A DUAL-LINEAR PREDICTOR APPROACH TO BLIND SOURCE EXTRACTION FOR NOISY MIXTURES

Wei Liu *, Danilo P. Mandic † and Andrzej Cichocki ‡

* Communications Research Group
Department of Electronic & Electrical Engineering, University of Sheffield, UK
† Communications and Signal Processing Research Group
Department of Electrical & Electronic Engineering, Imperial College London, UK
‡ Laboratory for Advanced Brain Signal Processing
Brain Science Institute, RIKEN, Japan

Abstract: A second-order statistics based dual-linear predictor structure is proposed for blind source extraction from noisy instantaneous mixtures. The noise component is assumed to be spatially and temporally white, but the variance information of noise is not required. A detailed proof of the proposed approach is provided and an adaptive algorithm is developed. Simulation results show that it can extract the source signals successfully.

Index Terms: Blind Source Extraction, Linear Predictor, Noisy Mixtures, Second-order Statistics.

1. INTRODUCTION

Blind source separation (BSS) is one of the most important and established research topics in the signal processing area [1, 2] and many algorithms have been proposed based on different statistical properties of the source signals. In BSS, normally all of the source signals are recovered simultaneously. Alternatively, we can extract only one or a subset of the sources at a time and then remove the extracted signals from the original mixtures (deflation) before performing the next extraction. This lead to the concept of blind source extraction [2].

Algorithms specifically designed for BSE can be roughly divided into two categories: i) those based on higher-order statistics (HOS) [3, 4]; ii) those based on second-order statistics (SOS) [5, 6, 7, 8], which assume that the sources are not correlated with each other and every source has a different temporal structure. The SOS approaches mainly employ an additional linear predictor within the BSE structure. An adaptive algorithm for this structure was proposed in [5], whereas a batch algorithm was derived for a one-step ahead predictor in [6]. In [8], a critical study of this approach is provided and a new, efficient BSE algorithm was derived by minimising the normalised mean square prediction error (MSPE). All of the algorithms were initially derived for the noise-free case. It is however realistic to assume environmental noise in real world applications. An example is the approach proposed in [9], where knowledge of the noise variance in the mixtures is assumed. However, when this information is unavailable or unreliable, the algorithm proposed in [9] will not work effectively.

In this paper, the case with spatially and temporally white noise will be studied and a novel algorithm will be developed based on a dual-linear predictor structure. The advantage of the proposed method is that there is no need to estimate the variance of the noise as in [9].

This paper is organised as follows. In Section 2, the dual-linear predictor structure for BSE will be provided and an adaptive algorithm will be derived with a detailed proof and analysis about its condition on which it can be applied to the BSE problem. Simulation results are shown in Section 3 and conclusions drawn in Section 4.

2. THE DUAL-LINEAR PREDICTOR STRUCTURE

2.1. The Structure

In a general BSS mixing model, the vector of observed mixtures \( \mathbf{x}[n] \) is given by

\[
\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n] + \mathbf{v}[n],
\]

(1)
$s[n] = [s_0[n] s_1[n] \cdots s_{L-1}[n]]^T$
$x[n] = [x_0[n] x_1[n] \cdots x_{M-1}[n]]^T$
$[A]_{m,t} = a_{m,t}, m = 0, \ldots, M-1, l = 0, \ldots, L-1 \ . \ (2)$

$v[n]$ is the additive noise vector.

We assume that the source signals are uncorrelated with each other and their correlation matrices are given by

$R_{ss}[0] = E\{s[n]s^T[n]\} = \text{diag}(\rho_0[0], \rho_1[0], \ldots, \rho_{L-1}[0]), \ (3)$

with $\rho_l[0] = E\{s_l[n] \cdot s_l[n]\}$, $l = 0, 1, \ldots, L - 1$, where $E\{\cdot\}$ denotes the statistical expectation operator, and

$R_{ss}[\Delta n] = E\{s[n]s^T[n-\Delta n]\} = \text{diag}(\rho_0[\Delta n], \rho_1[\Delta n], \ldots, \rho_{L-1}[\Delta n]) \ . \ (4)$

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

The noise component $v[n]$ is assumed to be uncorrelated with the source signals and its correlation matrix is given by

$R_{vv}[\Delta n] = E\{v[n]v^T[n-\Delta n]\} = \begin{cases} 0, & \text{for } \Delta n \neq 0 \\ \sigma_v^2 I, & \text{for } \Delta n = 0 \end{cases}, \ (5)$

where $\sigma_v^2$ is the variance of the noise.

For noise-free mixtures, a linear predictor can be employed to extract one of the sources, as shown in Figure 1, where the extracted signal $y[n]$ and the instantaneous output error $e[n]$ of the linear predictor with a length $P$ are given by

$y[n] = w^T x[n]$
$e[n] = y[n] - b^T y[n] \ , \ (6)$

where $w$ is the demixing vector and

$b = [b_1 b_2 \cdots b_P]^T$
$y[n] = [y[n-1] y[n-2] \cdots y[n-P]]^T \ . \ (7)$

As proved in [8], by minimising $J_0(w)$ with respect to $w$, the sources can be extracted successfully.

However, in the presence of noise, there will be a noise term in both the numerator and the denominator of (8) and the proof in [8] is not valid any more. To remove the effect of noise, we propose to exploit the white nature of the noise components and employ a dual-linear predictor structure as shown in Fig. 2, where a second linear predictor with coefficients vector $d$ of length $P_d$ is employed and the error signal $f[n]$ is given by

$f[n] = y[n] - d^T y_d[n] \ , \ (9)$

where

$y_d[n] = [y[n-1] y[n-2] \cdots y[n-P_d]]^T \ . \ (10)$

For the first linear predictor, the mean square prediction error (MSPE) $E\{e^2[n]\}$ is given by

$E\{e^2[n]\} = E\{y^2[n]\} - 2E\{y[n]b^T y[n]\} + E\{b^T y[n] y^T[n]b\}$
$= \sum_{p=0}^{P} b^2_p E\{y^2[n-p]\} - \sum_{p,q=0; p \neq q}^{P} s_{pq} b_p b_q E\{y[n-p]y[n-q]\}$
$= q_e E\{y^2[n]\} - \sum_{p,q=0; p \neq q}^{P} s_{pq} b_p b_q w^T R_{xx}[q-p]w$
$= q_e E\{y^2[n]\} - w^T \left( \sum_{p,q=0; p \neq q}^{P} s_{pq} b_p b_q R_{xx}[q-p] \right) w \ . \ (11)$
where \( q_c = \sum_{p=0}^{P} b_p^2 \) with \( b_0 = 1 \), and \( s_{pq} \) is 1 when \( p = 0 \) or \( q = 0 \), and \(-1\) otherwise. From (1), (4) and (5), we have

\[
R_{xx}[p - q] = A E\{s[n]s^T[n - (p - q)]\} A^T + E\{v[n]v^T[n - (p - q)]\} \nonumber
\]

\[
= A R_{ss}[p - q] A^T ,
\]

for \( p \neq q \). Then we have

\[
E\{e^2[n]\} = q_c E\{y^2[n]\} - w^T A \left( \sum_{p,q=0;p\neq q}^P s_{pq} b_p b_q R_{ss}[q - p] \right) A^T w \nonumber
\]

\[
= q_c E\{y^2[n]\} - g^T \left( \sum_{p,q=0;p\neq q}^P s_{pq} b_p b_q R_{ss}[q - p] \right) g \nonumber
\]

\[
= q_c E\{y^2[n]\} - g^T \hat{R}_{ss} g ,
\]

with \( g = A^T w \) denoting the global demixing vector and \( \hat{R}_{ss} \) is a diagonal matrix given by

\[
\hat{R}_{ss} = \sum_{p,q=0;p\neq q}^P s_{pq} b_p b_q R_{ss}[q - p] ,
\]

with its \( l - th \) diagonal element \( \hat{r}_l \) given by

\[
\hat{r}_l = \sum_{p,q=0;p\neq q}^P s_{pq} b_p b_q \rho_l[q - p] .
\]

Similarly, for the second linear predictor, we have

\[
E\{f^2[n]\} = a_c E\{y^2[n]\} - 2 E\{y[n]d^T y_d[n]\} + E\{d^T y_d[n] y_d^T[n] d\} \nonumber
\]

\[
= a_c E\{y^2[n]\} - g^T \left( \sum_{p,q=0;p\neq q}^{P_d} s_{pq} d_p d_q \hat{R}_{ss}[q - p] \right) g \nonumber
\]

\[
= a_c E\{y^2[n]\} - g^T \hat{R}_{ss} g ,
\]

with \( a_c = \sum_{p=0}^{P_d} d_p^2 \) with \( d_0 = 1 \) and \( \hat{R}_{ss} \) is a diagonal matrix given by

\[
\hat{R}_{ss} = \sum_{p,q=0;p\neq q}^{P_d} s_{pq} d_p d_q \hat{R}_{ss}[q - p] ,
\]

with its \( l - th \) diagonal element \( \hat{r}_l \) given by

\[
\hat{r}_l = \sum_{p,q=0;p\neq q}^{P_d} s_{pq} d_p d_q \rho_l[q - p] .
\]

### 2.2. The Proposed Cost Function

Note in the second term of both (13) and (16), there is not any noise component. Then we can construct a new cost function as follows

\[
J(w) = \frac{q_c E\{y^2[n]\} - E\{c^2[n]\}}{a_c E\{y^2[n]\} - E\{f^2[n]\}} = \frac{g^T \hat{R}_{ss} g}{g^T \hat{R}_{ss} g} .
\]

Now we impose another condition on the second linear predictor: suppose the coefficients \( d \) are chosen in such a way that all of the diagonal elements of \( \hat{R}_{ss} \) are of positive value. This is a difficult condition due to the blind nature of the problem. However, for a special case with \( P_d = 1 \) and \( d_1 = 1 \), i.e. a one step ahead predictor, we have

\[
\hat{R}_{ss} = 2 \hat{R}_{ss}[1] ,
\]

which is the correlation matrix of the source signals with a time lag of 1. Then the condition means each of the source signals should have a positive correlation with a delayed version of itself by lag 1. In reality, there are many signals having this correlation property and therefore can meet this requirement.

Since all of the diagonal elements of \( \hat{R}_{ss} \) are positive, we shall assume \( \hat{R}_{ss} = I \), i.e. \( \hat{r}_l = 1 \), \( l = 0, 1, \ldots , L - 1 \), as the differences in the diagonal elements can always be absorbed into the mixing matrix \( A \). This way, the diagonal elements \( \hat{r}_l \), \( l = 0, 1, \ldots , L - 1 \), of \( \hat{R}_{ss} \) in the numerator become the “normalised” autocorrelation values of each source signal and they are assumed to be different from each other. For the case with \( P_d = 1 \) and \( d_1 = 1 \), the “normalisation” here is not by \( E\{s_i^2[n]\} \), but by \( \hat{r}_l = 2 E\{s_i[n]s_i[n - 1]\} \).

Now we have

\[
J(w) = \frac{g^T \hat{R}_{ss} g}{\sqrt{g^T g}} ,
\]

where \( g = \frac{\tilde{g}}{\sqrt{g^T \tilde{g}}} \), which has a property \( g^T \tilde{g} = 1 \).

Now we have reached a similar cost function as the one given in [8, 9]. According to the proof provided there, we can draw the conclusion that when we minimize \( J(w) \) with respect to \( w \), this will result in a successful extraction of the source signal with the minimum “normalised” 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 [2]. This procedure is repeated until the last source signal is recovered.
2.3. Adaptive Algorithm

Applying the standard gradient descent method to $J(w)$, we have

\[ \nabla_w J = \frac{2}{(a_c E \{ y^2[n] \} - E \{ f^2[n] \})^2} (q_c E \{ y[n]x[n] \} - E \{ e[n] \hat{x}[n] \})(a_c E \{ y^2[n] \} - E \{ f^2[n] \}) - (q_c E \{ y^2[n] \} - E \{ e^2[n] \})(a_c E \{ y[n]x[n] \} - E \{ f[n] \hat{x}[n] \}) , \]

where

\[ \hat{x}[n] = x[n] - \sum_{p=1}^{P} b_p x[n-p] \]
\[ \hat{x}[n] = x[n] - \sum_{p=1}^{P_0} d_p x[n-p] . \] (23)

$E \{ e^2[n] \}$, $E \{ y^2[n] \}$ and $E \{ f^2[n] \}$ can be estimated respectively by

\[ \sigma_e[n] = \beta_e \sigma_e[n-1] + (1 - \beta_e)e^2[n] , \]
\[ \sigma_y[n] = \beta_y \sigma_y[n-1] + (1 - \beta_y)y^2[n] , \]
\[ \sigma_f[n] = \beta_f \sigma_f[n-1] + (1 - \beta_f)f^2[n] , \] (24)

where $\beta_e$, $\beta_y$ and $\beta_f$ are the corresponding forgetting factors with $0 \leq \beta_e, \beta_y, \beta_f < 1$.

Following standard stochastic approximation techniques [10], we obtain the following online update for $w[n]$

\[ w[n+1] = w[n] - \frac{2\mu}{(a_c \sigma_y - \sigma_f)^2} ((q_c y[n]x[n]) - (q_c \sigma_y - \sigma_e) \cdot (a_c \sigma_y - \sigma_f) - (q_c \sigma_y - \sigma_e) \cdot (a_c \sigma_y - \sigma_f)) . \] (25)

where $\mu$ is the learning rate. For the case with with $P_d = 1$ and $d_1 = 1$, we have $a_c = 2$ in (25), which will be used in our simulations.

3. SIMULATIONS

In the simulations, three source signals are used which are generated by passing three randomly generated white Gaussian signals through three different filters. The power of the sources is normalised to one. The correlation value of each of the source signals is checked to make sure it is positive and not close to zero for one sample shift. Fig. 3 shows the three source signals, denoted by $s_0$, $s_1$ and $s_2$, respectively.

The coefficients of the first linear predictor coefficients $b$ were randomly generated with a length of $P = 5$, and

\[ b = [-0.4548 \ -0.0053 \ 1.1957 \ -0.5590 \ -0.3617] . \] (26)

For the second linear predictor, $P_d = 1$, $d_1 = 1$, and $a_c = 2$.

The normalised correlation value $r_1$ for each source signal with this dual-linear predictor configuration is 0.0395, 0.2174 and 0.7949, respectively. As already proved, since the first source signal has the smallest correlation value of 0.0395, it will be extracted by minimizing the cost function.

The $3 \times 3$ mixing matrix $A$ is randomly generated and given by

\[ A = \begin{bmatrix} 0.9207 & 0.0299 & 0.3891 \\ 0.5165 & 0.3676 & 0.7733 \\ 0.7822 & -0.2735 & -0.5598 \end{bmatrix} . \] (27)

Its row vector is normalised to unity to make sure it is comparable to the noise variance, which is $\sigma_n^2 = 0.09$. The forgetting factors is $\beta_e = \beta_y = \beta_f = 0.975$ and the step-size $\mu = 0.0015$. A learning curve for this case is shown in Fig. 4, with the performance index defined as [2]

\[ PI = 10 \log_{10} \left( \frac{1}{L-1} \sum_{l=0}^{L-1} \max \{ g_l^2, \ldots, g_{L-1}^2 \} \right) , \] (28)

with $g = [g_0 \ g_1 \ \cdots \ g_{L-1}]$.

To show its performance in a more general context, we change the initial value of the demixing vector $w$ randomly each time to run the algorithm and the average learning curve over 1000 runs is given in Fig. 5. Both curves show a successful extraction of the source signal.
4. CONCLUSIONS

A dual-linear predictor structure has been proposed to blindly extract the source signals from their noisy mixtures. Due to the white nature of the added noise, its effect is removed implicitly in the cost function. Therefore the resultant adaptive algorithm can recover the corresponding source signals successfully without estimating the variance of noise, which has been illustrated by our simulations.

5. REFERENCES


