Tensor Analysis and Fusion of Multimodal Brain Images

This paper presents a general framework for tensor analysis of single-modality model inversion and multimodal data fusion using neuroimaging data as an example.

By Esin Karahan, Pedro A. Rojas-López, Maria L. Bringas-Vega, Pedro A. Valdés-Hernández, and Pedro A. Valdes-Sosa

ABSTRACT | Current high-throughput data acquisition technologies probe dynamical systems with different imaging modalities, generating massive data sets at different spatial and temporal resolutions—posing challenging problems in multimodal data fusion. A case in point is the attempt to parse out the brain structures and networks that underpin human cognitive processes by analysis of different neuroimaging modalities (functional MRI, EEG, NIRS, etc.). We emphasize that the multimodal, multiscale nature of neuroimaging data is well reflected by a multiway (tensor) structure where the underlying processes can be summarized by a relatively small number of components or "atoms." We introduce Markov–Penrose diagrams—an integration of Bayesian DAG and tensor network notation in order to analyze these models. These diagrams not only clarify matrix and tensor EEG and fMRI time/frequency analysis and inverse problems, but also help understand multimodal fusion via multiway partial least squares and coupled matrix-tensor factorization. We show here, for the first time, that Granger causal analysis of brain networks is a tensor regression problem, thus allowing the atomic decomposition of brain networks. Analysis of EEG and fMRI recordings shows the potential of the methods and suggests their use in other scientific domains.

KEYWORDS | Autoregressive processes; Bayesian models; Bayesian statistics; EEG/fMRI; electroencephalography; Granger causality; magnetic resonance imaging; multidimensional systems; multimodal data; N-PLS; PARAFAC; tensor decomposition; tensor network

I. INTRODUCTION

Without doubt this is the era of "big data." Technological advances in data acquisition technology are spurring the generation of unprecedented massive data sets, thus posing a permanent challenge to data analysts. Recent international efforts are marshalling the use of a bewildering array of different technologies to acquire high-throughput multimodal information about real-world systems. Examples of the systems and modalities probed are the Internet [1], geophysical data [2], and the human genome [3].

The human brain is today perhaps the most challenging biological object under study and has been pushed recently to the forefront of public awareness [4]–[6]. This interest stems from the fact that identification of the brain structures involved in cognitive processes would not only yield essential understanding about the human condition, but would also provide leverage to deal with the staggering global burden of disease for brain disorders [7]. The challenge of data sets stems from the fact that the brain is a highly nested set of interacting dynamical systems—from the subcellular level to the whole system. In turn, each level is actively being probed with an impressive arsenal of different measurement and imaging modalities. Some examples of system level measurements are high precision postmortem anatomy [8], diffusion weighted imaging

E. Karahan is with the Institute of Biomedical Engineering, Bogazici University, Istanbul 34684, Turkey (e-mail: esin.karahan@boun.edu.tr).
P. A. Rojas-López and **P. A. Valdés-Hernández** are with the Cuban Neuroscience Center, Cubanacan, Playa, Havana 10600, Cuba (e-mail: pedro.rojas@cneuro.edu.cu; multivac@cneuro.edu.cu).
M. L. Bringas-Vega is with the Key Laboratory for NeuroInformation of Ministry of Education, Center for Information in Medicine, University of Electronic Science and Technology of China, Chengdu 610054, China (e-mail: maria@uestc.edu.cn; maluisabringas@yahoo.com).
P. A. Valdes-Sosa is with the Key Laboratory for NeuroInformation of Ministry of Education, Center for Information in Medicine, University of Electronic Science and Technology of China, Chengdu 610054, China, and the Cuban Neuroscience Center, Cubanacan, Playa, Cuba (e-mail: pedro@uestc.edu.cn; peter@cneuro.edu.cu).
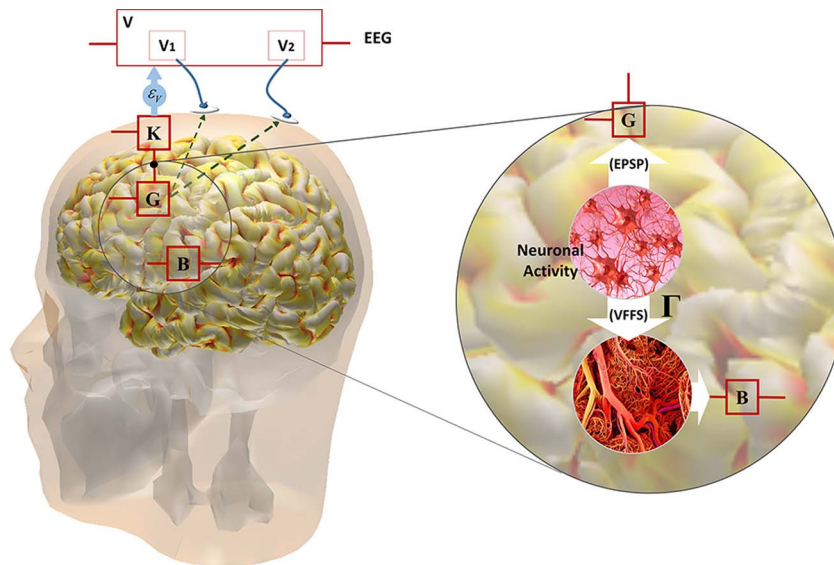
Fig. 1. *Neural origin of EEG and the BOLD signals. Neural activity in a cortical patch (amplified to the right of the figure) generates two streams of events. On the one hand, it induces an ensemble of postsynaptic potentials (EPSPs) that creates primary current density G in gray matter. The second stream of events is the emission of Γ, a vasoactive feedforward signal (VFFS), that triggers a delayed and sluggish increase in blood flow resulting in an excess of oxy-hemoglobin which produces the BOLD signal B (fMRI). As shown in the left part of the figure, the cortical patch is embedded in a volume conductor (the head). As a consequence, G is projected at electrodes (V1 and V2) placed at the scalp as determined by the lead field matrix K, which encapsulates the effects of the volume conductor. The generation of V is further specified by the use of the Markov–Penrose notation defined in this review [Fig. 4(d)]. Modified from [23]. A detailed model concerning the generation of B can be found in Fig. 4(e).*

(DWI) [9], [10], functional magnetic resonance imaging (fMRI) [11], electroencephalography (EEG) [12], and near-infrared spectroscopy (NIRS) [13]. These are just some of the techniques that are rapidly populating publicly available databases (see, for example, those listed at www.incf.org). For an overview of the methods providing data see [14].

In this review, brain information will serve as an example of the complexities encountered in analyzing big data multimodal sets, as well as to illustrate possible strategies (tensor models) to cope with these issues. These approaches, of course, may be applied to other knowledge domains.

Typical issues that arise when analyzing neuroscience data are as follows.

- Brain data are highly multidimensional and multimodal. Each imaging modality is always an indirect measurement of the underlying dynamical systems that are of interest. Thus, we are required not only to solve multiple inverse problems (one for each measurement modality) but also to carry out multimodal fusion.

- Each modality is recorded at different spatial and temporal resolutions and reflects different physiological processes. This poses challenging problems for multimodal data fusion.

- To compound the complexity of brain data, the analysis is not only required to identify specific

components and functions of the system but also to elucidate their interactions (i.e., to identify patterns of brain connectivity). This objective derives from consensus that neural computations are not the activity of relatively isolated neural masses but rather the coordinated activity of dynamically changing networks that involve huge amounts of neurons [15].

Though the methods we describe here for inverse problems and multimodal fusion are quite general, we will confine examples and detailed discussions to two types of brain imaging modalities: EEG and fMRI [blood oxygen level dependent (BOLD)]. The choice of these is based on their common physiological basis and complementarity as we will explain shortly.

As can be seen from Fig. 1, both types of measurements arise from neural activation which produces, on the one hand, primary current densities ($\mathbf{G}$) that are reflected on the scalp as EEG ($\mathbf{V}$), and, on the other hand, as a vasoactive feedforward signal (VFFS, $\mathbf{\Gamma}$) that produces a rush of blood rich in oxy-hemoglobin that is measurable as a local BOLD signal ($\mathbf{B}$) with magnetic resonance imaging being the best known method of fMRI. The physical models describing the generation of $\mathbf{V}$ and $\mathbf{B}$ from $\mathbf{G}$ and $\mathbf{\Gamma}$ are known as the respective forward problems.

It is precisely due to the effect of these forward models that we must solve modality-specific inverse problems in order to uncover the underlying brain activation and

connectivity. In the case of EEG, the objective is to overcome the spatial smearing of **G** that is due to volume conduction. For the EEG frequency ranges often studied, the quasi-static approximation [16] guarantees that both forward and inverse problems are linear. EEG frequency components, or rhythms, have been found to be important biomarkers of normal and abnormal brain states. A review of the basic facts about brain oscillations can be found in [17].

In the case of fMRI, the inverse problem compensates for the temporal blurring of the BOLD signal produced by the hemodynamic response function. Strictly speaking, the BOLD forward and inverse problems are nonlinear [18]. Nevertheless, a useful first approximation is to linearize these equations (see [19] for a recent example).

Thus, in this review, we will be dealing with a dual set of linear forward and inverse problems. It is not a coincidence that much of the EEG and fMRI analysis literature reads as exercises in the matrix analysis of ill-posed linear inverse problems. This has been beneficial due to the large and rapidly expanding toolkit of matrix-based method analysis. As we will show in Section III, the indeterminacies of matrix decompositions have required additional assumptions in order to obtain well-defined models tuned to the data. One such approach is independent component analysis (ICA) which has become quite popular in both EEG and fMRI analyses [20]–[22].

Coming back to the generative models depicted in Fig. 1, it is clear that EEG and fMRI have complementary strengths. EEG inverse solutions have low spatial resolution, while BOLD is weak in temporal granularity. Thus, these modalities seem ideally matched for image fusion in order to create a brain mapping modality that simultaneously has high spatial and temporal resolution [23]–[25]. As with the modality-specific inverse problems, a majority of the modeling strategies for this type of fusion are based on matrix methods [24], [26]–[29]. In particular, ICA methods for EEG/fMRI fusion are also very popular and have been recently reviewed in [30].

An alternative to purely matrix-based EEG/fMRI inverse solutions and fusion methods is to take full advantage of the inherent structure of the multimodal brain data, which many times is actually recorded as a multidimensional array or a tensor. It was recognized early [31] that tensors could be, under mild conditions, uniquely decomposed, a property applied with great advantage to the analysis of EEG in the seminal work of [32] and [33]. Following this lead, tensor-based data analysis has been vigorously developed in the past few years [34]–[37]. The applications to neuroimaging data seem quite promising and have been extensively reviewed in [38]–[41]. Of course, tensor methods have penetrated many fields. Several examples of their use in multimodal fusion outside brain science can be found in [42].

Rather than replicating the material covered in the excellent reviews mentioned above, our purpose here is to illustrate how tensor methods can enrich the statistical methodology underlying EEG/fMRI electrophysiological inverse solutions, brain connectivity, and image fusion. In order to do so, we integrate and generalize the work initiated in [23] and [43]–[49]. To facilitate visual representation of models, we introduce the Markov–Penrose diagram (the M–P diagram), a combination of Penrose diagrams with notation from the theory of directed acyclic graphs (DAGs) that we use to represent Bayesian tensor models and clarify inferential steps.

We will proceed as follows. We set down the basic definitions and notations for tensors and their operations in Section II. This will allow us to review matrix-based EEG/fMRI analysis methods in Section III and to argue for the need of multilinear approaches to the field. We then apply the ideas and notations of Section II successively to the EEG analysis (Section IV), brain connectivity (Section V), and EEG/fMRI fusion (Section VI). We introduce nonlinear extensions of tensor methods in Section VII. We discuss algorithms in Section VIII, where we mention software packages for tensor-based problems. Finally, in Section IX, we present some conclusions.

II. TENSOR DIAGRAMS AND OPERATIONS

Tensors are a generalization of vectors and matrices to higher dimensions, being embodied in different programming languages as multidimensional arrays. The order of a tensor is the number of its dimensions. We refer the reader to Tables 1 and 2 for a summary of the tensor notation and operations used in this review.

A. Penrose Diagrams

Tensor objects and operations may be represented by Penrose diagrams (the P diagram) that were inspired by Penrose's original work in [50] on theoretical physics and later adapted, with modifications, for tensor networks. Examples of P diagrams can be found in [41], [51], and [52]. Expressions in usual mathematical notation for vectors, matrices, and tensors as well as their corresponding Penrose diagrams are shown in Fig. 2(a)–(c). In this type of diagram, tensors are denoted as nodes and each dimension is depicted as a line (link) leaving the node. The model order of a tensor can be read off from its Penrose diagram as the number of dangling lines. Note that any tensor may have any number of additional "singleton" or virtual one-element dimensions (e.g., $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is the same tensor as $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N \times 1 \times \cdots \times 1}$). Tensors which have random values as elements are shown as circular nodes. Other nodes with special properties are depicted in Fig. 2. Those with constant/deterministic elements are shown as rectangular nodes [Fig. 2(d)]. If all of the elements of a tensor are nonnegative, this property (or for that matter any other relation) is depicted within the node outline [Fig. 2(e)]. A tensor with orthogonality imposed on

Table 1 Tensor Concepts

	Definition	Figure
x	Scalar	
$\mathbf{x} \in \mathbb{R}^{I \times 1}$	Vector	Fig. 2(a)
$\mathbf{X} \in \mathbb{R}^{I \times J}$	Matrix	Fig. 2(b)
$\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$	N Dimensional Tensor	Fig. 2(c)
$\mathbf{X}(i,:)$	i^{th} row of a matrix $\mathbf{X} \in \mathbb{R}^{I \times J}$	
$\mathcal{X}(:,:,k)$	Frontal slice (or face) of a three dimensional tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$	
$\mathcal{X} \geq 0$	Nonnegative tensor, if $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, $\mathcal{X}(i,j,k) \geq 0$ for $i \leq I, j \leq J, k \leq K$	Fig. 2(e)
$\mathbf{X}^T, \mathbf{X}^H$	Transpose, Conjugate (Hermitian) transpose	
$\mathbf{I}, \mathcal{I}$	Identity matrix, tensor	
$log, \mathbf{SVD}$	Natural logarithm, Singular Value Decomposition	

a given dimension will be shown with a rectangular box on that dimension [Fig. 2(f)].

Any junction between lines of two or more tensors denotes an operation between them that may result in a lower order tensor. A list of tensor operations used in this review, though of course there are many more, is given in Table 2. P diagrams of the related tensor operations are given in Fig. 3.

For clarity, we sometimes emphasize a particular tensor, the result of several operations, by surrounding it with triangular or rectangular shapes, as shown in Fig. 3(e). Note that the dimensionality of the resultant tensor will be clear from the lines leaving the shape.

B. Markov–Penrose Diagrams

Missing from the usual P diagrams are the concepts of probabilistic dependency. Probabilistic graphical models are extremely useful for making the conditional dependence in a statistical model explicit [53], in particular directed graphical models (DAGs) or Bayesian networks for graphical models.

We emphasize the need for probabilistic models which includes prior distributions in order to deal with the issues arising from high dimensionality of tensorial models, such as indeterminacy. In this review, we consider a Bayesian approach, in which we represent entities of our model according to probabilistic functions and define probabilistic relations between them in a Bayesian network fashion. In this way, we can interpret the results of fitting the variables of these models as maximum *a posteriori* (MAP) estimates, thus linking both Bayesian and tensorial frameworks.

A junction between the links of two tensors in a P diagram signifies undirected arithmetic operations [as in Fig. 3(a)]. By contrast, an arrow between two variable nodes in a DAG signifies directed conditional dependence.

To specify statistical models for tensors we need both types of links. To summarize, P diagrams are useful for the visualization of tensors and tensorial operations whereas DAGs are useful for showing the probability models.

This has prompted us to define a notation that incorporates both types of links, as shown in Fig. 4(a)–(c). Here a directed arrow denotes a conditional probabilistic dependency between tensors.

Functional penalties $\pi(\mathbf{x})$ on nodes (such as those that will be described in Section III) are depicted inside a special box shape node from which an arrow is directed to the constrained node. Note that this implies that the node variable is distributed as $p(\pi(\mathbf{x})) = C \exp(-\pi(\mathbf{x}))$ [see Fig. 4(c)].

To contrast DAG with the M–P diagram refer to Fig. 4(d) and (e) that displays two different graphical representations for the forward-matrix-based models that will be introduced for EEG and fMRI in Section III. It is clear that the DAG notation is operation blind; we can see the dependence between variables, but not the actual arithmetic. In M–P notation, we can show the same conditional dependence structure as in DAG notation, but we can also infer the formula for the model in this graph using the Penrose notation for operations.

There are other types of decompositions that might be applied to these models, such as Tucker decompositions [36], that are easily shown with this notation but, for reasons of space, will not be included in this review.

Note that the inclusion in some DAG-type arrow of an error term ε indicates measurement noise. Of course, in reality, these can be drawn from any probability distribution but, for ease of exposition, as in the matrix case, will be considered as a tensor of adequate dimensions containing identically and independent Gaussian variates.

The close formal affinity between certain types of tensor networks and graphical models studied in the machine learning community has already been pointed out in [41, Table 2]. Critch [54] and Critch and Morton [55] also highlight the similarities between particular types of tensor networks (depicting matrix product states) with hidden Markov models. Nevertheless, the links in these two types of structures are of a fundamentally different mathematical nature, and a distinctive notation was not considered.

Table 2 Tensor Operations

| General Tensor Operators	Type	Definition	Figure
Tensor contraction is a generalization of matrix multiplication for tensors.			
$\mathcal{Z} = \mathcal{X} \bullet_{\{J_1, \cdots, J_M\}} \mathcal{Y}$	Binary	Contraction of $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_M}$ and $\mathcal{Y} \in \mathbb{R}^{J_1 \times \cdots \times J_M \times K_1 \times \cdots \times K_P}$ over common dimensions $J_1, \ldots, J_M$ gives $\mathcal{Z} \in \mathbb{R}^{I_1 \times \cdots \times I_N \times K_1 \times \cdots \times K_P}$: $$\mathcal{Z}(i_1, \cdots, i_N, k_1, \cdots, k_P) = \sum_{j_1 \cdots j_M = 1}^{J_1 \cdots J_M} \mathcal{X}(i_1, \cdots, i_N, j_1, \cdots, j_M) \mathcal{Y}(j_1, \cdots, j_M, k_1, \cdots, k_P)$$	Fig. 3 (a), (b)
$\mathcal{Z} = [\![\mathcal{X}_m]\!]_{m=1:M}^{\{J_1, \cdots, J_N\}}$	Set	Contraction of a series of tensors defined as $\mathcal{X}_m \in \mathbb{R}^{I_m \times J_1 \times \cdots J_N}$ for $m = 1, \ldots, M$ is equal to: $$\mathcal{Z}(i_1, \cdots, i_M) = \sum_{j_1 \cdots j_N = 1}^{J_1, \cdots J_N} \mathcal{X}_1(i_1, j_1, \cdots, j_N) \cdots \mathcal{X}_M(i_M, j_1, \cdots, j_N)$$	
Tensor concatenation is the merging of two tensors whose dimensions should be the same except for the concatenation index.			
$\mathcal{Z} = \mathcal{X} |_{\{J|K\}} \mathcal{Y}$	Binary	Concatenation of $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots I_{n-1} \times J \times I_{n+1} \cdots \times I_N}$ and $\mathcal{Y} \in \mathbb{R}^{I_1 \times \cdots I_{n-1} \times K \times I_{n+1} \cdots \times I_N}$ on the Jth and Kth dimensions gives $\mathcal{Z} \in \mathbb{R}^{I_1 \times \cdots I_{n-1} \times (J+K) \times I_{n+1} \cdots \times I_N}$.	Fig. 3 (c) ,(d)
$\mathcal{Z} = [\mathcal{X}_m]_{m=1:M}^{\{J_1| \ \ |J_M\}}$	Set	Concatenation of a series of tensors defined as $\mathcal{X}_m \in \mathbb{R}^{I_1 \times \cdots I_{n-1} \times J_m \times I_{n+1} \cdots \times I_N}$ for $m = 1, \ldots, M$ over J_mth dimensions will give $\mathcal{Z} \in \mathbb{R}^{I_1 \times \cdots I_{n-1} \times (J_1 + \ldots + J_M) \times I_{n+1} \cdots \times I_N}$.	
Tensor convolution is a generalization of convolution for tensors.			
$\mathcal{Z} = \mathcal{X} *_{\{j_1, \cdots, j_M\}} \mathcal{Y}$	Binary	Convolution of $\mathcal{X} \in \mathbb{R}^{J_1 \times \ \times J_M}$ with $\mathcal{Y} \in \mathbb{R}^{J_1 \times \cdots \times J_M}$ over specified dimensions $J_1, \ldots, J_M$ gives $\mathcal{Z} \in \mathbb{R}^{J_1 \times \ \times J_M}$: $$\mathcal{Z}(j_1, \cdots, j_M) = \sum_{j_1', \cdots, j_M' = 1}^{J_1, \cdots J_M} \mathcal{X}(j_1', \cdots, j_M') \mathcal{Y}(j_1 - j_1', \cdots, j_M - j_M')$$	
$\mathcal{Z} = \langle \mathcal{X}_m \rangle_{m=1:M}^{\{J_1, \cdots, J_N\}}$	Set	Convolution of a series of tensors defined as $\mathcal{X}_m \in \mathbb{R}^{J_1 \times \cdots \times J_N}$ for $m = 1, \ldots, M$ is equal to: $$\mathcal{Z}(j_1, \cdots, j_N) = \sum_{j_1', \cdots, j_N' = 1}^{J_1, \cdots J_N} \mathcal{X}_1(j_1', \cdots, j_N') \cdots \mathcal{X}_M(j_1 - j_1', \cdots, j_N - j_N')$$	
Kruskal operator			
$\mathcal{X} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$ or $\mathcal{X} = \mathbf{A} \bullet_{\{R\}} \mathbf{B} \bullet_{\{R\}} \mathbf{C}$		Constructs a tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ from matrices $\mathbf{A} \in \mathbb{R}^{I \times R}, \mathbf{B} \in \mathbb{R}^{J \times R}, \mathbf{C} \in \mathbb{R}^{K \times R}$ with the same number of columns R, [36]. The PARAFAC model is defined in terms of this operator, R being the number of **atoms** and each of the matrices characterizing a certain **signature**. (All of the factor matrices are normalized to unit norm except for the first matrix.)	Fig. 5
t-Operators			
Tensor-product (t-product) is a special type of product defined in [101].			
$\mathcal{Z} = \mathcal{X} *_{\{K\}} \mathcal{Y}$		t-product of three dimensional tensors $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ and $\mathcal{Y} \in \mathbb{R}^{J \times L \times K}$ is equal to $\mathcal{Z} \in \mathbb{R}^{I \times L \times K}$ and defined by using convolution operation: $$\mathcal{Z}(:,:,k) = \sum_{k'=1}^{K} \mathcal{X}(:,:,k') \bullet_{\{J\}} \mathcal{Y}(:,:,k-k')$$	
t-SVD gives an optimal approximation of a tensor in the Frobenius norm of the difference, [101].			
$\mathcal{X} = \mathcal{U} *_{\{K\}} \mathcal{D} *_{\{K\}} \mathcal{V}^T$		t-SVD of $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ is $\mathcal{X} = \mathcal{U} *_{\{K\}} \mathcal{D} *_{\{K\}} \mathcal{V}^T$ where $\mathcal{U} \in \mathbb{R}^{I \times I \times K}, \mathcal{V} \in \mathbb{R}^{J \times J \times K}$ are t-orthogonal tensors and $\mathcal{D} \in \mathbb{R}^{I \times J \times K}$ is a tensor with diagonal faces.	
Tensor Nuclear Norm (TNN)			
$\|\mathcal{X}\|_\circledast$		**TNN** of $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ is $\|\mathcal{X}\|_\circledast = \sum_{i=1}^{\min(I,J)} \sum_{k=1}^{K} \breve{\mathcal{D}}(i,i,k)$	
$\mathrm{MatVec}(\mathcal{X})$		For $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, $\mathrm{MatVec}(\mathcal{X}) = \left(\mathcal{X}(:,:,1) \quad \mathcal{X}(:,:,2) \quad \cdots \quad \mathcal{X}(:,:,K) \right)^T$	
$\mathrm{tplz}(\mathcal{X})$		For $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ $$\mathrm{tplz}(\mathcal{X}) = \begin{pmatrix} \mathcal{X}(:,:,1) & \mathcal{X}(:,:,2)^H & \cdots & \mathcal{X}(:,:,K)^H \\ \mathcal{X}(:,:,2) & \mathcal{X}(:,:,1) & \cdots & \mathcal{X}(:,:,K-1)^H \\ \vdots & \ddots & \ddots & \vdots \\ \mathcal{X}(:,:,K) & \mathcal{X}(:,:,K-1) & \ddots & \mathcal{X}(:,:,1) \end{pmatrix}$$	
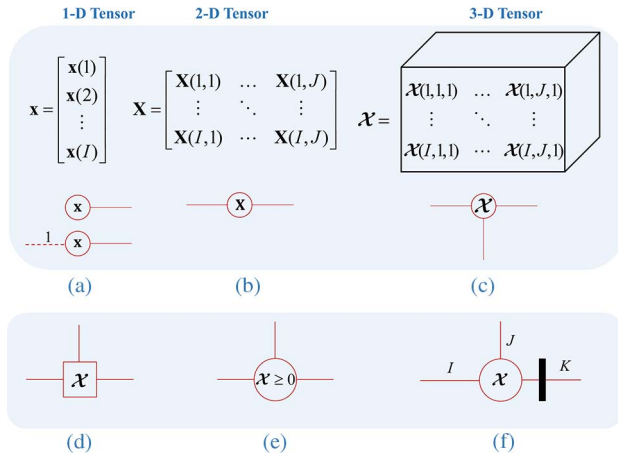
Fig. 2. *Penrose diagrams. (a)–(d) Graphical depictions of tensors of different orders (i.e., the number of dimensions) as circles, when referring to unobserved variables, and squares, when denoting constant tensors. The symbol used to denote the tensor is contained in a circle or a square. Note that the number of dimensions of a tensor is shown as the lines that leave it. When necessary, the number of elements of a dimension will be indicated besides the link that denotes it. Singleton dimensions (dimensions with only one level) are shown as dashed lines with number 1. (e) Nonnegativity and other relations for tensor elements indicated with the usual mathematical notation in a circle or a square. (f) Orthogonality constraints are shown as squared bars on the orthogonal dimensions.*

Note that we will be concentrating our analysis on linear tensor models for the sake of simplicity. There is a rapidly growing field of nonlinear methods for tensor analysis which can be incorporated into our framework. We give some examples when appropriate and outline a tensor approach to nonlinear neural mass models in Section VII.

III. MATRIX-BASED EEG/fMRI ANALYSIS

There is a huge literature on the physiological origin of EEG and fMRI. The interested reader can consult [56]–[58] for recent reviews on EEG, as well as [59] for fMRI. Equally large is the work dedicated to neurovascular coupling and EEG–fMRI fusion [60], [61]. The known facts have generated many different types of models. We here only choose, for purpose of illustration, the simplest models for the studied phenomena in order to illustrate matrix and tensor techniques. Of course, current biophysical modeling of these phenomena varies from the simple model we use for illustration to those based on nonlinear random differential equations and neural masses [62], [23]. Of great theoretical and practical use is the family of DCM [63], [64]. In fact the field is now moving toward neural field models whose mathematical underpinning are random partial evolution equations. We will touch upon these in Section VII.

We now formalize the EEG forward and inverse problems outlined in Section I and presented in Fig. 1. Both

EEG and fMRI are vector valued time series, collected into matrices (i.e., space × time).

Let us represent the recorded EEG by the matrix $\mathbf{V} \in \mathbb{R}^{I_E \times I_T}$, where I_E denotes the number of electrodes in the recording sensor array placed on the scalp, and I_T denotes the total number of observations obtained during the recording epoch. The sampling rate of EEG is typically in the range of around 1 kHz.

BOLD measurements are available for all brain voxels in an image, but we will assume that standard preprocessing
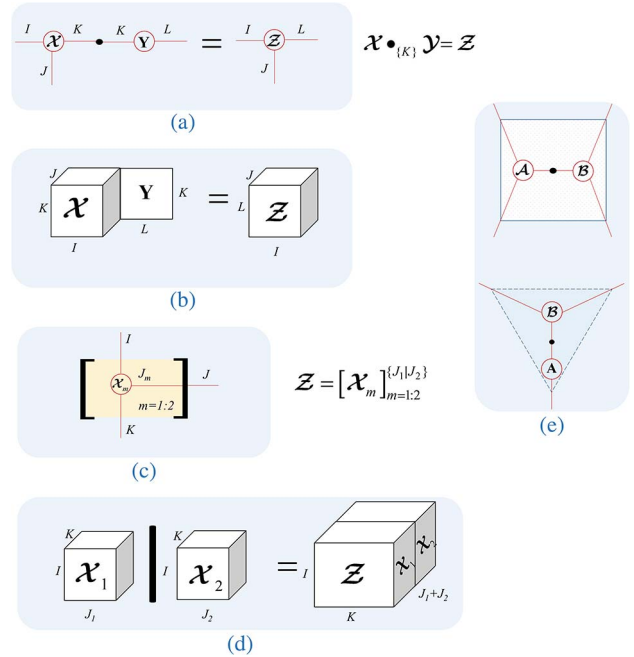

Fig. 3. *Tensor operation with Penrose (P) diagrams. (a) and (b) Graphical definitions of the contraction operator of $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ with $\mathcal{Y} \in \mathbb{R}^{K \times L}$ giving $\mathcal{Z} \in \mathbb{R}^{I \times J \times L \times K}$. This operation is denoted with a black dot in a junction between the lines representing the dimension K on both tensors. See Table 2 for a detailed formulation. (a) P diagram. (b) Three-dimensional representation of contraction of the same tensors. (c) and (d) Graphical definitions of the tensor concatenation operator is exemplified between two tensors of the identical number of dimensions and elements per dimension with at most one mode being different, in this case concatenation $\mathcal{X}_1 \in \mathbb{R}^{I \times J_1 \times K}$ and $\mathcal{Y}_2 \in \mathbb{R}^{I \times J_2 \times K}$ gives $\mathcal{Z} \in \mathbb{R}^{I \times (J_1+J_2) \times K}$. As can be noted, the resulting tensor preserves all the dimensions of the original except by the differing dimensions which are augmented with their sum. We have defined two types of operators for concatenation: the first is a binary operator, defined for two tensors, and the second is the concatenation operation on a set of tensors. See Table 2 for a detailed formulation. (c) On the left, we have the P diagram for the concatenation operator on a set of tensors, in this case $\mathcal{X}_1$ and $\mathcal{X}_2$, which are indexed by the m variable as noted in the lower right side where we define the limits of the indexed set. It is to be noted that in this operation the dimension in which concatenation occurs is the one showing the set index, in this case J_m, and outside of the bracket denoting the operation, the name of the dimension changes to $J = \sum_{m=1}^{2} J_m$. Mathematical formulation is shown on the right. (d) 3-D visual representation of the binary concatenation. (e) Tensors resulting from an operation involving other tensors can be surrounded by shapes (e.g., triangles, squares) to highlight the resulting tensor, rather than its parts.*
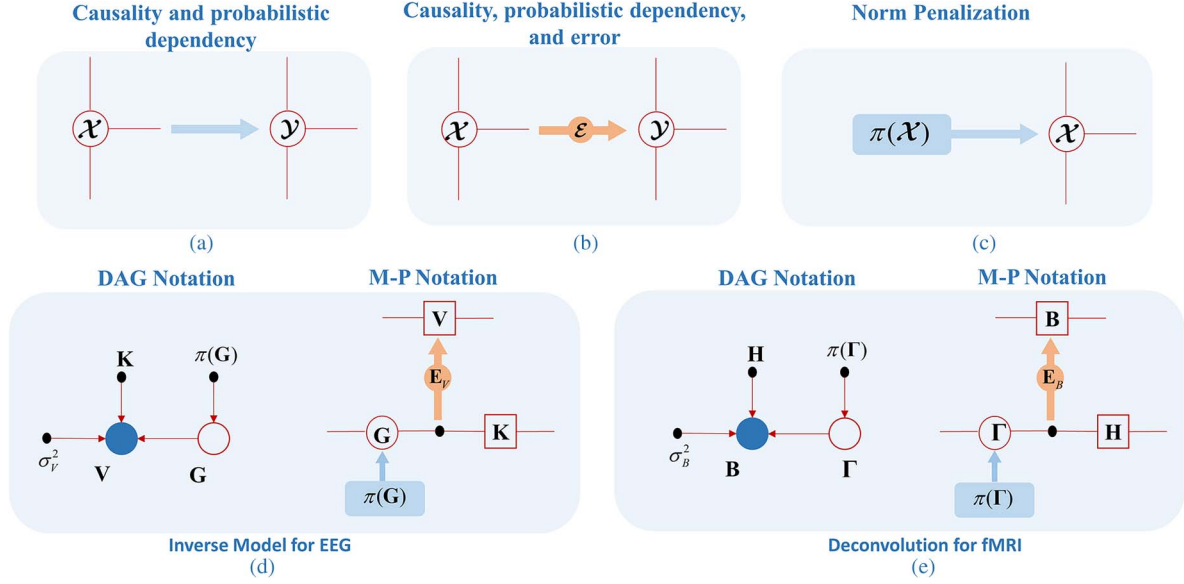
Fig. 4. *Markov–Penrose (M–P) diagrams. (a) An arrow between two tensors denotes probabilistic causality between them. (b) If we add an $\mathcal{E}$ circle to the arrow, we indicate additive error term. (c) If the emitter of the arrow is a probabilistic density function, this denotes a prior distribution for the variable it is pointing to. This is quite similar to the usual DAG notation. (d) Usual DAG diagrams [53] for the generative models of EEG (left) and its equivalent version in the M–P notation (right). The M–P diagram explicitly states that EEG matrix V is the contraction of the lead field matrix K with the primary current density matrix G corrupted by sensor noise E_V. This example model also includes a prior probability distribution for G. (e) Similar DAG and M–P diagrams for the generative model of the BOLD signal. Here the vasoactive feedforward signal Γ is contracted with the hemodynamic response function H (temporal convolution) producing the BOLD signal B. A prior for Γ is also shown.*

techniques have allowed us to project this activity to a more restricted space. This can be either a standard 2-D grid distributed on the cerebral cortex, or a larger set defined for a 3-D grid that spans cerebral cortex and thalamus. The distinction between the two cases will be clear from the context but in both cases with a total number of voxels I_{Cx}. This activity is sampled at much slower rates than EEG (typically in the order of 1 Hz) which results in a total number of observations $I_{T\delta} \ll I_T$. We therefore denote the recorded fMRI by the matrix $\mathbf{B} \in \mathbb{R}^{I_{Cx} \times I_{T\delta}}$. For a list of dimensions of matrices and tensors refer to Table 3 and for the EEG/fMRI related matrices and tensors used in this review to Table 4.

We also assume for EEG and fMRI recorded concurrently that the time samples for the fMRI are obtained at integer multiples of those for EEG.

A. Matrix EEG Inverse Problem

The discretized version of the forward problem for EEG is

$$\mathbf{V} = \mathbf{K}\mathbf{G} + \mathbf{E}_V \tag{1}$$

where $\mathbf{G} \in \mathbb{R}^{I_{Cx} \times I_T}$ denotes the primary current density defined over the same cortical grid as fMRI and sampled at

Table 3 Dimensions of Matrices and Tensors

Dimension	Definition
I_E / $I_{E\delta}$	Number of EEG scalp electrodes *
I_{Cx} / $I_{Cx\delta}$	Number of sources (EEG or BOLD) on a cortical surface grid
I_{Wm}	Number of voxels in white matter tracts
I_T / $I_{T\delta}$	Number of time points
I_F / $I_{F\delta}$	Number of frequency points
I_W	Number of subjects
I_{lag}	Number of past time points (time lags) in autoregressive models
N	Number of dimensions of a tensor
R	Number of atoms (components) in tensor decompositions
*Subscript δ is used for subsampling of that dimension	

Table 4 EEG and fMRI Related Matrices and Tensors

Symbol	Definition	Dimension
Constants		
K	EEG lead field	$I_E \times I_{Cx}$
G	Primary current density matrix (generators) of EEG	$I_{Cx} \times I_T$
H	Hemodynamic response matrix	$I_T \times I_{T\delta}$
Γ	VFFS	$I_{Cx} \times I_T$
L	Laplacian matrix	$I_E \times I_E$ or $I_{Cx} \times I_{Cx}$
Data Tensors		
V	EEG space/time data matrix	$I_E \times I_T$
B	fMRI matrix	$I_{Cx} \times I_{T\delta}$
FA	DTI-FA data matrix	$I_W \times I_{Wm}$
B$_t$	Data matrix of GC	$I_{Cx} \times I_{T\delta}$
$\mathcal{B}$	Time lagged data tensor of GC	$I_{lag} \times I_{Cx} \times I_{T\delta}$
Estimated Tensors		
$\mathcal{S}_T$	Space/time/frequency tensor for spectral estimate of EEG	$I_E \times I_{T\delta} \times I_{F\delta}$
$\mathcal{S}_S$	Space/subject/frequency tensor for spectral estimate of EEG	$I_E \times I_W \times I_{F\delta}$
$\mathcal{A}$	GC connectivity tensor	$I_{Cx} \times I_{Cx} \times I_{lag}$
Signatures of tensor decompositions		
M$_V$	Spatial signature of EEG atoms over electrodes	$I_E \times R$
M$_G$	Spatial signature of EEG atoms over the cortical grid	$I_{Cx} \times R$
M$_B$	Spatial signature of fMRI atoms over the cortical grid	$I_{Cx} \times R$
M$_{FA}$	Spatial signature of FA atoms over the white matter grid	$I_{Wm} \times R$
T$_V$	Temporal signature of EEG for atoms	$I_{T\delta} \times R$
T$_B$	Temporal signature of fMRI atoms for sampled time points	$I_{T\delta} \times R$
F$_V$	Spectral signature of EEG atoms	$I_{F\delta} \times R$
W$_B$	Subject signature of fMRI atoms	$I_W \times R$
W$_{FA}$	Subject signature for FA atoms	$I_W \times R$
M$_r$	Spatial signature for receiver voxels in GC atoms	$I_{Cx} \times R$
M$_s$	Spatial signature for sender voxels in GC atoms	$I_{Cx} \times R$
T	Temporal signature for GC atoms	$I_{lag} \times R$

the same time points as EEG. $\mathbf{K} \in \mathbb{R}^{I_E \times I_{Cx}}$ is the lead-field matrix which summarizes volume conduction effects in the head.[1] We will henceforth assume that the sensor error $\mathbf{E}_V \in \mathbb{R}^{I_E \times I_T}$ is a matrix with entries that are identically distributed zero mean Gaussian variates. Since we are modeling only sensor noise, this has been found to be a good approximation. This assumption allows us to measure the fit of models to the data by means of the Frobenius norm $\| \bullet \|_2^2$.[2]

Estimation of $\mathbf{G}$, also known as electrophysiological source imaging (ESI), is a well-known ill-posed problem. Therefore, a solution by naive minimization of the functional $\| \mathbf{V} - \mathbf{KG} \|_2^2$ is not possible and, in fact, it does not have a unique solution. Uniqueness may be obtained by adding prior anatomical and physiological information to the problem formulation.

According to this approach, estimation of $\mathbf{G}$ involves finding the argument $\hat{\mathbf{G}}$ that minimizes the following augmented functional:

$$\hat{\mathbf{G}} = \arg\min_{\mathbf{G}} \| \mathbf{V} - \mathbf{KG} \|_2^2 + \pi(\mathbf{G}). \qquad (2)$$

---

[1]Note that this is a linear operator under the quasi-static approximation for Maxwell's equations [172].

[2]Correlated or non-Gaussian error terms can easily be dealt with, but would complicate model expressions unnecessarily.

The penalization term $\pi(\mathbf{G})$ applied in ESI is generally a combination of different matrix norms. Let us consider some examples.

- One of the best known examples of source imaging is LORETA [65], in which the penalization takes the form $\pi(\mathbf{G}) = \|\mathbf{LG}\|_2^2$, which encourages estimation of spatially smooth sources on the cortex.
- Another variant is VARETA [66], which uses the penalty term $\pi(\mathbf{G}) = \|\mathbf{LG}\|_1$ with an $L1$-norm to impose both spatial smoothness and sparseness.
- More recently in [49], the following penalty term was proposed: $\pi(\mathbf{G}) = \lambda_1\|\mathbf{G}\|_1 + \lambda_2\|\mathbf{LG}\|_2^2$. This penalty achieves an optimal balance between spatial sparsity and smoothness of cortical sources by data-driven hyperparameters λ_1 and λ_2.

A comparison of these and other types of source imaging may be found in [48]. In that same paper, spatio–temporal orthogonal nonnegative independent component analysis (STONNICA) was proposed as a solution for the EEG inverse problem. This model not only illustrates more complex penalty terms, but also shows how additional constraints might be useful to find interpretable sources. A tensor extension is described in Section IV.

STONNICA is different in two ways from usual applications of ICA to EEG source localization [20].

1) It is based on a variant of ICA for which components are forced to be orthogonal and nonnegative (ONN–ICA) [67].
2) STONNICA identifies the components in source space directly (ICA tomography). By contrast, other types of ICA source/localization first carry out ICA in sensor space and then localize the extracted components (tomography of ICA).

The parameter estimates for this model are obtained as

$$
\begin{aligned}
(\hat{\mathbf{M}}_{\mathbf{V}}, \hat{\mathbf{T}}_{\mathbf{V}}) = \arg\min_{\mathbf{M}_{\mathbf{V}}, \mathbf{T}_{\mathbf{V}}} &\left\{ \frac{1}{2}\|\mathbf{V} - \mathbf{KM}_{\mathbf{V}}\mathbf{T}_{\mathbf{V}}\|_2^2 \right. \\
&\left. + \lambda_1\|\mathbf{M}_{\mathbf{V}}\|_1 + \lambda_2\|\mathbf{LM}_{\mathbf{V}}\|_2^2 \right\} \\
\text{s.t.} \quad \mathbf{M}_{\mathbf{V}}^T\mathbf{M}_{\mathbf{V}} = \mathbf{I}, &\qquad \mathbf{M}_{\mathbf{V}} \geq 0 \quad (3)
\end{aligned}
$$

where $\mathbf{M}_{\mathbf{V}} \in \mathbb{R}^{I_{Cx} \times R}$ is the spatial signature or cortical distribution of the ONN–ICA components. The ONN condition is equivalent to specifying spatially nonoverlapping EEG sources. This requirement, in addition to the smooth Lasso type constraints [44], [46], [49], results in the identification of sparse isolated clustered components that were used to identify distinct cognitive processes involved in face processing.

B. Matrix fMRI Inverse Problem

As mentioned before, the generation of the BOLD signal from the VFFS is best described by a set of nonlinear differential equations [18]. Therefore strictly speaking, the forward and inverse fMRI problems should be solved using neural mass models based on nonlinear random differential equations, as reviewed in detail in [23]. But this would detract from our objective of simplicity in illustrating matrix and tensor techniques.

We will, therefore, resort to a useful linear generative model for the BOLD signal, which is the convolution of the vasoactive feedforward signal $\gamma(g,t)$ at point g of the cortical grid with the hemodynamic response $h(t)$

$$
b(g,t) = \int h(t-\tau)\gamma(g,t)d\tau + \varepsilon_b(g,t). \quad (4)
$$

If this continuous time model is discretized over time at the $I_{T\delta}$ sampling times of the fMRI and the convolution operation stated as a matrix product, the resulting fMRI forward model is

$$
\mathbf{B} = \mathbf{\Gamma H} + \mathbf{E}_B \quad (5)
$$

where $\mathbf{\Gamma} \in \mathbb{R}^{I_{Cx} \times I_T}$ is the vasoactive feedforward signal matrix and $\mathbf{H} \in \mathbb{R}^{I_T \times I_{T\delta}}$ is the hemodynamic response matrix. Note that $\mathbf{H}$ is obtained by subsampling rows from the general square symmetric Toeplitz matrix $\{h(t_i - t_j)\}_{1 \leq i,j \leq I_T}$ defined at the finer time resolution.

Using similar procedures as those outlined in the discussion of ESI, the deconvolution of the fMRI may also be stated as an inverse problem

$$
\hat{\mathbf{\Gamma}} = \arg\min_{\mathbf{\Gamma}} \|\mathbf{B} - \mathbf{\Gamma H}\|_2^2 + \pi(\mathbf{\Gamma}). \quad (6)
$$

Glover [68] was the first to propose this type of deconvolution using a Wiener filter, that is, $\pi(\mathbf{\Gamma}) = \|\mathbf{\Gamma}\|_2^2$. Later Valdés-Sosa *et al.* [23] proposed the use of the penalty $\pi(\mathbf{\Gamma}) = \|\mathbf{\Gamma L}\|_2^2$ (a "LORETA-style" inverse problem) in the context of EEG/fMRI fusion, a topic which we now turn to.

C. Matrix-Based EEG/fMRI Fusion

From an examination of (2) and (6) it is clear that in order to carry out EEG/fMRI fusion a link must be established between $\mathbf{G}$ and $\mathbf{\Gamma}$. A first attempt was carried out in [69], where both quantities were assumed to be proportional to each other. Under these conditions, a form of matrix EEG/fMRI fusion procedure was developed that was formulated as

$$
\hat{\mathbf{G}} = \arg\min_{\mathbf{G}} \|\mathbf{V} - \mathbf{KG}\|_2^2 + \alpha\|\mathbf{B} - \mathbf{GH}\|_2^2 + \pi(\mathbf{G}). \quad (7)
$$

Using LORETA-type penalties, this method was capable of accurately localizing separate components of the somatosensory evoked magnetic response. Also note that this was the first formulation of a symmetrical type of EEG/

fMRI fusion, since neither modality has *a priori* priority over the other—their relative weight is determined by the data chosen constant α. Symmetrical data fusion approaches use complementary information of both modalities to estimate the common source of neural activity [69], [70] whereas in asymmetrical fusion methods one modality is given priority to guide the other one. A review can be found in [71].

Another type of matrix-based fusion is extracting the features of two modalities such as contrast maps for fMRI and averaged time courses of EEG and apply fusion methods to second level statistics to reveal the between subject variance [25]. For example, in joint ICA, data feature matrices are concatenated and decomposed by assuming that the data sets share the same mixing matrix/modulation profiles but different source components. In [72], the temporal data feature matrix of EEG ($\mathbf{V}$) constituted by the average ERP time courses over subjects and contrast maps of fMRI over subjects ($\mathbf{B}$) are jointly decomposed to find the temporal ($\mathbf{T_V}$) and spatial components ($\mathbf{M_B}$) as

$$\mathbf{V}|\mathbf{B} = \mathbf{A}(\mathbf{T_V}|\mathbf{M_B})$$

$$\text{s.t.} \quad \pi(\mathbf{T_V}) = \prod_{r=1}^{R} \pi(\mathbf{T_V}(r, :))$$

$$\pi(\mathbf{M_B}) = \prod_{r=1}^{R} \pi(\mathbf{M_B}(r, :))$$

where $\mathbf{A}$ is the common modulation profile of subjects and R is the number of components of ICA.

D. Beyond Matrix-Based Methods

As we have just discussed, matrix-based methods have been very useful for EEG/fMRI inverse solutions and their fusion.

Nevertheless, many EEG and fMRI data sets and model constructs are not best represented as matrices. Consider a set of EEG spectra recorded on a common set of sensors, for a number of subjects, for different experimental conditions. This is actually the data that can be arranged as a multidimensional array (channel, frequency, subject, condition).

Instead of reshaping the data into matrices for use with more standard methods, tensor decomposition leverages the natural format of the data in order to discover hidden structures (see [73]). It is therefore not surprising to see the increasing number of applications of tensor methods in the analysis of brain data [39], [74].

IV. TENSOR EEG ANALYSIS

A. Parallel Factor Analysis of Scalp EEG

One of the main advantages of tensor-based analyses is the ability to represent large multidimensional arrays in terms of much simpler structures. The best known such representation is the canonical decomposition or parallel factor analysis decomposition (PARAFAC), a tensor equivalent to principal component analysis. As mentioned before, for tensors of order equal or larger than 3, this decomposition is unique under rather mild conditions [31], [36].

For the sake of concreteness, we will illustrate PARAFAC in the context of time/frequency decompositions of EEG, which naturally lead to three-way tensors. Calculating the wavelet or Gabor spectrum for each channel $\mathbf{V}(i, :), i = 1, \ldots, I_E$, yields a matrix $\boldsymbol{\mathcal{S}}_T(i, :, :)$ for given time frames and frequencies. The spectra for all channels may be shaped into a three-way tensor $\boldsymbol{\mathcal{S}}_T$ of size $I_E \times I_{T\delta} \times I_{F\delta}$, where I_E is the number of sensors, $I_{T\delta}$ is the number time samples, and $I_{F\delta}$ is the number of frequency samples.

The PARAFAC model decomposes $\boldsymbol{\mathcal{S}}_T$ into R components or "atoms" which can be expressed in several alternative forms. In scalar notation, it is

$$\boldsymbol{\mathcal{S}}_T(i_E, i_{T\delta}, i_{F\delta}) = \sum_{r=1}^{R} \mathbf{M_V}(i_E, r) \mathbf{T_V}(i_{T\delta}, r) \mathbf{F_V}(i_{F\delta}, r)$$
$$+ \boldsymbol{\mathcal{E}}(i_E, i_{T\delta}, i_{F\delta}) \quad (8)$$

where i_E, $i_{T\delta}$, and $i_{F\delta}$ are indices for space, time, and frequency, respectively, and $\boldsymbol{\mathcal{E}}$ denotes noise. $\mathbf{M_V}(:, r)$, $\mathbf{T_V}(:, r), \mathbf{F_V}(:, r)$ are the spatial, temporal, and spectral signatures, respectively, for atom r.

To make clear the M–P notation, we reexpress (8) in terms of tensor operations

$$\boldsymbol{\mathcal{S}}_T = \mathbf{M_V} \bullet_{\{R\}} \mathbf{T_V} \bullet_{\{R\}} \mathbf{F_V} + \boldsymbol{\mathcal{E}}$$
$$= [\![\mathbf{M_V}, \mathbf{T_V}, \mathbf{F_V}]\!] + \boldsymbol{\mathcal{E}}. \quad (9)$$

Many multilinear models, in particular PARAFAC components, have a scale indeterminacy which does not affect interpretation. By convention, we will assume that factors are all normalized except the first one in the Kruskal operation.

This is the model applied in [43] which isolated several rhythmic components of EEG. There are several variations on this basic PARAFAC model that can improve interpretability. For example, we may impose sparseness, smoothness, nonnegativity, and orthogonality for the spatial signatures $\mathbf{M_V}(:, r)$ similar to that of (3).

The resulting model—a PARAFAC/ICA—is depicted in M–P format as in Fig. 5. In practical terms, this model may be estimated as follows:

$$(\hat{\mathbf{M}}_\mathbf{V}, \hat{\mathbf{T}}_\mathbf{V}, \hat{\mathbf{F}}_\mathbf{V}) = \underset{\mathbf{M_V}, \mathbf{T_V}, \mathbf{F_V}}{\arg\min} \left\{ \frac{1}{2} \|\boldsymbol{\mathcal{S}}_T - [\![\mathbf{M_V}, \mathbf{T_V}, \mathbf{F_V}]\!]\|_2^2 \right.$$
$$\left. + \lambda_1 \|\mathbf{M_V}\|_1 + \lambda_2 \|\mathbf{LM_V}\|_2^2 \right\}$$
$$\text{s.t.} \quad \mathbf{M_V}^T \mathbf{M_V} = \mathbf{I}, \qquad \mathbf{M_V} \geq 0, \quad \mathbf{F_V} \geq 0. \quad (10)$$
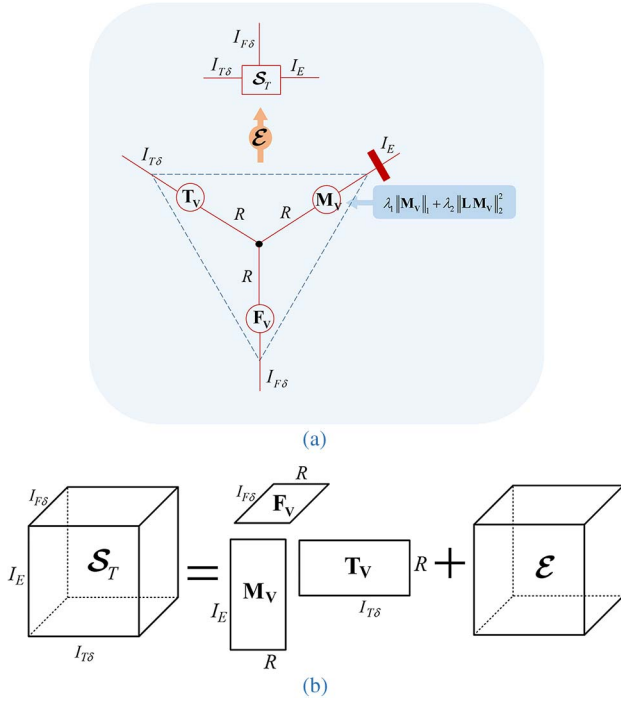
(a)

(b)

Fig. 5. *M–P diagram of a PARAFAC/ICA decomposition. (a) PARAFAC/ICA decomposition for a third-order tensor $\mathcal{S}_T$. Atoms (components) are latent variables shown with circles and $\mathcal{S}_T$ is the observed variable. Here $\mathbf{M_V}$ is required to be orthogonal, nonnegative, sparse, and smooth. (b) The more traditional 3-D representation of PARAFAC.*

As will be seen in Section IV-B, this model is easily combined with EEG source estimation.

A review of related work on combining tensor methods for EEG analysis can be found in [75]. The combination of ICA with PARAFAC has also been developed in [76] and [77]. A slightly different approach—applied to fMRI data—in [78] is to impose the PARAFAC structure during ICA extraction.

A characteristic of the PARAFAC model that underlies our presentation is that, as a consequence of the outer product structure of signatures, reconstructed atoms may appear as "oval blobs." While useful as a first approximation, this is a limitation that may be taken care of by including nonlinear crossed terms as variables or by using more complex multilinear models (e.g., Tucker decomposition). This is a topic that will be the subject of future research.

B. Tensor-Based EEG Inverse Problems

In Section III-A, we described STONNICA, a type of ICA in source space that was applied to time domain electrophysiological data. We have already seen that EEG time/frequency decompositions are best described in a tensor format and that PARAFAC decompositions may reveal interesting latent structures. It is therefore natural

to integrate PARAFAC and STONNICA, a model that is estimated as follows:

$$\mathcal{S}_T = [\![\mathbf{KM_G}, \mathbf{T_V}, \mathbf{F_V}]\!] + \mathcal{E}$$

for which the estimator will be

$$\left(\hat{\mathbf{M}}_\mathbf{G}, \hat{\mathbf{T}}_\mathbf{V}, \hat{\mathbf{F}}_\mathbf{V}\right) = \underset{\mathbf{M_G}, \mathbf{T_V}, \mathbf{F_V}}{\arg\min} \left\{ \frac{1}{2} \|\mathcal{S}_T - [\![\mathbf{KM_G}, \mathbf{T_V}, \mathbf{F_V}]\!]\|_2^2 \right.$$
$$\left. + \lambda_1 \|\mathbf{M_G}\|_1 + \frac{1}{2}\lambda_2 \|\mathbf{LM_G}\|^2 \right\}$$
$$\text{s.t.} \quad \mathbf{M_G^T M_G} = \mathbf{I}, \qquad \mathbf{M_G} \geq 0, \quad \mathbf{F_V} \geq 0. \quad (11)$$

$\mathbf{F_V}$ is the spectral signature of the generator sources of EEG. The Markov–Penrose diagram of this model is shown in Fig. 6(a).

The application of this model to the analysis of the resting state EEG is shown in Fig. 6(b). Note that three atoms were identified, all predominant in the occipital cortex. The frequency signatures are shown in the top left of that figure, showing that the components were primarily related to the different types of alpha rhythm.

Related work on combining PARAFAC and inverse solutions can be found in [79].

V. BRAIN CONNECTIVITY AS A TENSOR REGRESSION

We now address the use of tensor methods to evaluate brain connectivity. Of the different types shown in Fig. 7 we will concentrate on effective connectivity. This is the direct causal activation of one neural mass by another mediated by axonal pathways.

This topic has been reviewed in depth in [80], where it is argued that current work in this area is a fusion of several strands of research.

- One strand is based on graphical representations of causal models and the conditions under which causal inference is possible. This structural approach is best exemplified by the work of Pearl [81] who identifies the essential role of interventions to determine causality and in effect formalizes their application.
- A second line of work is that of Wiener [82], Akaike [83], Granger [84], and Schweder [85] in which predictability of one time series by another is used to lagged-based measures of statistical dependence. These measures of influence that apply to continuous time series, point processes, linear models, and nonlinear models were termed Wiener–Akaike–Granger–Schweder (WAGS) influence measures in [80].
- A third strand in the integration of WAGS theory with the structural approach previously mentioned
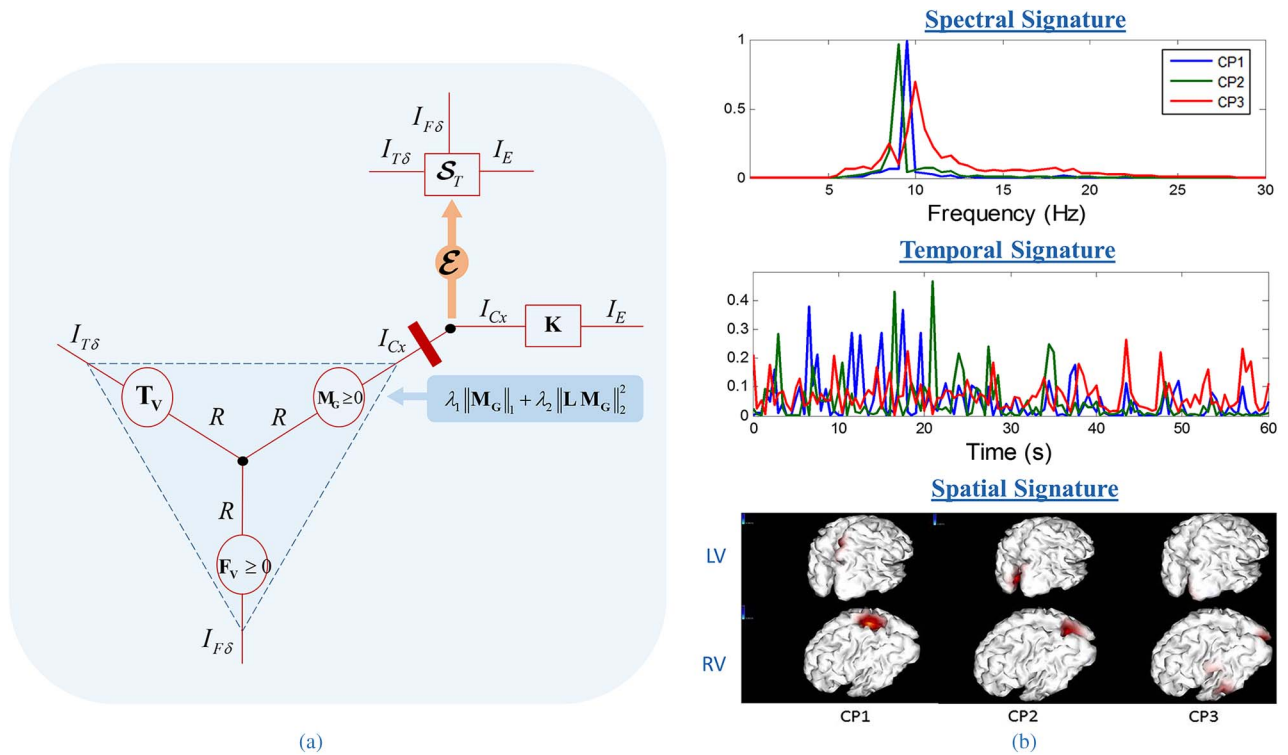
(a)

Spectral Signature

Temporal Signature

Spatial Signature

(b)

Fig. 6. *PARAFAC/STONNICA. (a) M–P diagram indicating the tensor operation flow for tensor STONNICA, in this case, a PARAFAC decomposition of an inverse solution, where $\mathcal{S}_T$ is the signal observed in the scalp and $\mathbf{K}$ is the lead field matrix. The atoms of the decomposition are denoted as spectral signature ($\mathbf{F_V}$ in the diagram), temporal signature ($\mathbf{T_V}$), and spatial localization of sources ($\mathbf{M_G}$), for the latter of which multiple regularizations are imposed. (b) An example of the application of this procedure on the resting state EEG from one subject. Tensor STONNICA is applied on the cross-spectral density of EEG for all segments in the frequency range 0.5–30 Hz in 0.5-Hz steps. MNI-based head model is used for the calculation of lead field. Spectral signatures show three distinct atoms at 9, 9.5, and 10 Hz, all of them being alpha atoms. Temporal signatures reveal coexistence of these rhythms at the same time in different magnitudes. Spatial signatures in source space are localized in occipital, striatal, and parastriatal areas. Note that three atoms are well separated in frequency and spatial localization. 3 Frequency domain STONNICA was presented at the XXVII Annu. Int. Meeting Human Brain Mapping, 2011: M. Bringas, I. Pedroso, V. Perez, J. Sanchez-Bornot, and P. Valdes-Sosa, "Resting state frequency domain Tomographic ICA."*
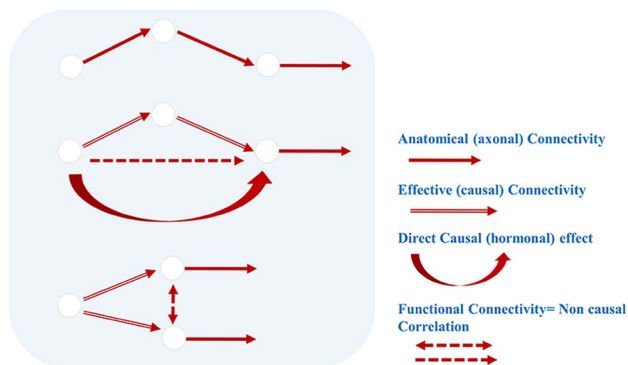

Fig. 7. *Several related concepts of brain connectivity are summarized. Anatomical connectivity is defined as the existence of axonal pathways that link two distinct neural masses. If this pathway is activated there is effective connectivity between these neural masses. On the other hand, if one only measures a correlation between the activities of two neural masses, one is measuring functional connectivity. It is effective connectivity that is of interest in defining functional neural networks.*

has been carried out for discrete time by Eichler [86] and for continuous time [87].

- Structural and WAGS influences are amplified by the biophysical modeling (c.f., by random differential equations) as best illustrated by dynamic causal modeling [88].

Biophysical causality explicitly examines the state–space equations that model not only system dynamics but also the observation equation [80]. From the discussion of the EEG forward problem in Section III-A, it becomes evident that volume conduction is a serious problem affecting the interpretation of influence measures. For a recent attempt to overcome the complexities of EEG Granger causality (GC), see [89].

Our specific application of GC to fMRI should be viewed in the light of the debate on the applicability of lag-based measures of influence for fMRI [90], which are based on the following arguments:

- fMRI has a very low temporal resolution that is limited by the rather large repetition time ($\sim$2 s) of most image acquisition pulse sequences;

- the low-pass behavior of the hemodynamic response blurs the temporal profile of information flow between neuronal populations;
- the variability of the hemodynamic response function (HRF) across brain regions has been considered a factor that might potentially bias GC estimates.

While the jury is still out on the applicability of lag-based methods for fMRI, recent work has offset, to a certain degree, the most pessimistic points of view. In [91], it was shown that GC is invariant to HRF convolution with higher temporal sampling and low measurement noise, a conclusion bolstered by [92]. Extensive simulations by Rodrigues and Andrade suggest that the optimal sampling frequency is around 100 ms [93]. This is also argued for experimentally by Lin *et al.*, who employed a fast fMRI sequence of 100 ms [94]. It is this data that we have used in this paper. In any case, the specific fMRI analysis shown should be taken as an illustration of tensor GC and further study of the validity of lag based methods for influence are warranted, especially by embedding them in formulation [95], [96].

Since our purpose is to show that the estimation of GC (influence) measures for fMRI data can be profitably approached using tensor methods we exclude EEG from this section and will limit our examples on a data set of BOLD measurements.

This is fast fMRI data (sampled at 10 Hz) from one of the subjects reported in [94]. The subject recorded had to respond with the corresponding hand to right or left visual hemi-field stimuli. Data from 1100 voxels were recorded from the visual (V), parietal (PCC), premotor (PreM), somatosensory (S), and motor (M) regions of interest (ROIs).

A. WAGS Influence or Granger Causality

This influence measure is based upon the multivariate autoregressive model (MAR). For a review and freely available toolbox, see [97]. The algorithms described in the cited paper and many others are useful only when analyzing a quite small number of time series. We now set out the matrix formulation of the problem.

However, as pointed out in [80] and [98], for brain imaging data, a high-dimensional MAR is needed to search for the influence fields that are the spatial maps of the influence of one brain area on the rest of the brain—something for which a Bayesian formulation is needed.

To formalize these ideas, we remind the reader that the BOLD signal is denoted by $\mathbf{B} \in \mathbb{R}^{I_{Cx} \times I_{T\delta}}$. Then, the spatial MAR is

$$\mathbf{b}_t = \sum_{q=1}^{I_{\text{lag}}} \mathbf{A}_q \mathbf{b}_{t-q} + \varepsilon_t \qquad (12)$$

where $\mathbf{b}_t = \mathbf{B}(:, t)$, I_{lag} is the number of past lags included in the model, and ε_t is the innovation noise.

The autoregressive matrices $\mathbf{A}_q \in \mathbb{R}^{I_{Cx} \times I_{Cx}}$ quantify the influence of the past time series on all others, including themselves. If a coefficient $\mathbf{A}_q(i, j) \neq 0$, we will say that time series j (Granger) influences time series i after q lags.

Thus, GC measures are essentially tests of the null hypothesis for coefficients of the MAR model.

One of the main problems is that for fMRI the number of time series I_{Cx} is much larger than the length of the time series $I_{T\delta}$. A MAR model contains $I_{\text{lag}} \cdot I_{Cx}^2 + (I_{Cx}^2 + I_{Cx})/2$ variables. This reveals (12) as a high-dimensional $p \gg n$ regression problem for which the usual multivariate statistical techniques fail.

To avoid the high-dimensional scenario, the problem can be reduced in size, first by averaging the BOLD signal over voxels for preselected ROI, and then carrying out bivariate GC for all pairs of ROI—correcting for multiple comparisons. The results of this analysis for the data set used in this section are shown in Fig. 10(a) in which an outflow of influence from V spreads out to other ROIs.

An alternative approach is to use methods for high-dimensional data regression presented in [99]. This penalized MAR approach was first presented in [98] and reviewed in [47]. Here all voxels of interest (the whole brain if necessary) may be included in the regression but statistical procedures that lead to variable selection were carried out. However, only matrix-based models were dealt with.

We now change our viewpoint completely and look at the penalized MAR model in a tensor format.

B. Granger Causality Viewed as a Tensor Regression

If we collect $\mathbf{b}_t$ from (12) for all time samples $t = I_{\text{lag}} + 1, \ldots, I_{T\delta} + I_{\text{lag}}$, we define matrix $\mathbf{B}_{t-q} \in \mathbb{R}^{I_{Cx} \times I_{T\delta}}$ as follows:

$$\mathbf{B}_{t-q} = \left[\mathbf{b}_{I_{\text{lag}}+1-q}, \ldots, \mathbf{b}_{I_T+I_{\text{lag}}-q} \right]^T.$$

Then, we concatenate matrices $\mathbf{B}_{t-q}$ constructed for all $q = 1, \ldots, I_{\text{lag}}$ to obtain a data tensor $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{I_{\text{lag}} \times I_{Cx} \times I_{T\delta}}$.

The autoregressive coefficients are also essentially a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I_{Cx} \times I_{Cx} \times I_{\text{lag}}}$ that is obtained by concatenating $\mathbf{A}_q$ matrices along the I_{lag} dimension. We state the concatenation operations explicitly as follows:

$$\boldsymbol{\mathcal{B}} = [\mathbf{B}_{t-q}]_{q=1:I_{\text{lag}}}^{\{1|\cdots|1\}}, \quad \boldsymbol{\mathcal{A}} = [\mathbf{A}_q]_{q=1:I_{\text{lag}}}^{\{1|\cdots|1\}}.$$

Note that we make use of the property of tensors that adding singleton dimensions to a tensor will not change the dimension. The concatenation operation for $\boldsymbol{\mathcal{B}}$ is depicted in Fig. 8.
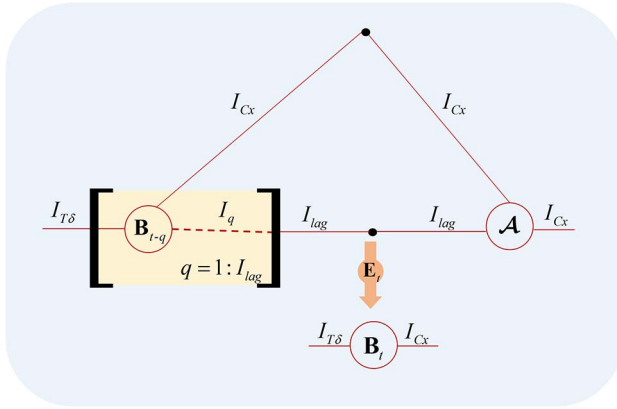
Fig. 8. *Representation of Granger causality as a tensor regression problem. The time series B_{t-q} of size $I_{Cx} \times I_{T\delta}$ is constituted by lagging a sample matrix B_t at $q = 1, \ldots, I_{lag}$. Concatenation of B_{t-q} over I_{lag} singleton dimensions gives the lagged data tensor $\mathcal{B}$ with size $I_{lag} \times I_{Cx} \times I_{T\delta}$. $\mathcal{B}$ is contracted with connectivity tensor $\mathcal{A}$ over spatial and temporal lag dimensions resulting the time series in voxels B_t.*

Thus, the MAR model is a tensor regression expressed by means of a tensor contraction

$$\mathbf{B}_t = \mathcal{A} \bullet_{\{I_{Cx}, I_{lag}\}} \mathcal{B} + \mathbf{E}_t. \qquad (13)$$

The M–P diagram for this type of model is given in Fig. 8. Note that $\mathcal{A} \in \mathbb{R}^{I_{Cx} \times I_{Cx} \times I_{lag}}$ is a 3-D tensor in the same figure.

In recognition of the tensor nature of the MAR in (13), we propose the following estimation procedure:

$$\hat{\mathcal{A}} = \arg\min_{\mathcal{A}} \left\{ \left\| \mathbf{B}_t - \mathcal{A} \bullet_{\{I_{Cx}, I_{lag}\}} \mathcal{B} \right\|_2^2 + \pi(\mathcal{A}) \right\}. \quad (14)$$

For example, in [47], the penalty $\pi(\mathcal{A}) = \lambda_1 \|\mathcal{A}\|_1 + \lambda_2 \|\mathcal{L} \bullet_{\{I_{Cx}\}} \mathcal{A}\|^2$ was used, as well as a number of other variants.

C. Granger Causality With t-Products

A refinement of the model just defined is to use tensor norms as penalization functions. Such an approach takes advantage of the t-product to implement the Levinson–Durbin estimation of MAR [100].

The t-product and related tensor operations were introduced by Kilmer *et al.* [101], and the key concept is summarized in Table 2. These operations allow 3-D tensors to be treated as though they were matrices which fit the tensor GC problem since three dimensions are adequate to define emitter, receiver, and temporal signatures of nodes.

We further need the following definitions.
- The sample covariance tensor $\mathcal{R} \in \mathbb{R}^{I_{Cx} \times I_{Cx} \times (I_{lag}+1)}$ is $\mathcal{R}(i_{Cx}, i_{Cx}, q) = (1/I_{T\delta}) \sum_{i_{T\delta}=1}^{I_{T\delta}} \mathcal{B}(q, i_{Cx}, i_{T\delta}) \mathbf{B}_t(i_{Cx}, i_{T\delta})$, where each block $\mathcal{R}(:, :, q)$ is an $I_{Cx} \times I_{Cx}$ cross-covariance matrix.
- $\mathcal{R}_1 = \mathcal{R}(:, :, 0 : I_{lag} - 1); \mathcal{R}_2 = \mathcal{R}(:, :, 1 : I_{lag}).$

The naïve solution to the classical Levinson–Durbin equation in our notation is

$$\mathrm{MatVec}(\mathcal{A}) = \mathrm{tplz}(\mathcal{R}_1)^{-1} \mathrm{MatVec}(\mathcal{R}_2). \qquad (15)$$

It is well known that this type of solution is not numerically stable. Therefore, one approach is to regularize the estimate of the covariance matrix $\mathcal{R}_1$ using the t-operations defined in Table 2. The specific estimator is

$$\hat{\mathcal{R}}_1 = \arg\min_{\Lambda} \left\{ \|\mathcal{R}_1 - \Lambda\|_2^2 + \lambda \|\Lambda\|_\circledast \right\} \qquad (16)$$

where the penalty term is the tensor nuclear norm (TNN) defined in Table 2.

The estimator $\hat{\mathcal{R}}_1$ can be explicitly found by shrinking the t-singular values in $\mathcal{D}$ by λ with the ρ function of [102]

$$\hat{\mathcal{R}}_1 = \mathcal{U} *_{\{I_{lag}\}} \rho(\mathcal{D}) *_{\{I_{lag}\}} \mathcal{V}^T.$$

Then, $\mathcal{A}$ is estimated as

$$\hat{\mathcal{A}} = \mathcal{V} *_{\{I_{lag}\}} \rho(\mathcal{D}) *_{\{I_{lag}\}} \mathcal{U}^T *_{\{I_{lag}\}} \mathcal{R}_2.$$

In fact, this operation was improved by using the circulant embedding defined in [103].

This model was applied to the data set analyzed in this section, and the resulting connectivity diagram is presented in Fig. 10(b). The method was able to deal with high-dimensional data having more than 1000 nodes and 20 lags with stable numerical results. It is also interesting to note that this estimate of connectivity seems to be much more sensitive than the simple bivariate approach.

D. Granger Causality With PARAFAC/ICA

As seen, the tensor GC uncover a much richer set of connections than simpler methods. However, basic neuroscience suggests that effective connectivity implies a structured sparsity of $\mathcal{A}$, which is desirable since it prunes many of the spurious connections characteristic of some functional connectivity measures. Structured sparsity can be achieved by positing a PARAFAC/ICA structure for the connectivity tensor. We will define a node as a sender if it

influences another set of nodes, and as a receiver if its activity is caused by other nodes.

Signatures of $\mathcal{A}$ are estimated by

$$
\begin{aligned}
(\hat{\mathbf{M}}_\mathbf{s}, &\hat{\mathbf{M}}_\mathbf{r}, \hat{\mathbf{T}}) \\
&= \underset{\mathbf{M}_\mathbf{s}, \mathbf{M}_\mathbf{r}, \mathbf{T}}{\arg\min} \left\{ \frac{1}{2} \left\| \mathbf{B}_t - \mathcal{A} \bullet_{\{I_{Cx}, I_{\text{lag}}\}} \mathcal{B} \right\|_2^2 + \lambda_1 \|\mathbf{M}_\mathbf{s}\|_1 \right. \\
&\qquad + \frac{1}{2} \lambda_2 \|\mathbf{L}\mathbf{M}_\mathbf{s}\|^2 + \lambda_3 \|\mathbf{M}_\mathbf{r}\|_1 + \frac{1}{2} \lambda_4 \|\mathbf{L}\mathbf{M}_\mathbf{r}\|^2 \\
&\qquad \left. + \lambda_5 \|\mathbf{T}\|_1 + \frac{1}{2} \lambda_6 \|\mathbf{L}\mathbf{T}\|^2 \right\}
\end{aligned}
$$

$$
\text{s.t.} \quad \mathcal{A} = [\![\mathbf{M}_\mathbf{s}, \mathbf{M}_\mathbf{r}, \mathbf{T}]\!], \qquad \mathbf{M}_\mathbf{s} \geq 0,
$$
$$
\mathbf{M}_\mathbf{s}^T \mathbf{M}_\mathbf{s} = \mathbf{I}, \quad \mathbf{M}_\mathbf{r} \geq 0, \qquad \mathbf{M}_\mathbf{r}^T \mathbf{M}_\mathbf{r} = \mathbf{I} \qquad (17)
$$

where $\mathbf{M}_\mathbf{r}$ is the spatial signature for receiving nodes $\mathbf{M}_\mathbf{s}$, is the spatial signature for sender nodes, and $\mathbf{T}$ is the temporal signature for causal lags.

In this model, the identifiability is enhanced by enforcing nonnegativity, orthogonality, smoothness, and sparseness for the spatial signatures and a smooth Lasso-type constraint for the lag signature. In other words, these constraints tend to estimate smooth patches of voxels on the cortex. Orthogonality and nonnegativity constraints guarantee that spatial factors can have only one nonnegative element in each row which can be interpreted as the cluster centroids [104], [105]. In this way, the connected spatial regions are confined to be nonoverlapping patches. This model is the generalization of clustering in which connectivity tensor is decomposed into sum of rank one triclusters [106].

The atomic decomposition of the 3-D connectivity tensor for the model of (17) favors a parsimonious model where the number of parameters to be estimated is $(2I_{Cx} + I_{\text{lag}})R$, with R being the model order of PARAFAC. The M–P diagram is shown in Fig. 9.

For the application of the GC–PARAFAC on the fMRI data set analyzed in this section, a time period of 500 ms corresponding to five time frame lags was selected as the temporal factor. A graph Laplacian matrix is used as the smoother matrix $\mathbf{L}$ in (17). The model order of PARAFAC was set to 3. Fig. 10(c) shows the existence of strong bottom–up and weak top–down connections between VC, PCC, M, and S. There is also lateral information flow from left to right visual areas.

We wish to note that the work in this section was encouraged by the tensor formulation of the state–space model set out in [96]. However, this is, to our knowledge, the first time the estimation of the well-known MAR model has been posed as a tensor regression problem. Although this formulation of MAR as a tensor problem is quite obvious, clearly stating the connection between MAR formulations and tensor models may help the introduction of concepts that are popular in multidimensional analysis
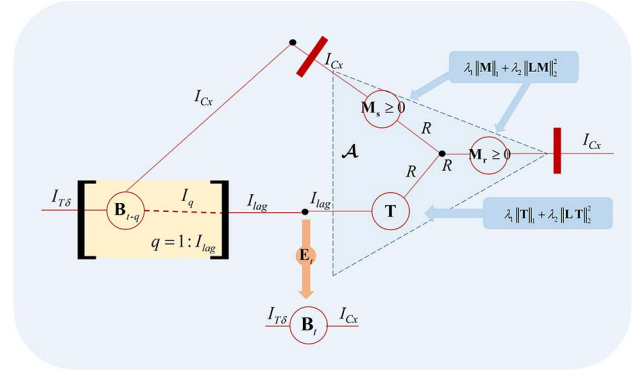

Fig. 9. *Granger causality as a PARAFAC decomposition. The connectivity tensor indicating voxel-to-voxel causal effect (denoted as tensor $\mathcal{A}$ in Fig. 8) applied a PARAFAC decomposition of order R into three component matrices: $\mathbf{M}_\mathbf{r}$ for receiver nodes, $\mathbf{M}_\mathbf{s}$ for sender nodes, and $\mathbf{T}$ for temporal lags. For the sake of interpretation, multiple priors are used for the estimation of components: sparsity, smoothness, orthogonality for spatial factors, and sparsity and smoothness for temporal lags. Since the connectivity tensor may take negative values, the temporal factor is kept to be real valued with the purpose of sign convention.*

for the analysis of the autoregressive processes. Two techniques immediately come to mind.

- As we have just shown, various tensor decompositions may aid in the analysis of the autoregressive coefficient tensor, of which PARAFAC is just the simplest. It can easily be seen that any of the various tensor decompositions can be applied with advantage to provide new insights into analyzing influence between time series [107], [108].
- In addition to the regularization of the signatures of the connectivity tensor that we have employed, different regularizations of the connectivity tensor, for example, the multiway nuclear norm [109], [110], may be employed to improve algorithms and interpretability.

VI. TENSOR EEG/fMRI FUSION

As reviewed in [23], symmetrical fusion of EEG/fMRI is an "equal opportunity" combination of modalities in which the relative influence of each modality to the final solution is not specified *a priori* but rather selected by the data. Fusion can be of two types: model driven and data driven. We will concentrate on data-driven approaches as an extension of the matrix-based approaches of Section III-C. We recall Fig. 1 and note that we will assume that the VFFS $\mathbf{\Gamma}$ and the primary current density $\mathbf{G}$ are related to each other in a simple fashion.

Fusion can be carried out by using tensor techniques that link the two modalities along a common dimension, usually temporal or spatial. Another possibility for multisubject data is to use subject identity as the common link.
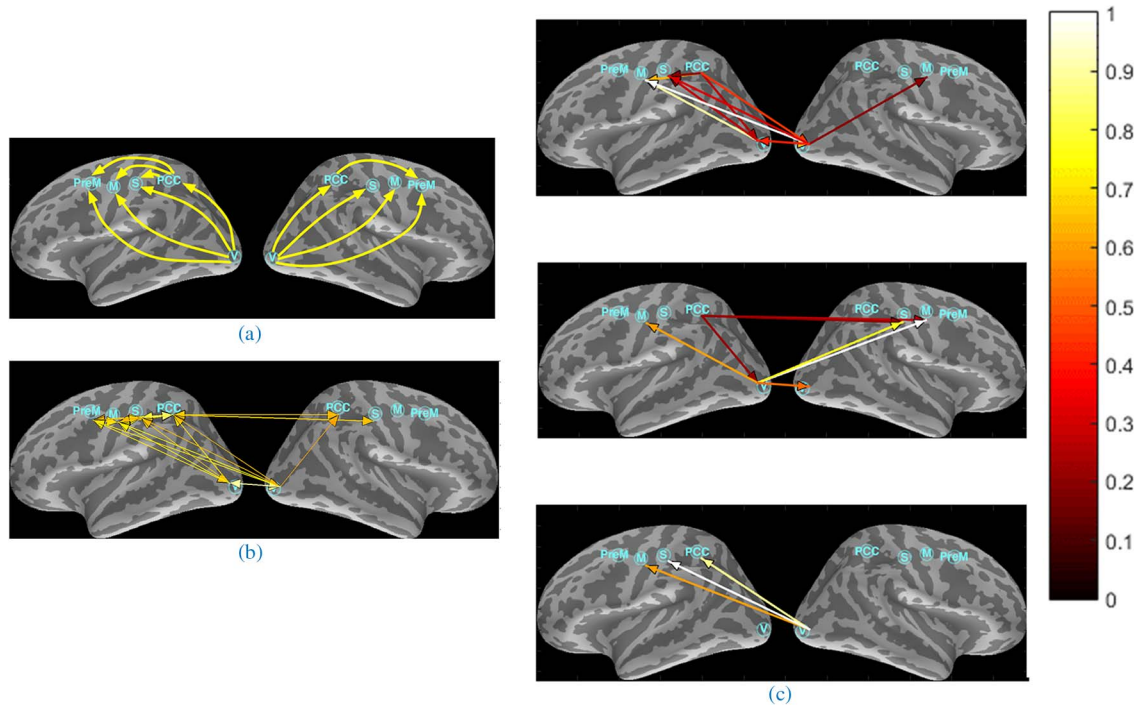
Fig. 10. *Granger causality on real data. The data corpus described in Section V was used for analyzing brain connectivity by means of different techniques, which offer comparable results. The arrows denote directional dominant flows of Granger causality between the visual (**V**), parietal (**PPC**), premotor (**PreM**), somatosensory (**S**), and motor (**M**) cortex. (a) The original results were published in [94] and extracted from [94, Fig. 2]. This is the dominant information flow calculated from the difference between two unidirectional Granger estimates among the ROIs. Only connections that have a p-value ≥ 0.05 are shown. (b) Results using the t-product technique described in Section V make the connectivity analysis more valuable revealing the bidirectional and interhemispheric interaction between ROIs. A major interaction was found on the left hemisphere which could be related to the right-handed condition of the subject. (c) The resulting three spatial atoms of the connectivity tensor retrieved by the GC analysis with PARAFAC decomposition. Each atom of receiver (**M_r**) and sender signatures (**M_s**) is grouped according to ROIs and the sum of each ROI is taken. Connectivity maps are generated for each atom by using directed arrows that are pointed from the cortical regions of senders which have a value greater than zero to positively active regions of the corresponding receiver signature. The temporal atoms, encoded in matrix **T**, showed an ascending connectivity influence, peaking at the first lag (100 ms) and slowly decaying afterwards. Note that in all of these results there is a predominance in causal directionality emerging from the **V** and **PPC** cortex to the rest of the areas. Magnitude of the connectivity is symbolized by the color bar on the right of the figure.*

In this section, we will give three examples of fusion: EEG/fMRI along the temporal domain, EEG/DTI along the subject domain, and EEG/fMRI along the spatial domain. We now explain these examples.

A. Multiway Partial Least Squares (N–PLS)

Multiway partial least squares is a tensor-based method in which both dependent and independent variables constituted as tensors are decomposed simultaneously, while the signatures or factors of the shared dimension are required to have maximal covariance [111].

Tensor PLS has been extended by using kernel methods and Tucker decomposition to model the nonlinear relations between two tensors [112]. For a review on N–PLS, see [113].

In [45], to find the BOLD correlates of EEG rhythms, the time-varying EEG spectrum $\mathcal{S}_T$ is decomposed into spatial $\mathbf{M_V}$, temporal $\mathbf{T_V}$, and spectral $\mathbf{F_V}$ signatures, and the fMRI data matrix $\mathbf{B}$ is decomposed into spatial $\mathbf{M_B}$

and temporal $\mathbf{T_B}$ signatures such that temporal factors $\mathbf{T_V}$ and $\mathbf{T_B}$ will have the maximum covariance. We therefore set the regression model between temporal signatures as

$$\mathbf{T_B} = \mathbf{T_V}\mathbf{C} + \mathbf{E}_{T_B}$$
$$\text{s.t.} \quad \mathcal{S}_T = [\![\mathbf{M_V}, \mathbf{T_V}, \mathbf{F_V}]\!] \text{ and } \mathbf{B} = [\![\mathbf{M_B}, \mathbf{T_B}]\!]. \quad (18)$$

Source localization is performed on the corresponding spatial signatures of EEG $(\mathbf{M_V})$ to find the generators of each component. Then, the source signatures of EEG denoted by $\mathbf{M_G}$ are estimated through the forward model defined by

$$\mathbf{M_V} = \mathbf{K}\mathbf{M_G} + \mathbf{E}_{M_V}. \quad (19)$$

The M–P diagram of this type analysis is shown in Fig. 11.

This method was applied on the resting state EEG/fMRI data of a single subject originally collected in [114]. Since this has been a well-studied data set, we will only briefly mention the results that can be found in detail in [45].

The time-varying EEG spectra $\boldsymbol{\mathcal{S}}_T \in \mathbb{R}^{I_E \times I_{T\delta} \times I_{F\delta}}$ is estimated for $I_E = 16$ channels in the frequency range of 0.5–50 Hz for $I_{F\delta} = 124$ frequency points and for $I_{T\delta} = 105$ time points. fMRI data are acquired in six adjacent slices that cut through occipital lobe and thalamus.

The N–PLS analysis described in (18) is applied by maximizing the covariance between temporal signatures. The scatterplot of the identified atoms shows this correspondence. Both topographic and source spatial signatures of EEG and spatial signature of fMRI are shown in Fig. 12.

Three atoms were extracted exhibiting spectral peaks in the alpha (α), theta (θ), and gamma (γ) range of EEG. There was a significant covariance between EEG and fMRI components only for α and θ, which are the only ones we will comment on. Note in Fig. 12 that both the scalp and source topographies are clearly delimited as being in the occipital region (α) and the frontal lobes (θ). The θ atom showed frontal negative activity, meaning that the increased BOLD signal corresponds to decreased spectral activity. Note that the spatial signature of the α atom is positive for the thalamus and negative for the visual cortex. These negative associations between BOLD and EEG activity in cortex have been interpreted as due to desynchronization of neural activity with greater thalamo–cortical input; a hypothesis that has received support with a large scale neural model described in [23]. In addition, it is shown in [43] that α is suppressed during mental calculation while the opposite is true for θ.

Identifying the coupling between EEG rhythms and resting state fMRI is a topic of current interest. An example for this type of work is [115]. Thus, it seems that tensor methods could be of use for the study of these interrelated phenomena.

A recent application of N–PLS to a different pair of modalities was to the joint decomposition of EEG and DTI functional anisotropy (FA). This was carried out in order to explore the neuroanatomical determinants of the inter-individual variability of the peak frequency of the EEG scalp alpha rhythm (8–12 Hz). Here the common dimension for both modalities was subject identification with an N of 200.

The data for each modality were encoded by an array with the common "subject dimension." The first modality is related to white matter architecture, as measured by DTI–FA. The voxels of the FA images in which the white matter probability is lower than 0.5 and FA is below 0.1 were masked out, leaving $I_{Wm} = 24\,764$ voxels. The masked FA images were vectorized in rows and concatenated over subject dimension $I_W = 200$ to build a 2-D matrix $\mathbf{FA} \in \mathbb{Z}^{I_W \times I_{Wm}}$. Since $\mathbf{FA} \in [0,1]$, they were logit-transformed to approximate normal distribution. The second modality is EEG scalp spectrum estimate, organized in a 3-D array $\boldsymbol{\mathcal{S}}_S \in \mathbb{R}^{I_E \times I_W \times I_{F\delta}}$, where $I_E = 76$ is the number of channels (a subset of EEG channels of the 10–20 system) and $I_{F\delta} = 58$ is the number of frequencies in the range of 0.39–29.68 Hz. For the details on the MRI and EEG acquisition and preprocessing, see [116]. Here both arrays were scaled and their grand mean subtracted. Inverse source localization was performed on the scalp spatial signatures of EEG using LORETA and MNI-based head models.
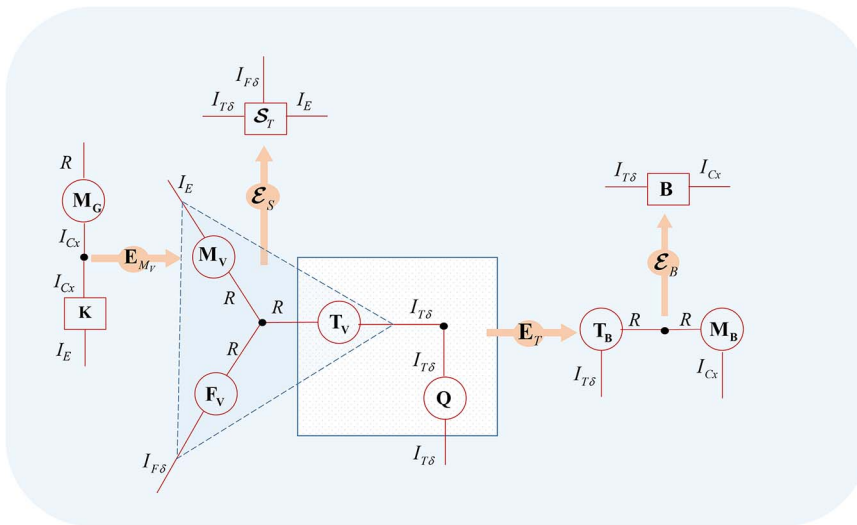

Fig. 11. *The M–P diagram for N–PLS. EEG data $\boldsymbol{\mathcal{S}}_T$ is decomposed using an* R *order PARAFAC into temporal* ($\mathbf{T_V}$)*, spectral* ($\mathbf{F_V}$)*, and spatial* ($\mathbf{M_V}$) *atoms, the latter of which is complemented by the source localization performed via the lead field matrix* K*, resulting in a spatial cortical atom* $\mathbf{M_G}$*. A similar* R *order PARAFAC decomposition is done on the fMRI data* B*, with temporal* ($\mathbf{T_B}$) *and spatial* ($\mathbf{M_B}$) *atoms in a way that covariance of temporal factors* $\mathbf{T_V}$, $\mathbf{T_B}$ *is maximized via matrix* Q*.*
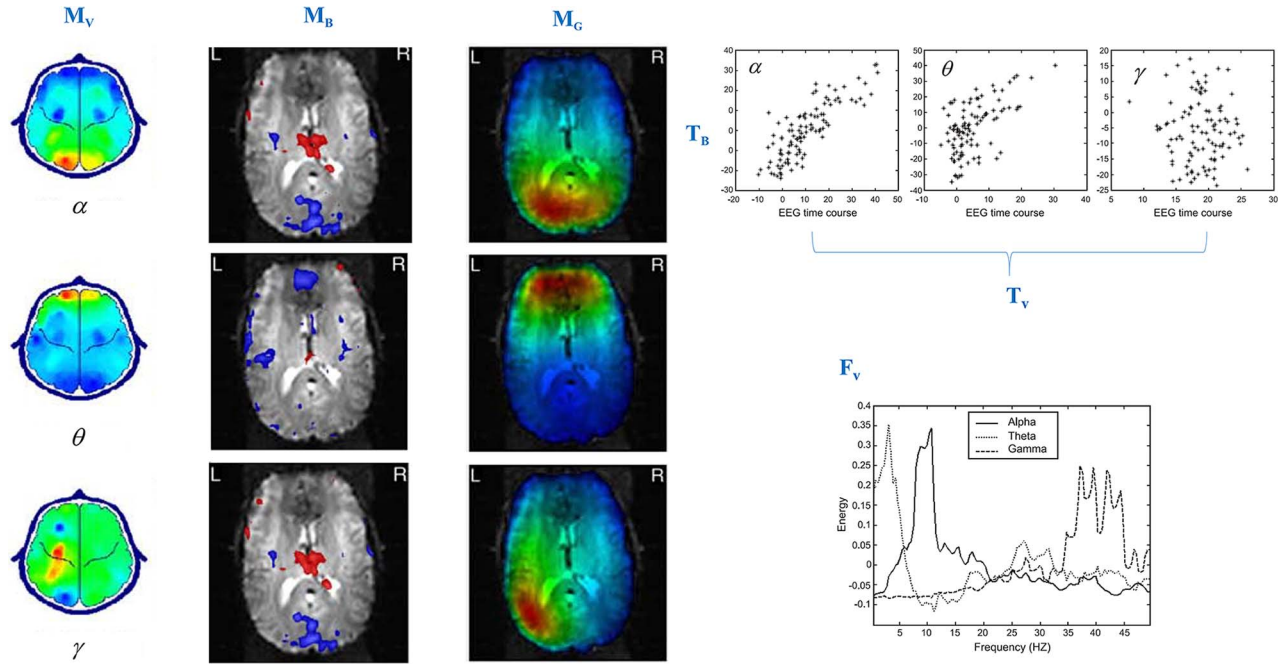
Fig. 12. *Multiway partial least squares results of EEG/fMRI fusion. Topographic maps of EEG* ($\mathbf{M_V}$) *and the corresponding source maps of EEG* ($\mathbf{M_G}$) *and spatial signature of fMRI* ($\mathbf{M_B}$) *are shown in columns. For all spatial maps, red color indicates positive values and blue color indicates negative values. Scatter plot of the fMRI temporal signatures* ($\mathbf{T_B}$) *against EEG temporal signatures* ($\mathbf{T_V}$) *is demonstrated for three atoms. Spectral signature of EEG* ($\mathbf{F_V}$) *is also shown. Signatures are named as alpha* (α)*, theta* (θ)*, and gamma* (γ) *due to the resemblance of their spectral characteristics to EEG rhythms, as revealed in the plot of the spectral signature of EEG. The EEG spatial signature shows activation in occipital regions for alpha, frontal regions for theta, and pario-temporal regions for gamma atoms. On the other hand, for alpha atom, the fMRI spatial signature shows negative activation in occipital and superior temporal regions and positive activation in the thalamus and insula. Theta atom of the fMRI spatial signature shows negative activation in anterior cingulate and occipital regions. Gamma atom of the fMRI spatial signature shows similar activity with the alpha atom. Scatter plot of the fMRI temporal signature against the EEG temporal signature significantly revealed positive correlation in the alpha atom. Correlation of these signatures for theta and gamma atoms is found insignificant. The EEG spectral signature showed three distinct peaks at alpha, theta, and gamma frequencies. Note the topographically distinct pattern of relations BOLD signals and spectral components that indicates stable relations between fMRI, resting state components, and EEG oscillations. The figure is adapted from [45].*

For the correlation of EEG and FA data, N–PLS is used in which the covariance between subject signatures is maximized. We determined that three atoms were adequate for the description of the EEG data set. Fig. 13 shows the spectral signature $\mathbf{F_V}$, the spatial signature $\mathbf{M_V}$, and inverse source constructions $\mathbf{M_G}$ of the atoms obtained from the N–PLS.

The spectral and spatial signatures associated to the atoms allowed us to identify them as follows.

1) An atom with a spectrum dominated by "slow wave activity," with essentially a $1/f$ decay. Its distribution was frontal. This atom might correspond with the EEG process identified as the "ξ process" described in [117] and [118].

2) An "alpha" (α) atom with a spectral peak very similar to those shown for other in Figs. 6 and 12. Its localization was occipital, as expected.

3) An "alpha contrast" atom, with a spectrum that resembles the derivative of that α atom, also with an occipital topography.

It is interesting that the first two components are similar to those described for the $\xi - \alpha$ model in [117] and [118], which was obtained by individual parameterized spectral models for EEG of a different set of 211 normal subjects.

The alpha contrast atom seems to reflect the individual-to-individual fluctuations of the peak alpha frequency across the sample. Interestingly, the subject signatures of both modalities for this atom have the highest and significant correlation ($R = 0.7701$, $p \sim 10^{-8}$), suggesting that white matter architecture indeed is correlated with the interindividual variation of the alpha rhythm. The spatial signature of the FA data set $\mathbf{M_{FA}}$ for this atom, shown in Fig. 13, is loaded on the major tracts, especially thalamo–cortical, in accordance with the results in [116]. This type of result is important in order to test models of the origin of the alpha rhythm.

B. Coupled Matrix-Tensor Factorization (CMTF)

We propose a new data fusion framework based on a joint decomposition of EEG and fMRI along the common

spatial profile. We extend the matrix-based EEG/fMRI fusion in (7) to coupled tensor decompositions of the EEG tensor $\boldsymbol{\mathcal{S}}_T$ and the fMRI data matrix $\mathbf{B}$. In Section IV-B, we showed that source signatures of EEG can be identified from $\boldsymbol{\mathcal{S}}_T$ by using PARAFAC. The EEG tensor $\boldsymbol{\mathcal{S}}_T \in \mathbb{R}^{I_E \times I_{T\delta} \times I_{F\delta}}$ is decomposed into source spatial $\mathbf{M}_\mathbf{G}$, temporal $\mathbf{T}_\mathbf{V}$, and spectral $\mathbf{F}_\mathbf{V}$ signatures, and the fMRI data matrix $\mathbf{B} \in \mathbb{R}^{I_{Cx} \times I_{T\delta}}$ is decomposed into spatial $\mathbf{M}_\mathbf{B}$ and temporal $\mathbf{T}_\mathbf{B}$ signatures, in which source spatial signatures are coupled during decomposition.

Earlier fusion algorithms were predicated on the idea that the support of the EEG active regions had a complete coincidence with that of the fMRI. Unlike conventional CMTF algorithms in which the common dimension is considered to be completely coupled, we distinguish also a discriminative subspace [119] where the signatures of both modalities are not overlapping. This enables us to deal with the cases in which EEG and fMRI sources may have a spatial mismatch [120]. Coupled and uncoupled spatial profiles are obtained for each modality.

Assume that $\mathbf{M}_\mathrm{eeg}$ is the source spatial factor of $\boldsymbol{\mathcal{S}}_T$ and $\mathbf{M}_\mathrm{fmri}$ is the spatial factor of $\mathbf{B}$, then in the proposed framework these factors will be: $\mathbf{M}_\mathrm{eeg} = \mathbf{M}_\mathbf{C}|_{\{R_C|R_G\}}\mathbf{M}_\mathbf{G}$ and $\mathbf{M}_\mathrm{fmri} = \mathbf{M}_\mathbf{C}|_{\{R_C|R_B\}}\mathbf{M}_\mathbf{B}$ where subscript $\boldsymbol{C}$ is for the common part and subscript $\boldsymbol{G}(\boldsymbol{B})$ is for the discriminant factor of EEG (fMRI). R_C is the number of common atoms, R_B is the number of discriminative atoms of fMRI, and R_G is the number of discriminative atoms of EEG. In this way, different model orders can be assigned to the decomposition of $\boldsymbol{\mathcal{S}}_T$ and $\mathbf{B}$ as long as the number of common components is kept the same, i.e., the column number of $\mathbf{M}_\mathbf{C}$.

Modality-specific and coupled signatures are estimated by

$$
\begin{aligned}
&\left(\hat{\mathbf{M}}_\mathbf{C}, \hat{\mathbf{M}}_\mathbf{G}, \hat{\mathbf{T}}_\mathbf{V}, \hat{\mathbf{F}}_\mathbf{V}, \hat{\mathbf{M}}_\mathbf{B}, \hat{\mathbf{T}}_\mathbf{B}\right) \\
&= \underset{\mathbf{M}_\mathbf{C}, \mathbf{M}_\mathbf{G}, \mathbf{T}_\mathbf{V}, \mathbf{F}_\mathbf{V}, \mathbf{M}_\mathbf{B}, \mathbf{T}_\mathbf{B}}{\arg\min} \\
&\times \left\{ \frac{1}{2} \left\| \boldsymbol{\mathcal{S}}_T - \left[\!\left[\mathbf{K}\left(\mathbf{M}_\mathbf{C}|_{\{R_C|R_G\}}\mathbf{M}_\mathbf{G}\right), \mathbf{T}_\mathbf{V}, \mathbf{F}_\mathbf{V} \right]\!\right] \right\|_2^2 \right. \\
&\left. + \gamma \frac{1}{2} \left\| \mathbf{B} - \left[\!\left[\left(\mathbf{M}_\mathbf{C}|_{\{R_C|R_B\}}\mathbf{M}_\mathbf{B}\right), \mathbf{T}_\mathbf{B} \right]\!\right] \right\|_2^2 \right\}. \quad (20)
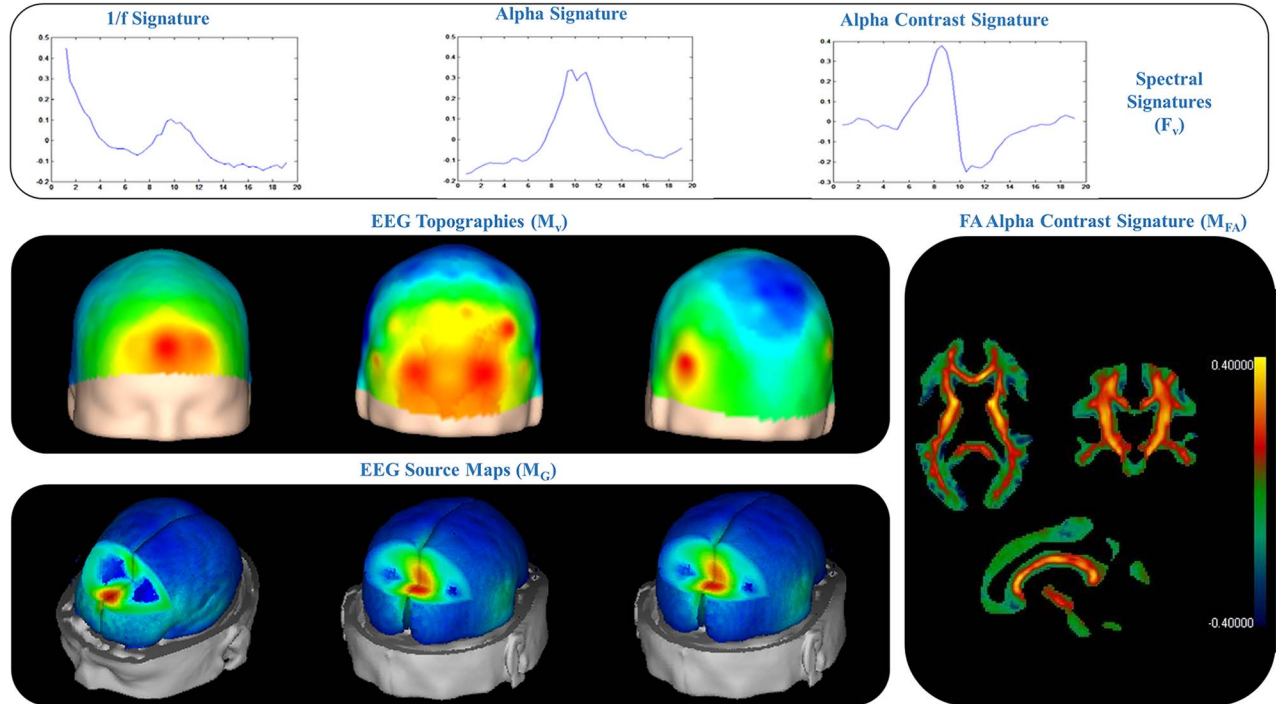\end{aligned}
$$

Fig. 13. *Multiway partial least squares for DTI–FA and EEG fusion. Spectral signatures of EEG ($\mathsf{F_V}$) and the spatial signature of FA ($\mathsf{M_{FA}}$) are shown. Both scalp distribution ($\mathsf{M_V}$) and its corresponding brain electrical tomography of the derivation signatures ($\mathsf{M_G}$) are demonstrated under the corresponding spectral signatures. The frequency signatures of the three EEG atoms show three distinct patterns. The first one is dominated by a low $1/f$ behavior. We will therefore identify this signature as the "slow wave signature." The second one is a pure alpha peak. We will therefore identify this atom as the "alpha signature." The most interesting signature, the third one, seems to be a contrast between low and high alpha. We will therefore identify this signature as the "alpha contrast signature." The slow wave signature shows a frontal source near the theta signature described in [45]. As clearly seen, the spectral alpha signature colocalizes with the source of the alpha rhythm, as described in [45]. The new alpha contrast signature has a localization very similar to the alpha signature. We only showed the corresponding spatial signature of the FA to alpha contrast signature. Major tracts, especially talamo–cortical, are enhanced in this signature in accordance with [116].*
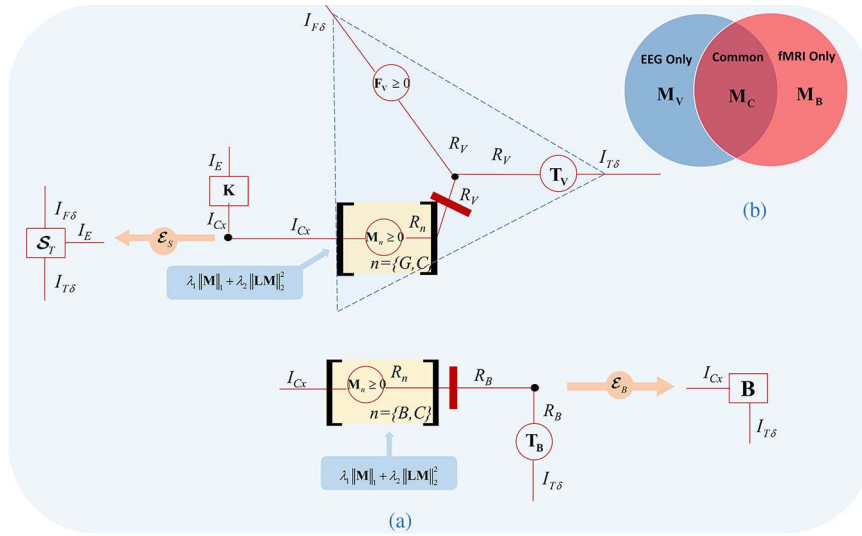
Fig. 14. *Coupled matrix-tensor factorization. (a) M–P diagram for the coupled matrix-tensor factorization for EEG/fMRI fusion. The EEG tensor $\mathcal{S}_T$ and the fMRI matrix $\mathbf{B}$ are decomposed simultaneously on common and discriminant spatial subspaces to encompass different physiological sources. The spatial signature $\mathbf{M}$ involves common component $\mathbf{M_C}$ and two uncommon $\mathbf{M_G}, \mathbf{M_B}$ components. The fMRI spatial signature is $(\mathbf{M_C}|_{\{R_C|R_B\}} \mathbf{M_B})$ and the temporal signature is $\mathbf{T_B}$. For EEG, the spatial signature of the generators is $(\mathbf{M_C}|_{\{R_C|R_G\}} \mathbf{M_G})$, the temporal signature is $\mathbf{T_V}$, and the spectral signature is $\mathbf{F_V}$. By incorporating the lead field matrix $\mathbf{K}$, the model extends the decomposition of EEG to source space. M–P diagrams of EEG and fMRI are separated for a better visualization. (b) Explicit representation for common and discriminative subspaces. Note that the common subspace is represented with $\mathbf{M_C}$.*

Furthermore, we impose nonnegativity, orthogonality, smoothness, and sparsity constraints on spatial factors to ensure uniqueness. The corresponding M–P diagram is shown in Fig. 14, and new parameters with constraints are found by

$$
\begin{aligned}
&(\hat{\mathbf{M}}_\mathbf{C}, \hat{\mathbf{M}}_\mathbf{G}, \hat{\mathbf{T}}_\mathbf{V}, \hat{\mathbf{F}}_\mathbf{V}, \hat{\mathbf{M}}_\mathbf{B}, \hat{\mathbf{T}}_\mathbf{B}) \\
&= \underset{\mathbf{M_C}, \mathbf{M_G}, \mathbf{T_V}, \mathbf{F_V}, \mathbf{M_B}, \mathbf{T_B}}{\arg\min} \\
&\times \left\{ \frac{1}{2} \left\| \mathcal{S}_T - \left[\!\left[\mathbf{K}\left(\mathbf{M_C}|_{\{R_C|R_G\}} \mathbf{M_G} \right), \mathbf{T_V}, \mathbf{F_V} \right]\!\right] \right\|_2^2 \right. \\
&\quad + \gamma \frac{1}{2} \left\| \mathbf{B} - \left[\!\left[\left(\mathbf{M_C}|_{\{R_C|R_B\}} \mathbf{M_B} \right), \mathbf{T_B} \right]\!\right] \right\|_2^2 \\
&\quad + \lambda_1 \|\mathbf{M_C}\|_1 + \frac{1}{2}\lambda_2 \|\mathbf{LM_C}\|^2 + \lambda_3 \|\mathbf{M_G}\|_1 \\
&\quad \left. + \frac{1}{2}\lambda_4 \|\mathbf{LM_G}\|^2 + \lambda_5 \|\mathbf{M_B}\|_1 + \frac{1}{2}\lambda_6 \|\mathbf{LM_B}\|^2 \right\}
\end{aligned}
$$

$$
\begin{aligned}
\text{s.t.} \quad & \left(\mathbf{M_C}|_{\{R_C|R_G\}} \mathbf{M_G} \right)^T \left(\mathbf{M_C}|_{\{R_C|R_G\}} \mathbf{M_G} \right) = \mathbf{I}, \\
& \left(\mathbf{M_C}|_{\{R_C|R_B\}} \mathbf{M_B} \right)^T \left(\mathbf{M_C}|_{\{R_C|R_B\}} \mathbf{M_B} \right) = \mathbf{I}, \\
& \mathbf{M_C} \geq 0, \quad \mathbf{M_G} \geq 0, \quad \mathbf{M_B} \geq 0, \quad \mathbf{F_V} \geq 0. \quad (21)
\end{aligned}
$$

The model in (21) can also be interpreted as the estimation of neuronal activity through two sources of infor-

mation with multiple priors. The γ parameter takes into account the scale difference between EEG and fMRI.

We applied the proposed algorithm on a simultaneously recorded EEG–fMRI data [121]. In this experiment, flashing light stimuli in 13 frequencies in the range of 6–42 Hz were presented in a block design paradigm. For this analysis, data segments in the resting periods of the 6-Hz stimulation session of one subject are used. Further information about the acquisition and preprocessing of the data can be found in [121].

The EEG was filtered with a high-pass filter with a cutoff frequency at 60 Hz and segmented in 2981-ms duration segments to match the repetition time of fMRI ($TR = 2981$ ms). The Thomson multitaper method was used to calculate the power spectrum of each segment [122]. We extracted the resting periods of the whole experiment and used them for further analysis. This resulted in an EEG tensor $\mathcal{S}_T \in \mathbb{R}^{I_E \times I_{T\delta} \times I_{F\delta}}$: with $I_E = 31$ channels, $I_{T\delta} = 38$ time points, and $I_{F\delta} = 58$ frequency points. The lead field was computed using a realistic head model with three homogenous isotropic conductor boundaries based on the MNI brain atlas.

The fMRI data were normalized to standard MNI space. The voxels on the cortical grid of EEG source space were extracted. Grand mean scaling over the session for the voxels inside the mesh was performed, and BOLD values were normalized to obtain a percentage change. In the end, we had an fMRI data matrix $\mathbf{B} \in \mathbb{R}^{I_{Cx} \times I_{T\delta}}$ with $I_{Cx} = 5124$

cortical grid points and $I_{T\delta} = 38$ time points. In total, three atoms were extracted from constrained CMTF model: one for the common atom, one for the individual EEG atom, and one for the individual fMRI atom.

Fig. 15(I) shows the spatial, temporal, and spectral signatures of the common atom. Since the two data sets were coupled only in spatial dimension, two temporal signatures for each modality were obtained—but these show only irregular activity and will not be described further.

The common spatial signature $\mathbf{M_C}$ shows a clear activation in occipital areas with an EEG spectral signature peaking at 10 Hz, which identifies it with the α activity found by other techniques and shown in Figs. 6, 12, and 13, in line with [45]. Pearson's correlation coefficient between the temporal signatures of EEG and fMRI of this common component is found to be -0.3346 with a p-value of 0.04, which is not corrected for serial correlations.

The discriminant fMRI atom is shown in Fig. 15(II). The spatial signature showed activation mostly in inferior frontal areas of left and right hemispheres, inferior parietal and middle temporal areas of the right hemisphere, precuneus and caudate. When the model order of the fMRI is increased these regions are distributed on separate atoms (results not shown). Therefore, it seems that the discriminant atom of fMRI might be the result of the interaction of several of the reported resting state networks.

The discriminant EEG atom shows a $1/f$ decay in the spectral signature with diffused activations in the inferior and middle frontal areas, as well as the temporal areas of both hemispheres, which identify it with the $\boldsymbol{\xi}$ EEG process mentioned in Section VI-A [see Fig. 15(III)(a)–(b)].

It seems interesting that the two last techniques, one applied to EEG/DTI data and the other to EEG/fMRI data, seem to support the $\boldsymbol{\xi} - \alpha$ model proposed in [117].

In this section, we presented a PARAFAC-based CMTF method for the decomposition of EEG and fMRI on the common and discriminative subspaces by considering the discrepancy in the neural origins. CMTF is not limited to PARAFAC decomposition and can be modified by using other decomposition methods, e.g., the Tucker method, to account for the interactions between the signatures [39], [123], [124]. CMTF differs from the linked ICA [125] in the sense that statistical independence of the spatial signatures is not required and common profiles can be divided into two subspaces. Recently, scalable and fast algorithms for CMTF have been developed and applied on the decomposition of fMRI and behavioral data [126].

C. Other Fields of Application of Tensor Fusion

By integrating different measurements of a phenomenon we can overcome the lack of precision they offer, providing a complementary vision of things we cannot measure directly.

Many examples of multimodal data integration of different fields are available, such as dynamic clustering by the combination of text and image information on the web [1]; speech recognition using complementary hand gesture [127]; multisensory image fusion, that is, merging relevant information proceeding from many images that comes from multiple sensors [128]; and human emotion recognition by exploring data arising from speech, face image, and thermal image [129]. Multimodal fusion for biological data is found in [3] for integration of gene expression data with text information and in [130] for metabolomics.

VII. NONLINEAR TENSOR MODELS

We have already mentioned, in passing, nonlinear extensions of tensor methods. Here we comment on a possible tensor formulation for neural mass models.

There are several reasons to consider nonlinear equations in EEG and fMRI modeling. As emphasized in [80], all dynamic imaging modalities must be cast into a state–space form, with both dynamical equations that model the evolution of brain dynamics (which is what we are interested in) and observation equations that explain how these brain dynamics are reflected as time series measured by the different imaging technologies. In particular, we have the following.

- Brain dynamics at the mesoscopic level that we are dealing with are best represented by neural mass modeling [62] or neural field modeling [131].
- The observation equation for the fMRI is highly nonlinear [18], [132].
- Though the EEG forward model is linear with the quasi-static approximation [16], an assumption we will follow here, there is some doubt that this condition might not hold for high-frequency activity. If this is so, even this forward model will be nonlinear.

In order to make ideas concrete, we will illustrate ideas with the state–space formulation of the EEG forward problem with a neural mass model dynamical equation. This is a slight extension of (1)

$$\mathbf{v}_t = \mathbf{K}\mathbf{g}_t + \mathbf{e}_{v,t} = \mathbf{K}\mathbf{O}\mathbf{x}_t + \mathbf{e}_{v,t}$$
$$\dot{\mathbf{g}} = f(\mathbf{x}; \boldsymbol{\theta}) + \boldsymbol{\varepsilon}$$

where $\mathbf{v}_t$ is the vector of EEG measurements at time t, $\mathbf{x}$ are the complete state–space variables that characterize the neural dynamics, $\mathbf{K}$ is the lead field, and $\mathbf{O}$ is an observation matrix that transforms state variables of the neural mass model $\mathbf{x}$ to current density $\mathbf{g}$. $\boldsymbol{\varepsilon}$ denotes white noise. The function f is nonlinear. In symbol $\boldsymbol{\theta}$, we lump
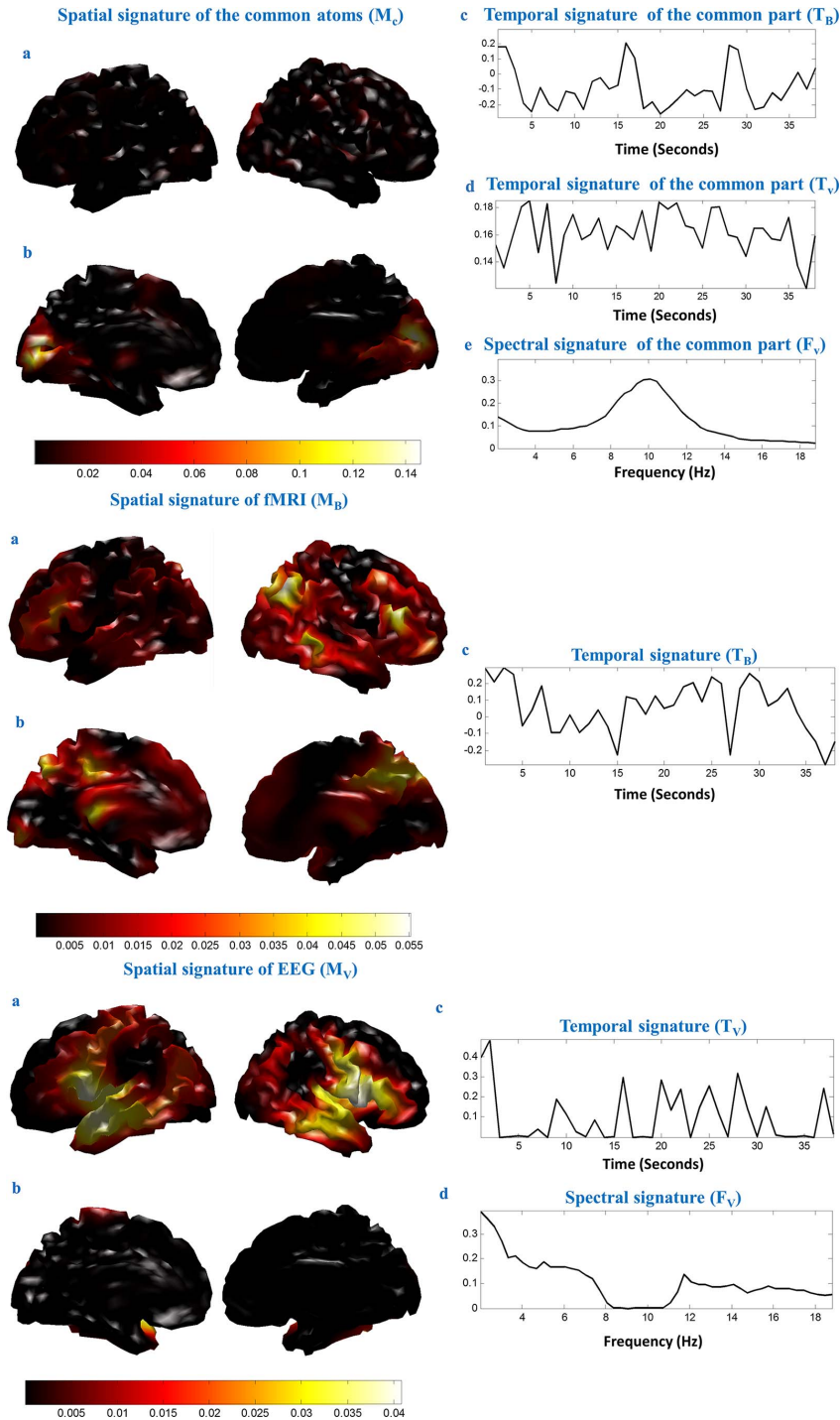
Fig. 15. *Coupled matrix-tensor factorization on the EEG–fMRI data. (I) Common atom extracted from CMTF. (a) and (b) The spatial signature* $(\mathbf{M}_C)$ *shows the distribution of activation of the spatial signature on the lateral and medial views of left and right hemispheres. Activity is localized in the occipital cortex. (c) The fMRI temporal signature of the common atom* $[\mathbf{T}_B(:,1)]$*. (d) The EEG temporal signature of the common atom* $[\mathbf{T}_V(:,1)]$*. (e) The EEG spectral signature of the common atom* $[\mathbf{F}_V(:,1)]$*. The 10-Hz peak in the EEG spectral signature indicates an alpha band activity. (II) Discriminant fMRI atom. (a) and ;(b) The spatial signature of the discriminant fMRI atom* $[\mathbf{M}_B(:,2)]$ *projected on the lateral and medial views of the left and right hemispheres. (c) The temporal course of the discriminant fMRI atom* $[\mathbf{T}_B(:,2)]$*. fMRI activity is diffused mostly in the frontal and temporal regions. (III) Discriminant EEG atom. (a) and (b) The spatial signature of the discriminant EEG atom projected on the lateral and medial views of the left and right hemispheres* $[\mathbf{M}_G(:,2)]$*. A diffused activity is revealed. (c) The temporal signature* $[\mathbf{T}_V(:,2)]$*. (d) The spectral signature* $[\mathbf{F}_V(:,2)]$*. Energy of the spectral signature decreases toward higher frequencies showing the* ξ *process. Spatial distribution is diffused over temporal and inferior frontal areas. All of the signatures are normalized to the unit norm.*

together all the parameters of neural masses and their connectivity. Estimation of $\mathbf{x}$ and $\boldsymbol{\theta}$ is a classic continuous to discrete filtering problem that can be dealt with by applying Kalman filtering [62] or with the variational techniques described by [133].

For the purpose of illustration suppose that we have the observation equation equal to the identity function, i.e., the dynamical equations are the observed variables. Then, our purpose will be to estimate $\boldsymbol{\theta}$ from the following stochastic differential equations model:

$$\dot{\mathbf{x}} = f(\mathbf{x}; \boldsymbol{\theta}) + \boldsymbol{\varepsilon}.$$

In [134], it was shown that this equation may be discretized with a stochastic version of the exponential Euler integrator as

$$\mathbf{x}_{t+\Delta t} = \mathbf{J}(t)^{-1}\left(e^{\mathbf{J}(t)\Delta t} - \mathbf{I}\right)f(\mathbf{x}_t; \boldsymbol{\theta}) + \omega_t$$

where $\mathbf{J}(t)$ is the Jacobian of function f at time t, Δt is the time step of integration, and ω_t is a noise term defined in the cited reference.[4] In some circumstances, this equation can be further simplified to

$$\mathbf{x}_{t+\Delta t} = \mathbf{B}(\mathbf{x}_t; \boldsymbol{\theta})\mathbf{x}_t + \omega_t \tag{22}$$

which is simply a nonlinear regression. This naturally leads to a tensor formation.

For multiple instants $t = 1, \ldots, I_T$, concatenation of the vector $\mathbf{x}_{t+\Delta t}$ is equal to the matrix $\mathbf{X} \in \mathbb{R}^{I_s \times I_T}$ defined as $\mathbf{X} = [\mathbf{x}_{t+\Delta t}]_{t=1:I_T}^{\{1|\cdots|1\}}$. $\mathbf{B}$'s are concatenated to obtain the parameter tensor $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{I_s \times I_s \times I_T \times I_\tau}$ as $\boldsymbol{\mathcal{B}} = [\mathbf{B}]_{t=1:I_T}^{\{1|\cdots|1\}}$. Here $I_\tau = 1$ is used for the time step. Finally, the past values of the state variables are $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_s \times I_T \times I_\tau}$ defined by $\boldsymbol{\mathcal{X}} = [\mathbf{x}_t]_{t=1:I_T}^{\{1|\cdots|1\}}$. The tensor formulation of (22) is expressed by the tensor convolution as follows:

$$\mathbf{X}(i_s, i_t) = \sum_{p=1}^{1} \boldsymbol{\mathcal{B}}(i_s, i_s, i_\tau, p) \bullet_{\{i_s, i_\tau\}} \boldsymbol{\mathcal{X}}(i_s, i_\tau, i_t - p)$$
$$\mathbf{X} = \boldsymbol{\mathcal{B}} *_{\{I_T\}} \boldsymbol{\mathcal{X}} + \Omega$$

which is of the type of regression problem we have been dealing with, albeit of a highly nonlinear character. Note that the convolution here contains contraction over two modes.

[4]As discussed in the cited paper, it should be noted that this is not the formula to be used for actual calculations due to numerical inaccuracy.

It is interesting to note that integration schemes of the type just described have also been extended to equations with multiplicative noise, to delay differential equations, and to partial differential stochastic equations. Stochastic partial differential equations have been the subject of recent work with tensor representations that are quite promising [135], [136].

VIII. ALGORITHMS AND SOFTWARE FOR TENSOR-BASED PROBLEMS

Throughout this review, we have concentrated on the graphical properties of tensor diagrams and on the modeling advantages of tensor models. In this section, we focus on a more practical subject, that is, the methods and algorithms used for fitting such models. We explain the choices we have made motivated by the main numerical difficulties. We then describe major software packages that are available to implement tensor models.

A. Optimization of the Model Functional

Though there are many techniques for tensor network decomposition, we present only the PARAFAC decomposition. Besides clarity of interpretation, a major reason for this selection is the ease with which it may be augmented with multiple penalizers that enhance interpretability and numerical stability. In fact, the alternating least squares algorithm (ALS), often used to estimate the PARAFAC model, is the easily implementable and flexible algorithm. ALS is an iterative algorithm in which, at each step, only one signature is estimated, considering all others to be fixed. Since each step is a linear regression, penalization methods can be incorporated naturally. ALS has been improved with line search at each step [137], though it can converge slowly, especially when the components are collinear.

Other methods such as gradient-based optimization methods [138] and generalized Schur decomposition [139] have been developed as an alternative to overcome the limitations of ALS. In addition, probabilistic methods for general tensor factorizations are presented in [140] and [141]. These are alternatives to the algorithmic choices we have made.

Our approach in this paper has been to use a modification of ALS, the hierarchical alternating least squares (HALS) algorithm in which, at each step of ALS, only one of the components of a factor is estimated, fixing other signatures of all atoms [142]. In particular, we have implemented a gradient-based HALS for the estimation of parameters for which orthogonality and nonnegativity are required: orthogonality can easily be imposed columnwise (refer to [143] for details).

In addition, since we use a combination of penalties with different forms of regression and decomposition, we separate our optimization problem into its additive components and combine the solutions by means of the

alternating direction method of multipliers (ADMM) [144]–[146].

The distributed nature of ADMM is also especially geared to solve very large problems in big data since multiple penalization problems may merged into a single global optimization which is solved piecewise. An example of this splitting behavior of ADMM is illustrated in (14). Influence fields may be required to be smooth both in incoming (receiver) and outgoing (emitter) directions with a different profile. An analytical solution with these two priors is not possible due to the penalization on two of the dimensions of the connectivity tensor. ADMM shows its strength by splitting this optimization into two tractable subproblems. A similar approach can be adopted in (17) and (21) for assigning multiple priors on components of a PARAFAC decomposition.

Since estimation of the factor matrices of the PARAFAC decomposition is a nonconvex optimization problem, an algorithm may reach different solutions with different starting points. In the ALS method, the objective function decreases at each step of the algorithm, but there is no guarantee that global minimum may be reached. We address the problem heuristically by means of the following approaches:

- running the algorithm with multiple random initial values;
- alternatively using, as a starting point, the eigenvectors of the unfolded tensor to be fitted [147];
- using a combination of the previous two schemes;
- estimating all runs and retaining the one with the best fit;
- in the case of models with nonnegative factor matrices, using a modified nonnegative double singular value decomposition proposed in [148] for the initial eigenanalysis;
- GC analysis with PARAFAC requires a decomposition at each step of the ADMM algorithm; in this analysis, a "warm start" is used with initial values for the decomposition set as the factor estimates from the previous step.

However, these techniques do not guarantee convergence to the global minima; this is an area of increased current research [149], [150]. A definitive solution might be obtained by approximating the models here with alternative convex ones.

B. Selection of Model Parameters

Multiple penalization methods entail tuning of a large amount of hyperparameters, with scant knowledge about the actual effect of parameter changes.

We have adopted the approach to carry out parameter selection based on the Bayesian information criterion (BIC). This requires the estimation of the degrees of freedom (DoF), which reflects model complexity, for a given set of parameters. While general formulations of DoF exist [151] and explicit formulas are available for many of the most common regularization problems [152], [153], the complicated nature (e.g., nondifferentiability) of some regularizers makes it very difficult to find an accurate mathematical expression. For this means, general purpose nonparametric methods have been developed such as in [154]. DoF for tensor problems are calculated according to the penalization function applied on each signature, e.g., for a smooth Lasso type of penalization, the DoF formulated by [152] is used.

The hyperparameters chosen are those that minimize BIC, a search over a suitable grid being carried out. One of the most extreme cases in terms of the number of hyperparameters is the CMTF. This is the CMTF objective function in (21) with ten hyperparameters: six for smooth Lasso-type penalization, three for the orthogonality constraints on spatial signatures, and one for the variance difference between two modalities. On the one hand, hyperparameters for orthogonality constraints are determined as described in [143]. The rest of the parameters are found by minimizing the BIC formulations spelled out in the Appendix.

C. Selection of the Number of Atoms

PARAFAC-type decompositions require the determination of the number of components or atoms, which is also known as the tensor rank. There is not any straightforward algorithm to determine the rank of a specific given tensor; in fact, this problem is NP-hard [155]. Furthermore, best rank approximation of a tensor may not exist [156]. Nevertheless, practical methods and heuristics have been developed in order to automatically determine the number of components in the PARAFAC model such as Corcondia [157], [158]. For the determination of the model order of PARAFAC-based models defined in this paper, we used Corcondia and evaluated the explained variance for a different number of components. For the selection of the number of common components (R_C) of the CMTF model, a heuristic approach is applied by decomposing two data sets independently. After deciding for an initial value of R_C based on spatial signatures, several model orders are tested and the best model is selected according to the maximum fit and observation of the factors. We also checked the collinearity of the factors and the convergence of the algorithm.

D. Software

There are available software sources for tensor decompositions and tensor operations. The MATLAB Tensor Toolbox offers classes for tensor operations and has several tensor decomposition algorithms [159]. The N-Way Toolbox supports constrained decompositions and N–PLS [160]. CuBatch is a MATLAB-based software for both tensor and classical data analysis, and validation tools providing graphical outputs [161]. Tensorlab provides several decomposition algorithms, including structured data fusion of tensors and complex optimization tools [162].

TDALAB and Tensorbox have a graphical user interface supporting several tensor decomposition types [163], [164]. The MATLAB CMTF Toolbox provides constrained and unconstrained algorithms for the joint decomposition of tensors with different orders by using PARAFAC [165]. ERPWAVELAB is a MATLAB-based toolbox for multichannel time/frequency analysis of EEG and MEG by various tensor decompositions, including PARAFAC, shifted PARAFAC, and Tucker [166]. The TT Toolbox includes functions for basic operations with tensor-train tensors that are the low-parametric representations of high-dimensional tensors [167]. Python implementation of the TT Toolbox is also available [168]. The Tensor Toolbox in Python includes the decomposition of tensors in tensor-train format and spectral tensor-train format [169]. For the construction and manipulation of tensors in the hierarchical Tucker format, the Hierarchical Tucker Toolbox is available [170]. The TensorReg Toolbox for MATLAB provides sparse PARAFAC and Tucker regression functions [171].

Many of the algorithms developed by the authors can be found at http://www.research.cneuro.cu/categories/tensor and http://neurosignal.boun.edu.tr/software/tensor.

IX. CONCLUSION

In this paper, we have presented a general framework for tensor analysis of single modality model inversion and multimodal data fusion. We have introduced the Markov–Penrose (M–P) diagrams to unify graphical tensor notations with that for directed acyclic graphical (DAG) description of Bayesian statistical models. Using these diagrams, different approaches for the solution of inverse problems of EEG and fMRI have been described as well as models for their fusion in common domains. Additionally, we have proposed a tensor MAR for modeling the causal brain networks and have reviewed symmetrical fusion methods with the proposed notation. We have reviewed algorithms and software packages for implementation of tensor-based problems.

EEG and fMRI are mediated by different physiological processes from neural activation leading to differences in their spatial and temporal resolutions. Biophysical models have been addressed to fuse electrical and metabolic signals in mesoscale. Due to the indirect nature of these signals, inverse problems for each modality should be solved to cover the interactions between modalities which are intrinsically ill-posed in their nature. The examples shown with simulations and real data support the usefulness of this type of approach.

As the amount of neuroimaging data increases tremendously, methods dealing with this problem should be developed. Statistical methods based on tensors embrace the high dimensionality of the multimodal data. The M–P tensor notation based on DAGs and Penrose diagrams, which unify mathematical models for connectivity and multimodal fusion, may play a heuristic role in suggesting new ways to analyze data, not only in neuroscience, but possibly in other fields. ∎

APPENDIX

We will use the BIC formula given as

$$\mathrm{BIC} = \log(\hat{\sigma}^2) + df \frac{\log(N)}{N}$$

where N is the number of observations, df is the DoF, and $\hat{\sigma}^2$ is the error variance estimated from the residual sum of squares (RSS) as $\hat{\sigma}^2 = \mathrm{RSS}/N$.

BIC formulations for coupled and uncoupled components of the spatial signatures are given as

$$
\begin{aligned}
&\mathrm{BIC}(\mathbf{M_C}) \\
&= \log\left(\frac{\left\| \boldsymbol{\mathcal{S}}_T - \left[\!\left[\mathbf{K}\left(\mathbf{M_C}\big|_{\{R_C|R_G\}}\mathbf{M_G}\right), \mathbf{T_V}, \mathbf{F_V} \right]\!\right] \right\|_2^2}{(n_\mathbf{V} + n_\mathbf{B})} \right.\\
&\quad \left. + \frac{\gamma \left\| \mathbf{B} - \left[\!\left[\left(\mathbf{M_C}\big|_{\{R_C|R_B\}}\mathbf{M_B}\right), \mathbf{T_B} \right]\!\right] \right\|_2^2}{(n_\mathbf{V} + n_\mathbf{B})} \right) \\
&\quad + df(\mathbf{M_C}) \log(n_\mathbf{V} + n_\mathbf{B})/(n_\mathbf{V} + n_\mathbf{B}) \\
&\mathrm{BIC}(\mathbf{M_G}) \\
&= \log\left(\frac{\left\| \boldsymbol{\mathcal{S}}_T - \left[\!\left[\mathbf{K}\left(\mathbf{M_C}\big|_{\{R_C|R_G\}}\mathbf{M_G}\right), \mathbf{T_V}, \mathbf{F_V} \right]\!\right] \right\|_2^2}{n_\mathbf{V}} \right) \\
&\quad + df(\mathbf{M_V}) \log(n_\mathbf{V})/n_\mathbf{V} \\
&\mathrm{BIC}(\mathbf{M_B}) \\
&= \log\left(\frac{\left\| \mathbf{B} - \left[\!\left[\left(\mathbf{M_C}\big|_{\{R_C|R_B\}}\mathbf{M_B}\right), \mathbf{T_B} \right]\!\right] \right\|_2^2}{n_\mathbf{B}} \right) \\
&\quad + df(\mathbf{M_B}) \log(n_\mathbf{B})/n_\mathbf{B}.
\end{aligned}
$$

$n_\mathbf{V}$ and $n_\mathbf{B}$ are the number of elements in $\boldsymbol{\mathcal{S}}_T$ and $\mathbf{B}$, respectively, and df is the DoF computed as in [152]. Hyperparameters λ_1 to λ_6 and γ in (21) are found as the minimum of the BIC multidimensional arrays given above.

REFERENCES

[1] W. Lu, L. Li, T. Li, H. Zhang, and J. Guo, "Web multimedia object clustering via information fusion," in *Proc. Int. Conf. Document Anal. Recognit.*, 2011, no. 1, pp. 319–323.

[2] F. Janoos, H. Denli, and N. Subrahmanya, "Multi-scale graphical models for spatio-temporal processes," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Cambridge, MA, USA: MIT Press, 2014, pp. 316–324.

[3] E. Zeng, C. Yang, T. Li, and G. Narasimhan, "Clustering genes using heterogeneous data sources," *Int. J. Knowl. Disc. Bioinf.*, vol. 1, no. 2, pp. 12–28, 2010.

[4] BRAIN Initiative, 2015. [Online]. Available: http://braininitiative.nih.gov/

[5] Human Brain Project, 2015. [Online]. Available: https://www.humanbrainproject.eu/

[6] T. Jiang, "Brainnetome: A new-ome to understand the brain and its disorders," *NeuroImage*, vol. 80, pp. 263–272, 2013.

[7] H. A. Whiteford et al., "Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010," *Lancet*, vol. 382, no. 9904, pp. 1575–1586, 2013.

[8] K. Amunts et al., "BigBrain: An ultrahigh-resolution 3D human brain model," *Science*, vol. 340, no. 6139, pp. 1472–1475, 2013.

[9] S. N. Sotiropoulos et al., "Advances in diffusion MRI acquisition and processing in the Human Connectome Project," *NeuroImage*, vol. 80, pp. 125–143, 2013.

[10] K. Uğurbil et al., "Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project," *NeuroImage*, vol. 80, pp. 80–104, 2013.

[11] N. K. Logothetis, "What we can do and what we cannot do with fMRI," *Nature*, vol. 453, no. 7197, pp. 869–878, Jun. 2008.

[12] C. M. Michel et al., "EEG source imaging," *Clin. Neurophysiol.*, vol. 115, no. 10, pp. 2195–2222, 2004.

[13] D. A. Boas, C. E. Elwell, M. Ferrari, and G. Taga, "Twenty years of functional near-infrared spectroscopy: Introduction for the special issue," *NeuroImage*, vol. 85, pp. 1–5, 2014.

[14] A. W. Toga and J. C. Mazziotta, Eds., *Brain Mapping: The Methods*. New York, NY, USA: Academic, 2002.

[15] O. Sporns, *Networks of the Brain*. Cambridge, MA, USA: MIT Press, 2011.

[16] P. L. Nunez and R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG*, 2nd ed. Oxford, U.K.: Oxford Univ. Press, 2006.

[17] G. Buzsaki, *Rhythms of the Brain*. Oxford, U.K.: Oxford Univ. Press, 2006.

[18] K. J. Friston, A. Mechelli, R. Turner, and C. J. Price, "Nonlinear responses in fMRI: The Balloon model, Volterra kernels, and other hemodynamics," *NeuroImage*, vol. 12, no. 4, pp. 466–477, Oct. 2000.

[19] K. J. Friston et al., "Granger causality revisited," *NeuroImage*, vol. 101, pp. 796–808, Jul. 5, 2014.

[20] S. Makeig, S. Debener, J. Onton, and A. Delorme, "Mining event-related brain dynamics," *Trends Cogn. Sci.*, vol. 8, no. 5, pp. 204–210, 2004.

[21] S. Makeig and J. Onton, "ERP features and EEG dynamics," in *The Oxford Handbook of Event-Related Potential Components*, E. S. Kappenman and S. J. Luck, Eds. Oxford, U.K.: Oxford Univ. Press, 2011.

[22] V. D. Calhoun and T. Adalı, "Multisubject independent component analysis of fMRI: A decade of intrinsic networks, default mode, and neurodiagnostic discovery," *IEEE Rev. Biomed. Eng.*, vol. 5, pp. 60–73, Jan. 2012.

[23] P. A. Valdés-Sosa et al., "Model driven EEG/fMRI fusion of brain oscillations," *Human Brain Mapping*, vol. 30, no. 9, pp. 2701–2721, Sep. 2009.

[24] K. Uludağ and A. Roebroeck, "General overview on the merits of multimodal neuroimaging data fusion," *NeuroImage*, vol. 102, pp. 3–10, May 2014.

[25] J. Sui, T. Adalı, Q. Yu, J. Chen, and V. D. Calhoun, "A review of multivariate methods for multimodal fusion of brain imaging data," *J. Neurosci. Methods*, vol. 204, no. 1, pp. 68–81, Feb. 2012.

[26] N. M. Correa, T. Eichele, T. Adalı, Y.-O. Li, and V. D. Calhoun, "Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI," *NeuroImage*, vol. 50, no. 4, pp. 1438–1445, May 2010.

[27] Y. Levin-Schwartz, V. D. Calhoun, and T. Adalı, "Data-driven fusion of EEG, functional and structural MRI: A comparison of two models," in *Proc. 48th Annu. Conf. Inf. Sci. Syst.*, 2014, doi:10.1109/CISS.2014.6814108.

[28] J. Sui et al., "Three-way (N-way) fusion of brain imaging data based on mCCA+jICA and its application to discriminating schizophrenia," *NeuroImage*, vol. 66, pp. 119–132, Feb. 2013.

[29] R. F. Silva, S. Plis, T. Adalı, and V. D. Calhoun, "A statistically motivated framework for simulation of stochastic data fusion models applied to multimodal neuroimaging," *NeuroImage*, vol. 102, pp. 92–117, Apr. 2014.

[30] X. Lei, P. A. Valdés-Sosa, and D. Yao, "EEG/fMRI fusion based on independent component analysis: Integration of data-driven and model-driven methods," *J. Integr. Neurosci.*, vol. 11, no. 3, pp. 313–337, Sep. 2012.

[31] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multimodal factor analysis," *UCLA Working Papers Phonetics*, vol. 16, pp. 1–84, 1970.

[32] A. Field and D. Graupe, "Topographic component (parallel factor) analysis of multichannel evoked potentials: Practical issues in trilinear spatiotemporal decomposition," *Brain Topography*, vol. 3, no. 4, pp. 407–423, 1991.

[33] J. Möcks, "Topographic components model for event-related potentials and some biophysical considerations," *IEEE Trans. Biomed. Eng.*, vol. 35, no. 6, pp. 482–484, Jun. 1988.

[34] R. Bro, *Multi-Way Analysis in the Food Industry: Models, Algorithms, and Applications*. Amsterdam, The Netherlands: Universiteit van Amsterdam, 1998.

[35] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multiway Data Analysis and Blind Source Separation*. New York, NY, USA: Wiley, 2009.

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

[37] M. Mørup, "Applications of tensor (multiway array) factorizations and decompositions in data mining," *Wiley Interdiscip. Rev. Data Mining Knowl. Disc.*, vol. 1, no. 1, pp. 24–40, Jan. 2011.

[38] E. Acar and B. Yener, "Unsupervised multiway data analysis: A literature survey," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 1, pp. 6–20, Jan. 2009.

[39] A. Cichocki, "Tensor decompositions: A new concept in brain data analysis?" May 2013. [Online]. Available: http://arxiv.org/abs/1305.0395

[40] A. Cichocki, "Era of big data processing: A new approach via tensor networks and tensor decompositions," Mar. 2014. [Online]. Available: http://arxiv.org/abs/1403.2048

[41] A. Cichocki, "Tensor networks for big data analytics and large-scale optimization problems," 2014. [Online]. Available: http://arxiv.org/abs/1407.3124

[42] E. Acar, M. A. Rasmussen, F. Savorani, T. Næs, and R. Bro, "Understanding data fusion within the framework of coupled matrix and tensor factorizations," *Chemom. Intell. Lab. Syst.*, vol. 129, pp. 53–63, Jun. 2013.

[43] F. Miwakeichi et al., "Decomposing EEG data into space-time-frequency components using parallel factor analysis," *NeuroImage*, vol. 22, no. 3, pp. 1035–1045, Jul. 2004.

[44] E. Martínez-Montes, J. M. Sánchez-Bornot, and P. A. Valdés-Sosa, "Penalized parafac analysis of spontaneous EEG recordings," *Stat. Sin.*, vol. 18, no. 4, pp. 1149–1464, 2008.

[45] E. Martínez-Montes, P. A. Valdés-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen, "Concurrent EEG/fMRI analysis by multiway partial least squares," *NeuroImage*, vol. 22, no. 3, pp. 1023–1034, Jul. 2004.

[46] J. M. Sanchez-Bornot, E. Martínez-Montes, A. Lage-Castellanos, M. Vega-Hernández, and P. A. Valdés-Sosa, "Uncovering sparse brain effective connectivity: A voxel-based approach using penalized regression," *Stat. Sin.*, vol. 18, no. 4, pp. 1501–1518, 2008.

[47] P. A. Valdés-Sosa et al., "Granger causality on spatial manifolds: Applications to neuroimaging," in *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*. New York, NY, USA: Wiley-VCH, 2006, pp. 1–53.

[48] P. A. Valdés-Sosa, M. Vega-Hernández, J. M. Sánchez-Bornot, E. Martínez-Montes, and M. A. Bobes, "EEG source imaging with spatio-temporal tomographic nonnegative independent component analysis," *Human Brain Mapping*, vol. 30, no. 6, pp. 1898–1910, Jun. 2009.

[49] M. Vega-Hernández, E. Martínez-Montes, J. M. Sanchez-Bornot, A. Lage-Castellanos, and P. A. Valdés-Sosa, "Penalized least squares methods for solving the EEG inverse problem," *Stat. Sin.*, vol. 18, no. 4, pp. 1535–1551, 2008.

[50] R. Penrose, "Applications of negative dimensional tensors," in *Combinatorial Mathematics and Its Applications*. London, U.K.: Academic, 1971, pp. 221–244.

[51] T. Huckle, K. Waldherr, and T. Schulte-Herbrüggen, "Computations in quantum tensor networks," *Linear Algebra Appl.*, vol. 438, no. 2, pp. 750–781, Jan. 2013.

[52] S. Holtz, T. Rohwedder, and R. Schneider, "The alternating linear scheme for tensor optimization in the tensor train format," *SIAM J. Sci. Comput.*, vol. 34, no. 2, pp. 683–713, 2012.

[53] C. M. Bishop, *Pattern Recognition and Machine Learning.* New York, NY, USA: Springer-Verlag, 2006.

[54] A. Critch, *Algebraic Geometry of Hidden Markov and Related Models.* Berkeley, CA, USA: Univ. California Press, 2013.

[55] A. Critch and J. Morton, "Algebraic geometry of matrix product states," *Symmetry Integrability Geometry, Methods Appl.*, vol. 10, Sep. 2014. [Online]. Available: http://dx.doi.org/10.3842/SIGMA.2014.095

[56] D. L. Schomer and F. H. L. da Silva, Eds., *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 6th ed. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2010.

[57] G. T. Einevoll, C. Kayser, N. K. Logothetis, and S. Panzeri, "Modelling and analysis of local field potentials for studying the function of cortical circuits," *Nature Rev. Neurosci.*, vol. 14, no. 11, pp. 770–785, Oct. 2013.

[58] F. Lopes da Silva, "EEG and MEG: Relevance to neuroscience," *Neuron*, vol. 80, no. 5, pp. 1112–1128, Dec. 2013.

[59] R. B. Buxton, *Introduction to Functional Magnetic Resonance Imaging Principles and Techniques*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[60] J. Riera, "Biophysical models: Neurovascular coupling, cortical microcircuits, and metabolism," in *Encyclopedia of Computational Neuroscience.* New York, NY, USA: Springer-Verlag, 2015, pp. 399–411.

[61] M. Ullsperger and S. Debener, *Simultaneous EEG and fMRI.* Oxford, U.K.: Oxford Univ. Press, 2010.

[62] P. A. Valdes, J. C. Jimenez, J. Riera, R. Biscay, and T. Ozaki, "Nonlinear EEG analysis based on a neural mass model," *Biol. Cybern.*, vol. 81, no. 5–6, pp. 415–424, Nov. 1999.

[63] K. E. Stephan, N. Weiskopf, P. M. Drysdale, P. A. Robinson, and K. J. Friston, "Comparing hemodynamic models with DCM," *NeuroImage*, vol. 38, no. 3, pp. 387–401, Nov. 2007.

[64] A. Marreiros, K. Stephan, and K. Friston, "Dynamic causal modeling," *Scholarpedia*, vol. 5, no. 7, p. 9568, 2010.

[65] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann, "Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain," *Int. J. Psychophysiol.*, vol. 18, pp. 49–65, 1994.

[66] J. Bosch-Bayard *et al.*, "3D statistical parametric mapping of EEG source spectra by means of variable resolution electromagnetic tomography (VARETA)," *Clin. EEG Neurosci.*, vol. 32, no. 2, pp. 47–61, Apr. 2001.

[67] M. D. Plumbley, "Algorithms for nonnegative independent component analysis," *IEEE Trans. Neural Netw.*, vol. 14, no. 3, pp. 534–543, May 2003.

[68] G. H. Glover, "Deconvolution of impulse response in event-related BOLD fMRI," *NeuroImage*, vol. 9, no. 4, pp. 416–429, Apr. 1999.

[69] N. J. Trujillo-Barreto, E. Martínez-Montes, L. Melie-García, and P. A. Valdés-Sosa, "A symmetrical Bayesian model for fMRI and EEG/MEG neuroimage fusion," *Int. J. Bioelectromagn.*, vol. 3, no. 1, 2001.

[70] J. Daunizeau *et al.*, "Symmetrical event-related EEG/fMRI information fusion in a variational Bayesian framework," *NeuroImage*, vol. 36, no. 1, pp. 69–87, May 2007.

[71] T. Murta, M. Leite, D. W. Carmichael, P. Figueiredo, and L. Lemieux, "Electrophysiological correlates of the BOLD signal for EEG-informed fMRI," *Human Brain Mapping*, vol. 36, no. 1, pp. 391–414, 2015.

[72] V. D. Calhoun, L. Wu, K. A. Kiehl, T. Eichele, and G. D. Pearlson, "Aberrant processing of deviant stimuli in schizophrenia revealed by fusion of fMRI and EEG data," *Acta Neuropsychiatr.*, vol. 22, no. 3, pp. 127–138, 2010.

[73] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, Jan. 2000.

[74] H. Becker *et al.*, "EEG extended source localization: Tensor-based vs. conventional methods," *NeuroImage*, vol. 96, pp. 143–157, Aug. 2014.

[75] F. Cong *et al.*, "Tensor decomposition of EEG signals: A brief review," *J. Neurosci. Methods*, vol. 248, pp. 59–69, 2015.

[76] X. F. Gong, C. Wang, Y. Hao, and Q. Lin, "Combined independent component analysis and canonical polyadic decomposition via joint diagonalization," in *Proc. IEEE China Summit/Int. Conf. Signal Inf. Process.*, 2014, no. 1, pp. 804–808.

[77] M. De Vos, D. Nion, S. Van Huffel, and L. De Lathauwer, "A combination of parallel factor and independent component analysis," *Signal Process.*, vol. 92, no. 12, pp. 2990–2999, 2012.

[78] C. F. Beckmann and S. M. Smith, "Tensorial extensions of independent component analysis for multisubject FMRI analysis," *NeuroImage*, vol. 25, no. 1, pp. 294–311, Mar. 2005.

[79] X. F. Gong and Q. H. Lin, "Spatially constrained parallel factor analysis for semi-blind beamforming," in *Proc. 7th Int. Conf. Natural Comput.*, 2011, vol. 1, pp. 416–420.

[80] P. A. Valdés-Sosa, A. Roebroeck, J. Daunizeau, and K. J. Friston, "Effective connectivity: Influence, causality and biophysical modeling," *NeuroImage*, vol. 58, no. 2, pp. 339–361, Apr. 2011.

[81] J. Pearl, "Statistics and causal inference: A review," *Test*, vol. 12, no. 2, pp. 281–345, 2003.

[82] N. Wiener, "The theory of prediction," in *Modern Mathematics for Engineers*, E. F. Beckenbach, Ed. New York, NY, USA: McGraw-Hill, 1956.

[83] H. Akaike, "On the use of a linear model for the identification of feedback systems," *Ann. Inst. Stat. Math.*, vol. 20, no. 1, pp. 425–439, 1968.

[84] C. W. J. Granger, "Economic processes involving feedback," *Inf. Control*, vol. 6, no. 1, pp. 28–48, 1963.

[85] T. Schweder, "Composable markov processes," *J. Appl. Probab.*, vol. 7, no. 2, pp. 400–410, 1970.

[86] M. Eichler, "A graphical approach for evaluating effective connectivity in neural systems," *Philos. Trans. Roy. Soc. Lond. B., Biol. Sci.*, vol. 360, no. 1457, pp. 953–967, 2005.

[87] D. Commenges and A. Gégout-Petit, "A general dynamical statistical model with causal interpretation," *J. Roy. Stat. Soc. Ser. B, Stat. Methodol.*, vol. 71, no. 3, pp. 719–736, 2009.

[88] K. Friston, "Hierarchical models in the brain," *PLoS Comput. Biol.*, vol. 4, no. 11, 2008, doi:10.1371/journal.pcbi.1000211.

[89] M. Vinck *et al.*, "How to detect the Granger-causal flow direction in the presence of additive noise?" *NeuroImage*, vol. 108, pp. 301–318, 2015.

[90] S. M. Smith *et al.*, "Network modelling methods for FMRI," *NeuroImage*, vol. 54, no. 2, pp. 875–891, Jan. 2011.

[91] A. K. Seth, P. Chorley, and L. C. Barnett, "Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling," *NeuroImage*, vol. 65, pp. 540–555, Jan. 2013.

[92] X. Wen, G. Rangarajan, and M. Ding, "Is granger causality a viable technique for analyzing fMRI data?" *PLoS One*, vol. 8, no. 7, Jul. 2013, Art. ID. e67428.

[93] J. Rodrigues and A. Andrade, "Lag-based effective connectivity applied to fMRI: A simulation study highlighting dependence on experimental parameters and formulation," *NeuroImage*, vol. 89, pp. 358–377, Apr. 2014.

[94] F. H. Lin *et al.*, "Increasing fMRI sampling rate improves Granger causality estimates," *PLoS One*, vol. 9, no. 6, pp. 26–29, 2014.

[95] L. Barnett and A. K. Seth, "Granger causality for state-space models," *Phys. Rev. E*, vol. 91, no. 4, Apr. 2015, Art. ID. 040101.

[96] G. R. Murthy, "Concurrent cyber physical systems: Tensor state space representation," Mar. 2013. [Online]. Available: http://arxiv.org/abs/1303.1597

[97] A. K. Seth, A. B. Barrett, and L. Barnett, "Granger causality analysis in neuroscience and neuroimaging," *J. Neurosci.*, vol. 35, no. 8, pp. 3293–3297, 2015.

[98] P. A. Valdés-Sosa *et al.*, "Estimating brain functional connectivity with sparse multivariate autoregression," *Philos. Trans. Roy. Soc. Lond. B, Biol. Sci.*, vol. 360, no. 1457, pp. 969–981, May 2005.

[99] P. Bühlmann and S. van de Geer, *Statistics for High-Dimensional Data*, 1st ed. Berlin, Germany: Springer-Verlag, 2011.

[100] J. Songsiri, *Graphical Models of Time Series: Parameter Estimation and Topology Selection.* Los Angeles, CA, USA: Univ. California Press, 2010.

[101] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, 2013.

[102] O. Semerci, N. Hao, M. E. Kilmer, and E. L. Miller, "Tensor-based formulation and nuclear norm regularization for multienergy computed tomography," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1678–1693, Apr. 2014.

[103] C. R. Dietrich and G. N. Newsam, "Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix," *SIAM J. Sci. Comput.*, vol. 18, no. 4, pp. 1088–1107, 1997.

[104] H. Huang, C. Ding, D. Luo, and T. Li, "Simultaneous tensor subspace selection and clustering: The equivalence of high order SVD and K-means clustering," in

Proc. 14th ACM SIGKDD Int. Conf. Knowl. Disc. Data Mining, 2008, p. 327.

[105] C. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 45–55, Jan. 2010.

[106] E. E. Papalexakis, N. D. Sidiropoulos, and R. Bro, "From K-means to higher-way co-clustering: Multilinear decomposition with sparse latent factors," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 493–506, Jan. 2013.

[107] L. Sorber, M. Van Barel, and L. De Lathauwer, "Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-$(L_r, L_r, 1)$ terms, and a new generalization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 695–720, Apr. 2013.

[108] G. Zhou and A. Cichocki, "Fast and unique tucker decompositions via multiway blind source separation," *Bull. Polish Acad. Sci. Technol. Sci.*, vol. 60, no. 3, pp. 389–405, 2012.

[109] M. Signoretto, L. De Lathauwer, and J. A. K. Suykens, "Nuclear norms for tensors and their use for convex multilinear estimation," Univ. Leuven, Leuven, Belgium, 2010.

[110] M. Signoretto, Q. Tran Dinh, L. Lathauwer, and J. A. K. Suykens, "Learning with tensors: A framework based on convex optimization and spectral regularization," *Mach. Learn.*, vol. 94, no. 3, pp. 303–351, May 2013.

[111] R. Bro, "Multiway calibration. Multilinear PLS," *J. Chemom.*, vol. 10, no. 1, pp. 47–61, Jan. 1996.

[112] Q. Zhao, G. Zhou, T. Adalı, L. Zhang, and A. Cichocki, "Kernel-based tensor partial least squares for reconstruction of limb movements," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2013, pp. 3577–3581.

[113] Q. Zhao, L. Zhang, and A. Cichocki, "Multilinear and nonlinear generalizations of partial least squares: An overview of recent advances," *WIREs Data Mining Knowl. Disc.*, vol. 4, no. 2, pp. 104–115, Mar. 2014.

[114] R. I. Goldman, J. M. Stern, J. Engel, and M. S. Cohen, "Simultaneous EEG and fMRI of the alpha rhythm," *NeuroReport*, vol. 13, no. 18, pp. 2487–2492, Dec. 2002.

[115] D. Mantini, M. G. Perrucci, C. Del Gratta, G. L. Romani, and M. Corbetta, "Electrophysiological signatures of resting state networks in the human brain," *Proc. Nat. Acad. Sci.*, vol. 104, no. 32, pp. 13 170–13 175, 2007.

[116] P. A. Valdés-Hernández *et al.*, "White matter architecture rather than cortical surface area correlates with the EEG alpha rhythm," *NeuroImage*, vol. 49, no. 3, pp. 2328–2339, Feb. 2010.

[117] P. A. Valdés-Sosa *et al.*, "Frequency domain models of the EEG," *Brain Topography*, vol. 4, no. 4, pp. 309–319, 1992.

[118] R. D. Pascual-Marqui, P. A. Valdes-Sosa, and A. Alvarez-Amador, "A parametric model for multichannel EEG spectra," *Int. J. Neurosci.*, vol. 40, no. 1–2, pp. 89–99, 1988.

[119] W. Liu, J. Chan, J. Bailey, C. Leckie, and K. Ramamohanarao, "Mining labelled tensors by discovering both their common and discriminative subspaces," in *Proc. SIAM Conf. Data Mining*, 2013. [Online]. Available: http://dx.doi.org/10.1137/1.9781611972832.68

[120] M. J. Rosa, J. Daunizeau, and K. J. Friston, "EEG-fMRI integration: A critical review of biophysical modeling and data analysis approaches," *J. Integr. Neurosci.*, vol. 9, no. 4, pp. 453–476, Dec. 2010.

[121] A. Bayram *et al.*, "Simultaneous EEG/fMRI analysis of the resonance phenomena in steady-state visual evoked responses," *Clin. EEG Neurosci.*, vol. 42, no. 2, pp. 98–106, Apr. 2011.

[122] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, Sep. 1982.

[123] Y. K. Yılmaz, A. T. Cemgil, and U. Simsekli, "Generalised coupled tensor factorisation," in *Advances in Neural Information Processing Systems 24*.   Cambridge, MA, USA: MIT Press, 2011, pp. 2151–2159.

[124] L. Sorber, M. Van Barel, and L. De Lathauwer, "Structured data fusion," *IEEE J. Sel. Top. Signal Process.*, vol. 9, no. 4, pp. 586–600, Jun. 2015.

[125] A. R. Groves, C. F. Beckmann, S. M. Smith, and M. W. Woolrich, "Linked independent component analysis for multimodal data fusion," *NeuroImage*, vol. 54, no. 3, pp. 2198–2217, Feb. 2011.

[126] E. E. Papalexakis, C. Faloutsos, T. M. Mitchell, and N. D. Sidiropoulos, "Turbo-SMT: Accelerating coupled sparse matrix-tensor factorizations by 200×," in *Proc. SIAM Int. Conf. Data Mining*, 2014, pp. 118–126, doi:10.1137/1.9781611973440.14.

[127] M. Miki, N. Kitaoka, C. Miyajima, T. Nishino, and K. Takeda, "Improvement of multimodal gesture and speech recognition performance using time intervals between gestures and accompanying speech," *EURASIP J. Audio Speech Music Process.*, vol. 2014, no. 1, 2014, doi:10.1186/1687-4722-2014-2.

[128] S. Li and B. Yang, "Hybrid multiresolution method for multisensor multimodal image fusion," *IEEE Sens. J.*, vol. 10, no. 9, pp. 1519–1526, Sep. 2010.

[129] Y. Yoshitomi, S. I. Kim, T. Kawano, and T. Kitazoe, "Effect of sensor fusion for recognition of emotional states using voice, face image and thermal image of face," in *Proc. IEEE Int. Workshop Robot Human Interactive Commun.*, 2000, pp. 178–183.

[130] K. Van Deun, T. F. Wilderjans, R. A. van den Berg, A. Antoniadis, and I. Van Mechelen, "A flexible framework for sparse simultaneous component based data integration," *BMC Bioinf.*, vol. 12, no. 1, p. 448, Jan. 2011.

[131] G. Deco, V. K. Jirsa, P. A. Robinson, M. Breakspear, and K. Friston, "The dynamic brain: From spiking neurons to neural masses and cortical fields," *PLoS Comput. Biol.*, vol. 4, no. 8, Aug. 2008, Art. ID. e1000092.

[132] B. Yeşilyurt, K. Uğurbil, and K. Uludağ, "Dynamics and nonlinearities of the BOLD response at very short stimulus durations," *Magn. Resoning Imaging*, vol. 26, no. 7, pp. 853–862, Sep. 2008.

[133] K. J. Friston, N. Trujillo-Barreto, and J. Daunizeau, "DEM: A variational treatment of dynamic systems," *NeuroImage*, vol. 41, no. 3, pp. 849–885, Jul. 2008.

[134] R. Biscay, J. C. Jimenez, J. J. Riera, and P. A. Valdes, "Local linearization method for the numerical solution of stochastic differential equations," *Ann. Inst. Stat. Math.*, vol. 48, no. 4, pp. 631–644, Dec. 1996.

[135] W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*, vol. 42.   Berlin, Germany: Springer-Verlag, 2012.

[136] S. Dolgov, A. Litvinenko, and H. G. Matthies, "Polynomial chaos expansion of random coefficients and the solution of stochastic partial differential equations in the tensor train format," 2015, pp. 1–29. [Online]. Available: http://arxiv.org/abs/1503.03210

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

[138] E. Acar, D. M. Dunlavy, and T. G. Kolda, "A scalable optimization approach for fitting canonical tensor decompositions," *J. Chemom.*, vol. 25, no. 2, pp. 67–86, Feb. 2011.

[139] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 2, pp. 295–327, Jan. 2004.

[140] Y. K. Yılmaz and A. T. Cemgil, "Algorithms for probabilistic latent tensor factorization," *Signal Process.*, vol. 92, no. 8, pp. 1853–1863, Aug. 2012.

[141] M. Schmidt and S. Mohamed, "Probabilistic non-negative tensor factorisation using Markov chain Monte Carlo," in *Proc. 17th Eur. Signal Process. Conf.*, 2009, pp. 1918–1922.

[142] A. Cichocki, R. Zdunek, and S. Amari, "Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization," in *Proc. Int. Conf. Independent Compon. Anal.*, 2007, pp. 169–176.

[143] K. Kimura, Y. Tanaka, and M. Kudo, "A fast hierarchical alternating least squares algorithm for orthogonal nonnegative matrix factorization," in *Proc. 6th Asian Conf. Mach. Learn.*, 2014, pp. 129–141.

[144] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[145] A. P. Liavas and N. D. Sidiropoulos, "Parallel algorithms for constrained tensor factorization via the alternating direction method of multipliers," Sep. 2014. [Online]. Available: http://arxiv.org/abs/1409.2383

[146] S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low-n-rank tensor recovery via convex optimization," *Inverse Probl.*, vol. 27, no. 2, Feb. 2011, Art. ID. 025010.

[147] A. Smilde, R. Bro, and P. Geladi, *Multi-Way Analysis With Applications in the Chemical Sciences*.   Chichester, U.K.: Wiley, 2004.

[148] C. Boutsidis and E. Gallopoulos, "SVD based initialization: A head start for nonnegative matrix factorization," *Pattern Recognit.*, vol. 41, no. 4, pp. 1350–1362, Apr. 2008.

[149] A. Uschmajew, "Local convergence of the alternating least squares algorithm for canonical tensor approximation," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 2, pp. 639–652, Jan. 2012.

[150] L. Wang, M. T. Chu, and B. Yu, "Orthogonal low rank tensor approximation: Alternating least squares method and its global convergence," *SIAM J. Matrix Anal. Appl.*, vol. 36, no. 1, pp. 1–19, Jan. 2015.

[151] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Stat.*, vol. 9, no. 6, pp. 1135–1151, 1981.

[152] M. Hebiri, "Regularization with the smooth-Lasso procedure," 2008. [Online]. Available: http://arxiv.org/abs/0803.0668

[153] H. Zou, T. Hastie, and R. J. Tibshirani, "On the 'degrees of freedom' of the Lasso," *Ann. Stat.*, vol. 35, no. 5, pp. 2173–2192, Oct. 2007.

[154] J. Ye, "On measuring and correcting the effects of data mining and model selection," *J. Amer. Stat. Assoc.*, vol. 93, no. 441, pp. 120–131, 1998.

[155] J. Håstad, "Tensor rank is NP-complete," *J. Algorithms*, vol. 11, no. 4, pp. 644–654, 1990.

[156] V. de Silva and L. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1084–1127, Jan. 2008.

[157] R. Bro and H. A. L. Kiers, "A new efficient method for determining the number of components in PARAFAC models," *J. Chemom.*, vol. 17, no. 5, pp. 274–286, Jun. 2003.

[158] M. Mørup and L. K. Hansen, "Automatic relevance determination for multi-way models," *J. Chemom.*, vol. 23, no. 7–8, pp. 352–363, Jul. 2009.

[159] B. W. Bader and T. G. Kolda, "MATLAB tensor toolbox," 2015.

[160] C. A. Andersson and R. Bro, "The N-Way Toolbox for MATLAB," *Chemom. Intell. Lab. Syst.*, vol. 52, no. 1, pp. 1–4, Aug. 2000.

[161] S. Gourvénec *et al.*, "CuBatch," 2005.

[162] L. Sorber, M. Van Barel, and L. De Lathauwer, "Tensorlab," 2014.

[163] G. Zhou and A. Cichocki, "TDALAB: Tensor decomposition laboratory," 2013.

[164] A. H. Phan, P. Tichavsk, and A. Cichocki, "Tensor-Box: A Matlab package for tensor decomposition," 2013.

[165] E. Acar, "The MATLAB CMTF Toolbox," 2014.

[166] M. Mørup, "ERPWAVELAB," 2006.

[167] I. V. Oseledets, "TT Toolbox," 2009.

[168] I. V. Oseledets, T. Saluev, D. V. Savostyanov, and S. V. Dolgov, "ttpy," 2014.

[169] D. Bigoni, "Tensor toolbox," 2015.

[170] C. Tobler and D. Kressner, "Hierarchical Tucker toolbox," 2013.

[171] H. Zhou, "TensorReg toolbox for Matlab," 2013.

[172] R. Plonsey and D. B. Heppner, "Considerations of quasi-stationarity in electrophysiological systems," *Bull. Math. Biophys.*, vol. 29, pp. 657–664, 1967.

ABOUT THE AUTHORS

Esin Karahan received the B.S. degree in electrical and electronics engineering and the M.S. degree in biomedical engineering from Bogazici University, Istanbul, Turkey, in 2005 and 2007, respectively. Currently, she is working toward the Ph.D. degree at the Institute of Biomedical Engineering, Bogazici University.

Currently, she is a Research Assistant at the Institute of Biomedical Engineering, Bogazici University. Her research interests include developing multiway analysis methods for neuroimaging data, statistical signal processing, source localization, and brain connectivity analysis.

Pedro A. Rojas-López received the B.S. degree in computer science from the University of Havana, Havana City, Cuba, in 2013.
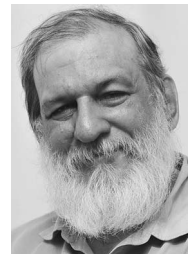
He participated in software production for improving image quality at the Integrated Technological Research Complex, Havana, Cuba, in 2012. Since September 2013, he has been working at the Neuroscience Center of Cuba, Cubanacan, Playa, Cuba, where his research interests include EEG source localization, brain connectivity analysis, and statistical signal processing. His primary recent work has been in Granger causality and EEG/fMRI fusion using tensor methods.

Maria L. Bringas-Vega received the B.S. degree in psychology from the University of Havana, Havana City, Cuba, in 1974 and the Ph.D. in psychophysiology from the National Center for Scientific Research of Cuba, La Habana, Cuba, in 1991.

She was a Senior Researcher and Head of the Neuropsychology Department, International Center for Neurological Restoration Havana, Havana, Cuba, from 1995 to 2014. She is a Senior Professor of Neuropsychology at the Cuban Neuroscience Center and currently a Professor of Neuroinformatics at the University of Electronic Sciences and Technology of China, Chengdu, China, where she has been working in application of neuroimaging and neuroinformatics methods to neurorestoration.

Pedro A. Valdés-Hernández received the B.S degree in nuclear physics from the University of Havana, Havana City, Cuba, in 2002. Currently, he is working toward the Ph.D. degree at the Cuban Neuroscience Center, Havana, Cuba.

He is now in charge of the Magnetic Resonance Unit and is the Head of the Neuroimaging Group. His research interests are in EEG forward and inverse modeling, functional magnetic resonance imaging, biophysical neural modeling, preclinical neuroimaging, etc.

Pedro A. Valdes-Sosa received the M.D. degree and the B.S. degree in mathematics from the University of Havana, Havana City, Cuba, in 1972 and 1973, respectively, and the Ph.D. degree in neuroscience from the National Center for Scientific Research of Cuba, La Habana, Cuba, in 1974.

He is the cofounder and has been the General Vice-Director for Research of the Cuban Neuroscience Center since 1990. Currently, he is also a Distinguished Professor of Neuroinformatics at the University of Science and Technology of China, Chengdu, China. He was named Senior Researcher in 1980 at the National Center for Scientific Research of Cuba. He is also a Second Degree Medical Specialist in Neurophysiology (1981), Senior Professor of the Higher Institute of Medical Sciences of Havana (2005), and was awarded the Doctor in Science (higher scientific degree) in 2011. He has published more than 240 papers.

Dr. Valdes-Sosa has been a member of the Cuban Academy of Sciences since 2000. He is on the Editorial Board of *PLOS One*, *Brain Topography*, and *Frontiers in Brain Imaging Methods*.