Nonlinear Blind Source Separation Using Kernel Multi-set Canonical Correlation Analysis

Hua-Gang Yu, Gao-Ming Huang, and Jun Gao
College of Electronic Engineering, Naval Univ. of Engineering, Wuhan, China
Email: yuhuagang103@163.com, hgaom@163.com, gaojunnj@163.com

Abstract—To solve the problem of nonlinear blind source separation (BSS), a novel algorithm based on kernel multi-set canonical correlation analysis (MCCA) is presented. Combining complementary research fields of kernel feature spaces and BSS using MCCA, the proposed approach yields a highly efficient and elegant algorithm for nonlinear BSS with invertible nonlinearity. The algorithm works as follows: First, the input data is mapped to a high-dimensional feature space and perform dimension reduction to extract the effective reduced feature space, translate the nonlinear problem in the input space to a linear problem in reduced feature space. In the second step, the MCCA algorithm was used to obtain the original signals.

Index Terms—nonlinear blind source separation, kernel feature spaces; multi-set canonical correlation analysis; reduced feature space; joint diagonalization

I. INTRODUCTION

In recent years, blind source separation (BSS) raises great interest. In fact, BSS plays an important role in many diverse application areas, such as radio communications, radar, sonar, seismology, image processing, speech processing (cocktail party problem) and biomedical signal analysis where multiple sensors are involved. In the last years, linear BSS has become relatively well established signal processing and data analysis techniques [1]. On the other hand, nonlinear BSS is technique that is still largely under development, and has the potential to become rather powerful tools [2].

Kernel-based methods have also been considered for solving the nonlinear BSS problem [3], [4]. The data are first implicitly mapped to high-dimensional feature space, and the effective reduced feature space in feature space is extracted, translate the nonlinear problem in the input space to a linear problem in reduced feature space.

Canonical correlation analysis (CCA) is a classical tool in multivariate statistical analysis to find maximally correlated projections between two data sets, since it was proposed by H. Hotelling [5]. There are detail descriptions of CCA in [6], which has been widely used in many modern information processing fields, such as for test of independence [7], blind equalization of MIMO channels [8] and BSS [9],[10]. With CCA, the objective is to find a transformation matrix which is applied to mixtures and maximizes the autocorrelation of each of the recovered signals (the outputs of the transformation matrix). By maximizing this autocorrelation, the original uncorrelated source signals will be recovered. This approach rests on the idea that the sum of any uncorrelated signals has an autocorrelation whose value is less or equal to the maximum value of individual signals [11]. The algorithm based on CCA which computation burden is little and the flexibility is strong, can satisfy the demand of engineering application. Based on the idea of maximize generalized relativity measurement, J. Kettenring present-ed several ways to generalize CCA to more than two sets of variables, namely MCCA (Multi-set CCA) [12]. By assuming that sources are not correlated with the others and every source has a different temporal structure, which is a mild condition that can be easily, satisfied in practical applications, a MCCA linear BSS algorithm was proposed in [13]. This paper proposed a novel nonlinear BSS algorithm based on kernel feature spaces and MCCA. The algorithm can adapt to nonlinear BSS with invertible nonlinearity. The details of the new method will be described in the following sections. The paper is organized as follows: Section II presents the detail analysis of linear BSS algorithm based on MCCA. In the Section III, the new nonlinear BSS algorithm based on kernel MCCA will be analyzed. Then some simulations of the algorithm proposed in this paper are conducted in the Section IV. Section V is the conclusion.

II. LINEAR BSS ALGORITHM BASED ON MCCA

A. Linear BSS problem formulation

Consider the following instantaneous linear mixture model:

$$\boldsymbol{x}(t) = \boldsymbol{H}\boldsymbol{s}(t) + \boldsymbol{n}(t). \qquad (1)$$

where $\boldsymbol{x}(t) = (x_1(t), x_2(t), \text{L } x_m(t))^T$ $(t = 1, \text{L } T)$ is the m-dimensional vector of mixed signals observed by m sensors, $\boldsymbol{n}(t)$ is the additive noise vector, $\boldsymbol{H}$ is the unknown $m \times n$ mixing matrix, $\boldsymbol{s}(t) = (s_1(t), s_2(t), \text{L } s_n(t))^T$ $(t = 1, \text{L } T)$ is the n-dimensional vector of source signals (which is also unknown and $m \geq n$), and the superscript ^{T} denotes the transpose operator. BSS methods aim at estimating the source signals $\boldsymbol{s}(t)$. Suppose the source signals are not correlated with the others and every source has a different temporal structure, and the additive noise vector $\boldsymbol{n}(t)$ is statistically independent of $\boldsymbol{s}(t)$.

The task of BSS is to estimate the mixing matrix $\boldsymbol{H}$ (or is pseudo-inverse, $\boldsymbol{W} = \boldsymbol{H}^{\#}$ that is referred to as the demixing matrix), given only a finite number of observation data $\boldsymbol{x}(t)$ $(t = 1, \text{L } T)$. And obtain the estimated source signals $\boldsymbol{y}(t) = (y_1(t), y_2(t), \text{L } y_n(t))^T$ $(t = 1, \text{L } T)$:

$$\boldsymbol{y}(t) = \boldsymbol{W}\boldsymbol{x}(t) . \qquad (2)$$

We just consider the case of $m = n$, for less sources than mixtures ($m > n$) the BSS problem is said to be over-determined, and it is easily reduced to a square BSS problem by selecting m mixtures or applying some more sophisticated preprocessing like PCA.

B. MCCA

CCA is a multivariate statistical technique similar in spirit to principal component analysis (PCA). While PCA works with a single random vector and maximizes the variance of projections of the data, CCA works with a pair of random vectors (or in general with a set of m random vectors) and maximizes correlation between sets of projections.

Given two random vectors, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, of dimension p_1 and p_2. The first canonical correlation can be defined as the maximum possible correlation between the two projections $\boldsymbol{u}_1 = \boldsymbol{\alpha}_1^{(1)T}\boldsymbol{x}_1$ and $\boldsymbol{v}_1 = \boldsymbol{\alpha}_1^{(2)T}\boldsymbol{x}_2$ of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ [6]:

$$\rho_1(\boldsymbol{x}_1, \boldsymbol{x}_2) = \max_{\boldsymbol{\alpha}_1^{(1)}, \boldsymbol{\alpha}_1^{(2)}} corr\left(\boldsymbol{\alpha}_1^{(1)T}\boldsymbol{x}_1, \boldsymbol{\alpha}_1^{(2)T}\boldsymbol{x}_2\right)$$

$$= \max_{\boldsymbol{\alpha}_1^{(1)}, \boldsymbol{\alpha}_1^{(2)}} \frac{\boldsymbol{\alpha}_1^{(1)T}\boldsymbol{C}_{12}\boldsymbol{\alpha}_1^{(2)}}{\left(\boldsymbol{\alpha}_1^{(1)T}\boldsymbol{C}_{11}\boldsymbol{\alpha}_1^{(1)}\right)^{\frac{1}{2}}\left(\boldsymbol{\alpha}_1^{(2)T}\boldsymbol{C}_{22}\boldsymbol{\alpha}_1^{(2)}\right)^{\frac{1}{2}}} \qquad (3)$$

where $\boldsymbol{C}_{11} = \boldsymbol{x}_1\boldsymbol{x}_1^T, \boldsymbol{C}_{12} = \boldsymbol{x}_1\boldsymbol{x}_2^T, \boldsymbol{C}_{21} = \boldsymbol{x}_2\boldsymbol{x}_1^T, \boldsymbol{C}_{22} = \boldsymbol{x}_2\boldsymbol{x}_2^T$.

After finding the first pair of optimal vectors $\boldsymbol{\alpha}_1^{(1)}$ and $\boldsymbol{\alpha}_1^{(2)}$, we can proceed to find the second pair $\boldsymbol{\alpha}_2^{(1)}$ and $\boldsymbol{\alpha}_2^{(2)}$ which maximizes the correlation and at the same time ensures that the new pair of combinations $\{\boldsymbol{u}_2, \boldsymbol{v}_2\}$ is uncorrelated with the first set $\{\boldsymbol{u}_1, \boldsymbol{v}_1\}$. This process is repeated until we find all the $\min(p_1, p_2) = p$ pairs of optimal vectors $\boldsymbol{\alpha}_i^{(1)}$ and $\boldsymbol{\alpha}_i^{(2)}$, $i = 1, 2, \cdots, p$.

Since the choice of rescaling is therefore arbitrary, Normalizing the vectors $\boldsymbol{\alpha}_i^{(1)}$ and $\boldsymbol{\alpha}_i^{(2)}$ by letting $\boldsymbol{\alpha}_i^{(1)T}\boldsymbol{C}_{11}\boldsymbol{\alpha}_i^{(1)} = 1$ and $\boldsymbol{\alpha}_i^{(2)T}\boldsymbol{C}_{22}\boldsymbol{\alpha}_i^{(2)} = 1$, we see that CCA reduces to the following Lagrangian [6]:

$$L(\lambda_i, \boldsymbol{\alpha}_i^{(1)}, \boldsymbol{\alpha}_i^{(2)}) = \boldsymbol{\alpha}_i^{(1)T}\boldsymbol{C}_{12}\boldsymbol{\alpha}_i^{(2)} - \frac{\lambda_i^{(1)}}{2}(\boldsymbol{\alpha}_i^{(1)T}\boldsymbol{C}_{11}\boldsymbol{\alpha}_i^{(1)} - 1)$$
$$- \frac{\lambda_i^{(2)}}{2}(\boldsymbol{\alpha}_i^{(2)T}\boldsymbol{C}_{22}\boldsymbol{\alpha}_i^{(2)} - 1) \qquad (4)$$

Taking derivatives in respect to $\boldsymbol{\alpha}_i^{(1)}$ and $\boldsymbol{\alpha}_i^{(2)}$, the CCA problem can be obtained by solving the following generalized Eigen values problem

$$\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21}\boldsymbol{\alpha}_i^{(1)} = \lambda_i^2\boldsymbol{C}_{11}\boldsymbol{\alpha}_i^{(1)}$$
$$\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{\alpha}_i^{(2)} = \lambda_i^2\boldsymbol{C}_{22}\boldsymbol{\alpha}_i^{(2)} . \qquad (5)$$

There's a generalized Eigen problem of the form $\boldsymbol{Ax} = \lambda\boldsymbol{Bx}$. We can therefore find the coordinate system that optimizes the correlation between corresponding coordinates by first solving for the generalized eigenvectors of (5) to obtain $\boldsymbol{\alpha}_i^{(1)}$ and $\boldsymbol{\alpha}_i^{(2)}$.

According to the ways in [6] we can generalize CCA problem to the minimization of the total distance.

For CCA, two matrices $\boldsymbol{A}^{(1)}$ and $\boldsymbol{A}^{(2)}$ contain the vectors $(\boldsymbol{\alpha}_1^{(1)}, \cdots, \boldsymbol{\alpha}_p^{(1)})$ and $(\boldsymbol{\alpha}_1^{(2)}, \cdots, \boldsymbol{\alpha}_p^{(2)})$. The CCA problem can also be express as

$$\max_{\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}} \quad tr(\boldsymbol{A}^{(1)T}\boldsymbol{C}_{12}\boldsymbol{A}^{(2)})$$
$$s.t. \qquad \boldsymbol{A}^{(k)T}\boldsymbol{C}_{kk}\boldsymbol{A}^{(k)} = \boldsymbol{I}$$
$$\boldsymbol{\alpha}_i^{(k)T}\boldsymbol{C}_{kl}\boldsymbol{\alpha}_j^{(l)} = 0 \qquad (6)$$
$$k, l = \{1, 2\}, l \neq k, i, j = 1, \cdots, p, j \neq i$$

where $\boldsymbol{I}$ is an identity matrix with size $q \times q$.

The canonical correlation problem can be transformed into a distance minimization problem where the distance between two matrices is measured by the Frobenius norm:

$$\min_{\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}} \quad \left\|\boldsymbol{x}_1\boldsymbol{A}^{(1)} - \boldsymbol{x}_2\boldsymbol{A}^{(2)}\right\|_F$$
$$s.t. \qquad \boldsymbol{A}^{(k)T}\boldsymbol{C}_{kk}\boldsymbol{A}^{(k)} = \boldsymbol{I} \qquad . \qquad (7)$$
$$\boldsymbol{\alpha}_i^{(k)T}\boldsymbol{C}_{kl}\boldsymbol{\alpha}_j^{(l)} = 0$$
$$k, l = \{1, 2\}, l \neq k, i, j = 1, \cdots, p, j \neq i$$

We can give a similar definition for MCCA. Given m multivariate random variables in matrix form $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_K\}$. We are looking for the linear combinations of the columns of these matrices in the matrix form $\boldsymbol{A}^{(1)}, \cdots, \boldsymbol{A}^{(K)}$ such that they give the optimum solution of the problem:

$$\min_{\boldsymbol{A}^{(1)}, \cdots, \boldsymbol{A}^{(K)}} \quad \sum_{k,l=1, k \neq l}^{K} \left\|\boldsymbol{x}_k\boldsymbol{A}^{(k)} - \boldsymbol{x}_l\boldsymbol{A}^{(l)}\right\|_F$$
$$s.t. \qquad \boldsymbol{A}^{(k)T}\boldsymbol{C}_{kk}\boldsymbol{A}^{(k)} = \boldsymbol{I}$$
$$\boldsymbol{\alpha}_i^{(k)T}\boldsymbol{C}_{kl}\boldsymbol{\alpha}_j^{(l)} = 0 \qquad (8)$$
$$k, l = 1, \cdots, K, l \neq k, i, j = 1, \cdots, q, j \neq i$$

which is the sum of the squared Euclidean distances between all of the pairs of the column vectors of the matrices $\boldsymbol{x}_k\boldsymbol{A}^{(k)}, k = 1, \cdots, K$.

C. BSS Algorithm Analysis

We start by specifying the signal. It is assumed that the sources are spatially uncorrelated, the correlation matrix of the sources $\boldsymbol{R}_{ss}[\tau] = \text{E}\{\boldsymbol{s}(t)\boldsymbol{s}^T(t - \tau)\}$ is a diagonal matrix, where $\text{E}\{\cdot\}$ denotes the statistical expectation operator. For nonzero correlation lags, we have

$$R_{ss}[\tau] = E\{s(t)s^T(t-\tau)\} = diag\{\lambda_0(\tau), L, \lambda_m(\tau)\} \quad (9)$$

with $\lambda_i(\tau) \neq 0$ for some nonzero delays τ.

First, we choose the observed signals $x(t)$ as x_1 and $x(t-\tau)$ as x_2. Then the Eigen value problem in (4) becomes

$$R_{xx}(\tau)R_{xx}^{-1}(0)R_{xx}(\tau)\alpha_i^{(1)} = \lambda_i^2 R_{xx}(0)\alpha_i^{(1)}$$
$$R_{xx}(\tau)R_{xx}^{-1}(0)R_{xx}(\tau)\alpha_i^{(2)} = \lambda_i^2 R_{xx}(0)\alpha_i^{(2)} \quad (10)$$

where $R_{xx}(\tau)$ and $R_{xx}(0)$ are the correlation matrices of the mixed signals.

In the context of BSS, the vectors $\alpha_i^{(1)}$ and $\alpha_i^{(2)}$ in (5) are the same and denote them by $w_i = \alpha_i^{(1)} = \alpha_i^{(2)}$. Which is the demixing vector applied to the mixed signals. The Separate signals can be obtained

$$y_i(t) = w_i^T x(t). \quad (11)$$

Equation (10) can be rewritten as

$$R_{xx}^{-1}(0)R_{xx}(\tau)R_{xx}^{-1}(0)R_{xx}(\tau)w_i = \lambda_i^2 w_i \quad (12)$$

which can be further simplified into

$$R_{xx}(\tau)w_i = \lambda_i R_{xx}(0)w_i, i = 1, \cdots, m. \quad (13)$$

Joint the sub problems into one, we have

$$R_{xx}(\tau)W = \Lambda R_{xx}(0)W \quad (14)$$

where the demixing matrix W contains the vectors $(w_1, \cdots, w_m)$, $\Lambda = diag(\lambda_1, \cdots, \lambda_m)$ is a diagonal matrix.

So the CCA problem for $x(t)$ and $x(t-\tau)$ can be simplified into a generalized Eigen value decomposition problem, and can be used for BSS [9].

For MCCA, we choose the vector $x_1, x_2, \ldots, x_K$ as $x(t-\tau_1), x(t-\tau_2), \ldots, x(t-\tau_K)$. Compare the (7) and (8). The MCCA problem for $x(t-\tau_1), x(t-\tau_2), \ldots, x(t-\tau_K)$ can be simplified into a joint diagonalization problem. We can find a joint diagonalizer W of $\{R_{xx}[\tau_i - \tau_j], i, j = 1, \cdots, K, i \neq j\}$ using the joint approximate diagonalization method in [14],[15], which satisfies

$$W R_{xx}[\tau_i - \tau_j]W^T = D_{i,j} \quad (15)$$

where $\{D_{i,j}\}$ is a set of diagonal matrices. The separate signals are computed as $y(t) = W^T x(t)$. We call this BSS algorithm MCCA.

III. KERNELIZING MCCA METHOD FOR NONLINEAR BSS

We generalize the MCCA linear BSS algorithm described above to the invertible nonlinear setting depicted in Fig. 1 [4]. This underlying mixing process
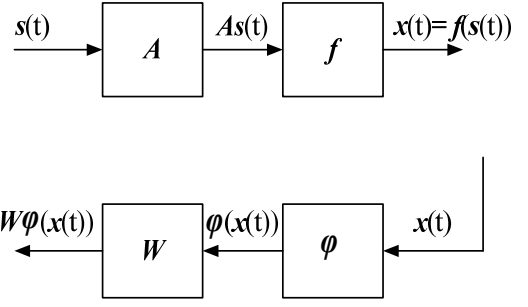

Figure 1. Invertible nonlinear model.

$x(t) = f(s(t))$ is more general than the post-nonlinear mixture model proposed in [16] because the nonlinearity f is not restricted to be component wise.

By projecting the input data $x(t)$ to some high-dimensional feature space via a nonlinear mapping φ and then find the weight matrix W that maximizes the correlation between separate signals. Obviously, if $\varphi = f^{-1}$ and $W = A^{-1}$, then the sources are perfectly recovered and $y(t) = s(t)$ [3].

The basic idea of kernel-based method allows constructing very powerful nonlinear variants of existing linear scalar product based algorithms by mapping the data $x(t)$ ($t = 1, L\ T$) implicitly into some kernel feature space F through some mapping $\varphi : \Re^m \to F$. Performing a simple linear algorithm in F, then corresponds to a nonlinear algorithm in input space. All can be done efficiently and never directly but implicitly in F by using the famous kernel trick $K_{ij} = K(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$. Kernels offer a great deal of flexibility, as they can be generated from other kernels. In the kernel, the data appears only through entries in the Gram matrix. Therefore, this approach gives a further advantage as the number of tunable parameters and updating time does not depend on the number of attributes being used.

A. Constructing reduced kernel feature space

However, a straight forward application of the kernel trick to BSS has so far failed as, after kernelizing, the BSS algorithm has to be applied to a T dimensional problem which is numerically neither stable nor tractable. We need to specify how to handle its possibly high dimensionality. As in Fig. 2, two methods that obtain an orthogonal basis in feature space with reduced dimension are described [3].

A method to obtain the low dimensional subspace is clustering [4]. Denoting the mapped points by $\varphi_x := [\varphi(x(1)), L, \varphi(x(T))]$ and $\varphi_v := [\varphi(v_1), L, \varphi(v_d)]$. We assume that the columns of φ_v constitute a basis of the column space of φ_x, formally expressed as

$$span(\varphi_v) = span(\varphi_x) \quad and \quad rank(\varphi_v) = d \quad (16)$$

And φ_v being a basis implies that the matrix $\varphi_v^T \varphi_v$ has full rank and its inverse exists, an orthonormal basis can be defined as $\Xi := \varphi_v (\varphi_v^T \varphi_v)^{-1/2}$. This basis enables us to

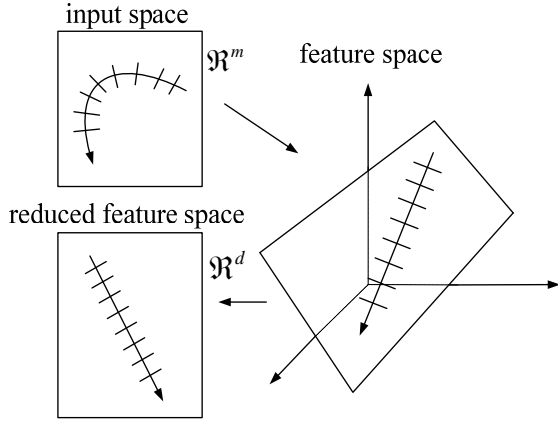parameterize the signals in feature space as real valued *d*-dimensional signals

Figure 2. Mapping the input data to reduced feature space.

$$\psi_x(t) := \Xi^T \varphi(x(t)) = (\varphi_v^T \varphi_v)^{-1/2} \varphi_v^T \varphi(x(t)) \quad (17)$$

By employing the kernel trick

$$(\varphi_v^T \varphi_v)_{ij} = \varphi(v_i)^T \varphi(v_j) = K(v_i, v_j)$$
$$(\varphi_v^T \varphi_x)_{it} = \varphi(v_i)^T \varphi(x(t)) = K(v_i, x(t)) \quad (18)$$

where $i, j = 1, \mathrm{L}, d$, $t = 1, \mathrm{L}, T$. Substituting (18) into (17) changes it to

$$\psi_x(t) := \begin{bmatrix} K(v_1, v_1) & \mathrm{L} & K(v_1, v_d) \\ \mathrm{M} & \mathrm{O} & \mathrm{M} \\ K(v_d, v_1) & \mathrm{L} & K(v_d, v_d) \end{bmatrix}^{-\frac{1}{2}}$$
$$\cdot \begin{bmatrix} K(v_1, x(t)) \\ \mathrm{M} \\ K(v_d, x(t)) \end{bmatrix} \quad (19)$$

Another more direct method to obtain the low dimensional subspace is KPCA (kernel principal component analysis) [4]. For simplicity, we assume that the data is centered in feature space. To perform KPCA we need to find eigenvectors $E = [e_1, \mathrm{L}, e_T]$ and Eigen values $\lambda_1 \geq \mathrm{L} \geq \lambda_T$ of the covariance matrix $\frac{1}{T} \varphi_x^T \varphi_x$. Now, let Λ be the diagonal matrix with the Eigen values along the diagonal and let

$$\left(\frac{1}{T} \varphi_x^T \varphi_x\right) E = E\Lambda \quad (20)$$

which can also be expressed as

$$\left(\frac{1}{T} \varphi_x \varphi_x^T\right)(\varphi_x E) = (\varphi_x E)\Lambda \quad (21)$$

So $\lambda_1, \mathrm{L}, \lambda_T$ are the Eigen values of $\frac{1}{T} \varphi_x \varphi_x^T$ with corresponding Eigen vectors $\varphi_x E$. Normalizing the first *d* eigenvectors yields a *d*-dimensional orthonormal basis

$$\Xi := \varphi_x E_d (T\Lambda)^{-1/2} \quad (22)$$

where using $(T\Lambda)^{-1/2}$ to ensure orthonormality and $E_d := [e_1, \mathrm{L}, e_d]$. So the input data corresponding in the reduced feature space can be described as

$$\psi_x(t) := \Xi^T \varphi(x(t)) = (T\Lambda)^{-1/2} E_d^T \varphi_x^T \varphi(x(t))$$
$$= \frac{1}{\sqrt{T}} \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & & 0 \\ & \mathrm{O} & \\ 0 & & \frac{1}{\sqrt{\lambda_d}} \end{bmatrix} \begin{bmatrix} e_1^T \\ \mathrm{M} \\ e_d^T \end{bmatrix} \cdot \begin{bmatrix} K(x(1), x(t)) \\ \mathrm{M} \\ K(x(T), x(t)) \end{bmatrix} \quad (23)$$

which are calculated conveniently using the kernel trick.

B. Nonlinear BSS

By mapping the signals $x(t)$ from input space onto signals $\psi_x(t)$ in reduced kernel feature space, the nonlinear problem in input space has been transformed to a linear problem in reduced kernel feature space. So we can apply MCCA method to solve this linear BSS problem in reduced kernel feature space.

For the signals in reduced kernel feature space

$$\psi_x(t) := \Xi^T \varphi(x(t)) \in \mathfrak{R}^d \quad (24)$$

We choose the vectors $\psi_x(t - \tau_1), \psi_x(t - \tau_2), \mathrm{K}, \psi_x(t - \tau_K)$ as the components $x_1, x_2, \ldots, x_K$ in MCCA. The MCCA problem for $\psi_x(t - \tau_1), \psi_x(t - \tau_2), \mathrm{K}, \psi_x(t - \tau_K)$ can be simplified into a joint diagonalization problem. We can find a joint diagonalizer W of $\{R_{\psi\psi}[\tau_i - \tau_j], i, j = 1, \cdots, K, i \neq j\}$ using the joint approximate diagonalization method in [14],[15], which satisfies

$$W R_{\psi\psi}[\tau_i - \tau_j] W^T = \Sigma_{i,j} \quad (25)$$

where $\{\Sigma_{i,j}\}$ is a set of diagonal matrices. The extracted d nonlinear components are $y(t) = W^T \psi_x(t)$, among which are the components of interest. Harmeling and Ziehe, *et al.* [4] have defined a selection procedure to find the original sources automatically. By repeating the algorithm with the same parameters (kernel choice, *d*), but instead of sending $x(t)$ into the feature space we start with the *d* dimensional demixed results $y(t)$, map them to the feature space, reduce the dimensionality and demix with MCCA, which yield $y'(t)$. The sought-after components of $y(t)$ are the ones that are matched best by the components $y'(t)$ of the second run of MCCA in feature   space.
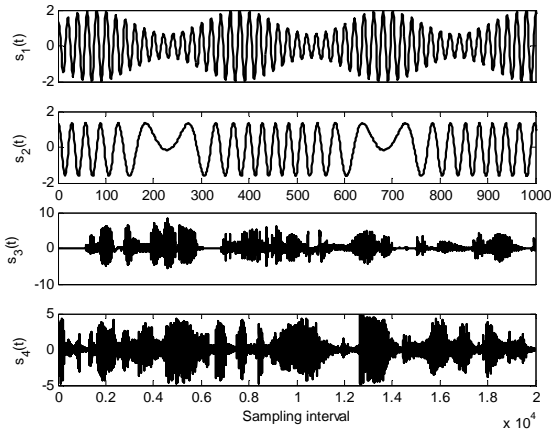
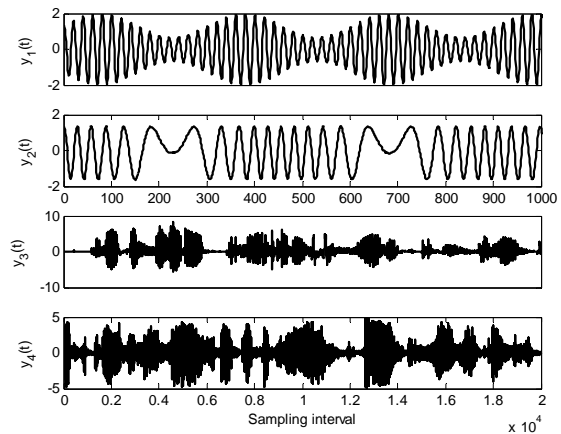Figure 3.   Source signals.

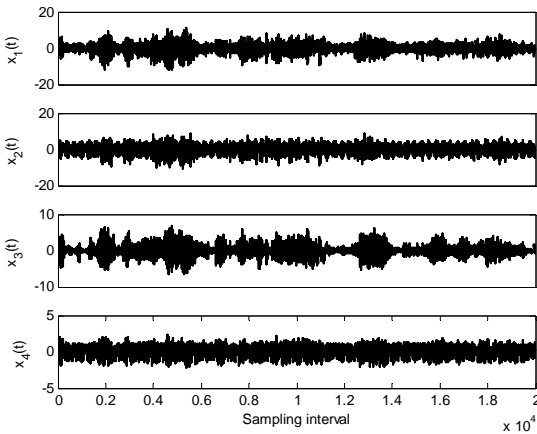Figure 5.   Separate signals by MCCA
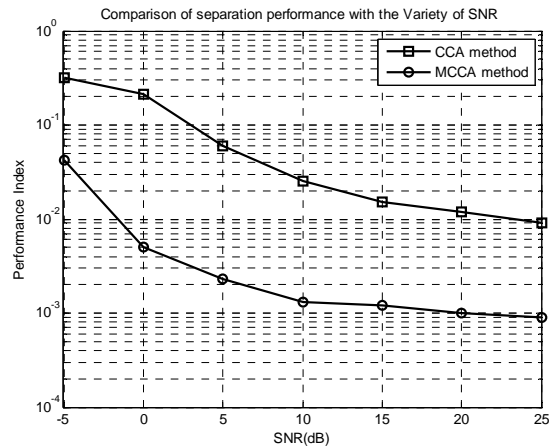

Figure 4.   Linear mixing signals

Figure 6.   Comparison of performance of two algorithms as a function of the signal to noise ratio(SNR)

IV. COMPUTER SIMULATIONS

All simulations were conducted in the MATLAB environment running on an ordinary PC with double 2.2GHZ CPU and 1GB size of memory. To make the demonstration more convincing, the blind separation algorithm was tested on mixtures of realistic speech signals (super-Gaussian signals) and modulated signals (sub-Gaussian signals).

A. Simulations for linear BSS by MCCA

In order to measure the performance of algorithms, we use the performance index (PI) as in [17] defined by

$$PI = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left\{ \left(\sum_{k=1}^{n} \frac{|g_{ik}|}{\max_j |g_{ij}|} - 1 \right) \right.$$

$$\left. + \left(\sum_{k=1}^{n} \frac{|g_{ki}|}{\max_j |g_{ji}|} - 1 \right) \right\} \quad (26)$$

where g_{ij} is the (i, j)-element of the global system matrix $\mathbf{G} = \mathbf{HW}$ and $\max_j g_{ij}$ represents the maximum value among the elements in the ith row vector of $\mathbf{G}$, $\max_j g_{ji}$ does the maximum value among the elements in the ith column vector of $\mathbf{G}$. When the perfect separation is achieved, the performance index is zero. In practice, the values of performance index around 10^{-2} gives quite a good performance.

As shown in Fig.3, the source signals are two speech source signals and two modulated signals (a AM signal and a FM signal), which are mixed by random matrix $\mathbf{H}$, the sampling points are 20000. The mixing signals are shown as in Fig.4. And the separation results by MCCA BSS algorithm are shown as in Fig.5, the cost time for separation is 0.1563s. It can effectively complete the separate processing and have a good performance of real time.

A lot of experiments have been conducted by the changing the impact of additive white Gaussian noise. Results as in Fig.4 show typical performance of the MCCA BSS algorithm is more accuracy and stabilization than CCA BSS algorithm. When the signal noise ratio (SNR) above zero, the performance index of MCCA BSS algorithm is less than 10^{-2} , and gives quite a good performance.
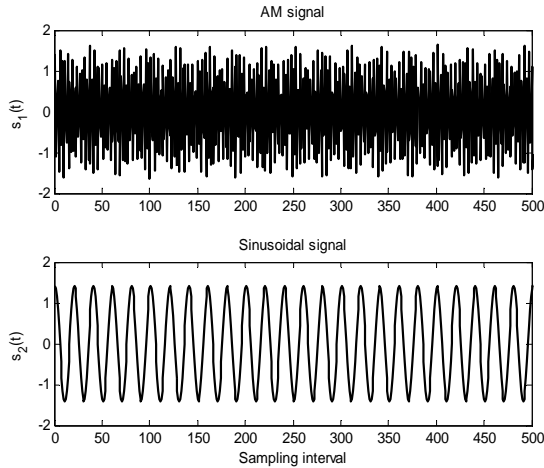

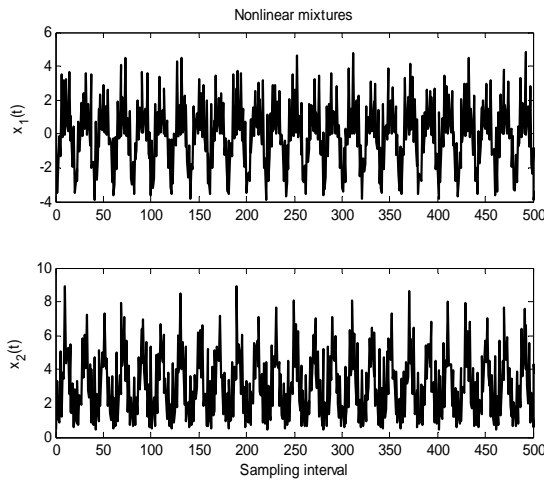Figure 7.   Source signals(one AM signal and one sinusoidal signal).

Figure 8.   Nonlinear mixture

B. Simulations for nonlinear BSS by kernel MCCA

In order to quantificationally evaluate the separation effect, here we choose the resemble coefficient ξ_{ij} of separation signal and source signal to test the resemble degree of separating signal $\boldsymbol{y}(t)$ and expecting signal $\boldsymbol{s}(t)$. The resemble coefficient ξ_{ij} can be described as [18]
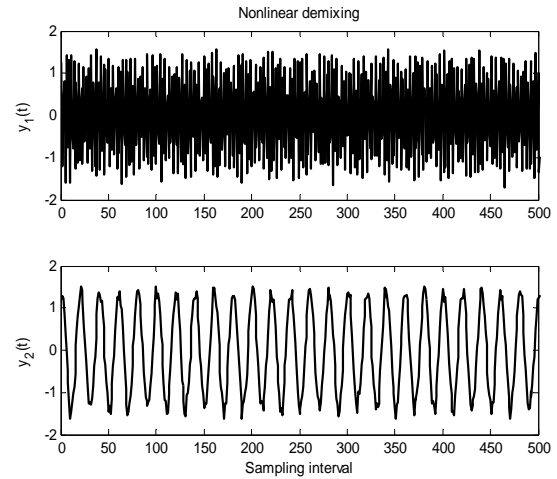
Figure 9.   Separate signals by kernel MCCA

$$\xi_{ij} = \xi(\boldsymbol{y}_i(t), \boldsymbol{s}_j(t)) = \frac{\left|\sum_{t=1}^{T} \boldsymbol{y}_i(t)\boldsymbol{s}_j(t)\right|}{\sqrt{\sum_{t=1}^{T} \boldsymbol{y}_i^2(t)\sum_{t=1}^{T} \boldsymbol{s}_j^2(t)}} \quad (27)$$

where $0 \le \xi_{ij} \le 1$. When $\boldsymbol{y}_i(t) = c \cdot \boldsymbol{s}_j(t)$ (c is a constant), then $\xi_{ij} = 1$. It allows separate results exit difference on amplitude. When $\boldsymbol{y}_i(t)$ and $\boldsymbol{s}_j(t)$ are mutual independent, $\xi_{ij} = 0$. When $\xi_{ij} < 0$, it means separate signals and source signals are reverse order.

First simulation is to sub-Gaussian signals. As shown in Fig. 7, the source signals are one sinusoidal signal and one FM signal:

$$\boldsymbol{s}_1(t) = (6 + \sin(2\pi \times 66t / f_s)) \cdot (\cos(2\pi \times 1222t / f_s))$$
$$\boldsymbol{s}_2(t) = (\cos(2\pi \times 100t / f_s)) \quad (28)$$

where $t = 1L\ 2000, f_s = 2000$. These source signals are nonlinearly mixed by

$$\boldsymbol{x}_1(t) = e^{s_1(t)} - e^{s_2(t)}$$
$$\boldsymbol{x}_2(t) = e^{-s_1(t)} + e^{-s_2(t)} \quad (29)$$

Employing a Gaussian RBF kernel,

$$K\left(x_i, x_j\right) = \exp(-\left\|x_i - x_j\right\|^2 \big/ 2\sigma^2) \quad (30)$$

and set $d = 20$, using Clustering method to construct the reduced feature space. The mixing signals are shown as in Fig. 8. And the separation results by this proposed kernel MCCA nonlinear BSS algorithm are shown as in

Fig. 9. The resemble coefficients of separate signals and source signals are [0.9913 0.0142; 0.0117 0.9935]. It can effectively complete the separate processing and have a good performance.

The second simulation is to super-Gaussian signals. As shown in Fig. 10, the source signals are two speech source signals (a man and a woman) $s_1(t)$=dr1_fetbo_sa2 and $s_2(t)$=dr2_mcewo_sx182, were randomly taken from the TIMIT database [3].These source signals are non-linearly mixed by

$$x_1(t) = -(s_2(t)+1)\cos(\pi s_1(t))$$
$$x_2(t) = 1.5(s_2(t)+1)\sin(\pi s_1(t)) \tag{31}$$

Employing a Gaussian RBF kernel, and set $d = 20$, using KPCA method to construct the reduced feature space. The mixing signals are shown as in Fig. 11. And the separation results by this proposed kernel MCCA nonlinear BSS algorithm are shown as in Fig. 12. The resemble coefficient matrix of separate signals and source
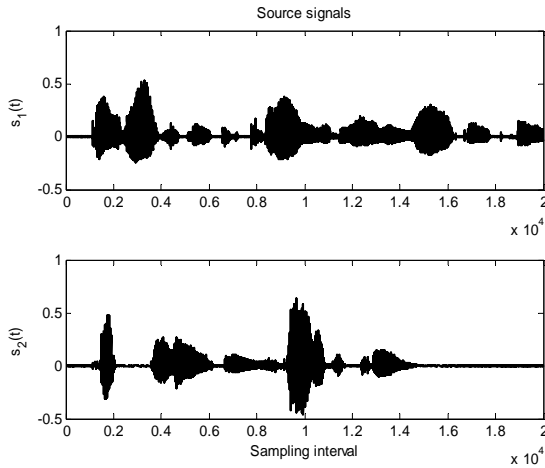

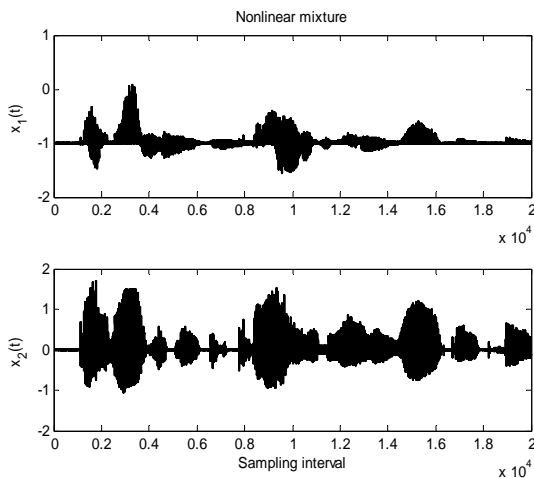Figure 10. Source signals(two speech signals).

Figure 11. Nonlinear mixture

signals is [0.9736 0.0342; 0.0136 0.9927]. It can complete the separate processing successfully.
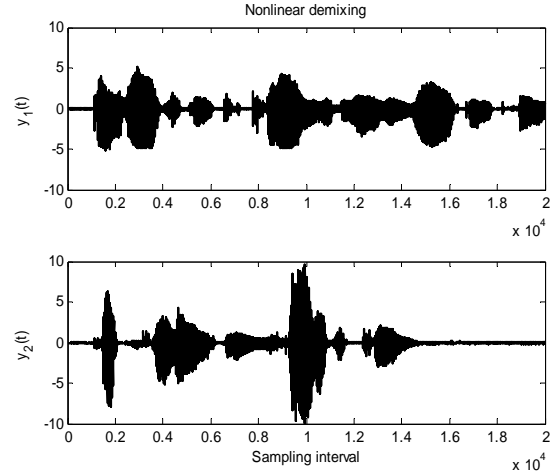

Figure 12. Separate signals by kernel MCCA

V. Conlusion

This paper proposed an algorithm for nonlinear BSS. It extends the MCCA linear BSS algorithm to the non-linear domain using the "kernel trick". In the first step of the method, the input data are mapped to high-dimensional feature space by kernel trick and the effective reduced feature space is extracted by Clustering or KPCA method, translating the nonlinear problem in the input space to a linear problem in reduced feature space. In the second step, by using MCCA linear BSS algorithm, the linear BSS problem can be efficient solved. MCCA linear BSS relies only on second-order statistics of the observation signals, the computation burden is relatively light, can complete the blind separation of super-Gaussian and sub-Gaussian signals at the same time, which make it have a broad application foreground. The experimental results show that the kernel MCCA can complete the nonlinear separate processing successfully.

References

[1] S. Amari, A. Cichocki, H. -M. Park and S. -Y. Lee, "Blind source separation and independent component analysis: a review," *Neural Information Processing-Letters and Reviews*, vol. 6, No. 1, pp. 1-57, Jan. 2005.
[2] C. Jutten, J. Karhunen, "Advances in Nonlinear Blind Source separation," *Proc. of 4th Int. Symposium on ICA and BSS (ICA2003)*, pp. 245-256, Nara, Japan, 2003.

[3] D. Martinez, A. Bray, "Nonlinear blind source separation using kernels," *IEEE Trans. Neural Networks*, vol.14, No. 1, pp. 228–235, Jan. 2003.

[4] S. Harmeling, A. Ziehe, M. Kawanabe, K. -R. Muller, "Kernel-based nonlinear blind source separation," *Neural Computation*, vol.15, No. 1, pp. 1089-1124, 2003.

[5] H.Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, pp. 312–377, 1936.

[6] D. R. Hardoon, S. Szedmak and J. Shawe-Taylor, "Canonical correlation analysis: an overview with application to learning methods," *Neural Computation*, vol. 16, pp. 2639–2664, 2004.

[7] S. Y. Huang, M. H. Lee and C. K. Hsiao, "Nonlinear measures of association with kernel canonical correlation analysis and applications," *Journal of Statistical Planning and Inference*, vol. 139, pp. 2162–2174, 2009.

[8] J. Via, I. Santamaria, and J. Perez, "Deterministic CCA-Based Algorithms for Blind Equalization of FIR-MIMO Channels," *IEEE Transactions on Signal Processing*, vol. 55, pp. 3867-3878, 2007.

[9] W. Liu, D. P. Mandic, and A. Cichocki, "Analysis and Online Realization of the CCA Approach for Blind Source Separation," *IEEE Transactions on Neural Networks*, vol. 18, pp. 1505-1510, 2007.

[10] Y. -O Li, T. Adali, W. Wang, and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 57, pp. 3918-3929, 2009.

[11] M. Borga and H. Knutsson, "A canonical correlation approach to blind source separation," Dept. Biomed. Eng., Linkoping Univ., Linkoping, Sweden, Tech. Rep. LiU-IMT-EX-0062, Jun. 2001.

[12] J. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, pp. 433-451, 1971.

[13] H. G. Yu, G. M. Huang, J. Gao, "Multiset Canonical Correlation Analysis Using for Blind Source Separation," *The 3rd International Conference on Networks Security, Wireless Communications and Trusted Computing*, in press.

[14] A. Ziehe, P. Laskov, G. Nolte, and K.-R. Múller, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation," *J. Mach. Learn. Res.*, vol. 5, pp. 777–800, 2004.

[15] S. Degerine and E. Kane, "A Comparative Study of Approximate Joint Diagonalization Algorithms for Blind Source Separation in Presence of Additive Noise," *IEEE Transactions on Signal Processing*, vol. 55 pp. 3022-3031, 2007.

[16] A. Talab, C. Jutten, "Source separation in post-nonlinear mixtures," *IEEE Transactions on Neural Networks*, vol. 47, no. 10, pp. 2807-2820, Oct. 1999.

[17] A. Cichocki. and S. Amari, Adaptive blind signal and image processing. New York, USA: Wiley, 2002, pp. 161-162.

[18] G. M. Huang, L. X. Yang, Z. Y. He, "Blind Source Separation Based on Generalized Variance," *ISNN2006, LNCS*, Vol.3971, pp. 496-501, 2006.

Hua-Gang Yu was born in 1984. He received the B.S. and M.E. degrees in communication engineering from Naval University of Engineering, Wuhan, China, in 2006 and 2008, respectively. He is currently working towards the Ph.D. degree in communication and information engineering at Naval University of Engineering. His current research interests in the area of blind signal processing, passive detection and wireless communications.

Gao-Ming Huang was born in 1972. He received the B.S. and M.E. degree in electronic warfare from Naval Electronic College of Engineering, Nanjing, China, in 1995 and 1998, respectively, and Ph.D. degrees in signal processing from dongnan university, Nanjing, China, in 2006. In 2001, he joined the Department of Communication Engineering of Naval Electronic College of Engineering as an Assistant. He became an Associate Professor in 2005 and a Professor in 2009. He is currently a professor with the Department of command automation, Naval University of Engineering, Wuhan, China. His research interests are electronic warfare, blind signal processing and passive detection.

Jun Gao was born in 1957. He received the B.S. degree in communication engineering from Naval Electronic College of Engineering, Taiyuan, China, in 1982, and the M.E. and Ph.D. degrees in electrical engineering from Beijing Institute of Technology, Beijing, China, in 1986 and 1989, respectively. In 1982, he joined the Department of Communication Engineering of Naval Electronic College of Engineering as an Assistant. He became an Associate Professor in 1992 and a Professor in 1996. He is currently a professor with the Department of Communication Engineering, Naval University of Engineering, Wuhan, China. His research interests are signal processing and digital communications.