Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Signal Processing 86 (2006) 1580–1585

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

# A normalised kurtosis-based algorithm for blind source extraction from noisy measurements

Wei Liu[a,*], Danilo P. Mandic[b]

[a]Communications Research Group, Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield S1 3JD, UK
[b]Communications and Signal Processing Research Group, Department of Electrical and Electronic Engineering, Imperial College London, UK

Received 12 December 2004; received in revised form 24 July 2005
Available online 23 September 2005

**Abstract**

In blind extraction of independent sources, the normalised Kurtosis is a normally used cost function for the cases without the initial prewhitening. The applications of this method are, however, limited to noise-free mixtures, which is not realistic. We therefore address this issue and propose a new cost function based on the normalised Kurtosis, which makes this class of algorithms suitable for noisy environments, a typical situation in practice. The proposed method is justified by a theoretical analysis and the performance of the derived algorithm is demonstrated by simulations.


*Keywords:* Blind source extraction; Noisy measurements; Kurtosis

## 1. Introduction

In blind source separation (BSS), solutions based on measurements of non-Gaussianity are well established and understood [1,2]. One important class of these methods is the kurtosis-based sequential blind source extraction (BSE) [3–5], which has been derived for independent sources. Two types of cost functions exist in this setting. The first one rests on a direct minimisation/maximisation of the kurtosis, which is meaningful only if the variance of the extracted source signal is bounded. This is usually achieved by first preprocessing the data by means of prewhitening and then normalising the demixing vector. The other type 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 weight normalisation operations. This has an advantage that an online application of this algorithm is easier to implement and exhibits more reliable performance.

Despite of being theoretically well justified, a major problem with the existing approaches is that, a vast majority of the previous research in the area of BSS has been conducted with the assumption of no additive noise. However, noisy measurements are common in practice and this needs to be taken into account when designing BSS algorithms.

As for the kurtosis-related BSE methods, the case with noisy measurements has been discussed when using kurtosis as a cost function [1], where the effect of Gaussian noise can be removed in the prewhitening stage. However, solutions for cases with noise have not been proposed for the normalised

*Corresponding author. Fax: +44 11 4222 5834.
*E-mail addresses:* w.liu@sheffield.ac.uk (W. Liu), d.mandic@imperial.ac.uk (D.P. Mandic).

0165-1684/$ - see front matter 
doi:10.1016/j.sigpro.2005.09.001

kurtosis-based algorithms, where, since there is no preprocessing stage, we need to consider the effect of noise directly within the cost function. In this paper, we propose a new cost function with a rigorous proof, which is different from that given in [4] for the noise-free case. An adaptive algorithm is derived correspondingly and its successful performance is demonstrated by simulations.

## 2. The proposed cost function

In the BSS/BSE setting, the vector of observed mixtures $\mathbf{x}[n]$ is generally given by

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

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

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

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

To extract one of the independent sources, we apply a demixing operation given by vector $\mathbf{w}$ to the mixtures $\mathbf{x}[n]$, which yields the extracted source $y[n]$, given by

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

where

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

denotes the global demixing vector, which is related to the quality of the performance of the BSE algorithm.

### 2.1. The normalised kurtosis-based cost function for noisy mixtures

By definition, the kurtosis of $y[n]$ is given by [1]

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

Since the kurtosis of a Gaussian random variable is zero, from (3), the kurtosis $\mathrm{kt}(y)$ has the same value as in the case with zero noise and can be expressed as

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

where

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

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

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

where $\mathbf{R}_s$ is the diagonal correlation matrix of the sources and $\mathbf{R}_v$ 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 $\mathbf{R}_s = \mathbf{I}$ [2]. Thus, Eq. (8) becomes

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

Note that, due to the noise term $\mathbf{w}^{\mathrm{T}}\mathbf{R}_v\mathbf{w}$ 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 somehow remove the noise term before normalising the kurtosis. To achieve this, we propose a new cost function.

If the signal we want to extract has a positive kurtosis, this new cost function we need to maximise can be expressed as

$$J(\mathbf{w}) = \frac{\mathrm{kt}(y)}{(E\{y^2\} - \mathbf{w}^{\mathrm{T}}\mathbf{R}_v\mathbf{w})^2} = \frac{\tilde{\mathbf{g}}^{\mathrm{T}}\mathbf{K}_s\tilde{\mathbf{g}}}{(\mathbf{g}^{\mathrm{T}}\mathbf{g})^2} = \hat{\mathbf{g}}^{\mathrm{T}}\mathbf{K}_s\hat{\mathbf{g}}, \tag{10}$$

where

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

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 $\hat{\mathbf{g}}^{\mathrm{T}}(-\mathbf{K}_s)\hat{\mathbf{g}}$. 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, in the sequel we shall only consider the case with the positive kurtosis.

### 2.2. The optimisation problem for BSE of noisy mixtures

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

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

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

Before we consider the optimisation problem of our BSE, we first consider the following general maximisation problem with respect to a vector $\bar{\mathbf{g}}$ for a fixed positive value of $c \leqslant 1$

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

where the cost function $\hat{J}(\mathbf{w})$ is defined as

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

It is not difficult to prove that, in general, the solution for (13) is a vector $\bar{\mathbf{g}}_{\text{opt}}$ with only one nonzero element strictly equal to $\pm c$ at the position corresponding to the largest diagonal element (kurtosis) $\text{kt}(s_k)$ of the matrix $\mathbf{K}_s$ [2]. When $\bar{\mathbf{g}} = \bar{\mathbf{g}}_{\text{opt}}$, we have

$$\hat{J}(\mathbf{w}) = \bar{\mathbf{g}}^{\mathrm{T}}\mathbf{K}_s\bar{\mathbf{g}} = c^2\text{kt}(s_k). \tag{15}$$

As $c$ increases, $\hat{J}(\mathbf{w})$ will increase correspondingly, and when $c = 1$, we have the maximum value of $\hat{J}(\mathbf{w}) = \text{kt}(s_k)$.

With these results in mind, now consider the following maximisation problem:

$$\max_{\hat{\mathbf{g}}} \hat{\mathbf{g}}^{\mathrm{T}}\mathbf{K}_s\hat{\mathbf{g}} \quad \text{subject to } \hat{\mathbf{g}} \text{ given in (11).} \tag{16}$$

Clearly, this is equivalent to searching for the maximum value of $\hat{J}(\mathbf{w})$ in a subspace of the $\bar{\mathbf{g}}$ for $c \leqslant 1$ defined in (11). Therefore, the maximum value of $J(\mathbf{w})$ cannot be larger than that of $\hat{J}(\mathbf{w})$ for $c \leqslant 1$. Furthermore, the maximum value of $J(\mathbf{w})$ is equal to that of $\hat{J}(\mathbf{w})$ for $c \leqslant 1$ only if there exists such a $\hat{\mathbf{g}}$ so that both the requirements $\hat{\mathbf{g}}^{\mathrm{T}}\hat{\mathbf{g}} = 1$ and $\hat{\mathbf{g}} = \bar{\mathbf{g}}_{\text{opt}}|_{c=1}$ can be satisfied simultaneously. In fact, from (11), it can be seen 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 equal to unity and the condition $\hat{\mathbf{g}} = \bar{\mathbf{g}}_{\text{opt}}|_{c=1}$ is also satisfied. In addition, this is the *only* choice which satisfies both of the above requirements.

When $\hat{\mathbf{g}} = \bar{\mathbf{g}}_{\text{opt}}|_{c=1}$, the corresponding global demixing vector $\mathbf{g}$ will be a vector $\mathbf{g}_{\text{opt}}$ 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 maximising $J(\mathbf{w})$ with respect to $\mathbf{w}$, instead of $\hat{\mathbf{g}}$, we next need to prove that there exists a $\mathbf{w}_{\text{opt}}$ which results in $\mathbf{g}_{\text{opt}}$.

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

$$\mathbf{w}_{\text{opt}} = \mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{g}_{\text{opt}}. \tag{17}$$

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

## 3. The adaptive algorithm for BSE in the presence of noise

To derive the adaptive algorithm based on (10), we need to know the correlation matrix $\mathbf{R}_v$, which is normally unavailable. However, in most of the cases, we can assume $\mathbf{R}_v = \sigma^2\mathbf{I}$. When $M > L$, the parameter $\sigma^2$ represents the smallest eigenvalue of the correlation matrix $\mathbf{R}_x$ of the mixed signals. Hence, we can use a subspace method or an adaptive principal component analysis algorithm to estimate it [1,6]. For $M = L$, it is difficult to estimate $\sigma$, unless we have some additional knowledge about the system, for example, the period when there are no source signals present, so that we can calculate the correlation matrix of the noise using those measurements. In the following analysis, we will assume $\mathbf{R}_v = \sigma^2\mathbf{I}$.

First, to simplify the derived algorithm, after each update, we perform a normalisation of the demixing vector $\mathbf{w}$, given by

$$\mathbf{w}[n] \leftarrow \mathbf{w}[n]/\sqrt{\mathbf{w}^{\mathrm{T}}[n]\mathbf{w}[n]}. \tag{18}$$

Thus, the cost function (10) changes into

$$J(\mathbf{w}) = \frac{E\{y^4\} - 3(E\{y^2\})^2}{(E\{y^2\} - \sigma^2)^2}. \tag{19}$$

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

$$\nabla_{\mathbf{w}}J = 4\frac{(E\{y^2\} - \sigma^2)(E\{y^3\mathbf{x}\} + 3\sigma^2E\{y\mathbf{x}\}) - (E\{y^4\} - 3\sigma^4)E\{y\mathbf{x}\}}{(E\{y^2\} - \sigma^2)^3}. \tag{20}$$

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

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \mu\phi(y[n])\mathbf{x}[n], \tag{21}$$

where $\mu$ is the stepsize and

$$\phi(y[n]) = \frac{\beta y[n]}{(m_2(y)-\sigma^2)^3}[(m_2(y)-\sigma^2)y^2[n] + 3\sigma^2 m_2(y) - m_4(y)]. \tag{22}$$

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

$$m_q(y)[n] = (1-\lambda)m_q(y)[n-1] + \lambda|y[n]|^q, \tag{23}$$

where $\lambda$ 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 (22) becomes

$$\phi(y[n]) = \frac{\beta y[n]}{m_2^3(y)}(m_2(y)y^2[n] - m_4(y)), \tag{24}$$

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

## 4. Simulations and results

The simulations were based on three source signals with binary, uniform and Gaussian distributions, respectively, as shown in Fig. 1. Their corresponding kurtosis values were [−2.000, −1.2246, 0.0742]. As the two non-Gaussian signals have negative kurtosis values, we have $\beta = -1$. In theory, by minimisation of the normalised kurtosis of the extracted signal, we will recover the first source, since it has the smallest kurtosis value.

A $4 \times 3$ mixing matrix was randomly generated and is given by

$$\mathbf{A} = \begin{bmatrix} 0.9575 & 0.5207 & 0.9248 \\ -0.9356 & -0.4131 & 0.6338 \\ -0.4264 & -0.9840 & 0.6787 \\ 0.5103 & 0.3774 & -0.1707 \end{bmatrix}. \tag{25}$$

To measure the quality of extraction of the presented algorithm, we employ the performance index (PI) defined as [2]

$$\mathrm{PI} = 10\log_{10}\left(\frac{1}{L-1}\left(\sum_{l=0}^{L-1}\frac{g_l^2}{\max\{g_0^2, g_1^2, \ldots, g_{L-1}^2\}} - 1\right)\right). \tag{26}$$

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

The additive noise was white Gaussian with a variance of $\sigma^2 = 0.04$. Since we have one more mixture than the number of sources, we can use this degree of freedom to estimate the noise variance $\sigma^2$.

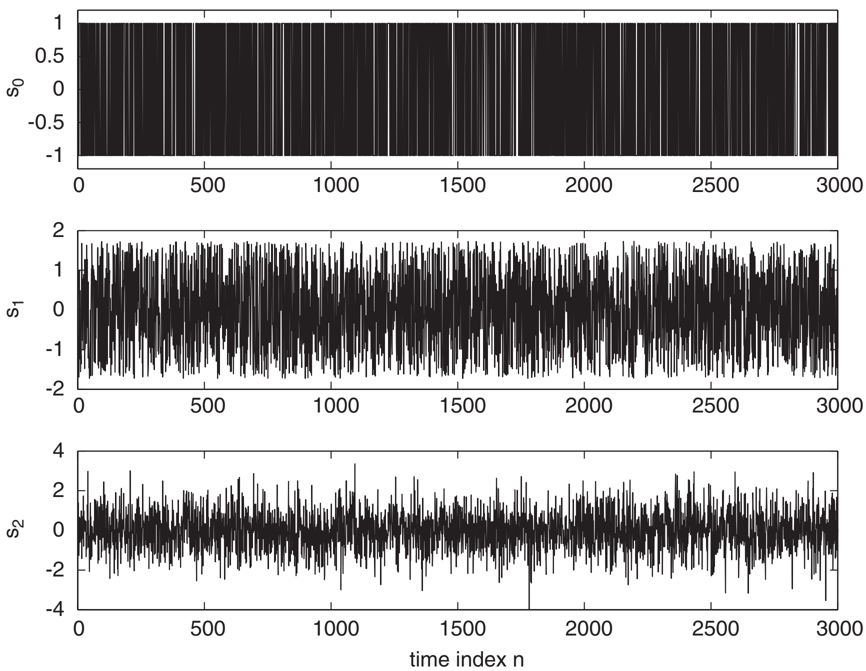


Fig. 1. The three source signals with binary, uniform and Gaussian distributions, respectively.

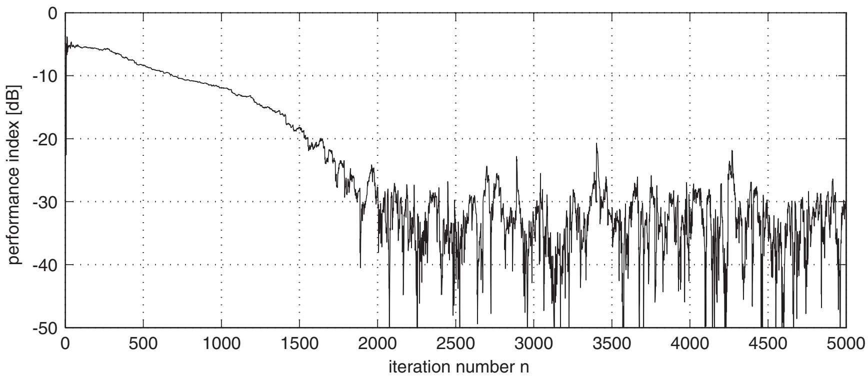


Fig. 2. The performance index using the proposed algorithm.

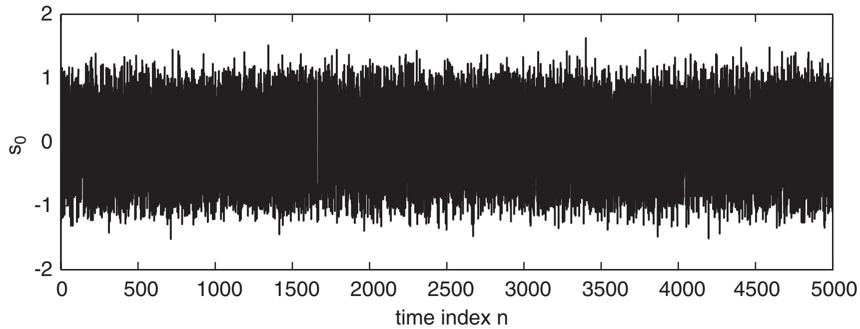


Fig. 3. The extracted source signal using the proposed algorithm.

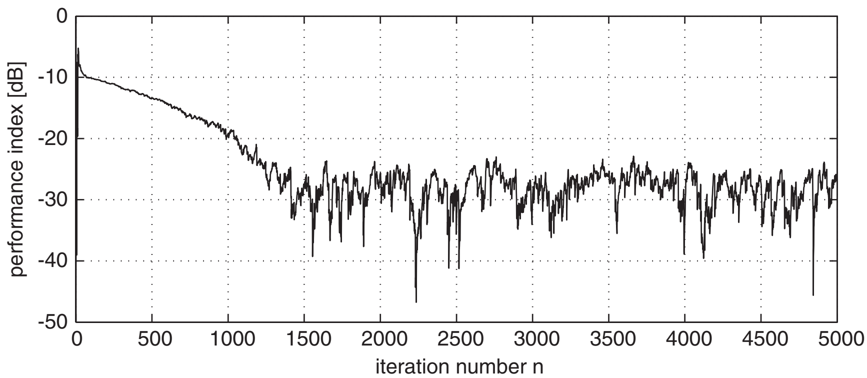


Fig. 4. The performance index using the existing noise-free algorithm.

During the adaptation, the forgetting factor was $\lambda = 0.02$ and the stepsize $\mu = 0.006$. As shown in Fig. 2, the PI reached a level of about $-35$ dB, indicating a successful extraction. The waveform of the extracted signal is given in Fig. 3, and it matches the first source signal except for the effect of noise.

To further illustrate the performance of the proposed algorithm, for comparison, the PI of the algorithm shown in (24) designed for the noise-free case [4], with the same noise level, same mixing matrix, same initial demixing vector $\mathbf{w}$, same forgetting factor and same stepsize, is given in Fig. 4. We see that for the steady state, the PI of this algorithm is about 5 dB worse than that of the proposed one.

In the next experiment, the steady-state PI value of the proposed algorithm is compared to that of the existing noise-free algorithm for different

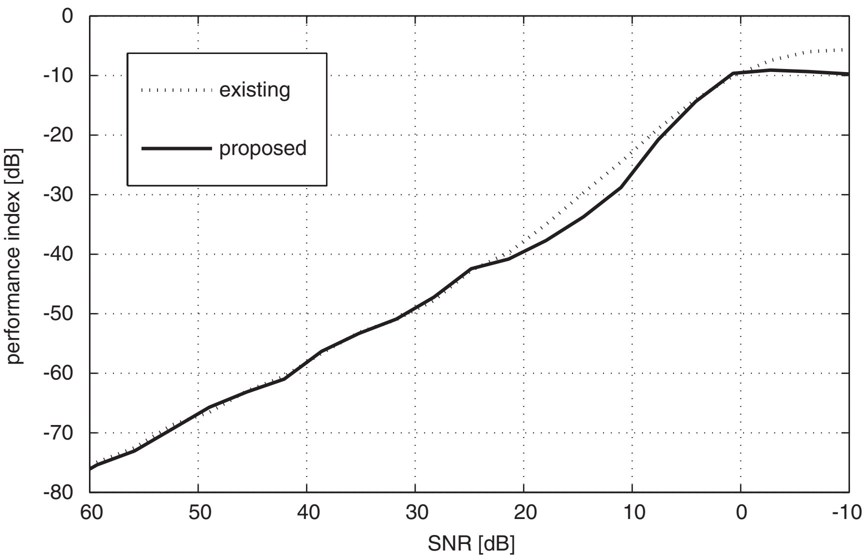


Fig. 5. The steady-state PI value of the two algorithms with respect to different SNRs.

signal-to-noise ratios (SNRs) and the results are shown in Fig. 5. Since the proposed algorithm takes the same form as the existing noise-free algorithm when there is no noise present ($\sigma = 0$), we should expect a very similar performance for low noise levels, as shown in the figure. When the noise level increases, starting from about SNR $= 25$ dB up to about SNR $= 5$ dB, the proposed approach outperforms the existing one and there is an approximately 5 dB difference between them for a major part of this range. When the noise level increases further, with the SNR smaller than about 5 dB, both of the two algorithms fail to extract the source signals due to the high noise level, although the proposed one has a more consistent performance for this very high noise level situation.

## 5. Conclusions

We have proposed a novel BSE 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 the knowledge of the correlation matrix of the noise. A proof of this method is provided and the performance of the proposed adaptive algorithm is verified by simulations.

## References

[1] A. Hyvarinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, New York, 2001.

[2] A. Cichocki, S. Amari, Adaptive Blind Signal and Image Processing, Wiley, New York, 2003.

[3] N. Delfosse, P. Loubaton, Adaptive blind separation of independent sources: a deflation approach, Signal Processing 95 (1995) 59–83.

[4] A. Cichocki, R. Thawonmas, S. Amari, Sequential blind signal extraction in order specified by stochastics properties, IEE Electron. Lett. 33 (1997) 64–65.

[5] A. Hyvarinen, E. Oja, A fast fixed-point algorithm for independent component analysis, Neural Comput. 9 (1997) 1483–1492.

[6] H.L. Van Trees, Optimum Array Processing, Part IV of Detection, Estimation, and Modulation Theory, Wiley, New York, USA, 2002.