DRO
Deakin University's Research Repository

This is the published version:

Xie, Shengli, Zhou, Guoxu, Yang, Zuyuan and Fu, Yuli 2009, On blind separability based on the temporal predictability method, *Neural Computation*, vol. 21, no. 12, pp. 3519-3531.

Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30059309

Reproduced with the kind permission of the copyright owner.

Copyright : 2009, MIT Press Journals

dro.deakin.edu.au
Deakin University CRICOS Provider Code: 00113B

On Blind Separability Based on the Temporal Predictability Method

Shengli Xie
adshlxie@scut.edu.cn
Guoxu Zhou
zhouguoxu@gmail.com
Zuyuan Yang
yangzuyuan@yahoo.com.cn
Yuli Fu
fuyuli@scut.edu.cn
School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510640, China

This letter discusses blind separability based on temporal predictability (Stone, 2001; Xie, He, & Fu, 2005). Our results show that the sources are separable using the temporal predictability method if and only if they have different temporal structures (i.e., autocorrelations). Consequently, the applicability and limitations of the temporal predictability method are clarified. In addition, instead of using generalized eigendecomposition, we suggest using joint approximate diagonalization algorithms to improve the robustness of the method. A new criterion is presented to evaluate the separation results. Numerical simulations are performed to demonstrate the validity of the theoretical results.

1 Introduction

Blind source separation (BSS) aims at recovering the underlying sources from their mixtures (observations) (Cichocki & Amari, 2002), where "blind" means that both the sources and the mixing parameters are unknown. Its linear instantaneous model is

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t),\ \ t = 1, 2, \ldots, T. \tag{1.1}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ is the mixing matrix, $\mathbf{s}(t)$ is the source vector, and $\mathbf{x}(t)$ is the observation vector at time instant t, respectively. It is known that when $m > n$, principal component analysis (PCA) can be used to reduce the dimensionality of the observations. For this reason, we consider only the

Neural Computation **21, 3519–3531** (2009) © 2009 Massachusetts Institute of Technology

case that $m = n$ and $\mathbf{A}$ is invertible in this letter. Then BSS can be achieved by finding an unmixing matrix $\mathbf{W}$ such that

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) = \mathbf{W}\mathbf{A}\mathbf{s}(t) = \mathbf{P}\mathbf{D}\mathbf{s}(t), \tag{1.2}$$

where $\mathbf{P}$ is a permutation matrix and $\mathbf{D}$ is a diagonal matrix. In other words, the sources are recovered up to a permutation factor and a scaling factor.

Many BSS methods have been proposed so far. These methods include minimum mutual information (MMI) (Bell & Sejnowski, 1995; Amari, Cichocki, & Yang, 1996), maximum likelihood (ML) (Cardoso, 1997), and maximization of nongaussianity (Hyvarinen, 1999). They are generally referred to as independent component analysis (ICA), which is based on the assumption that the sources are mutually independent (Common, 1994; Hyvarinen, Karhunen, & Oja, 2001; Cichocki & Amari, 2002; Stone, 2002). Another category of BSS methods is based on second-order statistics (SOS) of sources, which can be found in Tong, Liu, Soon, and Huang (1991), Molgedey and Schuster (1994), Belouchrani, AbedMeraim, Cardoso, and Moulines (1997), Ziehe and Muller (1998), Nuzillard and Nuzillard (2003), and Blaschke, Berkes, and Wiskott (2006). A novel method based on temporal predictability was also proposed, where sources and their mixtures were believed to have distinct temporal predictability (Stone, 2001). Like the traditional SOS methods, the temporal predictability method needs only to assume that the sources are uncorrelated. It does not need to estimate the probability density functions and can separate supergaussian signals and subgaussian signals simultaneously. These features make it an attractive method in BSS (Stone, Porrill, Porter, & Wilkinson, 2002; Hu et al., 2005; Song et al., 2006; Jia & Qian, 2007; Ye & Li, 2007). The temporal predictability method is based on the conjecture that the temporal predictability of any signal mixture is less than (or equal to) that of any of its component source signals. However, the conjecture is not rigorous. An essentially equivalent concept of covariance rate was proposed by Xie, He, and Fu (2005). It was proved that the covariance rate of a mixture signal is between the maximal and minimal covariance rates of the sources.

The main contribution of this letter is the in-depth separability analysis of BSS based on temporal predictability. Our results show that the sources are separable by the temporal predictability method if and only if they have different temporal structures (i.e., autocorrelations). Then the validity of the temporal predictability method is proved theoretically. In addition, joint approximate diagonalization algorithms are suggested, which significantly improves the robustness of the method. A new criterion is also proposed to evaluate the reliability of separation without knowledge of the sources and the mixing matrix. Finally, all theoretical results are verified by numerical simulations.

In this letter, we assume that the sources are stationary ergodic and uncorrelated. Also, the sources have zero mean and unit variance without loss of generality, as this condition can always be met by appropriately normalizing the sources.

2 Temporal Predictability and Covariance Rate

Given a signal $s(t)$, a new random process is defined as (Stone, 2001; Xie et al., 2005)

$$f_s^{(\lambda)}(t) = s(t) - \bar{s}(t);\\ \bar{s}(t) = \lambda\bar{s}(t-1) + (1-\lambda)s(t-1),\ \bar{s}(1) = s(1),\ 0 \le \lambda \le 1.^1 \tag{2.1}$$

Then, the temporal predictability of $s\,(t)$ is defined as (Stone, 2001)

$$r_s = \log\left\{\sum_{t=1}^{T}\left[f_s^{(\lambda_L)}(t)\right]^2 \Big/ \sum_{t=1}^{T}\left[f_s^{(\lambda_S)}(t)\right]^2\right\},\\ \lambda_L = 2^{-1/h_L},\ \lambda_S = 2^{-1/h_S} \tag{2.2}$$

where $0 \le h_S \ll h_L$ are parameters. Later, Xie et al. (2005) modified formula 2.2 and defined so-called covariance rate as follows:

$$\gamma_s = \operatorname{cov}\left(f_s^{(\lambda_L)}(t), f_s^{(\lambda_L)}(t)\right)/\operatorname{cov}\left(f_s^{(\lambda_S)}(t), f_s^{(\lambda_S)}(t)\right), \tag{2.3}$$

where $\operatorname{cov}(f_s^{(\lambda_L)}(t), f_s^{(\lambda_L)}(t))$ and $\operatorname{cov}(f_s^{(\lambda_S)}(t), f_s^{(\lambda_S)}(t))$ are the covariance of $f_s^{(\lambda_L)}(t)$ and $f_s^{(\lambda_S)}(t)$, respectively. After simple calculations we have

$$f_s^{(\lambda)}(t) = \sum_{k=1}^{t-1} \lambda^{t-k-1}[s(k+1) - s(k)],\ \forall\lambda \in [0, 1]. \tag{2.4}$$

We see $f_s^{(\lambda)}(t)$ is nothing but a weighted sum of the difference signal of $s(t)$. From equation 2.4, the relation between temporal predictability and the covariance rate is clarified:

[1] **Remark**. The initial value of $\bar{s}_1$ affects the properties of $f_s^{(\lambda)}(t)$ slightly. As shown by the theorem 1, if $\bar{s}(1) = s(1)$, then $E[f_s^{(\lambda)}(t)] = 0$ holds regardless of whether the mean value of $s(t)$ is subtracted beforehand. Otherwise if $\bar{s}(1) = 0$, $f_s^{(\lambda)}(t)$ has an attractive property that $f_s^{(0)}(t) = s(t) - s(t-1)$ and $f_s^{(1)}(t) = s(t)$. However, in this case, the mean value of the signal should be subtracted first to ensure a successful separation.

Theorem 1. *For the random process defined in equation 2.1, we have* $E[f_s^{(\lambda)}(t)] = 0$ *for any* $\lambda \in [0, 1]$. *Furthermore,* $\gamma_s > 0$ *and* $r_s = \log \gamma_s$.

Proof. From equation 2.4, $E[f_s^{(\lambda)}(t)] = \sum_{k=1}^{t-1} \lambda^{t-k-1}[E[s(k+1)] - E[s(k)]] = 0$. Then $\mathrm{cov}(f_s^{(\lambda)}(t), f_s^{(\lambda)}(t)) = E[f_s^{(\lambda)}(t) - E(f_s^{(\lambda)}(t))]^2 = E[f_s^{(\lambda)}(t)]^2$, so, $\gamma_s > 0$ and $r_s = \log \gamma_s$.

From theorem 1, we know that there is a one-to-one correspondence between the temporal predictability and the covariance rate of a signal. Thus, for simplicity, henceforth we refer to the covariance rate rather than temporal predictability.

According to equation 2.1, we further define that $\mathbf{F}_\mathbf{s}^{(\lambda)}(t) = [f_{s_1}^{(\lambda)}(t), f_{s_2}^{(\lambda)}(t), \ldots, f_{s_n}^{(\lambda)}]^\mathrm{T}$. For simplicity, the autocovariance matrices of $\mathbf{F}_\mathbf{s}^{(\lambda)}(t)$ and $\mathbf{s}(t)$ are denoted by $\mathbf{C}_\mathbf{s}^\lambda$ and $\mathbf{C}_\mathbf{s}$, respectively. (Because the signals have zero mean, the covariance equals the correlation matrix. For this reason, we do not make any distinction between the two in this letter.) The covariances of $f_{s_i}^{(\lambda)}(t)$ and $s_i(t)$ are denoted by $c_{s_i}^\lambda$ and c_{s_i}, respectively. Then if the sources are uncorrelated, the following properties hold (Xie et al., 2005):

i. $\mathbf{C}_\mathbf{s}^\lambda = diag(c_{s_1}^\lambda, c_{s_2}^\lambda, \cdots, c_{s_n}^\lambda)$
ii. If $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$, then $\mathbf{C}_\mathbf{x}^\lambda = \mathbf{A}\mathbf{C}_\mathbf{s}^\lambda\mathbf{A}^\mathrm{T}$,

where $\mathbf{C}_\mathbf{s}^\lambda = diag(c_{s_1}^\lambda, c_{s_2}^\lambda, \ldots, c_{s_n}^\lambda)$ is a diagonal matrix whose diagonal entries are $c_{s_i}^\lambda$, $i = 1, 2, \ldots, n$. Xie et al. gave a theorem to present the modified conjecture: the covariance rate of a signal mixture is between the maximal and minimal covariance rates of its sources—$r_{s_1} \leq r_{x_i} \leq r_{s_n}$ (assuming that $r_{s_1} \leq r_{s_2} \cdots \leq r_{s_n}$)—or, equivalently, $\log \gamma_{s_1} \leq \log \gamma_{x_i} \leq \log \gamma_{s_n}$. From the analysis, the temporal predictability method and the covariance rate method share the same cost function:

$$\max_{\mathbf{w}} \ \gamma(\mathbf{w}^\mathrm{T}\mathbf{x}(t)) = \frac{\mathbf{w}^\mathrm{T}\mathbf{C}_\mathbf{x}^{\lambda_L}\mathbf{w}}{\mathbf{w}^\mathrm{T}\mathbf{C}_\mathbf{x}^{\lambda_S}\mathbf{w}}. \tag{2.5}$$

From the fact that $r_{s_1} \leq r_{x_i} \leq r_{s_n}$, the max operator in equation 2.5 can also be replaced by the min operator. Obviously the generalized eigenvectors of $(\mathbf{C}_\mathbf{x}^{\lambda_L}, \mathbf{C}_\mathbf{x}^{\lambda_S})$ are the stationary points of equation 2.5. Note also that the generalized eigenvector matrix plays the role of jointly diagonalizing the matrices of $\mathbf{C}_\mathbf{x}^{\lambda_L}$ and $\mathbf{C}_\mathbf{x}^{\lambda_S}$. Motivated by this and the expression of $\mathbf{C}_\mathbf{x}^\lambda = \mathbf{A}\mathbf{C}_\mathbf{s}^\lambda\mathbf{A}^\mathrm{T}$, joint approximate diagonalization algorithms can be used to improve the robustness of the method. This will be detailed in section 4.

3 Separability Analysis

Let $\mathbf{w}$ be a generalized eigenvector of $(\mathbf{C}_{\mathbf{x}}^{\lambda_L}, \mathbf{C}_{\mathbf{x}}^{\lambda_S})$, that is, $\mathbf{w}$ is a stationary point of equation 2.5. The following theorem shows when $\mathbf{w}^{\mathrm{T}}\mathbf{x}(t)$ is a true source:

Theorem 2 (separability theorem). *If $\left(\mathbf{C}_{\mathbf{x}}^{\lambda_L}, \mathbf{C}_{\mathbf{x}}^{\lambda_S}\right)$ has k ($k \leq n$) distinct generalized eigenvalues, then a total of k sources can be separated. Furthermore, $\mathbf{w}_j^{\mathrm{T}}\mathbf{x}(t)$ must be a source if μ_j is a distinct generalized eigenvalue of $\left(\mathbf{C}_{\mathbf{x}}^{\lambda_L}, \mathbf{C}_{\mathbf{x}}^{\lambda_S}\right)$, where $\mathbf{w}_j$ is the generalized eigenvector associated with μ_j, $j = 1, 2, \ldots, k$.*

The proof is given in the appendix. (We call an eigenvalue μ_j *distinct* if it is not a repeated generalized eigenvalue.)

Theorem 2 is parallel to the identifiability theorems for the traditional SOS methods. (See theorem 2 in Tong et al., 1991, and theorem 2 in Belouchrani et al., 1997, respectively.) It gives not only criteria of full or partial separability but also a separation method. From the theorem, if repeated eigenvalues exist, some sources cannot be recovered by the associated eigenvectors. To see this, recall that the eigenvectors associated with repeated eigenvalues are not unique. However, the eigenvalues of $(\mathbf{C}_{\mathbf{x}}^{\lambda_L}, \mathbf{C}_{\mathbf{x}}^{\lambda_S})$ are affected by the values of λ_L, λ_S. Is it possible to avoid repeated eigenvalues by the elaborate selection of parameters λ_L, λ_S? This question is answered by the following three theorems.

Theorem 3. *$(\mathbf{C}_{\mathbf{x}}^{\lambda_L}, \mathbf{C}_{\mathbf{x}}^{\lambda_S})$ has k distinct generalized eigenvalues if and only if there are k sources which have distinct covariance rates.*

The proof is given in the appendix.

From theorem 3, if two sources have equal covariance rates, the corresponding eigenvalues are equal. But when will two signals have equal covariance rates? From theorem 1, after some simple calculations we have

$$c_s^{\lambda} = L_0(c_s(\tau)) + \sum_{k=1}^{2(t-2)} L_k(c_s(\tau))\lambda^k, \;\; \lambda \in [0, 1], \tag{3.1}$$

where $L_k(c_s(\tau))$ are linear combinations of $c_s(\tau)$, $c_s(\tau) = E[s(t)s(t+\tau)]$ and $\tau = 1, 2, \ldots, T-1$ denotes time lags. From equation 3.1, we have:

Theorem 4. *$\gamma_{s_i}(\lambda_L, \lambda_S) = \gamma_{s_j}(\lambda_L, \lambda_S)$ holds for any $0 \leq \lambda_L, \lambda_S \leq 1$ if and only if $c_{s_i}(\tau) = c_{s_j}(\tau)$ for all $\tau = 1, 2, \ldots, T-1$.*

The proof is presented in the appendix.

Although $f_s^{(\lambda)}(t)$ depends on the parameter λ, the equivalence of covariance rates of two signals is simply caused by the fact that the two sources have the same temporal structure. If there exists a time lag τ_0 such that $c_{s_i}(\tau_0) \neq c_{s_j}(\tau_0)$, there must be a pair of λ_L, λ_S, which makes s_i and s_j have different covariance rates; otherwise, they have equal covariance rates for any pair of λ_L, λ_S. Consequently, the following result yields immediately:

Corollary 1. *The sources are separable in the sense of covariance rate if and only if they have different temporal structure.*

From theorem 4 and the corollary, covariance rate and temporal predictability are essentially two measures of the temporal structure of a signal. First,

$$\gamma_{s_i}(1, 0) = \frac{E[(s_i(t) - s_i(t-1))^2]}{E[(s_i(t) - s_i(1))^2]}.$$

Note that $s_i(t) - s_i(1)$ is a shifted version of the original source but the waveform is maintained. The term $E[(s_i(t) - s_i(t-1))^2]$ measures the degree of invariance of the source. In this case, minimizing the covariance rate is coincident with linear slow feature analysis (SFA) (Blaschke et al., 2006). If $0 < \lambda < 1$, $c_{s_i}^{\lambda}$ is a weighted sum of the autocovariance $c_{s_i}(\tau)$ of s_i. Temporal predictability, covariance rate, and linear SFA are all based on the second-order statistics, and they extract variant or invariant features of temporally varying signals. In linear determined or overdetermined cases, this kind of feature is generally sufficient to extract a source that has a distinct variant feature. However, in traditional SOS-based algorithms, the time lags are required to be appropriately selected for the existence of equivalent time-delay correlations of sources (Tong et al., 1991; Belouchrani et al., 1997; Blaschke et al., 2006). This problem can be avoided in the temporal predictability method, since any two sources have different covariance rates as long as they have different temporal structure. Consequently, the temporal predictability method is expected to be more reliable than the traditional SOS methods theoretically.

4 Algorithms and Evaluation

As mentioned in section 2, the generalize eigendecomposition (GE) procedure in the temporal predictability method can be replaced by any joint approximate diagonalization (JAD) algorithm. Given a set of matrices of $\{\mathbf{R}_k = \mathbf{A}\Lambda_k\mathbf{A}^{\mathrm{T}} | k = 1, 2, \ldots, K\}$, JAD is a problem of finding a matrix $\mathbf{W}$, named a diagonalizer, which makes $\mathbf{W}\mathbf{R}_k\mathbf{W}^{\mathrm{T}}$ as diagonal as possible for all k (Yeredor, 2002; Ziehe, Laskov, Nolte, & Muller, 2004; Vollgraf & Obermayer, 2006). In the case where $\mathbf{A}$ is of full column rank, a diagonalizer $\mathbf{W}$ is generally an estimation of $\mathbf{A}^{-1}$ up to a permutation and scaling of the columns. State-of-the-art JAD algorithms, such as FFDIAG and QDiag, are quite efficient (Ziehe et al., 2004; Vollgraf & Obermayer, 2006).

Taking the covariance rate method into account, from property ii in section 2, we know that $\mathbf{C}_{\mathbf{x}}^{\lambda_k} = \mathbf{A}\mathbf{C}_{\mathbf{s}}^{\lambda_k}\mathbf{A}^{\mathrm{T}}$, where $\mathbf{C}_{\mathbf{s}}^{\lambda_k}$ is a diagonal matrix for all $k \in \mathcal{K}$, $\mathcal{K} = \{1, 2, \ldots, K, K \geq 2\}$. Thus, the unmixing matrix can be obtained by jointly diagonalizing the set of matrices $\{\mathbf{C}_{\mathbf{x}}^{\lambda_k} | k \in \mathcal{K}\}$. Naturally this method should be more robust than the GE-based method, because it can make use of reasonably comprehensive information about SOS and avoid the failure caused by inappropriate setting of λ_L and λ_S.

Regarding the evaluation of separation results, currently, signal-to-noise ratio (SNR) is a widely used performance index (Cichocki & Amari, 2002),

$$\mathrm{SNR}(s, y) = 10 \log \frac{E[s^2]}{E[(y - s)^2]} \ (dB), \tag{4.1}$$

where y is an estimation of s, and s, y are normalized to be of zero mean and unit variance. SNR compares output signals with sources directly. However, the sources are unknown in BSS. Therefore, it is meaningful to design a performance index that can evaluate separation results only from the observation signals.

For the methods based on temporal predictability, the separability of sources depends on the distinctness of the generalized eigenvalues of $(\mathbf{C}_{\mathbf{x}}^{\lambda_L}, \mathbf{C}_{\mathbf{x}}^{\lambda_S})$. From this fact, a new performance index is suggested. We first define the relative distance between x and y as follows:

$$d(x, y) = 1 - \exp\left(-10 \times \frac{2|y - x|}{|x| + |y|}\right), \tag{4.2}$$

where $|y - x|$ is the absolute distance between x and y and $\frac{|x|+|y|}{2}$ specifies the order of magnitude. Obviously $d(x, y) \geq 0$, $d(x, y) = d(y, x)$, $d(x, y) = 0$ if and only if $x = y$. This definition is reasonable in practice. For example, the absolute distance between 1.1 and 1.2 and that of 100.1 and 100.2 is the same. However, the latter is more likely to be equal if the influence of noise is taken into account. Then $d(1.1, 1.2) \approx 0.5809 > 0.0099 \approx d(100.1, 100.2)$. Suppose that $x \in \mathcal{R}$. The distinction index of x in $\mathcal{R}$ is then defined by

$$d(x, \mathcal{R}) = \min_{y \in \mathcal{R}\backslash\{x\}} d(x, y). \tag{4.3}$$

Empirically, x can be distinguished from the entries of $\mathcal{R}\backslash\{x\}$ when $d(x, \mathcal{R}) > 0.4$ (For the case that $x > 0$, there often is a good distinction between x and y if $y < (1 - 5\%)x$ or $y > (1 + 5\%)x$, that is, $d(x, y) > 0.4$ approximately). The distinction index can be used to measure the reliability of a separation:

Table 1: Separability of Two Groups of Sources.

		$y1$	$y2$	$y3$
Group I	SNR(dB)	5.3684	3.7598	8.4652
	RI	0.1243	0.1228	0.1391
Group II	SNR(dB)	49.8595	40.1868	53.4512
	RI	0.7984	0.7485	0.9950

Definition 1. *The reliability index (RI) that characterizes the level of how* $\mathbf{w}_i^T\mathbf{x}(t)$ *is a source is defined as* $d(\mu_i, \mathcal{R})$, *where* $\mathcal{R} = \{\mu_i | i = 1, 2, \ldots, n\}$ *and* $\mu_i, \mathbf{w}_i$ *are the generalized eigenvalues and the associated eigenvectors of* $(\mathbf{C}_\mathbf{x}^{\lambda_L}, \mathbf{C}_\mathbf{x}^{\lambda_S})$, *respectively.*

A larger value of *RI* means that the corresponding output is more likely to be a source. Different from SNR, *RI* is evaluated from the generalized eigenvalues of $(\mathbf{C}_\mathbf{x}^{\lambda_L}, \mathbf{C}_\mathbf{x}^{\lambda_S})$; thus, only the observation is needed. In JAD, we can simply let $\lambda_L = \min_{k\in\mathcal{K}} \lambda_k$ and $\lambda_S = \max_{k\in\mathcal{K}} \lambda_k$, or select a typical pair from $\mathcal{K}$.

It is worthwhile to note that if some values of μ_i are zero or nearly zero, where the corresponding signals have particularly small energy, *RI* may fail to measure the reliability of the separation. To see this, note that $d(x, y) \approx 1$ holds for any $y \neq 0$ if $x = 0$. Therefore, we further suggest checking the value of μ_i to avoid this exception, even if $d(\mu_i, \mathcal{R})$ is reasonably large.

5 Experiments

As mentioned by Stone (2001), the temporal predictability method succeeds in many examples. Here another three experiments are presented to illustrate the special properties discussed in this letter. In each run, a new mixing matrix **A** (see equation 1.1) is generated by the randn function in Matlab.

5.1 Experiments on Separability. There are two groups of signals. Group I consists of three uniformly distributed signals, and group II consists of three speech signals. The number of samples is 160,000. The unmixing matrix is estimated by GE algorithm with $\lambda_L = 1$ and $\lambda_S = 0.1$. The results are shown in Table 1.

From Table 1, we see that the sources in group I have not been recovered. The SNR of each estimated signal is less than 10 dB, and the corresponding RI is also very small. The sources in group II are separated successfully. The SNR of each estimated signal is higher than 40 dB, and the corresponding RI is greater than 0.7. We see that RI is able to measure the accuracy of a separation.

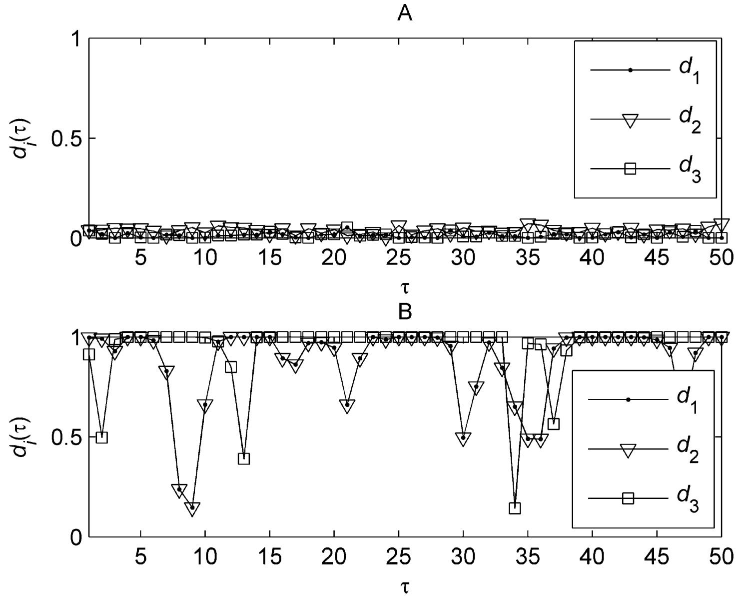

Figure 1: Distinction indices of $c_{s_i}(\tau)$, where $d_i(\tau) = d(c_{s_i}(\tau), C(\tau))$, with $C(\tau) = \{c_{s_i}(\tau)|i = 1, 2, 3\}$, $\tau = 0, 1, \ldots, 50$, $i = 1, 2, 3$.

We now investigate the reason that the sources in group II can be separated while those in group I cannot. Figure 1 is the plot of the distinction index of autocovariance of each source at time lag τ, where $\tau = 1, 2, \ldots, 50$. From Figure 1, by comparison, we see that the uniformly distributed signals in group I have almost the same autocovariance at each time lag (distinction indices are less than 0.1), but the speech signals in group II always have distinct autocovariance (many distinction indices are nearly 1). This is consistent with theorem 4.

5.2 Partial Extraction. The sources consist of two uniformly distributed signals and four sine-wave signals chosen from the ICALAB benchmarks named acsin4d.mat (Cichocki, Amari, Siwek, & Tanaka, 2007) . The number of samples is 1001. The unmixing matrix is estimated by GE algorithm with $\lambda_L = 1$ and $\lambda_S = 0.1$. The RIs of a typical separation are 0.0103, 0.0103, 0.6088, 0.6307, 0.9731, and 1.0000, respectively, from which we can infer that the last four signals are the sources while the first two are not. This is consistent with the fact that the corresponding SNRs (dB) are 2.5383, 2.9742, 61.1167, 61.5457, 65.4071, and 73.6396, respectively.

5.3 Experiments on the JAD-Based Covariance Rate Method. Six sources are considered here: $s1 = sign(\cos(2\pi 155t))$, $s2 = \sin(2\pi 800t)$,

Table 2: Correlations Between Sources and Recovered Signals Using FFDIAG.

Signal Recovered	Source Signals $s1$	$s2$	$s3$	$s4$	$s5$	$s6$
$y1$	0.0002	**−1.0000**	−0.0001	0.0000	0.0000	0.0135
$y2$	**1.0000**	0.0001	0.0002	0.0000	0.0004	−0.0220
$y3$	0.0000	0.0001	**−1.0000**	0.0000	0.0000	0.0068
$y4$	0.0009	−0.0001	0.0000	0.0000	**0.9989**	0.0223
$y5$	−0.0003	−0.0002	−0.0001	−0.0017	−0.0447	**0.9994**
$y6$	0.0000	0.0000	0.0000	**−1.0000**	0.0001	0.0071

Note: Each source signal has a high correlation with only one recovered signal, indicated by the bold figures.

Table 3: Comparison Between the GE-Based Algorithm and the JAD-Based Algorithm.

Method	SNR(dB)					
SOBI	26.4109	29.5410	54.4508	60.9902	68.4412	76.1652
GE(0,1)	7.8066	8.3362	53.1191	62.5875	66.0564	77.0640
GE(0.3,1)	22.3316	24.5952	41.7050	44.9689	61.8432	71.9787
JAD-CR	31.0081	33.2351	54.5361	59.7855	69.3151	79.1490

$s3 = \sin(2\pi 90t)$, $s4 = \sin(2\pi 9t)\sin(2\pi 300t)$, $s5 = \sin(2\pi 300t + 6\cos(2\pi 60t)$, and $s6$ is a uniformly distributed signal between -1 and 1. $t = 1 : 0.001 : 10$ (in Matlab code). Five covariance matrices are generated by setting $\lambda = 0.1, 0.3, \ldots, 0.9$. The FFDIAG algorithm is employed to diagonalize the set of covariance matrices (Ziehe et al., 2004). Correlations between the sources and the recovered signals are shown in Table 2.

A comparison of the JAD-based covariance rate method (JAD-CR) with the GE-based method and the classical SOS method, that is, the SOBI method (Belouchrani et al., 1997), is presented. The SOBI algorithm uses five time-lagged correlation matrices for joint diagonalization, where the time lags are 0,1,. . .,4, respectively. All the obtained results have been averaged over 1000 Monte Carlo trials and are shown in Table 3, where GE(λ_L, λ_S) denotes the generalized eigendecomposition of $(\mathbf{C}_{\mathbf{x}}^{\lambda_L}, \mathbf{C}_{\mathbf{x}}^{\lambda_S})$.

From Table 3, we can see that the values of λ_L and λ_S affect the accuracy of separation when GE is employed. However, for a given set of signals and fixed λ_L and λ_S, we find that the RIs and SNRs almost do not change, even if the mixing matrix changes. Also, JAD improves the separation robustness at the cost of more time. Currently JAD algorithms can almost achieve the same efficiency as the GE algorithm. Thus, for the JAD-based covariance rate method, the extra time is spent mainly on generating more signals to calculate the covariance matrices. From Table 3, we see that the separation result is worthy of this extra time cost.

One may question why JAD-CR does not outperform the SOBI algorithm evidently. In fact, if the time lags are appropriately selected, the SOBI algorithm is simply equivalent to the temporal predictability method. Otherwise, it will perform worse than the temporal predictability method. For the example in this experiment, the time lags are easy to set, and thus both JAD-CR and SOBI can perform well.

6 Conclusion

A comprehensive theoretical analysis of BSS was conducted based on temporal predictability. Both full and partial separability conditions were given, and the reason of inseparability was also investigated. Furthermore, to improve the performance of the results, JAD algorithms were suggested to replace the GE algorithm. All of these theoretical and technical results make the temporal predictability method more reliable and valid.

Appendix

Proof of Theorem 2. Let μ_i, $i = 1, 2, \ldots, n$ are the generalized eigenvalues of $(\mathbf{C}_{\mathbf{x}}^{\lambda_L}, \mathbf{C}_{\mathbf{x}}^{\lambda_S})$, and the associated eigenvector matrix is $\mathbf{W}$. Thus, $\mathbf{C}_{\mathbf{x}}^{\lambda_L}\mathbf{W} = \mathbf{C}_{\mathbf{x}}^{\lambda_S}\mathbf{W}\mathbf{M}$ holds, where $\mathbf{M} = diag(\mu_1, \mu_2, \ldots, \mu_n)$. From property ii, $\mathbf{A}\mathbf{C}_{\mathbf{s}}^{\lambda_L}\mathbf{A}^{\mathrm{T}}\mathbf{W} = \mathbf{A}\mathbf{C}_{\mathbf{s}}^{\lambda_S}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{M}$ holds. So

$$\left(\mathbf{C}_{\mathbf{s}}^{\lambda_S}\right)^{-1}\mathbf{C}_{\mathbf{s}}^{\lambda_L}\mathbf{A}^{\mathrm{T}}\mathbf{W} = \mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{M}. \tag{A.1}$$

Let $\mathbf{U} = \mathbf{A}^{\mathrm{T}}\mathbf{W}$, $\mathbf{\Gamma} = (\mathbf{C}_{\mathbf{S}}^{\lambda_S})^{-1}\mathbf{C}_{\mathbf{S}}^{\lambda_L} = diag(\gamma_{s_1}, \gamma_{s_2}, \ldots, \gamma_{s_n})$. Then equation A.1 can be rewritten as

$$\mathbf{\Gamma}\mathbf{U} = \mathbf{U}\mathbf{M} \text{ or } \gamma_i u_{ij} = u_{ij}\mu_j \text{ for any } i,\ j. \tag{A.2}$$

Suppose that μ_j is distinct. Note that $\mathbf{U}$ is invertible. Thus, there is at least one nonzero entry in the *j*th column of $\mathbf{U}$. Without loss of generality, assume that $u_{ij} \neq 0$. From equation A.2, $\gamma_i = \mu_j$.

1. u_{ij} is the one and only nonzero entry in the *i*th row of $\mathbf{U}$. To see this, assume that there is another entry, named $u_{ik} \neq 0$, yielding $\mu_k = \gamma_i = \mu_j$, which contradicts the assumption that μ_j is distinct.
2. u_{ij} is also the one and only nonzero entry in the *j*th column of $\mathbf{U}$. Suppose that there is another nonzero entry in the *j*th column of $\mathbf{U}$: $u_{kj} \neq 0$ for some $k \neq i$. Since $u_{kj} \neq 0$ and μ_j is distinct, from point 1, u_{kj} is the only nonzero entry in the *k*th row. Consequently, both the

ith row and the kth row of $\mathbf{U}$ have one and only one nonzero entry in the jth column, which contradicts the invertibility of $\mathbf{U}$.

3. From point 2, $\mathbf{w}_j^{\mathrm{T}}\mathbf{A} = k\mathbf{e}_l$, where k is a nonzero scalar and $\mathbf{e}_l$ is equal to $\mathbf{0}$ except that the lth entry is 1. In other words, $\mathbf{w}_j^{\mathrm{T}}\mathbf{A}\mathbf{s}(t)$ must be a proportion of a source signal.

Proof of Theorem 3. From the proof of theorem 2, u_{ij} is the one and only entry in the ith row and jth column of $\mathbf{U}$. Thus, $\gamma_i = \mu_j$.

Proof of Theorem 4. From equation 3.1, $c_{s_i}^{\lambda} = c_{s_j}^{\lambda}$ holds for any $0 \leq \lambda \leq 1$ if $c_{s_i}(\tau) = c_{s_j}(\tau)$ for all τ. Thus $\gamma_{s_i}(\lambda_L, \lambda_S) = \gamma_{s_j}(\lambda_L, \lambda_S)$.

Conversely, assume that $\gamma_{s_i}(\lambda_L, \lambda_S) = \gamma_{s_j}(\lambda_L, \lambda_S)$ holds for any $0 \leq \lambda_L, \lambda_S \leq 1$. Note that $c_{s_i}^{\lambda}$ is a polynomial with respect to λ and the coefficient associated with λ^k is a linear combination of $c_{s_i}(\tau)$. Note also that $c_{s_i}^{\lambda} = c_{s_j}^{\lambda}$ holds for all $0 \leq \lambda \leq 1$ if and only if their coefficients associated with λ^k are equal for all k. By straightforward calculation, it follows that $c_{s_i}(\tau) = c_{s_j}(\tau)$ for all τ. Let $\gamma_{s_i}(\lambda_L, \lambda_S) = \gamma_{s_j}(\lambda_L, \lambda_S)$ and λ_S is fixed. Thus $c_{s_i}^{\lambda_L} = c_{s_j}^{\lambda_L}$ and consequently, $c_{s_i}(\tau) = c_{s_j}(\tau)$ for all τ.

Acknowledgments

We thank the anonymous reviewers for their insightful comments and valuable suggestions. The work is supported by Key Program of National NSFC (grant U0635001, U0828003), National 973 Project of China (grant 2010CB731800), and National NSFC (grant 60874061, 60674033).

References

Amari, S. I., Cichocki, A., & Yang, H. (1996). A new learning algorithm for blind signal separation of sources. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in neural information processing, 8* (pp 757–763). Cambridge, MA: MIT Press.

Bell, A., & Sejnowski, T. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation, 7*(6), 1129–1159.

Belouchrani, A., AbedMeraim, K., Cardoso, J. F., & Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing, 45*(2), 434–444.

Blaschke, T., Berkes, P., & Wiskott, L. (2006). What is the relation between slow feature analysis and independent component analysis? *Neural Computation, 18*(10), 2495–2508.

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

Cichocki, A., & Amari, S. (2002). *Adaptive blind signal and image processing: learning algorithms and applications*. Hoboken, NJ: Wiley.

Cichocki, A., Amari, S., Siwek, K., & Tanaka, T. (2007). ICALAB toolboxes. Available online at http://www.bsp.brain.riken.jp/ICALAB.

Common, P. (1994). Independent component analysis: A new concept? *Signal Processing, 36*(3), 287–314.

Hu, D. W., Yan, L. R., Liu, Y. D., Zhou, Z. T., Friston, K. J., Tan, C. L., et al. (2005). Unified SPM-ICA for fMRI analysis. *Neuroimage, 25*(3), 746–755.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks, 10*(3), 626–634.

Hyvarinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. Hoboken, NJ: Wiley.

Jia, S., & Qian, Y. T. (2007). Spectral and spatial complexity-based hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing, 45*(12), 3867–3879.

Molgedey, L., & Schuster, H. G. (1994). Separation of a mixture of independent signals using time-delayed correlations. *Physical Review Letters, 72*(23), 3634–3637.

Nuzillard, D., & Nuzillard, J. M. (2003). Second-order blind source separation in the Fourier space of data. *Signal Processing, 83*(3), 627–631.

Song, Q. Y., Yin, F., Chen, H. F., Zhang, Y., Hu, Q. L., & Yao, D. Z. (2006). A novel unified SPM-ICA-PCA method for detecting epileptic activities in resting-state fMRI. In *Advances in natural computation (Pt. 2*, pp. 627–636. Berlin: Springer.

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

Stone, J. V. (2002). Independent component analysis: An introduction. *Trends in Cognitive Sciences, 6*(2), 59–64.

Stone, J. V., Porrill, J., Porter, N. R., & Wilkinson, I. D. (2002). Spatiotemporal independent component analysis of event-related fMRI data using skewed probability density functions. *Neuroimage, 15*(2), 407–421.

Tong, L., Liu, R. W., Soon, V. C., & Huang, Y. F. (1991). Indeterminacy and identifiability of blind identification. *IEEE Transactions on Circuits and Systems, 38*(5), 499–509.

Vollgraf, R., & Obermayer, K. (2006). Quadratic optimization for simultaneous matrix diagonalization. *IEEE Transactions on Signal Processing, 54*(9), 3270–3278.

Xie, S. L., He, Z. S., & Fu, Y. L. (2005). A note on Stone's conjecture of blind signal separation. *Neural Computation, 17*(2), 321–330.

Ye, M., & Li, X. (2007). An efficient measure of signal temporal predictability for blind source separation. *Neural Processing Letters, 26*(1), 57–68.

Yeredor, A. (2002). Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation. *IEEE Transactions on Signal Processing, 50*(7), 1545–1553.

Ziehe, A., Laskov, P., Nolte, G., & Muller, K. R. (2004). A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation. *Journal of Machine Learning Research, 5*, 777–800.

Ziehe, A., & Muller, K. R. (1998). TDSEP—an efficient algorithm for blind separation using time structure. In *Proc. of the 8th International Conference on Artificial Neural Networks (ICANN'98)* (pp. 675–680). Berlin: Springer-Verlag.

Received October 22, 2008; accepted March 27, 2009.

Copyright of Neural Computation is the property of MIT Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.