Parallel Factor Analysis in Sensor Array Processing

Nicholas D. Sidiropoulos, *Senior Member, IEEE*, Rasmus Bro, and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—This paper links multiple invariance sensor array processing (MI-SAP) to *parallel factor* (PARAFAC) analysis, which is a tool rooted in psychometrics and chemometrics. PARAFAC is a common name for low-rank decomposition of three- and higher way arrays. This link facilitates the derivation of powerful identifiability results for MI-SAP, shows that the uniqueness of single- and multiple-invariance ESPRIT stems from uniqueness of low-rank decomposition of three-way arrays, and allows tapping on the available expertise for fitting the PARAFAC model. The results are applicable to both data-domain and subspace MI-SAP formulations. The paper also includes a constructive uniqueness proof for a special PARAFAC model.**

I. Introduction

DIRECTION-of-arrival (DOA) estimation is a key problem in radar as well as in emitter localization in mobile wireless communications. In the latter, DOA estimation enables beamforming for interference suppression and user signal separation, and it is also useful in emergency (911) situations. DOA estimation is central to the sensor array processing problem, and it has sparked considerable research interest for more than two decades [19]. Sensor array processing tools range from nonparametric Fourier-based methods and conventional beamforming to parametric high-resolution direction finding techniques like MUSIC [26] and ESPRIT [25].

The single-parameter (azimuth only) case is simpler than the multiple-parameter case, e.g., azimuth and elevation or even additional parameters such as frequency and polarization. ESPRIT is particularly appealing in the single-parameter case, where it provides a solution based on eigenvalue decomposition. The basic idea behind ESPRIT applied to direction finding is to deploy two identical displaced (but otherwise arbitrary) subarrays. This induces rotational invariance in the baseband data that can be exploited to recover the sought-after DOA's. An important feature of ESPRIT is that it does not require subarray calibration information. Obtaining array calibration data is an expensive proposition, and drifts can render the collected data unreliable.

In the multiple-parameter case, things become considerably more complicated. In effect, one needs an array with displacement structure in more than one dimension, and applying ESPRIT separately in each dimension leaves much to be desired. On the one hand, even if azimuth and elevation parameters are independently resolved, the association problem (i.e., pairing corresponding azimuth and elevation estimates) remains. More importantly, this approach does not fully capitalize on the multidimensional invariance structure present in the data.

Several authors have considered ways around the association problem, see e.g., [36] and [42]. The weighted subspace fitting (WSF) viewpoint of Viberg and Ottersten [37] led to the single-parameter multiple-invariance subspace fitting (SSF) formulation of ESPRIT [32] and then to the multiple-parameter multiple-invariance SSF formulation of ESPRIT in [31]. In parallel, several authors have investigated suboptimal but computationally efficient solutions for the multiple-parameter multiple-invariance case (see, e.g., [12] and [41]).

ESPRIT ideas have revolutionized sensor array signal processing. Interestingly, a general principle underlying ESPRIT has flourished independently in other scientific fields and disciplines, where it is commonly referred to in a variety of ways, including parallel proportional profiles, trilinear decomposition, canonical decomposition, and *parallel factor* (PARAFAC) analysis. PARAFAC has been first introduced as a data analysis tool in psychometrics,[1] where it has been used, e.g., for "individual differences multidimensional scaling" but in phonetics, exploratory data analysis, statistics, arithmetic complexity, and other fields and disciplines. R. A. Harshman [14]–[16], and Caroll and Chang [7] developed the PARAFAC model. Nowadays, much of the research in the area is conducted in the context of chemometrics,[2] where it is used for spectrophotometric, chromatographic, and flow injection analyses. This body of work includes tools and concepts that are very useful in deriving identifiability results plus decades of expertise on the algorithmic side of things. The term PARAFAC is widely adopted in chemometrics, and we adopt it here as well.

In an earlier paper [29], we have linked PARAFAC to the DS-CDMA blind multiuser detection problem. In this paper, we wed PARAFAC ideas to multiple-invariance sensor array processing, with emphasis on identifiability results. Relative to the CDMA problem in [29], array invariance replaces spreading as the third dimension, and the information-bearing source signals can now be uncoded.

Manuscript received January 21, 1999; revised March 27, 2000. Part of this paper was presented at the 32nd Asilomar Conference on Signals, Systems, and Computers, November 1-4, 1998, Monterey, CA. This work was supported by NSF CAREER CCR 9733540, NSF CCR 9805350, the Nordic Industry Foundation Project P93-149, and the FØTEK Foundation through Prof. L. Munck. The associate editor coordinating the review of this paper and approving it for publication was Dr. Alex B. Gershman.

N. D. Sidiropoulos was with the Department of Electrical Engineering, University of Virginia, Charlottesville, VA 22903 USA. He is now with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: nikos@ece.umn.edu).

R. Bro is with the Chemometrics Group, Food Technology, Department of Dairy and Food Science, Royal Veternary and Agricultural University, Frederiksberg, Denmark.

G. B. Giannakis is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA.

Publisher Item Identifier S 1053-587X(00)05989-4.

[1]The first seed of PARAFAC ideas appeared as early as 1944 in a paper by R. B. Cattell in *Psychometrika* [9].

[2]Data analysis and its applications in chemistry.

PARAFAC is a common name for low-rank decomposition of three- and higher way arrays. As such, it falls under multiway analysis or linear algebra for multiway arrays. Multiway analysis has also been used in the context of higher order statistics (HOS)-based independent component analysis (ICA) of non-Gaussian source mixtures [10], [22]. Our approach herein is deterministic; it requires neither statistical independence nor non-Gaussianity.

A. Organization

The rest of this paper is structured as follows. Section II contains the MI-SAP data model and provides necessary MI-SAP preliminaries along with a summary of previously known MI-SAP identifiability results. Section III introduces the PARAFAC model and explains how it arises in the context of MI-SAP. Section IV contains the core contributions of this paper, which are in terms of MI-SAP identifiability results. Section V discusses algorithmic issues, whereas Section VI presents simulation results. Conclusions are drawn in Section VII.

II. MI-SAP Preliminaries

Consider a sensor array consisting of I elements receiving signals from M narrowband sources in the far field. The discrete-time baseband-equivalent model for the noisy array outputs can be written as follows:

$$\mathbf{z}(t) = \mathbf{G}\mathbf{s}(t) + \mathbf{v}(t) \tag{1}$$

where $\mathbf{G}$ is the $I \times M$ array response, $\mathbf{s}(t)$ is an $M \times 1$ vector of source signals, and $\mathbf{v}(t)$ models measurement noise. Throughout, we assume that $\mathbf{G}$ is tall $(I > M)$ and full rank, which is necessary for identifiability. Collecting N snapshots,

$$\mathbf{Z} = \mathbf{G}\mathbf{S} + \mathbf{V} \tag{2}$$

where $\mathbf{S}$ is the $M \times N$ source signal matrix, and $\mathbf{V}$ is the $I \times N$ noise matrix. Suppose that the array contains P displaced but otherwise identical subarrays of K sensors each (an extension of the usual ESPRIT scenario, which corresponds to $P = 2$). Let $\mathbf{J}_p$ denote a $K \times I$ selection matrix that extracts the K rows corresponding to the pth subarray; then (see, e.g., [12], [31], [32], [41])

$$\begin{bmatrix} \mathbf{X}_0 \\ \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_{P-1} \end{bmatrix} := \begin{bmatrix} \mathbf{J}_0 \\ \mathbf{J}_1 \\ \vdots \\ \mathbf{J}_{P-1} \end{bmatrix} \mathbf{G}\mathbf{S} = \begin{bmatrix} \mathbf{A}\boldsymbol{\Phi}_0 \\ \mathbf{A}\boldsymbol{\Phi}_1 \\ \vdots \\ \mathbf{A}\boldsymbol{\Phi}_{P-1} \end{bmatrix} \mathbf{S} \tag{3}$$

and

$$\begin{bmatrix} \tilde{\mathbf{X}}_0 \\ \tilde{\mathbf{X}}_1 \\ \vdots \\ \tilde{\mathbf{X}}_{P-1} \end{bmatrix} := \begin{bmatrix} \mathbf{J}_0 \\ \mathbf{J}_1 \\ \vdots \\ \mathbf{J}_{P-1} \end{bmatrix} \mathbf{Z} = \begin{bmatrix} \mathbf{X}_0 \\ \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_{P-1} \end{bmatrix} + \begin{bmatrix} \mathbf{V}_0 \\ \mathbf{V}_1 \\ \vdots \\ \mathbf{V}_{P-1} \end{bmatrix} \tag{4}$$

where

$\mathbf{A}$	$K \times M$ subarray response;
$\boldsymbol{\Phi}_p$	$M \times M$ diagonal depending on source parameters, pth subarray displacement, and frame of reference;
$\mathbf{V}_p = \mathbf{J}_p\mathbf{V}$	measurement noise for the pth subarray.

Throughout this paper, $\tilde{}$ denotes noisy data. Note that in general, $K + P - 1 \le I \le KP$, with left equality for subarrays that share $K - 1$ elements and right equality for subarrays that do not overlap.

As an example, consider a 5×5 uniform square array with interelement spacing $\lambda/2$ in both dimensions, $M = 3$ sources, and $P = 4$ square subarrays corresponding to the bottom-left $(p = 0)$, bottom-right $(p = 1)$, top-left $(p = 2)$, and top-right $(p = 3)$ 4×4 blocks $(K = 16)$. Then, assuming $p = 0$ is the reference subarray [31]

$$\begin{aligned} \boldsymbol{\Phi}_0 &= \mathbf{I} \\ \boldsymbol{\Phi}_1 &= \text{diag}(e^{-j\pi \sin\theta_{1,a}}, e^{-j\pi \sin\theta_{2,a}}, e^{-j\pi \sin\theta_{3,a}}) \\ \boldsymbol{\Phi}_2 &= \text{diag}(e^{-j\pi \cos\theta_{1,a} \sin\theta_{1,e}}, e^{-j\pi \cos\theta_{2,a} \sin\theta_{2,e}} \\ &\qquad \cdot e^{-j\pi \cos\theta_{3,a} \sin\theta_{3,e}}) \\ \boldsymbol{\Phi}_3 &= \boldsymbol{\Phi}_1\boldsymbol{\Phi}_2 \end{aligned}$$

where $\mathbf{I}$ denotes the $M \times M$ identity matrix, and $\theta_{1,a}, \theta_{2,a}, \theta_{3,a}$, and $\theta_{1,e}, \theta_{2,e}, \theta_{3,e}$ are the azimuth and elevation angles for the three sources (in rads).

The core problem that we deal with in this paper can now be stated as follows.

PROBLEM: Given $\mathbf{Z}$, M, find (estimate) $\mathbf{A}$, $\boldsymbol{\Phi}_p, p = 0, 1, \cdots, P-1$, and $\mathbf{S}$.

Returning to (1) and assuming that $\mathbf{v}(t)$ is spatially white

$$\mathbf{R}_{zz} := E\{\mathbf{z}(t)\mathbf{z}^H(t)\} = \mathbf{G}\mathbf{R}_{ss}\mathbf{G}^H + \sigma^2\mathbf{I}$$

where

$E\{\cdot\}$	expectation operator;
^{H}	Hermitian transpose (* is reserved for conjugation);
$\mathbf{R}_{ss}$	emitter signal covariance;
σ^2	noise variance.

If $\mathbf{R}_{ss}$ is full rank (M), then [26] span$(\mathbf{G})$ = span$(\mathbf{E})$, where $\mathbf{E}$ is an $I \times M$ matrix whose columns are the M eigenvectors of $\mathbf{R}_{zz}$ corresponding to the M largest eigenvalues. Hence, $\mathbf{E} = \mathbf{G}\mathbf{T}$, where $\mathbf{T}$ is $M \times M$ nonsingular. This is the basis of subspace approaches to signal parameter estimation. For the multiple-invariance structure at hand, (3) implies that [32]

$$\begin{bmatrix} \mathbf{E}_0 \\ \mathbf{E}_1 \\ \vdots \\ \mathbf{E}_{P-1} \end{bmatrix} := \begin{bmatrix} \mathbf{J}_0 \\ \mathbf{J}_1 \\ \vdots \\ \mathbf{J}_{P-1} \end{bmatrix} \mathbf{E} = \begin{bmatrix} \mathbf{A}\boldsymbol{\Phi}_0 \\ \mathbf{A}\boldsymbol{\Phi}_1 \\ \vdots \\ \mathbf{A}\boldsymbol{\Phi}_{P-1} \end{bmatrix} \mathbf{T} \tag{5}$$

which means that the signal subspace inherits the invariance structure present in the raw data. In practice, $\mathbf{R}_{zz}$ is estimated using

$$\hat{\mathbf{R}}_{zz} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{z}(t_n)\mathbf{z}^H(t_n) = \frac{1}{N}\mathbf{Z}\mathbf{Z}^H.$$

An estimate $\hat{\mathbf{E}}$ of $\mathbf{E}$ is obtained from the eigenvalue decomposition of $\hat{\mathbf{R}}_{zz}$ or, equivalently, the first M left singular vectors of $\mathbf{Z}$. In the noiseless case

$$\hat{\mathbf{R}}_{zz} = \frac{1}{N}\mathbf{G}\mathbf{S}\mathbf{S}^H\mathbf{G}^H.$$

Let $\hat{\mathbf{E}}$ denote the corresponding (noiseless, finite-sample) estimate of $\mathbf{E}$. If $\mathbf{S}$ is fat $(N \geq M)$ and full rank, then span$(\hat{\mathbf{E}}) =$ span$(\mathbf{G})$; else, if $\mathbf{S}$ is tall/loses rank, then span$(\hat{\mathbf{E}}) \subseteq$ span$(\mathbf{G})$, and $\hat{\mathbf{E}} = \mathbf{GT}$, but now, $\hat{\mathbf{E}}$ is $I \times r_{\mathbf{S}}$, and $\mathbf{T}$ is $M \times r_{\mathbf{S}}$ of rank $r_{\mathbf{S}}$. Hence

$$\begin{bmatrix} \hat{\mathbf{E}}_0 \\ \hat{\mathbf{E}}_1 \\ \vdots \\ \hat{\mathbf{E}}_{P-1} \end{bmatrix} := \begin{bmatrix} \mathbf{J}_0 \\ \mathbf{J}_1 \\ \vdots \\ \mathbf{J}_{P-1} \end{bmatrix} \hat{\mathbf{E}} = \begin{bmatrix} \mathbf{A}\boldsymbol{\Phi}_0 \\ \mathbf{A}\boldsymbol{\Phi}_1 \\ \vdots \\ \mathbf{A}\boldsymbol{\Phi}_{P-1} \end{bmatrix} \mathbf{T}. \tag{6}$$

In the noisy case, $\tilde{\mathbf{E}} = \hat{\mathbf{E}} + \breve{\mathbf{E}}$, and

$$\begin{aligned} \begin{bmatrix} \tilde{\mathbf{E}}_0 \\ \tilde{\mathbf{E}}_1 \\ \vdots \\ \tilde{\mathbf{E}}_{P-1} \end{bmatrix} &:= \begin{bmatrix} \mathbf{J}_0 \\ \mathbf{J}_1 \\ \vdots \\ \mathbf{J}_{P-1} \end{bmatrix} \tilde{\mathbf{E}} \\ &= \begin{bmatrix} \mathbf{A}\boldsymbol{\Phi}_0 \\ \mathbf{A}\boldsymbol{\Phi}_1 \\ \vdots \\ \mathbf{A}\boldsymbol{\Phi}_{P-1} \end{bmatrix} \mathbf{T} + \begin{bmatrix} \mathbf{J}_0 \\ \mathbf{J}_1 \\ \vdots \\ \mathbf{J}_{P-1} \end{bmatrix} \breve{\mathbf{E}}. \end{aligned} \tag{7}$$

A special case will be of interest as well. For linearly and uniformly displaced subarrays as those depicted in Fig. 1, it holds that $\boldsymbol{\Phi}_p = \boldsymbol{\Phi}_1^p$ (using the zeroth subarray as reference).

Both data-domain (4) and subspace models (7) incorporate errors. Assuming Gaussian noise, the least squares/weighted least squares principle is appropriate, leading, e.g., to data-domain least squares fitting

$$\min_{\mathbf{A},\boldsymbol{\Phi},\mathbf{S}} \left\| \begin{bmatrix} \tilde{\mathbf{X}}_0 \\ \vdots \\ \tilde{\mathbf{X}}_{P-1} \end{bmatrix} - \begin{bmatrix} \mathbf{A}\boldsymbol{\Phi}_0 \\ \vdots \\ \mathbf{A}\boldsymbol{\Phi}_{P-1} \end{bmatrix} \mathbf{S} \right\|_F^2 \tag{8}$$

and the corresponding weighted subspace fitting (WSF) problem [31], [32], [37], [39]

$$\min_{\mathbf{A},\boldsymbol{\Phi},\mathbf{T}} \left\| \begin{bmatrix} \tilde{\mathbf{E}}_0 \\ \vdots \\ \tilde{\mathbf{E}}_{P-1} \end{bmatrix} \mathbf{W}^{1/2} - \begin{bmatrix} \mathbf{A}\boldsymbol{\Phi}_0 \\ \vdots \\ \mathbf{A}\boldsymbol{\Phi}_{P-1} \end{bmatrix} \mathbf{T} \right\|_F^2 \tag{9}$$

where $\mathbf{W}^{1/2}$ is a weighting matrix, and $\boldsymbol{\Phi}$ holds all the diagonals. The problems in (8) and (9) have common structure from a regression viewpoint. Depending on the particular problem at hand, constraints on some or all of the matrices may be imposed, e.g., known displacement structure, training symbols, etc.

A. Previous Identifiability Results for MI-SAP

It is helpful to clarify that by *identifiability*, we mean uniqueness of all spatial source parameters as well as temporal source signals given a finite set of sensor measurements in the absence of noise and up to inherently unresolvable source permutation and scaling ambiguities. Other notions of identifiability are possible, e.g., [18] defines identifiability in the sense of distributions (i.e., distinct model parameterizations give rise to distinct distributions of the measurements), which is different from what is meant herein.

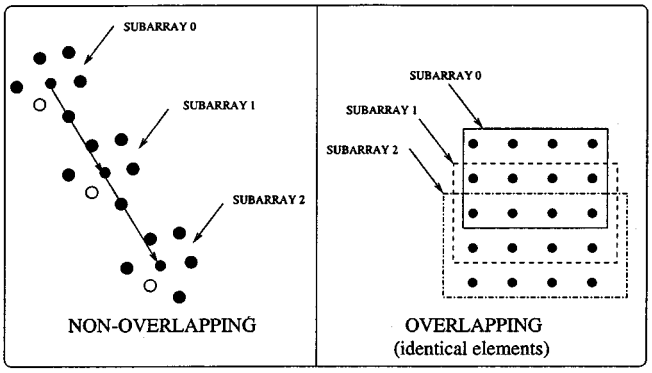

Fig. 1. Linearly and uniformly displaced subarrays.

- As stated earlier, $\mathbf{G}$ tall $(I > M)$ and full rank is necessary for identifiability [40].
- $P = 2$(single-invariance ESPRIT): $K \geq M$, $N \geq M$, full rank $\mathbf{A}$, $\mathbf{S}$, and displacement $< \lambda/2$ are required for identifiability (e.g., cf. [27] and [38]).
- P linearly and uniformly displaced subarrays

$$\begin{bmatrix} \hat{\mathbf{E}}_0 \\ \hat{\mathbf{E}}_1 \\ \vdots \\ \hat{\mathbf{E}}_{P-1} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{A}\boldsymbol{\Phi}_1 \\ \vdots \\ \mathbf{A}\boldsymbol{\Phi}_1^{P-1} \end{bmatrix} \mathbf{T}$$

and

$$\hat{\boldsymbol{\mathcal{E}}}_0 := \begin{bmatrix} \hat{\mathbf{E}}_0 \\ \hat{\mathbf{E}}_1 \\ \vdots \\ \hat{\mathbf{E}}_{P-2} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{A}\boldsymbol{\Phi}_1 \\ \vdots \\ \mathbf{A}\boldsymbol{\Phi}_1^{P-2} \end{bmatrix} \mathbf{T} =: \boldsymbol{\mathcal{A}}\mathbf{T}$$

$$\hat{\boldsymbol{\mathcal{E}}}_1 := \begin{bmatrix} \hat{\mathbf{E}}_1 \\ \hat{\mathbf{E}}_2 \\ \vdots \\ \hat{\mathbf{E}}_{P-1} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\boldsymbol{\Phi}_1 \\ \mathbf{A}\boldsymbol{\Phi}_1^2 \\ \vdots \\ \mathbf{A}\boldsymbol{\Phi}_1^{P-1} \end{bmatrix} \mathbf{T} = \boldsymbol{\mathcal{A}}\boldsymbol{\Phi}_1\mathbf{T}. \tag{10}$$

If $\mathbf{A}$ is already tall/full rank, then the single-invariance ESPRIT result applies [32, p. 874]. However, $\boldsymbol{\mathcal{A}}$ may well be tall/full rank even if $\mathbf{A}$ is not, resulting in significantly relaxed identifiability conditions. This has been left open in [32]. We will fill this gap in the sequel.

- General case: Identifiability results for the general case of arbitrary displacement structure are missing from the literature. We will provide such general results herein.

As alluded to in the introduction, the tools we will use toward our stated goals come from parallel factor analysis. How parallel factor analysis relates to multiple-invariance sensor array processing is explained next, including the necessary background.

III. Parallel Factor Analysis

We introduce some notation that will be useful in the sequel. Let $\boldsymbol{\Phi}$ be the $P \times M$ matrix whose $(p+1)$st row is the diagonal of $\boldsymbol{\Phi}_p$, and let $\mathbf{D}_p(\boldsymbol{\Phi})$ denote the diagonal matrix containing the $(p+1)$st row of $\boldsymbol{\Phi}$: $\mathbf{D}_p(\boldsymbol{\Phi}) = \boldsymbol{\Phi}_p$, and

$$\mathbf{X}_p = \mathbf{A}\mathbf{D}_p(\boldsymbol{\Phi})\mathbf{S}, p = 0, \cdots, P-1. \tag{11}$$

Letting $x_{k,n,p}$ stand for the (k,n) element of $\mathbf{X}_p$ (the signal part of the output of the kth antenna of the pth subarray at time n), we have

$$x_{k,n,p} = \sum_{m=0}^{M-1} a_{k,m}\phi_{p,m}s_{m,n} \tag{12}$$

where $a_{k,m}$ stands for the (k,m)-element of $\mathbf{A}$, and similarly for the others. Equation (12) expresses $x_{k,n,p}$ as a sum of triple products; it is variably known as the trilinear model, trilinear decomposition, triple product decomposition, canonical decomposition, or *parallel factor* (PARAFAC) analysis of $x_{k,n,p}$.

Define a $K \times N \times P$ *three-way array* $\underline{\mathbf{X}}$ with typical element $x_{k,n,p}$. A rank-1 three-way array $\underline{\boldsymbol{\epsilon}}$ is the "outer product" of three vectors $\mathbf{a}, \boldsymbol{\phi}, \mathbf{s}$; its typical element $\epsilon_{k,n,p}$ can be written as

$$\epsilon_{k,n,p} = a_k\phi_p s_n.$$

Equation (12) is an M-component triple product decomposition of $\underline{\mathbf{X}}$. The *rank* of $\underline{\mathbf{X}}$ is defined as the minimum number of rank-1 three-way components needed to decompose $\underline{\mathbf{X}}$ [20]. The above definition is consistent with matrix (two-way array) rank, which is the minimum number of rank-1 matrices needed to decompose a given matrix.

Notice how the assumed shift invariance is crucial in establishing trilinearity of the MI-SAP model; hopping through corresponding subarray elements involves picking up a factor that depends on the chosen subarrays and source azimuth/elevation parameters but not the particular subarray elements (k). In the absence of shift invariance, the three-way data in $\underline{\mathbf{X}}$ no longer obey a PARAFAC model of rank M.

Equation (11) can be viewed as "slicing" the 3-D array $\underline{\mathbf{X}}$ in a series of "slabs" (2-D arrays) perpendicular to the invariance dimension, i.e., $\mathbf{X}_p = [x_{:,:,p}]$, that is, the $K \times N$ 2-D slice of $\underline{\mathbf{X}}$ corresponding to the given p. Two additional slicings are possible and useful in understanding the regression algorithm in Section V.

$$\mathbf{Y}_k = \mathbf{S}^T\mathbf{D}_k(\mathbf{A})\boldsymbol{\Phi}^T, k = 0, 1, \cdots, K-1 \tag{13}$$

where the $N \times P$ matrix $\mathbf{Y}_k := [x_{k,:,:}]$. Similarly

$$\mathbf{Z}_n = \boldsymbol{\Phi}\mathbf{D}_n(\mathbf{S}^T)\mathbf{A}^T, n = 0, 1, \cdots, N-1 \tag{14}$$

where the $P \times K$ matrix $\mathbf{Z}_n := [x_{:,n,:}]$.

IV. Identifiability

We will need the following definition.

Definition 1: Consider a matrix $\mathbf{B} \in \mathbb{C}^{I\times J}$. If $r_{\mathbf{B}} := rank(\mathbf{B}) = r$, then $\mathbf{B}$ contains a collection of r linearly independent columns. Moreover, if *all* $\ell \leq J$ columns of $\mathbf{B}$ are linearly independent but there exists a collection of $\ell+1$ linearly dependent columns of $\mathbf{B}$ (or $\ell = J$), then $\mathbf{B}$ has *Kruskal*-rank (*k-rank*) $k_{\mathbf{B}} = \ell$.

The concept is implicit in the work of Kruskal [20], but the term was later coined by Harshman and Lundy [17]. Note that unlike regular matrix rank, the concept of k-rank is asymmetric, as it pertains to columns: A fat matrix $\mathbf{B} \in \mathbb{C}^{I\times J}$, $I < J$ may have full rank I (= row rank =column rank) but k-rank one if it contains two colinear columns or even zero if it contains an all-zero column. On the other hand, the k-rank of $\mathbf{B}^T$ will be full $(= I)$ in this case. A square/tall full-rank matrix has k-rank equal to its rank. In general, $k_{\mathbf{B}} \leq rank(\mathbf{B})$, $\forall \mathbf{B}$.

PARAFAC uniqueness has a long history [14]–[16], [27], but Kruskal's result [20], [21] is the deepest.

Theorem 1 [20]: Consider the set of P matrices $\mathbf{X}_p = \mathbf{A}\mathbf{D}_p(\boldsymbol{\Phi})\mathbf{S}$, $p = 0, 1, \cdots, P-1$, where $\mathbf{A} \in \mathbb{R}^{K\times M}$, $\boldsymbol{\Phi} \in \mathbb{R}^{P\times M}$, $\mathbf{S} \in \mathbb{R}^{M\times N}$, and M denotes the common dimension. If

$$k_{\mathbf{A}} + k_{\boldsymbol{\Phi}} + k_{\mathbf{S}^T} \geq 2M + 2 \tag{15}$$

then $\mathbf{A}$, $\boldsymbol{\Phi}$, and $\mathbf{S}^T$ are unique up to permutation and scaling of columns, meaning that any other triple $\overline{\mathbf{A}}, \overline{\boldsymbol{\Phi}}, \overline{\mathbf{S}}^T$ that gives rise to the data $\mathbf{X}_p, p = 0, 1, \cdots, P-1$ is related to $\mathbf{A}, \boldsymbol{\Phi}, \mathbf{S}^T$ via

$$\overline{\mathbf{A}} = \mathbf{A}\boldsymbol{\Pi}\boldsymbol{\Delta}_1, \quad \overline{\boldsymbol{\Phi}} = \boldsymbol{\Phi}\boldsymbol{\Pi}\boldsymbol{\Delta}_2, \quad \overline{\mathbf{S}}^T = \mathbf{S}^T\boldsymbol{\Pi}\boldsymbol{\Delta}_3 \tag{16}$$

where $\boldsymbol{\Pi}$ is a permutation matrix, and $\boldsymbol{\Delta}_{1,2,3}$ are diagonal scaling matrices satisfying

$$\boldsymbol{\Delta}_1\boldsymbol{\Delta}_2\boldsymbol{\Delta}_3 = \mathbf{I}. \tag{17}$$

The result holds for $M = 1$, irrespective of condition (15), as long as $\underline{\mathbf{X}}$ does not contain an identically zero 2-D slice along any dimension. ■

Radar and communication systems employing in-phase/quadrature processing lead to discrete-time baseband-equivalent models involving complex arrays. Kruskal, in [20], states and proves his uniqueness of trilinear decomposition results explicitly under the working assumption of real-valued arrays. Splitting into real and imaginary parts destroys trilinear structure, and three-way array rank is sensitive to the domain of numbers used in the decomposition [21]. Care should therefore be exercised in treating the complex case. The proof of the following result can be found in our earlier paper [29].

Theorem 2 [29]: Given $\mathbf{X}_p = \mathbf{A}\mathbf{D}_p(\boldsymbol{\Phi})\mathbf{S}, p = 0, 1, \cdots, P-1$, $\mathbf{A} \in \mathbb{C}^{K\times M}$, $\boldsymbol{\Phi} \in \mathbb{C}^{P\times M}$, and $\mathbf{S} \in \mathbb{C}^{M\times N}$, if

$$k_{\mathbf{A}} + k_{\boldsymbol{\Phi}} + k_{\mathbf{S}^T} \geq 2M + 2 \tag{18}$$

then $\mathbf{A}$, $\boldsymbol{\Phi}$, and $\mathbf{S}^T$ are unique up to permutation and (complex) scaling of columns. ■

As it stands, the condition applies to the data-domain MI-SAP formulation (3). The following result shows that nothing is lost in terms of identifiability by going to the corresponding subspace formulation (6).

Proposition 1—Subspace Identifiability: If $k_{\mathbf{A}} + k_{\boldsymbol{\Phi}} + k_{\mathbf{S}^T} \geq 2M + 2$, then the subspace model in (6) is identifiable, i.e., given $\hat{\mathbf{E}}_p$, $p = 0, 1, \cdots, P-1$, $\mathbf{A}$, $\boldsymbol{\Phi}$, and $\mathbf{T}^T$ are unique up to permutation and scaling of columns. ■

Proof: Let $\mathbf{Z} = \mathbf{E}\boldsymbol{\Sigma}\mathbf{U}^H$ be the SVD of $\mathbf{Z}$, where $\mathbf{E}$ is $I \times r_{\mathbf{S}}$, $\boldsymbol{\Sigma}$ is $r_{\mathbf{S}} \times r_{\mathbf{S}}$ diagonal containing the $r_{\mathbf{S}}$ nonzero singular values of $\mathbf{Z}$, and $\mathbf{U}$ is $N \times r_{\mathbf{S}}$. Now, $\mathbf{E} = \mathbf{G}\mathbf{T}$, and $\mathbf{Z} = \mathbf{G}\mathbf{S}$, from which it follows that $\mathbf{G}\mathbf{S} = \mathbf{G}\mathbf{T}\boldsymbol{\Sigma}\mathbf{U}^H$. With $\mathbf{G}$ tall/full rank, this implies that $\mathbf{S} = \mathbf{T}\boldsymbol{\Sigma}\mathbf{U}^H$ or $\mathbf{S}^T = \mathbf{U}^*\boldsymbol{\Sigma}\mathbf{T}^T$. Since $\mathbf{U}$ is tall/full rank, Sylvester's inequality implies that $r_{\mathbf{T}} = r_{\mathbf{S}}$; furthermore, since *every* m columns of $\mathbf{S}^T$ arise out of the corresponding m columns of $\mathbf{T}^T$, Sylvester's inequality shows that

any m columns of $\mathbf{S}^T$ are linearly independent if and only if the corresponding m columns of $\mathbf{T}^T$ are linearly independent; hence, $k_{\mathbf{T}^T} = k_{\mathbf{S}^T}$. The k-rank of $\mathbf{T}^T$ is inherited from $\mathbf{S}^T$. The result then follows by applying Theorem 2 to (6). Note that $\mathbf{S}$ can be recovered up to permutation and scaling of rows from $\mathbf{T}$ and the row subspace of $\mathbf{Z}$. ■

Carroll *et al.* [8] have in fact considered the relationship between the data domain and subspace fitting problems for what amounts to the special case of nonoverlapping subarrays extracted from $\mathbf{G}$, without reference to identifiability, however.

There exists a very interesting link between k-rank and the concept of *rank ambiguity* that appears in the study of sensor arrays (e.g., cf. [34], [35], and references therein). It is well known [40] that in order to uniquely determine the DOA's of ν uncorrelated sources from the array outputs, one requires an array manifold that is free of rank-ν ambiguities, meaning that *every* $\nu + 1$ steering vectors (corresponding to distinct DOA's) drawn from the manifold are linearly independent. Let $\overline{\nu}$ be the maximum such ν for a given manifold;[3] the associated antenna array can resolve up to $\overline{\nu}$ uncorrelated sources. Let $\overline{\nu}_{sub}, \overline{\nu}_{dis}$ be the resolvability bounds for the subarray manifold and the displacement manifold, respectively. Clearly

$$k_{\mathbf{A}} \geq \overline{\nu}_{sub} + 1, \quad k_{\mathbf{\Phi}} \geq \overline{\nu}_{dis} + 1.$$

Now, suppose that the source signal matrix $\mathbf{S}$ is fat $(N \geq M)$ and full rank (M). Then, the k-rank identifiability condition (18) becomes

$$k_{\mathbf{A}} + k_{\mathbf{\Phi}} \geq M + 2$$

which is guaranteed *a fortiori* if the DOA's of the different sources are distinct and

$$\overline{\nu}_{sub} + \overline{\nu}_{dis} \geq M.$$

The latter inequality is an intuitively pleasing *subarray synthesis* result; the resolving power of the reference subarray and the displacement subarray add up. See [34] for a discussion of issues related to the design of antenna arrays with a specified $\overline{\nu}$.

A random matrix whose columns are drawn independently from an absolutely continuous distribution has full rank with probability one. Interestingly, it also has full k-rank because any combination of columns can be thought of as another random matrix with columns drawn independently from an absolutely continuous distribution. This holds even if the elements across a given column are dependent random variables. In our present context, for *source-wise* independent source signals, $k_{\mathbf{S}^T} = \min(M, N)$, and therefore, (18) becomes

$$k_{\mathbf{A}} + k_{\mathbf{\Phi}} + \min(M, N) \geq 2M + 2$$

which implies that M sources can be identified with as few as $N = 2$ snapshots, provided $K \geq M$, $P \geq M$, and the subarray and displacement matrices are full rank (M).

A. DOA Recovery Considerations

Thus far, we have ignored the issue of DOA recovery from the matrix estimates. If the reference subarray is uncalibrated, as is usually assumed in ESPRIT approaches, then one relies on the structure of $\mathbf{\Phi}$ to recover azimuth and elevation estimates. Unlike DOA determination from array *outputs*, unambiguous azimuth/elevation recovery from the columns of $\mathbf{\Phi}$ requires only mild conditions on the displacement structure, which can be easily satisfied by array design. This is best illustrated by example. In the case of the $\lambda/2$ uniform square array in Section II, the displacement steering vector has the following structure:

$$\begin{bmatrix} 1 \\ e^{-j\pi \sin\theta_{m,a}} \\ e^{-j\pi \cos\theta_{m,a} \sin\theta_{m,e}} \\ e^{-j\pi \sin\theta_{m,a}} e^{-j\pi \cos\theta_{m,a} \sin\theta_{m,e}} \end{bmatrix}$$

which means that unambiguous azimuth/elevation recovery from the columns of $\mathbf{\Phi}$ is possible, provided the sources are confined within plus or minus 90° in both azimuth and elevation. In particular, this does not require that $P > M$, which is required for unambiguous DOA determination from displacement subarray outputs.

If the reference subarray is calibrated, then the steering vectors from both $\mathbf{A}$ and $\mathbf{\Phi}$ can be used to recover azimuth/elevation estimates.

B. One Angle in Common

It is interesting to investigate whether the identifiability condition (18) can hold in the presence of two sources with one angle in common. Consider a 3×3 uniform square array with interelement spacing $\lambda/2$, and $P = 4$ square subarrays corresponding to the bottom-left ($p = 0$ reference), bottom-right ($p = 1$), top-left ($p = 2$), and top-right ($p = 3$) 2×2 blocks ($K = 4$). Let there be $M = 2$ sources at the same azimuth but different elevations. Let $\phi_1 := e^{-j\pi \sin\theta_a}$, $\phi_2 := e^{-j\pi \cos\theta_a \sin\theta_{1,e}}$, and $\phi_3 := e^{-j\pi \cos\theta_a \sin\theta_{2,e}}$. Then

$$\mathbf{A} = \mathbf{\Phi} = \begin{bmatrix} 1 & 1 \\ \phi_1 & \phi_1 \\ \phi_2 & \phi_3 \\ \phi_1\phi_2 & \phi_1\phi_3 \end{bmatrix}.$$

If $\phi_3 \neq \phi_2$, then the first and the third row are linearly independent, and therefore, $\mathbf{A}$ and $\mathbf{\Phi}$ are full rank as well as full k-rank: $k_{\mathbf{A}} = k_{\mathbf{\Phi}} = 2$. For $\mathbf{S}$ fat and full rank, $k_{\mathbf{A}} + k_{\mathbf{\Phi}} + k_{\mathbf{S}^T} = 6 = 2 \times 2 + 2$, and hence, (18) is satisfied, and uniqueness is guaranteed.

C. Regarding Coherence

For coherent sources (rank-deficient source covariance matrix), the Cauchy–Schwartz inequality implies that the sample paths are colinear with probability one. The k-rank of a matrix with colinear nonzero columns is one, which implies that the sum of k-ranks in (18) can never exceed two times the number of sources plus two; hence, PARAFAC uniqueness apparently fails. However, we have the following.

- A recent partial uniqueness result of Bro *et al.* [3] (in the context of applications in chemistry) can be applied to show that the coherent source signals are still identifiable (albeit under more restrictive conditions). However, there is partial rotational freedom for buckets of columns of the subarray and displacement matrices associated with

[3] $\overline{\nu}$ depends on array geometry and other antenna design factors [34], [35].

the colinear sources, e.g., a pair of steering vectors corresponding to two coherent sources can be rotated without affecting the data.

- Full uniqueness can be restored for fully coherent sources if Vandermonde structure is available in either the subarray or (by symmetry) the displacement matrix, using smoothing ideas. This requires several new results on the k-rank of certain structured matrices, as well as optimizing the smoothing factor from the identifiability perspective. The final results include improved identifiability conditions for PARAFAC models exhibiting Vandermonde structure along one or two modes; these are reported in [30].

D. Subarrays of Different Sizes

One limitation of the PARAFAC approach is that it cannot handle subarrays of different sizes. This situation can arise, for example, if the full array is uniform rectangular with more rows than columns and one extracts two maximal row-overlap subarrays and two maximal column-overlap subarrays [11].

E. Linearly and Uniformly Displaced Subarrays

Let us now return to the special case of linearly and uniformly displaced subarrays in (10). It is of interest to determine relaxed conditions under which the matrix $\boldsymbol{\mathcal{A}}$ is full rank. We have the following result.

Lemma 1: (*Full Rank of Khatri–Rao Product*). Consider

$$\boldsymbol{\Phi} \odot \mathbf{A} := [\boldsymbol{\phi}_1 \otimes \mathbf{a}_1, \cdots, \boldsymbol{\phi}_M \otimes \mathbf{a}_M] = \begin{bmatrix} \mathbf{A}\mathbf{D}_0(\boldsymbol{\Phi}) \\ \mathbf{A}\mathbf{D}_1(\boldsymbol{\Phi}) \\ \vdots \\ \mathbf{A}\mathbf{D}_{P-1}(\boldsymbol{\Phi}) \end{bmatrix}$$

where $\mathbf{A}$ is $K \times M$, $\boldsymbol{\Phi}$ is $P \times M$, $\otimes$ stands for the Kronecker product, $\odot$ stands for the Khatri–Rao (column-wise Kronecker) product, and $\boldsymbol{\phi}_m$, $\mathbf{a}_m$ are the columns of $\boldsymbol{\Phi}$ and $\mathbf{A}$. If $k_{\mathbf{A}} + k_{\boldsymbol{\Phi}} \geq M + 1$, then $\boldsymbol{\Phi} \odot \mathbf{A}$ is full column rank M. ■

Remark 1: Note that $K \geq k_{\mathbf{A}}$, and $P \geq k_{\boldsymbol{\Phi}}$; hence , $k_{\mathbf{A}} + k_{\boldsymbol{\Phi}} \geq M+1$ implies $K+P \geq M+1 \Rightarrow KP \geq M$; therefore, the condition includes the requirement that $\boldsymbol{\Phi} \odot \mathbf{A}$ is tall.

Proof: Define $\boldsymbol{\Xi} := \boldsymbol{\Phi} \odot \mathbf{A}$ for notational simplicity, and assume the contrary, i.e., that the columns of $\boldsymbol{\Xi}$ are, in fact, linearly dependent. Then, there exist $\mu_1, \cdots, \mu_M$ not all equal to zero such that

$$\mu_1 \boldsymbol{\xi}_1 + \cdots + \mu_M \boldsymbol{\xi}_M = \mathbf{0}_{KP \times 1}$$

where $\boldsymbol{\xi}_m$ is the mth column of $\boldsymbol{\Xi}$. Through algebraic manipulation, it can be shown that the above can be rearranged into the following more convenient form:

$$\mathbf{A} \begin{bmatrix} \mu_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mu_M \end{bmatrix} \boldsymbol{\Phi}^T = \mathbf{0}_{K \times P}. \tag{19}$$

Let us suppose that $w \leq M$ of the μ's are nonzero. Let $\tilde{\mathbf{A}}$ be constructed out of the corresponding w columns of $\mathbf{A}$, and let $\tilde{\boldsymbol{\Phi}}^T$ be constructed out of the corresponding w rows of $\text{diag}([\mu_1 \cdots \mu_M])\boldsymbol{\Phi}^T$. Then

$$\begin{aligned} r &:= rank\left(\mathbf{A} \begin{bmatrix} \mu_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mu_M \end{bmatrix} \boldsymbol{\Phi}^T\right) \\ &= rank(\tilde{\mathbf{A}}\tilde{\boldsymbol{\Phi}}^T) \geq rank(\tilde{\mathbf{A}}) + rank(\tilde{\boldsymbol{\Phi}}^T) - w \end{aligned} \tag{20}$$

by Sylvester's inequality. However, by definition of k-rank

$$\begin{aligned} rank(\tilde{\mathbf{A}}) &\geq \min(w, k_{\mathbf{A}}), \\ rank(\tilde{\boldsymbol{\Phi}}) &\geq \min(w, k_{\boldsymbol{\Phi}}). \end{aligned} \tag{21}$$

Therefore, from (20) and (21), we obtain

$$r \geq \min(w, k_{\mathbf{A}}) + \min(w, k_{\boldsymbol{\Phi}}) - w.$$

We have the following cases.

1) $w \leq \min(k_{\mathbf{A}}, k_{\boldsymbol{\Phi}})$: $\Rightarrow r \geq w$.
2) $\min(k_{\mathbf{A}}, k_{\boldsymbol{\Phi}}) \leq w \leq \max(k_{\mathbf{A}}, k_{\boldsymbol{\Phi}})$: $\Rightarrow r \geq \min(k_{\mathbf{A}}, k_{\boldsymbol{\Phi}})$.
3) $w \geq \max(k_{\mathbf{A}}, k_{\boldsymbol{\Phi}})$: $\Rightarrow r \geq k_{\mathbf{A}} + k_{\boldsymbol{\Phi}} - w \geq M + 1 - w$.

Note that $k_{\mathbf{A}} + k_{\boldsymbol{\Phi}} \geq M + 1 \Rightarrow \min(k_{\mathbf{A}}, k_{\boldsymbol{\Phi}}) \geq 1$. For $1 \leq w \leq M$, all cases above lead to $r \geq 1$, but r is the rank of the matrix in (19); therefore, $r = 0$. This constitutes a contradiction, and thus, the proof is complete. ■

Lemma 1 is the heart of the following theorem.

Theorem 3—PARAFAC/Vandermonde Uniqueness: Consider the data model

$$\mathbf{X}_p = \mathbf{A}\mathbf{D}_p(\boldsymbol{\Phi})\mathbf{S}, p = 0, 1, \cdots, P-1 \tag{22}$$

and suppose that $\boldsymbol{\Phi} \in \mathbb{C}^{P \times M}$ is Vandermonde with distinct nonzero generators and that $\mathbf{S} \in \mathbb{C}^{M \times N}$ is fat and full rank. If $k_{\mathbf{A}} + \min(P-1, M) \geq M+1$, then $\mathbf{A}, \boldsymbol{\Phi}, \mathbf{S}^T$ are unique up to permutation and scaling of columns. Under the same conditions, the corresponding subspace model in (10) is also unique. ■

Proof: We prove it for the subspace model in (10); the proof for the data-domain model follows along similar lines. Recall that $\mathbf{S} \in \mathbb{C}^{M \times N}$ fat and full rank implies $\mathbf{T}$ square and full rank, and consider

$$\hat{\boldsymbol{\mathcal{E}}}_0 = \boldsymbol{\mathcal{A}}\mathbf{T}, \quad \hat{\boldsymbol{\mathcal{E}}}_1 = \boldsymbol{\mathcal{A}}\boldsymbol{\Phi}_1\mathbf{T}$$

with $\boldsymbol{\mathcal{A}} = \boldsymbol{\Phi}^{(P-1)} \odot \mathbf{A}$, where $\boldsymbol{\Phi}^{(P-1)}$ contains the first $P-1$ rows of $\boldsymbol{\Phi}$. $\boldsymbol{\Phi}^{(P-1)}$ is Vandermonde with distinct nonzero generators (the diagonal elements of $\boldsymbol{\Phi}_1$). It is shown in [30] that a Vandermonde matrix with distinct nonzero generators has full k-rank, i.e., $k_{\boldsymbol{\Phi}^{(P-1)}} = \min(P-1, M)$. Then, the condition $k_{\mathbf{A}} + \min(P-1, M) \geq M+1$ and Lemma 1 imply that $\boldsymbol{\mathcal{A}}$ is tall and full rank. This brings us back to the familiar case of single-invariance ESPRIT [25], [27], at which point, different routes can be taken. One way to obtain $\boldsymbol{\mathcal{A}}, \boldsymbol{\Phi}_1, \mathbf{T}$ is as follows. Define

$$\begin{aligned} \mathbf{C}_{00} &:= \hat{\boldsymbol{\mathcal{E}}}_0^H \hat{\boldsymbol{\mathcal{E}}}_0 = \mathbf{T}^H \boldsymbol{\mathcal{A}}^H \boldsymbol{\mathcal{A}} \mathbf{T} \\ \mathbf{C}_{01} &:= \hat{\boldsymbol{\mathcal{E}}}_0^H \hat{\boldsymbol{\mathcal{E}}}_1 = \mathbf{T}^H \boldsymbol{\mathcal{A}}^H \boldsymbol{\mathcal{A}} \boldsymbol{\Phi}_1 \mathbf{T}. \end{aligned}$$

TABLE I
DATA DOMAIN RECOVERY

Data Domain : Eqn. (22)
$\boldsymbol{\mathcal{A}} \stackrel{ps}{=} \boldsymbol{\mathcal{X}}_0 (\boldsymbol{\Gamma}_d \mathbf{R}_{00})^{\dagger}$
$\mathbf{S}^T \stackrel{ps}{=} \mathbf{R}_{00}^T \boldsymbol{\Gamma}_d^T$
$\boldsymbol{\Phi}_1 \boldsymbol{\Gamma}_d \mathbf{R}_{00} = \boldsymbol{\Gamma}_d \mathbf{R}_{01}$ (GED)
$\mathbf{R}_{00} := \boldsymbol{\mathcal{X}}_0^H \boldsymbol{\mathcal{X}}_0$
$\mathbf{R}_{01} := \boldsymbol{\mathcal{X}}_0^H \boldsymbol{\mathcal{X}}_1$
$\boldsymbol{\mathcal{X}}_0 := \left[\mathbf{X}_0^T, \cdots, \mathbf{X}_{P-2}^T\right]^T = \boldsymbol{\mathcal{A}}\mathbf{S}$
$\boldsymbol{\mathcal{X}}_1 := \left[\mathbf{X}_1^T, \cdots, \mathbf{X}_{P-1}^T\right]^T = \boldsymbol{\mathcal{A}}\boldsymbol{\Phi}_1\mathbf{S}$
$\boldsymbol{\mathcal{A}} := \boldsymbol{\Phi}^{(P-1)} \odot \mathbf{A}$

TABLE II
SUBSPACE DOMAIN RECOVERY

Subspace Domain : Eqn. (10)
$\boldsymbol{\mathcal{A}} \stackrel{ps}{=} \widehat{\boldsymbol{\mathcal{E}}}_0 (\boldsymbol{\Gamma}_s \mathbf{C}_{00})^{-1}$
$\mathbf{T}^T \stackrel{ps}{=} \mathbf{C}_{00}^T \boldsymbol{\Gamma}_s^T$
$\boldsymbol{\Phi}_1 \boldsymbol{\Gamma}_s = \boldsymbol{\Gamma}_s \mathbf{C}_{01} \mathbf{C}_{00}^{-1}$ (EVD)
$\mathbf{C}_{00} := \widehat{\boldsymbol{\mathcal{E}}}_0^H \widehat{\boldsymbol{\mathcal{E}}}_0$
$\mathbf{C}_{01} := \widehat{\boldsymbol{\mathcal{E}}}_0^H \widehat{\boldsymbol{\mathcal{E}}}_1$
$\widehat{\boldsymbol{\mathcal{E}}}_0 := \left[\widehat{\mathbf{E}}_0^T, \cdots, \widehat{\mathbf{E}}_{P-2}^T\right]^T = \boldsymbol{\mathcal{A}}\mathbf{T}$
$\widehat{\boldsymbol{\mathcal{E}}}_1 := \left[\widehat{\mathbf{E}}_1^T, \cdots, \widehat{\mathbf{E}}_{P-1}^T\right]^T = \boldsymbol{\mathcal{A}}\boldsymbol{\Phi}_1\mathbf{T}$
$\boldsymbol{\mathcal{A}} := \boldsymbol{\Phi}^{(P-1)} \odot \mathbf{A}$

The matrix $\mathbf{T}^H \boldsymbol{\mathcal{A}}^H \boldsymbol{\mathcal{A}}$ is square and full rank. Let $\boldsymbol{\Gamma}_s$ denote its inverse. Then

$$\boldsymbol{\Phi}_1 \boldsymbol{\Gamma}_s \mathbf{C}_{00} = \boldsymbol{\Gamma}_s \mathbf{C}_{01}.$$

$\mathbf{C}_{00}$ (and $\mathbf{C}_{01}$) is $M \times M$, of rank M, and hence

$$\boldsymbol{\Phi}_1 \boldsymbol{\Gamma}_s = \boldsymbol{\Gamma}_s \mathbf{C}_{01} \mathbf{C}_{00}^{-1}.$$

The rows of $\boldsymbol{\Gamma}_s$ can therefore be determined up to permutation and scaling from the left eigenvectors of $\mathbf{C}_{01}\mathbf{C}_{00}^{-1}$. From $\boldsymbol{\Gamma}_s$, $\mathbf{T}$ can be recovered as $\mathbf{T}^T \stackrel{ps}{=} \mathbf{C}_{00}^T \boldsymbol{\Gamma}_s^T$, whereas $\boldsymbol{\mathcal{A}}$ can be recovered as $\boldsymbol{\mathcal{A}} \stackrel{ps}{=} \hat{\boldsymbol{\mathcal{E}}}_0 (\boldsymbol{\Gamma}_s \mathbf{C}_{00})^{-1}$, where $\stackrel{ps}{=}$ means equal up to permutation and scaling of columns. $\mathbf{A}$ can then be recovered up to permutation and scaling of columns from $\boldsymbol{\mathcal{A}} = \boldsymbol{\Phi}^{(P-1)} \odot \mathbf{A}$. The computational steps are summarized in Tables I and II. In the presence of noise, the estimates can be refined using the algorithm described in the next section. ■

Remark 2: If a ULA is employed as reference subarray, then $\mathbf{A}$ is Vandermonde as well, and therefore, $k_{\mathbf{A}} = \min(K, M)$, assuming distinct generators. In this case, the k-rank condition in the statement of Theorem 3 becomes $\min(K, M) + \min(P - 1, M) \geq M + 1$.

V. TRILINEAR ALTERNATING LEAST SQUARES REGRESSION (TALS)

Let us now return to the least squares fitting problem in (8) and (9). Without loss of generality, let us consider the data-domain formulation in (8).

We are given the noisy data $\underline{\tilde{\mathbf{X}}}$ and wish to estimate $\mathbf{A}, \boldsymbol{\Phi}$, and $\mathbf{S}$. Using Khatri–Rao product notation and letting $\tilde{\mathbf{X}} := [\tilde{\mathbf{X}}_0{}^T, \cdots, \tilde{\mathbf{X}}_{P-1}{}^T]^T$, where $\tilde{\mathbf{X}}_p$, $p = 0, 1, \cdots, P-1$ are the noisy slabs along the invariance dimension, (8) can be written as

$$\min_{\mathbf{A}, \boldsymbol{\Phi}, \mathbf{S}} ||\tilde{\mathbf{X}} - (\boldsymbol{\Phi} \odot \mathbf{A})\mathbf{S}||_F^2. \tag{23}$$

It follows that the conditional least squares update of $\mathbf{S}$ given interim estimates $\hat{\mathbf{A}}, \hat{\boldsymbol{\Phi}}$ is given by

$$\hat{\mathbf{S}}_{CLS} := (\hat{\boldsymbol{\Phi}} \odot \hat{\mathbf{A}})^{\dagger} \tilde{\mathbf{X}} \tag{24}$$

where $(\cdot)^{\dagger}$ stands for pseudo-inverse. Stacking the data matrices in (13) leads to the following equivalent way of writing (8):

$$\min_{\mathbf{A}, \boldsymbol{\Phi}, \mathbf{S}} ||\tilde{\mathbf{Y}} - (\mathbf{A} \odot \mathbf{S}^T)\boldsymbol{\Phi}^T||_F^2 \tag{25}$$

where $\tilde{\mathbf{Y}} := [\tilde{\mathbf{Y}}_0{}^T, \cdots, \tilde{\mathbf{Y}}_{K-1}{}^T]^T$, and $\tilde{\mathbf{Y}}_k$, $k = 0, 1, \cdots, K-1$ are the noisy slabs along the subarray element dimension. It follows that

$$\hat{\boldsymbol{\Phi}}_{CLS}{}^T := (\hat{\mathbf{A}} \odot \hat{\mathbf{S}}^T)^{\dagger} \tilde{\mathbf{Y}}. \tag{26}$$

Finally, stacking the data matrices in (14)

$$\hat{\mathbf{A}}_{CLS}{}^T := (\hat{\mathbf{S}}^T \odot \hat{\boldsymbol{\Phi}})^{\dagger} \tilde{\mathbf{Z}} \tag{27}$$

where $\tilde{\mathbf{Z}} := [\tilde{\mathbf{Z}}_0{}^T, \cdots, \tilde{\mathbf{Z}}_{N-1}{}^T]^T$ and $\tilde{\mathbf{Z}}_n$, $n = 0, 1, \cdots, N-1$ are the noisy slabs along the temporal dimension.

TALS may be initialized by single-invariance ESPRIT when applicable or randomly otherwise. In the present context, TALS may also be initialized using any suboptimal multiple-parameter multiple-invariance algorithm, e.g., [12], [31], [41]—there are clearly many possibilities. Given initial estimates, TALS proceeds by updating $\hat{\mathbf{A}}$, $\hat{\boldsymbol{\Phi}}$, and $\hat{\mathbf{S}}$ in a round-robin fashion using (24), (26), and (27). Note that the conditional update of any given matrix may either improve or maintain, but cannot worsen, the current fit. Global monotone convergence to (at least) a local minimum follows directly from this observation.

Many algorithms based on the ALS principle have been proposed in the signal processing literature and aimed at solving a wide variety of estimation problems. The so-called ILSE algorithm is one well-known example [33]. All algorithms of the ALS genre build on the same basic "piecemeal" optimization approach. Aside from global monotone convergence, different ALS algorithms have different properties, especially regarding whether the global minimum is reached, rate of convergence, and complexity. ILSE, for example, utilizes a computationally demanding enumeration step, which is the fundamental complexity bottleneck. ILSP [33] is actually not an ALS algorithm because it uses a two-step finite-alphabet update procedure that is not optimal in the conditional LS sense, and hence, convergence is not guaranteed.

The per-iteration complexity of TALS is equal to the cost of computing a matrix pseudo-inverse. Overall complexity depends on the number of iterations, which varies depending on problem-specific parameters and the given batch of data. Typical runtime for the problems considered herein is between 0.01 s and a few seconds.

TALS was first utilized to fit the PARAFAC model by Harshman [14]–[16] and Caroll and Chang [7]. The bare-bones TALS algorithm outlined above does not take advantage of the structure of $\boldsymbol{\Phi}$. Other structures may also be available,

e.g., finite-alphabet or constant-modulus of the elements of $\mathbf{S}$, training symbols embedded in $\mathbf{S}$, known DOA for one or more sources, etc. If one has an algorithm that solves the vector regression problem $\min_{\mathbf{x}\in\mathcal{C}} \|\mathbf{y}-\mathbf{x}\|_2^2$, for arbitrary constraint set $\mathcal{C}$, then one can enforce $\mathcal{C}$ on the columns of any one of the three matrices $(\mathbf{A}, \boldsymbol{\Phi}, \mathbf{S}^T)$ in a LS fashion [4], thereby maintaining monotone convergence. Hence, arbitrary parametric or nonparametric *source-wise* constraints can be easily incorporated, and this is one of the nice features of ALS (cf. [4] for a treatment of unimodality and non-negativity constraints in the context of TALS). Another feature of ALS is its conceptual simplicity and the fact that there is no need to tune parameters to guarantee convergence. Finally, TALS can be extended to fit quadrilinear and more generally multilinear models.

TALS can be used to fit the data-domain problem in (8) or the WSF problem in (9). The Gauss–Newton (GN)-type algorithms of [31] and [32], which were developed for the WSF problem, may also be applied to (8). A GN approach for fitting the PARAFAC model has first been proposed in [13]. The question of which algorithmic approach is the most efficient has not yet been thoroughly investigated and further depends on the characteristics of the problem and the particular batch of data at hand. We use the COMFAC implementation of TALS described in [29]. COMFAC is basically a faster implementation of TALS that employs data compression using the *Tucker3* three-way model.

VI. Simulation

Given that our identifiability results and algorithms apply to both data domain and subspace problem formulations, an important question is which formulation to choose. A secondary consideration is which algorithm to choose. Since the latter depends on many factors, including data conditioning and whether/which constraints are enforced during the iterations, we confine ourselves to the following two goals: i) to illustrate identifiability results and ii) to report our experimental findings on the data domain versus subspace issue.

The permutation and scale ambiguity is inherently unresolvable from output data only and a common denominator to all blind identification/source separation methods. In the present sensor array processing context, recovering the source symbol, direction finding, and subarray response matrices within a common permutation matrix ambiguity simply amounts to shuffling the sources. If signal copy is the ultimate objective, then the permutation ambiguity can be resolved by resorting to *a priori* or embedded information, e.g., known DOA's or user ID bits, respectively. The (generally complex) scale ambiguity is relevant for digital communications applications, but it can be resolved using automatic gain control (AGC) and differential encoding/decoding [24, p. 187] or phase estimation. Note that column-wise scale ambiguity in the estimated $\boldsymbol{\Phi}$ is irrelevant because the extraction of azimuth and elevation information involves normalization with respect to the reference subarray, which effectively amounts to dividing the elements across one column with the top element. For the purpose of performance evaluation only, the permutation ambiguity is resolved using a greedy least squares matching algorithm. SNR is defined in terms of the observation model in (2) as $10log_{10}(\|\mathbf{GS}\|_F^2/\|\mathbf{V}\|_F^2)$. Complex Gaussian source signals are used throughout.

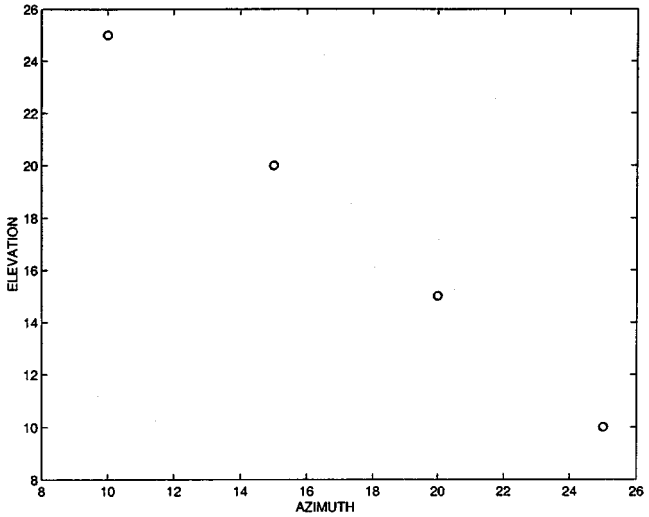

Fig. 2. Azimuth-elevation scatter, SNR = 54 dB.

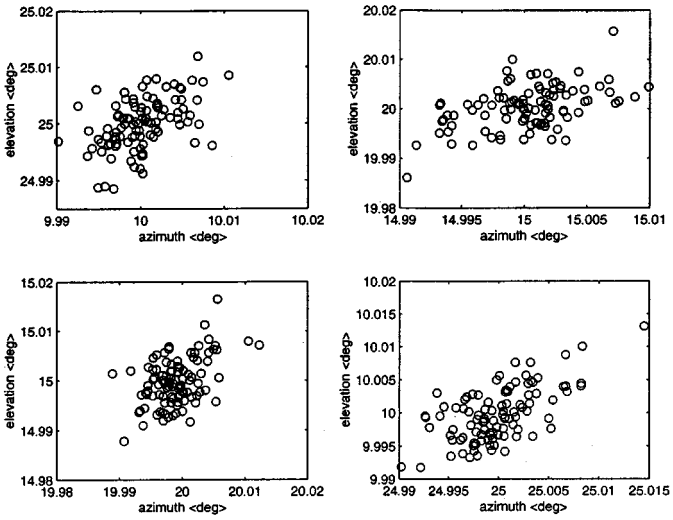

Fig. 3. Azimuth-elevation scatter zoom-in, SNR = 54 dB.

The first experiment is designed to illustrate performance in a high SNR scenario. Figs. 2 and 3 present azimuth-elevation scatter diagrams for $M = 4$ sources with (azimuth, elevation) equal to (10,25), (15,20), (20,15), and (25,10)°, using data-domain LS fitting as in (8) on a 5×5 uniform square array, $\lambda/4$ sensor spacing, three maximal-overlap six-element subarrays, $N = 100$ symbols, and SNR = 54 dB. Figs. 4 and 5 present the same for a 10×10 array, three maximal-overlap 81-element subarrays, $N = 300$ symbols, and SNR = 12 dB.

In order to compare data domain (8) versus (weighted) subspace fitting (9), we simulated the experimental setup in [31, ex. 1]. This example entails $M = 2$ uncorrelated sources, with (azimuth, elevation) equal to (−10,10), (−5,7)°, a 5×5 uniform square array, $P = 4$ maximal-overlap 16-element subarrays, $\lambda/2$ sensor spacing, SNR of 13 dB (10 dB per source),

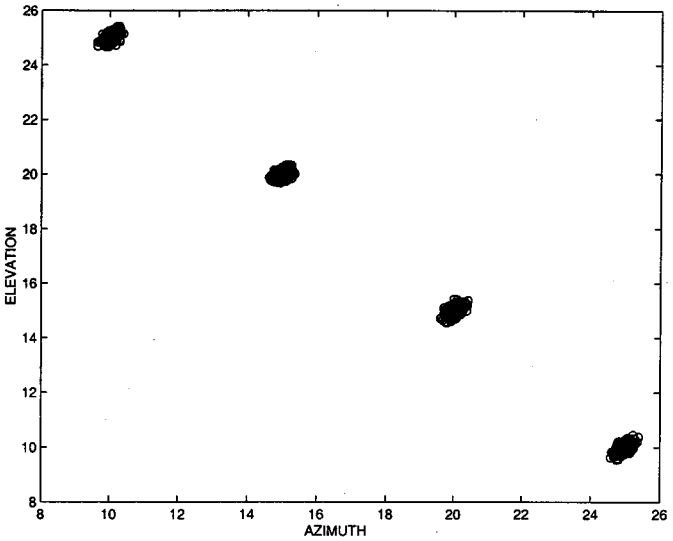

Fig. 4. Azimuth-elevation scatter, SNR $= 12$ dB.

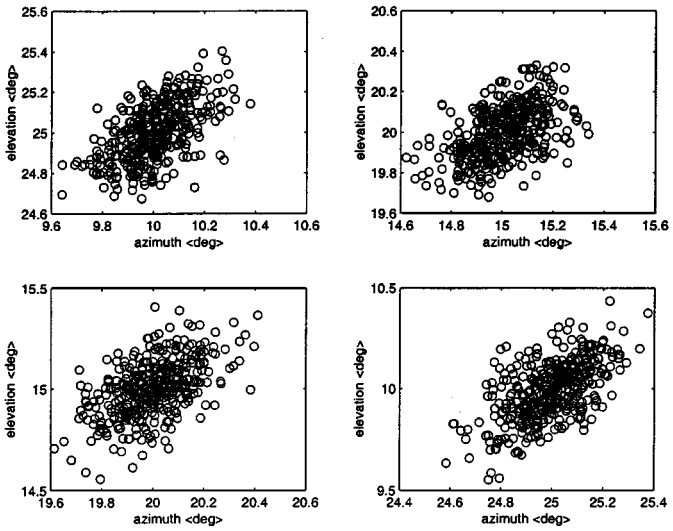

Fig. 5. Azimuth-elevation scatter zoom-in, SNR $= 12$ dB.

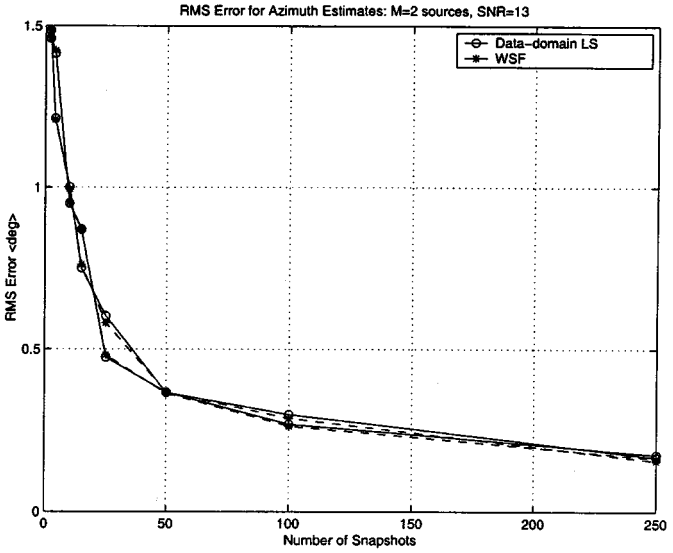

Fig. 6. Azimuth RMSE: Two uncorrelated sources, $(-10^\circ, 10^\circ)$, $(-5^\circ, 7^\circ)$, SNR $= 13$ dB.

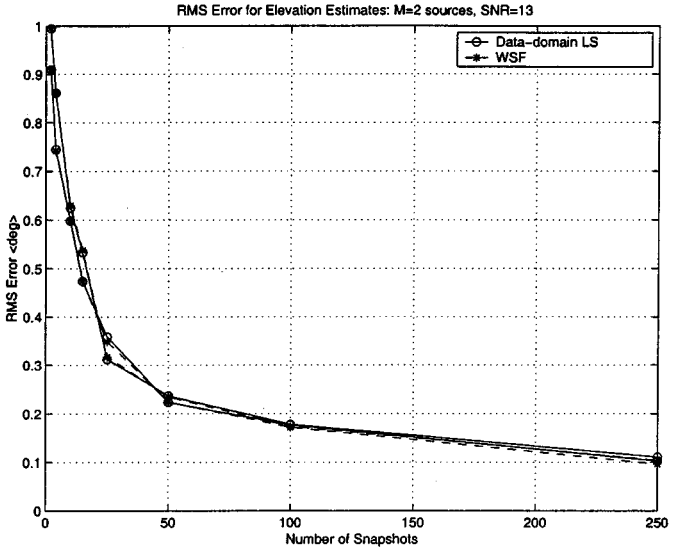

Fig. 7. Elevation RMSE: Two uncorrelated sources, $(-10^\circ, 10^\circ)$, $(-5^\circ, 7^\circ)$, SNR $= 13$ dB.

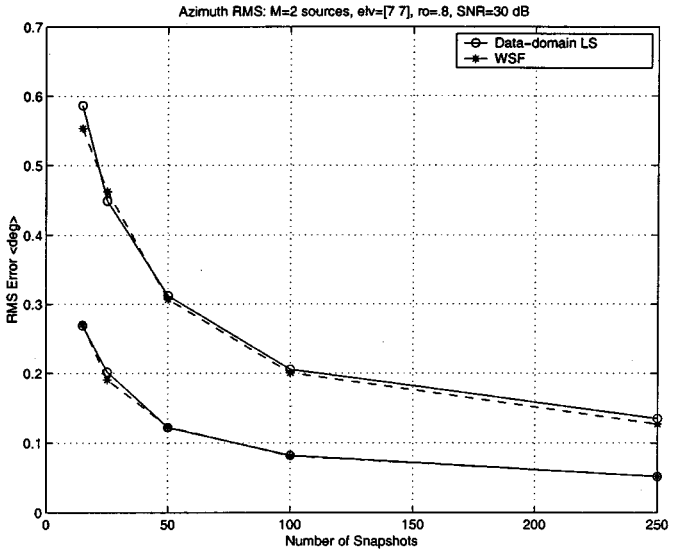

Fig. 8. Azimuth RMSE: Two 0.8-correlated sources, $(-10^\circ, 7^\circ)$, $(-5^\circ, 7^\circ)$, SNR $= 30$ dB.

250 Monte-Carlo trials, and $N = 25, 50, 100, 250$ symbol snapshots. The WSF weighting matrix in (9) was chosen according to [31]; see also [23], [37], and [39]. No weighting was used in (8). Single-invariance ESPRIT was used to initialize both (8) and (9). The results are reported in Figs. 6 and 7 in terms of azimuth and elevation root mean square error, respectively, for both (8) and (9) versus sample size. A more difficult scenario involving two correlated sources with correlation coefficient 0.8, common elevation of 7°, and SNR of 30 dB was also simulated. The remaining parameters are identical to those in Figs. 6 and 7. The results are reported in Figs. 8 and 9. Note that (8) and (9) provide comparable results, even in the relatively difficult scenario of Figs. 8 and 9. However, the situation can change quite drastically in tough cases. For the same setup as in Figs. 8 and 9, but this time taking SNR down to 23 dB, data-domain fitting behaves considerably worse than WSF, as shown in Fig. 10. The erratic behavior of the data domain LS curve is

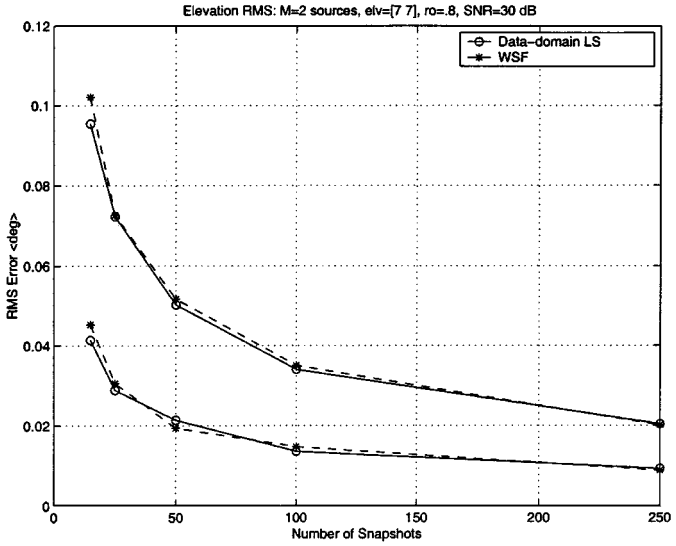

Fig. 9. Elevation RMSE: Two 0.8-correlated sources, $(-10^\circ, 7^\circ)$, $(-5^\circ, 7^\circ)$, SNR $= 30$ dB.

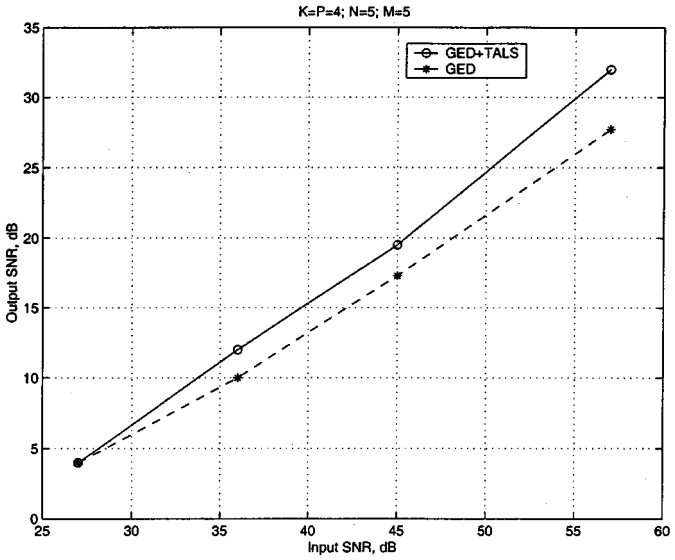

Fig. 11. Average output ($\hat{\mathbf{S}}$) SNR versus input SNR for $M = 5$ independent complex Gaussian sources, $K = P = 4$, $N = 5$, 1000 Monte Carlo trials.

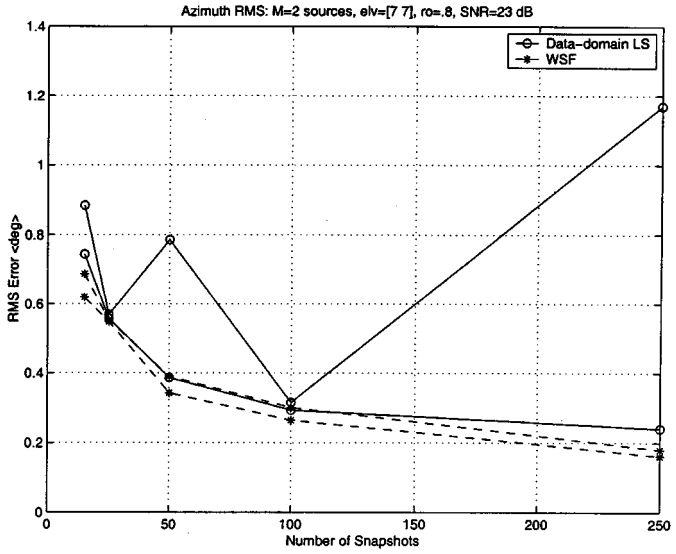

Fig. 10. Azimuth RMSE: Two 0.8-correlated sources, $(-10^\circ, 7^\circ)$, $(-5^\circ, 7^\circ)$, SNR $= 23$ dB.

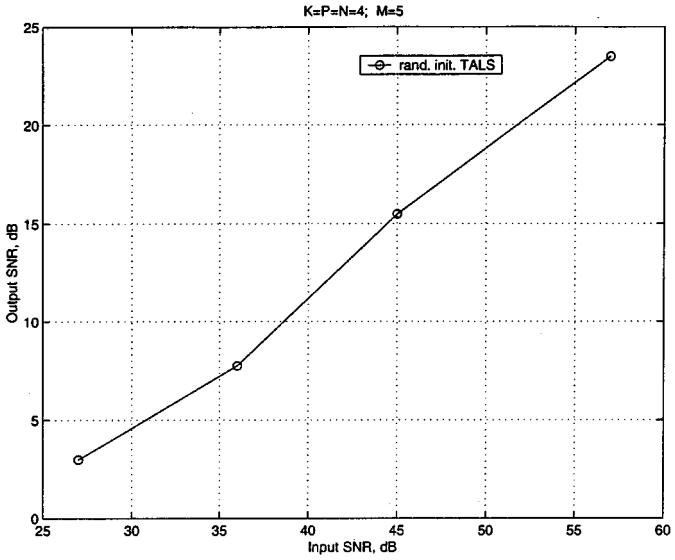

Fig. 12. Average output ($\hat{\mathbf{S}}$) SNR versus input SNR for $M = 5$ independent complex Gaussian sources, $K = P = N = 4$, 1000 Monte Carlo trials.

a manifestation of bad runs leading to local minima. This can happen in tough cases involving a combination of close-to-colinear sources and/or low SNR and/or closely spaced DOA's. The difference between data domain LS fitting and WSF can be attributed to the difference between deterministic and stochastic ML [39], to which WSF is asymptotically equivalent for proper $\mathbf{W}$. Our experimental results are consistent with [39], especially [39, Fig. 5(a), p. 2445]. Although we do not pursue this thread herein, note that weighting can also be applied in the data domain to improve performance, and a few reinitializations can help alleviate problems with local minima (but at the cost of runtime complexity).

The last two simulation experiments are meant to illustrate the validity of the identifiability conditions of Theorem 2/Proposition 1, and Theorem 3. Theorem 3 is first. Fig. 11 depicts source signal estimation results for $K = 4, P = 4$ nonoverlapping subarrays, $N = 5$ snapshots, and $M = 5$ sources, using randomly drawn complex Gaussian $\mathbf{A}$, $\mathbf{S}$, $\boldsymbol{\Phi}_1$, and $\mathbf{V}$ for a total of 1000 Monte Carlo trials. Both eigendecomposition and data-domain LS results are depicted. Output SNR improves roughly linearly with increasing input SNR, and acceptable estimation results can be obtained even with these small sample sizes in all three dimensions, provided input SNR is high enough—a clear manifestation of identifiability. Of course, better results can be obtained by increasing K, P, or N.

Taking N down to $4 < M$ means that Theorem 3 is no longer applicable, but Theorem 2 (and Proposition 1) is. Fig. 12 depicts source signal estimation results for $K = P = N = 4$ and, otherwise, the same setup as Fig. 11. This time data-domain LS is randomly initialized. The results are obviously worse than

Fig. 11 but still consistent with increasing SNR, which provides experimental validation of Theorem 2.

VII. Conclusions

This paper has established a link between MI-SAP and PARAFAC analysis. What PARAFAC brings to the table is primarily in the form of strong identifiability results. Conceptually, the link highlights the fact that single- and multiple-invariance ESPRIT uniqueness stems from the uniqueness of low-rank decomposition of three-way arrays. Interestingly, the link works both ways: Theorem 3 was motivated by multiple-invariance single-parameter ESPRIT, and it is the first constructive proof of PARAFAC uniqueness that is applicable when two of the three modes are smaller than the number of factors (sources). Proposition 1 also shows that uniqueness extends to subspace models in a lossless fashion.

Acknowledgment

The authors wish to acknowledge the influential work of Dr. J. Kruskal of the Mathematical Sciences Center at Bell Labs. N. Sidiropoulos further acknowledges personal communication (and a revised copy of [20] including clarifications) with Dr. Krushal. The authors would also like to acknowledge helpful feedback from anonymous reviewers.

References

[1] R. Bro, "PARAFAC: Tutorial and applications," *Chemometrics Intell. Lab. Syst.*, vol. 38, pp. 149–171, 1997.

[2] R. Bro and C. A. Andersson, "Improving the speed of multi-way algorithms. Part II: Compression," *Chemometrics Intell. Lab. Syst.*, to be published.

[3] R. Bro, R. Harshman, and N. D. Sidiropoulos, "Rank-deficient models for multi-way data Part I: Theory," and "Rank-deficient models for multi-way data Part II: Flow injection analysis," , submitted for publication.

[4] R. Bro and N. D. Sidiropoulos, "Least squares regression under unimodality and nonnegativity constraints," *J. Chemometrics*, vol. 12, pp. 223–247, 1998.

[5] 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.

[6] ——, "A fast least squares algorithm for separating trilinear mixtures," in *Proc. Int. Workshop Ind. Component Anal. Blind Signal Separation*, Aussois, France, Jan. 11–15, 1999.

[7] J. D. Carroll and J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.

[8] J. D. Carroll, S. Pruzansky, and J. B. Kruskal, "CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters," *Psychometrika*, vol. 45, pp. 3–24, 1980.

[9] R. B. Cattell, "Parallel proportional profiles and other principles for determining the choice of factors by rotation," *Psychometrika*, vol. 9, pp. 267–283, 1944.

[10] P. Comon, "Blind channel identification and extraction of more sources than sensors," in *Proc. SPIE*, San Diego, CA, July 19–24, 1998.

[11] M. Haardt and J. A. Nossek, "Simultaneous Schur decomposition of several nonsymmetric matrices to achieve automatic pairing in multidimensional harmonic retrieval problems," *IEEE Trans. Signal Processing*, vol. 46, pp. 161–169, Jan. 1998.

[12] M. Haardt, M. D. Zoltowski, C. P. Mathews, and J. A. Nossek, "2D unitary ESPRIT for efficient 2D parameter estimation," in *Proc. ICASSP*, vol. 3, Detroit, MI, May 8–12, 1995, pp. 2096–2099.

[13] C. Hayashi and F. Hayashi, "A new algorithm to solve the PARAFAC model," *Behaviormetrika*, vol. 11, pp. 49–60, 1982.

[14] R. A. Harshman, "Foundation of the PARAFAC procedure: Model and conditions for an 'explanatory' multi-mode factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, Dec. 1970.

[15] ——, "Determination and proof of minimum uniqueness conditions for PARAFAC1," *UCLA Working Papers in Phonetics*, vol. 22, pp. 111–117, 1972.

[16] ——, "PARAFAC2: Mathematical and technical notes," *UCLA Working Papers in Phonetics*, vol. 22, pp. 33–44, 1972.

[17] 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.

[18] B. Hochwald and A. Nehorai, "Identifiability in array processing models with vector-sensor applications," *IEEE Trans. Signal Processing*, vol. 44, pp. 83–95, Jan. 1996.

[19] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Signal Processing Mag.*, pp. 67–95, July 1996.

[20] 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.

[21] ——, "Rank decomposition, and uniqueness for 3-way and N-way arrays," in *Multiway Data Analysis*, R. Coppi and S. Bolasco, Eds. Amsterdam, The Netherlands: North-Holland, 1988, pp. 7–18.

[22] L. de Lathauwer, "Signal processing based on multilinear algebra," Doctoral thesis, Katholieke Univ. Leuven, Leuven, Belgium, Sept. 1997.

[23] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Trans. Signal Processing*, vol. 40, pp. 590–600, Mar. 1992.

[24] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[25] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invriance techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 984–995, July 1989.

[26] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 276–280, Mar. 1986.

[27] P. H. Schonemann, "An algebraic solution for a class of subjective metrics models," *Psychometrika*, vol. 37, p. 441, 1972.

[28] 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.

[29] ——, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. Signal Processing*, vol. 48, pp. 810–823, Mar. 2000.

[30] N. D. Sidiropoulos and X. Liu, "Identifiability conditions for deterministic blind beamforming in the presence of incoherent multipath with small delay spread," in *Proc. EUSIPCO*, Tampere, Finland, Sept. 5–8, 2000.

[31] A. Swindlehurst and T. Kailath, "Algorithms for azimuth-elevation direction finding using regular array geometries," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, pp. 145–156, Jan. 1993.

[32] A. Swindlehurst, B. Ottersten, R. Roy, and T. Kailath, "Multiple invariance ESPRIT," *IEEE Trans. Signal Processing*, vol. 40, pp. 868–881, Apr. 1992.

[33] 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.

[34] K.-C. Tan and Z. Goh, "A detailed derivation of arrays free of higher rank ambiguties," *IEEE Trans. Signal Processing*, vol. 44, pp. 351–359, Feb. 1996.

[35] K.-C. Tan, S. S. Goh, and E.-C. Tan, "A study of the rank-ambiguity issues in direction-of-arrival estimation," *IEEE Trans. Signal Processing*, vol. 44, pp. 880–887, Apr. 1996.

[36] A.-J. van der Veen, P. B. Ober, and E. F. Deprettere, "Azimuth and elevation computation in high resolution DOA estimation," *IEEE Trans. Signal Processing*, vol. 40, pp. 1828–1832, July 1992.

[37] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, pp. 1110–1121, May 1991.

[38] B. Ottersten, M. Viberg, and T. Kailath, "Performance analysis of the total least squares ESPRIT algorithm," *IEEE Trans. Signal Processing*, vol. 39, pp. 1122–1135, May 1991.

[39] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, pp. 2436–2449, Nov. 1991.

[40] M. Wax and I. Ziskind, "On unique localization of multiple sources by passive sensor arrays," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 37, pp. 996–1000, July 1989.

[41] K. T. Wong and M. D. Zoltowski, "Closed-form multi-dimensional multi-invariance ESPRIT," *Proc. ICASSP*, vol. 5, pp. 3489–3492, Apr. 21–24, 1997.
[42] M. Zoltowski and D. Stavrinides, "Sensor array signal processing via a procrustes rotations based eigenanalysis of the ESPRIT data pencil," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, p. 832, June 1989.
[43] N. D. Sidiropoulos and X. Liu, "PARAFAC methods for deterministic blind beamforming: Identifiability," , submitted for publication.

Nicholas D. Sidiropoulos (M'92–SM'99) received the diploma degree 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, 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 also was 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, University of Virginia, Charlottesville, in July 1997 as an Assistant Professor. He is currently an Associate Professor in the Department of Electrical and Computer Engineering at the University of Minnesota, Minneapolis. His current research interests are primarily in multi-way 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 Signal Processing Society and currently serves as Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. He received the NSF/CAREER award (Signal Processing Systems Program) in June 1998.

Rasmus Bro received the M.Sc. degree in analytical and medical chemistry from the Technical University, Lyngby, Denmark, in 1994. He received the Ph.D. degree in chemometrics from the University of Amsterdam, Amsterdam, The Netherlands.

From 1994 to 1998, he was first a Research Assistant and then Assistant Professor (chemometrics/data analysis) with the Chemometrics Group, Food Technology, Department of Food and Dairy Science, Royal Veterinary and Agricultural University, Frederiksberg, Denmark. In the spring of 1995, he spent two months in the Process Analysis Group at the Department of Analytical Chemistry, University of Amsterdam. Since 1998, he has worked as a Research Associate Professor at the Royal Veterinary and Agricultural University in the Chemometrics Group. From October 1997 to March 1998, he worked intermittantly as a Guest Researcher at MATFORSK (The Food Research Institute), Aas, Norway. His main research interests are mathematical modeling in a broad sense, mostly associated with so-called multiway modeling and constrained programming.

Georgios B. Giannakis (F'96) received the Diploma in electrical engineering from the National Technical University of Athens, Athens, Greece in 1981. From September 1982 to July 1986, he was with the University of Southern California (USC), Los Angeles, where he received the M.Sc. degree in electrical engineering in 1983, the M.Sc. degree in mathematics in 1986, and the Ph.D. degree in electrical engineering in 1986.

After lecturing for one year at USC, he joined the University of Virginia (UVA), Charlottesville, in 1987, where he became a Professor of electrical engineering in 1997, Graduate Committee Chair, and Director of the Communications, Controls, and Signal Processing Laboratory in 1998. Since January 1999, he has been with the University of Minnesota, Minneapolis, as a Professor of electrical and computer engineering. His general interests lie in the areas of signal processing and communications, estimation and detection theory, time-series analysis, and system identification—subjects on which he published more than 100 journal papers. Specific areas of expertise have included (poly)spectral analysis, wavelets, cyclostationary, and non-Gaussian signal processing with applications to sensor array and image processing. Current research focuses on transmitter and receiver diversity techniques for equalization of single- and multiuser communication channels, mitigation of rapidly fading channels, compensation of nonlinear amplifier effects, redundant filterbank transceivers for block transmissions, multicarrier, and wideband communication systems.

Dr. Giannakis was awarded the School of Engineering and Applied Science Junior Faculty Research Award from UVA in 1988 and the UVA-EE Outstanding Faculty Teaching Award in 1992. He received the IEEE Signal Processing Society's 1992 Paper Award in the Statistical Signal and Array Processing (SSAP) area and the 1999 Young Author Best Paper Award (with M. K. Tsatsanis). He co-organized the 1993 IEEE Signal Processing Workshop on Higher-Order Statistics, the 1996 IEEE Workshop on Statistical Signal and Array Processing, and the first IEEE Signal Processing Workshop on Wireless Communications in 1997. He guest (co-)edited two special issues on high-order statistics (*International Journal of Adaptive Control and Signal Processing* and the EURASIP journal *Signal Processing*) and the January 1997 Special Issue on Signal Processing for Advanced Communications of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS, a Secretary of the Signal Processing Conference Board, a member of the SP Publications Board, and a member and vice-chair of SSAP Technical Committee. He now chairs the Signal Processing for Communications Technical Committee and serves as Editor-in-Chief of the IEEE SIGNAL PROCESSING LETTERS. He is also a member of the IEEE Fellow Election Committee, the IMS, and the European Association for Signal Processing.