A Probabilistic Short-length Linear Predictability Approach to Blind Source Separation

Mahdi Khosravy[1] , Mohammad Reza Alsharif[1] and Katsumi Yamashita[2]

[1]Department of Information Engineering, Faculty of Engineering, University of the Ryukyus,
1 Senbaru, Nishihara, Okinawa 903-0213, Japan
Tel: +81-98-895-8681 Fax: +81-98-895-8727
[2]Graduate School of Engineering, Osaka Prefecture University,
1-1 Gakuen-cho, Sakai, Osaka, Japan
E-mail : [1]k078661@eve.u-ryukyu.ac.jp, [1]asharif@ie.u-ryukyu.ac.jp, [2]yamashita@eis.osakafu-u.ac.jp

Abstract: A merit function based on short length linear predictability of signal in an objective probabilistic algorithm is defined and used for blind source separation (BSS) of linear mixtures of signals. In BSS literatures, it has been conjectured that linear mixture of statistically independent source signals will result in a set of signals which each of them has less predictability than (or equal to) that of any of component source signals. We have used this property to extract source signals by finding an un-mixing matrix that maximizes the proposed merit function of predictability for each recovered signal. This method which is called Probabilistic Short-length Linear Predictability BSS (PSLP-BSS), its performance has been driven with many tests performed with mixtures of different kinds (speech, audio, image, constructed mathematical signals like saw tooth and sinusoidal). In all cases, correlation between each of source signals and each of extracted signals shows near-perfect performance of the method. The proposed BSS doesnt require any assumption regarding the probability density function of source signals. It has been demonstrated that PSLP-BSS can separate signal mixtures in which each mixture is a linear combination of source signals with gaussian, super-gaussian and sub-gaussian probability density functions. However, the method is adapted to temporal structure of recovered signals. Since, the un-mixing matrix that is concluded by proposed merit function can be obtained as the solution to a generalized eigenvalue routine, signals can be extracted simultaneously using the fast eigen value.

Key words: blind signal separation, predictability, gradient ascent, eigen value

1. Introduction

The separation of blind sources from mixed observed data is a fundamental signal processing problem that is known as Blind Source Separation (BSS). Typically, the observations are obtained at the output of a set of sensors. Each sensor receives a different combination of the source signals. Blind means that, the source signals are not observed and also no information is available about the mixture. Solution to BSS is based on realistic assumption as its basic principle that different physical mixing processes generate signals with properties different from source signals. The increment or decrement of such measurable property is the main key difference between observation and source signals. Usually BSS methods are based on a cost(merit) function of signals which measures the desired property of signals, and they look for a separation matrix

which maximizes the cost function, or minimize it. The minimization or maximization depends on the desired property for signals as well as merit function defined to measure the property. This paper defines a probabilistic short-length predictability (PSLP) measurement of signal as a merit function for separation of blind sources.

In BSS literatures among other principles for extraction of source signals, predictability[1] as well as complexity[2] have received much less attention than others like independence[3], non-gaussianity[4], Infomax [5]. Complexity pursuit suggest that any observed mixture of source signals is always more complex than each of source signals. And source signals which each of them has generated by a different physical source are simpler than mixture signals. Based on this conjecture, complexity pursuit seeks minimally complex source signals as key principle of separation.

While complexity pursuit methods is derived directly from information-theoretic[6], such methods tend to be more general and more complicated than is required for many BSS problems. In contrast, temporal predictability[1] minimizes a very simple measure of complexity.

Here, based on probability theory, we present a simple and fast short length measure of predictability. The proposed function, by using the most and least probable combinations as prediction values in measurement of predictability, performs near perfect efficiency in BSS. The performance of new method has been evaluated with many tests performed over mixtures of different kinds (speech, audio, image, constructed mathematical signals like saw tooth and sinusoidal).

Appropriate probabilistic choice of expectation weights in prediction, has brought us the advantage of fast measurement of predictability by means of a short averaging window. Furthermore, since the un-mixing matrix that is concluded by proposed merit function can be obtained as the solution to a generalized eigenvalue routine, source signals can be extracted simultaneously using the fast eigen value routine. So by using the proposed method higher performance speed will obtain at two points: first in short length proposed method of predictability measurement , and second in capability the of method for extension to fast eigenvalue problem.

Almost all BSS algorithms sometimes in optimizing the appropriate weights for maximization or minimization of merit function have the problem of being caught in a local minimum or maximum. But since the proposed merit function is composed of quadratic forms, like the classic pre-

dictability merit function it has exactly one global maximum, and the proposed BSS doesn't suffer from such problem. Finally as another advantage, the method is adapted to temporal structure of recovered signals.

The structure of the paper is as follows: section 2 explains general formulation of a BSS problem. section 3 presents the proposed measurement of predictability and BSS based on proposed method is explained in section 4. section 5 discusses the performance of the proposed BSS in separation of signal and image mixtures of different kinds sources. Finally section 6 concludes the paper.

2. BSS Problem Formulation

Here we face the general form of BSS problem. M unknown source signals upon transmission through a medium has been linearly mixed together and mixture signals are collected from M sensors. We don't know any thing about source signals and also mixing process. So, in matrix notation the $M \times 1$ vector $\mathbf{X}$ of observed signals ($[x_i]_{M \times 1}$), is multiplication of $M \times 1$ vector $\mathbf{S}$ of unknown source signals ($[s_i]_{M \times 1}$), by Unknown $M \times M$ mixing matrix $\mathbf{A}$:

$$\mathbf{X} = \mathbf{A} \times \mathbf{S}$$

Given only the observation signals $\mathbf{X}$, the solution of BSS problem, seeks for the best $M \times M$ matrix W as un-mixing matrix to obtain extracted signals as much as possible close to unknown source signals as follows

$$\mathbf{Y} = \mathbf{W} \times \mathbf{X}$$

where $\mathbf{Y}$ is $M \times 1$ vector of extracted signals ($[y_i]_{M \times 1}$). Each row of $\mathbf{W}$ is the $1 \times M$ un-mixing vector W_i related to each extracted signal y_i.

3. Probabilistic Measurement of Predictability

The merit function that we have defined for blind source separation is based on linear prediction model[7]. it measures predictability of recovered signals y_is based on a finite length of prior sample values. It performs by using two short length linear combination of past samples as prediction values, which both are probabilistic expectations; the most probable expectation $\widehat{y}(n)$ and the least probable expectation $y^{\vee}(n)$.

For obtaining $\widehat{y}(n)$, we choose each sample weight proportional to its probability p_i of being equal to current sample. The most probability weight belongs to the closest sample ($y(n-1)$), and by going farther from $y(n)$ through samples, their probability weights in prediction should be decreased. Experimentally we have observed that exponentially decrement leads to a better result. The most probable expectation as a prediction of current sample value can be estimated by the following short term weighted average:

$$\widehat{y}(n) = \alpha \sum_{i=1}^{M-1} p_i y(n-i) \qquad (1)$$

$$\sum_{i=1}^{M-1} \alpha p_i = 1$$

where αp_i is weighted probability of true prediction for i^{th} prior sample ($p_i = P(y(n) = y(n-i))$) and α is normalization factor of weights.

Here we want to estimate the least probable expectation $y^{\vee}(n)$, with the same preceding samples values as used for $\widehat{y}(n)$, but different probability weights. Therefore, we consider the finite length of expectation window as a precondition to our probabilistic estimation. And we choose each sample weight proportional to its probability of not being equal to the current sample value.

By such choice of weights the expectation value, as a prediction would be the weighted average with the most probable difference from the current sample. Therefore by using the short term window, the least probable expectation as a prediction of current sample value is estimated as follows:

$$y^{\vee}(n) = \beta \sum_{i=1}^{M-1} q_i y(n-i) \qquad (2)$$

$$q_i = 1 - p_i \quad , \quad \sum_{i=1}^{M-1} \beta q_i = 1$$

where βq_i is weighted probability of false prediction for i^{th} preceding sample ($q_i = P(y(n) \neq y(n-i)) = 1 - p_i$) and β is normalization factor of weights.

Based on above concept of the most and the least probable prediction values, here we interpret short length predictability of signal. It means that signal has a property that by applying the probability weights of the most probable expectation, mean square error of prediction will have its minimum value and by applying the probability weights of the least probable expectation, mean square error will have its maximum value. The probabilistic short length predictability function for a m-sampled signal is defined as:

$$F(y) = log \frac{e^{\vee}}{\widehat{e}} = log \frac{\sum_{n=1}^{m}(y^{\vee}(n) - y(n))^2}{\sum_{n=1}^{m}(\widehat{y}(n) - y(n))^2} \qquad (3)$$

Where $e^{\vee}$ and $\widehat{e}$ are respectively mean square errors for the least probable expectation and the most probable expectation with the same short length expectation window.

4. PSLP-BSS

Consider an extracted signal y_i formed by multiplication of corresponding $1 \times M$ component vector W_i by the set of M observed signals mixtures $\mathbf{X}$. Given that $y_i = W_i \mathbf{X}$, equation (3) can be rewritten as

$$F(y_i) = log \frac{e_i^{\vee}}{\widehat{e}_i} = log \frac{W_i C^{\vee} W_i^t}{W_i \widehat{C} W_i^t} \qquad (4)$$

where $C^{\vee}$ is a $M \times M$ covariance matrix of the error signals of the least probable expectation of signal mixtures and $\widehat{C}$ is a $M \times M$ covariance matrix of the error signals of the most probable expectation as prediction signal. The covariance matrix of error of the most probable and the least probable expectations between $i^t h$ and $j^t h$ m-sampled mixtures are as follows

$$C_{ij}^{\vee} = \sum_{n=1}^{m} ((x_i(n) - x_i^{\vee}(n))(x_j(n) - x_j^{\vee}(n))) \qquad (5)$$

$$\widehat{C}_{ij} = \sum_{n=1}^{m} ((x_i(n) - \widehat{x}_i(n))(x_j(n) - \widehat{x}_j(n))) \quad (6)$$

where $x^\vee$ and $\widehat{x}$ are the least and the most probable expectation signals of signal mixture. For a given set of signal mixtures $C^\vee$ and $\widehat{C}$ need to be computed only once. The derivative term of F, (for gradient ascent learning) is as follows

$$\nabla_{W_i} F = \frac{2W_i}{e_i^\vee} C^\vee - \frac{2W_i}{\widehat{e}_i} \widehat{C} \quad (7)$$

Since the function F is a ratio of quadratic forms, for gradient ascent the local minimum or maximum is not an issue of problem.

4.1 Extension to generalized eigen problem

Since, the solution of BSS problem will be the point that the gradient of F is zero, from equation (7), it can be obtained,

$$\widehat{C}^{-1} C^\vee W_i = W_i \left(\frac{e_i^\vee}{\widehat{e}_i} \times I \right) \quad (8)$$

Where I is identity matrix. It can be easily seen that the equation (8) has the form of $AV = VD$. This is the form of a generalized eigenvalue problem. So BSS solutions for W_i can be obtained as eigen vectors of $\widehat{C}^{-1} C^\vee$ matrix with corresponding eigen values $d_i = \frac{e_i^\vee}{\widehat{e}_i}$. Therefore by solving the generalized eigenvalue routine and obtaining **W** that contains all un-mixing vectors $W_i(i = 1, \ldots, M)$, all source signals can be simultaneously extracted.

The form of equation (8) is cause of form of the merit function whose derivative leads to eigenvalue equation. Indeed it makes separation of blind sources so fast by the bless of fast solution of eigenvalue problem.

5. Results and Discussion

The PSLP-BSS performance has been driven with many tests performed with mixtures of different kinds, such as speech signals, audio signals, gray-scale images and constructed mathematical signals like saw tooth and sinusoidal. In our simulation the length of prediction window for both of the most and the least probable expectations, is only 8 samples, and probability of equality of current oncoming signal sample value and its i_{th} prior sample value is $p_i = 2^{-i}$. Although values taken for the length of prediction window and p_i are not the most optimum, in all cases, correlation between each of source signals and each of extracted signals shows near perfect performance of the method. Furthermore, it has been demonstrated that PSLP-BSS can separate signal mixtures in which each mixture is a linear combination of source signals with gaussian, super-gaussian and sub-gaussian probability density functions. Fig.1 shows three audio source signals, three linear mixture of them and extracted signals by PSLP-BSS. The permutation matrix (multiplication matrix of the mixing and separator matrices as a quantitative measure of separation) in this example test is as follows (normalized form);
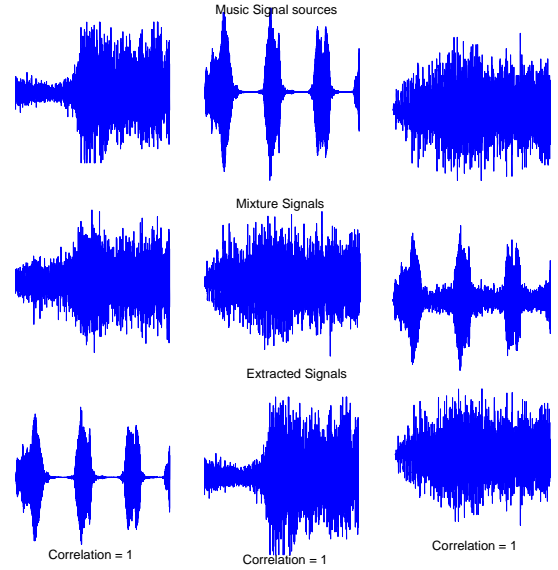

Figure 1. Audio source signals (top), their linear mixture signals (middle) and separated audio signals (bottom).

Table 1. Correlation between audio source signals (S1, S2 & S3) and extracted signals related to fig.1.

	Source#1	Source#2	Source#3
Extracted#1	0.0011	**1.0000**	0.0024
Extracted#2	**1.0000**	0.0009	0.0068
Extracted#3	0.0065	0.0000	**1.0000**

$$G = A \times W = \begin{bmatrix} 0.0011 & \underline{1.0000} & 0.0024 \\ \underline{1.0058} & 0.0005 & -0.0063 \\ 0.0066 & -0.0026 & \underline{0.9960} \end{bmatrix}$$

For G matrix, the closer underlined elements to one, and the closer other elements to zero indicate the more successful separation. besides the permutation matrix, table.1 shows the correlation between audio source signals and extracted signals related to fig.1. As it can be seen, the PSLP-BSS performance for separation audio mixtures is almost perfect. We have applied the proposed BSS for separation of mixtures of gray-scale images too. It separates blind image sources as well as blind signal sources with a high performance. Fig.2 shows the PSLP-BSS performance in separation linear mixtures of gray-scale images. Table.2 approves its near perfect performance by correlation between source and extracted images.

We also applied PSLP-BSS over mixture signals made of source signals with different probability distribution functions. Fig.3 shows one of experiments related to sources with different PDFs. In this figure sources are an audio signal, a sine wave and a white noise signal whose PDFs are respectively super-gaussian, sub-gaussian and gaussian. Table.3 convey the high performance of BSLP-BSS in separation of these sources and it approves its independence of any assumption regarding the probability density functions of sources.
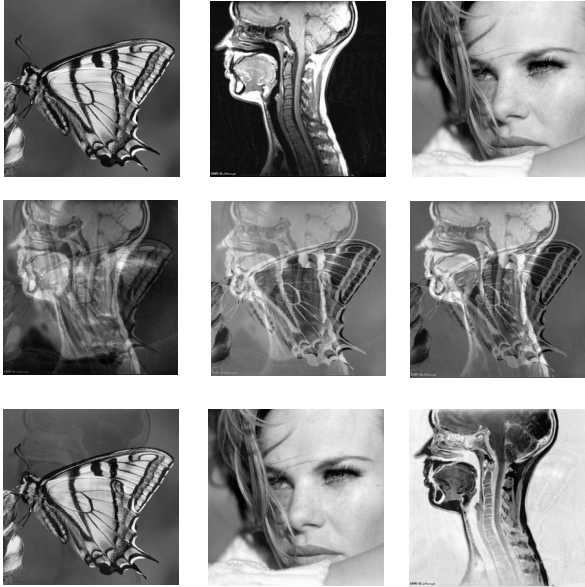
Figure 2. Source images (top), their linear mixture images (middle) & separated images by PLSP-BSS (bottom)

Table 2. Correlation between source Images (S1, S2 & S3) and extracted images related to fig.2.

	Source#1	Source#2	Source#3
Extracted#1	**0.9914**	0.0416	0.0171
Extracted#2	0.0190	0.0974	**0.9994**
Extracted#3	0.1304	**0.9943**	0.0298

6. Conclusion

A merit function based on short length linear predictability of signal in an objective probabilistic algorithm is defined and used for blind source separation of linear mixtures of signals as well as images. Its high performance over a lot of signals and images of different kinds has been approved.

Since, the method has been extended to generalized eigenvalue problem, fast and simultaneous separation of sources has become possible. In the other hand, because of appropriate probabilistic choice of expectation weights in prediction, it can use a short length window in predictability measurement, and it is faster than predictability measurement suggested in [1]. So proposed method obtains higher performance speed at two points: first by capability of using short length window for predictability measurement , second by the means of extension of the proposed method to fast eigenvalue routine.

BSS based on proposed merit function can extract signals with different PDFs from the mixture and it doesn't require any assumption for PDF of source signals. However, the most optimum weights and length of short-length prediction window are issues of our future work.
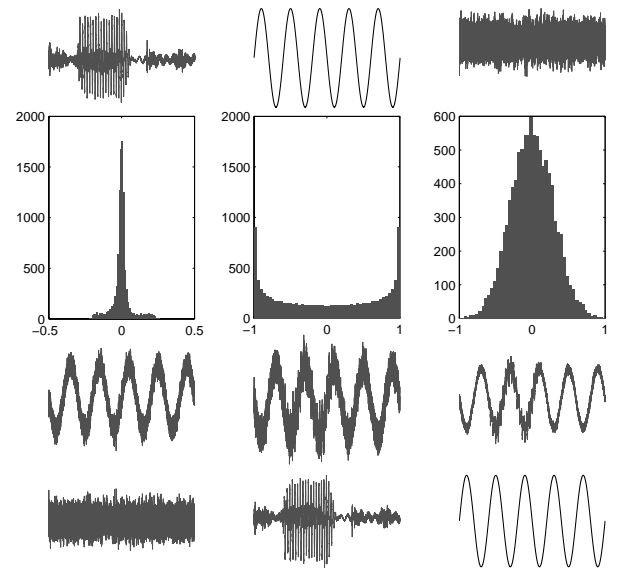

Figure 3. Three source signals with different PDFs(top) and their PDFs: resp. super gaussian, sub-gaussian & gaussian (second row), their linear mixture signals (third row) and extracted signals by PSLPBSS(bottom) .

Table 3. Correlation between source signals with different PDFs (S1, S2 & S3) and extracted signals from their linear mixture.

	Source#1 super gaussian	Source#2 sub-gaussian	Source#3 gaussian
Extracted#1	0.0014	0.0016	**0.9999**
Extracted#2	**1.0000**	0.0006	0.0083
Extracted#3	0.0006	**1.0000**	0.0146

References

[1] JV. Stone, *Blind source separation using temporal predictability*, Neural computation, 13(7), pp. 1559−1574, 2001.

[2] A. Hyvarinen, *Complexity pursuit: Combining non-gaussinity and autocorrelations of signal separation*, ICA2000, Helsinki, Finland, pp. 175−180, 2000.

[3] C. Jutten, & J. Herault, *Independent component analysis versus PCA*, In Proc. EUSIPCO, pp. 643−646, 1988.

[4] J. Friedman, *Exploratory projection pursuit*, Journal of the American Statistical Association, 82(397), 249−266, 1987.

[5] AJ. Bell, & TJ. Sejnowski, *An information-maximization approach to blind separation and blind deconvolution*, Neural computation,vol. 7, 1129−1159, 1995.

[6] A. Hyvarinen, *Complexity pursuit: Separating interesting components from time series. Neural computation*, 13, 883898, 2001.

[7] SV. Vaseghi, Linear Prediction Models, chapter. 8 of *Advanced digital signal processing and noise reduction*, John Wiley & Sons,Ltd ,Third edition, 2006.