BLIND SOURCE SEPARATION OF NONLINEARLY CONSTRAINED MIXED SOURCES USING POLYNOMIAL SERIES REVERSION

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

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

ABSTRACT

A novel polynomial-based neural network is proposed for nonlinear blind source separation. We focus our research on a recently presented mono-nonlinearity mixture where a linear mixing matrix is slotted into two mutually inverse nonlinearities. In this paper, we generalize the mono-nonlinearity mixing system to the situation where different nonlinearities are applied to the source signals. The theory of Series Reversion is merged with the neural network demixer to perform two layers of mutually inverse nonlinearities. The corresponding parameter learning algorithm for the proposed polynomial-based neural network demixer is also presented. Simulations have been carried out to verify the efficacy of the proposed approach. We demonstrate that the proposed network can successfully recover the original source signals in a blind mode under nonlinear mixing conditions.

1. INTRODUCTION

Rapid developments during the last decade have seen Independent Component Analysis (ICA) emerge as one of the most powerful tools in Blind Signal Separation [1-3]. Generally, the problem of the blind separation of independent sources involves a set of observations $\mathbf{x}=\begin{bmatrix} x_1(t) & x_2(t) & \cdots & x_p(t)\end{bmatrix}^T$ which are generated from a set of unknown independent components $\mathbf{s}=\begin{bmatrix} s_1(t) & s_2(t) & \cdots & s_q(t)\end{bmatrix}^T$ according to

$$x_i = f_i\left(s_1, s_2, \ldots, s_q\right) \tag{1}$$

where f_i is an unknown differentiable bijective mapping, $i=1,2,\ldots,p$ and *t* is the time or sample index. A technique known as Independent Component Analysis (ICA) is exploited to estimate both the mixing mappings f_i's and the original sources $s_i(t)$, $i=1,2,\ldots,q$. A popular assumption by most ICA algorithm is that the mixing mapping takes the form of the linear combination, i.e. $f_i\left(s_1, s_2, \ldots, s_q\right)=m_{i1}s_1+m_{i2}s_2+\cdots+m_{iq}s_q$. However, for many practical problems, mixed signals are more likely to subject to some kind of nonlinear distortions due to sensory or environmental limitations [2]. Hence, the search for an algorithm tailored specifically for nonlinear blind source separation has become increasingly important at both theoretical and practical levels.

The contribution of this paper is as follows: Firstly, a multi-nonlinearly constrained system is proposed as the mixing and demixing model. The model generalizes the original mono-nonlinearity model previously presented in [2]. The proposed model is a more general description than the post-nonlinear systems [5] and provides a better representation of a nonlinear mixture. Secondly, a new polynomial-based neural network demixer is proposed and developed as the separation system to estimate the unknown source signals. Finally, the theory of Series Reversion is incorporated into the derivation of the parameter learning algorithm to account for the special structure of the demixing network.

2. NONLINEAR ICA MODEL

A mono-nonlinearity mixing model derived from the theory of functional analysis was proposed in [2] to provide a general description of the mixing system in the following form:

$$\mathbf{x} = f\left(\mathbf{M}f^{-1}(\mathbf{s})\right) \tag{2}$$

where $\mathbf{M}=\begin{bmatrix}\mathbf{m}_1 & \mathbf{m}_2 & \cdots & \mathbf{m}_p\end{bmatrix}^T$ with dimension $p\times q$ and $\mathbf{m}_i=\begin{bmatrix} m_{i1} & m_{i2} & \cdots & m_{iq}\end{bmatrix}^T$. In this paper, we assume that the number of sources is equal to that of observations, i.e. $p=q=N$. This model is structured in the form of one linear mixing matrix sandwiched between two layers of nonlinearities, one of which is the inverse function of the other. The term 'mono-nonlinearity' represents the condition where an identical nonlinear distortion is applied to each source signal. However, there is no guarantee that this condition is always fulfilled in practice. In fact, the channels between observations and sources are arbitrarily distorted due to the uncertainty of the environment. Hence, to preserve the special relationship between the two layers in (2), we represent the 'multi-nonlinearity' constrained mixing system by the following model:

$$\mathbf{x} = \mathbf{D}_{\mathbf{f}} * (\mathbf{M}\mathbf{D}_{\mathbf{f}^{-1}} * \mathbf{s}) \tag{3}$$

where

$$\begin{aligned}
&\mathbf{D}_{\mathbf{f}} = \operatorname{diag}\left[f_1 \quad f_2 \quad \cdots \quad f_N\right] \\
&\mathbf{D}_{\mathbf{f}^{-1}} = \operatorname{diag}\left[f_1^{-1} \quad f_2^{-1} \quad \cdots \quad f_N^{-1}\right] \\
&\mathbf{D}_{\mathbf{g}} * \mathbf{u} = \operatorname{diag}\left[g_1 * u_1 \quad g_2 * u_2 \quad \cdots \quad g_N * u_N\right] \\
&g_i * u_i = g_i(u_i)
\end{aligned}$$

This model will reduce to the mono-nonlinearity mixing model when $f_1 = f_2 = \cdots = f_N = f$ and can further represent a linear mixture as a special case if $\{f_i\}_{i=1}^N$ is linear.

A demixing system for (3) can be described by the inverse of the mixing system where the original sources are estimated as follows:

$$\hat{\mathbf{s}} = \mathbf{D}_{\mathbf{f}^{-1}}^{-1} * (\mathbf{M}^{-1}\mathbf{D}_{\mathbf{f}}^{-1} * \mathbf{x}) \tag{4}$$

Using the identity $\mathbf{D}_{\mathbf{g}}^{-1} = \mathbf{D}_{\mathbf{g}^{-1}}$, (4) can be rewritten as

$$\hat{\mathbf{s}} = \mathbf{D}_{\mathbf{f}} * (\mathbf{W}\mathbf{D}_{\mathbf{f}}^{-1} * \mathbf{x}) \tag{5}$$

where $\mathbf{W}$ is the demixing matrix. Given the observed signals, the aim is to estimate $\{f_i\}_{i=1}^N$ and $\mathbf{W}$ such that the resulting transformed signals are mutually as independent as possible and statistically as close as possible to the source signals.

3. POLYNOMIAL-BASED NEURAL NETWORK FOR NONLINEAR ICA

In current literature, popular nonlinear network demixers such as SOM, GTM, RBF [4] and MLP with sigmoidal nonlinearity [7] are inherently nonlinear because of the fixed nonlinearities in the hidden neurons. However, the fixed rigidness of the nonlinearity will lead to the oversized and overfitted network and inevitably increase computational complexity [2]. Instead of using a fixed form of nonlinearity in the hidden neurons, we propose to design a demixer whereby its intrinsic nonlinearity can be flexibly controlled.

3.1. Polynomial-based Network as the Nonlinear ICA Demixing System

The Weierstrass Approximation Theorem states that for every continuous function $\phi:[\alpha,\beta]\to\mathbf{R}$, there always exists a polynomial series $p(u)=\sum_{m=0}^{M}\lambda_m u^m$, parameterized by $\boldsymbol{\theta}=\left\{M,\ \{\lambda_m\}_{m=0}^{M}\right\}$, which can uniformly approximate ϕ with arbitrary accuracy. Therefore, a feedforward polynomial-based network shown in Figure 1 is proposed to reflect the model in (5). The hidden layer neurons in the proposed network perform the polynomial series to approximate the mixing mapping functions $\{f_i\}_{i=1}^N$ and $\left\{f_i^{-1}\right\}_{i=1}^N$. The outputs of the demixing system assume the following form

$$\begin{aligned}
&\mathbf{y}_{[3]} = \mathbf{D}_{\mathbf{f}} * \mathbf{y}_{[2]} = \sum_{m=0}^{M_1}\left(\mathbf{a}_m \circ \mathbf{y}_{[2]}^m\right) \\
&\mathbf{y}_{[2]} = \mathbf{W}\mathbf{y}_{[1]} \\
&\mathbf{y}_{[1]} = \mathbf{D}_{\mathbf{f}^{-1}} * \mathbf{x} = \sum_{n=1}^{M_2}\left(\mathbf{b}_n \circ \left(\mathbf{x}-\mathbf{a}_0\right)^n\right)
\end{aligned} \tag{6}$$

where $\mathbf{y}_{[i]} = \left[y_{[i,1]} \ \cdots \ y_{[i,N]}\right]^T$, $\mathbf{a}_m = \left[a_{[m,1]} \ \cdots \ a_{[m,N]}\right]^T$, $\mathbf{b}_n = \left[b_{[n,1]} \ \cdots \ b_{[n,N]}\right]^T$, $y_{[j,i]}$ denotes the i^{th} output of the j^{th} layer in the demixer, $\left\{a_{[m,i]}\right\}_{m=0;i=1}^{m=M_1;i=N}$ and $\left\{b_{[n,i]}\right\}_{n=1;i=1}^{n=M_2;i=N}$ are the coefficients, M_1 and M_2 represents the order of the series expansion and '$\circ$' denotes the Hadamard product.

3.2. Series Reversion

As shown in Figure 1, the implementation of the proposed demixer requires the inverse function of the polynomial series. It is possible to express the inverse function of a polynomial in a closed form when the order of the forward function is 4 or less. However, computing the inverse function becomes difficult and intractable as the order increases. The theory of the Series Reversion provides an alternative solution and further establishes the foundation for computing the inverse function of a general polynomial expansion. In this paper, instead of presenting the theorem formally, we provide a paraphrase of the main theorem in [6] with further derivation to our proposed demixing system.

Theorem 1: If the function $g(.)$ has a polynomial expression as $g(u)=\sum_{m=0}^{M_1}\lambda_m u^m$, then its inverse function can be given by the similar form of $g^{-1}(u)=\sum_{n=1}^{+\infty}\gamma_n\left(u-\lambda_0\right)^n$ and the coefficients computed from

$$\gamma_n = \sum_{k_2,k_3,\ldots}\left[(-1)^{\sum_{i=2}^{M_1}k_i}\frac{\left(n-1+\sum_{i=2}^{M_1}k_i\right)!}{n!\prod_{i=2}^{M_1}\left(k_i!\right)}\left(\prod_{i=1}^{M_1}\lambda_i^{k_i}\right)\right] \tag{7}$$

where $k_2+2k_3+3k_4+\cdots=n-1,\ k_i\geq 0,\ i=2,3,4,\ldots$ and $k_1=-\left(n+\sum_{i=2}^{M_1}k_i\right)$. In addition, the differential of γ_n with respect to λ_m's takes the form of

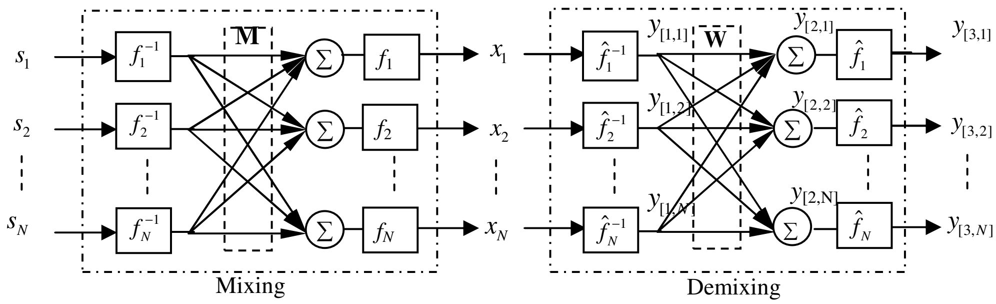

Figure 1: Multi-nonlinearity Constrained Mixing Model and Polynomial-based Nonlinear ICA Demixer

$$d\gamma_n = \sum_{m=1}^{M_1}\left(\sum_{k_2,k_3,\ldots}(-1)^{\sum_{i=2}^{M_1}k_i}\frac{(-k_1-1)!}{n!\prod_{i=2}^{M_1}(k_i!)}\left(\prod_{\substack{i=1\\ i\neq m}}^{M_1}\lambda_i^{k_i}\right)k_m\lambda_m^{k_m-1}\right)d\lambda_m \quad (8)$$

Hence, the derivative of the reverse series with respect to the coefficients in the forward polynomial $\frac{\partial g^{-1}}{\partial \lambda_m}$ can be easily derived from $\frac{\partial g^{-1}}{\partial \lambda_m} = \sum_{n=1}^{M_2}\left(\frac{\partial g^{-1}}{\partial \gamma_n}\frac{\partial \gamma_n}{\partial \lambda_m}\right)$.

3.3. Gradient Based Parameter Learning Algorithm

The primary goal of the demixing system is to obtain a set of signals as independent as possible. The cost function based on the Kullback-Leibler Divergence (KLD) in [1] is commonly used in blind signal separation problems. However, in nonlinear ICA, the preservation of independence is not strong enough to ensure signal separability and this inadvertently results in non-uniqueness of solutions. Therefore, to reduce the indeterminacy of non-unique solutions, the cost function is modified by incorporating a set of signal constraints into the original KLD cost function as follows:

$$J = -\log\left|\det\frac{d\mathbf{y}_{[3]}}{d\mathbf{x}^T}\right| - \sum_{i=1}^{N}\log\left(p_i(y_{[3,i]})\right) + \underbrace{\sum_{i=1}^{N}\beta_i f_i^{(c)}(y_{[3,i]}, s_i)}_{\text{constraints}} \quad (9)$$
$$f_i^{(c)}(y_{[3,i]}, s_i) = \sum_{j=1}^{D}\left[cum\left(y_{[3,i]}, j\right) - cum\left(s_i, j\right)\right]^2$$

where β_i's denotes a set of constants that control the weight of the additional constraints; $cum(u, j)$ represents the j^{th} order cumulant of u and D is the maximum order of the cumulant. In fact, these constraints imply the use of *a priori* information about the source distributions which is intended to match the outputs of the demixer to the original source signals in terms of cumulants. Given the structure of the demixer expressed in (6), the derivative of the cost function with respect to the parameters can therefore be derived as (10)-(12).

$$\frac{\partial J}{\partial \mathbf{W}}\mathbf{W}^T = \mathrm{I} - \left(\sum_{m=1}^{M_1} m(m-1)\,\mathrm{diag}\left(\mathbf{a}_m \circ \mathbf{y}_{[2]}^{m-2}\right)\right)\tilde{\phi}_1\mathbf{y}_{[2]}^T - \left(\sum_{m=1}^{M_1} m\,\mathrm{diag}\left(\mathbf{a}_m \circ \mathbf{y}_{[2]}^{m-1}\right)\right)\boldsymbol{\psi}\mathbf{y}_{[2]}^T \quad (10)$$

$$\frac{\partial J}{\partial \mathbf{a}_0} = \left[\mathrm{I} - \mathrm{diag}^{-1}\left(\tilde{\phi}_1\right)\mathbf{W}^T\mathrm{diag}^{-1}\left(\tilde{\phi}_2\right)\right]\boldsymbol{\psi} + \mathrm{diag}^{-1}\left(\tilde{\phi}_2\right)\mathbf{W}^T\delta\tilde{\phi}_1 - \left(\sum_{n=1}^{M_2} n(n-1)\,\mathrm{diag}\left(\mathbf{b}_n \circ (\mathbf{x}-\mathbf{a}_0)^{n-2}\right)\right)\tilde{\phi}_2 \quad (11)$$

$$\frac{\partial J}{\partial\,\mathrm{diag}(\mathbf{a}_m)} = \tilde{\phi}_1\left(m\mathbf{y}_{[2]}^{m-1}\right)^T + \mathrm{diag}\left[\boldsymbol{\psi}\left(\mathbf{y}_{[2]}^{m}\right)^T\right] + \tilde{\phi}_2\left(\sum_{n=1}^{M_2} n\tilde{\boldsymbol{\xi}}_{[n,m]} \circ (\mathbf{x}-\mathbf{a}_0)^{n-1}\right)^T + \kappa\mathbf{W}^T\delta\,\mathrm{diag}\left(\tilde{\phi}_1\right) - \mathrm{diag}^{-1}\left(\tilde{\phi}_1\right)\mathbf{W}^T\kappa\,\mathrm{diag}(\boldsymbol{\psi}) \quad (12)$$

where $\delta = \sum_{m=1}^{M_1} m(m-1)\,\mathrm{diag}\left(\mathbf{a}_m \circ \mathbf{y}_{[2]}^{m-2}\right)$, $\kappa = \mathrm{diag}\left(\sum_{n=1}^{M_2}\tilde{\xi}_{[n,m,i]} \circ (\mathbf{x}-\mathbf{a}_0)^n\right)$

$$\tilde{\phi}_1 = \left[-1\Big/\left(\sum_{m=1}^{M_1} m a_{[m,1]} y_{[2,1]}^{m-1}\right) \cdots -1\Big/\left(\sum_{m=1}^{M_1} m a_{[m,N]} y_{[2,N]}^{m-1}\right)\right]^T$$

$$\tilde{\phi}_2 = \left[-1\Big/\left(\sum_{n=1}^{M_2} n b_{[n,1]}\left(x_1 - a_{[0,1]}\right)^{n-1}\right) \cdots -1\Big/\left(\sum_{n=1}^{M_2} n b_{[n,N]}\left(x_N - a_{[0,N]}\right)^{n-1}\right)\right]^T$$

$$\boldsymbol{\xi}_{[n,m]} = \left[\tilde{\xi}_{[n,m,1]} \cdots \tilde{\xi}_{[n,m,N]}\right]^T$$

$$\tilde{\xi}_{[n,m,i]} = \sum_{k_2,k_3,\ldots}\left[(-1)^{\sum_{i=2}^{M_1}k_i}\frac{\left(n-1+\sum_{i=2}^{M_1}k_i\right)!}{n!\prod_{i=2}^{M_1}(k_i!)}\left(\prod_{\substack{v=1\\ v\neq j}}^{M_1} a_{[v,i]}^{k_v}\right)k_m a_{[m,i]}^{k_m-1}\right]$$

$$\boldsymbol{\psi} = \tilde{\boldsymbol{\psi}} + \boldsymbol{\beta} \circ \mathbf{f}^{(\mathbf{c})},\quad \boldsymbol{\beta} = \left[\beta_1 \cdots \beta_N\right]^T$$

$$\tilde{\boldsymbol{\psi}} = -\left[\frac{d\left[\log\left(p_1(y_{[3,1]})\right)\right]}{dy_{[3,1]}} \cdots \frac{d\left[\log\left(p_N(y_{[3,N]})\right)\right]}{dy_{[3,N]}}\right]^T$$

$$\mathbf{f}^{(\mathbf{c})} = \left[\frac{d\left[f_1^{(c)}\left(y_{[3,1]}, s_1\right)\right]}{dy_{[3,1]}} \cdots \frac{d\left[f_N^{(c)}\left(y_{[3,N]}, s_N\right)\right]}{dy_{[3,N]}}\right]^T$$

By inserting (10)-(12) into (13)-(14), the gradient descent based learning algorithm can be obtained.

$$\mathbf{W}(t+1) = \mathbf{W}(t) - \mu_{\mathbf{W}}\frac{\partial J}{\partial \mathbf{W}}\mathbf{W}^T\mathbf{W}(t) \quad (13)$$

$$\mathbf{a}_m(t+1) = \mathbf{a}_m(t) - \mu_{\mathbf{a}_m} \frac{dJ}{d\mathbf{a}_m} \quad ; \quad m = 0,1,\ldots,M_1 \tag{14}$$

4. RESULT

Five subgaussian signals are generated synthetically as the original sources and expressed as $\mathbf{s}(t) = [\text{Binary signal};\ \sin(1600\pi t);\ \sin\left(600\pi t + 6\cos(120\pi t)\right);\ \sin(180\pi t);\ \text{Uniform-distributed signal}]^T$. The source signals are then mixed according to (3) where $\mathbf{M}$ is a 5×5 random mixing matrix and $\mathbf{D_f} = \text{diag}\left[\tanh \quad \sinh^{-1} \quad \tanh \quad \sinh^{-1} \quad \tanh\right]$. The learning rates for the weights and the coefficients $\mathbf{a}_m$ are set to $\mu_{\mathbf{W}} = 0.001$ and $\mu_{\mathbf{a}_m} = 0.00003$, respectively. In order to assess the performance of the proposed algorithms, we compare the proposed method with existing algorithms (Linear ICA [1], RBF [4] and FMLP Network [7]) based on the performance index expressed as

$$\rho = \frac{1}{NT}\sum_{t=1}^{T}\sum_{n=1}^{N}\left[\frac{s_i(t)}{\sqrt{E\left(s_i^2\right)}} - \frac{y_{[3,i]}(t)}{\sqrt{E\left(y_{[3,i]}^2\right)}}\right]^2 \tag{13}$$

where T represents the length of the source signals. The source signals, signals recovered by Linear ICA method and the proposed network, the performance index of the tested algorithms are shown in Figure 2. We have also simulated the RBF and FMLP demixers with different number of hidden neurons respectively but no substantial improvement of results has been obtained. A Monte-Carlo experiment of 100 trials has been conducted for the RBF and FMLP demixer and in each simulation, the convergence of the RBF and the MLP demixers have been monitored to ensure that both demixers do not converge to local minima. In Figure 2(d), the proposed approach has demonstrated its efficacy in separating signals under the nonlinear mixture. The success is consecutively followed by the MLP and RBF but the separation results achieved by the linear method falls far from optimal and this indicates the crucial need for nonlinear separation techniques.

5. CONCLUSION

This paper proposes a new algorithm for separating nonlinearly mixed signals based on the multi-nonlinearity constrained mixing model. The hidden neurons activation function uses a set of finite order polynomials as a means to compensate for the nonlinear distortions and to regulate the overfitting of the demixer network. The theory of Series Reversion is integrated into the proposed neural network structure to provide a tractable computation of the inverse series. Simulation results have successfully shown that the proposed method has significantly outperformed other linear and nonlinear algorithms in terms of accuracy and convergence speed.

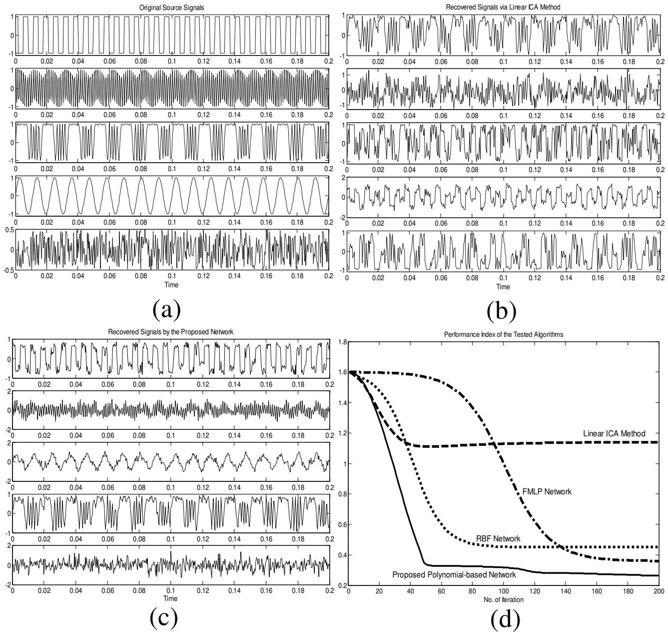

Figure 2: (a) Original sources.
(b) Recovered signals via Linear ICA method.
(c) Recovered signals via the proposed network.
(d) Performance index of the tested algorithms.

6. REFERENCES

[1] S.I. Amari and A. Cichocki, "Adaptive blind signal processing - Neural network approaches", *Proceedings of the IEEE*, **86**(10), pp. 2026-2048, Oct. 1998.

[2] W.L. Woo and S.S. Dlay, "Neural Network Approach to Blind Separation of Mono-nonlinearly Mixed Sources", *IEEE Trans. on Circuits and System-1*, **52**(6), pp. 1236-1247, 2005.

[3] A. Hyvarinen and P. Pajunen, "Nonlinear Independent Component Analysis: Existence and Uniqueness Results", *Neural Network*, 12(3), pp. 429-439, 1999.

[4] Y. Tan, J. Wang, and J. Zurada, "Nonlinear Blind Source Separation using a Radial Basis Function Network", *IEEE Trans. on Neural Network*, **12** (1), pp. 124-134, 2001.

[5] A. Taleb and C. Jutten, "Source Separation in Post-Nonlinear Mixtures", *IEEE Transactions on Signal Processing*, 47 (10), pp. 2807-2820, 1999.

[6] P.Gao,W.L.Woo and S.S. Dlay, "Weierstrass approach to nonlinear blind signal separation" in *Proc. of 4th Int. Symp. on Communication Systems, Networks and Digital Signal Processing*, Newcastle,UK, July 2004, pp. 588-591.

[7] W.L. Woo and S. Sali, "General Multilayer Perceptron Demixer Scheme for Nonlinear Blind Signal Separation", *IEE Proc. on Vision, Image and Signal Processing*, 149(5), pp. 253-262, 2002.