4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), April 2003, Nara, Japan

IDENTIFIABILITY AND SEPARABILITY OF LINEAR ICA MODELS REVISITED

Jan Eriksson and Visa Koivunen

Signal Processing Laboratory
Helsinki University of Technology
P.O. Box 3000, FIN-02015 HUT, Finland
{jan.eriksson,visa.koivunen}@hut.fi

ABSTRACT

We prove theorems that ensure identifiability, separability and uniqueness of linear ICA models. The currently used conditions in ICA community are hence extended to wider class of mixing models and source distributions. Examples illustrating the above concepts are presented as well.

1. INTRODUCTION

In this paper we are concerned with conditions ensuring that the mixing system may be identified and sources separated in a linear instantaneous ICA (Independent Component Analysis) model. The fundamental problems of identifiability, separability and uniqueness of ICA models are addressed. We give formal definitions to each of these concepts. A theorem is given for each of the above conditions and proofs are constructed. These theorems generalize the results given in earlier presented theorems addressing the same problems [Com94, CL96, Car98, DLDMV99, TJ99]. In particular, some of the restricting assumptions made in those theorems may be relaxed. As a result, the identifiability, separability or uniqueness can be ensured for wider class of ICA models and source distributions. The proofs given in this paper stem from the results in [KLR73, Com94, TJ99].

This paper is organized as follows. In section 2, the concepts of identifiability, separability and uniqueness are discussed. Various representations for the linear ICA model are considered as well. In section 3, a theorem for ensuring the identifiability of mixing system in the linear ICA model is given. In section 4, theorem on separability is presented. Separability ensures that the source signals may be recovered up to some ambiguities. In section 5, the uniqueness of the ICA model is addressed. It is a very relevant concept, in particular for underdetermined source separation problems. Finally, section 6 concludes the paper and proofs of the theorems are given in an Appendix.

2. DEFINITIONS AND PROBLEM STATEMENT

A general linear instantaneous Independent Component Analysis (ICA) model may be described by the equation

$$\mathbf{X} = \mathsf{A}\mathbf{S}, \tag{1}$$

where $(\mathbf{s}_1, \ldots, \mathbf{s}_m)^T = \mathbf{S}$ are unknown real valued independent non-degenerate random variables (*sources*), A is a constant $p \times m$ unknown mixing matrix, $p \geq 2$, and $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_p)^T$ is the observed random vector (*sensors*). In general, nothing is assumed about the ranks and the number of columns, thus the model includes both the traditional linear ICA as well as the ICA with more sources than sensors (ICA with overcomplete basis [HKO01]). The couple $(\mathsf{A}, \mathbf{S})$ is called a representation of $\mathbf{X}$. Since $\mathbf{X} = \mathsf{A}\mathbf{S} = (\mathsf{A}\Lambda\mathsf{P})(\mathsf{P}^{-1}\Lambda^{-1}\mathbf{S})$ for any diagonal matrix Λ (with nonzero diagonals) and permutation matrix P, $\mathbf{X}$ can never have completely unique representation. We are interested in conditions that guarantee (at least partially) the uniqueness of the representation up to these ambiguities. Further, if any two columns, say α_i and α_j, of A are multiples of each other, i.e. $\alpha_i = a\alpha_j$ for some $a \in \mathbb{R}$, then $\mathbf{X}$ has a representation with $m-1$ source variables. Also, suppose that columns that are multiples of each other were allowed. Then if any of source variables has an infinitely divisible distribution, then $\mathbf{X}$ would have representations for any given $\hat{m} \geq m$. Infinitely divisible distribution means that the characteristic function of the random variable can be written as product of characteristic functions for any given n. Such distributions include normal, Poisson and Gamma. Therefore, we also assume that representations are reduced in the sense that no two columns in mixing matrices are multiples of each other. Notice that this means that in two representations of $\mathbf{X}$ any column of one representations is multiple of exactly one column or none columns in another representation.

It is easily shown that if $\mathbf{X} = \mathsf{A}\mathbf{S} + \mathsf{a} = \mathsf{B}\mathbf{R} + \mathsf{b}$ for some constant (deterministic) vectors a and b, then linear manifolds generated by the columns of A and B coincide and further the vector $\mathsf{a} - \mathsf{b}$ belongs to this common manifold. Therefore, $\mathrm{rank}[\mathsf{A}] = \mathrm{rank}[\mathsf{B}]$, and adding constants to the model (1) gives no additional generality. Also models of the form $\mathsf{A}\mathbf{S} + \mathsf{B}\mathbf{R}$, where $\mathbf{S}$ and $\mathbf{R}$ are independent and consist of independent variables, can be represented as $(\mathsf{A}\ \mathsf{B})(\mathbf{S}^T\ \mathbf{R}^T)^T$. Thus the usual "noisy" ICA models $\mathbf{X} = \mathsf{A}\mathbf{S} + \mathbf{N}$, where $\mathbf{N}$ is a multinormal distribution (i.e. linear transformation of independent normal variables), are just special cases of the general linear model (1).

The model (1) is defined to be

1. *identifiable* or the structure is (essentially) unique, if in every representation $(\mathsf{B}, \mathbf{R})$ of $\mathbf{X} = \mathsf{A}\mathbf{S}$, every column of A is multiple of some column of B and vice versa,
2. *unique* if the model is identifiable and further variables $\mathbf{S}$ and $\mathbf{R}$ have the same distribution for some permutation up to changes of location and scale, and
3. *separable*, if for every full row rank matrix W such that $\mathsf{W}\mathbf{X}$ has independent components, $\Lambda\mathsf{P}\mathbf{S} = \mathsf{W}\mathbf{X}$ for some diagonal matrix Λ and permutation matrix P.

Any separable model will be shown to be unique. However, the opposite is not true. In order to illustrate the definitions, consider the following examples (claims are later proved).

Example 1. *If components of $\mathbf{S}$ are i.i.d. normally distributed, then also $\Lambda\mathsf{U}\mathbf{S}$ has independent components for any orthogonal matrix U and diagonal matrix Λ. Therefore, any multinormal mixing is not identifiable.*

Example 2. *As an example of a model, which is identifiable but is not separable nor unique, consider independent non-normal variables $\mathbf{s}_k$, $k = 1, \dots 4$. Let $\mathbf{n}_1$ and $\mathbf{n}_2$ be standard normal and independent. Then also $\mathbf{n}_1 + \mathbf{n}_2$ and $\mathbf{n}_1 - \mathbf{n}_2$ are independent. Now*

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \mathbf{s}_3 + \mathbf{n}_1 \\ \mathbf{s}_4 + \mathbf{n}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{s}_1 + \mathbf{s}_3 + \mathbf{s}_4 + \mathbf{n}_1 + \mathbf{n}_2 \\ \mathbf{s}_2 + \mathbf{s}_3 - \mathbf{s}_4 + \mathbf{n}_1 - \mathbf{n}_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} \mathbf{s}_1 + \mathbf{n}_1 + \mathbf{n}_2 \\ \mathbf{s}_2 + \mathbf{n}_1 - \mathbf{n}_2 \\ \mathbf{s}_3 \\ \mathbf{s}_4 \end{pmatrix}.$$

Example 3. *Suppose random variables $\mathbf{s}_1$, $\mathbf{s}_2$, and $\mathbf{s}_3$ with non-vanishing characteristic functions are non-normal and independent. Then*

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \mathbf{s}_3 \end{pmatrix}$$

is unique but not separable.

3. IDENTIFIABILITY

In this section, the requirements for the identifiability of an ICA model are given in a form of a theorem. This theorem relaxes the requirements imposed on cumulants in [Car98, Com94, DLDMV99]. In identifiable ICA model the coefficients of the mixing matrix A may be determined from the mixture $\mathbf{X}$ alone up to scaling of columns. The fact that this is also possible in cases where we have more sources than sensors is recognized by several authors in the ICA community. Algorithms for identifying some mixtures using higher-order statistics were presented in [Car91, DLDMV99] and gradient based method in [Ama99]. It was shown in [Tal01] that the special case of two by n in the first part of theorem 3.1 is identifiable. An algorithm for estimating the mixture was presented, too. Conditions in the latter case in our theorem are the same as in the usual case of linear ICA model separation as will be seen later.

Theorem 3.1 (Identiafiability of linear ICA model). *The model (1) is identifiable, if*

(i) all source variables are non-normal, or

(ii) A is of full column rank and at most one source variable is normal .

Proof. See Appendix.

It is now recognized that mixtures in Example 2 and in Example 3 are indeed identifiable. The theorem is further illustrated in the following example.

Example 4. *In order to see why in general not a single normal variable is allowed for identifiability, consider independent non-normal variables $\mathbf{s}_1, \mathbf{s}_2$, and standard normal variables $\mathbf{n}_1$ and $\mathbf{n}_2$. Now*

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 + 2\mathbf{n}_1 \\ 2\mathbf{n}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{s}_1 + \mathbf{s}_2 + 2\mathbf{n}_1 \\ \mathbf{s}_1 + 2\mathbf{n}_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \mathbf{s}_1 + \mathbf{n}_1 + \mathbf{n}_2 \\ \mathbf{s}_2 \\ \mathbf{n}_1 - \mathbf{n}_2 \end{pmatrix}.$$

4. SEPARABILITY

Separability considers the traditional linear ICA model and recovery of the sources, and the following theorem is well-known. It was proved in [Com94, CL96] for random variables with finite second order moments. The theorem given in this section relaxes the requirements imposed on the existence of 2nd order moments. Although the theorem follows from the results in [TJ99], to our best knowledge no explicit proof of the theorem in its more general form has been presented anywhere.

Theorem 4.1 (Separability of linear ICA model). *The model (1) is separable if and only if* A *is of full column rank and at most one source variable is normal.*

Proof. See Appendix.

It is now seen that mixtures in Example 2 and in Example 3 are not separable.

It should be noted that Theorem 4.1 holds with respect to the definitions of representation and separability. For instance, it is assumed that the class of mixing matrices is the class of all real valued matrices where no columns are multiples of each other. By further restricting this class of matrices, it is well possible that the model becomes separable for some class of random variables such that more than one normal variable is allowed in mixtures. Also, allowing separation of only independent sums of source variables, the separation of that type becomes possible [CL96] for matrices that are not of full column rank.

5. UNIQUENESS

It was already seen that if the number of sources is greater than the number of sensors and the sources are non-normal, it is still possible to identify the mixing matrix from the knowledge of $\mathbf{X}$ alone, although it is not possible to recover the source $\mathbf{S}$. However, the question arises that if it were possible to determine also the distribution of $\mathbf{S}$ in such cases, i.e. if the models are unique. Then we could reconstruct the original signals $\hat{\mathbf{S}} = (\mathbf{S}_1, \ldots, \mathbf{S}_n)$, that is the sample of the random variable $\mathbf{S}$, in probabilistic sense e.g. by maximizing the likelihood of the observation given then mixing matrix A and the density of $\mathbf{S}$ (which give the likelihood function). This type of separation problem is termed *overcomplete ICA* (or underdetermined source separation). There are few approaches available, e.g. [LS00, LLS99, IH01, VEP01]), mainly in Bayesian framework. Uniqueness has also been implicitly used in the the case when all sources are assumed to be discrete [PK97, CG99]. However, to our knowledge, theoretical justification only exists for the discrete case [TJ99], and there are no other known conditions that guarantee, for instance, the uniqueness of the likelihood function. The theorem given here extends the results of [TJ99] to cases where sources are not necessarily discrete.

To state our uniqueness theorem, we introduce an operator $\odot$ on matrices defined as matrix columnwise Kronecker product $\otimes$, i.e. if $\alpha_1, \ldots, \alpha_m$ and $\beta_1, \ldots, \beta_m$ are columns of A and B respectively, then $\mathsf{A} \odot \mathsf{B} = (\alpha_1 \otimes \beta_1 \; \cdots \; \alpha_m \otimes \beta_m)$. The power $(\mathsf{A}\odot)^r \mathsf{A}$ is given naturally by $\mathsf{A} \odot \cdots \odot \mathsf{A}$ (includes r times $\odot$). For instance, denoting by A the matrix in Example 3,

$$\mathsf{A} \odot \mathsf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}. \tag{2}$$

Theorem 5.1 (Uniqueness of linear ICA model). *The model (1) is unique if any of the following assertions hold.*

(i) The model is separable.

(ii) None of the characteristic functions of source variables has a component of the form $\exp(\mathcal{P}(u))$, *where* $\mathcal{P}(u)$ *is a polynomial of degree at least* 2.

(iii) All source variables are non-normal with non-vanishing characteristic functions, and $\operatorname{rank}[\mathsf{A}\odot\mathsf{A}] = m$.

Proof. See Appendix.

Since the characteristic function of a discrete distribution does not contain an exponent factor with a polynomial of degree more than one, the assertion (ii) covers the discrete distributions. This was proved in [TJ99].

The rank of the matrix in Equation (2) is three, and it is seen by the assertion (iii) that the mixture in Example 3 is unique.

Since the number of rows in $\mathsf{A} \odot \mathsf{A}$ can be at most $p(p+1)/2$, the maximum value of m given p in (iii) is obtained from $p(p-1)/2 < m \leq p(p+1)/2$. For instance, for ten sources it may be enough to have four sensors. It is easily shown that these numbers are attainable, i.e. there exist matrices that fulfill the requirement. Further, the condition seems to be similar to the nonsingularity of a square matrix, i.e. the maximum number is usually achieved for “randomly generated” matrices.

6. CONCLUSION

We presented proofs of theorems that ensure identifiability, separability and uniqueness of linear ICA models. Identifiability is concerned with estimating the mixing system whereas separability addresses the problem of recovering the source signals. Uniqueness in relevant in case of recovering sources in underdetermined ICA models. The proofs extend the currently used conditions in ICA community to wider class of mixing models and source distributions.

A. PROOFS

For the proofs we cite the following theorem (Theorem 10.3.1 in [KLR73]). Theorem similar to part (i) was derived in [TJ99].

Theorem A.1. *Let* $(\mathsf{A}, \mathbf{S})$ *and* $(\mathsf{B}, \mathbf{R})$ *be two representations of a* p*-dimensional random vector* $\mathbf{X}$*, where* A *and* B *are constant matrices of orders* $p \times m$ *and* $p \times n$ *respectively, and* $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_m)^T$ *and* $\mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_n)^T$ *are random vectors with independent components. Then the following assertions hold.*

(i) If the i*:th column of* A *is not multiple of any column of* B*, then* $\mathbf{s}_i$ *is normal.*

(ii) If the i*:th column of* A *is multiple of the* j*:th column of* B*, then the logarithms of the characteristic functions of* $\mathbf{s}_i$ *and* $\mathbf{r}_j$ *differ by a polynomial in a neighborhood of the origin.*

Notice that the proof of Theorem 10.3.1(ii) in [KLR73], as stated, is not correct. However, it is easily fixed by noting that the same construction as used in the proof of Lemma 10.2.2. produces matrices C_1 and C_2 such that the first column of C_1 is multiple of the first column of C_2, but not of any other columns. Thus Lemma 10.2.5. can be used to prove the theorem.

Remark that it does *not* follow from the second assertion that $\mathbf{s}_i = \mathbf{r}_j + \mathbf{n}$ for some independent normal variable $\mathbf{n}$ (where constants are considered as degenerate normals). In fact, A.A. Goldberg (c.f. [Luk83]) has shown that for any polynomial $\mathcal{P}(t)$ such that $\mathcal{P}(0) = 0$, there exist characteristic functions $\psi_1(t)$ and $\psi_2(t)$ such that $\psi_1(t) = \psi_2(t) \exp\big(\mathcal{P}(t)\big)$ for all $t \in \mathbb{R}$. We are not aware of any classification of these type of random variables.

Also the following simple lemma is needed.

Lemma A.1. *If random variables* $\mathbf{r}_1$ *and* $\mathbf{r}_2$ *are independent, then* $\mathbf{r}_1$ *is independent of* $\mathbf{r}_1 + \mathbf{r}_2$ *if and only if* $\mathbf{r}_1$ *is degenerate.*

Proof. The "if" part is obvious. To the other direction, note that the joint characteristic function of $\mathbf{r}_1$ and $\mathbf{r}_1 + \mathbf{r}_2$ is

$$\psi_{\mathbf{r}_1, \mathbf{r}_1 + \mathbf{r}_2}(s, t) = \psi_{\mathbf{r}_1}(s + t)\psi_{\mathbf{r}_2}(t) \tag{3}$$

by independence of $\mathbf{r}_1$ and $\mathbf{r}_2$. Since $\mathbf{r}_1$ and $\mathbf{r}_1 + \mathbf{r}_2$ are independent

$$\psi_{\mathbf{r}_1, \mathbf{r}_1 + \mathbf{r}_2}(s, t) = \psi_{\mathbf{r}_1}(s)\psi_{\mathbf{r}_1 + \mathbf{r}_2}(t) = \psi_{\mathbf{r}_1}(s)\psi_{\mathbf{r}_1}(t)\psi_{\mathbf{r}_2}(t). \tag{4}$$

Consider degenerate distribution as Gaussian with zero variance. By the Darmois–Skitovich theorem, $\mathbf{r}_1$ is Gaussian, and thus $\psi_{\mathbf{r}_1}(s) = \exp(i\mu s - \frac{1}{2}\sigma^2 s^2)$. Combining the equations (3) and (4) then gives $\exp(-\sigma^2 ts)\psi_{\mathbf{r}_2}(t) = \psi_{\mathbf{r}_2}(t)$, which is only possible if $\sigma^2 = 0$, i.e. $\mathbf{r}_1$ is degenerate. □

Proof of Theorem 3.1.

(i) Since there are no normal variables, by Theorem A.1(i) every column has to be multiple of some column in another representation, i.e. the model is identifiable.

(ii) By Theorem A.1(i) the columns corresponding to non-normal variables are identifiable in two representations $(\mathsf{A}, \mathbf{S})$ and $(\mathsf{B}, \mathbf{R})$ of $\mathbf{X}$. Since $\operatorname{rank}\big[\mathsf{A}\big] = \operatorname{rank}\big[\mathsf{B}\big] = m$ there must be either $m-1$ or m such columns in both representations. Since every column is multiple of exactly one or none in the other, representation must *simultaneously* have exactly one or none normal variables. In the latter case there is nothing to prove by part (i), so we take without a loss of generality that in both representations the first $m-1$ columns are the same.

Now $\mathsf{A}^{\#}\mathbf{X} = \mathbf{S} = \mathsf{A}^{\#}\mathsf{B}\mathbf{R} = \left(\begin{smallmatrix} \mathsf{I}_{m-1} \\ \mathsf{0}_{m-1}^T \end{smallmatrix}\ \mathsf{A}^{\#}\beta_m\right)\mathbf{R}$, where $\#$ denotes the pseudoinverse, i.e. $A^{\#} = (A^T A)^{-1} A^T$, β_m is the last column of B, I_m denotes the $m \times m$ identity matrix and $\mathsf{0}_m$ is the vector of m zeros. Since sources are non-degenerate, the last component b of $\mathsf{A}^{\#}\beta_m$ is nonzero. Since each component of $\mathbf{S}$ is independent, then by Lemma A.1 the normal component $\mathbf{r}_m$ of $\mathbf{R}$ is degenerate if any of $m-1$ first entries of $\mathsf{A}^{\#}\beta_m$ is nonzero. Therefore, $\mathbf{S} = (\mathbf{r}_1, \ldots, \mathbf{r}_{m-1}, b\mathbf{r}_m)^T$. Now $\mathbf{X} = (\mathsf{A}_{m-1}\ \alpha_m)(\mathbf{r}_1, \ldots, \mathbf{r}_{m-1}, b\mathbf{r}_m)^T = (\mathsf{A}_{m-1}\ \beta_m)(\mathbf{r}_1, \ldots, \mathbf{r}_m)^T$, where A_{m-1} denotes the first $m-1$ columns of A and α_m is the last. Therefore, $\alpha_m b \mathbf{r}_m = \beta_m \mathbf{r}_m$, which is only possible if $b\alpha_m = \beta_m$, i.e. the last columns are multiples.

□

Proof of Theorem 4.1. Suppose a matrix W of full column rank separates the model (1), i.e. $\mathbf{R} = \mathsf{W}\mathbf{X}$ has independent components. Then $\mathbf{X} = \mathsf{W}^{\#}\mathbf{R} = \mathsf{A}\mathbf{S}$ has two representations. Then by Theorem 3.1(ii) $\mathsf{W}^{\#} = \mathsf{A}\Lambda\mathsf{P}$ for some diagonal matrix Λ and permutation matrix P. Therefore $\mathbf{S} = \mathsf{A}^{\#}\mathbf{X} = \Lambda\mathsf{P}\mathbf{R}$.

To the other direction, if A is not full column rank, there can not exist a mapping W such that $\operatorname{rank}\big[\mathsf{W}\mathsf{A}\big] = m$, where m is the dimension of the linear manifold of $\mathbf{S}$. If two normal variables are allowed, then it is seen by using the construction of Example 1 that for any mixing there exists infinitely separation matrices W such that $\mathsf{W}\mathbf{X}$ has independent components but different distribution for each W. □

For the proof of the third case in the uniqueness theorem, the following theorem (Theorem $A.3.3$ in [KLR73]) is needed.

Theorem A.2. *Let* $(\mathsf{A}, \mathbf{S})$ *be a representation of a* p*-dimensional random vector* $\mathbf{X}$ *(with non-vanishing characteristic function), where* A *is a known* $p \times m$ *matrix and* r *is the integer such that* $\operatorname{rank}\big[(\mathsf{A}\odot)^r\mathsf{A}\big] = m > \operatorname{rank}\big[(\mathsf{A}\odot)^{r-1}\mathsf{A}\big]$*. Then the characteristic function of each* $\mathbf{s}_i$ *is determined up to a factor* $\exp\big(\mathcal{P}_{i,r}(t)\big)$*, where* $\mathcal{P}_{i,r}(t)$ *is a polynomial of degree at most* r*.*

Proof of Theorem 5.1.

(i) The identifiability follows from Theorem 4.1 and Theorem 3.1(ii). It is also seen from the proof of Theorem 3.1(ii) that the distribution of $\mathbf{S}$ is unique up to scaling and permutation. Thus the model is unique.

(ii) There can be no normal variables, and therefore the model is identifiable by Theorem 3.1(i). Now the logarithms of the characteristic functions of the source variables in two representations differ by a polynomial by Theorem A.1(ii). However, by the assumption this polynomial can be at most degree 1, that is, the source variables have the same distribution up to changes of location and scale.

(iii) By Theorem 3.1(i) the model is identifiable, and the uniqueness now follows from Theorem A.2 by choosing $r = 1$.

□

B. REFERENCES

[Ama99] S. Amari. Natural gradient learning for over- and under-complete bases in ICA. *Neural Computation*, 11(8):1875–1883, November 1999.

[Car91] J-F. Cardoso. Super-symmetric decomposition of the fourth-order cumulant tensor. blind identification of more sources than sensors. In *Proc. Int. Conf. on Acustics, Speech, and Signal Processing*, pages 3109–3112, 1991.

[Car98] J-F. Cardoso. Blind signal separation: Statistical principles. *Proceedings of the IEEE*, 86(10):2009–2025, October 1998.

[CG99] P. Comon and O. Grellier. Non-linear inversion of underdetermined mixtures. In *Proc. First Int. Workshop on ICA and Signal Separation*, pages 461–465, Aussois, France, January 1999.

[CL96] X-R. Cao and R-w. Liu. General approach to blind source separation. *IEEE Trans. on Signal Processing*, 44(3):562–571, March 1996.

[Com94] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, April 1994.

[DLDMV99] L. De Lathauwer, B. De Moor, and J. Vandewalle. ICA techiniques for more sources than sensors. In *Proc. of the IEEE SP Workshop on Higher-Order Statistics*, pages 121–124, Caesarea, Israel, June 1999.

[HKO01] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, New York, 2001.

[IH01] M. Inki and A. Hyvärinen. Two methods for estimating overcomplete independent component bases. In *Proc. of 3rd Int. Conf. on ICA and Blind Signal Separation*, pages 343–348, San Diego, California, December 2001.

[KLR73] A.M. Kagan, Yu.V. Linnik, and C.R. Rao. *Characterization Problems in Mathematical Statistics*. Probability and Mathematical Statistics. John Wiley & Sons, 1973.

[LLS99] T-W. Lee, M.S. Lewicki, and T.J. Sejnowski. Blind source separation of more sources than mixtures using overcomplete representations. *IEEE Signal Processing Letters*, 6(4):87–90, April 1999.

[LS00] M.S. Lewicki and T.J. Sejnowski. Learning overcomplete representations. *Neural Computations*, 12(2):337–363, February 2000.

[Luk83] E. Lukacs. *Developments in Characteristic Function Theory*. Griffin, 1983.

[PK97] P. Pajunen and J. Karhunen. Blind separation of binary sources with less sensors than sources. In *Proc. of 1997 Int. Conf. on Neural Networks*, pages 1994–1997, Houston, Texas, June 1997.

[Tal01] A. Taleb. An algorithm for the blind identification of n independent signals with 2 sensors. In *Proc. Int. Symp. on Signal Processing and its Applications*, pages 5–8, Kuala Lumpur, Malaysia, August 2001.

[TJ99] A. Taleb and C. Jutten. On underdetermined source separation. In *Proc. 24th Int. Conf. on Acustics, Speech, and Signal Processing*, volume 3, pages 1445–1448, Phoenix, Arizona, March 1999.

[VEP01] L. Vielva, D. Erdoğmuş, and J.C. Príncipe. Underdetermined blind source separation using a probabilistic source sparcity model. In *Proc. of 3rd Int. Conf. on ICA and Blind Signal Separation*, pages 675–679, San Diego, California, December 2001.