Stability Analyses of Information-Theoretic Blind Separation Algorithms in the Case Where the Sources Are Nonlinear Processes

Masashi Ohata and Kiyotoshi Matsuoka

***Abstract*—A basic approach to blind source separation is to define an index representing statistical dependency among the output signals of the separator and minimize it with respect to the separator's parameters. The most natural index might be mutual information among the output signals of the separator. In the case of convolutive mixture, however, since the signals must be treated as time series, it becomes very complicated to concretely express the mutual information as a function of the parameters. To cope with this difficulty, in most of the conventional methods, the source signals are assumed to be independent identically distributed (i.i.d.) or linear. Based on this assumption, some simpler indices are defined, and their minimization is made by such an iterative calculation as the gradient method. In actual applications, however, the sources are often not linear processes. This paper discusses what will happen when those algorithms postulating the linearity of the sources are applied to the case of nonlinear sources. An analysis of local stability derives a couple of conditions guaranteeing that the separator stably tends toward a desired one with iteration. The obtained results reveal that those methods, which are based on the minimization of some indices related to mutual information, do not work well when the sources signals are far from linear.**

***Index Terms*—Blind source separation, convolutive mixture, gradient-type algorithm, mutual information, stability analysis.**

I. Introduction

Blind source separation (BSS) is a method for recovering a set of source signals from their mixtures without any knowledge about the mixing process. It has been receiving a great deal of attention from various fields as a new signal processing technique. In view of the level of complexity, the mixing process can be classified into two types: instantaneous mixture and convolutive mixture. While early works for BSS dealt with the former type [1], [2], [6], recent works are mainly concerned with the latter type [3], [7], [8], [10], [11], [13], which is much more difficult from the theoretical and computational points of view. This paper deals with a class of BSS algorithms for convolutive mixture of sources.

A basic approach to BSS is to define an index representing statistical dependency among the output signals of the separator and minimize it with respect to the parameters of the separator. The most natural index might be mutual information among the output signals [11], [13]. In the case of convolutive mixture, however, since the signals must be treated as time series, it is very complicated to concretely express the mutual information as a function of the separator's parameters. To cope with this difficulty, in some conventional methods based on the minimization of mutual information, the source signals are assumed to be i.i.d. or linear processes. A time series is called *linear* if it can be transformed to an i.i.d. signal with a linear filter. Based on this assumption, some simpler indices are introduced, and the minimization is made by such an iterative calculation as the gradient method.

In actual applications of BSS, however, the sources are not necessarily linear. The question dealt with in this paper is: "Do those algorithms work well even if the sources are not linear?" As far as we know, there is no report that discusses this issue. The problem is equivalent to discussing stability of the algorithms as described in [2], [4], [6], and [9]. This paper shows some conditions guaranteeing that the separator stably tends to a desired one with the iteration. The obtained result will reveal that some conventional algorithms (based on the minimization of some indices related to mutual information) do not work well when the sources signals are far from linear.

This paper is structured as follows. Section II gives a formulation of BSS problems. Section III introduces two evaluation functions for BSS. Section IV shows a stability analysis for a finite-dimensional process, and Section V extends it to an infinite-dimensional process. Section VI shows an example to demonstrate the validity of the analyses. Section VII discusses the relation between our results and the conventional results [2], [11]. Section VIII is devoted to the conclusion.

II. Mixing Process and the Demixing Process

Let us consider a situation where statistically independent random signals $s_i(t)$ $(i = 1, \ldots, N)$ are generated by N sources and their mixtures are observed through N sensors. It is assumed that every source signal $s_i(t)$ is a stationary random process with zero mean, and the sensors' outputs $x_i(t)$ $(i = 1, \ldots, N)$ are given by a linear mixing process

$$\mathbf{x}(t) = \sum_{\tau=0}^{\infty} \mathbf{A}(\tau)\mathbf{s}(t-\tau) = \mathbf{A}(z)\mathbf{s}(t) \tag{1}$$

where $\mathbf{s}(t) \triangleq [s_1(t), \ldots, s_N(t)]^T$, $\mathbf{x}(t) \triangleq [x_1(t), \ldots, x_N(t)]^T$, and $\mathbf{A}(z) = \sum_{\tau=0}^{\infty} \mathbf{A}(\tau) z^{-\tau}$. Here, z represents the time-shift

Manuscript received August 29, 2000; revised October 16, 2001. The associate editor coordinating the review of this paper and approving it for publication was Kristie L. Bell.

M. Ohata is with the Department of Mechanical and Control Engineering, Kyushu Institute of Technology, Kitakyushu, Japan (e-mail: ohata@cntl.kyutech.ac.jp).

K. Matsuoka is with the Department of Brain Science and Engineering, Kyushu Institute of Technology, Kitakyushu, Japan (e-mail: matsuoka@brain.kyutech.ac.jp).

Publisher Item Identifier S 1053-587X(02)00409-9.

operator $\left(z^{-1}\mathbf{s}(t)=\mathbf{s}(t-1)\right)$ and will also be used as a complex variable. It is known that at most one source signal is allowed to be Gaussian to realize BSS.

The word "mixing" means that $\mathbf{A}(z)$ is not a diagonal matrix or, more precisely, $\mathbf{A}(z)$ cannot be diagonalized by multiplying any permutation matrix. In what follows, we will not refer to the matter of permutation to avoid immaterial discussion. For the mixing process, we assume two conditions: $\sum_{\tau=0}^{\infty}\|\mathbf{A}(\tau)\|<\infty$ and $\det\mathbf{A}(z)\neq 0$ for $|z|=1$, where $\|\cdot\|$ denotes the Frobenius norm of a matrix. The first condition states that the mixing process is stable, and the second one implies that $\mathbf{A}(z)$ is invertible, although the inverse $\mathbf{A}^{-1}(z)$ may not be a causal system.

To recover the source signals from the sensor signals, we consider a demixing process (which will be referred to as the separator) of the form

$$\mathbf{y}(t)=\sum_{\tau=-\infty}^{\infty}\mathbf{W}(\tau)\mathbf{x}(t-\tau)=\mathbf{W}(z)\mathbf{x}(t) \tag{2}$$

where $\mathbf{y}(t)\triangleq[y_1(t),\ldots,y_N(t)]^T$, and $\mathbf{W}(z)\triangleq\sum_{\tau=-\infty}^{\infty}\mathbf{W}(\tau)\,z^{-\tau}$. If the mixing process $\mathbf{A}(z)$ is known beforehand, the source signals can be recovered or separated by setting $\mathbf{W}(z)=\mathbf{A}^{-1}(z)$, of course. The essential difficulty of BSS is that we are placed in a restricted condition where $\mathbf{A}(z)$ or $\mathbf{A}^{-1}(z)$ must be estimated from the observed data $\mathbf{x}(t)$ only. Besides, the impulse response $\{\mathbf{W}(\tau)\}$ might need to take a noncausal form in general, i.e., $\mathbf{W}(\tau)\neq\mathbf{O}$ $(\tau<0)$. This causes the design of adaptive BSS algorithms to be somewhat complicated, but this issue is of no concern in this paper.

In BSS of a convolutive mixture, the definition of the source signals has an indeterminacy. Namely, if $s_1(t),\ldots,s_N(t)$ are source signals, their arbitrarily linear-filtered signals $b_1(z)s_1(t),\ldots,b_N(z)s_N(t)$ can also be considered to be source signals because they are also mutually independent; the mixing process is then $\mathbf{A}(z)\mathrm{diag}\left\{b_1^{-1}(z),\ldots,b_N^{-1}(z)\right\}$. There is no way to distinguish between them because the only information we are given *a priori* is the fact that the sources are mutually independent and that the mixing process is a linear one.

A source signal $s_i(t)$ is called *linear* if it can be expressed as $s_i(t)=c_i(z)e_i(t)$, where $c_i(z)$ is a linear filter and $e_i(t)$ is an i.i.d. signal. Conventional methods usually assume this linearity of the sources, and the separator is designed to provide $e_i(t)$. In the context of blind separation of linearly mixed signals, there is no substantial difference between "i.i.d. sources" and "linear sources." As opposed to most of the conventional works, this paper deals with the generic case that the sources are not necessarily linear.

III. Couple of Indices Evaluating Mutual Independence Among the Separator's Outputs

In what follows, we write a time series $\{\ldots,\mathbf{y}(-1),\mathbf{y}(0),\mathbf{y}(1),\ldots\}$ as $\mathbf{y}(\cdot)$ to discriminate it from $\mathbf{y}(t)$ at an instant of time t. Moreover, we will sometimes use such words as "mutual independence of $\mathbf{y}(\cdot)$" and "mutual information of $\mathbf{y}(\cdot)$" instead of using the lengthy words "mutual independence among $y_1(\cdot),\ldots,y_N(\cdot)$" and "mutual information among $y_1(\cdot),\ldots,y_N(\cdot)$."

A basic strategy of BSS is to determine $\mathbf{W}(z)$ so that the separator's output $\mathbf{y}(\cdot)$ will be mutually independent. The most natural index representing the independence may be the mutual information defined by

$$I\left(\mathbf{W}(z)\right)\triangleq\sum_{i=1}^{N}h\left[y_i(\cdot)\right]-h\left[\mathbf{y}(\cdot)\right] \tag{3}$$

where $h[y_i(\cdot)]$ and $h[\mathbf{y}(\cdot)]$ are marginal and joint entropy *rates* of $\mathbf{y}(\cdot)$. The entropy rate is defined as an average entropy per unit time $h[\mathbf{y}(\cdot)]\triangleq\lim_{T_0\to\infty}H\left[\mathbf{y}(1),\ldots,\mathbf{y}(T_0)\right]/T_0$, where $H[\ldots]$ denotes the joint entropy of $[\ldots]$.

The output $\mathbf{y}(\cdot)$ of the separator is independent iff $I\left(\mathbf{W}(z)\right)$ takes the minimum, which is zero. Therefore, it might seem that a desired separator could be obtained by minimizing $I\left(\mathbf{W}(z)\right)$, utilizing such an iterative optimization as the gradient method

$$\begin{aligned}\Delta\mathbf{W}(\tau)&=-\alpha\frac{\partial I\left(\mathbf{W}(z)\right)}{\partial\mathbf{W}(\tau)}\\&=-\alpha\left(\sum_{i=1}^{N}\frac{\partial h\left[y_i(\cdot)\right]}{\partial\mathbf{W}(\tau)}-\frac{\partial h\left[\mathbf{y}(\cdot)\right]}{\partial\mathbf{W}(\tau)}\right).\end{aligned} \tag{4}$$

Actually, however, this approach has a serious difficulty. In order to express $\partial h\left[y_i(\cdot)\right]/\partial\mathbf{W}(\tau)$ concretely, we need to introduce a parametric model $\mathcal{M}^{(1)}\left(y_i(\cdot),\Theta_i\right)$ for $y_i(\cdot)$ and to estimate the parameter set Θ_i from a sample data of $y_i(\cdot)$ (or $\mathbf{x}(\cdot)$). However, since $y_i(\cdot)$ is neither Gaussian nor white, the model would be very complicated to build. [On the other hand, it is simple to calculate $\partial h\left[\mathbf{y}(\cdot)\right]/\partial\mathbf{W}(\tau)$, at least formally. According to [13], the entropy rate $h\left[\mathbf{y}(\cdot)\right]$ is given by $h\left[\mathbf{y}(\cdot)\right]=\oint_{|z|=1}\log\left|\det\mathbf{W}(z)\right|z^{-1}dz+h\left[\mathbf{x}(\cdot)\right]$. Therefore, we have $\partial h\left[\mathbf{y}(\cdot)\right]/\partial\mathbf{W}(\tau)=\partial\oint_{|z|=1}\log\left|\det\mathbf{W}(z)\right|z^{-1}dz/\partial\mathbf{W}(\tau)$, requiring no statistical estimation.]

If the sources are i.i.d. or linear, we can replace the entropy rate $h\left[y_i(\cdot)\right]$ in (3) with the entropy $H\left[y_i(t)\right]$ $(=-E\left[\log p_i(y_i(t))\right]$, where $p_i(u)$ is the pdf of $y_i(t))$

$$\begin{aligned}I^{(1)}\left(\mathbf{W}(z)\right)&\triangleq\sum_{i=1}^{N}H\left[y_i(t)\right]-h\left[\mathbf{y}(\cdot)\right]\\&=I\left(\mathbf{W}(z)\right)+\sum_{i=1}^{N}g(y_i(\cdot)).\end{aligned} \tag{5}$$

Function $g\left(y_i(\cdot)\right)\triangleq H\left[y_i(t)\right]-h\left[y_i(\cdot)\right]$ is, so to say, the mutual information of $y_i(\cdot)$; it is non-negative and takes zero only when $y_i(\cdot)$ is i.i.d. Obviously, if the source signals are linear, minimization of $I^{(1)}\left(\mathbf{W}(z)\right)$ will give a desired separator, for which $\mathbf{y}(\cdot)$ is mutually independent and $y_i(\cdot)$ $(i=1,\ldots,N)$ becomes i.i.d. In this case, we have only to consider a parametric model $\mathcal{M}^{(2)}\left(y_i(t),\theta_i\right)$ for $y_i(t)$ and not for $y_i(\cdot)$.

If the pdf $q_i(u)$ of the i.i.d. source signal $s_i(t)$ is known beforehand, we can consider a further simpler index [5]. We replace $H[y_i(t)]$ $(= -E[\log p_i(y_i(t))])$ with $-E[\log q_i(y_i(t))]$ as

$$\begin{aligned} I^{(2)}(\mathbf{W}(z)) &\triangleq \sum_{i=1}^{N} (-E[\log q_i(y_i(t))]) - h[\mathbf{y}(\cdot)] \\ &= I(\mathbf{W}(z)) + \sum_{i=1}^{N} \left(g(y_i(\cdot)) + E\left[\log \frac{p_i(y_i(t))}{q_i(y_i(t))}\right]\right). \end{aligned} \quad (6)$$

In this equation, $E[\log p_i(y_i(t))/q_i(y_i(t))]$ is the divergence of $q_i(u)$ with respect to $p_i(u)$. It is obvious that $I^{(2)}(\mathbf{W}(z))$ takes the minimum, which is zero, for $\mathbf{W}(z)$, yielding $\mathbf{y}(t) = \mathbf{s}(t)$. Term $E[\log q_i(y_i(t))]$ can easily be estimated, using a set of sample data of $y_i(t)$ as $E[\log q_i(y_i(t))] \approx \sum_{t=1}^{L} \log q_i(y_i(t))/L$. In this case, no parameter estimation is required (other than $\mathbf{W}(z)$). Since the pdf of each source signal is unknown in actual situations, the pdf $q_i(u)$ is just a roughly conjectured model of a source. However, it is known that a desired separator can be obtained by minimizing $I^{(2)}(\mathbf{W}(z))$, even if $q_i(u)$ are not exactly identical to the real pdf of the sources [2].

Thus, for linear sources, we can obtain a desired separator by minimizing $I^{(m)}(\mathbf{W}(z))$ $(m = 1, 2)$ with respect to $\mathbf{W}(z)$. In other words, adopting the gradient algorithm $\Delta\mathbf{W}(\tau) = -\alpha\partial I^{(m)}(\mathbf{W}(z))/\partial\mathbf{W}(\tau)$, the separator $\mathbf{W}(z)$ converges stably to a desired solution if the modification rate α is sufficiently small and the initial value of the separator is appropriately chosen. In the case that the sources are nonlinear, however, we might have to perform minimization of $I(\mathbf{W}(z))$, but this is cumbersome for the reason mentioned above. Now, we encounter a question: "Does minimization of $I^{(m)}(\mathbf{W}(z))$ give a desired separator even if the sources are not linear?" This paper will answer this question.

As shown before, the definition of source signals has a certain indeterminacy, and hence, so does the definition of a *desired* separator. In the following, we will eliminate the indeterminacy. Let $\mathcal{D}$ be the set of all the separators that make $\mathbf{y}(\cdot)$ mutually independent. Then, for $\mathbf{W}(z) \in \mathcal{D}$, indices $I^{(1)}(\mathbf{W}(z))$ and $I^{(2)}(\mathbf{W}(z))$ reduce to

$$I'^{(1)}(\mathbf{W}(z)) = \sum_{i=1}^{N} g(y_i(\cdot)) \quad (7)$$

$$I'^{(2)}(\mathbf{W}(z)) = \sum_{i=1}^{N} \left(g(y_i(\cdot)) + E\left[\log \frac{p_i(y_i(t))}{q_i(y_i(t))}\right]\right). \quad (8)$$

We define *the desired separator* $\mathbf{W}^{*(m)}(z)$ $(m = 1, 2)$ as a minimum point of $I'^{(m)}(\mathbf{W}(z))$ in $\mathcal{D}$ and denote the output of the desired separator by $\mathbf{y}^{*(m)}(t)$, i.e.,

$$\mathbf{y}^{*(m)}(t) = \left[y_1^{*(m)}(t), \ldots, y_N^{*(m)}(t)\right]^T \triangleq \mathbf{W}^{*(m)}(z)\mathbf{x}(t). \quad (9)$$

Signal $y_i^{*(m)}(t)$ or $y_i^{*(m)}(\cdot)$ can be considered to be a source signal *normalized* in accordance with an index with which we are concerned. Note that this definition of the normalized sources depends only on a property of the sources [and $q_i(u)$ for $I^{(2)}(\mathbf{W}(z))$] and not on the mixing process $\mathbf{A}(z)$.

Now, the aforementioned question has become more specific: "Under what condition is the desired separator $\mathbf{W}^{*(m)}(z)$ a minimum point of $I^{(m)}(\mathbf{W}(z))$?" The condition will be expressed in terms of the normalized source signals defined above. It should, however, be noted that we will only discuss *local* minimality of the desired separator in the following.

IV. Analysis for a Finite-Dimensional Process

A. Matrix Representation of the Process

The sensor signal $\mathbf{x}(\cdot)$ is nonperiodic, in general, of course. In this section, however, to simplify mathematical treatment, we consider the case that $\mathbf{x}(\cdot)$ is a stationary but periodic random signal (with a very long period T_0). Here, the word "stationary" means that $\mathbf{x}(\cdot)$ is stationary in the sense of *ensemble*. On the other hand, the word "periodic" means that any *sample* of $\mathbf{x}(\cdot)$ is periodic with a fixed period T_0, i.e., $\mathbf{x}(t+T_0) = \mathbf{x}(t)$ for every t. Since the set of periodic functions forms a finite-dimensional (T_0-dimensional) vector space, the separator can be expressed in a matrix form as follows. It should be noted that this section is just a preliminary step to derive our main result in Section V, where the source signals are nonperiodic.

Due to the periodicity of the sensor signal, (2) can be written as

$$\begin{aligned} \mathbf{y}(t) &= \sum_{\tau=0}^{T_0-1} \sum_{n=-\infty}^{\infty} \mathbf{W}(\tau + nT_0)\mathbf{x}(t - \tau - nT_0) \\ &= \sum_{\tau=0}^{T_0-1} \sum_{n=-\infty}^{\infty} \mathbf{W}(\tau + nT_0)\mathbf{x}(t - \tau) \\ &= \sum_{\tau=0}^{T_0-1} \bar{\mathbf{W}}(\tau)\mathbf{x}(t - \tau) \end{aligned} \quad (10)$$

where $\bar{\mathbf{W}}(\tau) \triangleq \sum_{n=-\infty}^{\infty} \mathbf{W}(\tau + nT_0)$. Obviously, the separator's output $\mathbf{y}(t)$ is also stationary and periodic. Define the following vectors and matrices:

$$\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{iT_0} \end{bmatrix} \triangleq \begin{bmatrix} x_i(1) \\ \vdots \\ x_i(T_0) \end{bmatrix}$$

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT_0} \end{bmatrix} \triangleq \begin{bmatrix} y_i(1) \\ \vdots \\ y_i(T_0) \end{bmatrix}$$

$$\mathbf{x} \triangleq \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{bmatrix}, \quad \mathbf{y} \triangleq \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}$$

$$\mathbf{W}_{ij} \triangleq \begin{bmatrix} \bar{w}_{ij}(0) & \bar{w}_{ij}(T_0-1) & \cdots & \bar{w}_{ij}(1) \\ \bar{w}_{ij}(1) & \bar{w}_{ij}(0) & \cdots & \bar{w}_{ij}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \bar{w}_{ij}(T_0-1) & \bar{w}_{ij}(T_0-2) & \cdots & \bar{w}_{ij}(0) \end{bmatrix}$$

$$\mathbf{W} \triangleq \begin{bmatrix} \mathbf{W}_{11} & \ldots & \mathbf{W}_{1N} \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{N1} & \ldots & \mathbf{W}_{NN} \end{bmatrix}.$$

Vectors $\mathbf{x}_i$ and $\mathbf{y}_i$ are finite-dimensional representations of $x_i(\cdot)$ and $y_i(\cdot)$, and note that matrix $\mathbf{W}_{ij}$ is a circulant matrix. Then, the input–output relation of the separator can be expressed simply as

$$\mathbf{y} = \mathbf{W}\mathbf{x}. \tag{11}$$

Corresponding to $I(\mathbf{W}(z))$, $I^{(1)}(\mathbf{W}(z))$, and $I^{(2)}(\mathbf{W}(z))$, we define

$$I(\mathbf{W}) = \sum_{i=1}^{N} H[\mathbf{y}_i] - H[\mathbf{y}] \tag{12}$$

$$I^{(1)}(\mathbf{W}) = \sum_{i=1}^{N}\sum_{t=1}^{T_0} H[y_{it}] - H[\mathbf{y}] \tag{13}$$

$$I^{(2)}(\mathbf{W}) = -\sum_{i=1}^{N}\sum_{t=1}^{T_0} E[\log q_i(y_{it})] - H[\mathbf{y}]. \tag{14}$$

The desired separator $\mathbf{W}^{*(m)}$ $(m = 1, 2)$ can be defined in a similar way to $\mathbf{W}^{*(m)}(z)$; it is the separator that minimizes $I^{(m)}(\mathbf{W})$ with $I(\mathbf{W}) = 0$. Corresponding to (9), the normalized source signals $\mathbf{y}_i^{*(m)}$ are defined as

$$\mathbf{y}^{*(m)} = \begin{bmatrix} \mathbf{y}_1^{*(m)} \\ \vdots \\ \mathbf{y}_N^{*(m)} \end{bmatrix} \triangleq \mathbf{W}^{*(m)}\mathbf{x}. \tag{15}$$

Later, we will give the condition for $\mathbf{W}^{*(m)}$ to (locally) minimize function $I^{(m)}(\mathbf{W})$. For simplicity of notation, superscript $^{(m)}$ will often be omitted.

Finally, we note that due the stationarity of $y_i(\cdot)$, the pdf of y_{it} is equivalent for every t; therefore, we can write it as $p_i(u)$. Similarly, the pdf of variable y_{it}^* is denoted as $p_i^*(u)$. From this stationarity, we have such trivial equations as $H[y_{i1}] = \cdots = H[y_{iT_0}]$ and $E[\log q_i(y_{i1})] = \cdots = E[\log q_i(y_{iT_0})]$.

B. Analysis of $I^{(1)}(\mathbf{W})$

Let us define two kinds of parameters that specify certain statistical properties of the normalized source signal $\mathbf{y}_i^*$. One set of parameters is the covariance matrix $\mathbf{R}_i$ of $\mathbf{y}_i^*$:

$$\mathbf{R}_i \triangleq E\left[\mathbf{y}_i^*\mathbf{y}_i^{*T}\right] = \begin{bmatrix} r_i(0) & r_i(1) & \cdots & r_i(T_0-1) \\ r_i(T_0-1) & r_i(0) & \cdots & r_i(T_0-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_i(1) & r_i(2) & \cdots & r_i(0) \end{bmatrix}$$

where $r_i(\tau)$ is the correlation function of $y_i^*(\cdot)$. Due to the periodicity of $y_i^*(\cdot)$, the correlation function is not only symmetric but also periodic; $r_i(\tau) = r_i(-\tau) = r_i(T_0 - \tau)$.

As another parameter, we introduce an index representing the degree of non-Gaussianity of y_{it}^*:

$$\eta_i \triangleq -E\left[\frac{d^2 \log p_i^*(y_{it}^*)}{du^2}\right] = \int_{-\infty}^{\infty} \frac{(\dot{p}_i^*(u))^2}{p_i^*(u)} du = -\int_{-\infty}^{\infty} \ddot{p}_i^*(u) \log p_i^*(u) du \tag{16}$$

where $(\dot{\ }) \triangleq d(\ )/du$. In the above equation, we have assumed that $\dot{p}_i^*(u) \to 0$ and $\dot{p}_i^*(u)\log p_i^*(u) \to 0$ $(u \to \pm\infty)$; such regularity conditions are assumed to hold throughout the paper. According to Schwarz's inequality, we have

$$\begin{aligned}\sqrt{\eta_i}\sqrt{\sigma_i^2} &= \sqrt{\int_{-\infty}^{\infty}\left(\frac{\dot{p}_i^*(u)}{p_i^*(u)^{1/2}}\right)^2 du} \times \sqrt{\int_{-\infty}^{\infty}\left(u p_i^*(u)^{1/2}\right)^2 du} \\ &\geq \int_{-\infty}^{\infty} \dot{p}_i^*(u) u du \\ &= \int_{-\infty}^{\infty} p_i^*(u) du \\ &= 1\end{aligned}$$

where σ_i^2 is the variance of y_{it}^*. Thus, we have

$$\eta_i\sigma_i^2 \geq 1. \tag{17}$$

The equality in (17) holds only when $\dot{p}_i^*(u)/p_i^*(u)^{1/2} \propto u p_i^*(u)^{1/2}$ or $d\log p_i^*(u)/du \propto u$, i.e., when $p_i^*(u)$ is a Gaussian distribution [6].

Now, we show a theorem that states local minimality of $I^{(1)}(\mathbf{W})$ at $\mathbf{W} = \mathbf{W}^{*(1)}$, which is expressed in terms of the two kinds of parameters.

Theorem 1: Matrix $\mathbf{W}^{*(1)}$ (locally) minimizes $I^{(1)}(\mathbf{W})$ if and only if the following matrix is positive definite for every pair of i and j $(i \neq j)$:

$$\mathbf{M}_{ij}^{(1)} \triangleq \begin{bmatrix} \eta_i\mathbf{R}_j & \mathbf{I} \\ \mathbf{I} & \eta_j\mathbf{R}_i \end{bmatrix} \tag{18}$$

where $\mathbf{I}$ is the identity matrix.

Proof: We consider a small perturbation $d\mathbf{W}$ from $\mathbf{W}^*$ $\left(= \mathbf{W}^{*(1)}\right)$, i.e., $\mathbf{W} = \mathbf{W}^* + d\mathbf{W}$ is in the vicinity of $\mathbf{W}^*$. Then, the output $\mathbf{y}$ of the separator becomes

$$\mathbf{y} = (\mathbf{W}^* + d\mathbf{W})\mathbf{x} = \mathbf{y}^* + d\mathbf{y} \tag{19}$$

where $d\mathbf{y} \triangleq d\mathbf{W}\mathbf{x}$. In order to prove the theorem, we show that under the premise in the theorem, the function

$$\begin{aligned} f(d\mathbf{W}) &\triangleq I^{(1)}(\mathbf{W}^* + d\mathbf{W}) - I^{(1)}(\mathbf{W}^*) \\ &= \sum_{i=1}^{N}\sum_{t=1}^{T_0}(H[y_{it}] - H[y_{it}^*]) \\ &\quad - (H[\mathbf{y}] - H[\mathbf{y}^*]) \end{aligned} \tag{20}$$

is a positive definite (precisely speaking, positive semi-definite) quadratic form with respect to $d\mathbf{W}$.

Following [2], we put $d\mathbf{V} = d\mathbf{W}\mathbf{W}^{*-1}$, which is of the form

$$d\mathbf{V} = \begin{bmatrix} d\mathbf{V}_{11} & d\mathbf{V}_{12} & \cdots & d\mathbf{V}_{1N} \\ d\mathbf{V}_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & d\mathbf{V}_{(N-1)N} \\ d\mathbf{V}_{N1} & \cdots & d\mathbf{V}_{N(N-1)} & d\mathbf{V}_{NN} \end{bmatrix}$$

and whose block element $d\mathbf{V}_{ij}$ is a circulant matrix because every block element of $d\mathbf{W}$ and $\mathbf{W}^*$ is circulant. Then, (19) reads

$$\mathbf{y} = \left(\bar{\mathbf{I}} + d\mathbf{V}\right)\mathbf{y}^* = \mathbf{y}^* + d\mathbf{y} \tag{21}$$

where $\bar{\mathbf{I}}$ is the $NT_0 \times NT_0$ identity matrix, and $d\mathbf{y} = d\mathbf{V}\mathbf{y}^*$. Hereafter, we treat $f(d\mathbf{W})$ as a function with respect to $d\mathbf{V}$; $f(d\mathbf{V}) = f(d\mathbf{W})$.

It can be shown that $f(d\mathbf{V})$ can be decomposed as $f(d\mathbf{V}) = f'(\{d\mathbf{V}_{ij}\}) + f''(\{d\mathbf{V}_{ij}\})$; the first term and the second one contain only $\{d\mathbf{V}_{ii}\}$ and $\{d\mathbf{V}_{ij};\, i \neq j\}$, respectively. According to the definition of the normalized sources, $f'(\{d\mathbf{V}_{ij}\})$ is positive (semi-)definite. Therefore, we have only to show that $f''(\{d\mathbf{V}_{ij}\})$, or $f(d\mathbf{V})$ with $d\mathbf{V}_{ii} = \mathbf{O}$ $(i = 1, \ldots, N)$, is positive definite. Hereafter, we assume $d\mathbf{V}_{ii} = \mathbf{O}$. Then, $d\mathbf{y}_i$ induced by $d\mathbf{V}$ is

$$d\mathbf{y}_i = \sum_{j=1, j\neq i}^{N} d\mathbf{V}_{ij}\mathbf{y}_j^*. \tag{22}$$

Notice that $\mathbf{y}_i^*$ is independent of $\mathbf{y}_j^*$ $(j \neq i)$ and, hence, independent of $d\mathbf{y}_i$. The covariance matrix of $d\mathbf{y}_i$ is $E\left[d\mathbf{y}_i d\mathbf{y}_i^T\right] = \sum_{j=1, j\neq i}^{N} d\mathbf{V}_{ij}\mathbf{R}_j d\mathbf{V}_{ij}^T$, and we have

$$\begin{aligned} E\left[dy_{it}^2\right] &= \frac{1}{T_0}\mathrm{tr}\left(E\left[d\mathbf{y}_i d\mathbf{y}_i^T\right]\right) \\ &= \frac{1}{T_0}\sum_{j=1, j\neq i}^{N} \mathrm{tr}\left(d\mathbf{V}_{ij}\mathbf{R}_j d\mathbf{V}_{ij}^T\right). \end{aligned} \tag{23}$$

The following proof has three steps: In Step 1, we calculate $f_1(d\mathbf{V}) \triangleq \sum_{i=1}^{N}\sum_{t=1}^{T_0}\left(H[y_{it}] - H[y_{it}^*]\right)$; in Step 2, we calculate $f_2(d\mathbf{V}) \triangleq H[\mathbf{y}] - H[\mathbf{y}^*]$; in Step 3, we examine positive definiteness of $f(d\mathbf{V}) = f_1(d\mathbf{V}) - f_2(d\mathbf{V})$.

Step 1) In the beginning, let us express the pdf $p_i(u)$ of y_{it} using the pdf $p_i^*(u)$ of y_{it}^*. The conditional probability $p_i(u|dy_{it})$ of y_{it} when given dy_{it} is $p_i^*(u - dy_{it})$; therefore, $p_i(u)$ can be obtained by taking its expectation with respect to dy_{it}:

$$p_i(u) = E\left[p_i(u|dy_{it})\right] = E\left[p_i^*(u - dy_{it})\right]. \tag{24}$$

Expanding $p_i^*(u - dy_{it})$ up to the second-order term of dy_{it}, we have

$$p_i^*(u - dy_{it}) = p_i^*(u) - \dot{p}_i^*(u)dy_{it} + \frac{1}{2}\ddot{p}_i^*(u)dy_{it}^2. \tag{25}$$

Taking expectation with respect to dy_{it}, we find

$$p_i(u) = p_i^*(u) + \frac{1}{2}\ddot{p}_i^*(u)E\left[dy_{it}^2\right]. \tag{26}$$

Here, we have used the assumption that dy_{it} has zero mean.

From (26), we have

$$\log\frac{p_i^*(u)}{p_i(u)} = \log\left(1 - \frac{1}{2}\frac{\ddot{p}_i^*(u)}{p_i(u)}E\left[dy_{it}^2\right]\right) \tag{27}$$

or

$$-\log p_i(u) + \log p_i^*(u) = -\frac{1}{2}\frac{\ddot{p}_i^*(u)}{p_i(u)}E\left[dy_{it}^2\right]. \tag{28}$$

Multiplying both the sides with $p_i(u)$ and performing integration, we have

$$\begin{aligned} &-\int p_i(u)\log p_i(u)du + \int p_i(u)\log p_i^*(u)du \\ &= -\frac{1}{2}\int \ddot{p}_i^*(u)du E\left[dy_{it}^2\right] = 0. \end{aligned} \tag{29}$$

Using (26) again, the second term in the left side is given by

$$\begin{aligned} \int p_i(u)\log p_i^*(u)du &= \int p_i^*(u)\log p_i^*(u)du \\ &+ \frac{1}{2}\int \ddot{p}_i^*(u)\log p_i^*(u)du\, E\left[dy_{it}^2\right]. \end{aligned} \tag{30}$$

Combining (29) and (30), we obtain

$$\begin{aligned} &-\int p_i(u)\log p_i(u)du + \int p_i^*(u)\log p_i^*(u)du \\ &= -\frac{1}{2}\int \ddot{p}_i^*(u)\log p_i^*(u)du\, E\left[dy_{it}^2\right]. \end{aligned} \tag{31}$$

From (16) and (23) and the definition of entropy, we have

$$H[y_{it}] - H\left[y_{it}^*\right] = \frac{1}{2T_0}\eta_i \sum_{j=1, j\neq i}^{N} \mathrm{tr}\left(d\mathbf{V}_{ij}\mathbf{R}_j d\mathbf{V}_{ij}^T\right). \tag{32}$$

Thus, we find

$$\begin{aligned} f_1(d\mathbf{V}) &= \sum_{i=1}^{N}\sum_{t=1}^{T_0}\left(H[y_{it}] - H\left[y_{it}^*\right]\right) \\ &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1; j\neq i}^{N} \eta_i \mathrm{tr}\left(d\mathbf{V}_{ij}\mathbf{R}_j d\mathbf{V}_{ij}^T\right). \end{aligned} \tag{33}$$

Step 2) It is easy to show that

$$\begin{aligned} f_2(d\mathbf{V}) &= H[\mathbf{y}] - H\left[\mathbf{y}^*\right] \\ &= \log\det\left(\bar{\mathbf{I}} + d\mathbf{V}\right) \\ &= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} \mathrm{tr}\left(d\mathbf{V}_{ji}d\mathbf{V}_{ij}\right). \end{aligned} \tag{34}$$

Step 3) Function $f(d\mathbf{V}) = f_1(d\mathbf{V}) - f_2(d\mathbf{V})$ becomes

$$\begin{aligned} f(d\mathbf{V}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1, j>i}^{N} \mathrm{tr}\Big(&\eta_i d\mathbf{V}_{ij}\mathbf{R}_j d\mathbf{V}_{ij}^T \\ &+ \eta_j d\mathbf{V}_{ji}\mathbf{R}_i d\mathbf{V}_{ji}^T + d\mathbf{V}_{ij}d\mathbf{V}_{ji} + d\mathbf{V}_{ji}d\mathbf{V}_{ij}\Big). \end{aligned} \tag{35}$$

Since $d\mathbf{V}_{ij}$ and $\mathbf{R}_j$ are both circulant matrices, the multiplication of them is commutative. Accordingly $d\mathbf{V}_{ij}\mathbf{R}_j d\mathbf{V}_{ij}^T = \left(\mathbf{R}_j d\mathbf{V}_{ij}\right)d\mathbf{V}_{ij}^T = d\mathbf{V}_{ij}^T\mathbf{R}_j d\mathbf{V}_{ij}$ holds, and we have

$$\begin{aligned} f(d\mathbf{V}) &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1, j>i}^{N} \mathrm{tr}\Big(\eta_i d\mathbf{V}_{ij}^T\mathbf{R}_j d\mathbf{V}_{ij} \\ &\qquad + \eta_j d\mathbf{V}_{ji}\mathbf{R}_i d\mathbf{V}_{ji}^T \\ &\qquad + d\mathbf{V}_{ij}^{rmT} d\mathbf{V}_{ji}^T + d\mathbf{V}_{ji}d\mathbf{V}_{ij}\Big) \\ &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1, j>i}^{N} \mathrm{tr}\left(\left[d\mathbf{V}_{ij}^T \quad d\mathbf{V}_{ji}\right]\mathbf{M}_{ij}^{(1)}\begin{bmatrix} d\mathbf{V}_{ij} \\ d\mathbf{V}_{ji}^T \end{bmatrix}\right). \end{aligned} \tag{36}$$

It follows from this equation that $f(d\mathbf{V})$ is a positive definite function if and only if matrices $\mathbf{M}_{ij}^{(1)}$ $(i \neq j)$ are all positive definite.

■

C. Analysis of $I^{(2)}(\mathbf{W})$

In this case, the positive definiteness of $I^{(2)}(\mathbf{W})$ around $\mathbf{W}^*$ $\left(= \mathbf{W}^{*(2)}\right)$ is given in terms of the covariance matrices $\mathbf{R}_{ij}$ and parameters κ_i defined by

$$\kappa_i \triangleq -E\left[\frac{d^2 \log q_i(y_{it}^*)}{du^2}\right]. \quad (37)$$

Usually, $q_i(u)$ is chosen such that $d \log q_i(u)/du$ is a monotonically decreasing function, and then κ_i is positive.

Theorem 2: Matrix $\mathbf{W}^{*(2)}$ (locally) minimizes $I^{(2)}(\mathbf{W})$ if and only if the following matrix is positive definite for every pair of i and j $(i \neq j)$:

$$\mathbf{M}_{ij}^{(2)} \triangleq \begin{bmatrix} \kappa_i \mathbf{R}_j & \mathbf{I} \\ \mathbf{I} & \kappa_j \mathbf{R}_i \end{bmatrix}. \quad (38)$$

Proof: [Most of the variables are used without explanation because they are given in Section IV-B or can be defined similarly.] What we have to do is to show that the following function is a positive definite function:

$$\begin{aligned} f(d\mathbf{V}) =& I^{(2)}(\mathbf{W}^* + d\mathbf{W}) - I^{(2)}(\mathbf{W}^*) \\ =& -\sum_{i=1}^{N}\sum_{t=1}^{T_0}\left(E\left[\log q_i(y_{it})\right] - E\left[\log q_i\left(y_{it}^*\right)\right]\right) \\ & - \left(H[\mathbf{y}] - H\left[\mathbf{y}^*\right]\right). \end{aligned} \quad (39)$$

In this equation, $E\left[\log q_i(y_{it})\right] - E\left[\log q_i(y_{it}^*)\right]$ can be calculated as follows:

$$\begin{aligned} &\int \left(p_i(u) - p_i^*(u)\right) \log q_i(u) du \\ &= \frac{1}{2}\int \ddot{p}_i^*(u) \log q_i(u) du\, E\left[dy_{it}^2\right] \text{ [from (26)]} \\ &= -\frac{1}{2}\int \dot{p}_i^*(u) \frac{d \log q_i(u)}{du} du\, E\left[dy_{it}^2\right] \text{ (partial integration)} \\ &= \frac{1}{2}\int p_i^*(u) \frac{d^2 \log q_i(u)}{du^2} du\, E\left[dy_{it}^2\right] \text{ (partial integration)} \\ &= -\frac{1}{2T_0}\kappa_i \sum_{j=1, j\neq i}^{N} \operatorname{tr}\left(d\mathbf{V}_{ij}\mathbf{R}_j d\mathbf{V}_{ij}^T\right) \text{ [from (23) and (37)].} \end{aligned}$$

Here, we have assumed that $\dot{p}_i^*(u) \log q_i(u) \rightarrow 0$ and $p_i^*(u) \log q_i(u)/du \rightarrow 0$ $(u \rightarrow \pm\infty)$.

Thus, we have

$$\begin{aligned} f(d\mathbf{V}) =& \frac{1}{2T_0}\sum_{i=1}^{N}\sum_{t=1}^{T_0}\kappa_i \sum_{j=1,j\neq i}^{N} \operatorname{tr}\left(d\mathbf{V}_{ij}\mathbf{R}_j d\mathbf{V}_{ij}^T\right) \\ & + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} \operatorname{tr}\left(d\mathbf{V}_{ji} d\mathbf{V}_{ij}\right) \\ =& \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1,j>i}^{N} \operatorname{tr}\Big(\kappa_i d\mathbf{V}_{ij}\mathbf{R}_j d\mathbf{V}_{ij}^T \\ & + \kappa_j d\mathbf{V}_{ji}\mathbf{R}_i d\mathbf{V}_{ji}^T \\ & + d\mathbf{V}_{ij} d\mathbf{V}_{ji} + d\mathbf{V}_{ji} d\mathbf{V}_{ij}\Big). \end{aligned} \quad (40)$$

In the same way as Theorem 1, we can show that this function is positive definite if and only if matrices $\mathbf{M}_{ij}^{(2)}$ $(i \neq j)$ are all positive definite. ■

V. Analysis of an Infinite-Dimensional Process

In Section IV, we derived a stability condition of the iterative algorithm for the case that the source signals are periodic, stationary processes. Letting the period T_0 tend to infinity, we can obtain a stability condition for a nonperiodic, stationary process. To perform that, we first express the condition $\mathbf{M}_{ij}^{(m)} > 0$ in another form using the eigenvenvalues of the covariance matrix $\mathbf{R}_i$.

Since covariance matrix $\mathbf{R}_i$ is a circulant matrix, its eigenvalues $\lambda_i(n)$ $(n = 1, \ldots, T_0)$ can explicitly be expressed, using the elements $r_i(\tau)$ of $\mathbf{R}_i$, as

$$\lambda_i(n) = \sum_{\tau=0}^{T_0-1} r_i(\tau)\omega_n^\tau \quad (41)$$

where $\omega_n \triangleq \exp\left(-j2\pi n/T_0\right)$, and j is the imaginary unit. Note that $\lambda_i(n)$ is a non-negative real number. The corresponding (unit) eigenvectors are $\mathbf{h}_n = 1/\sqrt{T_0}\left[\omega_n^0, \ldots, \omega_n^{T_0-1}\right]^T$ [12]. Let $\mathbf{H}$ be a matrix defined by

$$\begin{aligned} \mathbf{H} \triangleq& \left[\mathbf{h}_1 \ \cdots \ \mathbf{h}_{T_0}\right] \\ =& \frac{1}{\sqrt{T_0}}\begin{bmatrix} \omega_1^0 & \omega_2^0 & \cdots & \omega_{T_0}^0 \\ \omega_1^1 & \omega_2^1 & \cdots & \omega_{T_0}^1 \\ \vdots & \vdots & \cdots & \vdots \\ \omega_1^{T_0-1} & \omega_2^{T_0-1} & \cdots & \omega_{T_0}^{T_0-1} \end{bmatrix}. \end{aligned} \quad (42)$$

It can easily be shown that this is a unitary matrix; $\mathbf{H}\mathbf{H}^\dagger = \mathbf{I}$, where $\mathbf{H}^\dagger$ represents the conjugate transpose of $\mathbf{H}$. Then, matrix $\mathbf{R}_i$ can be written as $\mathbf{R}_i = \mathbf{H}\boldsymbol{\Lambda}_i\mathbf{H}^\dagger$, where $\boldsymbol{\Lambda}_i = \operatorname{diag}\left\{\lambda_i(1), \ldots, \lambda_i(T_0)\right\}$.

Using $\mathbf{H}$, matrix $\mathbf{M}_{ij}^{(1)}$ is represented as

$$\begin{aligned} \mathbf{M}_{ij}^{(1)} =& \begin{bmatrix} \eta_i \mathbf{H}\boldsymbol{\Lambda}_j\mathbf{H}^\dagger & \mathbf{I} \\ \mathbf{I} & \eta_j \mathbf{H}\boldsymbol{\Lambda}_i\mathbf{H}^\dagger \end{bmatrix} \\ =& \begin{bmatrix} \mathbf{H} & \mathbf{O} \\ \mathbf{O} & \mathbf{H} \end{bmatrix} \mathbf{U}_{ij} \begin{bmatrix} \mathbf{H}^\dagger & \mathbf{O} \\ \mathbf{O} & \mathbf{H}^\dagger \end{bmatrix} \end{aligned} \quad (43)$$

where

$$\mathbf{U}_{ij} \triangleq \begin{bmatrix} \eta_i \boldsymbol{\Lambda}_j & \mathbf{I} \\ \mathbf{I} & \eta_j \boldsymbol{\Lambda}_i \end{bmatrix}. \quad (44)$$

Since $\mathbf{M}_{ij}^{(1)}$ is thus unitarily equivalent to $\mathbf{U}_{ij}$, the positive definiteness of $\mathbf{M}_{ij}^{(1)}$ is equivalent to that of $\mathbf{U}_{ij}$. This leads to another form of condition for $\mathbf{W}^{*(1)}$ to give a local minimum of $I^{(1)}(\mathbf{W})$. Namely, the condition is $\mathbf{U}_{ij} > 0$ or

$$\eta_i\eta_j\lambda_i(n)\lambda_j(n) > 1 \quad (45)$$

for every n $(= 1, \ldots, T_0)$.

Since $r_i(\tau)$ is a periodic function with period T_0, the eigenvalue $\lambda_i(n)$ can be rewritten as

$$\begin{aligned} \lambda_i(n) =& \sum_{\tau=0}^{T_0-1} r_i(\tau) e^{-j2\pi n\tau/T_0} \\ =& \sum_{\tau=[-T_0/2]}^{[T_0/2]} r_i(\tau) e^{-j2\pi n\tau/T_0} \end{aligned} \quad (46)$$

where $[u]$ denotes the maximum integer not greater than u. Thus far, we have assumed that $y_i^*(\cdot)$ is a periodic (hence finite-dimensional) random process. Now, we regard the periodic process as an approximation of a nonperiodic (infinite-dimensional) random process whose correlation function $\bar{r}_i(\tau)$ is equivalent to $r_i(\tau)$ for $[-T_0/2] \le \tau \le [T_0/2]$ and is otherwise zero for a sufficiently large T_0. Fixing $f \triangleq n/T_0$ and letting T_0 and n tend to infinity, we obtain from (46)

$$\lim_{T_0, n\to\infty} \lambda_i(n) = \sum_{\tau=-\infty}^{\infty} \bar{r}_i(\tau)e^{-j2\pi f\tau} \triangleq \Phi_i(f). \qquad (47)$$

Function $\Phi_i(f)$ is nothing but the power spectrum of the original (nonperiodic) process. Thus, we obtain the following theorem.

Theorem 3: The desired separator $\mathbf{W}^{*(1)}(z)$ (locally) minimizes $I^{(1)}(\mathbf{W}(z))$ if and only if the following inequality holds for every pair of i and j $(i \ne j)$ and for every frequency f:

$$\eta_i\eta_j\Phi_i(f)\Phi_j(f) > 1. \qquad (48)$$

In addition, for $I^{(2)}(\mathbf{W}(z))$, we can obtain a similar theorem.

Theorem 4: The desired separator $\mathbf{W}^{*(2)}(z)$ (locally) minimizes $I^{(2)}(\mathbf{W}(z))$ if and only if $\kappa_i > 0$ $(i = 1, \ldots, N)$, and the following inequality holds for every pair of i and j $(i \ne j)$ and for every frequency f:

$$\kappa_i\kappa_j\Phi_i(f)\Phi_j(f) > 1. \qquad (49)$$

It should be noted that the conditions in these theorems do not depend on the mixing process $\mathbf{A}(z)$. Local minimality of $\mathbf{W}^{*(m)}(z)$ is relevant only to certain statistical properties of the source signals.

VI. Example

Here, we show how Theorem 4 can be used in an actual implementation of BSS. For the minimization of $I^{(2)}(\mathbf{W}(z)) \triangleq \sum_{i=1}^{N}(-E[\log q_i(y_i(t))]) - h[\mathbf{y}(\cdot)]$, Amari *et al.* [3] proposes the following on-line algorithm:

$$\Delta\mathbf{W}(\tau) = \alpha\left\{\mathbf{W}(\tau) - \varphi(\mathbf{y}(t))\sum_{r}\mathbf{y}^T(t-\tau+r)\mathbf{W}(r)\right\} \qquad (50)$$

where $\varphi(\mathbf{y}(t)) = [\varphi_1(y_1(t)), \ldots, \varphi_N(y_N(t))]^T$, and $\varphi_i(u) = -d\log q_i(u)/du$. When some samples of the source signals $s_i(t)$ have been given, we can determine whether or not those signals can be separated from their mixtures by the above iterative calculation, as follows.

Procedure:

Step 1: Consider a (temporal) decorrelator $y_i(t) = w(z)s_i(t)$, which is a special case of (2) $(N = 1)$. Minimize $I_i^{(2)}(w(z)) \triangleq -E[\log q_i(y_i(t))] - h[y_i(\cdot)]$, using the algorithm (50) with $N = 1$. Then, the resultant output $y_i(t)$ gives the normalized source $y_i^*(t)$ associated with $I^{(2)}(\mathbf{W}(z))$.

Step 2: For the obtained normalized sources, calculate κ_i and $\Phi_i(f)$, and check the stability condition (49).

(The reader might think that the samples of the source signals cannot be obtained before source separation is made, but the same situation also occurs when the source signals are linear.)

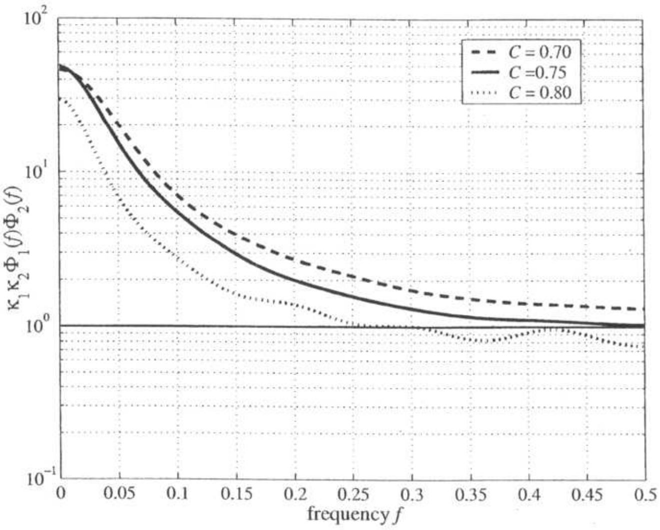


Fig. 1. Value of $\kappa_1\kappa_2\Phi_1(f)\Phi_2(f)$ for $C = 0.70, 0.75$, and 0.8.

Here, we show an example. The mixing process is a two-input, two-output system, and for the two sources, we consider binary-valued random signals generated by a Markov chain given by

$$\Pr\{s_i(t) = \pm 1 | s_i(t-1) = \pm 1\} = \frac{(1+C)}{2}$$
$$\Pr\{s_i(t) = \mp 1 | s_i(t-1) = \pm 1\} = \frac{(1-C)}{2} \qquad (51)$$

where parameter C takes a value between 0 and 1. When C is equal to 0, source $s_i(t)$ is linear (i.i.d.). As the parameter C increases, nonlinearity of $s_i(t)$ is enhanced. For $q_i(u)$, we use $q_i(u) \propto e^{-u^4/4}$ (or $\varphi_i(u) = u^3$), which is a sub-Gaussian distribution. Performing Step 1, we find (a couple of samples of) the normalized source signals. Using these, we can estimate κ_i and $\Phi_i(f)$.

Theorem 4 states that the value of $\kappa_1\kappa_2\Phi_1(f)\Phi_2(f)$ gives an index evaluating the stability of the process. If the value of this index is greater (smaller) than unity for every (some) f, then the desired separator will be stable (unstable). Fig. 1 shows the value of $\kappa_1\kappa_2\Phi_1(f)\Phi_2(f)$ as a function of frequency f; only the cases of $C = 0.70, 0.75$, and 0.80 are shown. From this, we can predict that the algorithm (50) will not give a desired solution when C exceeds around 0.75.

Next, we see what actually happens when BSS is performed using the algorithm (with $N = 2$). Since the stability does not depend on the mixing process, we here set $\mathbf{A}(z) = \mathbf{I}$ and use the normalized source signals for the sources. Then, the desired separator is $\mathbf{I}$. We define the (square) distance between the actual separating matrix $\mathbf{W}(z)$ and the desired separator $\mathbf{I}$ as

$$D(\mathbf{W}(z), \mathbf{I}) \triangleq \|\mathbf{W}(0) - \mathbf{I}\|^2 + \sum_{\tau\ne 0}\|\mathbf{W}(\tau)\|^2. \qquad (52)$$

Fig. 2 shows the final distance $D(\mathbf{W}(z), \mathbf{I})$ after a sufficiently large number of iterations; the initial value of $\mathbf{W}(z)$ is $\mathbf{I}$. This figure shows that the desired separator $\mathbf{W}^*(z) = \mathbf{I}$ becomes suddenly unstable when C exceeds around 0.75. This coincides well with the prediction obtained from the stability condition.

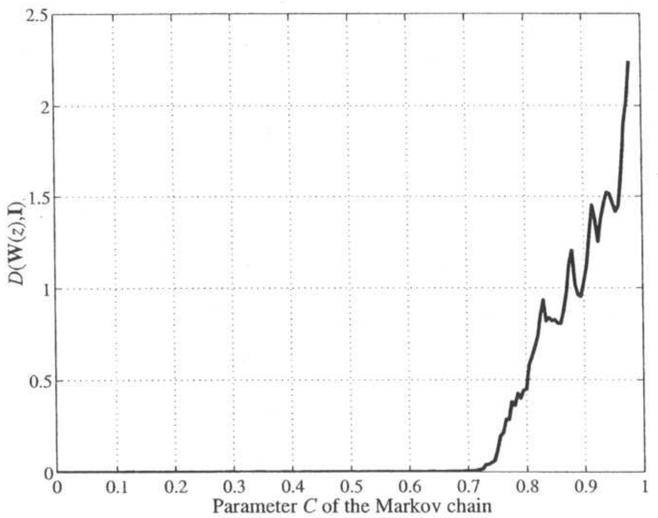


Fig. 2. Final distance between the desired separator $\mathbf{I}$ and the actual separator $D(\mathbf{W}(z), \mathbf{I})$ after a sufficiently large number of iterations.

VII. Discussion

When a desired separator is sought by the calculation $\Delta\mathbf{W}(\tau) = -\alpha\partial I^{(1)}(\mathbf{W}(z))/\partial\mathbf{W}(\tau)$, Theorem 3 gives the stability condition of the iterative process. If the sources are linear, their normalized source signals $y_i^*(\cdot)$ become white, i.e., $\Phi_i(f) \equiv r_i(0) = \sigma_i^2$ for every f. Therefore, the stability condition (48) reads

$$\eta_i\eta_j\sigma_i^2\sigma_j^2 > 1 \quad (i \neq j). \tag{53}$$

This inequality always holds as long as the sources are non-Gaussian because then, $\eta_i\sigma_i^2 > 1$. In the nonlinear case, however, $\Phi_i(f)$ is not constant with f; therefore, $\eta_i\eta_j\Phi_i(f)\Phi_j(f) \leq 1$ can occur for some f.

Let us see what is happening in $I^{(1)}(\mathbf{W}(z)) = I(\mathbf{W}(z)) + \sum_{i=1}^N g(y_i(\cdot))\ (\geq 0)$. If the sources are linear, both of the first and second terms become null for $\mathbf{W}(z) = \mathbf{W}^*(z)$, implying that $I^{(1)}(\mathbf{W}(z))$ never decreases for any perturbation from $\mathbf{W}^*(z)$. In the nonlinear case, however, $\sum_{i=1}^N g(y_i^*(\cdot))$ is strictly positive. Therefore, $\sum_{i=1}^N g(y_i^*(\cdot))$ can possibly take a smaller value by perturbing $\mathbf{W}(z)$ from $\mathbf{W}^*(z)$ in a direction toward the outside of $\mathcal{D}$ (which was defined in Section III), although $I(\mathbf{W}(z))$ becomes larger. As a result, there is a possibility that in total, $I^{(1)}(\mathbf{W}(z))$ takes a smaller value than $I^{(1)}(\mathbf{W}^*(z))$. In this situation, $\mathbf{W}^*(z)$ cannot be a local minimum point of $I^{(1)}(\mathbf{W}(z))$.

As for $I^{(2)}(\mathbf{W}(z))$, the situation is a little more complicated. When $q_i(u)$ has not been appropriately chosen, the source signals normalized in our sense are not necessarily i.i.d., even in the linear case. If the normalized source signal $y_i^{*(2)}(t)$ happen to be i.i.d., then the following inequality holds:

$$\kappa_i\Phi_i(f) \equiv \kappa_i\sigma_i^2 > 1. \tag{54}$$

This inequality comes from the stability condition given by Douglas *et al.* [11] and the fact that the normalization of the source signals is made to be stable with respect to "scale." In this case, inequality (49) is automatically satisfied. Such a condition can be seen in [11], which deals with linear sources. If the normalized sources are not i.i.d., the expression of the stability condition contains the power spectra, and not just the variances, as given in (49). Thus, including the case that the sources are linear, Theorem 4 can be regarded a generalization of the conventional result [2], [11].

VIII. Conclusion

The conventional BSS algorithms for convolutive mixture are usually designed to whiten the separator's output not only spatially but also temporally (in the sense of non-Gaussian statistics) and for the very reason they do not work well when the source signals are far from i.i.d. (or linear). In this paper, we have presented some conditions guaranteeing that certain gradient-type algorithms based on mutual information are stable in the vicinity of a desired solution. We also have shown an example to show its applicability. The result suggests that it is very important and challenging to design an algorithm that performs BSS without the temporal whitening.

References

[1] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," *Adv. Neural Inform. Process. Syst.*, vol. 8, pp. 757–763, 1996.

[2] S. Amari, T.-P. Chen, and A. Cichocki, "Stability analysis of adaptive blind source separation," *Neural Networks*, vol. 10, no. 8, pp. 1345–1351, 1997.

[3] S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. IEEE Int. Workshop Wireless Commun.*, 1997, pp. 101–104.

[4] S. Amati, T.-P. Chen, and A. Cichocki, "Nonholonomic orthogonal learning algorithms for blind source separation," *Neural Comput.*, vol. 12, no. 6, pp. 1463–1484, 2000.

[5] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995.

[6] J. F. Cardoso, "On the stability of some source separation algorithms," *J. VLSI Signal Process. Syst.*, vol. 26, no. 1, pp. 7–14, 2000.

[7] S. Choi and A. Cichocki, "An unsupervised hybrid network for blind separation of independent non-Gaussian source signals in multipath environment," *J. Commun. Networks*, vol. 1, no. 1, pp. 19–25, 1999.

[8] S. Choi, S. Amari, A. Cichocki, and R. Liu, "Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels," in *Proc. Int. Workshop Independent Component Anal. Blind Signal Separation*, Aussois, France, 1999, pp. 371–376.

[9] S. Cruces, A. Cichocki, and L. Castedo, "An iterative inversion approach to BSS," *IEEE Trans. Neural Networks*, vol. 11, pp. 1423–1437, June 2000.

[10] Y. Deville and N. Charkani, "Analysis of the stability of time-domain source separation algorithms for convolutively mixed signals," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Munich, Germany, 1997, pp. 1836–1839.

[11] S. C. Douglas, A. Cichocki, and S. Amari, "Multichannel blind separation and deconvolution of sources with arbitrary distributions," in *Proc. IEEE Workshop Neural Networks Signal Process.*, 1997, pp. 436–445.

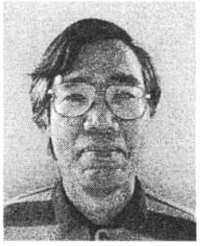
[12] F. A. Graybill, *Matrices With Applications in Statistics*. London, U.K.: Wadsworth, 1983.

[13] D.-T. Pham, "Mutual information approach to blind separation of stationary sources," in *Proc. Int. Workshop Independent Component Anal. Blind Signal Separation*, Aussois, France, 1999, pp. 215–220.

Masashi Ohata received the B.S., M.S., and Ph.D. degrees in control engineering from Kyushu Institute of Technology, Kitakyushu, Japan, in 1994, 1996 and 2001, respectively.

He is a research associate with the Department of Mechanical and Control Engineering, Kyushu Institute of Technology. His research interests are neural network and signal processing (especially, blind source separation).

Dr. Ohata is a member of the Society of Instruments and Control Engineering of Japan and the Institute of Electronics, Information, and Communication Engineers of Japan.



Kiyotoshi Matsuoka received the B.S. degree in mechanical engineering from Kyushu Institute of Technology, Kitakyushu, Japan, in 1971 and the M.S. and Ph.D. degrees in mechanichal engineering from the University of Tokyo, Tokyo, Japan, in 1976.

From 1976 to 1978, he was a lecturer with the Department of Control Engineering, Kyushu Institute of Technology, where he is currently a Professor. His research interests include neural networks, bioengineering, and artificial intelligence.

Dr. Matsuoka is a member of the Society of Instruments and Control Engineering of Japan, the Neural Network Society of Japan, and the International Neural Network Society.