Blind source extraction based on a linear predictor

W. Liu, D.P. Mandic and A. Cichocki

Abstract: A rigorous analysis of the blind source extraction (BSE) approach based on a linear predictor is provided. It is shown that by minimising the mean squared prediction error (MSPE), as originally proposed, it is only possible to reach a solution subject to an arbitrary orthogonal transformation. To remove this ambiguity, a new cost function based on the normalised MSPE is introduced which by design provides a unique solution to this class of BSE problems. Depending on whether the pre-whitening operation is required or not, a novel class of BSE algorithms are derived and approaches with both fixed and adaptive linear predictor coefficients are considered. The proposed algorithms are justified by both the analysis and simulation results.

1 Introduction

Recently, due to its wide application in the areas of biomedical engineering, sonar, radar, speech enhancement, telecommunications, and so on, blind source separation (BSS) has been studied extensively. It is a technique to recover the original sources from all kinds of their mixtures, without the knowledge of the mixing process and the sources themselves. Many methods have been proposed, based on different assumptions, such as independent sources [1–4], spatial decorrelation and temporal correlation of the sources [5, 6], and non-stationarity of the sources [7, 8]. In BSS, we can either simultaneously recover all the source signals from their mixtures, or extract only one or a subset of the sources at a time. The latter case is also referred to as blind source extraction (BSE). An obvious advantage of BSE over BSS is its simplicity in cases where we are only interested in a small subset of the source signals. Algorithms specifically designed for BSE include those based on high-order statistics (HOS) [1, 9–11] and second-order statistics (SOS) [12–16]. A comprehensive overview of BSE algorithms, is given in [17].

In this paper, we provide a rigorous analysis for a class of BSE algorithms using a linear predictor [12–15] and propose solutions in order to mitigate some of the problems associated with this approach. This class of algorithms assumes that the sources are not correlated with each other and every source has a different temporal structure. Fig. 1 shows this architecture, where the extracted signal $y[n]$ and the instantaneous output error $e[n]$ of the linear predictor with a length P are given by

$$\begin{aligned} y[n] &= \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}[n] \\ e[n] &= y[n] - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}[n] \end{aligned} \tag{1}$$

where $\boldsymbol{x}[n]$ denotes the mixture vector at time instant n, $\boldsymbol{w}$ is a weight vector belonging to the demixing matrix, and $\boldsymbol{b}$ is the coefficient vector of the linear predictor, given by

$$\begin{aligned} \boldsymbol{b} &= [b_1 b_2 \cdots b_P]^{\mathrm{T}} \\ \boldsymbol{y}[n] &= [y[n-1] y[n-2] \cdots y[n-P]]^{\mathrm{T}} \\ \boldsymbol{x}[n] &= [x_0[n] x_1[n] \cdots x_{M-1}[n]]^{\mathrm{T}} \end{aligned} \tag{2}$$

In the analysis of this approach, it has been assumed that, as long as the source signals exhibit different temporal structures, they can be extracted successfully by minimising the mean square prediction error (MSPE) $E\{e^2[n]\}$ [17] subject to various constraints. However, for those algorithms which constrain only the length of the demixing vector [12–15], the success rate of the extraction performed in this way is fairly low, although it improves significantly with pre-whitening.

Here, we further address the issue of the success rate and provide a critical study of this BSE structure and the associated learning algorithms. A rigorous analysis shows that, by simply minimising the mean square prediction error $E\{e^2[n]\}$, we will not be able to guarantee a successful extraction. It is also shown that the reason for this lies in the ambiguity of the power levels of the source signals. To circumvent this ambiguity, we propose a novel cost function and associated learning algorithms, which impose the required constraints intrinsically and in a natural way. The analysis also shows that by performing pre-processing in the form of pre-whitening of the sensor data, together with normalisation of the demixing vector $\boldsymbol{w}[n]$ as necessary measures, the previously proposed methods of this kind can extract the sources successfully.

doi:10.1049/iet-spr:20060263

Paper first received 29th August 2006 and in revised form 8th January 2007

W. Liu is with the Communications Research Group, Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield, S1 3JD, UK

D.P. Mandic is with the Communications and Signal Processing Research Group, Department of Electrical and Electronic Engineering, Imperial College London, UK

A. Cichocki is with the Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Wako-shi, Saitama 351-0198, Japan and is also with the Department of Electrical Engineering, Warsaw University of Technology, 00-661-Warsaw, Poland

E-mail: w.liu@sheffield.ac.uk

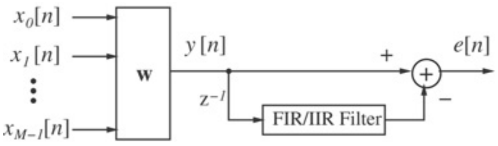

Fig. 1 *BSE architecture based on a linear predictor*

2 BSE using a linear predictor

In Fig. 1, the vector $\boldsymbol{x}[n]$ of mixed signals is given by

$$\boldsymbol{x}[n] = \boldsymbol{A}\boldsymbol{s}[n] \tag{3}$$

where $\boldsymbol{A}$ is the $M \times L$ mixing matrix and $\boldsymbol{s}[n]$ is the $L \times 1$ zero-mean source signal vector given by

$$\begin{aligned} &\boldsymbol{s}[n] = [s_0[n]s_1[n]\cdots s_{L-1}[n]]^{\mathrm{T}} \\ &[\boldsymbol{A}]_{m,l} = a_{m,l}, \quad m = 0, \ldots, M-1, \quad l = 0, \ldots, L-1 \end{aligned} \tag{4}$$

We assume that the sources are stationary, spatially uncorrelated and can be modelled by an AR process. The correlation matrix for zero lag is expressed as

$$\begin{aligned} \boldsymbol{R}_{ss}[0] &= E\{\boldsymbol{s}[n]\boldsymbol{s}^{\mathrm{T}}[n]\} \\ &= \mathrm{diag}\{\rho_0[0], \rho_1[0], \ldots, \rho_{L-1}[0]\} \end{aligned} \tag{5}$$

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

$$\begin{aligned} \boldsymbol{R}_{ss}[\Delta n] &= E\{\boldsymbol{s}[n]\boldsymbol{s}^{\mathrm{T}}[n-\Delta n]\} \\ &= \mathrm{diag}\{\rho_0[\Delta n], \rho_1[\Delta n], \ldots, \rho_{L-1}[\Delta n]\} \end{aligned} \tag{6}$$

with $\rho^l[\Delta n] \neq 0$ for some non-zero delays Δn.

The linear predictor in the BSE structure can be an FIR or IIR filter, adaptive or with fixed weights [17]. Without loss of generality, we shall assume a P-tap FIR filter with coefficient vector $\boldsymbol{b}$. The MSPE of the structure from Fig. 1 can be expressed as [17]

$$\begin{aligned} E\{e^2[n]\} &= E\{y^2[n]\} - 2E\{y[n]\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}[n]\} \\ &\quad + E\{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}[n]\boldsymbol{y}^{\mathrm{T}}[n]\boldsymbol{b}\} \\ &= E\{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}[n]\boldsymbol{x}^{\mathrm{T}}[n]\boldsymbol{w}\} \\ &\quad - 2\sum_{p=1}^{P} b_p \boldsymbol{w}^{\mathrm{T}} E\{\boldsymbol{x}[n]\boldsymbol{x}^{\mathrm{T}}[n-p]\}\boldsymbol{w} \\ &\quad + \sum_{p,q=1}^{P} b_p b_q \boldsymbol{w}^{\mathrm{T}} E\{\boldsymbol{x}[n-p]\boldsymbol{x}^{\mathrm{T}}[n-q]\}\boldsymbol{w} \\ &= \boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_{xx}[0]\boldsymbol{w} - 2\sum_{p=1}^{P} b_p \boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_{xx}[p]\boldsymbol{w} \\ &\quad + \sum_{p,q=1}^{P} b_p b_q \boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_{xx}[q-p]\boldsymbol{w} \end{aligned} \tag{7}$$

where $\boldsymbol{R}_{xx}[\Delta n] = E\{\boldsymbol{x}[n]\boldsymbol{x}^{\mathrm{T}}[n-\Delta n]\}$. From (3), we have

$$\begin{aligned} E\{\boldsymbol{x}[n]\boldsymbol{x}^{\mathrm{T}}[n-\Delta n]\} &= \boldsymbol{A}E\{\boldsymbol{s}[n]\boldsymbol{s}^{\mathrm{T}}[n-\Delta n]\}\boldsymbol{A}^{\mathrm{T}} \\ &= \boldsymbol{A}\boldsymbol{R}_{ss}[\Delta n]\boldsymbol{A}^{\mathrm{T}} \end{aligned} \tag{8}$$

Since, by assumption, the source signals are not correlated with each other, using this result, the MSPE $E\{e^2[n]\}$ now becomes [17]

$$\begin{aligned} E\{e^2[n]\} &= \boldsymbol{w}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{R}_{ss}[0]\boldsymbol{A}^{\mathrm{T}}\boldsymbol{w} - 2\sum_{p=1}^{P} b_p \boldsymbol{w}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{R}_{ss}[p]\boldsymbol{A}^{\mathrm{T}}\boldsymbol{w} \\ &\quad + \sum_{p,q=1}^{P} b_p b_q \boldsymbol{w}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{R}_{ss}[q-p]\boldsymbol{A}^{\mathrm{T}}\boldsymbol{w} \\ &= \boldsymbol{w}^{\mathrm{T}}\boldsymbol{A}\hat{\boldsymbol{R}}_{ss}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{w} \end{aligned} \tag{9}$$

where $\hat{\boldsymbol{R}}_{ss}$ is a diagonal matrix given by

$$\hat{\boldsymbol{R}}_{ss} = \boldsymbol{R}_{ss}[0] - 2\sum_{p=1}^{P} b_p \boldsymbol{R}_{ss}[p] + \sum_{p,q=1}^{P} b_p b_q \boldsymbol{R}_{ss}[q-p] \tag{10}$$

It should be noted that the diagonal elements of the matrix $\hat{\boldsymbol{R}}_{ss}$ represent the mean square prediction errors introduced by the corresponding source signals. They are, in general, different from each other for a specific power level, but change accordingly when the power levels of the sources vary. Due to this ambiguity of the power levels of the source signals, the differences in the diagonal elements of $\hat{\boldsymbol{R}}_{ss}$ can be conveniently absorbed into the mixing matrix $\boldsymbol{A}$ and we can therefore always assume $\hat{\boldsymbol{R}}_{ss}$ to be the identity matrix $\boldsymbol{I}$. Thus, (9) becomes

$$E\{e^2[n]\} = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{I}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{w} \tag{11}$$

For an arbitrary mixing matrix $\hat{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{F}$, where $\boldsymbol{F}$ is an arbitrary orthogonal matrix, we will have the same mean square error $E\{e^2[n]\}$, no matter whether the norm of the demixing vector is constrained or not.

Based on the above analysis, we can conclude that, by minimising $E\{e^2[n]\}$ with respect to the demixing vector $\boldsymbol{w}$, we can only obtain a result subject to an arbitrary orthogonal transformation. Essentially, this method is similar to the principal component analysis [17] and therefore it cannot guarantee successful extraction of the source signals.

In the next section, based on these observations, we propose a new cost function within this framework, which will remove the ambiguities associated with the previous approaches, and therefore greatly improve the success rate and performance.

3 Proposed algorithms for BSE with a linear predictor

3.1 General algorithm using the normalised MSPE

From the above analysis, we can see that the problem with minimising the cost function $E\{e^2[n]\}$ is the ambiguity associated with the prediction error (or the power level) for each source signal. Since the power levels of the source signals are unknown and arbitrary, they can have the same mean square prediction error for the same linear filter model. To remove this ambiguity, we propose a novel cost function, which is based on a normalised version of the MSPE, given by

$$J(\boldsymbol{w}) = \frac{E\{e^2[n]\}}{E\{y^2[n]\}}$$

The idea behind this approach is that, the normalised MSPE of a signal will not change with its power level and this normalised MSPE is normally different from those associated with the other signals. Notice that predictability is a fundamental property of a signal, and the normalised prediction error is an inherent characteristic of the signal, just like the normalised kurtosis [9, 17].

From

$$E\{y^2[n]\} = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_{xx}[0]\boldsymbol{w} = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{R}_{ss}[0]\boldsymbol{A}^{\mathrm{T}}\boldsymbol{w} \tag{12}$$

and following these arguments, the proposed cost function can now be expressed as

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{A}\hat{\boldsymbol{R}}_{ss}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{R}_{ss}[0]\boldsymbol{A}^{\mathrm{T}}\boldsymbol{w}} \tag{13}$$

Since the matrices $\hat{\boldsymbol{R}}_{ss}$ and $\boldsymbol{R}_{ss}[0]$, in general, cannot be an identity matrix at the same time, multiplying the mixing matrix $\boldsymbol{A}$ by an orthogonal matrix will change the value of the cost function, therefore the power level ambiguity is removed.

Let $\boldsymbol{g} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{w}$ denote the global demixing vector. Then (13) becomes

$$J(\boldsymbol{w}) = \frac{\boldsymbol{g}^{\mathrm{T}}\hat{\boldsymbol{R}}_{ss}\boldsymbol{g}}{\boldsymbol{g}^{\mathrm{T}}\boldsymbol{R}_{ss}[0]\boldsymbol{g}} \tag{14}$$

As pointed out before, without loss of generality, we shall assume that $\boldsymbol{R}_{ss}[0] = \boldsymbol{I}$, as the differences in the diagonal elements of $\boldsymbol{R}_{ss}[0]$ can always be absorbed into the mixing matrix $\boldsymbol{A}$. This way, the diagonal elements of $\hat{\boldsymbol{R}}_{ss}$ become the normalised MSPEs and they are assumed to be different from each other, otherwise the scheme will not be able to extract all the sources successfully. This yields

$$J(\boldsymbol{w}) = \frac{\boldsymbol{g}^{\mathrm{T}}\hat{\boldsymbol{R}}_{ss}\boldsymbol{g}}{\boldsymbol{g}^{\mathrm{T}}\boldsymbol{g}} \tag{15}$$

Let us define a new vector $\hat{\boldsymbol{g}} = \boldsymbol{g}/(\boldsymbol{g}^{\mathrm{T}}\boldsymbol{g})$, which has a property $\hat{\boldsymbol{g}}^{\mathrm{T}}\hat{\boldsymbol{g}} = 1$. Then the cost function becomes

$$J(\boldsymbol{w}) = \hat{\boldsymbol{g}}^{\mathrm{T}}\hat{\boldsymbol{R}}_{ss}\hat{\boldsymbol{g}} \tag{16}$$

Consider the optimisation problem formulated as

$$\min_{\hat{\boldsymbol{g}}} \hat{\boldsymbol{g}}^{\mathrm{T}}\hat{\boldsymbol{R}}_{ss}\hat{\boldsymbol{g}} \quad \text{subject to} \quad \hat{\boldsymbol{g}}^{\mathrm{T}}\hat{\boldsymbol{g}} = 1 \tag{17}$$

The solution to this problem is a vector $\hat{\boldsymbol{g}}_{opt}$ with only one non-zero element, which is strictly equal to unity at the position corresponding to the smallest diagonal element of the matrix $\hat{\boldsymbol{R}}_{ss}$ [17]. As $\hat{\boldsymbol{g}} = \boldsymbol{g}/(\sqrt{\boldsymbol{g}^{\mathrm{T}}\boldsymbol{g}})$, the corresponding global demixing vector $\boldsymbol{g}_{opt}$ will be the same as $\hat{\boldsymbol{g}}_{opt}$ except that the non-zero element in $\boldsymbol{g}_{opt}$ is an arbitrary constant c. Since we are minimising $J(\boldsymbol{w})$ with respect to $\boldsymbol{w}$, instead of $\hat{\boldsymbol{g}}$, we need to prove that there exists a $\boldsymbol{w}_{opt}$ which results in $\hat{\boldsymbol{g}}_{opt}$.

In fact, from $\boldsymbol{g} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{w}$, when $\boldsymbol{A}$ is of full rank and the number of mixtures M is larger or equal to the number of sources L, $\boldsymbol{w}_{opt}$ can be obtained using the pseudo-inverse of $\boldsymbol{A}^{\mathrm{T}}$ as

$$\boldsymbol{w}_{opt} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{g}_{opt} \tag{18}$$

For the case $M < L$, in general, we cannot find a $\boldsymbol{w}_{opt}$ which satisfies the equation $\boldsymbol{g} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{w}$, except for perhaps some special forms of $\boldsymbol{A}$. However, since $\boldsymbol{A}$ is unknown, it is difficult to know whether this $\boldsymbol{w}_{opt}$ exists or not.

Since the possible minimum value of $J(\boldsymbol{w})$ is reached only when $\boldsymbol{g} = \boldsymbol{g}_{opt}$, as long as there exists such a $\boldsymbol{w} = \boldsymbol{w}_{opt}$ so that $\boldsymbol{g} = \boldsymbol{g}_{opt}$, we can state that when we minimize $J(\boldsymbol{w})$ with respect to $\boldsymbol{w}$, this will result in a successful extraction of the source signal with a minimum normalised MSPE.

To derive the weight updates, applying the standard gradient descent method to $J(\boldsymbol{w})$, we have

$$\nabla_{\boldsymbol{w}}J = \frac{2}{E\{y^2[n]\}}\left(E\{e[n]\hat{\boldsymbol{x}}[n]\} - \frac{E\{e^2[n]\}}{E\{y^2[n]\}}E\{y[n]\boldsymbol{x}[n]\}\right) \tag{19}$$

where

$$\hat{\boldsymbol{x}}[n] = \boldsymbol{x}[n] - \sum_{p=1}^{P} b_p\boldsymbol{x}[n-p] \tag{20}$$

The MSPE $E\{e^2[n]\}$ and the power of the output of the demixing stage $E\{y^2[n]\}$ can be estimated recursively by

$$\begin{aligned} \sigma_e^2[n] &= \beta_e\sigma_e^2[n-1] + (1-\beta_e)e^2[n] \\ \sigma_y^2[n] &= \beta_y\sigma_y^2[n-1] + (1-\beta_y)y^2[n] \end{aligned} \tag{21}$$

where β_e and β_y are the corresponding forgetting factors.

Following some standard stochastic approximation techniques [18], from (19), we obtain the following online update equation for $\boldsymbol{w}[n]$

$$\boldsymbol{w}[n+1] = \boldsymbol{w}[n] - \frac{\mu}{\sigma_y^2[n]}\left(e[n]\hat{\boldsymbol{x}}[n] - \frac{\sigma_e^2[n]}{\sigma_y^2[n]}y[n]\boldsymbol{x}[n]\right) \tag{22}$$

where μ is the learning rate.

To avoid the critical case where the norm of $\boldsymbol{w}[n]$ becomes too small, after each update, we normalise it to unit length, which yields

$$\boldsymbol{w}[n+1] \leftarrow \boldsymbol{w}[n+1]/\sqrt{\boldsymbol{w}^{\mathrm{T}}[n+1]\boldsymbol{w}[n+1]} \tag{23}$$

3.2 Algorithm with pre-whitening

Alternatively, instead of the proposed approach based on minimising the normalised prediction error, we can constrain $E\{y^2[n]\}$ to a constant, which can be achieved in two ways. One is to add a new cost function $(1 - E\{y^2[n]\})^2$ and we minimise both of them at the same time [17]. But the problem is how much weight we give to the new cost function, and this weight will affect the performance of the derived algorithm. Another way is, we first pre-whiten the observed mixtures $\boldsymbol{x}[n]$, so that $\boldsymbol{R}_{xx}[0] = \boldsymbol{I}$, and then normalise the demixing vector $\boldsymbol{w}[n]$ according to (23). Under this condition, we have

$$E\{y^2[n]\} = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{R}_{xx}[0]\boldsymbol{w} = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} = 1 \tag{24}$$

and

$$J(\boldsymbol{w}) = \frac{E\{e^2[n]\}}{E\{y^2[n]\}} = E\{e^2[n]\} \tag{25}$$

Therefore, in this way the constraint to $E\{y^2[n]\}$ is achieved independently from the minimisation of the cost function, and we can perform BSE by simply minimising the standard MSPE.

To show this, it can be seen that from (12), we have

$$E\{y^2[n]\} = \boldsymbol{g}^{\mathrm{T}}\boldsymbol{R}_{ss}[0]\boldsymbol{g} \tag{26}$$

Assume $\boldsymbol{R}_{ss}[0] = \boldsymbol{I}$. From (24), we have

$$\boldsymbol{g}^{\mathrm{T}}\boldsymbol{g} = 1 \tag{27}$$

Then the optimisation problem with pre-whitening and normalisation can be formulated as

$$\min_{\boldsymbol{g}} \boldsymbol{g}^{\mathrm{T}}\hat{\boldsymbol{R}}_{ss}\boldsymbol{g} \quad \text{subject to} \quad \boldsymbol{g}^{\mathrm{T}}\boldsymbol{g} = 1 \tag{28}$$

Performing an analysis similar to that in the previous subsection, it can be shown that this scheme can indeed successfully extract the source signals.

From this platform, by simply minimising the MSPE, we can obtain the previously proposed BSE methods [12–15]. We have shown that, in this scenario, for those algorithms to work, the pre-whitening operation is a necessary preprocessing step, and not an optional one as often indicated (explicitly or implicitly) in the literature. Due to the fact that pre-whitening is usually performed before the actual extraction, and normalisation of the demixing vector (23) during the extraction, we did not find obvious problems with the performance of the previously proposed algorithms [12–15]. However, without pre-whitening, those methods will usually fail, as they are, in principle, simply minimising the MSPE.

In this case, for a fixed linear predictor, applying the standard gradient descent method to $E\{e^2[n]\}$, we can derive the following online update rule

$$\boldsymbol{w}[n+1] = \boldsymbol{w}[n] - \mu e[n]\hat{\boldsymbol{x}}[n] \quad (29)$$

where $\hat{\boldsymbol{x}}[n]$ is given in (20). This update is followed by the normalisation of the demixing vector by (23). Compared to (22), the update equation (29) appears much simpler. However, it requires a pre-whitening operation, which is not convenient and may be difficult to implement online. The algorithm proposed in (22), on the other hand, does not require any preprocessing, and is more suitable for online implementation.

It should be noted that after pre-whitening, by simply minimising the mean square error given in (7) subject to the normalisation of the demixing vector, the derived algorithm (29) can be considered as an adaptive version of finding the eigenvector of the weighted sum of both $\boldsymbol{R}_{xx}[0]$ and $\boldsymbol{R}_{xx}[n]$, $n = 1, \ldots, P$, as shown in (7). In this sense, this algorithm (29) is closely related to the SOBI (second-order blind identification) algorithm [5], where after pre-whitening, a demixing matrix is found by diagonalising the correlation matrices $\boldsymbol{R}_{xx}[n]$ at non-zero time lags. However, for the first proposed algorithm in (22), the pre-whitening step is not necessary, which is the major difference from the SOBI algorithm.

3.3 Choice of the coefficients of the linear predictor

From the above analysis, which proves the existence of the solution for the normalised MSPE-based cost function, we can see that for a chosen linear predictor with coefficient vector $\boldsymbol{b}$, as long as the source signals have different normalised MSPEs, such sources can be extracted successfully. This is achieved by minimising $J(\boldsymbol{w})$ using the proposed adaptive algorithms, which guarantees source extraction, however, the actual BSE performance is dependent on the specific choice of the coefficients within $\boldsymbol{b}$.

By intuition, the larger the relative differences of the normalised MSPEs of the source signals, the better the extraction performance. Indeed, in an ideal scenario, we wish to choose a linear predictor for which the normalised MSPEs of the source signals have the largest relative differences. However, due to the blind nature of the problem, it is impossible to find such an optimal linear predictor analytically. Alternatively, following the ideas from adaptive filtering, we may opt to minimise $J(\boldsymbol{w})$ with respect not only to the demixing vector $\boldsymbol{w}$, but also the coefficients of the linear predictor [12–15].

To illustrate this approach within the proposed framework of Section 3, we apply the normalised least mean square (NLMS) algorithm to perform updates of vector $\boldsymbol{b}$ [18], which for the cases with and without pre-whitening gives, respectively

$$\boldsymbol{b}[n+1] = \boldsymbol{b}[n] + \frac{\mu_b}{\boldsymbol{y}^{\mathrm{T}}[n]\boldsymbol{y}[n]} e[n]\boldsymbol{y}[n] \quad (30)$$

and

$$\boldsymbol{b}[n+1] = \boldsymbol{b}[n] + \frac{\mu_b}{(\sigma_y^2[n])(\boldsymbol{y}^{\mathrm{T}}[n]\boldsymbol{y}[n])} e[n]\boldsymbol{y}[n] \quad (31)$$

Although by using (30) and (31) we can achieve a smaller prediction error e[n], it is not immediately obvious whether this way we can also achieve the desired larger relative difference of the normalised MSPEs of the source signals. A simulation result in the next section illustrates the time variation of the relative differences of normalised MSPEs when adjusting the coefficients $\boldsymbol{b}$ adaptively.

As in the case with a fixed coefficient vector $\boldsymbol{b}$, our argument is that based on the analysis from Section 3, every source signal will have a different normalised MSPE (although the relative differences of the values of the resulting normalised MSPEs may not be very distinct), which guarantees a successful extraction.

4 Simulations

We performed experiments on four benchmark signals $s_0, \ldots, s_3$ taken from the file ABio7.mat provided by the ICALAB toolbox [17], as shown in Fig. 2. The coefficients of a randomly generated linear predictor with a length of $P = 20$ are given by

$$\begin{aligned}\boldsymbol{b} = [&0.8904\ -0.2785\ -0.8312\ \ 0.8970\ \ 0.3817\\ &0.2310\ \ 0.7900\ -0.9749\ -0.3982\ \ 0.9302\\ &-0.7958\ \ 0.1920\ -0.0620\ \ 0.4347\ \ 0.7177\\ &-0.6288\ -0.0753\ \ 0.8066\ -0.9558\ \ 0.5208]\end{aligned} \quad (32)$$

The normalised prediction errors of the four signals from Fig. 2 were, respectively, {9.0987, 4.0003, 2.4263, 0.6681}. Following the analysis from Sections 2 and 3, the signal with the smallest normalised prediction error will be extracted, which in this case is the fourth signal s_3.

In the first set of simulations, we illustrate the performance of the algorithm from (29). For this algorithm to

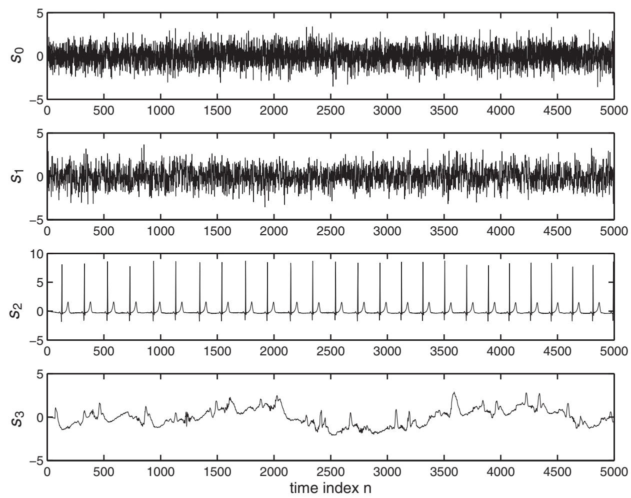

Fig. 2 *Source signals used in simulations*

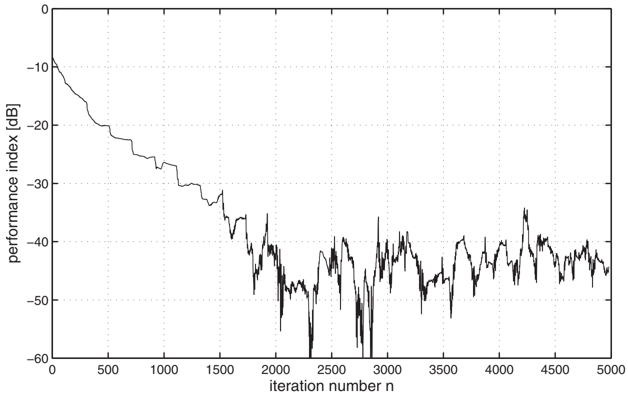

Fig. 3 *Performance index using method expressed in (29)*

work, we must pre-whiten the observed mixtures and normalise the demixing vector after each update. To measure the demixing effect of the algorithm, we employ the performance index defined as [17]

$$PI = 10\log_{10}\left(\frac{1}{L-1}\left(\sum_{l=0}^{L-1}\frac{g_l^2}{\max\{g_0^2, g_1^2, \ldots, g_{L-1}^2\}} - 1\right)\right) \tag{33}$$

with $\boldsymbol{g} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{w} = [g_0\, g_1 \cdots g_{L-1}]$.

The learning curve of the performance index with a stepsize $\mu = 3$ is shown in Fig. 3. The curve was obtained by averaging 100 trials of independent simulations, each with a randomly generated mixing matrix and a randomly generated initial value of the demixing vector. Observe that this way we were able to extract source s_3 successfully, which conforms to the results of the above analysis.

We next performed simulations for the proposed method without the pre-whitening given in (22). The forgetting factors were $\beta_e = \beta_y = 0.975$ and the stepsize $\mu = 0.0025$. A learning curve for this case is shown in Fig. 4 and it was also obtained by averaging 100 trials of independent simulations. As the performance index reaches a level of between -30 dB and -40 dB, we can say that the signal s_3 has been extracted successfully, as shown in Fig. 5. Comparing the extracted signal in Fig. 5 with the original s_3, it can be seen that $\hat{s}_3$ is a scaled version of s_3, which conforms to the principle that in BSS for instantaneous mixtures, the source signal can only be recovered subject to an arbitrary scalar.

The change of the linear predictor output $e[n]$ during the adaptation is shown in Fig. 6. At the initial stage of

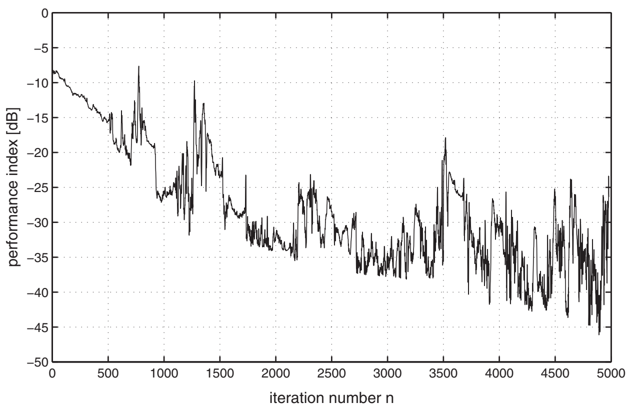

Fig. 4 *Performance index using method expressed in (22)*

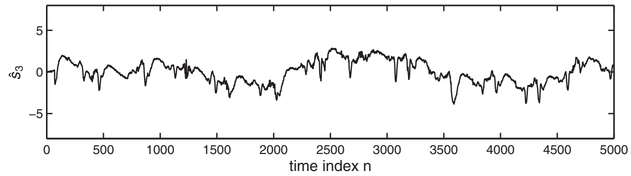

Fig. 5 *Source signal s_3 extracted using method expressed in (22)*

adaptation, the output signal $y[n]$ of the demixing vector $\boldsymbol{w}$ was still a mixture of all the four source signals, and after the filtering operation of the linear predictor, the output $e[n]$ did not look like any of the individual source signals, which can be verified by the values of $e[n]$ for the first 500 iterations in Fig. 6. With the adaptation of the demixing vector, the performance index reached a very small value, which means that the signal $y[n]$ has become one of the source signals and the linear predictor output $e[n]$ is simply a filtered version of the extracted signal $y[n]$ and in this case it is s_3. We can verify this by checking the similarity between the original signal s_3 and that of Fig. 6 for the part about $n > 1000$.

As mentioned in Section 3, although the algorithm based on pre-whitening seems simpler than the one based on the normalised MSPE, it is more difficult to implement it online. A comparison of the results of Figs. 3 and 4 shows the difference in the performance of the algorithms (22) and (29). For the one given in (22), the normalisation of the MSPE is performed by an approximation of the powers of $e[n]$ and $y[n]$, therefore it is not so precise as the one in (29), where the pre-whitening operation and normalisation of the demixing vector ensure a constant variance of the demixing vector output. As a result, in Fig. 4 we see a larger magnitude of oscillations of the steady-state performance index than in Fig. 3.

Finally, to further illustrate the effect of an adaptive linear predictor $\boldsymbol{b}$ within the analysed BSE structure, we next perform a comprehensive statistical analysis of this method and support it by numerical examples. As mentioned in Section 3.3, by adapting the coefficients of the linear predictor, we can achieve a smaller prediction error. The question is whether this leads to a larger relative difference of the normalised MSPEs of the source signals. Let us illustrate this on an example, where for simplicity, we consider the algorithm from Section 3.2. To analyse the time-varying behaviour of the relative differences of he normalised MSPEs during adaptation, we analysed the index PE, which shows the ratio between the two smallest normalised MSPEs.

Denote the normalised MSPEs of the source signals at the nth iteration by $pe_0[n], pe_1[n], \ldots, pe_{L-1}[n]$. The relative difference index PE at the nth iteration is calculated in the following way

$$PE = \frac{\min_2\{pe_0[n], \ldots, pe_{L-1}[n]\}}{\min\{pe_0[n], \ldots, pe_{L-1}[n]\}} \tag{34}$$

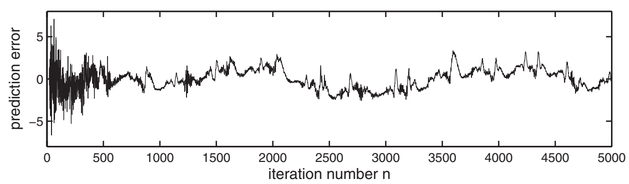

Fig. 6 *Prediction error $e[n]$ using method expressed in (22)*

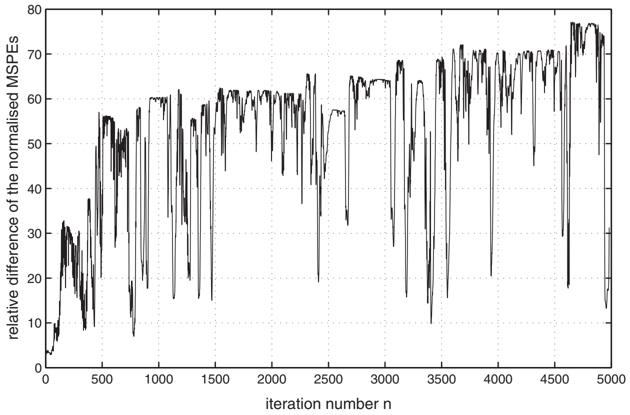

Fig. 7 *Change of relative difference of the normalised MSPEs with an adaptive linear predictor*

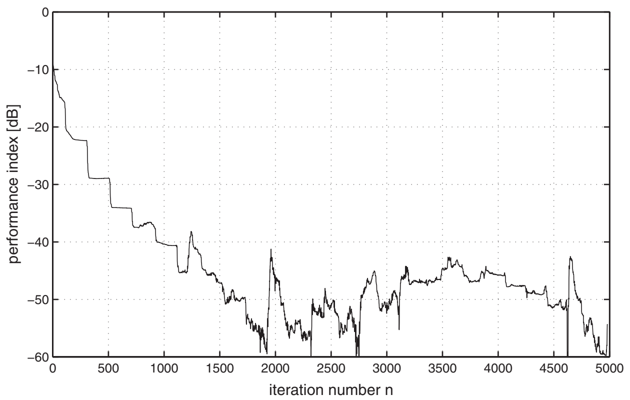

Fig. 8 *Performance index using the adaptive linear predictor approach*

where the symbol $\min_2$ denotes the second smallest value among the L normalised MSPEs. This is a natural measure of the minimum distance between the smallest normalised MSPE and the other MSPEs. Since the BSE algorithm extracts one of the sources at each stage, a larger PE indicates that one of the sources has a much smaller normalised MSPE than the remaining ones, and hence it is easier to extract it.

Fig. 7 shows the change of the PE during the adaptation of $\boldsymbol{b}$. Although the value of the PE shows a large variation between about 4 and 70, the PE is always greater than unity and on average it exhibits an increasing trend, which indicates that we can achieve a larger PE and hence an improved performance when using an adaptive linear predictor. The performance index of the BSE with prewhitening operation based on an adaptive linear predictor is shown in Fig. 8. Compared to the learning curve of Fig. 3, the adaptive linear predictor approach clearly outperforms the fixed one. This can be explained by the fact that the average PE value of the adaptive linear predictor approach is much larger than the one with a fixed linear predictor, which is approximately $2.4263/0.6681 \simeq 4$.

5 Conclusions

The problem of BSE based on a linear predictor structure has been addressed. The conditions for the existence of the solution have been highlighted, and the corresponding analysis has shown that based on the minimisation of the MSPE it is not possible to reach a unique solution to this problem. It has been further shown that this is due to the ambiguity associated with the power levels of the sources. To achieve a unique solution to this problem, a novel cost function has been introduced, based on which two novel adaptive BSE algorithms have been derived, each with either fixed or adaptive linear predictor coefficients. Simulation results have shown the usefulness of the proposed approach.

6 Acknowledment

The work of W. Liu and D.P. Mandic has been supported by EPSRC, GR/S71149/01.

7 References

1 Hyvarinen, A., and Oja, E.: 'A fast fixed-point algorithm for independent component analysis', *Neural Comput.*, 1997, **9**, pp. 1483–1492

2 Davis, M.E.: 'Identifiability issues in noisy ICA', *IEEE Signal Process. Lett.*, 2004, **11**, (5), pp. 470–473

3 Erdogmus, D., Hild, II K.E., Rao, Y.N., and Principe, J.C.: 'Minimax mutual information approach for independent components analysis', *Neural Comput.*, 2004, **16**, (6), pp. 1235–1252

4 Hesse, C.W., and James, C.J.: 'The FastICA algorithm with spatial constraints', *IEEE Signal Process. Lett.*, 2005, **12**, (11), pp. 792–795

5 Belouchrani, A., Abed-Meraim, K., Cardoso, J.F., and Moulines, E.: 'A blind source separation technique using second-order statistics', *IEEE Trans. Signal Process.*, 1997, **45**, (2), pp. 434–444

6 Degerine, S., and Malki, R.: 'Second-order blind separation of sources based on canonical partial innovations', *IEEE Trans. Signal Process.*, 2000, **48**, pp. 629–641

7 Pham, D.T., and Cardoso, J.F.: 'Blind separation of instantaneous mixtures of nonstationary sources', *IEEE Trans. Signal Process.*, 2001, **49**, (9), pp. 1837–1848

8 Jafari, M.G., Alty, S.R., and Chambers, J.A.: 'New natural gradient algorithm for cyclostationary sources', *IEE Proc., Vis. Image Signal Process.*, 2004, **151**, pp. 62–68

9 Cichocki, A., Thawonmas, R., and Amari, S.: 'Sequential blind signal extraction in order specified by stochastics properties', *Electron. Lett.*, 1997, **33**, pp. 64–65

10 Li, Y.Q., and Wang, J.: 'Sequential blind extraction of instantaneously mixed sources', *IEEE Trans. Signal Process.*, 2002, **50**, pp. 997–1006

11 Liu, W., and Mandic, D.P.: 'A normalised kurtosis based algorithm for blind source extraction from noisy measurements', *Signal Process.*, 2006, **86**, pp. 1580–1585

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

13 Barros, A.K., and Cichocki, A.: 'Extraction of specific signals with temporal structure', *Neural Comput.*, 2001, **13**, pp. 1995–2003

14 Cichocki, A., Rutkowski, T., and Siwek, K.: 'Blind signal extraction of signals with specified frequency band'. Proc. IEEE Workshop on Neural Networks for Signal Processing, September 2002, pp. 515–524

15 Mandic, D.P., and Cichocki, A.: 'An online algorithm for blind extraction of sources with different dynamical structures'. Proc. 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation, Nara, Japan, April 2003, pp. 645–650

16 Rutkowski, T.M., Yokoo, M., Yagi, K., Kameda, Y., Minoh, M., and Mandic, D.P.: 'Identification and tracking of active speaker's position in noisy environments'. Proc. International Workshop on Acoustic Echo and Noise Control, Kyoto, Japan, September 2003, pp. 283–286

17 Cichocki, A., and Amari, S.: 'Adaptive blind signal and image processing' (Wiley, New York, 2003)

18 Haykin, S.: 'Adaptive filter theory' (Prentice Hall, Englewood Cliffs, 1996, 3rd edn.)