

# A Factor Analysis Framework for Power Spectra Separation and Multiple Emitter Localization

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

**Abstract**—Spectrum sensing for cognitive radio has focused on detection and estimation of aggregate spectra, without regard for latent component identification. Unraveling the constituent power spectra and the locations of ambient transmitters can be viewed as the next step towards situational awareness, which can facilitate efficient opportunistic transmission and interference avoidance. This paper focuses on power spectra separation and multiple emitter localization using a network of multi-antenna receivers. A *PARAllel FACtor analysis* (PARAFAC)-based framework is proposed, which offers an array of attractive features, including identifiability guarantees, ability to work with asynchronous receivers, and low communication overhead. Dealing with corrupt receiver reports due to shadowing or jamming can be a practically important concern in this context, and addressing it requires new theory and algorithms. A robust PARAFAC formulation and a corresponding factorization algorithm are proposed for this purpose, and identifiability of the latent factors is theoretically established for this more challenging setup. In addition to pertinent simulations, real experiments with a software radio prototype are used to demonstrate the effectiveness of the proposed approach.

**Index Terms**—Spectrum estimation, spectra separation, emitter localization, tensor factorization, nonnegativity, robust estimation, cognitive radio.

## I. INTRODUCTION

**C**OGNITIVE radio can help resolve the problem of spectrum scarcity, by exploring and judiciously exploiting transmission opportunities in space, time, and frequency. *Spectrum sensing* is the first step towards this end, enabling secondary spectrum reuse while limiting collisions and persistent interference to licensed users [2], [3].

There is rich literature on spectrum sensing viewed as a set of parallel detection problems, one per frequency bin; see [4]

Manuscript received January 15, 2015; revised May 12, 2015; accepted June 18, 2015. Date of publication August 03, 2015; date of current version November 12, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Weifeng Su. The work of X. Fu, N. D. Sidiropoulos, and J. H. Tranter was supported in part by ARO STIR W911NF-15-1-0384. A conference version of part of this work was presented at ICASSP 2014 [1].

X. Fu was with the Department of Electronic Engineering, the Chinese University of Hong Kong. He is now with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: xfu@umn.edu).

N. D. Sidiropoulos and J. H. Tranter are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: nikos@umn.edu; trant004@umn.edu).

W.-K. Ma is with the Department of Electronic Engineering, the 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.2015.2464194

for a recent tutorial. Wideband spectrum sensing generally requires high sampling rates, implying expensive analog-to-digital converters (ADCs) that consume considerable amount of energy and can hardly fit in portable devices. Exploiting frequency-domain sparsity, compressive spectrum sensing can obtain accurate spectrum estimates at sub-Nyquist sampling rates, without frequency sweeping [5]. Cooperative spectrum sensing schemes that use compressive sensing have been considered in [6], [7], where the spectrum is estimated locally, then consensus on globally fused sensing outcomes is reached.

Whereas most work on spectrum sensing (e.g., [4]–[7]) has focused on reconstructing the signal's *Fourier spectrum* (i.e., the Fourier transform of the signal itself), in cognitive radio and certain other applications only the *power spectrum* (PS) (i.e., the Fourier transform of the signal's autocorrelation) is needed—there is no reason to reconstruct or demodulate the time-domain signal itself [8]–[10]. It was shown in [9] that the sampling rate requirements can be considerably relaxed by exploiting a low-order correlation model, without even requiring spectrum sparsity. The main idea in this line of work is that power measurements are linear in the autocorrelation function, hence a finite number of autocorrelation lags can be estimated by building an over-determined system of linear equations. This autocorrelation-based parametrization also underpins recent work in so-called *frugal sensing* [11]–[13], dealing with power spectrum estimation from 1-bit measurements.

Since the ambient radio frequency (RF) spectrum varies in space and with time, spectrum sensing is only the first step towards *situational awareness*—mapping out the RF environment to enable highly efficient reuse in space and time by unraveling the *transmitted* power spectra and locations of active emitters within range, including possibly misbehaving users or jammers. While a lot effort has been invested in spectrum sensing, including distributed and collaborative detection methods and their performance analysis, very limited progress has been made to date towards bringing RF cognition to the next stage. This paper takes a step in this direction: it tackles the problem of separating the underlying power spectra and localizing the transmitters from measured power spectra mixtures. Knowing the individual power spectrum and the location of each transmitter is useful for a number of reasons. For instance, combining the power spectra 'atoms' with location information allows intelligent beamforming (see Fig. 1 for an example) and more realistic spatial power spectrum interpolation. Transmitter power spectra and locations are also valuable from a network security point of view, as this information can be used to authenticate legitimate users and to detect intruders and jammers. Unraveling

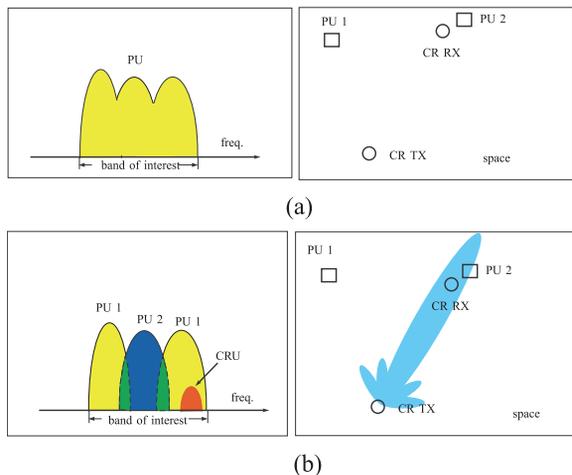


Fig. 1. Motivation for spectra separation and transmitter localization. Primary user 1 (PU1) is engaged in two-way communication with another node (not shown) using the same set of frequencies to receive and transmit in time-division duplex (TDD) mode. PU2 is likewise communicating with another node (not shown). (a) Using aggregate spectrum sensing, the cognitive radio units (CRU) see the band of interest fully occupied. Since PU2 and CR receivers are co-located, beamforming cannot be used for spatial interference avoidance. (b) If the individual PU power spectra and node locations can be estimated, on the other hand, the CR transmitter can modulate its signal in the band occupied by PU1 and beamform towards the CR receiver/PU2.

the measured spectra into their constituent components is therefore well-motivated, but also challenging—primarily because it falls under the ‘blind’ signal processing and separation regime.

To the best of our knowledge, the only work that has explicitly considered the problem of separating the transmitted spectra from multiple received spectral mixtures is [14]; there, the problem is formulated, and then treated, as non-negative matrix factorization (NMF) problem. This is an interesting point of view, however NMF is not unique in general, hence one cannot guarantee identifiability of the latent spectra this way—unless they happen to be few, sparse, and ‘random’, and identification conditions are hard to check. Another difficulty is that NMF does not admit a simple algebraic solution, even under relatively ideal conditions. Another related piece of work is *spectrum cartography* [15], [16], which aims to predict the ambient spectrum at any point in space, based on a limited set of measurements taken at different points in space. The authors of [15], [16] proposed using suitable basis expansion and dictionary learning (DL) models to *interpolate* the entire space-frequency function from the available measurements. Presumably, if one interpolates the spectrum at a point where a transmitter is actually present, one would ideally hope to recover that transmitter’s spectrum. Such an approach is not geared for latent spectra identification, however, as the setup is under-determined. The objective of [15], [16] is to *approximate* spectral mixtures at different points in space, using a basis expansion model and a DL parametrization (which is not unique). **Contributions:** The first contribution of this paper lies in formulating the joint power spectra separation and multiple emitter localization problem as a *PARALLEL FACTOR ANALYSIS* (PARAFAC) problem *in the temporal correlation domain*, under the classic autocorrelation-based parametrization of power spectrum analysis. Through this formulation, the identifiability of both the power spectrum and the direction-of-arrival

(DOA) of each transmitter can be guaranteed, as the *loading factors* of the PARAFAC model are unique (up to scaling and column-permutation ambiguities) under certain fairly mild conditions. The specific PARAFAC model that is proposed here involves two real non-negative loading matrices, corresponding to the sought power spectra and path losses, and a constant-modulus complex one associated with DOAs. This structure can be exploited in iterative model-fitting algorithms to improve estimation performance in difficult scenarios. Moreover, the proposed formulation does not require synchronization among the receivers, which can save considerable amounts of signaling overhead in practice.

The second major contribution addresses a practically important concern: handling corrupt receiver reports due to shadowing, jamming, or sensor failure. Systematically corrupted data represent a serious challenge for any method, and addressing it in our present context requires new theory and algorithms. A robust PARAFAC formulation and a corresponding factorization algorithm are proposed for this purpose. The proposed robust formulation aims at joint latent factor identification and corrupt receiver (report) detection. Identifiability of the latent factors is theoretically established for this more challenging setup. An attractive feature of the proposed robust factorization algorithm is that it is easily implementable. Judicious simulations as well as real experiments with a software radio prototype using Universal Software Radio Peripheral (USRP) radios by Ettus Research (<http://www.ettus.com>) are used to demonstrate the effectiveness of the proposed approach. Both the simulations and the laboratory experiment show that the proposed approach is promising in dealing with the problem of interest.

A preliminary version of part of this work was presented at ICASSP 2014 [1]. The ICASSP 2014 paper only considered a special case (corresponding to ESPRIT, rather than full-fledged PARAFAC), and did not include the robust formulation. This journal version brings in a more general and flexible problem setup, including the robust formulation and associated identifiability proof and algorithm, extensive simulations, plus laboratory experiments. A related but distinct approach was considered in [17]. There, we considered latent spectra identification for the case of synchronized but spatially uncalibrated single-antenna receivers, which is different from the setup here and in [1].

*Notation:*  $T$ ,  $H$  and  $*$  denote transpose, Hermitian transpose and conjugate, respectively;  $\dagger$  denotes the Moore-Penrose pseudoinverse;  $\otimes$ ,  $\odot$ , and  $\circ$  represent the Kronecker product, the Khatri-Rao product, and the Hadamard product, respectively;  $\text{rank}(\cdot)$  and  $\text{krank}(\cdot)$  are the matrix rank and Kruskal rank, respectively;  $[\mathbf{X}]_{m,:}$  and  $[\mathbf{X}]_{:,n}$  represent the  $m$ -th row and the  $n$ -th column of matrix  $\mathbf{X}$ , respectively;  $\mathbf{X}(\mathcal{Y}, :)$  denotes the submatrix of  $\mathbf{X}$  consisting of the rows indexed by  $\mathcal{Y}$ ;  $\|\mathbf{x}\|_p$  and  $\|\mathbf{X}\|_F$  denote the vector  $p$ -norm ( $p \geq 1$ ) and the matrix Frobenius norm, respectively;  $\|\mathbf{X}\|_{\text{col}=0}$  denotes a columnwise zero-norm which counts the number of non-zero columns of  $\mathbf{X}$ ;  $\text{vec}(\mathbf{X})$  denotes the operator that concatenates the columns of  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$  such that  $\text{vec}(\mathbf{X}) = [\mathbf{x}_1^T, \dots, \mathbf{x}_n^T]^T$ ;  $\text{Diag}(x_1, \dots, x_n)$  denotes a diagonal matrix with  $x_1, \dots, x_n$  as its diagonal elements;  $\mathcal{R}(\mathbf{X})$  denotes the range space of  $\mathbf{X}$ .

## II. BASIC SETUP AND SIGNAL MODEL

Consider a scenario where there are  $K$  primary or secondary transmitters (or, sources for brevity) in the far field of a network of  $N$  cognitive radio receivers (sensors), each equipped with  $M$  receive antennas and down-conversion chains. Let  $x_k(t) \in \mathbb{C}$ , for  $t = 0, 1, \dots$ , denote the  $k$ -th transmitted signal. We assume that  $x_k(t)$  is wide-sense stationary (WSS) for  $k = 1, \dots, K$ , and its bandwidth is relatively narrow compared to its carrier frequency so that a flat fading channel model can be adopted (e.g., typically, a signal with less than 20 MHz bandwidth in the 2–5 GHz band can be considered as narrowband; otherwise we may apply our approach to sub-bands using a bank of band-pass filters). We also assume line-of-sight (LOS) propagation, which is a reasonable model in several scenarios of practical interest—see [18] for an overview of propagation models in cognitive radio. Beyond cognitive radio, spectra separation and emitter localization is of interest in airborne search-and-rescue, law enforcement, and military operations, where surveillance receivers are mounted on airplanes or drones with a clear LOS to the ground. We further assume that the array baselines of the receivers are aligned, e.g., using a compass, so that the source signals impinge from the same directions on all receivers. This configuration is illustrated in Fig. 2. We assume that the down-conversion chains at each receiver have been synchronized—this can be easily done locally at each receiver. The received signals at the  $n$ th receiver can then be expressed as

$$\mathbf{y}_n(t) = \mathbf{\Phi}_n \mathbf{A}_n \mathbf{x}(t) + \mathbf{v}_n(t), \quad t = 0, 1, 2, \dots, \quad (1)$$

where  $\mathbf{y}_n(t) = [y_{n,1}(t), \dots, y_{n,M}(t)]^T \in \mathbb{C}^M$ ,  $\mathbf{x}(t) = [x_1(t), \dots, x_K(t)]^T \in \mathbb{C}^K$  denotes the vector of transmitted signals,  $\mathbf{v}_n(t) = [v_{n,1}(t), \dots, v_{n,M}(t)]^T \in \mathbb{C}^M$  denotes the corresponding noise vector,  $\mathbf{A}_n = \text{Diag}(a_{n,1}, \dots, a_{n,K})$  is a diagonal matrix whose  $k$ th diagonal element  $a_{n,k} \in \mathbb{C}$  contains the path loss and phase shift from transmitter  $k$  to receiver  $n$ , and  $\mathbf{\Phi}_n \in \mathbb{C}^{M \times K}$  is a matrix which is generally related to the geometry of the antenna placement and the DOAs of the received wavefronts at sensor  $n$ ; e.g., assuming that a uniform linear array (ULA) is employed at each receiver,

$$\mathbf{\Phi}_n = \begin{bmatrix} 1, & \dots, & 1 \\ e^{j\phi_1}, & \dots, & e^{j\phi_K} \\ \vdots & \vdots & \vdots \\ e^{j(M-1)\phi_1}, & \dots, & e^{j(M-1)\phi_K} \end{bmatrix},$$

where  $\phi_k = -2\pi d \sin(\theta_k)/\lambda$ ,  $\theta_k \in [-\pi, \pi]$  is the DOA of source  $k$  that is the same at different sensors under the aligned baselines model,  $\lambda$  is the wavelength corresponding to the carrier frequency, and  $d$  denotes the distance between neighboring receive antennas. Notice that by aligning the receiver baselines,  $\mathbf{\Phi}_1, \dots, \mathbf{\Phi}_N$  are identical (cf. Fig. 2). Under the signal model (1), our objective is to estimate  $\theta_1, \dots, \theta_K$  and the power spectra of  $x_1(t), \dots, x_K(t)$  simultaneously.

In practice, there will be small propagation delay differences depending on the particular transmitter and receiver considered. Thus, to be more accurate,  $\mathbf{x}(t)$  should be replaced by  $\mathbf{x}_n(t) = [x_1(t - \delta_{1,n}), \dots, x_K(t - \delta_{K,n})]^T$  in the above equation. However, the effect of  $\{\delta_{k,n}\}$  will disappear when we cross-correlate the outputs of any given receiver, by virtue of WSS and the fact

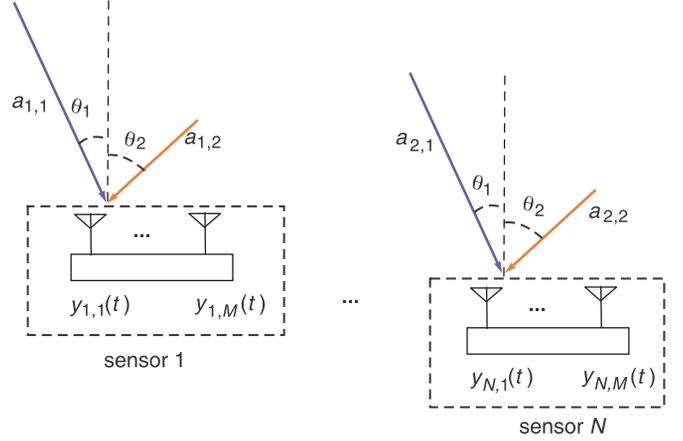


Fig. 2. Illustration of the scenario. The sensors are placed in a relatively far field of the sources such that the DOAs of the sources at different sensors can be considered identical.

that different transmissions are uncorrelated. We thus suppress these delays for brevity.

## III. PROPOSED APPROACH

To estimate the DOAs and the power spectra of the sources, we propose to formulate the problem in the temporal correlation domain. Assume that there are two antennas that are located at the positions  $rd$  and  $sd$  of the baseline of receiver  $n$ , where  $r, s \in \mathbb{Z}_+$  are integers—e.g., for a ULA, we have  $r, s \in \{0, 1, \dots, M-1\}$ . Let us define  $q = r - s$ , and let  $m(r)$  and  $m(s)$  for  $1 \leq m(r), m(s) \leq M$  be the indices of the antennas located at  $rd$  and  $sd$ , respectively. Then, by cross-correlating the received signals at these two antennas at receiver  $n$ , we obtain

$$c_n(q, \tau) = \mathbb{E} \left\{ y_{n,m(r)}(t) y_{n,m(s)}^*(t - \tau) \right\}, \quad (2)$$

where  $\tau \in \mathbb{Z}$  is the index of time lag and  $y_{n,m(r)}(t) = \sum_{k=1}^K a_{n,k} e^{j r \phi_k} x_k(t) + v_{n,m(r)}(t)$  under our signal model. Since  $x_k(t)$ 's are uncorrelated to each other, and the signal and the noise components are also uncorrelated, we obtain

$$c_n(q, \tau) = \sum_{k=1}^K |a_{n,k}|^2 e^{j q \phi_k} r_k(\tau) + \mathbb{E} \{ v_{n,m(r)}(t) v_{n,m(s)}^*(t - \tau) \},$$

where  $r_k(\tau) = \mathbb{E} \{ x_k(t) x_k^*(t - \tau) \}$ . Assume that the noise  $v_{n,m}(t)$  at each antenna of sensor  $n$  is white Gaussian, both temporally and spatially, with zero mean and variance  $\sigma_n^2$ , i.e.,  $v_{n,m}(t) \sim \mathcal{CN}(0, \sigma_n^2)$  for  $m = 1, \dots, M$ . Then, we have

$$\mathbb{E} \{ v_{n,m(r)}(t) v_{n,m(s)}^*(t - \tau) \} = \sigma_n^2 \delta(\tau) \delta(q), \quad (3)$$

where  $\delta(x)$  denotes the Kronecker delta, i.e.,  $\delta(x) = 1$  when  $x = 0$  and  $\delta(x) = 0$  otherwise. Thus, we can compactly express  $c_{n,q}(\tau)$  as

$$c_n(q, \tau) = \sum_{k=1}^K |a_{n,k}|^2 r_k(\tau) + \sigma_n^2 \delta(\tau) \delta(q). \quad (4)$$

Now, we apply the discrete-time Fourier transform (DTFT) on  $\{c_n(q, \tau)\}_{\tau=-\infty}^{\infty}$  so that we have

$$\begin{aligned} C_n(q, \omega) &= \sum_{\tau=-\infty}^{\infty} c_n(q, \tau) e^{-j\omega\tau} \\ &= \sum_{k=1}^K |a_{n,k}|^2 e^{-jq\phi_k} S_k(\omega) + \sigma_n^2 \delta(q). \end{aligned} \quad (5)$$

The frequency axis of  $C_n(q, \omega)$  can be discretized to  $F$  samples, and for this purpose let us denote  $G_n(q, f) = C_n(q, \frac{2\pi(f-1)}{F})$ , where  $f = 1, \dots, F$ ; the  $G_n(q, f)$ 's can be readily obtained using the discrete Fourier transform (DFT).

Suppose that there are  $L$  different  $q$ 's that result from different combinations of  $r$  and  $s$  and are denoted by  $q_1, \dots, q_L$  in ascending order, i.e.,  $q_1 < q_2 < \dots < q_L$ . Given the obtained  $G_n(q_\ell, f)$  for  $\ell = 1, \dots, L$ , we construct  $\mathbf{G}_\ell \in \mathbb{C}^{N \times F}$  for  $\ell = 1, \dots, L$  such that

$$[\mathbf{G}_\ell]_{n,f} = G_n(q_\ell, f), \quad f = 1, \dots, F, \quad (6)$$

It is readily shown that one can compactly express  $\mathbf{G}_\ell$  as

$$\mathbf{G}_\ell = \mathbf{B} \mathbf{D}_\ell(\Phi) \mathbf{S}^T + \boldsymbol{\eta} \mathbf{1}^T \delta(q_\ell), \quad (7)$$

where  $\mathbf{B} \in \mathbb{R}^{N \times K}$  such that  $[\mathbf{B}]_{n,k} = |a_{n,k}|^2$ ,  $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_K] \in \mathbb{R}^{F \times K}$ ,  $\mathbf{s}_k = [S_k(0), \dots, S_k(\frac{2\pi(F-1)}{F})]^T$  denotes the discretized power spectrum of source  $k$ ,  $\boldsymbol{\eta} = [\sigma_1^2, \dots, \sigma_N^2]^T$  holds all the noise variances at different sensors,  $\Phi = [\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_K] \in \mathbb{C}^{L \times K}$ ,  $\boldsymbol{\phi}_k = [e^{jq_1\phi_k}, \dots, e^{jq_L\phi_k}]^T \in \mathbb{C}^{L \times 1}$ , and

$$\mathbf{D}_\ell(\Phi) = \text{Diag}(e^{jq_\ell\phi_1}, \dots, e^{jq_\ell\phi_K})$$

denotes a diagonal matrix with the  $\ell$ th row of  $\Phi$  on its diagonal. Notice that  $\Phi$  contains all the DOA information on the exponents of its row elements, and that  $\mathbf{S}$  contains the desired power spectra of the sources as its columns. Hence, our problem amounts to estimating  $\mathbf{S}$  and  $\Phi$  from  $\{\mathbf{G}_\ell\}_{\ell=1}^L$ .

One should notice that  $\mathbf{G}_\ell$  is corrupted by the noise term  $\boldsymbol{\eta} \mathbf{1}^T$  when  $q_\ell = 0$ . In practice, there are many ways to (approximately) cancel the impact of  $\boldsymbol{\eta}$ . For instance,  $\sigma_n^2$  can sometimes be estimated in advance (when the transmitters are silent) and thus canceled [19]. Another approach can be found in [1], where we have provided an orthogonal complement projection-based noise removal method, which is applicable to the case where  $N \geq K$ . Here, we instead opt for a simple way of estimating  $\boldsymbol{\eta}$ , which is effective according to our simulations and real experiments. Specifically, we assume that *there exists a frequency  $f_0$ , such that  $[\mathbf{S}]_{f_0,:} = \mathbf{0}$* . The above assumption is based on the observation that the band of interest is usually not fully occupied by the transmitters—for spectrum sensing, this assumption is often mild. Consequently, the noise term can be estimated by

$$\hat{\boldsymbol{\eta}} = [\mathbf{G}_{\ell_0}]_{:,f_0},$$

where  $\ell_0 \in \{1, \dots, L\}$  satisfies  $q_{\ell_0} = 0$  and

$$f_0 = \arg \min_{f=1, \dots, F} \|[\mathbf{G}_{\ell_0}]_{:,f}\|_2,$$

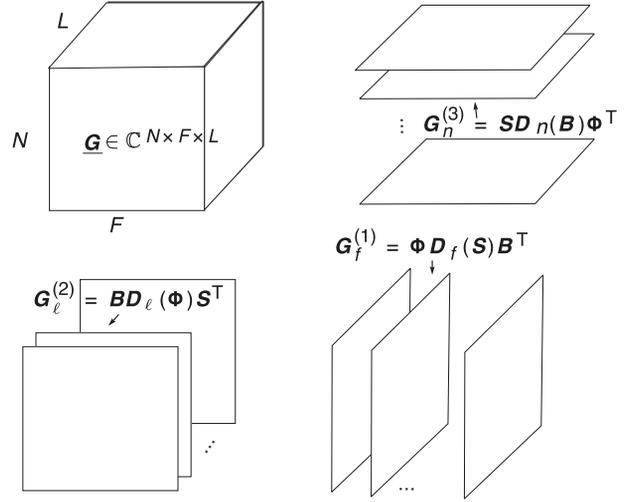


Fig. 3. Illustration of the slice representations of the three-way tensor  $\underline{\mathbf{G}}$ .

since  $\mathbf{G}_{\ell_0} = \mathbf{B} \mathbf{S}^T + \boldsymbol{\eta} \mathbf{1}^T$ , which is non-negative and the minimum-norm columns correspond to the noise floor. Hence, by subtracting  $\hat{\boldsymbol{\eta}} \mathbf{1}^T$  from  $\mathbf{G}_{\ell_0}$ , i.e., by letting

$$\mathbf{G}_{\ell_0} := \mathbf{G}_{\ell_0} - \hat{\boldsymbol{\eta}} \mathbf{1}^T,$$

we obtain a set of matrices

$$\{\mathbf{G}_\ell = \mathbf{B} \mathbf{D}_\ell(\Phi) \mathbf{S}^T\}_{\ell=1}^L. \quad (8)$$

In the following, we will assume that the noise has been removed and shall focus on the noiseless model in (8).

#### IV. PARALLEL FACTOR ANALYSIS IN TEMPORAL CORRELATION-DOMAIN

One may notice that (8) is a *slice/slab representation* of a three-way tensor [20], [21]. Specifically, let us define a tensor  $\underline{\mathbf{G}} \in \mathbb{C}^{N \times F \times L}$ , whose  $(n, f, \ell)$ -th element is defined by

$$\begin{aligned} [\underline{\mathbf{G}}]_{n,f,\ell} &= [\mathbf{G}_\ell]_{n,f} = \sum_{k=1}^K |a_{n,k}|^2 e^{jq_\ell\phi_k} S_k \left( \frac{2\pi(f-1)}{F} \right), \\ &= \sum_{k=1}^K [\mathbf{B}]_{n,k} [\Phi]_{\ell,k} [\mathbf{S}]_{f,k}. \end{aligned}$$

The three slice representations of  $\underline{\mathbf{G}}$  are [20]:

$$\mathbf{G}_f^{(1)} = [\underline{\mathbf{G}}]_{:,f,:} = \Phi \mathbf{D}_f(\mathbf{S}) \mathbf{B}^T, \quad (9a)$$

$$\mathbf{G}_\ell^{(2)} = [\underline{\mathbf{G}}]_{:,:,\ell} = \mathbf{B} \mathbf{D}_\ell(\Phi) \mathbf{S}^T, \quad (9b)$$

$$\mathbf{G}_n^{(3)} = [\underline{\mathbf{G}}]_{n,:,:} = \mathbf{S} \mathbf{D}_n(\mathbf{B}) \Phi^T; \quad (9c)$$

see Fig. 3 for an illustration. Note that sometimes it is convenient to represent a tensor using its *matrix unfoldings*, which are defined as

$$\underline{\mathbf{G}}^{(1)} = [\text{vec}(\mathbf{G}_1^{(1)}), \dots, \text{vec}(\mathbf{G}_F^{(1)})], \quad (10a)$$

$$\underline{\mathbf{G}}^{(2)} = [\text{vec}(\mathbf{G}_1^{(2)}), \dots, \text{vec}(\mathbf{G}_L^{(2)})], \quad (10b)$$

$$\underline{\mathbf{G}}^{(3)} = [\text{vec}(\mathbf{G}_1^{(3)}), \dots, \text{vec}(\mathbf{G}_N^{(3)})]. \quad (10c)$$

From (9), (10), and the vectorization property, i.e.,

$$\text{vec}(\mathbf{A}\text{Diag}(\mathbf{b})\mathbf{C}^T) = (\mathbf{C} \odot \mathbf{A})\mathbf{b},$$

we obtain the following compact expressions

$$\underline{\mathbf{G}}^{(1)} = (\mathbf{B} \odot \Phi) \mathbf{S}^T \quad (11a)$$

$$\underline{\mathbf{G}}^{(2)} = (\mathbf{S} \odot \mathbf{B}) \Phi^T \quad (11b)$$

$$\underline{\mathbf{G}}^{(3)} = (\Phi \odot \mathbf{S}) \mathbf{B}^T \quad (11c)$$

From the above, we see that the collected data  $\{\mathbf{G}_\ell\}_{\ell=1}^L$  is identical to the second slice representation in (9b). Hence, extracting  $\mathbf{B}$ ,  $\mathbf{S}$  and  $\Phi$  from the tensor  $\{\mathbf{G}_\ell\}_{\ell=1}^L$  is a rank decomposition problem commonly known as PARAllel FACtor (PARAFAC) analysis [19], [22]–[24]. A remarkable property of tensors is that this decomposition is essentially unique, under rather mild conditions. For example, note that deliberately picking  $q_1, \dots, q_L$  to be consecutive integers<sup>1</sup> induces Vandermonde structure on  $\Phi$ , in which case the following theorem applies:

**Theorem 1** [24]: *Assume that  $\Phi \in \mathbb{C}^{L \times K}$  is a Vandermonde matrix, and that*

$$\text{krank}(\mathbf{B}) + \min\{L + \text{krank}(\mathbf{S}), 2K\} \geq 2K + 2. \quad (12)$$

*Then, the tensor  $\underline{\mathbf{G}} \in \mathbb{C}^{N \times F \times L}$  with elements  $[\underline{\mathbf{G}}]_{n,f,\ell} = \sum_{k=1}^K [\mathbf{B}]_{n,k} [\Phi]_{\ell,k} [\mathbf{S}]_{f,k}$  admits unique decomposition in  $K$  rank-one components, i.e., the loading matrices  $\mathbf{B}$ ,  $\mathbf{S}$  and  $\Phi$  are unique up to scaling and a common column permutation (or, simply, essentially unique). Specifically, if  $\tilde{\mathbf{B}}$ ,  $\tilde{\mathbf{S}}$  and  $\tilde{\Phi}$  generate  $\underline{\mathbf{G}}$ , then it must hold that*

$$\tilde{\mathbf{B}} = \mathbf{B}\mathbf{\Pi}\mathbf{\Sigma}_1, \quad \tilde{\Phi} = \Phi\mathbf{\Pi}\mathbf{\Sigma}_2, \quad \tilde{\mathbf{S}} = \mathbf{S}\mathbf{\Pi}\mathbf{\Sigma}_3,$$

*where  $\mathbf{\Pi}$  is a column permutation matrix, and  $\mathbf{\Sigma}_i \in \mathbb{C}^{K \times K}$  are diagonal scaling matrices satisfying*

$$\mathbf{\Sigma}_1 \mathbf{\Sigma}_2 \mathbf{\Sigma}_3 = \mathbf{I}.$$

Theorem 1 means that  $\mathbf{B}$ ,  $\mathbf{S}$  and  $\Phi$  are identifiable from  $\underline{\mathbf{G}}$ —notably,  $\mathbf{B}$ ,  $\mathbf{S}$  and  $\Phi$  do not even need to be full-column rank matrices, which means that we can handle the case of more transmitters than receivers, for example. Even stronger almost-sure identifiability results are available—if  $\mathbf{B}$ ,  $\mathbf{S}$  and  $\Phi$  are drawn from a continuous joint distribution, see [25]–[27].

### A. Decomposition Algorithms

To identify the loading factors, many PARAFAC algorithms can be applied, e.g., those in [19], [25], [28], [29]. The so-called *trilinear alternating least squares* (TALS) algorithm is particularly appealing in our context, because it can readily incorporate constraints such as nonnegativity to improve estimation accuracy. In the next section, we will propose a new PARAFAC algorithm that is robust to corrupt data slabs, which can be considered as an extension of TALS.

As a special case of the above PARAFAC model, when  $\text{rank}(\mathbf{B}) = \text{rank}(\mathbf{S}) = K$  and  $\theta_1, \dots, \theta_K$  are distinct, the

<sup>1</sup>As an example, suppose that each sensor uses a ULA with  $M = 3$ . We see that  $0 \leq r, s \leq 2$  and thus  $r - s$  can form a set of  $q_\ell$ 's such that  $q_\ell \in \{-2, -1, 0, 1, 2\}$ . For specially-designed non-uniform antenna arrays, a larger set of consecutive differences can be obtained—see Remark 3.

classic *Estimation of Signal Parameters by Rotational Invariance Techniques* (ESPRIT) algorithm [30] can be applied to estimate  $\mathbf{S}$  and  $\theta_1, \dots, \theta_K$ . To implement ESPRIT, one may construct

$$\tilde{\mathbf{G}}_1 = \begin{bmatrix} \mathbf{G}_1 \\ \vdots \\ \mathbf{G}_{L-1} \end{bmatrix} = \begin{bmatrix} \mathbf{B}\mathbf{D}_1(\Phi) \\ \vdots \\ \mathbf{B}\mathbf{D}_{L-1}(\Phi) \end{bmatrix} \mathbf{S}^T = \tilde{\mathbf{B}}\mathbf{S}^T, \quad (13)$$

and

$$\tilde{\mathbf{G}}_2 = \begin{bmatrix} \mathbf{G}_2 \\ \vdots \\ \mathbf{G}_L \end{bmatrix} = \tilde{\mathbf{B}}\mathbf{D}\mathbf{S}^T, \quad (14)$$

where  $\mathbf{D} = \text{Diag}(e^{j\phi_1}, \dots, e^{j\phi_K})$ . The signal model in (13)–(14) can be handled by ESPRIT, and the desired parameters can be estimated using a single eigen-decomposition; see [1], [30]. Compared to TALS, the computational burden of ESPRIT is much lighter, which is a desired property in practice, particularly for real-time applications. Also, ESPRIT can be used to initialize TALS, when applicable.

### B. Remarks on the Proposed Approach

Before turning to the next section, we would like to make several remarks on the proposed PARAFAC-based formulation:

**Remark 1:** The proposed formulation does not require synchronization among the receivers; only local synchronization of the down-conversion chains of each individual multi-antenna receiver is required. To see this, let us assume that sensor  $n$  begins to work at time index  $t_n$  for  $n = 1, \dots, N$ , and that  $t_1 \neq \dots \neq t_N$ . As the down-conversion chains at sensor  $n$  are synchronized, antennas  $m(r)$  and  $m(s)$  work with the same time index  $t_n$ . Hence, one can see that the building block of the proposed formulation, i.e.,  $c_n(q, \tau)$ , is insensitive to the value of  $t_n$ —we have

$$c_n(q, \tau) = \mathbb{E} \left\{ y_{n,m(r)}(t_n) y_{n,m(s)}^*(t_n - \tau) \right\},$$

which is the same as in (6), by virtue of WSS. Notice that synchronization among the different receivers could entail considerable communication overhead and extra computations.

**Remark 2:** Those familiar with array processing and blind source separation (BSS) may readily think about using BSS approaches (e.g., *multiple invariance sensor array processing* (MI-SAP) [22]) to separate the raw signals, then estimate the power spectrum of each separated signal. This approach has several drawbacks relative to the one we propose here. First, it requires synchronization across the different multi-antenna receivers; second, it entails much higher communication overhead to send raw signals (as opposed to *locally computed* correlations) to the fusion center. Furthermore, the temporal correlation parametrization is naturally suited for estimating power spectra, and it comes with the side-bonus of non-negative  $\mathbf{B}$  and  $\mathbf{S}$ , properties that can be exploited to enhance identifiability and separation performance.

By noticing that  $\mathbf{G}_{\ell_0} = \mathbf{B}\mathbf{S}^T$ , it is also tempting to apply non-negative matrix factorization (NMF) to estimate  $\mathbf{S}$ , i.e.,

$$\hat{\mathbf{S}} = \arg \min_{\mathbf{B} \geq \mathbf{0}, \mathbf{S} \geq \mathbf{0}} \|\mathbf{G}_{\ell_0} - \mathbf{B}\mathbf{S}^T\|_F^2, \quad (15)$$

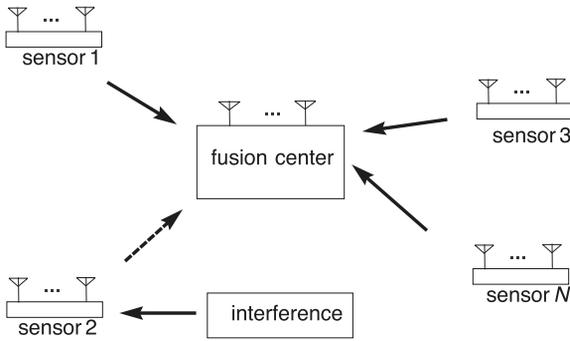


Fig. 4. A scenario where receiver 2 suffers from local interference and gives highly corrupted data to the fusion center.

since both  $\mathbf{B}$  and  $\mathbf{S}$  are non-negative. The above NMF formulation is similar to the idea in [14]. Nonetheless, NMF does not yield a unique solution in general [31], and thus the identifiability of  $\mathbf{S}$  is not guaranteed from the NMF formulation.

**Remark 3:** So far, we have implicitly focused on using the same ULA at each receiver. The number of antennas and the array geometry come into play *indirectly*, in that they determine the number of rows of the Vandermonde matrix  $\Phi \in \mathbb{C}^{L \times K}$ ; the higher  $L$  is, the higher the number of co-channel spectra that can be successfully separated (cf. Theorem 1). It is important to stress that  $L$  is *not* the number of antennas,  $M$ , but the number of distinct spatial correlation lags that can be estimated from the antenna array outputs, i.e., the number of distinct differences  $q = r - s$ , where  $r$  and  $s$  are associated with the antenna deployment geometry as described before. For example,  $r$  and  $s$  are consecutive integers from 0 to  $M - 1$  when a ULA is employed at each receiver, and thus  $L = 2M - 1$  in this case. The important observation here is that using non-uniform linear arrays, such as *two-level nested arrays* [32],  $r$  and  $s$  can be chosen from a set of non-consecutive integers with special structure which yields  $L = \mathcal{O}(M^2)$  while preserving the Vandermonde structure of  $\Phi$ . Another very interesting possibility is to use a so-called *sparse ruler* design [9].

## V. JOINT MODEL FITTING AND DETECTION OF CORRUPT DATA

In practice, there are scenarios where it is likely that some receivers will be reporting highly corrupted data to the fusion center. This can happen if a receiver is shadowed (in which case the LOS assumption is grossly violated), or subject to significant local (e.g., microwave) interference or no-listen-while-you-talk limitations—see [33]–[36] for further motivation and Fig. 4 for an illustration. In such a situation, directly applying TALS or other generic PARAFAC solvers as proposed in the last section may yield unexpectedly bad solutions, since corrupt data are *outliers* that inflict large deviations from the assumed PARAFAC model. The goal of this section is to robustify the formulation proposed in the previous section, and to devise a robust algorithm that can identify the desired latent variables and the receivers which are reporting corrupt data (or, *outlier receivers* for simplicity) simultaneously.

### A. The Proposed Formulation

Let us begin with the slice representations of three-way tensors, i.e., (9a)–(9c). Recall that, under the proposed formulation,

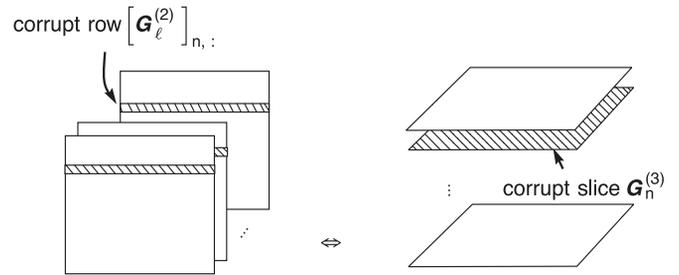


Fig. 5. The shadowed rows on the left constitute the physically corrupted data contributed by receiver 2.  $\mathbf{G}_n^{(3)}$  on the right is the corresponding corrupted slice, comprising the said rows.

$\{\mathbf{G}_\ell^{(2)}\}_{\ell=1}^L$  are the collected data, and  $[\mathbf{G}_\ell^{(2)}]_{n,:}$  is solely contributed by sensor  $n$  for  $\ell = 1, \dots, L$ . The key observation here is that  $\{[\mathbf{G}_\ell^{(2)}]_{n,:}\}_{\ell=1}^L$  form the whole slice  $\mathbf{G}_n^{(3)}$ . Hence, if the  $n$ -th receiver fails to provide reliable data, the whole slice  $\mathbf{G}_n^{(3)}$  is a corrupted slice (cf. Fig. 5). This observation will be exploited in the sequel.

Let  $\mathcal{A}$  denote the index set of the receivers reporting corrupt data, where  $\mathcal{A} \subset \mathcal{N}$  and  $\mathcal{N} = \{1, \dots, N\}$ ; we also denote the complement set as  $\mathcal{A}_c$  such that  $\mathcal{A}_c = \mathcal{N}/\mathcal{A}$ . By modeling outliers using an additive noise term, we have the following model,

$$\mathbf{G}_n^{(3)} = \mathbf{S}\mathbf{D}_n(\mathbf{B})\Phi^T + \mathbf{N}_n, \quad n = 1, \dots, N, \quad (16)$$

where  $\mathbf{N}_n \in \mathbb{C}^{L \times F}$  is the local interference to receiver  $n$  such that

$$\begin{cases} \mathbf{N}_n \neq \mathbf{0}, & n \in \mathcal{A}, \\ \mathbf{N}_n = \mathbf{0}, & n \notin \mathcal{A}. \end{cases} \quad (17)$$

Under this model,  $\{\mathbf{G}_n^{(3)}\}_{n \in \mathcal{A}}$  forms a set of *slice outliers* which do not obey the signal model (9c). Recall that we have

$$\underline{\mathbf{G}}^{(3)} = \left[ \text{vec}(\mathbf{G}_1^{(3)}), \dots, \text{vec}(\mathbf{G}_N^{(3)}) \right].$$

Hence, under the signal model in (16), we have

$$\underline{\mathbf{G}}^{(3)} = (\Phi \odot \mathbf{S})\mathbf{B}^T + \mathbf{N}, \quad (18)$$

where  $\mathbf{N} = [\mathbf{n}_1, \dots, \mathbf{n}_N]$  and  $\mathbf{n}_n = \text{vec}(\mathbf{N}_n)$  is a column sparse matrix. We propose to estimate  $\mathbf{S}$  and  $\Phi$  by solving the following fitting problem,

**Corrupt Slices–Robust PARAFAC Criterion :**

$$\begin{aligned} \min_{\mathbf{Q}, \mathbf{S}, \mathbf{B}, \Phi} & \quad \left\| \underline{\mathbf{G}}^{(3)} - (\Phi \odot \mathbf{S})\mathbf{B}^T - \mathbf{Q} \right\|_F^2 \\ \text{s.t.} & \quad \|\mathbf{Q}\|_{\text{col-0}} \leq P, \end{aligned} \quad (19)$$

where  $P \in \mathbb{Z}_+$  is a pre-defined parameter, indicating an upper bound on the number of outlier receivers—this upper bound can be rather loose in practice, as we will see. The idea is to use a column-sparse  $\mathbf{Q}$  to cancel the corrupted slices; if the remaining slices still comprise an identifiable PARAFAC model, then we can perhaps hope for ‘oracle’ estimation performance from the proposed strategy. We should remark that  $\mathbf{Q}$  is not necessarily an estimate of  $\mathbf{N}$ —it is an optimization variable that is adopted to cancel the impact of the corrupted slices. In fact, the formulated criterion in (19) can accommodate different types of interference, not necessarily restricted to additive interference; e.g.,

even when the corrupted slices do not have any signal component, the formulation in (19) is still applicable.

At this point, we should mention that the idea of modeling outliers as a column-sparse matrix was previously presented in [37] for subspace identification as a heuristic. Here, we take one step further—we prove that (19) is non-heuristic under our formulation from an identifiability perspective:

**Theorem 2:** *Let  $(\Phi^*, \mathbf{S}^*, \mathbf{B}^*, \mathbf{Q}^*)$  be an optimal solution to Problem (19), and  $\mathcal{A}^*$  and  $\mathcal{A}_c^*$  denote the index set of the non-zero columns of  $\mathbf{Q}^*$  and its complement, respectively. Assume that  $|\mathcal{A}| \leq P$ , that the elements of  $\mathbf{B}$  are drawn from some absolutely continuous distribution, that*

$$\mathcal{R}(\mathbf{N}) \cap \mathcal{R}(\Phi \odot \mathbf{S}) = \{\mathbf{0}\}, \quad (20)$$

and that

$$\min\{N - 2P, K\} + \min\{L + \text{krank}(\mathbf{S}), 2K\} \geq 2K + 2. \quad (21)$$

Then, under the signal model (16), the following hold with probability one:

- 1) The latent variables  $\Phi^*$ ,  $\mathbf{S}^*$  and  $\mathbf{B}^*(\mathcal{A}_c^*, :)$  are identical to  $\Phi$ ,  $\mathbf{S}$  and  $\mathbf{B}(\mathcal{A}_c^*, :)$  up to permutation and scaling ambiguities, respectively; i.e., we have  $\mathbf{B}^*(\mathcal{A}_c^*, :) = \mathbf{B}(\mathcal{A}_c^*, :)\mathbf{\Pi}\Sigma_1$ ,  $\Phi^* = \Phi\mathbf{\Pi}\Sigma_2$ , and  $\mathbf{S}^* = \mathbf{S}\mathbf{\Pi}\Sigma_3$ , where  $\mathbf{\Pi}$  and  $\Sigma_i$  are defined as before.
- 2) The corrupt slices can be detected, i.e.,  $\mathcal{A} \subseteq \mathcal{A}^*$ .

The proof of Theorem 2 is relegated to Appendix A. Notice that the conditions stated in the above theorem are in general mild: The matrix  $\mathbf{B}$  can be considered following the Chi-squared distribution in wireless communications as the matrix  $\mathbf{A}$  is usually modeled as being circularly symmetric Gaussian distributed [38]; also, random interferences can hardly reside in the range space of  $\Phi \odot \mathbf{S}$ . Hence, by Theorem 2, Problem (19) is a sound criterion for joint model fitting and outliers detection; that is, if the number of receivers reporting uncorrupted data is large enough, the power spectra, the DOAs, and the outlier receivers can be identified via solving Problem (19).

### B. Corrupt Slices-Robust Alternating Least Squares

Problem (19) is a hard optimization problem. To tackle it, we propose a block coordinate descent (BCD)-based algorithm—we alternate between solving the partial problems w.r.t. each of  $\mathbf{Q}$ ,  $\mathbf{S}$ ,  $\mathbf{B}$ ,  $\Phi$  while keeping the others fixed. We begin with the partial minimization w.r.t.  $\mathbf{Q}$ :

$$\min_{\|\mathbf{Q}\|_{\text{col-0}} \leq P} \left\| \left( \underline{\mathbf{G}}^{(3)} - (\Phi \odot \mathbf{S})\mathbf{B}^T \right) - \mathbf{Q} \right\|_F^2,$$

The above aims at finding the projection of  $\underline{\mathbf{G}}^{(3)} - (\Phi \odot \mathbf{S})\mathbf{B}^T$  onto a column sparse set. This problem is non-convex, but a closed-form solution exists as reported in [37]. To be specific, denote

$$\zeta_n = \left[ \underline{\mathbf{G}}^{(3)} - (\Phi \odot \mathbf{S})\mathbf{B}^T \right]_{:,n}, \quad n = 1, \dots, N.$$

Then, the solution  $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_N]$  is constructed with the following columns:

$$\mathbf{q}_n := \begin{cases} \zeta_n, & n \in \{n_1, \dots, n_P\}, \\ \mathbf{0}, & n \in \mathcal{N}/\{n_1, \dots, n_P\}, \end{cases} \quad (22)$$

### Algorithm 1: Corrupt Slices-Robust Alternating Least Squares (CSR-ALS)

---

**input** :  $\{\underline{\mathbf{G}}_\ell\}_{\ell=1}^L$ ; initialization:  $(\mathbf{B}, \Phi, \mathbf{S})$ ;  $P$ ;  $\epsilon > 0$ .

- 1  $\underline{\mathbf{G}}^{(3)} = \left[ \text{vec} \left( \underline{\mathbf{G}}_1^{(3)} \right), \dots, \text{vec} \left( \underline{\mathbf{G}}_N^{(3)} \right) \right]$ ;
- 2  $i = 0$ ;
- 3  $f^{(0)} = \infty$ ;
- 4 **repeat**
- 5      $i = i + 1$ ;
- 6      $\zeta_n := \left[ \underline{\mathbf{G}}^{(3)} - (\Phi \odot \mathbf{S})\mathbf{B}^T \right]_{:,n}$ , for  $n = 1, \dots, N$ ;
- 7     sort  $q_1, \dots, q_N$  such that  $\|\zeta_{n_1}\|_2 \geq \dots \geq \|\zeta_{q_N}\|_2$ ;
- 8      $\mathbf{Q} := [\mathbf{q}_1, \dots, \mathbf{q}_N]$  such that
 
$$\mathbf{q}_n := \begin{cases} \zeta_n, & n \in \{q_1, \dots, q_P\}, \\ \mathbf{0}, & n \in \mathcal{N}/\{q_1, \dots, q_P\}; \end{cases}$$
- 9      $\underline{\mathbf{E}}^{(3)} := \underline{\mathbf{G}}^{(3)} - \mathbf{Q}$ ;
- 10     obtain  $\underline{\mathbf{E}}^{(1)}$  and  $\underline{\mathbf{E}}^{(2)}$  by Eq. (26);
- 11      $\mathbf{S} := \left( (\mathbf{B} \odot \Phi)^\dagger \underline{\mathbf{E}}^{(1)} \right)^T$ ;
- 12      $\Phi := \left( (\mathbf{S} \odot \mathbf{B})^\dagger \underline{\mathbf{E}}^{(2)} \right)^T$ ;
- 13      $\mathbf{B} := \left( (\Phi \odot \mathbf{S})^\dagger \underline{\mathbf{E}}^{(3)} \right)^T$ ;
- 14      $f^{(i)} := \left\| \left( \underline{\mathbf{G}}^{(3)} - (\Phi \odot \mathbf{S})\mathbf{B}^T \right) - \mathbf{Q} \right\|_F^2$ ;
- 15 **until**  $|f^{(i)} - f^{(i-1)}| < \epsilon$ ;

**output**:  $\hat{\mathbf{S}} = \mathbf{S}$ ,  $\hat{\Phi} = \Phi$ ,  $\hat{\mathbf{Q}} = \mathbf{Q}$ .

---

where  $n_i$  denotes the index of the  $i$ th largest value of  $\{\|\zeta_1\|_2, \dots, \|\zeta_N\|_2\}$ . After updating  $\mathbf{Q}$ , the problem w.r.t.  $\mathbf{S}$ ,  $\mathbf{B}$ ,  $\Phi$  is

$$\min_{\mathbf{S}, \mathbf{B}, \Phi} \left\| \left( \underline{\mathbf{G}}^{(3)} - \mathbf{Q} \right) - (\Phi \odot \mathbf{S})\mathbf{B}^T \right\|_F^2. \quad (23)$$

Let us define

$$\underline{\mathbf{E}}^{(3)} := \underline{\mathbf{G}}^{(3)} - \mathbf{Q} \in \mathbb{C}^{LF \times N}. \quad (24)$$

Then, Problem (23) can be viewed as a PARAFAC factorization problem of  $\underline{\mathbf{E}}^{(3)}$ . Consequently,  $\mathbf{S}$ ,  $\mathbf{B}$  and  $\Phi$  can be updated following the rule of TALS; i.e., these three loading factors can be updated by solving the following three least squares problems:

$$\mathbf{S} := \arg \min_{\mathbf{S}} \left\| \underline{\mathbf{E}}^{(1)} - (\mathbf{B} \odot \Phi)\mathbf{S}^T \right\|_F^2, \quad (25a)$$

$$\Phi := \arg \min_{\Phi} \left\| \underline{\mathbf{E}}^{(2)} - (\mathbf{S} \odot \mathbf{B})\Phi^T \right\|_F^2, \quad (25b)$$

$$\mathbf{B} := \arg \min_{\mathbf{B}} \left\| \underline{\mathbf{E}}^{(3)} - (\Phi \odot \mathbf{S})\mathbf{B}^T \right\|_F^2, \quad (25c)$$

where  $\underline{\mathbf{E}}^{(1)} \in \mathbb{C}^{NL \times F}$  and  $\underline{\mathbf{E}}^{(2)} \in \mathbb{C}^{FN \times L}$  are defined by

$$\begin{aligned} \left[ \underline{\mathbf{E}}^{(3)} \right]_{(f-1)L+\ell, n} &= \left[ \underline{\mathbf{E}}^{(1)} \right]_{(\ell-1)N+\ell, f}, \\ &= \left[ \underline{\mathbf{E}}^{(2)} \right]_{(n-1)F+n, \ell}; \end{aligned} \quad (26)$$

see lines 11–13 of Algorithm 1.

Unlike TALS, which only alternates among  $\mathbf{S}$ ,  $\mathbf{B}$  and  $\Phi$ , we cyclically update the four blocks  $\mathbf{Q}$ ,  $\mathbf{S}$ ,  $\mathbf{B}$  and  $\Phi$  using (22) and (25a)–(25c) until some convergence criterion is satisfied—the additional step in (22) can be interpreted as an iterative outliers

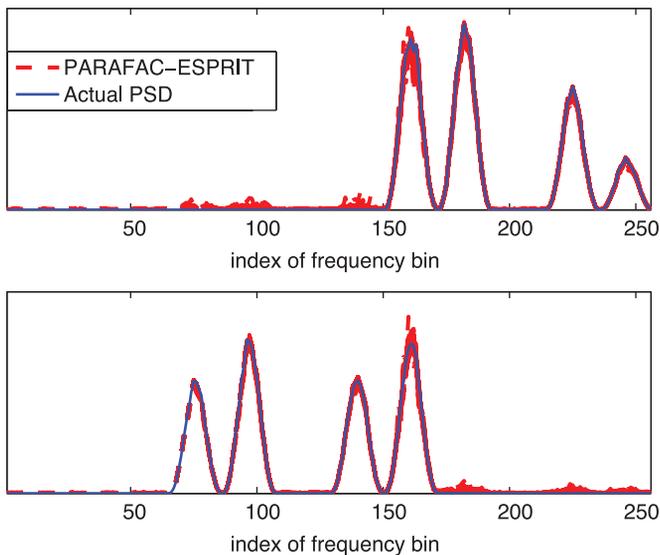


Fig. 6. The estimated power spectra by the proposed PARAFAC approach for 100 trials.  $(N, M, K) = (4, 3, 2)$ ;  $\theta_1 = -8^\circ$ ,  $\theta_2 = 5^\circ$ ; SNR = 10 dB; sample size =  $4 \times 10^5$ .

cancellation procedure. The detailed algorithm is presented in Algorithm 1. As the partial problems w.r.t.  $\mathbf{Q}$ ,  $\mathbf{S}$ ,  $\Phi$  and  $\mathbf{B}$  are all optimally solved, the cost function of Problem (19) can only decrease or stay the same each time one of  $\mathbf{Q}$ ,  $\mathbf{S}$ ,  $\Phi$ , or  $\mathbf{B}$  is updated. Since the cost is bounded below (by zero), it follows that the cost sequence will converge.

## VI. SIMULATIONS

In this section, we use computer simulations to demonstrate the effectiveness of the proposed algorithms. In each simulation trial, the band of interest consists of 12 subchannels. These subchannels are randomly allocated to several sources, and each source takes up to four subchannels. In each occupied subchannel, the corresponding source signal is generated by filtering a random signal that follows the i.i.d. zero-mean unit-variance circularly symmetric Gaussian distribution by a sinc-shape filter with a random scaling. The channel response  $a_{n,k}$  and the noise  $v_{n,m}(t)$  are also randomly generated in each simulation trial, both following the i.i.d. zero-mean circularly symmetric Gaussian distribution, with variances being one and  $\sigma^2$ , respectively. The signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = 10 \log_{10} \left( \frac{\mathbb{E} \left\{ \frac{1}{N} \sum_{n=1}^N \|\mathbf{A}_n \mathbf{x}(t)\|_2^2 \right\}}{\sigma^2} \right).$$

The denoising technique described in Section III is applied. We assume that ULAs are employed at the receivers, and  $d/\lambda = 1/2$ . In all the simulations, the band of interest is discretized into  $N_F = 256$  frequency bins. The simulation results are obtained by averaging 100 trials.

Figs. 6–7 show the estimated source power spectra by the proposed PARAFAC-based formulation and the NMF-based formulation (cf. (15)). There are two sources, and each figure comprises two panels (one for each source). For this experiment, the source power spectra remain fixed for ease of visualiza-

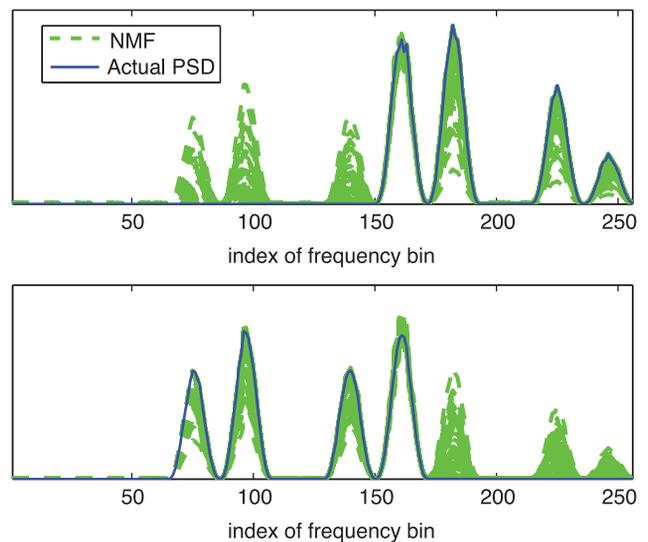


Fig. 7. The estimated power spectra by NMF for 100 trials.  $(N, M, K) = (4, 3, 2)$ ;  $\theta_1 = -8^\circ$ ,  $\theta_2 = 5^\circ$ ; SNR = 10 dB; sample size =  $4 \times 10^5$ .

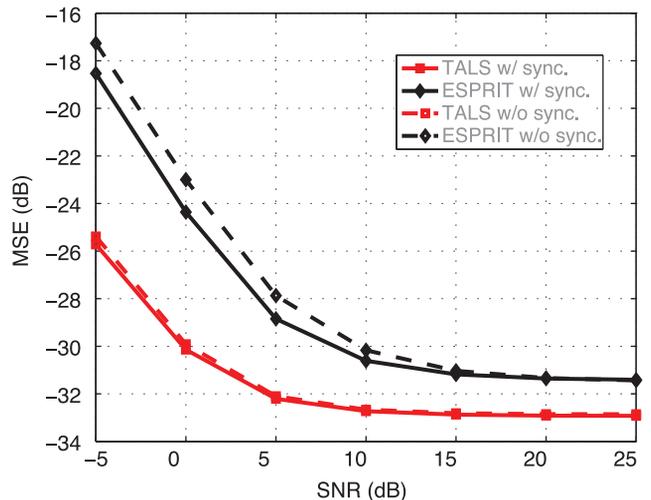


Fig. 8. The MSEs of the estimated power spectra by the PARAFAC algorithms under various SNRs.  $(N, M, K) = (4, 4, 3)$ ;  $\theta_1 = -8^\circ$ ,  $\theta_2 = 5^\circ$ ,  $\theta_3 = 20^\circ$ ; sample size =  $4 \times 10^5$ .

tion, while the particular signal realization, the channel and the noise are all randomly generated for each trial. The estimated power spectra for all 100 trials are overlaid on top of each other in each panel, to enable easy visual assessment of estimation variance, including leakage (‘crosstalk’) from one source to the other. We apply ESPRIT for solving the proposed PARAFAC problem formulation, and the fast hierarchical alternating least squares (HALS) algorithm proposed by Cichocki and Phan [39] for solving the NMF problem. Notice that HALS was empirically found to outperform many state-of-the-art NMF algorithms—see [40]. We see that PARAFAC-ESPRIT identifies the power spectra of both sources fairly well for all the trials. However, NMF fails to yield clearly separated spectra in many trials. This is not surprising, as the solution to the NMF problem (15) is not unique in general [31].

Fig. 8 shows the power spectra estimation accuracy of two PARAFAC-based algorithms, i.e., ESPRIT and TALS, estimated using Monte-Carlo simulation. The mean-square-error

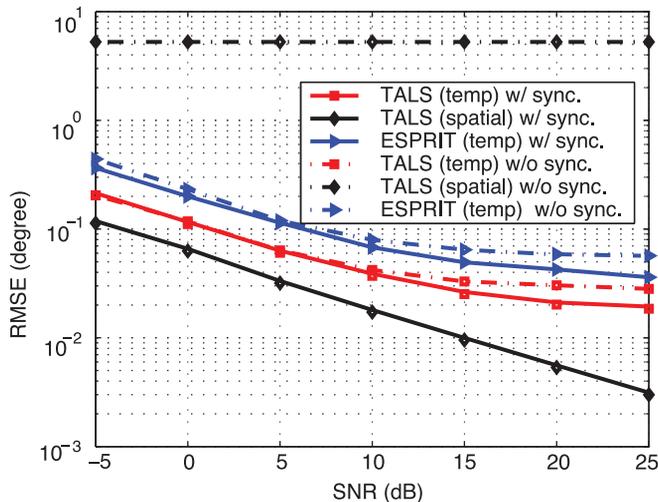


Fig. 9. The RMSEs of the estimated power spectra by the PARAFAC algorithms under various SNRs.  $(N, M, K) = (4, 4, 3)$ ;  $\theta_1 = -8^\circ$ ,  $\theta_2 = 5^\circ$ ,  $\theta_3 = 20^\circ$ ; sample size =  $4 \times 10^5$ .

(MSE) of the estimated power spectra is adopted as the performance measure, defined as

$$\text{MSE} = \min_{\substack{\boldsymbol{\pi} \in \Pi, \\ c_1, \dots, c_K \in \{\pm 1\}}} \frac{1}{K} \sum_{k=1}^K \left\| \frac{\mathbf{s}_k}{\|\mathbf{s}_k\|_2} - c_k \frac{\hat{\mathbf{s}}_{\pi_k}}{\|\hat{\mathbf{s}}_{\pi_k}\|_2} \right\|_2^2,$$

where  $\Pi$  is the set of all permutations of  $\{1, 2, \dots, K\}$ ; and  $\mathbf{s}_k$  and  $\hat{\mathbf{s}}_k$  are the true power spectrum of source  $k$  and the corresponding estimate, respectively. The sign variables  $c_1, \dots, c_K$  are included because we do not enforce nonnegativity of the estimated power spectra—although nonnegativity constraints can be easily incorporated in TALS, and will improve the estimation performance. There is no easy way to incorporate nonnegativity in algebraic ESPRIT-type algorithms. Notice that in this simulation we do not fix the shapes of the power spectra as has been done in the last simulation. We see that both algorithms yield reasonable estimates of the power spectra, but TALS is naturally more accurate. ESPRIT requires much less execution time since it is a closed-form solution. TALS requires 0.0289 seconds, while ESPRIT 0.00054 seconds for this setup, on average. It is also worth noting that the MSEs of TALS and ESPRIT are essentially not affected by using asynchronous sensors, which verifies the claim that our formulation needs no synchronization among the sensors.

Fig. 9 shows the root-mean-square errors (RMSEs) of the estimated DOAs by the temporal correlation-domain TALS and ESPRIT under the same settings as in Fig. 8; the RMSE is defined as

$$\text{RMSE} = \sqrt{\text{E} \left\{ \frac{1}{K} \sum_{k=1}^K |\theta_k - \hat{\theta}_k|^2 \right\}},$$

where  $\hat{\theta}_k$  and  $\theta_k$  are the estimated and the true DOAs of source  $k$ , respectively. As a baseline for comparison, the results of spatial correlation-domain TALS are also presented [22]; this

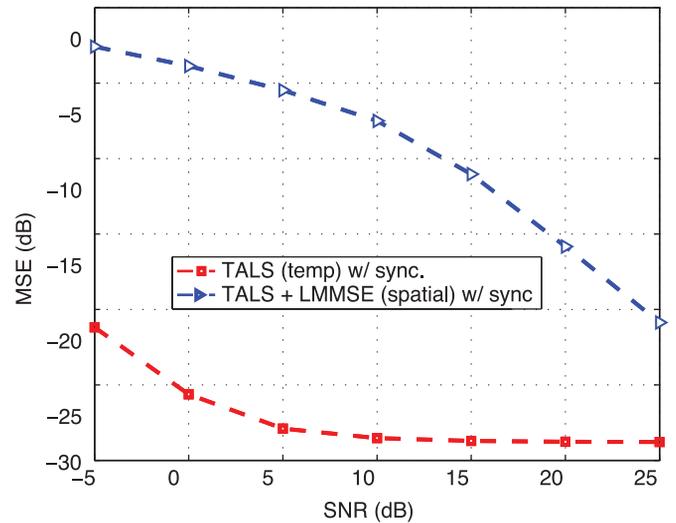


Fig. 10. The MSEs of the estimated power spectra by the temporal and spatial approaches. For the spatial PARAFAC, the LMMSE approach is applied for unmixing the raw sources at receiver 1; the estimated source signals are then used to estimate the power spectra.  $(N, M, K) = (4, 4, 3)$ ;  $\theta_1 = -8^\circ$ ,  $\theta_2 = 5^\circ$ ,  $\theta_3 = 20^\circ$ ; sample size =  $4 \times 10^5$ .

is the classical *multiple invariance sensor array processing* (MI-SAP) approach to direction finding. One can see that the spatial correlation-domain TALS exhibits superior RMSEs when synchronized sensors are employed. However, when there are timing mismatches among the sensors, this approach completely fails to yield reasonable estimates of the DOAs. On the other hand, the proposed temporal correlation-domain PARAFAC algorithms are very robust with respect to receiver asynchronism.

As mentioned in Remark 2, when the receivers are synchronized, using the estimated DOAs (or more precisely,  $\Phi_n$ ) by the spatial PARAFAC formulation, we may unmix the raw source signals at any receiver, e.g., by the spatial PARAFAC formulation, we may unmix the raw source signals at any receiver, e.g., using a linear minimum mean squared error (LMMSE) approach. Then, these unmixed sources can be used for estimating the power spectra. There are two drawbacks to this alternate approach: it entails very high communication overhead, and the accuracy of the estimated power spectra is not comparable to the proposed approach (see Fig. 10).

In Tables I–II, some further simulation results are presented. Specifically, Table I shows the MSEs and RMSEs yielded by the proposed temporal correlation-domain TALS and ESPRIT under different sample sizes. One can see that TALS outperforms ESPRIT as in the last simulation. Also, both algorithms prefer larger sample sizes—using larger sample sizes, the temporal correlations can be estimated more accurately and thus better estimates of the spectra and the DOAs can be expected. In Table II, we show the simulation results for an under-determined case, i.e., an  $N < K$  case. Notice that ESPRIT is not applicable under this setting as it requires that  $\text{rank}(\mathbf{B}) = K$ . We see that TALS yields reasonable estimates of both the power spectra and the DOAs when  $\text{SNR} \geq 0$  dB. We also see that the results given by TALS are almost identical with and without synchronization among the receivers.

TABLE I

MSEs (dB) AND RMSEs (DEGREE) OF THE ESTIMATED POWER SPECTRA AND DOAs BY THE ALGORITHMS, RESPECTIVELY.  $(N, K, M) = (4, 3, 4)$   
 $\theta_1 = -8^\circ, \theta_2 = 5^\circ, \theta_3 = 20^\circ$ ; SNR = 15 dB

Algorithm	measure	$T$			
		$10^5$	$4 \times 10^5$	$8 \times 10^5$	$10^6$
TALS (temp.)	MSE	-22.873	-29.9662	-35.6097	-38.7914
	RMSE	0.1167	0.0523	0.0387	0.0327
ESPRIT (temp.)	MSE	-19.0484	-25.502	-29.3231	-30.9283
	RMSE	0.2142	0.0905	0.0647	0.0555

TABLE II

MSEs (dB) AND RMSEs (DEGREE) OF THE ESTIMATED POWER SPECTRA AND DOAs BY THE ALGORITHMS, RESPECTIVELY.  $(N, K, M) = (3, 4, 3)$   
 $\theta_1 = -32^\circ, \theta_2 = -12^\circ, \theta_3 = 8^\circ, \theta_4 = 35^\circ$ ; SAMPLE SIZE =  $4 \times 10^5$

Algorithm	measure	SNR(dB)			
		-10	0	10	20
TALS (temp.) w/ sync.	MSE	-5.0057	-18.938	-28.4661	-27.47
	RMSE	8.3493	1.4022	0.7513	1.0968
TALS (temp.) w/o sync.	MSE	-5.0243	-19.2248	-26.6213	-27.3874
	RMSE	8.3709	1.405	0.7499	1.0988

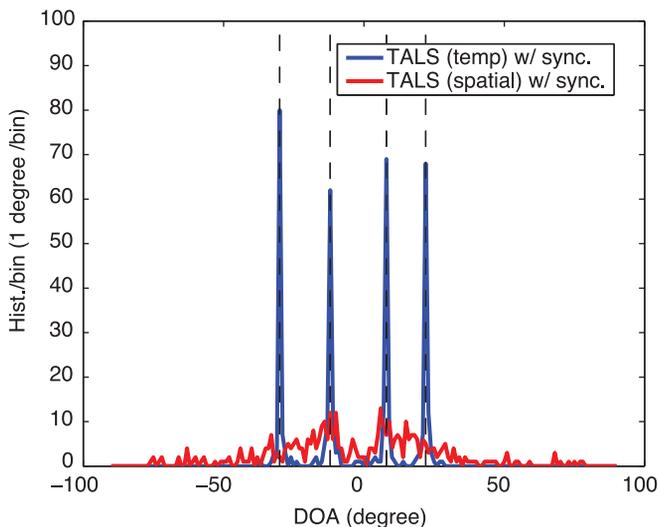


Fig. 11. Histogram of temporal correlation-domain and spatial correlation-domain TALS; SNR = 0 dB;  $(N, K, M) = (4, 3, 2)$ ; sample size =  $4 \times 10^5$ ; the dot lines are the true DOAs.

In Fig. 11, we use a particular case to demonstrate the identifiability improvement using the proposed temporal correlation-domain PARAFAC formulation, compared to the spatial domain one as in [22]. In this simulation, we set  $(N, K, M) = (3, 4, 2)$  and use synchronized sensors. The histograms of the estimated DOAs from 100 trials are plotted in this figure. One can see that the spatial PARAFAC almost fails completely, while the temporal approach works well in this case. The reason is that the row dimension of the Vandermonde loading matrix under our formulation is  $L = 3$  since  $q = r - s \in \{-1, 0, 1\}$ . However, under the MI-SAP formulation, this dimension equals  $M = 2$ . By carefully examining the condition in Theorem 1, we see that using the spatial domain formulation we cannot fulfill this condition, while using our temporal domain formulation we

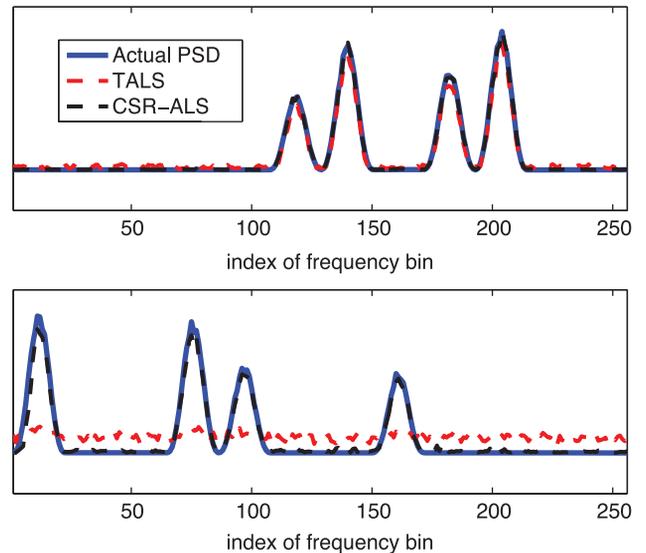


Fig. 12. The estimated power spectra by TALS and CSR-ALS respectively. SNR = 0 dB; SOR = -15 dB;  $(N, K, M) = (8, 2, 3)$ ;  $(\theta_1, \theta_2) = (-32^\circ, -18^\circ)$ ;  $\mathcal{A} = \{3, 4\}$ .

can. This verifies our claim in Remark 3 that the proposed formulation can improve the identifiability.

In the remaining simulations, we evaluate the effectiveness of CSR-ALS when there are some receivers reporting corrupt data to the fusion center. To quantify the level of corruption, we define the *signal-to-outlier ratio* (SOR) as

$$\text{SOR(dB)} = 10 \log_{10} \left( \frac{(1/N) \|(\Phi \odot \mathbf{S}) \mathbf{B}^T\|_F^2}{(1/|\mathcal{A}|) \|\mathbf{N}\|_F^2} \right).$$

The real and imaginary parts of the active columns of  $\mathbf{N}$  are both generated following the uniform distribution between zero and one and scaled to satisfy the specified SORs.

Fig. 12 presents an illustrative example using two sources. We set SOR = -15 dB, SNR = 0 dB,  $\mathcal{A} = \{3, 4\}$  and  $P = 3$ . The proposed CSR-ALS algorithm is initialized using plain TALS. In this example, we see that TALS fails to give a reasonable estimate of the second spectrum, while the estimated spectra by CSR-ALS match the actual spectra very well. The 2-norms of  $\hat{\mathbf{q}}_n$ 's are presented in Fig. 13. As one can see,  $\|\hat{\mathbf{q}}_3\|_2$  and  $\|\hat{\mathbf{q}}_4\|_2$  take the largest values, which implies that CSR-ALS correctly detects the receivers which are reporting corrupt data.

Table III shows the results of a Monte Carlo simulation. Here, besides TALS, we also provide a PARAFAC algorithm that is robust to outlying elements [41] as another baseline. This baseline algorithm employs an  $\ell_1$  fitting criterion as its cost function and thus is referred to as the  $\ell_1$  PARAFAC algorithm in the sequel. We set  $(N, K, M) = (10, 2, 4)$ ,  $(\theta_1, \theta_2) = (-32^\circ, -18^\circ)$ ,  $|\mathcal{A}| = 2$ , and SNR = 15 dB. The outlier receivers are randomly chosen in each trial. We let  $P = 4$  in this simulation. One can see that CSR-ALS exhibits much lower MSEs and RMSEs compared to TALS and  $\ell_1$  PARAFAC under all SORs; particularly, when the SOR = -15 dB and -10 dB, CSR-ALS still yields reasonable estimates of the DOAs and the power spectra, which is quite pleasing. On the other hand, the robust  $\ell_1$  PARAFAC algorithm does not work quite well under the simulated scenario. This might be because  $\ell_1$  PARAFAC is designed

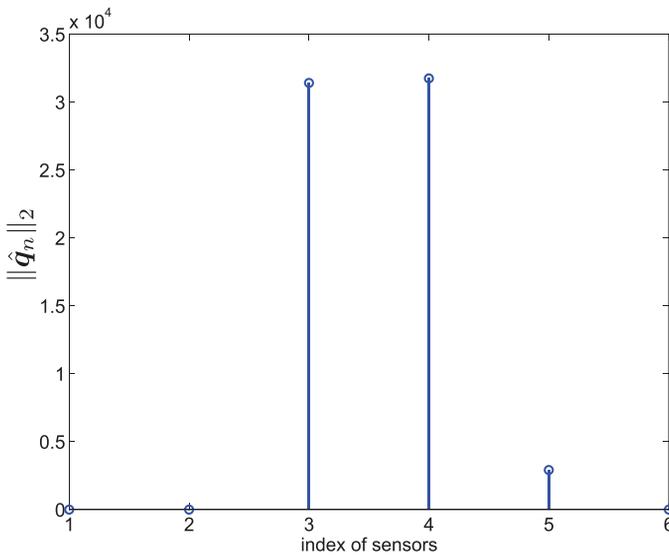


Fig. 13. The values of  $\{\|\hat{\mathbf{q}}_n\|_2\}_{n=1}^N$ . SNR = 0 dB; SOR = -15 dB;  $(N, K, M) = (8, 2, 3)$ ;  $(\theta_1, \theta_2) = (-32^\circ, -18^\circ)$ ; sample size =  $4 \times 10^5$ ;  $\mathcal{A} = \{3, 4\}$ .

TABLE III

MSEs (dB) AND RMSEs (DEGREE) OF THE ESTIMATED POWER SPECTRA AND DOAs BY THE ALGORITHMS UNDER VARIOUS SORS.  $(N, K, M) = (10, 2, 4)$ ;  $(\theta_1, \theta_2) = (-32^\circ, -18^\circ)$ ;  $P = 4$ ; SNR = 15 dB; SAMPLE SIZE =  $4 \times 10^5$

Algorithm	measure	SOR (dB)				
		-15	-10	-5	0	5
TALS	MSE	-2.4287	-3.9096	-8.3432	-15.4394	-21.5253
	RMSE	14.1418	12.0946	7.1292	3.1836	1.3593
$\ell_1$ PARAFAC	MSE	-2.5592	-3.8637	-8.355	-15.7056	-22.3924
	RMSE	14.4672	12.6886	7.428	3.213	1.2954
CSR-ALS	MSE	-12.24	-14.6863	-16.339	-22.4289	-39.3838
	RMSE	4.1743	3.3875	2.9917	1.2436	0.0154

TABLE IV

MSEs (dB) AND RMSEs (DEGREE) OF THE ESTIMATED POWER SPECTRA AND DOAs BY THE ALGORITHMS UNDER DIFFERENT NUMBER OF OUTLIER RECEIVERS.  $(N, K, M) = (12, 3, 4)$ ;  $(\theta_1, \theta_2, \theta_3) = (-32^\circ, -18^\circ, 5^\circ)$ ;  $P = 6$ ; SOR = -5 dB; SNR = 15 dB; SAMPLE SIZE =  $4 \times 10^5$

Algorithm	measure	$ \mathcal{A} $				
		1	2	3	4	5
TALS	MSE	-19.2945	-13.273	-10.0091	-8.9461	-7.7199
	RMSE	1.1994	2.8405	4.7601	5.0001	6.1171
$\ell_1$ PARAFAC	MSE	-24.3228	-15.5017	-10.3897	-8.5657	-6.7885
	RMSE	1.0852	2.7077	4.9433	5.2793	6.6878
CSR-ALS	MSE	-36.9257	-37.0093	-37.4261	-36.6984	-17.8684
	RMSE	0.0476	0.0482	0.0464	0.055	2.3447

for dealing with outliers at the individual element-level—it is not specialized for handling slice-level outliers.

Table IV summarizes simulation results for different numbers of outlier receivers. We set  $(N, K, M) = (12, 3, 4)$  and  $(\theta_1, \theta_2, \theta_3) = (-32^\circ, -18^\circ, 5^\circ)$ , and increase the number of outliers from one to four. The parameter  $P$  is set to be  $P = 6$  in all cases, irrespective of the true number of outliers. The SOR and SNR are fixed to be -5 dB and 15dB, respectively. We see that CSR-ALS is much more robust to different numbers of outliers compared to TALS and  $\ell_1$  PARAFAC. Particularly, more than 10 dB MSE reduction can be observed in these simulation results.

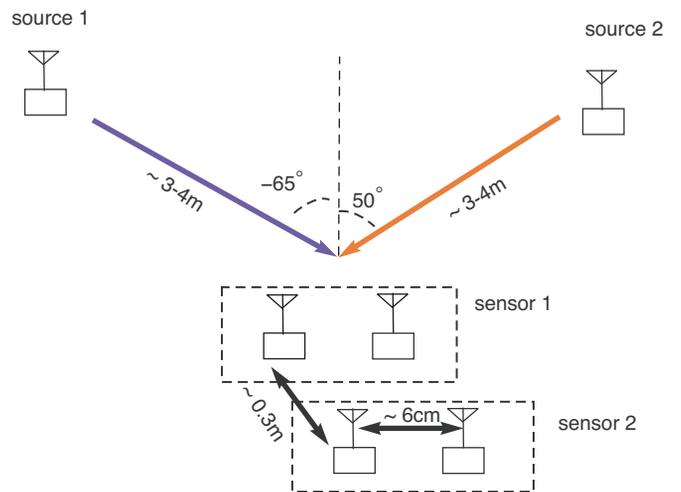


Fig. 14. Experimental layout of two transmitter, two receiver network. The two single-antenna radios inside each dashed box are synchronized with each other to act as a dual-antenna receiver; the two dashed boxes are *not* synchronized with one another.

## VII. LABORATORY EXPERIMENT

We have also conducted a laboratory experiment using software defined radios, for proof-of-concept. The experiment was carried out in the communications laboratory of the Department of Electrical and Computer Engineering at the University of Minnesota. Universal Software Radio Peripheral (USRP) radios by Ettus Research were used as both transmitters and receivers.

The experiment comprised  $K = 2$  transmitters and  $N = 2$  receivers, with each receiver equipped with  $M = 2$  antennas. In order to model a receiver with two antennas, each receiver was implemented using two single-antenna USRP radios. The trick here is to synchronize these two single-antenna radios with a common clock (to align carrier frequencies and sample time across these two devices) so that they act as one. Notice that this is equivalent to synchronizing the two down-conversion chains for a single dual-antenna receiver, as described in the introduction. We did not synchronize the two different receiving radio pairs—see Fig. 14 for an illustration. The two receivers were separated by 30 cm, and each pair of antennas within each receiver was separated by a distance of 6 cm (so that  $d/\lambda$  will be fixed to 0.5 if the band of interest is centered at 2.5 GHz). The primary transmitters were placed approximately 3–4 m away from the receivers at angles of around  $\theta_1 = -65^\circ$  and  $\theta_2 = 50^\circ$  with line-of-sight propagation to each antenna, as shown in Fig. 14.

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. See Fig. 15 for a plot of the power spectrum of the actual signal received at antenna  $y_{2,1}(t)$ . Fig. 16 illustrates the separated spectra that are obtained by applying ESPRIT and TALS,

<sup>2</sup>We should mention that such measured angles are subject to measuring errors, since the transmitters were not really placed in ‘far field’ due to laboratory space constraints. Nevertheless, the experiment results are reasonably close to the measured angles.

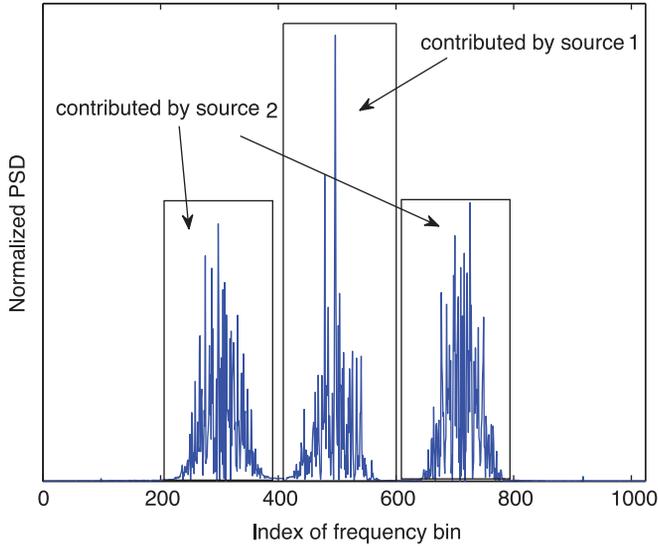


Fig. 15. The measured power spectrum using  $y_{2,1}(t)$ .

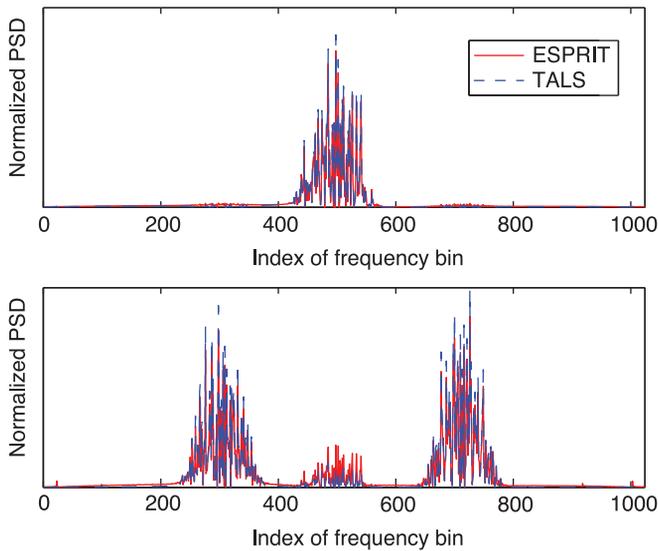


Fig. 16. The separated power spectra by ESPRIT and TALS, respectively.

TABLE V  
THE ESTIMATED AVERAGE MRRS AND DOAs BY ESPRIT, TALS, AND NMF  
RESPECTIVELY

Algorithm	Measure	Average Result
ESPRIT	SMR	0.1572
	DOAs	$(-67.1438^\circ, 47.3884^\circ)$
TALS	SMR	0.1014
	DOAs	$(-67.1433^\circ, 53.0449^\circ)$
NMF	SMR	0.2537
	DOAs	-

respectively. We can observe that both algorithms can separate the power spectra reasonably well. Particularly, TALS yields visually better estimation of the second power spectrum (ESPRIT shows more residual leakage from the first spectrum to the second).

Table V summarizes the results of multiple laboratory experiments (averaged over 10 measurements), to illustrate the consistency and effectiveness of our proposed framework. In order

to establish a metric for the performance of our power spectra separation, we define the side-lobe to main-lobe ratio (SMR) as our performance measurement. Specifically, let  $\mathcal{S}_1$  and  $\mathcal{S}_2$  denote the frequency index sets occupied by source 1 and source 2, respectively. We define

$$\text{SMR} = \frac{1}{2} \left( \frac{\|\hat{\mathbf{s}}_1(:, \mathcal{S}_2)\|_1}{\|\hat{\mathbf{s}}_1(:, \mathcal{S}_1)\|_1} + \frac{\|\hat{\mathbf{s}}_2(:, \mathcal{S}_1)\|_1}{\|\hat{\mathbf{s}}_2(:, \mathcal{S}_2)\|_1} \right);$$

notice that  $\text{SMR} \in [0, 1]$ , and since the power spectra from source 1 and source 2 do not overlap,  $\mathcal{S}_1$  and  $\mathcal{S}_2$  are disjoint, which is necessary for the SMR metric as defined above to be meaningful. Note that lower SMRs signify better spectra separation performance. We observe that the average SMRs of the ESPRIT and TALS algorithms are reasonably small, while NMF exhibits approximately double SMR on average. The estimated average DOAs are also presented in Table V; one can see that both ESPRIT and TALS yield similar estimated DOAs. It should be noted that power spectra separation was consistently achieved over numerous trials with varying geometry of source-receiver placement; DOA estimates exhibited somewhat greater variation in accuracy.

## VIII. CONCLUSION

The problem of joint power spectra separation and source localization has been considered in this paper. Working in the temporal correlation domain, this problem has been formulated as a PARAFAC decomposition problem. This novel formulation does not require synchronization across the different multi-antenna receivers, and it can exhibit better identifiability than conventional spatial correlation-domain sensor array processing approaches such as MI-SAP. Robustness issues have also been considered, and identifiability of the latent factors (and the receivers reporting corrupted data) was theoretically established in this more challenging setup. A robust PARAFAC algorithm has been proposed to deal with this situation, and extensive simulations have shown that the proposed approaches are effective. In addition to simulations, real experiments with a software radio prototype were used to demonstrate the effectiveness of the proposed approach.

## APPENDIX A PROOF OF THEOREM 2

Let  $v_{\text{obj}}$  be the objective value of Problem (19). It is evident  $v_{\text{obj}} \geq 0$ . To achieve the lower bound, one needs to find  $\mathbf{q}_n^*$ ,  $\Phi^*$ ,  $\mathbf{S}^*$ ,  $\beta_n^*$  such that

$$v_n^* = \left\| \begin{bmatrix} \mathbf{G}^{(3)} \\ \cdot, n \end{bmatrix} - (\Phi^* \odot \mathbf{S}^*) \beta_n^* - \mathbf{q}_n^* \right\|_2 = 0, \quad \forall n, \quad (27)$$

where  $\beta_n^*$  denote the transpose of the  $n$ th row of  $\mathbf{B}^*$  for  $n = 1, \dots, N$ . We now consider the values of  $v_{\text{obj}}$  under several scenarios, classified by the relationship between  $\mathcal{A}$  and  $\mathcal{A}^*$ .

*Case 1:* Assume that there is an optimal solution  $\mathbf{N}^*$  such that  $\mathcal{A} \subseteq \mathcal{A}^*$ . Denote the complement of  $\mathcal{A}^*$  by  $\mathcal{A}_c^*$ , i.e.,  $\mathcal{A}^* \cap \mathcal{A}_c^* = \emptyset$  and  $\mathcal{A}^* \cup \mathcal{A}_c^* = \mathcal{N}$ . Let  $\tilde{\mathbf{G}}^{(3)} = \mathbf{G}^{(3)}(:, \mathcal{A}_c^*)$ ;  $\tilde{\mathbf{G}}^{(3)}$  can also be represented as

$$\tilde{\mathbf{G}}^{(3)} = (\Phi \odot \mathbf{S}) \mathbf{B}(\mathcal{A}_c^*, :)^T. \quad (28)$$

Notice that  $\mathbf{B}(\mathcal{A}_c^*, :) \in \mathbb{C}^{|\mathcal{A}_c^*| \times K}$  and  $|\mathcal{A}_c^*| \geq N - P$ . By the assumption that  $\mathbf{B}$  is randomly drawn from a jointly continuous distribution, we have

$$\text{krank}(\mathbf{B}(\mathcal{A}_c^*, :)) \geq \min\{N - P, K\}.$$

which, combined with (21) and Theorem 1, leads to the conclusion that the representation in (28) is essentially unique. Hence, to make  $\nu_n^* = 0$  for all  $n \in \mathcal{A}_c^*$ , one must have  $\Phi^*, \mathbf{S}^*$  and  $\mathbf{B}^*(\mathcal{A}_c^*, :)$  being essentially identical to  $\Phi, \mathbf{S}$  and  $\mathbf{B}(\mathcal{A}_c, :)$ , respectively.

On the other hand, for  $n \in \mathcal{A}^*$ , one can always make (27) hold by letting

$$\mathbf{q}_n^* = \left[ \underline{\mathbf{G}}^{(3)} \right]_{:,n} - (\Phi^* \odot \mathbf{S}^*) \beta_n^*, \quad \forall n \in \mathcal{A}^*. \quad (29)$$

Hence, solutions that achieve  $v_{\text{obj}} = 0$  are solutions satisfying 1)-2) under this scenario.

*Case 2:* Assume that there is an optimal solution  $\mathbf{N}^*$  such that  $\mathcal{A} \not\subseteq \mathcal{A}^*$ . Under this scenario, there are three different index sets which need to be considered, namely,  $\mathcal{A}_1 = \mathcal{N} / \{\mathcal{A}^* \cup \mathcal{A}\}$ ,  $\mathcal{A}_2 = \mathcal{A}^*$  and  $\mathcal{A}_3 = \mathcal{A} / \{\mathcal{A}^* \cap \mathcal{A}\}$ ; note that  $\mathcal{N} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3$ .

By noticing that  $|\mathcal{A}_1| \geq N - 2P$ , and

$$\text{krank}(\mathbf{B}(\mathcal{A}_1, :)) \geq \min\{N - 2P, K\},$$

we conclude that  $\nu_n^* = 0$  for all  $n \in \mathcal{A}_1$  can be achieved only if  $\Phi^*, \mathbf{S}^*$  are identical to  $\Phi$  and  $\mathbf{S}$  up to column permutation and scaling by the same reason as in the last scenario. Also, as one can always let  $\mathbf{q}_n^*$  satisfy (29),  $\nu_n^*$  for  $n \in \mathcal{A}^*$  can be forced to be zero. However, by (20),  $\nu_n^*$  can never be zero for any  $n \in \mathcal{A}_3$ . Specifically, as  $\mathbf{n}_n \notin \mathcal{R}(\Phi \odot \mathbf{S}) = \mathcal{R}(\Phi^* \odot \mathbf{S}^*)$  and for any  $n \in \mathcal{A}_3$  we have

$$\left[ \underline{\mathbf{G}}^{(3)} \right]_{:,n} = (\Phi \odot \mathbf{S}) \beta_n + \mathbf{n}_n,$$

one can never find a  $\beta_n^*$  for any  $n \in \mathcal{A}_3$  such that

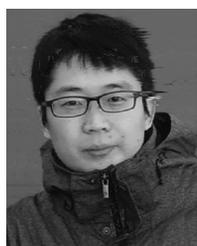
$$\left[ \underline{\mathbf{G}}^{(3)} \right]_{:,n} - (\Phi^* \odot \mathbf{S}^*) \beta_n^* = \mathbf{0}.$$

As conclusion of Case 1–2, optimal solutions have to satisfy 1)–2). This completes the proof.

## REFERENCES

- [1] 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, pp. 7799–7803.
- [2] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, 2007.
- [3] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Phys. Commun.*, vol. 4, no. 1, pp. 40–62, 2011.
- [4] 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, 2012.
- [5] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. ICASSP*, 2007, vol. 4, pp. 1357–1360.
- [6] Z. Fanzhi, C. Li, and Z. Tian, "Distributed compressive spectrum sensing in cooperative multihop cognitive networks," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 37–48, Feb. 2011.
- [7] 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.
- [8] Y. L. Polo, Y. Wang, A. Pandharipande, and G. Leus, "Compressive wide-band spectrum sensing," in *Proc. ICASSP*, 2009, pp. 2337–2340.
- [9] D. D. Ariyananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4775–4789, 2012.
- [10] M. A. Lexa, M. E. Davies, J. S. Thompson, and J. Nikolic, "Compressive power spectral density estimation," in *Proc. ICASSP*, May 2011, pp. 3884–3887.
- [11] 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.
- [12] O. Mehanna and N. D. Sidiropoulos, "Adaptive thresholding for distributed power spectrum sensing," *Proc. IEEE ICASSP*, pp. 4459–4463, 2013.
- [13] 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, Nov. 2015.
- [14] Z. Hu, R. Ranganathan, C. Zhang, R. Qiu, M. Bryant, M. C. Wick, and L. Li, "Robust non-negative matrix factorization for joint spectrum sensing and primary user localization in cognitive radio networks," presented at the IEEE Waveform Divers. Design Conf., Kauai, HI, USA, Jan. 22–27, 2012.
- [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, Oct. 2011.
- [16] S.-J. Kim, E. Dall'Anese, and G. B. Giannakis, "Cooperative spectrum sensing for cognitive radios using kriged kalman filtering," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 24–36, Feb. 2011.
- [17] X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Tensor-based power spectra separation and emitter localization for cognitive radio," in *Proc. IEEE 8th SAM Workshop*, 2014, pp. 421–424.
- [18] A. Molisch, L. Greenstein, and M. Shafi, "Propagation issues for cognitive radio," *Proc. IEEE*, vol. 97, no. 5, pp. 787–804, May 2009.
- [19] 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.
- [20] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
- [21] R. Bro, "PARAFAC. Tutorial and applications," *Chemometrics Intell. Lab. Syst.*, vol. 38, no. 2, pp. 149–171, 1997.
- [22] 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, Aug. 2000.
- [23] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 810–823, Mar. 2000.
- [24] 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.
- [25] L. D. Lathauwer, "A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 28, no. 3, pp. 642–666, Aug. 2006.
- [26] L. Chiantini and G. Ottaviani, "On generic identifiability of 3-tensors of small rank," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 3, pp. 1018–1037, 2012.
- [27] A. Stegeman, J. M. Ten Berge, and L. De Lathauwer, "Sufficient conditions for uniqueness in candecomp/parafac and indscal with random component matrices," *Psychometrika*, vol. 71, no. 2, pp. 219–229, 2006.
- [28] D. Nion and L. D. Lathauwer, "An enhanced line search scheme for complex-valued tensor decompositions. Application in DS-CDMA," *Signal Process.*, vol. 88, no. 3, pp. 749–755, 2008.
- [29] 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.
- [30] B. Ottersten, M. Viberg, and T. Kailath, "Performance analysis of the total least squares ESPRIT algorithm," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1122–1135, 1991.
- [31] 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.
- [32] P. Pal and P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4167–4181, 2010.
- [33] E. Dall'Anese, J. A. Bazerque, and G. B. Giannakis, "Group sparse lasso for cognitive network sensing robust to model uncertainties and outliers," *Phys. Commun.*, vol. 5, no. 2, pp. 161–172, 2012.
- [34] T. M. Taher, M. J. Misurac, J. L. LoCicero, and D. R. Ucci, "Microwave oven signal interference mitigation for Wi-Fi communication systems," in *Proc. 5th IEEE CCNC*, 2008, pp. 67–68.
- [35] A. Nasri and R. Schober, "Adaptive  $l_p$ -norm metric for secondary BICM-OFDM systems," in *Proc. IEEE GLOBECOM*, 2009, pp. 1–6.

- [36] A. W. Min, K.-H. Kim, and K. G. Shin, "Robust cooperative sensing via state estimation in cognitive radio networks," *Proc. IEEE DySPAN*, pp. 185–196, 2011.
- [37] T.-H. Chan, A. Ambikapathi, W.-K. Ma, and C.-Y. Chi, "Robust affine set fitting and fast simplex volume max-min for hyperspectral end-member extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3982–3997, 2013.
- [38] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [39] A. Cichocki and A.-H. Phan, "Fast local algorithms for large scale non-negative matrix and tensor factorizations," *IEICE Trans. Fund. Electron., Commun., Comput. Sci.*, vol. 92, no. 3, pp. 708–721, 2009.
- [40] K. Huang and N. D. Sidiropoulos, "Putting nonnegative matrix factorization to the test: A tutorial derivation of pertinent Cramér-Rao bounds and performance benchmarking," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 76–86, 2014.
- [41] 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, Aug. 2005.



**Xiao Fu** (S'12–M'15) received his 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. In 2014, he received his Ph.D. degree in electronic engineering from the Chinese University of Hong Kong (CUHK), Hong Kong. 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, United States. His research interests include signal processing and machine learning, with a recent emphasis on factor analysis and its applications.

Dr. Fu was an awardee of the Oversea Research Attachment Programme (ORAP) 2013 of the Engineering Faculty, CUHK, which sponsored his visit to the Department of Electrical and Computer Engineering, University of Minnesota, from September 2013 to February 2014. He received a Best Student Paper Award at ICASSP 2014.



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



**John H. Tranter** (S'12) received the B.E.E. degree in 2014, and is currently pursuing the M.S.E.E. degree (both at the University of Minnesota, Minneapolis, MN, USA), while serving as a Research Assistant in the Department of Electrical and Computer Engineering. He also has served as a Teaching Assistant, and is the instructor for the Communication Systems Laboratory. In addition, he received the Doctor of Musical Arts degree from the University of Minnesota in 2008, where in addition to his engineering studies, he is an Affiliate Professor

of Music.

His research interests include signal processing and cognitive radio, with a recent emphasis on beamforming design. He received a Best Student Paper Award at ICASSP 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, both in electronic engineering, from The Chinese University of Hong Kong (CUHK), 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, R.O.C. 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 and communications, with a recent emphasis on optimization, MIMO transceiver designs and interference management, blind signal processing theory, methods and applications, and hyperspectral unmixing in remote sensing.

Dr. Ma is currently serving or has served as Associate Editor and Guest Editor of several journals, which include IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE SIGNAL PROCESSING LETTERS, SIGNAL PROCESSING, IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS and IEEE SIGNAL PROCESSING MAGAZINE. He was a tutorial speaker in 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). Dr. Ma's students have won ICASSP Best Student Paper Awards in 2011 and 2014, respectively, and he is co-recipient of a WHISPERS 2011 Best Paper Award. He received Research Excellence Award 2013–2014 by CUHK.