Letters

Analysis and Online Realization of the CCA Approach for Blind Source Separation

Wei Liu, Danilo P. Mandic, and Andrzej Cichocki

***Abstract*—A critical analysis of the canonical correlation analysis (CCA) approach in blind source separation (BSS) is provided. It is proved that by maximizing the autocorrelation functions of the recovered signals we can separate the source signals successfully. It is further shown that the CCA approach represents the same class of generalized eigenvalue decomposition (GEVD) problems as the matrix pencil method. Finally, online realizations of the CCA approach are discussed with a linear-predictor-based algorithm studied as an example.**

***Index Terms*—Blind source separation (BSS), canonical correlation analysis (CCA), linear predictor, matrix pencil, second-order statistics (SOS).**

I. Introduction

Blind source separation (BSS) has been studied extensively and has become one of the most important and established research topics in the signal processing area [1]. There are mainly two classes of solutions to the BSS problem: those based on higher order statistics (HOS) and those based on second-order statistics (SOS). Within the SOS-based methods, for stationary sources, we usually assume the following: 1) sources are not correlated with one other and 2) every source has a different temporal structure or normalized spectrum. To recover such sources, the basic idea is to find a matrix which diagonalizes the covariance matrices of the mixed signals at different time lags. This diagonalization is normally realized in two steps. The first step is the prewhitening of the data, by which the general mixing matrix is reduced to an orthogonal matrix; in the second step, we find the inverse of this orthogonal matrix by diagonalizing an appropriately chosen covariance matrix at a nonzero time lag [2], or by jointly diagonalizing a number of covariance [3] or partial autocorrelation matrices [4].

Instead of the classical two-step solution to the SOS approach, we can also solve the BSS problem in one single step by the canonical correlation analysis (CCA) approach [5]–[7]. With CCA, the objective is to find a transformation matrix which is applied to the mixtures and maximizes the autocorrelation of each of the recovered signals (the outputs of the transformation matrix). In theory, 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, as proved in [8]. Although some examples have been given to show the validity of this CCA approach [8], there is no rigorous proof showing that by maximizing the autocorrelation function, we can indeed recover the source signals successfully. We hereby address this issue and provide a critical analysis of this approach. Our analysis shows that not only there is a close relationship between the CCA approach and the linear predictor approach [9]–[11], but also that the maximization of the autocorrelation value is equivalent to finding the generalized eigenvectors within the matrix pencil approach [12]–[14]. Finally, online realizations of the approach are discussed and a blind source extraction (BSE) algorithm [15] based on the linear predictor structure with fixed coefficients is studied as an example [11].

Manuscript received June 13, 2006; revised November 25, 2006; accepted December 28, 2006. The work of W. Liu and D. P. Mandic was supported by the EPSRC under the Grant GR/S71149/01.

W. Liu is with the Communications Research Group, Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield S1 3JD, U.K. (e-mail: w.liu@sheffield.ac.uk).

D. P. Mandic is with the Communications and Signal Processing Research Group, Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2BT, U.K. (e-mail: d.mandic@imperial.ac.uk).

A. Cichocki is with the RIKEN Brain Science Institute, Wako-shi, Saitama 351-0198, Japan and also with the Systems Research Institute, Polish Academy of Sciences, Warsaw 01-447, Poland (e-mail: a.cichocki@riken.jp).

Digital Object Identifier 10.1109/TNN.2007.894017

II. Overview and Analysis of the CCA Approach

A. Overview of the CCA Approach

In CCA [16], two sets of variables with a joint distribution are considered. The correlation between a linear combination of the variables of the first set and a linear combination of the variables of the second set is first analyzed. The two linear combinations are found by maximizing this correlation. Then, a second linear combination in each set is found such that the second set of linear combinations is uncorrelated with the first set of linear combinations and the correlation between the second set of linear combinations is also maximized. This procedure continues until there is no more such linear combinations left.

Suppose $\mathbf{z}$ is a zero-mean random vector with q components. We partition $\mathbf{z}$ into two subvectors $\mathbf{z}_1$ (with q_1 components) and $\mathbf{z}_2$ (with q_2 components), given by

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix}. \tag{1}$$

For convenience, we assume $q_1 \leq q_2$. The covariance matrix $\mathbf{\Sigma} = \mathbf{z}\mathbf{z}^T$, which is assumed to be positive definite, is partitioned accordingly into four blocks as

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} \tag{2}$$

where $\mathbf{\Sigma}_{11} = \mathbf{z}_1\mathbf{z}_1^T$, $\mathbf{\Sigma}_{12} = \mathbf{z}_1\mathbf{z}_2^T$, $\mathbf{\Sigma}_{21} = \mathbf{z}_2\mathbf{z}_1^T$, and $\mathbf{\Sigma}_{22} = \mathbf{z}_2\mathbf{z}_2^T$.

We can now construct linear combinations of the variables in each of the subvectors

$$\begin{aligned} a_0 &= \boldsymbol{\alpha}_0^T \mathbf{z}_1 \\ b_0 &= \boldsymbol{\beta}_0^T \mathbf{z}_2 \end{aligned} \tag{3}$$

where $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ are vectors containing the combination coefficients.

The problem of finding the two vectors $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ that maximize the correlation between a_0 and b_0 can be formulated as

$$\max_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0} J_0(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) \tag{4}$$

with

$$J_0(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) = \frac{\boldsymbol{\alpha}_0^T \mathbf{\Sigma}_{12} \boldsymbol{\beta}_0}{\sqrt{(\boldsymbol{\alpha}_0^T \mathbf{\Sigma}_{11} \boldsymbol{\alpha}_0)(\boldsymbol{\beta}_0^T \mathbf{\Sigma}_{22} \boldsymbol{\beta}_0)}}. \tag{5}$$

After finding the first pair of optimal vectors $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$, we can proceed to find the second pair $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ which maximizes the correlation and at the same time ensures that the new pair of combinations

$\{a_1, b_1\}$ is uncorrelated with the first set $\{a_0, b_0\}$. This process is repeated until we find all the $\min(q_1, q_2) = q_1$ pairs of optimal vectors $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i, i = 0, 1, \ldots, q_1 - 1$. It has been shown that the vectors $\boldsymbol{\alpha}_i$ can be obtained by solving the following generalized eigenvalue problem

$$\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}_i = \lambda_i^2 \boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}_i. \tag{6}$$

The vectors $\boldsymbol{\beta}_i$ can be found in the same way by exchanging the subscripts of the matrices in (6).

We now apply the CCA approach to blind separation of the instantaneous mixed signals $\mathbf{x}[n]$, which are modeled as weighted sums of individual sources $\mathbf{s}[n]$, given by

$$\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n] \tag{7}$$

with

$$\begin{aligned} \mathbf{s}[n] &= [s_0[n] s_1[n] \cdots s_{L-1}[n]]^{\mathrm{T}} \\ \mathbf{x}[n] &= [x_0[n] x_1[n] \cdots x_{M-1}[n]]^{\mathrm{T}} \\ [\mathbf{A}]_{m,l} &= a_{m,l}, m = 0, \ldots, M-1, \qquad l = 0, \ldots, L-1 \end{aligned} \tag{8}$$

where n is the time index, L is the number of sources, M is the number of mixtures, and $\mathbf{A}$ is the mixing matrix. We assume the sources are spatially uncorrelated with the correlation matrix expressed as

$$\begin{aligned} \mathbf{R}_{ss}[0] &= E\{\mathbf{s}[n]\mathbf{s}^T[n]\} \\ &= \mathrm{diag}\{\rho_0[0], \rho_1[0], \ldots, \rho_{L-1}[0]\} \end{aligned} \tag{9}$$

with $\rho_m[0] = E\{s_m[n] \cdot s_m[n]\}, m = 0, 1, \ldots, L-1$, where $E\{\cdot\}$ denotes the statistical expectation operator. For nonzero correlation lags, we have

$$\begin{aligned} \mathbf{R}_{ss}[\Delta n] &= E\{\mathbf{s}[n]\mathbf{s}^T[n - \Delta n]\} \\ &= \mathrm{diag}\{\rho_0[\Delta n], \rho_1[\Delta n], \ldots, \rho_{L-1}[\Delta n]\} \end{aligned} \tag{10}$$

with $\rho_l[\Delta n] \neq 0$ for some nonzero delays Δn.

First, we choose the vector $\mathbf{x}[n]$ as $\mathbf{z}_1$ in CCA and $\mathbf{x}[n - \Delta n]$ as $\mathbf{z}_2$. Then, the eigenvalue problem in (6) becomes

$$\mathbf{R}_{xx}[\Delta n]\mathbf{R}_{xx}^{-1}[0]\mathbf{R}_{xx}[-\Delta n]\boldsymbol{\alpha}_i = \lambda_i^2 \mathbf{R}_{xx}[0]\boldsymbol{\alpha}_i \tag{11}$$

where $\mathbf{R}_{xx}[\Delta n]$ and $\mathbf{R}_{xx}[0]$ are the correlation matrices of the mixed signals.

From (7), we have

$$\mathbf{R}_{xx}[\Delta n] = \mathbf{A}\mathbf{R}_{ss}[\Delta n]\mathbf{A}^T. \tag{12}$$

As $\mathbf{R}_{ss}[\Delta n]$ is diagonal, we have

$$\mathbf{R}_{xx}[\Delta n] = \mathbf{R}_{xx}^T[-\Delta n] = \mathbf{R}_{xx}^T[\Delta n]. \tag{13}$$

Then, (11) becomes

$$\mathbf{R}_{xx}[\Delta n]\mathbf{R}_{xx}^{-1}[0]\mathbf{R}_{xx}[\Delta n]\boldsymbol{\alpha}_i = \lambda_i^2 \mathbf{R}_{xx}[0]\boldsymbol{\alpha}_i. \tag{14}$$

In the context of BSS, the two vectors $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ are the same, and we denote them by

$$\mathbf{w}_i = \boldsymbol{\alpha}_i = \boldsymbol{\beta}_i \tag{15}$$

which is the ith demixing vector applied to the mixed signals. The corresponding output, that is the extracted ith signal, becomes

$$y_i[n] = \mathbf{w}_i^T \mathbf{x}[n]. \tag{16}$$

B. Analysis of the CCA Approach for BSS

Equation (14) can be rewritten as

$$\mathbf{R}_{xx}^{-1}[0]\mathbf{R}_{xx}[\Delta n]\mathbf{R}_{xx}^{-1}[0]\mathbf{R}_{xx}[\Delta n]\boldsymbol{\alpha}_i = \lambda_i^2 \boldsymbol{\alpha}_i \tag{17}$$

which can be further simplified into

$$\mathbf{R}_{xx}^{-1}[0]\mathbf{R}_{xx}[\Delta n]\boldsymbol{\alpha}_i = \lambda_i \boldsymbol{\alpha}_i. \tag{18}$$

Multiplying both sides with $\mathbf{R}_{xx}[0]$ and replacing $\boldsymbol{\alpha}_i$ by $\mathbf{w}_i$, we have

$$\mathbf{R}_{xx}[\Delta n]\mathbf{w}_i = \lambda_i \mathbf{R}_{xx}[0]\mathbf{w}_i \tag{19}$$

which represents the generalized eigenvalue decomposition (GEVD) problem introduced also by the matrix pencil method [12]–[14]. From this point of view, the CCA approach can be justified indirectly by the validity of the matrix pencil approach. However, to provide further insight into the CCA approach, a direct proof is necessary.

1) Proof of the CCA Approach: Existence of the Solution: Note that in the context of BSS, the maximization problem in (4) becomes

$$\max_{\mathbf{w}_0} J_0(\mathbf{w}_0) \tag{20}$$

with

$$J_0(\mathbf{w}_0) = \frac{\mathbf{w}_0^T \mathbf{R}_{xx}[\Delta n]\mathbf{w}_0}{\mathbf{w}_0^T \mathbf{R}_{xx}[0]\mathbf{w}_0} = \frac{\mathbf{w}_0^T \mathbf{A}\mathbf{R}_{ss}[\Delta n]\mathbf{A}^T \mathbf{w}_0}{\mathbf{w}_0^T \mathbf{A}\mathbf{R}_{ss}[0]\mathbf{A}^T \mathbf{w}_0}. \tag{21}$$

We need to prove that maximization of $J_0(\mathbf{w}_0)$ with respect to $\mathbf{w}_0$ will lead to a successful extraction of one of the source signals.

To achieve this, let $\mathbf{g}_0 = \mathbf{A}^T \mathbf{w}_0$ denote the first global mixing vector. Then, $J_0(\mathbf{w}_0)$ in (21) changes into

$$J_0(\mathbf{w}_0) = \frac{\mathbf{g}_0^T \mathbf{R}_{ss}[\Delta n]\mathbf{g}_0}{\mathbf{g}_0^T \mathbf{R}_{ss}[0]\mathbf{g}_0}. \tag{22}$$

Without the loss of generality, we will assume $\mathbf{R}_{ss}[0] = \mathbf{I}$, as the differences in the diagonal elements of $\mathbf{R}_{ss}[0]$ can always be absorbed into the mixing matrix $\mathbf{A}$. This way, the diagonal elements of $\mathbf{R}_{ss}[\Delta n]$ become the normalized autocorrelation values of each source signal and they are assumed to be different from each other. Now, we have

$$J_0(\mathbf{w}_0) = \hat{\mathbf{g}}_0^T \mathbf{R}_{ss}[\Delta n]\hat{\mathbf{g}}_0 \tag{23}$$

where $\hat{\mathbf{g}}_0 = (\mathbf{g}_0)/(\sqrt{\mathbf{g}_0^T \mathbf{g}_0})$, which has a property $\hat{\mathbf{g}}_0^T \hat{\mathbf{g}}_0 = 1$.

Next, consider the optimization problem formulated as

$$\max_{\hat{\mathbf{g}}_0} \hat{\mathbf{g}}_0^T \mathbf{R}_{ss}[\Delta n]\hat{\mathbf{g}}_0 \quad \text{subject to} \quad \hat{\mathbf{g}}_0^T \hat{\mathbf{g}}_0 = 1. \tag{24}$$

The solution to this problem is a vector $\hat{\mathbf{g}}_{0,\mathrm{opt}}$ with only one nonzero element, which is strictly equal to unity at the position corresponding to the largest diagonal element of the matrix $\mathbf{R}_{ss}[\Delta n]$ [1]. The corresponding global mixing vector $\mathbf{g}_{0,\mathrm{opt}}$ will be the same as $\hat{\mathbf{g}}_{0,\mathrm{opt}}$ except that the nonzero element in $\mathbf{g}_{0,\mathrm{opt}}$ is an arbitrary constant c. In this case, the corresponding output $y_0[n]$ will be a scaled version of one of the source signals.

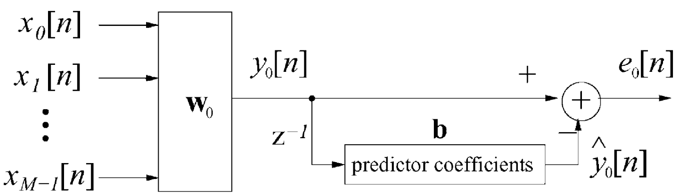

Fig. 1. Linear predictor structure for BSE.

However, since we are maximizing $J(\mathbf{w}_0)$ with respect to $\mathbf{w}_0$, instead of $\hat{\mathbf{g}}_0$, we need to prove that there exists a $\mathbf{w}_{0,\mathrm{opt}}$ which results in $\hat{\mathbf{g}}_{0,\mathrm{opt}}$.

From $\mathbf{g}_0 = \mathbf{A}^T\mathbf{w}_0$, when $\mathbf{A}$ is of full rank and the number of mixtures M is larger or equal to the number of sources L, $\mathbf{w}_{0,\mathrm{opt}}$ can be obtained using the pseudoinverse of $\mathbf{A}^T$ as

$$\mathbf{w}_{0,\mathrm{opt}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{g}_{0,\mathrm{opt}}. \tag{25}$$

For the case $M < L$, in general, we cannot find a $\mathbf{w}_{0,\mathrm{opt}}$ which satisfies the equation $\mathbf{g}_0 = \mathbf{A}^T\mathbf{w}_0$, except for some special forms of $\mathbf{A}$.

Since the possible maximum value of $J_0(\mathbf{w}_0)$ is reached only when $\mathbf{g}_0 = \mathbf{g}_{0,\mathrm{opt}}$, as long as there exists such a $\mathbf{w}_0 = \mathbf{w}_{0,\mathrm{opt}}$ so that $\mathbf{g}_0 = \mathbf{g}_{0,\mathrm{opt}}$, we can state that when we maximize $J_0(\mathbf{w}_0)$ with respect to $\mathbf{w}_0$, this will result in a successful extraction of the source signal with the maximum normalized autocorrelation value.

After extracting the first source signal, we may use a deflation approach to remove it from the mixtures, and then, subsequently perform the next extraction [1]. This procedure is repeated until the last source signal is recovered.

III. Adaptive Realization of the CCA Approach

The correlation $\mathbf{R}_{ss}[\Delta n]$ considered in the cost function $J_0(\mathbf{w}_0)$ is for a fixed single distance Δn and an obvious disadvantage with this choice is that, when the correlation functions of the source signals for that given distance are the same, this method will fail to extract the desired sources [8]. To improve the robustness of this method, we can make a slight modification to $J_0(\mathbf{w}_0)$. Instead of maximizing the correlation between $y_i[n]$ and $y_i[n-\Delta n]$, we maximize the correlation between $y_i[n]$ and a weighted sum of $y_i[n-p]$, $p = 1, 2, \ldots, P$ to give the new cost function [8]

$$\hat{J}_0(\mathbf{w}_0) = \frac{E\{y_0[n]\hat{y}_0[n]\}}{E\{y_0^2[n]\}} \tag{26}$$

where

$$\hat{y}_0[n] = \mathbf{b}^T\mathbf{y}_0[n] \tag{27}$$

with

$$\begin{aligned}\mathbf{b} &= [b_1\ b_2 \cdots b_P]^{\mathrm{T}}\\ \mathbf{y}_0[n] &= [y[n-1]\ y[n-2] \cdots y[n-P]]^{\mathrm{T}}.\end{aligned} \tag{28}$$

Applying the standard gradient-descent method to the cost function $\hat{J}_0(\mathbf{w}_0)$, we can easily derive the corresponding adaptive algorithm.

Now, if we consider $\mathbf{b}$ as the coefficients of a linear predictor, shown in Fig. 1, and $\hat{y}_0[n]$ as the prediction of $y_0[n]$ based on the previous inputs $\mathbf{y}_0[n] = [y[n-1]\ y[n-2] \cdots y[n-P]]^{\mathrm{T}}$, a further modification to this new cost function can be made by considering the linear prediction error

$$e_0[n] = y_0[n] - \hat{y}_0[n] = y_0[n] - \mathbf{b}^T\mathbf{y}_0[n] \tag{29}$$

and the subsequent minimization of the normalized mean-square prediction error (MSPE) [11]

$$\bar{J}_0(\mathbf{w}_0) = E\left\{\frac{e_0^2[n]\}}{E\{y_0^2[n]}\right\}. \tag{30}$$

Note that

$$\begin{aligned}E\{e_0^2[n]\} &= E\{y_0^2[n]\} - 2E\{y_0[n]\mathbf{b}^T\mathbf{y}_0[n]\} + E\{\mathbf{b}^T\mathbf{y}_0[n]\mathbf{y}_0^T[n]\mathbf{b}\}\\ &= \sum_{p=0}^{P} b_p^2 E\left\{y_0^2[n-p]\right\} - 2\sum_{p=1}^{P} b_p E\{y_0[n]y_0[n-p]\}\\ &\quad + \sum_{p,q=1;p\neq q}^{P} b_p b_q E\{y_0[n-p]y_0[n-q]\}\end{aligned} \tag{31}$$

with $b_0 = 1$. As the source signals are stationary, we have

$$\begin{aligned}E\{y_0^2[n-p]\} &= E\left\{y_0^2[n]\right\}\\ E\{y_0[n-p]y_0[n-q]\} &= E\{y_0[n]y_0[n+p-q]\}.\end{aligned} \tag{32}$$

Then, (31) becomes

$$\begin{aligned}E\left\{e_0^2[n]\right\} &= \sum_{p=0}^{P} b_p^2 E\left\{y_0^2[n]\right\} - 2\sum_{p=1}^{P} b_p E\{y_0[n]y_0[n-p]\}\\ &\quad + \sum_{p,q=1;p\neq q}^{P} b_p b_q E\{y_0[n]y_0[n+p-q]\}\end{aligned} \tag{33}$$

and we have

$$\begin{aligned}\bar{J}_0(\mathbf{w}_0) &= \frac{E\{e_0^2[n]\}}{E\{y_0^2[n]\}}\\ &= \sum_{p=0}^{P} b_p^2 - \frac{1}{E\left\{y_0^2[n]\right\}}\\ &\quad \times \left(2\sum_{p=1}^{P} b_p E\{y_0[n]y_0[n-p]\}\right.\\ &\qquad \left. - \sum_{p,q=1;p\neq q}^{P} b_p b_q E\{y_0[n]y_0[n+p-q]\}\right).\end{aligned} \tag{34}$$

Obviously, minimizing $\bar{J}_0(\mathbf{w}_0)$ is equivalent to maximizing the term within the parentheses of (34), which represents the correlation between $y_0[n]$ and a weighted sum of $y_0[n-p]$, $p = 1, 2, \ldots, P$. Therefore, $\bar{J}_0(\mathbf{w}_0)$ is of the same nature as $\hat{J}_0(\mathbf{w}_0)$. From (12) and (16), $E\{e_0^2[n]\}$ can be rewritten as [11]

$$\begin{aligned}E\left\{e_0^2[n]\right\} &= \mathbf{w}_0^T\mathbf{R}_{xx}[0]\mathbf{w}_0 - 2\sum_{p=1}^{P} b_p \mathbf{w}_0^T\mathbf{R}_{xx}[p]\mathbf{w}_0\\ &\quad + \sum_{p,q=1}^{P} b_p b_q \mathbf{w}_0^T\mathbf{R}_{xx}[q-p]\mathbf{w}_0\\ &= \mathbf{g}_0^T\hat{\mathbf{R}}_{ss}\mathbf{g}_0\end{aligned} \tag{35}$$

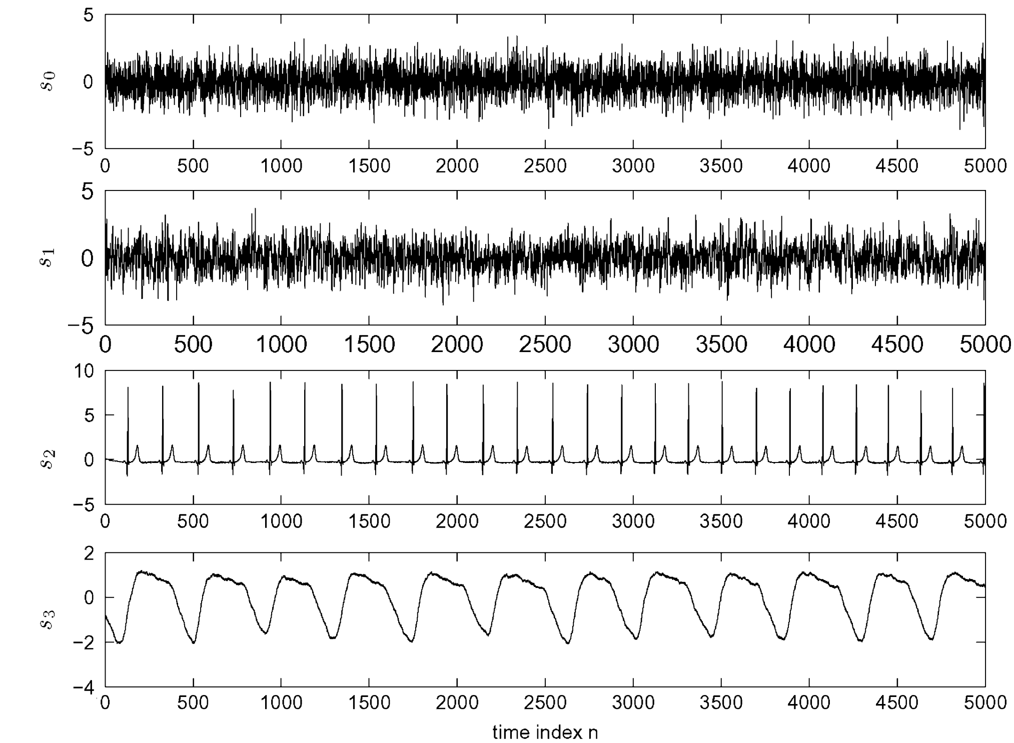

Fig. 2. Four source signals used in simulations.

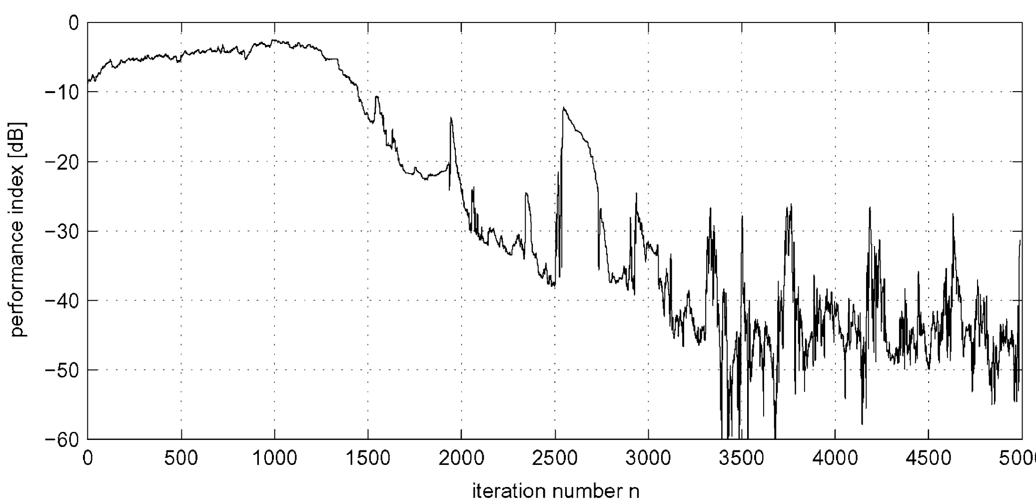

Fig. 3. Learning curve for the performance index using the adaptive extraction method.

with

$$\hat{\mathbf{R}}_{ss} = \mathbf{R}_{ss}[0] - 2\sum_{p=1}^{P} b_p \mathbf{R}_{ss}[p] + \sum_{p,q=1}^{P} b_p b_q \mathbf{R}_{ss}[q-p] \quad (36)$$

where $\hat{\mathbf{R}}_{ss}$ is a diagonal matrix and its diagonal elements represent the MSPEs introduced by the corresponding source signals. Then, the cost function $\bar{J}_0(\mathbf{w}_0)$ becomes

$$\bar{J}_0(\mathbf{w_0}) = \frac{\mathbf{g}_0^T \hat{\mathbf{R}}_{ss} \mathbf{g}_0}{\mathbf{g}_0^T \mathbf{R}_{ss}[0] \mathbf{g}_0}. \quad (37)$$

Similar to (23), we have

$$\bar{J}_0(\mathbf{w}_0) = \hat{\mathbf{g}}_0^T \hat{\mathbf{R}}_{ss} \hat{\mathbf{g}}_0. \quad (38)$$

Now, the diagonal elements of $\hat{\mathbf{R}}_{ss}$ represent the MSPEs of the sources with unit power, i.e., their normalized MSPEs. They are assumed to be different from one another. Using the proof provided in Section II-B, we can state that by minimizing the normalized MSPE (30), the source signal with the smallest normalized MSPE will be extracted successfully.

Applying the standard gradient-descent method to $\bar{J}_0(\mathbf{w}_0)$, we have

$$\nabla_{\mathbf{w_0}} \bar{J}_0 = \frac{2}{E\{y_0^2[n]\}} \left(E\{e_0[n]\hat{\mathbf{x}}[n]\} - \frac{E\left\{e_0^2[n]\right\}}{E\left\{y_0^2[n]\right\}} E\{y_0[n]\mathbf{x}[n]\} \right) \quad (39)$$

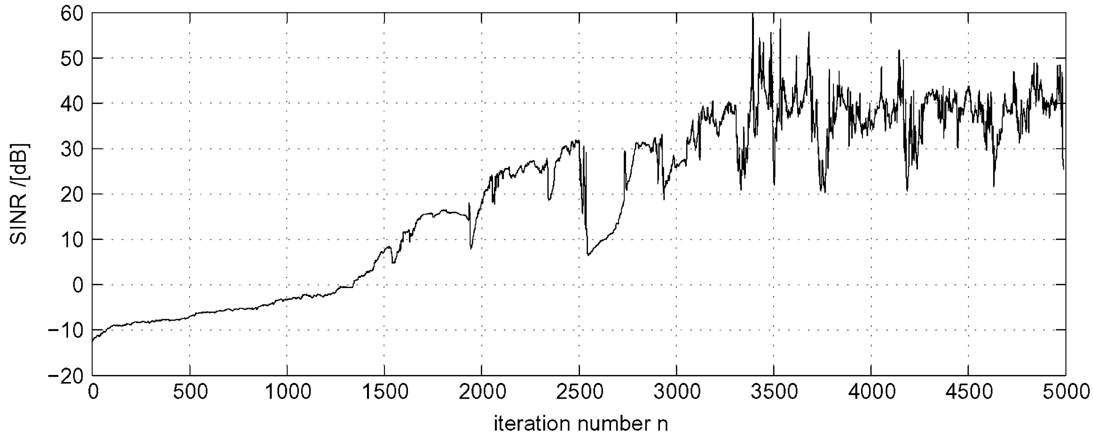

Fig. 4. Learning curve for a varying SIR.

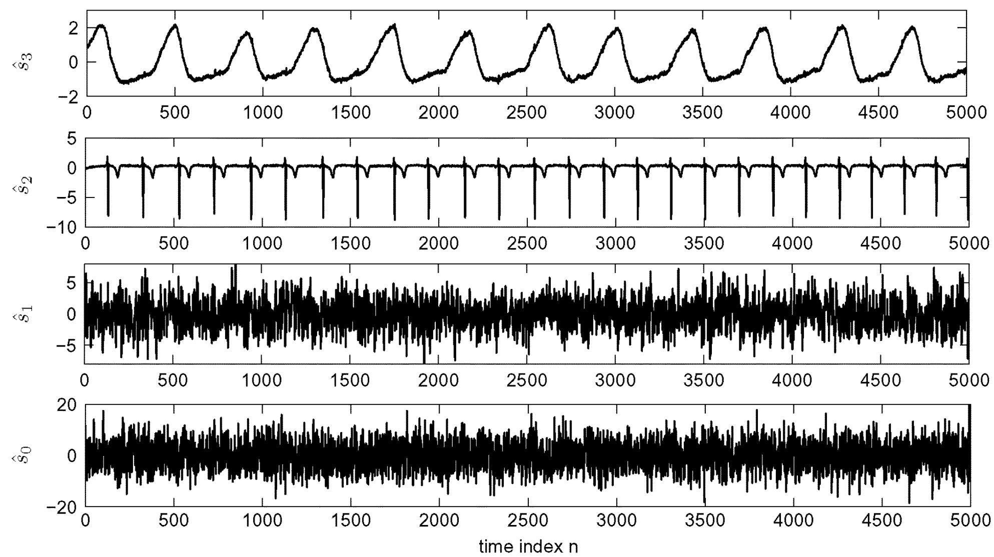

Fig. 5. Four extracted source signals.

where

$$\hat{\mathbf{x}}[n] = \mathbf{x}[n] - \sum_{p=1}^{P} b_p \mathbf{x}[n-p]. \tag{40}$$

The MSPE $E\{e_0^2[n]\}$ and the power of the extracted signal $E\{y_0^2[n]\}$ can be estimated recursively by

$$\begin{aligned} \sigma_{e_0}^2[n] &= \beta_{e_0}\sigma_{e_0}^2[n-1] + (1-\beta_{e_0})e_0^2[n], \\ \sigma_{y_0}^2[n] &= \beta_{y_0}\sigma_{y_0}^2[n-1] + (1-\beta_{y_0})y_0^2[n] \end{aligned} \tag{41}$$

where β_{e_0} and β_{y_0} are the corresponding forgetting factors.

Following some standard stochastic approximation techniques [17], we obtain the following online update for $\mathbf{w}_0[n]$:

$$\mathbf{w}_0[n+1] = \mathbf{w}_0[n] - \frac{\mu}{\sigma_{y_0}^2[n]} \left(e_0[n]\hat{\mathbf{x}}[n] - \frac{\sigma_{e_0}^2[n]}{\sigma_{y_0}^2[n]} y_0[n]\mathbf{x}[n] \right) \tag{42}$$

where μ is the learning rate. The total number of additions for each update is $M(P+3)+P+1$ and for the multiplications it is $M(P+3)+P+10$. To avoid the critical case where the norm of $\mathbf{w}_0[n]$ becomes too small, after each update, we can normalize it to unit length.

After obtaining the first extracted signal $y_0[n]$, we can remove it from the mixtures $\mathbf{x}[n]$ and apply the same algorithm to extract the second signal $y_1[n]$, and so on, until we have extracted all of the signals.

IV. Simulations

To illustrate the validity of this approach, we performed experiments on four benchmark signals $s_0, \ldots, s_3$ taken from the file ABio7.mat provided by the ICALAB toolbox [1], as shown in Fig. 2. The coefficients of a randomly generated linear predictor with a length of $P = 10$ are given by

$$\begin{aligned} \mathbf{b} = [&0.8032 \quad -0.3060 \quad -0.7430 \quad -0.7584 \quad -0.8896 \\ &1.1447 \quad -0.9456 \quad 0.5927 \quad 0.1641 \quad 1.6832]. \end{aligned} \tag{43}$$

The normalized prediction errors of the four signals from Fig. 2 were, respectively, $\{8.9657, 7.5055, 7.0566, 0.1612\}$. The mixing matrix $\mathbf{A}$ was randomly generated and given by

$$\mathbf{A} = \begin{bmatrix} 0.6360 & 0.9210 & -0.3073 & 0.3117 \\ -2.3134 & 1.7274 & 2.3179 & 1.0916 \\ -0.6808 & -1.0141 & 0.4437 & 1.2855 \\ -1.9456 & -1.0939 & -0.3019 & 0.4671 \end{bmatrix}. \tag{44}$$

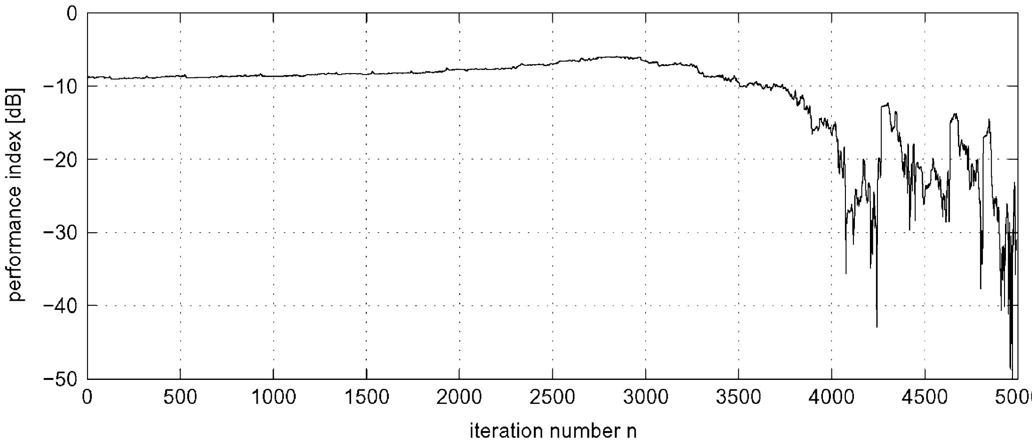

Fig. 6. Performance index learning curve for extracting the signal s_2.

From the analysis in Sections II and III, by minimizing the normalized MSPE, the signal with the smallest normalized MSPE will be extracted, which in this case is the fourth signal s_3.

To verify this, we conducted an experiment for which the learning curve with $\beta_{e_0} = \beta_{y_0} = 0.975$ and $\mu = 0.00012$ is shown in Fig. 3, whereby the performance index was defined as [1]

$$\mathrm{PI} = 10\log_{10}\left(\frac{1}{M-1}\left(\sum_{m=0}^{M-1}\frac{g_m^2}{\max\{g_0^2, g_1^2, \ldots, g_{M-1}^2\}} - 1\right)\right) \tag{45}$$

with $\mathbf{g} = \mathbf{A}^T\mathbf{w_0} = [g_0\ g_1 \cdots g_{M-1}]$. To further illustrate the performance of the proposed approach, the learning curve for the varying signal-to-interference ratio (SIR) within the extracted output is shown in Fig. 4.

As the performance index reaches a level around −40 dB and the SIR around 40 dB, we can say the signal s_3 has been extracted successfully, as shown by $\hat{s}_3$ in Fig. 5. The monotonic trend in Figs. 3 and 4 indicates the successful online training. To extract the second signal, we first removed s_3 from the original mixtures using a simple adaptive deflation algorithm [1], and then, applied the same algorithm to the new mixtures to extract the second one, and in this case, it was s_2, denoted by $\hat{s}_2$ in Fig. 5. The performance index learning curve is given by Fig. 6. Note in the second extraction, we still have four mixtures, i.e., $M = 4$, but the number of sources has been reduced to $L = 3$. We repeated this process until all of the source signals had been extracted, as shown by the remaining diagrams in Fig. 5.

V. Conclusion

We have revisited the CCA approach in the context of BSS. Although the validity of this approach can be justified indirectly by the matrix pencil method, to give further insight, we have provided a detailed direct proof for the existence of the solution. Two online realizations of this approach have been proposed, and, as an example, a BSE algorithm based on a linear predictor with fixed coefficients has been thoroughly analyzed.

References

[1] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. New York: Wiley, 2003.

[2] L. Tong, V. C. Soon, R. Liu, and Y. Huang, “AMUSE: A new blind identification algorithm,” in *Proc. IEEE Int. Symp. Circuits Syst.*, New Orleans, LA, May 1990, vol. 3, pp. 1784–1787.

[3] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and Moulines, “A blind source separation technique using second-order statistics,” *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[4] S. Degerine and R. Malki, “Second-order blind separation of sources based on canonical partial innovations,” *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 629–641, Mar. 2000.

[5] S. V. Schell and W. A. Gardner, “Programmable canonical correlation analysis: A flexible framework for blind adaptive spatial filtering,” *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 2898–2908, Dec. 1995.

[6] J. Galy and C. Adnet, “Canonical correlation analysis: A blind source separation using non-circularity,” in *Proc. IEEE Workshop Neural Netw. Signal Process.*, Australia, Dec. 2000, vol. 1, pp. 465–473.

[7] O. Friman, M. Borga, P. Lundberg, and H. Knutsson, “Exploratory fMRI analysis by autocorrelation maximization,” *NeuroImage*, vol. 16, pp. 454–464, Jun. 2002.

[8] 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.

[9] A. Cichocki and R. Thawonmas, “On-line algorithm for blind signal extraction of arbitrarily distributed, but temporally correlated sources using second order statistics,” *Neural Process. Lett.*, vol. 12, pp. 91–98, 2000.

[10] A. K. Barros and A. Cichocki, “Extraction of specific signals with temporal structure,” *Neural Comput.*, vol. 13, pp. 1995–2003, 2001.

[11] W. Liu, D. P. Mandic, and A. Cichocki, “A class of novel blind source extraction algorithms based on a linear predictor,” in *Proc. IEEE Int. Symp. Circuits Syst.*, Kobe, Japan, May 2005, pp. 3599–3602.

[12] C. Chang, Z. Ding, S. Yau, and F. Chan, “A matrix-pencil approach to blind separation of nonwhite signals in white noise,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Seattle, WA, May 1998, vol. 4, pp. 2485–2488.

[13] A. M. Tome, “Blind source separation using a matrix pencil,” in *Proc. IEEE-INNS-ENNS Int. Joint Conf. Neural Netw.*, Como, Italy, Jul. 2000, vol. 3, pp. 339–342.

[14] A. M. Tome, “An iterative eigendecomposition approach to blind source separation,” in *Proc. 3rd Int. Conf. Independent Compon. Anal. Blind Signal Separation*, San Diego, CA, Dec. 2001, pp. 424–428.

[15] S. A. Cruces-Alvarez, A. Cichocki, and S. Amari, “From blind signal extraction to blind instantaneous signal separation: Criteria, algorithms, and stability,” *IEEE Trans. Neural Netw.*, vol. 15, no. 4, pp. 859–873, Jul. 2004.

[16] T. W. Anderson, *An Introduction to Multivariate statistical Analysis*, 2nd ed. New York: Wiley, 1984.

[17] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.