Blind Separation of Nonlinear Mixing Signals Using Kernel with Slow Feature Analysis

Usama S. Mohmmad¹, Hany Saber²
¹Department of Electrical Engineering, Assiut University, Assiut 71516, Egypt
²Department of Electrical Engineering, South Valley University, Aswan, Egypt
Corresponding author Usama@aun.edu.ee

Abstract— This paper describes a hybrid blind source separation approach (HBSSA) for nonlinear mixing model (NL-BSS). The proposed hybrid scheme combines simply the kernel-feature spaces separation technique (KTDSEP) and the principle of the slow feature analysis (SFA). The nonlinear mixed data is mapped to high dimensional feature space using kernel-based method. Then, the linear blind source separation (BSS) based on the slow feature analysis (SFA) is used to extract the most slowness vectors among the independent data vectors. The proposed scheme is based on the following four key features: 1) estimating an orthonormal bases, 2) mapping the data into the subspace using this orthonormal bases, 3) applying linear BSS on the mapping data to make the data vectors in the feature spaces are independent, 4) applying the principle of slow feature analysis on the mapping data to select the desired signals. The SFA provides the dimension reduction according to the most independent and slowing variable signals. Moreover, the orthonormal bases estimation in the wavelet domain is introduced in this work to reduce the complexity of the KTDSEP algorithm. The motivation of using the wavelet transform, in estimating the orthonormal bases, is based on the fact that the low frequency band in the wavelet domain contains the significant power of the signal. The advantages of the proposed method are the fast estimation of the orthonormal bases and the dimension reduction of the estimating data vectors. Performed computer simulations have shown the effectiveness of the idea, even in presence of strong nonlinearities and synthetic mixture of real world data. Our extensive experiments have confirmed that the proposed procedure provides promising results.

Index Term-- Nonlinear blind source separation, slow feature analysis, independent slow feature analysis, slow feature analysis, kernel base algorithm, and kernel trick

I. INTRODUCTION
The problem of blind source separation (BSS) consists on the recovery of independent sources from their mixture. This is important in several applications like speech enhancement, telecommunication, biomedical signal processing, etc. Most of the work on BSS mainly addresses the cases of instantaneous mixing (BSS) can be solved by using independent component analysis (ICA) [4]. The goal of the ICA is to separate the signals by finding independent component from the data signal. It is important to note that if x and y are two independent random variables, any of their functions f(x) and f(y) are also independent. An even more serious problem is that in the nonlinear case, x and y can be mixed and still be statistically independent. Several authors [6-10] have addressed the important issues on the existence and uniqueness of solutions for the nonlinear ICA and BSS problems. In general, the ICA is not a strong enough constraint for ensuring separation in the nonlinear mixing case. There are several known methods that try to solve this nonlinear BSS problem. They can roughly be divided into algorithms with a parametric approach and algorithms with nonlinear expansion approach. With the parametric model the nonlinearity of the mixture is estimated by parameterized nonlinearities. In [8] and [11], neural network is used to solve this problem. In the nonlinear expansion approach the observed mixture is mapped into a high dimensional feature space and afterwards a linear method is applied to the expanded data. A common technique to turn a nonlinear problem into a linear one is introduced in [12].

II. Nonlinear mixture model
A generic nonlinear mixture model for blind source separation can be described as follows:

\[ X(t) = f(S(t)) \]  

Where \( f \) is an unknown vector of real functions. The linear instantaneous mixing (BSS) can be solved by using independent component analysis (ICA) [4]. The goal of the ICA is to separate the signals by finding independent component from the data signal. It is important to note that if x and y are two independent random variables, any of their functions f(x) and f(y) are also independent. An even more serious problem is that in the nonlinear case, x and y can be mixed and still be statistically independent. Several authors [6-10] have addressed the important issues on the existence and uniqueness of solutions for the nonlinear ICA and BSS problems. In general, the ICA is not a strong enough constraint for ensuring separation in the nonlinear mixing case. There are several known methods that try to solve this nonlinear BSS problem. They can roughly be divided into algorithms with a parametric approach and algorithms with nonlinear expansion approach. With the parametric model the nonlinearity of the mixture is estimated by parameterized nonlinearities. In [8] and [11], neural network is used to solve this problem. In the nonlinear expansion approach the observed mixture is mapped into a high dimensional feature space and afterwards a linear method is applied to the expanded data. A common technique to turn a nonlinear problem into a linear one is introduced in [12].

\[ X(t) = A \cdot f(S(t)) \]  

Where \( S(t) \) represents the statistically independent sources array while \( X(t) \) is the array containing the observed signals and \( f \) is unknown multiple-input and multiple-output (MIMO) mapping which called the nonlinear mixing transform (NMT). In order for the mapping to be invertible, we assume that the nonlinear mapping is monotone. We make the assumption here, for simplicity and convenience, that the dimensions of \( X \) and \( S \) is equal.

An important special case of the nonlinear mixture is the so-called post-nonlinear (PNL) mixture

\[ X(t) = f(AS(t)) \]  

Where \( f \) is an invertible nonlinear function that operates componentwise and \( A \) is a linear mixing matrix, more detailed

\[ X_i(t) = f\left( \sum_{j=1}^{m} a_{ij} S_j(t) \right), i=1, \ldots , n \]  

Where \( a_{ij} \) are the elements of the mixing matrix A. The PNL model was introduced by Taleb and Jutten[13]. It represents an important subclass of the general nonlinear model and has therefore attracted the interest of several researchers [14]-[18].
Applications are found, for example, in the fields of telecommunications, where power efficient wireless communication devices with nonlinear class C amplifiers are used [19] or in the field of biomedical data recording, where sensors can have nonlinear characteristics [20].

In this work, a new method to solve the PNL-BSS problem or the general nonlinear BSS is proposed. In the first step, the nonlinear mixed data is mapped to higher dimensional feature space $f_k$ and linear blind source separation algorithm is applied on the mapped data. In the second step, the principle of slow feature analysis is used on the linearly separated data to find the most slow and independent vectors in this feature space. In the following Sections, the kernel-feature spaces separation technique (KTDSEP) and the principle of slow feature analysis (SFA) will be discussed.

II. KERNEL FEATURE SPACES AND NONLINEAR BLIND SOURCE SEPARATION

In kernel-based learning, the data is mapped to a kernel feature space of a dimension that corresponds to the number of training data points. Suppose that the data is $X_i$ (i=1,………..,N), so the idea is to map the data $X_i$ into some kernel feature space $f_k$ by some mapping $\Phi : R^d \rightarrow f_k$. Performing a simple linear algorithm in $f_k$ corresponds to a nonlinear algorithm in the input space will solve the separation problem. Essential ingredients to kernel based learning are: (1) support vector machine (SVM) [21] [22] theory that can provide a relation between the complexity of the function class in use and the generalization error, and (2) the famous kernel trick

$$K(x,y)=\Phi(x) \cdot \Phi(y)$$

(6)

Which efficiently computes the scalar product. This trick is essential if $f_k$ is an infinite dimensional space. Even though $f_k$ might be infinite dimensional the subspace, where the data lies is maximally N-dimensional. However, the data typically forms an even smaller subspace in $f_k$ [23]. To map the nonlinear mixing data $X_1, X_N \in \mathbb{R}^d$ into the feature space $f_k$ the kernel trick in Equation (6) must be used, but the high dimensionality mapping data will be. In [21] the idea of estimating the orthonormal bases for subspace $f_k$ is proposed and these bases is used to map the nonlinear mixed data to the feature space $f_k$ for some further points $v_1, v_d \in \mathbb{R}^n$ from the same space, that will later generate a bases in $f_k$. The mapping method will be as follows:

Consider the mapping points as $\Phi_x = [\Phi(x_1) \ldots \Phi(x_N)]$, and $\Phi_r = [\Phi(v_1) \ldots \Phi(v_d)]$.

Assume that the columns of $\Phi_x$, constitute a base of the column space of $\Phi_x$, so

$$\text{span}(\Phi_x) = \text{span}(\Phi_r) \quad \text{and} \quad \text{rank}(\Phi_x^T \Phi_r) = d$$

(7)

Moreover, $\Phi v$ being bases implies that the matrix $\Phi_v \Sigma \Phi_v^T$ has full rank and its inverse exists. Now an orthonormal bases can be defined as follows:

$$\Psi = \Phi_v (\Phi_v^T \Phi_v)^{-1/2}$$

(8)

The column space of which is identical to the column space of $\Phi_v$. Consequently, the bases $\Psi$ enable us to parameterize all vectors that lie in the column space of $\Phi x$ by some vectors in $\mathbb{R}^d$. The space that is spanned by $\Psi$ is called parameter space. The orthonormal bases equation (7) enables the working in $\mathbb{R}^d$ the span of $\Psi$, which is extremely valuable since $d$ depends solely on the kernel function and the dimensionality of the input space. Therefore, $d$ is independent of $N$.

To map the mixed data from input space to feature space $f_k$ the kernel trick is used to perform the following:

$$(\Phi_x^T \Phi_x)_{ij} = \Phi(v_i)^T \Phi(v_j) = k(v_i, v_j)$$

with $i, j = 1, \ldots, d$

$$(\Phi_x^T \Phi_x)_{ij} = \Phi(v_i)^T \Phi(x_j) = k(v_i, x_j)$$

with $i = 1, \ldots, d, j = 1, \ldots, N$

(9)

Finally the mapping matrix will be

$$\Omega_x = \Psi^T \Phi_x = (\Phi_v^T \Phi_v)^{-1} \Phi_v^T \Phi_x$$

(10)

Which is also a real valued $d \times N$ matrix and the matrix $(\Phi_v^T \Phi_v)^{-1/2}$ is symmetric. Two main problems facing the kernel based algorithm (KTDSEP):

- Selecting the points $v_1, \ldots, v_d$ to constructed the orthonormal bases $\Psi$
- The number of output signals from KTDSEP algorithm will be $d$ signals and we need only $n$ signals where $n$ is smaller than $d$

In the traditional KTDSEP algorithms, the random sampling method and the k-mean clustering method are used to find the orthonormal bases. These two techniques increase the complexity of the algorithm. Moreover, the KTDSEP method cannot automatically find $n$ signals out of $d$ signals. The proposed solution by the KTDSEP method [24] is based on the correlation among the original signals and the output signals. In the blind source separation field the original signals is not available, so this technique is not practical. Another proposed solution [21] is done by applying the KTDSEP on the signals two times. This solution will increase the computational complexity of the algorithm.

In this work, after mapping the data, the linear blind source separation algorithm is used. The temporal decorrelation BSS (TDSEP) [24] is used as linear blind source separation algorithm. In the next Sections, the details of our proposed solution will be discussed.

III. SLOW FEATURE ANALYSIS AND NONLINEAR BLIND SOURCE SEPARATION

Slow Feature Analysis (SFA) is a method that extracts slowly varying signals from a given observed signals [25, 26]. Consider an input signal $x(t) = [x_1(t), \ldots, x_N(t)]^T$, the objective of SFA is to find a nonlinear input-output function $g(x) = [g_1(x), \ldots, g_N(x)]^T$ such that the components of $u(t) = g(x(t))$ are varying as slowly as possible. The variance of the first derivative is used as a measure of the slowness. The optimization problem will be as follows:

Minimize the objective function

$$\text{Minimize} \quad \text{var}(u'(t))$$
\[ \Delta(U_i(t)) = \begin{bmatrix} U_i^2(t) \end{bmatrix} \]  

(11)

Successively for each \( U_i(t) \) under the following constraints:

\[ \langle u_i(t) \rangle = 0 \]  

(a)

\[ \langle (u_i(t))^2 \rangle = 1 \]  

(b)

\[ \langle u_i(t) u_j(t) \rangle = 0 \quad \forall j < i \]  

(c)

Where \( \langle \rangle \) denotes averaging over time. Constraints (a), (b) and (c) ensure that the solution will not be the trivial solution \( (U_i(t)= \text{const}) \). Constraint (c) provides uncorrelated output signal components and thus guarantees that different components carry different information.

Note that slowly varying signal components are easier to predict and should therefore have strong correlations in time.

In the proposed work, the principle of slow feature analysis is used to overcome the main drawback of the kernel base algorithm (KTDSEP) to pickup \( n \) signals out of \( d \) signals. The slow feature analysis algorithm is motivated by the fact that the linear signals are the slowest varying signals among all the signals [12]. Therefore, the slow feature analysis can be used to pickup \( n \) slow varying signals (linear) from \( d \) signals.

The slow feature analysis algorithm can be summarized in the following steps:

i. Nonlinear expanding the input data.

ii. The nonlinear expanding signal is whitened using eigenvalue decomposition of the zero time-log correlation matrices.

iii. Derivative the whitened signal is calculated according to

\[ y(t) \approx y(t+1) - y(t) \]

iv. The rotating matrix \( Q \) is calculated using the joint approximate diagonalization (JAD).

v. The eigenvalue and the eigenvectors of \( Q \) are evaluated.

vi. The dimensional reduction of the signals by is achieved using the normalized eigenvectors that corresponding to the smallest eigenvalue of the rotation matrix \( Q \).

IV. THE PROPOSED HYBRID BLIND SOURCE SEPARATION APPROACH (HBSSA)

One of the most popular methods to solve the problem of the nonlinear mixing model (NL-BSS) is the mapping of the nonlinear mixing data into high dimensional feature space \( f_i \). Then the problem can be handled as a linear problem under some constraints [12], [21]. As mentioned in Section II, the kernel base algorithm (KTDSEP) is depended on mapping the nonlinear mixed data into high dimensional space.

The proposed separation method combines simply the kernel-feature Spaces separation technique (KTDSSEP) and the principle of the slow feature analysis (SFA). The estimation of the orthonormal bases is done in the wavelet domain to reduce the complexity of the KTDSEP algorithm. The motivation of using the wavelet transform, in estimating the orthonormal bases, is based on the fact that the low frequency band in the wavelet domain contains the significant power of the signal. Moreover, the SFA is used to select the desired signals from the output signals of the KTDSEP algorithm. The SFA algorithm is motivated by the fact that the linear signals are the slowest varying signals among all the signals. The orthonormal bases is started by selecting the data vector \( v_1, \ldots, v_d \) and testing these vector using Equation (6). This vector is selected using the k-mean clustering in KTDSEP. The wavelet transform is applied on the nonlinear mixed data \( n \)-times to select the \( v_1, \ldots, v_d \) from lowest frequency band. Then, equation (7) is used to find the orthonormal bases \( \Psi \). The proposed method can be summarized as the follows:

1. The wavelet transformed (using wavelet packet) is applied on the nonlinear mixed data \( n \)-times.
2. The points \( v_1, \ldots, v_d \) are selected from the lowest frequency band in the wavelet domain.
3. Equation (7) is used to construct the orthonormal bases \( \Psi \).
4. The kernel trick is used to map the mixed data into the nonlinear feature space \( f_i \) and find \( d \times N \) matrix

\[ \Psi = \Xi \Phi = (\Phi_1^T \Phi_2)^{-1} \Phi_1^T \Phi_2 \]

5. The output data \( \psi_k \) are pre-whitened.
6. Evaluate \( n \) time shifted covariance matrices among the vectors of pre-whitened \( \psi_k \).
7. The joint approximate diagonalization criteria is used with the n time shifted covariance matrices to estimate the \( d \times d \) separating matrix \( \mathbf{B} \).
8. The Slow Feature Analysis is applied on the \( d \times d \) separation matrix \( \mathbf{B} \) to get \( n \times d \) rotation matrix (\( Q_{ASE} \)) as following:

- Derivative the whitened signals \( \psi_k \), step (5).
- The rotating matrix \( Q \) is calculated using the joint approximate diagonalization (JAD).
- The eigenvalue and the eigenvectors of \( Q \) are evaluated.
- Applying the dimensional reduction to find \( n \) out of \( d \) signals by using the \( n \) normalized eigenvectors that corresponding to the smallest \( n \) eigenvalue of the rotation matrix \( Q \) and find \( Q_{ASE} = Q_{(md)} \times B_{(sd)} \).

9. Multiply the \( Q_{ASE} \) by the output \( d \times N \), from step (5), to get \( n \times d \) output signals.

Steps 5, 6 and 7 are refer to the TDSEP algorithm and step 8 is the principle of SFA algorithm.

V. ASSUMPTIONS AND DEFINITIONS

In the proposed algorithm, there are some assumptions are considered as follows: (1) the source signals must be independent, (2) the number of mixed signals are greater than or equal to the number of the source signals, (3) there are at most one source signal have Gaussian distribution, (4) the nonlinear function \( f \) is invertible function, and (5) the mixed matrix must be of full rank.

VI. SIMULATION RESULTS AND IMPLEMENTATION

Several types of signals are used to evaluate the performance of the proposed technique. To put the results in a comparison form, the same examples in [21] and we introduced a new example.

**Experiment 1:** In the first experiment, two sinusoidal signals with different frequencies are used:

\[ S(t) = [S_1(t) \quad S_2(T)]^T \]

where \( S_1(t) = \sin(100\pi t) \) and \( S_2(t) = \sin(42\pi t) \),
with \( t = 1, \ldots, 2000 \). These source signals are nonlinearly mixed according to the following equation:

\[
X_1(t) = e^{s_1(t)} - e^{s_2(t)} \\
X_2(t) = e^{-s_1(t)} + e^{-s_2(t)}
\]  

(12)

Here, a polynomial kernel of degree 5 is used,

\[
k(a, b) = (a^T b + 1)^5
\]  

(13)

In this experiment, 21 points \((v_1, \ldots, v_{21})\) are selected. So we have \( \psi \) with size of 21 x 2000. When TDSEP is applied on \( \psi \), 21 different data vectors are generated and the slow feature analysis generates 2 x 2000 data vectors. Fig. 1 shows the original, the mixing and the estimating sources. Moreover, the scattering plot of the three signals is shown in Fig 2.

Experiment 2: In this experiment, two speech signals, \( S(t) = [S_1(t), S_2(T)]^T \), of 2000 samples are mixed. The nonlinear mixing is done using Equation (14). Fig. 3 shows the original, the mixing and the estimating sources. The scattering plot of the original signals, the mixed signals and the estimating signals are shown in Fig 4.

\[
X_1(t) = - (S_2(t) + 1) \cos(\pi S_1(t)) \\
X_2(t) = 1.5 (S_2(t) + 1) \sin(\pi S_1(t))
\]  

(14)

Gaussian radial base function (RBF) Equation (15) is employed with \( \sigma = 0.5 \).

\[
k(a, b) = e^{-\frac{|a-b|^2}{2\sigma}}
\]  

(15)

In this experiment, 17 points \((v_1, \ldots, v_{20})\) is selected so the \( \psi \) has a size of 17 x 2000 is estimated. When TDSEP is applied on \( \psi \) we will have 17 different data vectors. The slow feature analysis is used to get 2 x 2000 data vectors from 17 x 2000 different data vectors.

Experiment 3: In this experiment, the same speech signals in experiment 2 are used. The nonlinear twisted function \([12],\) Equation (16), is used. Fig. 5 shows the original, the mixing and the estimating sources. The scattering plot of the original signals, the mixed signals and the estimating signals are shown in Fig 6.

\[
X_1(t) = (S_2(t) + 3S_1(t) + 6) \cos(1.5S_1(t)) \\
X_2(t) = (S_2(t) + 3S_1(t) + 6) \sin(1.5S_1(t))
\]  

(16)

The same RBF in Equation (15) with \( \sigma = 0.5 \) is employed in this experiment. 20 points \((v_1, \ldots, v_{20})\) is selected so we will get \( \psi \) of size 20 x 2000. When TDSEP is applied on \( \psi \) we will have 20 different data vectors. The slow feature analysis is used to get 2 x 2000 data vectors from 20x2000 different data vectors.

Fig. 1. The original data are the first two signals on the top, the mixing data are in the middle and the recovering data are in the last two rows.

Fig. 2. Scatter plot of: (a) the original data (b) the mixed data. (c) the estimated data

Fig. 3. the original data are the first two signals on the top, the mixing data are in the middle and the recovering data are in the last two rows.

Fig. 4. (a) Scatter plot of the original data. (b) Scatter plot of the mixed data. (c) the estimated data.
The separation algorithm (KTDSEP). The slow feature analysis principle is used in selecting the $n$ separated signals from $d$ output signals. Moreover, a new approach to construct the orthonormal bases by selecting the bases points from the lowest frequency band in the wavelet transformed is introduced. The advantage in performance of HBSSA lies in the complexity reduction in the orthonormal bases estimation process and the fast selection of the desired signals from the output signals of the KTDSEP algorithm.

**REFERENCES**


