Wideband multiple diversity tensor array processing

Francesca Raimondi, Rodrigo Cabral Farias, Olivier J.J. Michel, Pierre Comon

Wideband multiple diversity tensor array processing

Francesca E. D. Raimondi, Rodrigo Cabral Farias, Olivier Michel, Pierre Comon

***Abstract*—This paper establishes a tensor model for wideband coherent array processing including multiple physical diversities. A separable coherent focusing operation is proposed as a pre-processing step in order to ensure the multi-linearity of the interpolated data. We propose an ALS algorithm to process tensor data, taking into account the noise correlation structure introduced by the focusing operation. We show through computer simulations that the estimation of DoA and polarization parameters improves compared to existing narrowband tensor processing and wideband MUSIC. The performance is also compared to Cramér-Rao bounds of the wideband tensor model.**

***Index Terms*—localization, DoA estimation, interpolation, coherent, wideband, antenna array processing, tensor, MUSIC, High Resolution, polarization, CP decomposition**

I. Introduction

Direction of Arrival (DoA) estimation is a central problem in array signal processing, and in particular in telecommunications, seismology, speech, biomedical engineering, and astronomy. We are interested in acquisition systems composed of multiple sensors that receive source signals from different directions [1], [2]. In order to estimate DoA, existing techniques such as beamforming, MUSIC (MUltiple SIgnal Classification) [3], root-MUSIC [4] and ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) [5], take advantage of the space diversity embedded in the acquisition system. In particular, the time delay of a narrowband wave received at a given sensor can be expressed in baseband by a simple multiplication by a complex exponential carrying the information on the DoA. In addition to the space diversity, ESPRIT assumes the presence of two identical subarrays, displaced from each other by a space shift.

The space shift diversity was extended to more than one displacement through tensor decomposition of narrow-band waves, through a deterministic approach that allows to separate and estimate more sources than sensors in each subarray [6]. Tensor analysis requires at least three diversities to allow a multilinear decomposition[1] [7], that guarantees uniqueness under mild conditions.

Another element of diversity is given by wave polarization, whenever each sensor has multiple components [8], [9], [10], [11]. The above mentioned tensor approach was extended to include polarization diversity in [12], [13], [14], and propagation speed diversity of seismic waves in [15]. A major advantage of deterministic tensor analysis involving time or frequency as diversities, lies in its ability to estimate source waveforms together with DoAs.

MUSIC, ESPRIT and tensor decomposition for array processing originally addressed the narrow-band case. In fact, in wideband, signal subspace and steering matrices vary with frequency, thus requiring a focusing operation onto the same reference subspace. This approach, named *Coherent Signal Subspace* (CSS), may require an approximate prior estimate of DoA to form focusing matrices [16], [17], [18], or virtual arrays obtained by spatial interpolation [19], [20]. The latter extend the MUSIC algorithm to wideband processing for Uniform Linear Arrays (ULA). To our knowledge, only recently coherent tensor analysis was extended to process wideband waves for ULAs in [21], where a spatial interpolation technique was adopted. For sake of clarity, Table I illustrates the state of the art and our contributions: rows refer to theoretical approaches, and columns to physical diversities taken into account.

This paper aims at establishing a tensor model for high resolution wideband array processing with multiple antenna diversities (space, space shift, polarization and spatial gain patterns), under the assumption of plane waves in the far field. The multilinear coherent subspace preprocessing, required in the wideband case, introduces a correlation in the noise structure. An algorithm for tensor decomposition taking into account the noise covariance is proposed, and the performance is evaluated via the Cramér-Rao Bound (CRB).

This paper is organized as follows: Section II introduces the physical model and main issues of wideband array processing; Section III illustrates the main antenna diversities that can be exploited in tensor processing; Section IV illustrates the effect of wideband coherent processing on the tensor formulation; tensor notations and multilinear decompositions are introduced in Section V; the CRB is computed in Section VI; Section VII addresses algorithmic issues; Section VIII reports computer results for the particular case of space shift and polarization diversities, in comparison with wideband MUSIC and with the multilinear CRB.

II. From narrow-band to wide-band

A. Observation model

We consider R radiating sources in the far-field[2], s_r, $1 \leq r \leq R$, arriving from directions defined by unit vectors $\boldsymbol{d}(\theta_r)$, θ_r being a pair of angles in 3D, or a single angle in 2D. These sources impinge on an arbitrary array of L sensors located at

F. E. D. Raimondi, R. Cabral Farias, P. Comon and O. Michel are with GIPSA-Lab, CNRS, Univ. Grenoble Alpes, 38000 Grenoble, France (e-mail: firstname.lastname@gipsa-lab.grenoble-inp.fr).

This work was supported by the ERC Grant AdG-2013-320594 "DECODA".

[1]This decomposition is often referred to as *Canonical Polyadic* (CP), and sometimes as CanDecomp or PARAFAC.

[2]By "far-field assumption", it is meant that the wavefront curvature and dissipation are neglected.

Diversity	Time	Space	Freq.	Space Shift	Polar
MUSIC [3], [22]	s, $\times$	$\times$			
Vector Sensor MUSIC [3], [9], [10]	s, $\times$	$\times$			$\times$
WB MUSIC [23], [17], [24], [19], [20], [25]	s	$\times$	$\times$		
WB Vector Sensor MUSIC [26]	s	$\times$	$\times$		$\times$
ESPRIT ($M = 2$) [5]	s, $\times$	$\times$		$\times$	
WB ESPRIT [18], [27]	s	$\times$	$\times$	$\times$	
Tensor ESPRIT [6], [28], [29]	$\times$	$\times$		$\times$	
Vector Sensor CP [12], [13], [14]	$\times$	$\times$			$\times$
WB Tensor CP [21]		$\otimes$	$\otimes$	$\otimes$	
Gain Pattern CP	$\otimes$	$\otimes$	$\otimes$		
WB Vector Sensor CP		$\otimes$	$\otimes$		$\otimes$

TABLE I
MULTIPLE DIVERSITY IN ARRAY PROCESSING: s = AVERAGE OF THE COVARIANCE MATRIX THROUGH MULTIPLE SNAPSHOTS; WB = WIDEBAND; $\times$ = WHAT EXISTS IN LITERATURE, $\otimes$ = OUR CONTRIBUTION

$\boldsymbol{v}$	vector: in bold lower case
v_i	i-th element of $\boldsymbol{v}$
$\boldsymbol{A}$	matrix: in bold upper case
$\boldsymbol{I}_L$	identity matrix of size $L \times L$
$\boldsymbol{a}_i$	i-th column of $\boldsymbol{A}$
$\bar{\boldsymbol{g}}$	vector $\boldsymbol{g}$ without its first entry
$\tilde{\boldsymbol{x}}$	vector related to a virtual array
A_{ij}	$\{i,j\}$ element of $\boldsymbol{A}$
$\boldsymbol{A}^{\mathsf{T}}$	transpose of $\boldsymbol{A}$
$\boldsymbol{A}^*$	complex conjugate of $\boldsymbol{A}$
$\boldsymbol{A}^{\mathsf{H}}$	conjugate transpose of $\boldsymbol{A}$
$\boldsymbol{A}^\dagger$	Moore-Penrose pseudoinverse of $\boldsymbol{A}$
$\boldsymbol{v}^{\mathsf{T}}\boldsymbol{u}$	scalar product between real vectors $\boldsymbol{v}$ and $\boldsymbol{u}$
$\boldsymbol{v} \otimes \boldsymbol{u}$	outer (tensor) product between two vectors
$\boldsymbol{A} \boxtimes \boldsymbol{B}$	Kronecker product between $\boldsymbol{A}$ and $\boldsymbol{B}$
$\boldsymbol{A} \odot \boldsymbol{B}$	Hadamard (element-wise) product
$\boldsymbol{\mathcal{T}} \bullet_d \boldsymbol{A}$	mode-d product
$\parallel \boldsymbol{v} \parallel_{\boldsymbol{D}}$	$\sqrt{\boldsymbol{v}^{\mathsf{H}}\boldsymbol{D}^{-1}\boldsymbol{v}}$, weighted Euclidean norm
$\boldsymbol{\mathcal{T}}$	tensor: in bold calligraphic font
$\mathcal{T}_{ijk}$	$\{i,j,k\}$ element of $\boldsymbol{\mathcal{T}}$
$\mathrm{vec}\{\boldsymbol{\mathcal{T}}\}$	vectorization of $\boldsymbol{\mathcal{T}}$
$\dot{\boldsymbol{d}}(\theta)$	derivative $\partial \boldsymbol{d}/\partial\theta$

TABLE II
NOTATIONS

positions $\boldsymbol{p}_\ell$, $1 \le \ell \le L$. The signal received at the ℓ-th sensor at time t can be modeled as:

$$x_\ell[t] = \sum_{r=1}^{R} g_\ell(\theta_r) s_r[t - \tau_\ell(\theta_r)] + n_\ell[t] \tag{1}$$

where $s_r[t]$ is the signal transmitted by the r-th source, $t \in \{1, 2, \dots T\}$, $g_\ell(\theta_r)$ is the gain of the ℓ-th sensor, assumed to be independent of frequency, and $n_\ell[t]$ is an additive noise.

The delay of arrival $\tau_\ell(\theta)$ is directly related to sensor locations and DoA's via the expression

$$\tau_\ell(\theta) = \boldsymbol{p}_\ell^{\mathsf{T}} \boldsymbol{d}(\theta)/c \tag{2}$$

c being the wave propagation speed and $(^{\mathsf{T}})$ transposition.

At this stage, it is important to fix a scaling indeterminacy, present in all blind approaches. Observe that changing the pair $(g_\ell(\theta_r), s_r)$ into $(\rho g_\ell(\theta_r), s_r/\rho)$, $\forall \rho \neq 0$, leads to the same model. So we may fix $g_1(\theta) = 1$, $\forall\theta$, without restricting the generality, which means that the first sensor is treated as omnidirectional.

Denote by ω the radial frequency, and by $x(\omega)$ the Fourier Transform (FT) of $x[t]$. In this framework, model (1) can be equivalently rewritten in the frequency domain:

$$x_\ell(\omega) = \sum_{r=1}^{R} g_\ell(\theta_r)\, e^{-\jmath\omega\tau_\ell(\theta_r)} s_r(\omega) + n_\ell(\omega) \tag{3}$$

where $\jmath = \sqrt{-1}$, $x_\ell(\omega) = \mathcal{F}\{x_\ell[t]\}$ is the Fourier transform of $x_\ell[t]$ and g_ℓ are *a priori* unknown real functions unless otherwise specified, and $n_\ell(\omega)$ refers to a circular Gaussian white noise process at the ℓ-th sensor. The noise processes at different sensors are considered to be identically distributed and uncorrelated from one another. Alternatively, in vector form, (3) becomes

$$\boldsymbol{x}(\omega) = \boldsymbol{A}(\omega, \boldsymbol{\theta})\, \boldsymbol{s}(\omega) + \boldsymbol{n}(\omega) \tag{4}$$

where the $L \times R$ matrix $\boldsymbol{A}(\omega, \boldsymbol{\theta})$ depends on the vector[3] of directions of arrival (DoA's), $\boldsymbol{\theta} = [\theta_1, .., \theta_R]^{\mathsf{T}}$. The r-th column of $\boldsymbol{A}(\omega, \boldsymbol{\theta})$, denoted $\boldsymbol{a}(\omega, \theta_r)$ in the remainder, is the value of the *array manifold* taken at $\theta = \theta_r$.

To summarize, we assume the following hypotheses:

- the first sensor ($\ell = 1$) is taken as origin, so that $\boldsymbol{p}_1 = \boldsymbol{0}$, and has a unit gain in all directions, *i.e.* $g_1(\theta) = 1$, $\forall\theta$;
- the sensor gains $g_\ell(\theta)$ are real (which is actually equivalent to assuming that their phase is known) and frequency-flat;
- sources $s_r[t]$ are deterministic;
- the wave propagation speed c does not depend on frequency (*i.e.* the medium is not dispersive)[4];
- noise is circular complex white Gaussian.

[3] In 3D, $\boldsymbol{\theta}$ is a $R \times 2$ matrix of angles; it is a $R \times 1$ vector in 2D.

[4] The principles would actually remain valid for dispersive media provided the function $c(\omega)$ is known.

B. Virtual arrays

High Resolution (HR) methods are based on a partition of the observation space into signal and noise subspaces, usually via low-rank approximations. The advantage of transforming the dynamical model (1) into model (3) is that the latter is formally static for every fixed radial frequency ω, which makes it easier to define and compute the above-mentioned subspaces. The drawback is that these subspaces change with frequency, which raises difficulties in the frame of a wideband coherent processing. As a consequence, in order to estimate coherently the DoA's, it is necessary to first steer them towards a subspace defined at a reference radial frequency ω_0 [17], [19], [20], [23].

For this purpose, the complex exponential in (3) needs to become constant, in order to constitute a coherent contribution. Since sensor gains do not depend on frequency, this steering operation leads to defining a virtual array for each radial frequency ω defined by sensor positions $\tilde{\boldsymbol{p}}_\ell(\omega)$, related to actual sensor positions $\boldsymbol{p}_\ell$ by:

$$[\omega\tilde{\boldsymbol{p}}_\ell(\omega) - \omega_0\boldsymbol{p}_\ell]^\mathsf{T}\boldsymbol{d}(\theta) \equiv 0,\ mod\ 2\pi c, \tag{5}$$

where ω_0 is fixed. This equation is satisfied if the term within brackets is orthogonal to $\boldsymbol{d}(\theta_r)$ for every DoA, $1 \leq r \leq R$; but this requires a prior knowledge of DoA's [17]. On the other hand, a sufficient condition is clearly that $\tilde{\boldsymbol{p}}_\ell(\omega) = \frac{\omega_0}{\omega}\boldsymbol{p}_\ell$, which this time does not depend on θ; this corresponds to the stretched arrays proposed in [19], [20], [23].

Remark: More generally, if gains depended on frequency, equation (5) would become

$$\exp\{\jmath[\omega\tilde{\boldsymbol{p}}_\ell(\omega) - \omega_0\boldsymbol{p}_\ell]^\mathsf{T}\boldsymbol{d}(\theta)\} = \frac{g_\ell(\omega,\theta)}{g_\ell(\omega_0,\theta)}. \tag{6}$$

Satisfying this equation would then require prior knowledge of DoAs.

C. Interpolation

Equation (5) defines virtual arrays with identical array manifolds, but does not provide the signals that would be measured by the latter. It is hence necessary to devise an interpolator to compute synthetic measurements. Moreover, to preserve the separation between noise and signal subspaces, it is suitable to map $\boldsymbol{a}(\omega,\theta)$ to the same manifold $\boldsymbol{a}(\omega_0,\theta)$ *via* a *multiplication* by some $L \times L$ matrix $\boldsymbol{U}(\omega,\omega_0,\theta)$, which we call focusing matrix. It turns out that this is possible exactly only for at most L distinct values of θ. There are thus two possibilities.

- Either the DoA's are unknown, and a focusing matrix $\boldsymbol{U}(\omega,\omega_0,\theta)$ can be computed so as to satisfy the following relationship only approximately [23]:

$$\boldsymbol{U}(\omega,\omega_0,\Theta_k)\,\boldsymbol{a}(\omega,\theta) \approx \boldsymbol{a}(\omega_0,\theta) \tag{7}$$

  in an angular sector $\theta \in \Theta_k$ of limited extent, where ω_0 is a fixed reference frequency. They can be obtained by solving in the Least Squares (LS) sense the linear system $\boldsymbol{U}(\omega,\omega_0,\Theta_k)\mathbb{A}(\omega,\Theta_k) \approx \mathbb{A}(\omega_o,\Theta_k)$, where Θ_k contains I discrete angular values θ_i, $1 \leq i \leq I$, and matrix $\mathbb{A}(\omega,\Theta_k)$ is built with the corresponding columns $\boldsymbol{a}(\omega,\theta_i)$; see [30] for further details on how to assess the accuracy of an array interpolator.
- Or the R DoA's are approximately known in advance, $R \leq L$, and it is possible to satisfy (7) exactly for these approximate DoA's; see [17] for more details.

D. Example: the covariance matrix approach

For the sake of completeness, we shall now describe the model described in [23], [17], as it is used in the Wideband MUSIC algorithm in Section VIII for the sake of comparison. In the state of the art [23], [17], the previously explained interpolation method is used as a preprocessing for subspace methods (such as MUSIC), in order to solve the problem of coherently averaging the estimated covariance matrices. In fact, this model is based on second order moments whereas a deterministic approach will be described along with a multilinear interpolation procedure in Section V. This is illustrated below, as the covariance matrix $\boldsymbol{R} = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\mathsf{H}]$ depends on radial frequency ω:

$$\boldsymbol{R}(\omega) = \boldsymbol{A}(\omega,\boldsymbol{\theta})\boldsymbol{R}_s(\omega)\boldsymbol{A}(\omega,\boldsymbol{\theta})^\mathsf{H} + \boldsymbol{R}_n(\omega) \tag{8}$$

As a consequence, in order to estimate the DoA's by a coherent average over different values of ω, it is necessary to first steer all estimated $\hat{\boldsymbol{R}}(\omega)$ towards a subspace defined at a reference radial frequency ω_0 [17], [19], [20], [23]. By doing so, in every angular sector Θ_k we can use contributions of all frequencies to build an averaged estimate of the covariance matrix:

$$\bar{\boldsymbol{R}}(\Theta_k) = \sum_\omega \boldsymbol{U}(\omega,\omega_0,\Theta_k)\,\boldsymbol{x}(\omega)\boldsymbol{x}(\omega)^\mathsf{H}\boldsymbol{U}(\omega,\omega_0,\Theta_k)^\mathsf{H} \tag{9}$$

From (4) and (7) we obtain

$$\bar{\boldsymbol{R}}(\Theta_k) \approx \sum_\omega \boldsymbol{A}(\omega_0,\boldsymbol{\theta})\boldsymbol{R}_s(\omega)\boldsymbol{A}(\omega_0,\boldsymbol{\theta})^\mathsf{H} + \bar{\boldsymbol{R}}_n(\omega)$$

where $\boldsymbol{R}_s(\omega)$ denotes the covariance of the sources at radial frequency ω and $\bar{\boldsymbol{R}}_n(\omega)$ the noise covariance transformed by matrix $\boldsymbol{U}(\omega,\omega_0,\Theta_k)$ in the considered sector, *i.e.* $\theta \in \Theta_k$. This summation corresponds to averaging observations received on virtual arrays, each being a stretched or shrunk version of a reference array at radial frequency ω_0.

III. Multiple diversities

In the previous section, space and time/frequency diversities were taken into account. Yet, direction of arrival estimation can be significantly improved if other modalities are considered, such as space shift and polarization. In this section, a multiway formulation including multiple diversities is presented.

A. Space shift and gain pattern diversity

Space shift diversity was initially exploited by the ESPRIT algorithm for two subarrays deduced from each other by a translation [5]. It has been extended to more than two subarrays in [6] giving rise to a third order tensor decomposition for array processing. The diversities involved are space, time, and space shift. It is also possible to consider diversities induced by polarized sensors, and by anisotropic sensor gains. We refer to the latter as *gain pattern diversity*.

Denote $\boldsymbol{\delta}_m$ the location of the m-th subarray. We may take the first subarray as origin, so that $\boldsymbol{\delta}_1 = \mathbf{0}$. In addition, the gain of sensor ℓ is assumed to be the same for any subarray m, and is given by $g_\ell(\theta)$.

Let $\zeta_m(\theta) = \boldsymbol{\delta}_m^{\mathsf{T}} \boldsymbol{d}(\theta)/c$ be the delay of arrival on subarray m, and denote by $\mathcal{F}\{\cdot\}$ the Fourier Transform operator applied to received signals in the time window $t \in [0, T]$. Then, in the presence of additive noise $\mathcal{N}$:

$$\begin{aligned} \mathcal{X}_{\ell m}(\omega) &= \mathcal{M}_{\ell m}(\omega) + \mathcal{N}_{\ell m}(\omega) \\ \mathcal{M}_{\ell m}(\omega) &= \sum_{r=1}^{R} A_{\ell r}(\omega)\, B_{mr}(\omega)\, S_r(\omega) \\ \text{with} \quad & \left\{ \begin{array}{lcl} \mathcal{X}_{\ell m}(\omega) & = & \mathcal{F}\{x_{\ell m}[t]\} \\ A_{\ell r}(\omega) & = & g_\ell(\theta_r)\, e^{-\jmath\omega\tau_\ell(\theta_r)} \\ B_{mr}(\omega) & = & e^{-\jmath\omega\zeta_m(\theta_r)} \\ S_r(\omega) & = & \mathcal{F}\{s_r[t]\} \end{array} \right. \end{aligned} \tag{10}$$

In the remainder, since received signals are real, we consider only positive frequencies (without any loss of information). After discretization, we have:

$$\mathcal{X}_{\ell m}(\omega_q) = \sum_{r=1}^{R} A_{\ell r}(\omega_q)\, B_{mr}(\omega_q)\, S_r(\omega_q) + \mathcal{N}_{\ell m}(\omega_q) \tag{11}$$

where frequency $0 < \omega_{min} \leq \omega_q \leq \omega_{max} < \pi$, and $1 \leq q \leq Q$. In what follows, (except in Section VI and in Appendix A-C) we will assume omnidirectional sensors, *i.e.* gain patterns are known with $g_\ell(\theta) = 1, \forall\theta, \forall\ell$.

B. Polarization diversity

If we have at our disposal a single antenna ($M = 1$) array of polarized elements, each capable of recording 3 orthogonal components[5], the observation model is similar to (10). If we omit measurement noise, it takes the form:

$$\mathcal{M}_{\ell p}(\omega_q) = \sum_{r=1}^{R} A_{\ell r}(\omega_q)\, K_{pr}\, S_r(\omega_q),\ 1 \leq p \leq P = 3 \tag{12}$$

where the R columns of the $3 \times R$ factor matrix $\boldsymbol{K}$ are detailed in [31], [7]:

$$\boldsymbol{k}_r = \boldsymbol{H}(\theta_r)\boldsymbol{W}(\alpha_r)\boldsymbol{w}(\beta_r)$$

Both angles $\alpha_r \in (-\pi/2, \pi/2]$ and $\beta_r \in [-\pi/4, \pi/4]$ characterize the polarization of the r-th source; see [31], [10] for general expressions of $\boldsymbol{H}(\theta)$, $\boldsymbol{W}(\alpha)$ and $\boldsymbol{w}(\beta)$. For instance Rayleigh waves considered in Section VIII belong to the model characterized by:

$$\begin{aligned} \boldsymbol{H}(\theta_r) &= \begin{bmatrix} \cos\phi_r \cos\psi_r & -\cos\phi_r \sin\psi_r \\ \sin\phi_r \cos\psi_r & -\sin\phi_r \sin\psi_r \\ \sin\psi_r & \cos\psi_r \end{bmatrix} \\ \boldsymbol{W}(\alpha_r) &= \begin{bmatrix} \cos\alpha_r & \sin\alpha_r \\ -\sin\alpha_r & \cos\alpha_r \end{bmatrix} \\ \boldsymbol{w}(\beta_r) &= \begin{bmatrix} \cos\beta_r \\ \jmath\sin\beta_r \end{bmatrix} \end{aligned} \tag{13}$$

[5]For electromagnetic waves, two 3-component sensors can be used, leading to an observation of dimension 6 [12].

where it is reminded that in 3D the DoA $\theta_r = (\phi_r, \psi_r)$ is formed by a couple of angles: azimuth ϕ_r and elevation ψ_r. Without restricting too much the generality, we assume that polarization does not depend on ω over the bandwidth of interest. The main difference between (11) and (12) is that the second factor matrix, $\boldsymbol{K}$, now depends on extraneous unknown angular parameters (α_r, β_r), instead of the subarray space shifts $\boldsymbol{\delta}_m$.

C. All together

It is clear that the aforementioned diversities can be modeled all together, at the price of an increased notational complexity. If frequency, space, space shift, and polarization diversities are considered simultaneously, the data measurements depend on 4 variables, the data array is of size $L \times M \times P \times Q$, with $P = 3$, and involves 4 unknown matrices to identify. Therefore, discretizing and merging models (11) and (12) yield

$$\mathcal{X}_{\ell mpq} = \sum_{r=1}^{R} A_{\ell r}(\omega_q)\, B_{mr}(\omega_q)\, K_{pr}\, S_{qr} + \mathcal{N}_{\ell mpq} \tag{14}$$

Traditional subspace processing usually breaks the multi-dimensionality of the multiple diversity model through a vectorization of the entire array manifold (including space, polarization and/or space shift) [3].

The tensor approach through CP decomposition fully takes advantage of the multilinearity of a model with at least three diversities [7]. However, in order to have a totally multilinear model in (14), we need to remove the dependency on ω_q from factor matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, *e.g.* through a bilinear interpolation. This matter will be addressed in Section IV, while tensor model and notations will be discussed in Section V.

IV. Tensor wideband processing

Equation (11), as a matrix equation for a given frequency, cannot yield a unique factorization. Hence such equations need to be treated simultaneously for all frequencies, through the tensor formalism. The bilinear interpolation we propose is precisely a means to coherently combine them in a manner that has a physical meaning. In this section, we aim at jointly exploiting models (11), (12) and (14) for several angular frequencies. As mentioned in Section III-C, the difficulty comes from the fact that these models are obviously not multilinear w.r.t. ω_q, because exponentials in $\boldsymbol{A}(\omega_q)$ and $\boldsymbol{B}(\omega_q)$ are a function of frequency.

A. Virtual arrays

To face this problem, the idea is to build virtual arrays, as described in (7); but we have now a second condition to satisfy, in addition to (5):

$$[\omega_q \tilde{\boldsymbol{\delta}}_m(\omega_q) - \omega_0 \boldsymbol{\delta}_m]^{\mathsf{T}} \boldsymbol{d}(\theta) \equiv 0,\ mod\ 2\pi c \tag{15}$$

It is clear that it is sufficient to define the location of virtual sensors by:

$$\tilde{\boldsymbol{p}}_\ell(\omega_q) = \frac{\omega_0}{\omega_q}\boldsymbol{p}_\ell, \quad \tilde{\boldsymbol{\delta}}_m(\omega_q) = \frac{\omega_0}{\omega_q}\boldsymbol{\delta}_m \tag{16}$$

to satisfy both (5) and (15). With these virtual arrays, one can associate time delays $\tilde{\tau}_\ell(\omega_q,\theta) = \tilde{\boldsymbol{p}}_\ell^{\mathsf{T}}(\omega_q)\boldsymbol{d}(\theta)$ and $\tilde{\zeta}_m(\omega_q,\theta) = \tilde{\boldsymbol{\delta}}_m^{\mathsf{T}}(\omega_q)\boldsymbol{d}(\theta)$, and corresponding steering vectors $\tilde{\boldsymbol{a}}(\omega_q,\theta) = \exp\{-\jmath\omega_q\tilde{\boldsymbol{\tau}}(\omega_q,\theta)\}$ and $\tilde{\boldsymbol{b}}(\omega_q,\theta) = \exp\{-\jmath\omega_q\tilde{\boldsymbol{\zeta}}(\omega_q,\theta)\}$. Once virtual arrays are defined, it is necessary to compute the virtual data these array would receive, and this is done by interpolation as explained in the two next sections.

B. Bilinear transform

Space is first partitioned in K angular sectors of limited extent, Θ_k, $1 \le k \le K$. Then each sector is discretized in I angular values: $\Theta_k = \{\theta_{k1},\ldots,\theta_{kI}\}$. In each sector Θ_k and for every discretized frequency ω_q, we define matrices $\mathbb{A}$ and $\mathbb{B}$ of size $L\times I$ and $M\times I$, respectively:

$$\begin{aligned}\mathbb{A}(\omega_q,\Theta_k) &= [\boldsymbol{a}(\omega_q,\theta_{k1}),\ldots,\boldsymbol{a}(\omega_q,\theta_{kI})]\\ \mathbb{B}(\omega_q,\Theta_k) &= [\boldsymbol{b}(\omega_q,\theta_{k1}),\ldots,\boldsymbol{b}(\omega_q,\theta_{kI})]\end{aligned}$$

We also build $\tilde{\mathbb{A}}$ and $\tilde{\mathbb{B}}$ in the same manner with vectors $\tilde{\boldsymbol{a}}(\omega_q,\theta_{ki})$ and $\tilde{\boldsymbol{b}}(\omega_q,\theta_{ki})$ respectively. For a fixed reference frequency ω_0 (generally chosen to be inside of the band [20]), two focusing matrix families can then be defined by solving for square matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, of respective dimensions $L\times L$ and $M\times M$, the linear systems below in the Least Squares (LS) sense:

$$\boldsymbol{U}(\omega_q,\omega_0,\Theta_k)\,\mathbb{A}(\omega_q,\Theta_k) \approx \tilde{\mathbb{A}}(\omega_0,\Theta_k) \tag{17}$$

$$\boldsymbol{V}(\omega_q,\omega_0,\Theta_k)\,\mathbb{B}(\omega_q,\Theta_k) \approx \tilde{\mathbb{B}}(\omega_0,\Theta_k) \tag{18}$$

for every angular sector Θ_k and every frequency ω_q. In what follows we only consider one angular sector Θ, and we hence refer to $\boldsymbol{U}(\omega_q,\omega_0,\Theta)$ and $\boldsymbol{V}(\omega_q,\omega_0,\Theta)$ with $\boldsymbol{U}_q$ and $\boldsymbol{V}_q$ respectively. As for the detection of the most relevant sectors Θ_k, the angular field of view of the array can be scanned and sectors of largest power can be detected through a simple optimum beamforming technique [16].

C. Bilinear interpolation

Measurements recorded on actual arrays with sensors located at $\boldsymbol{p}_\ell + \boldsymbol{\delta}_m$ at frequency ω_q are interpolated to yield virtual measurements recorded by sensors located at $\tilde{\boldsymbol{p}}_\ell(\omega_q) + \tilde{\boldsymbol{\delta}}_m(\omega_q)$ as defined in (16). More precisely, applying the bilinear interpolation (17-18) to model (11) leads to

$$\tilde{\mathcal{M}}_{\ell mq} \approx \sum_{r=1}^{R} \tilde{A}_{\ell r}(\omega_0)\,\tilde{B}_{mr}(\omega_0)\,S_{qr} \tag{19}$$

where the slice $\tilde{\boldsymbol{M}}_q$ of tensor $\tilde{\boldsymbol{\mathcal{M}}}$ is obtained through the bilinear transformation of slice $\boldsymbol{M}_q$: $\tilde{\boldsymbol{M}}_q = \boldsymbol{U}_q\boldsymbol{M}_q\boldsymbol{V}_q^{\mathsf{T}}$. Now, $\tilde{\boldsymbol{\mathcal{M}}}$ may be seen as a tensor of order 3, since frequency ω_q affects only the third mode, and Equation (19) is actually a CP decomposition model.

In the remainder of this section, only space, space-shift and polarization will be addressed (this means that sensors are assumed omnidirectional: $g_l(\theta)=1,\forall l,\forall\theta$, as stated in III-A).

However, a new difficulty appears, due to the fact that the bilinear transformation affects the color of the noise. This must be taken into account into the CP identification algorithm (cf. Section VII) and the computation of the performance bounds (cf. Section VI).

D. Noise correlation induced by interpolation

The observation model (19) can be expressed through Q matrix slices, $1\le q\le Q$:

$$\tilde{\boldsymbol{X}}_q = \tilde{\boldsymbol{M}}_q + \tilde{\boldsymbol{N}}_q$$

where $\tilde{\boldsymbol{N}}_q$ is colored by the transform:

$$\tilde{\boldsymbol{N}}_q = \boldsymbol{U}_q\boldsymbol{N}_q\boldsymbol{V}_q^{\mathsf{T}} \tag{20}$$

Yet from [32], we have $\operatorname{vec}\{\boldsymbol{U}\boldsymbol{N}\boldsymbol{V}^{\mathsf{T}}\} = (\boldsymbol{V}\boxtimes\boldsymbol{U})\operatorname{vec}\boldsymbol{N}$. This leads to the following noise vectorization:

$$\operatorname{vec}\tilde{\boldsymbol{N}}_q = \left(\boldsymbol{V}_q\boxtimes\boldsymbol{U}_q\right)\operatorname{vec}\boldsymbol{N}_q$$

Hence, for every frequency ω_q, the noise vector $\operatorname{vec}\tilde{\boldsymbol{N}}_q$ is circular complex Gaussian with covariance $\boldsymbol{\Sigma}_q = \sigma^2\boldsymbol{V}_q\boldsymbol{V}_q^{\mathsf{H}}\boxtimes\boldsymbol{U}_q\boldsymbol{U}_q^{\mathsf{H}}$. If we stack all vectors $\operatorname{vec}\tilde{\boldsymbol{N}}_q$ in a single vector $\tilde{\boldsymbol{n}}$, then the latter has a covariance $\boldsymbol{\Sigma}_{(1)}$ that is block-diagonal[6] with diagonal blocks $\boldsymbol{\Sigma}_q$:

$$\boldsymbol{\Sigma}_{(1)} = \begin{bmatrix}\boldsymbol{\Sigma}_1 & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \ddots & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Sigma}_Q\end{bmatrix} \tag{21}$$

This change in covariance structure will be taken into account for calculating the Fisher Information matrix (FIM) in Section VI, and into the CP identification algorithm in Section VII.

E. Interpolation in the presence of polarization

Instead of space shift, if we consider polarization as the second diversity as in (12), the bilinear transformation becomes $\tilde{\boldsymbol{M}}_q = \boldsymbol{U}_q\boldsymbol{M}_q\boldsymbol{I}_P^{\mathsf{T}}$, as polarization matrix $\boldsymbol{K}$ does not depend on frequency:

$$\tilde{\mathcal{M}}_{\ell pq} \approx \sum_{r=1}^{R} \tilde{A}_{\ell r}(\omega_0)\,K_{pr}\,S_{qr} \tag{22}$$

Moreover, the diagonal blocks of vectorized noise covariance are simplified to $\boldsymbol{\Sigma}_q = \sigma^2\,\boldsymbol{I}_P\boxtimes\boldsymbol{U}_q\boldsymbol{U}_q^{\mathsf{H}}$.

If we now include all the diversities as in model (14), then interpolation is given by the following multi-linear transformation:

$$\tilde{\boldsymbol{\mathcal{M}}}_q = \boldsymbol{\mathcal{M}}_q \bullet_1 \boldsymbol{U}_q \bullet_2 \boldsymbol{V}_q \bullet_3 \boldsymbol{I}_P$$

where index q may be seen as denoting the q-th slice of a fourth order tensor. The mode-d product indicated as $\bullet_d$ refers to the multiplication of a tensor and a matrix along its d-th mode (see [33], [34] for details on this contraction product)[7]. This yields to interpolated elements

$$\tilde{\mathcal{M}}_{\ell mpq} \approx \sum_{r=1}^{R} \tilde{A}_{\ell r}(\omega_0)\,\tilde{B}_{mr}(\omega_0)\,K_{pr}\,S_{qr} \tag{23}$$

Similarly, the noise covariance of the q-th vectorized slice becomes $\boldsymbol{\Sigma}_q = \sigma^2\,\boldsymbol{I}_P\boxtimes\boldsymbol{V}_q\boldsymbol{V}_q^{\mathsf{H}}\boxtimes\boldsymbol{U}_q\boldsymbol{U}_q^{\mathsf{H}}$.

[6]Notice that $\boldsymbol{\Sigma}_{(1)}$ also corresponds to the covariance matrix of the vectorization of the mode-1 tensor unfolding of $\boldsymbol{\mathcal{N}}$, $\operatorname{vec}\{\boldsymbol{N}_{(1)}\}$.

[7]Notice that for matrices $\boldsymbol{X}\bullet_1\boldsymbol{A}\bullet_2\boldsymbol{B} = \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}^{\mathsf{T}}$.

V. Tensor decomposition

A. Canonical Polyadic decomposition

The tensor model allows to express a structured phenomenon through a multidimensional array of order D, by finite sums and products of simpler constituents [7]. A *decomposable* D-way tensor $\boldsymbol{\mathcal{D}}$ can be defined as the outer product of D vectors: $\boldsymbol{\mathcal{D}} = \boldsymbol{a} \otimes \boldsymbol{b} \otimes \ldots \otimes \boldsymbol{s}$, where $\boldsymbol{a} \in \mathbb{C}^L, \boldsymbol{b} \in \mathbb{C}^M, \ldots, \boldsymbol{s} \in \mathbb{C}^Q$. Any order-$D$ tensor admits a decomposition into a sum of decomposable tensors

$$\boldsymbol{\mathcal{M}} = \sum_{r=1}^{R} \lambda_r \, \boldsymbol{\mathcal{D}}_r \tag{24}$$

where coefficients λ_r can always be chosen to be real positive, and decomposable tensors $\boldsymbol{\mathcal{D}}_r$ to have unit norm, *i.e.* the product $\|\boldsymbol{a}\| \, \|\boldsymbol{b}\| \ldots \|\boldsymbol{s}\| = 1$, but other choices are possible (cf. Section V-B). The minimal value of R such that this decomposition holds is called the tensor rank. When R is the rank, decomposition (24) is called Canonical-Polyadic (CP); if R is not too large, this decomposition is unique; see [35], [36] and references therein. This is the main motivation in resorting to tensors.

In practice, we admit that observations $\boldsymbol{\mathcal{X}}$, related to the model $\boldsymbol{\mathcal{M}}$, are corrupted by an additive noise $\boldsymbol{\mathcal{N}}$. In other words, the observation model becomes $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{M}} + \boldsymbol{\mathcal{N}}$, and the goal is to fit (24) with the data. The fact that the set of rank-R tensors is not closed when $R > 1$ may raise difficulties, which are out of the scope of the present paper; we refer to [7], [29] for further details. In what follows tensor rank R, corresponding to the number of sources, is assumed to be known. See [23], [37] for details on model order estimation).

B. CP factors, normalisation and scaling

Consider the case of a tensor of order $D = 4$, which is sufficient for our purposes. Once the bases of the linear spaces involved are fixed, a fourth order tensor $\boldsymbol{\mathcal{M}}$ is defined by its coordinates, denoted $\mathcal{M}_{\ell mpq}$. If its dimensions are $L \times M \times P \times Q$, then the CP decomposition (24) becomes

$$\boldsymbol{\mathcal{M}} = \sum_{r=1}^{R} \lambda_r \, \boldsymbol{a}_r \otimes \boldsymbol{b}_r \otimes \boldsymbol{k}_r \otimes \boldsymbol{s}_r \tag{25}$$

and can be rewritten in terms of array of coordinates:

$$\mathcal{M}_{\ell mpq} = \sum_{r=1}^{R} \lambda_r \, A_{\ell r} \, B_{mr} \, K_{pr} \, S_{qr}$$

where matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{K}$ and $\boldsymbol{S}$, are of dimension $L \times R$, $M \times R$, $P \times R$ and $Q \times R$, respectively. For practical purposes and without restricting the generality, we have preferred to impose the first row of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ to be formed of ones (in fact, as justified in sections II-A and III-A, $\boldsymbol{p}_1 = \boldsymbol{0}$, $\boldsymbol{\delta}_1 = \boldsymbol{0}$ and $g_1(\theta) = 1$), unit norm columns in matrix $\boldsymbol{K}$, and to pull the remaining $R \times R$ scaling matrix, $\mathrm{Diag}\{\lambda_r A_{1r} B_{1r}\}$, into matrix $\boldsymbol{S}$. This choice permits to compute performance bounds on the retrieval of CP factors more easily. To conclude, in the rest of the paper, we shall use a CP decomposition model of the form:

$$\mathcal{X}_{\ell mpq} = \sum_{r=1}^{R} A_{\ell r} \, B_{mr} \, K_{pr} \, S_{qr} + \mathcal{N}_{\ell mpq} \tag{26}$$

with the normalizations mentioned above.

There exist sufficient conditions ensuring uniqueness of the exact CP [6], establishing the maximum number of sources that can be estimated. Less stringent conditions guaranteeing almost surely a unique solution can be found in [7], [38]:

$$R(K + L + M + Q - 3) < KLMQ$$

This holds true when data are not corrupted by noise. In the presence of noise, the existence of a best approximant to the low-rank problem can be inferred through physical quantities named *coherences*, related to the maximum degree of correlation between any two sources along each mode [7].

VI. Cramér-Rao bounds

In order to assess the performance of the proposed approach, we can compare obtained parameter errors with the CRB, *e.g.* for DoA and polarization parameters. The CRB, which is the lower bound on the covariance $\boldsymbol{R}$ of any unbiased estimator, is given by the inverse of the Fisher Information Matrix (FIM); in other words[8], $\boldsymbol{R} \succeq \boldsymbol{F}^{-1}$.

The evaluation of the FIM requires the partial derivatives of the log-likelihood w.r.t. unknown parameters, given the tensor model previously described. In what follows we compute these derivatives and the general expression of the FIM.

For simplicity, we assume from now on that sources and antennas are coplanar, which permits to parameterize DoA's by a single angle, $\theta_r = \phi_r$. Vectors $\boldsymbol{p}_\ell$ and $\boldsymbol{d}(\theta)$ are thus in $\mathbb{R}^2$, and we shall use the notation $\boldsymbol{p}_\ell = [p_{\ell 1}, p_{\ell 2}]^\mathsf{T}$ and $\boldsymbol{d}(\theta) = [\cos\theta, \sin\theta]^\mathsf{T}$. The reasoning would be the same in 3D but with an increased notation complexity.

Even though we only consider omnidirectional and known sensor responses in the remainder of the paper, the derivation of the CRB in this section and in Appendix A-C addresses the more general case of unknown gain patterns. The presence of anisotropic gain patterns is herein treated as a nuisance, but could be exploited as a plain diversity as shown in a companion paper [39]. The CRB for known omnidirectional gain patterns can be deduced as a particular case for constant $g_\ell(\theta) = 1, \forall\theta, \forall\ell$.

A. Likelihood

First, the multi-dimensional $L \times M \times P \times Q$ array $\boldsymbol{\mathcal{X}}$ is stored in a vector $\boldsymbol{x}$ of dimension $LMPQ$, following a bijective map: $\boldsymbol{x} = \mathrm{vec}\, \boldsymbol{\mathcal{X}}$. We adopt the usual definition [32] where the vec operation takes indices in lexicographical order. Then we have the identity [32], [33]: $\mathrm{vec}\{\boldsymbol{a}\boldsymbol{b}^\mathsf{T}\} = \boldsymbol{b} \boxtimes \boldsymbol{a}$, for any pair of vectors $(\boldsymbol{a}, \boldsymbol{b})$. More generally:

$$\mathrm{vec}\{\boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c}\} \quad = \quad \boldsymbol{c} \boxtimes \boldsymbol{b} \boxtimes \boldsymbol{a}. \tag{27}$$

[8]This inequality means that matrix $\boldsymbol{R} - \boldsymbol{F}^{-1}$ is semi-definite positive.

Thus, applying the vec operator to equation (25) leads to:

$$\boldsymbol{x} = \sum_{r=1}^{R} \boldsymbol{s}_r \boxtimes \boldsymbol{k}_r \boxtimes \boldsymbol{b}_r \boxtimes \boldsymbol{a}_r + \boldsymbol{n} \tag{28}$$

with $\boldsymbol{x} = \operatorname{vec} \boldsymbol{\mathcal{X}}$ and $\boldsymbol{n} = \operatorname{vec} \boldsymbol{\mathcal{N}}$. Now since the noise vector $\boldsymbol{n}$ is circular Gaussian, with zero mean and covariance $\boldsymbol{\Sigma}$, the log-likelihood then takes the form, up to a constant additive term:

$$\Upsilon(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{s}) = -(\boldsymbol{x} - \boldsymbol{\mu})^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \tag{29}$$

where $\boldsymbol{\mu} = \sum_{r=1}^{R} \boldsymbol{s}_r \boxtimes \boldsymbol{k}_r \boxtimes \boldsymbol{b}_r \boxtimes \boldsymbol{a}_r$. We shall assume an isotropic *measurement* noise, but as we explained in Section IV, the noise covariance will be affected by wideband interpolation so that $\boldsymbol{\Sigma} \neq \sigma^2 \boldsymbol{I}$.

B. Fisher Information matrix (FIM)

When computing the FIM with respect to complex parameters, it is also necessary to include the complex conjugate parameter, in order to avoid a loss in information [40]. This means just appending $\partial \boldsymbol{f} / \partial \boldsymbol{z}^*$, which is nothing else but $(\partial \boldsymbol{f} / \partial \boldsymbol{z})^*$. Consequently, for each source, there are three real parameters, θ_r, α_r and β_r, and two complex parameters, $\boldsymbol{s}_r$ and $\boldsymbol{s}_r^*$. In addition to these, the $(L-1)R$ entries $g_{\ell r} = g_\ell(\theta_r)$, $2 \leq \ell \leq L$, are real nuisance parameters. For every r, we store them in a vector $\bar{\boldsymbol{g}}_r$ of size $L-1$ (obtained by just deleting[9] the first entry of $\boldsymbol{g}_r$). In turn, we stack all these vectors $\bar{\boldsymbol{g}}_r$ in a $(L-1)R$-dimensional vector denoted by $\boldsymbol{\gamma}$.

Now real and complex parameters must be treated differently [40]. With this goal, define $\boldsymbol{v} = [\boldsymbol{\theta}^{\mathsf{T}}, \boldsymbol{\alpha}^{\mathsf{T}}, \boldsymbol{\beta}^{\mathsf{T}}, \boldsymbol{\gamma}^{\mathsf{T}}]^{\mathsf{T}}$, a real parameter vector of dimension $(L+2)R$, and $\boldsymbol{\xi} = [\boldsymbol{s}_1^{\mathsf{T}}, \ldots, \boldsymbol{s}_R^{\mathsf{T}}]^{\mathsf{T}}$, a complex parameter vector of size RQ. Then, from (29), the FIM contains nine blocks [41]:

$$\boldsymbol{F} = \frac{1}{\sigma^2} \begin{bmatrix} 2\Re\{\boldsymbol{G}_{11}\} & \boldsymbol{G}_{12} & \boldsymbol{G}_{12}^* \\ \boldsymbol{G}_{12}^{\mathsf{H}} & \boldsymbol{G}_{22} & \boldsymbol{0} \\ \boldsymbol{G}_{12}^{\mathsf{T}} & \boldsymbol{0} & \boldsymbol{G}_{22}^* \end{bmatrix} \tag{30}$$

where $\boldsymbol{G}_{11}$ is related to the real parameters, and $\boldsymbol{G}_{22}$ is related to the complex ones. More precisely, blocks $\boldsymbol{G}_{ij}$ are defined by:

$$\begin{aligned} \boldsymbol{G}_{11} &= \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{v}}\right)^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{v}}\right) \\ \boldsymbol{G}_{22} &= \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\xi}}\right)^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\xi}}\right) \\ \boldsymbol{G}_{12} &= \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{v}}\right)^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\xi}}\right) \end{aligned}$$

where $\boldsymbol{\Sigma}$ is the covariance of noise $\boldsymbol{n}$. The total dimension of the FIM $\boldsymbol{F}$ is hence $R(L+2+2Q)$ if all diversities are present. With these definitions, the covariance matrix $\boldsymbol{R}$ of any unbiased estimator of $[\boldsymbol{v}^{\mathsf{T}}, \boldsymbol{\xi}^{\mathsf{T}}, \boldsymbol{\xi}^{\mathsf{H}}]$ satisfies $\boldsymbol{R} \succeq \boldsymbol{F}^{-1}$.

Note that in $\boldsymbol{G}_{11}$, we obtain the derivative w.r.t. the DoA's using the chain rule below:

$$\frac{\partial \boldsymbol{\mu}}{\partial \theta_r} = \frac{\partial \boldsymbol{\mu}}{\partial \bar{\boldsymbol{a}}_r} \frac{\partial \bar{\boldsymbol{a}}_r}{\partial \theta_r} + \frac{\partial \boldsymbol{\mu}}{\partial \bar{\boldsymbol{b}}_r} \frac{\partial \bar{\boldsymbol{b}}_r}{\partial \theta_r} + \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{k}_r} \frac{\partial \boldsymbol{k}_r}{\partial \theta_r}. \tag{31}$$

[9] As stated in Section II-A, the first entry of $\boldsymbol{g}_r$ is equal to 1, $\forall r$, and is hence not unknown.

Scalar factor 2 in (30) is justified by (39), because chain rules used to calculate $\boldsymbol{G}_{11}$ involve complex variables when computing real derivatives. The derivatives required to evaluate (30) and (31) are given in Appendix A.

VII. Algorithms

The goal of the two remaining sections is to illustrate the performance of tensor wideband processing and to compare it to the CRB. The ALSCOLOR algorithm described in this section is not meant to be a major contribution, even if its originality is to take into account the noise covariance structure. The design of efficient algorithms is indeed out of the scope of the present paper.

Tensor wideband processing will be executed in two stages: we first recover factor matrices through CP model fitting (assuming they are uncoupled), and then estimate the DoA's or polarization parameters from the estimated factor matrices. A refinement could also be executed as a post-processing, *e.g.* by a local ascent maximizing the log-likelihood in (29), but it generally brings negligible improvement. Hence it will not be reported in this paper. We first describe both steps of this procedure, in their respective orders: retrieval of the CP factors taking into account the noise correlation structure, and the estimation of DoA's and polarization parameters.

A. Estimation of factor matrices

CP fitting corresponds to minimizing the tensor approximation error for model (19). Notice that similar considerations can be made for models (22) and (23). Supposing that the interpolation error is negligible compared to the error introduced by measurement noise, we are interested in minimizing the following cost function with respect to $\tilde{\boldsymbol{a}}_r$, $\tilde{\boldsymbol{b}}_r$ and $\boldsymbol{s}_r$:

$$\begin{aligned} \Upsilon &= \frac{1}{\sigma^2} \left\| \operatorname{vec}\{\tilde{\boldsymbol{X}}_{(1)}\} - \sum_{r=1}^{R} \boldsymbol{s}_r \boxtimes \tilde{\boldsymbol{b}}_r \boxtimes \tilde{\boldsymbol{a}}_r \right\|^2_{\boldsymbol{\Sigma}_{(1)}} = \\ &= \frac{1}{\sigma^2} \left\| \operatorname{vec}\{\tilde{\boldsymbol{X}}_{(2)}\} - \sum_{r=1}^{R} \boldsymbol{s}_r \boxtimes \tilde{\boldsymbol{a}}_r \boxtimes \tilde{\boldsymbol{b}}_r \right\|^2_{\boldsymbol{P}_{12} \boldsymbol{\Sigma}_{(1)} \boldsymbol{P}_{12}^T} = \\ &= \frac{1}{\sigma^2} \left\| \operatorname{vec}\{\tilde{\boldsymbol{X}}_{(3)}\} - \sum_{r=1}^{R} \tilde{\boldsymbol{b}}_r \boxtimes \tilde{\boldsymbol{a}}_r \boxtimes \boldsymbol{s}_r \right\|^2_{\boldsymbol{P}_{13} \boldsymbol{\Sigma}_{(1)} \boldsymbol{P}_{13}^T} \end{aligned} \tag{32}$$

where we define the weighted matrix norm $\|\boldsymbol{X}\|^2_{\boldsymbol{\Sigma}} = \operatorname{trace}\{\boldsymbol{X}^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{X}\}$. Expression $\boldsymbol{X}_{(d)}$ refers to the mode-d unfolding of tensor $\boldsymbol{\mathcal{X}}$. Notice that $\boldsymbol{\Sigma}_{(1)}$ is the covariance matrix of $\tilde{\boldsymbol{x}}$, the vectorization of the first unfolding $\tilde{\boldsymbol{X}}_{(1)}$, and permutation matrix $\boldsymbol{P}_{ij}$ denotes the transfer matrix from the covariance of the mode-i unfolding to that of the mode-j unfolding.

The above mentioned minimization corresponds to maximum likelihood estimation and can be performed in several ways: gradient descent, conjugate gradient, Newton methods and coordinate descent methods [42]. For ease of implementation, we choose to minimize (32) using block coordinate descent, with blocks corresponding to factor matrices themselves. This is commonly known as Alternating Least Squares

Algorithm 1. **ALSCOLOR** : ALS for correlated noise

1: Inputs: $\hat{\boldsymbol{A}}^0$, $\hat{\boldsymbol{B}}^0$, $\hat{\boldsymbol{S}}^0$, the covariance matrix $\boldsymbol{\Sigma}_{(1)}$ given by (21), the interpolated data tensor $\tilde{\boldsymbol{\mathcal{X}}}$.

2: Initialize with $\hat{\boldsymbol{A}}^0$, $\hat{\boldsymbol{B}}^0$, $\hat{\boldsymbol{S}}^0$. Denote for all k: $\hat{\boldsymbol{a}}^k = \mathrm{vec}\{\hat{\boldsymbol{A}}^k\}$, $\hat{\boldsymbol{b}}^k = \mathrm{vec}\{\hat{\boldsymbol{B}}^k\}$, $\hat{\boldsymbol{s}}^k = \mathrm{vec}\{\hat{\boldsymbol{S}}^k\}$.

3: While convergence criterion not met, do $k := k+1$

$$\begin{cases} \boldsymbol{Z} = (\hat{\boldsymbol{S}}^k \odot \hat{\boldsymbol{B}}^k) \boxtimes \boldsymbol{I}_L \\ \hat{\boldsymbol{a}}^k = \left(\boldsymbol{Z}^{\mathsf{H}} \boldsymbol{\Sigma}_{(1)}^{-1} \boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^{\mathsf{H}} \boldsymbol{\Sigma}_{(1)}^{-1} \mathrm{vec}\{\tilde{\boldsymbol{X}}_{(1)}\} \end{cases} \tag{33}$$

Proceed similarly for $\hat{\boldsymbol{b}}^k$ and $\hat{\boldsymbol{s}}^k$, with $\boldsymbol{\Sigma}_{(2)} = \boldsymbol{P}_{12}\boldsymbol{\Sigma}_{(1)}\boldsymbol{P}_{12}^T$ and $\boldsymbol{\Sigma}_{(3)} = \boldsymbol{P}_{13}\boldsymbol{\Sigma}_{(1)}\boldsymbol{P}_{13}^T$, respectively.

4: Unvectorize estimates: $\hat{\boldsymbol{A}} = \mathrm{unvec}\{\hat{\boldsymbol{a}}^k\}$, $\hat{\boldsymbol{B}} = \mathrm{unvec}\{\hat{\boldsymbol{b}}^k\}$ and $\hat{\boldsymbol{S}} = \mathrm{unvec}\{\hat{\boldsymbol{s}}^k\}$; normalize columns of $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{B}}$ by their first element:

$$\begin{cases} \hat{\boldsymbol{a}}_r := \hat{\boldsymbol{a}}_r / \hat{A}_{1r} \\ \hat{\boldsymbol{b}}_r := \hat{\boldsymbol{b}}_r / \hat{B}_{1r} \end{cases}$$

and normalize $\hat{\boldsymbol{S}}$ accordingly: $\hat{\boldsymbol{s}}_r := \hat{A}_{1r}\hat{B}_{1r}\,\hat{\boldsymbol{s}}_r$.

NB: When dealing with model (22), where the third diversity is given by polarization instead of space-shift, $\boldsymbol{B}$ is replaced by $\boldsymbol{K}$ is the steps above, and normalization is achieved as:

$$\begin{cases} \hat{\boldsymbol{a}}_r := \hat{\boldsymbol{a}}_r / \hat{A}_{1r} \\ \hat{\boldsymbol{k}}_r := \hat{\boldsymbol{k}}_r / \parallel \hat{\boldsymbol{k}}_r \parallel \\ \hat{\boldsymbol{s}}_r := \hat{A}_{1r} \parallel \hat{\boldsymbol{k}}_r \parallel \hat{\boldsymbol{s}}_r \end{cases}$$

5: End

(ALS) [42], as each block coordinate update corresponds to the solution of a least squares problem. The proposed algorithm, which we name *ALSCOLOR*, is detailed in Alg. 1.

B. Estimation of parameters

Once column vectors $\hat{\boldsymbol{a}}_r$, $\hat{\boldsymbol{b}}_r$ and/or $\hat{\boldsymbol{k}}_r$ are recovered from the tensor approximation algorithm, we can estimate DoA θ_r through an exhaustive search within the considered angular sector Θ and/or polarization parameters β_r and/or α_r.

1) Space shift diversity: if there are space, space shift and frequency diversities (cf VIII-A), DoA's are estimated by minimizing a weighted least squares criterion of the form:

$$\hat{\theta}_r^{\boldsymbol{D}} = \arg\min_{\theta \in \Theta_i} \left\| \begin{pmatrix} \hat{\boldsymbol{a}}_r - \tilde{\boldsymbol{a}}_r(\theta) \\ \hat{\boldsymbol{b}}_r - \tilde{\boldsymbol{b}}_r(\theta) \end{pmatrix} \right\|_{\boldsymbol{D}}^2 \tag{34}$$

where $\boldsymbol{D}$ is a positive definite matrix and $\|\boldsymbol{u}\|_{\boldsymbol{D}}^2 = \boldsymbol{u}^{\mathsf{H}}\boldsymbol{D}^{-1}\boldsymbol{u}$. We have experimented several choices for $\boldsymbol{D}^{-1}$: (i) the Fisher information matrix[10] $\boldsymbol{F}_r$ related to vector $[\boldsymbol{a}_r^{\mathsf{T}}, \boldsymbol{b}_r^{\mathsf{T}}]^{\mathsf{T}}$ computed at $[\hat{\boldsymbol{a}}_r, \hat{\boldsymbol{b}}_r, \hat{\boldsymbol{s}}_r]$, (ii) its approximation by its two main diagonal blocks, (iii) $\boldsymbol{D} = \boldsymbol{I}$. The two diagonal blocks are not diagonal, so that the parameter estimation $\hat{\theta}_r^{\boldsymbol{I}}$ assuming that all components of $\hat{\boldsymbol{a}}_r$ and $\hat{\boldsymbol{b}}_r$ have the same variance is less accurate. On the other hand, the correlation between $\hat{\boldsymbol{a}}_r$ and $\hat{\boldsymbol{b}}_r$ has revealed to be small, so that estimates obtained with (i) and (ii) were very similar, and will be both denoted by $\hat{\theta}_r^{\boldsymbol{F}}$.

This objective is arbitrary and suboptimal, but permits to find an acceptable solution, as will be demonstrated in Section VIII. In fact, a refinement by gradient ascent does not bring any significant improvement.

2) Polarization diversity: If there are space, polarization and frequency diversities (cf VIII-B), we can also estimate polarization parameters. For one single polarization parameter β_r, we perform an exhaustive search over a 2D grid of DoA $\theta \in \Theta$ and polarization angles $\beta \in [-\pi/4, \pi/4]$. We proceed analogously, by searching $\Theta \cup [-\pi/4,\, \pi/4]$ for:

$$(\hat{\theta}_r^{\boldsymbol{D}}, \hat{\beta}_r^{\boldsymbol{D}}) = \arg\min_{\theta,\beta} \left\| \begin{pmatrix} \hat{\boldsymbol{a}}_r - \tilde{\boldsymbol{a}}_r(\theta) \\ \hat{\boldsymbol{k}}_r - \tilde{\boldsymbol{k}}_r(\theta,\beta) \end{pmatrix} \right\|_{\boldsymbol{D}}^2 \tag{35}$$

The same choices of matrix $\boldsymbol{D}$ have been experimented (see Section VIII).

Remark: the above mentioned minimization procedures are performed in a limited domain: $\theta \in \Theta, \beta \in [-\pi/4, \pi/4]$. In order to prevent this search from becoming too costly, we apply an iterative progressive refinement of the grid, where each iteration is a zoomed version of the previous one.

VIII. Computer results

This Section is meant to illustrate the advantage of wideband tensor processing for the problem of DoA and/or polarization estimation.

Monte Carlo simulations make use of multiple diversities, throughout the section (in addition to space and frequency):

1) VIII-A: space shifts of a reference subarray, as described in Subsection III-A.
2) VIII-B: polarization diversity, as described in Subsection III-B.

Impinging sources are coplanar with respect to the array: only azimuth is to be estimated, and not elevation, without any loss of generality. We simulated $R = 2$ sources, at sampling frequency $f_s = 1\,\mathrm{kHz}$, for an observation time of $64\,\mathrm{ms}$ and propagating at $c = 1800\,\mathrm{m/s}$, which is the propagation speed of seismic S-waves through ice [15]. Hence, $T = 64$ time samples are available. Sources in time domain are white Gaussian processes with unit variance. The number of working frequency bins is $Q = 14$ (within the selected frequency range from $f_{min} = 0.25\,f_s$ to $f_{max} = 0.45\,f_s$). As for the distance d_{ij} between two neighboring sensors ($\ell = i$ and $\ell = j$) of the same subarray, we set $d_{ij} \leq \lambda_0/2$ where λ_0 is the reference wavelength, $\lambda_0 = 2\pi c/\omega_0$. The reference frequency $f_0 = \omega_0/2\pi$ is chosen to be the central bin $f_0 = (f_{min}+f_{max})/2 = 0.35 f_s$. Correlation between sources, when present, is $\rho = 0.9$. Wideband processing is performed in a single sector Θ, assuming interesting sectors (*i.e.* sectors containing most of the signal power) have been previously detected through a former low resolution beamforming technique. Former detection of relevant sectors is beyond the scope of this paper and can be found in the literature [1], [2]. Interpolation is then performed within the angular sector $\Theta = [0, 25^\circ]$, with a discretization of 0.1°.

In order to understand the relevance of tensor wideband processing, we also present a narrowband alternative for

[10] See [29], [28] for further details on the computation of CRB for factor matrices.

comparison (referred to as *NB-ALS*): received signals are filtered around central frequency $f_0 = \omega_0/2\pi$ with narrowband filter bandwidth $BW = 0.015 f_s$, and then directly processed through tensor CP approximation [6], [29]. Estimation efficiency is also evaluated in comparison to the wideband MUSIC approach [17], and to the wideband multilinear CRB computed in Section VI. The performance criterion is the relative total *Mean Square Error* (MSE) of the DoA estimates and, whenever present, of the polarization ellipticity estimates:

$$\begin{cases} \text{MSE}(\boldsymbol{\theta}) = \frac{1}{\pi^2}\frac{1}{NR}\sum_{n=1}^{N}\sum_{r=1}^{R}(\hat{\theta}_{rn} - \theta_r)^2 \\ \text{MSE}(\boldsymbol{\beta}) = \frac{1}{\pi^2}\frac{1}{NR}\sum_{n=1}^{N}\sum_{r=1}^{R}(\hat{\beta}_{rn} - \beta_r)^2 \end{cases} \tag{36}$$

where $\hat{\theta}_{rn}$ and $\hat{\beta}_{rn}$ are the estimated DoA and ellipticity angle respectively of source r in the n-th Monte-Carlo trial, $N = 2500$ being the number of trials. As in [29], [28], the SNR definition below is assumed:

$$SNR = 10\log_{10}\frac{\mathbb{E}\left[\boldsymbol{\mu}^{\mathsf{H}}\boldsymbol{\mu}\right]}{\mathbb{E}\left[\boldsymbol{n}^{\mathsf{H}}\boldsymbol{n}\right]} = 10\log_{10}\frac{\parallel \boldsymbol{\mu} \parallel^2}{LMPQ\sigma^2} \tag{37}$$

Estimation algorithms evaluated herein are the following:

1) *NB-ALS*: standard narrowband ALS in time domain as in [6], [29], after narrowband filtering around f_0, and simple parameter estimation: parameters are estimated by taking the identity matrix for $\boldsymbol{D}$ in (34) and (35).
2) *WB-ALS*: WB preprocessing and standard ALS in frequency domain not taking into account the correlation structure of noise, and simple DoA estimation taking $\boldsymbol{D} = \boldsymbol{I}$ in (34) and (35).
3) *WB-ALSCOLOR*: WB preprocessing and ALS Alg. 1 for correlated noise, and simple DoA estimation taking $\boldsymbol{D} = \boldsymbol{I}$ in (34) and (35).
4) *WB-W-ALSCOLOR*: WB preprocessing and ALS Alg. 1 for correlated noise and weighted DoA estimation: parameters are estimated through (34) and (35) with FIM weighting $\boldsymbol{D} = \boldsymbol{F}_r$.
5) *WB-MUSIC* with covariance matrix expressed in (9) and, for polarized sensors, *WB vector MUSIC*, where the array manifold takes into account wave polarization parameters.

Notice that *WB preprocessing* refers to the multilinear interpolation described in Section IV.

A. Wideband and space shift

As described in Section III-A, the same sub-array is repeated in space through M translations (cf. Fig. 1):

1) $M = 5$ sub-arrays;
2) each sub-array is a square 2D array of $L = 4$ sensors;
3) the distance between two neighboring sensors is $\lambda_0/2$;
4) the distance between two neighboring sub-arrays is $\approx 4 * \lambda_0/2$.

Interpolation of the data tensor is bilinear, which means it is performed for matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ separately, as described in subsection IV-B. Therefore, both sensor positions in $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_L]^{\mathsf{T}}$ and space shifts in $\boldsymbol{\Delta} = [\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_M]^{\mathsf{T}}$ need to be known, and DoA can be estimated both from $\boldsymbol{A}$ and $\boldsymbol{B}$. On the other hand, the linear interpolation of wideband MUSIC is performed on the global $LM \times 2$ matrix $\boldsymbol{P}_{tot} = [\boldsymbol{p}_1 + \boldsymbol{\delta}_1, \ldots, \boldsymbol{p}_L + \boldsymbol{\delta}_1, \ldots, \boldsymbol{p}_1 + \boldsymbol{\delta}_M, \ldots, \boldsymbol{p}_L + \boldsymbol{\delta}_M]^{\mathsf{T}}$ of the whole acquisition system (cf. Fig. 1), without taking into account the bilinearity of the tensor model.

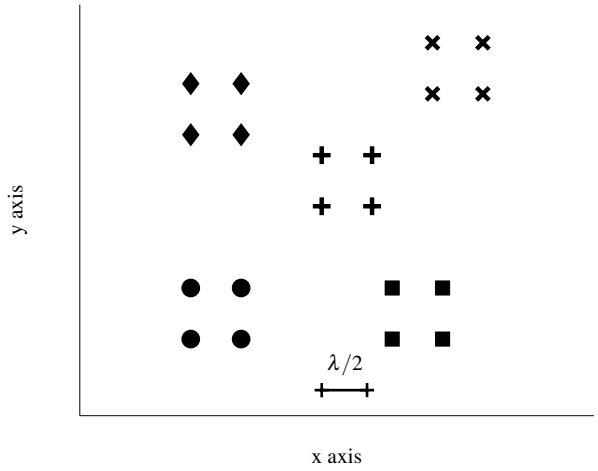


Fig. 1. Sensor position matrix - space shift diversity

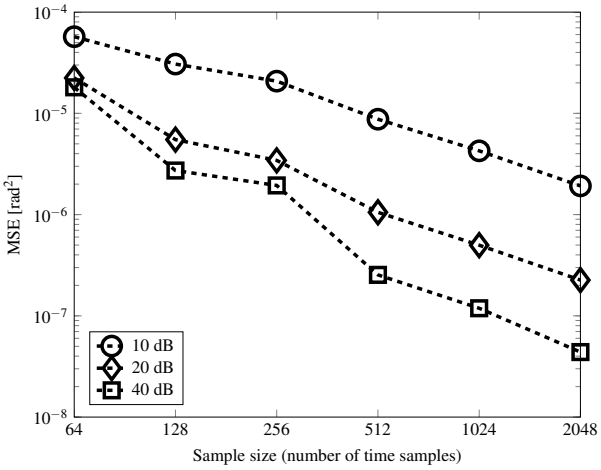
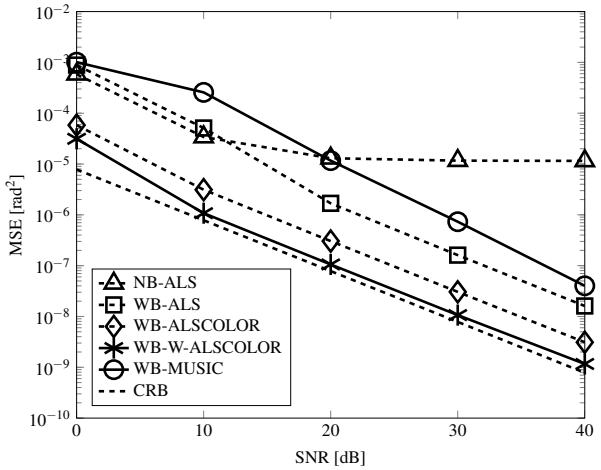


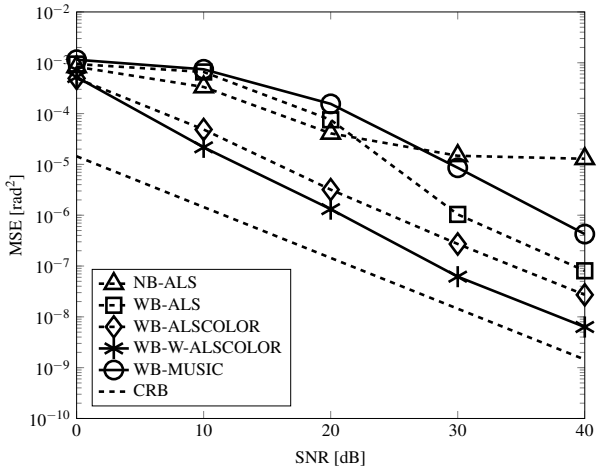
Fig. 2. MSE vs T - NB ALS - Narrowband Tensor CP

Fig. 2 shows the dependence of the narrowband tensor CP method on the number of available time samples, T: the MSE drastically drops for increasing data samples, because a larger data window allows a narrowband filtering of higher quality. In what follows, we fix the number of time samples to $T = 64$: we expect then that narrowband tensor CP will show a saturation starting from a certain SNR.

Fig. 3a shows the MSE on DoA estimation with respect to the SNR when $R = 2$ uncorrelated sources arrive from angles $\theta_1 = 5°$ and $\theta_2 = 20°$ respectively; Fig. 3b refers to a correlation coefficient of $\rho = 0.9$ between sources. ALSCOLOR (*i.e.* Alg. 1) significantly improves the standard ALS, since it takes into account the noise correlation structure introduced by interpolation. The reliability information contained in the FIM weighting $\boldsymbol{D} = \boldsymbol{F}_r$ on factor vectors in (34) also helps the estimation performance. In the present experimental conditions, wideband tensor methods outperform both MUSIC and narrowband CP. If the sources are correlated, the performance of all the algorithms slightly deteriorates correspondingly, and and the gap w.r.t. the CRB increases (cf. Fig. 3b).



(a) DoA estimation - Uncorrelated sources



(b) DoA estimation - Correlated sources: $\rho = 0.9$

Fig. 3. MSE vs SNR - space shift diversity

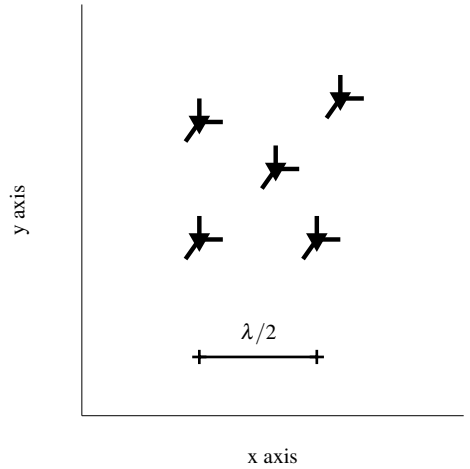
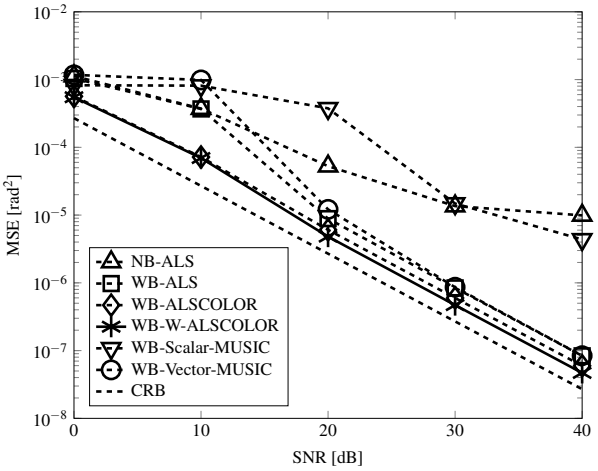
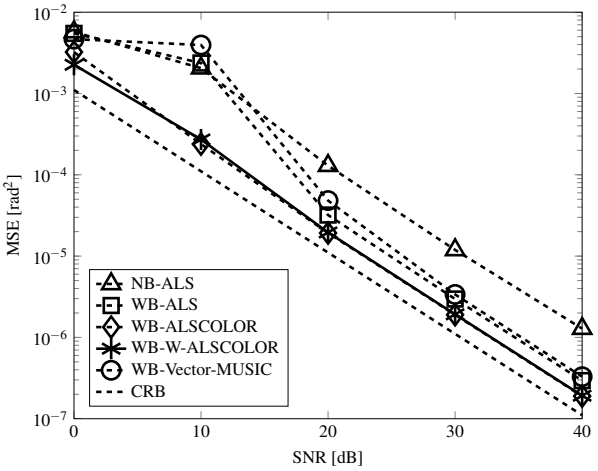


Fig. 4. 3C sensor position matrix (polarization diversity)



(a) DoA estimation



(b) Polarization estimation

Fig. 5. MSE vs SNR - polarization diversity - seismic Rayleigh waves

B. Wideband and polarization

Impinging signals are supposed to be elliptically polarized seismic sources, recorded by one three-component (3C) sensor array, or *vector-sensor array* as defined in [9] (cf. Fig. 4). We assume:

1) an array of $L = 5$ sensors;
2) the distance between two sensors is $\approx \lambda/2$;
3) $R = 2$ uncorrelated polarized sources arrive from angles $\theta_1 = 5°$ and $\theta_2 = 20°$ respectively.
4) Rayleigh waves (coplanar with the array), $\alpha = 0$, $\psi = 0$, $\beta \neq 0$. In particular, $\beta_1 = -20°, \beta_2 = 10°$.

Interpolation is linear, *i.e.* only for matrix $\boldsymbol{A}$, since polarization factor matrix $\boldsymbol{K}$ is independent from frequency, as described in subsection IV-E. We compare tensor methods with wideband *scalar MUSIC* (an incoherent average of P covariances), and wideband *vector MUSIC*, through a coherently averaged covariance matrix of size $PL \times PL$. Fig. 5a and Fig. 5b show the MSE with respect to the SNR for DoA estimation and for polarization estimation, respectively. The overall performance of the compared algorithms is similar to the space-shift case of Section VIII-A. However, the gain granted by FIM weighting $\boldsymbol{D} = \boldsymbol{F}_r$ is less significant, especially for β.

IX. Conclusion

The crucial difficulty of the tensor wideband approach stems from the need to focus every frequency contribution onto the same subspace. We show that multi-linear interpolation can

solve this issue: it allows to have a multi-linear model and thus to apply tensor decomposition techniques to wideband data. An important side effect is that the latter interpolation correlates the noise, which needs to be taken into account in the optimization process. In simulations, both with space shift diversity and polarization diversity, the proposed approach improves narrowband tensor processing and outperforms traditional subspace methods such as wideband MUSIC. We also show that, in addition to space and frequency, polarization can be used instead of space shift diversity, even in the wideband case.

APPENDIX A
DERIVATIVES REQUIRED FOR THE FIM

In order to evaluate the derivatives required for the computation of the FIM, and consequently for the CRB, we have to introduce the concept of complex derivatives.

A. Complex derivatives

A real function of a complex variable is never holomorphic, unless it is constant [43]. It is hence necessary to specify which convention is assumed. Even if a convention has been proposed much earlier in [43], we assume the definition proposed in [40] since it is now more widely used. Let $\boldsymbol{f}(\boldsymbol{z})$ be a real function from $\mathbb{C}^P$ to $\mathbb{R}^N$. Its complex derivative with respect to $\boldsymbol{z} \in \mathbb{C}^P$ is defined by:

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}} = \frac{1}{2}\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} - \frac{\jmath}{2}\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}} \tag{38}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ denote the real and imaginary parts of $\boldsymbol{z}$, respectively. This definition is consistent with [40], [29], [28]. This derivative is stored in a complex $N \times P$ matrix, regardless of the fact that $\boldsymbol{f}$ and $\boldsymbol{z}$ are row or column vectors, which allows easy-writing chain rules of the form: $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{a}}\frac{\partial \boldsymbol{a}}{\partial \boldsymbol{z}}$ with compatible matrix sizes.

Next, it is often convenient to compute some derivatives by the chain rule, which means that derivatives of a complex quantity w.r.t. a complex variable are needed. It turns out that quantities involved are holomorphic, so that this differential is meant in the usual sense of complex analysis.

If a real function $\boldsymbol{f}$ is to be derived w.r.t. a real variable $\boldsymbol{\theta}$, but using a chain rule involving a complex variable $\boldsymbol{z}$, then for consistency with (38) the following relation must apply [40], [29], [14]:

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\theta}} = 2\Re\left\{\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}}\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{\theta}}\right\} \tag{39}$$

In what follows, we calculate the derivatives required for the evaluation of the FIM in (30).

B. Derivatives w.r.t. $\boldsymbol{S}$

Matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ depend on angles of arrival θ_r, whereas $\boldsymbol{K}$ depends on angles[11] $(\theta_r, \alpha_r, \beta_r)$, $1 \le r \le R$. Let's start with matrix $\boldsymbol{S}$, which is unconstrained. The source signals are indeed treated as deterministic parameters, which allows to extract them. The derivative w.r.t $\boldsymbol{s}_r$ is [29]:

$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{s}_r} = \boldsymbol{I}_Q \boxtimes \boldsymbol{k}_r \boxtimes \boldsymbol{b}_r \boxtimes \boldsymbol{a}_r \in \mathbb{C}^{LMPQ \times Q} \tag{40}$$

C. Derivatives w.r.t. $\boldsymbol{A}$ and $\boldsymbol{B}$, and their parameters

Since the first row of $\boldsymbol{A}$ and $\boldsymbol{B}$ contains only ones, only the remaining rows need to be estimated. The submatrices formed of the remaining rows are denoted by $\overline{\boldsymbol{A}}$ and $\overline{\boldsymbol{B}}$, and their columns by $\bar{\boldsymbol{a}}_r$ and $\bar{\boldsymbol{b}}_r$, respectively. In other words, $\overline{\boldsymbol{A}} = \boldsymbol{J}_L\boldsymbol{A}$ where $\boldsymbol{J}_L$ is the L–$1 \times L$ line-selection matrix $\boldsymbol{J}_L = [\boldsymbol{0},\ \boldsymbol{I}_{L-1}]$. This is necessary to reduce the size of the FIM (otherwise it would be always rank deficient). We have:

$$\frac{\partial \boldsymbol{\mu}}{\partial \bar{\boldsymbol{a}}_r} = \boldsymbol{s}_r \boxtimes \boldsymbol{k}_r \boxtimes \boldsymbol{b}_r \boxtimes \boldsymbol{J}_L^{\mathsf{T}} \in \mathbb{C}^{LMPQ \times L-1}$$

Now, since functions $g_\ell(\theta)$ are unknown, but needed only at R points θ_r, $g_{\ell r} = g_\ell(\theta_r)$ are treated as extraneous fixed nuisance (unknown) parameters; this approach avoids the need for derivatives $\partial g_\ell(\theta)/\partial\theta$. For convenience, we denote by $\boldsymbol{g}_r$ the L-dimensional vector with entries $g_\ell(\theta_r)$, $1 \le \ell \le L$, and $\bar{\boldsymbol{g}}_r$ the vector obtained by removing the first entry of $\boldsymbol{g}_r$ (which is equal to 1 by definition). The case of known omnidirectional gain patterns can be deduced as a particular case for constant $g_\ell(\theta) = 1, \forall\theta, \forall\ell$. The following derivatives can be readily obtained:

$$\begin{aligned}
\frac{\partial \bar{\boldsymbol{a}}_r}{\partial \theta_r} &= -\frac{\jmath\omega}{c}\boldsymbol{J}_L\left(\boldsymbol{g}_r \odot \boldsymbol{P}\dot{\boldsymbol{d}}(\theta_r) \odot \exp\{-\jmath\omega\boldsymbol{\tau}_r\}\right) \in \mathbb{C}^{L-1} \\
\frac{\partial \bar{\boldsymbol{a}}_r}{\partial \bar{\boldsymbol{g}}_r} &= \boldsymbol{J}_L \exp\{-\jmath\omega\boldsymbol{\tau}(\theta_r)\}\boldsymbol{J}_L^{\mathsf{T}} \in \mathbb{C}^{L-1 \times L-1}
\end{aligned}$$

where the exponential is taken entry-wise; $\boldsymbol{\tau}_r \in \mathbb{C}^L$ is the vector containing entries $\tau_\ell(\theta_r)$, and $\boldsymbol{P}$ is the $L \times 2$ matrix with rows $\boldsymbol{p}_\ell$.

Similarly, we have:

$$\frac{\partial \boldsymbol{\mu}}{\partial \bar{\boldsymbol{b}}_r} = \boldsymbol{s}_r \boxtimes \boldsymbol{k}_r \boxtimes \boldsymbol{J}_M^{\mathsf{T}} \boxtimes \boldsymbol{a}_r \in \mathbb{C}^{LMPQ \times M-1}$$

The following derivative is readily obtained:

$$\frac{\partial \bar{\boldsymbol{b}}_r}{\partial \theta_r} = -\frac{\jmath\omega}{c}\boldsymbol{J}_M\left(\boldsymbol{\Delta}\dot{\boldsymbol{d}}(\theta_r) \odot \exp\{-\jmath\omega\boldsymbol{\zeta}_r\}\right) \in \mathbb{C}^{M-1}$$

where $\boldsymbol{\zeta}_r \in \mathbb{C}^M$ is the vector containing entries $\zeta_m(\theta_r)$, and $\boldsymbol{\Delta}$ is the $M \times 2$ matrix with rows $\boldsymbol{\delta}_m$.

The case of pattern diversity: if time, space shift and gain pattern diversities are available, but not space and polarization, then expressions are simpler. It is worth detailing this scenario because it has not been reported in the literature. In this case, matrix $\boldsymbol{A}$ contains only gains $g_{\ell r}$ and matrix $\boldsymbol{B}$ only delays, and one can prove that:

$$\begin{aligned}
\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{s}_r}\right)^{\mathsf{H}}\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{s}_k}\right) &= (\boldsymbol{b}_r^{\mathsf{H}}\boldsymbol{b}_k)(\boldsymbol{a}_r^{\mathsf{H}}\boldsymbol{a}_k)\,\boldsymbol{I}_Q \\
\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{a}_r}\right)^{\mathsf{H}}\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{a}_k}\right) &= (\boldsymbol{s}_r^{\mathsf{H}}\boldsymbol{s}_k)(\boldsymbol{b}_r^{\mathsf{H}}\boldsymbol{b}_k)\,\boldsymbol{I}_L \\
\left(\frac{\partial \boldsymbol{\mu}}{\partial \theta_r}\right)^{\mathsf{H}}\left(\frac{\partial \boldsymbol{\mu}}{\partial \theta_k}\right) &= \omega^2(\boldsymbol{s}_r^{\mathsf{H}}\boldsymbol{s}_k)(\boldsymbol{a}_r^{\mathsf{H}}\boldsymbol{a}_k)(\boldsymbol{b}_r^* \odot \boldsymbol{b}_k)^{\mathsf{T}}(\dot{\boldsymbol{\zeta}}_r \odot \dot{\boldsymbol{\zeta}}_k)
\end{aligned}$$

The effects of anisotropic gain patterns will be addressed in more details in a forthcoming paper [39].

[11] In a 3D space, $\theta_r = (\phi_r, \psi_r)$ since two angles are necessary. In a 2D space, we shall just use notation θ_r.

D. Derivatives w.r.t. $\boldsymbol{K}$ and its parameters

From (28) and (13), we have [14]

$$\begin{aligned}\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{k}_r} &= \boldsymbol{s}_r \boxtimes \boldsymbol{I}_3 \boxtimes \boldsymbol{b}_r \boxtimes \boldsymbol{a}_r \\ \frac{\partial \boldsymbol{\mu}}{\partial \alpha_r} &= \boldsymbol{s}_r \boxtimes \frac{\partial \boldsymbol{k}_r}{\partial \alpha_r} \boxtimes \boldsymbol{b}_r \boxtimes \boldsymbol{a}_r \\ \frac{\partial \boldsymbol{\mu}}{\partial \beta_r} &= \boldsymbol{s}_r \boxtimes \frac{\partial \boldsymbol{k}_r}{\partial \beta_r} \boxtimes \boldsymbol{b}_r \boxtimes \boldsymbol{a}_r\end{aligned}$$

In the 2D case when elevation $\psi = 0$, we have

$$\begin{aligned}\frac{\partial \boldsymbol{k}_r}{\partial \theta_r} &= \begin{bmatrix} -\sin\theta_r & 0 \\ \cos\theta_r & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{W}(\alpha_r)\boldsymbol{w}(\beta_r) \\ \frac{\partial \boldsymbol{k}_r}{\partial \alpha_r} &= \boldsymbol{H}(\theta_r) \begin{bmatrix} -\sin\alpha_r & \cos\alpha_r \\ -\cos\alpha_r & -\sin\alpha_r \end{bmatrix} \boldsymbol{w}(\beta_r) \\ \frac{\partial \boldsymbol{k}_r}{\partial \beta_r} &= \boldsymbol{H}(\theta_r)\,\boldsymbol{W}(\alpha_r) \begin{bmatrix} -\sin\beta_r \\ \jmath\cos\beta_r \end{bmatrix}\end{aligned}$$

REFERENCES

[1] W. Liu and S. Weiss, *Wideband beamforming: concepts and techniques*. John Wiley & Sons, 2010, vol. 17.

[2] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, 1996.

[3] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[4] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 8, Boston, 1983, pp. 336–339.

[5] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, no. 7, pp. 984–995, 1989.

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

[7] L.-H. Lim and P. Comon, "Blind multilinear identification," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1260–1280, Feb. 2014.

[8] A. Swindlehurst and M. Viberg, "Subspace fitting with diversely polarized antenna arrays," *IEEE Trans. on Ant. and Prop.*, vol. 41, no. 12, pp. 1687–1694, December 1993.

[9] A. Nehorai and E. Paldi, "Acoustic vector-sensor array processing," *IEEE Trans. Sig. Process.*, vol. 42, no. 9, pp. 2481–2491, 1994.

[10] ——, "Vector-sensor array processing for electromagnetic source localization," *IEEE Trans. Sig. Process.*, vol. 42, no. 2, pp. 376–398, 1994.

[11] D. Donno, A. Nehorai, and U. Spagnolini, "Seismic velocity and polarization estimation for wavefield separation," *IEEE Trans. Sig. Process.*, vol. 56, no. 10, pp. 4794–4809, 2008.

[12] X. Guo, S. Miron, D. Brie, S. Zhu, and X. Liao, "A CANDECOMP/PARAFAC perspective on uniqueness of DoA estimation using a vector sensor array," *IEEE Trans. Sig. Process.*, vol. 59, no. 7, pp. 3475–3481, 2011.

[13] X. Guo, S. Miron, and D. Brie, "The effect of polarization separation on the performance of CANDECOMP/PARAFAC-based vector sensor array processing," *Phys. Comm.*, vol. 5, no. 4, pp. 289–295, 2012.

[14] F. Raimondi and P. Comon, "Tensor decomposition of polarized seismic waves," in *GRETSI'2015*, Lyon, France, Sep. 8-11 2015, hal-01164363.

[15] F. Raimondi, P. Comon, O. Michel, S. Sahnoun, and A. Helmstetter, "Tensor decomposition exploiting diversity of propagation velocities: Application to localization of icequake events," *Signal Processing*, vol. 118, pp. 75–88, 2016.

[16] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Acoust. Speech Signal Process. Mag.*, vol. 5, no. 2, pp. 4–24, 1988.

[17] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 33, no. 4, pp. 823–831, 1985.

[18] B. Ottersten and T. Kailath, "Direction-of-arrival estimation for wide-band signals using the ESPRIT algorithm," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 2, pp. 317–327, 1990.

[19] H. Clergeot and O. Michel, "New simple implementation of the coherent signal subspace method for wide band direction of arrival estimation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* IEEE, 1989, pp. 2764–2767.

[20] J. Krolik and D. Swingler, "Focused wide-band array processing by spatial resampling," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 2, pp. 356–360, 1990.

[21] F. Raimondi, P. Comon, and O.Michel, "Wideband multilinear array processing through tensor decomposition," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2016.

[22] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound," *IEEE Trans. Acoust. Speech Sig. Process.*, vol. 37, no. 5, pp. 720–741, 1989.

[23] B. Friedlander and A. J. Weiss, "Direction finding for wide-band signals using an interpolated array," *IEEE Trans. on Signal Process.*, vol. 41, no. 4, pp. 1618–1634, 1993.

[24] H. Hung and M. Kaveh, "Focussing matrices for coherent signal-subspace processing," *IEEE Trans. Acoust. Speech and Signal Process.*, vol. 36, no. 8, pp. 1272–1281, 1988.

[25] M. Doran, E. Doron, and A. J. Weiss, "Coherent wide-band processing for arbitrary array geometry," *IEEE Trans. Signal Process.*, vol. 41, no. 1, p. 414, 1993.

[26] H. Chen and J. Zhao, "Coherent signal-subspace processing of acoustic vector sensor array for DoA estimation of wideband sources," *Sig. Process.*, vol. 85, no. 4, pp. 837–847, 2005.

[27] H. Hung and M. Kaveh, "Coherent wide-band esprit method for directions-of-arrival estimation of multiple wide-band sources," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, pp. 354–356, 1990.

[28] X. Liu and N. D. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 2074–2086, 2001.

[29] S. Sahnoun and P. Comon, "Joint source estimation and localization," *IEEE Trans. Signal Process.*, vol. 63, no. 10, pp. 2485–2495, 2015.

[30] B. Friedlander, "The root-MUSIC algorithm for direction finding with interpolated arrays," *Signal Process.*, vol. 30, no. 1, pp. 15–29, 1993.

[31] S. Anderson and A. Nehorai, "Analysis of a polarized seismic wave model," *IEEE Trans. Sig. Proc.*, vol. 44, no. 2, pp. 379–386, 1996.

[32] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. on Circuits and Systems*, vol. 25, no. 9, pp. 114–122, Sep. 1978.

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

[34] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, 2015.

[35] P. Comon, "Tensors: a brief introduction," in *IEEE Signal Process. Mag.*, vol. 31, 2014, pp. 44–53.

[36] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *J. Chemo.*, vol. 14, pp. 229–239, 2000.

[37] J. P. C. L. da Costa, F. Roemer, M. Haardt, and R. T. de Sousa, "Multi-dimensional model order selection," *EURASIP J. Advances Signal Process.*, vol. 2011, no. 1, p. 26, 2011.

[38] L. Chiantini, G. Ottaviani, and N. Vannieuwenhoven, "An algorithm for generic and low-rank specific identifiability of complex tensors," *SIAM J. matrix Ana. Appl.*, vol. 35, no. 4, pp. 1265–1287, 2014.

[39] F. Raimondi and P. Comon, "Tensor DoA estimation with directional elements," *submitted*, Dec. 2016.

[40] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. on Signal Process.*, vol. 55, no. 6, pp. 2740–2746, 2007.

[41] S. Sahnoun and P. Comon, "Tensor polyadic decomposition for antenna array processing," in *21st Int. Conf. Comput. Stat (CompStat)*, Geneva, Aug. 19-22 2014, pp. 233–240, hal-00986973.

[42] P. Comon, X. Luciani, and A. L. F. De Almeida, "Tensor decompositions, alternating least squares and other tales," *J. chemometrics*, vol. 23, no. 7-8, pp. 393–405, 2009.

[43] P. Comon, "Estimation multivariable complexe," *Traitement du Signal*, vol. 3, no. 2, pp. 97–101, Apr. 1986. [Online]. Available: http://hdl.handle.net/2042/1600

[44] A. J. Weiss and B. Friedlander, "Performance analysis of spatial smoothing with interpolated arrays," *IEEE Trans. Signal Process.*, vol. 41, no. 5, pp. 1881–1892, 1993.