CRAMÉR-RAO LOWER BOUND FOR LINEAR INDEPENDENT COMPONENT ANALYSIS

Zbyněk Koldovský[1,2], *Petr Tichavský*[1], *and Erkki Oja*[3]

[1]Institute of Information Theory and Automation,
P.O.Box 18, 182 08 Prague 8, Czech Republic

[2] Faculty of Nuclear Sciences and Physical Engineering
Czech Technical University, Trojanova 13, 120 00 Prague 2,

[3]Neural Networks Research Centre, Helsinki University of Technology
P.O.Box 5400, 02015 HUT, Finland

ABSTRACT

This paper derives a closed-form expression for the Cramér-Rao bound (CRB) on estimating the source signals in the linear independent component analysis problem, assuming that all independent components have finite variance. It is also shown that the fixed-point algorithm known as FastICA can approach the CRB (the estimate can be nearly efficient) in two situations: (1) when the distribution of the sources is not too much different from Gaussian, for symmetric version of the algorithm using any of the custom nonlinear functions (pow3, tanh, gauss), and (2) when the distribution of the sources is very different from Gaussian (e.g. has long tails) and the nonlinear function in the algorithm equals the score function of each independent component.

1. INTRODUCTION

The purpose of Blind Source Separation (BSS) is the extraction of a set of signals based merely on their mixtures. The instantaneous linear mixing model is presently a well-studied problem, in particular in Independent Component Analysis (ICA), one of the most successful methods for BSS [5]. We can express the mixing process as

$$\mathbf{X} = \mathbf{AS}, \tag{1}$$

where

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{d1} & \cdots & x_{dN} \end{pmatrix}$$

denotes a matrix of N samples of d mixed signals; similarly, $\mathbf{S}$ denotes a matrix of samples of the original signals s_{ij}. $\mathbf{A}$ is an unknown regular $d \times d$ mixing matrix. In this paper, we consider a model by which s_{ij} are mutually independent i.i.d. random variables with probability density functions (pdf) $f_i(s_{ij})$ $i = 1, \ldots, d$. Variables with the same pdf as s_{ij} for all $j = 1, \ldots, N$ will be denoted by s_i. The mutual independence of s_i is the basic assumption of ICA.

Thanks to an increasing attention to this problem, many algorithms have been developed in the last two decades, for instance, FastICA [3], JADE, and Infomax; for a review, see [5]. The differences between proposed algorithms are characterized by their accuracy, convergence, or computational demand: those are the natural questions when applying ICA in practice. Therefore, some theoretical analysis [2, 6, 9] or experimental comparisons have been provided [4]. In this article, we propose a Cramér-Rao lower bound for ICA, especially, for the elements of the gain matrix (defined below), which give us an algorithm independent lower bound for separation performance. It is also shown that the algorithm FastICA (the symmetric version) can approach this bound, when the original signals are nearly Gaussian.

2. CRAMÉR-RAO LOWER BOUND FOR ICA

Consider a vector of parameters $\boldsymbol{\theta}$ being estimated from a data vector $\mathbf{x}$, having probability density $f_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})$, using some unbiased estimator $\hat{\boldsymbol{\theta}}$. The Cramér-Rao lower bound (CRB) is the lower bound for the variance of $\hat{\boldsymbol{\theta}}$. Assume that $f_{\mathbf{x}|\boldsymbol{\theta}}$ is smooth and the following Fisher information matrix exists:

$$\mathbf{F}_{\boldsymbol{\theta}} = \mathrm{E}_{\boldsymbol{\theta}}\left[\frac{1}{f^2_{\mathbf{x}|\boldsymbol{\theta}}}\frac{\partial f_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\left(\frac{\partial f_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T\right] \tag{2}$$

Then, under some mild regularity condition [1], it holds

$$\operatorname{cov}\hat{\boldsymbol{\theta}} \geq \mathrm{CRB}_{\boldsymbol{\theta}} = \mathbf{F}_{\boldsymbol{\theta}}^{-1}.$$

Next, if $\boldsymbol{\varphi} = \boldsymbol{\varphi}(\boldsymbol{\theta})$ is a differentiable function of $\boldsymbol{\theta}$, then the Fisher information matrix for $\boldsymbol{\varphi}$ exists as well and is equal to

$$\mathbf{F}_{\boldsymbol{\varphi}} = \mathbf{J}_{\boldsymbol{\theta}}^{-1}\mathbf{F}_{\boldsymbol{\theta}}\mathbf{J}_{\boldsymbol{\theta}}^{-T}, \tag{3}$$

where $\mathbf{J}_{\boldsymbol{\theta}}$ is the Jacobian of the mapping $\boldsymbol{\varphi}(\boldsymbol{\theta})$. If the mapping is linear, or $\boldsymbol{\varphi}(\boldsymbol{\theta}) = \mathbf{M}\boldsymbol{\theta}$ for some regular matrix $\mathbf{M}$, then $\mathbf{J}_{\boldsymbol{\theta}} = \mathbf{M}$.

In the context of ICA, we first focus on deriving the CRB for estimation of the de-mixing matrix $\mathbf{W} = \mathbf{A}^{-1}$, i.e., the parameter vector is $\boldsymbol{\theta} = \text{vec}[\mathbf{W}]$.

2.1. Assumptions

The following assumptions will be considered throughout this paper:

$$\sigma_i^2 \stackrel{\text{def}}{=} \text{E}\left[s_i^2\right] = \int_R t^2 f_i(t)dt < +\infty \quad (4)$$

$$\kappa_i \stackrel{\text{def}}{=} \text{E}\left[\psi_i^2(s_i)\right] = \int_R \psi_i^2(t) f_i(t)dt < +\infty \quad (5)$$

$$\eta_i \stackrel{\text{def}}{=} \text{E}\left[s_i^2 \psi_i(s_i)^2\right] = \int_R t^2 \psi_i^2(t) f_i(t)dt < +\infty, \quad (6)$$

where $i = 1, \ldots, d$ and ψ_i denotes the score function of the corresponding pdf, i.e., $\psi_i(t) = -\frac{f_i'(t)}{f_i(t)}$. The mean value of the original signals is irrelevant in ICA, hence, we shall assume that it is zero. Variances σ_i^2 can be assumed to be equal to one, without any loss of generality.

2.2. The Fisher information matrix

From the independence of the original signals it follows that their mutual pdf is $f_{\mathbf{S}}(\mathbf{S}) = \prod_{i=1}^{d}\prod_{j=1}^{N} f_i(s_{ij})$. Then, using the transformation $\mathbf{X} = \mathbf{W}^{-1}\mathbf{S}$,

$$f_{\mathbf{x}}(\mathbf{X}) = |\det \mathbf{W}| f_{\mathbf{S}}(\mathbf{W}\mathbf{X}). \quad (7)$$

Incorporating this density into (2) the mn-th element of the $d^2 \times d^2$ Fisher information matrix, where $m = (i-1)d + j$, $n = (u-1)d + v$, and w_{ij} denotes the ij-th element of the matrix $\mathbf{W}$, is

$$\mathbf{F}_{mn} = \text{E}\left[\frac{|\det \mathbf{W}|^{-2}}{f_{\mathbf{S}}^2(\mathbf{S})} \frac{\partial f_{\mathbf{x}}}{\partial w_{ij}} \frac{\partial f_{\mathbf{x}}}{\partial w_{uv}}\right]. \quad (8)$$

The direct computation is rather lengthy and due to lack of space we refer to [8], where it is shown that

$$\frac{\partial f_{\mathbf{x}}}{\partial w_{ij}} = |\det \mathbf{W}| f_{\mathbf{S}}(\mathbf{S}) \left[a_{ji} + \sum_{\ell=1}^{N}\sum_{k=1}^{d} \frac{f_i'(s_{i\ell})}{f_i(s_{i\ell})} a_{jk} s_{k\ell}\right]$$

and

$$\mathbf{F}_{mn} = (N-1)^2 a_{ji} a_{vu} + N a_{ju} a_{vi} + \\ + \delta_{iu} N a_{ji} a_{vi}(\eta_i - 2) + \delta_{iu} N \kappa_i \sum_{\ell=1, \ell\neq u}^{d} a_{j\ell} a_{v\ell} \quad (9)$$

with κ_i, η_i defined in (5)-(6), δ_{iu} is the Kronecker's delta, and a_{ij} denotes the ij-th element of the mixing matrix $\mathbf{A}$. It can be shown, using (3), that

$$\mathbf{F}_{\boldsymbol{\theta}} = (\mathbf{A}^T \otimes \mathbf{I})\mathbf{F}_{\mathbf{I}}(\mathbf{A} \otimes \mathbf{I}), \quad (10)$$

where $\mathbf{F}_{\mathbf{I}}$ stands for the Fisher information matrix derived for a case when $\mathbf{A} = \mathbf{I}$ (identity matrix); $\otimes$ denotes the Kronecker product. Substituting $a_{ij} = \delta_{ij}$ into (9), it easily follows that

$$(\mathbf{F}_{\mathbf{I}})_{mn} = (N-1)^2 \delta_{ji}\delta_{vu} + N\delta_{ju}\delta_{vi} + \\ + N\Big(\delta_{ji}\delta_{vu}\delta_{vi}(\eta_i - \kappa_i) + \delta_{iu}\delta_{vj}\kappa_i\Big). \quad (11)$$

Some properties of the matrix will be shown in Appendix.

2.3. Accuracy of the estimation of $\mathbf{G} = \widehat{\mathbf{W}}\mathbf{A}$

Let $\widehat{\mathbf{W}}$ denote an estimator of the de-mixing matrix $\mathbf{W}$. Estimated signals $\hat{\mathbf{S}}$ are then $\hat{\mathbf{S}} = \widehat{\mathbf{W}}\mathbf{X} = \widehat{\mathbf{W}}\mathbf{A}\mathbf{S}$. It is interesting to compute the CRB for the elements of the so-called *gain* matrix $\mathbf{G} = \widehat{\mathbf{W}}\mathbf{A}$, because they characterize the residual presence of the $j-$th component in estimating the $i-$th independent component for $i, j = 1, \ldots, d$, $i \neq j$. Note that the new parameter vector $\boldsymbol{\theta}_{\mathbf{G}} = \text{vec}[\mathbf{G}]$ is just a linear function of the parameter $\boldsymbol{\theta}$, i.e., $\boldsymbol{\theta}_{\mathbf{G}} = \text{vec}[\widehat{\mathbf{W}}\mathbf{A}] = (\mathbf{A}^T \otimes \mathbf{I})\text{vec}[\widehat{\mathbf{W}}] = (\mathbf{A}^T \otimes \mathbf{I})\boldsymbol{\theta}$. Then, using (3) and (10), the Fisher information matrix of $\boldsymbol{\theta}_{\mathbf{G}}$ is

$$\mathbf{F}_{\mathbf{G}} = (\mathbf{W}^T \otimes \mathbf{I})\mathbf{F}_{\boldsymbol{\theta}}(\mathbf{W} \otimes \mathbf{I}) = \mathbf{F}_{\mathbf{I}}. \quad (12)$$

Note that $\mathbf{F}_{\mathbf{G}}$ is independent of the mixing matrix $\mathbf{A}$. The CRB for the $ij-$th element of $\mathbf{G}$ is

$$\text{var}(\mathbf{G}_{ij}) \geq \text{CRB}(\mathbf{G}_{ij}) = ({\mathbf{F}_{\mathbf{I}}}^{-1})_{mm}$$

where $m = (i-1)d + j$ and $i \neq j$. In Appendix A it is proved that for such m

$$({\mathbf{F}_{\mathbf{I}}}^{-1})_{mm} = \frac{1}{N}\frac{\kappa_j}{\kappa_i\kappa_j - 1}, \quad (13)$$

which gives us the desired lower bound

$$\text{var}(\mathbf{G}_{ij}) \geq \text{CRB}(\mathbf{G}_{ij}) = \frac{1}{N}\frac{\kappa_j}{\kappa_i\kappa_j - 1}. \quad (14)$$

The diagonal elements of $\mathbf{G}$ are not as important, they just reflect the accuracy of estimating the power of the components, or equivalently, the norm of rows of the de-mixing matrix. The signal-to-interference ratio of the $i-$th component can be defined as [9]

$$\text{SIR}_i = \frac{1}{\sum_{\substack{\ell=1 \\ \ell\neq i}}^{d} \text{E}[\mathbf{G}_{i\ell}^2]}. \quad (15)$$

3. TIGHTNESS OF THE BOUND

A CRB is called tight, if there exists an estimator which has variance equal to the bound. In this section we show that the CRB can be approached by the FastICA algorithm in some cases. In particular, the CRB will be compared with asymptotical variances of the gain matrix elements for this method, derived in [9].

Let $\mathbf{G}^{1U}$ and $\mathbf{G}^{SYM}$, respectively, be the gain matrix obtained by the one-unit and the symmetric variant of the algorithm using a nonlinear function $g(\cdot)$. It was shown in [9] that the normalized gain matrix elements $N^{1/2}\mathbf{G}^{1U}_{k\ell}$ and $N^{1/2}\mathbf{G}^{SYM}_{k\ell}$ have asymptotically Gaussian distribution $\mathcal{N}(0, V^{1U}_{k\ell})$ and $\mathcal{N}(0, V^{SYM}_{k\ell})$, where

$$V^{1U}_{k\ell} = \frac{\beta_k - \mu_k^2}{(\mu_k - \rho_k)^2} \tag{16}$$

$$V^{SYM}_{k\ell} = \frac{\beta_k - \mu_k^2 + \beta_\ell - \mu_\ell^2 + (\mu_\ell - \rho_\ell)^2}{(|\mu_k - \rho_k| + |\mu_\ell - \rho_\ell|)^2} \tag{17}$$

with $\mu_i = \mathrm{E}[s_i g(s_i)]$, $\rho_i = \mathrm{E}[g'(s_i)]$, $\beta_i = \mathrm{E}[g^2(s_i)]$, and $g'(\cdot)$ being the first derivative of $g(\cdot)$. It only has to be assumed that the above derivative and expectations exist.

Next, it can be shown that the asymptotic variance in (16) achieves its minimum for $g(\cdot)$ being equal to the score function of the distribution f_i, i.e., for $g(x) = \psi(x) = -f_i'(x)/f_i(x)$. In that case, $\mu_i = 1$, $\rho_i = \beta_i = \kappa_i$, and

$$\mathrm{var}[\mathbf{G}^{1U-opt}_{ij}] \approx \frac{1}{N} V^{1U}_{k\ell} = \frac{1}{N}\frac{1}{\kappa_i - 1} \tag{18}$$

$$\mathrm{var}[\mathbf{G}^{SYM-opt}_{ij}] \approx \frac{1}{N} V^{SYM}_{k\ell} = \frac{1}{N}\left(\frac{1}{4} + \frac{1}{2}\frac{1}{\kappa_i - 1}\right) \tag{19}$$

where κ_i was defined in (5). It can also be shown that for any distribution f_i it holds $\kappa_i \geq 1$ and the equality is attained only for the Gaussian distribution.

Comparison of (18) and (19) with (14) implies that the algorithm FastICA is nearly statistically efficient in two situations:

(1) One-unit version FastICA with the optimum nonlinearity is asymptotically efficient for $\kappa_i \to \infty$, regardless of the value of κ_j.

(2) Symmetric FastICA is nearly efficient for κ_i lying in a neighborhood of 1^+, provided that all independent components have the same probability distribution function, and g is equal to the joint score function.

Note, however, than in the latter case, as $\kappa_i \to 1^+$, the algorithm asymptotic variance goes to infinity, and the algorithm itself may fail, or its convergence might be slow. It happens because the task is badly conditioned: it is hard to separate components with nearly Gaussian distribution.

The performance of FastICA cannot approach the CRB in the following cases.

(1) Probability density function of the components is not smooth enough. An example is the uniform distribution.

(2) Probability density function of the components is smooth, but its score function is not continuous. In that case, the algorithm appears not to converge. An example is the generalized Gaussian distribution with parameter $\alpha \leq 1$.

(3) Components have distinct distribution functions.

Analytical performance of alternative ICA estimators is not known in closed form. The following simulation section studies the performance of two computationally extensive algorithms that are claimed to be more accurate than older algorithms: RADICAL [11] and NPICA [12]. We have tested implementations available on internet and compared their performance with the CRB as well. Their general superiority or tightness was not proven.

3.1. Separation of generalized Gaussian distributions

The generalized Gaussian distribution with the shape parameter α was introduced for BSS in [10]. For easy reference, its main properties are listed in Appendix B. Note that the distribution is Gaussian for $\alpha = 2$, it is Laplace distribution for $\alpha = 1$, it approaches the uniform distribution for $\alpha \to \infty$, and it has long tails for $\alpha \to 0$.

The score function of this distribution is proportional to $|x|^{\alpha-1}\mathrm{sign}(x)$. Hence $g(x) = |x|^{\alpha-1}\mathrm{sign}(x)$ is the theoretically optimum nonlinearity for this distribution. This function is, however, not continuous for $\alpha \leq 1$, and the resulting algorithm is not convergent in this region. For $\alpha \geq 1.5$ the FastICA algorithm with the optimum nonlinearity performs well and both the one-unit and symmetric versions of the algorithm are nearly efficient. This is shown in Figure 1, giving the SIR of eq. (15). The simulations are obtained from 50 independent separations of a signal of length $N = 1000$ with $d = 3$ components, all having the same distribution function.

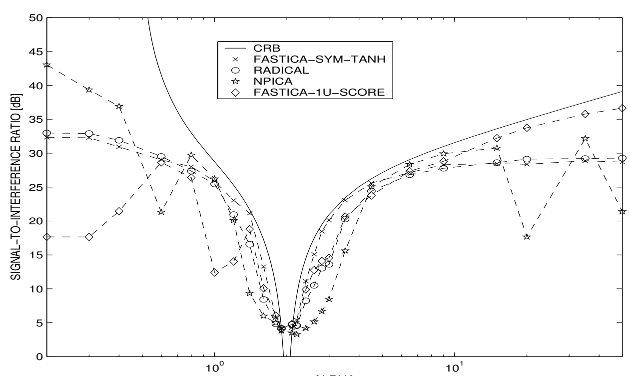

Fig. 1: Comparison of CRB with performance of four ICA techniques. The SIR or inverse variance is shown.

It is interesting to note that the mean square of the score function, κ_i, is finite for $\alpha > 1/2$ and infinite otherwise. It follows that the asymptotic variance of the optimum one-unit FastICA and the CRB is zero for $\alpha \leq 0.5$. It is believed that the estimation might be possible with the mean square error decaying faster than $1/N$, perhaps as $1/N^2$.

In the neighborhood of the point $\alpha = 2$ (Gaussian distribution), the symmetric FastICA with nonlinearity "tanh" appears to perform best of all the methods.

APPENDIX A: Computing $\mathbf{F_I}^{-1}$

Definition (11) can be rewritten as $\mathbf{F_I} = (N-1)^2\mathbf{F}_1 + N(\mathbf{P}+\mathbf{\Sigma})$, where mn-th element of $\mathbf{F}_1$, $\mathbf{P}$ and $\mathbf{\Sigma}$ are $\delta_{ji}\delta_{vu}$, $\delta_{ju}\delta_{vi}$, and $\delta_{ji}\delta_{vu}\delta_{vi}(\eta_i - \kappa_i) + \delta_{iu}\delta_{vj}\kappa_i$, respectively, for $m = (i-1)d + j$ and $n = (u-1)d + v$. Note that $\mathbf{F}_1$ is a rank-one matrix, $\mathbf{F}_1 = \mathbf{e}\mathbf{e}^T$, where $\mathbf{e} = \text{vec}(\mathbf{I})$. Applying the matrix inversion lemma gives

$$\mathbf{F_I}^{-1} = \frac{1}{N}\left[(\mathbf{P}+\mathbf{\Sigma})^{-1} - \frac{(\mathbf{P}+\mathbf{\Sigma})^{-1}\mathbf{e}\mathbf{e}^T(\mathbf{P}+\mathbf{\Sigma})^{-1}}{\frac{N}{(N-1)^2} + \mathbf{e}^T(\mathbf{P}+\mathbf{\Sigma})^{-1}\mathbf{e}}\right]$$

To compute the inversion $(\mathbf{P}+\mathbf{\Sigma})^{-1}$ note that $\mathbf{\Sigma}$ is diagonal,

$$\mathbf{\Sigma} = \text{diag}(\underbrace{\eta_1, \kappa_1, \ldots, \kappa_1}_{d}, \underbrace{\kappa_2, \eta_2, \kappa_2, \ldots, \kappa_2}_{d}, \ldots),$$

and $\mathbf{P}$ is a special permutation matrix such that $\mathbf{P}\text{vec}(\mathbf{M}) = \text{vec}(\mathbf{M}^T)$ for any $d \times d$ matrix $\mathbf{M}$. Moreover, $\mathbf{P}$ obeys $\mathbf{P}\mathbf{P} = \mathbf{I}$, and for any diagonal matrix $\mathbf{D} = \text{diag}(\mathbf{d})$ it holds that

$$\mathbf{P}\mathbf{D} = \mathbf{D}'\mathbf{P},$$

where $\mathbf{D}' = \text{diag}(\mathbf{P}\mathbf{d}) = \mathbf{P}\mathbf{D}\mathbf{P}$. These facts can be used to show that the inversion of $\mathbf{P}+\mathbf{\Sigma}$ can be written in the form $\mathbf{D}_1 + \mathbf{D}_2\mathbf{P}$ for suitable diagonal matrices $\mathbf{D}_1$ and $\mathbf{D}_2$. The equality

$$(\mathbf{P}+\mathbf{\Sigma})(\mathbf{D}_1 + \mathbf{D}_2\mathbf{P}) = \mathbf{I}$$

is fulfilled for $\mathbf{\Sigma}\mathbf{D}_1 + \mathbf{D}_2' = \mathbf{I}$ and $\mathbf{D}_1' + \mathbf{\Sigma}\mathbf{D}_2 = \mathbf{0}$. Hence

$$\mathbf{D}_1 = (\mathbf{\Sigma}'\mathbf{\Sigma} - \mathbf{I})^{-1}\mathbf{\Sigma}' \quad \text{and} \quad \mathbf{D}_2 = -\mathbf{\Sigma}^{-1}\mathbf{D}_1'$$

where $\mathbf{\Sigma}' = \mathbf{P}\mathbf{\Sigma}\mathbf{P}$ and $\mathbf{D}_1' = \mathbf{P}\mathbf{D}_1\mathbf{P}$. Finally, it can be shown that $(\mathbf{F_I}^{-1})_{mm} = N^{-1}(\mathbf{D}_1)_{mm}$ for $m = (i-1)d + j, i \neq j$. (13) easily follows. The detailed proof is in [8].

APPENDIX B: Generalized Gaussian distribution

The generalized Gaussian density function with parameter α, zero mean and variance one is defined as

$$f_\alpha(x) = \frac{\alpha\beta_\alpha}{2\Gamma(1/\alpha)} \exp\{-(\beta_\alpha|x|)^\alpha\} \tag{20}$$

where $\alpha > 0$ is a shape parameter, $\Gamma(\cdot)$ is the Gamma function and $\beta_\alpha = \sqrt{\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}}$.

The $k-$th absolute moment for the distribution is

$$\text{E}_\alpha\{|x|^k\} = \int_\infty^\infty |x|^k f_\alpha(x)dx = \frac{1}{\beta_\alpha^k}\frac{\Gamma\left(\frac{k+1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)} \tag{21}$$

The score function of the distribution is

$$\psi_\alpha(x) = -\frac{\frac{\partial f_\alpha(x)}{\partial x}}{f_\alpha(x)} = \frac{|x|^{\alpha-1}\text{sign}(x)}{\text{E}_\alpha[|x|^\alpha]} \tag{22}$$

Then, simple computations give

$$\kappa_\alpha = \text{E}_\alpha[\psi_\alpha^2(x)] = \begin{cases} \frac{\Gamma(2-\frac{1}{\alpha})\Gamma(\frac{3}{\alpha})}{[\Gamma(1+\frac{1}{\alpha})]^2} & \text{for} \quad \alpha > 1/2 \\ +\infty & \text{otherwise.} \end{cases}$$

4. REFERENCES

[1] R. C. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York, 1973.

[2] J.-F. Cardoso and B. H. Laheld, "Equivariant Adaptive Source Separation", *IEEE Trans. on Signal Processing*, vol. 44, pp. 3017-3030, 1996.

[3] A. Hyvärinen and E. Oja, "A Fast Fixed-Point Algorithm for Independent Component Analysis". *Neural Computation*, Vol. 9, No. 7, pp. 1483-1492, 1997.

[4] X. Giannakopoulos, J. Karhunen, and E. Oja, "Experimental Comparison of Neural Algorithms for Independent Component Analysis and Blind Separation," *Int. J. of Neural Systems*, vol. 9, pp. 651-656, 1999.

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

[6] E. Oja, "Convergence of the Symmetrical FastICA Algorithm", *Proc. 9th Int. Conf. on Neural Information Processing (ICONIP '02)*, Nov. 18-22, 2002.

[7] P. Tichavský and Z. Koldovský, "Optimal Pairing of Signal Components Separated by Blind Techniques", *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 119–122, 2004.

[8] P. Tichavský and Z. Koldovský, "Cramér-Rao Lower Bound for Independent Component Analysis", *Technical report* No. 2123,, ÚTIA, AV ČR, November 2004.

[9] P. Tichavský, Z. Koldovský, and E. Oja, "Performance Analysis of the FastICA Algorithm and Cramér-Rao Bounds for Linear Independent Component Analysis", submitted.

[10] K. Waheed and F.M. Salam, "Blind source recovery using an adaptive generalized Gaussian score function", Circuits and Systems, 2002. MWSCAS-2002. The 2002 45th Midwest Symposium on, 4-7 Aug. 2002 Vol.3, pp. 656-659.

[11] E.G. Learned-Miller, J.W. FIsher III., "ICA using spacings estimates of entropy", *J. Machine Learning Research* **4** (2003), pp. 1271-1295.

[12] R. Boscolo, H. Pan, V.P. Roychowdhury, "Independent component analysis based on nonparametric density estimation", *IEEE Tran. Neural Networks*, Vol. 15, No.1, January 2004, pp. 55-65.