INDEPENDENT VECTOR ANALYSIS, THE KOTZ DISTRIBUTION, AND PERFORMANCE BOUNDS

Matthew Anderson, Geng-Shen Fu, Ronald Phlypo, and Tülay Adalı

University of Maryland Baltimore County, Baltimore, MD 21250

ABSTRACT

The recent extensions of independent component analysis (ICA) to exploit source dependence across multiple datasets, termed independent vector analysis (IVA), have thus far only considered two multivariate source distribution models: the Gaussian and a second-order uncorrelated Laplacian distribution. In this paper, we introduce the use of the Kotz distribution family as a more flexible source distribution model which exploits both second and higher-order statistics. The Cramér-Rao lower bound (CRLB) for IVA performance prediction is shown to be analogous to the bound for blind source separation (BSS). Lastly, we provide an analytic expression for the CRLB when the sources follow the multivariate power exponential (MPE) subclass of distributions within the Kotz family.

***Index Terms*—** Independent vector analysis (IVA), joint blind source separation (JBSS), Kotz distribution, multivariate power exponential (MPE) distribution, multivariate generalized Gaussian distribution, Cramér-Rao lower bound (CRLB)

1. INTRODUCTION

Blind source separation (BSS) problems have been well studied and various algorithms have been developed and successfully applied in a vast array of applications [1, 2]. A generalization of the BSS problem to multiple datasets, termed joint blind source separation (JBSS), has been a more recent development. The recent interest in JBSS is motivated by various application domains such as when analyzing multisubject datasets in biomedical studies using functional magnetic resonance imaging or electroencephalography data, or when solving the convolutive independent component analysis (ICA) problem in the frequency domain using multiple frequency bins. A particular approach to JBSS, IVA with second-order uncorrelated multivariate Laplace distribution model (IVA-Lap), utilizes higher-order (greater than second-order) dependencies, but it does not exploit the linear dependencies expressed in the second-order statistics [3, 4]. These linear dependencies are explicitly exploited in [5] for defining IVA with multivariate Gaussian distribution model (IVA-Gauss) algorithms. Here, we introduce the Kotz family of distributions as a source distribution model. The Kotz family is shown to include the two source distributions used in already existing independent vector analysis (IVA) algorithms [3, 5]. The use of the Kotz distribution prior provides the ability to exploit both second and higher-order statistics jointly within one algorithm. Furthermore, we show the Cramér-Rao lower bound (CRLB) for IVA. The form of the CRLB for IVA is a generalization of the ICA CRLB given in [6, 7]. We provide a closed form expression for the CRLB when the sources come from a subclass of the Kotz distribution family called the multivariate power exponential (MPE).

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

2. JBSS & IVA PROBLEM FORMULATION

We begin by formulating the JBSS problem. There are K datasets, each containing V samples, formed from the linear mixture of N independent sources, $\mathbf{X}^{[k]} = \mathbf{A}^{[k]}\mathbf{S}^{[k]} \in \mathbb{R}^{N\times V}, 1 \le k \le K$. The entry in nth row and vth column of $\mathbf{S}^{[k]}$ is $s_n^{[k]}(v)$, the nth row of $\mathbf{S}^{[k]}$ is denoted by the column vector $\mathbf{s}_n^{[k]} = \left[s_n^{[k]}(1),\ldots,s_n^{[k]}(V)\right]^\mathsf{T} \in \mathbb{R}^V$, and the vth column of $\mathbf{S}^{[k]}$ is denoted by the column vector $\mathbf{s}^{[k]}(v) = \left[s_1^{[k]}(v),\ldots,s_N^{[k]}(v)\right]^\mathsf{T} \in \mathbb{R}^N$, where superscript T denotes transpose. The invertible mixing matrices, $\mathbf{A}^{[k]} \in \mathbb{R}^{N\times N}$, are unknown real-valued quantities to be estimated. The mixing matrices are not necessarily related.

The source vectors in each dataset can be concatenated to form $\mathbf{S} = \left[\left(\mathbf{S}^{[1]}\right)^\mathsf{T},\ldots,\left(\mathbf{S}^{[K]}\right)^\mathsf{T}\right]^\mathsf{T} \in \mathbb{R}^{NK\times V}$. Using this notation, we can denote the JBSS data model with a single equation, namely $\mathbf{X} = \mathbf{AS}$, where $\mathbf{A}$ is a block diagonal matrix or $\mathbf{A} = \oplus_{k=1}^K \mathbf{A}^{[k]}$. The nth source component matrix (SCM), $\mathbf{S}_n = \left[\mathbf{s}_n^{[1]},\ldots,\mathbf{s}_n^{[K]}\right]^\mathsf{T} \in \mathbb{R}^{K\times V}$, is independent of all other SCMs. Then the probability distribution function (pdf) of the concatenated source vector, $\mathbf{S}$, can be written as $p(\mathbf{S}) = \prod_{n=1}^N p_n(\mathbf{S}_n)$. A special case of the JBSS formulation, termed IVA, occurs when the V samples are independently and identically distributed (iid), so that a random vector, which we term as the source component vector (SCV), can be defined from a column of the SCM, i.e., $\mathbf{s}_n =$

$\left[\mathbf{s}_n^{[1]}, \ldots, \mathbf{s}_n^{[K]}\right]^\mathsf{T} \in \mathbb{R}^K$. This implies that for IVA it assumed that $p(\mathbf{S}) = \prod_{n=1}^N p_n(\mathbf{S}_n) = \prod_{n=1}^N \prod_{v=1}^V p_n(\mathbf{s}_n(v))$.

The JBSS solution finds K demixing matrices and the corresponding source estimates for each dataset, with the kth ones denoted as $\mathbf{W}^{[k]}$ and $\mathbf{Y}^{[k]} = \mathbf{W}^{[k]}\mathbf{X}^{[k]}$, respectively. The estimate of the nth component from the vth sample of the kth dataset is given by $y_n^{[k]}(v) = \left(\mathbf{w}_n^{[k]}\right)^\mathsf{T} \mathbf{x}^{[k]}(v)$, where $\left(\mathbf{w}_n^{[k]}\right)^\mathsf{T}$ is the nth row of $\mathbf{W}^{[k]}$. The estimate of the nth SCM is given as $\mathbf{Y}_n = \left[\mathbf{y}_n^{[1]}, \ldots, \mathbf{y}_n^{[K]}\right]^\mathsf{T}$.

3. IVA SOLUTION

The goal of IVA, the identification of the independent SCVs, can be achieved by minimizing the mutual information among the estimated source component vectors or equivalently by maximizing the likelihood function for the given observations $\mathbf{X}$. The mutual information for the SCVs is given as [3, 5]:

$$\mathcal{I}\{\mathbf{y}_1; \ldots; \mathbf{y}_N\} = \sum_{n=1}^N \mathcal{H}\{\mathbf{y}_n\} - \sum_{k=1}^K \log\left|\det \mathbf{W}^{[k]}\right| - \mathcal{H}\{\mathbf{x}\},$$

where we use $\mathcal{H}\{\mathbf{x}\} = -E\{\log p_\mathbf{x}(\mathbf{x})\}$ for Shannon (differential) entropy of the random vector $\mathbf{x}$. The entropy of the *mixture* data is inconsequential and thus discarded to define the IVA cost function:

$$\mathcal{J}(\mathbf{W}) \triangleq \sum_{n=1}^N \mathcal{H}\{\mathbf{y}_n\} - \sum_{k=1}^K \log\left|\det \mathbf{W}^{[k]}\right|. \quad (1)$$

The minimization of (1) can be performed using the general methods described in [4] and [5]. We utilize the *decoupled* optimization approach of the latter in this paper, which we summarize next. The benefits of the decoupled optimization approach includes an ability to tailor step-sizes and aids development of (quasi-) Newton algorithms.

Following [8], we define $\mathbf{h}_n^{[k]}$ to be a unit length vector such that $\widetilde{\mathbf{W}}_n^{[k]}\mathbf{h}_n^{[k]} = \mathbf{0}$, where $\widetilde{\mathbf{W}}_n^{[k]}$ is the $(N-1)\times N$ matrix formed by removing the nth row of the demixing matrix $\mathbf{W}^{[k]}$. Then, it can be shown that [9]:

$$\left|\det\left(\mathbf{W}^{[k]}\right)\right| = \left|\left(\mathbf{h}_n^{[k]}\right)^\mathsf{T} \mathbf{w}_n^{[k]}\right| \varpi_n^{[k]}, \quad (2)$$

where $\left(\varpi_n^{[k]}\right)^2 = \left|\det\left(\widetilde{\mathbf{W}}_n^{[k]}\left(\widetilde{\mathbf{W}}_n^{[k]}\right)^\mathsf{T}\right)\right|$. Clearly, $\varpi_n^{[k]}$ is invariant with respect to (wrt) $\mathbf{w}_n^{[k]}$. An efficient recursive method for computing $\mathbf{h}_n^{[k]}$ is given in [10].

Then, by substituting (2) into (1) as in [11], we have,

$$\begin{aligned}\mathcal{J}(\mathbf{W}) &= \sum_{m=1}^N \mathcal{H}\{\mathbf{y}_m\} - \sum_{l=1}^K \log\left|\left(\mathbf{h}_n^{[l]}\right)^\mathsf{T} \mathbf{w}_n^{[l]}\right| + \log \varpi_n^{[l]} \\ &= \mathcal{H}\{\mathbf{y}_n\} - \log\left|\left(\mathbf{h}_n^{[k]}\right)^\mathsf{T} \mathbf{w}_n^{[k]}\right| - C_n^{[k]}, \quad (3)\end{aligned}$$

where we note that $\mathcal{H}\{\mathbf{y}_m\}$ is independent of $\mathbf{w}_n^{[k]}$ for $m \neq n$ and we let $C_n^{[k]}$ be the new quantity containing all the terms that are invariant wrt $\mathbf{w}_n^{[k]}$. Then, the IVA cost function derivative wrt $\mathbf{w}_n^{[k]}$ is

$$\frac{\partial \mathcal{J}(\mathbf{W})}{\partial \mathbf{w}_n^{[k]}} = E\left\{\phi_n^{[k]}(\mathbf{y}_n)\mathbf{x}^{[k]}\right\} - \frac{\mathbf{h}_n^{[k]}}{\left(\mathbf{h}_n^{[k]}\right)^\mathsf{T} \mathbf{w}_n^{[k]}}, \quad (4)$$

where $\phi_n^{[k]}(\mathbf{y}_n)$ is the kth element of the multivariate score function $\boldsymbol{\phi}_n(\mathbf{y}_n) \triangleq -\partial \log p_n(\mathbf{y}_n)/\partial \mathbf{y}_n$.

The vector derivative of the likelihood function is used to iteratively update each source demixing row and results in nonorthogonal demixing matrices by using a gradient update rule followed by a renormalization of each demixing vector.

4. IVA PERFORMANCE BOUNDS

Here we provide the induced Cramér-Rao lower bound (iCRLB) for the estimation of the global demixing-mixing matrices, $\mathbf{G}^{[k]} \triangleq \mathbf{W}^{[k]}\mathbf{A}^{[k]}$. The calculation of the iCRLB uses the Hessian of the IVA cost function evaluated at the global minimum, which for the iid assumption is equivalent to the Fisher information matrix associated with estimating $\mathbf{G}$. Due to space constraints, we directly present the final form (computation is a generalization of the derivation in [5]).

The Hessian possesses a tractable structure for analysis with a form analogous that is block matrix expansion of the Hessian for ICA. Namely, the Hessian can be permuted into a block diagonal matrix where the first N block entries are of dimensions $K \times K$ and are given by $\mathbf{J}_n$. For this paper, the $\mathbf{J}_n$ matrices are not of interest. The remaining block entries are of dimensions $2K \times 2K$ and are given by

$$\mathbf{J}_{m,n} \triangleq \begin{bmatrix} \boldsymbol{\mathcal{K}}_{m,n} & \mathbf{I} \\ \mathbf{I} & \boldsymbol{\mathcal{K}}_{n,m} \end{bmatrix}, \ 1 \leq m < n \leq N, \quad (5)$$

where $\mathbf{I} \in \mathbb{R}^{K\times K}$ is the identity matrix,

$$\boldsymbol{\mathcal{K}}_{m,n} \triangleq \boldsymbol{\Gamma}_m \circ \mathbf{R}_n, \ 1 \leq m \neq n \leq N,$$

$\boldsymbol{\Gamma}_m \triangleq E\left\{\boldsymbol{\phi}_m(\mathbf{y}_m)\boldsymbol{\phi}_m^\mathsf{T}(\mathbf{y}_m)\right\}$, and $\circ$ denotes the Hadamard product. The form of the Hessian affirms that the separation performance can be analyzed by considering sources pairwise. Additionally, for the Gaussian distribution $\boldsymbol{\mathcal{K}}_{m,n} =$

$\mathbf{R}_m^{-1} \circ \mathbf{R}_n$, since $\boldsymbol{\Gamma}_m = E\left\{\mathbf{R}_m^{-1}\mathbf{y}_m\mathbf{y}_m^{\mathsf{T}}\mathbf{R}_m^{-1}\right\} = \mathbf{R}_m^{-1}$. In general, computing analytic expressions for $\boldsymbol{\Gamma}_m$ is nontrivial.

A common measure of BSS algorithm performance is the interference to source ratio (ISR). It assesses the amount of residual energy from each source that contributes to the energy of each *estimated* source. ISR is a convenient measure because it is indifferent to the inherent scaling ambiguity in BSS and it is an analytically tractable performance measure that readily admits to derivation of the iCRLB (on ISR) for a wide variety of BSS problems [5, 6, 7, 12]. For the general form we can show that the iCRLB for an element of the ISR matrix in IVA has the following form

$$\mathrm{ISR}_{m,n\neq m} \geq \tfrac{1}{V}\mathrm{tr}\left(\left(\boldsymbol{\mathcal{K}}_{m,n} - \boldsymbol{\mathcal{K}}_{n,m}^{-1}\right)^{-1}\right), \qquad (6)$$

where $\mathrm{tr}(\cdot)$ is the trace operator,

$$\mathrm{ISR}_{m,n} \triangleq \sum_{k=1}^{K} E\left\{\left(g_{m,n}^{[k]}\right)^2\right\},\ \ 1 \leq m \neq n \leq N,$$

gives the total energy of the nth source in the estimate of the mth source in each dataset when the true sources are normalized to have unit variance, and $g_{m,n}^{[k]} = \left\{\mathbf{G}^{[k]}\right\}_{m,n}$ is the entry in the mth row and nth column of $\mathbf{G}^{[k]}$. Notice that (6) extends the ICA bounds given in [6, 7]. As shown in [13], multiple datasets add a diversity in IVA that mirrors the role of sample-to-sample dependence that can be taken into account in ICA besides higher-order statistics [2, 12, 14].

5. IVA USING KOTZ DISTRIBUTION

The two existing IVA implementations use two different source distribution models. The first IVA algorithm [3], IVA-Lap, uses the multivariate score function given by $\boldsymbol{\phi}_n(\mathbf{y}_n) = \mathbf{y}_n/\sqrt{\mathbf{y}_n^{\mathsf{T}}\mathbf{y}_n}$. The second IVA algorithm [5], IVA-Gauss, uses the multivariate score function given by $\boldsymbol{\phi}_n(\mathbf{y}_n) = \mathbf{R}_n^{-1}\mathbf{y}_n$, where $\mathbf{R}_n$ is the covariance matrix associated with the actual SCV, i.e., $\mathbf{R}_n = E\left\{\mathbf{s}_n\mathbf{s}_n^{\mathsf{T}}\right\}$. This SCV model requires that the SCV covariance matrix is estimated as part of the IVA optimization procedure – note that the score function used in IVA-Lap requires no distribution parameters. In this section, we introduce the Kotz distribution family to greatly expand the available multivariate SCV distribution models for use in IVA.

The original introduction of the Kotz distribution [15] has been further analyzed in the more readily accessible work of [16]. The zero-mean K-dimensional real-valued Kotz distribution has the following pdf

$$p(\mathbf{x};\boldsymbol{\theta},\boldsymbol{\Sigma}) = \frac{\beta\lambda^{\nu}\Gamma(K/2)\left(\mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x}\right)^{\eta-1}}{\sqrt{\pi^K}\det\boldsymbol{\Sigma}\,\Gamma(\nu)}e^{-\lambda\left(\mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x}\right)^{\beta}},$$

where $\boldsymbol{\Sigma} \in \mathbb{R}^{K\times K}$ (dispersion matrix) is positive-definite, $\boldsymbol{\theta} = [\beta,\eta,\lambda]^{\mathsf{T}}$ denotes the scalar Kotz distribution parameters, namely, $\lambda > 0$ (kurtosis parameter), $\beta > 0$ (shape parameter), $\eta > (2-K)/2$ (hole parameter), , and we use $\nu \triangleq (2\eta + K - 2)/(2\beta) > 0$ for more compact notations. The gamma function is denoted by $\Gamma(\cdot)$. The covariance matrix for the Kotz distribution is given in [16],

$$\mathbf{R} \triangleq E\left\{\mathbf{x}\mathbf{x}^{\mathsf{T}}\right\} = \frac{\lambda^{-\beta^{-1}}}{K}\frac{\Gamma\left(\nu+\beta^{-1}\right)}{\Gamma(\nu)}\boldsymbol{\Sigma}. \qquad (7)$$

The set of distributions achieved by varying the Kotz parameters is vast and includes the MPE, which is sometimes called a multivariate generalized Gaussian distribution, when $\eta = 1$ and $\lambda = 1/2$. The hole parameter, η, can be used to make the mode of the distribution occur at $\mathbf{x} \neq \mathbf{0}$ when $\eta > 1$.

The multivariate score function associated with the Kotz distribution is,

$$\boldsymbol{\phi}(\mathbf{x}) = 2\left(1 - \eta + \lambda\beta\left(\mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x}\right)^{\beta}\right)\frac{\boldsymbol{\Sigma}^{-1}\mathbf{x}}{\mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x}}. \qquad (8)$$

By appropriately selecting the Kotz parameters we arrive at the score function used in IVA-Gauss: $\boldsymbol{\phi}(\mathbf{x}) = \mathbf{R}^{-1}\mathbf{x}$ when $\boldsymbol{\theta} = [1,1,1/2]^{\mathsf{T}}$ and $\boldsymbol{\Sigma} = \mathbf{R}$; and in IVA-Lap: $\boldsymbol{\phi}(\mathbf{x}) = \mathbf{x}/\sqrt{\mathbf{x}^{\mathsf{T}}\mathbf{x}}$ when $\boldsymbol{\theta} = [1/2,1,1/2]^{\mathsf{T}}$ and $\boldsymbol{\Sigma} = \mathbf{I}$.

The multivariate Kotz distribution possesses several desirable properties for IVA. It is a distribution which generalizes both IVA-Lap and IVA-Gauss, the score function is readily calculated, and it extends the set of sources for which JBSS can be performed. Thus, the Kotz distribution is appealing despite it requiring additional parameters $\boldsymbol{\theta}$ to estimate. The precise estimation of these parameters is difficult at best, yet fortunately for IVA, precise knowledge of these parameters is not critical for successful JBSS (as is similarly the case in ICA). One practical approach is to select the $\boldsymbol{\theta}$ from set of $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_P]$ that achieves the lowest cost. This approach is naturally accommodated in the decoupled vector optimization framework. Lastly, the second-order correlation information captured by the dispersion matrix, $\boldsymbol{\Sigma}$, can be estimated suboptimally via substitution of the sample average estimate for the covariance matrix, $\hat{\mathbf{R}} = 1/V\sum_{v=1}^{V}\mathbf{x}(v)\mathbf{x}^{\mathsf{T}}(v)$ into (7).[1]

5.1. Multivariate Power Exponential

The MPE distribution, a subset of the Kotz distribution family, is a general class in and of itself. Here, in this subsection, we provide analytic results for the MPE source model. It can be shown that for the MPE distributions, $\boldsymbol{\Gamma}_m = \kappa_m\mathbf{R}_m^{-1}$, where $\kappa_m = \kappa_{\mathrm{MPE}} \triangleq \left(\frac{2\beta}{K\Gamma\left(\frac{K}{2\beta}\right)}\right)^2\Gamma\left(\frac{K-2+4\beta}{2\beta}\right)\Gamma\left(\frac{K+2}{2\beta}\right)$, $\beta > 0$ and $K \geq 2$. Furthermore, it also can be shown that $\kappa_{\mathrm{MPE}} \geq 1$ with equality only if the SCV is Gaussian distributed, i.e., $\beta = 1$. The proof of this property is provided in the appendix.

[1] The code for implementing such an algorithm is available at http://mlsp.umbc.edu/resources.html.

When two sources follow MPE distributions, we have

$$\text{ISR}_{m,n} \geq \tfrac{1}{V}\text{tr}\left(\left(\kappa_m \mathbf{R}_m^{-1} \circ \mathbf{R}_n - \kappa_n^{-1}\left(\mathbf{R}_n^{-1} \circ \mathbf{R}_m\right)^{-1}\right)^{-1}\right).$$

A case of particular interest is when both sources have identity covariance matrices, then the iCRLB becomes simply $V^{-1}\left(\kappa_m - \kappa_n^{-1}\right)^{-1}$, which is finite as long as $\kappa_m \neq \kappa_n^{-1}$ and thus when $(\kappa_m, \kappa_n) \neq (1, 1)$ holds, i.e., as long as one source is non-Gaussian. This result is consistent with the nonidentifiability conditions provided in [5].

6. SIMULATIONS

In this section, we consider the performance of the algorithm presented in Section 5 using simulated datasets. The performance of the proposed IVA algorithm is compared with the iCRLB derived in Section 4.

For this experiment there are $N = 3$ SCVs, each generated from the same source distribution, namely, a zero-mean $K = 5$ dimensional MPE random vector with shape parameter, β, and the identity covariance matrix. The kth entry of each SCV is used as a latent source for the kth dataset. Entries of the random mixing matrices, $\mathbf{A}^{[k]}$, are from the standard normal distribution.

We compute the theoretical iCRLB for ISR and compare this value with the ISR achieved using the Kotz algorithm with a priori knowledge about the shape, kurtosis, and hole parameters. We then compute the total theoretical normalized ISR, defined as,

$$\text{ISR} \triangleq \sum_{m=1, n=1, m\neq n}^{N} V\, \text{ISR}_{m,n}.$$

We compare this theoretical ISR with the average ISR computed from 1000 independent trials of the algorithm as we vary the number of samples per dataset, V.

Due to the presence of local minima in the IVA cost function for non-Gaussian sources [5], the algorithm may converge to local minima. At local minima the sources are separated within a dataset but the SCVs are not successfully identified, i.e., the permutation ambiguity is unresolved. Thus we compare the iCRLB for the ISR with the median rather than the mean. From Fig.1, the performance of the IVA algorithm approaches the iCRLB as the sample size per dataset increases. The large degradation in achieved ISR at $\beta = 6$ indicates a sensitivity for nearly *uniform* multivariate sources when the sample size is small.

7. CONCLUSIONS

We have introduced a new family of source distributions for use in IVA, namely the Kotz distribution family, which includes the distributions used in existing IVA algorithms. The formulation for calculating the performance bounds of JBSS using iid samples is given for the general case. In addition, for the MPE family of distributions we provide the analytic expression for the performance bounds.

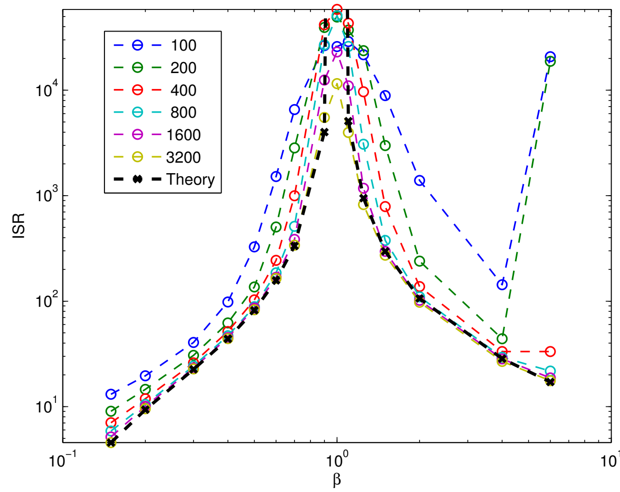

Fig. 1. The iCRLB theory for ISR as the shape parameter, β, varies is compared to the median ISR of 1000 trials for different numbers of iid samples, V.

A. APPENDIX

We make use of an inequality first shown in [17] involving Gurland's ratio,

$$T(u, v) \triangleq \frac{\Gamma(u)\Gamma(v)}{\Gamma^2((u+v)/2)}, \quad u, v > 0,$$

namely $T(u-\gamma, u+\gamma) \geq 1 + \frac{\gamma^2}{u-\gamma}, \quad u > |\gamma|$, where the equality holds only when $\gamma = 1$, [18].

For $K > 2$, let $z = K/2 - 1 > 0$, then

$$\begin{aligned}\kappa_{\text{MPE}} &= \frac{(\nu - 1/\beta)(\nu - 1/\beta + 1)}{\nu^2} T(\nu - 1/\beta, \nu + 1/\beta) \\ &\geq \frac{(\nu - 1/\beta)(\nu - 1/\beta + 1)}{\nu^2}\left(1 + \frac{\beta^{-2}}{\nu - 1/\beta}\right) \\ &= 1 + \frac{z}{(z+1)^2}\frac{(\beta-1)^2}{\beta}.\end{aligned}$$

Since $\beta > 0$, then $(\beta - 1)^2 \beta^{-1} \geq 0$ and thus $\kappa_{\text{MPE}} \geq 1$.

For $K = 2$, we have that

$$\frac{d\kappa_{\text{MPE}}}{d\beta} = \frac{2\Gamma(2\nu)}{\Gamma^2(\nu)}\left(\Psi(\nu+1) - \Psi(2\nu)\right),$$

where $\Psi(x) \triangleq d \ln \Gamma(x)/dx$ is the digamma function. Noting that the derivative is positive (negative) when $\Psi(\nu+1) > (<)\, \Psi(2\nu)$. Since $\Psi(x)$ is a monotonic nondecreasing function of x [19], then we must have that the arguments have the same relationship, i.e., $\beta \geq (<)\, 1$ implies $d\kappa_{\text{MPE}}/d\beta \geq (<)\, 0$ with equality only when $\beta = 1$.

8. REFERENCES

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Wiley-Interscience, 2001.

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

[3] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: an extension of ICA to multivariate components," in *Independent Component Analysis and Blind Signal Separation*, ser. Lecture Notes in Comput. Sci. Springer Berlin / Heidelberg, 2006, vol. 3889, pp. 165–172.

[4] 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.

[5] M. Anderson, T. Adalı, and X.-L. Li, "Joint blind source separation of multivariate Gaussian sources: Algorithms and performance analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1672–1683, Apr. 2012.

[6] P. Tichavský, Z. Koldovský, and E. Oja, "Performance analysis of the FastICA algorithm and Cramér-Rao bounds for linear independent component analysis," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1189–1203, Apr. 2006.

[7] E. Ollila, K. Hyon-Jung, and V. Koivunen, "Compact Cramér-Rao bound expression for independent component analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1421–1428, Apr. 2008.

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

[9] M. Anderson, X.-L. Li, P. 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 Process. (ICASSP)*, Mar. 2012, pp. 1885–1888.

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

[11] M. Anderson, X.-L. Li, and T. Adalı, "Nonorthogonal independent vector analysis using multivariate Gaussian model," in *Latent Variable Analysis and Signal Separation*, ser. Lecture Notes in Comput. Sci. Springer Berlin / Heidelberg, 2010, vol. 6365, pp. 354–361.

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

[13] T. Adalı, M. Anderson, and G. Fu, "IVA and ICA: Use of diversity in independent decompositions," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, Aug. 2012, pp. 61–65.

[14] G.-S. Fu, M. Anderson, R. Phlypo, and T. Adalı, "Algorithms for Markovian source separation by entropy rate minimization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2013, accepted.

[15] S. Kotz, "Multivariate distributions at a cross-road," *Statistical Distributions in Scientific Work*, 1975.

[16] S. Nadarajah, "The Kotz-type distribution with applications," *Statistics*, pp. 341–358, 2003.

[17] J. Gurland, "An inequality satisfied by the gamma function," *Scandinavian Actuarial Journal*, vol. 29, pp. 171–172, 1956.

[18] M. Merkle, "Gurland's ratio for the gamma function," *Computers & Mathematics with Applications*, vol. 49, no. 2–3, pp. 389–406, 2005.

[19] S. S. Dragomir, R. P. Agarwal, and N. S. Barnett, "Inequalities for beta and gamma functions via some classical and new integral inequalities," *Journal of Inequalities and Applications*, vol. 5, pp. 103–165, 2000.