Diversity in Independent Component and Vector Analyses: Identifiability, Algorithms, and Applications in Medical Imaging

Tülay Adalı, *Fellow, IEEE*, Matthew Anderson, and Geng-Shen Fu
Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County
Baltimore, MD 21250, USA
E-mail: adali@umbc.edu

Abstract— Starting with a simple generative model and the assumption of statistical independence of the underlying components, independent component analysis (ICA) decomposes a given set of observations by making use of the diversity in the data, typically in terms of statistical properties of the signal. Most of the ICA algorithms introduced to date have considered one of the two types of diversity: non-Gaussianity—*i.e.*, higher-order-statistics—*or*, sample dependence. A recent generalization of ICA, independent vector analysis (IVA), generalizes ICA to multiple data sets and adds the use of one more diversity, dependence across multiple data sets for achieving an independent decomposition, *jointly across multiple data sets*. Finally, both ICA and IVA, when implemented in the complex domain, enjoy the addition of yet another type of diversity, noncircularity of the sources—underlying components.

Mutual information rate provides a unifying framework such that all these statistical properties—types of diversity—can be jointly taken into account for achieving the independent decomposition. Most of the ICA methods developed to date can be cast as special cases under this umbrella, as well as the more recently developed IVA methods. In addition, this formulation allows us to make use of maximum likelihood theory to study large sample properties of the estimator, derive the Cramér-Rao lower bound (CRLB) and determine the conditions for the identifiability of the ICA and IVA models. In this overview paper, we first present ICA, and then its generalization to multiple data sets, IVA, both using mutual information rate, present conditions for the identifiability of the given linear mixing model and derive the performance bounds. We address how various methods fall under this umbrella and give examples of performance for a few sample algorithms compared with the performance bound. We then discuss the importance of approaching the performance bound depending on the goal, and use medical image analysis as the motivating example.

I. INTRODUCTION

Data-driven methods typically start with a simple latent variable model—of which the linear mixing has been the most common—and decompose a given set of V-dimensional P observations, typically arranged as a $P \times V$ observation matrix, into two matrices, a $P \times M$ mixing matrix and an $M \times V$ component/source matrix using a suitable cost. Since in this very general form, this is not a well defined problem, usually additional constraints are imposed on the mixing

and/or component matrices such as sparsity and non-negativity. ICA is based on the assumption of statistical independence of the underlying components, and because this is a strong assumption, it enables a solution subject to only scaling and permutation ambiguities. Independence is also a natural assumption in many problems and a set of features that are statistically *independent* can be easily used for many tasks. This is the reason for the popularity of ICA and its wide use in areas as diverse as biomedicine, communications, finance, geophysics, and remote sensing, see *e.g.*, [1]–[3]. In this overview paper, we use mutual information rate to provide a common umbrella for ICA such that the two most commonly used types of diversity to achieve ICA, dependence of samples and higher-order-statistics are both taken into account.

There are numerous applications where not only one set of observations but multiple data sets, which have some dependence among them, need to be *jointly analyzed*. Examples include analysis of medical data such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) collected from multiple subjects, remote sensing data such as hyperspectral images where each pixel provides spectral information over multiple frequency bands, analysis of muti-sensor or multi-modality data that provide complementary information, and multi-subject biometric data, among many others. In all of these cases, the underlying components within the data sets, and hence the observations themselves, exhibit statistical dependence, which is another form of diversity to exploit. One approach to analyze these multiple data sets is to perform an individual ICA on each data set separately. Since most applications require matching of the corresponding components from each data set, one should then use a permutation algorithm to align the estimated components/sources since the ordering of the sources cannot be determined by ICA. Such an approach becomes computationally prohibitive as the number of data sets and sources increases, but more importantly, it fails to take advantage of the additional diversity, statistical dependence across multiple data sets while performing the analysis. An approach for ICA of multiple data sets, called Group ICA, which is introduced in the context of fMRI analysis [4], temporally concatenates multiple data sets, and after a dimension reduction step, performs ICA on this concatenated data set and then reconstructs the estimates for each data set separately. As we demonstrate in this paper, while practical and useful, using a common subspace for performing

ICA is likely to lead to information loss. Multi-set canonical correlation analysis (MCCA) [5] alleviates the problem by making full use of all the available data and has found wide application, see *e.g.*, [6], [7]. It can be also shown to achieve joint blind source separation [8]. However, MCCA makes use of only second-order-statistics (SOS) and constrains the demixing matrix to be orthogonal, hence limiting the search space for the optimal solution.

IVA generalizes the ICA problem to multiple data sets in such a way that it allows making full use of the statistical dependence across multiple data sets, and can take not only SOS but higher-order-statistics into account as well, and includes MCCA as a special case. Using the IVA framework, one can exploit the statistical dependence of each source across multiple data sets leading to performance beyond what is achievable with single-set ICA algorithms applied separately to each data set. Additionally, IVA automatically aligns dependent sources across the data sets hence bypassing the need for the use of a second permutation algorithm for the task. The original formulation for IVA [9] assumes that sources across data sets have no second-order dependence, and uses a multivariate Laplacian model for the source component vector (SCV)—which is defined in Figure 3(a). In this overview paper, we present a more general formulation for IVA, show that just like ICA, IVA can be cast using mutual information rate and thus all three key statistical properties, sample dependence within a source, source dependence within an SCV as well as higher-order-statistics are taken into account [10]. We give the identifiability conditions and present results on large sample properties using maximum likelihood theory for both ICA and IVA, and in the process, discuss the parallels between the two approaches in terms of the role statistical dependence plays. We emphasize the fact that it is the SOS that determine identifiability for both ICA and IVA, and that the correlation structure defines the diversity needed for establishing an independent decomposition for both, and discuss the parallels for the two. The results for identifiability and large sample properties do consider another important diversity type, which is nonstationarity of the sources. Finally, application of ICA and IVA to medical image analysis is discussed highlighting the importance of diversity in these studies.

II. INDEPENDENT COMPONENT ANALYSIS (ICA)

We consider the basic noiseless ICA problem based on instantaneous mixing where there are as many sources as mixtures—the most common case, the overdetermined one, also the case in fMRI analysis, can be easily reduced to this form using order selection as in [11], [12]. The linear mixing model is then written as

$$\mathbf{x}(v) = \mathbf{A}\mathbf{s}(v), \quad 1 \leq v \leq V, \quad \mathbf{x}(v), \mathbf{s}(v) \in \mathbb{R}^N \quad (1)$$

where v is the sample index such as voxel, pixel, or time. The estimates are given by $\mathbf{u}(v) = \mathbf{W}\mathbf{x}(v)$, which can be also written in matrix form as $\mathbf{U} = \mathbf{W}\mathbf{X}$, where $\mathbf{u}_n^\top \in \mathbb{R}^V$ is the nth row of $\mathbf{U} = \mathbf{W}\mathbf{X}$, *i.e.*, $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_N]^\top$, and $\mathbf{X}, \mathbf{U} \in \mathbb{R}^{N \times V}$. Since we consider the more general case that includes sample dependence in the ICA formulation and

would like to keep the notation as simple as possible, we make the following definitions. We use $\mathbf{x}(v) \in \mathbb{R}^N$ to refer to the random vector that contains the N mixtures $x_n(v)$, $1 \leq n \leq N$, and $\mathbf{x}_n \in \mathbb{R}^V$ to denote the transpose of the nth row of the observation matrix $\mathbf{X} \in \mathbb{R}^{N \times V}$. When the reference is to a random quantity rather than observation, it will be clear from context.

In ICA, we assume that the sources $s_n(v)$ in $\mathbf{s}(v) = [s_1(v) s_2(v) \ldots s_N(v)]^\top$ are statistically independent, and make use of different properties of the signal, such as non-Gaussianity, sample dependence, geometric properties, or non-stationarity of the signal, *i.e.*, *diversity* in some form [1, Chapter 1]. Among those, the most commonly used type of diversity has been non-Gaussianity—higher-order-statistics (HOS)—of the sources. Most of the popular ICA algorithms such as Infomax [13], FastICA [14], and joint approximate diagonalization of eigenmatrices (JADE) [15] as well as many of the variants of maximum likelihood (ML) techniques with different approaches for approximating the source density, such as [16], [17], all fall under this umbrella. Even very recent surveys—such as [18]—primarily consider ICA algorithms within this group. As a result, in the community, most often, fundamental results such as those for identifiability always consider this more limiting view of achieving ICA. It has been hence commonly noted that *ICA can identify only a single Gaussian source*. As we note next, this is true *only when non-Gaussianity is the only form of diversity that is considered*.

Besides those making use of non-Gaussianity, another important group is algorithms that make use of linear dependence among the samples, hence SOS. These include the algorithm for multiple unknown signal extraction (AMUSE) [19], second-order blind identification (SOBI) [20], and weights-adjusted SOBI (WASOBI) [21] among others. In this case, we use a *random process* rather than a *random variable* model for the sources, and use $s_n(v)$ where v is an index such as time, pixel, or voxel. In this paper, we use v for voxel, as medical image analysis, *i.e.*, volume data will be our main motivating example.

Algorithms using only non-Gaussianity form a major portion of the ICA algorithms developed to date, while those using sample dependence come in second. An obvious question one may ask is *"Why not make use of both types of diversity, non-Gaussianity and sample dependence together, at the same time?"*. As one would expect, this approach leads to algorithms with better performance than those using only one type of diversity as demonstrated in [22]–[26]. In addition, use of these two types of diversity *jointly* allows for more relaxed conditions for the identifiability of the ICA model in (1).

This is our main goal in this section, to show how *mutual information rate* helps bring most of the ICA algorithms under one umbrella and helps determine identification conditions along with performance bounds so that the performance of various algorithms can be compared against this benchmark.

A. Cost Function

Mutual information is a natural cost for ICA since the goal is the maximization of independence among the source estimates

$\mathbf{u} = \mathbf{Wx}$, and has been used commonly when providing a general umbrella for approaches based on use of HOS. Here, using the random process notation as in (1), we write the nth source estimate as $u_n(v) = \mathbf{w}_n^\top \mathbf{x}(v)$, where $\mathbf{w}_n^\top$ is the nth row of the demixing matrix $\mathbf{W}$. We can then write the mutual information *rate* as

$$
\begin{aligned}
\mathcal{I}_r(\mathbf{W}) &= \sum_{n=1}^N H_r(u_n) - H_r(\mathbf{u}) \\
&= \sum_{n=1}^N H_r(u_n) - \log|\det \mathbf{W}| - H_r(\mathbf{x}) \quad (2)
\end{aligned}
$$

and take into account both HOS and sample dependence to achieve ICA. In (2), we used the Jacobian expression $p_s(\mathbf{u}) = p_s(\mathbf{Wx}) = p_X(\mathbf{x})|\det \mathbf{W}|^{-1}$, and hence the last term $H_r(\mathbf{x})$ is a constant with respect to $\mathbf{W}$, it can be replaced by C resulting in

$$
\mathcal{I}_r(\mathbf{W}) = \sum_{n=1}^N H_r(u_n) - \log|\det \mathbf{W}| - C \quad (3)
$$

where $H_r(u_n)$ is the entropy rate, which is $H_r(u_n) = \lim_{v \to \infty} H[u_n(1), \ldots, u_n(v)]/v$ and the entropy[1] is written as $H(u_n) = -E\{\log p_{s_n}(\mathbf{w}_n^\top \mathbf{x})\}$. When the process is stationary, we have $H_r(u_n) = \lim_{v \to \infty} H(u_n(v)|u_n(v-1), \ldots, u_n(1))$. Since entropy rate measures the per sample density of the average uncertainty of a random process, minimization of (3) makes use of both HOS—through the minimization of missing information, entropy—and sample dependence by making samples easier to predict by increasing sample dependence, *i.e.*, decreasing the entropy rate. The term $\log|\det \mathbf{W}|$ acts as a regularization term preserving the volume across the directions of source estimation. Since entropy is not scale invariant, *i.e.*, $H(x) \neq H(\alpha x)$ for $\alpha \neq 1$, without the regularization term, the cost function could be minimized by simply scaling the source estimates. Mutual information rate hence provides a broad umbrella under which one can study the properties of ICA algorithms by taking into account both HOS and sample dependence, the two types of diversity most commonly used for ICA.

When we constrain the demixing matrix to be orthogonal, *i.e.*, let $\mathbf{WW}^\top = \mathbf{I}$, we have $|\det(\mathbf{W})| = 1$, and the cost in (3) reduces to

$$
\mathcal{J}_r(\mathbf{W}) = \sum_{n=1}^N H_r(u_n) - C \quad (4)
$$

which maximizes the negentropy rate, the information-theoretic distance of a random process from that of a Gaussian for each source, under a variance constraint.

For a given set of observations, $\mathbf{X} \in \mathbb{R}^{N \times V}$, we can maximize the likelihood given by

$$
\mathcal{L}_{\text{ICA}}(\mathbf{W}) = \sum_{n=1}^N \log p_{s_n}(\mathbf{u}_n) + V \log|\det \mathbf{W}| \quad (5)
$$

where $\mathbf{u}_n \in \mathbb{R}^V$ is the transpose of the nth row of $\mathbf{U} = \mathbf{WX}$,

i.e., $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_N]^\top$. By the general asymptotic equipartition property [27], as $V \to \infty$, the maximization of likelihood function $\mathcal{L}_{\text{ICA}}(\mathbf{W})$ becomes equivalent to the minimization of the mutual information rate cost in (2). This is true if the probability density function (pdf) $p_{s_n}(\mathbf{u}_n)$ used in the ML formulation exactly matches the true pdf, which is implied when using mutual information rate as the cost. When there is a mismatch between the estimated pdf through likelihood and the true one, there is a bias that can be represented by the relative entropic—Kullback-Leibler—distance of the true density to the estimated. Using a flexible density model such as those employed by the two algorithms introduced in Section II-C, autoregressive mixture of Gaussians (AR-MOG) [24], and entropy rate bound minimization (ERBM) [25], decreases this bias. At this point, and for the performance discussion in the next section, we assume that the source pdf is known. In Section II-C, we discuss different ways of estimating the source pdf during adaptation, which lead to a number of different ICA algorithms that can all be studied under the mutual information rate minimization umbrella. In [17], a distinction is made between a true ML scheme that estimates the pdf and one that uses a fixed distribution where the latter is called a *quasi* ML procedure.

In the rest of the development, to simplify the discussion, we assume that all variables are zero mean so that the definitions of correlation and covariance matrices coincide.

B. Identification Conditions and the Performance Bound

Given the log likelihood in (5), we can compute the Fisher information matrix (FIM) using the expected value of its Hessian, which tells us how informative the given set of observations are for the estimation of the demixing matrix $\mathbf{W}$. The FIM also plays a key role in determining the identification conditions of the ICA model as well as the lower bound on the unbiased estimator, the Cramér-Rao lower bound (CRLB). We consider the FIM locally around the optimal point, $\mathbf{G} = \mathbf{AW} = \mathbf{I}$, hence have $u_n = s_n$. Due to the invariance of the induced CRLB[2] with respect to $\mathbf{G}$, the CRLB only depends on the statistics of the sources.

By making use of the independence of the sources, one can show that the FIM has a block diagonal structure with N scalars and $N(N-1)/2$ matrices that are 2×2. The scalar diagonal entries are all positive, and hence, the properties of the FIM are determined by the 2×2 matrices—pairwise interaction of sources—given by

$$
\mathbf{J}_{m,n}^{\text{ICA}} = \begin{bmatrix} \kappa_{m,n} & 1 \\ 1 & \kappa_{n,m} \end{bmatrix}, \quad 1 \leq m < n \leq N, \quad (6)
$$

where

$$
\kappa_{n,m} = \text{trace}\left(E\left\{\boldsymbol{\psi}(\mathbf{s}_n)\boldsymbol{\psi}^\top(\mathbf{s}_n)\right\} \mathbf{R}_m\right),
$$

$$
\boldsymbol{\psi}(\mathbf{s}_n) = -\frac{\partial \log p_{\mathbf{s}_n}(\mathbf{s}_n)}{\partial \mathbf{s}_n} \in \mathbb{R}^V, \text{ and } \mathbf{R}_n = E\{\mathbf{s}_n\mathbf{s}_n^\top\} \in \mathbb{R}^{V \times V}.
$$

Hence, the FIM is a function of the key source statistics, the two types of diversity the formulation in (5) takes into

---

[1] In the rest of the paper, we refer to differential entropy simply as entropy since discrete-valued random variables are not considered in this paper.

[2] Since the quantity being estimated is $\mathbf{W}$ rather than $\mathbf{G}$, the estimated bound is actually the *induced* CRLB following [28].

account, sample dependence and HOS. In addition, source nonstationarity is another type of diversity and the form in (6) considers the use of this third type of diversity as well since here the definitions are with respect to the complete source vector of dimensionality V, which matches the dimension of the samples in the given observation. The quantity $\psi(\cdot)$ is called the *score function* and is defined as the derivative with respect to the source estimate $\mathbf{u}_n$. This is a slightly different definition than in traditional ML theory where the score function is defined with respect to the parameter.

Since the blocks $\mathbf{J}_{m,n}^{\text{ICA}}$, $1 \leq m < n \leq N$ are the diagonal blocks of a covariance matrix, the FIM, they are positive semi-definite, and since FIM is block diagonal, they determine the condition for positive definiteness of the whole matrix. Evaluating the condition for which $\mathbf{J}_{m,n}^{\text{ICA}}$ becomes singular hence yields the non-identifiability condition for the ICA model—subject to the scaling and permutation ambiguities. It can be shown that $\mathbf{J}_{m,n}^{\text{ICA}}$ remains positive definite as long as there are no two sources that are Gaussian with proportional auto-covariance matrices, *i.e.*, we do not have two Gaussians, s_m and s_n in the mixture that satisfy $\mathbf{R}_m = \delta^2 \mathbf{R}_n$ [1, Chapter 4]. Hence in the presence of this simple correlation diversity, *i.e.*, when $\mathbf{R}_m \neq \delta^2 \mathbf{R}_n$, even Gaussian sources are separable using ICA when sample dependence and HOS are both considered. In addition, this result also includes use of nonstationarity as diversity—to keep the notation simple we have not included a time index in the definition of the auto-covariance matrices.

For algorithms that only take sample dependence into account however, for algorithms such as AMUSE, SOBI, and WASOBI, any two sources—not only Gaussians—with "similar" covariance matrices cannot be separated [19], [28]. Obviously, using these algorithms, i.i.d. sources cannot be separated either. When the sources are i.i.d., or when only HOS are taken into account implicitly assuming i.i.d. samples—as is the case in most of the ICA algorithms—then effectively, we have $\mathbf{R}_l = \sigma_l^2 \mathbf{I}$ for $l = n, m$. In this case, we can only identify a single Gaussian source since the correlation diversity is no longer available. This is the commonly known condition for the identifiability of the ICA model since the majority of ICA algorithms only exploit non-Gaussianity. However, as we note here, it is important to remember that this condition is true only for a specific case, and now there are effective algorithms that can take into account multiple types of signal diversity.

Using the expression in (6), we can write the CRLB as

$$\text{var}(w_{m,n}) \geq \frac{1}{V} \left(\kappa_{m,n} - \kappa_{n,m}^{-1} \right)^{-1}. \qquad (7)$$

Assuming that the mixtures are whitened such that $E\{\mathbf{X}\mathbf{X}^\top\} = \mathbf{I}$ and $\sigma_{s_n} = 1$, we can calculate the CRLB using the normalized interference-to-signal-ratio

$$\text{ISR} = \frac{1}{N(N-1)} \sum_{m,n=1,m\neq n}^{N} E\{g_{m,n}^2\} \qquad (8)$$

where $g_{m,n}$ are the entries of $\mathbf{G} = \mathbf{A}\mathbf{W}$, which we plot in Figure 1 along with performances of two algorithms introduced in the next section against this bound.

C. Algorithms

Mutual information rate in (3) can be minimized using relative/natural gradient updates [29], [30] as

$$\mathbf{W}(l+1) = \mathbf{W}(l) + \mu(\mathbf{I} - E\{\mathbf{\Psi}(\mathbf{U})\mathbf{U}^\top\})\mathbf{W}(l) \qquad (9)$$

where $\mathbf{\Psi}(\mathbf{U}) = [\psi_1(\mathbf{u}_1), \ldots, \psi_N(\mathbf{u}_N)]^\top \in \mathbb{R}^{N \times V}$, $\mu > 0$ is the step size, l is the iteration index, and the score function $\psi_n(\cdot)$ is defined in (6). In the update in (9), we include the complete source estimate matrix $\mathbf{U} = \mathbf{W}(l)\mathbf{X}$ as we consider sample correlation, rather than the commonly used random vector notation as in [29], [30]. The form of this update is the same as the one proposed in [31] based on nonlinear decorrelations, the original approach for achieving ICA [32].

The bound given in (7) assumes that the exact density of each source is known. In order to approach this bound, a number of density matching methods are proposed, in particular for the i.i.d. case, where the problem is simpler as we need to estimate a univariate rather than the multivariate score $\psi_n(\cdot)$ in (9). Solutions for the i.i.d. case include both parametric and nonparametric approaches as in efficient variant of FastICA (EFICA) [33] and non-parametric ICA (NP-ICA) [16], as well as a semi-parametric approach, ICA by entropy bound minimization (EBM) [34].

EBM uses an efficient entropy estimator where rather than estimating the entropy directly, an upper bound is estimated among a number of competing candidates determined by the maximum entropy principle and by a finite number of pre-specified measuring functions. Available prior information can be used in the selection of measuring functions, and even a simple selection of two odd and two even functions leads to a flexible algorithm that provides robust performance in a number of scenarios [34]. In [25], the flexible EBM density estimation strategy is combined with an invertible filter model such that both non-Gaussianity and sample dependence are taken into account to derive entropy rate bound minimization—originally introduced as full blind source separation—and hence to directly minimize (3). Other approaches that take both types of diversity into account are Markovian ICA [22] where the Markovian source model is adopted, autoregressive mixture of Gaussians [24], entropy rate minimization using an AR source model driven by GGD (ERM-ARG) [35], and MULTICOMBI [23] where either non-Gaussianity or sample dependence is taken into account by switching between the EFICA and WASOBI algorithms. All of these solutions assume stationarity of the sources.

The decoupling of the source estimates by assuming an orthogonal $\mathbf{W}$ introduced in Section II-A leads to negentropy rate as the cost, and greatly simplifies the score/density matching problem as the estimation for a given source then does not interact—and hence complicates—the estimation of others. This is the approach used in the FastICA algorithm [14], which is noted for its fast convergence. In [33], generalized Gaussian distribution (GGD)

$$p(s) \propto \exp(-|s|^{2\beta}) \qquad (10)$$

has been used as the source model to derive EFICA. Besides helping with density estimation, the assumption of
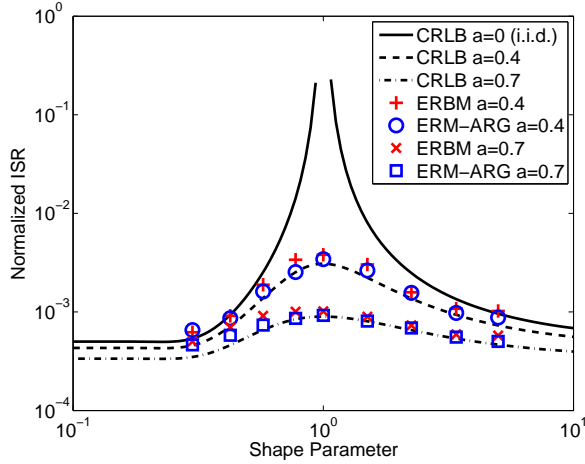
Fig. 1. Induced CRLB and performance of two algorithms as a function of shape parameter β (non-Gaussianity) for three levels of sample correlation, for AR coefficient $a = 0, 0.4$, and 0.7. Note the improvement in performance as the role of HOS (β moves away from 1) and as sample correlation (value of a) increase.
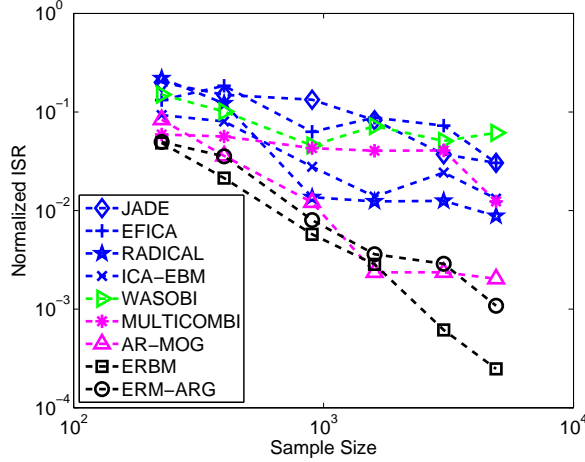

Fig. 2. Performance of nine algorithms using different types of diversity—either HOS or sample dependence, or both—in separation of a mixture of sources that come from a rich density. Note the best performance by algorithms making use of both types of diversity.

orthogonality provides a number of other advantages such as making second-order algorithms such as Newton-variants become more practical and allowing for easier implementation of constrained ICA [36]. However, this decoupling through constraining $\mathbf{W}$ to be orthogonal also limits the search space for the demixing matrices thus also limiting the achievable performance [37]. The decoupling approach given in [38], [39], transforms the matrix optimization problem to a series of vector optimization problems without having to constrain the matrix to be orthogonal. Here, it is also important to remember that the commonly used whitening step for the observations implies an orthogonal demixing matrix only when the number of samples $V \to \infty$, and hence does not guarantee an orthogonal demixing estimate unless it is embedded into the update mechanism. This decoupling approach is used in the EBM and ERBM algorithms that employ flexible density models for each source estimate. Finally algorithms that only make use of sample dependence, and hence SOS, such as AMUSE, SOBI, and WASOBI jointly diagonalize multiple

covariance matrices in order to determine an estimate for $\mathbf{W}$ rather than directly maximizing the likelihood (5). A second-order ICA approach based on ML with a Gaussian density model is given in [40], and it is shown that besides this ML-based algorithm, WASOBI approaches the CRLB as well, when the sources are stationary AR processes.

Example: Diversity, CRLB, and the Performance of two Algorithms: To demonstrate the role of diversity in attaining optimal performance while designing an algorithm, we consider a simple example, separation of two linearly mixed sources, an i.i.d. source drawn from a GGD (10) and a second source, a first-order AR process generated by a Gaussian process $\nu(v)$ such that $s(v) = as(v-1) + \nu(v)$. GGD assumes the form of a Gaussian for $\beta = 1$, is super-Gaussian when $0 < \beta < 1$ and sub-Gaussian when $\beta > 1$. Hence as β moves away from 1, the role of HOS increases, and similarly, the role of sample dependence increases as $|a| \to 1$. In Figure 1, we plot the CRLB given by (7) using the ISR (8). First note that for finite ISR, it suffices for one of the sources to have sample correlation—nonzero a—when both are Gaussian. The widely referenced and repeated condition for the real case that says "with ICA, one can identify only a single Gaussian" hence is true only when sample dependence is not taken into account— or is absent in that the samples are i.i.d., which rarely is the case in practice. In the same figure, we also show the performance of two algorithms that make use of both sample dependence and HOS: one that exactly matches the underlying source models, entropy rate minimization using AR model with a GGD driving process (ERM-ARG) [35] and the more flexible ERBM algorithm [25]. The results are shown for 1000 samples and 500 independent runs. While we observe that the exact match provides the best performance, the flexible ERBM does a decent job in approaching the bound as well and does not use prior information like the ERM-ARG.

Example: Performance Comparison in Separation of Natural Sources: In Figure 2, we show ISR of nine different algorithms in separation of ten artificially mixed images from [41] to demonstrate the performance of different algorithms in separation of sources that come from a rich class of distributions. Since for small sample sizes, there were a number of unstable runs, the results are plotted using median rather than the mean. The algorithms used in the comparison are JADE, EFICA, Robust Accurate Diirect ICA aLgorithm (RADICAL) and ICA-EBM that exploit the HOS, WASOBI that uses sample dependence, AR-MOG, ERM-ARG, and ERBM that use both, and finally MULTICOMBI that uses both but one at a time. The advantage of making use of both diversity jointly is clear as well as the superior performance of two algorithms that use flexible density models, AR-MOG and ERBM—though the performance of AR-MOG deteriorates with decreasing sample size due to its complexity.

III. INDEPENDENT VECTOR ANALYSIS (IVA)

In many applications, not only a single but multiple data sets with dependence among them need to be *jointly* analyzed. Examples include analysis of medical data such as fMRI and EEG from multiple subjects or at different conditions, data

from multiple frequency bins when solving the convolutive ICA problem in the frequency domain, and analysis of muti-sensor or fusion of multi-modality data with complementary information. IVA generalizes the ICA problem to multiple data sets so that one can take advantage of this additional type of diversity, the one across multiple data sets when achieving the decomposition.

Next, we show that IVA can be formulated using mutual information rate minimization like ICA but now with the addition of one more diversity, dependence among sources across data sets. Also, as in the case of ICA, we consider the general case that does not constrain the demixing matrices to be orthogonal, and as such, IVA generalizes CCA and MCCA [5] as well, both through incorporation of statistics higher than two and also by allowing a general nonorthogonal demixing matrix. Using this general formulation, we give the general conditions for identifiability of the IVA model as well as the performance bounds. The ICA result, as expected, becomes a special case when the number of data sets is set to one. We then present current algorithms for achieving IVA, and address the challenges in the area.

A. Cost Function

The IVA problem is defined similar to ICA except that we now have K data sets, each containing V samples, and formed from linear mixtures of N independent sources,

$$\mathbf{x}^{[k]}(v) = \mathbf{A}^{[k]}\mathbf{s}^{[k]}(v), 1 \leq k \leq K, \quad 1 \leq v \leq V \quad (11)$$

where $\mathbf{A}^{[k]} \in \mathbb{R}^{N \times N}$, $k = 1, \dots, K$ are invertible mixing matrices. The problem is finding K demixing matrices $\mathbf{W}^{[k]}$ such that sources for each data set can be estimated through $\mathbf{u}^{[k]}(v) = \mathbf{W}^{[k]}\mathbf{x}^{[k]}(v)$ for $k = 1, \dots, K$ as shown in Figure 3(a).

For K data sets $\mathbf{X}^{[k]} \in \mathbb{R}^{N \times V}$, we can recover the source estimates for each data set using $\mathbf{U}^{[k]} = \mathbf{W}^{[k]}\mathbf{X}^{[k]}$, and by defining augmented matrices $\mathbf{X}$ and $\mathbf{S}$, write the problem as

$$\begin{bmatrix} \mathbf{X}^{[1]} \\ \vdots \\ \mathbf{X}^{[K]} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{[1]} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}^{[K]} \end{bmatrix} \begin{bmatrix} \mathbf{S}^{[1]} \\ \vdots \\ \mathbf{S}^{[K]} \end{bmatrix} \Longleftrightarrow \mathbf{X} = \mathbf{A}\mathbf{S}$$

$$(12)$$

where

$$\mathbf{A} = \oplus \sum_{k=1}^{K} \mathbf{A}^{[k]}.$$

The estimates are given by $\mathbf{U} = \mathbf{W}\mathbf{X}$ and the demixing matrix is also block diagonal, $\mathbf{W} = \oplus \sum_{k=1}^{K} \mathbf{W}^{[k]}$. The decomposition is performed on this augmented matrix $\mathbf{X}$ so that the dependence of components of the source matrices $\mathbf{S}^{[k]}$ across data sets can be taken into account. In this model, the components within each $\mathbf{S}^{[k]}$ are assumed to be independent while we allow for dependence across corresponding components of $\mathbf{S}^{[k]}$ in multiple data sets. In fact, it is this additional dependence that IVA takes advantage of, and the following definition helps clarify the idea and is key for the whole development to follow.

We define the source component vector (SCV) for the nth SCV $\mathbf{s}_n$ as

$$\mathbf{s}_n(v) = \left[s_n^{[1]}(v), s_n^{[2]}(v), \dots, s_n^{[K]}(v) \right]^{\top} \in \mathbb{R}^K$$

i.e., by concatenating the nth source from each of the K data sets, or similarly, define the source component matrix (SCM) $\mathbf{S}_n$ shown in Figure 3(a), through concatenation of each row of $\mathbf{S}^{[k]}$ as $\mathbf{S}_n = [\mathbf{s}_n^{[1]}, \mathbf{s}_n^{[2]}, \dots, \mathbf{s}_n^{[K]}]^{\top}$. The SCV takes into account sample dependence through the inclusion of index v in its notation and we use both definitions, SCV and SCM, in the discussion to follow. As an example, in the fMRI analysis we introduce in Section IV, the nth SCM contains the spatial activation maps of the nth source, such as the motor component, for all K subjects in the study. One would expect the activation maps of different subjects to be statistically dependent, as for each subject, voxels at corresponding locations would show comparable levels of activation. This is the additional diversity that the general IVA formulation makes use of when achieving the decomposition. It is, however, important to note that while IVA makes use of this additional diversity, it does not require that it exists, and in its absence, reduces to individual ICAs on each data set. This additional diversity is also what helps with the resolution of permutation ambiguity among the sources estimated across the data sets. The identification condition we introduce in the next section specifies when all the sources in an IVA model can be identified, and is a quite relaxed condition. However, identification of sources does not imply that the sources will be aligned as well, and sources across data sets—components of each SCV—can be aligned only if the sources across the data sets are statistically dependent [10].

At this point it is also useful to note that the IVA formulation is a special case of the multidimensional ICA (MICA) problem [42], also defined as an independent subspace problem [43]. Though MICA is not necessarily defined for multiple data sets as IVA, we can use the augmented matrix definition in (12) to understand how MICA considers a more general model. For the MICA formulation, the mixing matrix is not assumed to have a block diagonal form and the number of components within each SCV can be different. Hence the problem is one of finding independent subspaces where in each, there might be different number of dependent components. There are many challenges for solving the general MICA problem. A major one among them is determining the number of subspaces and components within each SCV. A recent overview of MICA is given in [44] where an effective solution is offered for the multivariate-Gaussian case. Assuming the correct number of components within each SCV can be determined, MICA then identifies the independent subspaces and *does not identify* the individual components within each subspace, *i.e.*, the components within an SCV like IVA does. The IVA formulation provides enough additional restrictions to the MICA formulation so as to achieve identifiability of individual components while still creating a more general framework than ICA.

Since $\mathbf{W}$ has far fewer non-zero parameters than its full $KN \times KN$ dimension implies, we define $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{N \times N \times K}$, a three-dimensional array, to denote the set of parameters to be estimated.

Given the definition of an SCV, we formulate the IVA problem similar to that of ICA in (3) as

$$\mathcal{I}_r^{\text{IVA}}(\mathcal{W}) = \sum_{n=1}^{N} H_r\left(\mathbf{u}_n\right) - \sum_{k=1}^{K} \log\left|\det\left(\mathbf{W}^{[k]}\right)\right| - C \quad (13)$$

where the main difference is that we are now seeking to minimize the mutual information among SCVs rather than individual sources. To provide a clear view of the role of this additional diversity, we rewrite (13) as

$$\mathcal{I}_r^{\text{IVA}}(\mathcal{W}) = \sum_{n=1}^{N} \left(\sum_{k=1}^{K} H_r(u_n^{[k]}) - \mathcal{I}_r\left(\mathbf{u}_n\right) \right) - \sum_{k=1}^{K} \log\left|\det\left(\mathbf{W}^{[k]}\right)\right| - C. \quad (14)$$

Without the second term $\sum_{n=1}^{N} \mathcal{I}_r(\mathbf{u}_n)$, the expression in (14) is exactly equivalent to the sum of the cost in (3) across K data sets, hence performing independent ICAs on each data set. It is this second term, sum of mutual information within each SCV, $\sum_{n=1}^{N} \mathcal{I}_r(\mathbf{u}_n)$ that takes the diversity across data sets into account. The minimization of (14) hence increases mutual information among components of an SCV, thus making use of the natural dependence among data sets.

If we consider no sample dependence—hence the cost is mutual information ($\mathcal{I}$) rather than mutual information rate ($\mathcal{I}_r$)—and use the multivariate Gaussian model for the SCV, we have $H(\mathbf{u}_n) = (1/2)\log\left[(2\pi e)^K \prod_{k=1}^{K} \lambda_n^{[k]}\right]$ where $\lambda_n^{[k]}$ is the kth eigenvalue of the covariance matrix of the nth SCV, then (13) reduces to

$$\mathcal{I}^{\text{IVA-G}}(\mathcal{W}) = \frac{NK \log(2\pi e)}{2} + \frac{1}{2} \log\left(\prod_{n=1}^{N} \prod_{k=1}^{K} \lambda_n^{[k]} \right) - \sum_{k=1}^{K} \log\left|\det\left(\mathbf{W}^{[k]}\right)\right| - C. \quad (15)$$

This is exactly equivalent to the generalized variance method (GENVAR) cost function proposed for achieving multiset canonical correlation analysis [5] when we constrain the demixing matrices to be orthogonal hence eliminating the term $\sum_k \log\left|\det\left(\mathbf{W}^{[k]}\right)\right|$ but imposing a constraint on the sum of estimates, hence the eigenvalues. In [5], five cost functions are introduced for maximizing correlation among linearly transformed multiple data sets, which in our IVA formulation are the SCVs. The cost functions introduced in [5] all have the common objective of estimating $\mathbf{W}^{[k]}$ such that the SCV covariance matrix becomes as ill conditioned as possible, since this maximizes the correlation among the components within an SCV. Obviously the term $\log\left(\prod_{n=1}^{N} \prod_{k=1}^{K} \lambda_n^{[k]}\right)$ achieves this goal when we let the sum of the eigenvalues be constant, *i.e.*, constrain the demixing matrix. In (15), the term $\sum_k \log\left|\det\left(\mathbf{W}^{[k]}\right)\right|$ achieves this purpose and is written using the theoretically well justified cost of mutual information.

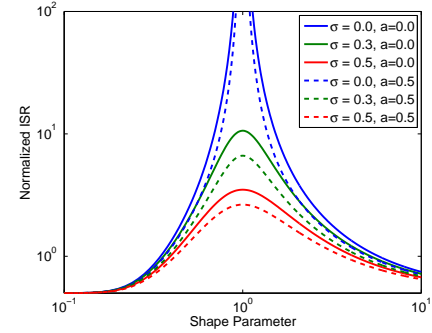For given $\mathbf{X}^{[k]}$, $k = 1, \ldots, K$, we can write the likelihood as

$$\mathcal{L}_{\text{IVA}}(\mathcal{W}) = \sum_{n=1}^{N} \log\left(p_n(\mathbf{U}_n)\right) + V \sum_{k=1}^{K} \log\left|\det\left(\mathbf{W}^{[k]}\right)\right|$$

where now the score function for the source component matrix
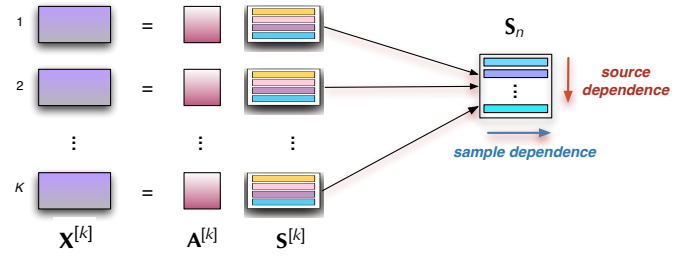
(a) IVA of multiple data sets and the two key signal properties available in addition to HOS: sample dependence and dependence among sources within a source component matrix $\mathbf{S}_n$

(b) Role of three types of diversity on performance in terms of induced CRLB (Normalized ISR)

Fig. 3.   For the IVA problem given in (a), note the improvement in performance shown in (b) as the values of σ (source dependence) and a (sample dependence) increase and shape parameter moves away from 1, *i.e.*, role of HOS increases.

$\mathbf{U}_n$ is written as

$$\mathbf{\Psi}_n^{\text{IVA}}\left(\mathbf{U}_n\right) = -\frac{\partial \log p_n\left(\mathbf{U}_n\right)}{\partial \mathbf{U}_n} \in \mathbb{R}^{K \times V}. \quad (16)$$

Now, we can proceed as in ICA to derive the performance bound and determine the conditions for the identifiability of the IVA model by working within ML theory.

B. Identification Conditions and the Performance Bound

We evaluate the FIM by the expected value of the Hessian of $\mathcal{L}_{\text{IVA}}$ at the optimal point $\mathbf{G} = \oplus_{k=1}^{K} \mathbf{W}_k \mathbf{A}_k = \oplus_{k=1}^{K} \mathbf{G}_k = \mathbf{I}$, which is now $KN^2 \times KN^2$ in dimension, since $\mathcal{W}$ has a total of KN^2 parameters. Since, the IVA formulation replaces the sources with SCMs that are mutually independent, and each SCM includes K components, the FIM is again block diagonal but now with N block matrices that are $K \times K$ and $N(N-1)/2$ matrices of dimension $2K \times 2K$. The properties are again determined by the latter blocks, those that describe the interaction of now the SCMs, the $2K \times 2K$ block matrices

$$\mathbf{J}_{m,n}^{\text{IVA}} \triangleq \begin{bmatrix} \mathcal{K}_{m,n} & \mathbf{I}_K \\ \mathbf{I}_K & \mathcal{K}_{n,m} \end{bmatrix} \in \mathbb{R}^{2K \times 2K}$$

where $\{\mathcal{K}_{m,n}\}_{k_1,k_2} = (1/V) E\left\{ \left(\psi_m^{\text{IVA},[k_1]}\right)^{\top} \mathbf{S}_n^{[k_1]} \left(\mathbf{S}_n^{[k_2]}\right)^{\top} \psi_m^{\text{IVA},[k_2]} \right\}$ when $m \neq n$, $\psi_n^{\text{IVA},[k]} = (\mathbf{\Psi}_n^{\text{IVA}})^{\top} \mathbf{e}_k$, where $\mathbf{\Psi}_n^{\text{IVA}}$ is given in (16), the subscript for the identity matrix $\mathbf{I}$ denotes its dimension, and $\mathbf{e}_k$ is the kth basis vector. Again, the FIM is a function of the key SCM statistics, and in this case, all three types of diversity—sample dependence, dependence within an SCM, and the HOS—that are considered in this

IVA formulation. Since the SCM is written for V samples, *i.e.*, is $K \times V$, nonstationarity is taken into account as well.

The FIM is a block diagonal matrix for this case as well and the identification condition for the IVA model is obtained by evaluating when $\mathbf{J}_{m,n}^{\mathrm{IVA}}$ remains positive definite. It is shown that [10], [45] identification of the IVA model in (11) is possible as long as no two SCMs have $\boldsymbol{\alpha}$-Gaussian components for which $\mathbf{R}_{m,\alpha} = (\mathbf{I}_V \otimes \mathbf{D}) \, \mathbf{R}_{n,\alpha} \, (\mathbf{I}_V \otimes \mathbf{D})$, for $1 \leq m \neq n \leq N$, where $\mathbf{D} \in \mathbb{R}^{K_\alpha \times K_\alpha}$ is any full rank diagonal matrix, K_α the number of $\boldsymbol{\alpha}$-Gaussian components, $\mathbf{R}_n = E\{\mathbf{S}_n \mathbf{S}_n^\top\}$, and $\otimes$ is the Kronecker product. An $\boldsymbol{\alpha}$-Gaussian component is defined as the subset of rows of an SCM that are independent from the others and have multivariate Gaussian distribution, and $\boldsymbol{\alpha}$ refers to the index of this subset within $\{1, \ldots, K\}$, and $\mathbf{R}_{n,\alpha}$ refers to the covariance matrix of the matrix formed from the $\boldsymbol{\alpha}$-Gaussian rows of an SCM. Hence, it is again a second-order condition that determines the identifiability of the model, and the major role played by the source covariance matrix in ICA is now replaced by the SCM covariance matrix. As in the case of ICA, the result holds for the use of nonstationarity as a diversity type.

A useful special case to consider is when the samples are i.i.d., which is equivalent to considering $V = 1$ so that $\mathbf{I}_V$ is a scalar and unity and we now consider SCVs where each entry is a random variable rather than SCMs or an SCV with entries that are random processes. This is the basic assumption in most ICA algorithms where only HOS are taken into account and it leads to practical and effective solutions that work well for most cases, including many where the samples are actually dependent. However, for ICA, with the i.i.d. assumption, we can only identify a single Gaussian source. For IVA, however, the condition for this case is more general and now we can identify the IVA model as long as there are no two $\boldsymbol{\alpha}$-Gaussian SCVs for which $\mathbf{R}_{m,\alpha} = \mathbf{D}\mathbf{R}_{n,\alpha}\mathbf{D} \in \mathbb{R}^{K_\alpha \times K_\alpha}$. Hence the identification of multiple Gaussians is possible with IVA provided that the covariance diversity is available, in the sense that covariance matrices of sources that are Gaussian across data sets are not essentially identical, *i.e.*, satisfy $\mathbf{R}_{m,\alpha} \neq \mathbf{D}\mathbf{R}_{n,\alpha}\mathbf{D} \in \mathbb{R}^{K_\alpha \times K_\alpha}$.

Finally, for $K = 1$, the condition reduces to that for ICA and we cannot identify any two Gaussians that have $\mathbf{R}_m = \delta^2 \mathbf{R}_n$, $\delta \neq 0$ where now the covariance is defined for a single source rather than an SCV. A comparison of these two conditions reveal the dual nature of the role of diversity in these two cases, diversity in the form of source dependence for IVA versus sample dependence in ICA. The diagonal matrix $\mathbf{D}$ for IVA and δ^2 for ICA are present in the conditions simply due to the inherent scaling ambiguity of the problem. The given identification conditions for the i.i.d. case coincide with those derived assuming a multivariate Gaussian model in [46], [47] since they are determined by second-order statistics.
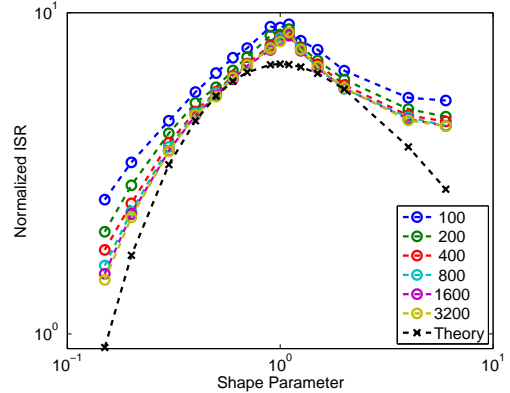
The CRLB for IVA is given by

$$\mathrm{var}\left(w_{m,n}^{[k]} \right) \geq \frac{1}{V} \mathbf{e}_k^\top \left(\boldsymbol{\mathcal{K}}_{m,n} - \boldsymbol{\mathcal{K}}_{m,n}^{-1} \right)^{-1} \mathbf{e}_k \qquad (17)$$

which has a similar form to (7), and again similarly, can be computed using the sum of ISR values, now defined as $\mathrm{ISR}_{m,n}^{[k]} = E\left\{ (g_{m,n}^{[k]})^2 \right\}$.
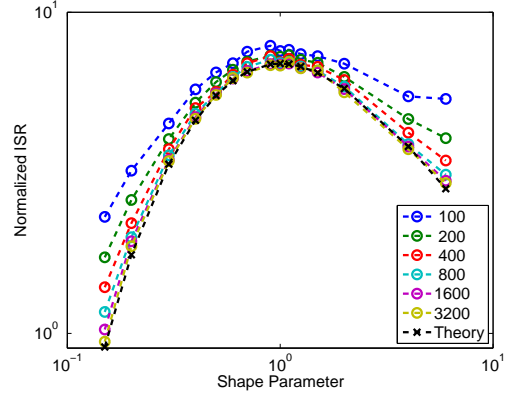
(a) Shape parameter β is assumed to be known

(b) Shape parameter is selected from $\beta = \{0.5, 2\}$

Fig. 4. Performance of two IVA algorithms that take source dependence and HOS into account for separation of three GGD sources of dimension $K = 5$ with shape parameter β and a random covariance matrix compared to the induced CRLB (normalized ISR) for different sample sizes.

Example: Role of Three Types of Diversity for IVA: The simple example shown in Figure 3(b) plots the CRLB in terms of ISR for two sources and two data sets, where the first set of sources—common to both data sets—is drawn from a multivariate GGD, which is Gaussian when the shape parameter $\beta = 1$, and has super-Gaussian marginals for $0 < \beta < 1$ and sub-Gaussian for $\beta > 1$. Second set of sources are an i.i.d. Gaussian and a first-order AR process $s(v) = as(v - 1) + \nu(v)$ where $\nu(v)$ is a white Gaussian process. Hence, the AR parameter a characterizes influence of sample correlation and the shape parameter β of non-Gaussianity, *i.e.*, HOS. Finally we introduce correlation for the first group of sources through a correlation coefficient σ. As observed in Figure 3(b), performance—as measured by the ISR—improves as sample source correlation and dependence across data sets—values of a and σ respectively—increase, and as the sources become more non-Gaussian—*i.e.*, as the value of β moves away from 1. We also note the condition for identifiability of the IVA model in that when the sources are all Gaussian ($\beta = 1$) and i.i.d., a finite ISR is still possible as long as there is correlation among the sources, in this example introduced only to the first set of GGD sources through σ.

C. Algorithms

In algorithm development, while it is desirable to consider together all types of diversity expressed in the cost (13),

current solutions available for the problem only take HOS and source dependence across data sets into account, primarily due to computational and modeling challenges. Hence, they minimize mutual information rather than mutual information rate. A generalization of joint diagonalization proposed in [48] is the only solution we know of that exploits sample dependence for the joint source separation problem in addition to the other two diversity, HOS and source dependence.

We write the relative/natural gradient updates for IVA to minimize the mutual information—hence not accounting for sample dependence—as

$$\mathbf{W}^{[k]}(l+1) = \mathbf{W}^{[k]}(l) + \mu(\mathbf{I} - E\{\boldsymbol{\psi}^{\text{IVA},[k]}(\mathbf{u}^{[k]})^{\top}\})\mathbf{W}^{[k]}(l)$$
(18)

where now the score function has the simpler form $\boldsymbol{\psi}^{\text{IVA},[k]} = -\left[\frac{\partial \log p_1(\mathbf{u}_1)}{\partial u_1^{[k]}}, \ldots, \frac{\partial \log p_N(\mathbf{u}_N)}{\partial u_N^{[k]}}\right]^{\top}$. Again, a key problem is the estimation of the score function, *i.e.*, the source pdf, during the adaptation. For IVA, as opposed to ICA, all solutions to date have emphasized parametric methods as nonparametric approaches can easily become prohibitive for the multidimensional case.

IVA is originally formulated for solving the convolutive ICA problem in the frequency domain [9], which is an application where resolution of the permutation ambiguity across frequency bins is critical to the success of the solution. Hence, the main application domain that is considered has been separation of acoustic sources resulting in an emphasis on models attractive for this case starting with the multivariate Laplace model [9], [49]. In [50], mixture of Gaussians is proposed where the noisy IVA problem and an online solution are considered as well. However these solutions fail to consider all-order statistical dependence within an SCV and in certain cases constrain the demixing matrix to be orthogonal/unitary as in [50], [51]. As discussed earlier, constraining the demixing matrix limits the performance and the decoupling trick introduced in Section II-C allows for advantages of orthogonality without having to constrain the matrix, and provides a number of additional advantages such as easier density matching, better convergence properties, and enabling easier derivation of second-order iterative algorithms. In (18), a single step size μ is used to update the entire demixing matrix while each row corresponds to a different source as in the case of ICA. In [47], a number of algorithms—including vector gradient descent and vector Newton algorithms—are derived using the decoupling trick so that the demixing matrices are not constrained to be orthogonal. They are then implemented using a multivariate Gaussian SCV model to derive a class of algorithms called IVA-G, and later using a the Kotz family [52] that includes the GGD, and hence Gaussian and Laplace as special cases. All of these solutions account for all-order statistical dependence for an SCV.

Iterative approaches to optimizing the IVA cost function are subject to similar convergence issues as iterative algorithms for ICA. It is shown that the Hessian matrix for the IVA cost with the multivariate Gaussian model always remains positive definite [47] and thus IVA-G has very desirable convergence properties. Hence, it is a good candidate for initializing the solution of
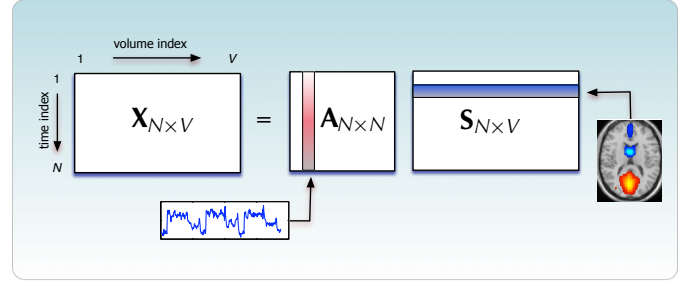

Fig. 5. Spatial ICA of fMRI data. Note the presence of both sample dependence and HOS as form of diversity.

other algorithms, and is used for initializing the solution of IVA with multivariate Laplace implemented as in [9] for the results we present in Section IV. For non-Gaussian sources, it is known that local minima exist in the cost function. These local minima correspond to demixing solutions which have different permutations across data sets [10]. Thus, even if a local minimum occurs, it is observed that the sources within each data set have been separated but the dependent sources across data sets are not aligned. This issue is addressed in [53] for the special case of spherical and super-Gaussian sources. *Example: Performance of Two IVA Solutions:* In Figure 4, we show the CRLB for separation of sources that are drawn from a multivariate GGD, and the performance of the IVA algorithm of [52] for different sample sizes. We implement two versions of the algorithm, one that estimates the covariance matrix but assumes that the true shape parameter β is known, and a second version that selects one of two $\beta = \{0.5, 2\}$ during the adaptation, which is a practical implementation. As expected, in both cases, the performance improves approaching the CRLB as the number of samples increase. In addition, while the first "clairvoyant" version of the algorithm in Figure 4(a), as expected, provides better performance, the second and practical implementation shown in Figure 4(a) provides quite satisfactory performance as well.

In terms of algorithms that only make use of linear dependence across multiple data sets, MCCA is the oldest, an extension of CCA [54] defined for two data sets. The algorithms given in [5] assume orthogonal demixing and are deflationary in nature such that each row of weights are estimated sequentially. IVA using multivariate Gaussian model also makes use of only linear dependence and can be derived within a ML framework [47], [55]. Since CCA can be achieved using generalized eigenvalue decomposition, it can also be posed as a diagonalization problem, which can be readily extended to achieve IVA using generalized joint diagonalization [48]. A review of extensions of CCA to include nonlinear dependences is given in [45].

IV. APPLICATION TO MEDICAL IMAGE ANALYSIS

FMRI has enabled us to directly study temporal and spatial changes in both the healthy and the diseased brain as a function of various stimuli, and has contributed greatly to our understanding of the most complex organ of the human body. Relatively low image contrast-to-noise ratio of the blood oxygenation level dependent fMRI signal, head movement,
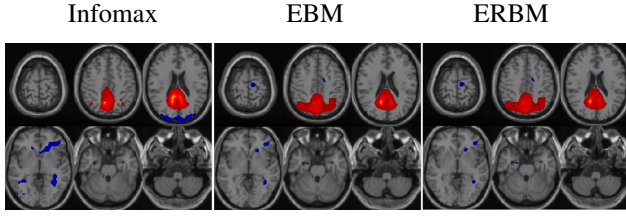
Fig. 6. Estimated t-maps for DMN using Infomax, EBM, and ERBM, and quantitative measures of their performance.

and undesired physiological sources of variability (cardiac, pulmonary) make detection of the activation-related signal changes difficult. The standard approach for the analysis of fMRI data has been correlating the time-series data with an assumed reference signal, *i.e.*, performing a simple linear regression as implemented in the popular statistical parametric mapping (SPM) software [56]. Even though it is robust, use of such a reference time course requires prior information, which most often is not reliable, and more importantly, in most cases it simply is not available. This is the case for data are acquired when subjects are at rest or performing naturalistic behavior such as watching a movie. Hence, following its first application to fMRI analysis [57], ICA has become an attractive solution, and is now widely used for fMRI analysis—for a recent review on ICA of fMRI, see [58].

Spatial ICA finds systematically non-overlapping, temporally coherent brain networks without constraining the temporal domain, hence can effectively recover *functional networks*. Functional connectivity refers to temporal correlations between spatially distinct regions of the brain, and ICA has been very effective in the study of networks of such intrinsic activity, since it naturally takes all the voxels into account when achieving the decomposition and provides a summary statistics for brain activity as well as its modulation across time. Besides, the linear superposition assumption holds for fMRI, see *e.g.*, [59], and the data-driven nature of ICA helps minimize unrealistic assumptions about the temporal domain and brain hemodynamics.

Figure 5 shows the application of ICA to fMRI analysis for finding *spatially independent components*, which has been by far the most common use of ICA for the problem. The observation matrix $\mathbf{X} \in \mathbb{R}^{T \times V}$ is formed by flattening the volume image data of V voxels at each time point. The time dimension is typically reduced from T, typically in the 100s, to N, a value around 30–60 to improve the estimation performance. Information-theoretical criteria (ITC) using principal component analysis (PCA) is most commonly employed for this step [11], *i.e.*, to determine the dimensionality of the signal subspace, usually with a correction for dependence among the samples (voxels) [12] so that a better estimate of ML can be used for the ITC. The spatially independent components—activation maps—form the sources, and the columns of the mixing matrix correspond to the temporal modulation of the corresponding source in the given time frame, $[1, T]$. In Figure 5, we show a sample time course after reconstruction to its original dimension T, and its corresponding Z-thresholded spatial activation map.

Example: Taking Sample Dependence and HOS in ICA of fMRI Data: The activation maps, the underlying independent sources, are typically super-Gaussian since they include heavy tails due to active voxels, those with high intensity values, and

(a) Multi-subject fMRI analysis with Group ICA

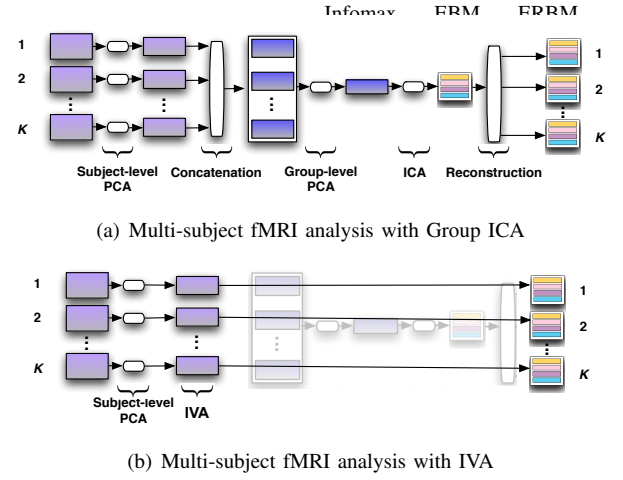(b) Multi-subject fMRI analysis with IVA

Fig. 7. Group ICA and IVA for multi-subject fMRI analysis. Note how IVA avoids projecting multi-subject data to a common space after subject-level PCA as well as the additional back-reconstruction step of Group ICA.

include sample dependence due to point spread function as well as low pass filtering, a common preprocessing step used for fMRI data. Hence, it would make sense to account for both types of diversity, HOS and sample dependence, when performing ICA of fMRI data. In Figure 6, we show the performance of three ICA algorithms in estimating the default mode network (DMN): (i) Infomax that uses a sigmoidal nonlinearity, a good match to super-Gaussian sources; (ii) EBM that uses a flexible density model, and (iii) ERBM that combines the flexible density estimation of EBM with a filtering approach to account for sample dependence. DMN is part of intrinsic networks, and one that has received much attention lately as it is regarded to be an important biomarker for different disorders. It activates preferentially when individuals focus on internal tasks, when the mind is wandering, and hence it is expected to be negatively correlated with the task time-course, when data are collected during a task. In this example, the data are collected from 20 subjects performing the auditory oddball task [60]. As shown in the Figure, all three algorithms have competitive performance, however both EBM and ERBM estimate more voxels than Infomax that correlate with the DMN mask. Also, when we perform a t-test on the multiple regression coefficients of the estimated time courses to determine their task-relatedness, ERBM yields the highest negative value for DMN, hence highest negative correlation with the task, indicating best performance using this metric [60].

Since the need to jointly analyze data from multiple subjects is inherent to most problems in medical data analysis, following the introduction of ICA for fMRI analysis [57] and its success, a simple but effective method, called Group ICA [4], is introduced for multi-subject fMRI data analysis. Group ICA performs a first-level dimension reduction at the individual subject level, and then temporally concatenates dimension-reduced subject data, to perform a second-level PCA to find a common subspace for data from all subjects. Then a single ICA is performed after which individual subject maps and time-courses are reconstructed as shown in Figure 7(a). There are a number of approaches for reconstructing the subject maps, which are evaluated and discussed in detail in [61]

(a) SimTB experimental set-up

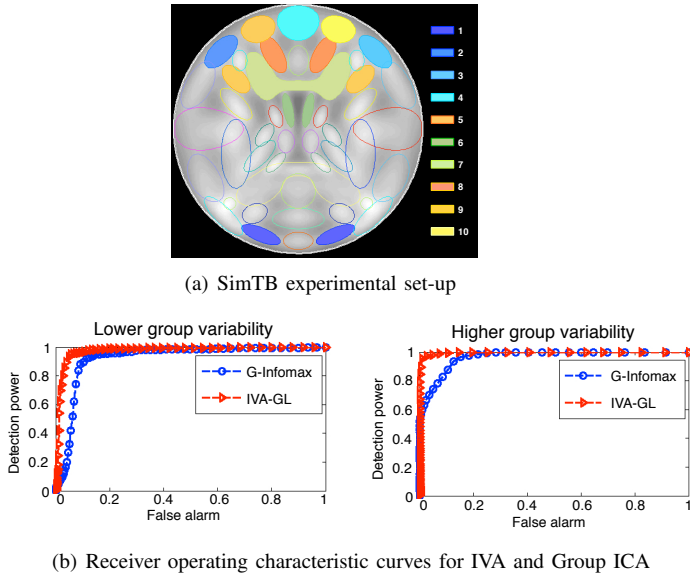(b) Receiver operating characteristic curves for IVA and Group ICA

Fig. 8. A group analysis study using fMRI-like data. IVA leads to better performance especially with increasing group variability, hence role of source dependence.
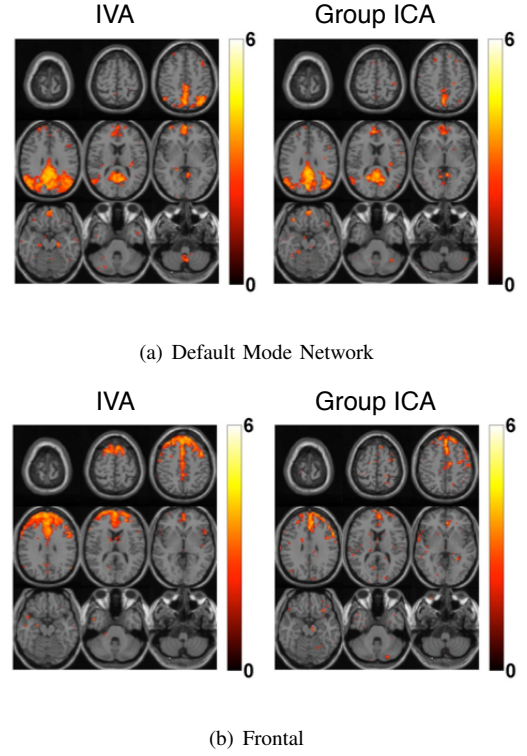

(a) Default Mode Network

(b) Frontal

Fig. 9. Sample estimated spatial maps for IVA and Group ICA for two components. Note the higher activation levels and spatial extent of the estimated maps using IVA.

as well as a number of ways to perform concatenation of data sets [62]. While robust and practical, the projection to a common subspace of data from different subjects can potentially cause loss of information in terms of subject variability. In the application of IVA to fMRI analysis shown in Figure 7(b), individual subject data are directly analyzed following the subject-level dimension reduction. Hence, all individual subject maps are estimated concurrently, and are aligned across subjects when there is dependence among them. This is typically the case for all components of interest, *i.e*, components corresponding to meaningful functional areas such as DMN and motor areas, since these naturally have statistical dependences across subjects. Components related to artifacts such as the motion artifact, however, might not be aligned for all subjects as these are less likely to have a similar dependence structure across subjects, and are more likely to be subject-specific. Next we demonstrate the advantage of IVA over the widely used Group ICA approach with two examples, one with simulated fMRI-like and a second one using real fMRI data.

Example: Capturing Subject Variability with IVA: To test the ability of IVA in capturing subject variability, we use the fMRI simulation toolbox, SimTB [63], and generate 10 components shown in Figure 8(a) for two groups, with 12 subjects in each. For each subject, components are randomly generated with small variations in terms of translation, rotation, and spread. For the first component, however, we introduce significant difference in terms of spread between the two groups of subjects, so as to simulate a typical difference in brain network volume change between the healthy and patient groups found in fMRI studies. The difference in spread is kept at two levels, one indicating a smaller difference between the groups, and a second one with greater difference in spread. Then, the performance is tested between the two approaches: Group ICA using Infomax with a nonlinearity matched to Laplacian pdf and IVA using a multivariate Laplacian model following

initialization with IVA-G, which we call IVA-GL [64]. In Figure 8(b), we show the receiver operating characteristics curves for the detection of the difference between the two groups at two levels of spread. To obtain the reference map, we perform a two-sample t-test between the two simulated groups, which is thresholded at 0.05 significance. By changing the threshold for t-values, we plot the receiver operating characteristics by counting the number of voxels within (N_{true}) and outside the reference map (N_{false}). The ratios of these values to the total number of voxels within the map and outside, respectively, yields the true positive—detection power—and false alarm values. IVA performs better than the Group ICA approach at both lower and higher group variability, and its performance improves when there is higher group variability. Because with higher group variability, diversity in terms of source dependence has more statistical power, improving the performance of the IVA approach.

Example: Performance of IVA and Group ICA with Real fMRI Data: To test the performance of IVA for the analysis of real fMRI data from subjects that exhibit significant variability, we used data collected from patients who suffered a stroke that primarily affected their motor areas. Data were collected while subjects performed a motor task that had alternating cycles of rest (30 sec) and task (24 sec), which was squeezing a ball. After standard preprocessing using SPM [56] as in [62], data from 10 subjects in two sessions, hence providing a total of 20 data sets, are analyzed using the two approaches for multi-subject analysis, Group ICA and IVA with the same algorithms as in the previous example, IVA-GL and Infomax with a nonlinearity matched to a Laplace pdf implemented using

Group ICA of fMRI Toolbox (GIFT) [65]. In Figure 9, we show the t-maps thresholded at a significance level of 0.05. As clearly observed from the figures, for both components, IVA leads to better estimation of the functional areas as indicated by more connected regions and better correlation with the masks generated for these areas as discussed in [66].

V. DISCUSSION

In this paper, we provided an overview of ICA by emphasizing two types of diversity, HOS and sample dependence, which have been the two most typically used for achieving ICA. By using mutual information rate, we showed that a number of ICA approaches can be brought under one umbrella. We then introduced IVA that extends ICA to multiple data sets, and presented a general formulation for IVA that adds a third type of diversity, dependence of sources across data sets to these two. We emphasized the parallels between ICA and IVA in the way diversity plays a role for both and discussed how a number of existing algorithms fit as special cases under this umbrella. Even though this is a rather broad umbrella, this has been a partial survey given the vast activity in the area. For example, algorithms that explicitly compute HOS such as JADE as well as those that make use of other types of diversity such as nonstationarity, noncircularity, and geometrical properties are not considered. Nonstationarity, however, is taken into account in our discussion on identifiability and it has been used with other types of diversity in algorithm development for example along with HOS in [67], and HOS, and sample dependence in [68].

Another important diversity type—which we could not discuss here due to space constraints—is non-circularity of the signals when ICA or IVA is implemented in the complex domain, which can be also studied under the mutual information rate umbrella [69]. Since it is the second-order-statistics that determine identifiability for ICA and IVA, again in this case, it is the impropriety, second-order noncircularity, that plays a key role. A random vector $\mathbf{x}$ is called second-order circular—or, proper—if its complementary covariance matrix $E\{\mathbf{xx}^\top\}$ vanishes. For improper signals, identification of i.i.d. Gaussians is possible if all circularity coefficients are distinct using strongly uncorrelating transform [70], and in addition, if we make use of HOS and sample dependence, it can be shown that the identification conditions we have given here become more relaxed. Then the ICA problem becomes non-identifiable *only* when there are Gaussian sources with *both* the covariance and complementary covariance matrices that are proportional, and proportional through a complex constant for the latter, as implied by the analyses in [71], [72]. A conjecture for IVA would be that the condition given here will also include the complementary covariances of SCVs. As one would expect, with the addition of each new type of diversity, identification becomes easier, a broader class of signals can be separated using ICA or IVA. A recent review of complex-valued ICA can be found in [69], and a comprehensive review of the field of blind source separation in the book [1].

ICA has found a fruitful application in fMRI analysis and IVA promises to be another attractive solution. ICA has been widely applied to fMRI and EEG analyses, two domains where the linear superposition assumption of ICA holds. A recent review [58] underlined the now wide use of ICA for fMRI analysis by showing the exponential growth in publications on the topic following its first application in 1998 [57]. An interesting recent claim was that ICA for fMRI has been successful because the widely used algorithms Infomax and FastICA—with kurtosis nonlinearity—select for super-Gaussian sources, hence it is sparsity that determines the final decomposition, and not independence [73]. A response to the article [74] showed that the examples in [73] were flawed and with the correct interpretation of underlying models in ICA, *it is indeed independence* that achieves a useful decomposition of the fMRI data. Still, it is worth noting that Infomax has been the most widely used algorithm for fMRI analysis, first, due to historical reasons—it was the first algorithm used—and then because its simple fixed score function is a good match for the fMRI sources and provides robustness. However, as the examples we give in this paper demonstrate, maximizing independence by using an algorithm with a flexible density matching mechanism can lead to improved performance. In addition, the fact that there is good support for the inherent linear superposition assumption of the basic ICA model of (1) suggests that the spatial maps can be regarded as hidden variables in the model, just like audio sources in a cocktail party problem. In these cases, maximizing the independence though flexible density matching, and making use of multiple types of diversity to approach the performance bound is meaningful. In applications such as data fusion, however, the approach is mostly exploratory in that the sources do not necessarily have physical meaning, they primarily help explain the data, relationships among modalities. Then, in this case, using a robust algorithm such as Infomax might be sufficient, as discussed in detail in a review on data-driven fusion [75].

The formulation of IVA we present here provides an attractive framework for joint blind source separation with numerous potential applications. These include those where MCCA has been applied such as medical data analysis and fusion, hyperspectral data analysis, blind equalization, and of course the first motivation for the IVA formulation, solution of the convolutive ICA problem. Among many others, multi-modality data fusion is an important application area for IVA as it would extend the successful application of MCCA [76] to include HOS without constraining the demixing matrix. IVA also presents number of challenges and interesting venues for future research. The well defined structure of IVA might allow a more flexible solution to the multidimensional ICA—also called subspace ICA—problem where components within each independent subspace are allowed to have dependences. Estimation of the density during adaptation—to truly approach the CRLB and improve performance—is a more difficult task than for ICA. The multivariate nature of the pdf makes the problem more challenging especially when the goal is not only modeling flexible marginals but also taking dependence among the components of an SCV into account. Hence, if successfully extended to the multivariate case, a flexible density model like EBM can achieve this desired balance, and potentially allow one to also account for sample dependence in the model and

estimate an SCM. This is not to say there are not challenges for ICA either. The performance of most ICA algorithms deteriorate when the number of sources increases as well as the noise level. The noisy and the undetermined cases still deserve much attention, and also the problem of nonlinear ICA. Hence, even though the field of blind source separation has now reached a maturity, there are still a good number of important challenges and problems that require our attention.

REFERENCES

[1] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Indepedent Component Analysis and Applications*. Academic Press, 2010.

[2] T. Adalı and S. Haykin, *Adaptive Signal Processing: Next Generation Solutions*. Hoboken, New Jersey: Wiley Interscience, 2010.

[3] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York, NY: Wiley, 2001.

[4] V. D. Calhoun, T. Adalı, J. J. Pekar, and G. D. Pearlson, "A method for making group inferences from functional MRI data using independent component analysis." *Human Brain Mapping*, vol. 14, no. 1, pp. 140–151, Nov. 2001.

[5] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, Dec. 1971.

[6] A. A. Nielsen, "Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 293–305, March 2002.

[7] J. Vía and I. Santamaría, "Adaptive blind equalization of SIMO systems based on canonical correlation analysis," in *Signal Processing Advances in Wireless Communications, 2005 IEEE 6th Workshop on*, June 2005, pp. 318–322.

[8] Y.-O. Li, W. Wang, T. Adalı, and V. D. Calhoun, "Joint blind source separation by multi-set canonical correlation analysis," *IEEE Trans. Signal Processing*, 2009.

[9] T. Kim, I. Lee, and T.-W. Lee, "Independent vector analysis: Definition and algorithms," in *Proc. 40th Asilomar Conf. Signals, Systems, Comput.*, 2006, pp. 1393–1396.

[10] M. Anderson, G.-S. Fu, R. Phlypo, and T. Adalı, "Independent vector analysis: Identification conditions and performance bounds," *IEEE Trans. Signal Processing*, accepted 2013.

[11] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387–392, April 1985.

[12] Y.-O. Li, T. Adalı, and V. D. Calhoun, "Estimating the number of independent components for fMRI data," *Human Brain Mapping*, vol. 28, no. 11, pp. 1251–1266, Nov. 2007.

[13] A. Bell and T. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, Nov. 1995.

[14] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Networks*, vol. 10, no. 3, pp. 626–634, May 1999.

[15] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc. Radar Signal Processing*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[16] R. Boscolo, H. Pan, and V. Roychowdhury, "Independent component analysis based on nonparametric density estimation," *IEEE Trans. Neural Networks*, vol. 15, no. 1, pp. 55–65, Jan. 2004.

[17] D. Pham and P. Garat, "Blind separation of mixtures of independent sources through a quasi maximum likelihood approach," *IEEE Trans. Signal Processing*, vol. 45, no. 7, pp. 1712–1725, July 1997.

[18] A. Hyvärinen, "Independent component analysis: Recent advances." *Philosophical Transactions of the Royal Society A*, vol. 371, no. 1984, pp. 1–19, Feb. 2013.

[19] L. Tong, R.-W. Liu, V. C. Soon, and Y.-F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. Circuits Syst.*, vol. 38, no. 5, pp. 499–509, May 1991.

[20] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[21] A. Yeredor, "Blind separation of Gaussian sources via second-order statistics with asymptotically optimal weighting," *IEEE Signal Proc. Letts.*, vol. 7, no. 7, pp. 197–200, July 2000.

[22] S. Hosseini, C. Jutten, and D. T. Pham, "Markovian source separation," *IEEE Trans. Signal Processing*, vol. 51, no. 12, p. 3009, Dec. 2003–3019.

[23] P. Tichavský, Z. Koldovský, and A. Yeredor, "A hybrid technique for blind separation of non-gaussian and time-correlated sources using a multicomponent approach," *IEEE Trans. Neural Networks*, vol. 19, no. 3, pp. 421–430, 2008.

[24] K. Hild, H. Attias, and S. Nagarajan, "An expectation-maximization method for spatio-temporal blind source separation using an AR-MoG source model," *IEEE Trans. Neural Networks*, vol. 19, no. 3, pp. 508–519, March 2008.

[25] X.-L. Li and T. Adalı, "Blind spatiotemporal separation of second and/or higher-order correlated sources by entropy rate minimization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Dallas, TX, March 2010, pp. 1934–1937.

[26] G.-S. Fu, R. Phlypo, M. Anderson, and T. Adalı, "Blind source separation by entropy rate minimization," *IEEE Trans. Signal Processing*, in review, 2013.

[27] T. M. Cover and J. A. Thomas, *Elements of Information theory*. Wiley, 1991.

[28] A. Yeredor, "Blind separation of Gaussian sources with general covariance structures: Bounds and optimal estimation," *IEEE Trans. Signal Processing*, vol. 58, no. 10, pp. 5057–5068, Oct. 2010.

[29] S.-I. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems, December 1995*, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds., vol. 8. The MIT Press, 1996, pp. 757–763.

[30] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.

[31] A. Cichocki, R. Unbehauen, and E. Rummert, "Robust learning algorithm for blind separation of signals," *Electronics Letts.*, vol. 30, no. 17, pp. pp. 1386–1387, Aug. 1994.

[32] J. Hérault and C. Jutten, "Space or time adaptive signal processing by neural networks models," in *Proc. Int. Conf. on Neural Networks for Computing*, Snowbird, Utah, 1986, pp. 206–211.

[33] Z. Koldovský, P. Tichavský, and E. Oja, "Efficient variant of algorithm FastICA for independent component analysis attaining the Cramér-Rao lower bound," *IEEE Trans. Neural Networks*, vol. 17, no. 5, pp. 1265–1277, Sep. 2006.

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

[35] G.-S. Fu, R. Phlypo, M. Anderson, X.-L. Li, and T. Adalı, "Algorithms for markovian source separation by entropy rate minimization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Vancouver, BC, May 2013, pp. 3248–3252.

[36] W. Lu and J. Rajapakse, "Constrained independent component analysis," in *Advances in Neural Information Processing Systems*, vol. 13, 2001, pp. 570–576.

[37] T. Adalı, H. Li, M. Novey, and J.-F. Cardoso, "Complex ICA using nonlinear functions," *IEEE Trans. Signal Processing*, vol. 56, no. 9, pp. 4356–4544, Sept. 2008.

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

[39] M. Anderson, X.-L. Li, P. A. Rodriguez, and T. Adalı, "An effective decoupling method for matrix optimization and its application to the ICA problem," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Kyoto, Japan, March 2012, pp. 1885–1888.

[40] A. Yeredor, "Performance analysis of the strong uncorrelating transformation in blind separation of complex-valued sources," *IEEE Trans. Signal Processing*, vol. 60, no. 1, pp. 478–483, Jan. 2012.

[41] A. Cichocki, S. Amari, K. Siwek, *et al.* ICALAB toolboxes. [Online]. Available: http://www.bsp.brain.riken.jp/ICALAB

[42] J.-F. Cardoso, "Multidimensional independent component analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, vol. 4, Seattle, WA, May 1998, pp. 1941–1944.

[43] A. Hyvärinen and P. Hoyer, "Emergence of phase- and shift-invariant features by decomposition of emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces," *Neural Computation*, vol. 12, no. 7, pp. 1705–1720, July 2000.

[44] D. Lahat, J.-F. Cardoso, and H. Messer, "Second-order multidimensional ICA: Performance analysis," *IEEE Trans. Signal Processing*, vol. 60, no. 9, pp. 4598–4610, Sept. 2012.

[45] M. Anderson, "Independent vector analysis: Theory, algorithms, and applications," Ph.D. dissertation, University of Maryland Baltimore County, May 2013.

[46] J. Via, M. Anderson, X.-L. Li, and T. Adalı, "Joint blind source separation from second-order statistics: Necessary and sufficient identifiability conditions," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 2520 – 2523.

[47] M. Anderson, X.-L. Li, and T. Adalı, "Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis," *IEEE Trans. Signal Processing*, vol. 60, no. 4, pp. 2049–2055, April 2012.

[48] X.-L. Li, T. Adalı, and M. Anderson, "Joint blind source separation by generalized joint diagonalization of cumulant matrices," *Signal Processing*, vol. 91, no. 10, pp. 2314–2322, Oct. 2011.

[49] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 1, pp. 70–79, Jan. 2007.

[50] J. Hao, I. Lee, T.-W. Lee, and T. J. Sejnowski, "Independent vector analysis for source separation using a mixture of Gaussians prior," *Neural Computation*, vol. 22, no. 6, pp. 1646–1673, 2010.

[51] H. Zhang, L. Li, and W. Li, "Independent vector analysis for convolutive blind noncircular source separation," *Signal Processing*, vol. 92, no. 9, pp. 2275–2283, 2012.

[52] M. Anderson, G.-S. Fu, R. Phlypo, and T. Adalı, "Independent vector analysis, the Kotz distribution, and performance bounds," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Vancouver, BC, May 2013, pp. 3243–3247.

[53] T. Itahashi and K. Matsuoka, "Stability of independent vector analysis," *Signal Processing*, vol. 92, no. 8, pp. 1809–1820, Aug. 2012.

[54] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, Dec. 1936.

[55] J. Via, M. Anderson, X.-L. Li, and T. Adalı, "A maximum likelihood approach for independent vector analysis of gaussian data sets," in *Proc. IEEE Int. Workshop on Machine Learning for Signal Processing (MLSP)*, Beijing, China, 2011.

[56] K. J. Friston, J. T. Ashburner, S. Kiebel, T. Nichols, and W. D. Penny, Eds., *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press, 2006.

[57] 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," *Human Brain Mapping*, vol. 6, no. 3, pp. 160–188, 1998.

[58] 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, Aug. 2012.

[59] A. M. Dale and R. L. Buckner, "Selective averaging of rapidly presented individual trials using fMRI," *Human Brain Mapping*, vol. 5, no. 5, pp. 329–340, 1997.

[60] W. Du, H. Li, X.-L. Li, V. Calhoun, and T. Adalı, "ICA of fMRI data: Performance of three ICA algorithms and the importance of taking correlation information into account," in *Proc. IEEE Int. Symp. Biomedical Imaging*, Chicago, IL, March-April 2011, pp. 1573–1576.

[61] E. Erhardt, S. Rachakonda, E. Bedrick, T. Adalı, and V. D. Calhoun, "Comparison of multi-subject ica methods for analysis of fMRI data," *Human Brain Mapping*, vol. 32, no. 12, pp. 2075–2095, Dec. 2011.

[62] V. D. Calhoun, J. Liu, and T. Adalı, "A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data," *NeuroImage*, vol. 45, no. 1, pp. S163–S172, March 2009.

[63] E. Erhardt, E. Allen, Y. Wei, T. Eichele, and V. D. Calhoun, "SimTB, a simulation toolbox for fMRI data under a model of spatiotemporal separability," *NeuroImage*, vol. 59, no. 4, pp. 4160–4167, Feb. 2012.

[64] S. Ma, R. Phlypo, V. D. Calhoun, and T. Adalı, "Capturing group variability using IVA: A simulation study and graph-theoretical analysis," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Vancouver, BC, May 2013, pp. 3128–3132.

[65] Group ICA of fMRI toolbox. [Online]. Available: http://mialab.mrn.org/software/gift/

[66] J. Laney, K. Westlake, E. Woytowicz, S. Ma, and T. Adalı, "Capturing subject variability in data-driven fMRI analysis: A graph-theoretical comparison," in *Proc. Conf. Information Sciences and Systems*, Princeton, NJ, March 2014.

[67] Z. Koldovský, J. Málek, P. Tichavský, Y. Deville, and S. Hosseini, "Blind separation of piecewise stationary non-Gaussian sources," *Signal Processing*, vol. 89, no. 12, pp. 2570–2584, Dec. 2009.

[68] H. Buchner, R. Aichner, and W. Kellermann, *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, ser. Editors, Y. Huang and J. Benesty. Boston/Dordrecht/London: Kluwer Academic Publishers, Feb. 2004, ch. BLIND SOURCE SEPARATION FOR CONVOLUTIVE MIXTURES: A UNIFIED TREATMENT, pp. 255–293.

[69] E. Moreau and T. Adalı, *Blind Identification and Separation of Complex-valued Signals*. London, UK and Hoboken, NJ: ISTE and Wiley, 2013.

[70] J. Eriksson and V. Koivunen, "Complex-valued ICA using second order statistics," in *Proc. IEEE Int. Workshop on Machine Learning for Signal Processing (MLSP)*, Saõ Luis, Brazil, Sept. 2004, pp. 183–192.

[71] B. Loesch and B. Yang, "Cramér-Rao bound for circular and noncircular complex independent component analysis," *IEEE Trans. Signal Processing*, vol. 61, no. 2, pp. 365–379, Jan. 2013.

[72] X.-L. Li and T. Adalı, "Blind separation of noncircular correlated sources using Gaussian entropy rate," *IEEE Trans. Signal Processing*, vol. 59, no. 6, pp. 2969–2975, June 2011.

[73] I. Daubechies, E. Roussos, S. Takerkarta, M. Benharrosha, C. Goldenb, K. D'Ardennea, W. Richtera, J. D. Cohena, and J. Haxbyaw, "Independent component analysis for brain fMRI does not select for independence," *Proc. National Academy of Sciences*, vol. 106, no. 26, pp. 10 415–10 422, June 2009.

[74] V. D. Calhoun, V. K. Potluru, R. Phlypo, R. F. Silva, B. A. Pearlmutter, A. Caprihan, S. M. Plis, and T. Adalı, "Independent component analysis for brain fMRI does indeed select for maximal independence," *PLoS ONE*, vol. 8, no. 8, 2013.

[75] V. D. Calhoun and T. Adalı, "Feature-based fusion of medical imaging data," *IEEE Trans. Info. Technology in Biomedicine*, vol. 13, no. 5, pp. 711–720, Sep. 2009.

[76] N. Correa, T. Adalı, and V. D. Calhoun, "Canonical correlation analysis for data fusion and group inferences: Examining applications of medical imaging data," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 39–50, July 2010.