An Improved Method for Blind Source Separation

Yong-jian ZHAO

School of Information Engineering, Shandong University (Wei hai), China

Keywords: Component, Separation, Kurtosis, Load, Property, Moment.

Abstract. In practical applications especially in biomedical signal processing, a large number of sensors is available but only one or a very few are desired. The simultaneous blind source separation (BSS) technique always introduces large computational load. A contrast function is formulated associated with normalized kurtosis. Furthermore, an improved learning algorithm is derived based on the standard gradient descent rule. In contrast to simultaneous BSS, the proposed method can provide more flexibility and has some potential advantages in terms of computational load. Computer simulations illustrate its performance.

Introduction

In recent decades, blind source separation (BSS) has received much research attention in various fields such as biomedical signal processing, data mining, image processing, etc. [1-3] Traditional BSS technique aims to simultaneously recover all source signals from their mixtures using some statistical properties of original sources [2-4]. In most cases especially in biomedical signal processing, a large number of sensors is available but only one or a very few are desired. For instance, in EEG or MEG, more than 20 sensor signals are observed while only one is interesting, the rest of which are regarded as interfering noise [5, 6]. Another typical example is the cocktail party problem; it is essential to extract voices of specific person instead of separating all original sources from a large array of microphones. In such applications, it is important to introduce an improved method so that we can extract only one or a very few original sources, which are potentially interesting or contain useful information. We start to overcome the shortcomings inherent in traditional BSS method. Such technique is called blind source extraction (BSE), which is a special class of BSS technique. In contrast to simultaneous BSS, such technique can provide more flexibility and has some potential advantages in terms of computational load.

Until now, several approaches [2-6] have been proposed for the BSE problem, which are based on either second-order or high order statistics of the data. For instance, Barros and Cichocki [2] firstly defined the prerequisites to extract the desired FECG using only prior information about its autocorrelation function. This method uses specific prior information about the desired source, namely, a suitable time delay at which the autocorrelation of the desired source has a peak while the rest have a small value. Since the desired FECG satisfies the constrained condition, it can be used for FECG extraction. However, the extracted signal is often contaminated with undesired sources and the algorithm is sensitive to the delay estimation error [5-7]. To circumvent this problem, Zhang [5, 6] developed an improved extraction algorithm associated with the augmented Lagrange function. Zhang’s algorithm owns better convergence stability. It must be mentioned that it is specially designed for linear instantaneous mixtures. Generally speaking, the nonlinear mixture models are more realistic and accurate than the linear mixture in many practical applications.

The Central Limit Theorem tells that the distribution of a sum of independent random variables tends to a Gaussian distribution [8-10]. That is to say, maximizing/minimizing the non-Gaussianity will produce an estimation of one original source. In recent decades, kurtosis has become a practical measure of non-Gaussianity. At present, kurtosis generally represents the preferred technique for blind source extraction [11-14]. After calculating the performance of kurtosis, a contrast function is introduced on the basis of normalized kurtosis. Furthermore, an improved learning algorithm is deduced based on the standard gradient descent approach. Someone can recover a desired source
signal from a large number of sensors. After a deflation process, he can also recover the source signals one by one. That is to say, the proposed method can provide more freedom in separation. Computer simulations illustrate the performance of the proposed method.

Proposed Learning Algorithm

Recently, BSS/BSE technique has enjoyed much theoretical and experimental attention [2, 6, 8-10]. Traditional BSS technique aims to simultaneously recover a set of statistically independent source signals with unknown coefficients are observed. Denote that unknown source signals \( s_i(t), i=1,\ldots,n \) which are mutually independent, which can be gastric slow waves, respiratory artifact and noise. The mixture of original sources, which are the sensor output \( x(k)=[x_i(k),\ldots,x_n(k)]^T \), can be depicted as follows:

\[
x(k) = As(k)
\]

(1)

Where \( A \in \mathbb{R}^{m \times n} \) is an unknown non-singular mixing matrix, \( k \) is the discrete time index, and \( s(k)=[s_i(k),\ldots,s_n(k)]^T \). Here \( s(k) \) is a vector of unknown zero-mean and unit-variant original source signals. In model (1), only the vectors \( x(k) \) can be observed and everything else is unknown. The original sources are mutually independent with non-Gaussian and nonlinear autocorrelation. The basic problem of BSS is to estimate the source signals \( s(k) \) using observations of the mixtures \( x(k) \) and some statistical properties of original sources. Traditional BSS technique aims to recover all source signals from their mixture simultaneously [1, 2, 6]. As the mixture of independent Gaussian sources mixed by an orthogonal mixing matrix remains independent, the BSS techniques are only applicable to non-Gaussian sources, or when at most only one source is Gaussian [9-11]. Under this condition, the extraction of one original source signal is equivalent to extracting an independent component from the mixtures. Assume that we want to extract a desired source signal \( y(k) \), an efficient technique is to introduce an iterative process to find a weight vector \( w \) so that \( y(k) = w^T x(k) = g^T s(k) \) is a good approximation to the desired source signal [4-6]. Here \( g^T = w^T A \) is the global demixing vector.

Non-Gaussianity is actually of paramount importance in BSS/BSE estimation [1, 4]. To utilize non-Gaussianity for signal separation, one must have a quantitative measure about non-Gaussianity of a random variable. As a first practical measure of non-Gaussianity, Kurtosis is the name given to the fourth-order cumulant of a random variable. Kurtosis is defined in the zero-mean case by

\[
kt = E\{y^4\} - 3(E\{y^2\})^2
\]

(2)

Where \( E\{\cdot\} \) expresses the statistical expectation operator.

Alternatively, the normalized kurtosis \( k_4(y) = \frac{E\{y^4\}}{[E\{y^2\}]^2} - 3 \) can be used. For whitened data \( E\{x^2\} = 1 \), both the versions of kurtosis reduces to \( kt(y) = k_4(y) = E\{y^4\} - 3 \). That is to say, for whitened signal, the fourth moment \( E\{x^4\} \) can be selected instead of kurtosis for characterizing the distribution of \( y \). Kurtosis has many useful properties. If \( x \) and \( y \) are two statistically independent random variables, one can deduce that

\[
kt(x+y) = kt(x) + kt(y)
\]

(3)

For any scalar parameter \( \delta \),

\[
kt(\delta y) = \delta^4 kt(y).
\]

(4)

Therefore, kurtosis is not linear with respect to its argument. Furthermore, kurtosis is the simplest statistical quantity for indicating the non-Gaussianity of a random variable. If \( x \) has a Gaussian
distribution, its kurtosis is zero. This is the sense in which kurtosis is “normalized” in contrast to the fourth moment, which is not zero for Gaussian variables. Normalized kurtosis owns the advantage that we do not need to perform the otherwise required pre-whitening and weight normalization operations. Therefore, to represent the stochastic properties of the source signals, we adopt the normalized kurtosis. A contrast function is formulated on the basis of normalized kurtosis:

$$\text{minimize } J(w) = -\frac{1}{4}|k_4(y)|$$

Subject to the constraint $\|w\| = 1$.

To minimize the absolute normalized value of kurtosis, we start from certain vector $w$, compute the direction in which the absolute value of the kurtosis of $y = w^T x$ is growing most strongly, on the basis of the available sample $x(1), \ldots, x(T)$ of mixture vector $x$, and then move the vector $w$ in that direction. Applying the standard gradient descent approach to (5), we derive a learning rule as follows:

$$\tilde{w}(t+1) = \tilde{w}(t) + \mu \cdot \text{sgn}(k_4(t)) f[y(t)] x(t)$$

$$w(t+1) = \tilde{w}(t+1) / \|\tilde{w}(t+1)\|$$

Where $\mu$ is the step size and the nonlinear function $f[y(t)]$ is deduced by

$$f[y(t)] = m_2 \frac{\partial k_4(y)}{\partial y} = y^3(t) - \frac{m_4(t)}{m_2(t)} y(t)$$

Where $m_p(t) \ (p = 2, 4)$ is the $p$th moment.

After the first source signal is recovered, we apply a deflation process to eliminate the recovered signal from its mixture. This means we should adopt a linear transformation given by:

$$\tilde{x}(k) = x(k) - w^T x(k) = x(k) - w^T A x(k)$$

The remained mixture then can undergo another process to recover the next signal. The procedure may be recursively applied to recover the source signals sequentially. In other words, one can recover the source signals in a prescribed order, which is important in practice.

**Computer Simulations**

In many applications especially in biomedical signal processing, a large number of sensor signals can be observed while only one is interested, the rest of which are considered interfering noise. In such applications, it is desirable to recover only the interested source signal instead of all sources. Therefore, we use biomedical signal to demonstrate our method. The ABio7 dataset, proposed by Barros [15], is a benchmark containing a set of typical biomedical sources, where each signal has 5000 samples with zero means and unit variances. Due to space limitation, we selected three source signals ($s_1$, $s_2$ and $s_3$) shown in Figure 1.

![Figure 1. Original biomedical signals.](image-url)
Signal $s_1$ is an ECG and the other two are common artifacts frequently observed while measuring the ECG: electrode and respiratory artifacts. A $3 \times 3$ mixing matrix $A$ was randomly generated whose rows were $A_1 = [0.81\ -0.46\ 0.73]$, $A_2 = [0.68\ 0.42\ 0.39]$ and $A_3 = [-0.53\ 0.79\ 0.61]$. Three original sources were mixed with the matrix $A$. The mixed results are depicted in Figure 2.

The separation results are shown in Figure 3.

We also ran typical BSS/BSE methods: BCBSE in [2] and MCBSE in [5, 6]. To measure the signal extraction quality, we used the following measure of the accuracy:

$$SI = -10 \times \log_2 [s(k) - \tilde{s}(k)]^2 \text{ (db)},$$

Where $s(k)$ is the desired source signal and $\tilde{s}(k)$ is the recovered signal. Obviously, the higher $SI$ is, the better the performance is. Generally speaking, the $SI$ value larger than 20db means a good extraction. To exactly compare the performance of these algorithms, the simulation was independently repeated 100 times so as to get the average and highest $SI$. The performance of extracting signal $s_1$ is depicted in Table 1. Here, we called our algorithm as NEWBSE. In contrast to BCBSE and EVBSE, NEWBSE generally had better accuracy. BCBSE showed the poorest accuracy, due to the fact that the recovered signal was always contaminated with some unwanted signals or noise. The superiority of our method is obvious.

<table>
<thead>
<tr>
<th>algorithm</th>
<th>BCBSE</th>
<th>MCBSE</th>
<th>NEWBSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>average $SI$</td>
<td>14.463</td>
<td>21.287</td>
<td>26.756</td>
</tr>
<tr>
<td>highest $SI$</td>
<td>15.65</td>
<td>25.287</td>
<td>28.856</td>
</tr>
</tbody>
</table>

**Summary**

Due to its potential application including many especially in biomedical signal processing, BSS has become an increasingly popular data analysis technique. Traditional BSS technique often recovers all the source signals simultaneously. In most cases, the number of sensors is large. If someone is only
interested in one or a very few original sources, such technique will introduce heavy computational load. To overcome the shortcomings inherent in the traditional BSS technique, an improved learning rule has been derived associated with normalized kurtosis. Someone can recover only the desired source signal from its mixture. After a deflation process, he can also recover the source signals one by one, which is important in practice. In contrast to the simultaneous BSS, such technique can provide more flexibility and has some potential advantages in terms of computational load.

Acknowledgement
This work is supported by the National Natural Science Foundation of China (11473019), Natural Science Foundation of Shandong (ZR2014AM015).

References