Nonlinear Single Channel Source Separation

A.M Darsono, Bin Gao, W.L. Woo and S.S. Dlay
School of Electrical, Electronic and Computer Engineering
Newcastle University, England, United Kingdom
Email: {a.m.darsono, bin.gao, w.l.woo, s.s.dlay}@ncl.ac.uk

Abstract—A new model of nonlinear single channel source separation is proposed in this paper. The proposed model is a linear mixture of the independent sources followed by an element-wise post-nonlinear distortion function. In addition, the paper develops a novel solution that efficiently compensates for the nonlinear distortion and performs source separation. The proposed solution is a two-stage process that consists of a Gaussianization transform and a maximum likelihood estimator for the sources. The paper also discusses the theory behind the proposed solution. Simulations have been carried out to verify the theory and evaluate the performance of the proposed algorithm. Results obtained have shown the effectiveness of the algorithm even in presence of the strong nonlinearity.

Keywords—Blind Source Separation, Independent Component Analysis, Gaussianization Transform and Maximum likelihood.

I. INTRODUCTION

Blind sources separation (BSS) has become one of the promising and exciting topics because of its solid theoretical foundations and potential applications in signal processing and advanced statistics [1-5]. BSS aims to recover unknown statistically independent sources from a set of observations. The extension from linear BSS to nonlinear BSS also been proposed [1][6][7] and the methods produces a good performance only if the numbers of observations equal to the numbers of sources. Single channel source separation (SCSS) is a branch of BSS family where the blind signal separation is achieved when only one single recording is available. Several approaches have been developed to solve the SCSS problem such as the computer auditory scene analysis (CASA) [2][3], underdetermined BSS [5][8] and nonnegative matrix factorization (NMF) [4][9]. CASA is an unsupervised method that separates the signal without using any prior knowledge about the corresponding source signals. For example in [2][3], after the observation signal is segmented into time-frequency cells using an appropriate transform such as the short-time Fourier transform (STFT) , the signal cues such as continuity, harmonicity, common onset and offset time are used to handle the separation problem. Another class of unsupervised method is the NMF which uses statistical characteristics such as nonnegativity and sparseness to separate the signal mixture [4][9]. Whereas CASA and NMF are unsupervised methods, the underdetermined BSS method [5][8] is the supervised technique where the separation processes rely on a priori knowledge of sources obtained during the training phase. The source is then projected onto a set of basis function for separation.

The above SCSS methods and other existing SCSS algorithms by far were based on linear model. Linear model is known for its simplicity and ease of implementation. These methods [2-5][8] show very good performance in separating the signal for linear mixture. However, in practical applications such as in speech recognition, music transcription or telecommunications, the transmitted signals are often received by nonlinear receiver such as carbon-button microphones [9] or antennas. In the real environments, the mixed signals are more likely to be nonlinear or subject to some kind of nonlinear distortions due to sensor sensitivity. Therefore, the assumption that the mixture is linear adopted by the existing SCSS approaches is violated and may not characterise the actual observed signals accurately. The need of an accurate representation of the distorted signals has resulted in the emergence of SCSS for nonlinear mixture model. So far, no method in SCSS has been proposed to solve the nonlinear problem.

The aim of this paper is to study and develop efficient solution to SCSS when the mixture has been nonlinearly distorted. The proposed solution is achieved in a two-stage approach, firstly, by compensating the nonlinear distortion and this followed by the unmixing process. The organization of the paper is as follow. Section II describes the post-nonlinear single channel model. Section III explains the proposed solution for two stage process. Section IV discusses the Gaussianization transform performances in compensates the nonlinearity. Experimental results are analyzed in Section V. Finally, Section VI concludes the paper.

II. NONLINEAR SINGLE CHANNEL MIXTURE

In this paper, we propose the nonlinear single channel mixture is modelled by the post-nonlinear (PNL) mixing model as described in the following:

\[ v(t) = f(A_1x_1(t) + A_2x_2(t) + \ldots + A_px_p(t)) \]  

where \( x_i(t) \) is the \( i \)-th sampled value of \( i \)-th source signal, \( A_i \) is the mixing gain of sources and \( f \) is an invertible nonlinear function. The goal is to recover \( x_i(t) \) given only single input \( y(t) \). For each source signal, a \( N \)-sample vector \( \mathbf{x}_i(t)=[x_i(t) \; x_i(t+1) \; \ldots \; x_i(t+N-1)]^T \) can be expressed as a linear combination of basis functions such that

\[ \mathbf{x}_i(t) = \sum_{k=1}^{N} a_{ik} s_{ik}(t) = \mathbf{A}_i s_i(t) \]  

where \( N \) is the number of basis functions, \( a_{ik} \) is the \( k \)-th basis function of \( i \)-th source, and \( s_{ik}(t) \) is the coefficient. The transform between \( \mathbf{x}_i(t) \) with coefficient vector, \( s_i(t) \) is assumed to be reversible with

\[ s_i(t) = \mathbf{W}_i \mathbf{x}_i(t) \]  

(3)
where inverse of the basis matrix, \( W_t = A_i^{-1} \).

The PNL model represents the important subclass of the general nonlinear model which is simpler and widely applicable. The PNL model that will be used is shown in fig. 1 where two independent sources are mixed together.

III. PROPOSED SOLUTION

In this work, the sources are estimated in a two-stage approach as shown in fig. 2. In the first stage, the nonlinearity of the observation \( v(t) \) is equalized using the nonlinear transform, \( g(v) = \hat{f}^{-1}(v) \) with the linearized signal is given by \( z(t) = g(v(t)) \). This is followed by the second stage where the linear separation algorithm based on maximum likelihood (ML) [5] will be used to obtain an estimated sources of \( \hat{x}_1(t) \) and \( \hat{x}_2(t) \).

A. Nonlinearity Compensation

In order for the signals in nonlinearly distorted mixture to be accurately separated, the nonlinearity of the mixture must be compensated. To this end, we propose a linearization technique known as the Gaussianization transform. The main impetus of using the proposed technique comes from the following principle: Firstly, we observe that the original sources in (1) are statistically independent and non-Gaussian. According to the central limit theorem, when the sources are mixed the resulting mixture tends toward a Gaussian distribution. However, the post-nonlinearity \( f(.) \) distorts the amplitude distribution of the linear mixture and subsequently transform it to a non-Gaussian distribution. Thus, the non-Gaussian behaviour of the observation \( v(t) \) to some extents is directly attributed to the nonlinearity \( f(.) \) in the mixture. As such, the nonlinearity can be compensated by finding a suitable transformation such that the output returns to a Gaussian distribution. Therefore, it is clear that by using this principle, the estimation problem can be readily split into two tasks where the first task is to compensate for the nonlinear distortion and the second task is to seek separation from the compensated signals where separation can be assumed to be linear [11],[10],[11]. In this paper, we constrained the number of sources to be two for illustration purpose only. In reality, the number of sources will be considerably more than two and in such case, our proposed method will work even more efficiently.

Although the nonlinearity \( f(.) \) is unknown, it is possible to determine the inverse function \( g(.) \) by finding a suitable transformation which convert the component \( v(t) \) to the Gaussian random variable. The goal is to find \( g(.) \) such that

\[
g_i(x_i) \sim N(0, \sigma_i^2) \quad \text{for} \quad i = 1,\ldots,N \tag{4}
\]

where \( \sigma_i^2 = 1 \) due to the usual scaling indeterminacies.

Consider \( F_i(v) \) denote the cumulative density function (CDF) as

\[
F_i(v) = \int_{-\infty}^{v} p_i(v) dt \tag{5}
\]

with \( p_i(v) \) denote probability density function of \( v_i \),

\[
p_i(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}.
\]

Assuming cdf of \( z_i \), \( F_i(z) \) is continuous and strictly increasing so that the inverse mapping exist, then \( F_i(z) \) can be express as,

\[
F_i(z) = F_i(v) \quad F_i(g(v)) = F_i(v) \quad g(v) = F_i^{-1}(v) \tag{6}
\]

Since the desired distribution of \( z(t) \) is Gaussian, and the cdf of Gaussian is expressed as \( \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{v^2}{2}} dt \),

the expression of \( F_i^{-1} \) can be replaced by the inverse of Gaussian cdf, \( \Phi(z)^{-1} \). Hence, the nonlinear mapping \( g(.) \) can be expressed as

\[
g = \Phi^{-1} \circ F_i \tag{7}
\]

If the mixed signal, \( y(t) \) are closely to Gaussian distribution, then the Gaussianized signal should estimate the signal perfectly with \( z(t) \approx y(t) \).

B. Source Estimation: Maximum Likelihood

After recovery of the nonlinearity, the linear SCSS approach based on maximum likelihood (ML) [5] [8] was used to solve the single channel source separation problem. Considered two sources signal \( x_1(t) \) and \( x_2(t) \), according to ICA, the sources are said to be statistically independent if only the probability density function (pdf),

![Figure 1: PNL mixing model for SCSS](image)

![Figure 2: Proposed two-stage nonlinear SCSS](image)
\[ p_{x_i(t),x_j(t)}(y(t)|\hat{x}_i(t),\hat{x}_j(t)) = p_{x_i(t)}(x_i(t))p_{x_j(t)}(x_j(t)) \]

where the number of samples, \( t = 1, \ldots, T \). The sources vectors are passed through the fixed basis filter \( W \), to generate set of basis coefficients,

\[ p(x_1(t), x_2(t), \ldots, x_T(t)|W) = \prod_{t=1}^{T} p(x_t(t)|W) \]

where \( x_t \) is a learning gain.

Equation (11) then become,

\[ \frac{\partial L}{\partial x_1(t)} = \sum_{i=1}^{T} \sum_{k=1}^{N} \left[ \frac{\partial \log p(s_{1k}(t))}{\partial x_1(t)} + \frac{\partial \log p(s_{2k}(t))}{\partial x_1(t)} \right] \]

(11)

Equation (11) then become,

\[ \frac{\partial L}{\partial x_1(t)} = \sum_{i=1}^{T} \sum_{k=1}^{N} \left[ \frac{\partial \log p(s_{1k}(t))}{\partial x_1(t)} \right] \]

(12)

where \( t_0 = t-n+1 \), \( \phi(s) = \frac{\partial \log p(s_{1k}(t))}{\partial s_{1k}(t)} \)

and \( w_{1kn} = W(k, n) \). Similar formulation are applied to the second source,

\[ \frac{\partial L}{\partial x_2(t)} = \sum_{i=1}^{T} \sum_{k=1}^{N} \left[ -\frac{\lambda_2}{\lambda_1} \phi(s_{2k}(t)) w_{1kn} + \frac{\lambda_1}{\lambda_2} \phi(s_{1k}(t)) w_{2kn} \right] \]

(13)

The update process of the sources can be written as

\[ x_i^*(t) = x_i(t) + \eta \frac{\partial L}{\partial x_i(t)} \]

(14)

where \( \eta \) is a learning gain.

Next step is to estimate the scaling factors, \( \lambda_i \) by finding the maximum the posteriori values. From (11), \( L \) was differentiating with respect to \( \lambda_i \) and substitute \( \lambda_2 = 1 - \lambda_1 \).\n
\[ \frac{\partial L}{\partial \lambda_i} = \sum_{i=1}^{T} \sum_{k=1}^{N} \left[ \frac{\partial \log p(s_{1k}(t))}{\partial \lambda_i} + \frac{\partial \log p(s_{2k}(t))}{\partial \lambda_i} \right] \]

\[ = \sum_{i=1}^{T} \sum_{k=1}^{N} \left[ -\phi(s_{1k}(t)) \frac{s_{1k}(t)}{\lambda_1} + \phi(s_{2k}(t)) \frac{s_{2k}(t)}{\lambda_2} \right] \]

(15)

where the partial derivative of \( s_{ik}(t) \) with respect to \( \lambda_i \) given by

\[ \frac{\partial s_{ik}(t)}{\partial \lambda_i} = \lambda_i s_{ik}(t) \frac{1}{\lambda_i} - \frac{s_{ik}(t)}{\lambda_i} \]

The update scaling factor of \( \lambda_1 \) and \( \lambda_2 \) can be written as,

\[ \lambda_1^* = h_2(\lambda_1 + \eta \lambda_1 \frac{\partial L}{\partial \lambda_1}) \]

(16)

and

\[ \lambda_2^* = 1 - \lambda_1^* \]

(17)

where \( \eta \) is a learning gain and \( h_2 \) is a limiting function.

Table 1 shows the summary of the proposed algorithm.

**IV. GAUSSIANIZATION TRANSFORM**

The motivation behind the Gaussianization transform is that the linearly mixed signals before nonlinear transformation are approximately Gaussian distributed. Here, the performance of Gaussianization to compensate the nonlinearity is evaluated. In fig. 3, the histogram of the signal is plotted along with a Gaussian model where its parameters are calculated from the data. The larger the deviation between the histogram and the Gaussian model signifies larger deviation from Gaussianity. Firstly, two non-gaussian distribution sources (e.g. piano sound and flute sound) as shown in fig. 3(a) and 3(b) respectively are linearly mixed. Theoretically it tends to be more Gaussian distributed as shown in fig. 3(c). The post-nonlinear (PNL) distortion is applied to the mixed signal using the following nonlinearity \( f(y) = 0.3y + \text{tanh}(3y) \).

This distortion causes the Gaussian distribution of the mixed signal to deviate from Gaussianity as in fig. 3(d). Without using any knowledge of the nonlinearity, the Gaussianization transform inverts the nonlinearly distorted signal and restore the distribution to the Gaussian pdf. The histogram of Gaussianized signal in

**TABLE I. ALGORITHM OF TWO-STAGE NONLINEAR SCSS**

<table>
<thead>
<tr>
<th>Stage</th>
<th>Observation: ( y(t) = \lambda_1 x_1(t) + \lambda_2 x_2(t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>First stage</td>
<td>The initial value of ( \lambda_1 ) and ( \lambda_2 ) are set at 0.5. For ( t = 1: T ), 1. Introduce the post-linearity distortion, ( f(.) ), ( z(t) = f(x(t)) ) 2. Find cdf, ( F_y ) and inverse cdf of Gaussian distribution, ( \Phi^{-1} ) 3. Estimate the nonlinearity mapping, ( g = \Phi^{-1} \circ F_y ) to find ( z(t) )</td>
</tr>
<tr>
<td>Second stage</td>
<td>Input: ( z(t) ) For ( t = 1: T ), ( k = 1: N ) and ( t = 1, 2 ) 1. Compute ( s_{ik}(t) = W(k, n) x_i(t) ) and ( \phi(s) = \frac{\partial \log p(s_{1k}(t))}{\partial s_{1k}(t)} ) 2. Update the sources signal, ( x_i^<em>(t) = x_i(t) + \eta \frac{\partial L}{\partial x_i(t)} ) and ( \lambda_1^</em> = h_2(\lambda_1 + \eta \lambda_1 \frac{\partial L}{\partial \lambda_1}) ) and ( \lambda_2^* = 1 - \lambda_1^* ) 3. Update the scaling factors, ( \lambda_1^* = h_2(\lambda_1 + \eta \lambda_1 \frac{\partial L}{\partial \lambda_1}) ) and ( \lambda_2^* = 1 - \lambda_1^* ) 4. Repeat steps 1 to 3 until convergence</td>
</tr>
</tbody>
</table>
fig. 3(e) proves the performance of the transformation. The line shows in the figure 3 was a histogram of a Gaussian model fit using the mean and variance of a signal.

In fig. 4, relationship between the signals of $y(t)$, $v(t)$ and $z(t)$ are shown using a scatter plot. A linear relationship between the gaussianized signal and mixed signal, $y(t)$ can be seen clearly in fig. 4(c). After successfully compensation, the gaussianized signal, $z(t)$ which is the estimation of $y(t)$ will be used for the separation process to estimate the source signals of $x_1(t)$ and $x_2(t)$.

V. EXPERIMENTS & ANALYSIS

An audio signal was used containing a piano and flute sound with sampling frequency of 22.05 kHz, mono sound with 16 bit. The objective is to separate the piano sound and the trumpet sound in the nonlinear mixture. The value of signal gain $\lambda_1$ and $\lambda_2$ both is fixed at 0.5. The number of iterations needed for the algorithms to converge is 200. Independent component analysis (ICA) algorithm (e.g. fastICA)[12] was used to obtain the basis

$W = A^{-1}$ and source coefficient density is modelled using generalized Gaussian parameter. The basis functions obtained in this simulation is based on [8] which used best characteristic features that being extracted from cross-correlation matrix of piano and flute sound

In fig. 5 the proposed algorithm shows the capability to separate the single mixture and recover the piano and flute sound very well in nonlinear mixture. Compare with fig. 6, a linear algorithm without the Gaussianization, the results are affected by the nonlinearity. To evaluate this, the performances of the algorithm have been measured using signal to distortion ratio (SDR) which measures an overall sound quality of the source separation. The Matlab toolbox and measurement details can be found in [13]. SDR value which is higher than 7dB can be considered as good because it shows that there is less distortion in the recovered signal and represents an acceptable perceptual measure.

Based on analysis in SDR as shown in table 2, overall, the proposed algorithm shows a very good separation results in a nonlinear environment with the good SDR values. As for nonlinear mixture which is our concern, the Gaussianization transform used to compensate the nonlinearity was proved to be effective where we can see that the result are better if the PNL algorithm is applied in nonlinear mixture.

In separation the single mixture of signal, the SCSS using ML algorithm achieved a good performance because the proposed algorithm using basis adapted by ICA learning rules [5] i.e. fastICA. This prior information from the basis and their corresponding pdfs are the key to obtaining a faithful MAP based inference algorithm. As in this experiment, the basis obtained using the best characteristic features which is already proved to be better in separation if the less number of bases used [8]. One of the most useful properties is that resulting in decompositions which are often intuitive and easy to interpret because they are sparse. For single channel using ML approach, the coefficients of the basis functions have the higher degree of sparseness.
Figure 6: Separation results of nonlinear mixture using linear algorithm

| TABLE II. NONLINEAR SCSS PERFORMANCE COMPARISON BY SDR (dB) |
|--------------------------|-------------|
|                         | Piano       | Flute       |
| Nonlinear mixture with linear algorithm | 20.1252     | 1.9402      |
| Nonlinear mixture with PNL algorithm      | 20.2825     | 6.1013      |

A. Experiments using Nonlinear Handset Model

In this section, an experiment result by using the different nonlinearity which is polynomial is presented. This nonlinearity corresponds to carbon-button handset mapper [14]. The nonlinearity is defined as a piecewise function with variable boundary points, given by

\[ f(y) = \begin{cases} 
  f_+ & \text{for } y > y_+ \\
  y(i) + y^3(i) & \text{for } y_- \leq y \leq y_+ \\
  f_- & \text{for } y > y_- 
\end{cases} \]  

where the output saturation levels, \( f_+ \) and \( f_- \) correspond to the input levels \( y_+ \) and \( y_- \) respectively. From Table 3, the results obtained show the same pattern as in Table 2 with the performance of PNL algorithm is always better than the linear algorithm. It proves that for nonlinearily distorted signal, by compensating the nonlinearity using Gaussianization, the separation in single channel can yield a good performance.

VI. CONCLUSION

A new model for nonlinear single channel mixture has been proposed along with a novel solution that combines the gaussianization transform and the time-domain maximum likelihood separation algorithm. Audio signal that contains piano and trumpet sound was successfully separated using the proposed algorithm. From the experiments, the proposed method shows significant performance with high SNR value and low MSE value in nonlinear mixture. Beside that, Gaussianization transform have performed very well in recovering the loss of signal information due to the nonlinearity.

| TABLE III. PERFORMANCE COMPARISON BY SDR FOR POLYNOMIAL NONLINEARITY (dB) |
|--------------------------|-------------|
|                         | Piano       | Flute       |
| Nonlinear mixture with linear algorithm | 17.8001     | 3.4641      |
| Nonlinear mixture with PNL algorithm      | 20.2009     | 3.7605      |

REFERENCES


