Brain Source Localization Using a Fourth-Order Deflation Scheme.

Laurent Albera, Anne Ferréol, Delphine Cosandier-Rimélé, Isabel Merlet, Fabrice Wendling

Brain source localization using a fourth order deflation scheme

Laurent Albera, *member, IEEE*, Anne Ferréol, Delphine Cosandier-Rimélé, Isabelle Merlet, Fabrice Wendling

***Abstract*—A high resolution method for solving potentially ill-posed inverse problems is proposed. This method named FO-D-MUSIC allows for localization of brain current sources with unconstrained orientations from surface electro- or magneto-encephalographic data using spherical or realistic head geometries. The FO-D-MUSIC method is based on i) the separability of the data transfer matrix as a function of location and orientation parameters, ii) the Fourth Order (FO) virtual array theory, and iii) the deflation concept extended to FO statistics accounting for the presence of potentially but not completely statistically dependent sources. Computer results display the superiority of the FO-D-MUSIC approach in different situations (very closed sources, small number of electrodes, additive Gaussian noise with unknown spatial covariance, ...) compared to classical algorithms.**

***Index Terms*—Sequential Source Localization, Backward Problem, EEG, MEG, MUSIC, Fourth Order Statistics.**

I. Introduction

ELECTROENCEPHALOGRAPHY (EEG) and magneto-encephalography (MEG) are two complementary techniques measuring, at the surface of the head, electrical potentials and magnetic fields produced by neuronal activity, respectively. The localization of the sources of this neuronal activity (during either cognitive or pathological processes) requires to solve the inverse problem; i.e. to localize sources only from surface recordings. In the general case, the EEG/MEG inverse problem is an ill-posed and under determined problem, as the number of sources is larger than the number of measurement points. To overcome this difficulty, localization techniques assume a finite number of sources to be localized. When this number is lower than the number of sensors positioned on the scalp, the problem is overdetermined and a unique solution can be obtained. It is noteworthy that the inverse problem is not specific to the field of neurophysiology, but it can be found in many other areas such as digital radiocommunications [12].

The solution of the inverse problem implies that a model of sources and a model of volume conductor are defined. In the study of cerebral activity, the current dipole is the most commonly used model for a source of electrical activity in the brain, as it is a biophysically-relevant representation of a small cortical area activity. As neuronal electromagnetic fields are sensitive to geometrical and electrical properties of the different head tissues (brain, bone, skin), the head can be modeled either by a set of nested concentric spheres with homogeneous and isotropic conductivities [20], or by realistically shaped models built from 3D anatomical data (Magnetic Resonance Imaging, MRI), with refined tissue conductivity values [10].

During the last three decades, many array processing methods were developed to estimate multidimensional parameters of sources such as localization parameters. In particular, among subspace approaches, the Second Order (SO) MUSIC (MUltiple SIgnal Classification) method [21] [22], can localize intracerebral sources in overdetermined contexts. Several variants were then proposed to improve the MUSIC performances.

On the one hand, Time MUSIC-like methods were reported, such as the extension of the original MUSIC algorithm to Fourth Order (FO) statistics proposed by Porat et al. [19]. The particularity of this algorithm is to deal with the case of underdetermined source mixtures. Among Time MUSIC-like algorithms, sequential approaches [17] [25] [15] should be mentioned. They are based both on the SO statistics and the deflation concept introduced to increase localization resolution. The RapMUSIC algorithm [15], a sequential method based on Ferrara's works [7], is of particular interest. This method takes advantage of the factored matrix formulation of the transfer relationship between the deep sources and the scalp data to reduce computing time by separating quasilinear from non-linear source parameter estimation.

On the other hand, Time-Frequency (TF) approaches were proposed as reported for example by Sekihara et al. [23] and Belouchrani et al. [3]. Their objective was to improve the resolution of the localization in the case of very closed sources with *spectral non-stationary* properties. Besides subspace methods, other localization methods applied to EEG and MEG data were reported. Readers may refer to the recent review by Michel et al. [14] for details.

In practice, the physiological signals of interest have non-zero higher order statistics. Nevertheless, most of the aforementioned array processing methods are based only on SO statistics. Therefore, they are restrictive and suboptimal as they do not take advantage of the information available at higher orders. Moreover, TF approaches are not useful for sources with identical TF supports. Besides, Time SO techniques cannot deal with underdetermined mixtures of sources or with a Gaussian noise of unknown coherence. Higher Order (HO) methods inherently account for these limitations. However, to date, there is no attempt to propose a FO method taking

L. Albera, D. Cosandier-Rimélé, I. Merlet and F. Wendling are with INSERM, U642, Rennes, F-35000, France; and with Université de Rennes 1, LTSI, Rennes, F-35000, France. (email:laurent.albera@univ-rennes1.fr). A. Ferréol is with the THALES Communications group, 146 Boulevard de Valmy, BP 82, F-92704 Colombes, France.

This work was supported in part by the Regional Council of Brittany (http://www.region-bretagne.fr) and it is protected by a patent intitled "*Procédé et dispositif d'identification de paramètres multidimensionnels : application à la localisation et la reconstruction d'activités électriques de profondeur au moyen d'observations de surface*", whose reference is WO 2007/057453.

advantage of the separability of the matrix transfer function between the input and output data and of the deflation concept.

The intent of the present paper is to describe a new FO MUSIC-like method addressing these issues. This method, referred to as FO-D-MUSIC, is based on i) the separability of the data transfer matrix as a function of location and orientation parameters and on ii) the FO virtual array theory [5], and accounts for the presence of potentially but not completely statistically dependent sources. Moreover the FO-D-MUSIC method uses the deflation concept which non-trivial extension to FO statistics is also presented in this paper. The paper is organized as follows. Assumptions about the noisy mixture of sourcesare introduced in section II. SO and FO statistic properties are presented in section III. Principles of the proposed algorithm are described in section IV which also provides some identifiability results (section IV-E). Finally, computer experiments are presented in section V.

II. Notations and hypotheses

A. The problem statement

We assume that K realizations of an N-dimensional random vector $\boldsymbol{x}$ are observed. Besides, vector $\boldsymbol{x}$ is given by:

$$\boldsymbol{x} = \boldsymbol{A}(\boldsymbol{\Theta})\ \boldsymbol{s} + \boldsymbol{\nu} \tag{1}$$

where $\boldsymbol{s} = [s_1, \cdots, s_P]^{\mathsf{T}}$ is a P-dimensional random vector, called source vector, which observations correspond to the time courses of the P current dipoles. Matrix $\boldsymbol{A}(\boldsymbol{\Theta}) = [\boldsymbol{a}(\boldsymbol{\theta}_1), \cdots, \boldsymbol{a}(\boldsymbol{\theta}_P)]$ is the $(N{\times}P)$ static mixing matrix, which depends on $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_P\}$, that is, the collection of the P multi-parameters of the sources. As far as the noise vector $\boldsymbol{\nu}$ is concerned, it is assumed to be Gaussian and statistically independent of the source vector. Moreover, some components of vector $\boldsymbol{s}$ can be statistically dependent, i.e. sources can be partially, but not completely correlated (in a wide-sense, at order 2 and 4). So, without loss of generality, it is possible to divide the P sources into J groups, with P_j sources in the j-th group ($1 \leq j \leq J$), in such a way that sources of same group are statistically dependent, while sources in different groups remain statistically independent. In particular, $J = P$ corresponds to P statistically independent sources whereas $J = 1$ corresponds to the case where all the sources are dependent. Of course, the P_j parameters are such that $P = \sum_{j=1}^{J} P_j$. Under these notations, the observation vector $\boldsymbol{x}$ can be rewritten as follows:

$$\boldsymbol{x} = \sum_{j=1}^{J} \boldsymbol{A}(\boldsymbol{\Theta}_j)\, \boldsymbol{s}_j + \boldsymbol{\nu} \tag{2}$$

where $\boldsymbol{A}(\boldsymbol{\Theta}_j)$ is the $(N \times P_j)$ submatrix of $\boldsymbol{A}$ corresponding to the j-th group of sources and $\boldsymbol{s}_j$ is the corresponding P_j-dimensional subvector of $\boldsymbol{s}$. It is noteworthy that the division of the P sources into J groups will be very useful in the following sections to find the identifiability conditions of the FO-D-MUSIC method, that is, the maximal number of sources which can be processed for a given number of observations.

In EEG (or MEG) applications, each source localization vector $\boldsymbol{a}(\boldsymbol{\theta})$ of the static mixing matrix represents electrical potential differences (or magnetic fields) generated from surface electrodes by a current dipole with a unit time course localized at a given position $\boldsymbol{\rho}$ for a given orientation $\boldsymbol{\phi}$. Recent empirical work on closed-form approximations for spherical and realistic head geometries (see [16] for more details) allow for the approximation of $\boldsymbol{a}(\boldsymbol{\theta})$ by the product of a $(N{\times}3)$ gain matrix $\boldsymbol{G}(\boldsymbol{\rho})$ and the orientation vector $\boldsymbol{\phi}$:

$$\boldsymbol{a}(\boldsymbol{\theta}) \approx \boldsymbol{G}(\boldsymbol{\rho})\ \boldsymbol{\phi} \tag{3}$$

where the multi-parameter vector $\boldsymbol{\theta} = [\boldsymbol{\rho}^{\mathsf{T}}\ \boldsymbol{\phi}^{\mathsf{T}}]^{\mathsf{T}}$ of the considered dipole includes the non-linear location parameter $\boldsymbol{\rho}$ and the quasilinear orientation parameter $\boldsymbol{\phi}$.

Although the method we developed can be applied to both EEG and MEG data, and to both spherical and realistic head models, the following results will be presented in the EEG context using a spherical head model. In other words the observed data are assumed, in the sequel, to be electrical potentials. The head is represented by three nested concentric spheres (brain, skull, and scalp), with conductivities chosen as constant and isotropic.

B. Gain matrix in an EEG context with spherical head model

In the case of a 3-shell spherical head model, the n-th row $\boldsymbol{G}_n(\boldsymbol{\rho})$ of the gain matrix $\boldsymbol{G}(\boldsymbol{\rho})$, has the following expression [4] [16]:

$$\boldsymbol{G}_n(\boldsymbol{\rho}) = \sum_{j=1}^{3} \lambda_j \left(\boldsymbol{h}(\boldsymbol{r}_1, \mu_j \boldsymbol{\rho}) - \boldsymbol{h}(\boldsymbol{r}_{N+1}, \mu_j \boldsymbol{\rho})\right)^{\mathsf{T}} \tag{4}$$

The $(N{+}1)$-th electrode is used as single common reference in order to compute N potential differences from the potentials recorded at the N other electrode locations. The (3×1) vector $\boldsymbol{h}(\boldsymbol{r}, \boldsymbol{\rho})$ is given by [16]:

$$\boldsymbol{h}(\boldsymbol{r}, \boldsymbol{\rho}) = \frac{\left(c_1(\boldsymbol{r}, \boldsymbol{\rho}) - c_2(\boldsymbol{r}, \boldsymbol{\rho})\, \boldsymbol{r}^{\mathsf{T}}\boldsymbol{\rho}\right) \boldsymbol{\rho} + c_2(\boldsymbol{r}, \boldsymbol{\rho})\, \|\boldsymbol{\rho}\|^2\, \boldsymbol{r}}{4\pi\sigma_3\, \|\boldsymbol{\rho}\|^2} \tag{5}$$

where σ_3 is the conductivity of the outermost layer of the 3-sphere head model, and parameters $c_1(\boldsymbol{r}, \boldsymbol{\rho})$ and $c_2(\boldsymbol{r}, \boldsymbol{\rho})$ are defined by:

$$\begin{aligned} c_1(\boldsymbol{r}, \boldsymbol{\rho}) &= 2\frac{(\boldsymbol{r}-\boldsymbol{\rho})^{\mathsf{T}}\boldsymbol{\rho}}{\|\boldsymbol{r}-\boldsymbol{\rho}\|^3} + \frac{1}{\|\boldsymbol{r}-\boldsymbol{\rho}\|} - \frac{1}{\|\boldsymbol{r}\|} \\ c_2(\boldsymbol{r}, \boldsymbol{\rho}) &= \frac{2}{\|\boldsymbol{r}-\boldsymbol{\rho}\|^3} + \frac{\|\boldsymbol{r}-\boldsymbol{\rho}\| + \|\boldsymbol{r}\|}{\|\boldsymbol{r}\|\|\boldsymbol{r}-\boldsymbol{\rho}\|\left(\|\boldsymbol{r}\|\|\boldsymbol{r}-\boldsymbol{\rho}\| + \|\boldsymbol{r}\|^2 - \boldsymbol{\rho}^{\mathsf{T}}\boldsymbol{r}\right)} \end{aligned} \tag{6}$$

Constants $\{\lambda_j\}_{1\leq j\leq 3}$ and $\{\mu_j\}_{1\leq j\leq 3}$ – the so-called "Berg parameters" [4] – are only dependent on the 3-sphere head model radii $\{R_j\}_{1\leq j\leq 3}$ and conductivities $\{\sigma_j\}_{1\leq j\leq 3}$. They should be fitted numerically by minimizing the right side of equation (5i") given by Zhang [27]. For instance, we computed the "Berg parameters" for specific radii [20] and conductivities [24] values, as shown in table I.

R_1 (cm)	R_2	R_3	σ_1 (S/cm)	σ_2	σ_3
8	8.5	9.2	3.3e-3	8.25e-5	3.3e-3
λ_1	λ_2	λ_3	μ_1	μ_2	μ_3
0.5979	0.2037	0.0237	0.6342	0.9364	1.0362

TABLE I

The "Berg parameters" for a specific 3-sphere head model

III. SO AND FO STATISTICS

A. Moments and cumulants

Recall that the first characteristic function $\Psi_{\boldsymbol{x}}^{(1)}$ of a random vector $\boldsymbol{x}$ always exists, is continuous and is defined by

$$\Psi_{\boldsymbol{x}}^{(1)}(\boldsymbol{u}) = \mathsf{E}[\exp(\mathrm{i}\boldsymbol{u}^{\mathsf{T}}\boldsymbol{x})] \tag{7}$$

where $\mathsf{E}[z]$ denotes the mathematical expectation of z. Since $\Psi_{\boldsymbol{x}}^{(1)}(\mathbf{0})=1$ and $\Psi_{\boldsymbol{x}}^{(1)}$ is continuous, then a small neighbourhood $\mathcal{U}$ of $\mathbf{0}$ exists, in which $\Psi_{\boldsymbol{x}}^{(1)}$ does not vanish. Denoting log the principal branch of the logarithm in the right half plane, we define the second characteristic function $\Psi_{\boldsymbol{x}}^{(2)}$ by

$$\forall \boldsymbol{u}\in\mathcal{U}, \qquad \Psi_{\boldsymbol{x}}^{(2)}(\boldsymbol{u}) = \log\Big(\Psi_{\boldsymbol{x}}^{(1)}(\boldsymbol{u})\Big) \tag{8}$$

Moments are the coefficients of the expansion of the first characteristic function $\Psi_{\boldsymbol{x}}^{(1)}$ about the origin, and *cumulants* are those of the second characteristic function $\Psi_{\boldsymbol{x}}^{(2)}$. More precisely, one defines the entries of r-th order moment and cumulant tensors of $\boldsymbol{x}$, respectively, as

$$M_{n_1,n_2,\cdots,n_r,\boldsymbol{x}} \overset{\text{def}}{=} (-\mathrm{i})^r \left.\frac{\partial^r \Psi_{\boldsymbol{x}}^{(1)}(\boldsymbol{u})}{\partial u_{n_1}\partial u_{n_2}\cdots\partial u_{n_r}}\right|_{\boldsymbol{u}=\mathbf{0}} \tag{9}$$

and

$$C_{n_1,n_2,\cdots,n_r,\boldsymbol{x}} \overset{\text{def}}{=} (-\mathrm{i})^r \left.\frac{\partial^r \Psi_{\boldsymbol{x}}^{(2)}(\boldsymbol{u})}{\partial u_{n_1}\partial u_{n_2}\cdots\partial u_{n_r}}\right|_{\boldsymbol{u}=\mathbf{0}} \tag{10}$$

Note that, using the Leonov-Shiryaev formula [13], it is possible to relate cumulant tensors to moment tensors; in particular, the r-th order cumulant tensor is related to moment tensors of order smaller than or equal to r. For instance, SO and FO cumulant tensors of a zero-mean random vector $\boldsymbol{x}$ can be computed from moment tensors of $\boldsymbol{x}$ in the following way:

$$\begin{aligned} C_{n_1,n_2,\boldsymbol{x}} &= M_{n_1,n_2,\boldsymbol{x}} \\ C_{n_1,n_2,n_3,n_4,\boldsymbol{x}} &= M_{n_1,n_2,n_3,n_4,\boldsymbol{x}} - M_{n_1,n_2,\boldsymbol{x}}M_{n_3,n_4,\boldsymbol{x}} \\ &\quad - M_{n_1,n_3,\boldsymbol{x}}M_{n_2,n_4,\boldsymbol{x}} - M_{n_1,n_4,\boldsymbol{x}}M_{n_2,n_3,\boldsymbol{x}} \end{aligned} \tag{11}$$

However, in practice, moments of the data are not exactly calculable and have to be estimated from L samples of data, in a way that is completely described in [1, section III-D] and which is not recalled here.

B. Moment and cumulant properties

Moment and cumulant tensors of a real random vector are *symmetric* since they are invariant under arbitrary index permutations.

Another important property of cumulants is that if at least two variables or groups of variables are statistically independent, then all cumulants involving these variables are null. The fact that this property is not shared by moments reinforces the interest in cumulants, especially in order to process more sources than observations, as explained in section IV.

For the sake of convenience, cumulants can be arranged in a symmetric matrix. Indeed, SO and FO cumulants of a random vector $\boldsymbol{x}$ can be standed in a $(N\times N)$ symmetric matrix $\boldsymbol{R_x}$, called *covariance matrix*, and in a $(N^2\times N^2)$ symmetric matrix $\boldsymbol{Q_x}$, called *quadricovariance matrix*, respectively:

$$\begin{aligned} R_{\boldsymbol{x}}(n_1,n_2) &= C_{n_1,n_2,\boldsymbol{x}} \\ Q_{\boldsymbol{x}}(N(n_1-1)+n_2, N(n_3-1)+n_4) &= C_{n_1,n_2,n_3,n_4,\boldsymbol{x}} \end{aligned} \tag{12}$$

where $R_{\boldsymbol{x}}(i,j)$ and $Q_{\boldsymbol{x}}(i,j)$ correspond to the (i,j)-th component of $\boldsymbol{R_x}$ and $\boldsymbol{Q_x}$ respectively. It is noteworthy that, contrary to the covariance matrix, the quadricovariance matrix is not generally positive definite.

Now, let's consider vector $\boldsymbol{x}$ given by model (1) where the noise vector $\boldsymbol{\nu}$ is assumed to be Gaussian and statistically independent of the source vector $\boldsymbol{s}$. Then the *multilinearity* property enjoyed by cumulants [13] gives the following relation between cumulants of $\boldsymbol{x}$ and cumulants of $\boldsymbol{s}$:

$$\begin{aligned} \boldsymbol{R_x} &= \textstyle\sum_{j=1}^{J} \boldsymbol{A}(\boldsymbol{\Theta}_j)\,\boldsymbol{R}_{\boldsymbol{s}_j}\boldsymbol{A}(\boldsymbol{\Theta}_j)^{\mathsf{T}} + \boldsymbol{R_\nu} \\ \boldsymbol{Q_x} &= \textstyle\sum_{j=1}^{J} (\boldsymbol{A}(\boldsymbol{\Theta}_j)\otimes\boldsymbol{A}(\boldsymbol{\Theta}_j))\boldsymbol{Q}_{\boldsymbol{s}_j}(\boldsymbol{A}(\boldsymbol{\Theta}_j)\otimes\boldsymbol{A}(\boldsymbol{\Theta}_j))^{\mathsf{T}} \end{aligned} \tag{13}$$

where $\boldsymbol{R}_{\boldsymbol{s}_j}$ and $\boldsymbol{Q}_{\boldsymbol{s}_j}$ are the $(P_j\times P_j)$ covariance and the $(P_j^2\times P_j^2)$ quadricovariance matrices, respectively, of the source vector $\boldsymbol{s}_j$, and where $\boldsymbol{R_\nu}$ is the noise covariance matrix. For any rectangular matrices $\boldsymbol{G}$ and $\boldsymbol{H}$, of size $(N_G\times P_G)$ and $(N_H\times P_H)$ respectively, the Kronecker matrix product $\boldsymbol{G}\otimes\boldsymbol{H}$ of size $(N_GN_H\times P_GP_H)$ is defined by:

$$\boldsymbol{G}\otimes\boldsymbol{H} = \begin{bmatrix} G(1,1)\boldsymbol{H} & \cdots & G(1,P_G)\boldsymbol{H} \\ \vdots & \ddots & \vdots \\ G(N_G,1)\boldsymbol{H} & \cdots & G(N_G,P_G)\boldsymbol{H} \end{bmatrix} \tag{14}$$

Although matrices $\boldsymbol{R_x}$ and $\boldsymbol{Q_x}$ can be written as a function of the static mixture $\boldsymbol{A}(\boldsymbol{\Theta}_j)$, they have different algebraic structures. Consequently, matrix $\boldsymbol{R_x}$ cannot be simply replaced by $\boldsymbol{Q_x}$ in source localization algorithms based on the covariance matrix. Indeed, given the algebraic structure of the quadricovariance matrix $\boldsymbol{Q_x}$ (13), the extension of a covariance-based method to FO statistics requires to elaborate an algorithm able to fully exploit the quadricovariance structure instead of the covariance one. This is the purpose of the next section.

IV. TOWARDS A FO MUSIC-LIKE APPROACH

A. Some additional assumptions

Besides the assumptions given in section II, the SO MUSIC-like methods, for instance MUSIC [7], S-MUSIC [17] and RapMUSIC [15], need the following hypotheses:

- **H1)** $P<N$;
- **H2)** Matrix $\boldsymbol{A}(\boldsymbol{\Theta})$ has a full rank equal to P;

However, when FO statistics are exploited, assumptions **H1** and **H2** can be replaced by the following ones:

- **A1)** $\forall j,\ 1\le j\le J, \quad P_j < N$;
- **A2)** $\forall j,\ 1\le j\le J$, matrices $\boldsymbol{A}(\boldsymbol{\Theta}_j)^{\otimes 2} \overset{\text{def}}{=} \boldsymbol{A}(\boldsymbol{\Theta}_j)\otimes\boldsymbol{A}(\boldsymbol{\Theta}_j)$ and $\boldsymbol{Q}_{\boldsymbol{s}_j}$ have a full rank equal to P_j^2;
- **A3)** $\mathcal{R}_J \overset{\text{def}}{=} \sum_{j=1}^{J} P_j^2 < N^2$;
- **A4)** The $(N^2\times\mathcal{R}_J)$ matrix $\boldsymbol{B}(\boldsymbol{\Theta}) \overset{\text{def}}{=} [\boldsymbol{A}(\boldsymbol{\Theta}_1)^{\otimes 2}\cdots\boldsymbol{A}(\boldsymbol{\Theta}_J)^{\otimes 2}]$ has a full rank equal to $\mathcal{R}_J$;

where J, P_j and $\boldsymbol{A}(\boldsymbol{\Theta}_j)$ were defined in section II. Note that under assumptions **A1** to **A4**, matrix $\boldsymbol{Q_x}$ given in equation (13) takes the following matrix form:

$$\boldsymbol{Q_x} = \boldsymbol{B}(\boldsymbol{\Theta})\ \boldsymbol{\mathcal{Q}_s}\,\boldsymbol{B}(\boldsymbol{\Theta})^{\mathsf{T}} \tag{15}$$

where $\mathcal{Q}_s$ is the $(\mathcal{R}_J \times \mathcal{R}_J)$ block diagonal matrix constructed from the J source quadricovariance matrices Q_{s_j}. The sparsity level of matrix $\mathcal{Q}_s$ is straight related to the ratio J/P. In particular, for sources that are all statistically dependent ($J = 1$), hypotheses **A1** to **A4** reduce to:

A'1) $P < N$;

A'2) Matrices $B(\Theta) = A(\Theta) \otimes A(\Theta)$ and $\mathcal{Q}_s$ have a full rank equal to P^2;

In such a case, the $(P^2 \times P^2)$ matrix $\mathcal{Q}_s$ is full and equal to the source quadricovariance matrix Q_s. On the other hand, when all the sources are statistically independent ($J = P$), hypotheses **A1** to **A4** reduce to:

A"1) $P < N^2$;

A"2) Matrix $B(\Theta) = A(\Theta) \oslash A(\Theta)$ and $\mathcal{Q}_s$ have a full rank equal to P;

where $\oslash$ denotes the columnwise Kronecker product operator, sometimes referred to as the Khatri-Rao product operator [11] and where the $(P \times P)$ matrix $\mathcal{Q}_s$ is now diagonal. Let's recall that for any rectangular matrices G and H, of size $(N_G \times P)$ and $(N_H \times P)$ respectively, the columns of the $((N_G N_H) \times P)$ matrix $G \oslash H$ are defined as $g_p \otimes h_p$, where g_p and h_p denote the columns of G and H respectively:

$$G \oslash H = [g_1 \otimes h_1 \;\; g_2 \otimes h_2 \;\; \cdots \;\; g_P \otimes h_P] \tag{16}$$

According to assumption **A'2** and the FO MUSIC metric, FO-D-MUSIC requires that $P^2 < N^2$ when all sources are statistically dependent, which means that $P < N$ since P and N are positive, hence assumption **A"2**. Consequently, in such a case, FO-D-MUSIC, as SO MUSIC-like methods, cannot process underdetermined mixtures of sources. However, when some sources are statistically independent, and more particularly when all sources are independent, FO-D-MUSIC may process underdetermined source mixtures, according to assumptions **A3** and **A"1**, respectively. Indeed, proof is given in section IV-D for P independent sources. In brief, statistical independence implies better performance when FO statistics are used, especially in terms of maximum number of processed sources.

B. From SO to FO MUSIC metric

At first sight, SO and FO MUSIC-like approaches share similarities. However, the extension of MUSIC [7] to FO statistics is not trivial since the covariance and the quadricovariance matrices have different algebraic structures. Consequently, the Eigen Value Decomposition (EVD) of the covariance matrix and the EVD of the quadricovariance matrix will give two different MUSIC metrics. Before presenting the FO MUSIC concept, let's recall the SO MUSIC one.

Let the EVD of the covariance matrix R_x be given by:

$$R_x = U_s \, \Lambda_s \, U_s^{\mathsf{T}} + U_\nu \, \Lambda_\nu \, U_\nu^{\mathsf{T}} \tag{17}$$

where Λ_s is the $(P \times P)$ real-valued diagonal matrix of the P strongest eigenvalues of R_x, U_s is the $(N \times P)$ matrix of the associated orthonormalized eigenvectors (called *SO signal eigenmatrix*), and U_ν is the $(N \times (N - P))$ matrix of the orthonormalized eigenvectors (called *SO noise eigenmatrix*) associated with the remaining eigenvalues of R_x. Indeed, since R_x is a real symmetrical matrix, it can be diagonalized using a real unitary similarity transformation, namely $U = [U_s \; U_\nu]$. Then each column of U_s is orthogonal to each column of U_ν. Moreover, $\mathrm{Span}\{U_s\} = \mathrm{Span}\{A(\Theta)\}$, that is, each column vector of $A(\Theta)$ is a linear combination of the SO signal eigenvectors. Therefore each column of $A(\Theta)$ is orthogonal to each column of U_ν. So, denoting by θ_p the location/orientation parameters of the p-th source and $a(\theta_p)$ the localizing vector appearing at the p-th column of matrix $A(\Theta)$, vectors $a(\theta_p)$ $(1 \le p \le P)$ are orthogonal to each column of U_ν. Thus the standard SO metric used in MUSIC [22], S-MUSIC [17] and IES-MUSIC [25] can be defined as follows:

$$I_1(\theta) = \frac{a(\theta)^{\mathsf{T}} \, U_\nu \, U_\nu^{\mathsf{T}} \, a(\theta)}{a(\theta)^{\mathsf{T}} \, a(\theta)} \tag{18}$$

Another way to define the SO MUSIC metric consists in using the SO signal eigenmatrix instead of the SO noise eigenmatrix. This was done by Mosher et Leahy [15] based on the principal angles concept [9], giving rise to the following metric:

$$I_1'(\theta) = \frac{a(\theta)^{\mathsf{T}} \, U_s \, U_s^{\mathsf{T}} \, a(\theta)}{a(\theta)^{\mathsf{T}} \, a(\theta)} \tag{19}$$

Thus the P global minima of I_1, or equivalently the P global maxima of I_1', correspond asymptotically to the P source multi-parameters θ_p. However, in the brain source localization context, this implies a six dimensional (6D) optimization, and therefore an extremely high computational complexity. In order to decrease this computational cost, Ferrara et Parks [7] and Mosher et Leahy [15] took advantage of the separability of the data transfer matrix as a function of non-linear and linear parameters. More particularly, when the SO signal eigenmatrix is used, like in RapMUSIC [15], the P source locations ρ_p can be found as the P global maxima of the following metric:

$$I_2'(\rho) = \lambda_{max}\{V(\rho)^{\mathsf{T}} \, U_s \, U_s^{\mathsf{T}} \, V(\rho)\} \tag{20}$$

which does not depend on the orientation parameter anymore, and where $\lambda_{max}\{B\}$ denotes the maximum eigenvalue of matrix B and $V(\rho)$ is the left singular matrix of $G(\rho)$. The P source orientations ϕ_p can then be derived from the eigenvectors corresponding to the P maximum eigenvalues in (20). This way, the source orientation parameters are deduced from the computation of the source location parameters, and so the 6D optimization is reduced to a 3D optimization.

Now let the EVD of the quadricovariance matrix Q_x be given by:

$$Q_x = E_s \, L_s \, E_s^{\mathsf{T}} + E_\nu \, L_\nu \, E_\nu^{\mathsf{T}} \tag{21}$$

where L_s is the $(\mathcal{R}_J \times \mathcal{R}_J)$ real-valued diagonal matrix of the non-zero eigenvalues of Q_x, E_s is the $(N^2 \times \mathcal{R}_J)$ matrix of the associated orthonormalized eigenvectors (called *FO signal eigenmatrix*), L_ν is the $((N^2 - \mathcal{R}_J) \times (N^2 - \mathcal{R}_J))$ real-valued diagonal matrix of the zero eigenvalues of Q_x, and E_ν is the $(N^2 \times (N^2 - \mathcal{R}_J))$ matrix of the associated orthonormalized eigenvectors (called *FO noise eigenmatrix*). Indeed, since Q_x is a real symmetrical matrix, it can be diagonalized using a real unitary similarity transformation, namely $E = [E_s \; E_\nu]$. Then each column of E_s is orthogonal to each column of E_ν.

However, if the space spanned by the column vectors of matrix $\boldsymbol{U_s}$ is equal to the space spanned by the column vectors of $\boldsymbol{A}(\boldsymbol{\Theta})$, a question remains for the space spanned by the column vectors of matrix $\boldsymbol{E_s}$? In fact, we can deduce from both matrix decompositions of $\boldsymbol{Q_x}$, given by (15) and (21) respectively, that $\mathrm{Span}\{\boldsymbol{E_s}\} = \mathrm{Span}\{\boldsymbol{B}(\boldsymbol{\Theta})\}$, that is, each column vector of $\boldsymbol{B}(\boldsymbol{\Theta})$ is a linear combination of the FO signal eigenvectors. Consequently, each column of $\boldsymbol{B}(\boldsymbol{\Theta})$ is orthogonal to each column of $\boldsymbol{E_\nu}$. So, all the vectors $\boldsymbol{a}(\boldsymbol{\theta}_p)^{\otimes 2} = \boldsymbol{a}(\boldsymbol{\theta}_p) \otimes \boldsymbol{a}(\boldsymbol{\theta}_p)$ $(1 \le p \le P)$ of $\boldsymbol{B}(\boldsymbol{\Theta})$ are orthogonal to each column of $\boldsymbol{E_\nu}$. Thus, we can built a FO metric from the FO noise projector $\boldsymbol{P_\nu} = \boldsymbol{E_\nu}\, \boldsymbol{E_\nu}^{\mathsf{T}}$ such as:

$$J_1(\boldsymbol{\theta}) = \frac{\boldsymbol{a}(\boldsymbol{\theta})^{\otimes 2^{\mathsf{T}}}\, \boldsymbol{P_\nu}\, \boldsymbol{a}(\boldsymbol{\theta})^{\otimes 2}}{\boldsymbol{a}(\boldsymbol{\theta})^{\otimes 2^{\mathsf{T}}}\, \boldsymbol{a}(\boldsymbol{\theta})^{\otimes 2}} \tag{22}$$

where the P roots correspond asymptotically to the P source multi-parameters $\boldsymbol{\theta}_p$. However, this computation needs a 6D optimization and it would be interesting to see if the separability of the data transfer matrix as a function of location and orientation parameters could be used like with the SO metric in order to decrease the computational cost. So let insert (3) in (22), then J_1 becomes:

$$J_1(\boldsymbol{\theta}) = \frac{(\boldsymbol{G}(\boldsymbol{\rho})\, \boldsymbol{\phi})^{\otimes 2^{\mathsf{T}}}\, \boldsymbol{P_\nu}\, (\boldsymbol{G}(\boldsymbol{\rho})\, \boldsymbol{\phi})^{\otimes 2}}{(\boldsymbol{G}(\boldsymbol{\rho})\, \boldsymbol{\phi})^{\otimes 2^{\mathsf{T}}}\, (\boldsymbol{G}(\boldsymbol{\rho})\, \boldsymbol{\phi})^{\otimes 2}} \tag{23}$$

where $(\boldsymbol{G}(\boldsymbol{\rho})\, \boldsymbol{\phi})^{\otimes 2} = (\boldsymbol{G}(\boldsymbol{\rho})\, \boldsymbol{\phi}) \otimes (\boldsymbol{G}(\boldsymbol{\rho})\, \boldsymbol{\phi})$. Using Kronecker product properties, J_1 can then be rewritten as follows:

$$J_1(\boldsymbol{\theta}) = \frac{\boldsymbol{\phi}^{\otimes 2^{\mathsf{T}}}\, \boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2^{\mathsf{T}}}\, \boldsymbol{P_\nu}\, \boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2}\, \boldsymbol{\phi}^{\otimes 2}}{\boldsymbol{\phi}^{\otimes 2^{\mathsf{T}}}\, \boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2^{\mathsf{T}}}\, \boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2}\, \boldsymbol{\phi}^{\otimes 2}} \tag{24}$$

where $\boldsymbol{\phi}^{\otimes 2} = \boldsymbol{\phi} \otimes \boldsymbol{\phi}$ and $\boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2} = \boldsymbol{G}(\boldsymbol{\rho}) \otimes \boldsymbol{G}(\boldsymbol{\rho})$ are the FO source orientation vector and the FO gain matrix, respectively. Therefore, using Gantmacher's work [8], criterion (24) can be concentrated with respect to $\boldsymbol{\rho}$, leading to:

$$J_2(\boldsymbol{\rho}) = \lambda_{min}\{\boldsymbol{K}(\boldsymbol{\rho})\} \tag{25}$$

where $\boldsymbol{K}(\boldsymbol{\rho}) = \left(\boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2^{\mathsf{T}}} \boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2}\right)^{-1} \boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2^{\mathsf{T}}} \boldsymbol{P_\nu} \boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2}$ and where $\lambda_{min}\{\boldsymbol{B}\}$ denotes the minimum eigenvalue of matrix $\boldsymbol{B}$. So parameters $\boldsymbol{\theta}_p = \left[\boldsymbol{\rho}_p^{\mathsf{T}}\ \boldsymbol{\phi}_p^{\mathsf{T}}\right]^{\mathsf{T}}$ can be obtained first by looking for the P roots $\boldsymbol{\rho}_p$ of the function in $\boldsymbol{\rho}$ defined by the minimum eigenvalue of matrix $\boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2^{\mathsf{T}}} \boldsymbol{P_\nu} \boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2}$ in the metric $\boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2^{\mathsf{T}}} \boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2}$, and then by computing the vector $\boldsymbol{\phi}_p^{\otimes 2}$ associated with the p-th source as the eigenvector corresponding to the minimum eigenvalue of matrix $\boldsymbol{K}\left(\boldsymbol{\rho}_p\right)$. An algorithm is proposed in section IV-C in order to deduce vector $\boldsymbol{\phi}$ from $\boldsymbol{\phi}^{\otimes 2}$. Consequently the orientation parameters are deduced from the location parameters. This way, the 6D-optimization problem is reduced to a 3D-optimization problem. Since the P source locations found are those for which matrix $\boldsymbol{K}\left(\boldsymbol{\rho}_p\right)$ has a zero minimum eigenvalue, they can also be computed as the source locations for which matrix $\boldsymbol{K}\left(\boldsymbol{\rho}_p\right)$ has a deficient rank, that is, a zero determinant. Consequently the computational cost can considerably be reduced if criterion J_2 is replaced by the following equivalent criterion:

$$J_3(\boldsymbol{\rho}) = \frac{\det\left\{\boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2^{\mathsf{T}}}\, \boldsymbol{P_\nu}\, \boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2}\right\}}{\det\left\{\boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2^{\mathsf{T}}}\, \boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2}\right\}} \tag{26}$$

where $\det\{\boldsymbol{V}\}$ denotes the determinant of matrix $\boldsymbol{V}$. On the one hand, note that a FO metric based on the correlation between $\mathrm{Span}\{\boldsymbol{B}(\boldsymbol{\Theta})\}$ and $\mathrm{Span}\{\boldsymbol{E_s}\}$ instead of the orthogonality between $\mathrm{Span}\{\boldsymbol{B}(\boldsymbol{\Theta})\}$ and $\mathrm{Span}\{\boldsymbol{E_\nu}\}$ could be built, giving birth to a natural extension of the SO metric (20) used in RapMUSIC [15] to FO statistics. A potential tool to measure correlation between two subspaces, as mentioned by Mosher et al. [15], is the principal angle technique [9] and could be used in order to obtain this new FO MUSIC metric. However, this metric could not be rewritten with less costly form such as (26). Consequently, at this stage, since a determinant computation is less costly than an EVD, especially for high matrix dimensions, criterion (26) is an attractive FO MUSIC-like metric. On the other hand, a simple algorithm scheme could be performed in order to decrease the computational cost of (26). It would consist in i) only computing the smallest (in number of elements) family of eigenvectors, either the noise one or the signal one, of $\boldsymbol{Q_x}$ using for instance the power method [9] and ii) deducing the FO noise projector $\boldsymbol{P_\nu}$ from the previous computation. More precisely, if $N^2 - \mathcal{R}_J \le \mathcal{R}_J$, then compute the FO noise eigenmatrix $\boldsymbol{E_\nu}$ and take $\boldsymbol{P_\nu} = \boldsymbol{E_\nu}\, \boldsymbol{E_\nu}^{\mathsf{T}}$, otherwise compute the FO signal eigenmatrix $\boldsymbol{E_s}$ and take:

$$\boldsymbol{P_\nu} = \mathbf{I}_{N^2} - \boldsymbol{E_s}\left(\boldsymbol{E_s}^{\mathsf{T}} \boldsymbol{E_s}\right)^{-1} \boldsymbol{E_s}^{\mathsf{T}} \tag{27}$$

C. From SO to FO deflation approach

This section reviews the concept of deflation used in S-MUSIC [17], IES-MUSIC [25] and RapMUSIC [15], and shows how it can be implemented when FO statistics are used, more particularly when criterion (26) is used. Again, this implementation is not trivial since matrices $\boldsymbol{R_x}$ and $\boldsymbol{Q_x}$ have different algebraic structures. Indeed, the SO deflation projector used in IES-MUSIC, or in S-MUSIC and RapMUSIC, cannot be applied when FO statistics are used. Moreover, the choice of this FO deflation projector, especially when sources are spatially correlated, is discussed hereafter.

As written in the previous subsection, a first idea would consist in searching for the P global maximizers of criterion (20) when SO statistics are used, or searching for the P roots of criterion (26) when FO statistics are preferred (in the sequel we will refer to both approaches as the MUSIC and FO-MUSIC algorithms respectively). Indeed, if the noise subspace projector was estimated perfectly, i.e. asymptotically, then the source locations would be directly found as the P global maximizers of (20) or as the P roots of (26) respectively. Nevertheless, for a finite number of samples, errors in our statistic estimate reduce (20) and (26) to a function with i) a single global optimum that corresponds for instance to the source of maximum Signal-to-Noise Ratio (SNR), and ii) $P-1$ local optima. Although the global optimum is easily identifiable, it is more difficult to find the P–1 remaining local optima because non-linear search techniques may miss shallow or adjacent peaks and return to a previous peak. Algorithms have been proposed to solve *peak-picking* problems [12], but they rapidly become complex and subjective as the number of sources and the dimensionality of vectors $\boldsymbol{\rho}_p$ increase [15]. In order to avoid this peak-picking problem a computation strategy based on the deflation concept was proposed in [15], [17], [25] when SO statistics are used. However, as the SO deflation approach cannot be applied to the FO MUSIC metric (26), we extended the deflation concept to FO statistics and more particularly to criterion (26), giving rise to the FO-D-MUSIC method.

In S-MUSIC and IES-MUSIC, the location, $\boldsymbol{\rho}_{\xi(1)}$, and the orientation, $\boldsymbol{\phi}_{\xi(1)}$, of the first source are determined at the same time by searching for the global minimum root of (18). The use of the bijective function ξ of $\{1, 2, \ldots, P\}$ into itself (i.e. a permutation) is necessary since the P source localizing vectors $\boldsymbol{a}(\boldsymbol{\theta}_p)$ may be found back, but only in the disorder. Indeed, as shown in equation (1), the order in which components of $\boldsymbol{s}$ and associated columns of $\boldsymbol{A}(\boldsymbol{\Theta})$ are set does not change the expression of $\boldsymbol{x}$. As far as the RapMUSIC and FO-D-MUSIC methods are concerned, the first source location, $\boldsymbol{\rho}_{\xi(1)}$, is determined by searching for the global maximizer of (20) and the global minimum root of (26) over a sufficiently densely sampled grid of the non-linear parameter space, respectively. Next, the orientation is derived from the source location in both approaches. On the one hand, RapMUSIC determines the first source orientation, $\boldsymbol{\phi}_{\xi(1)}$, as the normalized eigenvector corresponding to the global maximum eigenvalue in (20). On the other hand, FO-D-MUSIC finds the first FO source orientation vector, $\boldsymbol{\phi}_{\xi(1)}^{\otimes 2}$, as the normalized eigenvector corresponding to the global minimum eigenvalue of matrix $\left(\boldsymbol{G}(\boldsymbol{\rho}_{\xi(1)})^{\otimes 2^{\mathsf{T}}} \boldsymbol{G}(\boldsymbol{\rho}_{\xi(1)})^{\otimes 2}\right)^{-1} \boldsymbol{G}(\boldsymbol{\rho}_{\xi(1)})^{\otimes 2^{\mathsf{T}}} \boldsymbol{P}_{\boldsymbol{\nu}}\, \boldsymbol{G}(\boldsymbol{\rho}_{\xi(1)})^{\otimes 2}$. Then, the source orientation vector $\boldsymbol{\phi}_{\xi(1)}$ can be computed from $\boldsymbol{\phi}_{\xi(1)}^{\otimes 2}$, by i) reshaping it into an $(N \times N)$ matrix $\boldsymbol{F}_{\xi(1)}$ (the n-th column of $\boldsymbol{F}_{\xi(1)}$ is made up from the N consecutive elements of $\boldsymbol{\phi}_{\xi(1)}^{\otimes 2}$ as from the $[N(n-1)+1]$-th one), and ii) diagonalizing it. Indeed, the normalized eigenvector associated with the strongest eigenvalue of $\boldsymbol{F}_{\xi(1)}$, is, up to a sign factor, equal to $\boldsymbol{\phi}_{\xi(1)}$.

Once the first source has been localized, its contribution can be removed from the data and the second source multi-parameter vector, $\boldsymbol{\theta}_{\xi(2)}$, can be searched for: this defines the first step of the deflation scheme. More particularly, S-MUSIC builds the following orthogonal projecting matrix:

$$\boldsymbol{A}_1^{\perp} = \mathbf{I}_N - \boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)\, \boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)^{\mathsf{T}} / \left\|\boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)\right\|^2 \tag{28}$$

where $\boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big) \overset{\text{def}}{=} \boldsymbol{G}(\boldsymbol{\rho}_{\xi(1)})\, \boldsymbol{\phi}_{\xi(1)}$ and applies it to the source localizing vector $\boldsymbol{a}(\boldsymbol{\theta})$ before looking for the second source multi-parameter vector, $\boldsymbol{\theta}_{\xi(2)}$, from criterion (18). In IES-MUSIC, the projecting matrix is also applied to the source localizing vector, but it is not necessarily orthogonal and depends on a scalar-valued user parameter. Nevertheless, the optimal scalar is derived in [25] only for the case of two sources, which requires to know the localizing vectors of both sources. Thus, in practice IES-MUSIC needs to estimate both source localizing vectors with another method first. Moreover, S-MUSIC and IES-MUSIC are suboptimal since they remove the contribution of the first source only from the source localizing vector. Had they remove it from the data as well, they could increase the dimensionality of the noise subspace and therefore the estimation resolution at each step of the deflation scheme. In RapMUSIC the orthogonal projecting matrix (28) is applied both to the source localizing vector $\boldsymbol{a}(\boldsymbol{\theta})$ and to the SO signal eigenmatrix $\boldsymbol{U}_{\boldsymbol{s}}$ before looking for the second source location. It is noteworthy that this procedure allows to remove the contribution of the first source from the data when criterion (20) is used. Finally in FO-D-MUSIC, the contribution of the first source could be removed by applying the orthogonal projecting matrix (28) both to the source localizing vector $\boldsymbol{a}(\boldsymbol{\theta})$ and to vector $\boldsymbol{x}$. Nonetheless, this procedure would imply a new statistical estimation step such as the estimation of the quadricovariance matrix of the processed data, and therefore an increased computational cost. Indeed, it is better to remove the contribution of the first source from the initial quadricovariance matrix $\boldsymbol{Q}_{\boldsymbol{x}}$ instead of the data $\boldsymbol{x}$. However, contrary to the covariance matrix $\boldsymbol{R}_{\boldsymbol{x}}$, the quadricovariance matrix $\boldsymbol{Q}_{\boldsymbol{x}}$ cannot be multiplied on left and on right by $\boldsymbol{A}_1^{\perp}$ and $\boldsymbol{A}_1^{\perp^{\mathsf{T}}}$ respectively, in order to cancel the contribution of the first source. Indeed, the algebraic structure of $\boldsymbol{Q}_{\boldsymbol{x}}$ has to be studied to understand how the first source is involved on it. According to equation (13), the mathematical challenge then consists in cancelling all the column vectors of matrix $\boldsymbol{B}(\boldsymbol{\Theta})$ involving vector $\boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)$, that is, all the column vectors of $\boldsymbol{B}(\boldsymbol{\Theta})$ of the form $\boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)\otimes\boldsymbol{b}$ or $\boldsymbol{b}\otimes\boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)$ where $\boldsymbol{b}$ is a $(N\times 1)$ vector. When the first found source is statistically independent of all the other sources, that is $P_{\xi(1)} = 1$, the only column vector of $\boldsymbol{B}(\boldsymbol{\Theta})$ involving $\boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)$ is $\boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)^{\otimes 2}$. Therefore it can be cancelled using the following projecting matrix:

$$\boldsymbol{B}_1^{\perp} = \mathbf{I}_{N^2} - \boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)^{\otimes 2} \boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)^{\otimes 2^{\mathsf{T}}} / \left\|\boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)^{\otimes 2}\right\|^2 \tag{29}$$

However, when source $\xi(1)$ is dependent of one or several other sources, such as sources i and j for instance, the projecting matrix (29) is suboptimal. Vector $\boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)^{\otimes 2}$ of matrix $\boldsymbol{B}(\boldsymbol{\Theta})$ has to be cancelled as well as vectors $\boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)\otimes\boldsymbol{a}(\boldsymbol{\theta}_i)$, $\boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)\otimes\boldsymbol{a}(\boldsymbol{\theta}_j)$, $\boldsymbol{a}(\boldsymbol{\theta}_i)\otimes\boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)$ and $\boldsymbol{a}(\boldsymbol{\theta}_j)\otimes\boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)$. This may be achieved multiplying matrix $\boldsymbol{B}(\boldsymbol{\Theta})$ on the left by $\boldsymbol{A}_1^{\perp\otimes 2} = \boldsymbol{A}_1^{\perp} \otimes \boldsymbol{A}_1^{\perp}$. The proof straightly ensues from the algebraic structure of matrix $\boldsymbol{B}(\boldsymbol{\Theta})$ (see section IV-A) and properties of the Kronecker product. Consequently, in order to process the general case where source $\xi(1)$ is potentially correlated with other sources, the location parameters, $\boldsymbol{\rho}_{\xi(2)}$, associated with the $\xi(2)$-th source, are then found as the global minimizer of (26) replacing $\boldsymbol{G}(\boldsymbol{\rho})$ by $\boldsymbol{A}_1^{\perp\otimes 2}\, \boldsymbol{G}(\boldsymbol{\rho})$ and where $\boldsymbol{P}_{\boldsymbol{\nu}}$ will be no longer given from the EVD of matrix $\boldsymbol{Q}_{\boldsymbol{x}}$, but from the EVD of $\boldsymbol{A}_1^{\perp\otimes 2}\, \boldsymbol{Q}_{\boldsymbol{x}}\, \boldsymbol{A}_1^{\perp\otimes 2^{\mathsf{T}}}$. Due to matrix multiplication, the rank of this last matrix is now strictly smaller than $\mathcal{R}_J$. Indeed, we decreased the rank of $\boldsymbol{Q}_{\boldsymbol{x}}$ by removing the contribution of the first source from the initial statistical matrix $\boldsymbol{Q}_{\boldsymbol{x}}$, and consequently we increased the dimension of noise subspace. Besides, when source $\xi(1)$ is dependent of one or several other sources, the use of matrix $\boldsymbol{A}_1^{\perp\otimes 2}$ instead of the orthogonal projecting matrix $\boldsymbol{B}_1^{\perp}$ (29) allows for a greater increase of the dimension of noise subspace, which will lead to a best estimation of the second source location, $\boldsymbol{\rho}_{\xi(2)}$. Once the second source location has been found, the source orientation vector $\boldsymbol{\phi}_{\xi(2)}$ is computed in the same way as $\boldsymbol{\phi}_{\xi(1)}$, replacing $\xi(1)$ by $\xi(2)$, and the localizing vector $\boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(2)}\big) \overset{\text{def}}{=} \boldsymbol{G}(\boldsymbol{\rho}_{\xi(2)})\, \boldsymbol{\phi}_{\xi(2)}$ can be built. Eventually, it is possible to reduce the computational cost of the previous FO deflation process especially for a large number of observations and consequently large dimensions of $\boldsymbol{Q_x}$. Indeed, the second source location can be found as the global minimizer of (26) replacing $\boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2}$ by $\boldsymbol{A}_1^{\perp\otimes 2}\,\boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2}$ and redefining $\boldsymbol{P_\nu}$ by:

$$\boldsymbol{P_\nu} = \mathbf{I}_{N^2} - \boldsymbol{A}_1^{\perp\otimes 2}\,\boldsymbol{E_s}\left(\boldsymbol{E_s}^{\top}\boldsymbol{A}_1^{\perp\otimes 2^{\top}}\boldsymbol{A}_1^{\perp\otimes 2}\boldsymbol{E_s}\right)^{-1}\boldsymbol{E_s}^{\top}\boldsymbol{A}_1^{\perp\otimes 2^{\top}} \tag{30}$$

This way, the diagonalization of a ($N^2\times N^2$) matrix is (time-consuming for large values of N) is avoided.

Next, the S-MUSIC, IES-MUSIC, RapMUSIC and FO-D-MUSIC deflation approaches proceed all recursively up to estimate the P source parameter vectors $\boldsymbol{\theta}_p = \left[\boldsymbol{\rho}_p{}^{\top}\,\boldsymbol{\phi}_p{}^{\top}\right]^{\top}$. The IES-MUSIC needs a scalar-valued user parameter, which optimal value is only given for $P=2$ sources in [25]. An extension of the outlined algorithm to $P>2$ requires more effort and notation and is not considered in [25]. The S-MUSIC and RapMUSIC methods build the following projecting matrix once the $p-1$-th source localization has been achieved:

$$\left|\begin{array}{l}\forall\, j,\ 1\le j\le p-1,\ \boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(j)}\big) = \boldsymbol{G}(\boldsymbol{\rho}_{\xi(j)})\,\boldsymbol{\phi}_{\xi(j)}\\ \boldsymbol{A}_{p-1} = \left[\boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)\ \cdots\ \boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(p-1)}\big)\right]\\ \boldsymbol{A}_{p-1}^{\perp} = \mathbf{I}_N - \boldsymbol{A}_{p-1}\left((\boldsymbol{A}_{p-1})^{\top}\boldsymbol{A}_{p-1}\right)^{-1}(\boldsymbol{A}_{p-1})^{\top}\end{array}\right. \tag{31}$$

Then, S-MUSIC applies this matrix to the source localizing vector $\boldsymbol{a}(\boldsymbol{\theta})$ whereas RapMUSIC applies it both to the source localizing vector $\boldsymbol{a}(\boldsymbol{\theta})$ and to the SO signal eigenmatrix $\boldsymbol{U_s}$ before looking for the p-th source parameters. Finally in the FO-D-MUSIC algorithm, the p-th source localization step depends on two cases. If $p\le N$, then it mainly consists in minimizing criterion J_3 (26) replacing $\boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2}$ and $\boldsymbol{P_\nu}$ by $\boldsymbol{A}_{p-1}^{\perp\otimes 2}\,\boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2}$ and the FO noise projector of matrix $\boldsymbol{A}_{p-1}^{\perp\otimes 2}\,\boldsymbol{Q_x}\,\boldsymbol{A}_{p-1}^{\perp\otimes 2^{\top}}$, respectively, where $\boldsymbol{A}_{p-1}^{\perp\otimes 2} = \boldsymbol{A}_{p-1}^{\perp}\otimes\boldsymbol{A}_{p-1}^{\perp}$. Otherwise, if $p>N$, the previous procedure holds but matrix $\boldsymbol{A}_{p-1}^{\perp\otimes 2}$ is replaced by the following projecting matrix, $\boldsymbol{B}_{p-1}^{\perp}$:

$$\left|\begin{array}{l}\forall\, j,\ 1\le j\le p-1,\ \boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(j)}\big) = \boldsymbol{G}(\boldsymbol{\rho}_{\xi(j)})\,\boldsymbol{\phi}_{\xi(j)}\\ \boldsymbol{B}_{p-1} = \left[\boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(1)}\big)^{\otimes 2}\ \cdots\ \boldsymbol{a}\big(\boldsymbol{\theta}_{\xi(p-1)}\big)^{\otimes 2}\right]\\ \boldsymbol{B}_{p-1}^{\perp} = \mathbf{I}_{N^2} - \boldsymbol{B}_{p-1}\left((\boldsymbol{B}_{p-1})^{\top}\boldsymbol{B}_{p-1}\right)^{-1}(\boldsymbol{B}_{p-1})^{\top}\end{array}\right. \tag{32}$$

Indeed, according to (31), for $p=N+1$ matrix $\boldsymbol{A}_{p-1}^{\perp}$ is a zero square matrix whereas for $p>N+1$ matrix $\boldsymbol{A}_{p-1}^{\perp}$ is not defined. Consequently, matrix $\boldsymbol{A}_{p-1}^{\perp\otimes 2}$ cannot be used as soon as p is strictly greater than N. Note that the case for which $p>N$ is possible for FO-D-MUSIC since this algorithm, contrary to S-MUSIC, IES-MUSIC and RapMUSIC, can process underdetermined mixtures of sources (we will justify this assertion in section IV-E). As described in the previous paragraph, the diagonalization of matrix $\boldsymbol{A}_{p-1}^{\perp\otimes 2}\,\boldsymbol{Q_x}\,\boldsymbol{A}_{p-1}^{\perp\otimes 2^{\top}}$ for $p\le N$ (respectively $\boldsymbol{B}_{p-1}^{\perp}\,\boldsymbol{Q_x}\,(\boldsymbol{B}_{p-1}^{\perp})^{\top}$ for $p>N$) can be avoided in order to reduce the computational cost of the deflation scheme: $\boldsymbol{P_\nu}$ has to be constructed using equation (30) where $\boldsymbol{A}_1^{\perp\otimes 2}$ is replaced by $\boldsymbol{A}_{p-1}^{\perp\otimes 2}$.

D. Implementation of the FO-D-MUSIC algorithm

The different steps of the FO-D-MUSIC method are summarized below, when K observations of the stochastic vector $\boldsymbol{x}$ are available.

Step1 Fix p equal to one, estimate the FO statistics $C_{i_1,i_2,i_3,i_4,\boldsymbol{x}}$ from the K samples of $\boldsymbol{x}$ and compute an estimate $\widehat{\boldsymbol{Q}}_{\boldsymbol{x}}$ of the quadricovariance matrix $\boldsymbol{Q_x}$.

Step2 Build a set of matrices $\left\{\boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2}\right\}$ choosing a sufficiently densely sampled grid of vectors $\boldsymbol{\rho}$.

Step3 Compute the EVD of matrix $\widehat{\boldsymbol{Q}}_{\boldsymbol{x}}$, extract the estimates $\widehat{\boldsymbol{E}}_{\boldsymbol{s}}$ and $\widehat{\boldsymbol{E}}_{\boldsymbol{\nu}}$ of matrices $\boldsymbol{E_s}$ and $\boldsymbol{E_\nu}$ respectively, and compute the estimate $\widehat{\boldsymbol{P}}_{\boldsymbol{\nu}}$ of $\boldsymbol{P_\nu}$ according to the end of section IV-B.

Step4 Compute an estimate, $\widehat{J}_3$, of criterion J_3 (26) (using matrix $\widehat{\boldsymbol{P}}_{\boldsymbol{\nu}}$ instead of $\boldsymbol{P_\nu}$) over the suitably chosen grid, and search for its global minimum, $\widehat{\boldsymbol{\rho}}_{\xi(p)}$.

Step5 Compute vector $\widehat{\boldsymbol{\Phi}}_{\xi(p)}$ taking as solution the eigenvector corresponding to the minimum eigenvalue of matrix $\boldsymbol{G}(\widehat{\boldsymbol{\rho}}_{\xi(p)})^{\otimes 2^{\top}}\,\widehat{\boldsymbol{P}}_{\boldsymbol{\nu}}\,\boldsymbol{G}(\widehat{\boldsymbol{\rho}}_{\xi(p)})^{\otimes 2}$ in the metric $\boldsymbol{G}(\widehat{\boldsymbol{\rho}}_{\xi(p)})^{\otimes 2^{\top}}\boldsymbol{G}(\widehat{\boldsymbol{\rho}}_{\xi(p)})^{\otimes 2}$.

Step6 Extract the estimate $\widehat{\boldsymbol{\phi}}_{\xi(p)}$ of the source orientation vector $\boldsymbol{\phi}_{\xi(p)}$ from $\widehat{\boldsymbol{\Phi}}_{\xi(p)}$. In order to do this, first reshape it into a matrix $\widehat{\boldsymbol{F}}_{\xi(p)}$, and secondly compute the normalized eigenvector associated with the largest eigenvalue of $\widehat{\boldsymbol{F}}_{\xi(p)}$.

Step7 If the rank of matrix $\widehat{\boldsymbol{E}}_{\boldsymbol{s}}$ is not equal to one, that is, if the P sources are not all localized,

- **i)** increment p and build vector $\boldsymbol{a}(\widehat{\boldsymbol{\theta}}_{\xi(p-1)}) = \boldsymbol{G}(\widehat{\boldsymbol{\rho}}_{\xi(p-1)})\,\widehat{\boldsymbol{\phi}}_{\xi(p-1)}$,
- **ii)** compute matrix $\widehat{\boldsymbol{\Pi}}_{p-1}$ equal to $\widehat{\boldsymbol{A}}_{p-1}^{\perp\otimes 2}$ if $p\le N$ and to $\widehat{\boldsymbol{B}}_{p-1}^{\perp}$ otherwise (see section IV-C) where $\widehat{\boldsymbol{A}}_{p-1}^{\perp\otimes 2}$ and $\widehat{\boldsymbol{B}}_{p-1}^{\perp}$ are the estimates of $\boldsymbol{A}_{p-1}^{\perp\otimes 2}$ and $\boldsymbol{B}_{p-1}^{\perp}$ respectively using $\boldsymbol{a}(\widehat{\boldsymbol{\theta}}_{\xi(p-1)})$ instead of $\boldsymbol{a}(\boldsymbol{\theta}_{\xi(p-1)})$,
- **iii)** go back to **step4** replacing $\boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2}$ by $\widehat{\boldsymbol{\Pi}}_{p-1}\,\boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2}$ and where $\widehat{\boldsymbol{P}}_{\boldsymbol{\nu}}$ is achieved from (30) replacing $\boldsymbol{A}_1^{\perp\otimes 2}$ and $\boldsymbol{E_s}$ by $\widehat{\boldsymbol{\Pi}}_{p-1}$ and $\widehat{\boldsymbol{E}}_{\boldsymbol{s}}$, respectively,

else stop the procedure.

Note that this implementation requires neither the knowledge of number P nor its estimation, since the deflation procedure is stopped as soon as the rank of the estimated signal eigenmatrix $\widehat{\boldsymbol{E}}_{\boldsymbol{s}}$ is equal to one.

E. Identifiability of the FO-D-MUSIC method

From the previous sections, it appears that, under assumptions **A1** to **A4**, the FO-D-MUSIC method can localize P brain current sources from N surface observations. As this new algorithm may process underdetermined mixtures when some sources are statistically independent (see section IV-A), we limit the analysis to the latter case. Moreover, for the sake of simplicity, we assume that all the sources are statistically independent. In such a situation, hypotheses **A1** to **A4** reduce to **A"1** to **A"2**. Then vector $\boldsymbol{a}(\boldsymbol{\theta})^{\otimes 2}$ can be considered as an actual source localization vector but for a FO virtual array [5] of electrodes, that gives at each measurement time, $\mathcal{N}$ different virtual scalp data. $\mathcal{N}$ is directly related to the pattern of the actual sensors, to the geometry of the actual array of sensors and to the considered head model. Consequently, this means that $N^2 - \mathcal{N}$ components of all the vectors $\boldsymbol{a}(\boldsymbol{\theta})^{\otimes 2}$ are redundant components that bring no information. As a consequence, $N^2 - \mathcal{N}$ rows of the $\boldsymbol{B}(\boldsymbol{\Theta})$ matrix bring no information and are linear combinations of the others, which means that the rank of $\boldsymbol{B}(\boldsymbol{\Theta})$ cannot be greater than $\mathcal{N}$. In these conditions, matrix $\boldsymbol{B}(\boldsymbol{\Theta})$ may have a rank equal to P only if $P \leq \mathcal{N}$. Conversely, for a FO virtual array without any ambiguities up to order $\mathcal{N}-1$, P sources localized at P different positions generate a matrix $\boldsymbol{B}(\boldsymbol{\Theta})$ with a full rank P as long as $P \leq \mathcal{N}$. Thus the maximal number of statistically independent sources able to generate a matrix $\boldsymbol{B}(\boldsymbol{\Theta})$ with rank P is $\mathcal{N}$. However, when $P = \mathcal{N}$, an arbitrary vector $\boldsymbol{a}(\boldsymbol{\theta})^{\otimes 2}$ associated with an arbitrary set $\boldsymbol{\theta}$ of localization parameters is necessarily a linear combination of the source localization vectors $\boldsymbol{a}(\boldsymbol{\theta}_p)^{\otimes 2}$, $1 \leq p \leq \mathcal{N}$, since matrix $\boldsymbol{B}(\boldsymbol{\Theta})$ cannot have a rank greater than $\mathcal{N}$, and all the multi-parameter vectors $\boldsymbol{\theta}$ are then solutions of equation (22), which does not allow the source localization. Thus, a necessary condition for the localization of the sources to be the only solutions of equation (22) is that $P < \mathcal{N}$. This condition becomes sufficient for FO virtual arrays with no ambiguities. So we deduce that the algorithm which looks for the P minimizers of (22) is able to process up to $P = \mathcal{N} - 1$ sources, where $\mathcal{N}$ can be found as the maximum rank of matrix $\boldsymbol{B}(\boldsymbol{\Theta})$. However, when criterion (26) has to be rendered null instead of criterion (22) for the location of the sources only, and not for other locations, the (9×9) matrix $\boldsymbol{M}(\boldsymbol{\rho}) \stackrel{\text{def}}{=} \boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2^{\mathsf{T}}} \boldsymbol{P_\nu}\, \boldsymbol{G}(\boldsymbol{\rho})^{\otimes 2}$ has to be full rank when $\boldsymbol{\rho}$ does not correspond to a source's location. Using the definition of matrix $\boldsymbol{M}(\boldsymbol{\rho})$, this means that rank of matrix $\boldsymbol{P_\nu}$ cannot be lower than 9, which means that the rank of $\boldsymbol{E_\nu}$ has to be greater than or equal to 9. As, in the presence of P statistically independent sources such that $P < \mathcal{N}$, the rank of $\boldsymbol{E_\nu}$ is equal to $\mathcal{N} - P$ for a FO virtual array without any ambiguities up to order $\mathcal{N}-1$, the maximal number of sources that may be processed by FO-D-MUSIC method has to be lower than $\mathcal{N} - 9$. Conversely, for a FO virtual array without any ambiguities up to order $\mathcal{N}-1$, P sources having different locations with different orientations and such that $P \leq \mathcal{N} - 9$ are such that their locations are the only solutions that render null criterion (26). From the previous results, assuming a FO virtual array with $\mathcal{N}$ different virtual scalp data and with no ambiguities up to order $\mathcal{N} - 1$, we deduce that FO-D-MUSIC can process up to $P = \mathcal{N} - 9$ sources. Although the maximal number of potentially processed sources is a bit larger when criterion (22) is used, the minimization of (22) is in practice difficult to realize: it would need to optimize a 6-variables function while in criterion (26) a 3D minimization is sufficient. Note that exact computation of $\mathcal{N}$ as a function of N for a particular geometry of actual array of sensors and a particular head model would require more effort and notation, and is not considered in this work. However, looking at the algebraic structure of vector $\boldsymbol{a}(\boldsymbol{\theta}_p)^{\otimes 2}$ shows that $\mathcal{N}$ is smaller than $N(N-1)$ whatever the actual array is. Moreover, we estimated some values of $\mathcal{N}$ (reported in table II) using Matlab simulations. These values were computed from some values of N (6, 9, 18, ...) and from the head model described in section II-B. Mathematically, we built matrix $\boldsymbol{B}(\boldsymbol{\Theta}) = \boldsymbol{A}(\boldsymbol{\Theta}) \oslash \boldsymbol{A}(\boldsymbol{\Theta})$ from equations (3) to (6) and we estimated its maximum rank thanks to Matlab. Eventually, whereas the FO virtual array theory gives a theoretical justification of why FO-D-MUSIC can process underdetermined mixtures of independent sources, it also shows why FO-D-MUSIC performs better in the overderdetermined case when fine resolution is required. Indeed, as mentionned above, instead of using only N scalp measurements as the classical SO MUSIC-like methods, FO-D-MUSIC exploits $\mathcal{N}$ different virtual scalp data, where $\mathcal{N}$as illustrated in table II.

N	6	9	18	31	63	95
$\mathcal{N}$	21	45	171	496	2009	4340

TABLE II
ESTIMATED NUMBER OF FO VIRTUAL SCALP DATA AS A FUNCTION OF N.

V. COMPUTER RESULTS

In this section, the performance of the FO-D-MUSIC algorithm are compared with two classical SO MUSIC-like methods (namely MUSIC [7] and RapMUSIC [15]) in various situations using computer simulations. In addition, we decided to compute the performances of the FO-MUSIC method, which consists in searching simultaneously for the P "best" minimizers of criterion J_3, in order to show the contribution of the deflation scheme at the fourth order. As far as the head model is concerned, we used three nested concentric spheres with radius and conductivities values given in table I. One hundred twenty eight electrodes were placed on the scalp sphere using the 10-5 system [18]. Among them, only nineteen electrodes were used except in section V-E where we studied the effect of the number of surface observations by varying the number of electrodes. Besides, $P = 1$ or $P = 2$ independent sources were arranged in the xOz-plane. Note that the origin (O) of the head model was defined as the intersection of the O-Cz axis (z-axis), the O-T4 axis (x-axis) and the O-Fpz axis (y-axis). A physiologically-relevant model was used to compute realistic source temporal dynamics. This model consists in a network of coupled neuronal populations. It is described in previous reports [26] which showed that temporal dynamics of simulated signals closely resemble those actually recorded with intracranial electrodes in epileptic patients [6]. Briefly, the model is a lumped-parameter representation of a set of interconnected populations of neurons. Each population contains two subsets of neurons (main pyramidal cells and local interneurons) that interact via excitatory or inhibitory connections (postsynaptic interactions only). Populations can be coupled either uni- or bi-directionally via excitatory connections. Model output corresponds to the local field potential generated at each population of the network. Figure 1 shows the temporal dynamics for source 1 and 2 generated from the model, as well as five corresponding simulated surface EEG data. Note that the sources have an estimated *kurtosis* (FO cumulant normalized by the square of the variance) equal to 26.5 and 5.5 respectively. Besides, The source orientations, $\boldsymbol{\phi}_p$ ($1 \leq p \leq P$), were randomly fixed such as $\|\boldsymbol{\phi}_p\| = 1$.

We considered the background noise as Gaussian except for section V-D and as temporally and spatially white except for section V-C. In addition, we created a "0.1 mm"-spaced grid in the xOz-plane and computed the SO and FO gain matrices for each location on the grid. Eventually, the simulation results were averaged over $M = 200$ realizations. From one realization to another, both temporal dynamics and noise were changed while the mixing matrix stayed unchanged except for section V-A. SO and FO cumulants were estimated from 5000 data samples except for sections V-C and V-D where 10000 and 2500 samples were used, respectively. Simulations were performed using Matlab (V7.0, Release 14). As an example, when a grid of 100 location points is used, FO-D-MUSIC takes 360 milliseconds to localize two sources from eighteen surface observations, on a standard PC computer (64-bit processor, 4 GB RAM). Two criteria were used to quantify the quality of

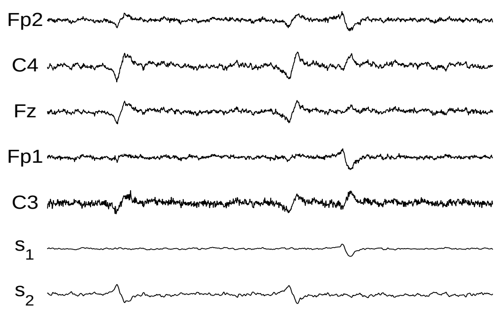


Fig. 1. Five selected surface EEG data from two deep sources (s_1 and s_2) having the same SNR equal to 10 dB.

the source localization. The first one is the Probability of Non-Localization (PNL), that is, the probability that the considered localization method does not succeed in finding exactly P solutions. For each localization method, the PNL criterion is defined by the ratio between the number of realizations for which all the sources are not localized and the total number of realizations M. The second one is the well-known averaged Root Mean Square Error (RMSE), computed for each source and for a given source localization method. More precisely, for a given number M' ($M' \leq M$) of realizations for which the considered localization method has succeeded in finding exactly P solutions, the averaged RMSE for source p associated with the localization estimation, $\mathit{RMSE}(\boldsymbol{\theta}_p)$, is defined by:

$$\mathit{RMSE}(\boldsymbol{\theta}_p) = \frac{1}{M'} \sum_{m=1}^{M'} \left(\min_{1 \leq j \leq P} \left\{ \left\| \boldsymbol{\theta}_p - \widehat{\boldsymbol{\theta}}_j^{(m)} \right\| \right\} \right) \tag{33}$$

where $\widehat{\boldsymbol{\theta}}_j^{(m)}$ is the j-th source parameter vector estimated during the m-th experiment. The minimization over the set $\{1, 2, \ldots, P\}$ of integers is necessary since the source parameter vectors may be recovered only in the disorder.

A. Effect of the dipole location on source localization

We studied in this section the behaviour of the MUSIC, RapMUSIC, FO-MUSIC and FO-D-MUSIC methods in the presence of a unique source. In fact, since we looked for only one source localization, the non-deflation method was equivalent to the deflation one, for a given order of statistics. Results are illustrated in figure 2 which displays the variations of the RMSE criterion at the output of the previous algorithms as a function of the source location (in centimetres) on the z-axis. They show that both FO MUSIC-like methods localize more precisely the source than both SO MUSIC-like algorithms, wherever the source is. The PNL criterion was close to zero for all the methods whatever the source location.

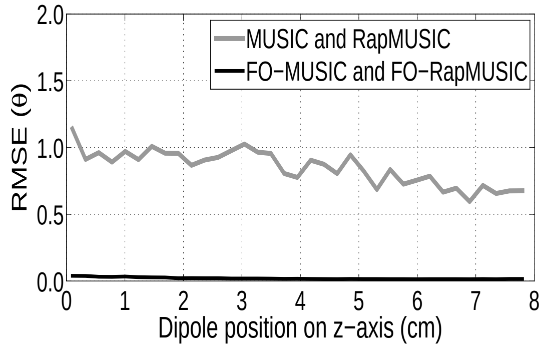


Fig. 2. Effect of dipole location on localization.

B. The case of poorly spatialy separated sources

Figure 3 presents the quantity $\mathit{RMSE}(\boldsymbol{\theta}_1)$+$\mathit{RMSE}(\boldsymbol{\theta}_2)$ at the output of the four methods as a function of the distance between two sources. One dipole was placed at position $\boldsymbol{\rho} = [0, 0, 0.8]^{\mathsf{T}}$ (in centimetres) while the other moved along the z-axis. It clearly appears on figure 3 that both FO approaches are quasi-insensitive to the distance between the dipoles contrary to SO methods. Indeed, the behaviour of the MUSIC algorithm is very affected as soon as the distance between sources decreases below 3.5 cm. Performances of the RapMUSIC are better than those of MUSIC for low distances, however they remain inferior to those given by both FO algorithms, whatever the distance is. As in section V-A, the PNL was quasi-zero for all the methods whatever the source distance.

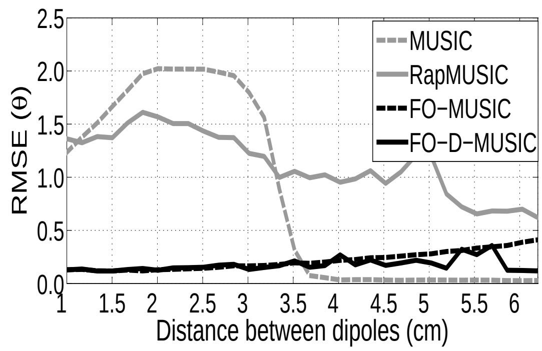


Fig. 3. Effect of distance between dipoles on localization.

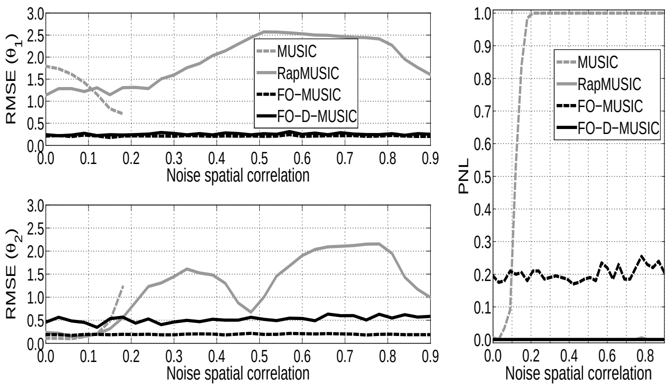
C. The case of colored noise

Both FO algorithms were compared to SO algorithms in the presence of a Gaussian noise with unknown spatial covariance. Two sources were positioned in depth such that their location vectors were given by $\boldsymbol{\rho}_1 = [0, 0, 2]^\mathsf{T}$ and $\boldsymbol{\rho}_2 = [0, 0, 4.4]^\mathsf{T}$ respectively. Figure 4 displays the variations of RMSE and PNL criteria for the four methods as a function of the noise spatial covariance factor ρ. Note that the Gaussian noise model employed in this simulation is the sum of an internal noise $\boldsymbol{\nu_{in}}$ and an external noise $\boldsymbol{\nu_{out}}$, of covariance matrices $\boldsymbol{R}_{\boldsymbol{\nu}}^{in}$ and $\boldsymbol{R}_{\boldsymbol{\nu}}^{out}$ respectively such that:

$$\boldsymbol{R}_{\boldsymbol{\nu}}^{in}(r, q) \stackrel{\text{def}}{=} \sigma^2 \delta[r-q]/2 \quad \boldsymbol{R}_{\boldsymbol{\nu}}^{out}(r, q) \stackrel{\text{def}}{=} \sigma^2 \rho^{|r-q|}/2 \quad (34)$$

where σ^2, ρ, $\boldsymbol{R}_{\boldsymbol{\nu}}(r, q) \stackrel{\text{def}}{=} \boldsymbol{R}_{\boldsymbol{\nu}}^{in}(r, q) + \boldsymbol{R}_{\boldsymbol{\nu}}^{out}(r, q)$ are the variance of total noise per sensor, the spatial covariance factor of noise and the (r, q)-th component of the total noise covariance matrix, respectively.

Figure 4(a) shows that both SO algorithms are sensitive to a Gaussian noise with unknown spatial covariance and are affected as soon as the noise spatial covariance increases beyond 0.2. Indeed, theoretically MUSIC and RapMUSIC require a perfect knowledge of the noise covariance [22]. On the contrary, FO-MUSIC and FO-D-MUSIC, since they use FO cumulants, are asymptotically insensitive to Gaussian noise, regardless of its space/time color. Computer results show that, although the PNL of RapMUSIC is quasi-zero, only FO-MUSIC and FO-D-MUSIC localize both sources with precision whatever the noise spatial covariance is. Nevertheless, for a given number of ten thousand samples, only FO-D-MUSIC among both FO methods succeeds in localizing both sources at each time (figure 4(b)).

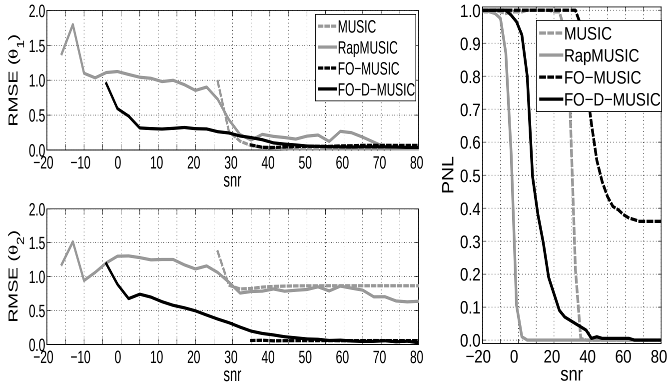


(a) RMSE criterion (b) PNL criterion

Fig. 4. Effect of colored noise on localization of 2 sources.

D. The case of non-Gaussian noise

Results show that the FO-D-MUSIC algorithm is unaffected by a Gaussian noise even when only a finite number of data samples are available (figure 4(a)). Therefore, we studied the behaviour of FO-D-MUSIC in the presence of an additive non-Gaussian noise. For that purpose, eye-blink artefacts and ElectroCardioGraphic (ECG) real signals were added to simulated background EEG signals, generated from the model [26]. This sum of signals was added to the mixture of two sources, located in depth ($\boldsymbol{\rho}_1 = [1, 0, 1]^\mathsf{T}$ and $\boldsymbol{\rho}_2 = [0.875, 0, 1.125]^\mathsf{T}$ respectively). The two sources were chosen close to each other to establish if the superiority of FO-D-MUSIC over SO MUSIC-like approaches was still valid in such a case with a non-Gaussian noise. Results are displayed in figure 5 where RMSE and PNL criteria are represented as a function of both sources SNR for the MUSIC, RapMUSIC, FO-MUSIC and FO-D-MUSIC algorithms. They show that, contrary to the FO MUSIC-like approaches, the SO ones do not succeed in localizing both sources with precision, even for a high SNR of 80 dB. Besides, unlike FO-MUSIC, FO-D-MUSIC succeeds in localizing both sources at each time as soon as the SNR increases beyond 40 dB. The fact that the PNL of FO-MUSIC does not really tend to zero as the SNR increases could be explained by the small number of samples (2500) used is this specific simulation. Even with a maximal SNR, FO-MUSIC might sometimes fail to find a solution (possibly because of errors in our FO estimates due to the small number of samples). This justifies the use of FO deflation scheme when a FO-MUSIC metric is considered. Finally, although the FO-D-MUSIC method seems to be the more efficient in this simulation, its convergence speed may be reduced by the presence of a non-Gaussian noise.



(a) RMSE criterion (b) PNL criterion

Fig. 5. Effect of non-Gaussian noise on localization of 2 sources.

E. Effect of the number of surface observations on source localization

In order to study the effect of the number of surface observations on the behaviour of the four previous MUSIC-like methods, two close sources were considered (location parameters equal to $\boldsymbol{\rho}_1 = [-1, 0, 3]^\mathsf{T}$ and $\boldsymbol{\rho}_2 = [-1.16, 0, 3.16]^\mathsf{T}$). In figure 6, the RMSE criterion at the output of RapMUSIC and FO-D-MUSIC is plotted against the number of surface observations (MUSIC and FO-MUSIC were not represented because the PNL criterion is close to one in these two cases). Contrary to FO-D-MUSIC, the RMSE criterion at the output of RapMUSIC needs at least sixty three surface observations (i.e. sixty four with the reference channel) to give accurate results and drops for a smaller number of observations. We recently reported [2] that FO-D-MUSIC encompasses MUSIC-like

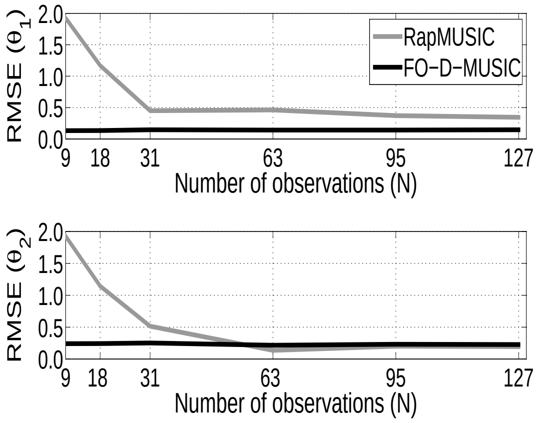


Fig. 6. Effect of the number of surface observations on localization of 2 sources.

methods when only ten surface electrodes were considered. Even with 127 EEG channels, RapMUSIC does not localize the first source as accurately as FO-D-MUSIC. As far as the PNL criterion at the output of RapMUSIC and FO-D-MUSIC is concerned, it was quasi-zero. This shows that the use of the deflation concept at second and fourth orders considerably increases the algorithm ability to localize both sources. In conclusion, the FO-D-MUSIC method outperforms FO-MUSIC as well as the classical SO MUSIC-like approaches, especially when sources are close to each other, independently from the number of observations used.

VI. Conclusion

In this paper, we propose a novel algorithm for brain current source localization, the FO-D-MUSIC method based on i) the separability of the data transfer matrix as a function of location and orientation parameters, ii) the Fourth Order (FO) virtual array theory, and iii) the deflation concept extended to FO statistics accounting for the presence of potentially but not completely statistically dependent sources. Although HO cumulants were considered for a long time as too difficult to estimate, they can be more useful than SO ones to solve inverse problems since they use, in a way, additional virtual sensors. This result was shown asymptotically in the presence of independent sources and reinforced by several simulations, performed for different numbers of samples, provided that a FO deflation scheme is used. Indeed, computer results showed the superiority of FO-D-MUSIC over FO-MUSIC (i.e. similar version of FO-D-MUSIC without the deflation scheme) and classical algorithms such as MUSIC [7] and RapMUSIC [15] for overdetermined mixtures of sources in different situations. In particular, unlike SO MUSIC-like algorithms the FO-D-MUSIC method remained unaffected by a Gaussian noise of unknown spatial covariance. Moreover, the FO deflation concept used in FO-D-MUSIC increases the probability of localizing all the sources. In addition, the FO-D-MUSIC approach shows its superiority specially when a fine resolution is required, for instance when sources are close to each other. Besides, at a similar level of performances, our method requires less surface observations than SO MUSIC-like approaches. Indeed, FO-D-MUSIC exhibits good performances for reduced number of surface observations and provides a reliable alternative when high-resolution EEG is unavailable. Our objective in the forthcoming work is i) to test the FO-D-MUSIC ability to localize more sources than surface observations, and ii) to evaluate it from real EEG data in epileptic patients in whom strong hypotheses about the localization of the epileptic zone are available.

References

[1] L. ALBERA, A. FERREOL, P. CHEVALIER, and P. COMON, “ICAR, a tool for blind source separation using fourth order statistics only,” *IEEE Transactions On Signal Processing*, 2005, to appear.

[2] L. ALBERA, A. FERREOL, D. COSANDIER-RIMELE, I. MERLET, and F. WENDLING, “Fourth order approaches for localization of brain current sources,” in *EMBC 06, 2006 IEEE 2006 International Conference of the Engineering in Medicine and Biology Society*, New York, USA, August 30 - September 3 2005.

[3] A. BELOUCHRANI, K. ABED-MERAIM, and M. G. AMIN, “Time-frequency music,” *IEEE Signal Processing Letters*, vol. 6, no. 5, pp. 109–110, May 1999.

[4] P. BERG and M. SHERG, “A fast method for forward computation of multiple-shell spherical head models,” *Electroencephalography and Clinical Neurophysiology*, vol. 90, no. 1, pp. 58–64, January 1994.

[5] P. CHEVALIER and A. FERREOL, “On the virtual array concept for the fourth-order direction finding problem,” *IEEE Transactions On Signal Processing*, vol. 47, no. 9, pp. 2592–2595, September 1999.

[6] D. COSANDIER-RIMELE, J.-M. BADIER, P. CHAUVEL, and F. WENDLING, “A physiologically plausible spatio-temporal model for depth-EEG signals recorded with intracerebral electrodes in human partial epilepsy,” *IEEE Transactions On Biomedical Engineering*, vol. 54, no. 3, pp. 380–388, March 2007.

[7] E. FERRARA and T. PARKS, “Direction finding with an array of antennas having diverse polarizations,” *IEEE Transactions On Antennas Propagation*, vol. 31, pp. 231–236, March 1983.

[8] F. R. GANTMACHER, *The Theory of Matrices, vol. 1*. New York: Chelsea, 1959.

[9] G. H. GOLUB and C. F. V. LOAN, *Matrix computations, second edition*. The Johns Hopkins University Press, Baltimore, MD, 1989.

[10] J. HAUEISEN, C. RAMON, M. EISELT, H. BRAUER, and H. NOWAK, “Influence of tissue resistivities on neuromagnetic fields and electric potentials studied with a finite element model of the head,” *IEEE Transactions on Biomdecial Engineering*, vol. 44, no. 8, pp. 727–735, August 1997.

[11] R. A. HORN and C. R. JOHNSON, *Topics in Matrix Analysis*. Cambridge University Press, New York, 1999.

[12] H. KRIM and M. VIBERG, “Two decades of signal processing: The parametric approach,” *IEEE Signal Processing Magazine*, vol. 13, pp. 67–94, July 1996.

[13] P. McCULLAGH, *Tensor Methods in Statistics*. Chapman and Hall, Monographs on Statistics and Applied Probability, 1987.

[14] C. M. MICHEL, M. M. MURRAY, G. LANTZ, S. GONZALEZ, L. SPINELLI, and R. GRAVE DE PERALTA, “EEG source imaging,” *Clinical Neurophysiology*, vol. 115, no. 10, pp. 2195–2222, October 2004.

[15] J. C. MOSHER and R. M. LEAHY, “Source localization using Recursively Applied and Projected (RAP) MUSIC,” *IEEE Transactions On Signal Processing*, vol. 47, no. 2, pp. 332–340, February 1999.

[16] J. C. MOSHER, R. M. LEAHY, and P. S. LEWIS, “EEG and MEG: Forward solutions for inverse methods,” *IEEE Transactions On Biomedical Engineering*, vol. 46, no. 3, pp. 245–259, March 1999.

[17] S. K. OH and C. K. UN, “A sequential estimation approach for performance improvement of eigenstructure-based methods in array processing,” *IEEE Transactions On Signal Processing*, vol. 41, pp. 457–463, 1993.

[18] R. OOSTENVELDA and P. PRAAMSTRA, “The five percent electrode system for high-resolution EEG and ERP measurements,” *Clinical Neurophysiology, Elsevier*, vol. 112, no. 4, pp. 713–719, April 2001.

[19] B. PORAT and B. FRIEDLANDER, “Direction finding algorithms based on high-order statistics,” *IEEE Transactions On Signal Processing*, vol. 39, no. 9, pp. 2016–2024, September 1991.

[20] S. RUSH and D. A. DRISCOLL, “EEG electrode sensitivity – An application of reciprocity,” *IEEE Transactions On Biomedical Engineering*, vol. 38, pp. 15–22, January 1969.

[21] R. O. SCHMIDT, “A signal subspace approach to multiple emitter location and spectral estimation,” Ph.D. dissertation, Stanford University, November 1981.

[22] ——, “Multiple emitter location and signal parameter estimation,” *IEEE Transactions On Antennas Propagation*, vol. 34, no. 3, pp. 276–280, March 1986, reprint of the original 1979 paper from the RADC Spectrum Estimation Workshop.

[23] K. SEKIHARA, S. NAGARAJAN, D. POEPPEL, and Y. MIYASHITA, “Time-frequency MEG-MUSIC algorithm,” *IEEE Transactions On Medical Imaging*, vol. 18, no. 1, pp. 92–97, January 1999.

[24] c. S.I. Gon J. D. MUNCK, J. VERBUNT, F. BIJMA, R. HEETHAAR, and F. L. D. SILVA, “In vivo measurement of the brain and skull resistivities using an eit-based method and realistic models for the head,” *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 6, pp. 754–767, June 2003.

[25] P. STOICA, P. HANDEL, and A. NEHORAI, “Improved sequential music,” *IEEE Transactions On Aerospace and Electronic Systems*, vol. 31, no. 4, pp. 1230–1239, October 1995.

[26] F. WENDLING, J. BELLANGER, F. BARTOLOMEI, and P. CHAUVEL, “Relevance of nonlinear lumped-parameter models in the analysis of depth-EEG epileptic signals,” *Biological Cybernetics*, vol. 83, pp. 367–378, 2000.

[27] Z. ZHANG, “A fast method to compute surface potentials generated by dipoles within multilayer anisotropic spheres,” *Physics in Medicine and Biology*,, vol. 40, no. 3, pp. 335–349, March 1995.

Laurent Albera was born in Massy, France, in 1976. After a DESS in Mathematics, He received in 2001 the DEA’s degree in Automatic and Signal Processing from the University of Science of Orsay (Paris XI). From 2001 to 2004, he was committed for a study within the framework of a Research Contract CIFRE (Convention Industrielle de Formation par la Recherche) between Thales Communications (formerly Thomson-CSF), the french National Centre for Scientific Research CNRS and the University of Nice Sophia-Antipolis. He received in 2003 the PH.D. degree in Sciences from the University of Nice Sophia-Antipolis, France. He is now Assistant Professor at the University of Rennes 1 and is affiliated with the research group LTSI (Laboratoire Traitement du Signal et de l’Image). His research interests include Multilinear Algebra and High-Order Statistics in order to perform Independent Component Analysis (ICA) and source localization.

Anne Ferréol was born in 1964 in Lyon, France. She received the Master Science degree from ICPI-Lyon and the Mastère degree from Ecole Nationale Supérieure des Télécommunications (ENST) in 1988 and 1989 respectively. She received in 2004 the PH.D. degree in Sciences from Ecole Normale Supérieur de Cachan. Since 1989 she has been working at THALES-Communications, in the array processing department. Her current interests concern DOA estimation and blind source separation.

Delphine Cosandier-Rimélé received the Engineering diploma in Electronics and Computer Engineering from the Ecole Nationale Supérieure des Sciences Appliquées et de Technologie (Lannion, France) in 2003, the M.S. degree and the PhD degree in Signal Processing and Telecommunications from the University of Rennes 1 (Rennes, France) in 2003 and 2007 respectively. Her research works deal with the reconstruction of surface and intracerebral EEG signals from physiologically-relevant models.

Isabelle Merlet was born in 1969 in Paris, France. She received in 1997 the PH.D Degree in Neurosciences from Lyon 1 University, France. Since 1997 she has been working at the Epilepsy departments of the Montreal Neurological Institute in Canada, and at the Neurological Hospital of Lyon in France. In 2005 she joined the INSERM laboratory U642 in Rennes. She is currently working on source localization, imaging, and signal processing in Epilepsy.

Fabrice Wendling was born in France in 1967. He received the Engineer Diploma in Biomedical Engineering from the University of Technology of Compiegne, France (1990), the Master’s degree in Bioengineering from the Georgia Institute of Technology, Atlanta, USA (1991), the PhD degree in Signal Processing (1996) and the Accreditation to supervise research (2003) from the University of Rennes, France. Since 1998, he has been working as a research scientist for the French NIH (INSERM) at the Laboratory of Signal and Image Processing (LTSI), in Rennes. Since 2001, he is responsible for a research project entitled ”Dynamics of neuronal systems in epilepsy and cognition” at the LTSI. His research works deal with the analysis of electrophysiological signal and the modeling of brain activity in the context of epilepsy.