NON-SPARSE APPROACH TO UNDERDETERMINED BLIND SIGNAL ESTIMATION

L.C. Khor, W.L. Woo and S.S. Dlay

School of Electrical, Electronic and Computer Engineering
University of Newcastle upon Tyne, United Kingdom

ABSTRACT

Conventional assumptions of square mixing matrix and negligible noise adopted in blind signal separation do not always correspond with real applications. Signal detection from a small number of sensors is often required in signal and image modeling and biomedical applications. This paper proposes a new algorithm to accurately estimate signals from underdetermined mixtures with less restrictions and assumptions compared with existing techniques. The strength of this algorithm is that it does not adopt the conventional assumptions on the mixing, signals and noise. The algorithm is capable of separating orthogonal and non-orthogonal mixtures of both sparse and non-sparse signals with additional Gaussian or non-Gaussian noise. This algorithm is also applicable to separating time-varying as well as instantaneous mixtures. Simulation results demonstrate the efficacy of the proposed algorithm for separation of time-varying mixtures in the presence of noise.

1. INTRODUCTION

In signal separation of linearly mixed signals, the mixing model is as below

$$\mathbf{x} = \mathbf{M}\mathbf{s} + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{s} = [s_1 \cdots s_m]^T$, $\mathbf{x} = [x_1 \cdots x_n]^T$, $\boldsymbol{\varepsilon}$ and $\mathbf{M}$ represent the original source signals, observed signals, noise and the mixing matrix respectively. Underdetermined mixture of signals where $m > n$, i.e. the number of source signals is greater than the number of observed signals are becoming more frequently addressed in Blind Signal Separation [1, 2]. A common feature in algorithms for solving underdetermined mixture is the exploitation of sparse representation of the source signals [3]. High sparsity is an essential requirement by these algorithms for good separability of source signals. However, the signals in some applications are non-sparse. Non-sparse signals are encountered in semireflective layer separation [4] and binary data in digital communication. In cases where the requirement of sparsity is not satisfied, non-sparse signals will need to be forced into a sparse representation via an appropriate transform prior to execution of the algorithm. Coefficients of wavelet and Fourier transformation do not produce sufficiently sparse signals to estimate signals. The solution must be the sparsest of all representations where a minimum number of non-zero coefficients possess a significant higher probability. This depends crucially on a measure that defines 'significant higher probability' and therein raises the conundrum of where the boundary between 'significant' and 'not significant' lies.

The measure of sparseness also needs to be carefully chosen in noisy data as slight additional noise will make the data completely non-sparse and kurtosis measures are to be avoided if it is not certain that the distribution is unimodal and symmetric [5]. The proposed algorithm avoids these pitfalls by not requiring sparseness of signals and is applicable for multimodal and non-symmetric distributions. Some algorithms disregard the presence of noise at the expense of accuracy. In practical scenarios, noise is usually present and should be taken into account when defining the mixture model which the formulation of the algorithm is based upon. Gaussian noise is often assumed in cases where the presence of noise is acknowledged. This assumption leads to inaccurate separation of mixtures with non-Gaussian noise. The proposed algorithm adopts a noise reduction model of wider coverage for both Gaussian and non-Gaussian noise.

2. FORMULATION OF ALGORITHM

The problem of blind signal separation of underdetermined mixtures can be solved by estimation of the unknown parameters $\mathbf{M}$ and $\mathbf{s}$ conditional upon $\mathbf{x}$. This problem is explicitly described by the following expression:

$$(\hat{\mathbf{M}}, \hat{\mathbf{s}}) = \arg\max_{\mathbf{M},\mathbf{s}} p(\mathbf{M}, \mathbf{s} | \mathbf{x}) \tag{2}$$

The estimation involves the process of formulating a prior density function of the original signals and mixing matrix. Multiplying these priors with the observed data conditional upon the original signals and mixing matrix, $p(\mathbf{x}|\mathbf{M},\mathbf{s})$ we obtain the posterior density function defined below:

$$p(\mathbf{M}, \mathbf{s} | \mathbf{x}) \propto p(\mathbf{x}|\mathbf{M},\mathbf{s})\, p(\mathbf{M})\, p(\mathbf{s}) \tag{3}$$

where $p(\mathbf{s})$ and $p(\mathbf{M})$ represent the prior probability density function for $\mathbf{s}$ and $\mathbf{M}$ respectively.

2.1. Mixing Matrix Estimation

The estimate for $\mathbf{M}$ can be obtained according to the following:

$$p(\hat{\mathbf{M}}|\mathbf{x}) = p(\hat{\mathbf{M}}) \int p(\mathbf{x}|\hat{\mathbf{M}}, \hat{\mathbf{s}})\, d\hat{\mathbf{s}} \tag{4}$$

The Generalized Gaussian Distribution (GGD) is adopted for the approximation of $p\left(\mathbf{x}\middle|\hat{\mathbf{M}},\hat{\mathbf{s}}\right)$:

$$p\left(\mathbf{x}-\hat{\mathbf{M}}\hat{\mathbf{s}}\right)=\frac{r}{2\lambda\Gamma\left(\frac{1}{\rho}\right)}\exp\left\{-\left(\frac{\left|\mathbf{x}-\hat{\mathbf{M}}\hat{\mathbf{s}}\right|}{\lambda}\right)^{\rho}\right\} \tag{5}$$

where $\Gamma(\bullet)$, ρ and $\lambda=\sqrt{\Gamma\left(\frac{1}{\rho}\right)\Big/\Gamma\left(\frac{3}{\rho}\right)}$ represent the standard Gamma function, parameter controlling density shape and generalized variance respectively. GGD allows the approximation of a wide range of noise statistics by varying ρ to represent Gaussian and non-Gaussian subject to the constraint $\rho>0$.

The estimate of ρ can be obtained via the first order variation of $\ln p(\hat{\boldsymbol{\varepsilon}})$ approximated from the following update equation for ρ:

$$\rho(t+1)=\rho(t)+\gamma(t)\Theta(t) \tag{6}$$

$$\Theta(t)=\frac{0.1\rho(t)+\left|\hat{\boldsymbol{\varepsilon}}(t)\right|^{\rho(t)}\left\{1-\ln\left(\left|\hat{\boldsymbol{\varepsilon}}(t)\right|^{\rho(t)}\right)\right\}}{\rho^{2}(t)} \tag{7}$$

where $\gamma(t)$ is the learning rate for ρ and $\hat{\boldsymbol{\varepsilon}}=\mathbf{x}-\hat{\mathbf{M}}\hat{\mathbf{s}}$. Taking the logarithm of (4), we can recast the estimate as:

$$\ln p\left(\hat{\mathbf{M}}\middle|\mathbf{x}\right)\propto\ln p\left(\hat{\mathbf{M}}\right)+\ln p\left(\hat{\mathbf{s}}\right)-\frac{1}{\lambda}\left\|\mathbf{x}-\hat{\mathbf{M}}\hat{\mathbf{s}}\right\|^{\rho}-\frac{1}{2}\ln|\mathbf{H}| \tag{8}$$

$$\mathbf{H}=\frac{1}{\lambda}\rho(\rho-1)\hat{\mathbf{M}}^{T}diag\left[\left|\mathbf{x}-\hat{\mathbf{M}}\hat{\mathbf{s}}\right|^{\rho-2}\right]\hat{\mathbf{M}}+\dot{\varphi}(\hat{\mathbf{s}}) \tag{9}$$

where $\dot{\varphi}(\hat{\mathbf{s}})=\frac{d}{d\hat{\mathbf{s}}d\hat{\mathbf{s}}^{T}}\ln p(\hat{\mathbf{s}})$. Following the natural gradient approach, we arrive at the update equation:

$$\begin{aligned}\hat{\mathbf{M}}(t+1)=\hat{\mathbf{M}}(t)+\eta\hat{\mathbf{M}}(t)\Big[&-\hat{\mathbf{M}}(t)^{T}\Sigma_{m}^{-1}\hat{\mathbf{M}}(t)\\&+\varphi(\hat{\mathbf{s}})\hat{\mathbf{s}}^{T}-\frac{1}{\lambda}\rho(\rho-1)\hat{\mathbf{M}}(t)^{T}\Big]\\&diag\left[\left|\mathbf{x}-\hat{\mathbf{M}}(t)\hat{\mathbf{s}}\right|^{\rho-2}\right]\hat{\mathbf{M}}(t)\mathbf{H}^{-1}+\underline{\phi}\hat{\mathbf{s}}^{T}\Big]\end{aligned} \tag{10}$$

where $\underline{\phi}=[\phi_{1}\ \cdots\ \phi_{k}]^{T}$ and $\phi_{m}=\frac{1}{2}\mathbf{H}_{mm}^{-1}\frac{d}{d\mathbf{s}_{m}}\nabla\nabla\ln p(\hat{\mathbf{s}})$.

2.2. Source Signal Estimation

Therefore, the estimate of **s** given **x** is defined as:

$$\hat{\mathbf{s}}=\arg\max_{\hat{\mathbf{s}}}\left[\ln p\left(\mathbf{x}-\hat{\mathbf{M}}\hat{\mathbf{s}}\right)+\ln p\left(\hat{\mathbf{s}}\right)\right] \tag{11}$$

The proposed algorithm is executed in the context of the Quasi-Newton update algorithm. The Quasi-Newton update retains the local convergence speed of the Newton method without the need to calculate the Hessian matrix required by the Newton method. Instead, it approximates the Hessian matrix using the past gradient information and iteratively updates the approximation using the current/latest gradient. This additional computation cost per iteration required by Quasi-Newton methods is compensated by a convergence far superior than gradient descent methods. The update for $\hat{\mathbf{s}}$ is then:

$$\begin{aligned}\hat{\mathbf{s}}(k+1)&=\hat{\mathbf{s}}(k)+\eta_{s}\left\{\frac{d^{2}J}{d\hat{\mathbf{s}}\hat{\mathbf{s}}^{T}}\right\}^{-1}\left[\frac{dJ}{d\hat{s}}\right]\\&=\hat{\mathbf{s}}(k)+\eta_{s}\left\{\hat{\mathbf{M}}^{T}\mathrm{diag}\left[\dot{\psi}(\hat{\boldsymbol{\varepsilon}})\right]\hat{\mathbf{M}}+\mathrm{diag}\left[\dot{\varphi}(\hat{\mathbf{s}})\right]\right\}^{-1}\\&\quad\left[-\hat{\mathbf{M}}^{T}\psi(\hat{\boldsymbol{\varepsilon}})+\varphi(\hat{\mathbf{s}})\right]\end{aligned} \tag{12}$$

where $\varphi(\hat{\mathbf{s}})=\left[\varphi(\hat{s}_{1})\ \varphi(\hat{s}_{2})\ \cdots\ \varphi(\hat{s}_{m})\right]^{T}$, $\varphi(\hat{s}_{i})\triangleq\frac{d}{d\hat{s}_{i}}\ln p(\hat{s}_{i})$, $\psi(\hat{\boldsymbol{\varepsilon}})=\left[\psi(\hat{\varepsilon}_{1})\ \psi(\hat{\varepsilon}_{2})\ \cdots\ \psi(\hat{\varepsilon}_{n})\right]^{T}$ and $\varphi(\hat{\boldsymbol{\varepsilon}})\triangleq\frac{d}{d\hat{\boldsymbol{\varepsilon}}}\ln p(\hat{\boldsymbol{\varepsilon}})$. $\mathbf{M}$ is not necessarily orthogonal and $\frac{d^{2}J}{d\hat{\mathbf{s}}\hat{\mathbf{s}}^{T}}=\hat{\mathbf{M}}^{T}\mathrm{diag}\left[\dot{\psi}(\hat{\boldsymbol{\varepsilon}})\right]\hat{\mathbf{M}}+\mathrm{diag}\left[\dot{\varphi}(\hat{\mathbf{s}})\right]$.

The prior density is modeled by Gaussian Mixture Models (GMM) which is capable of modeling both unimodal and multimodal densities:

$$p(\hat{\mathbf{s}})=\sum_{q=1}^{Q}\kappa_{q}p_{q}(\hat{\mathbf{s}}|\lambda_{q})=\sum_{q=1}^{Q}\frac{\kappa_{q}}{(2\pi)^{R/2}\left|\Sigma_{q}\right|^{1/2}}e^{\left\{-\frac{1}{2}(\hat{\mathbf{s}}-\boldsymbol{\mu}_{q})^{T}\Sigma_{q}^{-1}(\hat{\mathbf{s}}-\boldsymbol{\mu}_{q})\right\}} \tag{13}$$

Based on the GMM (13), we obtain the estimates:

$$\varphi(\hat{\mathbf{s}})=\frac{-\sum_{q=1}^{Q}\kappa_{q}p_{q}(\hat{\mathbf{s}}|\lambda_{q})\boldsymbol{\Sigma}_{q}^{-1}(\hat{\mathbf{s}}-\boldsymbol{\mu}_{q})}{p(\hat{\mathbf{s}})} \tag{14}$$

$$\dot{\varphi}(\hat{\mathbf{s}})=\frac{-\sum_{q=1}^{Q}\kappa_{q}\Sigma_{q}^{-1}p_{q}(\mathbf{s}|\lambda_{q})\Psi}{\left\{p(\mathbf{s})\right\}^{2}} \tag{15}$$

$$\Psi=\left\{p(\mathbf{s})\mathbf{I}-(\mathbf{s}-\boldsymbol{\mu}_{q})\left[p(\mathbf{s})(\mathbf{s}-\boldsymbol{\mu}_{q})^{T}\boldsymbol{\Sigma}_{q}^{-1}+\varphi^{T}(\mathbf{s})\right]\right\} \tag{16}$$

where R is the dimension of $\mathbf{s}$. κ_{q}, $\boldsymbol{\mu}_{q}$ and Σ_{q} are the mixture weight, mean and covariance matrix defining the GMM. Parameters of GMM are estimated from the EM algorithm which inherently satisfies the probabilistic constraints of the GMM. The parameters are adaptively estimated as:

$$\kappa_q(N) = \xi\, p\left(\lambda_q \left| \hat{\mathbf{s}}_N \right.\right) + (1-\xi)\,\kappa_q(N-1) \tag{17}$$

$$\mu_q = \frac{\hat{\mathbf{s}}_N\; p\left(\lambda_q \left| \hat{\mathbf{s}}_N \right.\right) + \sum_{n=1}^{N-1} \hat{\mathbf{s}}_n\; p\left(\lambda_q \left| \hat{\mathbf{s}}_n \right.\right)}{p\left(\lambda_q \left| \hat{\mathbf{s}}_N \right.\right) + \sum_{n=1}^{N-1} p\left(\lambda_q \left| \hat{\mathbf{s}}_n \right.\right)} \tag{18}$$

$$\boldsymbol{\Sigma}_q = \frac{\hat{\mathbf{s}}_N^2\; p\left(\lambda_q \left| \hat{\mathbf{s}}_N \right.\right) + \sum_{n=1}^{N-1} \hat{\mathbf{s}}_n^2\; p\left(\lambda_q \left| \hat{\mathbf{s}}_n \right.\right)}{p\left(\lambda_q \left| \hat{\mathbf{s}}_N \right.\right) + \sum_{n=1}^{N-1} p\left(\lambda_q \left| \hat{\mathbf{s}}_n \right.\right)} - \boldsymbol{\mu}_q^2 \tag{19}$$

where $\kappa_q(N-1) = \dfrac{1}{N-1}\sum_{n=1}^{N-1} p\left(\lambda_q \left| \hat{\mathbf{s}}_n \right.\right)$, $\xi = \dfrac{1}{N}$ and $0.05 \le \xi \le 0.2$.

The first and second order derivative of the noise term in (12) can consequently be calculated according to (20) and (21) as follows:

$$\psi\left(\hat{\boldsymbol{\varepsilon}}\right) = -\frac{\rho\left|\hat{\boldsymbol{\varepsilon}}\right|^{\rho-2}\hat{\boldsymbol{\varepsilon}}}{\lambda^{\rho}} \tag{20}$$

$$\dot{\psi}\left(\hat{\boldsymbol{\varepsilon}}\right) = -\frac{\rho\left|\hat{\boldsymbol{\varepsilon}}\right|^{\rho-2}}{\lambda^{\rho}}\left\{1 + \left(\rho-2\right)\left|\hat{\boldsymbol{\varepsilon}}\right|^{-2}\hat{\boldsymbol{\varepsilon}}^2\right\} \tag{21}$$

In summary, the algorithm is applicable for both stationary and time-varying signals by adaptively estimate $\hat{\mathbf{M}}$ from (10) and $\hat{\mathbf{s}}$ from (23):

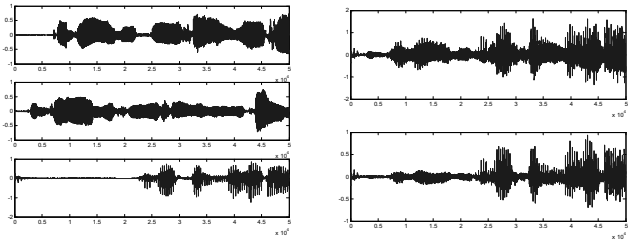
$$\hat{\mathbf{s}}(k+1) = \hat{\mathbf{s}}(k) - \eta_s \Xi(k) \tag{22}$$

$$\Xi = \left\{\hat{\mathbf{M}}^T \mathrm{diag}\left[\frac{\rho\left|\hat{\boldsymbol{\varepsilon}}\right|^{\rho-2}}{\lambda^{\rho}}\left\{1+\left(\rho-2\right)\left|\hat{\boldsymbol{\varepsilon}}\right|^{-2}\hat{\boldsymbol{\varepsilon}}^2\right\}\right]\hat{\mathbf{M}} - \mathrm{diag}\left[\frac{\sum_{q=1}^{Q}\kappa_q\Sigma_q^{-1}p_q(\hat{\mathbf{s}}|\lambda_q)\;\left\{p(\hat{\mathbf{s}})\mathbf{I} - (\hat{\mathbf{s}}-\boldsymbol{\mu}_q)\left[p(\hat{\mathbf{s}})(\hat{\mathbf{s}}-\boldsymbol{\mu}_q)^T\boldsymbol{\Sigma}_q^{-1} + \varphi^T(\hat{\mathbf{s}})\right]\right\}}{\left\{p(\hat{\mathbf{s}})\right\}^2}\right]\right\}^{-1}$$
$$\times\left[\mathbf{M}^T\left(\frac{\rho\left|\hat{\boldsymbol{\varepsilon}}\right|^{\rho-2}\hat{\boldsymbol{\varepsilon}}}{\lambda^{\rho}}\right) - \frac{\sum_{q=1}^{Q}\kappa_q p_q(\hat{\mathbf{s}}|\lambda_q)\boldsymbol{\Sigma}_q^{-1}(\hat{\mathbf{s}}-\boldsymbol{\mu}_q)}{p(\hat{\mathbf{s}})}\right] \tag{23}$$

3. RESULTS

To demonstrate the efficacy of the proposed algorithm, we compare its performance against the algorithm by Amari [6] and the FOCUSS algorithm [7]. Amari presented a basic natural gradient learning algorithm for underdetermined mixtures. However, this algorithm neglects the presence of noise. FOCUSS is a sparse technique that requires signals to be sparse and the mixing matrix is assumed known. Time-varying mixture of three original speech signals in Figure 1(a) with additional non-Gaussian noise at SNR=15 dB and the observed signals are presented in Figure 1(b). The signals separated by the algorithms are presented in Figure 1 (c), (d) and (e) for Amari, FOCUSS and the proposed algorithm respectively. In Figure 1 (c), it is observed that the first estimate is not clearly separated as it still contains mixtures of s_3. Also most of the signal in the first half of s_2 is lost in the separated signals and two of the estimates are similar. In Figure 1 (d), the first estimate is a poor estimate for s_2 where most of the signals in the first half are lost. The proposed algorithm clearly demonstrates its superior performance with all three signals clearly separated with a minimal amount of noise still present. Figure 2 presents the performance measure defined as

$$PI = \frac{1}{mT}\sum_{i=1}^{m}\sum_{t=1}^{T}\left(\frac{s_i(t)}{\sqrt{E\left[\left(s_i(t)\right)^2\right]}} - \frac{\hat{s}_i(t)}{\sqrt{E\left[\left(\hat{s}_i(t)\right)^2\right]}}\right)^2$$ for the separation of the speech signals Figure 3 depicts the performance index of the algorithms in estimation of mixtures of supergaussian, binary and gaussian signal. The figure shows the poor performance of Amari and FOCUSS in the estimation whereas the proposed algorithm demonstrates a much better performance under all levels of noise. This substantiates and reasserts the outstanding performance of the proposed algorithm in estimation of sparse and non-sparse signals.



(a) (b)

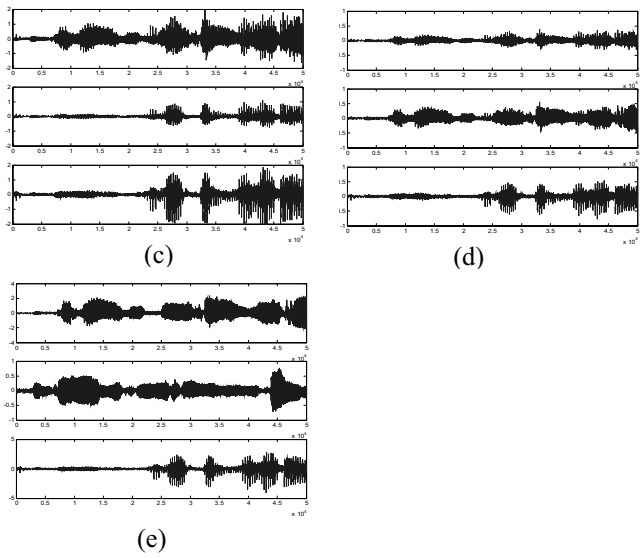


Figure 1: (a) Original Speech; (b) Observed Signals; (c) Estimation by Amari; (d) Estimation by FOCUSS; (e) Estimation by Proposed Algorithm

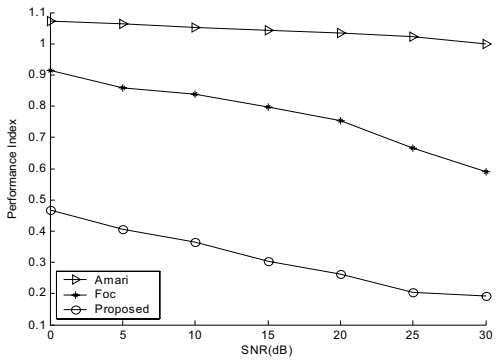


Figure 2: Performance Index of Algorithms

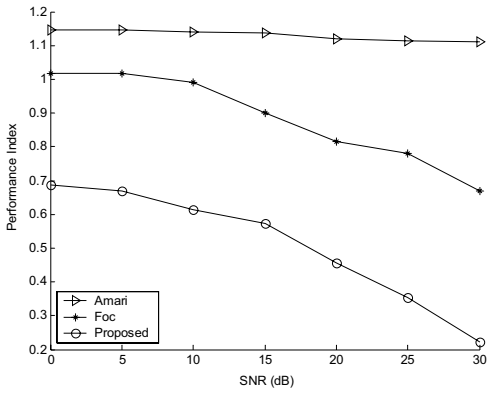


Figure 3: Error or estimation by (a) Amari; (b) FOCUSS; (c) Proposed Algorithm

4. CONCLUSION

This paper presents a new technique of robust blind signal estimation for underdetermined mixtures which produces superior accuracy and detailed results. The success of the algorithm is not dependent on the sparsity of signals, careful selection of sparseness measure and distribution of signals or noise. The estimate of the signals is greatly improved by the integration of a noise reduction procedure that extends to both Gaussian and non-Gaussian noise. Though the proposed algorithm's approach in modeling the signals and approximating noise is associated with high complexity cost, it is recompensed by the enhanced accuracy in the estimation of signals and robustness to noise. In blind separation of signals, an inaccurate density model of the signals affects the accuracy of the results asymptotically. Therefore, under limited sample size a more accurate model is crucial as an inaccurate model will lead to wrong results. With the increasing capacity of modern computers, the computational cost can be afforded in applications which require a great amount of detail and accuracy. Nevertheless, the complexity of the technique has been kept to a minimum without overly compromising accuracy. The GMM model adopted by the proposed algorithm in producing the results is composed of $Q = 2$ Gaussian density per mixture model and was still capable of producing good results. Nevertheless, this paper has mainly concentrated on introducing a superior technique for blind estimation applicable to both stationary and time-varying mixtures of sparse and non-sparse signals. This paper has presented a highly exciting solution for blind signal estimation of underdetermined mixtures in the form of an efficacious algorithm without the limitations and pitfalls of sparse techniques.

5. REFERENCES

[1] P. Comon, "Blind Identification and Source Separation in 2x3 Under-Determined Mixtures," *IEEE Transactions on Signal Processing*, vol. 52, pp. 11- 22, Jan 2004.

[2] A. Ferreol, L. Albera, and P. Chevalier, "Fourth order blind identification of underdetermined mixtures of sources (FOBIUM)," in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03), 2003, pp. 41-44.

[3] D. P. Wipf and B. D. Rao, "Sparse Bayesian Learning for Basis Selection," *IEEE Transactions on Signal Processing*, vol. 52, pp. 2153 - 2164, Aug. 2004.

[4] M. M. Bronstein, A. M. Bronstein, and M. Zibulevsky, "Separation of semireflective layers using Sparse ICA," in Proc. ICASSP 2003, 2003, pp. 733-736.

[5] J. Karvanen and A. Cichocki, "Measuring Sparseness of Noisy Signals," in 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), Nara, Japan, April 2003, pp. 125-130.

[6] S. Amari, "Natural Gradient Learning for Over- and Under-Complete Bases in ICA," *Neural Computation*, vol. 11, pp. 1875-1883, 1999.

[7] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-Delgado, "Subset Selection in Noise Based on Diversity Measure Minimization," *IEEE Transactions on Signal Processing*, vol. 51, pp. 760-770, March 2003.