A New Approach to Blind Source Separation with Global Optimal Property*

Yiu-ming Cheung
Department of Computer Science
Hong Kong Baptist University, Hong Kong
E-mail: ymc@comp.hkbu.edu.hk

Hailin Liu
Department of Applied Mathematics
Guangdong University of Technology, China
E-mail: lhliu21@163.com

ABSTRACT

This paper presents a new independency metric for blind source separation (BSS) problem. It is mathematically proved that the metric value of any linear combination of source signals is less than the largest one of sources under a loose condition. Further, the global optimization of this new metric is achieved by formulating it as a generalized eigenvalue problem. Subsequently, we guarantee to find out a correct de-mixing matrix through maximizing the proposed metric to separate the sources. The simulation results have shown its success in separating the linear combinations of sub-Gaussian and super-Gaussian sources with at most one Gaussian signal.

KEY WORDS

Blind Source Separation, Independency Metric, Independent Component Analysis, Generalized Eigenvalue Problem, Global Optimization.

1 Introduction

Since Jutten and Herault published their seminal work [1] in 1991, blind source separation (BSS) has been receiving wide attention in the fields of signal processing and neural networks because of their potential attractive applications in wireless communications, biomedicine, speech signal processing, earthquake reconnoitering, and so forth. In the literature, blind source separation with an instantaneous linear mixture has been formulated as an independent component analysis (ICA) problem: Suppose there are n channels of non-Gaussian source signals with at most one Gaussian one, denoted as s_1, s_2, …, s_n, which are statistically independent each other. The sources are instantaneously and linearly mixed by an unknown full-rank square matrix $\mathbf{A}$ and observed as:

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{1}$$

where $\mathbf{s} = [s_1, s_2, \ldots, s_n]^T$, $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$, and T is a transpose operation of a matrix. The objective of an ICA approach is to recover $\mathbf{s}$'s up to a constant scale and any permutation of indices through a set of observations $\{\mathbf{x}_i\}_{i=0}^N$ by finding out a de-mixing matrix $\mathbf{W}$ such that

$$\mathbf{y} = \mathbf{W}\mathbf{x}, \tag{2}$$

where $\mathbf{y} = [y_1, y_2, \ldots, y_n]^T$ is a recovered signal of $\mathbf{s}$.

In the past decade, a number of ICA algorithms based on different methodologies and theories have been proposed. For example, Bell and Sejnowski [2] presented the INFOMAX algorithm to maximize the mutual information between $\mathbf{x}$ and $\mathbf{y}$. Amari et al. [3] presented an MMI algorithm to minimize the mutual information between $\mathbf{y}$ and their components y_j's with using a natural gradient descent learning rule. Furthermore, Gaeta et al. [4] and Pham et al. [5] proposed the approaches based on maximum likelihood (ML) estimation, which was later to be shown [6] that this approach is equivalent to the INFOMAX. Girolami and Fyfe [7] used marginal negentropy as a projection index and showed that maximization of negentropy can reach the separation of the source signals. Lee et al. [8] have shown that INFOMAX, MMI, ML and Negentropy Maximization algorithms can be all unified in an information-theoretic framework. Further examples are nonlinear PCA algorithms for ICA [9, 10] and the cumulant-based algorithms [11], both of which are actually to approximately minimize the mutual information of the recovered signals. Empirical studies have shown the success of these algorithms in separating the sources. Unfortunately, the analysis of these algorithms' performance is intrinsically difficult upon the complicated nonlinearity of the contrast functions they use. To our best knowledge, from a theoretical viewpoint, it is still an open problem thus far when an ICA algorithm guarantees to separate the source signals, and when not.

In this paper, we therefore present a new ICA algorithm using a novel metric named *Independency Metric*, which is defined as a logarithm of a ratio of two covariances. The numerator is the covariance of a transform of $\mathbf{y}$'s component y, whereas the denominator is the covariance of y itself. It has been mathematically proved that the metric value of any linear combination of source signals is less than the largest one of sources under a loose condition. Since this metric is a quadratic form with respect to the de-mixing matrix $\mathbf{W}$, the global solution of $\mathbf{W}$ can be easily achieved by formulating the metric optimization as a generalized eigenvalue problem. We have given out a new ICA algorithm accordingly. In the literature, a related work

*The work described in this paper was supported by Faculty Research Grant of Hong Kong Baptist University with Project Number: FRG/01-02/II-24.

has been recently done by J.V. Stone [12]. In his paper, a metric named *Temporal Predictability* has been presented as a logarithm of a ratio of two prediction error terms. The numerator is the summation of long-term prediction errors of a $\mathbf{y}$'s component, while the numerator is the summation of its short-term prediction errors. Essentially, his work is based on the conjecture that, *given any set of statistically independent source signals, the temporal predictability of any signal mixture is less than (or equal to) that of any of its component source signals*. Unfortunately, although a number of experiments have reported its success, some experimental simulations have found that this conjecture is not always true. Instead, we have found that the temporal predictability of a signal mixture is larger than that of some source signals in some cases. For example, as shown in Table 1, the value of temporal predictability F of observation signal x is greater than that of source signals s_1 and s_2, respectively. Table 2 shows that $F(x)$ is greater than $F(s_2)$. In contrast, the truth of the proposed Independency Metric has been proved, but not a conjecture. The experimental simulations in Section 4 have further shown its success in separating the linear combinations of sub-Gaussian and super-Gaussian sources with at most one Gaussian signal.

Table 1. Temporal predictability F of the source and observation signals, where s_1 is a Gaussian signal with zero mean and unit variance, $s_2 = cos(t)$ is a Sub-Gaussian signal, and the observation $x = 0.3710\ s_1 + 0.8297\ s_2$.

No. of samples	5000	10000	20000	30000
$F(s_1)$	4.0735	4.1970	4.2501	4.2447
$F(s_2$	5.1386	6.4758	8.0768	8.6829
$F(x)$	5.2647	6.6670	8.4009	9.1150

Table 2. Temporal predictability of the source and observation signals, where s_1 and s_2 are the human speech signals recorded at $16k$ sampling rate respectively, and the observation $x = 0.5669\ s_1 - 1.2025\ s_2$.

No. of samples	10000	20000	30000	40000
$F(s_2)$	2.1437	1.7607	1.8414	1.8656
$F(x)$	2.2805	1.8553	1.9131	1.9220

2 A General Form of Independency Metric

Suppose the recovered signal $\mathbf{y}$ is from Eq.(2). Hence, from Eq.(1), we know that each component y of $\mathbf{y}$ is a linear mixture of n sources with:

$$\begin{aligned} y &= c_1 s_1 + c_2 s_2 + \ldots, + c_n s_n, \\ &= \mathbf{c}^T \mathbf{s}, \end{aligned} \quad (3)$$

where $\mathbf{c} = [c_1, c_2, \ldots, c_n]^T$ is an n-dimension nonzero vector. Suppose there exists a function g such that $g(s_1)$, $g(s_2)$, …, $g(s_n)$ are uncorrelated each other, and satisfy

$$g(y) = c_1 g(s_1) + c_2 g(s_2) + \ldots + c_n g(s_n) \quad (4)$$

and

$$\frac{cov(g(s_1))}{cov(s_1)}, \quad \frac{cov(g(s_2))}{cov(s_2)}, \ldots, \frac{cov(g(s_n))}{cov(s_n)} \quad (5)$$

are not equal each other. We then define a general form of Independency Metric as

$$L(y) = \frac{cov(g(y))}{cov(y)}. \quad (6)$$

Subsequently, we have the following Theorem:

Theorem 1 *Suppose the source signals s_1, s_2, …, s_n are uncorrelated each other, and there exists a function g satisfying Eq.(4) so that $g(s_1)$, $g(s_2)$, …, $g(s_n)$ are uncorrelated signals, and $L(s_1)$, $L(s_2)$, …, $L(s_n)$ are not equal each other. We denote*

$$L(s_{i_0}) = \max\{L(s_1), L(s_2), \ldots, L(s_n)\}. \quad (7)$$

For any recovered signal y described in Eq.(3), we then have

$$L(y) \le L(s_{i_0}). \quad (8)$$

When $y \ne a s_{i_0}$, where a is any non-zero constant, then

$$L(y) < L(s_{i_0}). \quad (9)$$

Proof:
Let $g(\mathbf{s}) = [g(s_1), g(s_2), \ldots, g(s_n)]^T$. Since $\mathbf{s}$ and $g(\mathbf{s})$ are both uncorrelated each other, we have the following diagonal matrices:

$$\begin{aligned} cov(\mathbf{s}) &= diag([cov(s_1), \ldots, cov(s_n)]) \quad (10) \\ cov(g(\mathbf{s})) &= diag([cov(g(s_1)), \ldots, cov(g(s_n))]) \end{aligned}$$

where $diag[a_1, a_2, \ldots, a_n]$ denotes the diagonal matrix whose j^{th} main diagonal element is a_j. From the definition of L, we thus have

$$L(s_i) = \frac{cov(g(s_i))}{cov(s_i)}, \quad i = 1, 2, \ldots, n. \quad (11)$$

That is,

$$cov(g(s_i)) = L(s_i) cov(s_i). \quad (12)$$

Following Eq.(10), we then have

$$cov(g(\mathbf{s})) = diag([L(s_1)cov(s_1), \ldots, L(s_n)cov(s_n)]). \quad (13)$$

Furthermore, based on Eq.(3), we have

$$\begin{aligned} cov(y) &= \mathbf{c}^T cov(\mathbf{s}) \mathbf{c} \\ cov(g(y)) &= \mathbf{c}^T cov(g(\mathbf{s})) \mathbf{c}. \end{aligned} \quad (14)$$

Subsequently, from Eq.(7), we obtain

$$\begin{aligned} cov(g(y)) &= \mathbf{c}^T diag[L(s_1)cov(s_1), \ldots, L(s_n)cov(s_n)]\mathbf{c} \\ &\le \mathbf{c}^T diag[L(s_{i_0})cov(s_1), \ldots, L(s_{i_0})cov(s_n)]\mathbf{c} \\ &= L(s_{i_0}) cov(y). \end{aligned} \quad (15)$$

Consequently, we have

$$L(y) = \frac{cov(g(y))}{cov(y)} \leq L(s_{i_0}), \tag{16}$$

where "=" is held if and only if $y = as_{i_0}$ with a being a non-zero constant. Hence, we can recover the source s_{i_0} by maximizing the following contrast function:

$$Q(\mathbf{w}) = \log L(y), \tag{17}$$

where $y = \mathbf{w}^T\mathbf{x}$. After extracting s_{i_0}, we can then extract the other source with the second largest Independency Metric value in the same way. Finally, we can acquire the correct de-mixing matrix $\mathbf{W}$, meanwhile recovering all sources.

2.1 A General ICA Algorithm via Maximizing Independency Metric

Since $L(s_j)$'s are not equal each other, without loss of generality, we assume

$$L(s_1) > L(s_2) > \ldots > L(s_n). \tag{18}$$

We denote the i^{th} column of $\mathbf{W}^T$ is $\mathbf{w}_i$. According to Theorem 1, we therefore have

$$Q(\mathbf{w}_1) = \log L(\mathbf{w}_1\mathbf{x}) \leq \log L(s_1). \tag{19}$$

Since A is an $n \times n$ nonsingular square matrix, thus

$$\max_{w_1 \neq 0} Q(\mathbf{w}_1) = \log L(s_1). \tag{20}$$

Suppose that $\mathbf{w}_0$ is an optimal solution of the following optimization problem:

$$\max_{\mathbf{w}_1 \neq 0} Q(\mathbf{w}_1). \tag{21}$$

As a result,

$$\max_{\mathbf{w}_1 \neq 0} Q(\mathbf{w}_1) = Q(\mathbf{w}_0). \tag{22}$$

In case of $\mathbf{w}_0\mathbf{x} \neq a\mathbf{s}_1$, we obtain $L(\mathbf{w}_0\mathbf{x}) < L(s_1)$ from Theorem 1. Consequently,

$$\max_{\mathbf{w}_1 \neq 0} Q(\mathbf{w}_1) = Q(\mathbf{w}_0) < \log L(s_1) = \max_{\mathbf{w}_1 \neq 0} Q(\mathbf{w}_1), \tag{23}$$

which leads to a contradiction. This implies that the source signal s_1 can be extracted through solving optimization problem in Eq.(20).

Since the function $Q(\mathbf{w}_1)$ is a logarithm of ratio of two quadratic forms, the optimal solutions of Eq.(20) must be a stable point. With some mathematical computations, we can finally obtain the gradient of $Q(\mathbf{w}_1)$:

$$\nabla Q(\mathbf{w}_1) = \frac{2cov(g(\mathbf{x}))\mathbf{w}_1}{\mathbf{w}_1^T cov(g(\mathbf{x}))\mathbf{w}_1} - \frac{2cov(\mathbf{x})\mathbf{w}_1}{\mathbf{w}_1^T cov(\mathbf{x})\mathbf{w}_1}. \tag{24}$$

Let $\nabla Q(\mathbf{w}_1) = 0$, we thus obtain

$$cov(g(\mathbf{x}))\mathbf{w}_1 = L(\mathbf{w}_1\mathbf{x})cov(\mathbf{x})\mathbf{w}_1. \tag{25}$$

Note that $cov(\mathbf{x})$ and $cov(g(\mathbf{x}))$ are all positive definite matric. Solving Eq. (25) actually becomes a generalized eigenvalue problem. Through solving it, we can obtain $\mathbf{w}_1$, which is an eigenvector corresponding to the maximum eigenvalue in Eq.(20).

According to the property of generalized eigenvalue problem, eigenvector corresponding to second large eigenvalue is $\mathbf{w}_2$, eigenvector corresponding to third large eigenvalue is $\mathbf{w}_3$, and so on. Finally, we can obtain a correct $\mathbf{W}$, whereby all of the source signals are recovered. In the next section, we will give out a specific g function, whereby the optimal solution of $\mathbf{W}$ is acquired.

3 A Detailed Implementation of Independency Metric

Suppose each recovered signal y is a function of time t. We then choose $g(y) = \int_0^t ydt$, which can be further approximated by

$$\int_0^t ydt = \sum_{i=1}^{t} \frac{y_i}{N}, \tag{26}$$

where N is the number of samples, and y_i's are the instances of y over the integral range. It can be seen that such a g function satisfies the requirement in Section 2. By putting it into Eq.(16) and Eq.(17), Eq.(24) subsequently becomes

$$\nabla Q(\mathbf{w}_1) = \frac{2cov(\int_0^t \mathbf{x}dt)\mathbf{w}_1}{\mathbf{w}_1^T cov(\int_0^t \mathbf{x}dt)\mathbf{w}_1} - \frac{2cov(\mathbf{x})\mathbf{w}_1}{\mathbf{w}_1^T cov(\mathbf{x})\mathbf{w}_1}. \tag{27}$$

Thus, let $\nabla Q(\mathbf{w}_1) = 0$, we then obtain

$$cov(\int_0^t \mathbf{x}dt)\mathbf{w}_1 = L(\mathbf{w}_1\mathbf{x})cov(\mathbf{x})\mathbf{w}_1, \tag{28}$$

which can therefore be solved by formulating it as a generalized eigenvalue problem as shown in Section 2.1.

4 Simulation Results

To investigate the performance of the proposed Independency Metric, two experiments were conducted in this section, in which sources signals are a combination of super-Gaussian and sub-Gaussian signals with at most one Gaussian signal.

4.1 Experiment 1

This experimental simulation comes from [12] (URL: $www.shef.ac.uk/\tilde{}pc1jvs/$). There are three independent source signals: a sub-Gaussian signal denoted as s_1 (a sine signal), a super-Gaussian signal s_2 (a speech sound), and a Gaussian signal s_3, and temporal structure was imposed on the signal by sorting its values in the ascending order (Details can be seen in [12]). In this experiment, the mixing matrix $\mathbf{A}$ was randomly generated, and the number of samples was $5,000$. With the de-mixing matrix $\mathbf{W}$ achieved by using Independency Metric, the final correlation between source signals $\mathbf{s}$ and recovered signals $\mathbf{y}$ is

$$Corr(\mathbf{s}, \mathbf{y}) = \begin{pmatrix} 0.0000 & \mathbf{1.0000} & 0.0000 \\ \mathbf{0.9998} & 0.0172 & 0.0022 \\ 0.0036 & 0.0413 & \mathbf{0.9991} \end{pmatrix}. \quad (29)$$

Figure 1(a) shows the first $1,000$ samples of observations, and Figure 1(b) shows the source signals (solid line), and its corresponding recovered signals (dot line) obtained via Independency Metric. It can be seen that maximizing Independency Metric has successfully separated those sources.

4.2 Experiment 2

In this experiment, we used two speech signals as sources, which was from a man and a woman, respectively. The sampling rate is 16kHz and are 10 seconds long (i.e., $160,000$ samples in total). These two source speeches were mixed by a randomly generated mixing matrix $\mathbf{A}$. The final correlation between source signals and recovered signals using Independency Metric is

$$Corr(\mathbf{s}, \mathbf{y}) = \begin{pmatrix} 0.0330 & \mathbf{0.9995} \\ \mathbf{0.9995} & -0.0330 \end{pmatrix}. \quad (30)$$

Figure 2(a) shows the first $1,000$ samples of each mixture, while Figure 2(b) gives out each source signal (solid line) and its corresponding recovered signal (dot line) obtained via Independency Metric. Once again, it can be seen that Independency Metric has successfully recovered the sources.

5 Conclusion

In this paper, we have presented a novel Independency Metric for ICA. It has been mathematically proved that the metric value of any linear combination of source signals is less than the largest one of sources under a loose condition. Further, the metric optimization can be formulated as a generalized eigenvalue problem, whereby an optimal solution is just those eigenvectors. Subsequently, we can guarantee to find out a correct de-mixing matrix through maximizing the proposed metric to separate the sources. The simulation results have shown its success in separating the linear combinations of sub-Gaussian and super-Gaussian sources with at most one Gaussian signal.

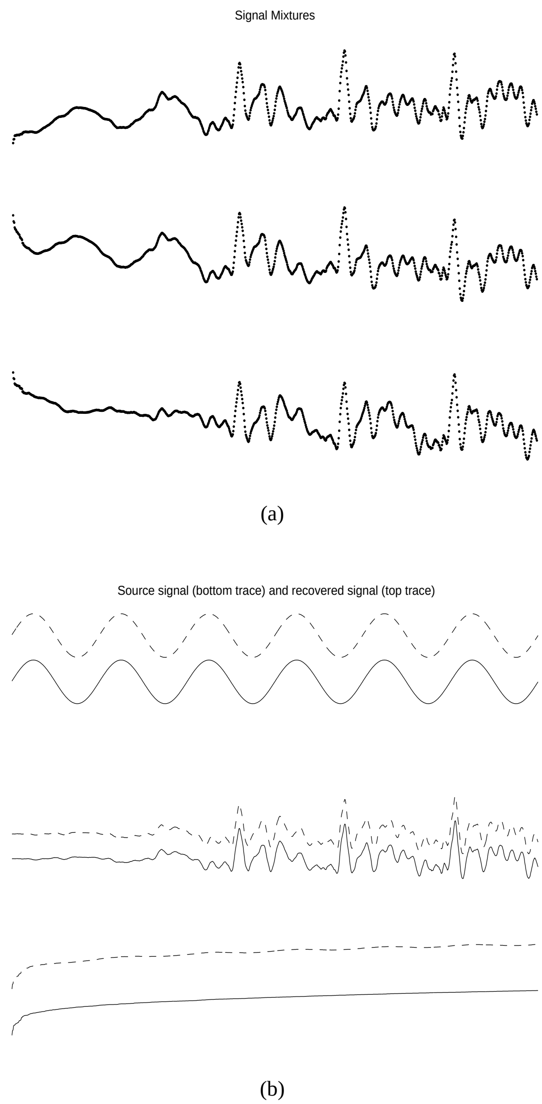
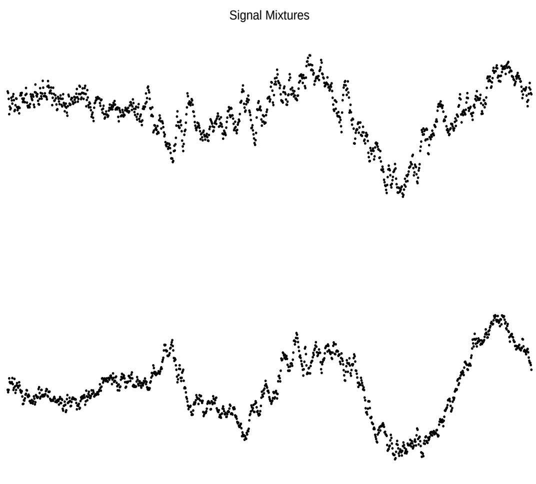


Figure 1. (a). The mixture of three source signals, and (b). the original sources and the recovered signals.

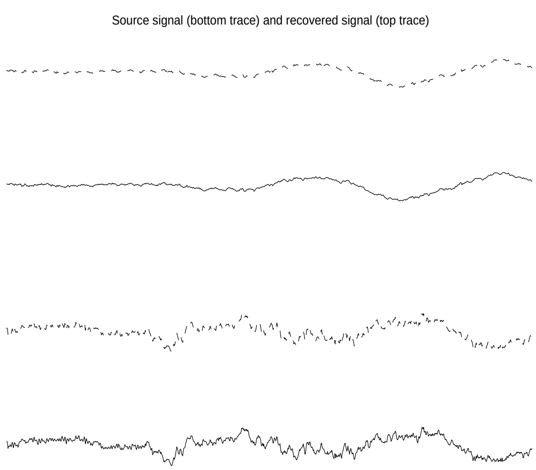
References

[1] C. Jutten and J. Herault, Blind separation of sources Part I: An adaptive algorithm based on neuromimetic architecture, *Signal Processing*, 24, 1991, 1–10.

[2] A. J. Bell and T. J. Sejnowski, An information approach to blind separation and blind deconvolution, *Neural Computation*, 7(6), 1995, 1129–1159.



(a)



(b)

Figure 2. (a). The mixture of two source signals, and (b). the original sources and the recovered signals.

[3] S. Amari, A. Cichocki, H. H. Yang, A new learning algorithm for blind signal separation, *Advances in Neural Information Processing Systems,* Cambridge, MA, 1997, 657–663.

[4] M. Gaeta and J.L. Lacoume, Source separation without prior knowledge: the maximum likelihood solution, *Proceedings of EUSIPO*, 1990, 621–624.

[5] D.T. Pham, P. Garrat and C. Jutten, Separation of a mixture of independent sources through a maximum likelihood approach, *Proceedings of EUSIPCO*, 1992, 771–774.

[6] J.F. Cardoso, Infomax and maximum likelihood for blind source separation, *IEEE Signal Processing Letters*, 4(4), 1997, 112–114.

[7] M. Girolami and C. Fyfe, Extraction of independent signal sources using a deflationary exploratory projection pursuit network with lateral inhibition, *IEE Proceedings on Vision, Image and Signal Processing Journal*, 14(5), 1997, 299–306.

[8] T.W. Lee, Independent component analysis — theory and applications, Kluwer Academic Publishers, 1998.

[9] J. Karhunen and J. Jourtsensalo, Representation and separtaion of signals using nonlinear PCA type learning, *Neural Networks*, 7, 1994, 113–127.

[10] E. Oja, The nonlinear PCA learning rule in independent component analysis, *Neurocomputing*, 17, 1997, 1211–1216.

[11] P. Comon, Independent component analysis, a new concept?, *Signal Processing*, 11, 1994, 287–314.

[12] J. V. Stone, Blind source separation using temporal predictability, *Neural Computation*, 13(7), 2001, 1559–1574.