POST-NONLINEAR UNDERCOMPLETE BLIND SIGNAL SEPARATION: A BAYESIAN APPROACH

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

University of Newcastle upon Tyne, UK
Email: chen.wei@ncl.ac.uk, l.c.khor@ncl.ac.uk, w.l.woo@ncl.ac.uk, s.s.dlay@ncl.ac.uk

ABSTRACT

The post-nonlinear undercomplete Blind Signal Separation problem is solved by a Bayesian approach in this paper. The proposed algorithm applies the Generalized Gaussian model to approximate the prior distribution probability and a Maximum a Posteriori (MAP) based learning algorithm to estimate the source signals, mixing matrix and the nonlinearity of the mixing process. The mixing nonlinearity is modeled by a Multilayer Perceptron (MLP) neural network. In our proposed algorithm, the source signals, mixing matrix and the parameters of the MLP are iteratively updated in an alternate manner until they converges to a fixed value. The noise variance is regarded as the hyper-parameter which is estimated in a closed form. Simulations based on real audio have been carried out to investigate the efficacy of the proposed algorithm. A performance gain of over 125% has been achieved when compared to linear approach.

1. INTRODUCTION

Independent Component Analysis (ICA), one of the more popular techniques employed in Blind Signal Separation (BSS), has received comparatively more attention over the last decade due to its simplicity and accuracy. Most of the solutions proposed to solve the problem of BSS aims to recover source signals from linear mixtures [1]. Nevertheless, nonlinear functions have been involved due to their existence in real applications [2,9,10]. A simple and effective post-nonlinear model proposed by Taleb and Jutten [3] composed of a linear mixing matrix and one layer of nonlinear distortion function. This model is particularly suited for problems where nonlinearity exists at the sensors. However, the algorithm is only applicable for nonlinear mixtures where the number of mixtures is equal to the number of sources.

In this paper, we consider a post-nonlinear undercomplete system which can be mathematically formulated as follows:

$$\mathbf{X} = f(\mathbf{AS}) + \mathbf{n} \tag{1}$$

where $\mathbf{S}$ and $\mathbf{X}$ are the source signals and mixed signals respectively; the mixing matrix $\mathbf{A}$ is a rectangular matrix with dimension $N_o \times N_s$ with $N_o \geq N_s$, $f(.)$ represent the post-nonlinear distortion function and $\mathbf{n} = \left[n_1, ..., n_{N_o}\right]^T$ represent Gaussian noises. The infrastructure of the model is shown in Fig. 1

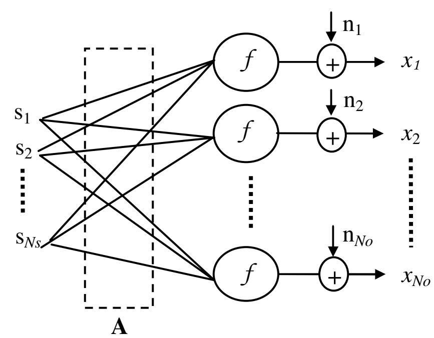

Fig. 1 Post-nonlinear undercomplete mixing model

The estimation of source signals can be implemented through two different approaches, the generative approach and the signal transformation approach. Generative approach involves estimating both the source signals and the mixing matrix, i.e. it aims to rebuild the mixing model and describe how the mixtures were generated; alternatively, the signal transformation approach aims to build a de-mixing system which attempt to cancel the mixing effect on the mixtures and subsequently recover the source signals. Both approach have has their own merits and shortcomings. The generative approach is able to handle more complicated cases, such as undercomplete mixtures where the number of sources is greater than the number of mixtures with the presence of noise, but it requires high computational complexity. Conversely, the signal transformation approach is simple and easy to compute but limited to noiseless and complete mixtures. In this paper, the generative approach is adopted since our model addresses both noisy and undercomplete mixtures.

The structure of this paper is organized as follows: the estimation of the mixing matrix and the source signals are firstly introduced in section 2; the nonlinear mismatch correction and noise variance estimation are presented in section 3; the simulation result and discussion are included in section 4 to verify the effectiveness of the proposed algorithm.

2. BAYESIAN FRAMEWORK AND MAP BASED ITERATIVE ESTIMATION

We introduce a Maximum a Posteriori (MAP) probability approach which estimates the joint probability distribution between $\mathbf{S}$ and $\mathbf{A}$ as follows

$$\left(\tilde{\mathbf{S}},\tilde{\mathbf{A}}\right)_{MAP} = \arg\max_{\mathbf{S},\mathbf{A}} P(\mathbf{S},\mathbf{A}|\mathbf{X}) \propto \arg\max_{\mathbf{S},\mathbf{A}} P(\mathbf{A}|\mathbf{X})P(\mathbf{S}|\mathbf{A},\mathbf{X}) \tag{2}$$

Maximizing the joint probability of $\mathbf{S}$ and $\mathbf{A}$ in (2) is a complicated process. To simplify the maximization process, we propose an approach where $P(\mathbf{A}|\mathbf{X})$ and $P(\mathbf{S}|\mathbf{A},\mathbf{X})$ are iteratively optimized according to the following equations:

$$\tilde{\mathbf{A}}_{MAP} = \arg\max_{\mathbf{A}} P(\mathbf{A}|\mathbf{X}) \tag{3}$$

$$\tilde{\mathbf{S}}_{MAP} = \arg\max_{\mathbf{S}} P(\mathbf{S}|\mathbf{A},\mathbf{X}) \tag{4}$$

The term $P(\mathbf{A}|\mathbf{X})$ in (3) can be expressed as

$$P(\mathbf{A}|\mathbf{X}) = \int P\left(\mathbf{A},\tilde{\mathbf{S}}|\mathbf{X}\right)d\tilde{\mathbf{S}} \propto \ln P(\mathbf{A}) + \ln \int P\left(\mathbf{X}|\mathbf{A},\tilde{\mathbf{S}}\right)P\left(\tilde{\mathbf{S}}\right)d\tilde{\mathbf{S}} \tag{5}$$

To approximate the source, mixing matrix and noise distribution in (5), we adopt the Generalized Gaussian Distribution (GGD) model for its computational simplicity, which can be mathematically represented as follows:

$$P(v) = \frac{p}{2\sqrt{2}\sigma\Gamma(1/p)}\exp\left(-\left(\frac{|v|}{\sqrt{2}\sigma}\right)^p\right) \tag{6}$$

where $P(v)$ is a GGD model with the variable v, standard deviation σ and shape factor p .Under the assumption of Gaussian noise with the same variance during mixing process, we obtain the following

$$P\left(\mathbf{X}|\mathbf{A},\tilde{\mathbf{S}}\right) = \left(\frac{1}{\sigma_{\mathbf{n}}\sqrt{2\pi}}\right)^{N_0}\exp\left(-\frac{\left\|\mathbf{X} - f\left(\mathbf{A}\tilde{\mathbf{S}}\right)\right\|^2}{2\sigma_{\mathbf{n}}^2}\right) \tag{7}$$

To simplify the computation of the integral term in (5), we exploit the Laplacian approximation and adopt the assumption that the integral follows the supergaussian distribution characterized by a sharp peak and heavy-tailed distribution. Under this assumption, we can substitute (7) into (5) and rewrite (5) as follows

$$\begin{aligned}&P(\mathbf{A}) + \left(\frac{1}{\sigma_{\mathbf{n}}\sqrt{2\pi}}\right)^{N_0}\int\exp\left(-\frac{\left\|\mathbf{X} - f\left(\mathbf{A}\tilde{\mathbf{S}}\right)\right\|^2}{2\sigma_{\mathbf{n}}^2}\right)P\left(\tilde{\mathbf{S}}\right)d\tilde{\mathbf{S}}\\ &\propto \ln P(\mathbf{A}) + \frac{1}{2}(N_0 - N_s)\ln\left(\frac{1}{2\pi\sigma_{\mathbf{n}}^2}\right) + \ln P\left(\tilde{\mathbf{S}}\right)\\ &-\frac{1}{2\sigma_{\mathbf{n}}^2}\left\|\mathbf{X} - f\left(\mathbf{A}\tilde{\mathbf{S}}\right)\right\|^2 - \frac{1}{2}\ln\det\left(\Lambda\left(\left\|\mathbf{X} - f\left(\mathbf{A}\tilde{\mathbf{S}}\right)\right\|^2\right)\right)\end{aligned} \tag{8}$$

where Λ corresponds to the Hessian function around the estimated $\tilde{\mathbf{S}}$ where $\tilde{\mathbf{S}}$ is estimated from (21) presented at the end of this section.

Due to limited space, this paper will only examine the last term $\Lambda\left(\left\|\mathbf{X} - f\left(\mathbf{A}\tilde{\mathbf{S}}\right)\right\|^2\right)$ in detail. Defining $\mathrm{H} = \Lambda\left(\left\|\mathbf{X} - f\left(\mathbf{A}\tilde{\mathbf{S}}\right)\right\|^2\right)$, the Hessian function becomes

$$\begin{aligned}\mathrm{H} &= \mathbf{A}^T\left(diag\left(\left\|\mathbf{X} - f\left(\mathbf{A}\tilde{\mathbf{S}}\right)\right\|\right)diag\left(f''\left(\mathbf{A}\tilde{\mathbf{S}}\right)\right) - diag\left(f'\left(\mathbf{A}\tilde{\mathbf{S}}\right)\right)^2\right)\mathbf{A}\\ &= \mathbf{A}^T diag(\Phi)\mathbf{A}\end{aligned} \tag{9}$$

where each element of H can be simply expressed as

$$h_{kl} = \sum_{i=1}^{N_o} a_{ik}a_{il}\phi_i \tag{10}$$

where $\phi_i = \left(x_i - f(\sum_{n=1}^{N_s} a_{in}\tilde{s}_n)\right)f''(\sum_{n=1}^{N_s} a_{in}\tilde{s}_n) - f'(\sum_{n=1}^{N_s} a_{in}\tilde{s}_n)^2$

In the absence of nonlinearity in the mixture, (8) reduces to its linear counterpart [4]

$$\ln P(\mathbf{A}) + \frac{1}{2}(N_0 - N_s)\ln\left(\frac{1}{2\pi\sigma_{\mathbf{n}}^2}\right) + \ln P\left(\tilde{\mathbf{S}}\right) - \frac{1}{2\sigma_{\mathbf{n}}^2}\left\|\mathbf{X} - f\left(\mathbf{A}\tilde{\mathbf{S}}\right)\right\|^2 - \frac{1}{2}\ln\left|\mathbf{A}\mathbf{A}^T\right|$$

with an additional term $P(\mathbf{A})$ introduced by MAP.

Differentiated with respect to $\mathbf{A}$, Hessian function in (9), which distinguishes the difference of the proposed nonlinear algorithm with its linear counterpart, is derived based on the chain rule and the identity [5] as follows

$$\frac{\partial \ln\det \mathrm{H}}{\partial \mathbf{A}} = \sum_{l=1}^{N_o}\sum_{k=1}^{N_o} h_{lk}^{-1}\varphi(i,k,l)\eta_i + \sum_{l=1}^{N_o}\sum_{k=1}^{N_o} h_{lk}^{-1}a_{ik}a_{il}\eta_i'\tilde{s}_j \tag{11}$$

where $\varphi(i,k,l) = \begin{cases} 2a_{ij} & j = k = l \\ a_{ik} & j = l \neq k \\ a_{il} & j = k \neq l \\ 0 & j \neq l \neq k \end{cases}$

and $\eta_i' = \left(x_i - f(\sum_{n=1}^{N_s} a_{in}\tilde{s}_n)\right)f'''(\sum_{n=1}^{N_s} a_{in}\tilde{s}_n) - 3f'(\sum_{n=1}^{N_s} a_{in}\tilde{s}_n)f''(\sum_{n=1}^{N_s} a_{in}\tilde{s}_n)$

where the first term in (11) is given by

$$\sum_{l=1}^{N_o}\sum_{k=1}^{N_o} h_{lk}^{-1}\varphi(i,k,l)\eta_i = 2diag(\eta)\mathbf{A}\left(\mathbf{A}^T diag(\eta)\mathbf{A}\right)^{-1} \tag{12}$$

Exploiting the identity $\mathbf{A}^{-T}\mathbf{A}^T = I$, we can express (12) as follows

$$\begin{aligned}\mathbf{A}^{-T}\mathbf{A}^T\sum_{l=1}^{N_o}\sum_{k=1}^{N_o} h_{lk}^{-1}\varphi(i,k,l)\eta_i &= \mathbf{A}^{-T}\mathbf{A}^T 2diag(\eta)\mathbf{A}\left(\mathbf{A}^T diag(\eta)\mathbf{A}\right)^{-1}\\ &= 2\mathbf{A}^{-T}\end{aligned} \tag{13}$$

The second term in (11) by

$$\sum_{l=1}^{N_o}\sum_{k=1}^{N_o} h_{lk}^{-1} a_{ik} a_{il} \eta_i' = diag(\eta')\left\langle \mathbf{A}H^{-1}\mathbf{A}^{\mathbf{T}}\right\rangle \tilde{\mathbf{S}}^T$$
$$= diag(\eta')\left\langle \mathbf{A}\left(\mathbf{A}^T diag(\eta)\mathbf{A}\right)^{-1}\mathbf{A}^{\mathbf{T}}\right\rangle \tilde{\mathbf{S}}^T$$
$$= diag(\eta')\left[\eta_1^{-1}\ldots\eta_{N_o}^{-1}\right]^T \tilde{\mathbf{S}}^T \quad (14)$$
$$= \left[\eta_1'\eta_1^{-1} \quad \cdots \quad \eta_{N_o}'\eta_{N_o}^{-1}\right]^T \tilde{\mathbf{S}}^T$$

where $\langle . \rangle$ and $diag(.)$ denotes the diagonal value of a matrix and the diagonalization process respectively.

Replacing (13) and (14) into (11), we obtain

$$\frac{\partial \ln \det \mathrm{H}}{\partial \mathbf{A}} = 2\mathbf{A}^{-T} + \left[\eta_1'\eta_1^{-1} \quad \cdots \quad \eta_{N_o}'\eta_{N_o}^{-1}\right]^T \tilde{\mathbf{S}}^T \quad (15)$$

Combining the derivation of nonlinear part (15) with derivation of the remaining linear parts in (8) [4], we can express the learning rule for $\mathbf{A}$ as

$$\mathbf{A(t+1) = A(t) + \mu_A \Delta A} \text{ where } \mathbf{\Delta A = L + N} \quad (16)$$

in which

$$\mathrm{L} = \frac{\partial \ln P(\mathbf{A})}{\partial \mathbf{A}} - \mathbf{A}^{-T}\left(\frac{\partial \ln P(\tilde{\mathbf{S}})}{\partial \tilde{\mathbf{S}}}\tilde{\mathbf{S}}^T + I\right) \quad (17)$$

$$\mathbf{N} = -\frac{1}{2}\left[\eta_1'\eta_1^{-1} \quad \cdots \quad \eta_{N_o}'\eta_{N_o}^{-1}\right]^T \tilde{\mathbf{S}}^T \quad (18)$$

We denote (17) as the linear component and (18) as the nonlinear component.

As for the estimation of source signals, we can formulate the cost function of the proposed generative network as

$$\tilde{\mathbf{S}} = \arg\max_{\mathbf{S}} P(\mathbf{S}|\mathbf{A},\mathbf{X}) \quad (19)$$

where

$$P(\mathbf{S}|\mathbf{A},\mathbf{X}) \propto -\frac{1}{2\sigma_n^2}\left\|\mathbf{X} - f(\mathbf{AS})\right\|^2 - \frac{1}{\sqrt{2}\sigma_s^{p_s}}\sum_{m=1}^{N_s}|s_m|^{p_s} \quad (20)$$

The derivative of the cost function with respect to $\mathbf{S}$ simply becomes

$$\nabla_{\mathbf{S}} P(\mathbf{S}|\mathbf{A},\mathbf{X}) = \frac{1}{\sigma_n^2}\mathbf{A}^T diag(f'(\mathbf{AS}))(\mathbf{X} - f(\mathbf{AS})) - \frac{p_s}{\sigma_s^{p_s}} diag\left(|s_m|\right)^{(p_s-2)}\mathbf{S} \quad (21)$$

Where the aim is to find an optimal solution where $\mathbf{S}$ satisfies the condition $\nabla_{\mathbf{S}} P(\mathbf{S}|\mathbf{A},\mathbf{X}) = 0$. This can be achieved through the adaptation of $\mathbf{S}$ by employing a gradient-based learning algorithm as follows:

$$\mathbf{S}(t+1) = \mathbf{S}(t) + \mu_{\mathrm{S}}\nabla_{\mathbf{S}} P(\mathbf{S}|\mathbf{A},\mathbf{X}) \quad (22)$$

where μ_{S} is the learning rate of the source estimation.

3. MINIMIZATION OF NONLINEAR MISMATCH AND NOISE VARIANCE ESTIMATION

Due to the lack of information on the mixture, a mismatch exists between the initial estimation of nonlinear function in (18) and the true function. To resolve this problem, a self-adaptive algorithm is proposed to approximate as similar as possible the true nonlinear function. It has been established by the Universal Approximation Theorem [6] that for every continuous function $f(.)$, there always exists a Multilayer Perceptron (MLP) which can uniformly approximate $f(.)$ in the form of

$$\delta(\mathbf{U},\mathbf{M}^{(1)},\mathbf{M}^{(2)},\boldsymbol{\beta}) = \mathbf{M}^{(1)}\tanh(\mathbf{M}^{(2)}\mathbf{U} + \boldsymbol{\beta}) \quad (23)$$

The estimation of accuracy for $\mathbf{S}$ and $\mathbf{A}$ will consequently degrade with the degree of mismatch of nonlinear function $f(.)$. Since the function $f(.)$ is a one-to-one mapping, the MLP indeed performs non-mixing nonlinear mapping. Hence, this suggests a straightforward approach for how the estimation of $\mathbf{A}$, $\mathbf{S}$ and $f(.)$ can be managed. Thus, this allows us to formulate a least square error criterion to minimize the mismatch between the true observed signal and the estimated observed signal as follows:

$$\{\mathbf{M}_n^{(1)},\mathbf{M}_n^{(2)},\beta_n\} = \arg\min_{\mathbf{M}_n^{(1)},\mathbf{M}_n^{(2)},\beta_n}\left\|x_n - \mathbf{M}_n^{(1)}\tanh(\mathbf{M}_n^{(2)}\sum_{m=1}^{N_s} a_{nm}s_m + \beta_n)\right\|^2$$
$$= \arg\min_{\mathbf{M}_n^{(1)},\mathbf{M}_n^{(2)},\beta_n}\left\|x_n - \mathbf{M}_n^{(1)}\tanh(\mathbf{M}_n^{(2)T}u_n + \beta_n^{\,T})\right\|^2 \quad (24)$$

where $\mathbf{M}_n^{(1)} = \left[m_{n,1}^{(1)},\cdots\cdots,m_{n,q}^{(1)}\right]$, $\mathbf{M}_n^{(2)} = \left[m_{n,1}^{(2)},\cdots\cdots,m_{n,q}^{(2)}\right]$, and $\beta_n = \left[\beta_{n,1},\cdots\cdots,\beta_{n,q}\right]$. Then the derivation of each parameter in (24) can be written as

$$\tilde{\mathbf{M}}_n^{(1)} = \arg\min_{\mathbf{M}_n^{(1)}}\left\|x_n - \mathbf{M}_n^{(1)}\tanh(\mathbf{M}_n^{(2)}u_n + \beta_n)\right\|^2$$
$$= -2\left(x_n - \mathbf{M}_n^{(1)}\tanh(\mathbf{M}_n^{(2)T}u_n + \beta_n^{\,T})\right)\tanh(\mathbf{M}_n^{(2)}u_n + \beta_n) \quad (25)$$

$$\begin{bmatrix}\tilde{\mathbf{M}}_n^{(2)}\\ \tilde{\beta}_n\end{bmatrix}$$
$$= \arg\min_{\tilde{\mathbf{M}}_n^{(2)},\tilde{\beta}_n}\left\|x_n - \mathbf{M}_n^{(1)}\tanh(\mathbf{M}_n^{(2)}u_n + \beta_n)\right\|^2 \quad (26)$$
$$= \left[-2\mathbf{M}_n^{(1)}\left(x_n - \mathbf{M}_n^{(1)}\tanh(\mathbf{M}_n^{(2)T}u_n + \beta_n^T)\right)\mathrm{diag}\left(\mathrm{sech}^2(\mathbf{M}_n^{(2)T}u_n + \beta_n^T)\right)\right]\begin{bmatrix}u_n\\ 1\end{bmatrix}$$

Once the coefficients of $\mathbf{M}$ and $\boldsymbol{\beta}$ of the MLP converge, the new estimate of nonlinear function is substituted into (16) and (21) to obtain refined estimates of $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{A}}$.

The noise variance $\sigma_{\mathbf{n}}^2$ in (8) can be regarded as a hyperparameter in Bayesian approach. Following Rajan and Rayner[7], the estimation of $\sigma_{\mathbf{n}}^2$ can be derived from

$$\arg\max_{\sigma_{\mathbf{n}}^2} P(\mathbf{X}|\mathbf{S},\mathbf{A},\sigma_{\mathbf{n}}^2) = 0 \quad (27)$$

and this gives a closed form estimate as

$$\sigma_{\mathbf{n}}^2 = \frac{\left\|\mathbf{X} - f(\mathbf{AS})\right\|^2}{(N_0 - N_s)} \quad (28)$$

4. RESULTS OF SIMULATION AND DISCUSSION

To demonstrate the incompetence of a linear BSS algorithm approach and the significance in the proposed algorithm of post-nonlinear undercomplete mixtures, we compare the effectiveness of the proposed algorithm with the well-known FOCUSS algorithm [8]. The noise is Gaussian distributed and is used to perturb the sensors. The initial estimate of the source signals are computed directly from (29)

$$\tilde{\mathbf{S}} = \mathbf{A}^{+}\mathbf{X} \tag{29}$$

where $\mathbf{A}^{+}$ is pseudoinverse of mixing matrix $\mathbf{A}$. The noise is Gaussian distributed and is used to perturb the sensors. While for the initial estimated source signals, they are computed directly from (35). In this experiment, the two audio waves shown in Fig.2 (a) correspond to the source signals. The source signals are transformed into three mixtures through (1) and depicted in Fig.2 (b). The mixing matrix is randomly generated from a Gaussian distribution. The actual post-nonlinear process $f(.)$ is set to $\tanh(.)$ and the estimated nonlinear process is assumed to be $(.)^{1/3}$ which means there is nonlinear distortion initially. Fig.1 (c) shows the recovered source signals by the algorithm proposed in this paper under SNR=20dB. Comparing Fig.2 (a) with Fig.2 (c), it clearly demonstrates the close resemblance between the original sources and the recovered source signals.

A performance index is introduced as a basis for comparison and can be defined as:

$$\mathrm{P} = 2\left(1 - \frac{1}{M}\sum_{i=1}^{N_s} |\rho_i|\right) \tag{29}$$

$$\rho_i = \frac{E\left[(s_i - E[s_i])^{*}(\tilde{s}_i - E[\tilde{s}_i])\right]}{\sqrt{E\left[|s_i - E[s_i]|^2\right]E\left[|\tilde{s}_i - E[\tilde{s}_i]|^2\right]}} \tag{30}$$

where ρ_i, $*$ and $|\cdot|$ are the normalized cross-correlation, complex conjugate and absolute operation respectively.

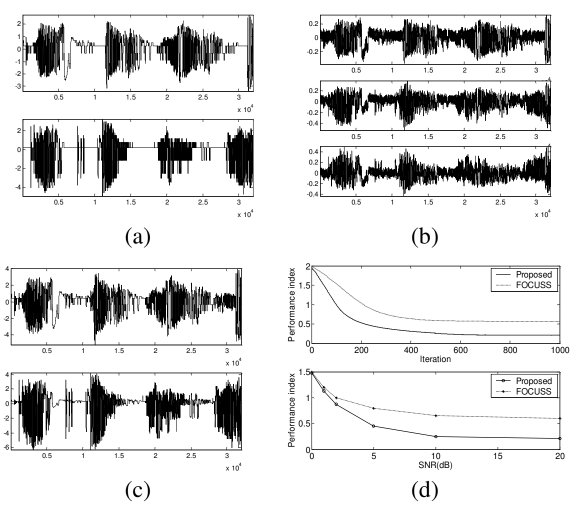

Fig.2. nonlinear undercomplete mixing using two speech signals. (a) Two source signals; (b) Three mixtures; (c) Two estimated source signals; (d) Performance index comparison.

Fig.2 (d) depicts the performance index under fixed SNR=20dB. The figure shows that both performance indices converge to a small fixed value (0.24 for post-nonlinear algorithm and 0.52 for FOCUSS) after 500 iterations. However, the performance of the proposed algorithm surpasses FOCUSS algorithm by over 100% under SNR=20dB. When SNR<5dB the presence of noise dominates and affects the performance of both algorithms significantly. However, a significant improvement in the performance of the proposed algorithm is observed as SNR increases. Furthermore, the gain in accuracy exceeds the FOCUSS algorithm by over 125% (0.21 for the proposed algorithm and 0.55 for FOCUSS).

5. CONCLUSION

This paper presents a novel algorithm to recover source signals from a set of blind nonlinear underdetermined mixtures. The algorithm is derived from a Bayesian framework and addresses simultaneously the problem of nonlinearity and undercomplete mixtures. Simulation results have demonstrated the efficiency of the proposed algorithm over linear algorithm in post-nonlinear mixtures.

REFERENCE

[1] J.F. Cardoso, Blind signal separation: Statistical principles, *Proc. of the IEEE*, Vol. 86, pp. 2009-2025, 1998

[2] W.L. Woo, S. Sali, General Multilayer Perceptron Demixer Scheme for Nonlinear Blind Signal Separation, *IEE Proc.Vis., Image & Signal Process*,149(5),pp.253-262, 2002

[3] A. Taleb, C. Jutten, Source Separation in Post-Nonlinear Mixtures, *IEEE Trans. Signal Processing*, Vol. 47, pp. 2807-2820, 1999

[4] S.J. Roberts, Independent component analysis: source assessment and separation: a Bayesian approach, *IEE Proc. Vis., Image & Signal Process*, Vol. 145, pp. 149-154, 1998

[5] C. Wei, W.L. Woo and S.S. Dlay, A MAP-based approach to blind nonlinear underdetermined mixture, *IEE Proc. Vis., Image & Signal Processing*, to be published.

[6] C.D. Cantrell, *Modern Mathematical Methods for Physicists and Engineers*,Cambridge University Press, 2000

[7] J.J. Rajan, P.J.W. Rayner, Model order selection for the singular-value decomposition and the discrete Karhunen-Loeve transform using a Bayesian approach, *IEE Proc. Vis., Image & Signal Processing*, Vol. 144, pp. 116-123, 1997

[8] B.D. Rao, K. Kreutz-Delgado, An Affine Scaling Methodology for Best Basis Selection, *IEEE Trans. Signal Processing*, Vol. 47, pp. 187-200, 1999

[9] C. Jutten, M. Babaie-Zadeh, and S. Hosseini, Three easy ways for separating nonlinear mixtures? *Signal Processing*, vol. 84, no. 2, pp. 217-229, 2004.

[10] L.C. Khor, W.L. Woo and S.S. Dlay, “Nonlinear Blind Signal Separation with Intelligent Controlled Learning”, IEE Proc. on Vision, Image and Signal Processing, 152(3), pp. 297-306, 2005.