A CCA Criterion Based Adaptive Algorithm for Blind Extraction of Specific Signal

Wei-Tao Zhang, Xiao-Guang Yuan, and Shun-Tian Lou

***Abstract*—In this paper, the canonical correlation approach for blind source separation is revisited. We proposed a novel criterion for source extraction, which is proved to be equivalent to the CCA criterion, whereas it has some advantages over the standard CCA criterion in developing fast learning algorithms. By optimizing the proposed criterion, we developed two learning algorithms for the extraction of specific signals. The first algorithm is based on the steepest descent technique, and the second one is a modified Newton algorithm. The experiment results demonstrate the effectiveness of the proposed learning algorithms, and show that the modified Newton algorithm converges much faster than the other extraction method.**

***Index Terms*—Blind source separation, canonical correlation analysis, steepest descent, Newton iteration.**

I. Introduction

Blind source separation, within the framework of independent component analysis (ICA) has attracted a great deal of attention in many engineering fields, such as speech enhancement, feature extraction, inter-user interference suppression in wireless communication, and biomedical signal analysis and so on. In ICA, The recorded multichannel signal reads

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)\,,\; t = 1,2,\ldots,T \tag{1}$$

where $\mathbf{s}(t)$ is the N dimensional source vector with zero mean, $\mathbf{x}(t)$ is the M dimensional mixed signal vector, $\mathbf{A}$ is a $M \times N$ ($M \geq N$) full column unknown mixing matrix. The goal is to recover the original source signals given the mixed signal without the prior knowledge of the sources and the mixing process.

Unlike the separation of all source signals, in this paper we focus our attention on the extraction of one specific source from the mixtures because the extraction approach provides more flexibility [1], it enables to extract the sources in a certain or prescribed order. Moreover, we can extract only the source of interest if some *a priori* knowledge of the source of interest can be exploited [2]. A number of blind extraction algorithms have been developed in the literature. The FastICA based on higher order statistics [3] is a typical sequential extraction algorithm for source separation. However, the higher order statistics are usually sensitive to outliers. An alternative is to exploit the second order statistics of sources with temporal structures. The extraction algorithm in [4] assumes that the source can be modeled by a stable autoregressive (AR) model. The sequential diagonalization algorithm [1] exploits the joint diagonalization structure of the source correlation matrices. Algorithms in [5], [6] assume that the sources have very special temporal structure, and they also implicitly use the periodicity of the sources. In addition, the canonical correlation analysis (CCA) approach is also proved to be a powerful tool for source extraction. However, most of these algorithms work in batch mode, and hence can not be used in real time applications [7], [8]. By minimizing the normalized mean square prediction error (MSPE) of the output, several online extraction algorithms have been developed with fixed [9], [10] or optimal learning prediction coefficients [4]. However, these online algorithms use a gradient descent technique and hence suffer from slow convergence rate.

In this paper, we presented a novel criterion with the merit of only global minimum for source extraction, and developed two learning algorithms for adaptive source extraction. The first one is the steepest descent optimization algorithm, the second one is a Newton like algorithm, where the standard Newton direction is properly modified, yielding a more robust quasi-Newton algorithm. Compared with the gradient learning algorithms, the proposed algorithms provide faster convergence speed.

II. The Cost Function

Let $\mathbf{w}$ be a M dimensional extraction vector, then the extracted signal can be written as

$$y(t) = \mathbf{w}^T\mathbf{x}(t) \tag{2}$$

The CCA based extraction schemes usually make the following assumption [7]-[10]:

Assumption 1: The source signals are spatially uncorrelated but temporally colored, i.e., the source correlation matrix is diagonal $\mathbf{R}_{\mathbf{s}}^{\tau} = \mathrm{diag}[r_1(\tau),\ldots,r_N(\tau)]$ with $r_i(\tau) = E[s_i(t)s_i(t-\tau)] \neq 0$, $i \in \mathcal{I}_N$ for some nonzero time delay τ, where $\mathcal{I}_N$ denotes the integer set $\{1,\ldots,N\}$.

The basic idea behind the CCA approach is that the sum of any uncorrelated signals has an autocorrelation that is less than or equal to the maximum value of individual signals [7]. Thus the CCA approach maximizes the normalized time

Manuscript received March 30, 2014; revised May 30, 2014. This work was supported in part by the NSFC of China under Grant 61201299, in part by the Fundamental Research Funds for the Central Universities under Grant K5051202044, and in part by the Research Fund for the Doctoral Program of Higher Education of China under Grant 20120203120007.

Wei-Tao Zhang, Xiao-Guang Yuan, and Shun-Tian Lou are with the School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: zhwt-work@foxmail.com, xgyuan@xidian.edu.cn, loushuntian@gmail.com).

DOI: 10.7763/IJIEE.2014.V4.466

delayed autocorrelation of the extracted signal $y(t)$ [7], i.e.

$$\max_{\mathbf{w}} \tilde{J}(\mathbf{w}) = \frac{E[y(t)y(t-\tau)]}{E[y^2(t)]} = \frac{\mathbf{w}^T\mathbf{R}_{\mathbf{x}}^{\tau}\mathbf{w}}{\mathbf{w}^T\mathbf{R}_{\mathbf{x}}^{0}\mathbf{w}} \tag{3}$$

where $\mathbf{R}_{\mathbf{x}}^{\tau} = E[\mathbf{x}(t)\mathbf{x}^T(t-\tau)]$ is the time delayed correlation matrix of the mixed signals and $\mathbf{R}_{\mathbf{x}}^{0} = E[\mathbf{x}(t)\mathbf{x}^T(t)]$ is the covariance matrix of the mixed signals. It can be easily seen that the CCA approach represents the same class of matrix pencil method. Let $(\lambda_i, \mathbf{v}_i)$ $i \in \mathcal{I}_M$ denotes the generalized eigenpair of matrix pencil $(\mathbf{R}_{\mathbf{x}}^{\tau}, \mathbf{R}_{\mathbf{x}}^{0})$, then the optimal solution is typically obtained by taking $\mathbf{w}$ to be the generalized eigenvector $\mathbf{v}_{\max}$ of matrix pencil $(\mathbf{R}_{\mathbf{x}}^{\tau}, \mathbf{R}_{\mathbf{x}}^{0})$ associated with the maximum generalized eigenvalue $\lambda_{\max}$.

It can be note that $(\mathbf{R}_{\mathbf{x}}^{\tau}, \mathbf{R}_{\mathbf{x}}^{0})$ and $(\mathbf{R}_{\mathbf{s}}^{\tau}, \mathbf{R}_{\mathbf{s}}^{0})$ are congruent matrix pencils if the mixing matrix $\mathbf{A}$ is assumed to be nonsingular [11], which indicates that $(\mathbf{R}_{\mathbf{x}}^{\tau}, \mathbf{R}_{\mathbf{x}}^{0})$ and $(\mathbf{R}_{\mathbf{s}}^{\tau}, \mathbf{R}_{\mathbf{s}}^{0})$ have the same generalized eigenvalues, i.e.,

$$\lambda_i = \frac{r_i(\tau)}{r_i(0)}, \ \forall i \in \mathcal{I}_N \tag{4}$$

Equation (4) physically shows that the uniqueness condition of CCA approach for source extraction is that the source signals have different normalized spectra. Moreover, equation (4) also implies that the signal with the maximum autocorrelation coefficient will be extracted.

Although the standard CCA criterion (3) can be used to develop learning algorithms for source extraction, the resulting algorithms usually suffer from slow convergence rate since the source signals are all assumed to be temporally colored. We present a new cost function as follows

$$\min_{\mathbf{w}} J(\mathbf{w}) = \mathbf{w}^T\mathbf{R}_{\mathbf{x}}^{0}\mathbf{w} - \log(\mathbf{w}^T\mathbf{R}_{\mathbf{x}}^{\tau}\mathbf{w}). \tag{5}$$

Straightforward matrix calculations show that the proposed cost function has the same global optimal solution as the standard CCA cost function, which guarantees the optimality of the resulting algorithm. In addition, the proposed cost function has completely different landscape with the standard CCA cost function, which will improve the convergence behavior of the resulting algorithm. Before presenting our algorithm, we would like to introduce an additional assumption below.

Assumption 2: Their exists a time delay τ such that the correlation matrix $\mathbf{R}_{\mathbf{x}}^{\tau}$ is positive definite.

The above assumption is of great significance because it makes the proposed criterion valid.

III. The Learning Algorithms

A. Steepest Descent Learning

In this section, we propose two learning algorithms for source extraction, the first one is devoted to the steepest descent learning, and the second one is a modified Newton learning algorithm.

Optimizing the proposed cost function with respect to the extraction vector $\mathbf{w}$ through the steepest descent technique is straightforward. The derivative $g(\mathbf{w})$ of $J(\mathbf{w})$ with respect to $\mathbf{w}$ can be written as

$$g(\mathbf{w}) = \nabla_{\mathbf{w}} J = 2\Big[\mathbf{R}_{\mathbf{x}}^{0}\mathbf{w} - (\mathbf{w}^T\mathbf{R}_{\mathbf{x}}^{\tau}\mathbf{w})^{-1}\mathbf{R}_{\mathbf{x}}^{\tau}\mathbf{w}\Big] \tag{6}$$

Note that the correlation matrices $\mathbf{R}_{\mathbf{x}}^{\tau}$ and $\mathbf{R}_{\mathbf{x}}^{0}$ can be recursively computed

$$\mathbf{R}_{\mathbf{x}}^{\tau}(t) = \beta\mathbf{R}_{\mathbf{x}}^{\tau}(t-1) + \mathbf{x}(t)\mathbf{x}^T(t-\tau)$$
$$\mathbf{R}_{\mathbf{x}}^{0}(t) = \beta\mathbf{R}_{\mathbf{x}}^{0}(t-1) + \mathbf{x}(t)\mathbf{x}^T(t)$$

where $0 < \beta \le 1$ is the forgetting factor. For steepest descent learning, the scale factor in equation (6) will be absorbed into the learning step size μ. We summarize the steepest descent learning algorithm in Table I.

TABLE I: The Steepest Descent Algorithm

Initialize $\mathbf{w}(0)$, $\mathbf{R}_{\mathbf{x}}^{0}(0)$ and $\mathbf{R}_{\mathbf{x}}^{\tau}(0)$ suitably.
For $t = 1, 2, \ldots$ do
$y(t) = \mathbf{w}^T(t-1)\mathbf{x}(t)$
$\mathbf{R}_{\mathbf{x}}^{\tau}(t) = \beta\mathbf{R}_{\mathbf{x}}^{\tau}(t-1) + \mathbf{x}(t)\mathbf{x}^T(t-\tau)$
$\mathbf{R}_{\mathbf{x}}^{0}(t) = \beta\mathbf{R}_{\mathbf{x}}^{0}(t-1) + \mathbf{x}(t)\mathbf{x}^T(t)$
$\mathbf{p}(t) = \mathbf{R}_{\mathbf{x}}^{\tau}(t)\mathbf{w}(t-1)$
$\mathbf{q}(t) = \mathbf{R}_{\mathbf{x}}^{0}(t)\mathbf{w}(t-1)$
$\kappa(t) = \mathbf{w}^T(t-1)\mathbf{p}(t)$
$\mathbf{g}(t) = \mathbf{q}(t) - \frac{1}{\kappa(t)}\mathbf{p}(t)$
$\mathbf{w}(t) = \mathbf{w}(t-1) - \mu\mathbf{g}(t)$

It should be pointed out that the selection of step size μ is a tradeoff between the convergence speed and the steady state accuracy. This is an inherit drawback of gradient based learning algorithms. To overcome this shortcoming, we propose a modified Newton learning algorithm below.

B. Modified Newton Learning

The Newton learning is given by

$$H(\mathbf{w})\Delta\mathbf{w} = -g(\mathbf{w}) \tag{7}$$

$$\mathbf{w}(t) = \mathbf{w}(t-1) + \Delta\mathbf{w} \tag{8}$$

where $H(\mathbf{w}) = \nabla_{\mathbf{w}}\nabla_{\mathbf{w}}^T J$ is the Hessian matrix of $J(\mathbf{w})$ with respect to $\mathbf{w}$, $\Delta\mathbf{w}$ represents the standard Newton direction. It is easy to show

$$\begin{aligned} H(\mathbf{w}) = 2\Big[\mathbf{R}_{\mathbf{x}}^{0} &+ 2(\mathbf{w}^T\mathbf{R}_{\mathbf{x}}^{\tau}\mathbf{w})^{-2}\mathbf{R}_{\mathbf{x}}^{\tau}\mathbf{w}\mathbf{w}^T\mathbf{R}_{\mathbf{x}}^{\tau T} \\ &-(\mathbf{w}^T\mathbf{R}_{\mathbf{x}}^{\tau}\mathbf{w})^{-1}\mathbf{R}_{\mathbf{x}}^{\tau}\Big] \end{aligned} \tag{9}$$

Moreover, the proper factorization of Hessian matrix shows that $H(\mathbf{w})$ is positive definite at the global minima point $\mathbf{v}_{\max}$. However this can not guarantee positive definiteness of $H(\mathbf{w})$ at each iteration, which is very important for Newton learning. We here propose to directly drop the last term in the right hand side of (9), yielding a modified Hessian matrix

$$H(\mathbf{w}) \approx 2\left[\mathbf{R}_{\mathrm{x}}^{0} + 2(\mathbf{w}^{T}\mathbf{R}_{\mathrm{x}}^{\tau}\mathbf{w})^{-2}\mathbf{R}_{\mathrm{x}}^{\tau}\mathbf{w}\mathbf{w}^{T}\mathbf{R}_{\mathrm{x}}^{\tau T}\right] \qquad (10)$$

Note that the modified Hessian matrix is always positive definite at each iteration step. The computation of $\Delta\mathbf{w}$ requires the calculation of the inverse of Hessian matrix, which is computationally demanding for online realization of the learning algorithm. A more efficient and numerically robust way is to apply the matrix inversion lemma to compute the inverse of Hessian matrix (10).

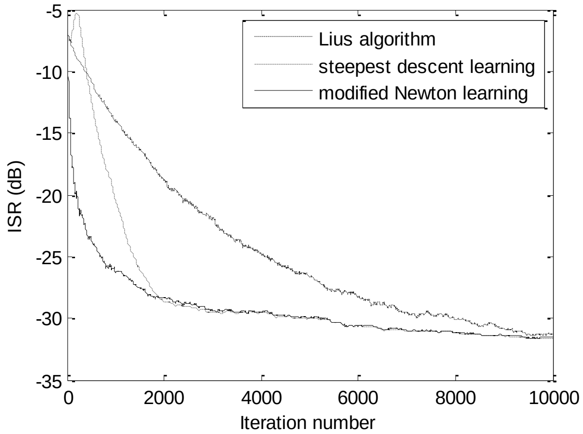

Fig. 1. Average ISR of the learning algorithms for extracting AR source. Each mixture is polluted by an additive white Gaussian noise with SNR=20dB.

TABLE II: THE MODIFIED NEWTON ALGORITHM

Initialize $\mathbf{w}(0)$, $\mathbf{Q}(0)$ and $\mathbf{R}_{\mathrm{x}}^{\tau}(0)$ suitably.

For $t = 1, 2, \ldots$ do

$$y(t) = \mathbf{w}^{T}(t-1)\mathbf{x}(t)$$

$$\mathbf{R}_{\mathrm{x}}^{\tau}(t) = \beta\mathbf{R}_{\mathrm{x}}^{\tau}(t-1) + \mathbf{x}(t)\mathbf{x}^{T}(t-\tau)$$

$$\mathbf{p}(t) = \mathbf{R}_{\mathrm{x}}^{\tau}(t)\mathbf{w}(t-1)$$

$$\mathbf{Q}(t) = \frac{1}{\beta}\left[\mathbf{Q}(t-1) - \frac{\mathbf{Q}(t-1)\mathbf{x}(t)\mathbf{x}^{T}(t)\mathbf{Q}(t-1)}{\beta + \mathbf{x}^{T}(t)\mathbf{Q}(t-1)\mathbf{x}(t)}\right]$$

$$\gamma(t) = \mathbf{w}^{T}(t-1)\mathbf{p}(t)$$

$$\mathbf{w}(t) = \frac{3\gamma(t)\mathbf{Q}(t)\mathbf{p}(t)}{\gamma^{2}(t) + 2\mathbf{p}^{T}(t)\mathbf{Q}(t)\mathbf{p}(t)}$$

Let $\mathbf{p} = \mathbf{R}_{\mathrm{x}}^{\tau}\mathbf{w}$, we have

$$H^{-1}(\mathbf{w}) = \frac{1}{2}\left[(\mathbf{R}_{\mathrm{x}}^{0})^{-1} - \frac{2(\mathbf{R}_{\mathrm{x}}^{0})^{-1}\mathbf{p}\mathbf{p}^{T}(\mathbf{R}_{\mathrm{x}}^{0})^{-1}}{\gamma^{2} + 2\mathbf{p}^{T}(\mathbf{R}_{\mathrm{x}}^{0})^{-1}\mathbf{p}}\right] \qquad (11)$$

where $\gamma \triangleq \mathbf{w}^{T}\mathbf{p}$. The substitution of (6) and (11) into (7) leads to the modified Newton direction

$$\begin{aligned}\Delta\mathbf{w} &= -H^{-1}(\mathbf{w})g(\mathbf{w}) \\ &= -\mathbf{w} + \frac{3\gamma(\mathbf{R}_{\mathrm{x}}^{0})^{-1}\mathbf{p}}{\gamma^{2} + 2\mathbf{p}^{T}(\mathbf{R}_{\mathrm{x}}^{0})^{-1}\mathbf{p}}\end{aligned} \qquad (12)$$

Then the modified Newton learning algorithm is given by

$$\begin{aligned}\mathbf{w}(t) &= \mathbf{w}(t-1) + \Delta\mathbf{w}\Big|_{\mathbf{w}=\mathbf{w}(t-1)} \\ &= \frac{3\gamma(t)(\mathbf{R}_{\mathrm{x}}^{0})^{-1}\mathbf{p}(t)}{\gamma^{2}(t) + 2\mathbf{p}^{T}(t)(\mathbf{R}_{\mathrm{x}}^{0})^{-1}\mathbf{p}(t)}\end{aligned} \qquad (13)$$

Let $\mathbf{Q} = (\mathbf{R}_{\mathrm{x}}^{0})^{-1}$, then $\mathbf{Q}$ can be also recursively computed by using the matrix inversion lemma. We summarize the modified Newton algorithm in Table II.

The proposed learning algorithms are devoted to extracting one single source, if all the sources have to be recovered, one can sequentially extract the sources by using the deflation technique [3].

Fig. 2. ECG signals taken by abdominal (the first 5 signals) and thoracic (the last 3 signals) measurements from a pregnant woman.

IV. EXPERIMENTAL RESULTS

The performance of the proposed learning algorithms is evaluated on both the synthetic data and real world signals. The first example serves to show the convergence speed and extraction accuracy of the learning algorithms for extracting the specific autoregressive (AR) source. The second example is devoted to an experiment on the extraction of real ECG data. The results are compared with standard CCA criterion based gradient learning algorithms (Liu's algorithm [9]).

In order to assess the performance of the blind extraction algorithms, we use the interference to signal ratio (ISR) as the performance measure, which is defined as [9]

$$\mathrm{ISR} = 10\log_{10}\left[\frac{1}{N-1}\left(\frac{\sum_{i=1}^{N} c_i^2}{\max\{c_i^2\}} - 1\right)\right] \qquad (14)$$

where $\mathbf{c} = \mathbf{w}^{T}\mathbf{A} = [c_1, \ldots, c_N]$ is the global mixing separating vector. Clearly the vector $\mathbf{c}$ with one and only one nonzero entry indicates perfect extraction of the specific source (i.e., the source with maximum autocorrelation coefficient).

A. The Synthetic AR Signal Extraction

The sources are $N = 3$ temporally colored Gaussian signals, which are generated by three first order AR processes

$$s_i(t) = \alpha_i s_i(t-1) + e_i(t),\ i = 1, 2, 3$$

where α_i is the AR coefficient, which is drawn from a uniform distribution in the range $(0,1)$, $e_i(t)$ is a zero-mean white Gaussian driving sequence with variance $1-\alpha_i^2$ such that the resulting source signal has unit variance. $M=4$ mixtures are observed in white Gaussian noise background with variance σ^2. The signal to noise ratio (SNR) is defined as $10\log_{10}\sigma^{-2}$. The mixing matrix $\mathbf{A}$ is randomly generated, whose entries are drawn independently from a standard normal distribution.

It is important to make sure that the algorithms extract the same source signal. For the proposed algorithm, the extracted signal will be the one with the maximum autocorrelation coefficient. It can be readily shown that the autocorrelation coefficient of the sources are $r_i(\tau)=\alpha_i^\tau$, $i=1,2,3$. Therefore, the source signal generated with $\alpha_m=\max\{\alpha_1,\alpha_2,\alpha_3\}$ will be extracted by the proposed algorithm. In order to extract the same source signal, we use the optimal prediction filter with coefficients $[1,-\alpha_m]$ for Liu's algorithm in [9].

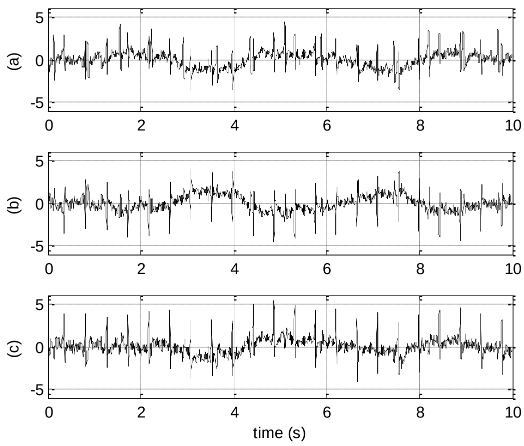

Fig. 3. The extracted FECG signals: (a) Liu's algorithm; (b) steepest descent algorithm; (c) modified Newton algorithm.

The settings for the extraction algorithms are as follows. The proposed algorithms use the time delay $\tau=1$. In order to obtain the better performance for the proposed steepest descent algorithm and Liu's algorithm, we use an exponentially decreased step size $\mu(t)=0.002\times 0.9998^t$. For the proposed modified Newton algorithm, without losing generality, we empirically set $\beta=1$ and use the initial values $\mathbf{w}(0)=[1,\ldots,1]^T$, $\mathbf{Q}(0)$ and $\mathbf{R}_\mathbf{x}^\tau(0)$ and $\mathbf{R}_\mathbf{x}^0(0)$ are identity matrix with proper size.

Fig. 1 plots the evolution of averaged ISR versus the iteration number over 500 independent trials. We see that as for the two gradient based learning algorithm, the proposed steepest descent algorithm outperforms the Liu's algorithm in terms of convergence speed. This is not surprising because they optimize the different criterion, which have absolutely different landscape. Among the three competitors, the proposed modified Newton algorithm converges much faster than the gradient ones, because it exhibits quadratic convergence rate in the neighborhood of the solution. However they achieve the same extraction accuracy because the proposed criterion has the same global optimal solution with the standard CCA criterion.

B. The Electrocardiogram Data Extraction

We also test our algorithm for the extraction of well-known electrocardiogram (ECG) data measured from a pregnant woman [12]. The data is shown in Fig. 2, which consists of 10 seconds of cutaneous potential recordings from $M=8$ sensors sampled at 250 Hz. One can see the strong and slow heart beating of mother and the weak and fast one of fetus. Our task is to extract both the fetus ECG (FECG) and mother ECG (MECG).

To extract the desired ECG signal, we use the periodicity of the desired signals. By computing the autocorrelation of the last recording, we find that the FECG and MECG signals have a periodicity of $P_F=112$ and $P_M=185$ samples respectively. To exploit this information, we set $\tau=P_F$ for the FECG extraction and $\tau=P_M$ for MECG. For gradient algorithms, the learning rate is $\mu=0.002$. The forgetting factor is $\beta=0.95$ for modified Newton algorithm.

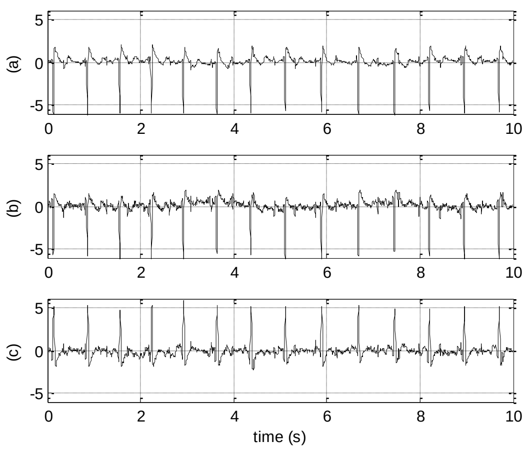

Fig. 4. The extracted MECG signals: (a) Liu's algorithm; (b) steepest descent algorithm; (c) modified Newton algorithm.

The extracted FECG and MECG signals are displayed in Fig. 3 and Fig. 4 respectively. We see that the fast heart beating of fetus can be clearly extracted by the proposed algorithms, whereas Liu's algorithm extracted the FECG with inferior extraction performance because the extracted FECG consists of visible MEGC signal. This may be due to its slow convergence rate. Observing the extracted FECG signal by the proposed algorithms, we see that the fetus heart rate is approximately 130~150 per minute, which is in accordance with the fact. As for the MECG, we see that three algorithms can successfully extract the MECG signals from the observations with the similar extraction performance.

V. Conclusion

In this paper, we proposed a novel CCA criterion for source extraction, which has a distinct landscape with the standard CCA criterion. This will lead to different behavior of learning algorithms. We also proposed the steepest descent learning and modified Newton learning algorithms to optimize the proposed criterion. The results on synthetic data and real world signals have demonstrated the good performance of the proposed learning algorithms.

REFERENCES

[1] X. L. Li and X. D. Zhang, "Sequential blind extraction adopting second-order statistics," *IEEE Signal Process. Lett.*, vol. 14, no. 1, pp. 58–61, 2007.

[2] R. Vollgraf, I. Schießl, and K. Obermayer, "Blind source separation of single components from linear mixtures," in *Proc. Int. Conf. Artificial Neural Networks*, 2001, pp. 509–514.

[3] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3 pp. 626–634, 1999.

[4] 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, no. 1, pp. 91–98, 2000.

[5] A. K. Barros and A. Cichocki, "Extraction of specific signals with temporal structure," *Neural Comput.*, vol. 13, no. 9, pp. 1995–2003, 2001.

[6] Z. L. Zhang and Z. Yi, "Robust extraction of specific signals with temporal structure," *Neurocomputing*, vol. 69, no. 7–9, pp. 888–893, 2006.

[7] M. Borga and H. Knutsson, "A canonical correlation approach to blind source separation," *Tech. Rep., Dept. Biomed. Eng.*, Linkoping Univ., Sweden, Jun. 2001.

[8] Y.-O. Li, T. Adali, W. Wang, and V. D. Calhoun, "Joint Blind Source Separation by Multiset Canonical Correlation Analysis," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3918–3929, 2009.

[9] W. Liu, D. P. Mandic, and A. Cichocki, "Analysis and online realization of the CCA approach for blind source separation," *IEEE Trans. Neural Netw.*, vol. 18, no. 5, pp. 1505–1510, 2007.

[10] W. Liu, D. P. Mandic and A. Cichocki, "Blind Source Extraction Based on a Linear Predictor," *IET Signal Processing*, vol. 1, no. 1, pp. 29–34, 2007.

[11] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.

[12] Dais．Database for the identification of systems．[Online] Available: http://homes.esat.kuleuven.be/~smc/daisy/daisydata.html

Wei-Tao Zhang was born in Shaanxi, China in 1983. He received the Ph.D. degree in control science and engineering from Xidian University, Xi'an, China, in 2011.

He is currently an associate professor with the School of Electronic Engineering, Xidian University. His research interests include blind signal processing, array processing, and biomedical engineering.

Xiao-Guang Yuan was born in Shanxi, China in 1982. He received the Ph.D. degree in circuit and system from Xidian University, Xi'an, China, in 2010.

He is currently a lecturer with the School of Electronic Engineering, Xidian University. His research interests include signal detection, signal processing and decision fusion.

Shun-Tian Lou was born in Zhejiang, China, in 1962. He received the B.Sc. degree in automatic control and M.Sc. degree in electronic engineering from Xidian University, Xi'an, China, and the Ph.D. degree in navigation guidance and control from Northwest Polytechnical University, Xi'an, China, in 1985, 1988, and 1999, respectively.

From 1999 to 2002, he was a postdoctoral fellow at the Institute of Electronic Engineering, Xidian University. Currently, he is a professor at the School of Electronic Engineering, Xidian University. His research interests are in the areas of signal processing, pattern recognition, and intelligent control using neural network and/or fuzzy system.