NON-ORTHOGONAL CONSTRAINED INDEPENDENT VECTOR ANALYSIS: APPLICATION TO DATA FUSION

Suchita Bhinge[1,*], *Qunfang Long*[1,*], *Yuri Levin-Schwartz*[1], *Zois Boukouvalas*[2], *Vince D. Calhoun*[3,4] *and Tülay Adalı*[1]
[1]University of Maryland, Baltimore County, Dept. of CSEE, MD, 21250
[2]University of Maryland, Baltimore County, Dept. of Mathematics and Statistics, MD, 21250
[3]University of New Mexico, Dept. of ECE, Albuquerque, NM, 87131
[4]The Mind Research Network, Albuquerque, NM, 87106

ABSTRACT

The existence of complementary information across multiple sensors has driven the proliferation of multivariate datasets. Exploitation of this common information, while minimizing the assumptions imposed on the data has led to the popularity of data-driven methods. Independent vector analysis (IVA), in particular, provides a flexible and effective approach for the fusion of multivariate data. In many practical applications, important prior information about the data exists and incorporating this information into the IVA model is expected to yield improved separation performance. In this paper, we propose a general formulation for non-orthogonal constrained IVA (C-IVA) framework that can incorporate prior information about either the sources or the mixing coefficients into the IVA cost function. A powerful decoupling method is the major enabling factor in this task. We demonstrate the improved performance of C-IVA over the unconstrained IVA model using both simulated as well as real medical imaging data.

Index Terms— Constrained optimization, data fusion, independent vector analysis, multivariate

1. INTRODUCTION

The use of multiple sensors has become common in many fields, since each sensor is expected to provide complementary information about the system under study [1, 2, 3, 4]. The desire to exploit this common information across datasets has driven the development of joint blind source separation (JBSS) techniques, such as independent vector analysis (IVA) [5]. IVA, a recent extension of independent component analysis (ICA) to multiple datasets, uses the similarities across datasets to achieve a powerful decomposition and has been used in many applications, see *e.g.*, [6, 7, 8, 9]. The success of IVA is due to its use of a simple generative model that enables flexibility while minimizing the assumptions placed on the data. In many applications, important prior information about the data is available and incorporating this information into the IVA framework is expected to improve the extraction of the true latent sources by providing a better model match and relaxing the independence assumption in the case of ICA and IVA [10].

There has been a fair amount of interest in incorporating prior knowledge into the ICA framework, see *e.g.*, [11, 12, 13], however these methods cannot be directly extended to the IVA framework since they would not account for source dependence across datasets, an important form of diversity exploited by IVA. On the other hand, a method was developed in [14] that takes advantage of this diversity, however, the mathematical framework is limited by the application, speech processing, and requires the demixing matrices to be orthogonal, thus limiting the solution space.

In this paper, we propose a general mathematical formulation for constrained independent vector analysis (C-IVA) that can incorporate any form of prior information about either the sources or the mixing vectors, without the requirement of orthogonality on the part of the demixing matrices. The new C-IVA framework combines the flexibility of data-driven methods and the robustness to noise and other artifacts, of model-based methods, in addition to exploiting dependence across datasets, making it a useful tool for source separation. In this work, we apply the IVA-Gaussian (IVA-G) algorithm [15] to the C-IVA framework and test its performance using simulated dataset and a multitask fusion dataset. C-IVA improves the estimation of the underlying sources for the simulated dataset and demonstrates significant increase in the classification of subjects as patients and healthy controls by improving the estimation of discriminative biomarkers.

The paper is organized as follows. Section 2 describes the general IVA model. In Section 3, we talk about decoupling of demixing vectors and a general framework for incorporating constraints into the IVA objective function. Section 4 describes the comparison results followed by the conclusion in Section 5.

2. INDEPENDENT VECTOR ANALYSIS

Given K datasets, $\mathbf{X}^{[k]} \in \mathbb{R}^{N \times P}$, where N is the number of observations and P is the number of samples, the IVA model

*Suchita Bhinge and Qunfang Long contributed equally to this work.

**This work was supported in part by NSF-CCF 1618551, NIH R01 EB 020407 and NIH R01 EB 005846.

is given by

$$\mathbf{X}^{[k]} = \mathbf{A}^{[k]}\mathbf{S}^{[k]}, \ k = 1, \ldots, K, \tag{1}$$

where $\mathbf{A}^{[k]} \in \mathbb{R}^{N\times N}$ is the mixing matrix and $\mathbf{S}^{[k]} \in \mathbb{R}^{N\times P}$ are the latent independent sources for the kth dataset. The goal of IVA is to estimate K demixing matrices $\mathbf{W}^{[k]}$ to compute source estimates using $\mathbf{Y}^{[k]} = \mathbf{W}^{[k]}\mathbf{X}^{[k]}$. This can be achieved by minimizing the mutual information (MI) between the SCVs and the cost function is given by

$$\mathcal{I}_{IVA}(\mathcal{W}) = \sum_{n=1}^{N} \mathcal{H}\left(\mathbf{y}_n\right) \sum_{k=1}^{K} \log|\det \mathbf{W}^{[k]}| - C = \sum_{n=1}^{N} \left\{ \sum_{k=1}^{K} \mathcal{H}\left(y_n^{[k]}\right) - \mathcal{I}\left(\mathbf{y}_n\right) \right\} - \sum_{k=1}^{K} \log|\det \mathbf{W}^{[k]}| - C, \tag{2}$$

where $\mathcal{W} = \left\{\mathbf{W}^{[1]}, \ldots, \mathbf{W}^{[K]}\right\}$, $\mathcal{H}(\cdot)$ is the entropy and $\mathcal{I}(\cdot)$ is the mutual information. The last term in (2), C, is the entropy of $\left[\mathbf{X}^{[1]}, \ldots, \mathbf{X}^{[K]}\right]$ and is constant with respect to $\mathbf{W}^{[k]}$, thus, can be ignored. The MI term, $\mathcal{I}(\cdot)$, in (2) plays an important role in exploiting the complementary information across all the datasets, since it maximizes the dependence within each source component vector (SCV), where an SCV is defined by concatenating the nth source from each of the K datasets, *i.e.*, $\mathbf{y}_n = [y_n^{[1]}, \ldots, y_n^{[K]}]^{\mathrm{T}}$. Without the MI term, the IVA cost function would be equivalent to performing ICA on each dataset separately [15].

Algorithm choice plays an important role in the estimation of the latent sources in ICA and IVA. For applications that have many samples, algorithms that exploit both second order statistics (SOS) and higher order statistics (HOS) [16] are a suitable match. However, for sample-poor cases, an algorithm that exploits SOS, such as IVA-G [15], is a preferable choice, since estimation of HOS is unreliable in this regime.

3. CONSTRAINED IVA

Since prior information about all the sources and demixing vectors is usually not available, it is impractical to attempt to exploit prior information directly in (2). This issue can be avoided by re-expressing (2) with respect to each row of the demixing matrix, $\mathbf{w}_n^{[k]}$ as

$$\mathcal{I}^d_{IVA}(\mathbf{w}_n^{[k]}) = \mathcal{H}(\mathbf{y}_n) - \log\left|\left(\mathbf{h}_n^{[k]}\right)^T \mathbf{w}_n^{[k]}\right|, \tag{3}$$

where $\mathbf{h}_n^{[k]}$ is a unit vector perpendicular to all rows of $\mathbf{W}^{[k]}$ except $\mathbf{w}_n^{[k]}$ [15, 17]. Decoupling primarily allows us to constrain the demixing vectors or sources individually without assuming that $\mathbf{W}^{[k]}$ is orthogonal and thus limiting the solution space. Additionally, decoupling enables adaptive updating of the learning parameter for each direction, which benefits the optimization for complicated objective functions.

Wewill now define the C-IVA framework using the decoupled objective function defined in (3). Given an inequality constraint function g_n, the MI objective function is optimized subject to the constraint

$$g_n(\mathbf{w}_n^{[k]}, \mathbf{r}_n^{[k]}) = \rho_n - \epsilon(\mathbf{w}_n^{[k]}, \mathbf{r}_n^{[k]}) \leq 0, \tag{4}$$

where $\mathbf{r}_n^{[k]}$ is the reference vector for $\mathbf{w}_n^{[k]}$ or $\mathbf{y}_n^{[k]}$, ϵ is a distance measure and ρ_n is the constraint threshold. The proposed C-IVA framework allows flexibility in the definition of the distance measure, ϵ. Typical definitions of the distance measure include inner product, mean square error, mutual information and correlation-type distances. In this paper, we use Pearson correlation as the distance measure

$$\epsilon(\mathbf{w}_n^{[k]}, \mathbf{r}_n^{[k]}) = \begin{cases} \text{corr}\left((\mathbf{y}_n^{[k]})^T \mathbf{r}_n^{[k]}\right), & \text{for } \mathbf{y}_n^{[k]} \\ \text{corr}\left((\mathbf{w}_n^{[k]})^T \mathbf{r}_n^{[k]}\right), & \text{for } \mathbf{w}_n^{[k]} \end{cases}$$

By using the Pearson correlation as a distance measure, we restrict $\epsilon(\mathbf{w}_n^{[k]}, \mathbf{r}_n^{[k]})$ to be between 0 and 1, steering $\rho \leq 1$. Thus, a higher value of ρ imposes a harder constraint on the decomposition, while a lower value of ρ reduces the power of the constraint on the decomposition.

Following the C-ICA framework defined in [10], the inequality constraint in (4) can be incorporated in (3) by defining the augmented Lagrangian optimization function and ignoring the terms independent of $\mathbf{w}_n^{[k]}$

$$\mathcal{I}_n^c(\mathbf{w}_n^{[k]}) = \mathcal{H}(\mathbf{y}_n) - \log\left|\left(\mathbf{h}_n^{[k]}\right)^T \mathbf{w}_n^{[k]}\right| - \frac{1}{2\gamma_n} \left\{ \left[\max\{0, \mu_n^{[k]} + \gamma_n g_n(\mathbf{w}_n^{[k]}, \mathbf{r}_n^{[k]})\}\right]^2 - (\mu_n^{[k]})^2 \right\}, \tag{5}$$

where $\mu_n^{[k]}$ is the Lagrangian multiplier and $\gamma_n > 0$ is a learning parameter. Using the vector gradient descent method, the gradient update function of (5) can be written as

$$\frac{\partial \mathcal{I}^c_{IVA}(\mathbf{w}_n^{[k]})}{\partial \mathbf{w}_n^{[k]}} = E\left\{\phi^{[k]}(\mathbf{y}_n)\mathbf{x}^{[k]}\right\} - \frac{\mathbf{h}_n^{[k]}}{(\mathbf{h}_n^{[k]})^T \mathbf{w}_n^{[k]}} - g_n'(\mathbf{w}_n^{[k]}, \mathbf{r}_n^{[k]})\hat{\mu}_n^{[k]}(\mathbf{r}_n^{[k]})^T,$$

where g_n' is the derivative of g_n with respect to $((\mathbf{w}_n^{[k]})^T \mathbf{r}_n^{[k]})$ and $\phi^{[k]}(\mathbf{y}_n) = -\partial \log p_1(\mathbf{y_1})/\partial y_1^{[k]}$. The Lagrange multiplier is updated in each iteration using gradient ascent

$$\hat{\mu}_n^{[k]} \leftarrow \max\left\{0, \gamma_n^{[k]} g_n(\mathbf{w}_n^{[k]}, \mathbf{r}_n^{[k]}) + \mu_n^{[k]}\right\}.$$

4. RESULTS AND DISCUSSION

An important application where the exploitation of prior information is expected to provide significant advantages is the fusion of functional magnetic resonance imaging (fMRI) data. In this case, a natural form of prior information is the designation of the subjects as either patients with schizophrenia or healthy controls and incorporating this information into the fusion analysis is expected to alleviate the issue inherent to the fusion of fMRI data using the effective tIVA model [18], the limited sample size. The fusion model for each dataset comprises of P-dimensional features extracted from each of the N subjects belonging to patients with schizophrenia and healthy controls, thus $\mathbf{X}^{[k]} \in \mathbb{R}^{N\times P}$. The aim is to obtain spatial maps, $\mathbf{S}^{[k]}$, that demonstrate a significant difference between controls and patients, whose corresponding column in the mixing matrix, also referred to as the subject covariation, can be approximated using a step-type response. Thus,

in order to incorporate prior information related to the class designations of the subjects into the IVA framework, we constrain one of the sources to have a step-type response. Since the commonality across multiple datasets is provided through the subject covariations, we use the transposed IVA (tIVA) model [9] that is given by transposing (1)

$$\mathbf{X}'^{[k]} = (\mathbf{X}^{[k]})^T = (\mathbf{S}^{[k]})^T(\mathbf{A}^{[k]})^T = \mathbf{A}'^{[k]}\mathbf{S}'^{[k]}. \quad (6)$$

Note that by transposing (1), the role of observations and samples are interchanged. In general, IVA-G algorithm [15] is preferable since it makes full use of SOS making it reliable for the tIVA model, in which the number of samples, *i.e.*, number of subjects, is limited. We compare the performance of tIVA and constrained tIVA (C-tIVA) using both simulated data as well as real multi-task fMRI data.

4.1. Simulation results

The simulation data is generated such that it resembles the fusion model. We generate 3 datasets such that $\mathbf{X}'^{[1]}$ is $500\times N$, $\mathbf{X}'^{[2]}$ and $\mathbf{X}'^{[3]}$ are $10^5\times N$. The sources, *i.e.*, subject covariations, $\mathbf{S}'^{[k]}$, are of size $10\times N$. Out of the 10 SCVs, one SCV has a step-type response, while the remaining nine SCVs are generated from a multivariate Gaussian (MG) distribution. The step-type response is mixed with white Gaussian noise, enabling the definition of the signal-to-noise ratio (SNR) as the ratio of the variance of the true step to the variance of the additive noise. The corresponding spatial maps, $\mathbf{A}'^{[k]}$ are generated from a uncorrelated multivariate Laplacian distribution, since it provides a good approximation for the spatial maps [19]. Following the performance of 100 runs of tIVA and C-tIVA, a two-sample t-test is performed on the estimated subject covariations from the converged runs to detect a step-type response. The subject covariation with the highest absolute t-statistic is selected and the corresponding spatial map, $\hat{\mathbf{a}}_m'^{[k]}$, is correlated with the true spatial map, $\mathbf{a}_1'^{[k]}$. The average of all the correlation values over converged runs is obtained as a function of different number of samples, N, and SNR values, see Figures 1(a-b). The results in Figure 1(a) show higher correlation coefficient values for C-tIVA when compared to tIVA when the number of samples is low, indicating that exploiting prior information is advantageous in the sample-poor case. Additionally, adding a constraint improves the estimation of the bimodal step-type response for an algorithm that assumes the sources have a unimodal distribution. Figure 1(b) indicates higher correlation coefficient values for C-tIVA when the noise level is high. This demonstrates that incorporating prior information into the IVA framework makes the decomposition more robust to noise.

In order to study if the introduction of a constraint into the IVA cost function results in artificial estimation of discriminative subject covariations, we explore the effect of the constraint, by adjusting the value of ρ in the case where there is no true step-type response. We generate all 10 SCVs from an MG distribution such that the SCVs do not have a step-type response and perform tIVA and C-tIVA on them, followed by performing a two-sample t-test on each dataset and

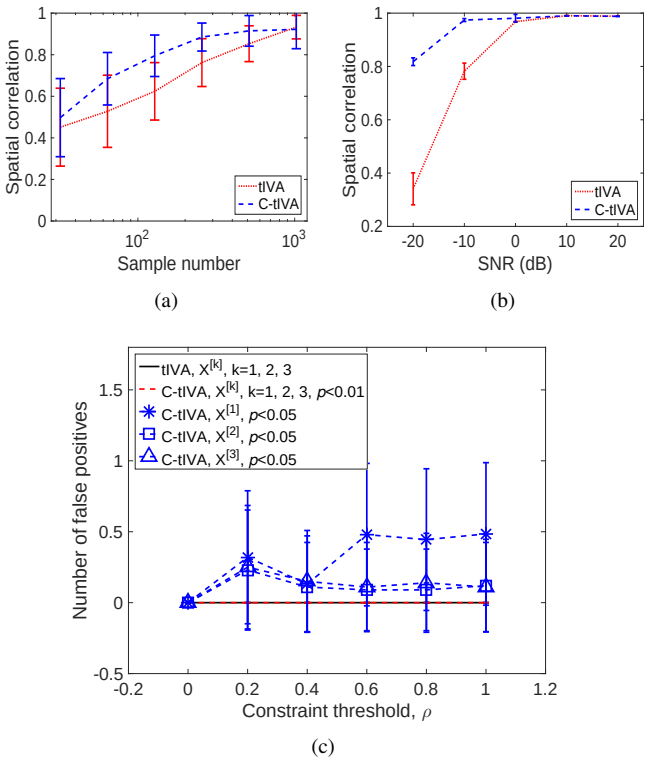

Fig. 1. (a) Average correlation between true and estimated spatial map with respect to number of samples N, using SNR of 0 dB and $\rho = 0.7$. (b) Average correlation with respect to SNR using 400 samples and $\rho = 0.7$. (c) Number of false positives with respect to ρ, using 400 samples at $p < 0.05$ and $p < 0.01$.

record the number of significant components (false positives) with respect to change in ρ. The results of this study are shown in Figure 1(c). From Figure 1(c), it is clear that using $p < 0.05$ as the significance threshold can result in selection of incorrect subject covariations, however this is not the case at $p < 0.01$.

4.2. Multitask fMRI data results

The multitask fMRI data used in this study was collected from 150 healthy controls and 121 patients with schizophrenia during the performance of three tasks: an auditory odd ball (AOD) task, the Sternberg item recognition paradigm (SIRP) task and a sensory motor (SM) task [20]. The AOD task involves the subject listening to three different tones: standard (1 kHz tones occurring with probability 0.82), novel (computer generated, complex sounds occurring with probability 0.09), and target (1.2 kHz tones with probability 0.09) played in a pseudo-random order. The subject is asked to press a button with the right thumb at the target tone. For the SIRP task, the subject is asked to memorize a set of 1, 3 or 5 integers between 0 and 9. A series of integers is displayed and the subject has to press a button with their right thumb if the displayed integer belonged to the pre-defined set. The SM task involves a series of tonal changes in an increasing and then decreasing manner, for which the subject is required

to press a button with their right thumb at each tonal change.

4.2.1. *Global difference maps*

Prior to performing tIVA, we perform order selection using [21] and use an estimated order of 24 for both tIVA and C-tIVA. IVA-G is performed on the signal subspace for 10 runs and we determine a 'best run' using [22]. A two-sample t-test, at $p < 0.01$, is performed to detect the spatial maps that demonstrate significant difference between patients and controls. The value for p is selected based on Figure 1(c), to ensure that the constraint is not forced on the decomposition identifying non-existing differences. For some decompositions multiple significant components are detected, which makes a direct comparison difficult. Thus, we generate a Global Difference Map (GDM) for each method in order to summarize all of the significant spatial maps as follows. For M statistically significant maps, at $p \leq 0.01$, $\hat{\mathbf{a}}_m^{\prime[k]}$, a GDM is generated using

$$\mathbf{a}_{GDM}^{\prime[k]} = \sum_{m=1}^{M} \frac{|T_m|}{\sum_{n=1}^{M} |T_n|} \hat{\mathbf{a}}_m^{\prime[k]},$$

where T_m is the t-statistic for the mth significant component.

Figure 2 shows the GDMs for tIVA and C-tIVA using $\rho = 0.3, 0.5$ and 0.7. By visual inspection, the GDM for tIVA and C-tIVA, at $\rho = 0.3$ are similar, which is reasonable, since for lower values of ρ, C-IVA framework is equivalent to IVA based on (5). For $\rho = 0.7$, the GDM show higher activation for patients in the somatosensory cortex, indicating that the patients had a difficulty during motor execution while performing the tasks. For the visual SIRP task, the controls display higher activation in the visuo-motor cortex, however no such activation is seen for lower values of ρ and tIVA-G. Additionally, higher auditory activation is seen in controls for the auditory SM task.

4.2.2. *Classification performance*

We measure the performance of tIVA and C-tIVA based on their ability to classify subjects as either patients or controls. The leave-p-out cross-validation technique is performed for 100 runs. For each run, two-thirds of the patients and controls are randomly selected to form a training set and the remaining one-third of the subjects form the testing set. GDMs for tIVA and C-tIVA are computed and are regressed on the training set to estimate global subject covariations for each dataset. These subject covariations are used to train a k-nearest neighbor, $k = 10$, classifier. Subject covariation estimates for testing are obtained by regressing the GDM on each subject of the testing set. Classification accuracy (CA) is then obtained for both methods, tIVA and C-tIVA, by counting the number of correct classifications.

In order to measure the discriminative power of both methods, a two-sample t-test is performed on the estimated subject covariations of the testing set and the p-value for each run is recorded. The mean result for CAs, $\overline{\text{CA}}$, median of p-value over 100 runs and the p-value resulting from a t-test performed on the CAs of tIVA and C-tIVA, p_{CA}, are shown in Table 1. The results indicate that $\overline{\text{CA}}$ increases as the constraint is introduced into the decomposition and p_{CA} indicates that the results are further improved as the constraint threshold is increased.

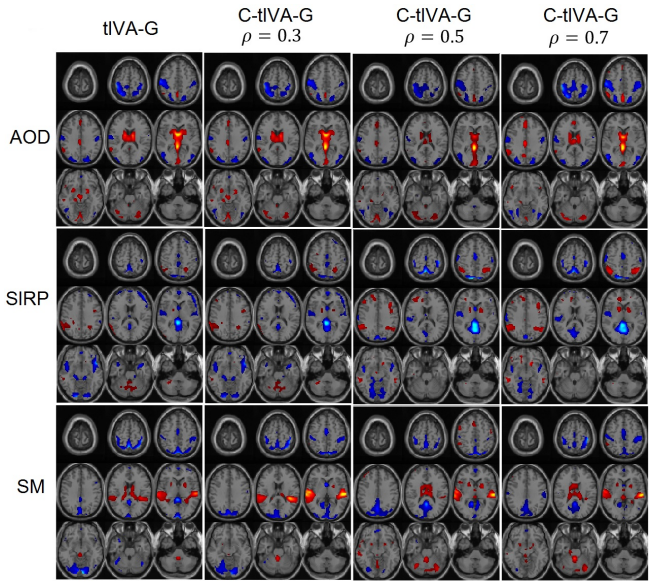

Fig. 2. GDMs for AOD, SIRP and SM obtained for tIVA and C-tIVA ($\rho = 0.3, 0.5, 0.7$). The spatial maps are z-maps thresholded at $z = 1.5$, where red indicates higher activation for controls and blue indicates higher activation for patients.

	ρ	p-value			$\overline{\text{CA}}$	p_{CA}
		AOD	SIRP	SM		
tIVA	–	2.51×10^{-4}	0.019	0.004	69.62	–
C-tIVA	0.3	1.27×10^{-4}	0.018	0.0021	70.4	0.085
C-tIVA	0.5	6.36×10^{-5}	0.013	8.95×10^{-4}	71.98	3.56×10^{-5}
C-tIVA	0.7	6.31×10^{-5}	0.014	6.11×10^{-4}	72.67	2.13×10^{-7}

Table 1. Comparison of tIVA with C-tIVA ($\rho = 0.3, 0.5, 0.7$) in terms of median of p-value over 100 runs, mean of accuracy values over 100 runs, $\overline{CA}$ and p-value resulting from a t-test performed on the CAs of tIVA and C-tIVA, p_{CA}.

5. CONCLUSION

In this paper, we propose a non-orthogonal constrained IVA framework in order to incorporate prior information into the decomposition. The proposed method is compared with regular IVA using simulated and real multitask data. C-IVA demonstrated superior performance over IVA in high levels of noise and when the data does not fully satisfy the modeling assumptions, such as the underlying assumption for source distribution and there are insufficient number of samples. C-IVA demonstrated superior classification performance indicating higher discriminative power than tIVA for multitask fMRI data. The p-values indicate an increasing difference in the classification accuracies between the two algorithms with respect to ρ, highlighting the importance of the constraint threshold.

6. REFERENCES

[1] C. Arndt and O. Loffeld, "Information gained by data fusion," in *Lasers, Optics, and Vision for Productivity in Manufacturing I*. International Society for Optics and Photonics, 1996, pp. 32–40.

[2] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 6–23, 1997.

[3] C. Barillot, D. Lemoine, L. Le Briquer, F. Lachmann, and B. Gibaud, "Data fusion in medical imaging: merging multimodal and multipatient images, identification of structures and 3D display aspects," *European journal of radiology*, vol. 17, no. 1, pp. 22–27, 1993.

[4] A. P. James and B. V. Dasarathy, "Medical image fusion: A survey of the state of the art," *Information Fusion*, vol. 19, pp. 4–19, 2014.

[5] 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.

[6] A. M. E. Engberg, K. W. Andersen, M. Mørup, and K. H. Madsen, "Independent vector analysis for capturing common components in fMRI group analysis," in *2016 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, June 2016, pp. 1–4.

[7] I. Lee, T. Kim, and T.-W. Lee, "Independent vector analysis for convolutive blind speech separation," in *Blind speech separation*. Springer, 2007, pp. 169–192.

[8] J.-H. Lee, T.-W. Lee, F. A. Jolesz, and S.-S. Yoo, "Independent vector analysis IVA: multivariate approach for fMRI group study," *Neuroimage*, vol. 40, no. 1, pp. 86–109, 2008.

[9] T. Adalı, Y. Levin-Schwartz, and V. D. Calhoun, "Multimodal data fusion using source separation: Two effective models based on ICA and IVA and their properties," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1478–1493, Sept 2015.

[10] P. A. Rodriguez, M. Anderson, X. L. Li, and T. Adalı, "General non-orthogonal constrained ICA," *IEEE Transactions on Signal Processing*, vol. 62, no. 11, pp. 2778–2786, June 2014.

[11] W. Lu and J. C. Rajapakse, "Constrained independent component analysis," *Adv. Neural Inf. Process. Syst.*, vol. 13, pp. 570–576, 2000.

[12] W. Lu and J. C. Rajapakse, "Approach and applications of constrained ICA," *IEEE transactions on neural networks*, vol. 16, no. 1, pp. 203–212, 2005.

[13] V. Calhoun, T. Adalı, M. Stevens, K. Kiehl, and 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.

[14] 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.

[15] M. Anderson, T. Adalı, and X.-L. Li, "Joint blind source separation with multivariate Gaussian model: algorithms and performance analysis," *2012 Signal Processing*, vol. 60, no. 4, pp. 1672–1683, 2012.

[16] 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 (ICASSP)*, 2013, pp. 3243–3247.

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

[18] T. Adalı, Y. Levin-Schwartz, and V. D. Calhoun, "Multimodal data fusion using source separation: Application to medical imaging," *Proc. IEEE*, vol. 103, no. 9, pp. 1494–1506, Sep. 2015.

[19] M. Ghasemi and A. Mahloojifar, "FMRI data analysis by blind source separation algorithms: A comparison study for non-gaussian properties," in *18th Iranian Conference on Electrical Engineering*, May 2010, pp. 13–17.

[20] R. L. Gollub, J. M. Shoemaker, M. D. King, T. White, S. Ehrlich, S. R. Sponheim, V. P. Clark, J. A. Turner, B. A. Mueller, V. Magnotta *et al.*, "The MCIC collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia," *Neuroinformatics*, vol. 11, no. 3, pp. 367–388, 2013.

[21] Y. Levin-Schwartz, Y. Song, P. J. Schreier, V. D. Calhoun, and T. Adalı, "Sample-poor estimation of order and common signal subspace with application to fusion of medical imaging data," *NeuroImage*, vol. 134, pp. 486–493, 2016.

[22] W. Du, S. Ma, G. S. Fu, V. D. Calhoun, and T. Adalı, "A novel approach for assessing reliability of ICA for FMRI analysis," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 2084–2088.