Publication III

Esa Ollila, Hyon-Jung Kim, and Visa Koivunen. 2008. Compact Cramér–Rao bound expression for independent component analysis. IEEE Transactions on Signal Processing, volume 56, number 4, pages 1421-1428.

Compact Cramér–Rao Bound Expression for Independent Component Analysis

Esa Ollila, *Member, IEEE*, Hyon-Jung Kim, and Visa Koivunen, *Senior Member, IEEE*

***Abstract*—Despite of the increased interest in independent component analysis (ICA) during the past two decades, a simple closed form expression of the Cramér–Rao bound (CRB) for the demixing matrix estimation has not been established in the open literature. In the present paper we fill this gap by deriving a simple closed-form expression for the CRB of the demixing matrix directly from its definition. A simulation study comparing ICA estimators with the CRB is given.**

***Index Terms*—Cramér–Rao lower bound, efficient estimator, FastICA, Fisher information, independent component analysis (ICA).**

I. Introduction

INDEPENDENT component analysis (ICA) is a relatively recent (see [1], [2]) technique of multivariate data analysis with the purpose of extracting unobserved source signals or *independent components* (ICs) from their observed linear *mixtures*. In (real-valued) linear instantaneous ICA model the observed random vector $\mathbf{x} = (x_1, \dots, x_d)^T$ of mixtures is generated by

$$\mathbf{x} = A\mathbf{s} \tag{1}$$

where $A = (\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_d)$ is unknown $d \times d$ *mixing matrix* of full rank and $\mathbf{s} = (s_1, \dots, s_d)^T$ is the unobserved random vector of ICs, i.e., the source signals. The goal is then to estimate, based on the i.i.d. sample $\mathbf{x}_1, \dots, \mathbf{x}_n$ from (1), the *demixing matrix* $W = (\mathbf{w}_1 \quad \cdots \quad \mathbf{w}_d)^T = A^{-1}$ which, subsequently, allows the estimation of the source vectors that generated the data by $\hat{\mathbf{s}}_1 = \hat{W}\mathbf{x}_1, \dots, \hat{\mathbf{s}}_n = \hat{W}\mathbf{x}_n$. At this point, neglect the scaling, sign and permutation ambiguity [1], [3] in the estimation of the *demixing vectors* $\mathbf{w}_1, \dots, \mathbf{w}_d$ (row vectors of W). These issues are addressed later in the paper. Several estimation methods have been proposed to solve the above problem, for instance FastICA and JADE (see [2] and [4] for reviews).

It is highly useful to have a lower bound for the statistical variability (accuracy) of an estimator. Cramér–Rao bound (CRB) provides a lower bound on the covariance matrix of any unbiased estimator of a parameter vector. CRB, which is the inverse of the Fisher information matrix (FIM), can be used e.g., to show that an unbiased estimator is uniformly minimum variance unbiased (UMVU) estimator. CRB is also related to asymptotic optimality theory.

Despite of the increased interest in ICA during the past two decades, a closed-form expression for the CRB for the demixing matrix estimation has not been established in the open literature. CRB is derived indirectly in [5]–[9] via asymptotic approximations of the likelihood or via asymptotic covariance matrix of the maximum-likelihood (ML) estimator of a transformed parameters such as the interference-to-signal ratio. In the present paper we fill this gap by deriving a simple, *compact closed-form expression* for the CRB of the vectorized parameter $\boldsymbol{\theta} = \mathrm{vec}(W^T)$ directly from its definition; see Theorems 1 and 2 of the present paper.

Remarkably, the CRB depends on the distribution of s_i's only through two scalars defined in (6) and (7) that are rather easy to calculate. This is in agreement with the earlier (asymptotic) results derived in [8]. CRB thus provides an easily computable performance criterion for ICA. Simple expressions for the $\mathbf{w}_i$-blocks of the inverse of FIM are derived, which, in turn provide the CRB for estimation of the demixing vectors $\mathbf{w}_i$. This is a useful e.g., as many ICA methods, such as the 1-unit FastICA, employ deflation approach, i.e., they do not estimate the demixing matrix W as a whole but a single demixing vector $\mathbf{w}_i$ (one by one, if wanted). In this paper we use different approach than earlier papers by exploiting matrix results e.g., involving Kronecker product, vec-operator and commutation matrix that enable the derivation of the inverse of FIM into a simple closed form. Two recent studies on the CRB can also be found from [10], [11]. Also in these papers, a compact closed-form expression for the CRB of the demixing matrix was not explicitly derived.

II. CRLB for ICA

Suppose i.i.d. observations $\mathbf{x}_1, \dots, \mathbf{x}_n$ are distributed as $\mathbf{x}$ having the pdf $f_{\boldsymbol{\theta}}(\mathbf{x})$ with parameter vector $\boldsymbol{\theta} \in \Theta$. The inverse of the FIM of $\boldsymbol{\theta}$

$$\mathcal{I}_{\boldsymbol{\theta}} = E[\nabla_{\boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(\mathbf{x}) \{\nabla_{\boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(\mathbf{x})\}^T] \tag{2}$$

gives, under regularity conditions[1], the CRB on the covariance matrix of an unbiased estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ in the sense that

$$\mathrm{cov}(\hat{\boldsymbol{\theta}}) \geq n^{-1}\mathcal{I}_{\boldsymbol{\theta}}^{-1}. \tag{3}$$

Above, for symmetric matrices B and C, the notation "$B \geq C$" implies that $B - C$ is positive semidefinite. The CRB (3)

Manuscript received November 17, 2006; revised August 27, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Erik G. Larsson.

E. Ollila and V. Koivunen are with the Signal Processing Laboratory, Helsinki University of Technology, FIN-02015 HUT, Finland (e-mail: esollila@wooster.hut.fi; visa@wooster.hut.fi).

H-J. Kim is with the Department of Mathematics and Statistics, University of Oulu, FIN-90014 University of Oulu, Finland (e-mail: hyon-jung.kim@oulu.fi).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2007.910484

[1]The regularity conditions are required for the interchange of certain differentiation and integration operators (see [12] for details)

thus implies, for example, that $\mathrm{var}(\hat{\theta}_i) \geq n^{-1}(\mathcal{I}_{\boldsymbol{\theta}}^{-1})_{ii}$, where $\hat{\theta}_i$ denotes the ith component of $\hat{\boldsymbol{\theta}}$ and $(\mathcal{I}_{\boldsymbol{\theta}}^{-1})_{ii}$ the (i,i)th element of $\mathcal{I}_{\boldsymbol{\theta}}^{-1}$. CRB is also related to asymptotic optimality theory in the sense that asymptotic covariance matrix of the ML estimator coincides with $\mathcal{I}_{\boldsymbol{\theta}}^{-1}$. Recall however that there may not exist an unbiased estimator that attains the CRB for all $\boldsymbol{\theta} \in \Theta$.

Next we recall the scaling, sign and permutation ambiquity of the ICA problem: if D is a $d \times d$ diagonal matrix and P is a $d \times d$ permutation matrix, then $\mathbf{x} = (AP^{-1}D^{-1})(DP\mathbf{s})$, where $DP\mathbf{s}$ has independent components as well. Therefore, components of $\mathbf{s}$ can be identified only up to multiplying constants and permutation. Therefore, scales of s_i's can be fixed, e.g., by imposing $\mathrm{var}(s_i) = 1, i = 1, \ldots, d$. This scaling convention is common in ICA and it renders A (respectively, W) unique up to permutation and sign of its columns (respectively, rows).

A. Assumptions

First we form the parameter vector

$$\boldsymbol{\theta} = \mathrm{vec}(W^T) = (\mathbf{w}_1^T, \ldots, \mathbf{w}_d^T)^T \in \mathbb{R}^{d^2} \quad (4)$$

where $\mathbf{w}_i \in \mathbb{R}^d$ are the row vectors of W and the "vec" is the well-known vectorizing operator ([13], p. 30), namely, if B is $n \times m$ matrix, then $\mathrm{vec}(B)$ is a nm-dimensional vector formed by stacking the column vectors of the matrix B on top of each other. The pdf of $\mathbf{x} = A\mathbf{s}$ is $f_{\boldsymbol{\theta}}(\mathbf{x}) = |\det(W)| \prod_{i=1}^d f_i(\mathbf{w}_i{}^T\mathbf{x})$, where f_i denotes the pdf of s_i. Use of matrix derivatives gives

$$\frac{\partial}{\partial W^T} \log f_{\boldsymbol{\theta}}(\mathbf{x}) = A - \mathbf{x}\,\varphi(W\mathbf{x})^T$$

where $\varphi(\mathbf{s}) = (\varphi_1(s_1), \ldots, \varphi_d(s_d))^T$ and $\varphi_i(s) = -f_i'(s)/f_i(s)$ is the *location score function* of the ith IC. The *Fisher score* of the parameter (4) in the ICA model can now be calculated by

$$\nabla_{\boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(\mathbf{x}) = \mathrm{vec}\left\{ \frac{\partial}{\partial W^T} \log f_{\boldsymbol{\theta}}(\mathbf{x}) \right\}. \quad (5)$$

The following assumptions on ith IC s_i for $i = 1, \ldots, d$ are made.

a) s_i has zero mean $E(s_i) = 0$ and unit variance $\mathrm{var}(s_i) = E(s_i^2) = 1$ and only one of the IC's $s_1, \ldots, s_d$ can have a Gaussian distribution.

b) The pdf f_i of s_i satisfy
   - b.1) f_i is continuous with contiguous support, $f_i(s) > 0$ and $f_i'(s) = (d/ds)f_i(s)$ exist $\forall s$ on the support of the density f_i;
   - b.2) $sf_i(s)$ tends to zero as s tends to the boundaries of the support of f_i.

c) The following variances:

$$\kappa_i = \mathrm{var}(\varphi_i(s_i)) = E\left[\varphi_i^2(s_i)\right] = -\int \varphi_i(s) f_i'(s) ds \quad (6)$$

$$\lambda_i = \mathrm{var}(\varphi_i(s_i)s_i) = E\left[\varphi_i^2(s_i)s_i^2\right] - 1 = -\int \varphi_i(s) f_i'(s) s^2\, ds - 1 \quad (7)$$

exist and are finite.

Rather surprisingly, the assumption of finite variance in a) turns out to be crucial for the existence of the FIM. Such a restriction necessarily excludes, for instance, the Cauchy distribution which does not possess finite variance. Due to indeterminacy of the scales of the s_i's, we have assumed in a), without any loss of generality, that IC's have unit variance. The mean of s_i is irrelevant and is, for ease of exposition, assumed to be zero. The necessity of at most one Gaussian component is a necessary restriction in ICA [1].

Assumption b.1) is mainly needed for the existence of the Fisher score (5). Assumption b.2) is not very restrictive and quite reasonable for densities with infinite support. b.2) implicitly implies that $f_i(s)$ tends to zero as s tends to the boundaries of the support of f_i, which subsequently implies that $E[\varphi_i(s_i)] = -\int f_i'(s)ds = 0$. Hence, b.2) may not often be satisfied for densities with finite or semi-finite support. Clearly, e.g., the (zero mean) uniform distribution and the exponential distribution do not satisfy b). Note that the zero mean Laplace distribution satisfies b.2) but it does not satisfy b.1) since it is not differentiable at $s = 0$. Nevertheless, Laplace distribution can be approximated to within arbitrary precision by a valid pdf that does satisfy b). Note that the assumption b) ensures that $E[s_i\varphi_i(s_i)] = 1$ and it is in fact a necessary condition for the Fisher score (5) to satisfy $E[\nabla_{\boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(\mathbf{x})] = \mathbf{0}$ [see Lemma 1a) of Appendix B], which is a basic assumption of CRB theory.

For finiteness of the variances in (6) and (7), the respective integrands in (6) and (7), i.e., $h_i(s) = \varphi_i(s)f_i'(s)$ and $g_i(s) = \varphi_i(s)f_i'(s)s^2 = h_i(s)s^2$ need to decay rapidly enough to zero as s tends to $\pm\infty$ in case of infinite support sources, or, be bounded in case of finite support sources. For example, the zero mean Rayleigh distribution which is commonly used in communications theory satisfies assumptions a) and b), but not c). It can be shown [Lemma 1b) of Appendix B] that $\kappa_i \geq 1$ with equality if and only if s_i is a Gaussian random variable and that $\lambda_i > 0$.

If $f_i''(s)$ (second derivative of f_i) exists at all s, then κ_i can be calculated by

$$\kappa_i = E[\varphi_i'(s_i)]$$

provided that d.1) $f_i'(s) \to 0$ as s tends to the boundaries of the support of f_i. Note that d.1) is satisfied for all infinite support sources. Thus, the assumption d.1) should be checked for distributions with finite or semi-finite support only. Similarly, if we assume that $f_i''(s)$ exists at all s, then

$$\lambda_i = E\left[\varphi_i'(s_i)s_i^2\right] + 1$$

provided that d.2) $f_i'(s)s^2 \to 0$ as s tends to the boundaries of the support of f_i. Note that d.1) implies d.2) if f_i has finite support, but not in the case of infinite or semi-finite support. These alternative formulae [proofs are given in Lemma 1c) of Appendix B] often provide an easier method to calculate the values of κ_i and λ_i.

B. FIM and Its Inverse

We may calculate the FIM (2) using the expression

$$\begin{aligned}\mathcal{I}_{\boldsymbol{\theta}} &= E[\mathrm{vec}\{A(I - \mathbf{s}\varphi(\mathbf{s})^T)\}\mathrm{vec}\{A(I - \mathbf{s}\varphi(\mathbf{s})^T)\}^T] \\ &= (I \otimes A)E[\mathrm{vec}\{I - \mathbf{s}\varphi(\mathbf{s})^T\} \\ &\quad \times \mathrm{vec}\{I - \mathbf{s}\varphi(\mathbf{s})^T\}^T](I \otimes A^T)\end{aligned} \quad (8)$$

where I denotes the identity matrix. Here we applied (5) and algebraic properties involving the vec transformation and the Kronecker product (c.f. Appendix A) and that $\mathbf{x}$ follows ICA model, i.e., $\mathbf{x} = A\mathbf{s}$ and $W\mathbf{x} = \mathbf{s}$.

Next theorem reveals the compact expression of FIM. The proofs of the theorems are given in Appendix C.

Theorem 1: In the ICA model (1) and under Assumptions a)-c), the FIM $\mathcal{I}_{\boldsymbol{\theta}}$ of $\boldsymbol{\theta} = \mathrm{vec}(W^T)$ is a $d^2 \times d^2$ block matrix with (i,j)-block being equal to $d \times d$ matrix:

$$\mathcal{I}_{\boldsymbol{\theta}}[i,j] = \begin{cases} \lambda_i \mathbf{a}_i \mathbf{a}_i^T + \kappa_i \sum_{\substack{l=1 \\ l\neq i}}^{d} \mathbf{a}_l \mathbf{a}_l^T & \text{if } i = j \\ \mathbf{a}_j \mathbf{a}_i^T & \text{if } i \neq j. \end{cases}$$

Remark 1: The whole $d^2 \times d^2$ matrix $\mathcal{I}_{\boldsymbol{\theta}}$ can be constructed using the above $d \times d$ blocks $\mathcal{I}_{\boldsymbol{\theta}}[i,j]$ via the formula (15) in the Appendix A.

Remark 2: The FIM of Theorem 1 is not in agreement with (34) of [10]. This is due to the fact that the pdf $f_X(\mathbf{x}_1,\ldots,\mathbf{x}_n)$ for the i.i.d. sample $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ in (32) of [10] has a superscript n missing from $|\det(W)|$ which subsequently leads to an inaccurate expression for the entries of FIM. To be more specific, in (34) of [10], the first term (containing the term $(n-1)^2$) is wrong. If that term is eliminated, then the element-wise expression (34) of [10] and the block-matrix expression of Theorem 1 are equivalent. Naturally, for $n = 1$, the expressions are equivalent without modifications.

Using Theorem 1, a simple and compact expression for the inverse of the FIM can now be presented.

Theorem 2: In the ICA model (1) and under Assumptions a)-c) and denoting $\boldsymbol{\theta} = \mathrm{vec}(W^T), \mathcal{I}_{\boldsymbol{\theta}}^{-1}$ exists and is a $d^2 \times d^2$ block matrix with (i,j)-block being equal to a $d \times d$ matrix:

$$\mathcal{I}_{\boldsymbol{\theta}}^{-1}[i,j] = \begin{cases} \frac{1}{\lambda_i}\mathbf{w}_i \mathbf{w}_i^T + \sum_{\substack{l=1 \\ l\neq i}}^{d} \frac{\kappa_l}{\kappa_i\kappa_l - 1}\mathbf{w}_l \mathbf{w}_l^T & \text{if } i = j \\ -\frac{1}{\kappa_i\kappa_j - 1}\mathbf{w}_j \mathbf{w}_i^T & \text{if } i \neq j. \end{cases}$$

Note that diagonal blocks $\mathcal{I}_{\boldsymbol{\theta}}^{-1}[i,i]$ give the CRB for an unbiased estimator $\hat{\mathbf{w}}_i$ of the demixing vector $\mathbf{w}_i$:

$$\mathrm{cov}(\hat{\mathbf{w}}_i) \geq \frac{1}{n}\mathcal{I}_{\boldsymbol{\theta}}^{-1}[i,i]$$

for $i = 1,\ldots,d$. Theorem 2 shows that the CRB depends on the distributions of s_i only through the scalars κ_i and λ_i for $i = 1,\ldots,d$. Theorem 2 also implies that only one of s_i's can be Gaussian: if the first and second component, say, are Gaussian, then $\kappa_1 = \kappa_2 = 1$ [Lemma 1 b) in Appendix B] and $\kappa_2/(\kappa_1\kappa_2 - 1)$ is not defined. Still, even in this case, any other block $\mathcal{I}_{\boldsymbol{\theta}}^{-1}[i,i]$ for $i \geq 3$ exists (since the denominators $\kappa_i\kappa_l - 1, i \neq l \in \{1,\ldots,d\}$, do not vanish), indicating that all the remaining rows of W expect the first two can be consistently estimated. That is, the presence of two Gaussian sources does not eliminate the possibility to recover the other sources.

In ICA, the performance of the separation is often investigated via

$$\hat{G} = (\,\hat{\mathbf{g}}_1 \quad \cdots \quad \hat{\mathbf{g}}_d\,)^T = \hat{W}A$$

since the estimated ith source is $\hat{s}_i = \hat{\mathbf{w}}_i^T\mathbf{x} = \hat{\mathbf{g}}_i^T\mathbf{s} = \sum_{j=1}^{d}\hat{g}_{ij}s_j$. Thus, $\hat{g}_{ij}$ and $\mathrm{var}(\hat{g}_{ij}) = E[\hat{g}_{ij}^2]$ for $i \neq j$ represent the magnitude and the average power of interference of jth source in the estimated ith source signal. Since $E[\hat{g}_{ii}] = 1$, the variance $\mathrm{var}(\hat{g}_{ii})$ reflects how accurately the presence of ith source itself is estimated. The CRB for $\hat{\boldsymbol{\vartheta}} = \mathrm{vec}(\hat{G}^T)$ is independent of the parameter W as it is a nonsingular linear transformation of $\hat{\boldsymbol{\theta}} = \mathrm{vec}(\hat{W}^T)$, i.e., $\hat{\boldsymbol{\vartheta}} = (I \otimes A^T)\hat{\boldsymbol{\theta}}$, where $\otimes$ denotes the *Kronecker product*: for any matrix A and B, $A \otimes B$ is a block matrix with (i,j)-block being equal to $a_{ij}B$. Therefore, $\mathrm{cov}(\hat{\boldsymbol{\vartheta}}) = (I \otimes A^T)\mathrm{cov}(\hat{\boldsymbol{\theta}})(I \otimes A)$, which by (3) and (8) indicate that

$$\mathrm{cov}(\hat{\boldsymbol{\vartheta}}) \geq n^{-1}(I \otimes A^T)\mathcal{I}_{\boldsymbol{\theta}}^{-1}(I \otimes A) = n^{-1}\mathcal{I}_I^{-1} \tag{9}$$

where $\mathcal{I}_I$ denotes the value of $\mathcal{I}_{\boldsymbol{\theta}}$ at $\boldsymbol{\theta} = \mathrm{vec}(I)$ (i.e., at $W = I$). Hence, $\mathrm{cov}(\hat{\mathbf{g}}_i) \geq n^{-1}\mathcal{I}_I^{-1}[i,i]$, and Theorem 2 gives the following bounds:

$$\delta_i = \sum_{\substack{j=1 \\ j\neq i}}^{d} \mathrm{var}(\hat{g}_{ij}) \geq n^{-1}\sum_{\substack{j=1 \\ j\neq i}}^{d} \frac{\kappa_j}{\kappa_i\kappa_j - 1} \text{ and}$$

$$\mathrm{var}(\hat{g}_{ii}) \geq n^{-1}\frac{1}{\lambda_i}$$

where δ_i may be interpreted as the average power of interfering source signals to the estimated ith source.

The fact that the CRB for elements of $\hat{G}$ is independent of A is in agreement with the equivariance property [14] shared by many ICA estimators. To be more specific, let $\hat{W} = \hat{W}(X_n)$ be an estimator of W based upon i.i.d. data set $X_n = (\mathbf{x}_1,\ldots,\mathbf{x}_n)$ from the ICA model (1). Thus, the $d \times n$ data matrix X_n can be factored as $X_n = AS_n$, where $S_n = (\mathbf{s}_1,\ldots,\mathbf{s}_n)$ is an i.i.d. data set distributed as $\mathbf{s}$. Equivariant estimator satisfies $\hat{W}(X_n) = \hat{W}(S_n)A^{-1}$ and thus $\hat{G} = \hat{W}(X_n)A = \hat{W}(S_n)$ is independent of A. This property is nicely reflected in the above derived bound (9) for $\hat{G}$. See also [15] for a similar result concerning the induced bound on $\hat{G}$.

III. Simulation Study

The performance of FastICA [16], [17] algorithm is next compared with the CRB via a simulation study. Over the past ten years, the FastICA algorithm has become a benchmark method of ICA due to its simplicity, fast computation and a user-friendly public-domain software.[2] The two variants of FastICA, the symmetric approach and the 1-unit (or deflation) approach and the possibility to choose the nonlinearity, provide a vast selection of FastICA estimators, which can have largely different statistical properties. We compare four different FastICA estimators: both the symmetric and the 1-unit FastICA estimators using nonlinearities "*pow3*" and "*tanh*." These estimators are hereafter referred by obvious acronyms POW3, TANH, 1u-POW3 and 1u-TANH. The nonlinearity *pow3* is the original [16] FastICA algorithm whereas *tanh* is described as a "good general purpose nonlinearity" in [17].

The simulation setup consists of $d = 3$ (zero mean and unit variance) infinite-support symmetric source signals: s_1 having Laplace distribution, s_2 having t_5-distribution and s_3 possessing logistic distribution. $m = 1000$ simulated samples of the source signals were generated using different sample lengths and each

[2]http://www.cis.hut.fi/projects/ica/fastica

sample was mixed by a randomly generated mixing matrix A. Although Laplace density $f_1(s) = (\sqrt{2})^{-1}\exp(-\sqrt{2}|s|)$ do not satisfy assumption b.1) (since it does not have a derivative at zero), it can be approximated to within arbitrary precision by a valid density that does. Moreover, by setting $\varphi_1(s) = \sqrt{2}\text{sign}(s)$, yields $\kappa_1 = 2$ and $\lambda_1 = 1$ for the Laplace-distributed source s_1.

Fig. 1 depicts the calculated *mean-squared error* $\text{MSE}(\hat{\mathbf{g}}_i)$ as a function of signal sample length for the estimated sources $\hat{s}_i = \hat{\mathbf{g}}_i^T\mathbf{x} (i \in \{1,2,3\})$. The MSE is calculated by

$$\begin{aligned}\text{MSE}(\hat{\mathbf{g}}_i) &= \text{Tr}\left\{\frac{1}{m}\sum_{k=1}^{m}\left(\hat{\mathbf{g}}_i^{(k)} - \mathbf{g}_i\right)\left(\hat{\mathbf{g}}_i^{(k)} - \mathbf{g}_i\right)^T\right\} \\ &= \text{MSE}(\hat{g}_{i1}) + \text{MSE}(\hat{g}_{i2}) + \text{MSE}(\hat{g}_{i3})\end{aligned}$$

where $\hat{\mathbf{g}}_i^{(k)} = A^T\hat{\mathbf{w}}_i^{(k)}$ and $\hat{\mathbf{w}}_i^{(k)}$ denotes estimate of $\mathbf{w}_i$ computed from the kth generated sample and $\mathbf{g}_i$ denotes the ith column of the identity matrix. Note that FastICA estimator, explicitly by its definition, constraints the solution for the ICA model with unit variance sources. Hence, the FastICA demixing matrix estimator is by default (without any additional normalization) suitable for comparison with the derived CRB. Nevertheless, it is still possible to solve the demixing matrix only up to permutation/sign-change of the rows. Hence, we need to match the computed value of $\hat{W}$ with the true W: each $\hat{W}^{(k)} = (\mathbf{w}_1^{(k)}\mathbf{w}_2^{(k)}\mathbf{w}_3^{(k)})^T$ computed by TANH or POW3 from the kth simulated sample are multiplied from left by a permutation and sign-change matrix P that produces the smallest value for $\|P\hat{G}^{(k)} - I\|$, where $\hat{G} = \hat{W}^{(k)}A$ and $\|\cdot\|$ is the Frobenius norm. Also, it is not known beforehand which one of the original sources is being estimated by 1u-TANH or 1u-POW3. It seems to depend largely on the value of the initial estimate.[3] Therefore, all the estimates $\hat{\mathbf{w}}_i^{(k)}$ computed by 1u-TANH or 1u-POW3 are sign corrected values of the estimates giving best match with the correct value of $\mathbf{w}_i$.

Fig. 1 shows that 1u-POW3 is performing the worst in all cases: especially it ranks clearly the lowest in separating logistic and t_5-distributed IC; for separating Laplacian IC its performance is close to POW3. The poor performance of 1u-POW3 with logistic IC can be explained by its poor separation ability for sources possessing kurtosis values even moderately close to a Gaussian distribution. The existence of sixth-order moments of the IC is needed for the existence of asymptotic variances of 1u-POW3 ([18], [17], [10]) and POW3 [10]. This explains why for t_5-distributed IC (which do not possess sixth-order moments) there is a slow increase in MSE for 1u-POW3 and POW3 for the largest values of signal length (although this trend would become more apparent for larger signal lengths than $n = 8000$ shown in the figure). TANH has the best performance in all cases although for Laplacian IC the performance difference with 1u-TANH is rather marginal. 1u-TANH however performs well only for the Laplacian IC. For the logistic source TANH reaches close to the CRB. This is not surprising as the nonlinearity *tanh* is the location score function (up to sign and scale differences) of the logistic distribution, and hence the optimal nonlinearity.

[3]E.g., for signal length $n = 750$ only 3% of 1u-POW3 or 1u-TANH estimates did not estimate $\mathbf{w}_2$ (the sign-corrected estimate was closer to $\mathbf{w}_1$ or $\mathbf{w}_3$ as measured by their angles) when the true value of $\mathbf{w}_2$ was given as an initial estimate, and, 0% failed for $n > 1500$.

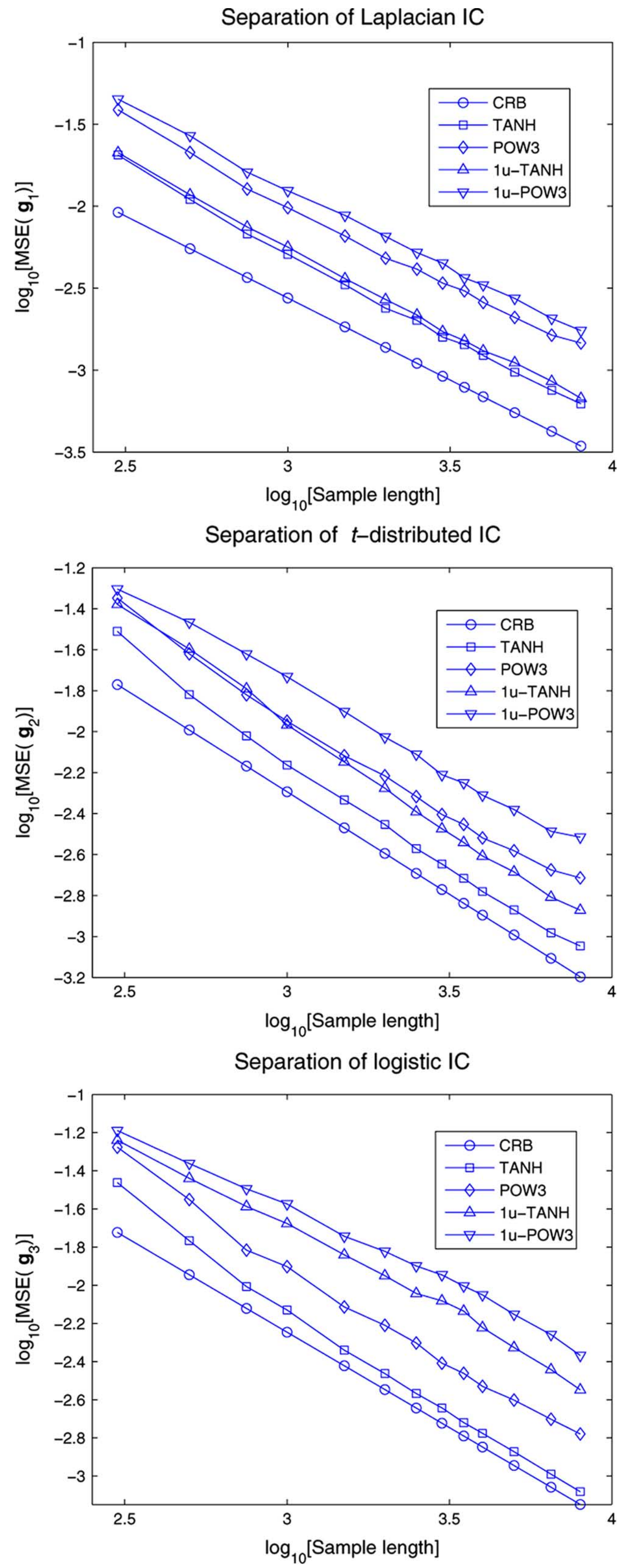

Fig. 1. Separation results in terms of $\text{MSE}(\hat{\mathbf{g}}_i)$ depicted as a function of sample length.

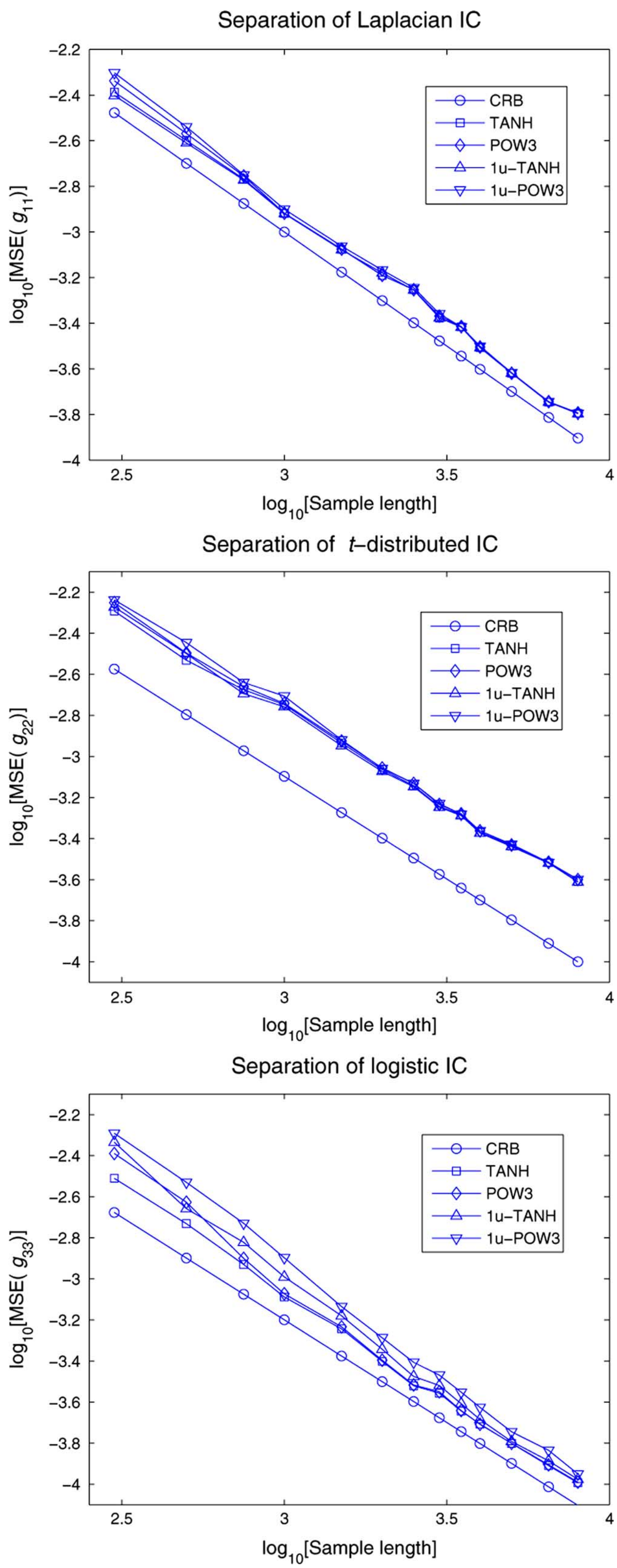

Fig. 2. Separation results in terms of $\text{MSE}(\hat{g}_{ii})$ depicted as a function of sample length.

Fig. 2 depicts the values of $\text{MSE}(\hat{g}_{ii})$ alone $(i \in \{1,2,3\})$. Recall that $\text{var}(\hat{g}_{ii})$ reflects how accurately the presence of ith source itself is estimated in $\hat{s}_i$, whereas $\text{MSE}(\hat{\mathbf{g}}_i)$ includes also the effects of interfering source signals. Fig. 2 clearly shows that there are very little (and for t_5 and Laplacian IC practically none) difference in the performance of the above FastICA estimators. Thus, the differences between the estimators are mainly due to their ability to cancel out the interfering source signals in the estimate of each source.

IV. Conclusion

Based on rather general assumptions on the distributions of the sources s_i, we derived, in Theorem 2, a simple and compact closed-form expression of the CRB for the demixing matrix estimation. The CRB depends on the distribution of s_i only through two scalars of (6) and (7). Hence, in most cases, it yields a practical and easily computable performance criterion for ICA as was demonstrated by our simulation study.

At the end, we wish to clarify that, this paper provides a *novel* compact closed-form expression for the CRB of the demixing matrix estimation based on elegant matrix manipulations. Tedious elementwise derivations used in many related papers are thus be avoided. In addition, the result corrects the error in deriving FIM in a recent related result [10]. We also think that our method of proof based on the novel use of matrix algebra can provide a useful machinery for CRB derivations for related multivariate signal processing models.

Appendix A
Relevant Matrix Algebra

Let L_{ij} denote a $d \times d$ matrix with a 1 in the (i,j) position and 0's elsewhere. Often we write L_i for L_{ii}. It is useful to note that

$$AL_{ij}A^T = \mathbf{a}_i\mathbf{a}_j^T, \quad L_{ij}L_{kl} = 0 \text{ for } j \neq k, \quad L_{ij}L_{jl} = L_{il}. \tag{10}$$

A *commutation matrix* K_d is a $d^2 \times d^2$ block matrix with (i,j)-block being equal to a $d \times d$ matrix that has a 1 at entry (j,i) and 0's elsewhere, that is

$$K_d = \sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d} L_{ij}\otimes L_{ji} + \sum_{i=1}^{d} L_i \otimes L_i. \tag{11}$$

Useful algebraic properties involving the 'vec'-operator and commutation matrix are [13]

$$\begin{aligned} \text{vec}(ABC) &= (C^T \otimes A)\text{vec}(B),\\ K_dK_d &= I,\\ K_d(A\otimes B) &= (B\otimes A)K_d \end{aligned} \tag{12}$$

where the first identity holds for all matrices A, B and C such that the product ABC is properly defined and the third identity holds for all $d \times d$-matrices A and B. Some useful rules of calculus involving the Kronecker product are also listed below [13]:

$$\begin{aligned} (A\otimes B)^T &= A^T \otimes B^T,\\ (A\otimes B)(C\otimes D) &= AC \otimes BD \end{aligned} \tag{13}$$

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1},$$
$$A \otimes (B + C) = A \otimes B + A \otimes C. \tag{14}$$

The first identity in (13) holds for all matrices A and B and second identity for all matrices A, B, C and D such that the products AC and BD are properly defined. The first identity in (14) holds for all nonsingular matrices A and B and second identity holds if B and C are of same size. For later use we note that any $d^2 \times d^2$ block matrix A may be written analytically using its $d \times d$ diagonal-blocks $A[i,i]$ and $d \times d$ off-diagonal blocks $A[i,j], i \neq j$, as follows:

$$A = \sum_{i=1}^{d} L_{ii} \otimes A[i,i] + \sum_{i=1}^{d} \sum_{\substack{j=1 \\ j \neq i}}^{d} L_{ij} \otimes A[i,j]. \tag{15}$$

Appendix B
Additional Lemmas

Lemma 1:

a) Under Assumptions a)–c)

$$E[s_i \varphi_j(s_j)] = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{16}$$

or equivalently, $E[\text{vec}\{\mathbf{s}\varphi(\mathbf{s})^T\}] = \text{vec}(I)$, or equivalently, $E[\nabla_{\boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(\mathbf{x})] = \mathbf{0}$.

b) Under Assumptions a)–c), $\kappa_i \geq 1$ with equality if and only if s_i is a Gaussian random variable. Furthermore, $\lambda_i > 0$, or equivalently, $\eta_i = E[s_i^2 \varphi_i^2(s_i)] > 1$.

c) Assume $f_i''(s)$ exists at all s on the support of the density of f_i. Then under Assumptions a)–c), $\kappa_i = E[\varphi_i'(s)]$ provided d.1) holds and $\lambda_i = E[\varphi_i'(s)s^2] + 1$ provided d.2) holds.

Proof:

a) for $i \neq j$: $E[s_i \varphi_j(s_j)] = E[s_i]E[\varphi_j(s_j)] = 0$ as s_i and s_j are independent and zero mean. The result $E[s_i \varphi_i(s_i)] = -\int s(f_i'(s))/(f_i(s)) f_i(s) ds = -\int s f_i'(s) ds = 1$ follows using integration by parts and Assumption b.2). This result is well-known (see, e.g., [5], [19], and [20]). Note that the expected value of the Fisher score (5) is

$$E[\nabla_{\boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(\mathbf{x})] = E[\text{vec}\{A(I - \mathbf{s}\varphi(\mathbf{s})^T)\}]$$
$$= (I \otimes A) E[\text{vec}\{I - \mathbf{s}\varphi(\mathbf{s})^T\}].$$

Thus, since matrix $(I \otimes A)$ is nonsingular (as A is nonsingular), $E[\nabla_{\boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(\mathbf{x})] = \mathbf{0}$ if and only if $E[\text{vec}\{I - \mathbf{s}\varphi(\mathbf{s})^T\}] = \mathbf{0}$, i.e., $E[\text{vec}\{\mathbf{s}\varphi(\mathbf{s})^T\}] = \text{vec}(I)$, i.e., (16) holds.

b) By the a)-part of the Lemma, $1 = E[s_i \varphi_i(s_i)]$. By correlation inequality: $1 = |E[s_i \varphi_i(s_i)]| \leq \sqrt{E[s_i^2]}\sqrt{E[\varphi_i^2(s_i)]} = \sqrt{\kappa_i}$ with equality if and only if $\varphi_i(s) \propto s$ (i.e., s_i is Gaussian). This result is not new (see, e.g., Appendix B in [5] or [19]). Next note that $\lambda_i = \text{var}(s_i \varphi_i(s_i)) = E[s_i^2 \varphi_i^2(s_i)] - (E[s_i \varphi_i(s_i)])^2 = \eta_i - 1 > 0$ since variance is positive for nondegenerate random variables and $\varphi_i(s)$ cannot be a constant function equal to zero in its entire support (i.e., the uniform distribution) due to assumption b.2).

c) Now $\kappa_i = E[\varphi_i^2(s_i)] = -\int \varphi_i(s)[(f_i'(s))/(f_i(s))] f_i(s) ds = -\int \varphi_i(s) f_i'(s) ds$, which, by integration by parts, equals $E[\varphi_i'(s)]$ provided that d.1) holds. Similarly, observe that $\eta_i = E[\varphi_i^2(s_i)s_i^2] = -\int \varphi_i(s) s^2 f_i'(s) ds$, which, by integration by parts, equals

$$\eta_i = -[\varphi_i(s) s^2 f_i(s)]_a^b$$
$$+ \int \{2\varphi_i(s)s + \varphi_i'(s)s^2\} f_i(s) ds$$
$$= 2 + E[\varphi_i'(s)s^2]$$

where a and b denotes the left and right boundary of the support of the density f_i. Here we used that $\int \varphi_i(s) s f_i(s) ds = 1$ due to a)-part of the Lemma and $[\varphi_i(s) s^2 f_i(s)]_a^b = -[f_i'(s)s^2]_a^b = 0$ provided that d.2) holds. Note that $\lambda_i = \eta_i - 1$. □

Lemma 2: Under Assumptions a)–c)

$$\Omega = E[\text{vec}\{\mathbf{s}\varphi(\mathbf{s})^T\} \text{vec}\{\mathbf{s}\varphi(\mathbf{s})^T\}^T]$$
$$= K_d + \text{vec}(I)\text{vec}(I)^T + J \tag{17}$$

where J is $d^2 \times d^2$ diagonal matrix

$$J = \sum_{i=1}^{d} (\eta_i - 2) L_i \otimes L_i + \sum_{i=1}^{d} \sum_{\substack{j=1 \\ j \neq i}}^{d} \kappa_i L_i \otimes L_j. \tag{18}$$

Proof: Ω is a $d^2 \times d^2$ block matrix whose (i,i)-block $\Omega[i,i]$ is a $d \times d$ matrix $\Omega[i,i] = E[\mathbf{s}\mathbf{s}^T \varphi_i^2(s_i)]$ which is a diagonal matrix since the components of $\mathbf{s}$ are independent and zero mean. Diagonal elements are $(\Omega[i,i])_{jj} = E[s_i^2 \varphi_i^2(s_i)] = \eta_i$ for $j = i$ and $(\Omega[i,i])_{jj} = E[s_j^2 \varphi_i^2(s_i)] = E[s_j^2]E[\varphi_i^2(s_i)] = \kappa_i$ for $i \neq j$ (as $E[s_j^2] = 1$). Thus, $\Omega[i,i] = \eta_i L_i + \kappa_i \sum_{j \neq i} L_j$.

The (i,j)-block $\Omega[i,j]$ of Ω for $i \neq j$ is a $d \times d$ matrix $\Omega[i,j] = E[\mathbf{s}\mathbf{s}^T \varphi_i(s_i) \varphi_j(s_j)]$ which has 1 at entry (i,j) and (j,i) since

$$(\Omega[i,j])_{ij} = (\Omega[i,j])_{ji}$$
$$= E[s_i s_j \varphi_i(s_i) \varphi_j(s_j)]$$
$$= E[s_i \varphi_i(s_i)] E[s_j \varphi_j(s_j)] = 1$$

[where the last identity follows from Lemma 1a)] and 0's elsewhere (since the components of $\mathbf{s}$ are independent with zero mean and $E[\varphi_i(s_i)] = 0$ for $i = 1, \ldots, d$). Thus, $\Omega[i,j] = L_{ij} + L_{ji}$.

Then by using (15) and the rule in (14) we may write Ω as a sum:

$$\Omega = \sum_{i=1}^{d} \eta_i L_i \otimes L_i + \sum_{i=1}^{d} \sum_{\substack{j=1 \\ j \neq i}}^{d} \kappa_i L_i \otimes L_j$$
$$+ \sum_{i=1}^{d} \sum_{\substack{j=1 \\ j \neq i}}^{d} \{(L_{ij} \otimes L_{ji}) + (L_{ij} \otimes L_{ij})\}$$
$$= K_d + \text{vec}(I)\text{vec}(I)^T + J$$

where the last identity follows by using (11) and $\text{vec}(I)\text{vec}(I)^T = \sum_i \sum_{j \neq i} L_{ij} \otimes L_{ij} + \sum_i L_i \otimes L_i$. □

Lemma 3:

$$
\begin{aligned}
&(K_d+J)^{-1}\\
&\quad=\sum_{i=1}^{d} L_i\otimes\left(\frac{1}{\lambda_i}L_i+\sum_{\substack{j=1\\ j\neq i}}^{d}\frac{\kappa_j}{\kappa_i\kappa_j-1}L_j\right)\\
&\qquad+\sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d}\frac{-1}{\kappa_i\kappa_j-1}L_{ij}\otimes L_{ji}
\end{aligned}
$$

where the $d^2\times d^2$ diagonal matrix J is defined in (18).

Proof: It is easy to verify that

$$
(K_d+J)^{-1}=(I-J^{-1}K_d)D^{-1}
$$

where D is a diagonal matrix $D=J-K_dJ^{-1}K_d$. Using properties (12) of the commutation matrix, we get

$$
\begin{aligned}
K_dJ^{-1}K_d&=K_d\left[\sum_{i=1}^{d}\frac{1}{\eta_i-2}L_i\otimes L_i+\sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d}\frac{1}{\kappa_i}L_i\otimes L_j\right]K_d\\
&=\sum_{i=1}^{d}\frac{1}{\eta_i-2}L_i\otimes L_i+\sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d}\frac{1}{\kappa_j}L_i\otimes L_j
\end{aligned}
$$

and thus

$$
\begin{aligned}
D&=\sum_{i=1}^{d}\left\{\eta_i-2-\frac{1}{\eta_i-2}\right\}L_i\otimes L_i\\
&\quad+\sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d}\left\{\kappa_i-\frac{1}{\kappa_j}\right\}L_i\otimes L_j\\
&=\sum_{i=1}^{d}\left\{\frac{(\eta_i-2)^2-1}{\eta_i-2}\right\}L_i\otimes L_i\\
&\quad+\sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d}\left\{\frac{\kappa_i\kappa_j-1}{\kappa_j}\right\}L_i\otimes L_j.
\end{aligned}
$$

Then note that

$$
\begin{aligned}
J^{-1}K_d&=\left[\sum_{i=1}^{d}\frac{1}{\eta_i-2}L_i\otimes L_i+\sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d}\frac{1}{\kappa_i}L_i\otimes L_j\right]\\
&\quad\times\left[\sum_{k=1}^{d}L_k\otimes L_k+\sum_{k=1}^{d}\sum_{\substack{l=1\\ l\neq k}}^{d}L_{kl}\otimes L_{lk}\right]\\
&=\sum_{i=1}^{d}\frac{1}{\eta_i-2}L_i\otimes L_i+\sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d}\frac{1}{\kappa_i}L_{ij}\otimes L_{ji}
\end{aligned}
$$

which follows using (10) and (13). Thus

$$
\begin{aligned}
I-J^{-1}K_d&=\sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d}L_i\otimes L_j\\
&\quad+\sum_{i=1}^{d}\left\{\frac{(\eta_i-2)-1}{\eta_i-2}\right\}L_i\otimes L_i\\
&\quad+\sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d}\left\{\frac{-1}{\kappa_i}\right\}L_{ij}\otimes L_{ji}
\end{aligned}
$$

which follows by replacing identity matrix I in the left-hand side of the equation by $\sum_i\sum_{j\neq i}L_i\otimes L_j+\sum_i L_i\otimes L_i$. It is now straightforward to verify, resorting to (10) and (14), that the product of the matrices $I-J^{-1}K_d$ and D^{-1} derived above gives the expression for $(K_d+J)^{-1}$ stated in the lemma. Recall that $\lambda_i=\eta_i-1$. □

Appendix C
Proofs

Proof of Theorem 1: Using Lemma 1a) and Lemma 2 gives

$$
\begin{aligned}
&E\left[\mathrm{vec}\{I-\mathbf{s}\varphi(\mathbf{s})^T\}\mathrm{vec}\{I-\mathbf{s}^T\varphi(\mathbf{s})^T\}\right]\\
&\quad=\mathrm{vec}(I)\mathrm{vec}(\mathrm{I})^{\mathrm{T}}+\Omega-E[\mathrm{vec}\{\mathbf{s}\varphi(\mathbf{s})^T\}]\mathrm{vec}(I)^T\\
&\quad\quad-\mathrm{vec}(I)E[\mathrm{vec}\{\mathbf{s}\varphi(\mathbf{s})^T\}]^T\\
&\quad=K_d+J
\end{aligned}
$$

where Ω is defined in (18). Plugging the above expression in (8), and the summation (11) in place of K_d yields

$$
\begin{aligned}
\mathcal{I}_{\boldsymbol{\theta}}&=(I\otimes A)(K_d+J)(I\otimes A^T)\\
&=(I\otimes A)\left[\sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d}L_{ij}\otimes L_{ji}+\sum_{i=1}^{d}\underbrace{(\eta_i-1)}_{=\lambda_i}L_i\otimes L_i\right.\\
&\qquad\left.+\sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d}\kappa_iL_i\otimes L_j\right](I\otimes A^T)\\
&=\sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d}L_{ij}\otimes AL_{ji}A^T+\sum_{i=1}^{d}\lambda_iL_i\otimes AL_iA^T\\
&\quad+\sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d}\kappa_iL_i\otimes AL_jA^T\\
&=\sum_{i=1}^{d}L_i\otimes\left(\lambda_i\mathbf{a}_i\mathbf{a}_i^T+\kappa_i\sum_{\substack{j=1\\ j\neq i}}^{d}\mathbf{a}_j\mathbf{a}_j{}^T\right)\\
&\quad+\sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d}L_{ij}\otimes\mathbf{a}_j\mathbf{a}_i^T
\end{aligned}
$$

which by (15) gives the stated claim. □

Proof of Theorem 2: Since $\mathcal{I}_{\boldsymbol{\theta}} = (I \otimes A)(K_d + J)(I \otimes A^T)$, it follows that

$$\mathcal{I}_{\boldsymbol{\theta}}^{-1} = (I \otimes W^T)(K_d + J)^{-1}(I \otimes W).$$

Then using Lemma 3 and recalling rules of calculus of Kronecker product stated in (13), it is straightforward to write $\mathcal{I}_{\boldsymbol{\theta}}^{-1}$ in the form claimed in the theorem. □

ACKNOWLEDGMENT

E. Ollila would like to thank the Academy of Finland for supporting this research. The authors would like to thank the anonymous reviewers for their helpful and valuable comments and suggestions.

REFERENCES

[1] P. Comon, "Independent component analysis—A new concept?," *Signal Process.*, vol. 36, pp. 287–314, 1994.

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

[3] J. Eriksson and V. Koivunen, "Identifiability, separability and uniqueness of linear ICA models," *IEEE Signal Process. Lett.*, vol. 11, no. 7, pp. 601–604, Jul. 2004.

[4] A. Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing*. New York: Wiley, 2002.

[5] O. Shalvi and E. Weinstein, "Maximum likelihood and lower bounds in system identification with non-Gaussian inputs," *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 328–339, 1994.

[6] D.-T. Pham and P. Garat, "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1712–1725, Jul. 1997.

[7] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proc. IEEE*, vol. 9, no. 10, pp. 2009–2025, 1998.

[8] D. Yellin and B. Friedlander, "Multichannel system identification and deconvolution: Performance bounds," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1410–1414, May 1999.

[9] Y. Lomnitz and A. Yeredor, "A blind ML-scheme for blind source separation," presented at the IEEE Workshop Statistical Signal Processing (SSP), St. Louis, MO, Sept. 28–Oct. 1, 2003, St. Louis.

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

[11] V. Vigneron and C. Jutten, C. G. Puntonet and A. Prieto, Eds., "Fisher information in source separation problems," in *Proc. 5th Int. Conf. Independent Component Analysis (ICA)*, Granada, Spain, Sep. 2004.

[12] P. J. Bickel and K. A. Doksum, *Mathematical Statistics–Basic Ideas and Selected Topics*. Upper Saddle River, NJ: Prentice-Hall, 2001, vol. I.

[13] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus*. Chichester, U.K.: Wiley, 1999.

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

[15] E. Doron, A. Yeredor, and P. Tichavský, "A Cramér–Rao-induced bound for blind separation of stationary parametric Gaussian sources," *IEEE Signal Process. Lett.*, vol. 14, no. 6, pp. 417–420, Jun. 2007.

[16] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.*, vol. 9, no. 7, pp. 1483–1492, 1997.

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

[18] A. Hyvärinen, "One-unit contrast functions for independent component analysis: A statistical analysis," in *Neural Networks for Signal Processing VII (Proc. IEEE Workshop on Neural Networks for Signal Processing)*, Amelia Island, FL, 1997, pp. 388–397.

[19] D.-T. Pham, "Blind separation of instantaneous mixture sources via an independent component analysis," *IEEE Trans. Signal Process.*, vol. 44, no. 11, pp. 2768–2779, Nov. 1996.

[20] D.-T. Pham, "Blind separation of instantaneous mixture of sources based on order statistics," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 363–375, Feb. 2000.

Esa Ollila (M'03) received the M.Sc. degree in mathematics from the University of Oulu, Finland, in 1998 and the Ph.D. degree in statistics (with honors) from the University of Jyväskylä, Finland, in 2002.

In 2002, he joined the Academy of Finland in the Signal Processing Laboratory at Helsinki University of Technology, Helsinki, Finland, where he is currently a Postdoctoral Fellow. His current research interests focus on statistical signal processing and robust and nonparametric statistical methods.

Hyon-Jung Kim received the Ph.D. degree in statistics from North Carolina State University.

Since 2001, she has been an Assistant Professor in the Department of Mathematical Sciences, University of Oulu. She is also a Research Associate at the Finnish Forest Research Institute (METLA). Her current research interests include independent component analysis, spatial statistics, and Bayesian methods.

Visa Koivunen (M'87–SM'98) received the D.Sc. (Tech.) degree (with honors) from the Department of Electrical Engineering, University of Oulu.

From 1992 to 1995, he was a Visiting Researcher at the University of Pennsylvania, Philadelphia. In 1996, he held a faculty position at the Department of Electrical Engineering, University of Oulu. From August 1997 to August 1999, he was an Associate Professor at the Signal Processing Laboratory, Tampere University of Technology. Since 1999, he has been a Professor of signal processing at the Department of Electrical and Communications Engineering, Helsinki University of Technology (HUT), Finland. He is one of the Principal Investigators in the SMARAD (Smart and Novel Radios) Center of Excellence in Radio and Communications Engineering nominated by the Academy of Finland. Since 2003, he has been also Adjunct Full Professor at the University of Pennsylvania, Philadelphia. During his sabbatical leave in 2006–2007, he was Nokia Visiting Fellow at the Nokia Research Center, as well as a Visiting Fellow at Princeton University, Princeton, NJ. His research interest include statistical, communications, and sensor array signal processing. He has published more than 200 papers in international scientific conferences and journals.

Dr. Koivunen received the Primus Doctor (best graduate) Award among the doctoral graduates in years 1989–1994. He coauthored the papers receiving the best paper award in IEEE PIMRC 2005, EUSIPCO 2006, and EuCAP 2006. He served as an Associate Editor for IEEE SIGNAL PROCESSING LETTERS. He is a member of the editorial board for the *Signal Processing* journal and the *Journal of Wireless Communication and Networking*. He is a member of the IEEE Signal Processing Society for the Communication Technical Committee (SPCOM-TC). He served as the general chair of the IEEE Signal Processing Advances in Wireless Communication (SPAWC) 2007 conference in Helsinki, Finland, in June 2007.