

IVA AND ICA: USE OF DIVERSITY IN INDEPENDENT DECOMPOSITIONS

Tülay Adalı, Matthew Anderson, and Gengshen Fu

University of Maryland Baltimore County, Baltimore, MD 21250

ABSTRACT

Starting with a simple linear 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. Most of the ICA algorithms introduced to date have made use of one of the two types of diversity, non-Gaussianity or sample dependence. We first discuss the main results for ICA in terms of identifiability and performance with these two types of diversity, and then introduce independent vector analysis (IVA), generalization of ICA for decomposition of multiple datasets at a time. We show that the role of diversity in this case parallels that in ICA, and discuss identifiability conditions and performance bounds in a maximum likelihood framework.

Index Terms— Source separation, maximum likelihood, identifiability and performance

1. INTRODUCTION

Independent component analysis (ICA) achieves successful decompositions of a given set of observations under the assumption of statistical independence and has been an active area of research with a wide array of applications [10]. A recent generalization of ICA, independent vector analysis (IVA) [3, 14, 15] decomposes multiple datasets *simultaneously* using the same generative model such that the source estimates are aligned across all the datasets. The problem arises in many domains such as medical imaging when analyzing multi-subject functional magnetic resonance imaging (fMRI) data [19] and when solving the convolutive ICA problem in the frequency domain across multiple frequency bins [14]. Both ICA and IVA make use of diversity in the datasets when achieving the decomposition, and the two types of diversity that are most frequently used are non-Gaussianity and sample dependence of the underlying components. As one would expect, the role of diversity—and hence the results for identifiability and performance—in the two approaches parallel each other closely. In this overview article, we first present a review of the main results for ICA emphasizing the role two important types of diversity plays in terms of performance and identifiability conditions under the umbrella of maximum likelihood (ML) theory, and then present new results for IVA that parallel those results and generalize them to the multivariate case. We emphasize the fact that *linear* dependence (correlation) is

a strong concept in that it is sufficient for the separation of sources under mild conditions both for ICA and IVA. However in terms of achievable performance, we can have significant gains by making use of non-Gaussianity, *i.e.*, higher-order-statistics along with sample dependence.

2. ICA

We start with the basic ICA problem where $\mathbf{x}, \mathbf{s} \in \mathbb{R}^N$ and write

$$\mathbf{x}(v) = \mathbf{A}\mathbf{s}(v), \quad 1 \leq v \leq V. \quad (1)$$

where \mathbf{A} is a full rank square mixing matrix, and hence we assume instantaneous mixing and as many observations x_n as sources/components s_n —which also includes the overdetermined case since one can easily reduce the problem to (1) for this case. We assume that the index v can be time, or a spatial or volume index, a voxel as in the case of fMRI analysis.

Given that the sources are mutually independent, one can achieve ICA and form the source estimates $\mathbf{u}(v) = \mathbf{W}\mathbf{x}(v)$ by estimating the demixing matrix \mathbf{W} making use of diversity in some form. The most popular approach has been the use of non-Gaussianity as the form of diversity, *i.e.*, using higher-order-statistics (HOS) to achieve the decomposition. Under this umbrella, one can either start with mutual information as the cost and arrive at the two most popular approaches, based on either maximum likelihood (ML) or maximization of negentropy to achieve the ICA decomposition, or can explicitly calculate HOS as in joint approximate diagonalization of eigenmatrices (JADE) [7] among others [9, 10]. In this overview, we concentrate on the former approach as it allows the study of large sample properties in an ML framework, and can also be used to study properties of another important class of ICA algorithms under the same umbrella, those that make use of sample dependence within a component/source as the source of *diversity*. We start our discussion with the general case and with the cost that takes advantage of sample dependence together with non-Gaussianity, *i.e.*, HOS, of the sources. The natural cost in this case is the mutual information rate [10, 17], which can be written as

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

where $H_r(u_n) = \lim_{v \rightarrow \infty} H[u_n(1), \dots, u_n(v)]/v$ is the entropy rate of the n th source estimate u_n . The entropy rate $H_r(\mathbf{x}) = \lim_{v \rightarrow \infty} H[\mathbf{x}(1), \dots, \mathbf{x}(v)]/v$ of the observations is a constant with respect to \mathbf{W} and thus the statistical dependence among the separated sources is naturally minimized by

This work is supported by the NSF grants NSF-III 1017718 and NSF-CIF 1117056.

minimizing the *total entropy rate of all source estimates*. The regularization term $\log |\det(\mathbf{W})|$ penalizes ill-conditioned matrices, and reduces (2) to maximization of negentropy rate as the cost, *i.e.*, minimization of sum of entropy rates under a variance constraint when \mathbf{W} is constrained to be orthogonal ($\mathbf{W}\mathbf{W}^T = \mathbf{I}$) so that the term is 0.

We assume that the sources are stationary, hence we have $H_r(u_n) = \lim_{v \rightarrow \infty} H(u_n(v) | u_n(v-1), u_n(v-2), \dots, u_n(1))$. Further, to simplify the discussion for the ICA-IVA connections in terms of sample dependence as the source of diversity, we let the extent of dependence be finite and given by K which yields

$$H_r(u_n) = H(\mathbf{u}_n) - H(\tilde{\mathbf{u}}_n) \quad (3)$$

where $\mathbf{u}_n = [u_n(v), u_n(v-1), \dots, u_n(v-K+1)]^T \in \mathbb{R}^K$, $\tilde{\mathbf{u}}_n = [u_n(v-1), \dots, u_n(v-K+1)]^T \in \mathbb{R}^{K-1}$ and the expectation is with respect to $p_{s_n}(\mathbf{u}_n)$. We include the index for the random vector \mathbf{u}_n and specifically write $\mathbf{u}_n(v)$ when we refer to a given observation.

For a given set of observations $\mathbf{X} \in \mathbb{R}^{N \times V}$ where $V \gg K$, we estimate the sources using $\tilde{\mathbf{U}}(v) = \mathbf{W}\tilde{\mathbf{X}}(v)$ where $\tilde{\mathbf{U}}(v) = [\mathbf{u}_1(v), \mathbf{u}_2(v), \dots, \mathbf{u}_N(v)]^T \in \mathbb{R}^{N \times K}$ and $\tilde{\mathbf{X}}$ is defined similarly. Using this notation for finite memory conditions we have the following log likelihood

$$\begin{aligned} \mathcal{L}(\mathbf{W}) &= \frac{1}{V-K+1} \sum_{v=K}^V \sum_{n=1}^N [\log p_{s_n}(\mathbf{u}_n(v)) - \log p_{s_n}(\tilde{\mathbf{u}}_n(v))] \\ &\quad + \log |\det \mathbf{W}| \\ &= \frac{1}{V-K+1} \sum_{v=K}^V \sum_{n=1}^N \log p_{s_n}(u_n(v) | u_n(v-1), \dots, \\ &\quad u_n(v-K+1)) + \log |\det \mathbf{W}| \end{aligned} \quad (4)$$

The resulting expression is the conditional log likelihood to be maximized for given V observations—and hence the sign in the cost (2) is reversed. When we assume independent and identically distributed (i.i.d.) samples, *i.e.*, ignore sample dependence, in this case, the asymptotic equipartition property directly yields from (2) the more commonly encountered form of likelihood in ICA formulations

$$\mathcal{L}(\mathbf{W}) = \frac{1}{V} \sum_{v=1}^V \sum_{n=1}^N \log p_{s_n}(u_n(v)) + \log |\det \mathbf{W}| \quad (5)$$

since non-Gaussianity is the more frequently used form of diversity in ICA. Here, $u_n(v) = \mathbf{w}_n^T \mathbf{x}(v)$, using the random vector notation for the observations (mixture) and \mathbf{w}_n^T is the n th row of \mathbf{W} . It is the form in (5) that leads to the popular Infomax [4], along with all the ML variations using different density models, *e.g.*, using adaptive scores [21] or entropy bound minimization (EBM) as in [18]—and when the demixing matrix is constrained to be orthogonal as in FastICA [12], and its variants such as efficient FastICA (EFICA) [16].

2.1. The gradient and the stationary point

Given the cost function, we can first study the stationary point (or the estimating functions [6]), *i.e.*, the solution of $\nabla \mathcal{I}_r(\mathbf{W}) =$

$\partial \mathcal{I}_r(\mathbf{W}) / \partial \mathbf{W} = \mathbf{0}$, or in the case of ICA, solution of the relative/natural gradient which yields a more convenient form for identifying the condition at the stationary point besides providing advantages in the update. It is the gradient post-multiplied by $\mathbf{W}^T \mathbf{W}$, which is a positive definite matrix and hence does not change the direction of the gradient, and eliminates the inversion of the demixing matrix in the update.

Using relative/natural gradient updates [1, 6], the likelihood function in (4) can be maximized using

$$\Delta \mathbf{W} = (\mathbf{I} - E\{\psi(\tilde{\mathbf{U}})\tilde{\mathbf{U}}^T\})\mathbf{W} \quad (6)$$

where the entries of $\psi(\tilde{\mathbf{U}}) \in \mathbb{R}^{N \times K}$, are given by

$$\{\psi(\tilde{\mathbf{U}})\}_{n,k} = -\frac{\partial \log p_{s_n}(\mathbf{u}_n)}{\partial \mathbf{u}_n^T \mathbf{e}_k}$$

where \mathbf{e}_k is the k th standard basis vector. When writing the update, we have ignored the second pdf in (4), the pdf of the $K-1$ -dimensional random vector, which can be included for potentially better performance. Note that for a *true* ML estimation, we also need to estimate the probability density function (pdf) $p_{s_n}(\cdot)$, which in this case is a multivariate pdf since sample dependence is taken into account. An important special case is given for the Gaussian source as $\psi_n(\mathbf{u}_n) = \mathbf{R}_n^{-1} \mathbf{u}_n$ where $\mathbf{R}_n = E\{\mathbf{u}_n \mathbf{u}_n^T\}$.

If we assume that the samples are i.i.d., then the entries of the score function are given by

$$\psi_n(u_n) = -\frac{\partial \log p_{s_n}(u_n)}{\partial u_n}$$

which can be estimated much more easily than the previous case as now we have the pdf of a univariate variable. It is also worth noting that in many instances instead of estimating the pdf, a simple nonlinearity selection mechanism can be implemented depending on the super and sub-Gaussianity of the sources, an approach that provides satisfactory performance for most scenarios, but is a *quasi* ML approach rather than a true ML approach.

What defines the stationary point $\Delta \mathbf{W} = \mathbf{0}$ is the *pairwise condition*, and for the general case given in (6), it is

$$E\{\psi_n(\mathbf{u}_n^T) \mathbf{u}_l\} = 0 \text{ for } n \neq l$$

which reduces to $E\{\psi_n(u_n) u_l\} = 0$ for i.i.d. samples, and has been the motivation behind the first approach to ICA using *nonlinear decorrelations* [8, 11]. Also, it is worth noting that for this case, if we were to only use second-order statistics and choose the linear Gaussian score function, since $\mathbf{R}_n = \sigma_n^2 \mathbf{I}$ where σ_n^2 is the variance of the n th source, the stable point then corresponds to $E\{u_k u_l\} = 0$ for $k \neq l$, and in this case, we can only achieve decorrelation.

2.2. The Hessian, CRLB, and identifiability

We assume that the mixtures are whitened and the sources have unit variance, which implies that the product $\mathbf{G} = \mathbf{W}\mathbf{A}$ should approximate a permutation matrix at the optimum and hence $u_n = s_n$. With these assumptions, the Cramér Rao lower bound (CRLB) and the induced CRLB, *i.e.*, the bounds for the parameters that are actually estimated, \mathbf{W} , and \mathbf{G} coincide, and hence we have $E\{g_{m,n}^2\} = \text{var}(w_{m,n})$.

For the local stability analysis of the ML cost in (4), we form the Hessian with respect to the global demixing matrix $\mathbf{G} = \mathbf{W}\mathbf{A}$, with entries $\partial^2 \mathcal{L} / \partial g_{m_1, n_1} \partial g_{m_2, n_2}$, where $g_{m,n}$

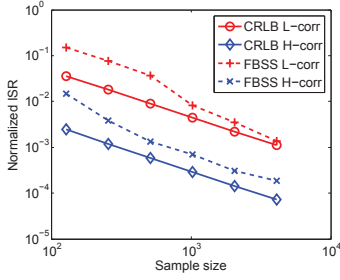


Fig. 1. CRLB and the performance of FBSS for correlated Gaussian sources with low and high levels of sample correlation

are the entries of \mathbf{G} . We evaluate the negative of the expected value of the Hessian at the optimum $\mathbf{G} = \mathbf{I}$, assuming perfect knowledge of the source distributions to obtain the Fisher information matrix (FIM). The positive definiteness of the FIM is determined by the 2×2 matrix [5]

$$\mathbf{J}_{m,n} = \begin{bmatrix} \kappa_{m,n} & 1 \\ 1 & \kappa_{n,m} \end{bmatrix}, \quad (7)$$

where $\kappa_{m,n} = \text{trace} \left(E \left\{ \psi_n(\mathbf{s}_n) \psi_n^T(\mathbf{s}_n) \right\} \mathbf{R}_m \right)$, *i.e.*, by the pairwise relationship of sources for $1 \leq m < n \leq N$ and $\mathbf{J}_{m,n}$ is positive definite *unless* there are any two sources that are Gaussian with proportional covariance matrices. Hence, one can achieve source separation unless there are two sources that are both Gaussian and have covariance matrices that satisfy $\mathbf{R}_n = \alpha \mathbf{R}_m$ for $\alpha \neq 0$, a condition that also guarantees identification of sources when only second-order statistics are used as in weights-adjusted second-order blind identification (WASOBI) algorithm [22]. When we ignore sample dependence and use the form in (5), we have $\mathbf{R}_n = \mathbf{I}$ for all sources (since we assume they all have unit variance) and can only make use of non-Gaussianity, *i.e.*, HOS, and hence in this case we can separate sources as long as there is only one Gaussian source in the mixture [9]—and obviously for both cases, the separation is possible only upto a scaling and permutation ambiguity that is inherent to the problem. The CRLB can be evaluated using the inverse of $\mathbf{J}_{m,n}$ as [5]

$$\text{var}(w_{m,n}) \geq \frac{1}{V} (\kappa_{m,n} - \kappa_{n,m}^{-1})^{-1}, \quad n \neq m \quad (8)$$

where for the Gaussian score, we have $\kappa_{m,n} = \text{Trace}(\mathbf{R}_m^{-1} \mathbf{R}_n)$.

In Figures 1 and 2, we show the CRLB for two cases: correlated Gaussian and i.i.d. super-Gaussian sources, realizations from a zero mean generalized Gaussian distribution (GGD) given by $p_s(s) = \frac{\beta}{\alpha \Gamma(1/2\beta)} e^{-(|s|/\alpha)^{2\beta}}$ where $\Gamma(\cdot)$ is the Gamma function and β is the shape parameter such that $\beta = 1$ corresponds to Gaussian. Parameter α is chosen so that the variance is unity. Results are shown for two cases: (i) two Gaussian sources with two different degrees of sample correlation, lower correlation where the span of correlation is two, and a second case where it is four and the first two correlation coefficients are the same as the first (lower

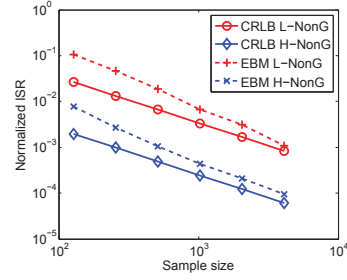


Fig. 2. CRLB and the performance of EBM with i.i.d. super-Gaussian sources with low and high levels of non-Gaussianity

correlation) case; and (ii) two i.i.d. sources with both shape parameters β first chosen as 0.7, and then as 0.4, *i.e.*, corresponding to lower and higher non-Gaussianity as measured by the distance from that of a Gaussian. We plot the normalized ISR given by $(1/N(N-1)) \sum_{\{m,n=1,m \neq n\}}^N E\{g_{m,n}^2\}$. In the figure, along with the CRLB for these cases, we also show the performance of two recent algorithms, EBM [18], which ignores sample dependence but uses a flexible density matching scheme, and full blind source separation (FBSS) [17], which uses mutual information rate in (2) as the cost and includes an effective model for sample dependence besides EBM-based density estimation. As observed in the figures, the CRLB decreases with increasing correlation or non-Gaussianity, and the two algorithms approach the bound as the number of samples increase.

3. IVA

The IVA problem is defined similar to ICA except that we now have K data sets, each containing V samples, formed from linear mixtures of N independent sources [15],

$$\mathbf{x}^{[k]}(v) = \mathbf{A}^{[k]} \mathbf{s}^{[k]}(v), \quad 1 \leq k \leq K, \quad 1 \leq v \leq V. \quad (9)$$

The v th sample of the zero-mean source vector, $\mathbf{s}^{[k]}(v) = [s_1^{[k]}(v), \dots, s_N^{[k]}(v)]^T \in \mathbb{R}^N$, is a realization of the random vector $\mathbf{s}^{[k]}$, and $\mathbf{A}^{[k]} \in \mathbb{R}^{N \times N}$, $k = 1, \dots, K$ are the invertible mixing matrices. The underlying assumption is that the individual source component vectors (SCVs) $\mathbf{s}_n^T = [s_n^{[1]}, \dots, s_n^{[K]}]$ are statistically independent, *i.e.*, the pdf of the concatenated source vector, $\mathbf{s}^T = [(\mathbf{s}^{[1]})^T, \dots, (\mathbf{s}^{[K]})^T] \in \mathbb{R}^{NK}$ can be written as $p(\mathbf{s}) = \prod_{n=1}^N p_n(\mathbf{s}_n)$. Note that in this case, different from ICA, we do not consider sample dependence, and in terms of notation, let \mathbf{s}_n refer to the n th SCV, which is a *random vector*, rather than the realization of a given K -dimensional random vector. We show that linear dependence as a form of diversity plays a similar role to that in ICA, *but* now in terms of the use of dependence within a source component vector, the vector that contains all the corresponding sources from K data sets.

The problem is now that of finding K demixing matrices $\mathbf{W}^{[k]}$ such that sources for each data set can be estimated through $\mathbf{u}^{[k]} = \mathbf{W}^{[k]} \mathbf{x}^{[k]}$ for $k = 1, \dots, K$. The estimate of the n th SCV is given as $\mathbf{u}_n^T = [u_n^{[1]}, \dots, u_n^{[K]}]$. The mixing matrices are potentially distinct for each data set and are not necessarily related. Note that as in the case of ICA, the sources and the mixing matrix can only be identified up to an arbitrary order and scaling ambiguity [9]. However, by posing the problem as a *joint* independent decomposition, the goal is to eliminate the global permutation ambiguity in that the arbitrary order of the sources within a data set is retained across all K datasets.

As the cost, we use mutual information rather than mutual information rate, as we are not taking the sample dependence into account. The goal hence is the minimization of mutual information among the SCVs—rather than individual (univariate) sources as in ICA—and is written as

$$\mathcal{I}_{\text{IVA}}(\mathbf{W}) = \sum_{n=1}^N H(\mathbf{u}_n) - \sum_{k=1}^K \log |\det(\mathbf{W}^{[k]})| - C \quad (10)$$

where C is $H(\mathbf{x}^{[1]}, \dots, \mathbf{x}^{[K]})$, a constant with respect to the demixing matrices. We can rewrite the cost as [3]

$$\begin{aligned} \mathcal{I}_{\text{IVA}}(\mathbf{W}) = & \sum_{n=1}^N \left(\sum_{k=1}^K H(u_n^{[k]}) - \mathcal{I}(\mathbf{u}_n) \right) \\ & - \sum_{k=1}^K \log |\det(\mathbf{W}^{[k]})| - C \end{aligned}$$

to clarify the ultimate goal in the decomposition for IVA, which is making sure that when we minimize the entropy of all components to achieve independence, we also maximize mutual information, *i.e.*, the dependence within each estimated SCV.

3.1. The gradient and the stationary point

The relative/natural gradient update of the demixing matrix for each data set takes the form

$$\Delta \mathbf{W}^{[k]} = \left(\mathbf{I} - E \left\{ \boldsymbol{\psi}^{[k]}(\mathbf{u}) \left(\mathbf{u}^{[k]} \right)^T \right\} \right) \mathbf{W}^{[k]} \quad (11)$$

as in ICA, and now the score vector is given by

$$\boldsymbol{\psi}^{[k]}(\mathbf{u}) = \left[\psi_1^{[k]}(\mathbf{u}_1), \dots, \psi_N^{[k]}(\mathbf{u}_N) \right]^T \in \mathbb{R}^N$$

where $\psi_n^{[k]}(\mathbf{u}_n) = -\partial \log p_{s_n}(\mathbf{u}_n) / \partial u_n^{[k]}$, *i.e.*, it is formed by selecting the k th entries from each of the N multivariate score functions, $\boldsymbol{\psi}(\mathbf{u}_n) = -\partial \log p_{s_n}(\mathbf{u}_n) / \partial \mathbf{u}_n \in \mathbb{R}^K$. It is important to note that $\boldsymbol{\psi}^{[k]}(\mathbf{u}_n)$ is a function of the multivariate random vector, *i.e.*, the n th estimated SCV, and in implementation, we have to estimate the multivariate pdf along with the demixing matrices to achieve a *true* ML solution. The important special case, Gaussian pdf, gives $\boldsymbol{\psi}(\mathbf{u}_n) = \mathbf{R}_n^{-1} \mathbf{u}_n$

whose k th entry is $\psi_n^{[k]}(\mathbf{u}_n) = \mathbf{e}_k^T \mathbf{R}_n^{-1} \mathbf{u}_n$ where \mathbf{e}_k is the k th basis vector and $\mathbf{R}_n = E\{\mathbf{u}_n \mathbf{u}_n^T\}$. In this case, the stationary point is given by

$$E\{\psi_n(u_n) u_l\} = 0 \text{ for } n \neq l, \text{ for } k = 1, \dots, K.$$

3.2. The Hessian, CRLB, and identifiability

For IVA, similarly we evaluate the Hessian at $\mathbf{G} = \mathbf{I}$ when the source estimates are given by s_n , and now the performance and identifiability conditions are determined by the *pairwise interactions of the SCVs* rather than sources as in ICA:

$$\mathbf{J}_{m,n} = \begin{bmatrix} \mathcal{K}_{m,n} & \mathbf{I} \\ \mathbf{I} & \mathcal{K}_{n,m} \end{bmatrix}, \quad 1 \leq m < n \leq N, \text{ where} \quad (12)$$

$$\{\mathcal{K}_{m,n}\}_{k_1, k_2} = E \left\{ \frac{\partial \psi_{m,n}^{[k_1]}(\mathbf{u}_m)}{\partial u_m^{[k_2]}} \right\} \{\mathbf{R}_n\}_{k_1, k_2}, \quad 1 \leq m, n \leq N$$

and $\{\mathbf{R}_n\}_{k_1, k_2} = E\{s_n^{[k_1]} s_n^{[k_2]}\}$. From (12), we can write the CRLB for the estimates of the demixing matrix entries as [2]

$$\text{var}(w_{m,n}^{[k]}) \geq \frac{1}{V} \mathbf{e}_k^T (\mathcal{K}_{m,n} - \mathcal{K}_{n,m}^{-1})^{-1} \mathbf{e}_k, \quad 1 \leq m \neq n \leq N. \quad (13)$$

For a single data set, we have $\mathcal{K}_{m,n} = E\left\{\frac{\partial \psi_m(u_m)}{\partial u_m}\right\} \sigma_n^2 = \kappa_m$, since $\sigma_n^2 = 1$, now a scalar quantity, derivative of univariate score function for $K = 1$. The CRLB then assumes the form in [20]

$$\text{var}(w_{m,n}) \geq \frac{1}{V} (\kappa_m - \kappa_n^{-1})^{-1}, \quad m \neq n$$

where the bound is derived for the widely employed case that only considers non-Gaussianity ignoring source dependence. This bound can also be recognized as a special case of (8), since for i.i.d. sources, $\mathbf{R}_m = \mathbf{I}$, and $\kappa_{m,n}$ in (7) is written as $E\{\psi^2(u_n)\} = -E\{\partial \psi(u_n) / \partial u_n\} = \kappa_n$ using the regularity condition $E\{\psi(u)\} = 0$, which we assume holds [13].

Another important special case for the IVA approach, which also leads to the conditions for identifiability, is obtained using the multivariate Gaussian prior for the SCV, hence ignoring the higher-order statistical information. In this case, we have [3]

$$\mathcal{K}_{m,n} = \mathbf{R}_m^{-1} \circ \mathbf{R}_n, \quad 1 \leq m \neq n \leq N \quad (14)$$

and the Fisher information matrix is a positive semidefinite block diagonal matrix, and is singular *if and only if*

$$\mathbf{R}_m^{-1} \circ \mathbf{R}_n - (\mathbf{R}_n^{-1} \circ \mathbf{R}_m)^{-1} \succeq 0$$

holds for any $1 \leq m \neq n \leq N$, *i.e.*, the expression is positive semi-definite—assuming nonsingular covariance matrices. Here, $\mathbf{A} \circ \mathbf{B}$ denotes the Hadamard or Schur element-wise matrix product of \mathbf{A} and \mathbf{B} . If any two SCVs both

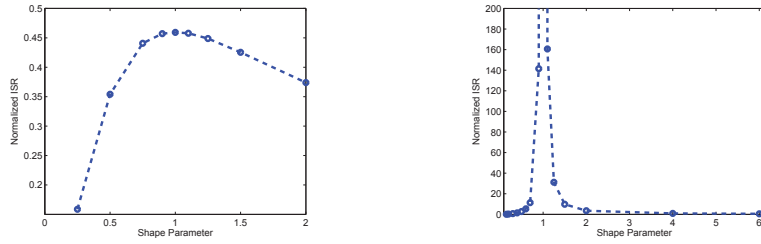


Fig. 3. CRLB as a function of the shape parameter, β for three SCVs with (left) random, and (right) identity covariance matrices. Note that the SCV is multivariate Gaussian for $\beta = 1$.

have block diagonal covariance matrices where at least two of the equivalent blocks of the covariance matrices are essentially identical, *i.e.*, they satisfy $\mathbf{R}_n = \mathbf{D}\mathbf{R}_m\mathbf{D}$ for $1 \leq m \neq n \leq N$, where \mathbf{D} is any full rank diagonal matrix, then the FIM becomes positive semidefinite. Note the similarity of this condition to the one for ICA when sample dependence is taken into account, which states that the sources cannot have proportional covariance matrices. Finally it is worth noting that the CRLB for IVA (13) with the multivariate Gaussian assumes a form similar to the one in (8), which is given in [22] for source separation using second-order statistics $\text{var}(w_{m,n}) \geq (\kappa_{m,n} - \kappa_{n,m}^{-1})^{-1} / V$.

We study the behavior of CRLB for two cases. First in Figure 3, we show the CRLB for three SCVs each with a randomly selected covariance matrix that admits a solution with the Gaussian multivariate prior for $K = 3$. In this case, the normalized ISR (induced CRLB) is defined as $\text{ISR}_{\text{norm}} = 1 / KN(N-1) \sum_{k=1}^K \sum_{\{m,n=1,m \neq n\}}^N (g_{m,n}^{[k]})^2$. We control the non-Gaussianity of the SCV, again using a shape parameter β for the multivariate GGD. The normalized ISR decreases as the SCVs become more non-Gaussian and is maximum when $\beta = 1$, for the multivariate Gaussian case. The second case shown in Figure 3 on the right uses the same SCVs but now without any dependence within an SCV. Note the divergence for multivariate Gaussian ($\beta = 1$) when there is no linear dependence to be used to achieve diversity. Even though the range of normalized ISR values in the figure makes it difficult to observe, the CRLB is significantly higher for the uncorrelated case compared to the first (correlated) case even when we move away from Gaussianity, *i.e.*, for values of β away from 1.

4. REFERENCES

- [1] S.-I. Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, 1998.
- [2] M. Anderson and T. Adali. IVA: Performance, identifiability, and algorithms using Kotz distributions. In *Proc. IEEE Int. Workshop on Machine Learning for Signal Processing (MLSP)*, Santander, Spain, 2012.
- [3] M. Anderson, X.-L. Li, and T. Adali. Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis. *IEEE Trans. Signal Processing*, 2012.
- [4] A. Bell and T. Sejnowski. An information maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.
- [5] J.-F. Cardoso. *Chapter on Likelihood in Handbook of Blind Source Separation*. Academic Press (P. Comon and C. Jutten editors), 2010.
- [6] J.-F. Cardoso and B. Laheld. Equivariant adaptive source separation. *IEEE Trans. Signal Processing*, 44:3017–3030, 1996.
- [7] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-Gaussian signals. *IEE Proc. Radar Signal Processing*, 140:362–370, 1993.
- [8] A. Cichocki and R. Unbehauen. Robust neural networks with online learning for blind identification and blind separation of sources. *IEEE Trans. Circuits Syst. I: Fund. Theory Apps.*, 43:894–906, 1996.
- [9] P. Comon. Independent component analysis – a new concept? *Signal Processing*, 36:287–314, 1994.
- [10] P. Comon and C. Jutten. *Handbook of Blind Source Separation*. Academic Press, 2010.
- [11] J. Héroult and C. Jutten. Space or time adaptive signal processing by neural networks models. In *Proc. Int. Conf. on Neural Networks for Computing*, pages 206–211, Snowbird, Utah, 1986.
- [12] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Networks*, 10(3):626–634, 1999.
- [13] S. M. Kay. *Fundamentals of Statistical Processing, Volume I: Estimation Theory*. Prentice Hall Signal Processing Series, Upper Saddle River, NJ, 1993.
- [14] 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.*, 15(1):70–79, Jan. 2007.
- [15] T. Kim, I. Lee, and T.-W. Lee. Independent vector analysis: Definition and algorithms. In *Proc. 40th Asilomar Conf. Signals, Systems, Comput.*, pages 1393–1396, 2006.
- [16] 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*, 17(5):1265–1277, 2006.
- [17] X.-L. Li and T. Adali. 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.
- [18] X.-L. Li and T. Adali. Independent component analysis by entropy bound minimization. *IEEE Trans. Signal Processing*, 58(10):5151–5164, Oct. 2010.
- [19] Y.-O. Li, T. Adali, W. Wang, and V. D. Calhoun. Joint blind source separation by multi-set canonical correlation analysis. *IEEE Trans. Signal Processing*, 57(10):3918–3929, Oct. 2009.
- [20] E. Ollila, K. Hyon-Jung, and V. Koivunen. Compact Cramér-Rao bound expression for independent component analysis. *IEEE Trans. Signal Processing*, 56(4):1421–1428, Apr. 2008.
- [21] D. Pham and P. Garat. Blind separation of mixtures of independent sources through a quasi maximum likelihood approach. *IEEE Trans. Signal Processing*, 45(7):1712–1725, 1997.
- [22] A. Yeredor. Blind separation of Gaussian sources with general covariance structures: Bounds and optimal estimation. *IEEE Trans. Signal Processing*, 58(10):5057–5068, Oct. 2010.