A NORMALISED KURTOSIS BASED BLIND SOURCE EXTRACTION ALGORITHM FOR NOISY MIXTURES

Wei Liu † *and Danilo P. Mandic* ‡

† Communications Research Group
Department of Electronic & Electrical Engineering
University of Sheffield, UK
wei.liu@sheffield.ac.uk

‡ Communications and Signal Processing Research Group
Department of Electrical & Electronic Engineering
Imperial College London, UK
d.mandic@imperial.ac.uk

ABSTRACT

We introduce an algorithm for blind source extraction (BSE) of independent sources in the presence of noise, without the need for initial prewhitening, for which the normalised kurtosis is used within the cost function. Unlike the previously proposed methods designed for noise-free mixtures, which is not realistic in practical applications, we address BSE for noisy mixtures and propose a novel cost function which caters for the effects of noise. The proposed method is justified by rigorous analysis and supported by simulations.

1. INTRODUCTION

For blind source separation, solutions based on different measurements of non-Gaussianity have been well established in the past decade [1, 2]. One of them is the kurtosis-based sequential blind source extraction (BSE) algorithm [3, 4, 5] designed for independent sources. There are two kinds of cost functions within this framework. One is based on the direct minimisation/maximisation of the kurtosis, which is meaningful only if the variance of the extracted source signal is bounded. This can be achieved by preprocessing of the data in the form of prewhitening and normalisation of the update of the demixing vector. Another kind of cost function is based on the normalised kurtosis [4], the advantage of which is that, we do not need to perform the otherwise required prewhitening and the normalisation operation, which makes this approach suitable for online applications.

A vast majority of the previous work in the area of blind source separation has relied on the underlying assumption of no additive noise. As for the kurtosis-related BSE methods, although the case with noisy measurements has been discussed when using kurtosis as a cost function [1], such an approach has not been introduced for the normalised kurtosis case. There is an important difference between the kurtosis and normalised kurtosis based methods. When we use kurtosis as a cost function, the effect of Gaussian noise can be removed in the prewhitening stage, however, for the normalised kurtosis case, we need to consider the effect of noise within the cost function, which is fundamentally different.

To that cause, we propose a new cost function and derive the corresponding adaptive algorithm for the extraction of independent sources from their noisy mixtures. A detailed proof of the existence of the solution for the normalised kurtosis method in the context of noisy measurements is also provided.

This paper is organized as follows. In Section 2, with an analysis of the normalised kurtosis in the case of noisy measurements, we provide the new cost function which accounts for the effect of noise, followed by the detailed proof. An adaptive algorithm derived by maximizing the cost function is given in Section 3. Simulation results are given in Section 4 and conclusions drawn in Section 5.

2. THE PROPOSED COST FUNCTION

The vector of the observed mixtures $\mathbf{x}[n]$ can be expressed as

$$\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n] + \mathbf{v}[n], \tag{1}$$

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

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

In this scenario, we have M sensors and L sources, and assume that the noise is i.i.d. white Gaussian and independent of the source signals.

To extract one of the independent sources, a demixing vector $\mathbf{w}$ is applied to $\mathbf{x}[n]$ to give the extracted signal $y[n]$

$$y[n] = \mathbf{w}^T \mathbf{x}[n] = \mathbf{g}^T \mathbf{s}[n] + \mathbf{w}^T \mathbf{v}[n] \ , \tag{3}$$

where

$$\mathbf{g}^T = \mathbf{w}^T \mathbf{A} = [g_0 \ g_1 \ \cdots \ g_{L-1}] \tag{4}$$

is the global demixing vector.

Recall that, by definition, the kurtosis of $y[n]$ is given by

$$kt(y) = E\{y^4\} - 3(E\{y^2\})^2 \ . \tag{5}$$

As the kurtosis of a Gaussian variable is zero, kurtosis $kt(y)$ remains unaltered in the presence of additive white Gaussian noise, and can be expressed as

$$kt(y) = \Sigma_{l=0}^{L-1} g_l^4 \ kt(s_l) = \tilde{\mathbf{g}}^T \mathbf{K}_s \tilde{\mathbf{g}} \ , \tag{6}$$

where

$$\begin{aligned}
\tilde{\mathbf{g}}^T &= [g_0^2 \ g_1^2 \ \cdots \ g_{L-1}^2] \\
\mathbf{K}_s &= \mathrm{diag}\{kt(s_0) \ kt(s_1) \ \cdots \ kt(s_{L-1})\} \ . \quad (7)
\end{aligned}$$

The normalised kurtosis is obtained when kurtosis $kt(y)$ is divided by the square of the variance $E\{y^2\}$, and is given by

$$(E\{y^2\})^2 = (\mathbf{g}^T \mathbf{R}_s \mathbf{g} + \mathbf{w}^T \mathbf{R}_v \mathbf{w})^2 \ , \tag{8}$$

where $\mathbf{R}_s = E\{\mathbf{s}[n]\mathbf{s}^T[n]\}$ is the diagonal correlation matrix of the sources and $\mathbf{R}_v = E\{\mathbf{v}[n]\mathbf{v}^T[n]\}$ is the correlation matrix of the noise. As the differences in the diagonal elements of $\mathbf{R}_s$ can be absorbed into the mixing matrix $\mathbf{A}$, we can always assume $\mathbf{R}_s = \mathbf{I}$. Thus, equation (8) simplifies into

$$(E\{y^2\})^2 = (\mathbf{g}^T \mathbf{g} + \mathbf{w}^T \mathbf{R}_v \mathbf{w})^2 \ . \tag{9}$$

Notice that due to the noise term in $(E\{y^2\})^2$, we cannot use the normalised kurtosis as the cost function in the same way as in [4]. Instead, we need to remove this noise term before performing kurtosis normalisation.

If the signal to be extracted has a positive kurtosis, we perform maximization of the modified cost function

$$\begin{aligned}
J(\mathbf{w}) &= \frac{kt(y)}{(E\{y^2\} - \mathbf{w}^T \mathbf{R}_v \mathbf{w})^2} \\
&= \frac{\tilde{\mathbf{g}}^T \mathbf{K}_s \tilde{\mathbf{g}}}{(\mathbf{g}^T \mathbf{g})^2} = \hat{\mathbf{g}}^T \mathbf{K}_s \hat{\mathbf{g}} \ , \quad (10)
\end{aligned}$$

where

$$\hat{\mathbf{g}}^T = \frac{1}{g_0^2 + g_1^2 + \cdots + g_{L-1}^2} [g_0^2 \ \ g_1^2 \ \cdots \ g_{L-1}^2] \ . \tag{11}$$

Otherwise, if the kurtosis of the signal we want to extract is negative, we can simply change the sign of $J(\mathbf{w})$, which will be absorbed eventually into the kurtosis matrix $\mathbf{K}_s$. In this case, the new cost function $J(\mathbf{w})$ becomes $\hat{\mathbf{g}}^T(-\mathbf{K}_s)\hat{\mathbf{g}}$, and in the modified kurtosis matrix $\hat{\mathbf{K}}_s = -\mathbf{K}_s$, the diagonal element corresponding to the interesting signal remains positive. Therefore, without loss of generality, we shall only consider the case with the positive kurtosis.

Suppose the source signal with the largest kurtosis is the $k-th$ source signal $s_k[n]$. From (11), we have

$$\| \hat{\mathbf{g}} \|_2^2 = \hat{\mathbf{g}}^T \hat{\mathbf{g}} \leq 1 \ , \tag{12}$$

where $\| \hat{\mathbf{g}} \|_2^2 = 1$ only if one of the g_l, $l = 0, \ldots, L-1$ is nonzero and the remaining ones are zero.

Consider now the following maximization problem for a fixed positive value of $c \leq 1$ and a general vector $\bar{\mathbf{g}}$

$$\max_{\bar{\mathbf{g}}} \hat{J}(\mathbf{w}) \qquad \text{subject to} \qquad \bar{\mathbf{g}}^T \bar{\mathbf{g}} = c^2 \ , \tag{13}$$

where

$$\hat{J}(\mathbf{w}) = \bar{\mathbf{g}}^T \mathbf{K}_s \bar{\mathbf{g}} \ . \tag{14}$$

It can be proved that in general the solution to this problem is a vector $\bar{\mathbf{g}}_{opt}$ with only one nonzero element strictly equal to $\pm c$ at the position corresponding to the largest diagonal element (kurtosis) $kt(s_k)$ of the matrix $\mathbf{K}_s$ [2]. With $\bar{\mathbf{g}} = \bar{\mathbf{g}}_{opt}$, we have $\hat{J}(\mathbf{w}) = \bar{\mathbf{g}}^T \mathbf{K}_s \bar{\mathbf{g}} = c^2 kt(s_k)$. As c increases, $\hat{J}(\mathbf{w})$ will increase correspondingly, and when $c = 1$, we have the maximum value $\hat{J}(\mathbf{w}) = kt(s_k)$.

Then, the maximization problem given by

$$\max_{\hat{\mathbf{g}}} \hat{\mathbf{g}}^T \mathbf{K}_s \hat{\mathbf{g}} \tag{15}$$

is obviously equivalent to searching for the maximum value of $\hat{J}(\mathbf{w})$ in a subspace of $\bar{\mathbf{g}}$ for $c \leq 1$ defined in (11). Observe that the maximum value of $J(\mathbf{w}) = \hat{\mathbf{g}}^T \mathbf{K}_s \hat{\mathbf{g}}$ cannot be larger than that of $\hat{J}(\mathbf{w})$ for $c \leq 1$. Furthermore, the maximum value of $J(\mathbf{w})$ is equal to that of $\hat{J}(\mathbf{w})$ for $c \leq 1$ only if we can find such a $\hat{\mathbf{g}}$ that both the requirements $\hat{\mathbf{g}}^T \hat{\mathbf{g}} = 1$ and $\hat{\mathbf{g}} = \bar{\mathbf{g}}_{opt} \mid_{c=1}$ can be satisfied simultaneously. In fact, from (11), we can see that, when $g_k^2 = \alpha^2$ ($\alpha > 0$) and $g_l = 0$, for all $l \neq k$, the norm of $\hat{\mathbf{g}}$ is exactly equal to unity and also we have $\hat{\mathbf{g}} = \bar{\mathbf{g}}_{opt} \mid_{c=1}$ satisfied. Moreover, this is also the only choice which satisfies both of the requirements.

When $\hat{\mathbf{g}} = \bar{\mathbf{g}}_{opt} \mid_{c=1}$, the corresponding $\mathbf{g}$ will be a vector $\mathbf{g}_{opt}$ with only one nonzero element $g_k = \pm\alpha$. In this case, $y[n] = \pm\alpha s_k[n]$, that is, the desired signal has been extracted. Since we are actually maximizing $J(\mathbf{w})$ with respect to $\mathbf{w}$, instead of $\hat{\mathbf{g}}$, to validate the approach, we need to prove that there exists a $\mathbf{w}_{opt}$ which results in $\mathbf{g}_{opt}$.

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

$$\mathbf{w}_{opt} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{g}_{opt} \ . \tag{16}$$

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

3. THE PROPOSED ADAPTIVE ALGORITHM

To derive an adaptive BSE algorithm based on (10), we need the knowledge of the correlation matrix $\mathbf{R}_v$, which is normally unavailable. For our analysis, in many cases it is reasonable to adopt $\mathbf{R}_v = \sigma^2\mathbf{I}$. When $M > L$, the parameter σ^2 represents the smallest eigenvalue of the correlation matrix $\mathbf{R}_x$ of the observed signals. A subspace method or an adaptive principal component analysis algorithm can be used for its estimation [1, 6]. For $M = L$, it is difficult to estimate σ, unless we have some additional knowledge about the system, for example, the period when there are no source signals present, so that the correlation matrix of the noise can be calculated using those measurements.

For convenience, after each update, we perform a normalisation of the demixing vector $\mathbf{w}$, given by

$$\mathbf{w}[n] \leftarrow \mathbf{w}[n]/\sqrt{\mathbf{w}^T[n]\mathbf{w}[n]} \ . \tag{17}$$

Thus, the cost function (10) changes into

$$J(\mathbf{w}) = \frac{E\{y^4\} - 3(E\{y^2\})^2}{(E\{y^2\} - \sigma^2)^2} \ , \tag{18}$$

To derive the updates of the demixing vector $\mathbf{w}$, we apply the standard gradient descent method to $J(\mathbf{w})$ and obtain

$$\begin{aligned} \nabla_{\mathbf{w}} J \ &= \ \frac{4}{(E\{y^2\} - \sigma^2)^3} \big((E\{y^2\} - \sigma^2)(E\{y^3\mathbf{x}\} \\ &\quad +3\sigma^2 E\{y\mathbf{x}\}) - (E\{y^4\} - 3\sigma^4)E\{y\mathbf{x}\}\big) \ . \end{aligned} \tag{19}$$

After some standard statistical approximations, we arrive at the following update equation

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \mu\phi(y[n])\mathbf{x}[n] \ , \tag{20}$$

where μ is the stepsize and

$$\begin{aligned} \phi(y[n]) \ &= \ \frac{\beta y[n]}{(m_2(y) - \sigma^2)^3} \big[(m_2(y) - \sigma^2)y^2[n] \\ &\quad +3\sigma^2 m_2(y) - m_4(y))\big] \ . \end{aligned} \tag{21}$$

The moments $m_q(y), q = 2, 4$ are estimated recursively by

$$m_q(y)[n] = (1-\lambda)m_q(y)[n-1] + \lambda|y[n]|^q \ , \tag{22}$$

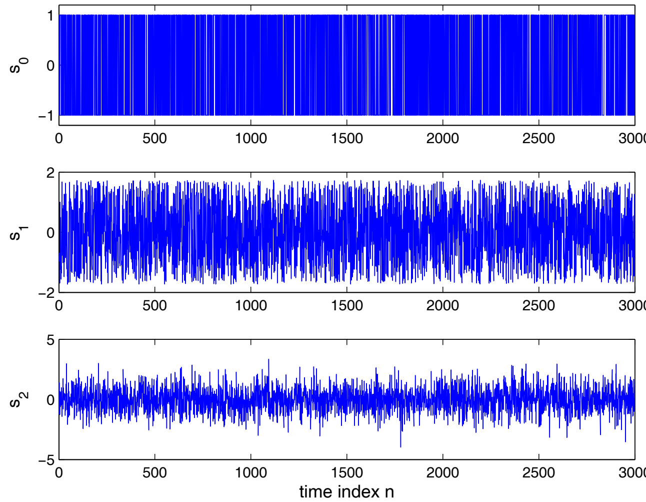

Fig. 1: The three source signals with binary (s_0), uniform (s_1) and Gaussian (s_2) distributions.

where λ denotes the forgetting factor and $\beta = 1$ for source extraction with positive kurtosis and $\beta = -1$ for negative kurtosis.

Note that in the noise-free case ($\sigma^2 = 0$), the expression (21) becomes

$$\phi(y[n]) = \frac{\beta y[n]}{m_2^3(y)}\left(m_2(y)y^2[n] - m_4(y)\right) \ , \tag{23}$$

which is exactly the algorithm proposed in [4] as desired, except for the constant $\frac{1}{m_2^2(y)}$, which can be absorbed into the stepsize.

4. SIMULATIONS AND RESULTS

The simulations were based on three source signals with the binary, uniform and Gaussian distribution, respectively, as shown in Fig. 1. Their corresponding kurtosis values were -1.9978, -1.2237, and 0.0151. Since the two non-Gaussian signals have negative kurtosis values, we have $\beta = -1$, and by minimizing the kurtosis of the extracted signal, in theory, we will extract signal s_0, as it has the smallest kurtosis value.

The 4×3 mixing matrix was randomly generated and is given by

$$\mathbf{A} = \begin{bmatrix} 0.9575 & 0.5207 & 0.9248 \\ -0.9356 & -0.4131 & 0.6338 \\ -0.4264 & -0.9840 & 0.6787 \\ 0.5103 & 0.3774 & -0.1707 \end{bmatrix} . \tag{24}$$

To measure the quality of BSE for the presented algo-

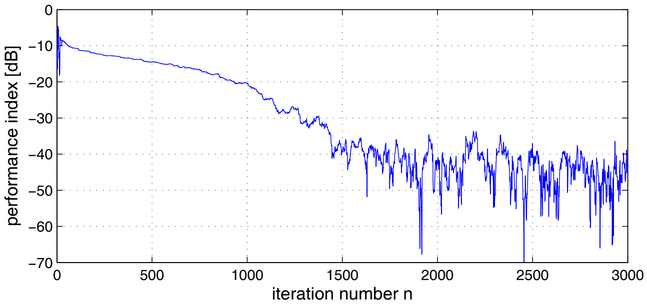

Fig. 2: The performance index for the proposed algorithm.

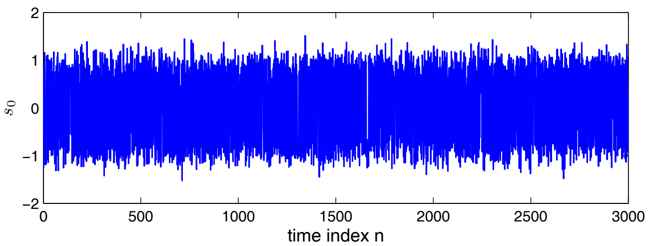

Fig. 3: The extracted source signal.

rithm, we employ the performance index (PI) defined as [2]

$$PI = 10\log_{10}\left(\frac{1}{L-1}\left(\sum_{l=0}^{L-1}\frac{g_l^2}{\max\{g_0^2, g_1^2, \ldots, g_{L-1}^2\}} - 1\right)\right) . \tag{25}$$

The smaller the value of PI, the better the quality of extraction.

The additive noise was white Gaussian with $\sigma = 0.1$. As we have one more mixture than the number of sources, we use this degree of freedom to estimate the noise variance σ^2. During the adaptation, the forgetting factor was $\lambda = 0.02$ and the stepsize $\mu = 0.004$. As shown in Fig. 2, the performance index reached a level of almost -40 dB, indicating a successful extraction. The waveform of the extracted signal is given in Fig. 3, which is actually the first source signal except for the effect of noise.

Next, the steady-state PI value of the proposed algorithm was compared to that of the existing noise-free algorithm given in (23) for different signal-to-noise ratios (SNRs) and the results are shown in Table. 1. Since the proposed algorithm takes the same form as the existing noise-free one for $\sigma = 0$, they have a very similar performance for very low noise levels, as shown in Table. 1. When the noise level increases, the proposed approach outperforms the existing one. For very high noise levels, the performance difference between the two algorithms becomes smaller and effectively they both fail to extract the source signals when the signal and noise power are identical ($SNR = 0.0$ dB).

Tab. 1: The steady-state PI value of the two algorithms with respect to different SNRs:

SNRs [dB]	proposed alg. [dB]	existing alg. [dB]
35.2	-51.7	-51.2
28.3	-44.6	-43.4
25.0	-40.9	-39.3
21.4	-37.2	-34.9
17.9	-33.5	-30.4
14.5	-30.0	-25.9
11.0	-26.0	-21.4
7.6	-22.1	-17.4
4.1	-17.5	-14.0
0.0	-10.4	-10.6

5. CONCLUSIONS

We have proposed a novel blind source extraction algorithm for noisy measurements based on the maximisation/minimisation of the normalised kurtosis. The effect of noise is removed from the previously proposed cost function provided we know the correlation matrix of the noise. A proof of this method is provided and the derived adaptive algorithm is verified by simulations.

6. REFERENCES

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

[2] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*, John Wiley & Sons, Inc., New York, 2003.

[3] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach," *Signal Processing*, vol. 95, pp. 59–83, 1995.

[4] A. Cichocki, R. Thawonmas, and S. Amari, "Sequential blind signal extraction in order specified by stochastics properties," *IEE Electronics Letters*, vol. 33, pp. 64–65, January 1997.

[5] A. Hyvarinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, pp. 1483–1492, 1997.

[6] H. L. Van Trees, *Optimum Array Processing, Part IV of Detection, Estimation, and Modulation Theory*, John Wiley & Sons, Inc., New York, U.S.A., 2002.