Subspace Identification Through Blind Source Separation

Moritz Grosse-Wentrup, *Student Member, IEEE,* and Martin Buss, *Member, IEEE*

***Abstract*—Given a linear and instantaneous mixture model, we prove that for blind source separation (BSS) algorithms based on mutual information, only sources with non-Gaussian distribution are consistently reconstructed independent of initial conditions. This allows the identification of non-Gaussian sources and consequently the identification of signal and noise subspaces through BSS. The results are illustrated with a simple example, and the implications for a variety of signal processing applications, such as denoising and model identification, are discussed.**

***Index Terms*—Blind source separation (BSS), consistency, denoising, identifiability, independent component (IC) analysis, independent components, model identification, noise, stability, subspace.**

I. Introduction

In real-world applications, algorithms for blind source separation (BSS) return differing sets of independent components (ICs) if they are applied to the same data set with varying initial conditions. It is generally agreed that only ICs that can be consistently reconstructed independent of initial conditions should be regarded as meaningful and that the signal-to-noise ratio (SNR) of the reconstructed data can be increased if all inconsistent ICs are discarded (see, e.g., [1]–[6]). While only consistent ICs are regarded as meaningful, to our knowledge, no interpretation of the consistency of ICs and no mathematical proof of the conditions for consistency have been published so far.

In this letter, we prove that for BSS algorithms based on mutual information, such as the widely used Infomax algorithm [7], only sources with non-Gaussian distribution are consistently reconstructed. Sources with Gaussian distribution, on the other hand, are arbitrarily mixed together, resulting in a reconstruction that is dependent on the initial conditions of the algorithm. Multiple iterations of BSS algorithms based on mutual information thus allow the identification of sources with non-Gaussian distribution. Reprojecting these on the observation space results in a representation of the original measurements with all Gaussian sources removed. This can also be interpreted as a signal and noise subspace identification. The columns of the reconstructed mixing matrix associated with consistent ICs span the signal subspace. Algorithms for BSS based on mutual information are thus capable of blindly identifying the rank and a basis of the signal subspace. This is of interest in a variety of signal processing applications, such as denoising and model identification, in which the signal subspace is not known *a priori*.

This letter is organized as follows. In Section II, we will introduce the instantaneous mixture model and show that BSS algorithms based on mutual information are capable of reconstructing non-Gaussian sources, even if more than one Gaussian source is present in the data. In Section III, we will prove that only non-Gaussian sources are consistent and show how the non-Gaussian sources can be identified. We will then illustrate the results with a simple example in Section IV and discuss the implications for a variety of signal processing applications in the conclusion.

Manuscript received August 8, 2005; revised September 25, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yue (Joseph) Wang.

The authors are with the Institute of Automatic Control Engineering, Technical University Munich, 80290 Munich, Germany (e-mail: MoritzGW@ieee.org; M.Buss@ieee.org).

Digital Object Identifier 10.1109/LSP.2005.861581

II. BSS by Minimization of Mutual Information

The linear BSS model is described by

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

with stationary random variables (r.v.s) $\boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{s} \in \mathbb{R}^N$ and the instantaneous, full-rank mixing matrix $A \in \mathbb{R}^{N\times N}$. It is thus assumed that there are as many sensors as sources. Furthermore

$$p(\boldsymbol{s}) = \prod_{i=1}^{N} p(s_i) \tag{2}$$

i.e., the s_i, $i = 1,\ldots,N$ are assumed to be statistically independent. Furthermore, $M < N$ sources are assumed to have Gaussian distribution. Without loss of generality, it is assumed that these are the sources s_i, $i = 1,\ldots,M$. Also without loss of generality, all sources are assumed to have zero mean.

The original sources are reconstructed from the measurements through a linear demixing matrix

$$\boldsymbol{y} = B\boldsymbol{x} \tag{3}$$

with $\boldsymbol{y} \in \mathbb{R}^N$ and $B \in \mathbb{R}^{N\times N}$. If $\boldsymbol{x}$ is sphered with the sphering matrix Γ such that $Cov(\Gamma\boldsymbol{x}\boldsymbol{x}^T\Gamma^T) = I$, the solution space is restricted to orthogonal matrices. We can then reformulate the problem by defining $C := B\Gamma A$ as

$$\boldsymbol{y} = B\Gamma\boldsymbol{x} = B\Gamma A\boldsymbol{s} = C\boldsymbol{s} \tag{4}$$

with C also orthogonal.

For algorithms based on minimizing the mutual information between the recovered sources y_i, C is found by minimizing

$$I(\boldsymbol{y}) = \sum_{i=1}^{N} H(y_i) - H(\boldsymbol{y}) \tag{5}$$

with $H(y_i) = -\int_{-\infty}^{\infty} p_{y_i}(u)\log(p_{y_i}(u))du$ as the differential entropy. This leads to

$$\min_C I(\boldsymbol{y}) = \min_C \left\{ \sum_{i=1}^{N} H(y_i) - H(\boldsymbol{y}) \right\} = \min_C \left\{ \sum_{i=1}^{N} H(y_i) - \log(|\det C|) - H(\boldsymbol{s}) \right\}. \tag{6}$$

Since $H(\boldsymbol{s})$ is independent of C and $\log(|\det C|) = 0$ because of the constraints on C, (6) simplifies to

$$\min_C \sum_{i=1}^{N} H(y_i). \tag{7}$$

Based on [8], we will now show that by finding a solution of (7), the original non-Gaussian sources can be reconstructed in spite of the presence of multiple Gaussian sources. With $F(C) := \sum_{i=1}^{N} H(y_i)$, the gradient of $F(C)$ under the orthogonality constraint becomes [9]

$$\nabla_m F(C) = \nabla F(C) - C\nabla F(C)^T C. \tag{8}$$

Since $CC^T = I$, solutions of (7) are given by

$$\nabla F(C)C^T = C\nabla F(C)^T. \tag{9}$$

Since

$$y_i = c_{i,1}s_1 + \cdots + c_{i,N}s_N \tag{10}$$

the distribution of y_i is given by

$$p_{y_i}(u) = p_{y_i}(c_{i,1}s_1 + \cdots + c_{i,N}s_N) = \frac{1}{c_{i,1}} p_{s_1}\left(\frac{u_1}{c_{i,1}}\right) * \cdots * \frac{1}{c_{i,N}} p_{s_N}\left(\frac{u_N}{c_{N,1}}\right). \tag{11}$$

Denoting $h(\boldsymbol{c}_i) := H(y_i)$ with $\boldsymbol{c}_i$, the ith row of C (9) becomes

$$\nabla h(\boldsymbol{c}_k) \cdot \boldsymbol{c}_l^T = \nabla h(\boldsymbol{c}_l) \cdot \boldsymbol{c}_k^T \tag{12}$$

for $k, l = 1, \ldots, N$, $k \neq l$. Combining

$$\frac{\partial H(y_i)}{\partial c_{i,j}} = -\int_{-\infty}^{\infty} (\log p_{y_i}(u) + 1) \frac{\partial p_{y_i}(u)}{\partial c_{i,j}} du \tag{13}$$

and (12) then results in

$$\int_{-\infty}^{\infty} (\log p_{y_k}(u) + 1) \times \left[\frac{\partial p_{y_k}(u)}{\partial c_{k,1}} \cdot c_{l,1} + \cdots + \frac{\partial p_{y_k}(u)}{\partial c_{k,N}} \cdot c_{l,N} \right] du = \int_{-\infty}^{\infty} (\log p_{y_l}(u) + 1) \times \left[\frac{\partial p_{y_l}(u)}{\partial c_{l,1}} \cdot c_{k,1} + \cdots + \frac{\partial p_{y_l}(u)}{\partial c_{l,N}} \cdot c_{k,N} \right] du \tag{14}$$

for $k, l = 1, \ldots, N$, $k \neq l$, which is fulfilled if

$$\frac{\partial p_{y_k}(u)}{\partial c_{k,1}} \cdot c_{l,1} + \cdots + \frac{\partial p_{y_k}(u)}{\partial c_{k,N}} \cdot c_{l,N} = 0 \tag{15}$$

for $k, l = 1, \ldots, N$, $k \neq l$.

The analysis of (15) can be simplified by transforming it into the frequency domain. With

$$\varphi_{y_i}(\omega) := \int_{-\infty}^{\infty} p_{y_i}(u) \exp^{-j\omega u} du \tag{16}$$

the characteristic function of p_{y_i}, (11) becomes

$$\varphi_{y_i}(\omega) = \varphi_{s_i}(c_{i,1}\omega) \times \cdots \times \varphi_{s_N}(c_{i,N}\omega). \tag{17}$$

Substituting (17) in (15) and dividing by $\varphi_{s_1}(c_{k,1}\omega) \times \cdots \times \varphi_{s_N}(c_{k,N}\omega)$ results in

$$\frac{\omega\varphi'_{s_1}(c_{k,1}\omega) \cdot c_{l,1}}{\varphi_{s_1}(c_{k,1}\omega)} + \cdots + \frac{\omega\varphi'_{s_N}(c_{k,N}\omega) \cdot c_{l,N}}{\varphi_{s_N}(c_{k,N}\omega)} = 0 \tag{18}$$

for $k, l = 1, \ldots, N$, $k \neq l$.

Now only if s_i is Gaussian does it hold that

$$\varphi'_{s_i}(\alpha\omega) = -\alpha\omega\varphi_{s_i}(\alpha\omega). \tag{19}$$

Since the first M sources are assumed to be Gaussian, (18) simplifies to

$$-\omega^2(c_{k,1}c_{l,1} + \cdots + c_{k,M}c_{l,M}) + \frac{\omega\varphi'_{s_{M+1}}(c_{k,M+1}\omega) \cdot c_{l,M+1}}{\varphi_{s_{M+1}}(c_{k,M+1}\omega)} + \cdots + \frac{\omega\varphi'_{s_N}(c_{k,N}\omega) \cdot c_{l,N}}{\varphi_{s_N}(c_{k,N}\omega)} = 0. \tag{20}$$

Since $\varphi'_{s_i}(c_{i,j}\omega)|_{c_{i,j}=0} = 0$ because all sources have zero mean, any matrix of the form

$$C = \begin{pmatrix} Q & 0 \\ 0 & P \end{pmatrix} \tag{21}$$

with $Q \in \mathbb{R}^{M \times M}$ any orthogonal matrix and $P \in \mathbb{R}^{(N-M)\times(N-M)}$ any permutation matrix fulfills (20). The set of resulting unmixing matrices is then given by

$$B = CA^{-1}\Gamma^{-1} = \begin{pmatrix} Q & 0 \\ 0 & P \end{pmatrix} A^{-1}\Gamma^{-1} \tag{22}$$

and the recovered sources are given by

$$\boldsymbol{y} = B\Gamma\boldsymbol{x} = B\Gamma A\boldsymbol{s} = CA^{-1}\Gamma^{-1}\Gamma A\boldsymbol{s} = \begin{pmatrix} Q & 0 \\ 0 & P \end{pmatrix} \boldsymbol{s}. \tag{23}$$

Hence, any matrix that returns a permuted version of the original non-Gaussian sources and a mixture of the Gaussian sources is a solution of (7). Furthermore, it is easy to check that for solutions of the above form $I(\boldsymbol{y}) = 0$, i.e., the mutual information between the recovered sources becomes zero.

This shows that algorithms based on the minimization of mutual information are indeed capable of recovering the original non-Gaussian sources, even in the presence of multiple Gaussian sources. However, it is not clear whether every solution of (7)—and thus every solution of (20)—has to be of the form in (21).

III. Identifying Non-Gaussian Sources

We will now prove that only non-Gaussian sources are consistent and show how these can be identified. Consider two different solutions B_1 and B_2 of (7). Inverting B_1 and B_2 leads to two pairs $(A_1, \boldsymbol{s}_1)$ and $(A_2, \boldsymbol{s}_2)$, both of which are representations of $\boldsymbol{x}$, i.e.,

$$\boldsymbol{x} = A_1 \boldsymbol{s}_1 = A_2 \boldsymbol{s}_2. \tag{24}$$

Then, the following holds.

Theorem 1: If column $\boldsymbol{a}_1^j$ of A_1 is not linearly dependent on any column of A_2, then s_1^j is Gaussian. This can be shown through the following argument, which is based on [10].

Proof: Consider the two representations $(\boldsymbol{A}_1, \boldsymbol{s}_1)$ and $(\boldsymbol{A}_2, \boldsymbol{s}_2)$ of $\boldsymbol{x}$. The characteristic function of $\boldsymbol{x}$ is then given by

$$\begin{aligned}\varphi_{\boldsymbol{x}}(\boldsymbol{\omega}) &= E\left\{\exp(j\boldsymbol{\omega}^T \boldsymbol{x})\right\}\\ &= E\left\{\exp(j\boldsymbol{\omega}^T \boldsymbol{A}_1 \boldsymbol{s}_1)\right\}\\ &= E\left\{\exp(j\boldsymbol{\omega}^T \boldsymbol{A}_2 \boldsymbol{s}_2)\right\}.\end{aligned} \tag{25}$$

Since the mutual information has no local minima [11], for any solution obtained through (7), the mutual information between the elements of the recovered sources $I(\boldsymbol{s}_{1/2}) = 0$. It is a well-known fact that the mutual information between the elements of a r.v. is zero if and only if the elements of the r.v. are independent. Consequently, the elements of $\boldsymbol{s}_{1/2}$ are independent, and we can write

$$\prod_{i=1}^{N} \varphi_{s_1^i}\left(\left(\boldsymbol{a}_1^i\right)^T \boldsymbol{\omega}\right) = \prod_{i=1}^{N} \varphi_{s_2^i}\left(\left(\boldsymbol{a}_2^i\right)^T \boldsymbol{\omega}\right) \tag{26}$$

with a_1^i as the ith column of A_1. With the definition of the second characteristic functions $\psi_x = \log(\varphi_x)$, (26) becomes

$$\sum_{i=1}^{N} \psi_{s_1^i}\left(\left(\boldsymbol{a}_1^i\right)^T \boldsymbol{\omega}\right) = \sum_{i=1}^{N} \psi_{s_2^i}\left(\left(\boldsymbol{a}_2^i\right)^T \boldsymbol{\omega}\right). \tag{27}$$

Assume that $\tilde{p}$ columns of $\boldsymbol{A}_1$ are linearly dependent on $\tilde{p}$ columns of $\boldsymbol{A}_2$. Without loss of generality, assume furthermore that these are the first $\tilde{p}$ columns, i.e., $\alpha_i \boldsymbol{a}_1^i = \boldsymbol{a}_2^i$, $i = 1, \ldots, \tilde{p}$, $\alpha_i \in \mathbb{R}$. Equation (27) then becomes

$$\begin{aligned}&\sum_{i=1}^{\tilde{p}} \psi_{s_1^i}\left(\left(\boldsymbol{a}_1^i\right)^T \boldsymbol{\omega}\right) + \sum_{i=\tilde{p}+1}^{N} \psi_{s_1^i}\left(\left(\boldsymbol{a}_1^i\right)^T \boldsymbol{\omega}\right)\\ &\quad = \sum_{i=1}^{\tilde{p}} \psi_{s_2^i}\left(\alpha\left(\boldsymbol{a}_1^i\right)^T \boldsymbol{\omega}\right) + \sum_{i=\tilde{p}+1}^{N} \psi_{s_2^i}\left(\left(\boldsymbol{a}_2^i\right)^T \boldsymbol{\omega}\right).\end{aligned} \tag{28}$$

Now in (28), replace $\boldsymbol{\omega}$ by $\boldsymbol{\omega} + \boldsymbol{\delta}_N$ such that $\boldsymbol{a}_1^N \boldsymbol{\delta} = 0$. Subtracting this from (28) results in

$$\begin{aligned}&\sum_{i=1}^{\tilde{p}} \Delta_{\epsilon_{i,N}} \psi_{s_1^i}\left(\left(\boldsymbol{a}_1^i\right)^T \boldsymbol{\omega}\right) + \sum_{i=\tilde{p}+1}^{N-1} \Delta_{\epsilon_{i,N}} \psi_{s_1^i}\left(\left(\boldsymbol{a}_1^i\right)^T \boldsymbol{\omega}\right)\\ &\quad = \sum_{i=1}^{\tilde{p}} \Delta_{\epsilon_{i,N}} \psi_{s_2^i}\left(\alpha\left(\boldsymbol{a}_1^i\right)^T \boldsymbol{\omega}\right)\\ &\qquad + \sum_{i=\tilde{p}+1}^{N} \Delta_{\xi_{i,N}} \psi_{s_2^i}\left(\left(\boldsymbol{a}_2^i\right)^T \boldsymbol{\omega}\right)\end{aligned} \tag{29}$$

with the finite difference operator

$$\Delta_h f(x) = f(x+h) - f(x) \tag{30}$$

and

$$\epsilon_{i,k} = \left(\boldsymbol{a}_1^i\right)^T \boldsymbol{\delta}_k \tag{31}$$

$$\xi_{i,k} = \left(\boldsymbol{a}_2^i\right)^T \boldsymbol{\delta}_k. \tag{32}$$

In (29), $\psi_{s_1^N}$ disappeared due to $\boldsymbol{a}_1^N \boldsymbol{\delta} = 0$. By induction, we can now eliminate all $\psi_{s_1^i}$ and $\psi_{s_2^i}$, except for $i = l$.

We thus obtain

$$\Delta_{\xi_{l,2N-1}}, \ldots, \Delta_{\xi_{l,1}} \psi_{s_2^l}\left(\left(\boldsymbol{a}_2^l\right)^T \boldsymbol{\omega}\right) = 0 \tag{33}$$

if $l = \tilde{p}+1, \ldots, N$, and

$$\Delta_{\epsilon_{l,2N-1}}, \ldots, \Delta_{\epsilon_{l,1}} \left[\psi_{s_1^l}\left(\left(\boldsymbol{a}_1^l\right)^T \boldsymbol{\omega}\right) - \psi_{s_2^l}\left(\alpha\left(\boldsymbol{a}_1^l\right)^T \boldsymbol{\omega}\right)\right] = 0 \tag{34}$$

if $l = 1, \ldots, \tilde{p}$.

Now the theorem of Marcinkiewicz [12] states that any random variable with only a finite number of nonzero cumulants has a Gaussian distribution. We can thus conclude the following.

- s_2^l, $l = \tilde{p}+1, \ldots, N$ are Gaussian.
- $\left[\psi_{s_1^l}\left(\left(\boldsymbol{a}_1^l\right)^T \boldsymbol{\omega}\right) - \psi_{s_2^l}\left(\alpha\left(\boldsymbol{a}_1^l\right)^T \boldsymbol{\omega}\right)\right]$, $l = 1, \ldots, \tilde{p}$ are second-order polynomials. However, nothing can be said about the distributions of s_1^l and s_2^l for $l = 1, \ldots, \tilde{p}$, except that they are not Gaussian.

■

This shows that only non-Gaussian sources are consistent. By running a BSS algorithm based on mutual information repeatedly on the same data set with different initial conditions and checking linear dependencies between the columns of the reconstructed mixing matrices, we can thus differentiate Gaussian and non-Gaussian sources.

IV. EXAMPLE

We will now illustrate the results of the previous two sections with a simple example. Consider the case $N = 3$ with one non-Gaussian source with sub-Gaussian distribution

$$s_1 = \sin t, \quad t \in \left[0, \frac{3}{2}\pi\right] \tag{35}$$

and two Gaussian sources with zero mean and variance one

$$s_2, s_3 \sim N(0,1) \tag{36}$$

each sampled with 5000 data points. Any mean of the sources can be subtracted from the measurements to make the sources zero mean, and any scaling can be arbitrarily traded between the sources and the mixing matrix. Without loss of generality, the Gaussian sources can thus be chosen to have zero mean and unit variance.

The sources are mixed according to

$$\boldsymbol{x} = A\left[s_1, s_2, s_3\right]^T \tag{37}$$

with a randomly generated full-rank nonorthogonal matrix

$$A = \begin{pmatrix} -0.1735 & 0.7240 & -0.1545\\ 0.3621 & 0.4088 & 0.7137\\ 0.9158 & -0.5556 & 0.6832 \end{pmatrix} \tag{38}$$

and the original sources are then reconstructed through

$$\boldsymbol{y} = B\boldsymbol{x} \tag{39}$$

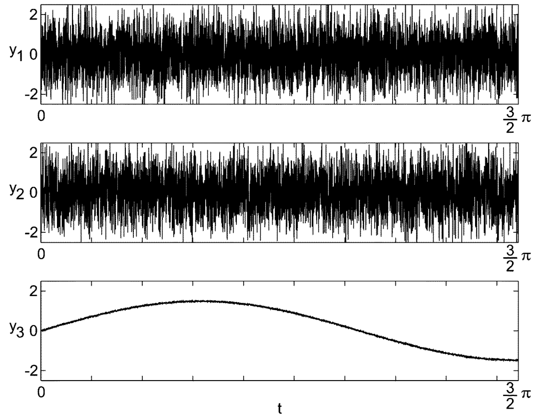

Fig. 1. Reconstructed sources.

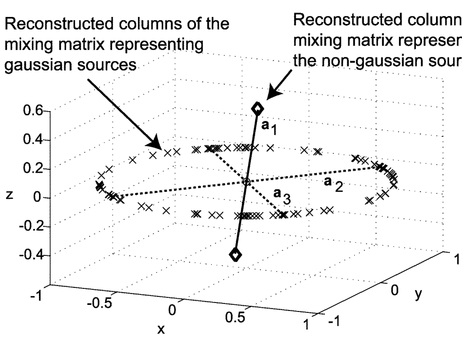

Fig. 2. Original ($\boldsymbol{a}_i, i = 1, \ldots, 3$) and reconstructed columns of the mixing matrix.

with B obtained by minimizing the mutual information of the elements of $\boldsymbol{y}$ with the extended Infomax algorithm [7]. The reconstructed signals are shown in Fig. 1 with normalized variance to remove scaling indeterminacies. As can be seen in the third panel, signal s_1 is reconstructed despite the presence of two sources with Gaussian distribution, which confirms the result of Section II.

We then performed 50 reconstructions of $\boldsymbol{y}$ using the extended Infomax algorithm with uniformly distributed initial conditions B_0. Inverting the resulting unmixing matrices, we obtained the representations $(A_i, \boldsymbol{s}_i)$, $i = 1, \ldots, 50$ of $\boldsymbol{x}$. Normalizing the columns of all matrices A_i to remove scaling indeterminacies and plotting these together with the original columns of (38) results in Fig. 2. As expected from the results of Section III, the column $\boldsymbol{a}_1$ of the non-Gaussian source is consistently reconstructed, while the reconstructed columns associated with the Gaussian sources are randomly distributed in the subspace spanned by the two columns $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$ of (38) associated with the Gaussian sources s_2 and s_3. Note that for better visualization, the data have been rotated such that the subspace spanned by $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$ coincides with the xy-plane.

V. Conclusion

We can summarize the results as follows. Algorithms for BSS based on minimizing the mutual information can reconstruct the original non-Gaussian sources, even in the presence of multiple Gaussian sources. Furthermore, for any two unmixing matrices B_1 and B_2 that minimize the mutual information of the recovered sources and that lead to the two representations $(\boldsymbol{A}_1, \boldsymbol{s}_1)$ and $(\boldsymbol{A}_2, \boldsymbol{s}_2)$ of $\boldsymbol{x}$, the following hold.

- If column $\boldsymbol{a}_1^l$ of $\boldsymbol{A}_1$ is linearly dependent on some column in $\boldsymbol{A}_2$, s_1^l has a non-Gaussian distribution.
- If column $\boldsymbol{a}_1^{\tilde{l}}$ in $\boldsymbol{A}_1$ is not linearly dependent on any column in $\boldsymbol{A}_2$, $s_1^{\tilde{l}}$ has a Gaussian distribution.

Hence, checking linear dependencies between the columns of any two solutions of BSS algorithms that minimize the mutual information between the recovered sources allows the identification of sources with Gaussian distribution.

For BSS based on mutual information, it thus turns out that ICs regarded as meaningful because of their consistency are ICs with a non-Gaussian distribution. Besides the implications this has for the interpretation of ICs in a variety of signal processing applications, these results are also of interest in the context of denoising and model identification. Excluding the nonconsistent and reprojecting the consistent ICs onto the observation space leads to a representation of the measurements $\boldsymbol{x}$ with all Gaussian sources removed. Since subspace identification based on BSS is not restricted to an orthogonal basis, a higher SNR than with denoising methods based on eigenvalue decomposition can be expected. Finally, since the number of consistently reconstructed ICs equals the rank of the signal subspace, the presented results offer a new approach for model identification, e.g., for estimating the number of sources in EEG/MEG source reconstruction.

References

[1] J. Himberg, A. Hyvaerinen, and F. Esposito, "Validating the independent components of neuroimaging time series via clustering and visualization," *Neuroimage*, vol. 22, pp. 1214–1222, 2004.

[2] J. Anemueller, T. Sejnowski, and S. Makeig, "Reliable measurement of cortical flow patterns using complex independent component analysis of electroencephalographic signals," *Independent Compon. Anal. Blind Signal Separation*, vol. 3195, pp. 1009–1016, 2004.

[3] F. Meinecke, A. Ziehe, M. Kawanabe, and K. Mueller, "A resampling approach to estimate the stability of one-dimensional or multidimensional independent components," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 12, pp. 1514–1525, Dec. 2002.

[4] D. Kreil and D. MacKay, "Reproducibility assessment of independent component analysis of expression ratios from DNA microarrays," *Comput. Funct. Genom.*, vol. 4, pp. 300–317, 2003.

[5] F. Westad and M. Kermit, "Cross validation and uncertainty estimates in independent component analysis," *Anal. Chimica Acta*, vol. 490, pp. 341–354, 2003.

[6] K. Mueller, R. Vigario, F. Meinecke, and A. Ziehe, "Blind source separation techniques for decomposing event-related brain signals," *Int. J. Bifurc. Chaos*, vol. 14, no. 2, pp. 773–791, 2004.

[7] A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1004–1034, 1995.

[8] R. Boscolo, H. Pan, and V. P. Roychowdhury, "Beyond Comon's identifiability theorem for independent component analysis," in *Proc. Artificial Neural Netw.*, vol. 2415, Lecture Notes in Computer Science, 2002, pp. 1119–1124.

[9] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.

[10] A. Taleb and C. Jutten, "On underdetermined source separation," in *Proc. IEEE Acoustics, Speech, Signal Processing Conf.*, vol. 3, Mar. 15–19, 1999, pp. 1445–1448.

[11] M. Babaie-Zadeh and C. Jutten, "A general approach for mutual information minimization and its application to blind source separation," *Signal Process.*, vol. 85, pp. 975–995, May 2005.

[12] J. Marcinkiewicz, "Sur une propriété de la loi de gauss," *Math. Zeit.*, vol. 44, pp. 612–618, 1938.