Adaptive constrained independent vector analysis: An effective solution for analysis of large-scale medical imaging data

Suchita Bhinge, Qunfang Long, Vince D. Calhoun, and Tülay Adalı

***Abstract*—There is a growing need for flexible methods for the analysis of large-scale functional magnetic resonance imaging (fMRI) data for the estimation of global signatures that summarize the population while preserving individual-specific traits. Independent vector analysis (IVA) is a data-driven method that jointly estimates global spatio-temporal patterns from multi-subject fMRI data, and effectively preserves subject variability. However, as we show, IVA performance is negatively affected when the number of datasets and components increases especially when there is low component correlation across the datasets. We study the problem and its relationship with respect to correlation across the datasets, and propose an effective method for addressing the issue by incorporating reference information of the estimation patterns into the formulation, as a guidance in high dimensional scenarios. Constrained IVA (cIVA) provides an efficient framework for incorporating references, however its performance depends on a user-defined constraint parameter, which enforces the association between the reference signals and estimation patterns to a fixed level. We propose adaptive cIVA (acIVA) that tunes the constraint parameter to allow flexible associations between the references and estimation patterns, and enables incorporating multiple reference signals, without enforcing inaccurate conditions. Our results indicate that acIVA can reliably estimate high-dimensional multivariate sources from large-scale simulated datasets, when compared with standard IVA. It also successfully extracts meaningful functional networks from a large-scale fMRI dataset for which standard IVA did not converge. The method also efficiently captures subject-specific information, which is demonstrated through observed gender differences in spectral power, higher spectral power in males at low frequencies and in females at high frequencies, within the motor, attention, visual and default mode networks.**

***Index Terms*—Blind source separation, fMRI analysis, high dimensional, semi-blind**

I. INTRODUCTION

NEUROIMAGING analysis has allowed for the identification of distinguishing characteristics of the human brain including gender [1], age [1], [2], addiction, and different brain disorders [3], [4], [5], [6], [7]. There is an increasing availability of data which provides additional power to help identify global signatures for a given population and patterns specific to each individual, group, condition, or modality [8]. With the increasing availability of data, there is a growing need of flexible methods that efficiently capture the global patterns while still preserving the subject-specific information from the large-scale datasets. Common approaches for fMRI analysis involve extraction of per subject features summarizing the overall temporal activity as in [9], [10], transformation of high-dimensional datasets to a lower dimensional feature space such as in group independent component analysis (GICA) [11], [12], multisubject dictionary learning [13], tensor decompositions [14], or using prior information such as regions-of-interest (ROI) or a task paradigm in a regression type analysis to analyze a single subject at a time. Some methods adopt a divide-and-conquer approach that divide the high dimensional problem into a series of smaller dimensional problems and combines the results [15], [16]. Although these approaches have successfully identified relevant biomarkers, they may not take advantage of all available information, are sensitive to ROI selection and do not exploit the complementary information across multiple subjects.

Independent component analysis (ICA) is a popular data-driven approach used to extract subject-specific time courses and spatial maps under the assumption of statistical independence, however it is limited to the analysis of a single dataset. GICA jointly analyzes the data from multiple subjects and estimates a common, global representation of the functional networks across subjects, however it might be limited in preserving subject variability. Independent vector analysis (IVA) is an extension of ICA to multiple datasets that jointly decomposes the multisubject data into subject-specific time courses and spatial maps. It has been shown in various studies that IVA is better in capturing the subject variability [17], [18], [19] compared with GICA and provides automatic source alignment across subjects by exploiting the source dependence across datasets. It presents a wide range of algorithms depending on the assumption of the latent multivariate source distribution and provides general identification conditions that allow for flexibility in the estimation of the underlying sources [19].

Although IVA provides a general framework for the analysis of multi-subject fMRI data, its performance depends on a number of factors such as number of samples [20], number of datasets [21], number of sources, and the level of correlation

Manuscript submitted December 19, 2019; accepted June 1, 2020. This work is supported in part by the National Institute of Biomedical Imaging and Bioengineering under Grant R01 EB 020407, National Science Foundation under Grant 1631838 and National Science Foundation-Computing and Communication Foundations under Grant 1618551.

S. Bhinge, Q. Long and T. Adalı are with the Department of Electrical and Computer Engineering, University of Maryland, Baltimore County, Baltimore, MD 21250 USA. V. D. Calhoun is with The Mind Research Network, Albuquerque, NM 87106 USA, and also with the Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131 USA.

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author. The material includes SupplementaryMaterial.pdf document. Contact suchita1@umbc.edu for further questions about this work.

between the marginals of the multivariate sources. In this paper, we first study how these factors affect the performance of IVA through simulations, and in particular, highlight an important point that has not been addressed before, the role of the level of correlation across the datasets, and discuss the strengths and weaknesses of IVA in detail. We analyze the effect of number of datasets and the level of correlation between the marginals of the multivariate sources on standard IVA through simulations, and show that the performance of IVA is largely affected when the sources have low to moderate levels of correlation when dimensionality increases. We link those to efficiency considerations in maximum likelihood estimation. In the analysis of fMRI data, the multivariate sources corresponding to spatial components that are typically super-Gaussian distributed [22] commonly exhibit low to moderate correlation across subjects, see supplementary material-S1. We identify this as a main concern for the limited nature of the adoption of IVA for the analysis of large-scale fMRI data, which has been unfortunate given its particular strengths for preserving subject variability.

Due to the increasing availability of large-scale fMRI datasets, the need for a method that can effectively estimate meaningful global functional networks, and simultaneously capture the individual-specific traits from those, is highly desirable. This is also one of the key steps that will enable precision medicine that needs to fully account for individual variability. In this work, we introduce an effective method for addressing the high dimensionality issue of IVA by incorporating reference information of the functional networks into the IVA framework, as a guidance in high-dimensional scenarios. Constrained IVA is an effective method that incorporates prior information regarding the sources or the columns of the mixing matrix into the IVA cost function [23], [24]. It relaxes the independence assumption by enabling a desirable balance between data-driven and model driven methods, and provides a model match through the use of accurate constraints. However, cIVA makes use of a user-defined constraint parameter that controls the degree of correspondence between reference signal and the estimated component. The successful performance of cIVA depends on the selection of prior information and the user-defined constraint parameter, *i.e.*, when the prior information is incorrect a lower constraint parameter must be used such that the prior information is not enforced on the decomposition [21], [23]. On the other hand, when the prior information is correct, a higher constraint parameter must be used such that the components are deterred from effect of noise and artifacts. However, in most practical applications, the selection of a constraint parameter becomes difficult since it is unknown whether the prior information is accurate, and becomes more complicated when prior information regarding multiple signals need to be incorporated.

We propose the adaptive cIVA (acIVA) method that adaptively controls the association between the reference signal and estimation patterns, and enables efficient incorporation of multiple reference signals into the IVA framework. The acIVA technique ensures that the prior information is used to guide the solution and does not enforce inaccurate constraints. In this paper, we study the performance of acIVA on simulated high dimensional datasets in terms of estimating the underlying dataset-specific sources, and demonstrate its successful performance to overcome a major weakness of standard IVA. We also apply acIVA to large-scale fMRI data acquired from 327 subjects and show that acIVA successfully preserves subject-variability compared to gICA, and identifies functionally relevant resting-state networks in a data-driven manner, whereas standard IVA did not converge on this dataset. The acIVA technique also captures gender differences in spectral power, where higher spectral power is observed in males at low frequencies ($<$ 0.05 Hz), and higher spectral power is observed in females at high frequencies (0.05 to 0.15 Hz).

The remainder of the paper is organized as follows. Section II introduces the IVA and acIVA models. Section III discusses the aspects of IVA that affect its performance and demonstrates the effect of high dimensionality on IVA. Section IV demonstrates the use of acIVA on high dimensional datasets using simulated datasets and Section V shows results from applying acIVA to a large-scale fMRI dataset. Section VI discusses the results and Section VII concludes the paper.

II. BACKGROUND

A. Independent vector analysis

IVA extends ICA to multiple datasets to jointly estimate components that are independent within each dataset and dependent across datasets. Given K datasets, each comprised of N components, $\mathbf{x}^{[k]} \in \mathbb{R}^N,\ k = 1, \ldots, K$, we have, $\mathbf{x}^{[k]} = \mathbf{A}^{[k]}\mathbf{s}^{[k]},\ \ k = 1, \ldots, K$, where $\mathbf{A}^{[k]} \in \mathbb{R}^{N \times N}$ is the mixing matrix. IVA estimates K demixing matrices, $\mathbf{W}^{[k]}$, to compute the source estimates, $\hat{\mathbf{s}}^{[k]} = \mathbf{W}^{[k]}\mathbf{x}^{[k]}$, by maximizing the likelihood function or equivalently minimizing the mutual information based cost function given as [19], [25],

$$\mathcal{J}_{\text{IVA}} = \sum_{n=1}^{N} \left[\sum_{k=1}^{K} \mathcal{H}\left(\hat{s}_n^{[k]}\right) - \mathcal{I}\left(\hat{\mathbf{s}}_n\right) \right] - \sum_{k=1}^{K} \log \left| \det \mathbf{W}^{[k]} \right|, \tag{1}$$

where $\mathcal{H}\left(\hat{s}_n^{[k]}\right)$ denotes the entropy of the nth source estimate for the kth dataset, and $\mathcal{I}\left(\mathbf{s}_n\right)$ denotes the mutual information among sources within the nth source component vector (SCV), $\hat{\mathbf{s}}_n^T = \left[\hat{s}_n^{[1]}, \ldots, \hat{s}_n^{[K]}\right]$. The covariance matrix of the nth SCV, $\mathbf{\Sigma}_n \in \mathbb{R}^{K \times K}$, is a positive definite matrix. The minimization of the cost function simultaneously weighs the independence within the dataset through the entropy term along with the log determinant term and dependence across the datasets through the mutual information term. SCV takes into account the dependence across the datasets and the nth SCV is formed by concatenating the nth component from all the K datasets as shown in Fig. 1. For a given set of observations, the IVA model can be written as $\mathbf{X}^{[k]} = \mathbf{A}^{[k]}\mathbf{S}^{[k]} \in \mathbb{R}^{N \times V}$, where $\mathbf{S}^{[k]} = \left[\mathbf{s}_1^{[k]}, \ldots, \mathbf{s}_N^{[k]}\right]^T$, $\mathbf{s}_n^{[k]} \in \mathbb{R}^V, n = 1, \ldots, N$ are latent sources and V is the number of samples. The estimated sources are obtained using $\hat{\mathbf{S}}^{[k]} = \mathbf{W}^{[k]}\mathbf{X}^{[k]}$ and the nth SCV is defined as $\mathbf{s}_n = \left[\mathbf{s}_n^{[1]}, \ldots, \mathbf{s}_n^{[K]}\right]^T \in \mathbb{R}^{K \times V}$.

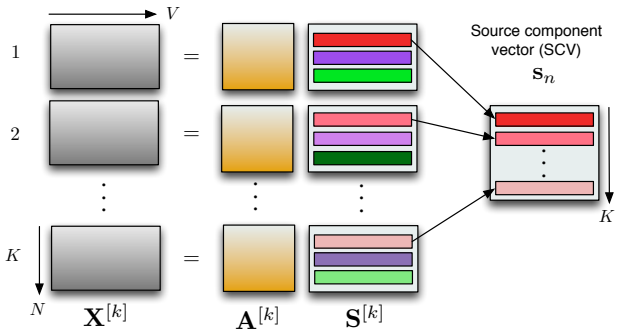

Fig. 1. Given a set of observations, the IVA model is given as $\mathbf{X}^{[k]} = \mathbf{A}^{[k]}\mathbf{S}^{[k]},\ k = 1, \ldots, K$, where $\mathbf{A}^{[k]}$ is the mixing matrix and the rows in $\mathbf{S}^{[k]}$, are the latent sources that are dependent across datasets. The nth SCV is formed by grouping the corresponding nth source from each dataset together.

B. Constrained IVA

Constrained IVA (cIVA) is a semi-blind source separation (BSS) algorithm that incorporates prior information regarding the sources or the columns of the mixing matrix into the IVA cost function, and helps relax the independence assumption which at times might not be well satisfied [23]. The cIVA method relaxes the independence assumption of IVA and provides a desirable balance between data-driven techniques that minimize the assumptions placed on the data, and model-driven techniques that make use of prior information, which if correct yield them robust to noise and artifacts. Let $d_l,\ l = 1, \ldots, L$, denote the lth reference sample belonging to a known vector $\mathbf{d}_l$ and L denote the number of references. The cIVA algorithm directs the estimation of the sources towards the reference signals through an additional constraint term, hence the IVA cost function is now given as,

$$\mathcal{J} = \mathcal{J}_{\text{IVA}} - \sum_{l=1}^{L} \frac{1}{2\gamma_l} \sum_{k=1}^{K} \left\{ \left[\max\{0, \mu_l^{[k]} + \gamma_l g(\hat{s}_l^{[k]}, d_l)\} \right]^2 - (\mu_l^{[k]})^2 \right\}, \tag{2}$$

where γ_l is the penalty parameter, $\mu_l^{[k]}$ is the Lagrange multiplier and $g(\hat{s}_l^{[k]}, d_l)$ is the inequality constraint function given as,

$$g(\hat{s}_l^{[k]}, d_l) = \rho_l - \epsilon(\hat{s}_l^{[k]}, d_l) \leq 0, \tag{3}$$

where $\hat{s}_l^{[k]} = \left(\mathbf{w}_l^{[k]}\right)^T \mathbf{x}^{[k]}$ is the estimated component, $\epsilon(\cdot, \cdot)$ is a function that defines the measure of similarity between the estimated source and reference sample, and ρ_l is the constraint parameter. Note that the number of constraints, L, is less than or equal to N, and the indices n and l can be used interchangeably for constrained sources or mixing vectors. In the following text we will use index l to refer to constrained sources and n to refer to the unconstrained sources. The definition of the constraint function as in (3) allows for the use of different distance functions such as the Euclidean distance, square error and mutual information.

For a given set of observations with V samples, the IVA model can be written as, $\mathbf{X}^{[k]} = \mathbf{A}^{[k]}\mathbf{S}^{[k]},\quad k = 1, \ldots, K$, where $\mathbf{X},\ \mathbf{S} \in \mathbb{R}^{N \times V}$, the constraint function is given by $g(\hat{\mathbf{s}}_l^{[k]}, \mathbf{d}_l) = \rho_l - \epsilon(\hat{\mathbf{s}}_l^{[k]}, \mathbf{d}_l) \leq 0$ and $\mathbf{d}_l \in \mathbb{R}^V$ is the reference signal. In this work, we define $\epsilon(\hat{\mathbf{s}}_l^{[k]}, \mathbf{d}_l)$ as the absolute value of the Pearson's correlation coefficient between the reference signal and the estimated source, hence $0 \leq \epsilon(\hat{\mathbf{s}}_l^{[k]}, \mathbf{d}_l) \leq 1$. The constraint parameter, ρ_l, acts as a lower bound for correlation between the reference signal, $\mathbf{d}_l$, and the estimated source to control the degree of correspondence between the two. A higher value of ρ_l enforces the estimated source to be exactly similar to the reference signal, not allowing the reference component to vary across datasets, whereas a lower value results in the estimated component to deviate from the reference signal making it to be prone to noise and other artifacts. Hence, the selection of ρ_l plays a crucial role in the performance of the cIVA algorithm. In Section III-A, we discuss an adaptive technique, namely, acIVA, that efficiently incorporates reference signals into the IVA cost function by tuning the constraint parameter, and its extension to multiple reference signals.

III. KEY ASPECTS OF IVA

Although IVA preserves the variability across datasets by estimating dataset-specific mixing matrix and sources, its performance depends on a number of factors such as number of samples [20], number of datasets [21] and sources. In this section, we study through simulations the role of these factors and potential strengths and weaknesses of IVA. The effect of sample size on IVA is studied in several studies such as in [20], [25], [26]. As with maximum likelihood based estimators, the performance of IVA improves when the sample size increases. On the other hand, with more datasets available to exploit the dependence structure in IVA, the performance of IVA improves with an increase in the number of datasets, due to the availability of more cross-dataset information [25]. For an infinite number of samples, the performance of IVA will keep on improving. However, in real world applications, there is a limitation to the number of available samples. Due to the limited number of samples available, IVA experiences a performance drop with increase in number of sources or datasets for a fixed number of samples [21]. In this paper, we note that this trade-off between the dimensionality and utilization of maximal information content also depends on the level of correlation among the SCVs and the number of sources, an important point that has not been addressed before to the best of our knowledge. In this section, we study the effect of level of correlation level on IVA and the effect of number of sources is studied in Section IV.

In order to demonstrate the effect of the number of datasets and the level of correlation on IVA, for fixed V and N, we generate K datasets, $\mathbf{x}^{[k]} \in \mathbb{R}^N$, such that $\mathbf{x}^{[k]} = \mathbf{A}^{[k]}\mathbf{s}^{[k]}$. The elements in the mixing matrices, $\mathbf{A}^{[k]}, k = 1, \ldots, K$ are randomly generated from a uniform distribution. The N SCVs are generated from a K-dimensional multivariate Laplacian distribution with a covariance structure $\mathbf{\Sigma}_n = \mathbf{Q}\mathbf{Q}^T$, where $\mathbf{Q} \in \mathbb{R}^{K \times K}$ is a randomly generated matrix. We vary the level of correlation within an SCV through the generation of the matrix $\mathbf{Q}$, where the elements in $\mathbf{Q}$ denoted as q_{ij}, are generated randomly from a normal distribution, $\mathcal{N}(0, 1)$ for Case 1, $\mathcal{N}(0.1, 0.3)$ for Case 2, a uniform distribution,

$\mathcal{U}(-0.2, 0.8)$ for Case 3, $\mathcal{U}(0, 1)$ for Case 4, and $\mathcal{U}(0.1, 0.3)$ for Case 5. The resulting distribution of correlation values in $\mathbf{\Sigma}_n$ for $K = 10$ is shown in Fig. 2(a). We can see that the correlation values demonstrate an increasing trend from Case 1 to Case 5 corresponding to increase in the level of correlation within an SCV.

After obtaining the SCVs, the sources for the kth dataset are obtained by grouping together the kth row from all the N SCVs, $\mathbf{s}^{[k]} = \left[s_1^{[k]}, \ldots, s_N^{[k]}\right]$. We generate $V = 1000$ samples and vary the number of datasets from 2 to 40. We obtain three independent observation sets of the datasets, $\mathbf{X}^{[k]}$, by randomly generating the mixing matrices and sources for each set. IVA using the IVA-L-SOS algorithm is performed for five runs on each set, and the performance of both methods is measured in terms of joint inter-symbol interference (jISI) [25]. The jISI metric measures the performance of the methods in terms of its ability to separate the sources ($0 \leq$ jISI ≤ 1), where 0 indicates better separation of underlying SCVs, *i.e.*, $\mathbf{W}^{[k]}\mathbf{A}^{[k]} = I,\ \forall k \in \{1, \ldots, K\}$. IVA-L-SOS simultaneously accounts for second and higher order statistics, and assumes the sources are correlated multivariate Laplacian distributed [21]. This assumption is a good model match for fMRI analysis since the sources tend to have a super-Gaussian distribution with correlation across subjects [12], [22]. The average of the jISI metric across all converged runs for IVA is shown in Fig. 2(b).

When there are sufficient samples available, the advantage of exploiting source dependence across datasets in IVA is observed in Fig. 2(b), as the performance improves with increase in number of datasets for a fixed number of sources and samples. However, this increasing trend is observed upto a certain limit, after which IVA is affected by the curse of dimensionality. The limit varies based on the level of correlation. The effect of dimensionality in IVA is observed when there is insufficient statistical power to provide reliable and meaningful estimates of the high dimensional multivariate probability distribution functions, due to the availability of limited samples. The high dimensional effect is more prominent in Cases 1 and 2. The low correlation structure of the SCVs for these cases have higher distances between the marginals resulting in the data points to be sparsely located in the multidimensional space, hence the estimation of such SCVs becomes even more difficult in high dimensional scenarios. For highly correlated SCVs, as in Case 5, the marginals are more densely located, which aids the estimation of these sources and yields a more efficient estimator in the maximum likelihood sense.

One of the discussion items for IVA has been its requirement for the sources to be highly dependent across datasets, affecting its ability to capture subject-specific information. However, note that for Cases 1 and 2, which have a low to moderate correlation structure, the jISI is similar to that of Cases 3 to 5 for lower number of datasets, which indicates that IVA does not require or enforce the sources across datasets to be highly correlated. In the analysis of fMRI datasets, the nth SCV typically corresponds to a functional network activated across multiple subjects. The structure of the multivariate distributions of these SCVs show high activations in a smaller group of voxels and lower activation values for a larger group of voxels. This results in fMRI sources to have low correlations across datasets, see supplementary material-S1, affecting their estimation significantly in high dimensional scenarios. This issue can be addressed by using references to estimate the SCVs. In the next section, we describe the general acIVA method to incorporate multiple reference signals into the IVA cost function.

A. Adaptive constrained IVA (acIVA)

The use of reference signals provides an effective way to relax the independence assumption and lead the optimization search towards a better solution. It also leverages the benefits of model-driven and data-driven techniques. However, use of a fixed value for the constraint parameter does not allow the model to efficiently estimate the local patterns for each dataset. Hence, we propose a procedure to adaptively tune the constraint parameter during the optimization of cIVA. The idea of the adaptive tuning technique is to find the highest value for $\rho_l^{[k]}$ from a sequence of possible values, $(\rho_i)_{i\in\mathbb{N}}$, that satisfies the condition in (3) for the lth constraint and kth dataset. For example, given a sequence $(3, 2.5, 5, 1.2)$, the highest lower bound for $\epsilon(\hat{s}_l^{[k]}, d_l) = 2.8$ from the sequence is $\hat{\rho}_l^{[k]} = 2.5$.

Algorithm 1 describes the steps in the acIVA technique. We define ρ_i as the ith element from the sorted sequence, $(\rho_i)_{i\in\mathbb{N}}$, L as the number of constraints, and $d_l,\ l = 1, \ldots, L$, as the reference signal used to constrain L out of N sources. We randomly initialize the demixing matrices, $\mathbf{W}^{[k]}$, set $\mu_l^{[k]} = 0$ and γ_l to a positive scalar value. At each iteration, we obtain an estimate of the sources, $\hat{s}_l^{[k]},\ k = 1, \ldots, K,\ l = 1, \ldots, L$ and update the Lagrange multiplier, $\mu_l^{[k]}$, as given in line 10 of Algorithm 1. The constraint function used to update the Lagrange multiplier is computed using the bound selected from lines 7 to 9. Starting with the index, i, set to the smallest element in the sequence, $(\rho_i)_{i\in\mathbb{N}}$, the algorithm increments the index i until it satisfies the condition on line 8, which checks if the $i+1$th bound exceeds $\left|\epsilon(\hat{s}_l^{[k]}, d_l)\right|$ while the ith bound is lower than $\left|\epsilon(\hat{s}_l^{[k]}, d_l)\right|$. The new value of the constraint parameter, $\hat{\rho}_l$ is set to ρ_i if the condition is met and is used to compute the gradient, $\partial\mathcal{J}/\partial\mathbf{w}_l^{[k]}$, as follows,

$$\partial\mathcal{J}/\partial\mathbf{w}_l^{[k]} = \partial\mathcal{J}_{\text{IVA}}/\partial\mathbf{w}_l^{[k]} - g'\left(\hat{s}_l^{[k]}, d_l\right)\mu_l^{[k]}(d_l), \quad (4)$$

where $g'\left(\hat{s}_l^{[k]}, d_l\right)$ is the derivative of $g\left(\hat{s}_l^{[k]}, d_l\right)$ with respect to $\mathbf{w}_l^{[k]}$, $\partial\mathcal{J}_{\text{IVA}}/\partial\mathbf{w}_l^{[k]}$ is the gradient of (1), and update the demixing vector as in line 13 followed by obtaining a new estimate of the sources. The process is repeated until the convergence criterion, following the one proposed in [23], is met. Hence, the constraint parameter gets closer to the true correlation between the source and the reference signal at every iteration, without enforcing the reference signal on the decomposition. In the next section, we will study the performance of standard IVA and acIVA in high dimensional scenarios for cases where the SCVs have moderately low correlation, as observed in fMRI data.

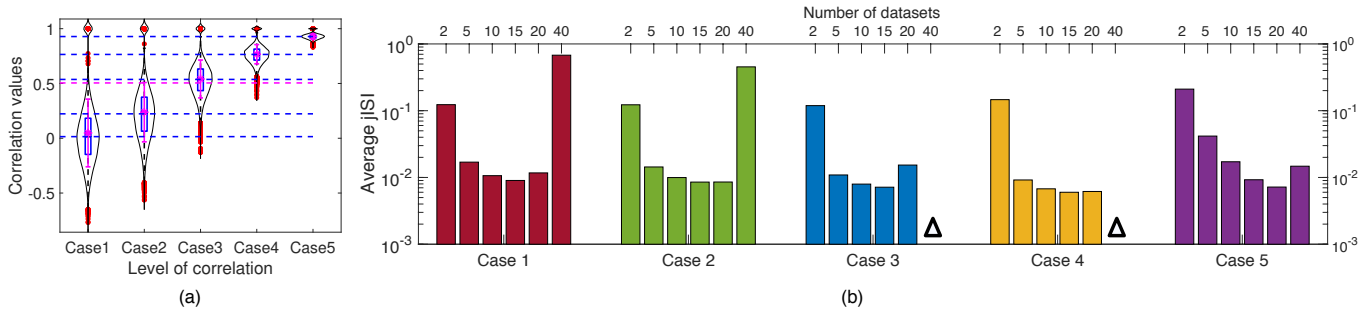

Fig. 2. (a) Distribution of correlation values in $\mathbf{\Sigma}_n = \mathbf{Q}\mathbf{Q}^T$ for Cases 1-5. The black curve is the smoothed distribution of correlation values within all SCVs. The box plot displays the median, the 25th and 75th percentiles of the correlation values with whiskers extending to the 99.3% confidence interval and some outliers beyond whisker. The mean and standard deviation are in magenta. Horizontal magenta refer to the global mean from all cases and horizontal blue lines refer to the median across all SCVs in each case. (b) Performance of IVA in terms of jISI with respect to number of datasets, K, for each case. The jISI metric is the average jISI computed across all converged runs. The level of correlation increases from Case 1 to Case 5 with Case 1 corresponding to low correlation within an SCV whereas Case 5 corresponds to high correlated SCVs. The number of sources and number of samples is fixed to $N = 85$ and $V = 1000$, respectively. For Case 3 and Case 4 at $K = 40$ no run converged in 1024 iterations, which is indicated by '$\triangle$', however with an increase in K we start observing an increasing trend.

Algorithm 1 acIVA

Input: A sequence, $(\rho_i)_{i\in\mathbb{N}}$, of possible values for $\rho_l^{[k]}$, sorted in ascending order
Randomly initialize K demixing matrices, $\left[\mathbf{W}^{[1]}, \ldots, \mathbf{W}^{[K]}\right]$ and set $\mu_n^{[k]} = 0, \gamma_n$ to be a positive scalar value
for $n = 1, \ldots, N$ **do**
  for $k = 1, \ldots, K$ **do**
    Compute $\hat{\mathbf{s}}^{[k]} = \mathbf{W}^{[k]}\mathbf{x}^{[k]},\ k = 1, \ldots, K$
    if $n \in \{1, \ldots, L\}$ **then**
      $i = 1$
      repeat $i = i + 1$
      until $\rho_i - \left|\epsilon(\hat{s}_l^{[k]}, d_l)\right| < 0 < \rho_{i+1} - \left|\epsilon(\hat{s}_l^{[k]}, d_l)\right|$
      $\hat{\rho}_l^{[k]} = \rho_i$
      $\mu_l^{[k]} = \max\left\{0, \gamma_l g\left(\hat{s}_l^{[k]}, d_l\right) + \mu_l^{[k]}\right\}$
      Compute $\partial\mathcal{J}/\partial\mathbf{w}_l^{[k]}$ using (4)
      $\mathbf{w}_n^{[k]} = \mathbf{w}_n^{[k]} + \partial\mathcal{J}/\partial\mathbf{w}_n^{[k]}$
    else
      Compute $\partial\mathcal{J}_{\text{IVA}}/\partial\mathbf{w}_n^{[k]}$
      $\mathbf{w}_n^{[k]} = \mathbf{w}_n^{[k]} + \partial\mathcal{J}_{\text{IVA}}/\partial\mathbf{w}_n^{[k]}$
Repeat 3 to 13 until convergence

IV. APPLICATION OF ACIVA TO SIMULATED DATA

In order to study the performance of acIVA over standard IVA on high dimensional datasets, we generate simulated datasets similar to the set-up described in Section III. We vary the number of datasets, K, from 2 to 60, and the number of sources, N, from 40 to 85, with a fixed sample size $V = 1000$. The SCVs are generated from a multivariate Laplacian distribution with a covariance structure, $\mathbf{\Sigma} = \mathbf{Q}\mathbf{Q}^T,\ q_{ij} = \mathcal{N}(0.1, 0.3)$. This structure results in low to moderately correlated sources that is a good match to the fMRI sources, since the fMRI sources are known to have a super-Gaussian distribution [22] with low to moderate level of correlation, see supplementary material-S1. The sources are linearly mixed using a randomly generated mixing matrix for each dataset to obtain the observation set, $\mathbf{X}^{[k]}$. We generate three different observation sets by randomly generating the mixing matrices and sources for each set. We obtain 20 runs of standard IVA and acIVA using the IVA-L-SOS algorithm. For acIVA half of the total number of sources are used as constraints, $L = N/2$, $\gamma_n = 3$ and the set $\mathcal{P}$ is defined as $0.1, \ldots, 0.9$. The reference signal for the constraint source is obtained by computing the mean of the SCV. We measure the performance of the two methods using jISI, which measures the estimation of the whole demixing matrices, and spatial correlation, which measures the estimation of the individual sources of interest. For standard IVA, we align the estimated sources with respect to the original sources, whereas for acIVA no additional alignment step is required to align the constrained sources. The average of the jISI metric and spatial correlation of the constrained sources across 20 runs and three observation sets for IVA and acIVA is shown in Fig. 3(a) and Fig. 3(b) respectively.

From Fig. 3, we observe that IVA performance improves with an increase in number of datasets upto a certain limit for different numbers of sources. This range can be defined as the best range in which there are sufficient samples available for IVA to accurately estimate the underlying parameters. In this range, however, the application of acIVA does not improve performance compared with IVA, indicating that this is the best IVA performance one can obtain, for the fixed choice of algorithm and number of samples. Although the increase in number of sources for a fixed number of datasets does not significantly affect the performance, it does determine the limit for better performance. The range of better performance becomes tighter as we increase the number of sources, after which we observe a degradation in performance due to the effect of high dimensionality. The application of acIVA in this high dimensional range however shows a significant improvement in performance, indicating that the use of prior information is providing reference to the search for a better solution in high dimensional scenarios. The spatial correlation of the estimated constrained sources and the corresponding ground truth shows a similar trend as the jISI results. A high spatial correlation also suggests that the acIVA method does not enforce constraints on the decomposition and efficiently

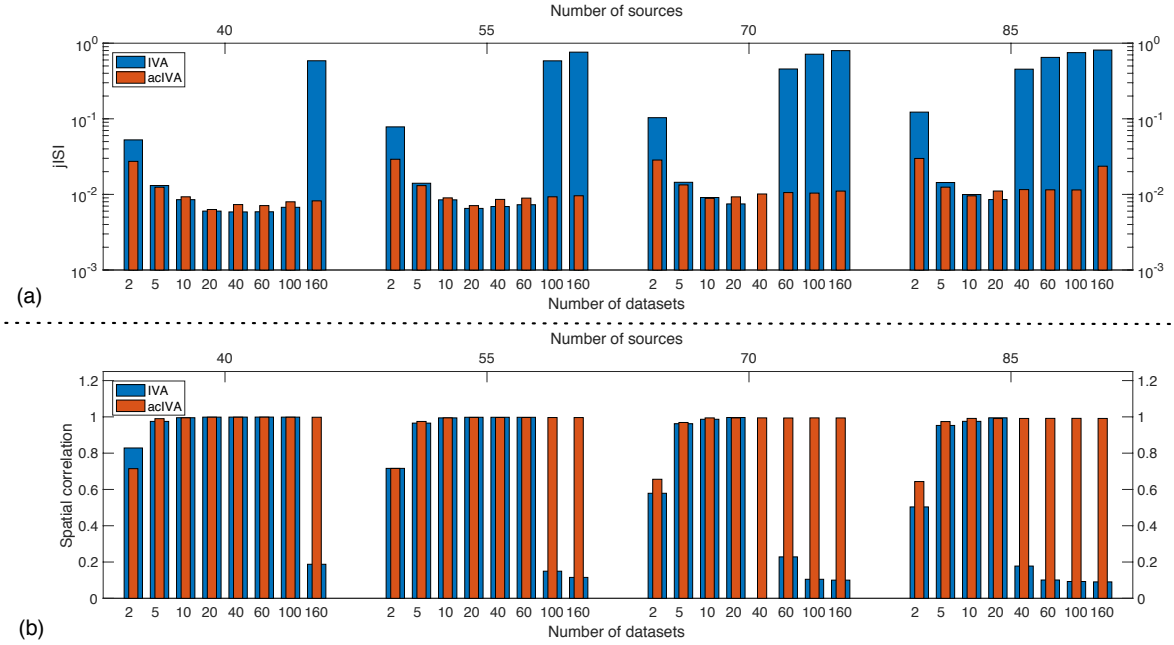

Fig. 3. Performance of IVA and acIVA in terms of jISI (a) and spatial correlation (b), with respect to number of sources, N and number of datasets, K. The number of sources, N, is varied from 40 to 85, the number of datasets, K, is varied from 2 to 60, and the number of samples, V, is fixed at 1000. For acIVA, half of the components are constrained, *i.e.*, $L = N/2$. For $K = 40$ and $N = 40$ none of the IVA runs converged in 1024 iterations. (a) The jISI is averaged across all converged runs for both acIVA and IVA. We observe degradation in performance of IVA for higher number of datasets, across different values of N whereas acIVA provides improvement in performance compared to IVA. (b) The spatial correlation of the constrained sources with ground truth is averaged across all converged runs. We see that acIVA was able to recover the constrained sources in high-dimensional scenarios, however IVA showed poor performance.

estimates the spatial components across datasets. The acIVA method is able to estimate the underlying sources significantly better than standard IVA in high dimensional scenarios.

V. APPLICATION OF ACIVA TO REAL FMRI DATA

We apply acIVA on a large-scale real fMRI dataset acquired from 327 subjects (164 female and 163 male). All images were collected on a 3-Tesla Siemens TIM Trio scanner with a 12-channel radio frequency coil. T2*-weighted functional images were acquired using a gradient-echo EPI sequence with TE = 29 milliseconds, TR = 2 seconds, flip angle = 75°, slice gap = 1.05 millimeters (mm), slice thickness = 3.5 mm, field of view = 240 mm, matrix size = 64 × 64, voxel size = 3.75 mm × 3.75 mm × 4.55 mm. The participants were asked to keep their eyes open during the scan and stare passively at a fixation point for 5 minutes, 4 seconds (152 volumes). Any additional volumes were discarded to match data quantity across participants. Images were realigned using INRIalign, and slice-timing corrected using the middle slice as the reference frame. Data are then spatially normalized into the standard Montreal Neurological Institute (MNI) space, resliced to 3 mm × 3 mm × 3 mm voxels, and smoothed using a Gaussian kernel with a full-width at half-maximum (FWHM) of 10 mm. Masking using the group ICA for fMRI toolbox (GIFT) was performed on each volume to remove the non-brain voxels and vectorized, resulting in an observation set for the kth subject as, $\bar{\mathbf{X}}^{[k]} \in \mathbb{R}^{152\times 58541}$.

The reference signals for acIVA are extracted using gICA [11], [12]. The gICA technique is shown in Fig. 4. Given K datasets from K subjects, gICA performs subject-specific

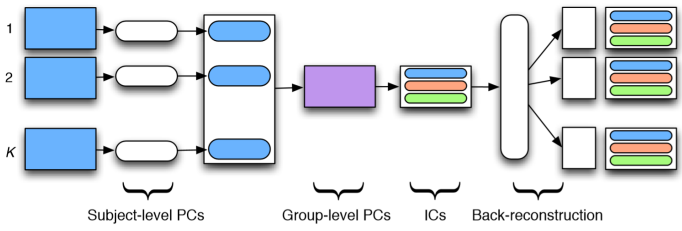

Fig. 4. Given K datasets, gICa performs subject-level PCA in order to obtain a signal subspace for each subject. The signal subspace from all subjects is concatenated to form a tall data matrix followed by applying a group-level PCA to extract a common signal subspace. ICA is applied on the group-level PCs in order to estimate independent components, which are back-reconstructed to obtain subject-specific time courses and spatial maps.

principal component analysis (PCA) in order to remove noise from each dataset, followed by a group-level PCA in order to extract a common signal subspace from all subjects. In order to account for higher order statistics, gICA performs ICA on the group-level principal components (PCs) and the estimated independent components (ICs) are back-reconstructed to obtain subject-specific spatial maps and time courses, as shown in Fig. 4. The number of components is estimated as the mean plus one standard deviation of the orders computed across all subjects. The mean and standard deviation of the orders estimated using the entropy rate based order selection by finite memory length model [27], which incorporates sample dependence into the information theoretic criteria, is 79.56 and 8.98 respectively. We select the final order as 90 (mean plus one standard deviation). ICA using the entropy-rate bound minimization algorithm [28] is applied on the subject datasets. Out of the 90 components, we select $L = 42$ group-level

independent components (ICs) corresponding to functionally relevant resting-state networks by visual inspection and these components are used as reference signals in acIVA.

We perform acIVA on male and female group separately using the group ICA components as reference signals. Each subject's data is dimension reduced using principal component analysis in order to select the most informative components and $N = 55$ uncorrelated components are used to form a dimension reduced dataset, $\mathbf{X}^{[k]} \in \mathbb{R}^{55\times 58541}$ for each subject. Five runs of acIVA using the IVA-L-SOS algorithm are obtained with different initialization and the best run is selected using the cross jISI method, which is an extension of the method proposed in [29] to multiple datasets. The estimated demixing matrices of the selected run are used to compute the sources. The estimated components are grouped into 8 domains: auditory (AUD), motor, parietal, fronto-parietal (FP), frontal, visual, default mode network (DMN) and cerebellum (CB). The components in each domain are shown in supplementary material-S2[1]. We also applied standard IVA on the datasets from each group, however, it did not converge in 1024 iterations. The acIVA technique converged in approximately 300 iterations.

A. Identification of functional networks

The use of reference signals in acIVA improves the identification of functional networks due to the adaptive nature allowing the functionally connected regions to interact in a more flexible and data-driven manner. It naturally groups the regions that have correlated and anti-correlated relationship with the reference signal. The default mode network (DMN) is a large-scale brain network of interconnected brain regions that form hubs and subsystems. It is commonly known to be activated when the person is in the resting-state with thoughts pertaining to oneself, others, the past and future, and hence is one of the widely explored network in various disorders. The core functional hubs of DMN are located in the medial prefrontal cortex (mPFC), posterior cingulate cortex (PCC), Precuneus and angular gyrus (AG). The acIVA method identified the core default mode network in IC 31, when only mPFC is used as the reference signal, as shown in Fig. 5. This IC also shows activation in the regions of the central executive network namely the dorsolateral prefrontal cortex (dlPFC) and the posterior parietal cortex (PPC), which has shown anti-correlation with the default mode network [30]. Similarly for IC 32, the ventrolateral prefrontal cortex (vlPFC) is extracted along with posterior DMN (PCC+AG) using acIVA for a corresponding reference signal showing activation in the AG. For IC 33, the anterior cingulate cortex (ACC) and regions are extracted along Precuneus. The ACC and INS regions form the salience network that plays a critical role in switching between DMN and CEN [30], [31]. The reference signal in IC 34 consists of the regions associated with the dorsal medial subsystem [32], namely the temporoparietal junction (TPJ), lateral temporal cortex (LTC), temporal pole (TempP). The dorsal medial subsystem also consists of dmPFC which is extracted using acIVA alongwith the reference signal. Precuneus and anterior mPFC (amPFC) have shown strong association with the dorsal medial subsystem and act as functional hubs for information transfer across the subsystem [32].

[1]Supplementary materials are available in the supporting documents /multimedia tab.

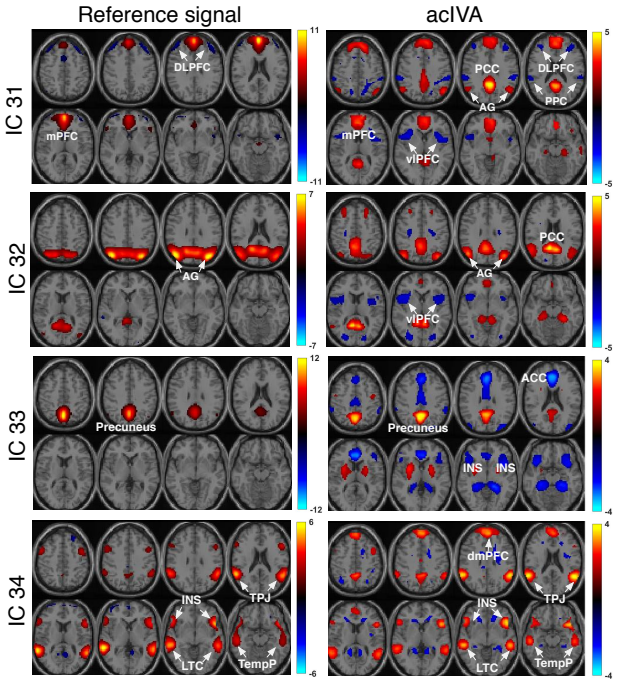

Fig. 5. Components estimated using acIVA and corresponding reference signal, which are the group-level components of GICA. Our results indicate that the proposed acIVA technique naturally groups together functional networks, *e.g.*, acIVA identified the whole default mode network (PCC, mPFC, AG) when only the anterior DMN (mPFC) was used as the reference signal. (Abbreviations: medial prefrontal cortex (mPFC), anterior cingulate cortex (ACC), posterior cingulate cortex (PCC), angular gyrus (AG), dorsolateral prefrontal cortex (dlPFC), posterior parietal cortex (PPC), ventrolateral prefrontal cortex (vlPFC), insular (INS), dorsal medial prefrontal cortex (dmPFC), temporoparietal junction (TPJ), lateral temporal cortex (LTC), temporal pole (TempP).

B. Performance of acIVA and gICA in terms of preserving subject-variability

In this section, we study the performance of acIVA in terms of its ability to preserve subject-specific information using two techniques, namely, 'variability maps' and capturing gender differences in spectral power, and compare its performance with the widely-used gICA technique for fMRI analysis [11], [12]. The acIVA technique computes subject-specific spatial maps and time courses, where the dependent components are grouped together to form an SCV, as discussed in Section II. In order to obtain subject-specific spatial maps and time courses for gICA, the group level ICs/reference signals are back-reconstructed using PCA-based back-reconstruction [22].

The variability map for each component is obtained as the standard deviation at each voxel computed across subjects. The results for the variability maps for the components associated with the DMN are shown in Fig. 6. The maps from acIVA demonstrate high standard deviation across subjects, at voxels corresponding to meaningful regions in the DMN and dorsal medial subsystem, whereas gICA demonstrates lower standard deviation at these voxels. This suggest that since gICA performs a significant dimensionality reduction step in the group-level PCA stage, most of the variability associated

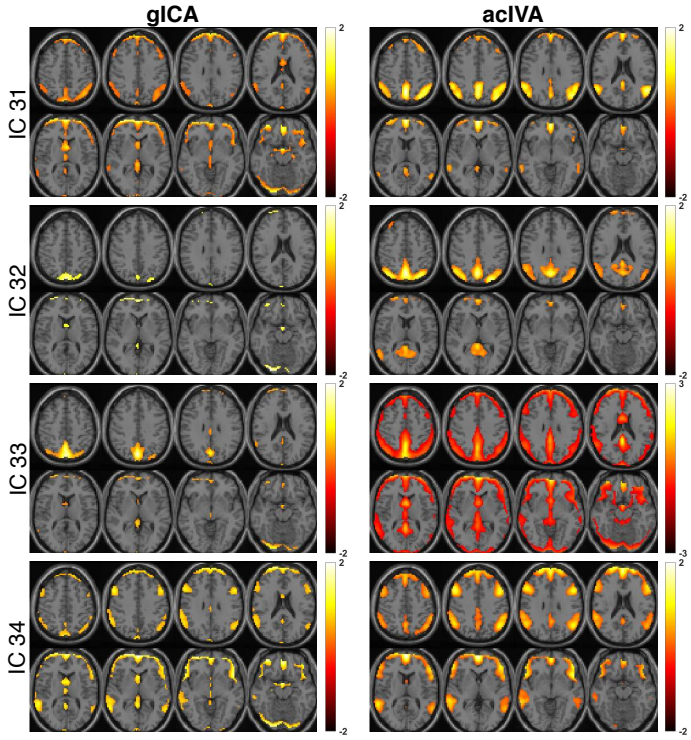

Fig. 6. Variability maps (threshold at 1) for acIVA and gICA. High standard deviation is observed at voxels corresponding to the DMN and dorsal medial subsystem for acIVA as compared to gICA.

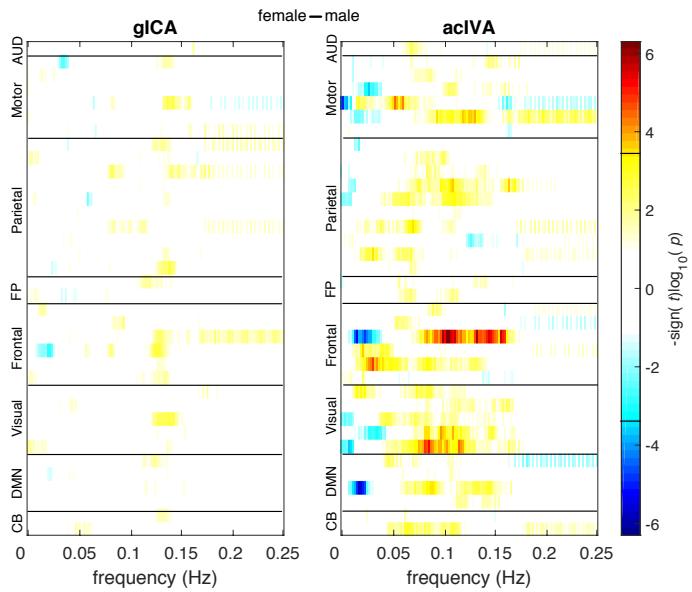

Fig. 7. Frequency range that demonstrates significant ($p < 0.05$) difference between male and female group, displayed as $-\text{sign}(t)\log_{10}(p)$. Hot colors demonstrate higher spectral power in female group whereas cold colors indicate high spectral power in male group. The line on the colorbar indicates the FDR-corrected threshold ($p = 0.05$) for significance. Higher spectral power is observed in the male group at low frequencies (< 0.05 Hz) for the motor, frontal, visual and DMN components, whereas females show higher spectral power at high frequencies (0.05 to 0.15 Hz). The acIVA technique better captures gender differences than gICA technique, indicating its ability to preserve subject variability.

with individual subjects is lost. Hence, the subject-specific components from gICA are mostly centered around the group ICs with a low standard deviation. However, since acIVA does not perform the group-level dimensionality reduction step, it is able to capture the variability associated with each subject resulting in a higher standard deviation.

We compare the performance of acIVA and gICA in terms of capturing gender differences in spectral power. We obtain the component power spectra for all subjects using the time courses from acIVA and gICA, and identify differences using a two-sample t-test between male and female group at each frequency level. The components showing significant difference ($p < 0.05$, corrected) are shown in Fig. 7. Significantly higher spectral power is observed in the male group at low frequencies (< 0.05 Hz) in few motor (ICs 5), frontal (IC 22) and DMN (IC 33) components. High spectral power in the motor (IC 5) and frontal component (IC 22) is also observed in a similar study [1]. Significantly higher spectral power is observed in the female group at higher frequencies (0.05 to 0.15 Hz) in the frontal component (IC 22) and visual component (IC 30). In general, our results show high spectral power in the female group in the frequency range 0.05 to 0.15 Hz across the motor, parietal, frontal, visual, DMN and cerebellum components, although not significant.

VI. DISCUSSION

IVA is a joint blind source separation technique that has been shown to efficiently capture subject variability from multisubject fMRI data [17], [18], [19]. It groups the subject-specific local patterns pertaining to a common, global attribute in a meaningful manner, in a sample sufficient regime, *i.e.*, when the number of samples are sufficient to estimate the multivariate SCVs exhibiting different levels of correlation. Its performance improves as we increase the number of datasets due to the availability of more information across datasets, provided we are within the 'sufficient' sample regime. However, with a further increase in dimensionality—number of datasets—, the performance of IVA drops for a fixed number of samples, indicating a tradeoff between the taking advantage of more information across datasets and curse of dimensionality. This drop in performance depends on the number of sources, *i.e.*, for a higher number of sources, a performance drop is observed for a lower number of datasets. Hence for IVA, high dimensionality is a function of the number of samples and sources. With an availability of fMRI scanners with a higher temporal resolution, there is an increase in number of volumes for each subject that captures more information. For this type of data, the estimated model order—number of sources—is expected to be higher resulting in degradation in performance at a lower value of K. In high dimensional scenarios, IVA is able to recover the highly correlated SCVs better than the low correlated SCVs since these SCVs carry more common information across datasets and are a better match to the model assumption. In this work, we propose the acIVA method to estimate the underlying high dimensional SCVs in order to overcome the limitations of IVA.

The acIVA algorithm efficiently incorporates prior information regarding the sources or the columns of the mixing matrix, into the IVA decomposition and we demonstrate its potential use in the analysis of large-scale datasets. Our results from Fig. 3 indicate that acIVA provides reliable and meaningful estimation of the underlying sources when there

are insufficient samples available with respect to the number of sources and datasets. Although acIVA demonstrated superior performance than standard IVA for higher number of datasets, it is not impervious to the effect of high dimensionality. As the number of dimensions approaches the number of samples or in the sample-poor regime, where the number of samples are less than the number of dimensions, the performance of acIVA is expected to drop, as observed in Fig. 3 at $N = 85$ and $K = 160$. However, this drop in acIVA performance is observed at a more extreme case as compared with standard IVA. Additionally, the jISI results from Fig. 3(a) show that the performance of acIVA is slightly affected in the sufficient sample regime for some cases, *e.g.*, $N = \{40, 55\}$, $K = \{40, 60\}$. The spatial correlation of the constrained sources for the corresponding points demonstrates similar performance as IVA, indicating that the constrained sources or the sources of interest are estimated accurately, however the estimation of the unconstrained sources is penalized to some degree. This effect can be reduced by incorporating prior information regarding the remaining sources. It should also be noted that the performance of acIVA depends on the number of constraints, i.e., acIVA performance might improve further if prior information regarding all the sources is incorporated.

Data-driven techniques are widely-used for the analysis of fMRI data due to their flexible nature that allows for identification of natural relationships that were not derived a-priori [33], [22]. On the other hand model-driven methods such as region-of-interest based methods make stronger assumptions regarding the nature of the decomposition, and are robust to noise and other artifacts. Typically model-driven methods extract functionally connected regions from resting-state fMRI data by identifying voxels that have correlated activation patterns with a pre-defined region-of-interest or a seed voxel. These methods usually outperform the data-driven techniques only when prior information is accurate [34]. The acIVA technique provides a desirable balance between data-driven and model-driven techniques, and takes advantage of the robustness properties of model-driven methods and flexible nature of data-driven techniques. A comparison of data-driven, model-driven and semi-BSS method on task-related fMRI data demonstrated robust performance of semi-blind ICA in the presence of noise and when the prior information is not completely accurate [34]. The acIVA technique adaptively tunes the constraint parameter in order to efficiently incorporate the prior information, *i.e.*, does not impose inaccurate prior information on the decomposition. The use of adaptive parameter tuning technique also allows for flexibility for the functionally connected regions to fully interact and does not enforce the sources to be exactly similar to the reference component, as shown in Fig. 5.

VII. CONCLUSION

In this paper, we emphasize the need for flexible methods that can scale to handle large-scale fMRI data. We discuss the benefits of IVA in terms of providing a flexible framework and in terms of its ability to better capture the subject-specific information. However, we demonstrate that the performance of IVA depends on a number of factors such as number of samples, number of datasets, number of sources and the level of correlation between the marginals of the multivariate sources. To summarize, the IVA performance degrades with decrease in number of samples and the level of correlation within the SCVs, and with an increase in datasets and number of sources. We discuss the acIVA technique and its extension to incorporate multiple reference signals that efficiently guides the solution and does not enforce inaccurate prior information on the IVA decomposition, thus achieving a desirable balance between data-driven and model-driven techniques. We demonstrate its ability to overcome the main challenges of IVA, namely, the effect of high dimensionality. We also apply acIVA on a large-scale resting-state fMRI data and show that acIVA efficiently estimates functionally connected brain regions.

REFERENCES

[1] 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," *Frontiers in Systems Neuroscience*, vol. 5, p. 2, 2011.

[2] L. Q. Uddin, K. S. Supekar, S. Ryali, and V. Menon, "Dynamic reconfiguration of structural and functional connectivity across core neurocognitive brain networks with development," *Journal of Neuroscience*, vol. 31, no. 50, pp. 18 578–18 589, 2011.

[3] E. Damaraju, E. A. Allen, A. Belger, J. Ford, S. McEwen, D. Mathalon, B. Mueller, G. Pearlson, S. Potkin, A. Preda *et al.*, "Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia," *NeuroImage: Clinical*, vol. 5, pp. 298–308, 2014.

[4] B. Rashid, E. Damaraju, G. D. Pearlson, and V. D. Calhoun, "Dynamic connectivity states estimated from resting fMRI identify differences among Schizophrenia, bipolar disorder, and healthy control subjects," *Frontiers in Human Neuroscience*, vol. 8, p. 897, 2014.

[5] M. Assaf, K. Jagannathan, V. D. Calhoun, L. Miller, M. C. Stevens, R. Sahl, J. G. O'Boyle, R. T. Schultz, and G. D. Pearlson, "Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients," *NeuroImage*, vol. 53, no. 1, pp. 247–256, 2010.

[6] Z. Fu, Y. Tu, X. Di, Y. Du, J. Sui, B. B. Biswal, Z. Zhang, N. de Lacy, and V. D. Calhoun, "Transient increased thalamic-sensory connectivity and decreased whole-brain dynamism in autism," *NeuroImage*, 2018.

[7] N. de Lacy and V. D. Calhoun, "Dynamic connectivity and the effects of maturation in youth with attention deficit hyperactivity disorder," *Network Neuroscience*, vol. 3, no. 1, pp. 195–216, 2018.

[8] S.-J. Kim, V. D. Calhoun, and T. Adalı, "Flexible large-scale fMRI analysis: A survey," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 6319–6323.

[9] Y. Levin-Schwartz, V. D. Calhoun, and T. Adalı, "Quantifying the interaction and contribution of multiple datasets in fusion: application to the detection of schizophrenia," *IEEE transactions on medical imaging*, vol. 36, no. 7, pp. 1385–1395, 2017.

[10] E. Acar, Y. Levin-Schwartz, V. D. Calhoun, and T. Adalı, "Tensor-based fusion of EEG and FMRI to understand neurological changes in schizophrenia," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2017, pp. 1–4.

[11] V. Calhoun, T. Adalı, G. Pearlson, and J. Pekar, "Group ICA of functional MRI data: Separability, stationarity, and inference," in *Proc. Int. Conf. on ICA and BSS San Diego, CA. p*, vol. 155, 2001.

[12] V. D. Calhoun, T. Adalı, G. D. Pearlson, and J. Pekar, "A method for making group inferences from functional MRI data using independent component analysis," *Human Brain Mapping*, vol. 14, no. 3, pp. 140–151, 2001.

[13] G. Varoquaux, A. Gramfort, F. Pedregosa, V. Michel, and B. Thirion, "Multi-subject dictionary learning to segment an atlas of brain spontaneous activity," in *Biennial International Conference on Information Processing in Medical Imaging*. Springer, 2011, pp. 562–573.

[14] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, 2015.

[15] A. H. Phan and A. Cichocki, "PARAFAC algorithms for large-scale problems," *Neurocomputing*, vol. 74, no. 11, pp. 1970–1984, 2011.

[16] S. K. Suter, M. Makhynia, and R. Pajarola, "TAMRESH–tensor approximation multiresolution hierarchy for interactive volume visualization," in *Computer Graphics Forum*, vol. 32, no. 3pt2. Wiley Online Library, 2013, pp. 151–160.
[17] J. Laney, K. P. Westlake, S. Ma, E. Woytowicz, V. D. Calhoun, and T. Adalı, "Capturing subject variability in fMRI data: A graph-theoretical analysis of GICA vs. IVA," *Journal of neuroscience methods*, vol. 247, pp. 32–40, 2015.
[18] A. M. Michael, M. Anderson, R. L. Miller, T. Adalı, and V. D. Calhoun, "Preserving subject variability in group fMRI analysis: Performance evaluation of GICA vs. IVA," *Frontiers in Systems Neuroscience*, vol. 8, p. 106, 2014.
[19] T. Adali, M. Anderson, and G.-S. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 18–33, 2014.
[20] M. Anderson, G.-S. Fu, R. Phlypo, and T. Adalı, "Independent vector analysis, the Kotz distribution, and performance bounds," in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 3243–3247.
[21] S. Bhinge, R. Mowakeaa, V. D. Calhoun, and T. Adalı, "Extraction of time-varying spatio-temporal networks using parameter-tuned constrained IVA," *IEEE Transactions on Medical Imaging*, vol. 38, no. 7, pp. 1715–1725, Jul. 2019.
[22] V. D. Calhoun and T. Adalı, "Multisubject independent component analysis of fMRI: A decade of intrinsic networks, default mode, and neurodiagnostic discovery," *IEEE Reviews in Biomedical Engineering*, vol. 5, pp. 60–73, 2012.
[23] S. Bhinge, Q. Long, Y. Levin-Schwartz, Z. Boukouvalas, V. D. Calhoun, and T. Adalı, "Non-orthogonal constrained independent vector analysis: Application to data fusion," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 2666–2670.
[24] A. H. Khan, M. Taseska, and E. A. Habets, "A geometrically constrained independent vector analysis algorithm for online source extraction," in *International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2015, pp. 396–403.
[25] M. Anderson, T. Adalı, and X.-L. Li, "Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1672–1683, 2012.
[26] Z. Boukouvalas, G.-S. Fu, and T. Adalı, "An efficient multivariate generalized gaussian distribution estimator: Application to IVA," in *2015 49th Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2015, pp. 1–4.
[27] G.-S. Fu, M. Anderson, and T. Adalı, "Likelihood estimators for dependent samples and their application to order detection," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4237–4244, 2014.
[28] G.-S. Fu, R. Phlypo, M. Anderson, X.-L. Li *et al.*, "Blind source separation by entropy rate minimization," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4245–4255, 2014.
[29] Q. Long, C. Jia, Z. Boukouvalas, B. Gabrielson, D. Emge, and T. Adalı, "Consistent run selection for independent component analysis: Application to fMRI analysis," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2581–2585.
[30] D. Sridharan, D. J. Levitin, and V. Menon, "A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks," *Proceedings of the National Academy of Sciences*, vol. 105, no. 34, pp. 12 569–12 574, 2008.
[31] V. Menon and L. Q. Uddin, "Saliency, switching, attention and control: a network model of insula function," *Brain Structure and Function*, vol. 214, no. 5-6, pp. 655–667, 2010.
[32] J. R. Andrews-Hanna, J. Smallwood, and R. N. Spreng, "The default network and self-generated thought: component processes, dynamic control, and clinical relevance," *Annals of the New York Academy of Sciences*, vol. 1316, no. 1, pp. 29–52, 2014.
[33] V. D. Calhoun and N. de Lacy, "Ten key observations on the analysis of resting-state functional MR imaging data using independent component analysis," *NeuroImaging Clinics*, vol. 27, no. 4, pp. 561–579, 2017.
[34] V. D. Calhoun, T. Adali, M. Stevens, K. 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.

Suchita Bhinge received the PhD degree in Electrical Engineering and M.S. degree in Electrical Engineering in 2020 and 2015 respectively, from University of Maryland, Baltimore County, Maryland, U.S.A, and B.S. in Electronics and Telecommunication from University of Pune, India in 2013. Her research focuses on developing unsupervised machine learning models by integrating concepts from matrix and tensor factorization, information theory and optimization methods. Her interests include matrix decomposition, machine learning, deep learning, medical imaging analysis, data fusion and video processing.

Qunfang Long received the B.S. degree in Electronic Science & Technology and M.S. degree in Microelectronics & Solid-State Electronic from the Wuhan University, Wuhan, China, in 2012 and 2015 respectively, and the M.S. degree in Electrical Engineering from the University of Maryland, Baltimore County in 2018. Her research interests include neuroimaging analysis, blind source separation, data fusion, and machine learning for signal processing.

Vince D. Calhoun is founding director of the tri-institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS) and a Georgia Research Alliance eminent scholar in brain health and image analysis where he holds appointments at Georgia State University, Georgia Institution of Technology and Emory University. He was previously the President of the Mind Research Network and Distinguished Professor of Electrical and Computer Engineering at the University of New Mexico. He is the author of more than 800 full journal articles and over 850 technical reports, abstracts and conference proceedings. Dr. Calhoun is a fellow of the Institute of Electrical and Electronic Engineers, The American Association for the Advancement of Science, The American Institute of Biomedical and Medical Engineers, The American College of Neuropsychopharmacology, and the International Society of Magnetic Resonance in Medicine. He served at the chair for the Organization for Human Brain Mapping from 2018-2019 is a past chair of the IEEE Machine Learning for Signal Processing Technical Committee. He currently serves on the IEEE BISP Technical Committee and is also a member of IEEE Data Science Initiative Steering Committee.

Tülay Adalı (S'89–M'92–SM'98–F'09) received the Ph.D. degree in Electrical Engineering from North Carolina State University, Raleigh, NC, USA, in 1992 and joined the faculty at the University of Maryland Baltimore County (UMBC), Baltimore, MD, the same year. She is currently a Distinguished University Professor in the Department of Computer Science and Electrical Engineering at UMBC. She has been active in conference and workshop organizations, in particular within the IEEE. She is currently serving as the Vice President for Technical Directions of the IEEE Signal Processing Society. Prof. Adali is a Fellow of the IEEE and the AIMBE, a Fulbright Scholar, and an IEEE Signal Processing Society Distinguished Lecturer. She is the recipient of a 2020 Humboldt Research Award and past recipient of a 2010 IEEE Signal Processing Society Best Paper Award, 2013 University System of Maryland Regents' Award for Research, and an NSF CAREER Award. Her current research interests are in the areas of statistical signal processing, machine learning, and their applications with emphasis on applications in medical image analysis and fusion.