROCESSING MAGAZINE. He was a tutorial speaker at EUSIPCO 2011 and ICASSP 2014. He is currently a Member of the Signal Processing Theory and Methods Technical Committee (SPTM-TC) and the Signal Processing for Communications and Networking Technical Committee (SPCOM-TC). He has received several awards, such as ICASSP Best Student Paper Awards in 2011 and 2014, respectively (by Dr. Ma's students), WHISPERS 2011 Best Paper Award, the Research Excellence Award 2013–2014 by CUHK, and the 2015 IEEE Signal Processing Magazine Best Paper Award.