



# Tensor decomposition exploiting diversity of propagation velocities: Application to localization of icequake events

Francesca Raimondi<sup>a,\*</sup>, Pierre Comon<sup>a</sup>, Olivier Michel<sup>a</sup>, Souleymen Sahnoun<sup>a</sup>, Agnes Helmstetter<sup>b</sup>

<sup>a</sup> CNRS, University Grenoble Alpes, Gipsa-Lab, F-38000 Grenoble, France

<sup>b</sup> ISTERre, CNRS, OSUG-C, 38400 Saint Martin d'Heres Cedex, France

## ARTICLE INFO

### Article history:

Received 20 January 2015

Received in revised form

18 June 2015

Accepted 19 June 2015

Available online 29 June 2015

### Keywords:

DoA estimation

Antenna array processing

Tensor

Seismic

Elastic waves

Diversity

## ABSTRACT

The problem of direction of arrival (DoA) estimation of seismic plane waves impinging on an array of sensors is considered from a new deterministic perspective using tensor decomposition techniques. In addition to temporal and spatial sampling, further information is taken into account, based on the different propagation speed of body waves (P and S) through solid media. Performances are evaluated through simulated data in terms of the Cramér–Rao bounds and compared to other reference methods such as ESPRIT and MUSIC, in the presence of additive Gaussian circular noise. The proposed approach is then applied to real seismic data recorded at the Argentière glacier, occurring at the interface between the ice mass and the underlying bedrock. MUSIC and ESPRIT rely on the estimation of the covariance matrix of received data, thus requiring a large number of time samples. Moreover, information about propagation speed diversity is not taken into account by existing models in array processing. The discovered advantage in terms of the average error in estimating the direction of arrival of body waves is noteworthy, especially for a low number of sensors, and in separating closely located sources. Additionally, an improvement of precision in processing real seismic data is observed.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

In many cases, the superimposition principle applies in practical problems, provided the nonlinearity domain is not reached (turbulence and saturation). This allows us to model physical phenomena as linear combinations of a few simpler ones. In this paper, we are interested in the decomposition of a multivariate function into a sum of functions whose variables separate. In particular, this simplified model is relevant in narrow-band antenna array processing in the far-field, which we consider in the present framework.

In the context of seismic monitoring, seismology aims at studying waves generated by rupture phenomena

taking place within a volume of interest (rock and ice). Although the most interesting events take place at a certain depth – mostly unknown – within the analyzed volume, acquisition systems and sensor arrays are most often located close to the surface. The main quantity to be measured is ground displacement (in the form of its derivative – velocity – or its second derivative – acceleration), produced by impinging elastic waves. The localization of the sources forms the first requirement of data analysis, in order to prevent damage provoked by seismic events, and to monitor the activity of complex structures such as glaciers or volcanoes. Seismic arrays, after being introduced in the 1960s, have made essential contributions to this problem. These arrays consist of numerous seismometers placed at discrete points in space in a well-defined configuration [1]: apart from an improvement of SNR by combining individual sensor recordings, they have

\* Corresponding author.

E-mail address: [francesca.raimondi@gipsa-lab.grenoble-inp.fr](mailto:francesca.raimondi@gipsa-lab.grenoble-inp.fr) (F. Raimondi).

been used to refine models of the Earth's interior, through classical tools such as beamforming, slant stacking techniques and frequency-wave number analysis.

In wider terms, direction of arrival (DoA) estimation is a central problem in array signal processing, concerning several areas of engineering including telecommunications, speech, astronomy, seismology, and medical applications. Array processing requires a set of multiple sensors placed at different positions in space, receiving source signals from different directions [2,3]. Among techniques aiming at estimating directions of arrival, some need to resort to an exhaustive search, like beamforming and MUSIC (Multiple Signal Classification) [4], whereas others do not, like root-MUSIC [5] and ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) [6]. Whenever propagation speed is considered as a constant  $v$ , traditional array processing only relies on a temporal and spatial sampling of the propagating wavefield. MUSIC algorithm is based on the spectral decomposition of the sample covariance matrix under the spatially white noise assumption (to be presented in Appendix B). This method has the advantage of being asymptotically *statistically efficient*, unlike beamforming techniques, despite a serious sensitivity to SNR and resolution limitations for correlated or closely spaced sources [3]. Moreover, the algorithm requires the perfect knowledge of the position of each sensor. ESPRIT applies to an array composed of two identical subarrays displaced one from the other according to an unknown translation vector, whereas the calibration of the array is needed otherwise. The concept of signal-subspace processing embodied by MUSIC and ESPRIT, originally derived for narrow band and stationary signals, can be generalized to the wide-band case [7–9].

A deterministic approach based on tensor decomposition has been introduced in [10], through the extension of the rotational invariance principle to more than one displacement. It provides the localization of more sources than sensors in each subarray, with less restrictive requirements for signal stationarity than the afore-mentioned statistical methods. The advantage of tensor decompositions lies in the need for shorter data records, since the estimation of statistical quantities from available samples is not a requirement anymore. Furthermore, like ESPRIT, it allows to estimate the impinging signals up to a scale factor, without resorting to a spatial matched filter. The tensor model puts forward parsimony and separability [11]:

1. Parsimony expresses a function  $g$  as a finite sum of simpler constituents

$$g = \sum_{r=1}^R \varsigma_r h_r \quad (1)$$

2. Separability decouples a function  $h$  that depends on multiple factors into a product of simpler constituents  $\phi_d$ ,  $d = 1, \dots, D$ , each one depending only on one factor  $\mathbf{x}_d$

$$h(\mathbf{x}_1, \dots, \mathbf{x}_D) = \prod_{d=1}^D \phi_d(\mathbf{x}_d)$$

In the field of array processing for source separation and DoA estimation,  $R$  refers to the number of sources impinging on an array, and  $D$  to the tensor order, i.e. the

dimension of multilinearity within the model

$$g(\mathbf{x}_1, \dots, \mathbf{x}_D) = \sum_{r=1}^R \varsigma_r \prod_{d=1}^D \phi_{rd}(\mathbf{x}_d)$$

Tensor decomposition derives from the need to solve the *inverse problem*, i.e., the identification of factors  $\phi_{rd}$  based on noisy measurements of  $g$ : as it will be hereafter discussed, the direction of arrival can be extracted after the resolution of this problem. For this purpose, the measurements are stored in a multidimensional array and decomposed into a sum of *rank one* terms [12,10]. A *decomposable* three-way tensor can be defined by a vector triplet<sup>1</sup>

$$d_{lmk} = a_l b_m c_k \text{ or equivalently } \mathcal{D} = \mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c}$$

Any order-3 tensor admits a decomposition into a sum of decomposable tensors

$$\mathcal{M} = \sum_{r=1}^R \varsigma_r \mathcal{D}(r) \quad (2)$$

where coefficients  $\varsigma_r$  can always be chosen to be real positive, and decomposable tensors  $\mathcal{D}(r)$  to have unit norm, i.e. for Euclidean norm,  $\|\mathcal{D}\| = \|\mathbf{a}\| \|\mathbf{b}\| \|\mathbf{c}\| = 1$ . The minimal value of  $R$  such that this decomposition holds is called the tensor rank: if  $R$  is not too large, the corresponding decomposition is unique and deserves to be referred to as Canonical Polyadic (CP); other terminologies include *rank decomposition* or *Candecomp/Parafac*. Hence, decomposable tensors have a rank equal to 1, by definition [12]. Now in terms of coordinates, tensor  $\mathcal{M}$  is represented by a  $L \times M \times K$  three-way array, which consequently decomposes as

$$M_{lmk} = \sum_{r=1}^R \varsigma_r A_{lr} B_{mr} C_{kr} \quad (3)$$

where the three factor matrices  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  have unit norm columns. This is equivalent to the general  $R$ -term trilinear model

$$\mathcal{M} = \sum_{r=1}^R \varsigma_r \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r \quad (4)$$

where each array  $\mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r$  is a rank-1 array.

This paper is aimed at exploiting *another type of diversity*, in addition to spatial and temporal sampling traditionally employed in array processing (cf. Section 3 for a detailed explanation of the concept of diversity): the propagation speed diversity<sup>2</sup> of body waves through solids, namely pressure (P) and shear (S) waves. Current array processing methods like [4,6] only focus on information conveyed by a single body wave, like the P wave, whereas the contents delivered by the other is somehow wasted. Our approach intends to exploit this information as a whole, whereas translational invariance used in [10] is no longer necessary.

This paper is organized as follows. Section 2 presents the physical model and the assumptions. Section 3.1 makes a synthesis of the main narrowband 2-D preexisting algorithms. Our deterministic method, exploiting the propagation speed

<sup>1</sup> Once bases in every linear space are fixed, tensors are defined by their array of coordinates. See [12] for details.

<sup>2</sup> Since the focus is on narrow-band processing, the distinction between group and phase propagation velocities is irrelevant.

diversity of body waves, is presented in Section 3. Section 4 is devoted to the description of the employed algorithms for tensor decomposition and DoA estimation. Simulated data and real seismic data are treated in Sections 5 and 6 respectively, where results are compared with traditional methods detailed in Section 3.1 and with a statistical performance study.

## 2. Physical modeling and assumptions

Matrix algebra notations (Table 1) and main assumptions are hereby introduced:

- A1 Far-field approximation:** The distance of the source from the receiving array is much greater than the array aperture. This is equivalent to assume a planar wavefront at the sensor level. Moreover, sensors and sources are considered point-like, as their size is negligible with respect to the source-to-sensor distance. We assume that sources and sensors are coplanar, which does not restrict the generality. In fact, there is a rotational indeterminacy around the axis defined by the linear array (we can thus work in a 2-D coordinate systems).
- A2 Narrow-band in base-band:** Signals of interest are the product of a varying amplitude (complex envelope) and a high-frequency signal (cf. Appendix A). We assume that the spectral supports of both parts do not overlap (this is sometimes referred to as the Bedrosian condition).<sup>3</sup> Under this condition, one can work in base-band with the complex envelope of the low-pass signal. For this type of signal, a time delay of the original signal is equivalent to a phase shift of the complex envelope.
- A3 Homogeneous and isotropic medium at the antenna level:** Ray-paths can be approximated by straight lines.
- A4 Dissipation at the antenna scale is excluded,** as the array dimension is negligible with respect to dissipation characteristic length.
- A5 The impulse responses of particle motion** is the only source of variability between impinging P and S waves.

In the context of beamforming applied to seismic events originating at depth, one has to deal with a double arrival of elastic body waves: P waves and S waves. They have different propagation velocities and a variable frequency content: the former tend to have a higher frequency content than the latter. The P wave arrives first and its particle motion is parallel to the propagation direction, whereas the S wave (the second arrival observed on seismic records) is transversal (its particle motion lies in the plane perpendicular to the propagation direction) [14] and is linearly polarized. Linear polarization refers to the fact that more coherent seismic energy is located in one principal direction of particle motion [15]. P waves and S waves have theoretical velocities  $v_1$  and  $v_2$ , respectively,

given by Ref. [16]:

$$\begin{cases} v_1 = \left( \frac{\lambda + 2\mu}{\rho} \right)^{1/2} \\ v_2 = \left( \frac{\mu}{\rho} \right)^{1/2} \end{cases}$$

where  $\lambda$  and  $\mu$  are Lamé's constants and  $\rho$  is the density of the medium. Since they are positive constants,  $v_1$  is greater than  $v_2$ . In particular, the velocity of the S wave ranges from zero up to 70% of the velocity of the P wave. For fluids  $\mu$  is zero,  $v_2 = 0$  and therefore S waves do not propagate.

For the physical model of wave propagation, the following properties are assumed:

- P1** Pressure (P) and Shear (S) waves propagate at velocities  $v_1$  and  $v_2$ , respectively, under the approximation of non-dispersive medium.

and in order to allow the multilinearity and separability of the final model:

- P2** P and S signals are filtered in a narrow-band around  $f_1$  and  $f_2$ , respectively, such that  $\frac{v_1}{f_1} = \frac{v_2}{f_2} = \lambda$  is a constant, where  $\lambda$  refers to the wavelength of impinging waves.
- P3** Base-band P and S signals propagating from the same source event and received in two different narrow bands around  $f_1$  and  $f_2$  are proportional
 
$$w_{bf}^{(2)}(t) = \alpha(f_1, f_2) w_{bf}^{(1)}(t)$$

See (A.1) in Appendix A for more details.

Furthermore, we can summarize the following notations, which will be subsequently used:

- N1** The first sensor composing the array is taken as the origin point for translation in space.
- N2** In order to avoid scale indeterminacies in tensor decomposition, the P wave  $w_{bf}^{(1)}(t)$  is fixed up to a scale factor (see Appendix A for more details).

Additional hypotheses or notations are progressively introduced when needed:

- H1** Identical sensor responses (calibration).
- H2** Incident signals are uncorrelated to noise.
- H3** The number of sources of interest,  $R$ , is smaller than the number of sensors  $L$ :  $R < L$ .
- H4** Noise spatial coherency is known. Therefore, one can always consider (thanks to spatial prewhitening) that the noise covariance matrix is proportional to identity:  $\mathbb{E}\{\mathbf{nn}^H\} = \sigma_n^2 \mathbf{I}$ , where  $\sigma_n$  may be unknown, after whitening is applied.
- H5** Noise is additive and Gaussian complex circular:  $\mathbb{E}\{\mathbf{nn}^T\} = \mathbf{0}$ .
- H6** The number of time samples  $M$  is greater than the number of sensors  $L$ :  $M > L$ .

<sup>3</sup> The Bedrosian theorem [13] is hereby stated: The Hilbert transform of the product of two complex valued functions  $f, g: \mathbb{R} \rightarrow L^2(\mathbb{R})$  with non-overlapping Fourier spectra ( $F(\omega) \approx 0$  for  $|\omega| > a$  and  $G(\omega) \approx 0$  for  $|\omega| < a$  where  $a$  is a positive constant) is given by the product of the low-frequency signal  $f$  and the Hilbert transform of the high-frequency signal  $g$ :  $H\{f(x)g(x)\} = f(x)H\{g(x)\}$ ,  $x \in \mathbb{R}$ .

**Table 1**

Matrix algebra notations.

$\mathbf{v}$	Vector $\mathbf{v}$
$v_i$	Element of $\mathbf{v}$
$\mathbf{v}^T \mathbf{u}$	Scalar product between real vectors $\mathbf{v}$ and $\mathbf{u}$
$\mathbf{v} \otimes \mathbf{u}$	Outer (tensor) product of $\mathbf{v}$ and $\mathbf{u}$
$\mathbf{A}$	Matrix $\mathbf{A}$
$\mathbf{a}_i$	$i$ -th column of $\mathbf{A}$
$A_{ij}$	Element of $\mathbf{A}$
$\mathbf{A}^T$	Transpose of $\mathbf{A}$
$\mathbf{A}^*$	Complex conjugate of $\mathbf{A}$
$\mathbf{A}^H$	Conjugate transpose of $\mathbf{A}$
$\mathbf{A}^\dagger$	Moore–Penrose pseudoinverse of $\mathbf{A}$
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product between $\mathbf{A}$ and $\mathbf{B}$
$\mathbf{A} \oslash \mathbf{B}$	Hadamard (element-wise) product
$\ \cdot\ _F$	Frobenius norm
$\mathcal{T}$	Three-way tensor $\mathcal{T}$
$T_{ijk}$	Element of $\mathcal{T}$

### 3. Tensor-based solution

The advantage of the tensor formalism lies in its ability to restore identifiability of parameters: the impinging signals, their directions of arrival, and the related connection between P and S waves. On the other hand, the multi-dimensional character of tensor models requires the presence of at least 3 types of diversity. We review below some diversities that may be available in antenna array processing.

1. *Time diversity*: Every base-band signal  $w$  is a function of time  $t$ . Moreover, it may be stationary or transient. Thus, recorded signal  $g^{(1)}$  is given by the sum of  $R$  simultaneous sources

$$g^{(1)}(t) = \sum_{r=1}^R w_r(t) \Leftrightarrow G_m^{(1)} = \sum_{r=1}^R B_r(t_m), \quad (5)$$

where  $m = 1, \dots, M$ . The right equation represents a discretization of the left one, i.e.  $B_r(t_m)$  corresponds to impinging wave  $w_r(t)$  after sampling.

2. *Space diversity*: The basis of traditional array processing consists in performing a spatial sampling, in addition to the temporal diversity. According to assumptions **A1**, **A2** and **H1**

$$g^{(2)}(t, \mathbf{b}) = \sum_{r=1}^R \exp\left\{i \frac{2\pi}{\lambda} \mathbf{b}^T \mathbf{d}_r\right\} w_r(t) \Leftrightarrow G_{lm}^{(2)} = \sum_{r=1}^R A_r(\mathbf{b}_r) B_r(t_m) \quad (6)$$

where  $l = 1, \dots, L$  refers to the sensor cardinality with respect to the reference,  $\lambda$  is the observed narrowband wavelength,  $\mathbf{b} \in \mathbb{R}^3$  is the recording position within the acquisition system and  $\mathbf{d}_r$  is the unit vector pointing to the  $r$ -th source.<sup>4</sup>

3. *Translational diversity*: (refer to [6] for one rotational invariance and to [10] for multiple roto-

translations)

$$g^{(3)}(t, \mathbf{b}, \boldsymbol{\delta}) = \sum_{r=1}^R \exp\left\{i \frac{2\pi}{\lambda} \mathbf{b}^T \mathbf{d}_r\right\} w_r(t) \exp\left\{i \frac{2\pi}{\lambda} \boldsymbol{\delta}^T \mathbf{d}_r\right\} \Leftrightarrow G_{lmk}^{(3)} = \sum_{r=1}^R A_r(\mathbf{b}_r) B_r(t_m) D_r(\boldsymbol{\delta}_k) \quad (7)$$

where  $k = 1, \dots, K$  is the index of the translation with respect to the reference array, and  $\boldsymbol{\delta}$  is the translational invariant repeating the array configuration over space. ESPRIT may be seen as a particular case of this diversity, when translation vector  $\boldsymbol{\delta}$  relates  $K=2$  identical sub-arrays, whereas the tensor model itself can be applied to  $K > 2$  identical configurations.

4. *Polarization diversity*: Expressions (5)–(7) refer to one-component (1C) sensors, but can be generalized to three-component (3C) sensors as well. Moreover, in the latter case, one can take advantage of an additional diversity related to polarization, as explained in [17,11]:

$$g^{(4)}(t) = \sum_{r=1}^R w_r(t) \mathbf{p}(\boldsymbol{\Phi}_r) \Leftrightarrow G_{mk}^{(4)} = \sum_{r=1}^R B_r(t_m) P_k(\boldsymbol{\Phi}_r) \quad (8)$$

where  $\mathbf{p}(\cdot)$  is the polarization response vector of the considered sensor and  $\boldsymbol{\Phi}_r = [\psi_r, \theta_r, \alpha_r, \beta_r]^T$  contains the polarization information of the  $r$ -th impinging wave-field with respect to the sensor position: the azimuth  $\psi$ , the elevation  $\theta$ , the orientation angle  $\alpha$  and the ellipticity angle  $\beta$ . In particular, the elevation  $\theta$  refers to the angle of arrival when **A1** is made. Discrete index  $k \in \{1, 2, 3\}$  refers to one out of three components of the 3C sensor vector  $\mathbf{p}$ .

5. *Repetition diversity*: It is a possible extension of dimensionality whenever we deal with multiple events describing the same physical phenomenon. The additional assumption of a linear relation between events from the same cluster has to be made, so that multiple events describing the same physical phenomenon are related to each other by a complex coefficient

$$g^{(5)}(j, t) = \sum_{r=1}^R \gamma_r(j) w_r(t) \Leftrightarrow G_{jm}^{(5)} = \sum_{r=1}^R \gamma_r(j) B_r(t_m) \quad (9)$$

where discrete index  $j = 1, \dots, J$  is the cardinality of the event of the cluster (1 being the first recorded occurrence and  $J$  the last one), and  $\gamma_r(\cdot)$  a (generally unknown) complex coefficient. Repetition diversity has been already used in other contexts, in [18] for instance.

6. *Propagation speed diversity*: It may be seen as equivalent to a frequency diversity between P and S waves, under assumption **A2** and properties **P1**, **P2** and **P3**. The complete model on which this paper is based will be subsequently developed on the basis of the linear decomposition below

$$g^{(6)}(f, t) = \sum_{r=1}^R \alpha_r(f) w_r(t) \Leftrightarrow G_{mk}^{(6)} = \sum_{r=1}^R B_r(t_m) C_r(f_k) \quad (10)$$

where again the second equation represents the discretization of the first one,  $f_k$  denote working frequencies for P and S waves ( $k=1$  for the P wave and  $k=2$  for the S wave), as defined in **P2**, chosen by the user as a function of propagation speeds  $v_1$  and  $v_2$ , and  $\alpha_r(\cdot)$  a complex unknown coefficient.

<sup>4</sup> For the sake of simplicity, we will refer to both the proper direction of arrival  $\mathbf{d}_r$  and the angle of arrival  $\theta_r$  with the acronym DoA.

The focus of this paper consists in integrating at the same time the spatial, temporal and speed diversity, respectively embodied by matrices  $\mathbf{A} \in \mathbb{C}^{L \times R}$ ,  $\mathbf{B} \in \mathbb{C}^{M \times R}$  and  $\mathbf{C} \in \mathbb{C}^{2 \times R}$ .

### 3.1. Traditional matrix-based solutions

Traditional 2-D solutions only employ the concepts of *time diversity* and *space diversity*: base-band signals are stored in a data matrix  $\mathbf{Y}$  of size  $L \times M$ , where  $L$  is the number of sensors and  $M$  the number of time samples. In the particular instance of **A1** and **A2**, each scalar element is given by

$$Y_{lm} = y_l(t_m) = \sum_{r=1}^R A_{lr} w_r(t_m) + n_l(t_m), \quad l = 1, \dots, L$$

with

$$A_{lr} = \exp\left\{i \frac{\omega}{v} (\mathbf{b}_l^T \mathbf{d}_r)\right\}$$

$\mathbf{b}_l \in \mathbb{R}^3$  is the sensor position within the acquisition system,  $\mathbf{d}_r$  is the unit vector pointing to the  $r$ -th source and  $\omega$  refers to the angular frequency of the narrowband propagating wavefields. Moreover,  $n_l(t_m)$  refers to additive noise, recorded by the  $l$ -th sensor at time  $t_m$ , satisfying **H2** and **H5**.

Vector  $\mathbf{y}(t_m)$  depends on two main factors, as in (6): the  $L \times R$  steering matrix  $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_R]$ , and the  $R \times 1$  source waveform vector  $\mathbf{s}(t_m) = [w_1(t_m) \ \dots \ w_R(t_m)]^T$ . This can be written in the following compact form:

$$\mathbf{y}(t_m) = \mathbf{A} \mathbf{s}(t_m) + \mathbf{n}(t_m), \quad m = 1, \dots, M$$

or

$$\mathbf{Y} = [\mathbf{Y}(t_1) \ \dots \ \mathbf{Y}(t_M)] = \mathbf{A} \mathbf{S} + \mathbf{N}$$

where  $\mathbf{S} = [\mathbf{s}(t_1), \dots, \mathbf{s}(t_M)]$  is the  $R \times M$  source waveform matrix. This formulation is the basis of MUSIC and ESPRIT algorithms, which are detailed in [Appendix B](#).

### 3.2. Speed diversity in tensor format

P and S waves are generated by the same physical source but propagate at two different velocities,  $v_1$  and  $v_2$  respectively. Thus, P and S signals received at position  $\mathbf{b}$  and at time  $t$  are respectively given by

$$\begin{cases} y^P(\mathbf{b}, t) = \sum_{r=1}^R w_r^P(t - \tau_r^P(\mathbf{b})) + n(t, \mathbf{b}) \\ y^S(\mathbf{b}, t) = \sum_{r=1}^R w_r^S(t - \tau_r^S(\mathbf{b})) + n(t, \mathbf{b}) \end{cases}$$

where  $\{\tau_r^P, \tau_r^S\}$  express the respective delays for P and S signals recorded at position  $\mathbf{b}$ .

The signal received at a point  $\mathbf{b}$  in space contains  $R$  sources of interest plus an additive noise: under **A1** and **A2**, the time delays of P and S waves correspond to the following phase shifts respectively:  $\psi_r^P(\mathbf{b}) = \frac{\omega_1}{v_1} (\mathbf{b}^T \mathbf{d}_r)$ ,  $\psi_r^S(\mathbf{b}) = \frac{\omega_2}{v_2} (\mathbf{b}^T \mathbf{d}_r)$ .

With property **P2**, records produced by the  $r$ -th source contain two delayed narrowband waves propagating around frequencies  $f_1, f_2$  at velocities  $v_1$  and  $v_2$ , yielding two resulting waves with the same wavelength  $\lambda$ . This model is equivalent to the general three-variable expression

$$z(t, \mathbf{b}, f) = \sum_{r=1}^R \alpha_r(f) w_r(t) \exp\{i \psi_r(\mathbf{b})\} + n(t, \mathbf{b}, f)$$

where coefficient  $\alpha_r(f)$  is the same as in (10) and  $\psi_r(\mathbf{b}) = \frac{2\pi}{\lambda} (\mathbf{b}^T \mathbf{d}_r)$ . Thus, in the absence of noise  $n(t, \mathbf{b}, f)$ , function  $z(t, \mathbf{b}, f)$  decomposes into a sum of  $R$  functions whose variables separate.

Discretization yields an array of finite dimensions  $L \times M \times 2$  ( $L$  sensors located at  $\mathbf{b}_\ell$ ,  $M$  time samples  $t_m$  and  $K=2$  propagation velocities  $v_k$ ), so that the data array writes

$$Z_{lmk} = \sum_{r=1}^R A_r(\mathbf{b}_\ell) B_r(t_m) C_r(f_k) + n_l(t_m, f_k)$$

where frequency  $f_k, k = 1, 2$ , is fixed according to property **P2** ( $\lambda = \frac{v_1}{f_1} = \frac{v_2}{f_2}$ ). In the case of a Uniformly spaced Linear Array (ULA), we just have that  $\mathbf{b}_\ell^T \mathbf{d}_r = (l-1)\Delta \sin \theta_r$ , where  $\theta_r$  refers to the angle of arrival of the  $r$ -th source  $\mathbf{d}_r$ , and  $\Delta$  is the distance between two consecutive sensors composing the ULA.

### 3.3. Exact CP decomposition

A CP decomposition is said to be *essentially* unique if there exists a unique set  $\{\zeta_r, \mathcal{D}(r), 1 \leq r \leq R\}$  such that equality holds in (2). However, even if the CP decomposition is unique, there exist several ways of writing (3). In fact, writing a decomposable tensor as the outer product of vectors is subject to scaling indeterminacies, which actually stem from multilinearity properties of tensors [12], since

$$\alpha \mathbf{a} \otimes \beta \mathbf{b} \otimes \gamma \mathbf{c} = \mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c}$$

if  $\alpha\beta\gamma = 1$ . Even if the above vectors are of unit Euclidean norm, there remain two scaling indeterminacies of unit modulus. As a consequence, attention should be paid to the difference between CP or *essential* uniqueness, and uniqueness of matrix factors appearing in (3).

There exist sufficient conditions ensuring CP uniqueness, e.g. the Kruskal condition [10]

$$\kappa_A + \kappa_B + \kappa_C \geq 2R + 2 \quad (11)$$

where the notation  $\kappa_A$  refers to the *Kruskal-rank*<sup>5</sup> of matrix  $\mathbf{A}$ . However, less stringent conditions guaranteeing almost surely a unique solution can be found, for instance [11,12,19]

$$R(K + L + M - 2) < LMK$$

In the present paper, we deal with simpler examples when  $R=1$  as for glacier data (cf. Section 6) and  $R=2$  relative to simulated data (cf. Section 5) so that essential uniqueness of a decomposition is always guaranteed.

<sup>5</sup> The Kruskal rank of a matrix  $\mathbf{A}$  is the largest number  $\kappa_A$  such that any subset of  $\kappa_A$  columns are linearly independent.



In practice, it is more convenient to fix *trivial* indeterminacies of unit modulus in Eq. (3). Therefore, as suggested in [20], we fix 2R parameters without restricting the generality. More precisely, based on properties described in Section 2, we assume the following:

1. From property **N1** we refer to the first sensor by index 1

$$\exp\{i\psi_{1r}\} = \exp\left\{\frac{2\pi}{\lambda}(\mathbf{b}_1^T \mathbf{d}_r)\right\} \forall r \implies \mathbf{A}_{1r} = 1 \forall r \quad (12)$$

2. From property **P3** we refer to the P wave by index 1

$$\alpha_r(f_1) = 1, \quad \forall r \implies \mathbf{C}_{1r} = 1 \quad \forall r, \quad (13)$$

cf. (3)

Note that after fixing the scaling indeterminacies, the number of remaining free parameters is now  $(L+M+K-2)R$  [20], and not  $(L+M+K-3)R$  as assumed in [10].

### 3.4. Presence of noise: maximum likelihood

Cramér–Rao Bound (CRB) represents the lower bound on the variance of any unbiased estimator of a deterministic parameter. We now detail the derivation of the CRB in order to assess the performance of the proposed method and then compare different methods with one another. If noise  $\mathbf{n}$  is considered as i.i.d. complex circularly symmetric Gaussian of known variance  $\sigma_n^2$ , Signal-to-Noise ratio (SNR) is defined as (cf. [21])

$$\text{SNR} = 10 \log_{10} \frac{\|\mathcal{M}\|_F^2}{LMK\sigma_n^2}$$

where the operator  $\|\cdot\|_F$  indicates Frobenius norm:  $\|\mathcal{M}\|_F = \sqrt{\text{tr}(\boldsymbol{\mu}^H \boldsymbol{\mu})}$  and  $\boldsymbol{\mu} = \text{vec}(\mathcal{M})$ . The problem of scale ambiguity of the CP decomposition is solved by the assumption of expressions (12) and (13) that the first row of matrices  $\mathbf{A}$  and  $\mathbf{C}$  is normalized to  $[1 \dots 1]_{1 \times R}$ . Therefore, the unknown complex parameter vector  $\boldsymbol{\vartheta} \in \mathbb{C}^{(L+M+K-2)R \times 1}$  has the following form:

$$\boldsymbol{\vartheta} = [\bar{\mathbf{a}}_1^T, \dots, \bar{\mathbf{a}}_R^T, \mathbf{b}_1^T, \dots, \mathbf{b}_R^T, \bar{\mathbf{c}}_1^T, \dots, \bar{\mathbf{c}}_R^T]$$

where  $\bar{\mathbf{a}}_r^{\text{def}} = [A_{2r}, \dots, A_{Lr}]$ ,  $\mathbf{b}_r^{\text{def}} = [B_{1r}, \dots, B_{Mr}]$ ,  $\bar{\mathbf{c}}_r^{\text{def}} = [C_{2r}, \dots, C_{Kr}]$ .

Given the additive model  $\mathcal{Z} = \mathcal{M} + \mathcal{N}$ , the likelihood function for a zero mean, circularly symmetric complex Gaussian noise  $\mathbf{n} = \text{vec}(\mathcal{N})$  with covariance  $\sigma_n^2 \mathbf{I}$  takes the form:

$$L(\mathbf{z}|\boldsymbol{\vartheta}) = \frac{1}{(\pi\sigma_n^2)^{LMK}} \exp\left\{-\frac{1}{\sigma_n^2}(\mathbf{z} - \boldsymbol{\mu}(\boldsymbol{\vartheta}))^H (\mathbf{z} - \boldsymbol{\mu}(\boldsymbol{\vartheta}))\right\},$$

where  $\mathbf{z} = \text{vec}(\mathcal{Z})$ . The corresponding log-likelihood function  $f(\boldsymbol{\vartheta}) = \log(L(\mathbf{z}|\boldsymbol{\vartheta}))$  can be written as

$$f(\boldsymbol{\vartheta}) = -LMK \log(\pi\sigma_n^2) - \frac{1}{\sigma_n^2}(\mathbf{z} - \boldsymbol{\mu}(\boldsymbol{\vartheta}))^H (\mathbf{z} - \boldsymbol{\mu}(\boldsymbol{\vartheta}))$$

The maximization of the likelihood amounts to minimizing the square of the  $\ell^2$  norm  $\|\mathbf{z} - \boldsymbol{\mu}(\boldsymbol{\vartheta})\|_2$ .

### 3.5. DoA Cramér–Rao bounds

Then, the complex Fisher information matrix (FIM) is given by (cf. [21,20])

$$\boldsymbol{\Phi}(\boldsymbol{\vartheta}) = \mathbb{E}\left\{\left(\frac{\partial f(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right)^H \left(\frac{\partial f(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right)\right\}$$

A straightforward derivation leads to

$$\frac{\partial f(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \frac{1}{\sigma_n^2} \left[ \mathbf{n}^T \frac{\partial \boldsymbol{\mu}^*}{\partial \boldsymbol{\vartheta}} + \mathbf{n}^H \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\vartheta}} \right]$$

where  $\mathbf{n} = \mathbf{z} - \boldsymbol{\mu}$ .

Since noise is circularly symmetric ( $\mathbb{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma_n^2 \mathbf{I}_{LMK}$  and  $\mathbb{E}\{\mathbf{n}\mathbf{n}^T\} = 0$ ), the FIM reduces to [20]

$$\boldsymbol{\Phi}(\boldsymbol{\vartheta}) = \frac{1}{\sigma_n^2} \left[ \left(\frac{\partial \boldsymbol{\mu}^*}{\partial \boldsymbol{\vartheta}}\right)^H \left(\frac{\partial \boldsymbol{\mu}^*}{\partial \boldsymbol{\vartheta}}\right) + \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\vartheta}}\right)^H \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\vartheta}}\right) \right]$$

The CRB of any unbiased estimator of complex vector parameter  $\boldsymbol{\vartheta}$  is given by the inverse of the FIM (see Appendix C for more details about CRB of the DoA estimation problem).

## 4. Algorithms

### 4.1. Computation of the CP decomposition

We hereby recall the expressions of the  $R$ -term trilinear model for the three-way array  $\mathcal{M}$ :

$$M_{lmk} = \sum_{r=1}^R \varsigma_r A_{lr} B_{mr} C_{kr}$$

$$\mathcal{M} = \sum_{r=1}^R \varsigma_r \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r \quad (14)$$

A direct decomposition algorithm has been proposed in [22] for  $K \geq 2$ : it takes as inputs the  $K$  matrix slabs  $\mathbf{M}_k = \mathcal{M}[:, :, k]$  and the number of factors  $R$ , and returns the estimates of matrices  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$ . Since the third dimension  $K$  investigated in the present paper refers to the propagation speed diversity, it corresponds to  $K=2$  ( $k=1$  for the P wave and  $k=2$  for the S wave). We thus state what follows for  $K=2$ , so as to ease its readability. The uniqueness of such a decomposition is guaranteed provided that the  $R$  column vectors corresponding to two of the ways are linearly independent, and the  $R$  column vectors associated with the third way have the property that no two are collinear (see below). The trilinear model (14) can be rewritten as  $K=2$  matrices  $\mathbf{M}_k$  of size  $L \times M$ , according to the *trilinear matrix equations* [23,22]

$$\begin{cases} \mathbf{M}_1 = \sum_{r=1}^R (\mathbf{a}_r \otimes \mathbf{b}_r) \mathbf{C}_{1r} = \sum_{r=1}^R \mathbf{a}_r \mathbf{C}_{1r} \mathbf{b}_r^T \\ \mathbf{M}_2 = \sum_{r=1}^R (\mathbf{a}_r \otimes \mathbf{b}_r) \mathbf{C}_{2r} = \sum_{r=1}^R \mathbf{a}_r \mathbf{C}_{2r} \mathbf{b}_r^T \end{cases}$$

which we denote in short

$$\mathbf{M}_k = \mathbf{A} \mathbf{C}_k \mathbf{B}^T, \quad k \in \{1, 2\} \quad (15)$$

where

$$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_R],$$

$$\begin{aligned}\mathbf{B} &= [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_R], \\ \mathbf{C}_1 &= \text{Diag}\{C_{11}, C_{12}, \dots, C_{1R}\} \\ \mathbf{C}_2 &= \text{Diag}\{C_{21}, C_{22}, \dots, C_{2R}\}\end{aligned}$$

and coefficients  $\varsigma_r$  were pulled in factor matrix  $\mathbf{B}$ .

The following more constraining identifiability conditions are employed, which concurrently imply the Kruskal condition in Eq. (11):

- IC1** The columns of  $\mathbf{A}$  are linearly independent i.e.  $\kappa_A = R$ .
- IC2** The columns of  $\mathbf{B}$  are linearly independent i.e.  $\kappa_B = R$ .
- IC3** Every pair of columns of  $\mathbf{C}$  is linearly independent i.e.  $\kappa_C = 2$ .

Each condition refers to one way of the array: the first two conditions state that there must be at least  $R$  factors present in two ways. The third requires that no two factors are linked by a proportional relationship along the other way.

The decomposition proposed in [22] exploits the comparison between the two following linear combinations, for every vector of weights  $\boldsymbol{\varrho}$

$$\begin{cases} \mathbf{M}(\boldsymbol{\varrho}) = \boldsymbol{\varrho}_1 \mathbf{M}_1 + \boldsymbol{\varrho}_2 \mathbf{M}_2 \\ \mathbf{C}(\boldsymbol{\varrho}) = \boldsymbol{\varrho}_1 \mathbf{C}_1 + \boldsymbol{\varrho}_2 \mathbf{C}_2 \end{cases}$$

As particular combination we denote  $\mathbf{M}_+ = \mathbf{M}_1 + \mathbf{M}_2$  and  $\mathbf{C}_+ = \mathbf{C}_1 + \mathbf{C}_2$ . The following result constitutes the core of the decomposition algorithm (see [22] for more details):

**Theorem 1.** *If conditions IC1 and IC2 hold,  $\mathbf{A}$  and  $\mathbf{B}$  have a left inverse and:*

1. the columns of matrix  $\mathbf{A}$  are the eigenvectors of the matrix  $\mathbf{M}(\boldsymbol{\varrho})\mathbf{M}_+^\dagger$ . The corresponding eigenvalues are the diagonal entries of  $\mathbf{C}(\boldsymbol{\varrho})\mathbf{C}_+^\dagger$ .
2. matrix  $\mathbf{B}$  can be obtained as  $\mathbf{B} = (\mathbf{C}_+^{-1}\mathbf{A}^\dagger\mathbf{M}_+)^T$ ;
3. the columns of matrix  $\mathbf{C}$  can be obtained thanks to the relation  $\mathbf{C}_k = \mathbf{A}^\dagger\mathbf{M}_k(\mathbf{B}^T)^\dagger$ .

**Proof.** [Sketch] From (15), we have  $\mathbf{M}_k = \mathbf{A}\mathbf{C}_k\mathbf{B}^T$ ,  $k \in \{1, 2\}$ , and  $\mathbf{M}_+ = \mathbf{A}\mathbf{C}_+\mathbf{B}^T$ . Hence for any  $k$ ,  $\mathbf{M}_k\mathbf{M}_+^\dagger\mathbf{A} = \mathbf{A}\mathbf{C}_k\mathbf{C}_+^{-1}$ .  $\square$

#### 4.2. Extraction of DoA from matrix factors

Once the decomposition is obtained, the parameters of the trilinear problem can be extracted as follows.

1. Direction of arrival  $\theta_r$  of the  $r$ -th source impinging on the ULA is calculated from

$$A_{lr} = \exp\left\{i\frac{2\pi}{\lambda}(\mathbf{b}_r^T \mathbf{d}_r)\right\} = \exp\left\{i\frac{2\pi}{\lambda}\Delta(l-1)\sin(\theta_r)\right\}$$

In order to estimate direction of arrival  $\boldsymbol{\theta}$ , matrix  $\mathbf{A}$  can be considered as a *Vandermonde* matrix in the absence of noise if the antenna is ULA, with kernel  $u_r = \exp\left\{i\frac{2\pi}{\lambda}\Delta\sin(\theta_r)\right\}$ . Direction of arrival is then estimated through a Least Squares (LS) solution or the more robust Total Least Squares (TLS) solution. Each  $r$ -th line of the Vandermonde matrix contains multiple

information about the direction of arrival  $\theta_r$ . If two overlapping blocks are taken,  $\bar{\mathbf{a}} = (1 \ u_r \dots u_r^{L-2})^T$  and  $\underline{\mathbf{a}} = (u_r \ u_r^2 \dots u_r^{L-1})^T$ , the following relationship should hold in the absence of noise:

$$u_r \bar{\mathbf{a}} = \underline{\mathbf{a}}$$

The Least Squares (LS) solution is given by  $\hat{u}_r^{LS} = \frac{\bar{\mathbf{a}}^H \underline{\mathbf{a}}}{\|\bar{\mathbf{a}}\|^2}$ . The Total Least Squares (TLS) solution (described in Appendix D) theoretically yields a better solution, but no difference has been observed in practice in our computer experiments. Once  $\hat{u}_r$  is computed, the  $r$ -th angle of arrival can be obtained as  $\hat{\theta}_r = \arcsin\left[\frac{\lambda}{2\pi\Delta}\text{angle}(\hat{u}_r)\right]$ , where  $\text{angle}(\cdot)$  is the phase angle determination in  $[-\pi, \pi]$ .

2. Signal  $w_r(t_m)$  of the  $r$ -th source is extracted up to a scaling factor directly from matrix  $\mathbf{B}$ :  $\hat{w}_r(t_m) \propto B_{mr}$ , for  $m = 1, \dots, M$ .
3. Complex multiplicative coefficient  $\alpha_r(f_k)$  is extracted directly from the factor  $\mathbf{C}$ :  $\hat{\alpha}_r(f_k) = C_{kr}$ .

#### 4.3. Refinement

An alternative to the exact tensor decomposition is represented by tensor approximation minimizing R-rank approximation error  $Y$ , as in [20]

$$Y(\mathbf{A}, \mathbf{B}, \mathbf{C}; \boldsymbol{\varsigma}) = \left\| \mathbf{z} - \sum_{r=1}^R \varsigma_r \mathbf{a}_r \boxtimes \mathbf{b}_r \boxtimes \mathbf{c}_r \right\|^2 = Y(\mathbf{x}; \boldsymbol{\varsigma}) \quad (16)$$

with  $\mathbf{z} = \text{vec}\{\mathcal{Z}\}$ ,  $\mathbf{x} = \text{vec}\{[\mathbf{A}^T, \mathbf{B}^T, \mathbf{C}^T]\}$  and  $\boldsymbol{\varsigma} = [\varsigma_1, \dots, \varsigma_R]$ . An iterative minimization problem (e.g. conjugate gradient descent method) in order to find  $(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}})$  could be initialized with the exact decomposition in [22], and is stated as

$$(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}) = \arg \min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} (Y(\mathbf{A}, \mathbf{B}, \mathbf{C}; \boldsymbol{\varsigma})) = \arg \min_{\mathbf{x}} (Y(\mathbf{x}; \boldsymbol{\varsigma}))$$

An alternative to the two-stage derivation of DoA estimates (tensor CP decomposition followed by an estimation of  $\boldsymbol{\theta}$  from matrix  $\mathbf{A}(\boldsymbol{\theta})$ ) lies in a direct parametrization of the objective function in Eq. (16) with DoAs with  $\mathbf{x} = \text{vec}\{[\mathbf{A}^T(\boldsymbol{\theta}), \mathbf{B}^T, \mathbf{C}^T]\}$

$$(\hat{\boldsymbol{\theta}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}) = \arg \min_{\boldsymbol{\theta}, \mathbf{B}, \mathbf{C}} (Y(\mathbf{A}(\boldsymbol{\theta}), \mathbf{B}, \mathbf{C}; \boldsymbol{\varsigma})) = \arg \min_{\mathbf{x}} (Y(\mathbf{x}(\boldsymbol{\theta}); \boldsymbol{\varsigma}))$$

This may allow performances to get closer to the CR bounds.

#### 5. Numerical simulations

Signals were simulated according to sampling conditions, which are typically those selected for Argenti re experiments (cf. [24,25]) and emulate the normal seismic activity of a glacier. Estimation efficiency was evaluated in comparison with two other narrowband algorithms, ESPRIT [6] and MUSIC [4,26], and with deterministic CRB as a benchmark [20]. The performance criterion is the *Root*

Mean Square Error (RMSE) of the DoA

$$\text{RMSE}(\boldsymbol{\theta}) = \sqrt{\frac{1}{NR} \sum_{n=1}^N \sum_{r=1}^R (\hat{\theta}_m - \theta_r)^2}, \quad (17)$$

where  $\hat{\theta}_m$  is the estimated DoA of source  $r$  in the  $n$ -th Monte-Carlo trial,  $N$  being the number of trials. The number of simultaneous sources was chosen to be  $R = 2$ . The SNR definition for P wave was

$$\text{SNR} = 10 \log_{10} \frac{\|\mathbf{p}\|_F^2}{LM\sigma_n^2}$$

where  $\|\mathbf{p}\|_F^2$  refers to the P wave energy. This is consistent with previous works, where the S wave data is traditionally thrown away even if received. We have:

1. We use values for P and S wave velocities through ice, as in [24,27]:  $v_1 = 3600 \text{ m s}^{-1}$  and  $v_2 = 1610 \text{ m s}^{-1}$ .
2.  $\frac{v_1}{v_2} = \frac{f_1}{f_2} \approx 2.24$ .
3.  $f_1 = 193 \text{ Hz}$ ,  $f_2 = 86 \text{ Hz}$ .
4. Time duration of simulated records: 201 ms.
5. Configuration of the array: ULA with inter-sensor distance  $\Delta = 10 \text{ m}$ . Three arrays were simulated, each with a different number of sensors:  $L \in \{3, 10, 30\}$ .
6. Simulated directions of arrivals:
 

Fig. 1: endfire DoA<sub>1</sub> =  $-60^\circ$ , DoA<sub>2</sub> =  $-40^\circ$   
 Fig. 2: broadside DoA<sub>1</sub> =  $-10^\circ$ , DoA<sub>2</sub> =  $+10^\circ$   
 Fig. 3(a): endfire DoA<sub>1</sub> =  $(-45 - x/2)^\circ$ , DoA<sub>2</sub> =  $(-45 + x/2)^\circ$   
 Fig. 3(b): broadside DoA<sub>1</sub> =  $(0 - x/2)^\circ$ , DoA<sub>2</sub> =  $(0 + x/2)^\circ$

where  $x$  refers to the  $x$ -axis of Fig. 3 (the overall angular separation between the two sources).

A simple source was simulated as approximately narrowband: a signal carried by a sinusoid modulated by a Kaiser window of envelope  $a_r(t_m)$

$$w_{kr}(t_m) = a_{kr}(t_m) \exp\{i\phi_{kr}(t_m)\} \exp\{i\omega_k t_m\}$$

where  $\phi_{kr}(t_m)$  are independently drawn in  $[0, 2\pi)$  for every  $t_m$ . Since narrowband  $\omega_k$  is known for P and S waves, simulated signal  $w_{kr}(t_m)$  is brought to the baseband through demodulation

$$w_{kr}(t_m) = a_{kr}(t_m) \exp\{i\phi_{kr}(t_m)\}$$

Monte Carlo simulations ( $N=200$  trials) show a superior performance of CP deterministic decomposition with respect to MUSIC and ESPRIT algorithms, in the sense of the RMSE defined in (17), especially for a low number of sensors composing the array. For larger number of sensors, performances of MUSIC and CP decomposition become comparable. In particular, Figs. 1 and 2 illustrate computer simulations for a variable number of sensors ( $L \in \{3, 10, 30\}$ ) composing the array and for different configurations of sources impinging from a *broadside* (perpendicular to the array) or *endfire* area (laterally to the array) perspective.

The following observations can be made.

- For low SNR values, the RMSE does not increase anymore at some point, as SNR decreases. This can be observed in particular for  $L=3$  in Figs. 1(a) and 2(a). This is a known saturation phenomenon [28], which is

due to the fact that estimated angles are bounded. For such low values of SNR, the CRB becomes meaningless.

- For larger numbers of sensors, e.g.  $L=10$  or  $L=30$ , the latter saturation phenomenon is not yet observed at SNR=5 dB, and one can observe a take-off of the RMSEs from their CRBs. This threshold effect is well known and expected at low SNR values [29].
- The performance of CP and ESPRIT methods do not reach the CRB for  $L=10$  or  $L=30$ . This loss can be explained by the 2-step nature of the estimation process, which consists of estimating first steering vectors, and then corresponding DoAs in a second stage (cf. Section 4.2).
- For  $L=10$  and  $L=30$ , MUSIC seems to perform better than ESPRIT. This is due to the fact that MUSIC finds the DoA by executing an exhaustive search, which is hence a one-stage procedure.

Figs. 3 and 4 represent the evolution of the RMSE (obtained with  $N=300$  trials) with respect to angular separation of the two simultaneous sources, for a fixed SNR of 20 dB: the tendency is an evident predominance in the performance of the CP decomposition algorithm, especially for closely spaced sources.

To conclude, the advantage of CP decomposition over ESPRIT is always present, whereas it outperforms MUSIC substantially when the number of sensors remains limited or when sources are close to each other (less than  $15^\circ$ ).

## 6. Results on real data

### 6.1. Windowing of P and S recorded signals

Since recorded P and S waves are consecutive and distinct in time, they need to be selected and cropped, so as to form two  $M \times L$  slices within the data sensor, where  $M$  is the number of data samples, and  $L$  the number of sensors. Whenever P and S complex envelopes are not aligned, the broadband general formulation provides for an additive model of a deterministic component and a stochastic component

$$\begin{cases} y^{(P)}(t - t_l^{(P)}) = w_P(t - t_l^{(P)}) + n_l(t - t_l^{(P)}) \\ y^{(S)}(t - t_l^{(S)} + \xi_{PS}) = w_S(t - t_l^{(S)} + \xi_{PS}) + n_l(t - t_l^{(S)} + \xi_{PS}) \end{cases}$$

where  $l = 1, \dots, L$ , and  $\xi_{PS}$  is the alignment error between P and S slices. The narrowband approximation (A2) at a given ostensible wavelength  $\lambda = \frac{v_1}{f_1} = \frac{v_2}{f_2}$  gives

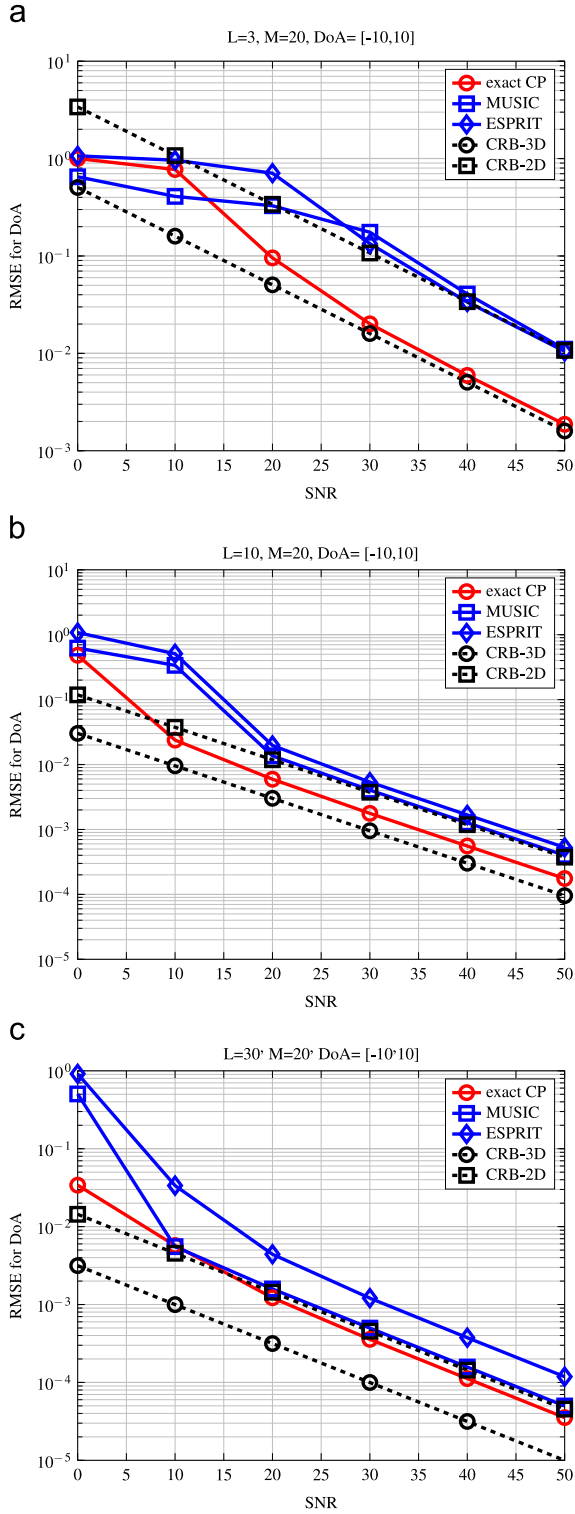
$$\begin{cases} y^{(P)}(t - t_l^{(P)}) \simeq \alpha_P w(t) \exp\{-i\psi_l\} + n_l(t - t_l^{(P)}) \\ y^{(S)}(t - t_l^{(S)} + \xi_{PS}) \simeq \alpha_S w(t + \xi_{PS}) \exp\{-i\psi_l\} + n_l(t - t_l^{(S)} + \xi_{PS}) \end{cases}$$

where  $\psi_l = \frac{2\pi}{\lambda} \Delta(l-1) \sin \theta$ , and proportional coefficients  $\alpha_P$  and  $\alpha_S$  directly derive from property P3.

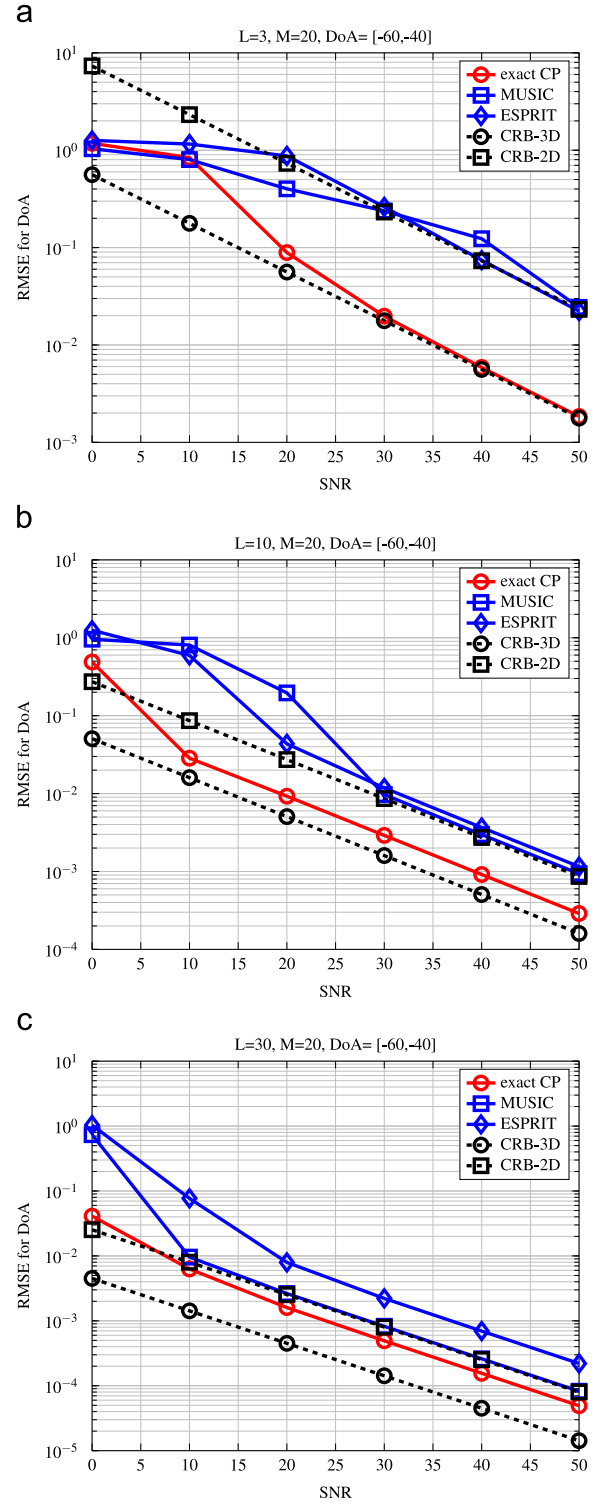
Alignment is pursued through cross-correlation between P and S narrowband complex envelopes after the detection process, which is theoretically justified provided signals are jointly stationary and decorrelated to noise  $n_l(t)$ :

$$\begin{aligned} R_l^{PS}(\tau + \xi_{PS}) &= \mathbb{E}\{y^{(P)}(t - t_l^{(P)}) y^{(S)}(t - t_l^{(S)} + \tau + \xi_{PS})\} \\ &= \alpha_P \alpha_S \exp\{-i2\psi_l\} R_l^{ww}(\tau + \xi_{PS}) \end{aligned}$$





**Fig. 1.** DoA RMSE (rad) vs SNR (dB) and for various numbers of sensors and  $M=20$ . Broadside perspective:  $\text{DoA} = [-10, 10]$ . (a) array of  $L = 3$  sensors, (b) array of  $L = 10$  sensors, and (c) array of  $L = 30$  sensors



**Fig. 2.** DoA RMSE (rad) vs SNR [dB] and for various numbers of sensors and  $M=20$ . Endfire perspective:  $\text{DoA} = [-60, -40]$ . (a) endfire perspective - array of  $L = 3$  sensors, (b) array of  $L = 10$  sensors, and (c) array of  $L = 30$  sensors

If the alignment is fulfilled on the reference sensor ( $l=1$ ),

$$R_1^{PS}(\tau + \xi_{PS}) = \alpha_P \alpha_S R_1^{WW}(\tau + \xi_{PS})$$

which attains its maximum for  $\tau_{max} = -\xi_{PS}$ . Once  $\tau_{max}$  is determined, the best alignment can be performed to the P and S waves.

## 6.2. A dataset of 27 events from the same cluster at the Argenti re glacier

The Argenti re glacier is a 10 km long glacier located in the French side of Mont Blanc massif, covering a 19 km<sup>2</sup> surface. It is characterized by high seismic activity, as stated in [30,31]. Temperate Alpine glaciers are characterized by ice remaining at melting point, deforming by two main mechanisms: plastic deformation of ice and basal sliding. The underneath flow of water plays an important role in basal sliding, because it modifies the water pressure. Furthermore, brittle behavior of ice is the reason of the sudden openings of crevasses and falls of seracs on the surface of a glacier, with indirect effects on glacial hydrology by means of water transfer from the surface to greater depth zones. An array composed of 9 velocimeters (model is Sismob IHR1C, 1 component, eigenfrequency of 2 Hz) pointing in the direction perpendicular to glacier motion at 1 kHz sampling frequency was placed on the glacier surface at 2400 m above sea level. The sensors within the array were spaced out at 10 m intervals, and the thickness of the glacier under the array is about 190 m [24,32]. Described deep events took place at 450 m from the array, at a depth of about 190 m from the surface [24]. The general case involves  $R$  far field sources impinging on the array, at a large unknown distance [24], in the presence of noise, thus allowing us to use the plane wave approximation **A1** of the wavefields impinging on the array. Signals resulting of seismic events within the glacier are assumed zero-mean, nonstationary and broadband stochastic processes over the observation time, and are assumed to be uncorrelated to environmental noise. The DoA technique

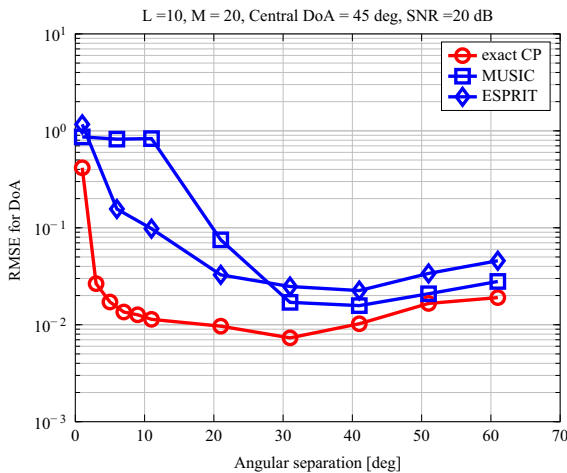


Fig. 3. RMSE (rad) vs angular separation (deg) – endfire perspective.

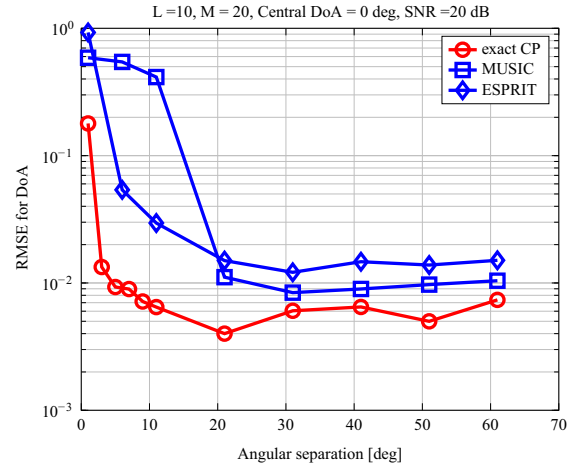


Fig. 4. RMSE (rad) vs angular separation (deg) – broadside perspective.

evaluated via simulated signals was applied to the 27 deep events recorded by the array during November 2012. These events were associated with the same cluster of deep events [24], on the basis of their waveform similarities through cross-correlation. We use values for P and S wave velocities through ice, as in [27,24]:  $v_1 = 3600 \text{ m s}^{-1}$  and  $v_2 = 1610 \text{ m s}^{-1}$ , according to Property **P1**. An example of one deep event recorded by the 9 sensors with P and S waves is provided in Fig. 7.

The frequency optimization was achieved throughout the dataset from the same cluster of events (see Fig. 5 for results), after preprocessing of recorded signals (see Appendix E for details). Median dominant frequency  $\bar{f}_1$  of P wave is then calculated. Thus, all the P waves from the cluster are filtered around  $\bar{f}_1 = 228 \text{ Hz}$  and all the S waves are filtered around  $\bar{f}_2 = \frac{v_2}{v_1} \bar{f}_1 = 102 \text{ Hz}$  before tensor decomposition. This result is coherent with spectral analysis of P and S waves. DoA estimation is performed over the dataset with different methods: MUSIC and ESPRIT over P and S waves, and joint tensor decomposition with speed diversity. Results are shown in Fig. 6: the dispersion of the DoA estimates through the joint CP decomposition of P and S waves is smaller than that of MUSIC and ESPRIT for P and S waves separately.

## 7. Conclusion and perspectives

Throughout this paper, we developed a tensor decomposition model for seismic data exploiting propagation speed diversity of P and S waves. A physical model was traced, followed by simulations and statistical comparisons with ESPRIT and MUSIC, and theoretical CRBs. Our approach was also tested on real data of seismic activity of an Alpine glacier, and different techniques were compared in terms of localization efficacy.

The strength of our method lies in the integration of the double information conveyed by P and S waves of distant events impinging in succession on the array: to the traditional dimensions of array processing (recording samples for time and sensor locations for space), we added

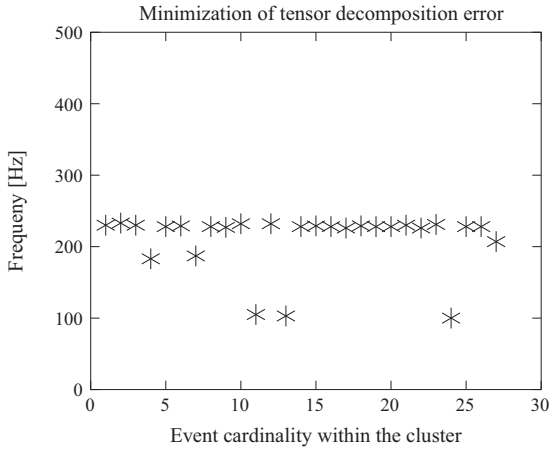


Fig. 5. Optimized working frequency for P wave.

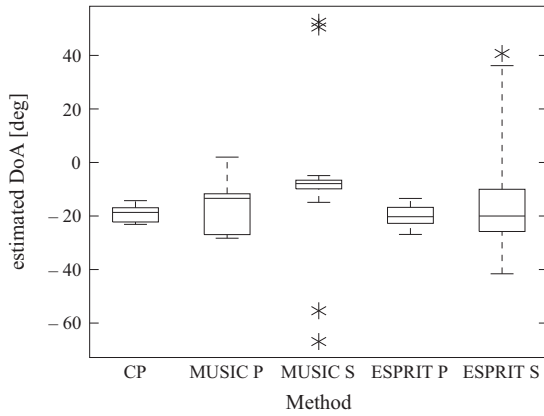


Fig. 6. DoA estimation throughout the cluster of deep events.

the dual content transmitted by the P and S waves, temporally distinct by virtue of the diversity of propagation speed. The effect of adding a way to the data array is evident in terms of estimation performances and localization precision of a cluster of real events originating from the same source, especially for short data durations and for arrays composed of a small number of sensors.

Further research is expected upon deterministic models of repetition and polarization diversities, in the context of array processing. In the matter of adaptation to real data, since A3 may become, unfortunately, unrealistic, a phase velocity heterogeneity among sensors may be another factor hindering performance of practical application. Furthermore, calibration errors concerning the configuration of sensors in the array may possibly be included into the model, with consequences upon the statistical performances of the proposed method.

## Acknowledgments

This work has been supported by the ERC AdG-2013-320594 Grant “DECODA”. We thank the Associate Editor

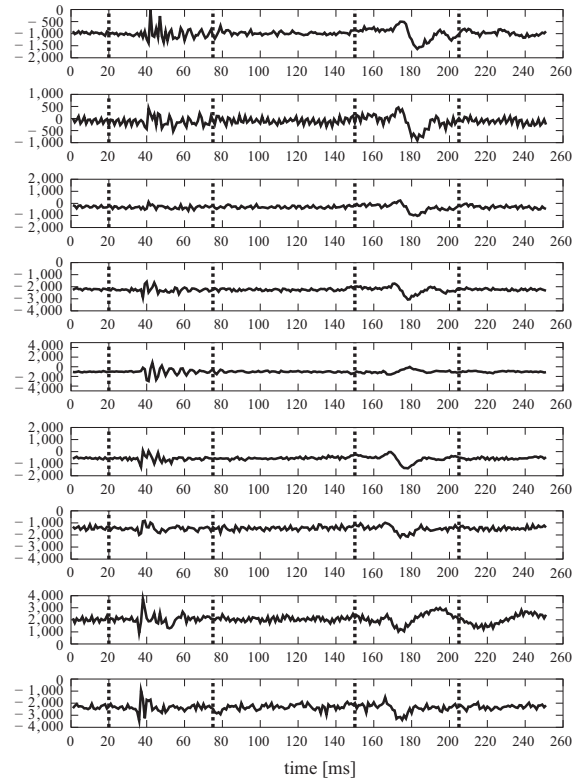


Fig. 7. Example of an event from the Argenti re glacier – 2012: record from the array with distinct P and S waves – vertical component.

and the five anonymous referees for their careful reading and helpful comments.

## Appendix A. Justification of P3

The resulting signal propagating from a given source  $r$  and impinging on a given sensor of the array is composed of two contributions before addition to background and instrumentation noise: the P phenomenon, generally followed by the S phenomenon. An additional assumption A5 identifies the impulse responses of particle motion as the only source of variability between received P and S waves

$$w^{(1)}(t) \approx h^{(1)}(t) * w(t) m(t)$$

$$w^{(2)}(t) \approx h^{(2)}(t) * w(t) m(t)$$

where  $h^{(1)}(t)$  is the impulse response of the given medium to the P wave and  $h^{(2)}(t)$  to the S wave. In particular, notice that the source signal can be decomposed into a product of a low-frequency part (the source complex envelope  $w(t)$ ) and a high-frequency contribution  $m(t)$

$$w^{(i)}(t) \approx h^{(i)}(t) * w(t) m(t)$$

This corresponds to the base-band complex signal

$$w_b^{(i)}(t) \approx h_b^{(i)}(t) * w(t) M_+ \exp\{-j2\pi\Delta f_i t\}$$

where  $\exp\{-j2\pi\Delta f_i t\}$  is a frequency shift due to different demodulation for P and S waves from working frequency  $f_i$

and  $f_2$  respectively, and  $M_+$  is the complex amplitude of  $m(t)$ . In the frequency domain, we have then

$$W_b^{(i)}(f) = \mathcal{F}\{w_b^{(i)}(t)\}(f) = H_b^{(i)}(f)M_+ \left( W(f) * \delta_{\Delta_{f_i}}(f) \right)$$

For a generally low-pass transfer function,  $H_b^{(i)}(f)$  can be considered as a constant  $H_b^{(i)}$ ,

$$W_b^{(i)}(f) = H_b^{(i)}M_+ W(f - \Delta_{f_i})$$

Since band-pass filtering around frequency  $f_i$  and a joint base-band translation are equivalent to a base-band translation from frequency  $f_i$  followed by a low-pass filtering with window  $\Pi(f)$ , we have

$$W_{bf}^{(i)}(f) \approx \Pi(f)H_b^{(i)}M_+ W(f - \Delta_{f_i})$$

Then we have in time domain, for ideal low-pass filtering  $\Pi(f) \approx \Pi$  in the support of interest

$$w_{bf}^{(i)}(t) \propto \left( \Pi H_b^{(i)}M_+ \right) w(t)$$

provided that the source complex envelope has a smooth spectrum (almost constant in the narrowband supports of interest), that is  $W(f - \Delta_{f_i}) \propto W(f - \Delta_{f_2}) \propto W(f)$ .

Finally, resulting P wave and S wave complex envelopes, after base-band translation and filtering, are related by a proportionality relationship

$$w_{bf}^{(2)}(t) \propto \frac{H_b^{(2)}}{H_b^{(1)}} w_{bf}^{(1)}(t) \quad (\text{A.1})$$

## Appendix B. Details of MUSIC and ESPRIT algorithms

### B.1. MUSIC

The analysis of the  $L \times L$  signal covariance matrix  $\mathbf{R}_{YY}$  is the foundation of the MUSIC algorithm [4]

$$\mathbf{R}_{YY} = \mathbb{E}\{\mathbf{Y}\mathbf{Y}^H\} = \mathbf{A}\mathbb{E}\{\mathbf{S}\mathbf{S}^H\}\mathbf{A}^H + \mathbb{E}\{\mathbf{N}\mathbf{N}^H\}$$

Under the assumptions **H2**, **H3** and **H4**,

$$\mathbf{R}_{YY} = \mathbf{A}\mathbf{R}_{SS}\mathbf{A}^H + \sigma_n^2\mathbf{I}$$

Since  $\text{rank } \mathbf{A}\mathbf{R}_{SS}\mathbf{A}^H = R$ , the eigenvalues  $\lambda_1 \geq \dots \geq \lambda_R > \lambda_L$  of  $\mathbf{R}_{YY}$  are

$$\begin{cases} \lambda_i > \sigma_n^2 & \text{for } i = 1, \dots, R \\ \lambda_i = \sigma_n^2 & \text{for } i = R+1, \dots, L \end{cases}$$

If the noise spatial covariance is not proportional to identity, but equal to  $\sigma_n^2\mathbf{G}$  where  $\mathbf{G}$  is known up to the scaling factor  $\sigma_n^2$ , the same reasoning applies where  $\lambda_i$  are generalized eigenvalues of the matrix pencil  $(\mathbf{R}_{YY}, \mathbf{G})$ . Let  $\mathbf{E}_S$  denote the  $L \times R$  matrix whose columns are the  $R$  first (generalized) eigenvectors of  $\mathbf{R}_{YY}$ , and  $\mathbf{E}_N$  the  $L \times (L-R)$  matrix containing the  $(L-R)$  noise eigenvectors:  $\mathbf{E}_S = [\mathbf{e}_1 | \dots | \mathbf{e}_R]$  and  $\mathbf{E}_N = [\mathbf{e}_{R+1} | \dots | \mathbf{e}_L]$ . Thus, the squared Euclidean distance from a vector  $\mathbf{x}$  to the signal subspace  $\text{Span}\{\mathbf{E}_S\} = \text{Span}\{\mathbf{A}\}$  is  $d^2 = \mathbf{x}^H \mathbf{E}_N \mathbf{E}_N^T \mathbf{x}$ . The MUSIC algorithm aims at finding the values of angle of arrival  $\theta$  for which the distance  $\mathbf{a}(\theta)^H \mathbf{E}_N \mathbf{E}_N^T \mathbf{a}(\theta)$  between the array manifold  $\mathcal{A}_\theta$  and the signal subspace is minimized. For this purpose, an exhaustive search is performed on a grid

$\{\theta_i, 1 \leq i \leq \Theta\}$  of arbitrary precision. This is equivalent to maximizing the so-called MUSIC “spectrum”

$$P_{MU}(\theta) = \frac{\mathbf{a}(\theta)^H \mathbf{a}(\theta)}{\mathbf{a}(\theta)^H \mathbf{E}_N \mathbf{E}_N^T \mathbf{a}(\theta)}$$

whose numerator is a normalization factor. Measurements in the presence of signal  $\mathbf{S}$  are used to estimate  $\mathbf{R}_{YY}$

$$\hat{\mathbf{R}}_{YY} = \frac{1}{M} \sum_{m=1}^M \mathbf{y}(t_m) \mathbf{y}^H(t_m) = \frac{1}{M} \mathbf{Y} \mathbf{Y}^H$$

If noise is absent, parameter estimates are given by the intersections of the array manifold  $\mathcal{A}_\theta$  and the signal subspace  $\text{Span}\{\mathbf{E}_S\}$ . In the presence of noise, there are no intersections with probability one [6]: parameter estimates are thus given by the  $R$  largest peaks of the MUSIC spectrum.

Under assumptions **H3**, **H4** and **H5**, the generic CRB for angles of arrival  $\boldsymbol{\theta} = [\theta_1, \dots, \theta_R]$  and noise covariance  $\sigma_n^2 \mathbf{I}$  is given by [26], if noise is circular Gaussian

$$\text{CRB}(\boldsymbol{\theta}) = \frac{\sigma_n^2}{2} \left\{ \sum_{m=1}^M \Re[\mathbf{S}^H(t_m) \mathbf{D}^H [\mathbf{I} - \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H] \mathbf{D} \mathbf{S}(t_m)] \right\}^{-1}, \quad (\text{B.1})$$

where

$$\mathbf{S}(t_m) = \begin{bmatrix} w_1(t_m) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & w_R(t_m) \end{bmatrix} \in \mathbb{C}^{R \times R}$$

$$\mathbf{D} = [d(\theta_1), \dots, d(\theta_R)] \in \mathbb{C}^{L \times R}$$

whose  $r$ -th element is expressed as the derivative of steering vectors with respect to directional elements

$$d(\theta_r) = \frac{d\mathbf{a}(\theta_r)}{d\theta_r} \in \mathbb{C}^{L \times 1}$$

### B.2. ESPRIT

The ESPRIT algorithm reduces the computational complexity, by exploiting the structure of the sensor array when available [6]: the array is composed of two identical known subarrays (here denoted with letters  $\mathbf{x}$  and  $\mathbf{y}$ ) displaced from each other by a (possibly unknown) displacement vector  $\boldsymbol{\delta}$ . Around frequency  $f_0$ , the observation model is

$$\begin{cases} \mathbf{x}(t_m) = \mathbf{A}\mathbf{s}(t_m) + \mathbf{n}_x(t_m), & m = 1, \dots, M \\ \mathbf{y}(t_m) = \boldsymbol{\Phi}\mathbf{A}\mathbf{s}(t_m) + \mathbf{n}_y(t_m), & m = 1, \dots, M \end{cases}$$

where, in the ULA configuration,  $\boldsymbol{\Phi} = \text{Diag}\{e^{j2\pi f_0 \boldsymbol{\delta} \sin \theta_1 / c}, \dots, e^{j2\pi f_0 \boldsymbol{\delta} \sin \theta_R / c}\} \in \mathbb{C}^{R \times R}$  is a unitary operator that relates the measurements from subarray  $\mathbf{X} \in \mathbb{C}^{(L-1) \times M}$  to those from subarray  $\mathbf{Y} \in \mathbb{C}^{(L-1) \times M}$ . The joint output vector is then defined as

$$\mathbf{z}(t_m) = \begin{bmatrix} \mathbf{x}(t_m) \\ \mathbf{y}(t_m) \end{bmatrix} = \bar{\mathbf{A}}\mathbf{s}(t_m) + \mathbf{n}_z(t_m)$$

$$\bar{\mathbf{A}} = \begin{bmatrix} \mathbf{A} \\ \boldsymbol{\Phi}\mathbf{A} \end{bmatrix}, \quad \mathbf{n}_z(t_m) = \begin{bmatrix} \mathbf{n}_x(t_m) \\ \mathbf{n}_y(t_m) \end{bmatrix} \quad (\text{B.2})$$

The signal subspace  $\text{Span}\{\mathbf{E}_S\}$  is obtained by the first  $R$  eigenvectors of the covariance of the measurements  $\mathbf{R}_{ZZ} = \mathbf{A}\mathbf{R}_{SS}\mathbf{A}^T + \sigma^2\mathbf{I}$ . From Eq. (B.2)  $\mathbf{E}_S$  can be further decomposed in the couple

$$\mathbf{E}_S = \begin{bmatrix} \mathbf{E}_X \\ \mathbf{E}_Y \end{bmatrix}$$

The solution to the problem  $\Psi\mathbf{E}_X = \mathbf{E}_Y$  is given by a Least Squares (LS) solution

$$\hat{\Phi} = \mathbf{E}_X^\dagger \mathbf{E}_Y$$

or by Total Least Squares (TLS) solution

$$[\mathbf{E}_X + \mathbf{Q}_X]\hat{\Phi} = [\mathbf{E}_Y + \mathbf{Q}_Y]$$

where residual matrices  $\mathbf{Q}_X$  and  $\mathbf{Q}_Y$  have minimum Frobenius norm and the eigenvalues of  $\Psi$  are equal to the diagonal elements of  $\Phi$ .

### Appendix C. Cramér–Rao bound of the DoA estimation

The parameter vector of the general model of  $R$  sources impinging on the array from angles of arrival  $\theta = [\theta_1, \dots, \theta_R]$  is expressed by

$$\begin{aligned} \theta &= [\theta_1, \dots, \theta_R; \mathbf{b}_1^T, \dots, \mathbf{b}_R^T, \bar{\mathbf{c}}_1^T, \dots, \bar{\mathbf{c}}_R^T; \mathbf{b}_1^H, \dots, \bar{\mathbf{c}}_R^H] \\ &= [\theta; \xi; \xi^*] \end{aligned}$$

with  $\xi = [\mathbf{b}_1^T, \dots, \mathbf{b}_R^T, \bar{\mathbf{c}}_1^T, \dots, \bar{\mathbf{c}}_R^T]$ . In particular, if the noise is circularly complex, the parameter vector and the FIM simplify to [20]

$$\theta = [\theta; \xi]$$

and

$$\Phi = \frac{1}{\sigma_n^2} \begin{pmatrix} 2\Re\{\mathbf{G}_{11}\} & \mathbf{G}_{12} & \mathbf{G}_{12}^* \\ \mathbf{G}_{12}^H & \mathbf{G}_{22} & \mathbf{0} \\ \mathbf{G}_{12}^T & \mathbf{0} & \mathbf{G}_{22}^* \end{pmatrix}$$

where  $\mathbf{G}_{ij} = \left(\frac{\partial \mu}{\partial \theta_i}\right)^H \left(\frac{\partial \mu}{\partial \theta_j}\right)$ ,  $(i, j) \in \{1, 2\} \times \{1, 2\}$ ,  $\theta_1 = \theta$  and  $\theta_2 = \xi$ . Using the chain rule and complex derivative formula (cf. [20]),

$$\begin{aligned} \frac{\partial \mu}{\partial \theta_r} &= \left(\frac{\partial \mu}{\partial \mathbf{a}_r^T}\right) \left(\frac{\partial \mathbf{a}_r}{\partial \theta_r}\right) \\ &= i\frac{2\pi}{\lambda} \cos \theta_r (\mathbf{I}_L \otimes \mathbf{b}_r \otimes \mathbf{c}_r)(\mathbf{a}_r \boxtimes \mathbf{v}_L) \equiv \phi_{\theta_r} \end{aligned}$$

where  $\mathbf{v}_L = [0, 1, \dots, L-1]$ . Thus,

$$\frac{\partial \mu}{\partial \theta} = [\phi_{\theta_1}, \dots, \phi_{\theta_R}] \in \mathbb{C}^{LMK \times R}$$

As for nuisance cross-terms of the FIM,

$$\frac{\partial \mu}{\partial \mathbf{b}_r^T} = (\mathbf{a}_r \otimes \mathbf{I}_{MK})(\mathbf{I}_M \otimes \mathbf{c}_r)\mathbf{J}_M \equiv \phi_{b_r} \in \mathbb{C}^{LMK \times (M-1)}$$

$$\frac{\partial \mu}{\partial \bar{\mathbf{c}}_r^T} = \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{I}_K \equiv \phi_{\bar{c}_r} \in \mathbb{C}^{LMK \times K}$$

with selection matrix  $\mathbf{J}_M = [\mathbf{0}_{(M-1),1} \mathbf{I}_{(M-1)}]$ . Thus,

$$\frac{\partial \mu}{\partial \xi} = [\phi_{b_1}, \dots, \phi_{b_R}, \phi_{\bar{c}_1}, \dots, \phi_{\bar{c}_R}] \in \mathbb{C}^{LMK \times R(K+M-1)}$$

Once the FIM is calculated, it can be inverted, so that the first leading  $R \times R$  block in matrix  $\Phi^{-1}$  corresponds to the CRB of unbiased DoA estimators.

### Appendix D. Estimation of DoA

The Total Least Squares (TLS) solution is given by the minimization problem

$$\hat{u}_r^{\text{TLS}} = \min_{\alpha, \beta} \|\alpha \underline{\mathbf{a}} - \beta \bar{\mathbf{a}}\|^2, \quad \text{s.t. } \|\alpha \beta\| = 1$$

where  $\alpha \underline{\mathbf{a}} - \beta \bar{\mathbf{a}} = \begin{bmatrix} \underline{\mathbf{a}} \\ \beta \end{bmatrix} - \begin{bmatrix} \bar{\mathbf{a}} \\ \alpha \end{bmatrix}$ . Thus,

$$\hat{u}_r^{\text{TLS}} = \min_{\alpha, \beta} [\alpha^* \beta^*] \begin{bmatrix} \|\underline{\mathbf{a}}\|^2 & -\underline{\mathbf{a}}^H \bar{\mathbf{a}} \\ -\bar{\mathbf{a}}^H \underline{\mathbf{a}} & \|\bar{\mathbf{a}}\|^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad (\text{D.1})$$

The corresponding solution is given by  $\hat{u}_r = -V_{12}V_{22}^{-1}$ , where matrix  $\mathbf{V}$  derives from the SVD of augmented matrix  $[\underline{\mathbf{a}} \ \bar{\mathbf{a}}] = \mathbf{U}\mathbf{S}\mathbf{V}^T$ .

Once  $\hat{u}_r$  is computed, another operation is necessary to estimate the  $r$ -th angle of arrival. The latter can be obtained as  $\hat{\theta}_r = \arcsin[\frac{\Omega}{\pi} \text{angle}(\hat{u}_r)]$ , where  $\Omega = \frac{\lambda}{2\Delta} \geq 1$  denotes the oversampling factor, and  $\text{angle}(\cdot)$  is the phase determination in  $[-\pi, \pi]$ .

**Remark 1.** If  $\hat{u}_r$  is badly estimated, it may happen that  $|\sin \theta_r|$  is larger than 1. This occurs if  $\text{angle}(\hat{u}_r)$  falls outside the interval  $[-\pi/\Omega, \pi/\Omega]$ . Should this be the case, for large  $L$  it can be helpful to redefine  $\underline{\mathbf{a}}$  and  $\bar{\mathbf{a}}$  as

$$\begin{aligned} \bar{\mathbf{a}} &= (1 \quad u_r \quad \dots \quad u_r^{L-3})^T \\ \underline{\mathbf{a}} &= (u_r^2 \quad u_r^3 \quad \dots \quad u_r^{L-1})^T \end{aligned}$$

The relation between  $\hat{u}_r$  and  $\hat{\theta}_r$  then changes and  $\text{angle}(\hat{u}_r)$  just needs to belong to the interval  $[-2\pi/\Omega, 2\pi/\Omega]$ , and the value of  $|\sin \theta_r|$  will be smaller than 1 as long as  $\Omega \leq 2$ , that is, as long as  $\Delta \geq \lambda/4$ .

### Appendix E. Preprocessing of seismic data

A time series  $x(t)$  can be considered a *pass-band* signal if its spectral support is limited and it does not include the origin:

$$\begin{aligned} &[-f_0 - W, -f_0 + W] \\ &\cup [f_0 - W, f_0 + W] \quad \text{with } \infty > f_0 > W > 0 \end{aligned}$$

Every *pass-band* signal can be associated with an *analytic* signal  $\hat{x}(t)$  whose support does not contain positive frequencies:

$$\hat{X}(f) = \sqrt{(2)} U_+(f) X(f) \longleftrightarrow \hat{x}(t) = \frac{1}{2} [x(t) + iH\{x(t)\}]$$

where  $U_+(\cdot)$  indicates the Heaviside step function, and  $H\{x(t)\}$  refers to the Hilbert transform of real signal  $x(t)$ .

Since a real record is characterized by even spectral symmetry, it can be represented by its analytical signal without loss of information.

The complex envelope of  $X(f)$  around frequency  $f_0$  is obtained from the *base-band* analytical signal by a mere translation in frequency

$$\tilde{X}(f) = \hat{X}(f + f_0) \longleftrightarrow \tilde{x}(t) = \hat{x}(t)e^{-i2\pi f_0 t}$$



For a given carrier at frequency  $f_0$ , a complex envelope around  $f_0$  is in bijection with a complex number representing the modulus and the phase of the carrier.

Thus, recorded data have to be filtered in order to satisfy the narrowband assumption **A2**. However, beside filtering, we did not deconvolve the signals from any instrument response, because the frequency content of the icequakes is much higher than the resonance frequency of the velocimeters (2 Hz). We thus consider that the instrument response is flat.

Moreover, an approximation has to be made with respect to the spectral content of signals, in case the ratio of the dominant frequencies does not reflect the ratio of the velocities of P and S waves, according to property **P2**. In order to maintain a constant wavelength  $\lambda$ , which is a necessary condition for separability of the multilinear model, a compromise needs to be reached in extracting the dominant frequencies, as a consequence of the ideal condition  $\lambda = \frac{v_1}{f_1} = \frac{v_2}{f_2}$  in property **P2**. Within the set of four parameters determining wavelength,  $\{v_1, v_2, f_1, f_2\}$  one only needs 3 degrees of freedom to determine optimal working conditions. P and S wave propagation velocities  $v_1$  and  $v_2$  are given by geophysical active analysis of known reflecting waves from the surface to the glacier bed [24,33]. The only parameter subject to optimization is then P or S wave frequency  $f_1$  or  $f_2$ . If  $f_1$  is chosen as the free parameter and optimized, then  $f_2$  directly derives from property **P2**

$$f_2 = \frac{C_2}{C_1} f_1 = \frac{C_2}{\lambda} \quad (\text{E.1})$$

Optimization is fulfilled by minimizing the Frobenius norm of tensor decomposition error

$$\hat{f}_1 = \underset{f_1}{\operatorname{argmin}} \left\| \mathbf{z}(f_1) - \sum_{r=1}^R \lambda_r \mathbf{a}_r(f_1) \boxtimes \mathbf{b}_r \boxtimes \mathbf{c}_r(f_1) \right\|^2$$

where  $\mathbf{z}(f_1)$  is the data array after narrowband filtering around  $f_1$  and  $f_1 \in (0 \text{ Hz}, f_s/2 = 500 \text{ Hz})$ , given sampling frequency  $f_s$ . Then, from Eq. (E.1), we have  $\hat{f}_2 = \frac{v_2}{v_1} \hat{f}_1$ .

## References

- [1] S. Rost, C. Thomas, Array seismology: methods and applications, *Rev. Geophys.* 40 (3) (2002) 2–11.
- [2] W. Liu, S. Weiss, Wideband Beamforming: Concepts and Techniques, vol. 17, John Wiley & Sons, Chichester, West Sussex, UK, 2010.
- [3] H. Krim, M. Viberg, Two decades of array signal processing research: the parametric approach, *IEEE Signal Process. Mag.* 13 (4) (1996) 67–94.
- [4] R.O. Schmidt, Multiple emitter location and signal parameter estimation, *IEEE Trans. Antenna Propag.* 34 (3) (1986) 276–280.
- [5] A. Barabell, Improving the resolution performance of eigenstructure-based direction-finding algorithms, in: ICASSP, vol. 8, Boston, 1983, pp. 336–339.
- [6] R. Roy, T. Kailath, ESPRIT-estimation of signal parameters via rotational invariance techniques, *IEEE Trans. Acoust. Speech Signal Process.* 37 (7) (1989) 984–995.
- [7] B. D. Van Veen, K. M. Buckley, Beamforming: A versatile approach to spatial filtering, *IEEE Acoust. Speech Signal. Process. Mag.* 5 (2) (1988) 4–24.
- [8] H. Wang, 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.* 33 (4) (1985) 823–831.
- [9] H. Clergeot, O. Michel, New simple implementation of the coherent signal subspace method for wide band direction of arrival estimation, in: ICASSP, Glasgow, 1989, pp. 2764–2767.
- [10] N.D. Sidiropoulos, R. Bro, G.B. Giannakis, Parallel factor analysis in sensor array processing, *IEEE Trans. Signal Process.* 48 (8) (2000) 2377–2388.
- [11] L.-H. Lim, P. Comon, Blind multilinear identification, *IEEE Trans. Inf. Theory* 60 (2) (2014) 1260–1280, open access. hal-00763275.
- [12] P. Comon, Tensors: a brief introduction, *IEEE Signal Process. Mag.* 31 (2014) 44–53.
- [13] E. Bedrosian, A product theorem for Hilbert transforms, Research Memorandum, RM-3439-PR, Rand Corporation.
- [14] S. Anderson, A. Nehorai, Analysis of a polarized seismic wave model, *IEEE Trans. Signal Process.* 44 (2) (1996) 379–386.
- [15] P.J. Oonincx, Automatic phase detection in seismic data using the discrete wavelet transform, CWI Technical Report PNA-R 9811 (1998).
- [16] R.E. Sheriff, L.P. Geldart, Exploration seismology, in: History, Theory, and Data Acquisition, vol. 1, Cambridge University Press, New York, NY, 1983.
- [17] X. Guo, S. Miron, D. Brie, S. Zhu, X. Liao, A Candecom/Parafac perspective on uniqueness of DoA estimation using a vector sensor array, *IEEE Trans. Signal Process.* 59 (7) (2011) 3475–3481.
- [18] M. Niknazar, H. Becker, B. Rivet, C. Jutten, P. Comon, Blind source separation of underdetermined mixtures of event-related sources, *Signal Process.* 101 (2014) 52–64 (hal-00952039).
- [19] L. Chiantini, G. Ottaviani, N. Vannieuwenhoven, An algorithm for generic and low-rank specific identifiability of complex tensors, *SIAM J. Matrix Anal. Appl.* 35 (4) (2014) 1265–1287.
- [20] S. Sahnoun, P. Comon, Tensor polyadic decomposition for antenna array processing, in: The 21st International Conference on Computational Statistics (CompStat), Geneva, 2014, pp. 233–240, hal-00986973.
- [21] X. Liu, N.D. Sidiropoulos, Cramér–Rao lower bounds for low-rank decomposition of multidimensional arrays, *IEEE Trans. Signal Process.* 49 (9) (2001) 2074–2086.
- [22] S. Leurgans, R. Ross, R. Abel, A decomposition for three-way arrays, *SIAM J. Matrix Anal. Appl.* 14 (4) (1993) 1064–1083.
- [23] J. B. Kruskal, Three-way arrays: Rank and uniqueness of trilinear decompositions, *Linear Algebra Appl.* 18 (1977) 95–138.
- [24] A. Helmstetter, B. Nicolas, P. Comon, M. Gay, Basal icequakes recorded beneath an Alpine glacier (Glacier d'Argentière, Mont Blanc, France): evidence for stick-slip motion? *J. Geophys. Res.* 120 (3) (2015) 379–401.
- [25] A. Helmstetter, B. Nicolas, P. Comon, M. Gay, Intermediate-depth icequakes and harmonic tremor in an alpine glacier (Glacier d'Argentière, Mont Blanc, France): evidence for hydraulic fracturing? *J. Geophys. Res.* 120 (3) (2015) 402–416.
- [26] P. Stoica, A. Nehorai, MUSIC, maximum likelihood, and Cramér–Rao bound, *IEEE Trans. Acoust. Speech Signal Process.* 37 (5) (1989) 720–741.
- [27] M. Vallon, Contribution à l'étude de la Mer de Glace - Alpes françaises (Ph.D. thesis), Faculté des Sciences de l'Université de Grenoble, May 1967.
- [28] H.L. Van Trees, Detection, Estimation, and Modulation Theory, Optimum Array Processing, John Wiley & Sons, New York, USA, 2004.
- [29] E. Chaumette, J. Galy, A. Quinlan, P. Larzabal, A new Barankin bound approximation for the prediction of the threshold region performance of maximum likelihood estimators, *IEEE Trans. Signal Process.* 56 (11) (2008) 5319–5333.
- [30] P.-F. Roux, Méthodes sismologiques pour l'étude de la fracturation dans les glaciers alpins: glaciers d'Argentière et du Gorner (Ph.D. thesis), Université de Savoie, 2008.
- [31] P.-F. Roux, D. Marsan, J.-P. Métaixian, G. O'Brien, L. Moreau, Micro-seismic activity within a serac zone in an alpine glacier (Glacier d'Argentière, Mont Blanc, France), *J. Glaciol.* 54 (184) (2008) 157–168.
- [32] C. Vincent, A. Soruco, D. Six, E. le Meur, Glacier thickening and decay analysis from 50 years of glaciological observations performed on glacier d'argentière, mont blanc area, france, *Ann. Glaciol.* 50 (50) (2009) 73–79.
- [33] F. Walter, N. Deichmann, M. Funk, Basal icequakes during changing subglacial water pressures beneath Gornergletscher, Switzerland, *J. Glaciol.* 54 (186) (2008) 511–521.