BLIND EXTRACTION OF NOISY EVENTS USING NONLINEAR PREDICTOR

Wai Yie Leong and Danilo P. Mandic

Communications and Signal Processing Group
Dept. of Electronics and Electrical Engineering
Imperial College London, SW7 2AZ, UK
{w.leong, d.mandic}@imperial.ac.uk

Wei Liu

Communications Research Group
Dept. of Electronic and Electrical Engineering
University of Sheffield, Sheffield S1 3JD
w.liu@sheffield.ac.uk

ABSTRACT

Existing blind source extraction (BSE) methods are limited to noise-free mixtures, which is not realistic. We therefore address this issue and propose an algorithm based on the normalised kurtosis and a nonlinear predictor within the BSE structure, which makes this class of algorithms suitable for noisy environments, a typical situation in practice. Based on a rigorous analysis of the existing BSE methods we also propose a new optimisation paradigm which aims at minimising the normalised mean square prediction error (MSPE). This makes redundant the need for preprocessing or orthogonality transform. Simulation results are provided which confirm the validity of the theoretical results and demonstrate the performance of the derived algorithms in noisy mixing environments.

Index Terms— Blind source separation, blind source extraction, adaptive nonlinear prediction, noisy mixtures

1. INTRODUCTION

Recently, due to its wide potential application in the areas including biomedical engineering, sonar, radar, speech enhancement, telecommunications, blind source separation (BSS) [8] has been studied extensively and has become one of the most important research topics in the signal processing area [3, 4, 2]. This is a technique which aims at recovering the original sources from all kinds of their mixtures, without the knowledge of the mixing process and the sources themselves. In blind source separation process, there are n sources $s_1(k), s_2(k), \ldots, s_n(k)$, which are passed through an unknown mixing system with added noise; by m sensors we acquire the received mixed signals $x_1(k)$, $x_2(k), \ldots, x_m(k)$. With appropriate separation algorithms, the original signals are then separated from their mixtures subject to the ambiguities of permutation and scaling. For instantaneous mixing, the mixtures are modelled as weighted sums of individual sources without dispersion or time delay, given by

$$\mathbf{x}(k) = \mathbf{A} \cdot \mathbf{s}(k) + \mathbf{vn}(k), \tag{1}$$

with $[\mathbf{A}]_{i,j} = a_{i,j},\ i = 1, \ldots, m,\ j = 1, \ldots, n$, $\mathbf{vn}(k)$ is the noise vector and $\mathbf{A}$ is the mixing matrix. We normally assume that the sources are zero-mean and the elements of $\mathbf{vn}(k)$ are white Gaussian and independent of the source signals.

In general, by BSS we obtain all the n sources simultaneously, but we can also choose to extract a single source or a subset of sources from their mixtures and repeat this process until we extract the last source or the last desired one from a subset of sources [7, 1, 11, 9, 10]. The BSS approach operating in this way is also called blind source extraction (BSE) [2]. Compared to the general simultaneous BSS for multiple sources, BSE provides us with more freedom in separation. We can design and employ different algorithms at different stages of the extraction, according to the features of the source signal we want to extract at a particular stage. By extracting only the set of signals of interest, we also save much of the unnecessary computation, especially when the spatial dimension of observed mixtures is large and the number of signals of interest is small.

This paper proposes an improvement on the existing BSE algorithms and provides efficient solutions for BSE of instantaneous noisy mixtures. Based on a rigorous analysis of the normalised mean square prediction error (MSPE) for a linear predictor based BSE method for noisy mixtures [10], we propose a novel higher-order statistical method based on the minimisation of normalised mean square nonlinear prediction error. This approach does not require prior knowledge of the noise variance, unlike methods for BSE of noisy mixtures, based on the removal of noise term directly from the cost function.

2. BLIND SOURCE EXTRACTION FOR NOISY MIXTURES

A general structure of the BSE process for extracting one single source at a time is shown in Fig. 1; where there are two principal stages: extraction and deflation [7]. The original mixtures first undergo the extraction stage to have one source recovered; after deflation, the effects of the extracted source are removed from the mixtures. These new ”deflated” mixtures then undergo the next extraction process to recover the second source; this process repeats until the last source of interest is recovered.

2.1. BSE with a Nonlinear Predictor in Noisy Environments

In a noisy environment, to extract one of the sources, we apply a demixing operation, given by $\mathbf{w}$, which yields

$$y_1(k) = \mathbf{w}_1^T \mathbf{x}_1(k) = \mathbf{g}_1^T \mathbf{s}_1(k) + \mathbf{w}_1^T \mathbf{vn}_1(k) \tag{2}$$

where $\mathbf{g}_1^T = \mathbf{w}_1^T \cdot \mathbf{A}$.

For independent sources, Liu et al. [10] proposed to remove the effect of noise by manipulating the cost function, based on

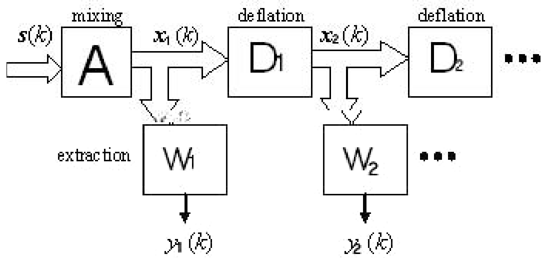

Fig. 1. A general structure of the blind source extraction (BSE).

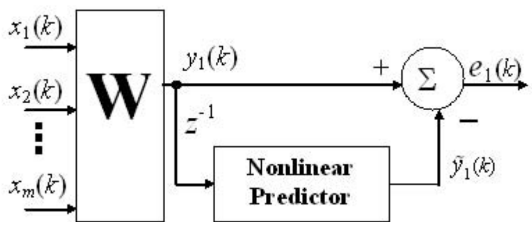

Fig. 2. A structure of the nonlinear predictor.

an estimate of the variance of this noise. The cost function used had the same generic form as that for the noise-free case, but the method required some prior knowledge of the noise variance. More specifically, as the kurtosis of a Gaussian random variable is zero, the kurtosis of an extracted signal, $kt(y_1(k))$ will be the same as in the case with zero noise. Therefore, it is convenient to apply normalised cost function

$$C_1(\mathbf{w}_1) = -\frac{\beta kt(y_1)}{4(E\{y_1^2\})^2} \;, \tag{3}$$

where $\beta = 1$ for the extraction of source signals with positive kurtosis and $\beta = -1$ for sources with negative kurtosis. For a zero-mean random variable y_1, the kurtosis is defined as [3]

$$kt(y_1) = E\{y_1^4\} - 3(E\{y_1^2\})^2 \;, \tag{4}$$

where $E\{\cdot\}$ denotes the statistical expectation operator. Notice however that kurtosis based algorithms are only applicable to independent non-Gaussian sources (or at most one Gaussian). We can therefore work towards relaxing this condition; consider the case with temporally correlated sources, including Gaussian ones. More specifically, assume

$$\begin{aligned} \mathbf{R}_{ss}(0) &= E\{\mathbf{s}(k)\mathbf{s}^T(k)\} \\ &= \mathrm{diag}\{\rho_0(0), \rho_1(0), \ldots, \rho_{n-1}(0)\}, \end{aligned} \tag{5}$$

with $\rho_i(0) = E\{s_i(k) \cdot s_i(k)\}$, $i = 0, 1, \ldots, n-1$, and

$$\begin{aligned} \mathbf{R}_{ss}(\Delta k) &= E\{\mathbf{s}(k)\mathbf{s}^T(k - \Delta k)\} \\ &= \mathrm{diag}\{\rho_0(\Delta k), \rho_1(\Delta k), \ldots, \rho_{n-1}(\Delta k)\} \end{aligned} \tag{6}$$

with $\rho_i(\Delta k) \neq 0$ for some nonzero delay Δk.

To circumvent problems associated with a linear predictor (and associated Gaussianity [6]) within standard BSE structure, following the practice from radar and laser research, one convenient way to deal with the noisy cases would be to employ a nonlinear predictor [5] within the BSE structure as shown in Fig. 2, where the weighted sum $y_1(k) = \mathbf{w}_1^T \cdot \mathbf{x}_1(k)$ is passed through a nonlinear predictor with a length P. In Fig.2, a standard extraction process with extracting coefficients $\mathbf{w}_1(k)$ is used in the first step to extract one signal (denoted by $y_1(k)$) from the mixture $\mathbf{x}_1(k)$. In the next step, a nonlinear adaptive finite impulse response (FIR) filter with coefficients $\mathbf{b}_1(k)$ and nonlinearity Φ is used to assist the extraction. The use of nonlinear predictor is particularly important to support the extraction process in eliminating the effects of the remaining noise [5]. In Fig.2, the filter output $\tilde{y}_1(k)$ is an estimate of the extracted signal $y_1(k)$ and the filter nonlinearity $\Phi(\cdot)$ is typically a sigmoid function. The estimation of the extracted signal $y_1(k)$ is naturally accompanied by a prediction error, defined by

$$\begin{aligned} e_1(k) &= y_1(k) - \tilde{y}_1(k) \\ &= \sum_{i=1}^{m} x_i(k) w_{1i}(k) \\ &\quad - \Phi\left(\sum_{p=1}^{P} b_{1p}(k) \sum_{i=1}^{m} x_i(k-p) w_{1i}(k-p)\right) \end{aligned} \tag{7}$$

where, for convenience, $\Phi(k)$ stands for $\Phi(\mathbf{b}_1^T(k)\mathbf{w}_1(k))$.

To adjust the filter coefficients $\mathbf{b}_1(k) = [b_{11}(k),\ b_{12}(k), \ldots,\ b_{1p}(k)]^T$, tap-delayed output $\mathbf{y}_1(k) = [y_1(k-1),\ y_1(k-2),\ \ldots,\ y_1(k-P)]^T$ and the extracting coefficients $\mathbf{w}_1(k) = [w_{11}(k),\ w_{12}(k),\ \ldots, w_{1m}(k)]^T$, we derive a gradient descent algorithm, which is based on minimisation of the normalised nonlinear prediction error $e_1(k)$. We therefore define the cost function for the BSE based on the structure from Fig.2 in terms of the normalised mean squared prediction error (MSPE) as

$$C_1(\mathbf{w}_1) = \frac{E\{e_1^2(k)\}}{E\{y_1^2(k)\}} \;. \tag{8}$$

where

$$\begin{aligned} E\{y_1^2(k)\} &= \mathbf{w}_1^T \mathbf{R}_{xx}(0) \mathbf{w}_1 \\ &= \mathbf{w}_1^T \mathbf{A} \mathbf{R}_{ss}(0) \mathbf{A}^T \mathbf{w}_1 \end{aligned} \tag{9}$$

Rewriting (7), the MSPE $E\{e_1^2(k)\}$ can be expressed as[1].

$$\begin{aligned} & E\{e_1^2(k)\} \\ =\ & [y_1(k) - \tilde{y}_1(k)]^2 \\ =\ & \left[\sum_{i=1}^{m} x_i(k) w_{1i}(k) - \Phi\Big(\sum_{p=1}^{P} b_{1p}(k) \sum_{i=1}^{m} x_i(k-p) w_{1i}(k-p)\Big)\right]^2 \end{aligned} \tag{10}$$

By minimizing the cost function $C_1(\mathbf{w}_1)$ with respect to the demixing vector $\mathbf{w}_1$, the global demixing vector $\mathbf{g}_1 = \mathbf{w}_1^T \cdot \mathbf{A}$ tends to have only one nonzero element and consequently only the source signal with the smallest normalised MSPE for the nonlinear predictor $\mathbf{b}$ will be extracted [6].

[1] More detail and justification can be found in [5]. Due to the space limitation the full analysis is not presented.

To derive a gradient descent adaptation for every element $b_{1p}(k)$, $p = 1, 2, \ldots, P$ of the filter coefficient vector $\mathbf{b}_1$ and every element $w_{1i}(k)$, $i = 1, 2, \ldots, m$ of the extracting coefficient vector $\mathbf{w}_1$ we have

$$b_{1p}(k+1) = b_{1p}(k) - \mu_b \nabla_{b_{1p}} C_1(\mathbf{w}_1(k), \mathbf{b}_1(k)) \tag{11}$$

where μ_b is the learning rate for the adaptation of $\mathbf{b}_1$.

The updates for the filter and the extracting coefficients now become

$$b_{1p}(k+1) = b_{1p}(k) + \mu_b(k) e_1(k) \Phi'(k) y_1(k-p) \tag{12}$$

which can be expressed in the vector form as

$$\mathbf{b}_1(k+1) = \mathbf{b}_1(k) + \mu_b(k) e_1(k) \Phi'(k) y_1(k) \tag{13}$$

where the $\Phi'(k)$ denotes its derivative at time instant k.

Applying the standard gradient descent method to minimise $C_1(\mathbf{w}_1)$, and using some stochastic approximations, we can obtain the following online update equation [10]

$$\mathbf{w}_1(k+1) = \mathbf{w}_1(k) - \frac{\mu_w}{\sigma_y^2(k)} \left(e_1(k) \hat{\mathbf{x}}_1(k) - \frac{\sigma_e^2(k)}{\sigma_y^2(k)} y_1(k) \mathbf{x}_1(k) \right) , \tag{14}$$

where μ_w is the learning rate, β_e and β_y are the corresponding forgetting factors.

$$\begin{aligned} \sigma_e^2(k) &= \beta_e \sigma_e^2(k-1) + (1-\beta_e) e_1^2(k) , \\ \sigma_y^2(k) &= \beta_y \sigma_y^2(k-1) + (1-\beta_y) y_1^2(k) , \end{aligned} \tag{15}$$

In deflation, the new deflated mixtures become

$$\hat{\mathbf{x}}_1(k) = \mathbf{x}_1(k) - \sum_{p=1}^{P} b_{1p} \mathbf{x}_1(k-p) , \tag{16}$$

This completes the derivation of the proposed BSE algorithm for extracting noisy signals.

2.2. Simulations

Fig. 3(a) shows three source signals, denoted by s_1 with binary distribution, s_2 with Gaussian distribution and s_3 a random waveform, where used in simulations. The signals s_1 and s_2 have positive kurtosis ($\beta = 1$). The length $P = 3$ nonlinear predictor is adopted in this experiment. Monte Carlo simulations with 5000 iterations of independent trials were performed. This way, the normalised prediction errors of the three signals were respectively $\{9.5492, 10.1327, 10.3047\}$. The 3×3 mixing matrix $\mathbf{A}$ was randomly generated and is given by

$$\mathbf{A} = \begin{bmatrix} 0.4974 & -0.1222 & 0.9032 \\ 0.2462 & -0.6966 & 0.6442 \\ 0.7976 & 0.3492 & 0.3445 \end{bmatrix} . \tag{17}$$

To further illustrate the proposed approach, the variance of the noise in (1)was set $\sigma_{vn}^2 = 0.1$. By minimising the normalised MSPE, we expect the signal with the smallest normalised prediction error to be extracted, which is the first signal s_1. The forgetting factors were $\beta_e = \beta_y = 0.1$ and the stepsize $\mu_w = \mu_b = 0.0017$. The learning curve for this case is shown in Fig. 4 for both the proposed nonlinear predictor and the normalised MSPE [10], with the performance index defined as [2]

	signal 1	signal 2	signal 3
Original Signal	1.0006	1.8105	2.7771
Noisy Mixture	2.4427	2.0387	2.4247
MPSE Linear Predictor [10]	2.3986	2.4727	1.6847
Proposed BSE	1.0611	1.0204	2.7663

Table 1. Kurtosis of the original sources and the kurtosis of the extracted signals using the proposed method and the normalised MPSE linear predictor method [10].

$$PI = 10 \log_{10} \left(\frac{1}{n-1} \left(\sum_{m=0}^{n-1} \frac{g_m^2}{\max\{g_0^2, g_1^2, \ldots, g_{n-1}^2\}} - 1 \right) \right) , \tag{18}$$

with $\mathbf{g} = \mathbf{A}^T \mathbf{w} = [g_0 \; g_1 \; \cdots \; g_{n-1}]$. As the performance index reached the level of around -16 dB, we can say the signal s_1 had been extracted successfully.

The waveform of the sequentially extracted signal by the proposed nonlinear predictor method is given in Fig.3(b). Based on the smallest predictor error, the proposed nonlinear predictor first extracted s_1 with binary distribution, followed by s_2 a random waveform and then s_3 with Gaussian distribution. These three extracted matched closely with the original source signals. If, instead of the proposed nonlinear predictor, the standard normalised MSPE [10] approach was used, it was unable to give satisfactory extraction performance, as shown in Fig.3(c). In addition, it can be seen that the proposed blind extraction algorithm provides, in general, better kurtosis matching of source and output signals (Table I).

To further illustrate the qualitative performance of the proposed approach, scatter plots of the original sources and the recovered output signals are displayed in Fig.5. These scatter plots show the degree of independence between the outputs, where each point on the diagram corresponds to one data vector. Conforming with the above results, the extracted output signals using the proposed method outperformed the normalised MSPE [10] based extraction.

3. CONCLUSIONS

We have addressed a special class of blind source separation (BSS) algorithms, namely blind source extraction (BSE), by which we can recover a single source or a subset of sources each time, instead of recovering all of the sources simultaneously. We have studied the BSE problem in noisy environments and proposed a new BSE algorithm based on minimisation of mean square nonlinear prediction error. Unlike the existing algorithms for noisy BSE, which remove the effects of noisy directly from the cost function, this approach does not require the knowledge of noise variance, or any preprocessing. Simulations have shown that the proposed algorithm can perform satisfactory extraction of the corresponding sources from noisy mixtures.

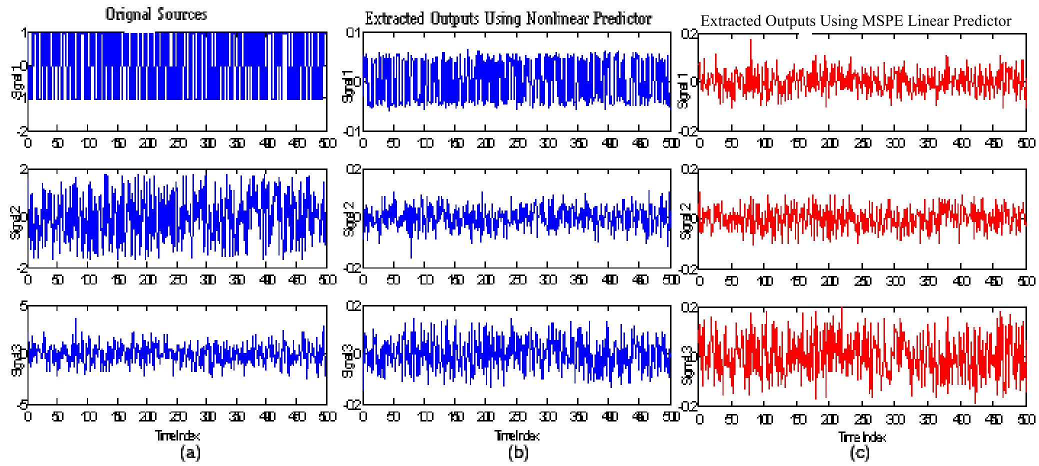

Fig. 3. Source signals used in simulations: (a) The original source signals, s_1 with binary distribution, s_2 a Gaussian distribution and s_3 with random waveform; (b) The extracted output signals based on nonlinear predictor, s_1 with binary distribution, s_2 a random waveform and s_3 with Gaussian distribution; (c) The extracted output signals based on normalised MSPE linear predictor, s_1 with random waveform, s_2 a random waveform and s_3 with Gaussian distribution;

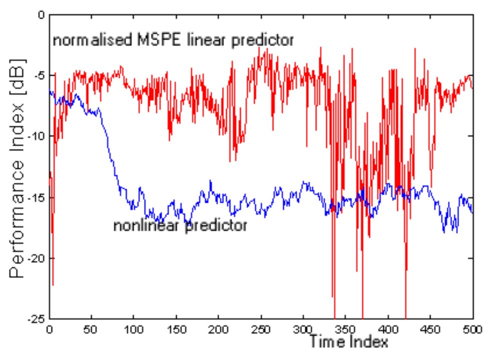

Fig. 4. The performance index using the proposed nonlinear predictor and normalised MPSE linear predictor [10].

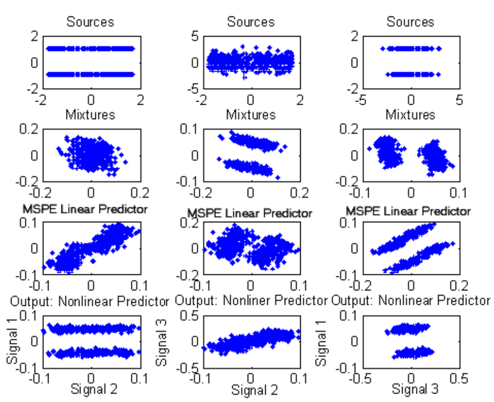

Fig. 5. Scatter plots comparing the independence level of the extracted signals.

4. REFERENCES

[1] A.Cichocki and R.Thawonmas. On-line algorithm for blind signal extraction of arbitrarily distributed, but temporally correlated sources using second order statistics. *Neural Processing Letters*, 12:91–98, 2000.

[2] A.Cichocki and S.I.Amari. *Adaptive Blind Signal and Image Processing*. Locally Adaptive Algorithms for ICA and Their Implementations. John Wiley Sons, Ltd, England, 2002.

[3] A.Hyvarinen, J.Karhunen, and E.Oja. *Independent Component Analysis*. John Wiley & Sons, Inc, Canada, 2001.

[4] S.C. Douglas. Chapter 7: Blind signal separation and blind deconvolution. *Handbook of Neural Network Signal Processing*, 2002.

[5] I.Pitas and A.N.Venetsanopoulos. *Nonlinear Digital Filters Principles and Applications*. Kluwer Academic Publishers, Boston, 1990.

[6] D.P. Mandic and J.Chambers. *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*. Wiley, 2001.

[7] N.Delfosse and P.Loubaton. Adaptive blind separation of independent sources: A deflation approach. *Signal Processing*, 49:59–83, 1995.

[8] P.Comon. Independent Component Analysis, a new concept? *Special Issue on Higher-Order Statistics, Signal Processing*, 36(3):287–314, April 1994.

[9] S.A.Cruces-Alvarez, A.Cichoki, and S.I.Amari. On a new blind signal extraction algorithm: Different criteria and stability analysis. *IEEE Signal Processing Letters*, 9:233–236, August 2002.

[10] W.Liu, D.P.Mandic, and A.Cichocki. Blind source extraction of instantaneous noisy mixtures using a linear predictor. In *The 2006 IEEE International Symposium on Circuits and Systems (ISCAS2006)*, pages 4199–4202, Kos, Greece, May 2006.

[11] Z.Malouche and O.Macchi. Adaptive unsupervised extraction of one component of a linear mixture with a single neuron. *IEEE Transactions on Neural Networks*, 9:123–138, January 1998.