Identifiability Results for Blind Beamforming in Incoherent Multipath with Small Delay Spread

Nicholas D. Sidiropoulos, *Senior Member, IEEE,* and Xiangqian Liu

***Abstract*—Several explicit identifiability results are derived for deterministic blind beamforming in incoherent multipath with small delay spread. For example, it is shown that if the *sum* of spatial and fractional sampling diversities exceeds two times the total number of paths, then identifiability can be guaranteed even for one symbol snapshot. The tools come from the theory of low-rank three-way array decomposition (commonly referred to as parallel factor analysis (PARAFAC) and data smoothing in one and two dimensions. New results regarding the Kruskal-rank of certain structured matrices are also included, and they are of interest in their own right.**

***Index Terms*—Array signal processing, delay estimation, direction-of-arrival estimation, matrix decomposition.**

I. Introduction

BLIND beamforming is the problem of reconstructing source signals given only output data from a sensor array without assuming knowledge of propagation parameters. Instead, blind beamforming algorithms rely on various structural properties of the problem to reconstruct the source signals *and* estimate relevant propagation parameters.

Blind beamforming has been under intense research scrutiny during the past decade, primarily due to its importance in wireless communications, radar, source localization, and system identification. Although the literature on the subject is broad and the available blind beamforming algorithms build on rather diverse principles, they can be coarsely classified as follows:

i) those that exploit known space/time manifold structure (e.g., ESPRIT [13] is probably one of the simplest and best well-known examples);
ii) those based on known source signal structure, e.g., finite alphabet, constant modulus, known pulse shape/spreading, or cyclostationarity (e.g., [1], [2], [16], [21]);
iii) those that exploit statistical independence of the sources (cf., [6] and references therein for a tutorial review).

A recent *IEEE Proceedings* paper by van der Veen [22] provides an excellent overview of algebraic methods for deterministic blind beamforming, which capitalize on i) and/or ii). Deterministic blind beamforming techniques make relatively strong structural assumptions but in return may guarantee highly accurate source signal and propagation parameter estimates in the high SNR regime with only a small number of sensors and/or available samples. This is particularly important in mobile wireless environments, wherein the number of sources and multipath rays can be large and propagation parameters can be changing rapidly relative to the signaling rate.

Identifiability pertains to the capability of recovering all relevant source signals and propagation parameters in the absence of noise and unmodeled dynamics. Identifiability is a natural prerequisite for a well-posed estimation problem because it signifies the existence of a unique desired solution under ideal operating conditions. Given identifiability, one would then turn to estimation (model fitting) *algorithms*. Eventually, *robustness* and *performance* of a given algorithm need to be studied.

Even if a particular model is identifiable, this does not necessarily mean that a given algorithm will identify the model parameters in the absence of noise. This may require additional algorithm-specific identifiability conditions. However, model identifiability lays out the least common denominator that is necessary for *any* algorithm to identify the model parameters.

Model identifiability issues related to deterministic blind beamforming are the focus of this paper. Its starting point has been the realization that PARAFAC is the core problem underlying [22]. The important twist is that existing PARAFAC uniqueness results do not directly apply to the deterministic blind beamforming problem because delay spread induces collinearity in the rows of the effective signal matrix, violating a basic premise behind all PARAFAC uniqueness results to date. Taking advantage of additional (Vandermonde) structure that is often present in the context of deterministic blind beamforming, we show how one can squeeze the most out of spatio-temporal smoothing to prove identifiability. Note that the link to PARAFAC also implies that existing least squares model fitting algorithms [3], [5], [18], [19] (which do not require additional algorithm-specific identifiability conditions) are directly applicable to deterministic blind beamforming.

A drawback of the work reported in [22] is that the working conditions are necessary but not sufficient for identifiability; rank issues remain. The most important of these rank issues has to do with nontrivial lower bounds on the rank of the Khatri–Rao product, and it is addressed herein in Lemma 1. Another issue has to do with the requirement that the source (user) signal matrix $\mathbf{S}$ be fat and full rank. In a rapidly fading environment (or, e.g., when data are collected over slots in a TDMA system), it is desirable to alleviate the requirement that $\mathbf{S}$ be fat and full rank. This is also true for finite-alphabet digital communication trans-

Manuscript received May 10, 1999; revised September 15, 2000. This work was supported by NSF/CAREER CCR-9733540 (revised award CCR-0096165) and NSF/Wireless CCR-9979295. A partial summary of results appears in the Proceedings of EUSIPCO, September 5–8, 2000, Tampere, Finland, and in the Proceedings of ICASSP, June 5–9, 1999, Istanbul, Turkey. The associate editor coordinating the review of this paper and approving it for publication was Prof. A. M. Zoubir.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: nikos@ece.umn.edu).

Publisher Item Identifier S 1053-587X(01)00072-1.

missions for which there is a generally small but positive probability of rank-deficiency (e.g., $\approx 10\%$ for i.i.d. 8×10 BPSK $\mathbf{S}$ and higher for correlated signal streams). We will lift such requirements herein.

The work reported in [23]–[25] assumes that the overall spatio-temporal channel matrix has already been estimated using training symbols and seeks to determine path angles and delays by exploiting the Kronecker structure of the spatio-temporal manifold. If the channel is known or has been estimated, then the source symbols can be readily estimated, and this is usually the end goal in digital communication applications (but path angles and delays can be useful in a 911 emergency situation). Similar rank issues appear in [23]–[25]. The work in [24] is of particular interest because it employs spatio-temporal smoothing. Starting from three dimensionality conditions that are necessary for identifiability, [24] attempts to optimize the choice of smoothing factors to maximize the number of resolvable rays. This has two drawbacks: 1) identifiability is not guaranteed for the resulting maximum since the objective is only an upper bound on the number of rays that could be resolvable, and ii) due to analytical difficulties, the maximization is carried out assuming continuous parameters, which can only be an approximation. In contrast, we optimize a sufficient condition (hence our maximum is achieved), and we do it by explicit integer optimization (hence, the result is exact). A concise statement of identifiability conditions pertaining to [22]–[25] can be found in the end of the Appendix.

The main contributions of this paper are in terms of the following:

- a backbone result on the so-called Kruskal-rank of the Khatri–Rao product (Lemma 1) that also provides a useful lower bound on the rank of the Khatri–Rao product.
- a variety of explicit identifiability results for deterministic blind beamforming in the presence of incoherent multipath with small delay spread. For example, we show that if the *sum* of spatial and fractional sampling diversities exceeds two times the total number of paths, then identifiability can be guaranteed, even for one snapshot.
- Improved PARAFAC/Vandermonde identifiability results (Theorems 2, 3). These are contributions to the theory of multi-way analysis. They can be directly applied to yield interesting identifiability conditions for most problems in [22]–[25] as well as other related problems in the recent literature.

A. Organization

The rest of this paper is structured as follows. Section II contains a compact derivation of the so-called *incoherent multipath with small delay spread* data model, following [22]. Although most models in [22] fall under our framework, we focus on this particular model because it strikes a good balance between plausibility and generality on one hand and simplicity of exposition on the other. Section III provides very brief background on PARAFAC and the uniqueness of trilinear low-rank decomposition. This is required to understand the proofs and derivations in this paper. Section IV presents our main results. Conclusions are drawn in Section V. Proofs are deferred to the Appendix.

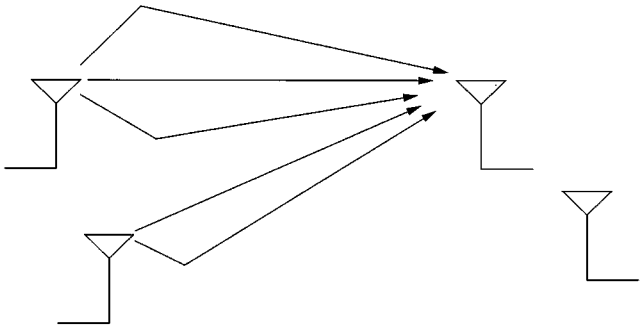

Fig. 1. Multipath propagation scenario.

B. Notation

Notation conventions used throughout this paper are standard. We will make use of the Kronecker product and the Khatri–Rao product. These are defined next.

Given two matrices $\mathbf{A}$ $(I \times F_1)$ and $\mathbf{B}$ $(J \times F_2)$, the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is the $IJ \times F_1F_2$ matrix defined as

$$\mathbf{A} \otimes \mathbf{B} := \begin{bmatrix} \mathbf{B}a_{1,1} & \cdots & \mathbf{B}a_{1,F_1} \\ \vdots & \vdots & \vdots \\ \mathbf{B}a_{I,1} & \cdots & \mathbf{B}a_{I,F_1} \end{bmatrix}.$$

Given two matrices $\mathbf{A}$ $(I \times F)$ and $\mathbf{B}$ $(J \times F)$ with the same number of columns, the Khatri–Rao product $\mathbf{B} \odot \mathbf{A}$ is the $IJ \times F$ matrix defined as

$$\mathbf{B} \odot \mathbf{A} := [\, \mathbf{b}_1 \otimes \mathbf{a}_1 \quad \cdots \quad \mathbf{b}_F \otimes \mathbf{a}_F \,] = \begin{bmatrix} \mathbf{A}\mathbf{D}_0(\mathbf{B}) \\ \mathbf{A}\mathbf{D}_1(\mathbf{B}) \\ \vdots \\ \mathbf{A}\mathbf{D}_{J-1}(\mathbf{B}) \end{bmatrix} \tag{1}$$

where $\mathbf{a}_f$ is the fth column of $\mathbf{A}$, and similarly for $\mathbf{b}_f$ (i.e., the Khatri–Rao product is a column-wise Kronecker product), and $\mathbf{D}_j(\cdot)$ is an operator that extracts the $(j+1)$th row of its matrix argument and constructs a diagonal matrix out of it.

II. Data Model for Blind Beamforming

Fig. 1 is a schematic of the communications scenario under consideration. A total of ρ sources are transmitting to a base station, which uses an array of M receive antennas. The baseband-equivalent multipath propagation model for one source with r_1 associated paths can be written as

$$\mathbf{x}(t) = \mathbf{h}(t) * s(t) \tag{2}$$

where

- $\mathbf{x}(t)$ vector consisting of the M receive antenna outputs at time t;
- $s(t)$ (scalar) source signal at time t;
- $*$ convolution;

and the $M \times 1$ impulse response vector $\mathbf{h}(t)$ can be modeled as

$$\mathbf{h}(t) = \sum_{j=1}^{r_1} \mathbf{a}(\theta_j)\beta_j g(t - \tau_j)$$

where

- θ_j direction of arrival of the jth ray;
- $\mathbf{a}(\theta)$ array steering vector;
- β_j complex path loss that collects the overall attenuation and phase shift;
- $g(t)$ pulse shape function;
- τ_j path delay for the jth ray.

In incoherent multipath with small delay spread, we assume that $g(t)$ is zero outside an interval $[0, L_g)$ with $L_g < 1 = T$ (the common normalized pulse period), and the delay spread is small enough for $L_g + \max(\tau_{ij}) < 1$, where τ_{ij} is the delay of the jth ray of the ith source. This means that every sample of the received signal is a combination of ρ and not more than ρ source symbols. Let r_i be the number of rays associated with source i, and let $r = \sum_{i=1}^{\rho} r_i$ be the total number of rays. Then, the received baseband signal at the output of the base station antenna array can be expressed as

$$\mathbf{x}(t) = \sum_{i=1}^{\rho} \mathbf{h}_i(t) * s_i(t) = \sum_{i=1}^{\rho} \sum_{j=1}^{r_i} \mathbf{a}(\theta_{ij})\beta_{ij} g(t - \tau_{ij}) s_i(t)$$

where $s_i(t)$ is the information-bearing signal of the ith source at time t (held constant over T), and it is assumed that all sources employ a common pulse shape. Suppose that excess bandwidth is available [22] for sample $x(t)$ at a rate of P times the symbol rate, and collect samples during N symbol periods. Then, construct an $MP \times N$ data matrix[1] $\mathbf{X}$

$$\mathbf{X} := \begin{bmatrix} \mathbf{x}(0) & \mathbf{x}(1) & \cdots & \mathbf{x}(N-1) \\ \mathbf{x}\left(\frac{1}{P}\right) & \mathbf{x}\left(1+\frac{1}{P}\right) & \cdots & \mathbf{x}\left(N-1+\frac{1}{P}\right) \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{x}\left(\frac{P-1}{P}\right) & \mathbf{x}\left(1+\frac{P-1}{P}\right) & \cdots & \mathbf{x}\left(N-1+\frac{P-1}{P}\right) \end{bmatrix}.$$

Under our assumptions, $g(t)$ is nonzero between $[0, L_g)$ and $L_g + \max(\tau_{ij}) < 1$, and thus, $g(t - \tau_{ij})$ is possibly nonzero only within $[0, 1)$ for any i and j. Define

$$\mathbf{g}(\tau) = \begin{bmatrix} g(0-\tau) & g\left(\frac{1}{P}-\tau\right) & \cdots & g\left(\frac{P-1}{P}-\tau\right) \end{bmatrix}^T$$

where $(\cdot)^T$ denotes matrix transpose, and construct a vector $\mathbf{h}_i$ with samples of $\mathbf{h}_i(t)$, which is the impulse response vector for the ith source:

$$\mathbf{h}_i = \begin{bmatrix} \mathbf{h}_i(0) & \mathbf{h}_i\left(\frac{1}{P}\right) & \cdots & \mathbf{h}_i\left(\frac{P-1}{P}\right) \end{bmatrix}^T.$$

$\mathbf{h}_i$ can be expressed as [22]

$$\mathbf{h}_i = \sum_{j=1}^{r_i} (\mathbf{g}_{ij} \otimes \mathbf{a}_{ij})\beta_{ij}, \qquad \mathbf{g}_{ij} := \mathbf{g}(\tau_{ij}), \qquad \mathbf{a}_{ij} := \mathbf{a}(\theta_{ij}).$$

Recall that $r = \sum_{i=1}^{\rho} r_i$ is the total number of rays for all sources. Let us conveniently index the rays (and corresponding angle, path loss, and delay parameters) from 1 to r, starting with all rays associated with the first source and then all rays associated with the second source, etc., and define

$$\begin{aligned} \mathbf{A}_\theta &= [\mathbf{a}(\theta_1) \cdots \mathbf{a}(\theta_r)] && : M \times r \\ \mathbf{\Gamma} &= \text{diag}[\beta_1 \cdots \beta_r] && : r \times r \\ \mathbf{G}_\tau &= [\mathbf{g}(\tau_1) \cdots \mathbf{g}(\tau_r)] && : P \times r \end{aligned}$$

and a selection matrix $\mathbf{J}$ that joins rays associated with a given source

$$\mathbf{J} := \begin{bmatrix} \mathbf{1}_{r_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{1}_{r_\rho} \end{bmatrix}$$

where $\mathbf{1}_m$ denotes an $m \times 1$ vector consisting of 1s. Then, it can be shown [22] that

$$\mathbf{H} := [\mathbf{h}_1 \cdots \mathbf{h}_\rho] = (\mathbf{G}_\tau \odot \mathbf{A}_\theta)\mathbf{\Gamma}\mathbf{J} \tag{3}$$

and

$$\mathbf{X} = (\mathbf{G}_\tau \odot \mathbf{A}_\theta)\mathbf{\Gamma}\mathbf{J}\mathbf{S} \tag{4}$$

where

$$\mathbf{S} = \begin{bmatrix} s_1(0) & s_1(1) & \cdots & s_1(N-1) \\ s_2(0) & s_2(1) & \cdots & s_2(N-1) \\ \vdots & \vdots & & \vdots \\ s_\rho(0) & s_\rho(1) & \cdots & s_\rho(N-1) \end{bmatrix}.$$

Equation (4) can also be written as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_0 \\ \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_{P-1} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_\theta \mathbf{D}_0(\mathbf{G}_\tau) \\ \mathbf{A}_\theta \mathbf{D}_1(\mathbf{G}_\tau) \\ \vdots \\ \mathbf{A}_\theta \mathbf{D}_{P-1}(\mathbf{G}_\tau) \end{bmatrix} \mathbf{\Gamma}\mathbf{J}\mathbf{S}. \tag{5}$$

If $g(t)$ is bandlimited and sampled at or above the Nyquist rate, then taking the DFT of each oversampled antenna output over a single symbol period, the following model is obtained [22]:

$$\overline{\mathbf{X}} = (\mathbf{F}_\phi \odot \mathbf{A}_\theta)\mathbf{\Gamma}\mathbf{J}\mathbf{S} \tag{6}$$

which can also be written as

$$\overline{\mathbf{X}} = \begin{bmatrix} \overline{\mathbf{X}}_0 \\ \overline{\mathbf{X}}_1 \\ \vdots \\ \overline{\mathbf{X}}_{P-1} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_\theta \mathbf{D}_0(\mathbf{F}_\phi) \\ \mathbf{A}_\theta \mathbf{D}_1(\mathbf{F}_\phi) \\ \vdots \\ \mathbf{A}_\theta \mathbf{D}_{P-1}(\mathbf{F}_\phi) \end{bmatrix} \mathbf{\Gamma}\mathbf{J}\mathbf{S} \tag{7}$$

where

$$\mathbf{F}_\phi = \begin{bmatrix} 1 & \cdots & 1 \\ \phi_1 & \cdots & \phi_r \\ \vdots & & \vdots \\ \phi_1^{P-1} & \cdots & \phi_r^{P-1} \end{bmatrix}$$

$$\phi_\ell = e^{-j(2\pi/P)\tau_\ell}, \qquad \ell = 1, \cdots, r.$$

The advantage of (7) versus (5) is that the former exhibits convenient structure in the sense that $\mathbf{F}_\phi$ is a Vandermonde matrix.

[1]The same symbol (e.g., $\mathbf{X}$) is occasionally used to denote different things for the benefit of simplicity and intuition; the meaning will always be clear from the context.

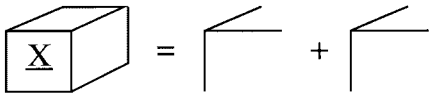

Fig. 2. Schematic of rank-two decomposition of a three-way array as a sum of two rank-one three-way factors.

We will return to (5) and (7) [equivalently, (4) and (6)] after introducing the reader to PARAFAC ideas plus some new preparatory results.

III. Background on PARAFAC and the Uniqueness of Trilinear Decompositions

Consider an $I \times J \times K$ *three-way array* $\underline{\mathbf{X}}$ with typical element $x_{i,j,k}$ and the F*-component trilinear decomposition*

$$x_{i,j,k} = \sum_{f=0}^{F-1} a_{i,f} b_{j,f} c_{k,f}. \tag{8}$$

Equation (8) expresses the three-way array $\underline{\mathbf{X}}$ as a sum of F rank-one three-way factors; a schematic of such a decomposition is given in Fig. 2. Analogous to the definition of matrix (two-way array) rank, the rank of a three-way array $\underline{\mathbf{X}}$ can be defined as the minimum number of rank-one (three-way) components needed to decompose $\underline{\mathbf{X}}$. The trilinear decomposition is also known as canonical decomposition, triple-product decomposition, and PARAFAC analysis [7]–[9], [11]. A fundamental (and quite surprising) transition takes place when one moves from matrices (two-way arrays) to three-way arrays: rank-F matrix decompositions are not unique for any $F > 1$, whereas low-rank three-way array decomposition (PARAFAC) is essentially unique for an interesting range of $F > 1$. In order to make this precise, we need a few definitions.

Define an $I \times F$ matrix $\mathbf{A}$ with typical element $\mathbf{A}(i, f) := a_{i,f}$, $J \times F$ matrix $\mathbf{B}$ with typical element $\mathbf{B}(j, f) := b_{j,f}$ and $K \times F$ matrix $\mathbf{C}$ with typical element $\mathbf{C}(k, f) := c_{k,f}$. Furthermore, define $J \times K$ matrices $\mathbf{X}_i$, $I \times K$ matrices $\mathbf{Y}_j$, and $I \times J$ matrices $\mathbf{Z}_k$ with corresponding typical elements $\mathbf{X}_i(j, k) := \mathbf{Y}_j(i, k) := \mathbf{Z}_k(i, j) := x_{i,j,k}$. Then, the model in (8) can be written in three different ways in terms of systems of simultaneous matrix equations (each of which can be interpreted as "slicing" the three-way array $\underline{\mathbf{X}}$ along one of the three dimensions):

$$\mathbf{X}_i = \mathbf{B}\mathbf{D}_i(\mathbf{A})\mathbf{C}^T, \qquad i = 0, 1, \cdots, I-1 \tag{9}$$

$$\mathbf{Y}_j = \mathbf{A}\mathbf{D}_j(\mathbf{B})\mathbf{C}^T, \qquad j = 0, 1, \cdots, J-1 \tag{10}$$

$$\mathbf{Z}_k = \mathbf{A}\mathbf{D}_k(\mathbf{C})\mathbf{B}^T, \qquad k = 0, 1, \cdots, K-1. \tag{11}$$

Stacking the data in, e.g., (9), into a $JI \times K$ matrix $\mathbf{U}$ and using the definition of the Khatri–Rao product in (1), it follows that

$$\mathbf{U} = (\mathbf{A} \odot \mathbf{B})\mathbf{C}^T. \tag{12}$$

Corresponding compact matrix representations can also be derived from either (10) or (11).

The following concept is key for PARAFAC uniqueness, and it will also play an essential role in the derivation of identifiability results for the deterministic blind beamforming problems considered herein.

Definition 1: Given $\mathbf{A} \in \mathbb{C}^{I \times F}$, $r_{\mathbf{A}} := \text{rank}(\mathbf{A}) = r$ iff it contains *at least a* collection of r linearly independent columns but no collection of $r+1$ linearly independent columns. The *Kruskal*-rank (or k-rank for short—which is a term coined by Harshman and Lundy [10]) of $\mathbf{A}$ is denoted by $k_{\mathbf{A}}$. $k_{\mathbf{A}} = k$ if *every* k columns of $\mathbf{A}$ are linearly independent, but either $k = F$, or there exists a collection of $k+1$ linearly dependent columns in $\mathbf{A}$ ($k_{\mathbf{A}} \leq r_{\mathbf{A}} \leq \min(I, F)$, $\forall\, \mathbf{A}$).

PARAFAC uniqueness has a rich history [9], [11], [12], [14]. Among the various results, the one due to Kruskal [11] is the deepest. Kruskal's result was recently generalized to the complex-valued case by Sidiropoulos *et al.* [18].

Theorem 1 ($\mathbb{R}$*-[11];* $\mathbb{C}$*-[18]):* Given $\mathbf{X}_i = \mathbf{B}\mathbf{D}_i(\mathbf{A})\mathbf{C}^T$, $i = 0, 1, \cdots, I-1$, $\mathbf{A} \in \mathbb{C}^{I \times F}$, $\mathbf{B} \in \mathbb{C}^{J \times F}$, $\mathbf{C} \in \mathbb{C}^{K \times F}$, if

$$k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2(F+1) \tag{13}$$

then $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are unique up to common permutation and (complex) scaling[2] of columns. ■

There exist several practical algorithms that compute trilinear decompositions. These range from generalized eigenvalue decomposition to iterative algorithms based on alternating least squares (ALS), Gauss–Newton, and related optimization techniques. The choice of algorithm depends on application-specific characteristics such as data dimensionality and rank (eigenvalue methods demand much stricter conditions than what is required by Theorem 1), noise level and color, etc. ALS combined with simple acceleration schemes, such as relaxation, is currently perhaps the best shot at a one-size-fits-all algorithm; cf., [3], [5], [18], and [19].

IV. Main Results

The following Lemmas relate to the k-rank of certain structured matrices and will prove useful in the sequel. Proofs can be found in the Appendix. The first Lemma shows that the k-rank of the Khatri–Rao product is almost sup-additive.

Lemma 1 (k*-rank of Khatri–Rao Product):* If neither $\mathbf{A}$ ($I \times F$) nor $\mathbf{B}$ ($J \times F$) contains a zero column (and, hence, $k_{\mathbf{A}} \geq 1$, $k_{\mathbf{B}} \geq 1$), then $k_{\mathbf{B} \odot \mathbf{A}} \geq \min(k_{\mathbf{A}} + k_{\mathbf{B}} - 1, F)$. ■

Remark 1: Note that since rank is greater than or equal to k-rank, Lemma 1 provides a useful lower bound on the rank of the Khatri–Rao product.

The second Lemma deals with the k-rank of Vandermonde matrices.

Lemma 2 (Vandermonde k*-rank Lemma):* An $m \times n$ Vandermonde matrix

$$\mathbf{V} := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_n^2 \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_1^{m-1} & \alpha_2^{m-1} & \cdots & \alpha_n^{m-1} \end{bmatrix}$$

[2]Meaning that it is possible to permute the columns of one matrix, provided the columns of the other two matrices are permuted in the same way, and that it is possible to scale for example the first column of any two matrices, provided the first column of the third matrix is counter-scaled such that the scale of the rank-one factor constructed from the outer product of the three columns remains unaffected. This is what is referred to as permutation and scale ambiguity in the sequel.

with distinct nonzero generators $\alpha_1, \alpha_2, \cdots, \alpha_n \in \mathbb{C}$ is not only full rank but is also full k rank: $k_{\mathbf{V}} = r_{\mathbf{V}} = \min(m, n)$. Note that the latter fails in general for the transpose of a Vandermonde matrix. ■

We now return to (4) and (6). Notice that both equations describe the same PARAFAC model viewed in different domains. However, due to the presence of the selection matrix $\mathbf{J}$, $\mathbf{\Gamma JS}$ has collinear rows, and hence, $(\mathbf{\Gamma JS})^T$ has collinear columns and, hence, is k-rank one. This means that Theorem 1 fails to prove identifiability of (4)–(6). We will prove identifiability by capitalizing on Vandermonde structure to build up the k-rank. This is the subject of the next two Theorems and their corollaries.

Theorem 2: Given $\mathbf{X}_i = \mathbf{B}\mathbf{D}_i(\mathbf{A})\mathbf{C}^T$, $i = 0, 1, \cdots, I-1$, $\mathbf{A} \in \mathbb{C}^{I\times F}$, $\mathbf{B} \in \mathbb{C}^{J\times F}$, $\mathbf{C} \in \mathbb{C}^{K\times F}$ with $\mathbf{A}$ Vandermonde with distinct nonzero generators, if

$$k_{\mathbf{B}} + \min(I + k_{\mathbf{C}}, 2F) \geq 2(F+1) \tag{14}$$

then $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are identifiable up to permutation and scaling of columns. ■

Remark 2: Condition (14) is clearly more relaxed than what is available without explicitly capitalizing on the Vandermonde structure of $\mathbf{A}$. Recall that the generic k-rank identifiability condition for the trilinear model is

$$k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2(F+1)$$

which, even after claiming full k-rank for $\mathbf{A}$ due to its Vandermonde structure, yields

$$k_{\mathbf{B}} + \min(I, F) + k_{\mathbf{C}} \geq 2(F+1).$$

Recalling that k-rank is less than or equal to rank, the latter may also be written as

$$k_{\mathbf{B}} + \min(I, F) + \min(k_{\mathbf{C}}, F) \geq 2(F+1)$$

from which it is clear that (14) is an improvement.

In the context of (6), $\mathbf{\Gamma JS}$ plays the role of $\mathbf{C}^T$, whereas $(\mathbf{F}_\phi, \mathbf{A}_\theta)$ play the role of $(\mathbf{B}, \mathbf{A})$ or $(\mathbf{A}, \mathbf{B})$, depending on which one is assumed to be Vandermonde.

Corollary 1: If $\mathbf{F}_\phi$ is Vandermonde with distinct nonzero generators (distinct path delays and P at or above Nyquist), the path angles are distinct, and any M spatial manifold vectors corresponding to distinct angles are linearly independent, then the model in (6) is identifiable, provided that

$$\min(r, M) + \min(2r, P+1) \geq 2r + 2. \tag{15}$$

As a special case of (15), if the fractional sampling factor $P \geq 2r-1$, then $M = 2$ antennas are sufficient for r paths, and hence, the system is capable of supporting many more paths than sensors, provided sufficient fractional sampling diversity (in terms of excess bandwidth) is available. This is not a stringent requirement (in theory, one only requires the generators of the Vandermonde matrix $\mathbf{F}_\phi$ to be nonzero and distinct), but performance will inevitably be affected when very limited excess bandwidth is available.

Thus far, we have capitalized on the Vandermonde structure of $\mathbf{F}_\phi$, but we can also exploit Vandermonde structure in $\mathbf{A}_\theta$, if available. This involves little additional effort, given the perfect symmetry of the trilinear model [cf., (9)–(11)].

Corollary 2: If $\mathbf{A}_\theta$ is Vandermonde (corresponding to a uniform linear array) with distinct nonzero generators (distinct path angles and no spatial aliasing), and $\mathbf{G}_\tau$ is full k-rank [for distinct path delays, this can be achieved by proper pulse-shape design, e.g., $g(t) = 1 - t$, $t \in (0, 1)$], then the model in (4) is identifiable, provided that

$$\min(r, P) + \min(2r, M+1) \geq 2r + 2. \tag{16}$$

Consider a special case of (16). If $M \geq 2r-1$, then fractional sampling rate $P = 2$ is sufficient for identifiability.

Remark 3: The latter result does not require Vandermonde structure on $\mathbf{F}_\phi$ and, hence, can be applied to the data in the original (time) domain without the need to take the DFT. Note that working in the time domain also bypasses the Nyquist sampling (or above) issue altogether.

Remark 4: Note the dual roles played by M (antenna diversity) and P (fractional sampling diversity) in the above corollaries. When both $\mathbf{F}_\phi$ and $\mathbf{A}_\theta$ are Vandermonde, duality simplifies to symmetry, as we will see shortly.

The proof of Theorem 2 (cf., the Appendix) employs one-dimensional (1-D) smoothing [15] and involves optimization of the smoothing factor from the viewpoint of proving the best identifiability result possible. If the data model exhibits exponential structure along two of the three modes (and, hence, two of the three matrices are Vandermonde), then one can, of course, smooth in one dimension and claim full k-rank of the matrix corresponding to the second exponential mode (cf., Lemma 2). However, one wonders if stronger identifiability results can be obtained via 2-D (spatio-temporal) smoothing. This is explored in the following Theorem, whose proof can be found in the Appendix.

Theorem 3: Given $\mathbf{X}_i = \mathbf{B}\mathbf{D}_i(\mathbf{A})\mathbf{C}^T$, $i = 0, 1, \cdots, I-1$, $\mathbf{A} \in \mathbb{C}^{I\times F}$, $\mathbf{B} \in \mathbb{C}^{J\times F}$, $\mathbf{C} \in \mathbb{C}^{K\times F}$, where $\mathbf{A}$ and $\mathbf{B}$ are Vandermonde matrices with distinct nonzero generators, if $I + J + k_{\mathbf{C}} \geq 2F + 2$, then $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are identifiable up to permutation and scaling of columns. ■

Remark 5: On the surface, the condition $I+J+k_{\mathbf{C}} \geq 2F+2$ appears to be more relaxed than (14), which, after claiming full k-rank of the Vandermonde matrix $\mathbf{B}$, yields

$$\min(J, F) + \min(I + k_{\mathbf{C}}, 2F) \geq 2F + 2.$$

However, observe that one may, without loss of generality, always assume that $J = \min(I, J)$, and some work shows that, in fact, we have the following lemma.

Lemma 3 (Identifiability Condition Equivalence): For $J = \min(I, J)$

$$\min(J, F) + \min(I + k_{\mathbf{C}}, 2F) \geq 2F + 2 \Leftrightarrow I + J + k_{\mathbf{C}} \geq 2F + 2.$$ ■

The interesting conclusion is that in contrast to 1-D smoothing, which is indeed helpful in deriving identifiability results, 2-D smoothing does not further improve things.

The following results are obtained as corollaries of Theorem 3.

Corollary 3: For $r > \rho$ (multiple rays per source), if $\mathbf{F}_\phi$ is Vandermonde with distinct nonzero generators (distinct path delays and P at or above Nyquist), $\mathbf{A}_\theta$ is Vandermonde (corresponding to a uniform linear array) with distinct nonzero generators (distinct path angles and no spatial aliasing), then the model in (6) is identifiable, provided that $M + P > 2r$, i.e., the sum of spatial and fractional sampling diversities strictly exceeds the total number of distinct paths, irrespective of the number of snapshots N.

Corollary 4: For $r = \rho$ (single ray per source), if $\mathbf{S}$ is tall ($N \geq \rho$) and full rank, then under the remaining assumptions in Corollary 3, the model in (6) is identifiable, provided $M+P > r+1$.

As alluded to earlier on, note the complete symmetry in the roles of M (antenna diversity) and P (fractional sampling diversity).

V. Conclusions

We have investigated identifiability issues in the context of deterministic blind beamforming in the presence of incoherent multipath with small delay spread. The viewpoint derives from the theory of low-rank decomposition of multiway arrays, known as PARAFAC analysis, which was shown to be the core problem underlying deterministic blind beamforming. Using a series of new Lemmas and spatio-temporal smoothing with optimized choice of smoothing factor(s) in one and two dimensions, we have proven a variety of relaxed and precise identifiability results, which improve on what is available in [22]. For example, we have shown that if the *sum* of spatial and fractional sampling diversities exceeds two times the total number of paths, then identifiability can be guaranteed. The link to PARAFAC also implies that generic PARAFAC least squares model fitting algorithms are directly applicable to deterministic blind beamforming.

In the context of PARAFAC theory, our results prove that improved uniqueness properties are applicable if a given trilinear model exhibits Vandermonde structure along one or two modes; therefore, the additional structure indeed pays off. This is a contribution to the theory of multiway analysis. Interestingly, in contrast to 1-D smoothing, which is indeed helpful in establishing identifiability results, 2-D smoothing does not seem to add anything in this respect.

Appendix

Proof of Lemma 1: We will need the following Lemmas.

Lemma 4 (Rank of Khatri–Rao Product [19]): Given $\mathbf{A}$ ($I \times F$) and $\mathbf{B}$ ($J \times F$), if $k_\mathbf{A} + k_\mathbf{B} \geq F + 1$, then $\mathbf{B} \odot \mathbf{A}$ is full column rank. ■

Lemma 5: Let $\mathbf{A}$ be an $I \times F$ matrix, and let $\tilde{\mathbf{A}}$ be an $I \times d$ matrix constructed using any $d \leq F$ columns of $\mathbf{A}$. Then

$$\min(d, k_\mathbf{A}) \leq k_{\tilde{\mathbf{A}}} \leq d \tag{17}$$

where k_A stands for the k-rank of $\mathbf{A}$. ■

Proof: Clearly, since $\tilde{\mathbf{A}}$ has d columns, $k_{\tilde{\mathbf{A}}} \leq d$. Every $k_\mathbf{A}$ columns of $\mathbf{A}$ are linearly independent, and hence, every $\min(d, k_\mathbf{A})$ columns of $\tilde{\mathbf{A}}$ are linearly independent, and the Lemma follows. ■

For a proof of Lemma 1, consider d columns of $\mathbf{B} \odot \mathbf{A}$ arising out of the corresponding d columns of $\mathbf{A}$ and $\mathbf{B}$. Let $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ denote matrices formed out of these columns of $\mathbf{A}$ and $\mathbf{B}$. According to Lemma 4, if $d+1 \leq k_{\tilde{\mathbf{A}}} + k_{\tilde{\mathbf{B}}}$, then the d columns of $\mathbf{B} \odot \mathbf{A}$ under consideration are guaranteed to be linearly independent. According to Lemma 5, this will be guaranteed *a fortiori*, provided

$$d + 1 \leq \min(d, k_\mathbf{A}) + \min(d, k_\mathbf{B}). \tag{18}$$

In light of (18), we consider the following cases.

- If $d \leq \min(k_\mathbf{A}, k_\mathbf{B})$, then $\min(d, k_\mathbf{A}) + \min(d, k_\mathbf{B}) = 2d \geq d + 1$.
- If $\min(k_\mathbf{A}, k_\mathbf{B}) \leq d \leq \max(k_\mathbf{A}, k_\mathbf{B})$, then $\min(d, k_\mathbf{A}) + \min(d, k_\mathbf{B}) = d + \min(k_\mathbf{A}, k_\mathbf{B}) \geq d + 1$.
- If $d \geq \max(k_\mathbf{A}, k_\mathbf{B})$, then $\min(d, k_\mathbf{A}) + \min(d, k_\mathbf{B}) = k_\mathbf{A} + k_\mathbf{B}$.

Therefore for $d + 1 \leq k_\mathbf{A} + k_\mathbf{B}$, i.e., $d \leq k_\mathbf{A} + k_\mathbf{B} - 1$, any d columns of $\mathbf{B} \odot \mathbf{A}$ are guaranteed to be linearly independent. Note that $\mathbf{B} \odot \mathbf{A}$ is $IJ \times F$; therefore

$$k_{\mathbf{B} \odot \mathbf{A}} \geq \min(k_\mathbf{A} + k_\mathbf{B} - 1, F) \tag{19}$$

and thus, the proof is complete. ■

Proof of Lemma 2: The interesting case is when $\mathbf{V}$ is fat, i.e., $m < n$. Any subset of columns of a Vandermonde matrix forms a Vandermonde matrix (but this is not necessarily true for subsets of rows of a Vandermonde matrix). Therefore, any selection of up to m columns of $\mathbf{V}$ will constitute a tall/square Vandermonde matrix with distinct nonzero generators, which is full column rank (e.g., [20, p. 274]). ■

Proof of Theorem 2: Consider the $IJ \times K$ matrix obtained by stacking the data $\mathbf{X}_i$

$$\mathbf{X} := \begin{bmatrix} \mathbf{B}\mathbf{D}_0(\mathbf{A}) \\ \mathbf{B}\mathbf{D}_1(\mathbf{A}) \\ \vdots \\ \mathbf{B}\mathbf{D}_{I-1}(\mathbf{A}) \end{bmatrix} \mathbf{C}^T.$$

Define m matrices of size $LJ \times K$

$$\overline{\mathbf{X}}^{(\mu)} := \mathbf{X}(J\mu + 1{:}\, J(L+\mu), :), \qquad \mu = 0, 1, \cdots, m-1$$

where $\mathbf{X}(J\mu+1{:}\, J(L+\mu), :)$ stands for rows $J\mu+1$ to $J(L+\mu)$ (inclusive) of matrix $\mathbf{X}$, m is known as the smoothing factor [15], and $L := I - m + 1$. Note that due to the Vandermonde structure of

$$\mathbf{A} = \begin{bmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_F \\ & \vdots & \\ a_1^{I-1} & \cdots & a_F^{I-1} \end{bmatrix}$$

it holds that

$$\overline{\mathbf{X}}^{(\mu)} = \begin{bmatrix} \mathbf{B}\mathbf{D}_\mu(\mathbf{A}) \\ \mathbf{B}\mathbf{D}_{\mu+1}(\mathbf{A}) \\ \vdots \\ \mathbf{B}\mathbf{D}_{L+\mu-1}(\mathbf{A}) \end{bmatrix} \mathbf{C}^T = \begin{bmatrix} \mathbf{B}\mathbf{D}_0(\mathbf{A}) \\ \mathbf{B}\mathbf{D}_1(\mathbf{A}) \\ \vdots \\ \mathbf{B}\mathbf{D}_{L-1}(\mathbf{A}) \end{bmatrix} \mathbf{\Phi}_\mathbf{A}^\mu \mathbf{C}^T$$

where

$$\mathbf{\Phi_A} = \text{diag}(a_1, a_2, \cdots, a_F).$$

Laying out these m matrices in parallel[3]

$$\overline{\mathbf{X}} := \begin{bmatrix} \overline{\mathbf{X}}^{(0)} & \cdots & \overline{\mathbf{X}}^{(m-1)} \end{bmatrix} = \begin{bmatrix} \mathbf{B}\mathbf{D}_0(\mathbf{A}) \\ \mathbf{B}\mathbf{D}_1(\mathbf{A}) \\ \vdots \\ \mathbf{B}\mathbf{D}_{L-1}(\mathbf{A}) \end{bmatrix} \begin{bmatrix} \mathbf{C}^T & \mathbf{\Phi_A}\mathbf{C}^T & \cdots & \mathbf{\Phi}_\mathbf{A}^{m-1}\mathbf{C}^T \end{bmatrix}.$$

Note that

$$\begin{bmatrix} \mathbf{C}^T & \mathbf{\Phi_A}\mathbf{C}^T & \cdots & \mathbf{\Phi}_\mathbf{A}^{m-1}\mathbf{C}^T \end{bmatrix}^T = \begin{bmatrix} \mathbf{C}\mathbf{D}_0(\mathbf{A}) \\ \mathbf{C}\mathbf{D}_1(\mathbf{A}) \\ \vdots \\ \mathbf{C}\mathbf{D}_{m-1}(\mathbf{A}) \end{bmatrix} =: \mathbf{A}^{(m)} \odot \mathbf{C}$$

where $\mathbf{A}^{(m)}$ is a Vandermonde matrix constructed out of the first m rows of $\mathbf{A}$. Similarly

$$\begin{bmatrix} \mathbf{B}\mathbf{D}_0(\mathbf{A}) \\ \mathbf{B}\mathbf{D}_1(\mathbf{A}) \\ \vdots \\ \mathbf{B}\mathbf{D}_{L-1}(\mathbf{A}) \end{bmatrix} =: \mathbf{A}^{(L)} \odot \mathbf{B}$$

where $\mathbf{A}^{(L)}$ is a Vandermonde matrix constructed out of the first L rows of $\mathbf{A}$. It follows that

$$\overline{\mathbf{X}} = \left(\mathbf{A}^{(L)} \odot \mathbf{B}\right)\left(\mathbf{A}^{(m)} \odot \mathbf{C}\right)^T$$

which is a special PARAFAC model with double Khatri–Rao structure. It follows from Theorem 1 that $\mathbf{B}$, $\mathbf{A}^{(L)}$ (and, therefore, $\mathbf{A}$ as well), and $\mathbf{A}^{(m)} \odot \mathbf{C}$ are unique up to permutation and scaling of columns (and, hence, the same holds for $\mathbf{C}$), provided that

$$k_\mathbf{B} + k_{\mathbf{A}^{(L)}} + k_{\mathbf{A}^{(m)} \odot \mathbf{C}} \geq 2(F+1).$$

Using Lemmas 1 and 2, this yields

$$k_\mathbf{B} + \min(L, F) + \min(k_{\mathbf{A}^{(m)}} + k_\mathbf{C} - 1, F) \geq 2(F+1).$$

Invoking Lemma 2 once more gives

$$k_\mathbf{B} + \min(L, F) + \min(\min(m, F) + k_\mathbf{C} - 1, F) \geq 2(F+1)$$

or equivalently

$$k_\mathbf{B} + \min(L, F) + \min(m + k_\mathbf{C} - 1, F) \geq 2(F+1).$$

Recall $m = I - L + 1$ to obtain

$$\Gamma(L) := k_\mathbf{B} + \min(L, F) + \min(I + k_\mathbf{C} - L, F) \geq 2(F+1).$$

Choosing $L = \min(I, F)$ leads to $\Gamma(\min(I, F)) = k_\mathbf{B} + \min(I, F) + \min(I + k_\mathbf{C} - \min(I, F), F)$. By considering different cases for F relative to the breakpoints $(I + k_\mathbf{C})/2$ and I, it is easy to see that $\Gamma(\min(I, F)) \geq k_\mathbf{B} + \min(2F, I + k_\mathbf{C})$. With $L = \min(I, F)$, the above identifiability condition becomes

$$k_\mathbf{B} + \min(2F, I + k_\mathbf{C}) \geq 2(F+1).$$

This completes the proof. Note again that no better choice of L can be found from the viewpoint of identifiability since $\Gamma(L)$ can be easily shown to be bounded from above by $k_\mathbf{B} + \min(2F, I + k_\mathbf{C})$; hence, with this choice of L, $\Gamma(L)$ actually achieves its upper bound. ■

Proof of Theorem 3: Consider $\mathbf{X}_i = \mathbf{B}\mathbf{D}_i(\mathbf{A})\mathbf{C}^T$, $i = 0, 1, \cdots, I-1$, $\mathbf{A} \in \mathbb{C}^{I \times F}$, $\mathbf{B} \in \mathbb{C}^{J \times F}$, $\mathbf{C} \in \mathbb{C}^{K \times F}$, where $\mathbf{A}$ and $\mathbf{B}$ are Vandermonde matrices with distinct nonzero generators

$$\mathbf{A} = \begin{bmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_F \\ & \vdots & \\ a_1^{I-1} & \cdots & a_F^{I-1} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & \cdots & 1 \\ b_1 & \cdots & b_F \\ & \vdots & \\ b_1^{J-1} & \cdots & b_F^{J-1} \end{bmatrix}.$$

Let

$$\mathbf{X} := \begin{bmatrix} \mathbf{B}\mathbf{D}_0(\mathbf{A}) \\ \mathbf{B}\mathbf{D}_1(\mathbf{A}) \\ \vdots \\ \mathbf{B}\mathbf{D}_{I-1}(\mathbf{A}) \end{bmatrix} \mathbf{C}^T$$

and capitalize on the Vandermonde structure of $\mathbf{A}$ by constructing a new[4] matrix $\overline{\mathbf{X}}$

$$\overline{\mathbf{X}} = \begin{bmatrix} \mathbf{B}\mathbf{D}_0(\mathbf{A}) \\ \mathbf{B}\mathbf{D}_1(\mathbf{A}) \\ \vdots \\ \mathbf{B}\mathbf{D}_{L_1-1}(\mathbf{A}) \end{bmatrix} \begin{bmatrix} \mathbf{C}^T & \mathbf{\Phi_A}\mathbf{C}^T & \cdots & \mathbf{\Phi}_\mathbf{A}^{m-1}\mathbf{C}^T \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{X}}_0 \\ \overline{\mathbf{X}}_1 \\ \vdots \\ \overline{\mathbf{X}}_{L_1-1} \end{bmatrix} \overline{\mathbf{C}}^T, \qquad (L_1 = I - m + 1)$$

where $\overline{\mathbf{X}}$ has size of $L_1 J \times mK$, and

$$\overline{\mathbf{C}}^T = \begin{bmatrix} \mathbf{C}^T & \mathbf{\Phi_A}\mathbf{C}^T & \cdots & \mathbf{\Phi}_\mathbf{A}^{m-1}\mathbf{C}^T \end{bmatrix}$$
$$\mathbf{\Phi_A} = \text{diag}(a_1, a_2, \cdots, a_F).$$

Next, take advantage of the Vandermonde structure of $\mathbf{B}$, and construct a matrix $\tilde{\mathbf{X}}$

$$\tilde{\mathbf{X}} = \begin{bmatrix} \tilde{\mathbf{X}}_0 \\ \tilde{\mathbf{X}}_1 \\ \vdots \\ \tilde{\mathbf{X}}_{L_1-1} \end{bmatrix} \begin{bmatrix} \overline{\mathbf{C}}^T & \mathbf{\Phi_B}\overline{\mathbf{C}}^T & \cdots & \mathbf{\Phi}_\mathbf{B}^{m'-1}\overline{\mathbf{C}}^T \end{bmatrix}$$

where $\tilde{\mathbf{X}}_i$ is a matrix containing the first L_2 rows of $\overline{\mathbf{X}}_i$, and $m' = J - L_2 + 1$. We can further write $\tilde{\mathbf{X}}$ as

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{B}^{(L_2)}\mathbf{D}_0(\mathbf{A}^{(L_1)}) \\ \mathbf{B}^{(L_2)}\mathbf{D}_1(\mathbf{A}^{(L_1)}) \\ \vdots \\ \mathbf{B}^{(L_2)}\mathbf{D}_{L_1-1}(\mathbf{A}^{(L_1)}) \end{bmatrix} \tilde{\mathbf{C}}^T \quad : L_1 L_2 \times m'mK$$

[3]This is not the same $\overline{\mathbf{X}}$ as in (6).

[4]This is not the same $\overline{\mathbf{X}}$ as in the proof of Theorem 2.

where

$$\tilde{\mathbf{C}}^T = \begin{bmatrix} \overline{\mathbf{C}}^T & \mathbf{\Phi_B}\overline{\mathbf{C}}^T & \cdots & \mathbf{\Phi_B}^{m'-1}\overline{\mathbf{C}}^T \end{bmatrix}$$
$$\mathbf{\Phi_B} = \mathrm{diag}(b_1, b_2, \cdots, b_F).$$

The generic identifiability condition for the PARAFAC model is $k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2(F+1)$. It follows that the k-rank identifiability condition for the 2-D smoothed model above is

$$k_{\mathbf{B}^{(L_2)}} + k_{\mathbf{A}^{(L_1)}} + k_{\tilde{\mathbf{C}}} \geq 2(F+1). \tag{20}$$

Recall

$$\begin{aligned} k_{\tilde{\mathbf{C}}} &= k_{\mathbf{B}^{(m')} \odot \overline{\mathbf{C}}} \\ &\geq \min(k_{\mathbf{B}^{(m')}} + k_{\overline{\mathbf{C}}} - 1, F) \\ &= \min(\min(m', F) + \min(m + k_{\mathbf{C}} - 1, F) - 1, F) \\ &= \min(m' + m + k_{\mathbf{C}} - 2, F) \\ &= \min(I + J + k_{\mathbf{C}} - L_1 - L_2, F). \end{aligned}$$

Condition (20) becomes

$$k_{\mathbf{B}^{(L_2)}} + k_{\mathbf{A}^{(L_1)}} + \min(I + J + k_{\mathbf{C}} - L_1 - L_2, F) \geq 2(F+1). \tag{21}$$

Since $\mathbf{A}^{(L_1)}$ and $\mathbf{B}^{(L_2)}$ are Vandermonde matrices with distinct nonzero generators, they have full k-rank. We therefore rewrite (21) as

$$\begin{aligned} \Gamma_2(L_1, L_2) &:= \min(L_1, F) + \min(L_2, F) \\ &\quad + \min(I + J + k_{\mathbf{C}} - L_1 - L_2, F) \\ &\geq 2(F+1). \end{aligned}$$

Select $L_1 = \min(I, F)$, $L_2 = \min(J, F)$, leading to

$$\begin{aligned} &\Gamma_2(\min(I, F), \min(J, F)) \\ &\quad = \min(I, F) + \min(J, F) \\ &\qquad + \min(I + J + k_{\mathbf{C}} - \min(I, F) - \min(J, F), F). \end{aligned}$$

Per the statement of the Theorem, we wish to prove that $I + J + k_{\mathbf{C}} \geq 2F + 2$ is sufficient for identifiability. Consider the following cases:

- $F \leq I$, $F \leq J$. Then

$$\begin{aligned} &\Gamma_2(\min(I, F), \min(J, F)) \\ &\quad = 2F + \min(I + J + k_{\mathbf{C}} - 2F, F) \geq 2F + 2 \end{aligned}$$

since $I+J+k_{\mathbf{C}} \geq 2F+2$ guarantees $I+J+k_{\mathbf{C}}-2F \geq 2$.

- $F \leq I$, $F > J$. Then

$$\Gamma_2(\min(I, F), \min(J, F)) = F + J + \min(I + k_{\mathbf{C}} - F, F).$$

However, $I+J+k_{\mathbf{C}} \geq 2F+2$ guarantees $I+k_{\mathbf{C}}-F \geq F - J + 2$ so we have

$$\Gamma_2(\min(I, F), \min(J, F)) \geq F + J + \min(F - J + 2, F).$$

Suppose $F - J + 2 \leq F$; then, $\Gamma_2(\min(I, F), \min(J, F)) \geq 2F + 2$. If, on the other hand, $F \leq F - J + 2$, then $\Gamma_2(\min(I, F), \min(J, F)) = 2F + J \geq 2F + 2$ since $J \geq 2$ (this is a trivial nondegeneracy condition).

- $F > I$, $F \leq J$. Then, $\Gamma_2(\min(I, F), \min(J, F)) \geq 2F + 2$ by symmetry (needs $I \geq 2$).
- $F > I$, $F > J$. Then

$$\Gamma_2(\min(I, F), \min(J, F)) = I + J + \min(k_{\mathbf{C}}, F).$$

Recall that $k_{\mathbf{C}}$ stands for the k-rank of $\mathbf{C}$, and hence, $k_{\mathbf{C}} \leq F$ to obtain

$$\Gamma_2(\min(I, F), \min(J, F)) = I + J + k_{\mathbf{C}} \geq 2F + 2.$$

We conclude that identifiability is guaranteed if $I+J+k_{\mathbf{C}} \geq 2F + 2$. Note that $\Gamma_2(L_1, L_2) \leq I + J + k_{\mathbf{C}}$, and hence, the condition cannot be further relaxed by means of smoothing. ■

Identifiability Conditions Pertaining to [22]–[25]:

- [22]:

— The $\rho \times N$ (ρ in our notation plays the role of d in [22]) signal matrix $\mathbf{S}$ is fat and full rank.

— The Khatri–Rao product $\mathbf{F}_\phi \odot \mathbf{A}_\theta$ that appears in [22, Eq. (24), p. 2000] is tall ($r \leq MP$) and full rank.

— The Khatri–Rao product $\mathbf{F}_\phi^{(P-m+1)} \odot \mathbf{A}_\theta$ that appears in [22, Eq. (25), p. 2001] (where $\mathbf{F}_\phi^{(P-m+1)}$ consists of the first $P-m+1$ rows of $\mathbf{F}_\phi$, and m is a temporal smoothing factor) is tall ($r \leq M(P - m + 1)$) and full rank. Note that this condition is stronger than (and, hence, implies that) $\mathbf{F}_\phi \odot \mathbf{A}_\theta$ is tall and full rank. In addition, note that there is a typographical error in [22, Eq. (25), p. 2001], where $\mathbf{F}_\phi^{(P-m+1)}$ should appear in place of $\mathbf{F}_\phi \odot \mathbf{A}_\theta$.

— The Khatri–Rao product $\mathbf{F}_\phi^{(m)} \odot \mathbf{T}^T\mathbf{J}^T$ that appears in [22, Eq. (25), p. 2001] (where $\mathbf{F}_\phi^{(m)}$ consists of the first m rows of $\mathbf{F}_\phi$) needs to be tall ($r \leq m\rho$) and full rank.

— The matrices A', A'' that appear in [22, Eq. (26), p. 2001] need to be tall and full rank.

- [23]–[25]:

— These assume that the overall spatio-temporal channel matrix has already been estimated using training symbols and seek to determine path angles and delays by exploiting the Kronecker structure of the spatio-temporal manifold.

— [24] (including precursor results in [23]). This requires similar conditions—$\mathbf{A}_{x\phi}$, $\mathbf{A}_{x\theta}$, $\mathcal{F}_\epsilon^T$ (notation of [24]) need be tall and full rank.

— [25]. The Khatri–Rao product $\mathbf{G}_\tau \odot \mathbf{A}_\theta$ is required to be tall and full column rank, whereas an effective "signal" matrix (actually holding fading coefficients corresponding to different slots in a TDMA system) must be fat and full rank, implying that one needs at least as many channel estimates (computed one per TDMA slot) as the total number of multipath rays. Additional conditions are needed to guarantee that the proposed eigenvalue solution will identify the model parameters. Note that [25] deals with the single-user case.

REFERENCES

[1] B. G. Agee, S. V. Schell, and W. A. Gardner, "Spectral self-coherence restoral: A new approach to blind adaptive signal extraction using antenna arrays," *Proc. IEEE*, vol. 78, pp. 753–767, Apr. 1990.

[2] K. Anand, G. Mathew, and V. U. Reddy, "Blind separation of multiple co-channel BPSK signals arriving at an antenna array," *IEEE Signal Processing Lett.*, vol. 2, pp. 176–178, Sept. 1995.
[3] R. Bro, "PARAFAC: Tutorial and applications," *Chemometrics Intell. Lab. Syst.*, vol. 38, pp. 149–171, 1997.
[4] R. Bro, N. D. Sidiropoulos, and G. B. Giannakis, "Optimal joint azimuth-elevation and signal-array response estimation using parallel factor analysis," in *Proc. 32nd Asilomar Conf. Signals, Syst., Comput.*, Monterey, CA, Nov. 1998.
[5] ——, "A fast least squares algorithm for separating trilinear mixtures," in *Proc. ICA—Int. Workshop Independent Component Anal. Blind Signal Separation*, Aussois, France, Jan. 11–15, 1999, pp. 289–294.
[6] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, pp. 2009–2025, Oct. 1998.
[7] J. D. Carroll and J. Chang, "Analysis of individual differences in multidimensional scaling via an N generalization of 'Eckart–Young' decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.
[8] R. B. Cattell, "Parallel proportional profiles and other principles for determining the choice of factors by rotation," *Psychometrika*, vol. 9, pp. 267–283, 1944.
[9] R. A. Harshman, "Foundations of the PARAFAC procedure: Model and conditions for an 'Explanatory' multi-mode factor analysis," *UCLA Working Papers Phonetics*, vol. 16, pp. 1–84, Dec. 1970.
[10] R. A. Harshman and M. E. Lundy, "The PARAFAC model for three-way factor analysis and multidimensional scaling," in *Research Methods for Multimode Data Analysis*, H. G. Law, C. W. Snyder, Jr., J. Hattie, and R. P. McDonald, Eds. New York: Praeger, 1984, pp. 122–215.
[11] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Applicat.*, vol. 18, pp. 95–138, 1977.
[12] ——, "Rank decomposition, and uniqueness for 3-way and N arrays," in *Multiway Data Analysis*, R. Coppi and S. Bolasco, Eds. Amsterdam, The Netherlands: North-Holland, 1988, pp. 7–18.
[13] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 984–995, July 1989.
[14] P. H. Schonemann, "An algebraic solution for a class of subjective metrics models," *Psychometrika*, vol. 37, p. 441, 1972.
[15] T. J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 806–811, Apr. 1985.
[16] J. J. Shynk and R. P. Gooch, "The constant modulus array for cochannel signal copy and direction finding," *IEEE Trans. Signal Processing*, vol. 44, pp. 652–660, Mar. 1996.
[17] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Deterministic waveform-preserving blind separation of DS-CDMA signals using an antenna array," in *Proc. IEEE SP Workshop Statist. Signal Array Process.*, Portland, OR, Sept. 14–16, 1998, pp. 304–307.
[18] ——, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. Signal Processing*, vol. 48, pp. 810–823, Mar. 2000.
[19] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Signal Processing*, vol. 48, pp. 2377–2388, Aug. 2000.
[20] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1997.
[21] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array—Part I: Algorithms," *IEEE Trans. Signal Processing*, vol. 44, pp. 1184–1197, May 1996.
[22] A.-J. van der Veen, "Algebraic methods for deterministic blind beamforming," *Proc. IEEE*, vol. 86, pp. 1987–2008, Oct. 1998.
[23] A.-J. van der Veen, M. C. Vanderveen, and A. J. Paulraj, "Joint angle and delay estimation using shift-invariance properties," *IEEE Signal Processing Lett.*, vol. 4, pp. 142–145, May 1997.
[24] ——, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Trans. Signal Processing*, vol. 46, pp. 405–418, Feb. 1998.
[25] M. C. Vanderveen, A.-J. van der Veen, and A. J. Paulraj, "Estimation of multipath parameters in wireless communications," *IEEE Trans. Signal Processing*, vol. 46, pp. 682–690, Mar. 1998.

Nicholas D. Sidiropoulos (M'92–SM'99) received the Diploma in electrical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland at College Park (UMCP), in 1988, 1990, and 1992, respectively.

From 1988 to 1992, he was a Fulbright Fellow and a Research Assistant at the Institute for Systems Research (ISR), UMCP. From September 1992 to June 1994, he served his military service as a Lecturer in the Hellenic Air Force Academy. From October 1993 to June 1994, he was a Member of Technical Staff, Systems Integration Division, G-Systems Ltd., Athens, Greece. He held Postdoctoral (from 1994 to 1995) and Research Scientist (from 1996 to 1997) positions at ISR-UMCP before joining the Department of Electrical Engineering, the University of Virginia, Charlottesville, in July 1997 as an Assistant Professor. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, the University of Minnesota, Minneapolis. His current research interests are primarily in multiway analysis and its applications in signal processing for communications and networking.

Dr. Sidiropoulos is a Member of the Signal Processing for Communications Technical Committee (SPCOM-TC) of the IEEE SP Society and currently serves as Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING, and IEEE SIGNAL PROCESSING LETTERS. He received the NSF/CAREER award (Signal Processing Systems Program) in June 1998.

Xiangqian Liu received the B.S. degree from Beijing Institute of Technology, Beijing, China, in 1997 and the M.S. degree from University of Virginia, Charlottesville, in 1999, both in electrical engineering. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis.

His research interests are in the area of signal processing for communications with focus on wireless multiuser communications.