Abstract—Demixing independent source signals from their nonlinear mixtures is a very important issue in many scenarios. This paper presents a novel method for blindly separating unobservable independent source signals from their nonlinear mixtures. The demixing system is modeled using a parameterized neural network whose parameters can be determined under the criterion of independence of its outputs. Compared to conventional gradient-based approaches, the GA-based approach for blind source separation is characterized by high accuracy, high robustness, and high convergence rate. Simulation results are discussed to demonstrate that the proposed GA-based approach is capable of separating independent sources from their nonlinear mixtures generated by a parametric separation model.

1 Introduction

The problem of source separation is to extract independent signals from their linear or nonlinear mixtures. Source separation may be achieved in different ways according to the amount of available prior information. So-called blind source separation (BSS) is to recover original source signals from their mixtures without any prior information on the sources themselves and the mixing parameters of the mixtures. BSS techniques have received extensive attention beyond the context of signal processing due to their very weak requirements on the sources and their mixtures. BSS techniques are divided into two categories: a) the parametric form of mixtures is known, b) signal sources are statistically independent, c) the number of sensors is equal to that of sources. Because of the weak conditions, from another point of view, the separation system may be seen as a black box, which receives mixtures at its inputs and provides the estimation of original sources at its outputs. However, the outputs of the separation system are not known a priori due to the nature of blind separation. As a result, we only expect the outputs of the system to be statistically independent. Thus a BSS algorithm is to adjust the internal parameters of the separation system so as to obtain the independence of its outputs. This implies the learning algorithm of a BSS system is of an unsupervised style. When the parameters of the separation system are correctly tuned, an estimation of sources can be obtained at the outputs regardless of the indeterminacies of permutation and scaling [2].

In spite of many difficulties in separating independent sources from nonlinear mixtures, several effective models and methods were recently proposed for nonlinear BSS. Deco [13] studied a volume-conserving nonlinear transforms for nonlinear BSS. Pajunen et al. [14] used Kohonen’s self-organizing map (SOM) to extract sources from nonlinear mixtures. It is a model-free method but suffers from the exponential growth of the network complexity and interpolation error in recovering continuous sources. Taleb et al. [15] proposed an entropy-based BSS algorithm in post nonlinear mixtures. Yang et al. [16] proposed an information backpropagation algorithm for inter-channel nonlinear mixtures in the sense of entropy maximization and mutual information minimization and adopted a sigmoidal nonlinear transformation of the nonlinear model based on the work by Burel [11]. These newly developed models are established on the basis of parametric models because it is very important for nonlinear BSS to obtain unique separating results when only the independence of sources are known a priori. All of these methods are developed based on the gradient-based methods to avoid computing some unknown quantities in an unsupervised manner. Therefore, these methods are susceptible to the local minima problem during the learning process and are thus limited in many practical applications.

On the other hand, different from the likelihood estimation of probability which can be performed with local minima, the BSS problem requires to obtain a global optimal solution. Furthermore, the learning objective functions of the BSS problem are multi-modal and highly nonlinear. But all of the existing learning algorithms of BSS systems are based on stochastic gradient methods such as back-propagation method, bigradient method, and natural gradient method [1, 2]. These conventional gradient optimization techniques may converge to “bad” solutions unless good initial values are provided, which is impossible in view the blind hypothesis. Therefore, it is a very important issue to develop new BSS algorithms on the basis of global optimization techniques, which is the very topic of this paper.

In order to overcome the local minimum problem in many existing methods, we here propose a BSS approach based on a genetic algorithm (GA). In this approach we first define some cost functions to measure the independence of the
outputs, which consist of higher-order statistics (HOS) of the outputs. Then, by using a GA to minimize the cost functions, we can obtain high accurate estimation of original sources at the outputs of the separation system. A GA is a type of structured stochastic search method that mimics the process of biological evolution. Unlike conventional gradient-based approaches to nonlinear optimization problems, GAs are not susceptible to problems with local minima that arise with multi-modal error surfaces, and GAs can be guaranteed to approach global minima under suitable circumstances. Furthermore, the objective functions to be optimized can be nonlinear and even discontinuous. Since the cost functions are nonconvex and the number of decision variables (i.e., the parameters of the separation system), is small for this particular application, we believe that the GA approach is more effective and efficient than various conventional gradient-based approaches.

Our objective in this paper is to present a novel procedure for separating original independent sources from their nonlinear mixtures using a GA and HOS of the outputs of a separation system. The proposed approach differs from previous ones in several aspects. The first aspect is to utilize a global optimization method to learn the unknown parameters of the separation system. The second aspect is to minimize a predetermined cost function that measure the independence of the outputs of the separation system and can be expressed by using HOS. The third aspect is that this method is able to handle both linear and nonlinear mixtures. Our simulation studies demonstrate that the GA-based BSS scheme is robust to estimation errors in HOS and can achieve global optimal solutions from any initial values of the separation system. Furthermore, since the number of parameters to be optimized in this problem is usually small, GAs are particularly effective and efficient for this kind of optimization problems.

2 Blind Source Separation

Based on the specific application and available prior information about source signals, the source mixing process can be modeled with various mathematical models. Here we discuss a general mixing model which can model most of actual mixing processes. Let unobservable source signals be \( s(t) = [s_1(t), s_2(t), \ldots, s_n(t)]^T \) with mutually independent and stationary components. We further assume that each source has a moment of any order with a zero mean. The mixture is generally expressed as

\[
x(t) = Af(Hs(t))
\]

where \( x(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T \), the subscript \( ^T \) denotes the transpose operator, \( f(\cdot) = [f_1(\cdot), f_2(\cdot), \ldots, f_n(\cdot)]^T \) is an unknown component-wise nonlinear transform function called nonlinear mixing function (NMF), and \( H, A \in \mathbb{R}^{n \times n} \) are unknown nonsingular mixing matrices which mimic the instantaneous mixtures of signals.

The unknown mixing and proposed separating systems for BSS are illustrated in Fig. 1. As shown in the figure, the unknown mixing system at the left part of Fig. 1 can be modeled as a cascaded instantaneous linear mixing, component-wise nonlinear transformation and linear mixing. Actually, this model is realistic since most of the practical mixing systems can be well described by Fig. 1 which contains both channel and inter-channel nonlinearities. At meantime, we also give the separating system at the right part of Fig. 1 which is an inverse procedure of the mixing process among which the nonlinear separating function (NSF) \( g = [g_1(\cdot), g_2(\cdot), \ldots, g_n(\cdot)]^T \) of the separating system is a parametric inverse function of the nonlinear mixing transform function \( f \).

The task of the separating system is to recover the unobservable original signals \( s(t) \) from the observable mixtures \( x(t) \) without any knowledge of the source signals \( s(t) \), the mixing matrices \( H, A \) or the parameters of the nonlinear transform function \( f(\cdot) \). Since we assume the nonlinear transform function \( f(\cdot) \) and its inverse \( g(\cdot) \) are of parametric network formulations, our problem is to estimate the parameters of demixing matrices \( B \) and \( W \) and the separating function for the outputs \( y(t) \) to approximate the original sources apart from the indetermination of a permutation and scaling due to the weakness of the hypotheses. That is to say the learning algorithm of this BSS problem is to tune the parameters of the separation system so as to achieve the independence of the outputs.

As shown in Fig. 1, the outputs of the separating system are

\[
y(t) = Wg[Bx(t)].
\]

Substituting eqn. (1) into eqn. (2), we can obtain

\[
y(t) = PDs(t)
\]

where \( P \) is a permutation matrix and \( D \) is a nonsingular and diagonal matrix. Eqn. (3) holds when the following condition is satisfied

\[
BA = I, \quad g(\cdot) = f^{-1}(\cdot), \quad WH = PD.
\]

This means that the components of the outputs \( y \) are independent.
3 Cost Functions

3.1 Independence Conditions and Constraints

In order to measure the independence of the outputs, it is very natural to use the probability criterion of random variables since stochastic independence of random variables is defined based on its probability distribution. Let \( p(y) \) be the probability density function (pdf) of random variable \( y \), \( p_y(y_i) \) be marginal pdf of \( y_i \), and \( p(y) \) be the joint pdf of random vector \( y \), we have the following conditions for the independence of random variables.

**Independence Conditions:** For the independent components of random vector \( y \), the following statements are mutually equivalent

(a) \( p(y) = \prod_{i=1}^{n} p_y(y_i) \).

(b) \( KL[p_y(y), p(y)] = \int p_y(y) \log(\frac{p_y(y)}{p(y)}) dy = 0 \).

(c) \( \int [p(y) - \prod_{i=1}^{n} p_y(y_i)]^2 dy = 0 \).

Although the above conditions are necessary and sufficient for the independence of the outputs of the separation system, they need the estimation of the probability or entropy of the outputs. In order to measure the independence of the outputs, it is very useful in the context of blind signal processing according to eqn. (5). If the components in \( y \) are related as a Fourier transform pair; i.e.,

\[
\Phi_y(v) = \int p_y(y) \exp(-j v^T y) dy = E[\exp(-j v^T y)]
\]  

where \( v = [v_1, \ldots, v_n]^T \) is a vector of variables in Fourier transform domain and \( j = \sqrt{-1} \) stands for an imaginary unit. In particular, for the single random variable, we have

\[
\Phi_y(v_i) = E[\exp(-j v_i y_i)] \quad \text{for the single random variable,}
\]

In what follows we will use two methods to re-express the independence conditions of the theorem. One uses the moment statistics of the nonlinear function transform of the outputs. The other uses the HOS of the outputs. In order to further facilitate learning, constrain the outputs, and achieve good separation results, we impose two conventional constraints on the outputs of the separation system:

\[
C_1 := \sum_{i=1}^{n} E(y_i) = 0,
\]

\[
C_2 := \sum_{i=1}^{n} [E(y_i^2) - 1]^2 = 0.
\]

Constraint (7) is very natural and can always be achieved by adjusting the bias of neurons in the output layer when the system is implemented in a neural network. Constraint (8) is used to limit the variance of the outputs to be unity which is very useful in the context of blind signal processing according to many simulation studies. So we will add these two constraints in the cost functions to be defined and call them basic constraints.

3.2 Dependence measure by nonlinear function moment

In view of eqns. (5)-(6), according to the probability theory on the independence of random variables, condition (a) can be equivalently expressed by using the characteristic function as

\[
\Phi_y(v) = \prod_{i=1}^{n} \Phi_{y_i}(v_i). \tag{9}
\]

Since the characteristic function of a random vector is equal to the product of that of each component of the outputs when they are mutually independent, if we take the Taylor expansion at both sides of eqn. (9), cross-moments must be zero. As a result, we get an idea that if all joint cross-moments of the outputs are forced to zero, the independence of the outputs can be achieved. This means that we can get an expression of the dependence measure by using moments which are easier to cope with. Unfortunately, it is impossible to take the moments of all orders into account since the exhaust computation is not feasible in practice. To amend it, one can use a nonlinear transform function of signals before the computation of its moments. In this way, we can indirectly take all the moments into account to separate nonlinear mixtures. Suppose we have \( n \) infinitely differentiable functions \( h_1(\cdot), \ldots, h_n(\cdot) \) and have the following Taylor expansion,

\[
h_i(y_i) = \sum_{j=0}^{\infty} \frac{h_i^{(j)}(0)}{j!} y_i^j = \sum_{j=0}^{\infty} c_{j,i} y_i^j, \quad i = 1, \ldots, n \quad \tag{10}
\]

where we define \( c_{j,i} = \frac{h_i^{(j)}(0)}{j!} \) to simplify notations. If we take the central moment of quantity \( \prod_{i=1}^{n} h_i(y_i) \), we can get

\[
M(h_1(y_1), \ldots, h_n(y_n)) = E(\prod_{i=1}^{n} h_i(y_i)) - \prod_{i=1}^{n} E(h_i(y_i)). \tag{11}
\]

Substituting eqn. (10) into eqn. (11), we have

\[
M(h_1(y_1), \ldots, h_n(y_n)) = \sum_{j_1=0}^{\infty} \cdots \sum_{j_n=0}^{\infty} c_{j_1,1} \cdots c_{j_n,n} M(y_1^{j_1}, \ldots, y_n^{j_n}). \tag{12}
\]

If the components in \( y \) are mutually independent, then we have \( M(y_1^{j_1}, \ldots, y_n^{j_n}) = 0 \) according to eqn. (11). From eqn. (12), we know that \( M(h_1(y_1), \ldots, h_n(y_n)) \) is null if all \( M(y_1^{j_1}, \ldots, y_n^{j_n}) = 0 \) for \( j_k = 1, \ldots, \infty \) where \( k = 1, \ldots, n \).
1, ⋯, n. Hence \( M(h_1(y_1), \cdots, h_n(y_n)) = 0 \) is a necessary condition for \( y_1, \cdots, y_n \) to be independent. Based on the above discussions, we can define the following cost function for BSS in case of two independent sources.

\[
C(\theta) = \sum_{i=1}^{n} \sum_{j \neq i}^{n} M(h_i(y_1), h_j(y_2))^2 + \alpha C_1 + \beta C_2 \tag{13}
\]

where \( \theta \) is the parameter vector of the separation system, \( \alpha \) and \( \beta \) are two positive constants weighting the dependence measure and the basic constraints. Even though this is only a necessary condition for the dependence of outputs, but it is very easy to deal with and good separation results are often obtained by minimizing this cost function, as will be further illustrated by our simulation results.

3.3 Dependence measure by higher-order moments

By using statement (c) to measure the dependence of random variables, we can get another expression of dependence by the higher-order moments of the outputs \( y \). Before we proceed, a smoothing window should be enforced on the difference of the joint pdf and product of marginal pdf of \( y \) for reducing the fluctuations on the coefficients of the expression and decreasing the errors of real data. Hence the measure of the dependence of random variables becomes

\[
D(y) = \|p_y(y) - \prod_{i=1}^{n} p_y(y_i)\|_w^2 = \int \left[ [p_y(y) - \prod_{i=1}^{n} p_y(y_i)] * w(y) \right]^2 \, dy \tag{14}
\]

with

\[
w(y) = \prod_{i=1}^{n} w(y_i), \tag{15}
\]

where \(*\) in eqn. (14) denotes the convolution operator, \( w(y_i) (i = 1, \cdots, n) \) are window functions that may take different forms such as rectangular window function, up-cosine function, or Gaussian window function and can be selected according to specific applications. By considering eqn. (5), taking the Fourier transform of eqn. (14), and at meantime considering the properties of the Fourier transforms, we can obtain

\[
\int \| \Phi_y(v) - \prod_{i=1}^{n} \Phi_{y_i}(v_i)\|^2 W(v) \, dv, \tag{16}
\]

where \( W(v) \) is the Fourier transform of the window function \( w(y) \) defined in eqn. (15).

Since most actual signals considered are bounded and the moments of bounded random variables always exist, we can take Taylor expansion around the origin of the characteristic function.

\[
\Phi_y(v) = \sum_{\alpha_1, \cdots, \alpha_n} \frac{1}{\alpha_1! \cdots \alpha_n!} \frac{\partial^{\alpha_1 + \cdots + \alpha_n} \Phi_y(0)}{\partial v_1^{\alpha_1} \cdots \partial v_n^{\alpha_n}} v_1^{\alpha_1} \cdots v_n^{\alpha_n}, \tag{17}
\]

where

\[
\Phi_{y_i}(v_i) = \sum_{\alpha_i = 0}^{\infty} \frac{1}{\alpha_i!} \frac{\partial^{\alpha_i} \Phi_y(0)}{\partial v_i^{\alpha_i}} v_i^{\alpha_i}. \tag{18}
\]

According to the convention that the partial derivative of order zero of a function is the function itself, we can obtain

\[
\Phi_y(v) - \prod_{i=1}^{n} \Phi_{y_i}(v_i) = \sum_{\alpha_1, \cdots, \alpha_n} Q_{\alpha_1, \cdots, \alpha_n} v_1^{\alpha_1} \cdots v_n^{\alpha_n} \tag{19}
\]

with

\[
Q_{\alpha_1, \cdots, \alpha_n} = \frac{1}{\alpha_1! \cdots \alpha_n!} \left\{ \frac{\partial^{\alpha_1 + \cdots + \alpha_n} \Phi_y(0)}{\partial v_1^{\alpha_1} \cdots \partial v_n^{\alpha_n}} - \sum_{i=0}^{\infty} \frac{\partial^{\alpha_i} \Phi_y(0)}{\partial v_i^{\alpha_i}} \right\} \tag{20}
\]

According to the definitions of a characteristic function and the moment of random variables, we can have the following relationship between the origin moment and the derivatives of the characteristic function at zero.

\[
\frac{\partial^{\alpha_1 + \cdots + \alpha_n} \Phi_y(0)}{\partial v_1^{\alpha_1} \cdots \partial v_n^{\alpha_n}} = (-j)^{\alpha_1 + \cdots + \alpha_n} E[y_1^{\alpha_1} \cdots y_n^{\alpha_n}], \tag{21}
\]

\[
\frac{\partial^{\alpha_i} \Phi_y(0)}{\partial v_i^{\alpha_i}} = (-j)^{\alpha_i} E[y_i^{\alpha_i}]. \tag{22}
\]

Thus

\[
Q_{\alpha_1, \cdots, \alpha_n} = \frac{1}{\alpha_1! \cdots \alpha_n!} (-j)^{\alpha_1 + \cdots + \alpha_n} M_{\alpha_1, \cdots, \alpha_n}, \tag{23}
\]

where \( M_{\alpha_1, \cdots, \alpha_n} \) stands for the \( (\alpha_1 + \cdots + \alpha_n) \)-th central moment of \( y \).

In view of eqns. (16)- (23), the expression of the measure of dependence can be rewritten as

\[
\int \left| \sum_{\alpha_1, \cdots, \alpha_n} Q_{\alpha_1, \cdots, \alpha_n} v_1^{\alpha_1} \cdots v_n^{\alpha_n} \right|^2 W(v) \, dv, \tag{24}
\]

which can also be transformed as

\[
\int \sum_{\alpha_1, \cdots, \alpha_n} \sum_{\beta_1, \cdots, \beta_n} Q_{\alpha_1, \cdots, \alpha_n} Q_{\beta_1, \cdots, \beta_n} v_1^{\alpha_1 + \beta_1} \cdots v_n^{\alpha_n + \beta_n} W(v) \, dv. \tag{25}
\]

By interchanging the sequence of summation and integral, we can obtain

\[
\sum_{\alpha_1, \cdots, \alpha_n} \sum_{\beta_1, \cdots, \beta_n} Q_{\alpha_1, \cdots, \alpha_n} Q_{\beta_1, \cdots, \beta_n} \int v_1^{\alpha_1 + \beta_1} \cdots v_n^{\alpha_n + \beta_n} W(v) \, dv. \tag{26}
\]

Finally, we can reach

\[
\sum_{\alpha_1, \cdots, \alpha_n} \sum_{\beta_1, \cdots, \beta_n} d_W(\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_n) M_{\alpha_1, \cdots, \alpha_n} M_{\beta_1, \cdots, \beta_n} \tag{26}
\]
with

\[ d_W(\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n) = (-j)^{\alpha_1+\cdots+\alpha_n-\beta_1-\cdots-\beta_n} R_W(\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n) \]  

(27)

where

\[ R_W(\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n) = \int d_1^{\alpha_1+\beta_1} \cdots d_n^{\alpha_n+\beta_n} W(v) dv \]  

(28)

is the \((\alpha_1+\beta_1+\cdots+\alpha_n+\beta_n)\)-th moment of the transformed window function \(W(v)\). It should be pointed out that the positive constant in eqn. (25) is disregarded without any effect on our conclusion.

In this way, we have obtained the exact expression of the dependence measure in eqn. (14) in terms of the HOS of outputs, which is much easier to cope with. Because eqn. (26) is always nonnegative, it becomes zero if and only if all the HOS \(M_{\alpha_1, \ldots, \alpha_n}\) are equal to zero. It also means that the independence of the outputs has been achieved by dealing with the HOS of outputs only. In addition, according to the property and smoothing effect of the window functions, it is easily shown that the coefficient \(d_W(\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n)\) gradually decreases to zero as the index of sum \(\alpha_1+\beta_1+\cdots+\alpha_n+\beta_n\) increases. This property allows us to use the finite sum of lower orders to approximate eqn. (26) with a satisfactory accuracy. For this purpose we can define the following cost function

\[ C(\theta) = \sum_{\alpha_1, \ldots, \alpha_n} \sum_{\beta_1, \ldots, \beta_n} d_W(\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n) \\
\cdot M_{\alpha_1, \ldots, \alpha_n} M_{\beta_1, \ldots, \beta_n} + \alpha C_1 + \beta C_2 \]  

(29)

where the basic constraints are also included in the cost function to help getting the correct separation in nonlinear mixtures. This criterion is exactly the same as the dependence criterion of outputs. So it is also a sufficient and necessary condition for the independence of outputs. As a result, minimizing eqn. (29) will result in the independence of outputs and reach the goal of BSS. However, the cost function in eqn. (29) includes the moments of the outputs in any order which are impossible to realize in specific calculation. Therefore, in our specific implementation, we only take finite moments (up to \(K\)-th moment) into account, then eqn. (29) becomes

\[ C(\theta) = \sum_{i, \alpha_1 < \alpha_i} \sum_{j, \beta_1 < \beta_j} d_W(\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n) \\
\cdot M_{\alpha_1, \ldots, \alpha_n} M_{\beta_1, \ldots, \beta_n} + \alpha C_1 + \beta C_2. \]  

(30)

By using the derivation results of eqns. (7), (8), (13) and (30), we approach our aforementioned goal of defining cost functions by HOS for BSS. The minimization of the cost function in eqn. (30) gives the correct separation results for linear or nonlinear mixtures in our parametric model.

### 4 GA-based Separation Approach

The GA-based BSS algorithm can be implemented as the following iterative procedure.

1. An initial population \(\{\hat{\theta}_i\}_{i=1}^N\) of size \(N\) is created from a random initial set of parameters. The encoding length of each parameter is 16 bits. By decoding the individual to get the parameter of the system, the fitness for each individual is evaluated using \(1/C(\theta)\).

2. Two mates are selected for reproduction with probabilities that are proportional to their fitness values using the roulette wheel selection.

3. The multi-point crossover operator with crossover probability \(P_c\) is applied to the two mates, and two offsprings are generated.

4. The mutation operator with probability \(P_m\) is applied to the newly generated offspring.

5. The fitness values for the offsprings are computed after they are decoded as the parameter sets of the parametric system.

6. Steps 2-5 are repeated until an entirely new population of individuals is generated.

7. The previous population is replaced with the new population incorporated with an elitism strategy.

8. If the stopping criterion is satisfied, go to step 11.

9. If generation number is greater than a predetermined value go to step 10, otherwise go to step 2.

10. Re-initialization of the population with the best individual in the current population survival, go to step 2.

11. Output the individual with the best fitness value and terminate the iterative procedure.

### 5 Simulation Results

In order to verify the validity and performance of the proposed algorithm, several computer simulations were conducted to test the GA-based approach to blind separation of independent sources from their nonlinear mixture.

**Example 1:** Consider the mixing case of two independent random signals: a random binary signal and a random signal with uniformly distribution on interval \((-1,1)\). The mixing matrices \(\mathbf{H}\) and \(\mathbf{A}\) are randomly chosen as

\[
\mathbf{H} = \begin{pmatrix} 0.6744 & 0.3248 \\ 0.2461 & 0.9217 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 0.9121 & 0.2292 \\ 0.4763 & 0.7348 \end{pmatrix}.
\]

The nonlinear transform function in each channel is the hyperbolic tangent function with unity amplitude and half-unity gain in eqn. (31).

\[
f_i(u_i) = -2\gamma \arctanh(-\delta_i u_i).
\]  

(31)

Fig. 2 depicts the two original sources and their mixtures.

The separation system adopts a two-layer neural network with piecewise linear neurons in input and output layers and sigmoid neurons in hidden units. Specifically, the sigmoid.
activation function are in the form of
\[
g_i(t_i) = \alpha_i \frac{1 - \exp(-\beta_i v_i)}{1 + \exp(-\beta_i v_i)}
\] (32)
where \(\alpha_i > 0\) and \(\beta_i > 0\) are respectively the amplitude and gain of the \(i\)-th neuron in the hidden layer. Thus we can derive the corresponding mixing nonlinear transform in the left part of Fig. 1.

For this parametric separation system, we have 12 parameters (i.e., matrices \(\mathbf{B}\) and \(\mathbf{W}\), and amplitude \(\alpha\) and gain \(\beta\) for two neurons in the hidden layer) to be tuned during the learning process. Because the amplitude parameter of the activation function can be absorbed into the demixing matrices, we can let \(\alpha_i = 1\) without loss of generality. Thus we have 10 parameters left to be determined.

In our GA-based method, we encoded each of the 10 parameters as 16-bit binary string. So each individual is represented by a 160-bit binary string. The population size \(N = 30\). The crossover probability \(P_c = 0.99\) while mutation probability \(P_m = 0.15\). In order to expedite convergence, a multiple-point crossover is used in our experiments for exploring the better solution in the solution space. Specifically, the number of crossovers between two mates is equal to that of decision parameters of the problem. Mutation operator with probability \(P_m\) is also employed to prevent from the premature of the GA learning process. We also assume the maximum generation for GA process to be 30. If the stopping criterion is not satisfied, the GA process will be restarted with the elitism strategy. In order to accelerate the convergence, we adopt a restart-up strategy in this experiment.

The data length of the two sources are only 150 points in this experiment. When tuning the parameters, these data are fed into the separation system repeatedly until the GA-based learning process converges. All the results given here were obtained by averaging over 50 different runs. Each run used the same mixture samples with a different randomly initialized population.

For the cost function of eqn. (13), we choose two sets of nonlinear functions. Case I is \(f_1(x_1) = \tanh(x_1)\) and \(f_2(x_2) = x_2^3\). Case II is \(f_1(x_1) = \tanh(x_1)\) and \(f_2(x_2) = \tanh(0.5x_2)\). Two cases gives similar results. Figs. 3 and 4 plot the separation results and evolution curves based on the cost functions in the two cases.

For the cost function of eqn. (30), the order of the Taylor expansion is chosen as up to four, i.e., \(K = 4\). We also employ a rectangular windowing function for the calculation of the cost function. Certainly, other type window functions such as Gaussian window and up-cosine window can also be adopted even though the selection of window functions are closely related to the application. Figs. 5-6 depict the evolution of the cost function in eqn. (30) and the separated signals.
by using our GA-based BSS approach with this cost function, respectively.

It turns out from our simulation results that the performance of the cost function by HOS is slightly better than that of the cost function by nonlinear function but the computational complexity of the former is substantially higher than that of the latter. So a compromise should be made between the two cases. If the accuracy and performance of a given problem is more important, the former is preferred, otherwise the latter is desirable to save limited computational resources.

**Example 2:** In order to further test the practical applicability of the proposed method, we consider a “cocktail party” problem. Two speakers, a man and a woman, are considered in this test. The sampling rate is 11.02 KHz and each sample is quantized as eight bits. Their speeches are mixed by randomly mixing matrices and monotonically nonlinear transformation. The mixtures, also digitalized as 8-bit per point, are used as the inputs of the separation system. All data are normalized in the range of $[-1, +1]$ for the sake of computation convenience. Fig. 7 shows the original speeches to be recovered and their mixtures.

We use a two-layer neural network with sigmoid activation function for the neurons in a hidden layer but other neurons are linear. Also, there are 10 decision parameters to be determined in this case. For the GA, we choose the population size

$$N = 30,$$ The crossover probability $P_c = 1.0$ while mutation probability $P_m = 0.25$. Similar to Example 1, a multiple-point crossover and elitism strategy are employed. There is no re-startup mechanism in this experiment. The evolution process of the cost functions are averaged on 5 runs. Fig. 8 shows the separated results by the proposed GA-based BSS approach based on the cost function defined in eqn. (30). The separating experiment based on the cost function in eqn. (13) is also carried out by using the proposed approach. To save space, the separated results are not shown due to its somewhat similarity to that in Fig. 8. It is shown from the graphs of Fig. 7 and Fig. 8 that our proposed method achieved the successful separation of the two speech signals from their nonlinear mixtures.

**6 Concluding Remarks**

A GA-based BSS approach has been developed for blind source separation from the nonlinear mixtures of independent sources. The proposed method overcomes the local minima problem occurred in the conventional gradient-based methods and can obtain global optimal solutions to nonlinear BSS problem from any initial conditions. Extensive simulation re-
sults have demonstrated the validity and performance of this GA-based BSS method. Apart from obtaining globally optimal separations, the proposed GA-based method is also characterized by high accuracy, high robustness, and fast convergence. Since the number of parameters to be optimized in BSS is usually small, the GA-based method is very suitable for this kind of problem. Another feature of this method is that it also suits the case of short available data. This is particular useful in some medical applications where source signals may appear in a very short time period.

Bibliography


