

Received April 16, 2020, accepted May 1, 2020, date of publication May 12, 2020, date of current version May 26, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2994276

Information Assisted Dictionary Learning for fMRI Data Analysis

MANUEL MORANTE^{1,2}, YANNIS KOPSINIS³, (Member, IEEE),
SERGIOS THEODORIDIS^{1,4}, (Life Fellow, IEEE), AND ATHANASSIOS PROTOPAPAS⁵

¹Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, 157 84 Athens, Greece

²Computer Technology Institutes and Press Diophantus (CTI), 265 04 Patras, Greece

³Libra MLI Ltd., Edinburgh EH6 6JG, U.K.

⁴Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong, Shenzhen 518172, China

⁵Department of Special Needs Education, University of Oslo, 0318 Oslo, Norway

Corresponding author: Manuel Morante (morante3@gmail.com)

This work was supported by the European Union's Seventh Framework Programme (H2020-MSCA-ITN-2014), through the Marie Skłodowska-Curie Fellowship within the Innovative Training Network (MacSeNet), under Grant agreement no. 642685.

ABSTRACT In this paper, the task-related fMRI problem is treated in its matrix factorization form, focusing on the Dictionary Learning (DL) approach. The proposed method allows the incorporation of a priori knowledge that is associated with both the experimental design and available brain atlases. Moreover, it can cope efficiently with uncertainties in the modeling of the hemodynamic response function. In addition, the method bypasses one of the major drawbacks of the DL methods; namely, the selection of the sparsity-related regularization parameters. Under the proposed formulation, the associated regularization parameters bear a direct relation to the number of the activated voxels for each one of the sources' spatial maps. This natural interpretation facilitates fine-tuning of the related parameters and allows for exploiting external information from brain atlases. The proposed method is evaluated against several other popular techniques, including the classical General Linear Model (GLM). The obtained performance gains are quantitatively demonstrated via a novel realistic synthetic fMRI dataset as well as real data from a challenging experimental design.

INDEX TERMS Dictionary learning, fMRI, semi-blind, sparsity, weighted norms.

I. INTRODUCTION

To perform actions/tasks, the brain relies on the simultaneous activation of many *Functional Brain Networks* (FBN), which are engaged in appropriate interaction to effectively execute the tasks. Such networks, potentially distributed over the whole brain, are defined as *segregated* regions exhibiting high functional connectivity. Connectivity is quantified via the underlying correlations among the associated activation/deactivation time patterns, referred to as *time courses* [1]. Functional Magnetic Resonance Imaging (fMRI) is the dominant data acquisition technique for the detection and study of FBNs [2]. fMRI measures the Blood Oxygenation Level-Dependent (BOLD) contrast [3], which tracks the evoked hemodynamic response of the brain to the corresponding neuronal activity. This process can be modeled as a convolution between the actual neuronal activation and a *person-dependent* impulse response function, called the

Hemodynamic Response Function (HRF). The fMRI captures 3D images with a typical resolution of $64 \times 64 \times 32$ voxels per image, acquired in sequences of 200 to 300 images per session (usually, one image every one or two seconds). The obtained fMRI measurements associated with each voxel comprise a mixture of the time courses corresponding to those FBNs that are active at the specific voxel. Moreover, beyond the brain-induced sources, additional machine-induced interfering sources are also present that contribute to the measured mixture. The ultimate goal of fMRI analysis is to detect the set of activated voxels, referred to as *spatial map*, in which each brain source of interest manifests itself in revealing the corresponding FBN.

In the case of block- or event-related experimental designs, i.e., when the subject is presented with a fixed set of conditions, the time courses associated with these experimental conditions are usually estimated as the convolution of the pre-defined stimuli of each condition with the canonical Hemodynamic Response Function (cHRF) [4]. Hereafter, such time courses are referred to as *task-related time courses*.

The associate editor coordinating the review of this manuscript and approving it for publication was Yunjie Yang.

A prominent approach for fMRI data analysis is via Blind Source Separation (BSS), which is usually performed with the aid of appropriate matrix factorization schemes [5]. In general, BSS methods aim to discover the different sources from the fMRI data without the necessity of any prior information regarding the experimental task. This feature makes BSS-based methods the dominant tool for the analysis of resting-state fMRI data, which lack any prior external task-related information. Independent Component Analysis (ICA) [6]–[8], and Dictionary Learning (DL) are the most popular paths in this direction. A drawback of ICA is the underlying independence assumption, which can be violated in fMRI, especially in the presence of high spatial overlapping, [9]–[11]. Unlike ICA, DL relies on sparsity that is a reasonable assumption for the neuronal brain activity [12]–[15].

However, DL approaches are not without shortcomings. The tuning of the associated regularization parameters is not easy in practice, and it is performed via cross-validation techniques; however, in real experiments, this is not possible due to the lack of ground truth data. Therefore, the only way to fine-tune parameters is via visual inspection of the results, a process that requires the subjective judgment of the user, which can be inconsistent and susceptible to errors. This may hamper the adoption of DL approaches in practice.

Alternatively, another candidate family of BSS for fMRI data analysis is the Non-negative Matrix Factorization (NMF) approach [16]–[18]. Unlike the aforementioned approaches, NMF methods impose a non-negativity constraint over the matrix factorization. However, the non-negativity constraint may not be valid in practice [19]. In the fMRI context, such a constraint would aim at eliminating the negative contribution of the BOLD signal response, leaving only the positive activations [16]; this is an undesired effect, taking in consideration the true nature of the hemodynamic response [3], [20]. Furthermore, NMF algorithms often require tuning of several regularization parameters, sharing the same limitations with standard DL techniques.

Conventional analysis of fMRI data relies on the General Linear Model (GLM), which assumes the prior availability of the task-related time courses [21]. This approach suffers from a critical limitation: It assumes that the HRF is known and fixed, whereas in reality the HRF may vary across subjects [22], as well as among brain locations [20]. In contrast, BSS methods make no assumption regarding the HRF and can reveal other brain-induced sources beyond the task-related ones. For example, they inherently model interfering artifacts, such as scanner-induced artifacts, uncorrected head-motion residuals, or other unmodeled physiological signals that may obscure the brain activity of interest.

Despite their advantages, BSS methods share a major drawback compared to GLM: when two or more task-related sources manifest themselves in highly overlapping brain regions, ICA (to a larger extent) and DL (to a smaller extent) can fail to discriminate them [23]. From a neuroscience

perspective, the presence of overlaps between FBNs is frequent in most of the typical experimental designs of interest. More specifically, several research groups have reported that conventional task-related FBNs such as motor, language, emotion, or auditory, exhibit considerable overlap with each other [24]–[26].

In an attempt to overcome the aforementioned fundamental drawbacks of the BSS methods against GLM, alternative approaches have been proposed [27], [28]. In the ICA case, the most relevant is to impose task-related information. Collectively, such methods are referred to as constrained ICA [29]–[34]. Although these often lead to enhanced performance, compared with their fully blind counterparts [33], they suffer from a critical limitation: the embedded constraint, e.g., the imposed task-related time courses, must not violate the independence assumption. This requirement poses stringent constraints either on the total number of allowable time sequences, e.g., [35], or on the nature of the imposed time courses, which need to be independent of each other [30]–[37]. Both restrictions heavily limit the applicability of constrained ICA in fMRI, since the most common case is to have experimental designs that comprise more than two BOLD sequences.

Furthermore, in contrast to many unconstrained ICA algorithms, which require a reduced number of relatively easy to tune parameters, all constrained ICA algorithms require extensive regularization parameter fine-tuning [34], based on cross-validation. Even the most recent constrained ICA technique, referred to as CSTICA [34], involves three regularization parameters and, as it is pointed out by the authors, the algorithm needs further improvement to “enable these parameters not to be determined by the experiments”. Besides constrained ICA, there are also NMF algorithms that allow incorporation of external information [17], [18]; yet, they suffer from similar drawbacks that limit their applicability in practice.

Recently, a DL method called *Supervised Dictionary Learning* (SDL) [38] was introduced, which allows the incorporation of external information from the task-related time courses with a rationale similar to GLM. As a result, SDL is greatly aided in the case of highly overlapping spatial maps and attains performance similar to that of GLM. However, SDL inherits from GLM two primary drawbacks: a) it builds upon the cHRF, which is fixed and, inevitably, different from the true one, and b) it adopts a regularized formulation of the DL, which inherits the difficulties associated with the tuning of the corresponding regularization parameter. In Section IV, Table 2, we provide a thorough comparison among all competitive approaches and their characteristics of interest in the fMRI case.

In this paper, a novel DL formulation of the fMRI BSS problem, referred to as Information Assisted Dictionary Learning (IADL) is proposed, which, among other merits, alleviates the two aforementioned critical disadvantages of SDL as well as those of the constrained ICA approaches. More specifically:

- A new sparsity constraint is adopted, which bears a physical interpretation that naturally complies with the segregated nature of FBNs. Unlike standard approaches, the proposed sparsity constraint establishes a bridge between the optimization parameters and the expected number of activated voxels of each source.
- The proposed sparsity constraint also offers the flexibility of simultaneously dealing with sparse and dense sources. Indeed, in real fMRI, in addition to sparse sources, dense sources may appear, usually related to physiological or machine-induced artifacts.
- A new semi-blind DL approach is proposed that incorporates task-related information. In contrast to the standard approaches, where any task-related information is fully governed by the canonical HRF, our novel formulation incorporates this information in a *relaxed* way, allowing the imposed time course to adjust to the subject (or subjects) at hand. Thanks to this relaxation, we implicitly accommodate discrepancies between the HRF and the cHRF, and we cope with distortions and inaccuracies regarding the convolutional model, e.g., due to nonlinear effects. In case no prior task-related information is available (e.g., resting-state fMRI data), the proposed method still benefits from the newly adopted sparsity constraint.
- A new, highly realistic synthetic dataset, is constructed, that allows conducting a thorough performance evaluation of the new method against state-of-the-art ICA- and DL-based techniques.

Notation: A lower case letter, x , denotes a scalar, a bold capital letter, \mathbf{X} , denotes a matrix, and a bold lower case letter, \mathbf{x} , denotes a vector with its i^{th} component denoted as x_i . The i^{th} row and the i^{th} column of a matrix, $\mathbf{X} \in \mathbb{R}^{M \times N}$, are represented as $\mathbf{x}^i \in \mathbb{R}^{1 \times N}$ and $\mathbf{x}_i \in \mathbb{R}^{M \times 1}$, respectively. Moreover, x_{ij} denotes the element “located” at row i and column j of the matrix \mathbf{X} .

II. NOVEL DL CONSTRAINTS TAILORED TO TASK-RELATED fMRI

A. PRELIMINARIES ON DL-BASED FMRI ANALYSIS

The data collected during an fMRI experiment form a two-dimensional data matrix as follows: Each of the acquired 3D images is unfolded and stored into a vector, $\mathbf{x} = [x_1, x_2, \dots, x_N] \in \mathbb{R}^{1 \times N}$, where N is the total number of voxels per image. Such vectors correspond to the sequence of, say, T successively obtained images and they are stacked as rows to form the data matrix $\mathbf{X} \in \mathbb{R}^{T \times N}$. Note that, in practice, before the formation of the data matrix, \mathbf{X} , several standardized preprocessing steps are performed to account for many detrimental effects related to the fMRI image acquisition process, such as slice-timing correction, head motion, realignment, normalization, etc.

From a mathematical point of view, the source separation problem can be described as a matrix factorization task of the data matrix, i.e.,

$$\mathbf{X} \approx \mathbf{D}\mathbf{S}, \quad (1)$$

where, following the dictionary learning jargon, $\mathbf{D} \in \mathbb{R}^{T \times K}$ is the dictionary matrix, whose columns represent different time courses, $\mathbf{S} \in \mathbb{R}^{K \times N}$ is the coefficient matrix, whose rows are the spatial maps associated with the corresponding time courses, and K is the number of sources. In general, such a matrix factorization can be obtained via the solution of a constrained optimization task:

$$(\hat{\mathbf{D}}, \hat{\mathbf{S}}) = \underset{\mathbf{D}, \mathbf{S}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{D}\mathbf{S}\|_F^2 \quad \text{s.t.} \quad \begin{array}{l} \mathbf{D} \in \mathcal{D} \\ \mathbf{S} \in \mathcal{L} \end{array}, \quad (2)$$

where \mathcal{D} , \mathcal{L} are two sets of admissible matrices, defined by an adopted set of appropriately imposed constraints and $\|\cdot\|_F$ denotes the Frobenius norm of a matrix.

The concept of signal sparsity refers to discrete signals that involve a sufficiently large number of zero values. The typical way to quantify sparsity is via the ℓ_0 -norm: given an arbitrary vector, $\mathbf{x} \in \mathbb{R}^N$, the ℓ_0 -norm (which, strictly speaking, is not a norm in its strict mathematical definition [5]) is defined as the number of non-zero components of a vector. In standard DL methods, the sparsity constraints are usually implemented in two ways: either each column of \mathbf{S} is separately constrained to be sparse, e.g., $\|\mathbf{s}_i\|_0 \leq \gamma_i$, where γ_i is the maximum number of the non-zero values of the i^{th} column of \mathbf{S} , or the full coefficient matrix, \mathbf{S} , is constrained to involve, at most, $\hat{\gamma}$ non-zeros [38]–[40].

The proposed method introduces new constraints on the spatial maps (i.e., on each row of the coefficient matrix, \mathbf{S}) and time courses (i.e., on the dictionary, \mathbf{D}), the design of which serves the specific needs of the task-related fMRI data analysis. These proposed constraints are discussed next.

B. INFORMATION-BEARING SPARSITY CONSTRAINTS ON THE SPATIAL MAPS

In the fMRI framework, sparsity appears to be a natural assumption for the segregated nature of the spatial maps of the FBNs. In other words, each row, say \mathbf{s}^i , of the coefficient matrix, \mathbf{S} , should have non-zero values only at entries that correspond to voxels activated by the corresponding time course \mathbf{d}_i . The smaller the area that a specific FBN occupies in the brain, the sparser the corresponding row of the coefficient matrix should be. At this point it is worth recalling that none of the DL methods applied in fMRI so far impose sparsity row-wise. However, imposing sparsity column-wise assumes that only a few sources are active in each voxel. Although this assumption is generally true, this piece of information is voxel-dependent, and it is hard to accommodate a suitable regularization parameter that simultaneously works optimally for all of the thousands of columns. On the other hand, imposing sparsity in the full coefficient matrix may be easier to handle, but it is a general constraint that fails to exploit relevant information regarding the segregated nature of each FBN.

In this paper, to the best of our knowledge, *it is the first time that the DL framework is extended to allow sparsity promotion along rows of the coefficient matrix*. Looking at (2), sparsity in the rows of the coefficient matrix can be imposed

using the following admissible set of constraints:

$$\mathcal{L}_0 = \left\{ \mathbf{S} \in \mathbb{R}^{K \times N} \mid \|\mathbf{s}^i\|_0 \leq \phi_i, i = 1, 2, \dots, K \right\}, \quad (3)$$

where ϕ_i is a user-defined parameter, which denotes the maximum number of non-zero elements of the i^{th} row of \mathbf{S} . In the fMRI case, the parameter ϕ_i has a clear physical interpretation and it corresponds to the total number of voxels that are active due to the i^{th} source. This corresponds to the number of voxels that the corresponding FBN occupies, an estimate of which can be obtained from brain atlases.

It is well known that the ℓ_0 -norm constraint results in an NP-hard optimization, and it is usually relaxed to its closest convex relative, namely, the ℓ_1 -norm [41]. The corresponding constrained set then becomes:

$$\mathcal{L}_1 = \left\{ \mathbf{S} \in \mathbb{R}^{K \times N} \mid \|\mathbf{s}^i\|_1 \leq \lambda_i, i = 1, 2, \dots, K \right\}, \quad (4)$$

where λ_i are *new* user-defined parameters to implicitly control the sparsity of the rows of the coefficient matrix. In contrast to the ϕ_i parameters, the new parameters, λ_i , are not directly related to the sparsity level. This makes them hard to tune in practice, unless cross-validation is an option, which is not the case in fMRI.

An additional novelty of IADL that allows us to overcome this obstacle is the application across rows of a *weighted version* of the ℓ_1 -norm. In particular, given an arbitrary vector $\mathbf{x} \in \mathbb{R}^N$, the weighted ℓ_1 -norm is defined as:

$$\|\mathbf{x}\|_{1, \mathbf{w}} = \sum_{i=1}^N w_i |x_i|, \quad (5)$$

where \mathbf{w} is a real positive vector given by

$$w_i = \frac{1}{|x_i| + \varepsilon} \quad i = 1, 2, \dots, N, \quad (6)$$

and $\varepsilon \in \mathbb{R}^+$ is a real positive number, which is introduced in order to avoid division by zero, providing enhanced numerical stability, [5], [42], [43], and can be set directly to a small value, e.g. 10^{-6} . Accordingly, the row-sparsity constraint now becomes:

$$\mathcal{L}_w = \left\{ \mathbf{S} \in \mathbb{R}^{K \times N} \mid \|\mathbf{s}^i\|_{1, \mathbf{w}^i} \leq \phi_i \quad i = 1, 2, \dots, K \right\}, \quad (7)$$

where \mathbf{w}^i is the vector of weights that correspond to the vector \mathbf{s}^i computed as in (6). The key point is that, similar to the ℓ_0 case, the constraint is imposed via the sparsity level ϕ_i and, consequently, ϕ_i can now be used as an upper bound. This is theoretically substantiated, since it has been shown that the weighted ℓ_1 -norm is bounded by the corresponding ℓ_0 -norm [43], as discussed in Section VI of the supplementary material. Moreover, as shown in Section VI-C of the supplementary material, for a given \mathbf{w} , \mathcal{L}_w remains *convex*.

Hereafter, an equivalent but conceptually easier to handle sparsity-related measure that is independent of the length of the vector, known as sparsity percentage, will be used interchangeably with sparsity level. Sparsity percentage expresses

the proportion of zeros within a vector, \mathbf{x} , and it is given by

$$\theta = \left(1 - \frac{\phi}{N} \right) \times 100, \quad (8)$$

where $\phi = \|\mathbf{x}\|_0$, and N is the total number of elements.

C. TASK-RELATED DICTIONARY SOFT CONSTRAINTS

The enhanced discriminative power of the GLM over BSS methods comes from the fact that, in the GLM, the task-related time courses are explicitly provided to GLM modeling via the estimated BOLD sequences [2]. Such information is left unexploited in the BSS framework. Hence, it seems reasonable to also incorporate this information into the BSS methods, leading naturally to a semi-blind formulation.

As stated in the introduction, in contrast to ICA techniques, DL-based methods can easily incorporate, in principle, any constraint in the time courses, since sparsity is not affected. This fact has been exploited in SDL through splitting the dictionary into two parts:

$$\mathbf{D} = [\Delta, \mathbf{D}_F] \in \mathbb{R}^{T \times K}, \quad (9)$$

where the first part, $\Delta \in \mathbb{R}^{T \times M}$, comprises fixed columns, which are set equal to the imposed task-related time courses, $\delta_1, \delta_2, \dots, \delta_M$. The second part, $\mathbf{D}_F \in \mathbb{R}^{T \times (K-M)}$, is left to vary, and is learned during the DL optimization phase. Nevertheless, SDL inherits the same drawback associated with the GLM: the constrained atoms of the fixed dictionary (columns of the matrix Δ) lead to improvement only if the imposed task-related time courses are sufficiently accurate. Otherwise, the task-related time courses will be mis-modeled and their contribution can introduce detrimental effects, leading to inaccurate results.

In this paper, we relax the strong *equality* requirement of SDL to a looser *similarity*-based distance-measuring norm constraint. Then, if part of the a priori information is inaccurate, e.g., the assumed HRF differs from the true one, the method can efficiently adjust the constrained atoms since they are not forced to remain fixed and equal to the pre-selected time courses. Moreover, the proposed modeling also accounts for multiple factors that potentially alter the functional shape of the task-related time courses across subjects and brain regions, such as vascular differences, partial volume imaging, brain activations, [22], hematocrit concentrations [44], lipid ingestion [45], and even nonlinear effects due to short interstimulus intervals [46].

Mathematically, the starting point is to split the dictionary into two parts:

$$\mathbf{D} = [\mathbf{D}_C, \mathbf{D}_F] \in \mathbb{R}^{T \times K}, \quad (10)$$

where, in contrast to SDL, the part $\mathbf{D}_C \in \mathbb{R}^{T \times M}$ has columns which are constrained to be *similar*—rather than equal—to the imposed task-related time courses. Then, we can define a new *convex* set of admissible dictionaries:

$$\mathcal{D}_\delta = \left\{ \mathbf{D} \in \mathbb{R}^{T \times K} \mid \begin{array}{l} \|\mathbf{d}_i - \delta_i\|_2^2 \leq c_\delta \quad i=1, \dots, M \\ \|\mathbf{d}_i\|_2^2 \leq c_d \quad i=M+1, \dots, K \end{array} \right\}, \quad (11)$$

where $\|\cdot\|_2$ denotes the Euclidean norm, \mathbf{d}_i is the i^{th} column of the dictionary \mathbf{D} and δ_i is the i^{th} a priori selected task-related time course. The constant c_δ is a user-defined parameter, which controls the *degree of similarity* between the constrained atoms and the imposed time courses; essentially, this reflects our confidence on how accurate the cHRF is for the subject under consideration. In particular, as it is further explained in Section I-C of the supplementary material, c_δ accounts for the natural variability the HRFs are expected to have among subjects, giving rise to consistent strategies for its tuning. Moreover, the free atoms have a bounded norm controlled by c_d , another user-defined parameter to avoid ill-conditioned phenomena. This can be fixed to 1 [47].

III. THE IADL ALGORITHM

In this section, we present an implementation of IADL that solves (2), incorporating the two proposed sets of constraints, namely, $\mathcal{D} = \mathcal{D}_\delta$ and $\mathcal{L} = \mathcal{L}_w$.

The simultaneous minimization for \mathbf{D} and \mathbf{S} is challenging due to both the non-convexity of the task in (2) and the potentially large size of the data matrix. The size of the latter is restrictive for optimization frameworks, which are computationally demanding. Therefore, we adopt the *Block Majorized Minimization* (BMM) rationale, which provides a robust framework for the solution of this type of optimization tasks, e.g., [48]. In this way, the BMM scheme simplifies the optimization task by adopting a two-step alternating minimization, under certain assumptions [49]. For simplicity, we algorithmically describe the proposed DL approach next, and the full mathematical derivation and convergence proof are both analytically provided in detail in Section VI of the supplementary material.

Put succinctly, the proposed DL algorithm follows the standard scheme of classical DL methods, which iteratively alternate between a sparse coding step and a dictionary update step. Concerning the sparse coding step, the corresponding recovery mechanism is a soft thresholding operator similar to the one that corresponds to the standard ℓ_1 -norm constraint, as explained in [43]. In this study, we used an efficient implementation of the algorithm of the weighted ℓ_1 -norm projector based on [50].

In Algorithm 1, we present the pseudo-code for solving (2), given the number of sources, K , an arbitrary set of estimates, $\mathbf{S}^{[0]}, \mathbf{D}^{[0]}$, M estimates of the task-related time courses, $\delta_1, \delta_2, \dots, \delta_M$, grouped by columns in the matrix Δ , the sparsity for each row, $\phi = [\phi_1, \phi_2, \dots, \phi_K]^T$, and the number of iterations, $Iter$. Observe that the free parameters of the algorithm that need to be tuned are: a) the number of sources, K , b) the maximum sparsity per row ϕ_i , c) the radius parameter c_δ and d) the parameters c_d and ε involved in equations (11) and (6), respectively. The last two can be directly fixed to 1 and a small value, say 10^{-6} , respectively, and their choice is not critical. The rest of the parameters can be straightforwardly tuned based on physical arguments, which can easily

Algorithm 1 Information Assisted DL

```

input :  $K, \mathbf{S}^{[0]}, \mathbf{D}^{[0]}, \Delta, \phi, Iter$ 
1  $\mathbf{B} = \mathbf{D}^{[0]}$ ;
2 for  $t = 0$  to  $Iter$  do
3    $c_S$  equal to (or larger than)  $\|\mathbf{B}^T \mathbf{B}\|$ ;
4    $\mathbf{A} = \frac{1}{c_S} [\mathbf{B}^T \mathbf{X} + (c_S \mathbf{I}_K - \mathbf{B}^T \mathbf{B}) \mathbf{S}^{[t]}]$ ;
5    $\mathbf{W} \leftarrow w_{ij} = \frac{1}{|a_{ij}| + \varepsilon}$  with  $(\varepsilon > 0)$ ;
6   for  $i = 1$  to  $K$  do
7     if  $\|\mathbf{a}^i\|_{1, \mathbf{w}^i} > \phi_i$  then
8        $\mathbf{a}^i = \mathcal{P}_{B_{\ell_1[\mathbf{w}^i, \phi_i]}}(\mathbf{a}^i)$ ;
9     end
10  end
11   $\mathbf{S}^{[t+1]} = \mathbf{A}$ ;

12   $c_D$  equal to (or larger than)  $\|\mathbf{A} \mathbf{A}^T\|$ ;
13   $\mathbf{B} = \frac{1}{c_D} [\mathbf{X} \mathbf{A}^T + \mathbf{D}^{[t]} (c_D \mathbf{I}_M - \mathbf{A} \mathbf{A}^T)]$ ;
14  for  $i = 1$  to  $M$  do
15    if  $\|\mathbf{b}_i - \delta_i\|^2 > c_\delta$  then
16      end
17       $\mathbf{b}_i = \frac{c_\delta^{\frac{1}{2}} (\mathbf{b}_i - \delta_i)}{\|\mathbf{b}_i - \delta_i\|} + \delta_i$ ;
18    end
19    for  $i = M + 1$  to  $K$  do
20      if  $\|\mathbf{b}_i\|^2 > c_d$  then
21         $\mathbf{b}_i = \frac{c_d^{\frac{1}{2}}}{\|\mathbf{b}_i\|} \mathbf{b}_i$ ;
22      end
23    end
24     $\mathbf{D}^{[t+1]} = \mathbf{B}$ ;
25 end
output:  $\mathbf{D} = \mathbf{D}^{[t+1]}, \mathbf{S} = \mathbf{S}^{[t+1]}$ 
26  $\mathcal{P}_{B_{\ell_1[\mathbf{w}^i, \phi_i]}}$  is the projection operator over the weighted  $\ell_1$ -norm ball,
 $B_{\ell_1[\mathbf{w}^i, \phi_i]} = \{\mathbf{x} \in \mathbb{R}^N \mid \|\mathbf{x}\|_{1, \mathbf{w}^i} \leq \phi_i\}$ , of weights  $\mathbf{w}^i$  and radius  $\phi_i$ .
This projection operator onto the weighted  $\ell_1$ -norm ball is derived in
closed form in [43].

```

be drawn from the fMRI study at hand. The algorithm is insensitive to the overestimation of K , which renders this parameter easily tunable (see discussion in Section I.A and Section V.C of the supplementary material). Also, the maximum sparsity per row is readily obtained from published brain atlases (see Section I.B and Section II in the supplementary material) and c_δ is easily tuned by considering the task-related time courses (see Section I.C in the supplementary material); the latter results from established HRF models and their expected variability. Furthermore, we provide a Matlab implementation for IADL-based analysis, which is freely available in the IADL's capsule¹ within the CodeOcean platform. It comprises automatic initialization and parameter self-tuning for c_δ . Accordingly, its default parameter setup

¹IADL's capsule: <https://codeocean.com/capsule/5510515/tree>

TABLE 1. Sparsity parameter θ used for IADL.

(a) Sparsity (%)			(b) True Sparsity		(c)
θ_1	θ_2	θ_3	N ^o	Sparsity (%)	Sp. (%) θ_r
95	95	85	1	95.28	90
90	90	85	11	91.60	95
94	90	85	14	94.57	95
95	95	95	6	97.04	95
94	90	90	3	95.53	95
93	90	90	2	95.33	90
92	90	90	13	94.51	95
91	90	90	5	93.30	95
90	90	90	10	92.67	90
89	85	85	8	91.82	90
88	85	85	12	91.53	90
87	85	85	4	88.25	85
86	85	85	7	88.07	80
85	85	85	19	86.14	70
80	80	80	9	85.51	50
80	80	80	15	71.95	30
75	75	75	20	71.84	20
75	75	75	18	1.99	10
70	70	70	16	1.00	0
60	70	70	17	1.00	0
10	10	10			
5	5	5			
0	0	0			
0	0	0			
0	0	0			
0	0	0			

Table (a) depicts the three different choices of sparsity θ_1 , θ_2 and θ_3 used in the experiments with the synthetic data. Table (b) depicts the sorted exact sparsity percentage values of the implemented realistic synthetic dataset. Note that the first three values in Tables (a) and (b) correspond to the assisted sources of interest. Table (c) shows the values used for the real fMRI analysis.

can be adopted out-of-the-box with any task-related fMRI dataset. At the same time, a more thoughtful specification of the parameters, e.g., incorporating the expected sparsity level of the task-related FBNs at hand, can further improve results.

IV. PERFORMANCE COMPARISON

A. PERFORMANCE RESULTS BASED ON SYNTHETIC DATA

In Section III of the supplementary material, a novel synthetic data set is presented. This highly realistic dataset emulates demanding experimental tasks, where some of the spatial maps substantially overlap each other. Therefore, this synthetic dataset allows us to effectively evaluate the performance of the proposed DL method, in comparison with the state-of-the-art of blind and semi-blind approaches, under more realistic settings.

The adopted performance measure, r , is associated with Pearson's correlation coefficient among the estimated and the true sources, and is described in detail in Section IV.A of the supplementary material.

The aim of this performance study is twofold: First, to study the effectiveness of the proposed approach in dealing with HRF mis-modeling; and second, to evaluate the decomposition performance of the algorithm with respect to the set of sources of interest. As benchmarks, the following

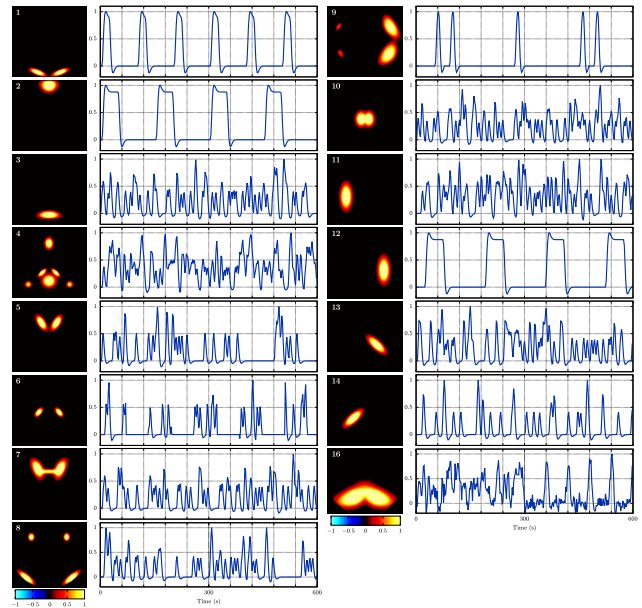


FIGURE 1. Visual representation of the synthetic spatial maps and their corresponding time courses generated with the canonical HRF. The intensity of the sources is normalized to facilitate visual inspection.

competitive algorithms are considered: a) McICA,² which is a constrained ICA algorithm [35] that allows assisting a source using task-related time courses, b) SDL, c) an Online DL algorithm (ODL) [51], which is included in SPAMS³ toolbox, and d) three ICA algorithms, namely, Infomax⁴ [52], a widely used ICA algorithm within the fMRI community, JADE⁵ [53], which we used as an initialization point for all DL algorithms, and CanICA⁶ [54], a state-of-the-art ICA-based algorithm for fMRI data analysis.

To emulate HRF variability, we generated six different “subjects”, through six different, yet realistic, synthetic HRFs. The selected HRFs are depicted in Fig. 2. Six different datasets were built, one for each subject, with the only difference among them being the HRF used to generate the brain-induced time courses. Sources 1, 11 and 14 (see Fig. 1)

²Morphologically constrained ICA is a constrained ICA algorithm that subtracts one particular signal of interest using given a priori information. In this paper we utilized the open implementation from <http://dsp.ucsd.edu/~zhilin/Software.html>

³The SPAMS (SPArse Modeling Software) is an open-source optimization toolbox for solving various sparse estimation problems. <http://spams-devel.gforge.inria.fr>

⁴Infomax is an algorithm for ICA relying on maximum likelihood. In this paper, we used an open-source implementation of this algorithm from the DTU Toolbox: <http://cogsys.imm.dtu.dk/toolbox/menu.html>

⁵The Joint Approximate Diagonalization of Eigenmatrices (JADE) is an algorithm for ICA that separates observed mixed signals into latent source signals by exploiting fourth-order moments. In this paper, we used the open-source implementation from J.F. Cardoso: <http://mikexcohen.com/lecturelets/eigen/jader.m>

⁶Canonical ICA (CanICA) is an ICA-based algorithm specifically designed for fMRI group analysis by identifying the common subspace to all the subjects containing the main activation patterns [54]. In this study, we utilized the CanICA available from the package Machine learning for Neuro-Imaging in Python (nilearn) <https://nilearn.github.io/>

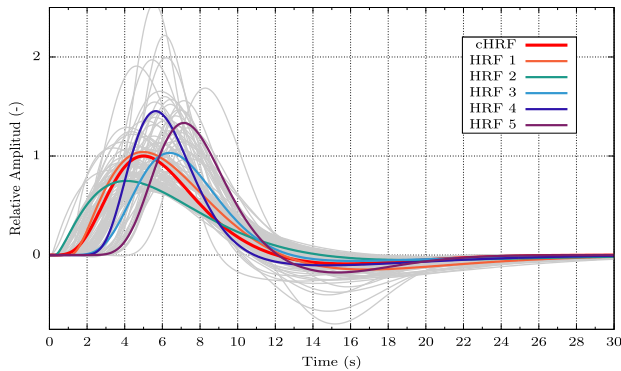


FIGURE 2. Graphic representation of 100 HRFs (gray) randomly generated from the two gamma distributions model. The red curve represents the canonical HRF (cHRF) and the rest of the colored HRFs stands for the five selected alternatives.

were chosen to be the task-related time courses, since they correspond to realistic scenarios that are often encountered in practice: Source 1 is easy to identify, since it barely spatially overlaps with other sources and corresponds to a block-event experimental design. Sources 11 and 14 are more challenging and exhibit notable overlap, emulating an event-related task (intervals shorter than 5 seconds, see Fig. 1). Consequently, we generated the imposed task-related time courses δ_1 , δ_2 and δ_3 , convolving the experimental conditions related to these sources with the cHRF, according to the standard procedure that is followed in GLM/SPM-based analysis.

Concerning the parametrization of the algorithms, K was overestimated by 20%, i.e., $K = 25$ rather than 20. This is typical in realistic scenarios since it is not possible to know the exact number of sources. All benchmark methods require an estimate of the number of sources. Thus, the same value was provided to all the algorithms. Moreover, as discussed in Section III, IADL requires to set up three extra quantities: a) the maximum sparsity for each task-related sources, b) the set of maximum sparsity values the rest of the sources can take and c) the parameter c_δ .

For (a), the imposed sparsity percentages for the task-related source 1, 11 and 14 are 95%, 90% and 94% respectively, which correspond to a slight overestimation compared to the true sparsity values. Sparsity set-up information related to this experiment is also listed in Table 1. In particular, the true sparsities of the task-related time courses are shown in Table 1.b and the corresponding imposed sparsities are shown in the corresponding rows of the first column of Table 1.a (sparsity set up scheme θ_1).

For (b), the exact used sparsity values are shown in the rest of the rows of θ_1 . To better grasp the relationships among imposed values and the true sparsity of non-task-related sources, both the true and the imposed ones are sorted with decreasing sparsity percentage. In Section IV-B, it will be shown that the proposed scheme is largely insensitive to the provided sparsity values.

Finally, for (c), we set $c_\delta = 0.2$, which is large enough to accommodate the variations among HRFs in the different synthetic datasets. This value has been calculated following the algorithmic procedure described in the Supplementary Section I-C.

SDL and ODL tune the sparsity constraint via a single regularization parameter, λ . In contrast to IADL, λ is not directly related to a certain sparsity level, and its optimum value is case-specific. Moreover, it does not bear a concise physical interpretation that could serve as a guide for its tuning. Also, the range of values of the optimum λ can vary significantly from case to case. It can take a very small value, e.g., 0.01 or a relatively large value, e.g., 10. In the synthetic dataset case, since the ground truth, i.e., the correct decomposition, is fully known, λ can be optimized through cross-validation. However, such a luxury is never available in practice, where real fMRI datasets need to be analyzed. In this study, we evaluate SDL using λ values that lead to the best performance according to specific criteria: The first criterion was to optimize SDL in terms of the mean performance of the time courses of all the sources, leading to a $\lambda_1 = 0.02$. The second criterion was to optimize with respect to the mean value of the full source performance (see Eq. S31 in the supplementary material) for the assisted sources only, leading to a $\lambda_2 = 2.8$. In both cases, we conducted the λ value optimization for the subject that corresponds to the cHRF. Observe the large gap between the obtained λ values, which is indicative of the sensitivity of the SDL approach to λ tuning.

On the other hand, the optimal value for the fully blind ODL algorithm was found to be $\lambda = 1.6$. SDL and ODL are initialized from ICA, similarly to IADL. Such an initialization leads to better performance compared to the original version presented in [38]. These observations agree with the results reported in [55], where the authors use ICA as an initialization point for their proposed DL algorithm to enhance its performance.

McICA requires fine-tuning a set of 4 regularization parameters. We observed that optimal selection of these parameters heavily depends on the particular synthetic subject, similarly to the SDL algorithm. Accordingly, we manually optimized these parameters via cross-validation aiming to achieve the best average performance over all the subjects.

Fig. 3.A and Fig. 3.B show the performance results with respect to the full source and the time course, respectively. The horizontal axis indicates the six synthetic subjects that correspond to different HRFs. Both figures comprise three inset graphs, each of which depicts performance with respect to three different sets of sources: (a) the task-related sources 1, 11 and 14, (b) the brain-like sources (1, 2, ..., 15), and (c) the whole dataset (1, 2, ..., 20) which includes both brain-like sources and artifacts. For completeness, the Supplementary Fig. S4 shows the individual performance of the studied methods for each assisted source and depicts some of the obtained time courses and their corresponding spatial maps.

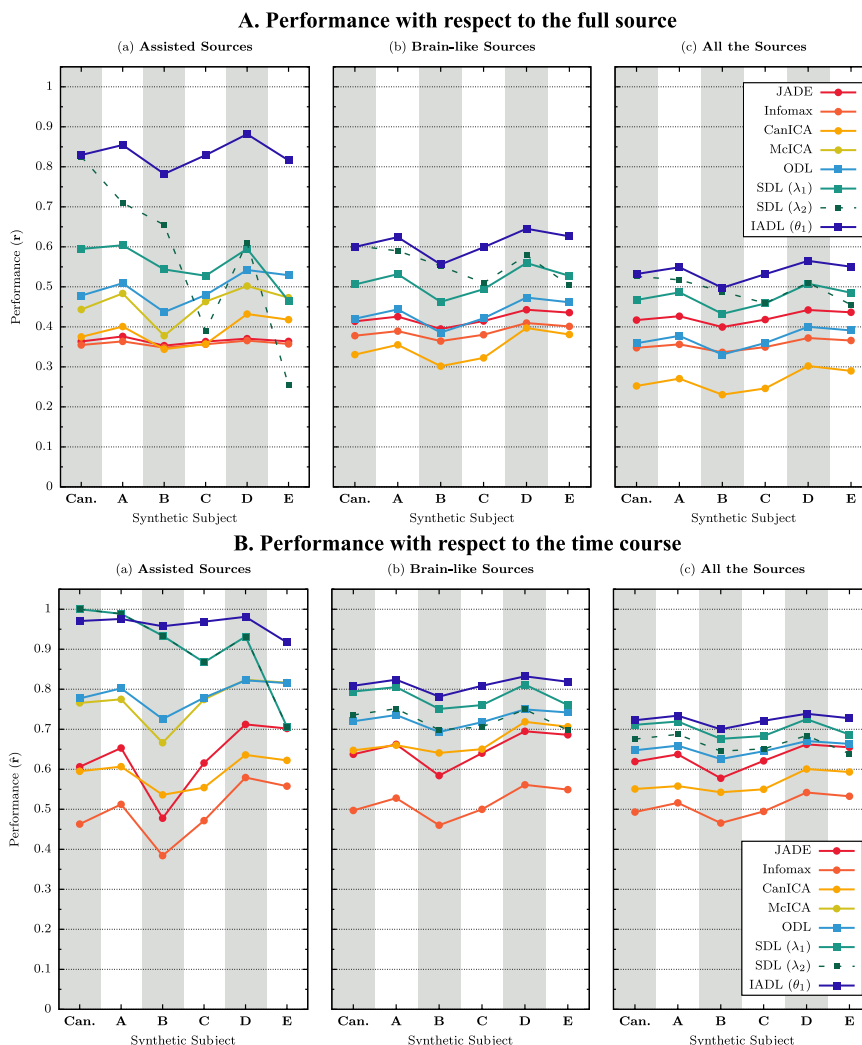


FIGURE 3. Performance comparison of different approaches with respect to the full source (A) and with respect the time courses only (B). The inset figures correspond to (a) the sources of interest [1, 11, 14], (b) the brain-like sources only, and (c) all sources (including artifacts).

Let us first focus on the two information-assisted DL algorithms, SDL and IADL, whose performance is indicated with green and dark blue curves, respectively. They are both assisted with the task-related time courses that correspond to the cHRF. In the SDL case, the solid and the dashed curves correspond to regularization parameter tuning equal to λ_1 and λ_2 respectively. Recall that these two cases lead to the best performance for the canonical subject in Fig. 3.B.(a) and Fig. 3.A.(c), respectively. In the inset Fig. 3.A.(a) and Fig. 3.B.(a), it is already observed that IADL manages to cope well with subject variability even in cases where the discrepancy between the canonical and the subject HRF is large, e.g., subject E (see Fig. 2). On the contrary, in Fig. 3.A.(a), SDL copes well only in the canonical case, for which the algorithm has been explicitly optimized. In Fig. 3.B.(a), SDL attains superior performance only in the canonical case—where the exact task-related time courses have been used—as well as in subject A which, as can be seen in Fig. 2, has an HRF

similar to the canonical one. Focusing on SDL performance for subjects B to E, it is apparent that the strategy to fix the assisted time courses can lead to high deterioration according to both performance measures.

In the cases where only brain-like and all the sources are considered (mid and rightmost subfigures), the proposed approach still outperforms SDL. Note that the time courses estimated by IADL are overall better than those of SDL even in the case of the canonical subject, in which SDL is fully optimized exploiting ground truth knowledge. In comparison to the fully blind methods, i.e., ODL, JADE, and Infomax, task-related assisted methods perform better. Concerning the performance of the blind methods, we observed that ODL works better than ICA-based approaches for the optimal selected λ value. However, note that in practice it is very hard—if possible at all—to optimize the associated λ parameter as done here with the synthetic dataset.

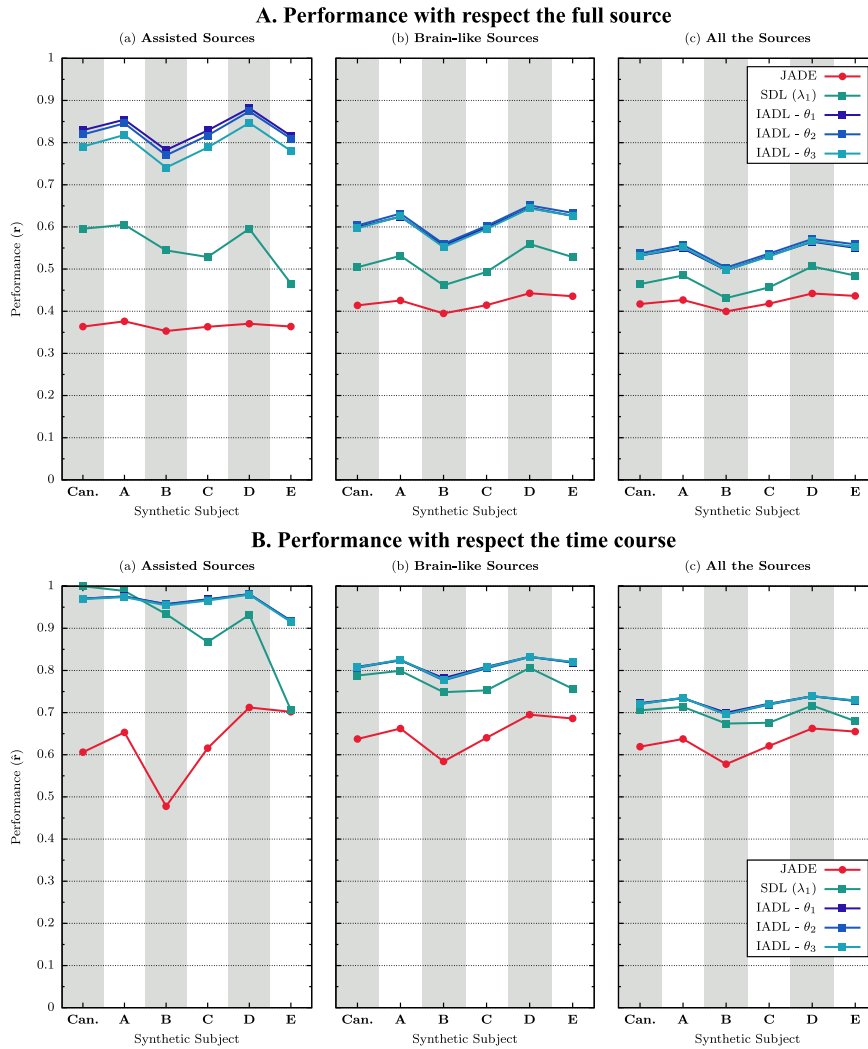


FIGURE 4. Performance evaluation of the IADL algorithm for different choices of sparsity. Fig. (A) shows performance with respect to the full sources and Fig. (B) with respect to the time courses only. The inset figures correspond to (a) the sources of interest [1, 11, 14], (b) the brain-like sources only, and (c) all sources (including artifacts) IADL performance using three different choices of sparsity. The figures also include the results from SDL and JADE from the Fig. 3 as a reference.

The yellow curves in Fig. 3 depict the performance of McICA. This particular constrained ICA algorithm provides estimates only of the assisted sources [35]; hence, results correspond only to the assisted brain-like sources. First, we observe that the McICA algorithm performs better than JADE, CanICA, and Infomax. On the other hand, our proposed IADL algorithm outperforms McICA for all the studied synthetic subjects.

Observe that CanICA (orange curves) exhibits performance that is similar to that of the other two ICA algorithms (see Fig. 3.B). This was expected because CanICA performs best in the multi-subject level [54], rather than in single-subject setups as the one we used so far, where it does not offer any particular advance. In section V, where we deal with multi-subject analysis, we confirm the superiority of CanICA over the other ICA methods examined here.

B. IADL ROBUSTNESS AGAINST SPARSITY PARAMETER MISTUNING

In this section, the tolerance of the proposed approach to the choice of maximum sparsity parameters, ϕ_i , is investigated. In the preceding experiment with synthetic data, it was found convenient to separately consider the sparsity constraints of the task-related time courses, which need to be explicitly set on a one-to-one basis. The sparsity constraints of the remaining sources only need a rough estimation of the sparsity percentage, as described in Section I-B of the supplementary material. This convention is followed in Table 1.a, where three such setups, denoted by θ_1 , θ_2 and θ_3 are listed. The first one is the closest, overall, to the true sparsities. In θ_2 , the sparsities of the non-task-related sources, i.e., the 4th up to the 25th, have been grossly assigned in a simplified way; 5 sources with sparsity percentage 90%, 5 sources with

80%, etc. In θ_3 , the sparsity percentage associated to the task-related sources have been largely relaxed by fixing the sparsity percentage value to 85% for all three of them (i.e., the 1st up to the 3rd). This figure roughly corresponds to the smallest sparsity percentage that can be found in all the FBNs and atlases that we have checked (a list of FBN sparsity percentages for several published brain atlases can be found in Section II of the supplementary material).

Fig. 4 shows the performance for these three sparsity setups. The first figure illustrates performance results with respect to the full sources, whereas the second one shows performance results concerning the time courses only. Besides the sparsity levels, the setup of the experiment is the same as the previous one. Performance curves of JADE and SDL are repeated here for reference. Observe that the proposed approach is remarkably robust to sparsity specifications. Indeed, in the analysis of the performance of the time courses (Fig. 4.B), there are no detrimental effects, whereas in the full-source case, θ_3 led to minor performance degradation in the estimates of the task-related sources. This result confirm that the requirement to explicitly set the sparsity for the task-related time courses is not an obstacle when using the proposed algorithm. Thus, if there is extra information concerning the maximum expected sparsity level of the task-related time courses, then this can be used since it can only help. Otherwise, the sparsity percentages for the task-related time courses are all set to a safely small value, e.g., 85%, and this will still be beneficial to the algorithm leading to reliable results.

C. COMPARISON BETWEEN IADL AND GLM

For completeness, we have performed a comparison between the proposed IADL and the standard GLM approach using the SPM12⁷ toolbox, where the design matrix comprises the three task-related time courses. For this study, we observed that SPM and IADL recover the assisted sources 1 and 3 correctly. However, for assisted source 2 the result of SPM is significantly inferior to that of IADL. Full details of this experiment can be found in Section V-B of the supplementary material.

D. COMPARISON BETWEEN IADL AND SEVERAL ALTERNATIVES FOR THE ANALYSIS OF fMRI DATA

To position IADL among all the alternative matrix factorization-based methods for fMRI, concerning all their features and analysis capabilities, we present a comprehensive comparison in Table 2. Among these different alternatives, observe that IADL complies with all the studied criteria apart from criterion **F**, namely to be able to explicitly specify the place/voxels within the brain that a FBN will appear through user-defined masks. However, note that IADL can also comply with this criterion with relatively mild modifications in the spatial map constraint, \mathcal{L}_w . Such

⁷Statistical Parametric Mapping (SPM). Wellcome Trust Centre for Neuroimaging, London, UK. <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>

TABLE 2. Comparison between several alternatives for fMRI data analysis.

Algorithm	Criteria								
	A	B	C	D	E	F	G	H	I
IADL	✓	✓	✓	✓	✓	✗	✓	✓	✓
ICA [6], [7], [8]	✗	✗	✓	✓	✓	✗	✗	✓	✓
CanICA [54]	✗	✗	✓	✓	✓	✗	✗	✓	✓
SPM [4]	✓	✗	✗	✓	✓	✗	✗	✓	✓
McICA [35]	✓	✓	✓	✗	✗	✗	✗	✓	✓
CSTICA [32], [34]	✓	✓	✓	✗	✗	✓	✗	✓	✗
CTICA [29], [30]	✓	✓	✓	✗	✗	✗	✗	✓	✗
ADL [56]	✓	✓	✓	✗	✗	✗	✗	✓	✓
CSICA [31]	✗	✗	✓	✗	✗	✓	✗	✗	✗
T-NMF [18]	✓	✓	✓	✗	✗	✗	✗	✗	✗
ODL [51]	✗	✗	✓	✗	✗	✗	✗	✗	✓
SDL [38]	✓	✗	✓	✗	✗	✓	✗	✗	✗
S-NMF [17]	✗	✗	✓	✗	✗	✓	✗	✗	✗
Sparse-SPM [28], [11]	✗	✗	✓	✗	✗	✗	✗	✗	✗

Criteria:

A – Can impose task-related time information in a strict way.

B – Can impose task-related time information in a relaxed/assisted way.

C – Provides blind estimation of brain-induced sources and other unmodeled physiological noise or artifacts.

D – Neither cross validation nor experimental visual inspection are needed for parameter fine tuning.

E – Free parameters are easy to tune and accept a physical interpretation within the fMRI context.

F – Allows to include spatial information from the sources of interest through user-defined masks.

G – Allows to include spatial information for all the sources in a relaxed/assisted way.

H – Can explicitly deal with dense sources.

I – An implementation of the code is freely available.

a development is beyond this paper, and will be presented elsewhere.

V. TASK-fMRI DATA ANALYSIS

The following study with real fMRI data aims to illustrate the advantages of the proposed approach in a realistic scenario, and to compare its performance with other standard techniques. In particular, for this study, in addition to the IADL algorithm we also employed SDL, the standard GLM implemented in SPM, and two ICA algorithms: ERBM from the GIFT⁸ toolbox and CanICA⁶ from the Nilearn toolbox.

A. fMRI DATA

For this study, we considered 900 subjects from the motor-task fMRI dataset of the WU-Minn Human Connectome Project [57], which is available at the HCP repository,⁹ and the acquisition parameters are summarized in the imaging protocols.¹⁰ This experiment follows a standard block paradigm, where a visual cue asks the participants to either tap their left/right fingers, squeeze their left/right toes or move their tongue. Each movement block lasted 12 seconds

⁸The Group ICA of fMRI Toolbox (GIFT) is an open Matlab software suitable for independent component analysis and blind source separation of group (and single subject) fMRI data: <http://mialab.mrn.org/software/gift/>.

⁹Human Connectome Project: <https://www.humanconnectome.org/>

¹⁰HCP 3T Imaging Protocol Overview: <http://protocols.humanconnectome.org/HCP/3T/imaging-protocols.html>

and is preceded by a 3 second of visual cue. In addition, there are 3 extra fixation blocks of 15 second each, as detailed in the Human Connectome Project protocols.¹¹

There are two main reasons for selecting this specific dataset: First, the FBNs related to this experimental design are well studied [38], [58]–[62], which facilitates the evaluation of the results. Second, this dataset is particularly challenging: The FBNs of interest exhibit significant asymmetries in their intensity [60] and some spatial maps exhibit high overlap, particularly within the cerebellar cortex [61]. Besides, the cerebral areas from the motor cortex are larger and exhibit a lower inter-subject variability than those from the cerebellum.

Finally, on top of the standard preprocessing pipeline already applied to the obtained datasets (see [59], [63]), we further smoothed each volume with a 4-mm FWHM Gaussian kernel.

B. METHODS & PARAMETER SETUP

For the GLM analysis, we used SPM12⁶, and we followed the same standard procedure as described in [59]. Put succinctly, we defined six task-related time courses, i.e., one per experimental condition: visual, right-hand, left-hand, right-foot, left-foot, and tongue. We estimated each task-related time course as the convolution between the cHRF and each experimental condition, where each experimental condition consisted of a succession of blocks with duration equal to its presentation time. Apart from the six task-related time courses, the design matrix also includes the temporal and spatial derivatives. Then, following the same approach as in [59], at the single-subject level, we computed a linear contrast to assess significant activity.

For the group study, we randomly split the dataset to generate a total number of twenty groups of 15, 30, and 60 subjects each. Then, we performed a group analysis for each combination of subjects to assess significant activity. We studied these three group sizes to evaluate the impact of the number of subjects on performance.

For the matrix factorization methods, in all cases the total number of sources was set equal to 20, which is a reasonable estimate of the expected number of sources (see further discussion in Section I.A of the supplementary material). For the ICA analysis, first, we used the software toolbox GIFT, which implements multiple ICA algorithms in the context of fMRI data analysis. In this study, the algorithms Fast-ICA, Infomax, Erica and ERBM were tested. To save space, we only reported the results of the ERBM algorithm [64], since it appeared to perform somewhat better compared to the rest. Finally, we performed a group fMRI analysis using CanICA⁶, a state-of-the-art ICA-based algorithm.

Concerning IADL and SDL, they both used the same six task-related time courses used in the SPM analysis. Table 1.c shows the sparsity percentage set in IADL. The first six

values correspond to the task-related time courses, i.e., visual, right/left-hand, right/left-foot, and tongue, in this order. The sparsity percentage of the rest of the sources (7th to 20th in Table 1.c) are gradually diminishing in a fashion similar to the one used in the analysis of synthetic data above (see Table 1.a). As discussed in the previous section, IADL is robust to sparsity overdetermination. Therefore, little difference in performance is expected if the sparsity percentage values are increased (e.g., all six first values in Table 1.c could be set to 90%. Parameter c_δ was set to 4.6, following the same automatic tuning approach as used in Section IV (see also the discussion in Section I.C of the supplementary material).

In SDL, the λ parameter, which determines the achieved sparsity, requires to be fine-tuned. Of course, due to the lack of a ground-truth, we can only implicitly optimize λ against the performance criteria that we used, namely, reproducibility and reliability, as we further explain in the next section. However, we observed that the optimal λ is heavily dependent on both the group size as well as on the particular selection of subjects. Accordingly, the λ value that we finally used, $\lambda = 200$, is the one that leads to the best mean performance across all group sizes.

The spatial maps used to evaluate the performance of the matrix factorization methods were computed via the pseudo-inverse approach, namely $\tilde{\mathbf{S}} = \mathbf{D}^+ \mathbf{X}$, following the discussion in Section I.E of the supplementary material. Then, we identified the spatial maps associated with our task of interest. For the blind methods, we determined the specific sources of interest by selecting those sources whose time courses presented the highest correlation with the studied tasks, in a similar fashion as described in Section I.D of the supplementary material. For all group analyses, we thresholded each spatial map for statistical significance at $z > 2.32$ ($p < 0.01$), an informative lower threshold as used in [59]. Similarly, all statistics were computed voxel-wise (not, for example, using cluster-based thresholding), to maximize simplicity of interpretation of the results.

C. PERFORMANCE EVALUATION

To quantitatively evaluate the performance of the different methods, we considered two different criteria: a) reproducibility and b) reliability.

1) REPRODUCIBILITY

Reproducibility refers to the similarity among different realizations with the same number of subjects, since the particular selection of subject may affect performance. In this study, we measured reproducibility among pairs of the twenty randomly generated groups using three complementary metrics:

- t - A metric that measures the one-to-one overlap among components [54].
- e - A metric that quantifies the match between the subspaces spanned by the maps from each component [54].
- J - The Jaccard overlap, a standard metric that quantifies the similarity between images, which has already been used to quantify group reproducibility in fMRI [62].

¹¹Task-fMRI 3T Imaging Protocol Details: <http://protocols.humanconnectome.org/HCP/3T/task-fMRI-protocol-details.html>

TABLE 3. Average reproducibility measures t , e , and J among the five studied motor tasks for the different studied group sizes of SPM, IADL, SDL, CanICA, and ERBM, where the value in parenthesis indicates the measured standard deviation of the obtained results among motor tasks. The table shows the average reproducibility of the cerebral and the cerebellar areas separately, to have a clear view of the behavior of the ROIs.

Cerebrum									
	15 Subjects			30 Subjects			60 Subjects		
	t (σ)	e (σ)	J (σ)	t (σ)	e (σ)	J (σ)	t (σ)	e (σ)	J (σ)
SPM	0.45 (0.03)	0.22 (0.03)	0.17 (0.02)	0.54 (0.04)	0.31 (0.04)	0.21 (0.02)	0.62 (0.03)	0.41 (0.04)	0.27 (0.02)
IADL	0.73 (0.04)	0.56 (0.06)	0.37 (0.04)	0.81 (0.04)	0.79 (0.07)	0.51 (0.04)	0.88 (0.03)	0.99 (0.08)	0.62 (0.04)
SDL	0.54 (0.12)	0.37 (0.11)	0.26 (0.06)	0.59 (0.14)	0.52 (0.19)	0.32 (0.096)	0.61 (0.13)	0.59 (0.18)	0.35 (0.10)
CanICA	0.62 (0.07)	0.59 (0.05)	0.43 (0.05)	0.66 (0.09)	0.68 (0.03)	0.51 (0.07)	0.70 (0.09)	0.73 (0.01)	0.57 (0.07)
ERBM	0.21 (0.08)	0.13 (0.08)	0.10 (0.04)	0.25 (0.08)	0.18 (0.06)	0.13 (0.04)	0.27 (0.08)	0.24 (0.07)	0.16 (0.05)

Cerebellum									
	15 Subjects			30 Subjects			60 Subjects		
	t (σ)	e (σ)	J (σ)	t (σ)	e (σ)	J (σ)	t (σ)	e (σ)	J (σ)
SPM	0.40 (0.04)	0.17 (0.04)	0.17 (0.02)	0.53 (0.04)	0.29 (0.04)	0.22 (0.02)	0.66 (0.04)	0.43 (0.05)	0.30 (0.04)
IADL	0.61 (0.07)	0.40 (0.08)	0.33 (0.05)	0.65 (0.09)	0.46 (0.10)	0.39 (0.06)	0.79 (0.04)	0.64 (0.06)	0.50 (0.05)
SDL	0.44 (0.12)	0.24 (0.10)	0.21 (0.07)	0.47 (0.11)	0.36 (0.11)	0.23 (0.07)	0.50 (0.12)	0.35 (0.13)	0.26 (0.09)
CanICA	0.33 (0.08)	0.18 (0.07)	0.18 (0.04)	0.50 (0.07)	0.35 (0.07)	0.29 (0.04)	0.59 (0.10)	0.48 (0.08)	0.37 (0.04)
ERBM	0.14 (0.07)	0.08 (0.04)	0.07 (0.03)	0.22 (0.08)	0.14 (0.07)	0.10 (0.04)	0.27 (0.07)	0.22 (0.07)	0.14 (0.04)

The details regarding the formal definition of these metrics can be found in Section I.B of the supplementary material.

It is essential to understand that reproducibility alone does not measure if the method works correctly or not, in the sense of correctly separating the different brain sources. Instead, it only provides information regarding how consistent the obtained results are among different realizations of the group analysis. Note that a method can systematically fail to separate a specific brain source, and it can still exhibit good reproducibility (i.e., consistently producing similar wrong results), for example, keeping two sources merged as a single one.

2) RELIABILITY

Reliability refers to the ability of the methods to consistently detect significant activity within the expected regions of interest (ROIs). Note that the ROIs related to the motor tasks have been well documented and studied [24], [38], [58]–[62]. Therefore, we can define a mask that approximately delineates the corresponding ROIs for each motor task.

To measure reliability, we first construct a conjunction map for each motor task. The conjunction map is a particular kind of spatial map, where each voxel indicates the number of subjects/groups that exhibited significant activity within that specific voxel. In this study, for each method, we determine the conjunction map among the twenty groups for each studied group size. Then, we normalize the lesion maps dividing by the total number of realizations. Thus, the normalized lesion maps have values from 0 to 1, where 0 means that no activity was detected in that voxel, and 1 means that all studied groups showed significant activity within that particular voxel.

Using the lesion maps and the defined ROIs for each motor task, we quantify the reliability of the method through two complementary metrics:

- OC - Overlap Consistency
- FPM - False Positive Rate

The OC measures the mean value of the detected significant active voxels within the ROI, that is, the mean value

of the active voxels of the conjunction map within the ROI. This value serves to evaluate the success of the method to find consistent activity within the expected ROIs. On the other hand, the FPR (also known as fall-out or false alarm rate) is defined as the ratio of the number of negative events wrongly categorized as positive (false positive) over the total number of actual negative events. We count as a false positive any activated voxel outside the ROI.

Ideally, the perfect method should exhibit a mean overlap consistency of 100% and an FPR of 0%, meaning that all the analyses obtained the same spatial maps within the expected ROI. Note that both the OC and the FPR are needed to provide a complete view of the reliability of the results. For example, one method may exhibit a good OC, but also a large FPR, in which case, the method is not reliable. Similarly, a lower FPR is not useful if the method does not have a good OC.

D. RESULTS

We analyzed the brain areas associated with each motor task from the cerebrum and the cerebellum separately. We divided the analysis of performance over these two main areas to provide a complete view of the behavior of the studied methods, since the motor areas within the cerebrum present different behavior than those from the cerebellum, as we further discuss in the next section.

Table 3 shows the obtained reproducibility measurement for each studied method. The table depicts the mean value of the three implemented metrics (t , e and J), and the value in parentheses stands for the standard deviation obtained among the different studied motor tasks (left/right hand, left/right foot, and tongue) for each studied group size. Similarly, Table 4 shows the reliability of each studied method and group size. The table shows the mean value of the OC and the FPR and the value in parentheses indicates the corresponding standard deviation among the five studied motor tasks.

For completeness, Fig. 5 shows the spatial maps from two randomly selected groups with 60 participants. We focused on the spatial maps from the group analysis

TABLE 4. Average reliability of SPM, IADL, SDL, CanICA, and ERBM for the different group sizes. The table shows mean values for OC and FPR among the five studied motor tasks from the analysis of the conjunction maps over the cerebral and cerebellar areas separately. The value in parenthesis indicates the standard deviation of the obtained measures among the studied motor tasks.

Cerebrum						
	15 Subjects		30 Subjects		60 Subjects	
	OC (σ)	FPR (σ)	OC (σ)	FPR (σ)	OC (σ)	FPR (σ)
SPM	67.7 (7.6)	6.51 (2.51)	68.3 (8.0)	6.08 (3.04)	67.6 (9.6)	6.18 (3.47)
IADL	78.7 (5.5)	1.75 (0.20)	81.6 (3.4)	1.20 (0.20)	88.3 (3.2)	0.95 (0.21)
SDL	73.2 (3.6)	3.79 (1.12)	73.4 (5.1)	3.19 (0.49)	73.6 (5.4)	2.12 (0.44)
CanICA	71.1 (30.3)	2.32 (1.40)	74.9 (27.5)	1.79 (1.23)	76.2 (32.5)	1.42 (1.00)
ERBM	35.7 (21.2)	3.27 (0.97)	38.0 (26.2)	2.98 (1.11)	42.3 (25.0)	2.84 (1.38)

Cerebellum						
	15 Subjects		30 Subjects		60 Subjects	
	OC (σ)	FPR (σ)	OC (σ)	FPR (σ)	OC (σ)	FPR (σ)
SPM	51.6 (7.0)	6.37 (1.98)	59.9 (5.8)	5.86 (2.40)	68.2 (4.9)	5.48 (2.57)
IADL	61.9 (3.8)	2.09 (0.47)	65.7 (4.6)	1.72 (0.23)	75.0 (4.6)	1.27 (0.18)
SDL	56.7 (6.8)	5.06 (1.36)	60.3 (9.03)	5.62 (1.72)	64.9 (6.9)	4.57 (2.04)
CanICA	31.4 (22.5)	1.21 (1.03)	39.3 (25.5)	0.85 (0.96)	45.6 (31.2)	0.48 (0.41)
ERBM	29.1 (16.5)	5.22 (1.33)	35.4 (25.0)	4.24 (1.72)	36.2 (26.1)	4.65 (2.16)

with 60 participants since, according to Table 3, all methods performed best with this group size. Both figures show the significant active voxels at $z > 2.32$ ($p < 0.01$), and each row presents the most representative slices for each motor task [59].

E. DISCUSSION

The quantitative study verifies that the proposed IADL algorithm exhibited the best reproducibility and reliability at the same time, followed by CanICA, SDL, SPM, and ERBM.

A closer inspection reveals some interesting facts: First, regarding the performance of the ICA algorithms, CanICA achieved considerably better performance compared to ERBM. In particular, CanICA exhibits excellent reproducibility among groups. These results agree with the expected behavior of CanICA, since, before the separation of the independent components, CanICA identifies the common subspace to all the subjects that contains the activation patterns, as detailed in [54]. However, CanICA exhibited a relatively poor OC with considerable variance among tasks, as Table 4 shows. The reason for this is that CanICA (as well as ERBM) had difficulties separating some of the motor areas. For example, the spatial maps in Fig. 5 show that CanICA failed to separate the motor areas from the feet correctly. Furthermore, the excellent reproducibility of CanICA indicates that the algorithm was systematically failing to separate these motor areas.

Second, SPM presents an excellent OC, especially over the cerebral areas. However, SPM exhibited a low reproducibility that is driven by the large presence of false-positive activations, as the large FPR in Table 4 indicates.

SDL appears to outperform SPM in all metrics. This is consistent with the quantitative analyses at the single-subject level performed in [38], but, to our knowledge, it has never been previously demonstrated using quantitative analyses at the group level. Compared with IADL, we can say that SDL's

performances lies roughly in the middle between IADL and SPM. Moreover, we observed that SDL fails to recover some motor areas in some realizations. For example, in Fig. 5.A, SDL missed the motor area corresponding to the right foot, whereas IADL and SPM both show significant activity within the expected ROI. Note that missing an area does not occur often, and it is driven by the particular set of subjects because, unlike CanICA, the relatively lower reproducibility of SDL does not allow us to generalize this observation. Besides, we also noticed that SDL presented more spurious activity across the brain compared to IADL and lower than SPM (see the FPR in Table 4).

With respect to parameter selection for the semi-blind methods, we should emphasize that the differences between IADL and SDL are substantial. First of all, parameter λ in SDL was explicitly tuned (through a tedious and time-consuming optimization process) to perform best in terms of the particular quantitative measures. The value obtained through this optimization process was $\lambda = 200$. This particular λ value lacks any interpretation and is considerably different from the λ value used in the analysis of synthetic data. In contrast, for the parameter set-up of the IADL algorithm, we followed the same guidelines as for the analysis of the synthetic data *without explicitly optimizing for the specific metrics*. Moreover, all parameters have a physical interpretation, as discussed in Section I of the supplementary material.

Differences between the main brain areas

In this study, we analyzed the performance of the studied methods over the cerebrum and the cerebellum separately. The main reason why we divided the analysis over these two main areas is because the FBNs of interest exhibit significant asymmetries, and some of the areas present high overlap. Furthermore, the areas of the motor cortex are large and present a relatively high intensity compared to those from the cerebellum.

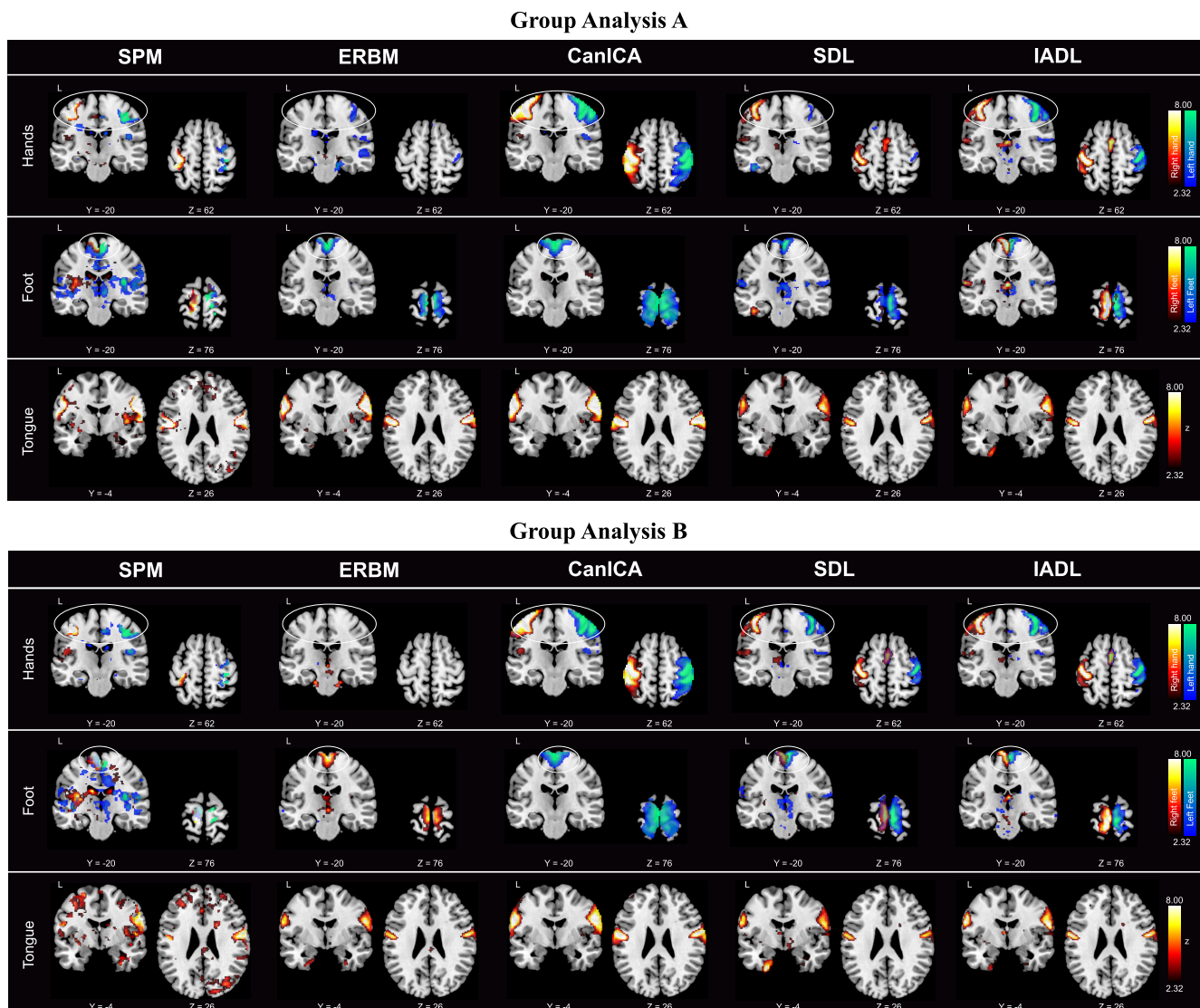


FIGURE 5. Significant active voxels ($z > 2.32$) for two randomly selected group analyses (A and B) out of a total of twenty analyses performed. Results correspond to studies with 60 subjects per group. Each row shows the most representative positions for each specific task: left/right hand, left/right feet and tongue.

The detailed quantitative analysis of the performance of the studied methods over these two main brain parts revealed that the cerebellar areas are considerably more challenging compared to the cerebral ones, as expected. In particular, we observed that the performance of all studied method drops over the cerebellar areas compared to the results from the motor cortex. Interestingly, IADL attains the best performance over the cerebellum, exhibiting a relatively high OC, low FPR, as well as good reproducibility, even for the group analysis with just 15 subjects.

Effect of group size

Our quantitative analysis revealed that the number of subjects has a tangible effect on the performance of various methods. In general, larger groups exhibit better performance compared to smaller groups, which complies with the

expected general behavior. This effect is particularly evident for the SPM method, which had the highest performance gain compared to the rest of the tested methods. Furthermore, we observe that the obtained results for SPM closely resemble the Jaccard overlap results regarding the effects of the number of subjects reported in [62] that also studied the same motor task fMRI dataset.

Algorithmic complexity

Regarding the computational cost of the proposed algorithm, we implemented an efficient approach based on [40], which allows avoiding computationally expensive matrix calculations, as we detail in Section V of the supplementary material. Thus, the most computationally expensive step of the proposed algorithm is the sparse projection (see line 8 in Algorithm 1). The sparse projection depends on the number

of voxels, N , and the number of non-zeros (sparsity percentage), ϕ , and it scales proportionally to the number of sources, K . For these reasons, we implemented an efficient algorithm based on [50] to perform the sparse projection. In the course of these group analyses, we did not observe significant differences in the computation time of the proposed algorithm compared to the rest of studied methods. Only SDL required considerably longer computation time compared to the rest.

VI. CONCLUSIONS

In this paper, we present a new Dictionary Learning method that naturally incorporates external information via two novel convex constraints: a) a sparsity constraint based on the weighted ℓ_1 -norm, which allows to set row-wise sparsity constraints that naturally encapsulate the sparsity of the spatial maps, and b) a similarity constraint over the dictionary, which integrates external a priori information that is available from the experimental task.

The proposed sparsity constraint constitutes a natural alternative to the standard ℓ_1 -norm regularization, allowing to incorporate external sparsity-related information that is available in brain atlases and bypasses the problem of selecting the regularization parameters by following cross-validation arguments that have no practical meaning when real data are involved. Furthermore, the incorporation of the task-related time courses from the experimental task enhances the performance decomposition of their corresponding sources. Moreover, the newly proposed constraints exhibit higher tolerance and robustness to mis-modeling, compared to alternative approaches.

The advantages and the enhanced performance obtained by the proposed method have been verified through detailed quantitative analyses with both realistic synthetic and task-related real fMRI datasets.

ACKNOWLEDGMENT

The authors would like to thank Prof. E. Kofidis, Dept. of Statistics and Insurance Science, University of Piraeus (Greece), and C. Chatzichristos, Dept. of Electrical Engineering (ESAT), Leuven (Belgium), whose comments contributed to improve the quality of the final manuscript.

REFERENCES

- [1] V. Perlberg and G. Marrelec, "Contribution of exploratory methods to the investigation of extended large-scale brain networks in functional MRI: Methodologies, results, and challenges," *Int. J. Biomed. Imag.*, vol. 2008, pp. 1–14, Mar. 2008.
- [2] R. A. Poldrack, J. A. Mumford, and T. E. Nichols, *Handbook of Functional MRI Data Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [3] S. A. Huettel, A. W. Song, and G. McCarthy, *Functional Magnetic Resonance Imaging*, vol. 1, 2nd ed. Sunderland, MA, USA: Sinauer Associates, 2009.
- [4] K. J. Friston, A. P. Holmes, and J. Ashburner. (1999). *Statistical Parametric Mapping (SPM)*. [Online]. Available: <http://www.fil.ion.ucl.ac.uk/spm/>
- [5] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*, 2nd ed. New York, NY, USA: Academic, 2020.
- [6] V. D. Calhoun and T. Adali, "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, 2012.
- [7] W. Du, Y. Levin-Schwartz, G.-S. Fu, S. Ma, V. D. Calhoun, and T. Adali, "The role of diversity in complex ICA algorithms for fMRI analysis," *J. Neurosci. Methods*, vol. 264, pp. 129–135, May 2016.
- [8] Y. Levin-Schwartz, V. D. Calhoun, and T. Adali, "Quantifying the interaction and contribution of multiple datasets in fusion: Application to the detection of schizophrenia," *IEEE Trans. Med. Imag.*, vol. 36, no. 7, pp. 1385–1395, Jul. 2017.
- [9] V. D. Calhoun, V. K. Potluru, R. Phlypo, R. F. Silva, B. A. Pearlmuter, A. Caprihan, S. M. Plis, and T. Adali, "Independent component analysis for brain fMRI does indeed select for maximal independence," *PLoS ONE*, vol. 8, no. 8, 2013, Art. no. e73309.
- [10] I. Daubechies, E. Roussos, S. Takerkart, M. Benharrosh, C. Golden, K. D'Ardenne, W. Richter, J. D. Cohen, and J. Haxby, "Independent component analysis for brain fMRI does not select for independence," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 26, pp. 10415–10422, Jun. 2009.
- [11] Y.-B. Lee, J. Lee, S. Tak, K. Lee, D. L. Na, S. W. Seo, Y. Jeong, and J. C. Ye, "Sparse SPM: Group sparse-dictionary learning in SPM framework for resting-state functional connectivity MRI analysis," *NeuroImage*, vol. 125, pp. 1032–1045, Jan. 2016.
- [12] K. D. Harris and T. D. Mrsic-Flogel, "Cortical connectivity and sensory coding," *Nature*, vol. 503, no. 7474, pp. 51–58, Nov. 2013.
- [13] S. Zhao, J. Han, X. Jiang, H. Huang, H. Liu, J. Lv, L. Guo, and T. Liu, "Decoding auditory saliency from brain activity patterns during free listening to naturalistic audio excerpts," *Neuroinformatics*, vol. 16, nos. 3–4, pp. 309–324, Oct. 2018.
- [14] M. Carroll, G. Cecchi, I. Rish, R. Garg, and A. Rao, "Prediction and interpretation of distributed neural activity with sparse models," *NeuroImage*, vol. 44, no. 1, pp. 112–122, Jan. 2009.
- [15] R. Jenatton, A. Gramfort, V. Michel, G. Obozinski, E. Eger, F. Bach, and B. Thirion, "Multiscale mining of fMRI data with hierarchical structured sparsity," *SIAM J. Imag. Sci.*, vol. 5, no. 3, pp. 835–856, Jan. 2012.
- [16] J. Xie, P. K. Douglas, Y. N. Wu, A. L. Brody, and A. E. Anderson, "Decoding the encoding of functional brain networks: An fMRI classification comparison of non-negative matrix factorization (NMF), independent component analysis (ICA), and sparse coding algorithms," *J. Neurosci. Methods*, vol. 282, pp. 81–94, Apr. 2017.
- [17] S. Ferdowsi, V. Abolghasemi, B. Makkiabadi, and S. Sanei, "A new spatially constrained NMF with application to fMRI," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 5052–5055.
- [18] S. Ferdowsi, V. Abolghasemi, and S. Sanei, "A constrained NMF algorithm for bold detection in fMRI," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Aug. 2010, pp. 77–82.
- [19] B. Feng, Z. L. Yu, Z. Gu, and Y. Li, "Analysis of fMRI data based on sparsity of source components in signal dictionary," *Neurocomputing*, vol. 156, pp. 86–95, May 2015.
- [20] G. K. Aguirre, E. Zarahn, and M. D'Esposito, "The variability of human, BOLD hemodynamic responses," *NeuroImage*, vol. 8, no. 4, pp. 360–369, Nov. 1998.
- [21] F. I. Karahanoğlu, C. Caballero-Gaudes, F. Lazeyras, and D. Van De Ville, "Total activation: FMRI deconvolution through spatio-temporal regularization," *NeuroImage*, vol. 73, pp. 121–134, Jun. 2013.
- [22] D. A. Handwerker, J. M. Ollinger, and M. D'Esposito, "Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses," *NeuroImage*, vol. 21, no. 4, pp. 1639–1651, Apr. 2004.
- [23] W. Zhang, J. Lv, X. Li, D. Zhu, X. Jiang, S. Zhang, Y. Zhao, L. Guo, J. Ye, D. Hu, and T. Liu, "Experimental comparisons of sparse dictionary learning and independent component analysis for brain network inference from fMRI data," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 1, pp. 289–299, Jan. 2019.
- [24] T. D. Hermansen, S. Ventegodt, and I. Kandel, "Human development XI: The structure of the cerebral cortex. Are there really modules in the brain?" *Sci. World J.*, vol. 7, pp. 1922–1929, Sep. 2007.
- [25] J. M. Fuster, "Cortex and memory: Emergence of a new paradigm," *J. Cognit. Neurosci.*, vol. 21, no. 11, pp. 2047–2072, Nov. 2009.
- [26] J. Xu, M. N. Potenza, V. D. Calhoun, R. Zhang, S. W. Yip, J. T. Wall, G. D. Pearlson, P. D. Worhunsky, K. A. Garrison, and J. M. Moran, "Large-scale functional network overlap is a general property of brain functional organization: Reconciling inconsistent fMRI findings from general-linear-model-based analyses," *Neurosci. Biobehav. Rev.*, vol. 71, pp. 83–100, Dec. 2016.
- [27] D. Hu, L. Yan, Y. Liu, Z. Zhou, K. J. Friston, C. Tan, and D. Wu, "Unified SPM-ICA for fMRI analysis," *NeuroImage*, vol. 25, no. 3, pp. 746–755, 2005.

- [28] K. Lee, S. Tak, and J. C. Ye, "A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1076–1089, May 2011.
- [29] W. Lu and J. C. Rajapakse, "Constrained independent component analysis," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 570–576.
- [30] V. D. Calhoun, T. Adali, M. C. Stevens, K. A. Kiehl, and J. J. Pekar, "Semi-blind ICA of fMRI: A method for utilizing hypothesis-derived time courses in a spatial ICA analysis," *NeuroImage*, vol. 25, no. 2, pp. 527–538, Apr. 2005.
- [31] Q.-H. Lin, J. Liu, Y.-R. Zheng, H. Liang, and V. D. Calhoun, "Semiblind spatial ICA of fMRI using spatial constraints," *Hum. Brain Mapping*, vol. 31, no. 7, pp. 1076–1088, 2010.
- [32] Z. Wang, M. Xia, Z. Jin, L. Yao, and Z. Long, "Temporally and spatially constrained ICA of fMRI data analysis," *PLoS ONE*, vol. 9, no. 4, 2014, Art. no. e94211. [Online]. Available: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0094211>
- [33] W. Lu and J. C. Rajapakse, "ICA with reference," *Neurocomputing*, vol. 69, nos. 16–18, pp. 2244–2257, Oct. 2006.
- [34] Y. Shi, W. Zeng, N. Wang, and L. Zhao, "A new constrained spatiotemporal ICA method based on multi-objective optimization for fMRI data analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 9, pp. 1690–1699, Sep. 2018.
- [35] Z.-L. Zhang, "Morphologically constrained ICA for extracting weak temporally correlated signals," *Neurocomputing*, vol. 71, nos. 7–9, pp. 1669–1679, Mar. 2008.
- [36] W. Lu and J. C. Rajapakse, "Approach and applications of constrained ICA," *IEEE Trans. Neural Netw.*, vol. 16, no. 1, pp. 203–212, Jan. 2005.
- [37] T. Rasheed, Y.-K. Lee, and T.-S. Kim, "Constrained spatiotemporal ICA and its application for fMRI data analysis," in *Proc. 13th Int. Conf. Biomed. Eng.*, in IFMBE Proceedings. Berlin, Germany: Springer, 2009, pp. 555–558.
- [38] S. Zhao, J. Han, J. Lv, X. Jiang, X. Hu, Y. Zhao, B. Ge, L. Guo, and T. Liu, "Supervised dictionary learning for inferring concurrent brain networks," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 2036–2045, Oct. 2015.
- [39] Y. Kopsinis, H. Georgiou, and S. Theodoridis, "fMRI unmixing via properly adjusted dictionary learning," in *Proc. 22nd Eur. Signal Process. Conf. (EUSIPCO)*, 2014, pp. 2075–2079.
- [40] M. Yaghoobi, T. Blumensath, and M. E. Davies, "Dictionary learning for sparse approximations via the majorization method," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2178–2191, Jun. 2009.
- [41] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [42] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, 2008.
- [43] Y. Kopsinis, K. Slavakis, and S. Theodoridis, "Online sparse system identification and signal reconstruction using projections onto weighted ℓ_1 balls," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 936–952, Mar. 2011.
- [44] J. M. Levin, B. D. Frederick, M. H. Ross, J. F. Fox, H. L. von Rosenberg, M. J. Kaufman, N. Lange, J. H. Mendelson, B. M. Cohen, and P. F. Renshaw, "Influence of baseline hematocrit and hemodilution on BOLD fMRI activation," *Magn. Reson. Imag.*, vol. 19, no. 8, pp. 1055–1062, Oct. 2001.
- [45] M. D. Noseworthy, J. Alfonsi, and S. Bells, "Attenuation of brain BOLD response following lipid ingestion," *Hum. Brain Mapping*, vol. 20, no. 2, pp. 116–121, Oct. 2003.
- [46] K. J. Friston, O. Josephs, G. Rees, and R. Turner, "Nonlinear event-related responses in fMRI," *Magn. Reson. Med.*, vol. 39, no. 1, pp. 41–52, Jan. 1998.
- [47] G. I. Allen and M. Weylandt, "Sparse and functional principal components analysis," 2013, *arXiv:1309.2895*. [Online]. Available: <http://arxiv.org/abs/1309.2895>
- [48] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.
- [49] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, Jan. 2013.
- [50] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projections onto the ℓ_1 -ball for learning in high dimensions," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 272–279.
- [51] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Jan. 2010.
- [52] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, Nov. 1995.
- [53] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc.-F, Radar Signal Process.*, vol. 140, no. 6, pp. 362–370, 1993.
- [54] G. Varoquaux, S. Sadaghiani, P. Pinel, A. Kleinschmidt, J. B. Poline, and B. Thirion, "A group model for stable multi-subject ICA on fMRI datasets," *NeuroImage*, vol. 51, no. 1, pp. 288–299, May 2010.
- [55] V. Abrol, P. Sharma, S. F. Roohi, A. K. Sao, and A. A. Kassim, "Fast and robust FMRI unmixing using hierarchical dictionary learning," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 714–718.
- [56] M. M. Moreno, Y. Kopsinis, E. Kofidis, C. Chatzichristos, and S. Theodoridis, "Assisted dictionary learning for FMRI data analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 806–810.
- [57] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, and K. Ugurbil, "The WU-minn human connectome project: An overview," *NeuroImage*, vol. 80, pp. 62–79, Oct. 2013.
- [58] A.-K. Seghouane and A. Iqbal, "Basis expansion approaches for regularized sequential dictionary learning algorithms with enforced sparsity for fMRI data analysis," *IEEE Trans. Med. Imag.*, vol. 36, no. 9, pp. 1796–1807, Sep. 2017.
- [59] D. M. Barch, G. C. Burgess, M. P. Harms, S. E. Petersen, B. L. Schlaggar, M. Corbetta, M. F. Glasser, S. Curtiss, S. Dixit, C. Feldt, D. Nolan, E. Bryant, T. Hartley, O. Footer, J. M. Bjork, R. Poldrack, S. Smith, H. Johansen-Berg, A. Z. Snyder, and D. C. Van Essen, "Function in the human connectome: Task-fMRI and individual differences in behavior," *NeuroImage*, vol. 80, pp. 169–189, Oct. 2013.
- [60] B. T. T. Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, B. Fischl, H. Liu, and R. L. Buckner, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *J. Neurophysiol.*, vol. 106, no. 3, pp. 1125–1165, Sep. 2011.
- [61] R. L. Buckner, F. M. Krienen, A. Castellanos, J. C. Diaz, and B. T. T. Yeo, "The organization of the human cerebellum estimated by intrinsic functional connectivity," *J. Neurophysiol.*, vol. 106, no. 5, pp. 2322–2345, Nov. 2011.
- [62] B. O. Turner, E. J. Paul, M. B. Miller, and A. K. Barbey, "Small sample sizes reduce the replicability of task-based fMRI studies," *Commun. Biol.*, vol. 1, no. 1, p. 62, Dec. 2018.
- [63] M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, D. C. Van Essen, and M. Jenkinson, "The minimal preprocessing pipelines for the human connectome project," *NeuroImage*, vol. 80, pp. 105–124, Oct. 2013.
- [64] G. Fu, R. Phlypo, M. Anderson, X. Li, and T. Adali, "Blind source separation by entropy rate minimization," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4245–4255, Aug. 2014.



MANUEL MORANTE was born in Granada, Spain, in 1992. He received the bachelor's degree in physics and the master's degree in nanotechnology, physics, and its applications from the University of Granada, in 2014 and 2015, respectively. He is currently pursuing the Ph.D. degree with the National and Kapodistrian University of Athens, Greece.

In 2015, he received a Marie Skłodowska-Curie Fellowship within the Innovative Training Network (MacSeNet) of the European Union's Seventh Framework Programme (H2020-MSCA-ITN-2014).



YANNIS KOPSINIS (Member, IEEE) received the Ph.D. degree from the Department of Informatics and Telecommunications, University of Athens, in 2004.

He is a Co-Founder of Libra MLI Ltd. He has gained a number of prestigious personal research grants. Moreover, he has worked for more than five years as a Senior Research Fellow of the School of Engineering and Electronics, The University of Edinburgh, U.K. He has published more than

50 articles in technical journals and conferences, and he has coauthored three book chapters. His work has received more than 1200 citations. He has worked on signal processing and machine learning theory for more than two decades in applications ranging from telecoms to audio and medical. His current research interests include in the areas of online machine learning, cooperative and federated learning, deep learning theory and methods, constrained optimization, and robust predictive analytics. He has received the Marie Curie IEF Fellowship and the Ramón y Cajal Fellowship from the University of Granada, Spain.



SERGIOS THEODORIDIS (Life Fellow, IEEE) is currently a Professor of signal processing and machine learning with the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, and he is the holder of a part-time Chair of The Chinese University of Hong Kong, Shenzhen. He has authored the book *Machine Learning: A Bayesian and Optimization Perspective* (Academic Press, 2015), and coauthored the best-selling book *Pattern Recognition* (Academic Press, Fourth Edition, 2009), the book *Introduction to Pattern Recognition: A MATLAB Approach* (Academic Press, 2010), and three books in Greek, two of them for the Greek Open University. His research interests include in the areas of adaptive algorithms, distributed and sparsity-aware learning, machine learning and pattern recognition, signal processing, learning for biomedical applications, and audio processing and retrieval.

Dr. Theodoridis has served as a Distinguished Lecturer for the IEEE Signal Processing as well as the Circuits and Systems Societies. He is a Fellow of the IET and the EURASIP, and a Corresponding Fellow of the Royal Society of Edinburgh (RSE). He currently serves as the Vice President of the IEEE Signal Processing Society. He was Otto Monstead Guest Professor with the Technical University of Denmark, in 2012, and holder of the Excellence Chair of the Department of Signal Processing and Communications, University Carlos III, Madrid, Spain, in 2011. He has served as the President of the European Association for Signal Processing (EURASIP), a member of the Board of Governors for the IEEE Circuits and Systems (CAS) Society, a member of the Board of Governors (Member-at-Large) of the IEEE SP Society, and the Chair of the Signal Processing Theory and Methods (SPTM) Technical Committee of the IEEE SPS. He was a recipient of the 2014 IEEE Signal Processing Society Education Award, the 2014 EURASIP Meritorious Service Award, and the 2017 EURASIP Athanasios Papoulis Award. He has coauthored seven articles that have received the Best Paper Awards, including the 2009 IEEE Computational Intelligence Society Transactions on Neural Networks Outstanding Paper Award and the 2014 *IEEE Signal Processing Magazine* Best Paper Award. He has served as an Editor-in-Chief for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was the Co-Editor of the book *Adaptive System Identification and Signal Processing Algorithms*, (Prentice Hall, 1993). He is the Editor-in-Chief of *Signal Processing* book series (Academic Press) and the Co-Editor-in-Chief of the *E-Reference Signal Processing* (Elsevier).



ATHANASSIOS PROTOPAPAS was born in Athens, Greece, in 1969. He received the B.S. degree in physics from The University of Patras, Greece, in 1993, and the M.S. degree in cognitive science and engineering and the Ph.D. degree in cognitive science from Brown University, Providence, RI, USA, in 1995 and 1997, respectively.

From 1996 to 1998, he was a Research Scientist at Scientific Learning Corp., San Francisco, CA, USA. From 1999 to 2010, he was a Principal Researcher at the Institute for Language and Speech Processing, Athens, and an Associate Professor of cognitive science at the National and Kapodistrian University of Athens, Athens, from 2010 to 2016. Since 2016, he has been a Professor with the Department of Special Needs Education, University of Oslo, Norway. His research interest focuses on the development of reading skills. He is an Associate Editor of the *Journal of Experimental Psychology: Human Perception and Performance* and an Academic Editor of *PLoS One*.

• • •