# **Hidden Markov Blind Source Separation** of Post-nonlinear Mixture

Jingyi Zhang, W.L. Woo, and S.S. Dlay School of Electrical, Electronic and Computer Engineering Newcastle University UNITED KINDOM Email: jingyi.zhang@ncl.ac.uk, w.l.woo@ncl.ac.uk, s.s.dlay@ncl.ac.uk

Abstract — In this paper, a novel solution is developed to solve the problem of separating noisy and post-nonlinearly distorted mixture. In the proposed work, the source signals are nonstationary and temporally correlated. A generative model based on Hidden Markov Model (HMM) is derived to track the nonstationarity of the source signal while the source signal itself is modeled by temporally correlated Generalized Gaussian Distribution (GGD) Model. The Maximum Likelihood (ML) approach is developed to estimate the parameters of the proposed model by using the Expectation Maximization (EM) algorithm and the source signals are estimated by Maximum a Posteriori (MAP) approach. The strength of the proposed approach lies in the tracking of the nonstationarity of the source signal by HMM and the temporal correlation by the autoregressive (AR) source model. This has resulted in high performance accuracy, fast convergence and efficient implementation of the estimation algorithm. Simulations have been investigated to verify the effectiveness of the proposed algorithm and the results have shown significant improvement has been obtained when compared with nonlinear algorithm without using HMM.

Keywords-Hidden Markov Model, Nonlinear signal processing, Blind source separation (BSS).

#### T INTRODUCTION

The need of accurate representation and separation of the nonlinear distorted signals has resulted in the emergence of nonlinear BSS. The self-organizing map (SOM) has been used in [1] but it suffers from both network complexity and interpolation errors for continuous phase signals. Neural network models [2-6] based on nonlinear ICA algorithms are more structured and reported to produce better results than SOM models. The post-nonlinear model proposed by Taleb [5] is suitable for the practical applications which involve the use of nonlinear sensors. For the separability of the post-nonlinear model, [5, 6] have provided the analysis. Similar approaches were later adopted in [7, 8] where the hidden neuron functions are spanned by polynomials. Recently, a new result is developed in [9] in which the nonlinearity is characterized by a class of strictly monotonic continuously differentiable functions for BSS of nonlinear mixtures. However, many practical applications have been found to involve highly nonstationary and temporally correlated signal, such as speech signals and biomedical signals. Hence, a post-nonlinear BSS model of nonstationary and temporally correlated source signals must be proposed for most practical applications. Harmeling et al [10] proposed a kernel-based algorithm for nonlinear mixture of  $\theta_s = \{R_i, \beta_i, \mathbf{a}_{i,l}\}$  denotes all the parameters in the GGD model.

temporally correlated source. However, the nonstationarity of the source signal is not used and the mixing environment in the algorithm is assumed to be noiseless.

In this paper, a noisy post-nonlinear mixing of nonstationary and temporally correlated source is considered where the observed output  $\mathbf{X}_t$  at time t can be expressed by the following

$$\mathbf{x}_t = \mathbf{f}(\mathbf{M}\mathbf{s}_t) + \mathbf{n}_t \tag{1}$$

The matrix M is the mixing matrix,  $s_t$  is the vector of the source signals at time t and f(.) is a layer of continuously differentiable nonlinear function. The additive noise  $\mathbf{n}_t$  is assumed to be Gaussian with zero mean and diagonal covariance matrix  $(1/\beta)\mathbf{I}$ . The objective is to provide optimal estimation of the source signals, the mixing process parameters and the parameters of the additive noise.

The contribution of this paper is to provide a novel solution to noisy post-nonlinear distorted mixture of nonstationary and temporally correlated sources. The nonstationarity of the source signal is tracked by HMM in the proposed generative model while the source signal itself is modeled by temporally correlated GGD model. The ML approach is developed to estimate the parameters in the proposed model where the postnonlinearity is approximated by a set of polynomials whose coefficients are updated as part of the mixing process parameters and the source signal is estimated by MAP approach. It is shown in Section IV that by exploiting the nonstationarity (in addition of the temporal correlation) of the source signals, significant improvement in source separation performance has been gained.

#### PROPOSED MODEL II.

In the proposed model, the distribution of source *i* is modeled by temporally correlated GGD model as following

$$p(s_{i,t} \mid \boldsymbol{\theta}_{s_i}) = \frac{R_i \beta_i}{2\Gamma(1/R_i)} \exp(-\left\{\beta_i \left| s_{i,t} - \overline{s}_{i,t} \right| \right\}^{R_i})$$
(2)

where  $\Gamma(.)$  is the standard gamma function,  $\beta_i$  is the generalized variance and  $R_i$  controls the shape of the distribution. The temporal correlation of the source signal is represented by AR process, expressed as

$$\overline{s}_{i,t} = \sum_{l=1}^{L} a_{i,t,l} s_{i,t-l} = \mathbf{a}_{i,t} \widetilde{\mathbf{s}}_{i,t}$$
(3)

where  $\mathbf{a}_{i,t} = [a_{i,t,1} \ a_{i,t,2} \ \cdots \ a_{i,t,L}]$ ,  $\tilde{\mathbf{s}}_{i,t} = [s_{i,t-1} \ s_{i,t-2} \ \cdots \ s_{i,t-L}]^{\mathrm{T}}$  and

The GGD model parameter  $\{R_i, \beta_i\}$  of each source at every time *t* is assumed to switch between  $K_s$  different sets to model the nonstationarity of the source signal. Hence, a HMM can be used in the proposed model. The HMM has *K* hidden states, which is given by  $K = K_s^{N_s}$  where  $N_s$  is the number of sources. In the HMM, the initial state probability vector is defined as  $\pi = [\pi_1 \ \pi_2 \ \cdots \ \pi_K]$  and the state transition matrix is defined as **A**. The parameters of the mixing process and the source model is concatenated into the parameter set  $\theta_o = \{\theta_s, \mathbf{f}(.), \mathbf{M}, \beta\}$ . Hence, the parameters of the proposed generative model, as a whole, are concatenated into the parameter set  $\theta = \{\theta_o, \mathbf{A}, \pi\}$ . In Section III, the proposed model is trained by the ML approach while the source signals are estimated by the MAP approach.

# III. LEARNING RULES

In this section, the ML approach is derived to update the parameters in the proposed model by EM algorithm and the MAP approach is also derived to estimate the source signal. The proposed algorithm iterates between model update and source estimation until convergence.

# (a) Model update by ML approach

The joint likelihood of the observation sequence  $\mathbf{x}_{1:T} = [\mathbf{x}_1^T \ \mathbf{x}_2^T \ \cdots \ \mathbf{x}_T^T]^T$  and hidden state sequence  $\mathbf{q}_{1:T} = [q_1 \ q_2 \ \cdots \ q_T]^T$  to be written as

$$p_{\boldsymbol{\theta}}(\mathbf{q}_{1:T}, \mathbf{x}_{1:T}) = p_{\boldsymbol{\theta}}(q_1) \prod_{t=2}^{T} p_{\boldsymbol{\theta}}(q_t \mid q_{t-1}) \prod_{t=1}^{T} p_{\boldsymbol{\theta}}(\mathbf{x}_t \mid q_t)$$
(4)

and the joint log likelihood can be then written as

$$\log p_{\theta}(\mathbf{q}_{1:T}, \mathbf{x}_{1:T}) = \log p_{\theta}(q_1) + \sum_{t=2}^{T} \log p_{\theta}(q_t | q_{t-1}) + \sum_{t=1}^{T} \log p_{\theta}(\mathbf{x}_t | q_t)$$
(5)

Following Baye's rule and Jensen's inequality, the joint log likelihood can be maximized by using the EM algorithm to maximize the *auxiliary function*  $F(\hat{\theta}, \theta)$  [11], which is the expectation of the joint log likelihood where the expectation is taken relative to the conditional distribution of hidden state  $p_{\hat{\theta}}(\mathbf{q}_{1:T} | \mathbf{x}_{1:T})$  with the parameters estimated in the previous iteration. Hence, the *auxiliary function* is written as

$$F(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = F(\hat{\boldsymbol{\pi}}, \boldsymbol{\pi}) + F(\hat{\mathbf{A}}, \mathbf{A}) + F(\hat{\boldsymbol{\theta}}_{o}, \boldsymbol{\theta}_{o})$$
(6)  

$$F(\hat{\boldsymbol{\pi}}, \boldsymbol{\pi}) = \sum_{\mathbf{q}_{1:T}} p_{\hat{\boldsymbol{\theta}}}(\mathbf{q}_{1:T} \mid \mathbf{x}_{1:T}) \log p_{\boldsymbol{\theta}}(q_{1})$$
  

$$F(\hat{\mathbf{A}}, \mathbf{A}) = \sum_{\mathbf{q}_{1:T}} p_{\hat{\boldsymbol{\theta}}}(\mathbf{q}_{1:T} \mid \mathbf{x}_{1:T}) \sum_{t=2}^{T} \log p_{\boldsymbol{\theta}}(q_{t} \mid q_{t-1})$$
  

$$F(\hat{\boldsymbol{\theta}}_{o}, \boldsymbol{\theta}_{o}) = \sum_{\mathbf{q}_{1:T}} p_{\hat{\boldsymbol{\theta}}}(\mathbf{q}_{1:T} \mid \mathbf{x}_{1:T}) \sum_{t=1}^{T} \log p_{\boldsymbol{\theta}}(\mathbf{x}_{t} \mid q_{t})$$
(7)

where  $\Sigma_{\mathbf{q}_{1:T}}$  denotes a sum over all possible hidden state sequences. The three terms can be maximized separately giving rise to parameter update equations for each part of the proposed model. The term of  $F(\hat{\theta}_a, \theta_a)$  can be re-arranged as

$$F(\hat{\boldsymbol{\theta}}_{o}, \boldsymbol{\theta}_{o}) = \sum_{k=1}^{K} \sum_{t=1}^{T} \gamma_{k}(t) \log p_{\boldsymbol{\theta}}(\mathbf{x}_{t} | q_{t})$$
(8)

where  $\gamma_k(t) = p(q_t = k | \mathbf{x}_{1:T}, \boldsymbol{\theta})$  is the probability of being in state *k* at time *t* given observation  $\mathbf{x}_{1:T}$ . Given  $\boldsymbol{\theta}$  and defining

$$\begin{aligned} \alpha_{k}(t) &= p(\mathbf{x}_{1:t}, q_{t} = k \mid \mathbf{\theta}) \quad , \quad \eta_{k}(t) = p(\mathbf{x}_{t+1:T}, q_{t} = k \mid \mathbf{\theta}) \\ \zeta_{i,j}(t) &= p(q_{t} = i, q_{t+1} = j \mid \mathbf{x}_{1:T}, \mathbf{\theta}) \quad , \quad b_{k}(t) = p(\mathbf{x}_{t} \mid q_{t} = k) \end{aligned} \tag{9}$$

 $\gamma_k(t)$  can be calculated by the forward-backward algorithm [11] as follows

$$\begin{aligned} \alpha_{i}(t) &= \sum_{j=1}^{K} \alpha_{k}(t) a_{ji} \ b_{j}(t) , \text{ with } \alpha_{i}(1) = \pi_{i} b_{i}(1) \\ \eta_{i}(t) &= \sum_{j=1}^{K} a_{ij} \ \eta_{j}(t+1) b_{j}(t+1) , \text{ with } \eta_{i}(T) = 1 \\ \xi_{i,j}(t) &= \alpha_{i}(t) a_{ij} b_{j}(t+1) \eta_{j}(t+1) \Big/ \sum_{i'=1}^{K} \sum_{j'=1}^{K} \alpha_{i'}(t) a_{ij'} b_{j'}(t+1) \eta_{j'}(t+1) \\ \gamma_{k}(t) &= \alpha_{k}(t) \eta_{k}(t) \Big/ \sum_{k'=1}^{K} \alpha_{k'}(t) \eta_{k'}(t) , \ b_{k}(t) = G(\mathbf{x}_{t} - \mathbf{\mu}_{k}; \mathbf{\Sigma}_{k}) \end{aligned}$$
(10)

where  $G(\mathbf{x}_t - \boldsymbol{\mu}_k; \boldsymbol{\Sigma}_k)$  denotes a multidimensional Gaussian distribution of  $\mathbf{x}_t$  with mean  $\boldsymbol{\mu}_k$  and covariance matrix  $\boldsymbol{\Sigma}_k$ , which can be obtained as

$$\boldsymbol{\mu}_{k} = \sum_{t=1}^{T} \gamma_{k}(t) \mathbf{x}_{t} / \sum_{t=1}^{T} \gamma_{k}(t) , \ \boldsymbol{\Sigma}_{k} = \sum_{t=1}^{T} \gamma_{k}(t) [\mathbf{x}_{t} - \boldsymbol{\mu}_{k}] [\mathbf{x}_{t} - \boldsymbol{\mu}_{k}]^{T} / \sum_{t=1}^{T} \gamma_{k}(t)$$
(11)

Maximizing  $F(\hat{\pi}, \pi)$  and  $F(\hat{A}, A)$  leads to the usual update equations for  $\pi_k$  and the entries  $a_{ij}$  of A [11], obtained as

$$_{k} = \gamma_{k}(1)$$
 ,  $a_{ij} = \sum_{t=1}^{T-1} \xi_{i,j}(t) / \sum_{t=1}^{T-1} \gamma_{i}(t)$  (12)

In this paper,  $p_{\theta}(\mathbf{x}_t | q_t)$  in (8) is expressed as

$$p(\mathbf{x}_t \mid \mathbf{\theta}) = \int p(\mathbf{x}_t \mid \mathbf{s}_t, \mathbf{\theta}) p(\mathbf{s}_t \mid \mathbf{\theta}_s) d\mathbf{s} = \int p(\mathbf{x}_t \mid \mathbf{s}_t, \mathbf{\theta}_o) p(\mathbf{s}_t \mid \mathbf{\theta}_s) d\mathbf{s} \quad (13)$$

By using Laplace approximation,  $p(\mathbf{x}_t | \boldsymbol{\theta})$  can be expressed as

$$p(\mathbf{x}_t \mid \mathbf{\theta}) \approx \frac{(2\pi)^{N_o/2}}{|\mathbf{H}_t|^{1/2}} p(\mathbf{x}_t \mid \hat{\mathbf{s}}_t, \mathbf{\theta}_o) p(\hat{\mathbf{s}}_t \mid \mathbf{\theta}_s)$$
(14)

where  $\hat{\mathbf{s}}_t$  is the estimated source signal by MAP approach and

$$p(\mathbf{x}_{t}|\hat{\mathbf{s}}_{t}, \mathbf{\theta}_{o}) = G(\mathbf{x}_{t} - \mathbf{f}(\mathbf{M}\hat{\mathbf{s}}_{t}); (1/\beta)\mathbf{I}) , p(\hat{\mathbf{s}}_{t}|\mathbf{\theta}_{s}) = \prod_{i=1}^{N_{s}} p(\hat{s}_{i,t}|\mathbf{\theta}_{s_{i}})$$
$$\mathbf{H}_{t} = -\frac{\partial^{2} \log p(\mathbf{x}_{t}|\hat{\mathbf{s}}_{t}, \mathbf{\theta}_{o})}{\partial \hat{\mathbf{s}}_{t} \partial \hat{\mathbf{s}}_{t}^{T}}$$
(15)

Hence,  $F(\hat{\theta}_o, \theta_o)$  in (13) can be expressed as

$$F(\hat{\boldsymbol{\theta}}_{o},\boldsymbol{\theta}_{o}) = \frac{1}{T} \sum_{k=1}^{K} \sum_{t=1}^{T} \gamma_{k}(t) \log p(\boldsymbol{\mathbf{x}}_{t} \mid \boldsymbol{\theta})$$
  
$$= \frac{1}{T} \sum_{k=1}^{K} \sum_{t=1}^{T} \gamma_{k}(t) \left[ \frac{N_{o} - N_{s}}{2} \log \beta - \frac{1}{2} \left[ \log \left| \boldsymbol{\mathbf{M}}^{\mathrm{T}} \boldsymbol{\mathbf{Q}}_{t} \boldsymbol{\mathbf{M}} \right| + \beta \left\| \boldsymbol{\mathbf{x}}_{t} - \boldsymbol{\mathbf{f}}(\boldsymbol{\mathrm{M}}\hat{\boldsymbol{s}}_{t}) \right\|^{2} \right]$$
  
$$+ \sum_{i=1}^{N_{s}} \left[ \log R_{i,k} + \log \beta_{i,k} - \log \left( 2\Gamma(1/R_{i,k}) \right) - \left( \beta_{i,k} \left| \hat{s}_{i,t} - \overline{s}_{i,t} \right| \right)^{R_{i,k}} \right] \right]$$
(16)

where we have divided  $F(\hat{\theta}_{o}, \theta_{o})$  by the signal length T and

 $\mathbf{Q}_t = diag \left[ \mathcal{Q}_{1,t} \quad \cdots \quad \mathcal{Q}_{N_o,t} \right]$ ,  $\mathcal{Q}_{i,t} = f_i'^2 - (x_i - f_i) f''$  (17) The update equation for every parameter of  $\boldsymbol{\theta}$  can now be obtained by using their partial derivative respectively as

$$\mathbf{M}_{n+1} = \mathbf{M}_n + \lambda \frac{\partial F(\mathbf{\theta}_o, \mathbf{\theta}_o)}{\partial \mathbf{M}} \Big|_{\mathbf{M} = \mathbf{M}_n}$$

$$\frac{1}{\beta} = \left(\sum_{k=1}^{K} \sum_{t=1}^{T} \left[ \gamma_{k}(t) \left\| \mathbf{x}_{t} - \mathbf{f}(\mathbf{M} \, \hat{\mathbf{s}}_{t}) \right\|^{2} \right] \right) / \left( (N_{o} - N_{s}) \sum_{k=1}^{K} \gamma_{k}(t) \right)$$

$$R_{i,k,n+1} = R_{i,k,n} + \lambda \frac{\partial F(\hat{\boldsymbol{\theta}}_{o}, \boldsymbol{\theta}_{o})}{\partial R_{i,k}} \Big|_{R_{i,k} = R_{i,k,n}}$$

$$\beta_{i,k,n+1} = \beta_{i,k,n} + \lambda \frac{\partial F(\hat{\boldsymbol{\theta}}_{o}, \boldsymbol{\theta}_{o})}{\partial \beta_{i,k}} \Big|_{\beta_{i,k} = \beta_{i,k,n}}$$

$$\mathbf{a}_{i,t,n+1} = \mathbf{a}_{i,t,n} + \lambda \frac{\partial F(\hat{\boldsymbol{\theta}}_{o}, \boldsymbol{\theta}_{o})}{\partial \mathbf{a}_{i,t}} \Big|_{\mathbf{a}_{i,t} = \mathbf{a}_{i,t,n}}$$
(18)

where  $\lambda$  is the learning rate which can be made adaptive and

$$\frac{\partial F(\hat{\boldsymbol{\theta}}_{o},\boldsymbol{\theta}_{o})}{\partial \mathbf{M}} = \frac{1}{T} \sum_{k=1}^{K} \sum_{t=1}^{T} \gamma_{k}(t) \left[ \left( \beta \mathbf{u}_{t} - \frac{1}{2} \overline{\mathbf{Q}}_{t} \mathbf{w}_{t} \right) \hat{\mathbf{s}}_{t}^{T} - 2 \mathbf{Q}_{t} \mathbf{M} \left( \mathbf{M}^{T} \mathbf{Q}_{t} \mathbf{M} \right)^{-1} \right] \\
\mathbf{W}_{t} = \mathbf{M} \left( \mathbf{M}^{T} \mathbf{Q}_{t} \mathbf{M} \right)^{-1} \mathbf{M}^{T}, \mathbf{w}_{t} = \left[ w_{1,t} \cdots w_{N_{o},t} \right]^{T}, w_{i,t} = \mathbf{W}_{t,ii} \\
\overline{\mathbf{Q}}_{t} = diag \left[ \overline{\mathcal{Q}}_{1,t} \cdots \overline{\mathcal{Q}}_{N_{o},t} \right], \overline{\mathcal{Q}}_{i,t} = 3f_{i}' f_{i}'' - (x_{i,t} - f_{i}) f''' \\
\mathbf{u}_{t} = \left[ u_{1,t} \cdots u_{N_{o},t} \right]^{T}, u_{i,t} = (x_{i,t} - f_{i}) f_{i}' \\
\frac{\partial F(\hat{\mathbf{\theta}}_{o}, \mathbf{\theta}_{o})}{\partial R_{i,k}} = \frac{1}{T} \sum_{t=1}^{T} \gamma_{k}(t) \left[ \frac{1}{R_{i,k}} + \frac{\Gamma'(1/R_{i,k})}{2\Gamma(1/R_{i,k})R_{i,k}^{2}} - \ln(\beta_{i,k}|\hat{s}_{i,t} - \overline{s}_{i,t}|) (\beta_{i,k}|\hat{s}_{i,t} - \overline{s}_{i,t}|)^{R_{i,k}} \right] \\
\frac{\partial F(\hat{\mathbf{\theta}}_{o}, \mathbf{\theta}_{o})}{\partial \beta_{i,k}} = \frac{1}{T} \sum_{t=1}^{T} \gamma_{k}(t) \left[ \frac{1}{\beta_{i,k}} - R_{i,k} \beta_{i,k}^{R_{i,k}-1} |\hat{s}_{i,t} - \overline{s}_{i,t}|^{R_{i,k}} \right] \\
\frac{\partial F(\hat{\mathbf{\theta}}_{o}, \mathbf{\theta}_{o})}{\partial \mathbf{a}_{i,t}} = \frac{1}{T} \sum_{k=1}^{K} \gamma_{k}(t) R_{i,k} \beta_{i,k}^{R_{i,k}} |\hat{s}_{i,t} - \overline{s}_{i,t}|^{R_{i,k}-1} \tilde{\mathbf{s}}_{i,t}^{T}$$
(19)

Following the Weierstrass Theorem [9, 10, 12], a set of polynomials is used to estimate the nonlinear functions **f**. Thus, every scalar function  $f_i$  of **f** can be approximated as

$$f_{i} = \sum_{z=0}^{Z} b_{i,z} (\underbrace{\mathbf{m}_{i} \hat{\mathbf{s}}_{t}}_{v_{i,t}})^{z} = \mathbf{b}_{i} \mathbf{v}_{i,t} , \quad f_{i}' = \sum_{z=1}^{Z} z b_{i,z} v_{i,t}^{z-1} = \mathbf{b}_{i} \mathbf{v}_{i,t}^{(1)}$$

$$f_{i}'' = \sum_{z=2}^{Z} b_{i,z} z (z-1) v_{i,t}^{z-2} = \mathbf{b}_{i} \mathbf{v}_{i,t}^{(2)}$$
(20)

where  $\mathbf{m}_i$  is the  $i^{th}$  row of  $\mathbf{M}$ ,  $\mathbf{v}_{i,t}^{(1)} = \begin{bmatrix} 0 \ 1 \ 2v_{i,t} \ \cdots \ Zv_{i,t}^{Z-1} \end{bmatrix}^T$ ,  $\mathbf{v}_{i,t} = \begin{bmatrix} 1 \ v_{i,t} \ \cdots \ v_{i,t}^Z \end{bmatrix}^T$ ,  $\mathbf{v}_{i,t}^{(2)} = \begin{bmatrix} 0 \ 0 \ 2 \ \cdots \ Z(Z-1)v_{i,t}^{Z-2} \end{bmatrix}^T$ ,  $\mathbf{b}_i = \begin{bmatrix} b_{i,0} \ \cdots \ b_{i,Z} \end{bmatrix}$ . The polynomial coefficients  $b_{i,z}$  can be updated as one of the model parameters by maximizing  $F(\hat{\mathbf{\theta}}_{\alpha}, \mathbf{\theta}_{\alpha})$ . Define

 $\mathbf{B} = diag \begin{bmatrix} \mathbf{b}_{1} \cdots \mathbf{b}_{N_{o}} \end{bmatrix}, \mathbf{v}_{t}^{\mathrm{T}} = \begin{bmatrix} \mathbf{v}_{1,t}^{\mathrm{T}} \cdots \mathbf{v}_{N_{o},t}^{\mathrm{T}} \end{bmatrix}, \mathbf{X}_{t} = diag \begin{bmatrix} x_{1,t} \cdots x_{N_{o},t} \end{bmatrix}$  $\mathbf{V}_{t}^{(1)} = diag \begin{bmatrix} \mathbf{v}_{1,t}^{(1)} \cdots \mathbf{v}_{N_{o},t}^{(1)} \end{bmatrix}, \mathbf{V}_{t}^{(2)} = diag \begin{bmatrix} \mathbf{v}_{1,t}^{(2)} \cdots \mathbf{v}_{N_{o},t}^{(2)} \end{bmatrix}, \mathbf{V}_{t} = diag \begin{bmatrix} \mathbf{v}_{1,t} \cdots \mathbf{v}_{N_{o},t} \end{bmatrix} (21)$ Hence, the update equation for **B** can be obtained as

$$\begin{aligned} \mathbf{B}_{n+1} &= \mathbf{B}_{n} + \lambda \frac{\partial F(\mathbf{\theta}_{o}, \mathbf{\theta}_{o})}{\partial \mathbf{B}} \Big|_{\mathbf{B}=\mathbf{B}_{n}} \\ \frac{\partial F(\mathbf{\hat{\theta}}_{o}, \mathbf{\theta}_{o})}{\partial \mathbf{B}} &= \frac{1}{T} \sum_{k=1}^{K} \sum_{t=1}^{T} \gamma_{k}(t) \Big[ \beta(\mathbf{X}_{t} - \mathbf{B}\mathbf{V}_{t}) \mathbf{V}_{t}^{\mathrm{T}} - \frac{1}{2} \Big[ \left( \mathbf{V}_{t}^{\mathrm{T}} \mathbf{B}^{\mathrm{T}} - \mathbf{X}_{t} \right) \mathbf{W}_{t} \mathbf{V}_{t}^{(2)^{\mathrm{T}}} \\ &+ \mathbf{W}_{t} \Big( 2\mathbf{B} \mathbf{V}_{t}^{(1)} \mathbf{V}_{t}^{(1)^{\mathrm{T}}} + \mathbf{V}_{t}^{(2)^{\mathrm{T}}} \mathbf{B} \mathbf{V}_{t}^{\mathrm{T}} \Big) \Big] \Big] (22) \end{aligned}$$

Thus, all the parameters in the proposed model are updated.

# (b) Source estimation by MAP approach

With the updated model parameters, the source signal can now be estimated by MAP approach, which can be expressed as

$$\hat{\mathbf{s}}_{t} = \underset{\mathbf{s}_{t}}{\operatorname{argmax}} \log p(\mathbf{s}_{t} | \mathbf{x}_{t}) = \underset{\mathbf{s}_{t}}{\operatorname{argmax}} \log \left( p(\mathbf{x}_{t} | \mathbf{s}_{t}) p(\mathbf{s}_{t}) \right)$$
(23)

where  $p(\mathbf{x}_t | \mathbf{s}_t)$  is expressed by  $p(\mathbf{x}_t | \mathbf{\theta})$  in (13). Hence, the source signal can be estimated by gradient ascent algorithm

$$\hat{\mathbf{s}}_{t} = \mathbf{s}_{t} + \lambda_{s} \frac{\partial \log\left(p(\mathbf{x}_{t} | \mathbf{s}_{t}) p(\mathbf{s}_{t})\right)}{\partial \mathbf{s}_{t}}$$

$$\frac{\partial \log\left(p(\mathbf{x}_{t} | \mathbf{s}_{t}) p(\mathbf{s}_{t})\right)}{\partial \mathbf{s}_{t}} = \frac{1}{T} \sum_{k=1}^{K} \gamma_{k}(t) \left[\beta \mathbf{M}^{\mathrm{T}} \mathbf{F}'(\mathbf{x}_{t} - \mathbf{f}) - \tilde{\beta}_{k} \mathbf{\Lambda}_{k,t}\right]$$

$$\mathbf{F}' = diag \left[f_{1}' \cdots f_{N_{o}}'\right] , \quad \tilde{\beta}_{k} = diag \left[\beta_{1,k}^{R_{1,k}} \cdots \beta_{N_{s},k}^{R_{N_{s},k}}\right]$$

$$\mathbf{\Lambda}_{k,t} = \left[R_{1,k} \left|\hat{s}_{1,t} - \overline{s}_{1,t}\right|^{R_{1},k-1} \cdots R_{N_{s},k} \left|\hat{s}_{N_{s},t} - \overline{s}_{N_{s},t}\right|^{R_{N_{s},k}-1}\right]^{\mathrm{T}} (24)$$

Hence, the proposed algorithm alternates between model update (a) and source estimation (b) until it converges.

## IV. RESULTS

In this section, two experimental simulations under different conditions have been designed to investigate the efficacy of the proposed approach under different environment.

# (a) Performance under nonstationary Gaussian sources

In this simulation, each source is generated with two different types of segments, one with small variance and the other one with large variance. The source signal is stationary Gaussian within every segment but nonstationary between different segments. For simplicity, the temporal correlation of source signal is modeled by an AR process with order 3, while the HMM used in the proposed model has 2 hidden states. The mixing matrix is randomly selected and the additive Gaussian noise is added to the mixture to obtain the required SNR. The function  $f_1(v_1) = \tanh(v_1)$  and  $f_2(v_2) = v_2 + v_2^3$  are selected as the post-nonlinear distortions. The function  $f_1(v_1)$  is bounded while  $f_2(v_2)$  unbounded and this selection is taken merely to investigate the performance of the proposed algorithm under two different forms of nonlinearity. The performance of the proposed algorithm without using HMM has also been plotted to show the improvement delivered by using HMM in the proposed model. The proposed algorithm converges fast after 10 iterations and it is clear from the obtained results that the proposed algorithm with HMM has out-performed and provides significantly better recovered signals. This because by adopting HMM in the proposed source model, for every time t  $\beta_i$  and  $R_i$  can switch between K different sets to model different variance of different source segments. The obtained results show the importance of adopting the source non-stationarity in the proposed model when the source signal is highly non-stationary and post-nonlinearly distorted. In Fig. 1(d), Mean Square Error (MSE) between the recovered and original source signals versus SNR is shown. The average MSE improvement by using HMM for SNR ranging from 5dB to 30dB is about 150%. The proposed algorithm without HMM performs much less than satisfactory since the source non-stationarity is not adopted and  $\beta_i$  and  $R_i$  are constrained to be constant for all segments. Hence, the source model can not model different variances of different source segments but only approximate the average for all segments.



Fig. 1: Signals in simulation 1 at SNR=25dB. (a) Original sources. (b) Recovered sources with HMM. (c) Recovered sources without HMM. (d) MSE versus SNR.



Fig. 2: Signals in speech separation at SNR=25dB (a) Original sources. (b) Recovered sources with HMM. (c) Recovered sources without HMM.

# (b) Performance under speech signals

In this simulation, the proposed algorithm is tested for two speech signals with length about 4 seconds. The mixing environment is set up as same as in simulation (a). A  $12^{th}$  order AR process is used in source GGD model and a HMM with 4 hidden states is used in the proposed model. Fig. 2 displays the original source, recovered source by proposed algorithm with HMM and without HMM at SNR=25dB from (a) to (c), respectively. It is clear that the proposed algorithm provides better recovered source with HMM in the proposed model. Hence, the obtained results again show the importance of using HMM to track the nonstationarity of the source signals when the source signal is highly nonstationary.

## V. CONCLUSIONS

A novel statistical approach of nonlinear blind source separation of nonstationary and temporally correlated sources has been proposed. The derivation of the proposed model tracks the nonstationarity of source signal by using HMM and the temporal correlation of source signal by using temporally correlated GGD model as source model. The ML approach is developed to update the model parameters by EM algorithm and the source signal is estimated by MAP approach. Results have shown high performance accuracy and fast convergence of the proposed algorithm. Significant improvement has been obtained in comparison with nonlinear algorithm that does not track the source signal nonstationarity.

#### REFERENCES

[1] J.K. Lin, D.G. Grier, and J.D. Cowan, "Source separation and density estimation by faithful equivariant SOM," in *Advances in Neural Inform. Process. Systems*, Cambridge, MIT Press, vol. 9, 1997.

[2] G. Burel, "Blind separation of sources: A nonlinear neural algorithm," *Neural Networks*, pp. 937-947, 1992.

[3] H. Valpola and J. Karhunen, "An unsupervised ensemble learning method for nonlinear dynamic state-space models," *Neural Computation*, vol. 14, pp. 2647-2692, 2002.

[4] H. Valpola, E. Oja, A. Ilin, and J. Karhunen, "Nonlinear blind source separation by variational Bayesian learning," *IEICE Trans. on Fundamentals of Elecs, Comms. and Comp. Sciences*, vol. E86-A, pp. 532-541, 2003.

[5] A. Taleb and C. Jutten, "Source separation in post-nonlinear mixtures," *IEEE Trans. on Signal Processing*, vol. 47, no. 10, pp. 2807-2820, Oct. 1999.

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

[7] T.W. Lee, "Blind source separation of nonlinear mixing models," *Neural Networks*, pp. 121-131, 1997.

[8] W.L. Woo and L. C. Khor, "Blind restoration of nonlinearly mixed signals using multilayer polynomial neural network," *IEE Proc. on Vision, Image and Signal Processing*, vol. 151, no. 1, pp. 51-61, Feb. 2004.

[9] W.L. Woo and S.S. Dlay, "Neural network approach to blind signal separation of mono-nonlinearly mixed sources," *IEEE Trans. on Circuits and System I*, vol. 52, no. 6, pp. 1236-1247, Jun. 2005.

[10] S. Harmeling, A. Ziehe, M. Kawananbe and K-R. Muller, "Kernel-based nonlinear blind source separation," Neural Computation, vol. 15, no. 5, pp. 1089-1124, 2003.

[11] L.R. Rabiner, "A toturial on hidden Markov models and selected applications in speech recognition," *Proc. of IEEE*, vol. 77, no. 2, pp. 257-286, Feb. 1989.

[12] J. Zhang, W.L. Woo and S.S. Dlay, "Blind source separation of postnonlinear convolutive mixture," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 15, no. 8, pp. 2311-2330, Nov. 2007.