

# Constrained Independent Vector Analysis with Reference for Multi-Subject fMRI Analysis

Trung Vu\*, Francisco Laport\*, Hanlu Yang, Vince D. Calhoun, and Tülay Adali

**Abstract**—Independent component analysis (ICA) is now a widely used solution for the analysis of multi-subject functional magnetic resonance imaging (fMRI) data. Independent vector analysis (IVA) generalizes ICA to multiple datasets (multi-subject data). Along with higher-order statistical information in ICA, it leverages the statistical dependence across the datasets as an additional type of statistical diversity. As such, IVA preserves variability in the estimation of single-subject maps but its performance might suffer when the number of datasets increases. Constrained IVA is an effective way to bypass computational issues and improve the quality of separation by incorporating available prior information. Existing constrained IVA approaches often rely on user-defined threshold values to define the constraints. However, an improperly selected threshold can have a negative impact on the final results. This paper proposes two novel methods for constrained IVA: one using an adaptive-reverse scheme to select variable thresholds for the constraints and a second one based on a threshold-free formulation by leveraging the unique structure of IVA. Notably, the proposed algorithms do not require all components to be constrained, utilizing free components to model interferences and components that might not be in the reference set. We demonstrate that our solutions provide an attractive solution to multi-subject fMRI analysis both by simulations and through analysis of resting state fMRI data collected from 98 subjects — the highest number of subjects ever used by IVA algorithms. Our results show that both proposed approaches obtain significantly better separation quality and model match while providing computationally efficient and highly reproducible solutions.

**Index Terms**—independent vector analysis, constrained IVA, multivariate Gaussian distribution, fMRI analysis.

## I. INTRODUCTION

INDEPENDENT component analysis (ICA) is a blind source separation (BSS) technique that decomposes a multivariate signal into statistically independent components. This data-driven approach has found fruitful applications in the

This work is supported in part by the grants NIH R01MH118695, NIH R01MH123610, NIH R01AG073949, NSF 2112455, NSF 2316420, and Xunta de Galicia ED481B 2022/012.

Trung Vu, Francisco Laport, Hanlu Yang, and Tülay Adali are with Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, MD 21250, USA (e-mail: {trungvv, flopez2, hyang3, adali}@umbc.edu). Francisco Laport is also with CITIC Research Center, University of A Coruña, Campus de Elviña, 15071 A Coruña, Spain.

Vince Calhoun is with the Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, and Emory University, Atlanta, GA 30303, USA (e-mail: vcalhoun@gsu.edu).

Trung Vu and Francisco Laport contributed equally to this work.

analysis of neuroimaging data including functional magnetic resonance imaging (fMRI) [1], electroencephalography (EEG) [2], and magnetoencephalography (MEG) [3] data. In fMRI studies, ICA can extract either spatially or temporally independent components corresponding to a single subject [4] while spatial ICA has been the dominant version [5].

In generalization to analysis of multi-subject data, the group ICA method [6] has been by far the most commonly used approach where the multi-subject fMRI data are temporally concatenated. Other approaches include spatial concatenation [7] and tensor organization [8]. Once the group data is created, typically through two levels of dimension reduction using principal component analysis, ICA is applied to extract group-independent components, which can be then used to compute the subject-specific components by back-reconstruction, dual or other flavors of regression [6], [9], [10]. The disadvantage of group ICA, however, is that it relies on the assumption of a common subspace among all subjects and hence, its ability to capture subject variability might be limited. Another powerful approach to multi-subject data analysis is independent vector analysis (IVA), which generalizes ICA to multi-subject fMRI analysis by exploiting the statistical dependencies across the subject datasets [11]–[13]. IVA has been shown to perform well in capturing subject variability [14], [15] and is competitive with ICA which has been extensively studied in this regard [16]. Additionally, as a by-product, dependent sources across the subjects are automatically aligned in IVA, avoiding the permutation ambiguity existing in ICA [17]. Notwithstanding, one major drawback of IVA is that its performance degrades when the number of datasets increases or when the level of variability among the subjects is very low [18], [19]. Other approaches to group studies include multiset canonical correlation analysis (MCCA) [20], MultiView ICA [21], joint ICA (jICA) [22], and tensor decomposition [8], [23].

ICA can be further improved in various ways by incorporating spatial constraints [24]. Following a similar strategy, to improve the performance of IVA, constrained IVA has been developed as an effective way to incorporate prior knowledge (often about the sources or the mixing matrices) while also addressing the aforementioned limitations. Similar to constrained ICA [25], [26], constrained IVA introduces (in)equality constraints to the cost function and utilizes the Lagrangian framework to solve the constrained optimization problem. A reliable set of constraints guides IVA algorithms to avoid sub-optimal solutions and increase the quality of source separation and of the estimated components by providing a better model match. There have been two major types of

prior information (references): rough templates of the sources [25], [27] or the mixing matrix [28], [29]. However, given the current emphasis on resting state fMRI data analyses, spatial constraints are more commonly used and are our focus here as well. If a reference is constructed properly, it is the one and only one that is closest to the desired signal in terms of a closeness measure. Therefore, when incorporated into the IVA framework, such references carry adequate information to distinguish the desired signal from artifacts and noise. One key challenge in constrained IVA is selecting a threshold parameter that enforces the closeness between the reference signal and the desired output. A feasible threshold depends on both the designed reference signal and the closeness measure. If the threshold is too small, the output may produce a sub-optimal solution. If the threshold is too large, the corresponding constraint might not be satisfied and cause the learning to become unpredictable. In the context of constrained ICA, Lu and Rajapakse [25] suggested using a small threshold initially and then gradually increasing the threshold. However, this method requires multiple runs of the algorithm, which is computationally expensive, especially for application to IVA. Recently, Bhinge *et al.* [19] studied constrained IVA with multivariate Laplace distributed sources (IVA-L-SOS) and proposed a heuristic scheme, named parameter-tuned constrained IVA (pt-cIVA), for selecting the constraint thresholds from a set of pre-defined values. This adaptive tuning scheme, while facilitating the automatic selection of threshold values, may lead to a sub-optimal solution where the threshold is much smaller than the desired value.

In this paper, we present two attractive solutions to multi-subject fMRI analysis via constrained IVA with reference. First, we address the aforementioned issue with pt-cIVA by introducing an adaptive-reverse scheme for threshold selection, named adaptive-reverse constrained IVA (ar-cIVA). By alternating between a conservative scheme and an assertive scheme, our proposed approach allows the threshold values to increase when the constraints are easily satisfied and to decrease when the constraints are too difficult to achieve. Furthermore, to eliminate the need for threshold selection, we propose a second method for constrained IVA that is threshold-free (tf-cIVA). The references are utilized as a regularization for the IVA cost function, in which not only the similarity between the reference and the corresponding source but also the similarity between that reference and the other sources are taken into account. In both methods, we leverage IVA with multivariate Gaussian sources to exploit second-order statistics (SOS) while utilizing the similarity between the sources and the references to account for higher-order statistics (HOS). Compared with IVA using a multivariate Laplace density model, both our approaches offer significantly faster runtime, with iteration complexity independent of the sample size. We demonstrate the effectiveness of the proposed approaches through a number of experiments on both simulated fMRI-like data and real fMRI data with 98 subjects. We emphasize that to the best of our knowledge, this is the highest number of subjects that have been used by the IVA framework when full correlation is taken into account as in our case. The results show that our methods significantly outperform the

unconstrained IVA method as well as existing constrained IVA methods. It is also promising that our approaches can be applied to large-scale data with a few hundred to thousand subjects, as well as other application domains in joint BSS.

The rest of this paper is organized as follows. Section II provides a brief review of independent vector analysis, the multivariate Gaussian sources, and reference-constrained IVA. Then, Sections III and IV present our two proposed approaches to constrained IVA, namely adaptive-reverse constrained IVA and threshold-free constrained IVA, followed by the implementation details described in Section V. In Sections VI and VII, we demonstrate the effectiveness of the proposed algorithms in simulated fMRI-like data and real fMRI data, respectively. Finally, Section VIII summarizes our work in this paper and discusses potential directions for future work.

## II. PRELIMINARIES

**Notation.** Throughout the paper, we use the notations  $\|\cdot\|_F$  and  $\|\cdot\|_2$  to denote the Frobenius norm and the spectral norm of a matrix, respectively. Additionally,  $\|\cdot\|$  is used on a vector to denote the Euclidean norm. Boldfaced symbols are reserved for vectors and matrices. The notation  $(\cdot)^\top$  denotes the transpose of a matrix. The  $t \times t$  identity matrix is denoted by  $I_t$ . The  $t$ -dimensional vector of all zeros and the  $t$ -dimensional vector of all ones are denoted by  $0_t$  and  $1_t$ , respectively. In addition, the  $i$ th vector in the natural basis of  $\mathbb{R}^t$  is denoted by  $e_i$ . The notation  $\otimes$  denotes the Kronecker product between two matrices and  $\text{vec}(\cdot)$  denotes the vectorization of a matrix by stacking its columns on top of one another. Given an  $n$ -dimensional vector  $\mathbf{x}$ ,  $x_i$  denotes its  $i$ th element and  $\text{diag}(\mathbf{x})$  denotes the  $n \times n$  diagonal matrix with the corresponding diagonal entries  $x_1, \dots, x_n$ . Similarly, for an  $m \times n$  matrix  $\mathbf{X}$ , the  $(i, j)$  entry of  $\mathbf{X}$  is denoted by  $X_{ij}$ .

### A. Independent Vector Analysis (IVA)

Consider  $K$  datasets (subjects), each formed by  $V$  samples (voxels) of linear mixtures of  $N$  independent sources

$$\mathbf{x}^{[k]}(v) = \mathbf{A}^{[k]} \mathbf{s}^{[k]}(v), \quad (1)$$

for  $k = 1, \dots, K$  and  $v = 1, \dots, V$ . Here,  $\mathbf{A}^{[k]} \in \mathbb{R}^{N \times N}$  is an invertible mixing matrix for the  $k$ th dataset and  $\mathbf{s}^{[k]}(v) = [s_1^{[k]}(v), \dots, s_N^{[k]}(v)]^\top$  is the  $v$ th sample of the corresponding source vector. By stacking the  $n$ th source component across  $K$  datasets, we introduce a key concept: the  $n$ th source component vector (SCV) as a  $K$ -dimensional random vector<sup>1</sup>

$$\mathbf{s}_n = [s_n^{[1]}, \dots, s_n^{[K]}]^\top.$$

An appropriate multivariate probability density function (pdf) of the SCV can take all order statistical information within and across the  $K$  datasets into account. The goal of IVA is to identify the independent SCVs via the estimation of  $K$  demixing matrices of the form  $\mathbf{W}^{[k]} = [\mathbf{w}_1^{[k]}, \dots, \mathbf{w}_N^{[k]}]^\top \in \mathbb{R}^{N \times N}$ . Denote  $\mathbf{y}^{[k]}(v) = \mathbf{W}^{[k]} \mathbf{x}^{[k]}(v)$  as the vector containing  $N$  estimated sources for the  $k$ th dataset. The  $n$ th

<sup>1</sup>For convenience, we use the same symbol  $s_n^{[k]}$  to denote the random variable. The samples of  $s_n^{[k]}$  are indicated by the index  $v$  in  $s_n^{[k]}(v)$ .

estimated SCV corresponding to the sample index  $v$  is given by  $\mathbf{y}_n(v) = [y_n^{[1]}(v), \dots, y_n^{[K]}(v)]^\top \in \mathbb{R}^K$ . Assuming the samples are independently and identically distributed (iid) and using the maximum likelihood principle [13], one can write the IVA cost as minimizing the negative log-likelihood w.r.t.  $\mathbf{W} = \{\mathbf{W}^{[k]}\}_{k=1}^K$  and  $\Sigma = \{\Sigma_n\}_{n=1}^N$

$$\mathcal{J}_{\text{IVA}}(\mathbf{W}, \Sigma) \triangleq -\frac{1}{V} \sum_{v,n} \log p_n(\mathbf{y}_n(v) | \Sigma_n) - \sum_k \log |\det \mathbf{W}^{[k]}|, \quad (2)$$

where  $p_n$  denotes the pdf of the  $n$ th SCV. In this formulation, there are no restrictions on the  $\mathbf{W}^{[k]}$  beyond being invertible.

The multivariate Gaussian distribution (MGD) provides an attractive solution to model SCV pdfs in terms of complexity and allows taking full SOS into account. Assuming each estimated SCV  $\mathbf{y}_n$  follows an MGD with zero mean and covariance matrix  $\Sigma_n \in \mathbb{R}^{K \times K}$ , the IVA cost in (2) can be rewritten as (3) in Table I, where  $\mathbf{X}^{[k]} = [\mathbf{x}^{[k]}(1), \dots, \mathbf{x}^{[k]}(V)]$  is the  $N \times V$  data matrix. In [12], Anderson *et al.* study the theoretical properties (e.g., local stability and identifiability conditions) of the IVA framework with MGD source model, i.e., IVA-G, and demonstrate its effectiveness in joint BSS.

In application to fMRI analysis, since underlying sources are more likely to be super-Gaussian [13], [18], the use of only SOS might come across as a limitation. For example, the multivariate Laplace distribution (MLD) has been shown to provide a better model match to fMRI sources [5], [19]. However, this approach is computationally expensive since its iteration complexity depends on the number of data samples. In multi-subject fMRI data analysis, IVA with MLD methods such as IVA-L-SOS and its constrained variants have only been applied to medium-scale settings of no more than 64 subjects and 20 components [30]. In this work, we use a larger dataset of 98 subjects and select an order of 60 components. We demonstrate how to guide the estimation by introducing reference signals and how the model match is maintained while still achieving computational efficiency. In subsequent sections, we will simply refer to the IVA-G cost in (3) as IVA cost for convenience.

### B. Constrained IVA with Reference

In constrained IVA, we consider a set of reference signals  $\{\mathbf{r}_n\}_{n=1}^M \subset \mathbb{R}^V (M \leq N)$  that can be used as prior constraints to guide the separation of sources. For the  $k$ th dataset, the  $n$ th estimated source is given by  $\mathbf{y}_n^{[k]} = [y_n^{[k]}(1), \dots, y_n^{[k]}(V)]^\top$ . The idea here is to ensure that  $\mathbf{r}_n$  has a higher correlation with its corresponding SCV than any other SCVs in the same dataset, i.e.,

$$\epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]}) > \epsilon(\mathbf{r}_n, \mathbf{y}_m^{[k]}) \quad \forall m \neq n, \quad (6)$$

where  $\epsilon : \mathbb{R}^V \times \mathbb{R}^V \rightarrow [0, 1]$  is some similarity measure,  $n = 1, \dots, M$ , and  $m = 1, \dots, N$ . As an example,  $\epsilon(\cdot)$  can be chosen as the absolute value of Pearson correlation

$$\epsilon(\mathbf{a}, \mathbf{b}) = |\text{corr}(\mathbf{a}, \mathbf{b})| = \frac{|\mathbf{a}^\top \mathbf{b}|}{\|\mathbf{a}\| \|\mathbf{b}\|}. \quad (7)$$

A common approach to implementing such constraints is via a pre-defined threshold parameter  $\rho$  [25], [26], [31]. By selecting an appropriate value of  $\rho$  such that

$$\epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]}) \geq \rho > \epsilon(\mathbf{r}_n, \mathbf{y}_m^{[k]}) \quad \forall m \neq n, \quad (8)$$

only one independent component is extracted as the closest one to the reference signal. Thus, the thresholding-constrained formulation is proposed in [31] as

$$\min_{\mathbf{W}, \Sigma} \mathcal{J}_{\text{IVA}}(\mathbf{W}, \Sigma) \quad \text{s.t. } \epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]}) \geq \rho_n \quad \forall n, \quad (9)$$

where  $n = 1, \dots, M$  and  $k = 1, \dots, K$ . The major disadvantage of formulation (9) is that the best values for the threshold parameters are often unknown in practice. If  $\rho_n$  is too small, the output may produce a different component. If  $\rho_n$  is too large, the estimate might not yield a desired component because the corresponding constraint causes the learning to become unpredictable. Hence, an ideal value of  $\rho_n$  is the one that is closest to the similarity between the reference  $\mathbf{r}_n$  and the true source  $\mathbf{s}_n^{[k]}$ , i.e.,  $\epsilon(\mathbf{r}_n, \mathbf{s}_n^{[k]})$ . To address this issue, an adaptive scheme to select  $\rho$ , pt-cIVA, has been proposed in [19]. The idea is to use a set of predefined thresholds  $\mathcal{P}$  and at each iteration, pick a value that is closest to the similarity value between the reference  $\mathbf{r}_n$  and the estimated sources  $\mathbf{y}_n^{[k]}$ :

$$\rho_n = \operatorname{argmin}_{\rho \in \mathcal{P}} \min_{1 \leq k \leq K} |\rho - \epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]})|. \quad (10)$$

While this heuristic was shown to improve the performance of constrained IVA [19], it may lead to a sub-optimal solution where  $\rho_n$  can be much smaller than  $\epsilon(\mathbf{r}_n, \mathbf{s}_n^{[k]})$ . Indeed, if at some iteration, the threshold is always selected such that it is smaller than or equal to  $\epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]})$ , the constraints in (9) will be automatically satisfied and will have no effect on increasing  $\epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]})$  in the next iteration. Another issue with formulation (9) is that the threshold  $\rho_n$  does not depend on  $k$  and hence, does not consider the case where  $\epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]})$  has a different threshold from  $\epsilon(\mathbf{r}_n, \mathbf{y}_n^{[l]})$ , for  $k \neq l$ . To accommodate the variability among the subjects within an SCV, one needs to impose different thresholds for different levels of closeness  $\epsilon(\cdot)$  in the constraints.

### III. ADAPTIVE-REVERSE CONSTRAINED IVA

We introduce an adaptive-reverse scheme for selecting the constraint thresholds that significantly improves the performance of pt-cIVA while maintaining the same computational complexity per iteration. First, we extend (9) to a more flexible constrained formulation that takes into account the subject variability across components

$$\min_{\mathbf{W}, \Sigma} \mathcal{J}_{\text{IVA}}(\mathbf{W}, \Sigma) \quad \text{s.t. } \epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]}) \geq \rho_n^{[k]} \quad \forall n, k. \quad (11)$$

It is emphasized that (9) uses the same threshold  $\rho_n$  for all  $K$  subjects in the  $n$ th component while (11) uses different thresholds  $\rho_n^{[k]}$  for each subject in the  $n$ th component. Second, we propose an adaptive scheme to select  $\rho_n^{[k]}$  that alternates between two principles: (i) choosing the smallest value that does not satisfy the constraint

$$\rho_n^{[k]} = \operatorname{argmin}\{\rho \in \mathcal{P} \mid \rho > \epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]})\}, \quad (12)$$

**TABLE I:** Formulas of the IVA cost function, the augmented Lagrange function, and its gradient.

$$\mathcal{J}_{\text{IVA}}(\mathbf{W}, \boldsymbol{\Sigma}) = \frac{NK}{2} \log(2\pi) + \frac{1}{2} \sum_{n=1}^N \log |\det(\boldsymbol{\Sigma}_n)| + \frac{1}{2} \sum_{n=1}^N \sum_{k,l=1}^K (\mathbf{e}_k^\top \boldsymbol{\Sigma}_n^{-1} \mathbf{e}_l) (\mathbf{w}_n^{[k]})^\top \left( \frac{1}{V} \mathbf{X}^{[k]} (\mathbf{X}^{[l]})^\top \mathbf{w}_n^{[l]} - \sum_{k=1}^K \log |\det(\mathbf{W}^{[k]})| \right) \quad (3)$$

$$\mathcal{L}_{\gamma, \boldsymbol{\rho}}(\mathbf{W}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) = \mathcal{J}_{\text{IVA}}(\mathbf{W}, \boldsymbol{\Sigma}) + \frac{1}{2\gamma} \sum_{n=1}^M \sum_{k=1}^K \left( \left( \max(0, \mu_n^{[k]} + \gamma(\rho_n^{[k]} - \epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]}))) \right)^2 - (\mu_n^{[k]})^2 \right) \quad (4)$$

$$\frac{\partial \mathcal{L}_{\gamma, \boldsymbol{\rho}}}{\partial \mathbf{w}_n^{[k]}} = \sum_{l=1}^K \left( \frac{1}{V} \mathbf{X}^{[k]} (\mathbf{X}^{[l]})^\top \mathbf{w}_n^{[l]} \mathbf{e}_l^\top \boldsymbol{\Sigma}_n^{-1} \mathbf{e}_k - \frac{\mathbf{d}_n^{[k]}}{(\mathbf{d}_n^{[k]})^\top \mathbf{w}_n^{[k]}} - \mathbb{I}_{n \leq M} \max(0, \mu_n^{[k]} + \gamma(\rho_n^{[k]} - \epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]}))) \right) \mathbf{X}^{[k]} \frac{\partial \epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]})}{\partial \mathbf{y}_n^{[k]}} \quad (5)$$

**Algorithm 1: Adaptive-Reverse Constrained IVA (ar-cIVA)**
**Input:**  $\{\mathbf{X}^{[k]}\}_{k=1}^K \subset \mathbb{R}^{N \times V}$ ,  $\{\mathbf{r}_n\}_{n=1}^M \subset \mathbb{R}^V$ ,  $\gamma$ ,  $\mu_{\max}$ 
**Output:**  $\mathbf{W}, \boldsymbol{\Sigma}$ 

```

1: Set the current scheme to (12)
2: repeat
3:   for  $n = 1, \dots, N$  do
4:     for  $k = 1, \dots, K$  do
5:       Compute  $\hat{\boldsymbol{\Sigma}}_n^{-1}$ 
6:       Update  $\mu_n^{[k]}$  using (15) and (16)
7:       if  $\mu_n^{[k]} \geq \mu_{\max}$  then
8:         Switch the current scheme to (13)
9:       else if  $\mu_n^{[k]} \leq 0$  then
10:        Switch the current scheme to (12)
11:      else
12:        Keep the current scheme
13:        Select  $\rho_n^{[k]}$  based on the current scheme
14:        Compute  $d\mathbf{w}_n^{[k]} = \partial \mathcal{L}_{\gamma, \boldsymbol{\rho}} / \partial \mathbf{w}_n^{[k]}$  using (5)
15:        Project  $\tilde{\mathbf{d}}\mathbf{w}_n^{[k]} = (\mathbf{I}_n - \mathbf{w}_n^{[k]} (\mathbf{w}_n^{[k]})^\top) d\mathbf{w}_n^{[k]}$ 
16:        Update  $\mathbf{w}_n^{[k]} = \mathbf{w}_n^{[k]} - \eta \frac{d\mathbf{w}_n^{[k]}}{\|\tilde{\mathbf{d}}\mathbf{w}_n^{[k]}\|}$ 
17:        Normalize  $\mathbf{w}_n^{[k]} = \frac{\mathbf{w}_n^{[k]}}{\|\mathbf{w}_n^{[k]}\|}$ 
18:        Update  $[\hat{\boldsymbol{\Sigma}}_n]_{kl}$  for  $l = 1, \dots, K$ 
19: until convergence

```

and (ii) choosing the largest value that satisfies the constraint

$$\rho_n^{[k]} = \text{argmax}\{\rho \in \mathcal{P} \mid \rho \leq \epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]})\}. \quad (13)$$

On the one hand, (12) creates an over-tight constraint that forces the value of  $\epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]})$  to increase after each iteration. On the other hand, (13) creates a feasible problem where each constraint is always satisfied. Using the appropriate principle at each iteration, the desired value of the threshold — that is close to  $\epsilon(\mathbf{r}_n, \mathbf{s}_n^{[k]})$  — can be recovered.

### A. Augmented Lagrangian Method with Decoupling

To solve (11) as an inequality-constrained optimization, we utilize the augmented Lagrangian method and a decoupling method that enables sequential updates of each row of individual demixing matrices. The augmented Lagrangian function is given in (4), where  $\boldsymbol{\mu} \in \mathbb{R}^{M \times K}$  is the Lagrange multiplier and  $\gamma > 0$  is the scalar penalty parameter as in [32] where the framework is used for ICA. It can be shown [33] that for sufficiently large  $\gamma$ , the solution of (4) coincides with the

solution of (11). At the  $i$ th iteration, we update the parameters to minimize  $\mathcal{L}_{\gamma, \boldsymbol{\rho}}$  based on their current values as follows

$$(\mathbf{W}^{i+1}, \boldsymbol{\Sigma}^{i+1}) = \text{argmin}_{\mathbf{W}, \boldsymbol{\Sigma}} \mathcal{L}_{\gamma, \boldsymbol{\rho}}(\mathbf{W}, \boldsymbol{\Sigma}, \boldsymbol{\mu}^i), \quad (14)$$

$$(\alpha_n^{[k]})^{i+1} = (\mu_n^{[k]})^i + \gamma((\rho_n^{[k]})^i - \epsilon(\mathbf{r}_n, (\mathbf{y}_n^{[k]})^{i+1})), \quad (15)$$

$$(\mu_n^{[k]})^{i+1} = \max(0, (\alpha_n^{[k]})^{i+1}), \quad (16)$$

where  $\mu_n^{[k]}$  is the  $(n, k)$ -entry of  $\boldsymbol{\mu}$ . In (14), the value of  $\boldsymbol{\Sigma}_n$  that minimizes  $\mathcal{L}_{\gamma, \boldsymbol{\rho}}(\cdot)$  is given by

$$\hat{\boldsymbol{\Sigma}}_n = \frac{1}{V} \sum_{v=1}^V \mathbf{y}_n(v) \mathbf{y}_n(v)^\top = \text{argmin}_{\boldsymbol{\Sigma}_n} \mathcal{J}_{\text{IVA}}(\mathbf{W}, \boldsymbol{\Sigma}). \quad (17)$$

Additionally, to update  $\mathbf{W}$ , we utilize the vector gradient method IVA-G-V in [12] and derive the gradient of the augmented Lagrange function as follows. First, we rewrite the term  $\log |\det \mathbf{W}^{[k]}|$  in (3) as the sum of two terms  $\log |(\mathbf{d}_n^{[k]})^\top \mathbf{w}_n^{[k]}| + \log (\det(\tilde{\mathbf{W}}_n^{[k]} (\tilde{\mathbf{W}}_n^{[k]})^\top)) / 2$  where  $\tilde{\mathbf{W}}_n^{[k]}$  is the  $(N-1) \times N$  matrix obtained by removing the  $n$ th row from  $\mathbf{W}^{[k]}$  and  $\mathbf{d}_n^{[k]} \in \mathbb{R}^N$  satisfies  $\tilde{\mathbf{W}}_n^{[k]} \mathbf{d}_n^{[k]} = \mathbf{0}_{N-1}$ . This technique [34] is often referred to as the *decoupling trick*, enabling the derivation of the gradient of  $\mathcal{J}_{\text{IVA}}$  w.r.t.  $\mathbf{w}_n^{[k]}$

$$\frac{\partial \mathcal{J}_{\text{IVA}}}{\partial \mathbf{w}_n^{[k]}} = \frac{1}{V} \sum_{v=1}^V \mathbf{x}^{[k]}(v) (\mathbf{y}_n(v))^\top \boldsymbol{\Sigma}_n^{-1} \mathbf{e}_k - \frac{\mathbf{d}_n^{[k]}}{(\mathbf{d}_n^{[k]})^\top \mathbf{w}_n^{[k]}}.$$

The advantage of this decoupling procedure is that one can avoid the dependence on the number of samples by recognizing that  $\mathbf{y}_n^{[k]}(v) = (\mathbf{w}_n^{[k]})^\top \mathbf{x}_n^{[k]}(v)$  and pre-computing the sample covariance matrix  $\hat{\mathbf{R}}_x^{kl} = \frac{1}{V-1} \mathbf{X}^{[k]} (\mathbf{X}^{[l]})^\top = \frac{1}{V-1} \sum_t \mathbf{x}^{[k]}(v) (\mathbf{x}^{[l]}(v))^\top$ . Second, summing the gradient of the cost function and the gradient of the constraint, we obtain the gradient of the augmented Lagrange function  $\mathcal{L}_{\gamma, \boldsymbol{\rho}}$  w.r.t to  $\mathbf{w}_n^{[k]}$  in (5), for  $n = 1, \dots, N$  and  $k = 1, \dots, K$ . In this formula,  $\mathbb{I}_{n \leq M}$  is the indicator of the event  $n \leq M$ . When the Pearson correlation is used as the similarity measure, the last term in (5) can be further simplified and can be computed independent of the sample size  $V$ . Finally, certain refinements as suggested in [35] (e.g., projecting the gradient onto the tangent space to the unit sphere, normalizing the gradient norm, and projecting the demixing vector back onto the unit sphere) are incorporated into the algorithm.

### B. Adaptive-Reverse Scheme for Constraint Thresholds

We can now define the adaptive-reverse scheme for selecting the values of  $\rho_n^{[k]}$ . In (15) and (16), we note that the value of  $\mu_n^{[k]}$  increases when the constraint  $\epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]}) \geq \rho_n^{[k]}$  is

violated and decreases when the constraint holds. Therefore, using the argmin scheme (12) will increase the value of the Lagrange multipliers toward  $+\infty$  while using the argmax scheme (13) will decrease their values toward 0. To combine the advantage of both approaches, we propose an adaptive-reverse scheme that determines the scheme based on the values of the Lagrange multipliers. In particular, when  $\mu_n^{[k]}$  exceeds a certain value  $\mu_{\max}$ , we switch from the argmin scheme to the argmax scheme. Conversely, when  $\mu_n^{[k]}$  goes down to 0, we switch from the argmax scheme to the argmin scheme. We summarize the adaptive-reverse for constrained IVA (ar-cIVA) in Algorithm 1.

#### IV. THRESHOLD-FREE CONSTRAINED IVA

In this section, we propose a novel formulation of constrained IVA that eliminates the need for threshold parameters. Our idea is to maximize the similarity between the reference  $\mathbf{r}_n$  and the corresponding estimated source component (corresponding-component), and at the same time, promote the dissimilarity between that reference  $\mathbf{r}_n$  and the other estimated component  $\mathbf{y}_m^{[k]}$  (cross-component), for all  $m \neq n$ . Thus, we introduce a regularization term

$$\mathcal{J}_{\text{ref}}(\mathbf{W}) = \sum_{n=1}^M \sum_{k=1}^K \left( \sum_{\substack{m=1 \\ m \neq n}}^M \epsilon^2(\mathbf{r}_n, \mathbf{y}_m^{[k]}) - \epsilon^2(\mathbf{r}_n, \mathbf{y}_n^{[k]}) \right). \quad (18)$$

The new objective function is hence a linear sum of the IVA cost function and the regularization

$$\mathcal{L}_{\lambda}(\mathbf{W}, \Sigma) = \mathcal{J}_{\text{IVA}}(\mathbf{W}, \Sigma) + \frac{\lambda}{2} \mathcal{J}_{\text{ref}}(\mathbf{W}), \quad (19)$$

where  $\lambda > 0$  is the regularization parameter. By selecting an appropriate value for  $\lambda$  (via parameter tuning), we can balance the trade-off between the IVA cost (minimizing the correlation between the source components) and the regularization term (maximizing the correlation between the components and the reference signals). Our formulation in (19) is similar to the multi-objective function optimization framework in [36]. In their work, Du and Fan introduced an improved version of constrained ICA by optimizing two conflicting cost functions: one that maximizes the independence among the components and one that maximizes the closeness between the components and their corresponding references. Nonetheless, compared with the approach in [36], our proposed method not only generalizes constrained ICA to constrained IVA but also introduces the cross-component similarity to the objective function. This promotes the solution in which there is one and only one independent component that is closest to each reference.

From (18), the gradient of  $\mathcal{J}_{\text{ref}}(\cdot)$  w.r.t.  $\partial \mathbf{w}_n^{[k]}$ , for  $n = 1, \dots, N$  and  $k = 1, \dots, K$ , is given by

$$\begin{aligned} \frac{\partial \mathcal{J}_{\text{ref}}}{\partial \mathbf{w}_n^{[k]}} = 2\mathbb{I}_{n \leq M} \left( \sum_{\substack{m=1 \\ m \neq n}}^M \epsilon(\mathbf{r}_m, \mathbf{y}_m^{[k]}) \frac{\partial \epsilon(\mathbf{r}_m, \mathbf{y}_n^{[k]})}{\partial \mathbf{w}_n^{[k]}} \right. \\ \left. - \epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]}) \frac{\partial \epsilon(\mathbf{r}_n, \mathbf{y}_n^{[k]})}{\partial \mathbf{w}_n^{[k]}} \right). \quad (20) \end{aligned}$$

---

#### Algorithm 2: Threshold-Free Constrained IVA (tf-cIVA)

---

**Input:**  $\{\mathbf{X}^{[k]}\}_{k=1}^K \subset \mathbb{R}^{N \times V}, \{\mathbf{r}_n\}_{n=1}^M \subset \mathbb{R}^V, \lambda$

**Output:**  $\mathbf{W}, \Sigma$

```

1: repeat
2:   for  $n = 1, \dots, N$  do
3:     for  $k = 1, \dots, K$  do
4:       Compute  $\hat{\Sigma}_n^{-1}$ 
5:       Compute  $d\mathbf{w}_n^{[k]} = \partial \mathcal{L}_{\lambda} / \partial \mathbf{w}_n^{[k]}$  based on (20)
6:       Update  $\mathbf{w}_n^{[k]}$  using  $d\mathbf{w}_n^{[k]}$ 
7:       Update  $[\hat{\Sigma}_n]_{kl}$  for  $l = 1, \dots, K$ 
8:   until convergence

```

---

Thus, the gradient of  $\mathcal{L}_{\lambda}(\cdot)$  is the sum of  $\partial \mathcal{J}_{\text{IVA}} / \partial \mathbf{w}_n^{[k]}$  and  $\lambda \partial \mathcal{J}_{\text{ref}} / \partial \mathbf{w}_n^{[k]}$ . The vector-gradient method to minimize  $\mathcal{L}_{\lambda}(\cdot)$ , named tf-cIVA, is described in Algorithm 2. Compared with the thresholded formulation for constrained IVA in (11), the regularized formulation does not require threshold parameters as well as other hyperparameters for the augmented Lagrange method (i.e.,  $\gamma$  and  $\mu_{\max}$ ).

#### V. IMPLEMENTATION AND EVALUATION

**Compared Methods.** We compare our new algorithms against the following methods: IVA-G-V [12] for unconstrained IVA, cIVA-fixed for constrained IVA with fixed threshold [31], and its adaptive thresholding version (pt-cIVA) [19]. The IVA-G-V algorithm for the unconstrained problem can be viewed as a baseline where no prior knowledge about the sources is used. For pt-cIVA, we note that the proposed version in [19] uses MLD for the SCVs, which is significantly slower than MGD. Indeed, the IVA-L-SOS versions do not finish within 2 weeks while the IVA-G versions run for a few hours, using the same setting in our simulation. Therefore, we reimplement pt-cIVA with MGD to make its computational time comparable with other methods. In addition, we use the set of pre-defined thresholds  $\mathcal{P}_{\text{pt}} = \{0.001, 0.1, 0.2, \dots, 0.9\}$  and the penalty parameter  $\gamma = 3$  as specified by the authors in [19]. For ar-cIVA, we use a finer set of pre-defined thresholds  $\mathcal{P}_{\text{ar}} = \{0.01, 0.02, \dots, 0.99\}$ .<sup>2</sup> Furthermore, we set the penalty parameter  $\gamma = 100$  and the cut-off value for the Lagrange multiplier  $\mu_{\max} = 1$ . For tf-cIVA, we use a grid-search approach to select the value of the regularization parameter over the logarithmic scale  $\{0.1, 1, 5, 10, 100, 1000, 10000\}$  to check the sensitivity of the algorithm with respect to the choice of  $\lambda$ . After evaluating of values for  $\lambda$  and observing the performance is stable for sufficiently large values, we pick the ones that yield the best results to report ( $\lambda = 1$  for the simulated data and  $\lambda = 100$  for the real data). All algorithms use the same initial step size  $\eta$  for the gradient updates and a decay scheme that decreases  $\eta$  by a factor of 0.95 when the objective function does not decrease at a certain iteration. In addition, the stopping criteria for all algorithms are based on

<sup>2</sup>We also tried the finer set of thresholds for pt-cIVA but there was no significant difference. Hence, we present the results with default options for pt-cIVA in this work.

the change in  $\mathbf{W}$  at each iteration [12]

$$\max_{k,n} \left\{ 1 - \left| (\mathbf{w}_n^{[k]})^i \top (\mathbf{w}_n^{[k]})^{i+1} \right| \right\} < \epsilon,$$

where  $\epsilon = 10^{-6}$  throughout this work.

**Evaluation metric.** To evaluate the performance of different IVA algorithms, we use the following metrics:

- Joint inter-symbol-interference (joint-ISI) is introduced in [12], which assumes that the ground-truth mixing matrix  $\mathbf{A}^{[k]}$  for the  $k$ th dataset is known. This is an extension of the normalized inter-symbol-interference (ISI) in the context of ICA [37], [38]. Let  $\mathbf{G}^{[k]} = \mathbf{W}^{[k]} \mathbf{A}^{[k]}$ , for  $k = 1, \dots, K$ , be the global demixing-mixing matrices and  $|\mathbf{G}^{[k]}|$  be the absolute matrix with the  $(m, n)$  entry being  $|G_{mn}^{[k]}|$ . The joint-ISI is defined as the ISI of the mean absolute value matrix  $\mathbf{G} = 1/K \sum_{k=1}^K |\mathbf{G}^{[k]}|$ , i.e.,  $\text{joint-ISI}(\mathbf{G}^{[1]}, \dots, \mathbf{G}^{[K]}) = \text{ISI}(\mathbf{G})$ , where

$$\text{ISI}(\mathbf{G}) = \frac{\sum_{i=1}^N \left( \frac{\sum_{j=1}^N |G_{ij}|}{\max_p |G_{ip}|} - 1 \right) + \sum_{j=1}^N \left( \frac{\sum_{i=1}^N |G_{ij}|}{\max_p |G_{pj}|} - 1 \right)}{2N(N-1)}.$$

When the sources are jointly separated for all datasets, the estimated global matrices for all datasets should be close to an identity matrix up to the same permutation. Thus, the joint-ISI closer to 0 indicates better performance.

- Cross joint inter-symbol-interference (cross-joint-ISI) measures the consistency of the components across  $R$  runs. Let  $\mathbf{W}_r^{[k]}$  be the  $k$ th demixing matrix of the  $r$ th run. In [39], the cross-joint-ISI of the  $i$ th run and the  $j$ th run is defined as

$$\text{cross-joint-ISI}_{ij}(\{\mathbf{W}_r^{[k]}\}_{r=1, k=1}^{R, K}) = \text{joint-ISI}(\mathbf{P}_{ij}^{[1]}, \dots, \mathbf{P}_{ij}^{[K]}),$$

where  $\mathbf{P}_{ij}^{[k]} = \mathbf{A}_i^{[k]} \mathbf{W}_j^{[k]}$  and  $\mathbf{A}_i^{[k]} = (\mathbf{W}_i^{[k]})^{-1}$ . The cross-joint-ISI of the  $i$ th run is computed by averaging all its pairwise cross-joint-ISI values

$$\text{cross-joint-ISI}_i = \frac{1}{R} \sum_{j=1, j \neq i}^R \text{cross-joint-ISI}_{ij}.$$

Note that cross-joint-ISI can be computed when there is no ground truth available as it only depends on the demixing matrices. On the other hand, joint-ISI requires the true demixing matrices in its evaluation.

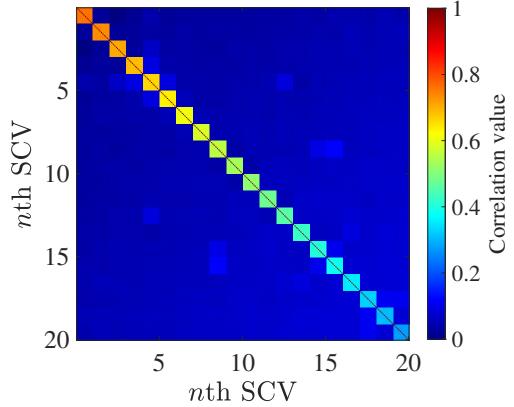
- Partial similarity factor (partial SF) measures the average of the squares of the correlation between the estimated source and the corresponding ground truth:

$$\text{Partial SF} = \left( \frac{1}{MK} \sum_{n=1}^M \sum_{k=1}^K (\epsilon(\mathbf{s}_n^{[k]}, \mathbf{y}_n^{[k]}))^2 \right)^{1/2}.$$

The index  $n$  runs from 1 to  $M$ , meaning that only source components with corresponding reference signals are used. A lower value of this metric indicates poor estimation of the sources as well as poor source alignment across the datasets.

## VI. HYBRID SIMULATION RESULTS

This section compares the performance of the two proposed methods with the three aforementioned IVA algorithms using simulated fMRI-like data by changing the number of subjects and reference signals. Our goal is to better understand



**Fig. 1:** A visualization of the full SCV covariance matrix using model (21) with  $N = 20, K = 20, V = 58515, \mu_0 = 0.1, \mu_1 = 0.2$ , and  $\boldsymbol{\varphi} = [\varphi_1, \varphi_2, \dots, \varphi_N]^\top$  as a linearly spaced vector in the range  $[0.3, 0.9]$ . From (21), one can compute the values of the 20 diagonal blocks in this matrix, ranging from 0.28 (bottom-right) to 0.76 (top-left).

the behavior of these algorithms in different types of fMRI datasets. In the next section, we demonstrate an application with a practical fMRI dataset.

**Extraction of reference signals.** We use reference signals extracted by NeuroMark, i.e., the Neuromark\_fMRI\_1.0 template [40], which includes 20 fMRI networks and is divided into seven functional domains based on their anatomical and functional properties: the subcortical (SC), auditory (AUD), sensorimotor (MOT), visual (VIS), cognitive control (CC), default mode (DMN) and cerebellar (CB) domains.<sup>3</sup> For convenience, we denote the set of  $N = 20$  reference signals by  $\{\mathbf{r}_n\}_{n=1}^N$ , each contains  $V = 58515$  samples  $\mathbf{r}_n = [r_n(1), r_n(2), \dots, r_n(V)]^\top$ . In addition, each reference signal is normalized to zero mean and unit variance. Finally, we note that there is a certain level of dependency among the reference signals, i.e., they are not absolutely independent.

**Hybrid source generation.** Given the reference signals, we generate observations of SCVs  $\{\mathbf{S}_n\}_{n=1}^N \subset \mathbb{R}^{K \times V}$  for  $K$  subjects as follows. First, we define a  $NK$ -dimensional random vector  $\mathbf{z}$  following multivariate Gaussian distribution with zero mean and covariance matrix

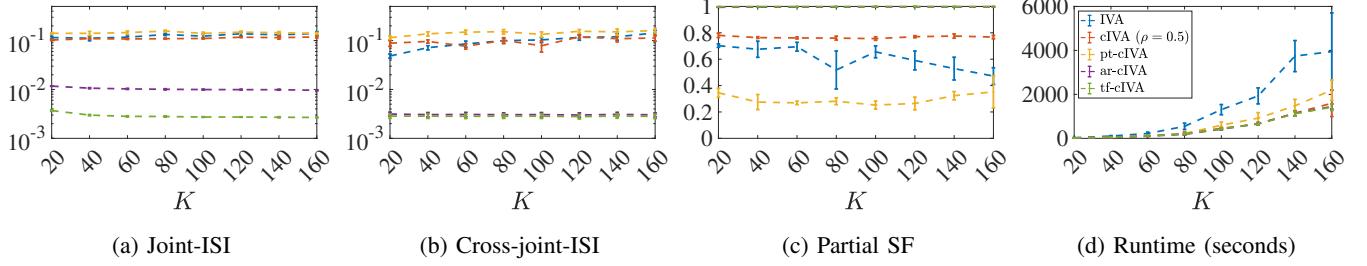
$$\boldsymbol{\Sigma}_z = (\mu_0 \mathbf{1}_N \mathbf{1}_N^\top + (\mu_1 - \mu_0) \mathbf{I}_N) \otimes \mathbf{1}_K \mathbf{1}_K^\top + (1 - \mu_1) \mathbf{I}_{NK},$$

where  $0 \leq \mu_0 \leq \mu_1 \leq 1$ . Second, we generate  $V$  samples of  $\mathbf{z}$  and partitioning the data matrix into  $N$  submatrices of dimension  $K \times V$ , i.e.,  $\mathbf{Z} = [\mathbf{Z}_1^\top, \dots, \mathbf{Z}_N^\top]^\top$ . Third, the  $n$ th source data matrix is formed by

$$\mathbf{S}_n = \sqrt{1 - \varphi_n^2} \mathbf{1}_K \mathbf{r}_n^\top + \varphi_n \mathbf{Z}_n \in \mathbb{R}^{K \times V}, \quad (21)$$

where  $\varphi_n \in [0, 1]$  controls how close the  $n$ th source is to the reference  $\mathbf{r}_n$ . Figure 1 depicts the SCV covariance matrices of the simulated fMRI-like data used in this section. Specifically,

<sup>3</sup>The original template contains a total of 53 references. In the hybrid simulation experiment with varying numbers of subjects, to reduce the runtime, we only use a subset of 20 references (with 2 references from AU and 3 references from each of the other 6 functional domains).



**Fig. 2:** Comparison of five different IVA algorithms for the simulated fMRI-like data with  $M = N = 20$  and  $V = 58515$ . The (a) joint-ISI, (b) cross-joint-ISI, (c) partial similarity factor, and (d) runtime are shown as functions of the number of subjects  $K$ . The error bars represent one standard deviation calculated over 20 runs. For each value of  $K$ , the same mixing matrix is used to generate the simulated fMRI-like data across 20 runs. All algorithms use the same initialization for each run. Note that in all four plots, the purple line and the green line almost overlap and in plot (d), the red line also overlaps these two.

we choose  $\varphi_n$  to control the correlation within each SCV varying in the range  $[0.3, 0.9]$ . While we think that this is the most interesting scenario with different levels of correlation within each SCV, we also tried other scenarios where the correlation within each SCV is consistently low (or high) and obtained similar conclusions. Due to the space limitation, we did not include these results here.

**Results.** We evaluate the performance of the five aforementioned IVA algorithms with regard to the changes in (i) the number of subjects  $K$  and (ii) the number of reference signals. In the first experiment, the number of subjects  $K$  is varied while the numbers of samples  $V$ , source components  $N$ , and reference signals  $M$  are fixed. As can be seen from Fig. 2, our two proposed algorithms, tf-cIVA (dashed green line) and ar-cIVA (dashed purple line), significantly outperform other algorithms in terms of joint-ISI, cross-joint-ISI, and partial SF. It is noteworthy that tf-cIVA slightly outperforms ar-cIVA since the former considers both maximizing the corresponding-component similarity and minimizing the cross-component similarity, while the latter only looks at the corresponding-component similarity. As the number of subjects  $K$  increases, the cross-joint-ISI of the unconstrained IVA (the blue dashed line in Fig. 2-b) increases while the partial SF (Fig. 2-c) decreases, indicating that this method becomes less reproducible in large-scale settings. This degradation is also noted in [19] as the curse of dimensionality in IVA. The same phenomenon is also observed for pt-cIVA (yellow dashed line), which verifies our earlier discussion on the conservative nature of this method in selecting thresholds. Interestingly, almost independent of the number of subjects, tf-cIVA and ar-cIVA yield consistently excellent performance. This is highlighted by the fact that our algorithms exploit both HOS (by effectively imposing constraints with reference signals) and SOS (via the Gaussian source model). In terms of runtime, Fig. 2-d shows that cIVA, ar-cIVA, and tf-cIVA are the fastest algorithms while IVA is the slowest.

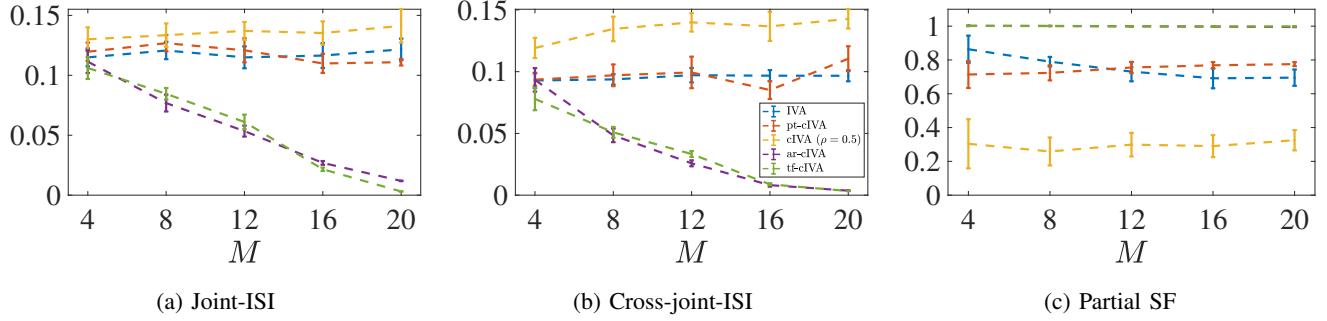
In the second experiment, we vary the number of reference signals  $M$  while fixing the number of components  $N$ , the number of subjects  $K$ , and the number of samples  $V$ . The performance of the five aforementioned algorithms is shown in Fig. 3. Overall, the two proposed algorithms, ar-cIVA (purple

dashed line) and tf-cIVA (green dashed line) yield the highest partial similarity factors that are almost equal to 1 across different values of  $M$ , indicating that the components corresponding to the  $M$  references (“constrained” components) are estimated accurately. On the other hand, their overall performances in terms of joint-ISI and cross-joint-ISI improve significantly over other algorithms as  $M$  increases, indicating the important role of the number of references used relative to the actual number of source components. For unconstrained IVA (blue dashed line), its performance is independent of  $M$  as no references are used. The performance of cIVA with  $\rho = 0.5$  (red dashed line) also improves as  $M$  increases from 4 to 16 references. However, when the number of references equals the number of components, we observe a slight decrease in its performance. This is because the correlation between the source and the reference signal varies in the range  $[0.31, 0.81]$  across components. Thus, for some components, the threshold  $\rho = 0.5$  cannot be satisfied. The pt-cIVA algorithm performs worst in terms of joint-ISI, cross-joint-ISI, and partial SF.

## VII. MULTI-SUBJECT FMRI DATA ANALYSIS

This section evaluates the performance of the different algorithms on real fMRI data. Our goal is to demonstrate that the proposed methods offer better model matches and more interpretable results.

**Data acquisition and preprocessing.** We use the resting state fMRI data set from the bipolar-schizophrenia network on intermediate phenotypes (B-SNIP) [41], [42]. Identical diagnostic and recruitment approaches were applied to all recruited subjects at multiple sites (Baltimore, Chicago, Dallas, Detroit, and Hartford). In particular, in this study, we employ the data collected from the Baltimore site and select  $K = 98$  subjects: 49 healthy controls (HCs) and 49 randomly selected schizophrenia patients (SZs). A single 5-minute run was captured for each subject. The individuals involved in the study were instructed to maintain an open-eyed state, concentrate on a crosshair presented on a display screen, and remain still throughout the scanning process. Moreover, a custom-built head-coil cushion was used to restrict head movements. Alertness during the scan was confirmed immediately afterward, and the procedure was repeated if needed.



**Fig. 3:** Comparison of five different IVA algorithms for the simulated fMRI-like data with  $N = 20$ ,  $K = 40$ , and  $V = 58515$  as the number of reference signals  $M$  increases. The error bars represent one standard deviation calculated over 20 runs. For each value of  $R$ , the mixing matrix remains the same across 20 runs, while the hybrid sources are generated independently for each run. All algorithms use the same initialization for each run. Note that in plot (c), the purple line and the green line almost overlap.

These instructions helped reduce head motion and prevented subjects from falling asleep. The fMRI data were captured by a 3-Tesla Siemens Triotim scanner with  $TE = 30$  ms,  $TR = 2.21$  s, flip angle =  $70^\circ$ , acquisition matrix =  $64 \times 64$  mm, and voxel size =  $3.4 \times 3.4 \times 3$  mm $^3$ . For each subject, 134 time points were obtained. We removed the first 3 time points to address the T-1 effect and each subject's image data was preprocessed including motion correction and slice-time correction. The corrected data were warped into the standard Montreal Neurological Institute (MNI) space through an echo-planar imaging template and then were resampled to  $3 \times 3 \times 3$  mm $^3$  isotropic voxels. The resampled fMRI data were further smoothed using a Gaussian kernel with a full width at half maximum (FWHM) equal to 6 mm. In addition, in order to remove non-brain voxels and flatten the data, each subject image was masked, yielding an observation vector of  $V = 58515$  voxels for each of the  $T = 131$  time points.

**Results.** We evaluate the performance of the five aforementioned IVA algorithms when applied to real fMRI data. For this purpose, as in the experiments developed for simulated fMRI-like data, we use the functional templates extracted by Neuromark, specifically the neuromark\_fMRI\_1.0 template [40], which is composed of 53 resting-state networks (RSNs) from seven different functional domains: SC (5 RSNs), AUD (2 RSNs), MOT (9 RSNs), VIS (9 RSNs), CC (17 RSNs), DMN (7 RSNs) and CB (4 RSNs). Each of these RSNs is employed as a reference signal by the IVA algorithms. In addition, since fMRI data might also contain signals not of interest such as motion-related signals, scanner-related signals, or noise due to magnetic resonance acquisition, among others, we utilize a larger number of components than reference signals to capture the interferences and components that would not necessarily match to a specific functional template. Without the flexibility to have more components than references ( $N > M$ ), the estimated sources might be noisy and inaccurate. In our experiments, we noted  $N = 60$  as a good number that balances the trade-off between the model complexity and flexibility.

The obtained results by the different IVA algorithms are shown in Fig. 4. The cross-joint-ISI values for 50 independent

runs are depicted in Fig. 4-a. Unconstrained IVA and pt-cIVA present higher values than the rest of the algorithms, achieving less consistent results across all the runs. It can also be seen that the fixed-threshold scheme, cIVA with  $\rho = 0.5$ , is the second-best algorithm in terms of cross-joint-ISI. However, if the threshold value is modified, cIVA with  $\rho = 0.3$ , its performance decreases, making it clear that the correct selection of a threshold value is of paramount importance for the performance of these algorithms. On the other hand, we can also observe that the proposed adaptive-reverse scheme, ar-cIVA, also offers low cross-joint-ISI values with small variance across runs and significantly outperforms the previously proposed adaptive cIVA algorithm (pt-cIVA). We would like to highlight that ar-cIVA selects the thresholds automatically across subjects and components while cIVA with a fixed  $\rho$  requires a careful selection of threshold values. Finally, we can see that the proposed algorithm tf-cIVA outperforms the rest of the algorithms, achieving the most consistent results with the lowest cross-joint-ISI values. The spatial maps of three different RSNs (AUD, DMN and VIS) obtained by the most consistent run of this algorithm (tf-cIVA) are shown in Fig. 4-c as an example.

Another useful measure to quantify the quality of estimation of the fMRI components is the power spectra of RSN time courses and the power ratio between low-frequency ( $< 0.1$  Hz) and high-frequency ( $> 0.15$  Hz) bands. Considering the frequencies of neural-activity related BOLD signals are generally below 0.15 Hz, low power ratio values are typically associated with cardiac and respiratory noise, while high power ratio values mostly indicate BOLD activity [43]. The power ratio for the most consistent run is depicted in Fig. 4-b. It is important to note that in the case of unconstrained IVA, only 26 out of the 60 estimated components are selected as meaningful after inspecting their spatial maps and power spectra values, hence its power ratio results are omitted in Fig. 4-b. In addition, for the sake of a clearer comparison of the constrained algorithms, three outliers from pt-cIVA (with power ratio values of 22.59, 26.92, and 29.57) were removed. For the other algorithms, the 53 estimated components related to the reference signals are

taken into account for their assessment. As we can observe, the proposed algorithm tf-cIVA achieves the highest median power ratio value (red line within each boxplot) followed by ar-cIVA. Since high power ratio values are associated with higher BOLD activity, the higher values obtained by tf-cIVA indicate better performance in the estimation of the components.

For a more detailed analysis of the algorithms, we also evaluate the correlations between components' time courses, i.e., the functional network connectivity (FNC) maps. The patterns revealed in FNC are consistent with known functional network relationships, for example, we note the anticorrelation between the DMN components and sensory-related networks (e.g., MOT and VIS) and the positive correlation within each functional domain [43]. The aggregated FNC matrices for HC subjects and the most consistent run for the three constrained algorithms with the highest median power ratio are shown in the lower diagonal of Fig. 5. The FNC obtained by the proposed algorithm tf-cIVA shows higher contrast than the rest of the algorithms. In particular, we can see a positive correlation between functional domains such as sensorimotor and visual, and a negative correlation between DMN and sensory-related networks. On the other hand, the FNC matrices obtained by the other algorithms do not show such a clear pattern, where the correlation value is, for most of the RSNs, closer to zero. The upper diagonal of Fig. 5 presents the difference mean FNC between the two groups (HC - SZ). We can appreciate that tf-cIVA shows stronger differences between the two groups than the other algorithms, where HCs show higher connectivity than SZ within MOT and VIS, consistent with recent studies [44], [45]. In addition, we also conduct a statistical analysis to compare the performance of the algorithms. For this purpose, a paired t-test of the HCs FNCs matrices is performed, and the results obtained by tf-cIVA are compared with those obtained by the fixed-threshold algorithm with the highest power ratio, cIVA ( $\rho = 0.3$ ), and with the two other adaptive algorithms, i.e., pt-cIVA and ar-cIVA. The resulting T-value maps are shown in Fig. 6, where the upper diagonal presents the networks with significant connectivity differences between the compared algorithms after the false discovery rate (FDR) correction of the p-values ( $< 0.05$ ) [46]. As previously mentioned, we expect the FNC to present high and positive correlation within each functional domain (the diagonal block). In this regard, it can be seen that tf-cIVA exhibits higher connectivity values within functional domains such as VIS, MOT, DMN, CC, or CB. It should be noted that, as illustrated in Fig. 6-c, ar-cIVA achieves comparable FNCs to tf-cIVA. Nonetheless, higher connectivity values are still observable in the diagonal block. Hence, the obtained FNCs and T-maps suggest that tf-cIVA provides a better model match to fMRI data and therefore more interpretable results [47].

We further analyze the group differences in FNCs between HC and SZ by applying a two-sample t-test to the results obtained by the algorithms. After FDR correction of the p-values, tf-cIVA is the only algorithm showing more than one significant connectivity difference between groups, i.e., cIVA ( $\rho = 0.5$ ) and ar-cIVA do not show any significant difference, while cIVA ( $\rho = 0.3$ ) and pt-cIVA show significant differences only between two RSNs. Fig. 7 demonstrates the mean FNC of

the networks that have a significant difference between groups after FDR correction of the p-values. The results show that SZ patients present weaker connectivity than HC within MOT and VIS domains. Also, the SZ group shows lower connectivity between domains such as MOT and AUD, and MOT and VIS. Some additional group differences can be observed between CB and MOT and VIS, and also between SC and MOT and VIS RSNs, consistent with recent studies [40], [44], [45]. These group differences observed in tf-cIVA results increase our confidence that the proposed algorithm results in a better model match and performance in preserving subject variability.

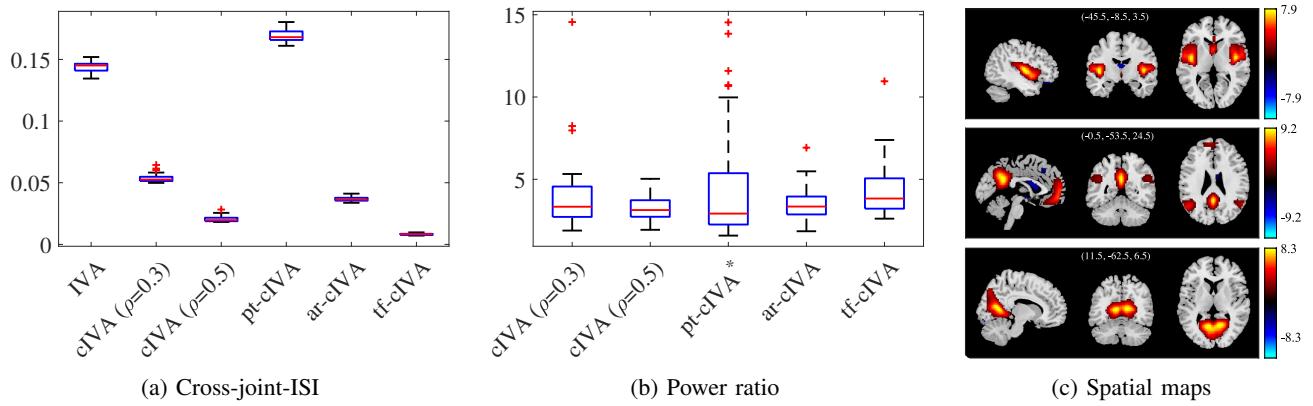
## VIII. DISCUSSION

We proposed two novel approaches for constrained IVA that alleviate the need for pre-specified thresholds, thus significantly increasing their utility for fMRI data analysis. We demonstrated that these methods yield fully interpretable network estimates and can effectively capture HOS, even though they are implemented with a multivariate Gaussian model. The multivariate Gaussian implementation along with the use of an effective constraint framework hence enables achieving a desirable balance between performance and computational complexity. An additional advantage of the constrained approach is that the permutation ambiguity of ICA/IVA is alleviated and post-analysis and sorting of components becomes a much easier task.

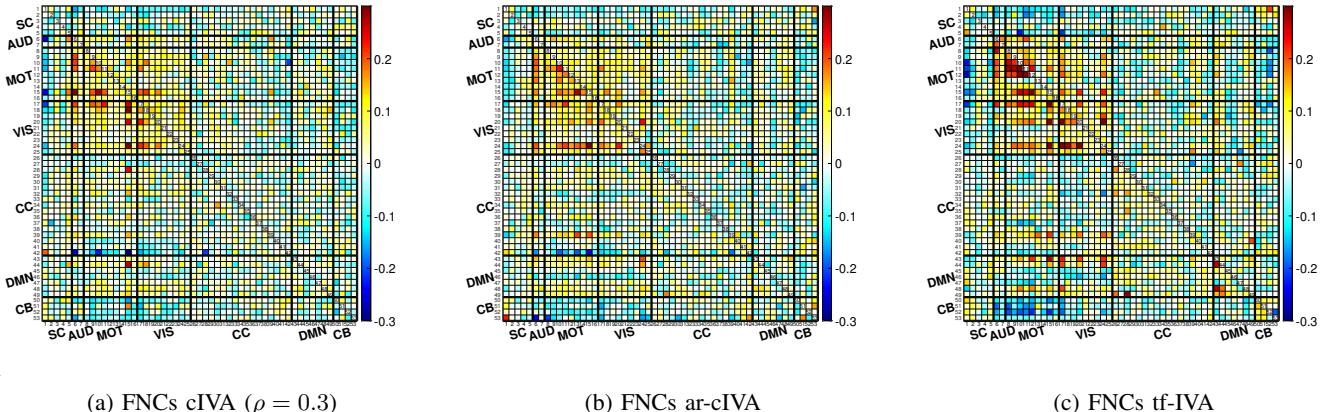
While we have demonstrated here an application to a dataset with 98 subjects, significantly higher than what has been used with IVA approaches that take HOS into account, and higher than results with IVA-G, the method is scalable to the analysis of thousands of subjects. Hence, the proposed methods enable large-scale analyses, including the identification of homogeneous subgroups, studies of large-scale dynamics, and replicability. The new methods are also applicable to other joint BSS applications such as those in remote sensing and video analysis.

## REFERENCES

- [1] M. J. McKeown, S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components," *Hum. Brain Mapp.*, vol. 6, no. 3, pp. 160–188, 1998.
- [2] J. Onton, M. Westerfield, J. Townsend, and S. Makeig, "Imaging human EEG dynamics using independent component analysis," *Neurosci. Biobehav. Rev.*, vol. 30, no. 6, pp. 808–822, 2006.
- [3] H. Luckhoo, J. R. Hale, M. G. Stokes, A. C. Nobre, P. G. Morris, M. J. Brookes, and M. W. Woolrich, "Inferring task-related networks using independent component analysis in magnetoencephalography," *Neuroimage*, vol. 62, no. 1, pp. 530–541, 2012.
- [4] V. D. Calhoun, T. Adali, G. Pearson, and J. J. Pekar, "Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms," *Hum. Brain Mapp.*, vol. 13, no. 1, pp. 43–53, 2001.
- [5] V. D. Calhoun and T. Adali, "Multisubject independent component analysis of fMRI: A decade of intrinsic networks, default mode, and neurodiagnostic discovery," *IEEE Trans. Biomed. Eng.*, vol. 5, pp. 60–73, 2012.
- [6] V. D. Calhoun, T. Adali, G. D. Pearson, and J. J. Pekar, "A method for making group inferences from functional MRI data using independent component analysis," *Hum. Brain Mapp.*, vol. 14, no. 3, pp. 140–151, 2001.
- [7] M. Svensén, F. Kruggel, and H. Benali, "ICA of fMRI group study data," *NeuroImage*, vol. 16, no. 3, pp. 551–563, 2002.

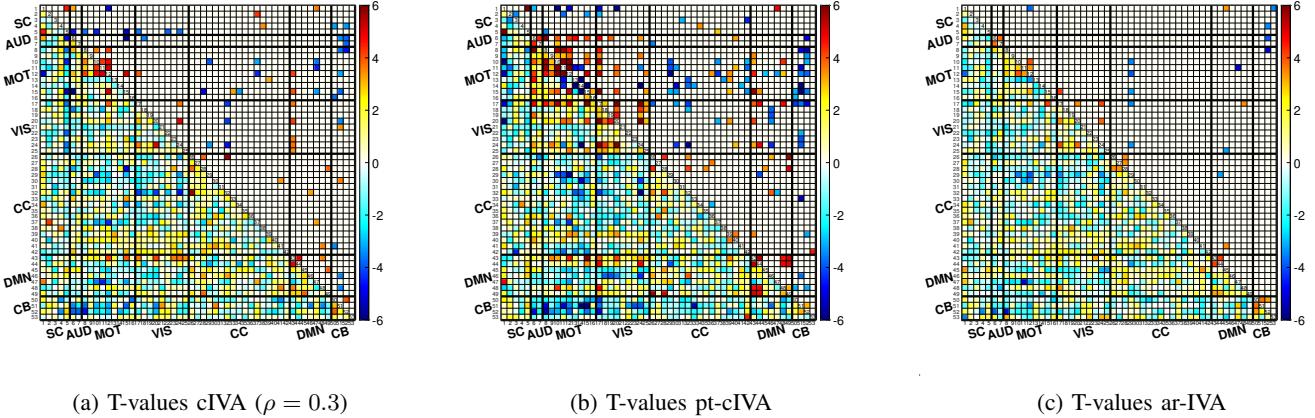


**Fig. 4:** Comparison of different IVA algorithms for real fMRI data with  $N = 60$ ,  $K = 98$ , and  $V = 58515$ . Plot (a) shows the cross-joint-ISI values for 50 independent runs for each algorithm. Plot (b) shows the power ratio for the most consistent run of the constrained algorithms. Note that in the case of pt-cIVA three outliers (with values of 22.59, 26.92, and 29.57) were removed for a clearer comparison. Plot (c) shows the average spatial maps across the  $K = 98$  subjects of three different RSNs: AUD, DMN, VIS (top to bottom). The coordinates (mm) of the peak activity are shown at the top of each spatial map.

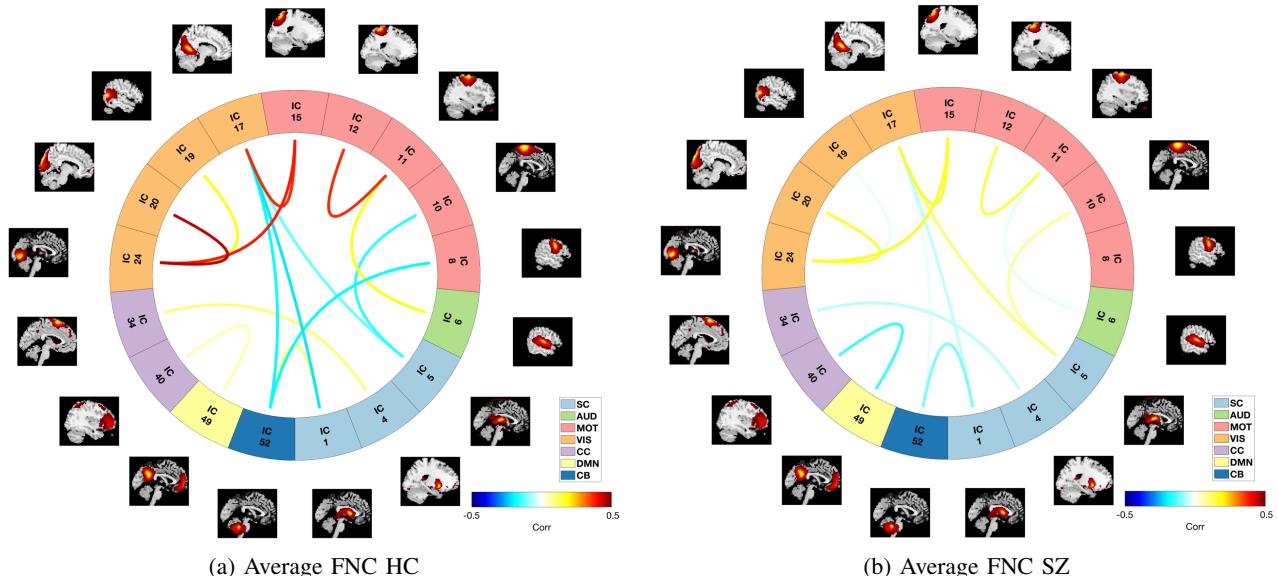


**Fig. 5:** Aggregated FNCs for HCs (lower diagonal) and mean difference FNCs (HC – SZ, upper diagonal) for the most consistent run. Pairwise Pearson correlation between RSNs time courses are first Fisher z-transformed and averaged across all HC/SZ subjects, then inverse z-transformed. The upper diagonal shows the difference between the mean FNC of HC and SZ, where red areas indicate higher connectivity in HC compared with SZ, and blue areas indicate lower connectivity in HC compared with SZ. The 53 components associated with the reference signals are considered for the analysis of the FNC matrix.

- [8] 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, 2005.
- [9] C. F. Beckmann, C. E. Mackay, N. Filippini, S. M. Smith *et al.*, “Group comparison of resting-state fMRI data using multi-subject ICA and dual regression,” *Neuroimage*, vol. 47, no. Suppl 1, p. S148, 2009.
- [10] E. B. Erhardt, S. Rachakonda, E. J. Bedrick, E. A. Allen, T. Adali, and V. D. Calhoun, “Comparison of multi-subject ICA methods for analysis of fMRI data,” *Hum. Brain Mapp.*, vol. 32, no. 12, pp. 2075–2095, 2011.
- [11] T. Kim, T. Eltoft, and T.-W. Lee, “Independent vector analysis: An extension of ICA to multivariate components,” in *Independent Component Analysis and Blind Signal Separation*. Springer, 2006, pp. 165–172.
- [12] M. Anderson, T. Adali, and X.-L. Li, “Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis,” *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1672–1683, 2011.
- [13] T. Adali, M. Anderson, and G.-S. Fu, “Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging,” *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 18–33, 2014.
- [14] A. M. Michael, M. Anderson, R. L. Miller, T. Adali, and V. D. Calhoun, “Preserving subject variability in group fMRI analysis: Performance evaluation of GICA vs. IVA,” *Front. Syst. Neurosci.*, vol. 8, p. 106, 2014.
- [15] J. Laney, K. P. Westlake, S. Ma, E. Woytowicz, V. D. Calhoun, and T. Adali, “Capturing subject variability in fMRI data: A graph-theoretical analysis of GICA vs. IVA,” *J. Neurosci. Methods*, vol. 247, pp. 32–40, 2015.
- [16] E. A. Allen, E. B. Erhardt, Y. Wei, T. Eichele, and V. D. Calhoun, “Capturing inter-subject variability with group independent component analysis of fMRI data: a simulation study,” *Neuroimage*, vol. 59, no. 4, pp. 4141–4159, 2012.
- [17] B. Loesch, F. Nesta, and B. Yang, “On the robustness of the multidimensional state coherence transform for solving the permutation problem of frequency-domain ICA,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* IEEE, 2010, pp. 225–228.
- [18] Q. Long, S. Bhinge, V. D. Calhoun, and T. Adali, “Independent vector analysis for common subspace analysis: Application to multi-subject fMRI data yields meaningful subgroups of schizophrenia,” *NeuroImage*, vol. 216, p. 116872, 2020.
- [19] S. Bhinge, R. Mowakeaa, V. D. Calhoun, and T. Adali, “Extraction of time-varying spatiotemporal networks using parameter-tuned constrained IVA,” *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1715–1725, 2019.
- [20] J. R. Kettenring, “Canonical analysis of several sets of variables,” *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.



**Fig. 6:** T-value maps showing the differences in the FNCs for HCs obtained by paired t-tests for tf-cIVA vs. cIVA ( $\rho = 0.3$ ), tf-cIVA vs. pt-cIVA, and tf-cIVA vs. ar-IVA. The lower diagonal of the maps shows the T-values before the FDR correction, while the upper diagonal shows the T-values after passing the FDR correction ( $p$ -value  $< 0.05$ ). Positive T-values indicate tf-cIVA shows a higher connectivity value than the compared algorithm. The 53 components associated with the references are considered for the paired t-test.



**Fig. 7:** Average FNC strength obtained by tf-cIVA across subjects for HC and SZ groups. A two-sample t-test is applied and only the RSNs with significant differences between groups after FDR correction are shown. For each RSN, the averaged values in SZ and HC groups are presented by the connecting lines. The outer circle depicts the ICs indices and their corresponding average spatial maps. The 53 components associated with the reference signals are considered for the two-sample t-test.

- [21] H. Richard, L. Gresele, A. Hyvärinen, B. Thirion, A. Gramfort, and P. Ablin, “Modeling shared responses in neuroimaging studies through MultiView ICA,” *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 19 149–19 162, 2020.
- [22] V. D. Calhoun, T. Adali, G. D. Pearson, and K. A. Kiehl, “Neuronal chronometry of target detection: fusion of hemodynamic and event-related potential data,” *Neuroimage*, vol. 30, no. 2, pp. 544–553, 2006.
- [23] A. H. Andersen and W. S. Rayens, “Structure-seeking multilinear methods for the analysis of fMRI data,” *NeuroImage*, vol. 22, no. 2, pp. 728–739, 2004.
- [24] M. S. Salman, Y. Du, D. Lin, Z. Fu, A. Fedorov, E. Damaraju, J. Sui, J. Chen, A. R. Mayer, S. Posse *et al.*, “Group ICA for identifying biomarkers in schizophrenia: ‘Adaptive’ networks via spatially constrained ICA show more sensitivity to group differences than spatio-temporal regression,” *NeuroImage Clin.*, vol. 22, p. 101747, 2019.
- [25] W. Lu and J. C. Rajapakse, “Approach and applications of constrained ICA,” *IEEE Trans. Neural Netw.*, vol. 16, no. 1, pp. 203–212, 2005.
- [26] P. A. Rodriguez, M. Anderson, X.-L. Li, and T. Adali, “General non-orthogonal constrained ICA,” *IEEE Trans. Signal Process.*, vol. 62, no. 11, pp. 2778–2786, 2014.
- [27] Z.-L. Zhang, “Morphologically constrained ICA for extracting weak temporally correlated signals,” *Neurocomputing*, vol. 71, no. 7–9, pp. 1669–1679, 2008.
- [28] 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, 2005.
- [29] M. De Vos, L. De Lathauwer, and S. Van Huffel, “Spatially constrained ICA algorithm with an application in EEG processing,” *Signal Process.*, vol. 91, no. 8, pp. 1963–1972, 2011.
- [30] Q. Long, S. Bhinge, V. D. Calhoun, and T. Adali, “Relationship between dynamic blood-oxygen-level-dependent activity and functional network

connectivity: Characterization of schizophrenia subgroups," *Brain Connect.*, vol. 11, no. 6, pp. 430–446, 2021.

[31] S. Bhinge, Q. Long, Y. Levin-Schwartz, Z. Boukouvalas, V. D. Calhoun, and T. Adali, "Non-orthogonal constrained independent vector analysis: Application to data fusion," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* IEEE, 2017, pp. 2666–2670.

[32] W. Lu and J. Rajapakse, "Constrained independent component analysis," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 13, 2000.

[33] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*. Academic press, 2014.

[34] X.-L. Li and X.-D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solution," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1803–1814, 2007.

[35] X.-L. Li and T. Adali, "Independent component analysis by entropy bound minimization," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5151–5164, 2010.

[36] Y. Du and Y. Fan, "Group information guided ICA for fMRI data analysis," *Neuroimage*, vol. 69, pp. 157–197, 2013.

[37] E. Moreau and O. Macchi, "A one stage self-adaptive algorithm for source separation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 3. IEEE, 1994, pp. III–49.

[38] S.-i. Amari, A. Cichocki, and H. Yang, "A new learning algorithm for blind signal separation," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 8, 1995.

[39] Q. Long, C. Jia, Z. Boukouvalas, B. Gabrielson, D. Emge, and T. Adali, "Consistent run selection for independent component analysis: Application to fMRI analysis," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* IEEE, 2018, pp. 2581–2585.

[40] Y. Du, Z. Fu, J. Sui, S. Gao, Y. Xing, D. Lin, M. Salman, A. Abrol, M. A. Rahaman, J. Chen *et al.*, "NeuroMark: An automated and adaptive ICA based pipeline to identify reproducible fMRI markers of brain disorders," *NeuroImage Clin.*, vol. 28, p. 102375, 2020.

[41] C. A. Tamminga, E. I. Ivleva, M. S. Keshavan, G. D. Pearlson, B. A. Clementz, B. Witte, D. W. Morris, J. Bishop, G. K. Thaker, and J. A. Sweeney, "Clinical phenotypes of psychosis in the bipolar-schizophrenia network on intermediate phenotypes (B-SNIP)," *Am. J. Psychiatry*, vol. 170, no. 11, pp. 1263–1274, 2013.

[42] C. A. Tamminga, G. Pearlson, M. Keshavan, J. Sweeney, B. Clementz, and G. Thaker, "Bipolar and schizophrenia network for intermediate phenotypes: Outcomes across the psychosis continuum," *Schizophr. Bull.*, vol. 40, no. Suppl.2, pp. S131–S137, 2014.

[43] E. A. Allen, E. B. Erhardt, E. Damaraju, W. Gruner, J. M. Segall, R. F. Silva, M. Havlicek, S. Rachakonda, J. Fries, R. Kalyanam *et al.*, "A baseline for the multivariate comparison of resting-state networks," *Front. Syst. Neurosci.*, vol. 5, p. 2, 2011.

[44] Y. Du, Z. Fu, Y. Xing, D. Lin, G. Pearlson, P. Kochunov, L. E. Hong, S. Qi, M. Salman, A. Abrol *et al.*, "Evidence of shared and distinct functional and structural brain signatures in schizophrenia and autism spectrum disorder," *Commun. Biol.*, vol. 4, no. 1, p. 1073, 2021.

[45] X. Meng, A. Iraji, Z. Fu, P. Kochunov, A. Belger, J. Ford, S. McEwen, D. H. Mathalon, B. A. Mueller, G. Pearlson *et al.*, "Multimodel order independent component analysis: A data-driven method for evaluating brain functional network connectivity within and between multiple spatial scales," *Brain Connect.*, vol. 12, no. 7, pp. 617–628, 2022.

[46] Y. Benjamini and D. Yekutieli, "False discovery rate-adjusted multiple confidence intervals for selected parameters," *Journal of the American Statistical Association*, vol. 100, no. 469, pp. 71–81, 2005.

[47] T. Adali, F. Kantar, M. A. B. S. Akhonda, S. Strother, V. D. Calhoun, and E. Acar, "Reproducibility in matrix and tensor decompositions: focus on model match, interpretability, and uniqueness," *IEEE Signal Process. Mag.*, vol. 39, no. 4, pp. 8–24, 2022.