Nonlinear independent component analysis by learning generalized adalines

Jiann-Ming Wu†, Yi-Cyun Yang

Abstract— **This work proposes the novel modeling of post-nonlinear mixtures of independent sources and a learning method for the reverse problem that addresses on concurrent estimation of model parameters and independent components subject to given multichannel observations. The proposed post-nonlinear mixture model is realized by a network of multiple generalized adalines(gadalines), where each weighted gadaline emulates a transmitting link that maps independent sources to single channel observations. Based on the post-nonlinear mixture assumption, learning multiple weighted gadalines for retrieving independent components is resolved by the leave-one-out approximation operated under the mean-field-annealing process. Each time for some selected single channel observations, the dominant independent component is refined to compensate for the error of approximating the selected channel by the remaining independent components. This work shows that interactive dynamics derived for gadaline optimization executed under the mean field annealing is accurate and reliable for blind separation of post-nonlinear mixtures of independent sources.**

Keywords— **Blind source separation, post-nonlinear mixtures, leave-one-out approximation, mean field annealing, gadaline optimization**

I. Introduction

Independent component analysis (ICA)[1]-[4] has been extensively applied to blind source separation (BSS) of real world signals, such as electrocardiograms(ECG), electroencephalograms(EEG), event related potential(ERP) and magnetic resonance images(MRI). The formation of multichannel signals has been typically modeled following the linear mixture assumption, under which multichannel observations, $\{\mathbf{x[t]}\}_\mathbf{t}$, are regarded as a sample from linear mixtures of independent sources,

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{1}$$

where $\mathbf{s} = (\mathbf{s_1}, \mathbf{s_2}, ..., \mathbf{s_d})^\mathbf{T}$ denotes collection of independent sources and $\mathbf{A}$ is an unknown invertible $d \times d$ mixing matrix. Under the linear mixture assumption independent sources can be theoretically recovered by the following demixing process,

$$\mathbf{y[t]} = \mathbf{W}\mathbf{x[t]},$$

where the product of $\mathbf{W}$ and $\mathbf{A}$ or its permutation is a diagonally dominant matrix. Independent component analysis has been well resolved by minimization of statistical criteria, such as the Kurtosis or Kullback-Leibler divergence[7][8], however its effectiveness could be only guaranteed under the linear mixture assumption.

Correspondence: Jiann-Ming Wu, Department of Applied Mathematics, National Dong Hwa University, Hualien,Taiwan. Tel. 8863-8633531, FAX : 8863-8633510, email: jmwu@mail. ndhu.edu.tw

The formation of multichannel observations is extended in this work to the following post-nonlinear(PNL) mixture model,

$$\begin{aligned} \mathbf{x} &= F(\mathbf{h} = \mathbf{A}\mathbf{s}) \\ &= (f_1(h_1 = \mathbf{a_1}\mathbf{s}), ..., \mathbf{f_d}(\mathbf{h_d} = \mathbf{a_d}\mathbf{s}))^T \end{aligned} \tag{2}$$

where F denotes a function vector and $\mathbf{a_i}$ denotes the ith row of the mixing matrix $\mathbf{A}$. Since each function element f_i in F comes behind a linear projection, it is termed as a post-nonlinear function[5][6]. Of independent sources PNL mixtures reduce to linear mixtures if each f_i is restricted to be linear. Provided without PNL function elements and the mixing matrix, PNL independent component analysis is aimed to recover independent sources subject to given multichannel observations that are assumed as PNL mixtures of independent sources. The PNL mixture model is more general than the linear mixture model for formation emulation of multichannel observations. Effective PNL independent component analysis is therefore expected more practical for blind separation of real world signals whose formation is probably beyond the scope well characterized by linear mixtures of independent sources.

Neural networks of weighted gadalines that are generalized from adalines of Widrow[9] have been proposed for data driven function approximation via supervised learning. The weighted gadaline that carries out an adaptive PNL projection[13] is employed to emulate the PNL transmitting link that maps independent sources to single channel observations. By emulation using weighted gadalines, the PNL ICA problem is translated to concurrent estimation of a gadaline network and independent components subject to multichannel observations which are assumed as a sample from PNL mixtures of independent sources.

Effective concurrent estimation of weighted gadalines and independent components requires highly accurate and reliable collective decisions on determination of tremendous unknowns. The learning method proposed for the PNLICA problem iteratively operates the leave-one-out approximation, which is realized by execution of interactive dynamics derived for single gadaline optimization, under the mean field annealing process[10]-[12]. Each time for some selected single channel observations, the dominant independent component is refined to compensate for the error of approximating the selected channel by the remaining independent components.

This paper is organized to give details of emulating the PNL mixture model using multiple weighted gadalines in the upcoming section, present the proposed learning

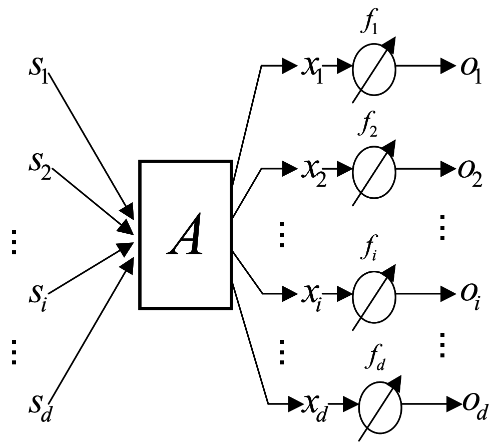

Fig. 1. The PNL mixture model

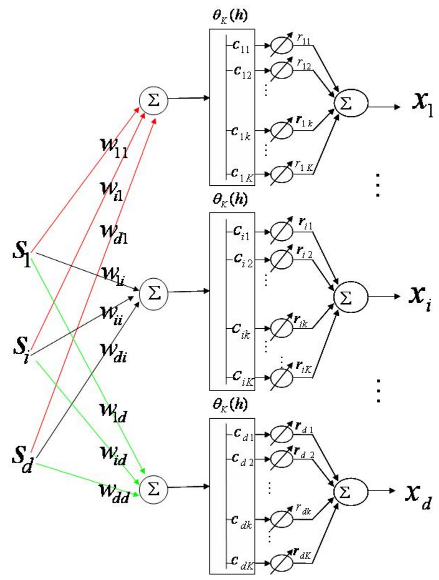

Fig. 2. A gadaline network for emulation of the PNL mixture model

method in section III, show numerical results and arrive at our conclusions in the final section.

II. Post-nonlinear mixture model

The block diagram in Figure 1 shows the PNL mixture model of equation (2) that is emulated by the gadaline network shown in Figure 2, where each weighted gadaline implements a PNL transmitting link from independent components to single channel observations.

Gadaline [13] is an abbreviation of the Generalized ADAptive LINear Element that is a multi-state neural processing element generalized from the adaline of Widrow [9]. As shown in figure 2, a gadaline is composed of a receptive field $\mathbf{w}$ and a K-state transfer function that maps the linear projection, $h=\mathbf{w^T s}$, to a unitary vector of K binary bits like

$$\theta_K(h;\mathbf{c}) = \left\{ \begin{array}{c} \mathbf{e_1^K} \text{ if } \mathbf{h} \in \mathbf{I_1} \\ \vdots \\ \mathbf{e_i^K} \text{ if } \mathbf{h} \in \mathbf{I_i} \\ \vdots \\ \mathbf{e_K^K} \text{ if } \mathbf{h} \in \mathbf{I_K} \end{array} \right\},$$

where $\mathbf{c} = (\mathbf{c_1}, \mathbf{c_2}, ..., \mathbf{c_K})^{\mathbf{T}}$ contains K distinct knots which partition the function domain into K disjoint intervals, each indicated by $I_i = \{h | i = \arg\min_k |h - c_k|\}$, and $\mathbf{e_i^K}$ denotes a unitary vector with the only active bit at the ith position. In response to an input vector, a weighted gadaline has an output that measures the product of a K-state activation and a posterior weight,

$$y = \mathbf{r^T} \boldsymbol{\theta}_\mathbf{K}(\mathbf{w^T s}; \mathbf{c}) \tag{3}$$

where $\mathbf{r} = (\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_K})^{\mathbf{T}}$. In the previous work [13], a weighted gadaline is organized to perform a post-nonlinear projection with adaptive post-nonlinearity to clamped paired data. The gadaline network shown in figure 2 serves as a post-nonlinear mixture model which is adaptive both in linear and nonlinear parts. For blind separation of post-nonlinear mixtures of independent sources, estimation of network parameters subject to given multichannel observations is resolved by the learning method proposed in the upcoming section.

III. Post-nonlinear ICA

A. Gadaline optimization

Gadaline optimization subject to given paired data, $\{(\mathbf{s[t]}, \mathbf{y[t]})\}_\mathbf{t}$, for a single transmitting link is introduced in this subsection. The first entry of the paired data denotes independent sources and the second entry denotes single channel observations. Gadaline optimization[13] that has been applied for data driven function approximation is aimed to optimize parameters of a weighted gadaline, including the receptive field $\mathbf{w}$, knot vector $\mathbf{c}$ and posterior weight $\mathbf{r}$ subject to constrains proposed by given paired data.

The output of a weighted gadaline in response to the projection $h = \mathbf{w^T s}$ is characterized by the following conditional probability density function(pdf),

$$q(y|h \in I_k) = f(y; r_k, \sigma_y^2), \tag{4}$$

where f denotes the normal pdf with mean r_k and variance σ_y^2. Combining equation (3) and equation (4), we have

$$q(y|h) = \sum_{k=1}^{K} \theta_K^T(\mathbf{w^T s[t]}; \mathbf{c}) \mathbf{e_k^K} \mathbf{f}(\mathbf{y}; \mathbf{r_k}, \boldsymbol{\sigma}_\mathbf{y}^\mathbf{2}) \tag{5}$$

Define

$$\boldsymbol{\delta}[\mathbf{t}] = \boldsymbol{\theta}_\mathbf{K}^\mathbf{T}(\mathbf{h[t]} = \mathbf{w^T s[t]}; \mathbf{c}).$$

Collection of projections on $\mathbf{w}$, denoted by

$$H_k = \{h[t] | \boldsymbol{\delta}[\mathbf{t}] = \mathbf{e_k^K}\},$$

is assumed as a sample from the normal pdf with the mean c_k and variance σ_h^2. Fitting the joint pdf of y and h to all $(h[t], y[t])$ with $h[t]$ belonging H_k leads to the following average log likelihood,

$$\begin{aligned} l_k &= \frac{1}{N_k} \log \prod_{t:\boldsymbol{\delta}[\mathbf{t}]=\mathbf{e_k^K}}^{N} q(y[t]|h[t]) f(h[t]; c_k, \sigma_h^2) \\ &= \frac{1}{N_k} \sum_{t:\boldsymbol{\delta}[\mathbf{t}]=\mathbf{e_k^K}}^{N} \log\{f(y[t]; r_k, \sigma_y^2) f(h[t]; c_k, \sigma_h^2)\}, \end{aligned}$$

where N_k denotes the size of the set H_k and $\delta_k[t]$ is the kth element of $\boldsymbol{\delta}[\mathbf{t}]$. Summing up all l_k leads to the following criterion,

$$\begin{aligned} & E(\boldsymbol{\delta}, \mathbf{w}, \mathbf{c}, \mathbf{r}, \boldsymbol{\sigma}) \\ = & -\sum_{k=1}^{K} \frac{N_k}{N} l_k \\ = & -\frac{1}{N} \sum_{k=1}^{K} \sum_{t:\boldsymbol{\delta}[\mathbf{t}]=\mathbf{e_k^K}}^{N} \log\{f(y[t]; r_k, \sigma_y^2) f(h[t]; c_k, \sigma_h^2)\} \\ = & -\frac{1}{N} \sum_{k=1}^{K} \sum_{t=1}^{N} \boldsymbol{\delta}[\mathbf{t}] \mathbf{e_k^K} \log\{\mathbf{f}(\mathbf{y}[\mathbf{t}]; \mathbf{r_k}, \boldsymbol{\sigma}_\mathbf{y}^\mathbf{2}) \mathbf{f}(\mathbf{h}[\mathbf{t}]; \mathbf{c_k}, \boldsymbol{\sigma}_\mathbf{h}^\mathbf{2})\} \\ = & -\frac{1}{N} \sum_{t=1}^{N} \sum_{k=1}^{K} \delta_k[t] \log f(y[t]; r_k, \sigma_y^2) \\ & -\frac{1}{N} \sum_{k=1}^{K} \sum_{t=1}^{N} \delta_k[t] \log f(h[t]; c_k, \sigma_h^2) \qquad (6) \end{aligned}$$

whose minimization subject to

$$\begin{aligned} \sum_{k=1}^{K} \delta_k[t] &= 1, \forall t \\ \delta_k[t] &\in \{0, 1\}, \forall t, k, \end{aligned}$$

induces a mathematical framework for gadaline optimization, where $\boldsymbol{\delta}$ denotes collection of all $\boldsymbol{\delta}[\mathbf{t}]$ and $\boldsymbol{\sigma} = \{\boldsymbol{\sigma}_\mathbf{h}, \boldsymbol{\sigma}_\mathbf{y}\}$.

Since of consisting of both continuous and discrete variables, the fitting criterion of equation (6) is minimized by a hybrid of the mean-field-annealing (MFA) and gradient decent methods. $\boldsymbol{\delta}$ is randomized according to the Boltzmann assumption[12] and all its elements are assumed to be statistically independent. It follows that the Hopfield-like free energy[10] that sums up individual entropies of all elements in $\boldsymbol{\delta}$ and the mean of the fitting criterion E can be expressed as follows,

$$\begin{aligned} & \psi(\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{c}, \mathbf{r}, \boldsymbol{\sigma}) \\ = & E(\mathbf{v}, \mathbf{w}, \mathbf{c}, \mathbf{r}, \boldsymbol{\sigma}) + \sum_{\mathbf{t=1}}^{\mathbf{N}} \sum_{\mathbf{k=1}}^{\mathbf{K}} \mathbf{v_k}[\mathbf{t}] \mathbf{u_k}[\mathbf{t}] \\ & -\frac{1}{\beta} \sum_{t=1}^{N} \log \left[\sum_{k=1}^{K} \exp(\beta u_k[t]) \right], \end{aligned}$$

where all $u_k[t]$ are auxiliary variables. Directly setting

$$\begin{cases} \frac{\partial \psi}{\partial u_k[t]} = 0 \\ \frac{\partial \psi}{\partial v_k[t]} = 0 \end{cases}, \forall t, k$$

$$\begin{cases} \frac{\partial E(\mathbf{v}, \mathbf{w}, \mathbf{c}, \mathbf{r}, \boldsymbol{\sigma})}{\partial c_k} = 0 \\ \frac{\partial E(\mathbf{v}, \mathbf{w}, \mathbf{c}, \mathbf{r}, \boldsymbol{\sigma})}{\partial r_k} = 0 \\ \frac{\partial E(\mathbf{v}, \mathbf{w}, \mathbf{c}, \mathbf{r}, \boldsymbol{\sigma})}{\partial \sigma_h} = 0 \\ \frac{\partial E(\mathbf{v}, \mathbf{w}, \mathbf{c}, \mathbf{r}, \boldsymbol{\sigma})}{\partial \sigma_y} = 0 \\ \frac{\partial E(\mathbf{v}, \mathbf{w}, \mathbf{c}, \mathbf{r}, \boldsymbol{\sigma})}{\partial \mathbf{w}} = \mathbf{0} \end{cases}$$

leads to the following updating rules that determine the saddle point of the free energy function,

$$\begin{aligned} u_k[t] &= -\frac{1}{2N\sigma_h^2} \left(\mathbf{w^T s}[\mathbf{t}] - \mathbf{c_k} \right)^2 \qquad (7) \\ & \quad -\frac{1}{2N\sigma_y^2} (y[t] - r_k)^2 \\ v_k[t] &= \frac{\exp(\beta u_k[t])}{\sum_{l=1}^{K} \exp(\beta u_l[t])} \qquad (8) \end{aligned}$$

$$\begin{aligned} c_k &= \frac{\sum_{t}^{N} v_k[t] h[t]}{\sum_{t=1}^{N} v_k[t]}, \qquad (9) \\ r_k &= \frac{\sum_{t=1}^{N} v_k[t] y[t]}{\sum_{t=1}^{N} v_k[t]}, \qquad (10) \\ \sigma_h &= \left[\frac{1}{T} \sum_{t=1}^{N} \sum_{k=1}^{K} \delta_k[t] \left(\mathbf{w^T s}[\mathbf{t}] - \mathbf{c_k} \right)^2 \right]^{1/2}, \qquad (11) \\ \sigma_y &= \left[\frac{1}{T} \sum_{t=1}^{N} \sum_{k=1}^{K} \delta_k[t] (y[t] - r_k)^2 \right]^{1/2}. \qquad (12) \\ \mathbf{w} &= \mathbf{B}^{+} \mathbf{b}. \qquad (13) \end{aligned}$$

where $\mathbf{B}^{+}$ denotes the pseudo inverse of $\mathbf{B}$ with elements

$$\begin{aligned} B_{ji} &= \sum_{t=1}^{N} s_j[t] s_i[t], \\ b_j &= \sum_{k=1}^{K} c_k \left(\sum_{t=1}^{N} v_k[t] s_j[t] \right). \end{aligned}$$

B. Leave-one-out approximation

The leave-one-out approximation operated under the mean field annealing process is employed for concurrent estimation of network parameters and independent components subject to given multichannel observations. Let $\mathbf{x}[\mathbf{t}]$ denote multichannel observations and all $\mathbf{x}[\mathbf{t}]$ be a sample from PNL mixtures of independent sources and $\mathbf{o}[\mathbf{t}]$ denote the instance of unknown independent sources, where t runs from 1 to N, and $\mathbf{x}[\mathbf{t}] = (\mathbf{x_1}[\mathbf{t}], \mathbf{x_2}[\mathbf{t}], ..., x_d[t])^T$.

The proposed learning method iteratively refines estimation of independent sources and network parameters. According to the PNL mixture model emulated in figure 2, the observation x_i of the ith channel could be approximated by post-nonlinear projection of estimated independent components other than the ith independent component and the obtained approximating error could be used to define the ith independent component. Since both independent components and the gadaline network are unknown to the learning method, their uncertainty is compensated by the mean-field-annealing process, which schedules the temperature-like parameter $\frac{1}{\beta}$ gradually from high to low values to emulate physical annealing.

Let n denote the iteration number and $\mathbf{s}[\mathbf{n},\mathbf{t}] = (\mathbf{s_1}[\mathbf{n},\mathbf{t}], ..., s_d[n,t])^{\mathbf{T}}$ denote the approximation to $\mathbf{o}[\mathbf{t}]$ maintained at the nth iteration by the learning method. Initially, n is set to zero and $\mathbf{s}[\mathbf{n},\mathbf{t}]$ is directly set to $\mathbf{x}[\mathbf{t}]$ for $n = 0$. At each iteration, the learning method approximates observations $\{x_i[t]\}_t$ of each channel i by the output of a weighted gadaline in responding inputs belonging $\{s_j[n,t]\}_{t,j\neq i}$. Following equation (6), parameters of the ith weighted gadaline, including $\mathbf{w}, \mathbf{c}, \mathbf{r}$ and $\boldsymbol{\sigma}$, and $\mathbf{v}$, are optimized by interactive dynamics (7-13), where the gadaline input $\mathbf{s}[\mathbf{t}]$ is set to have only elements in $\{s_j[n,t]\}_{j\neq i}$ and the gadaline output $y[t]$ is set to $x_i[t]$ for all t. The approximating error obtained by gadaline optimization is expressed by

$$e_i[n,t] = x_i[t] - \sum_k v_k[t] r_k, \tag{14}$$

which is employed to refine current $\{s_i[n,t]\}_t$ for deriving $\{s_i[n+1,t]\}_t$. Gadaline optimization is further applied to determine optimal scalar inputs to a weighted gadaline whose expected output is set to $e_i[n,t]$. For the purpose the following updating rule should be recruited to interactive dynamics (7-12),

$$s[t] = \frac{\sum_k c_k v_k[t]}{\sum_k v_k[t]}. \tag{15}$$

The obtained optimal scalar inputs are assigned to $\{s_i[n+1,t]\}_t$ to compensate for the error of approximating the ith channel by remaining independent components. The proposed learning method iteratively estimates independent sources one by one under the mean field annealing process until the halting condition holds. The learning method for blind separation of PNL mixtures of independent sources is summarized by the following stepwise procedure.

1. Set $n = 0$, $s_i[n,t] = x_i[t], \forall i, t$, and β to a sufficiently low value.
2. For each i, sequentially execute the following steps.
 (a) Set $\mathbf{s}[\mathbf{t}]$ to have all elements belonging $\{s_j[n,t]\}_{j\neq i}$ and $y[t] = x_i[t]$ for all t.
 (b) Find $\mathbf{v}, \mathbf{w}, \mathbf{c}, \mathbf{r}$ using equation (7)$\sim$(13).
 (c) Determine $e_i[n,t]$ by equation (14) $\forall t$.
 (d) Use interactive dynamics (7-12) and (15) to determine optimal inputs to a weighted gadaline whose expected output $y[t]$ is clamped to $e_i[n,t]$ and replace $s_i[n+1,t]$ with the obtained optimal input for all t.
3. If the halting condition holds exit, otherwise set $n = n+1$, $\beta = \frac{\beta}{0.995}$ and go to step 2.

IV. Numerical simulations and conclusions

The first example tests the proposed learning method for blind separation of linear mixtures of independent sources. The gadaline network shown in figure 2 reduces to a linear network if each knot vector $\mathbf{c}$ is fixed to have elements that partition the bounded range of projections, such as $[-5, 5]$, into equal lengthed intervals and each posterior weight is fixed to $\mathbf{c}$. For the linear case, the step 2d is simplified to replace $s_i[n+1,t]$ with $\phi(e_i[n,t])$, where

$$\phi(z) = \sum_{k=1}^{K} v_k[t] c_k,$$

and

$$\begin{cases} u_k[t] = (z - c_k)^2 \\ v_k[t] = \frac{\exp(-\beta u_k[t])}{\sum_{l=1}^{K} \exp(-\beta u_l[t])} \end{cases}.$$

At sufficiently low β, the transfer function ϕ maps all its inputs to zero for compensating high uncertainty of independent source estimation. After the parameter β is gradually increased by the mean field annealing process, the transfer function ϕ tends to act similar to a linear threshold function. For extremely large β, independent source estimation operates with confident certainty. The proposed learning method for the linear case is applied to process multichannel observations shown in figure 4, which are linear mixtures of independent sources of figure 3 through a random square matrix created by

$$A = I_d + (\text{rand}(d,d) - 0.5) \times 0.6 \tag{16}$$

where $\text{rand}(d,d)$ denotes a random matrix whose $d \times d$ elements are sampled from a uniform random variable ranging within [0 1], d denotes the channel number of independent sources, and I_d is an identity matrix. The obtained independent components shown in figure 5 well recover independent sources of figure 3.

Multichannel observations shown in figure 7 are PNL mixtures of independent sources of figure 6 created by

$$\begin{pmatrix} \sin(\frac{2\pi t}{800}) \\ \sin(\frac{2\pi t}{300} + 6\cos(\frac{2\pi t}{60})) \\ \sin(\frac{2\pi t}{90}) \end{pmatrix}, \text{ for } t = 1, ..., 800,$$

in the previous work[8]. The used PNL mixture model has its linear part generated by the rule (16) and PNL part realized by three hyper tangent functions. Without given the mixing matrix A and PNL functions, the proposed learning method is employed to estimate the PNL mixture model and independent components from multichannel observations of figure 7. As shown in figure 8, the three PNL functions estimated by the proposed learning method are invertible sigmoid-like functions. Through their inverse functions multichannel observations of figure 7 could be transformed to linear mixtures of independent sources. We further applied the proposed learning method to learn a linear mixture network for blind separation of the transformed data and attained independent components of figure 9 which shows well recovery of independent sources of figure 6.

Numerical results show the proposed learning method is effective for linear and post-nonlinear independent component analysis. The leave-one-out approximation operated under the mean field annealing process is feasible for concurrent estimation of the gadaline network and independent components. Each time for selected single channel observations, the dominant independent component is regarded absent from contributing to the selected channel and is refined to compensate for the approximating error of the selected channel by the remaining independent components. The idea is simple but its implementation needs accurate and reliable collective decisions on determination of tremendous discrete and continuous unknowns. This work shows execution of interactive dynamics derived for gadaline optimization under the mean field annealing process is potential for fitting the computational requirement of concurrent estimation of the gadaline network and independent components.

Effective learning of the PNL mixture model proposed in this work significantly extends the application domain of blind separation of real world signals since the traditional ICA algorithm is impractical for blind separation of PNL mixtures of independent sources. The PNL mixture assumption which is more general for formation emulation of real world signals than the linear mixture assumption. The proposed learning method translates PNLICA to individual sub-tasks of gadaline optimization and operates under the mean field annealing process for accurate neural computations. Its derivation avoids formulating and minimizing complicate statistical criteria, such as the Kurtosis or Kullback Leibler divergence[7][8], for quantifying dependency of random multivariates that are PNL mixtures of independent sources. Under the PNL mixture assumption, effective reduction of statistical dependency of independent components through direct minimization of the KL divergence is a complicate task that is still challenging researchers in the field of neural networks for nonlinear independent component analysis. Future works will focus on further verification of applying the proposed learning method for nonlinear independent component analysis of real world signals, such as electrocardiograms(ECG), electroencephalograms(EEG), event related potential(ERP) and magnetic resonance images(MRI).

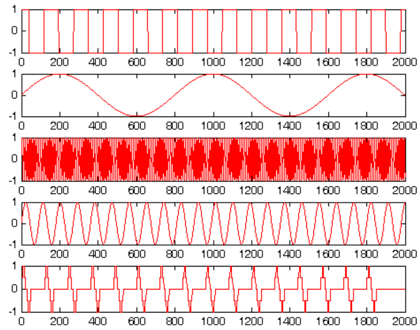

Fig. 3. Independent sources

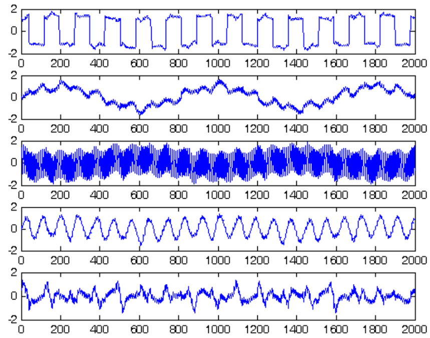

Fig. 4. Multichannel observations derived by linear mixtures of independent sources

References

[1] P. Comom, Independent component analysis, a new concept, Signal Process, vol. 36, no. 3, pp. 287-314, Apr 1994.

[2] J. F. Cardoso, Blind Signal Separation: Statistical Principles, PROCEEDINGS OF THE IEEE 86 (10): 2009-2025 OCT 1998

[3] J. F. Cardoso, High-order contrasts for independent component analysis, Neural Computation, vol. 11, pp. 157-192, 1999.

[4] R. Boscolo, H. Pan, and V. Roychowdhury, Independent component analysis based on nonparametric density estimation, IEEE Trans. Neural Networks, vol. 15, no. 1, pp. 55-65, 2004.

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

[6] L. B. Almeida, Linear and nonlinear ICA based on mutual information, MISEP method, Signal Processing, vol. 84, pp. 231-245, 2003.

[7] S. I. Amari, A. Cichocki and H. H. Yang, A new learning algorithm for blind signal separation, Advances in Neural Information Processing Systems, vol 8, pp. 757-763, 1996.

[8] J. Basak, and S. Amari, Blind source separation of a mixture of uniformly distributed source signal: a novel approach, Neural Computation, vol. 11, no. 4, pp. 1011-1034. 1999.

[9] B.Widrow, Generalization and information storage in networks of adaline neuron in self-organizing systems, M. Yovitz, G. Jacobi, and G. Goldstein, Eds. Washington, DC: Spartan, pp. 435-461, 1962.

[10] Peterson C. & Söderberg B., A new method for mapping optimization problems onto neural network, Int. J. Neural Syst. 1(3), 1989.

[11] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, Upper Saddle River, 1999.

[12] J. M. Wu, and S. J. Chiu, Independent component analysis using Potts models, IEEE Trans. Neural Networks, vol. 12, no. 2, pp. 202-211, Mar 2001.

[13] J. M. Wu, Z. H. Lin, and P. H. Hsu, Function approximation using generalized adalines, IEEE Trans. Neural Networks., vol. 17, no. 3, pp. 541-558, May 2006.

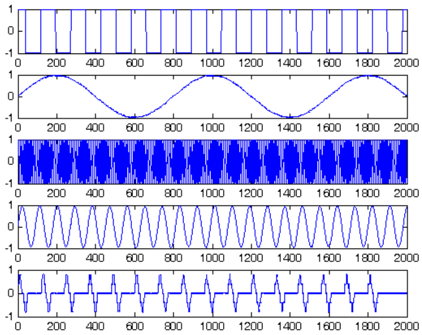

Fig. 5. Independent components obtained by the proposed learning method for a linear network

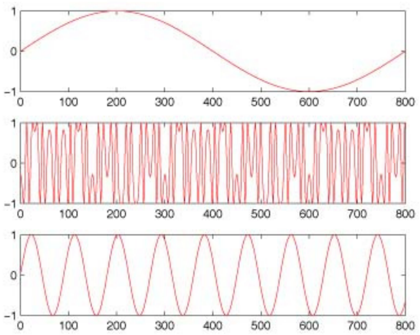

Fig. 6. Independent sources

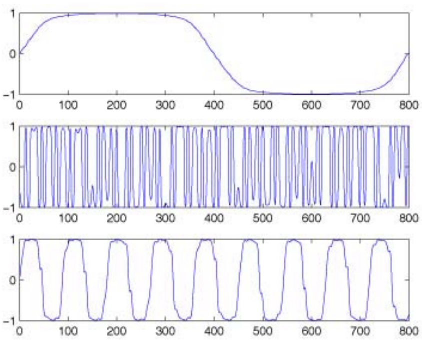

Fig. 7. Multichannel observations sampled from PNL mixtures of independent sources

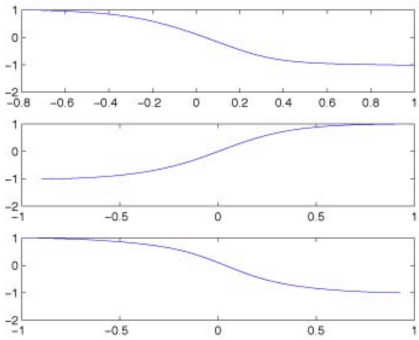

Fig. 8. PNL functions derived by the proposed learning method

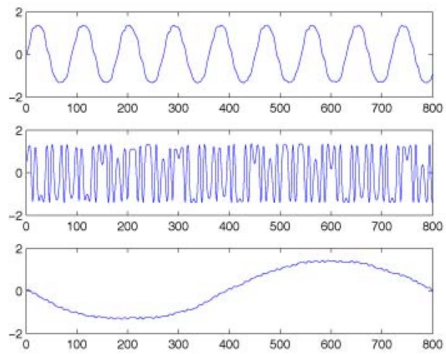

Fig. 9. Independent components estimated by the proposed learning method for blind separation of multichannel observations of figure 7