4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), April 2003, Nara, Japan

A Robust Algorithm for Blind Separation of Convolutive Mixture of Sources

KIYOTSHI MATSUOKA* and SATOHI NAKASHIMA**
*Department of Brain Science and Engineering
**Department of Control Engineering
Kyushu Institute of Technology
JAPAN
matsuoka@brain.kyutech.ac.jp

Abstract: - Conventional algorithms for blind source separation do not necessarily work well for real-world data. One of the reasons is that, in actual applications, the mixing matrix is often almost singular at some part of frequency range and it can cause a certain computational instability. This paper proposes a new algorithm to overcome this singularity problem. The algorithm is based on the minimal distortion principle proposed by the authors and in addition incorporates a kind of regularization term into it, which has a role to suppress the gain of the separator for the frequency range at which the mixing matrix is almost singular.

1 Introduction

Blind source separation (BSS) is a method for recovering a set of statistically independent signals from the observation of their mixtures without any prior knowledge about the mixing process. In view of the level of complexity the mixing processes can be classified into two types: instantaneous mixture and convolutive mixture. This paper deals with BSS of convolutive mixture in general.

In the authors' experience, although conventional methods for BSS are able to achieve separation for artificially synthesized data, they do not necessarily work well for real-world data. The results of separation are often unsatisfactory and, what is worse, they sometimes suffer from incomprehensible computational instability. We are sometimes faced with the following phenomenon. As we apply an on-line BSS algorithm to the data observed, in the beginning of the iteration the algorithm appears to behave as expected, but suddenly some instability occurs.

Although a lot of reasons are conceivable about why those conventional methods are not so effective in real situations, we want to focus on one issue in this paper, which can more or less occur almost anytime. Namely, we consider the situation that the transfer function matrix of the mixing process becomes almost singular (at some part of frequency range).

The task of BSS is to find a demixing matrix, which is basically the inverse of the mixing matrix, and to apply it to the signals observed. So, if the mixing matrix is nearly singular, then the norm of the demixing matrix becomes very large and some numerical instability can occur. Even if the inverse has been obtained successfully, the separator becomes too sensitive to noises inevitably coming into the sensors.

In actual applications there are a lot of situations where the mixing matrix is almost singular. In BSS of speech signals for example, if the microphones are located close to each other, the mixing matrix, i.e. the transfer function matrix from the speakers to the microphones, becomes almost singular, particularly for a low frequency range.

The purpose of this paper is to propose an approach for overcoming the above-mentioned problem. The algorithm derived is fundamentally based on the minimal distortion principle (MDP), an idea that was proposed by the authors to normalize the demixing matrix with certain favorable properties. However, similarly to other conventional methods, the original MDP cannot cope with the case where the mixing process is almost singular. In this paper, in order to solve the problem, we introduce a generalized form of the original MDP, which is referred to as ε-MDP. Based on ε-MDP, we derive a new BSS algorithm in which a kind of regularization term is incorporated to obtain a robust separator.

Mathematical Nomenclature

Here we give a list of mathematical notations appearing in the following sections. In this list, $\mathbf{X}$ represents a matrix (or vector) and $\mathbf{X}(z)=\sum_{k=-\infty}^{\infty}\mathbf{X}_k z^{-k}$ represents a transfer function matrix (or vector). Matrix $\mathbf{X}$ may be complex-valued while coefficients $\mathbf{X}_k$ of $\mathbf{X}(z)$ are all real-valued.

- Frequency transfer function $\mathbf{X}(e^{2\pi jf})$ associated with $\mathbf{X}(z)$ is denoted by $\tilde{\mathbf{X}}(f)$. If $\tilde{\mathbf{X}}(f)$ is nonsingular for every frequency f, then $\mathbf{X}(z)$ is said to be nonsingular.
- The conjugate transpose of matrix $\mathbf{X}$ is denoted by $\mathbf{X}^H$. The same notation is also used for transfer function matrix $\mathbf{X}(z)$ as $\mathbf{X}^H(z)\triangleq\mathbf{X}^T(z^{-1})$.
- $\operatorname{tr}\mathbf{X}$ stands for the trace of square matrix $\mathbf{X}$, i.e., the sum of the diagonal entries of $\mathbf{X}$.
- The Frobenius norm of matrix $\mathbf{X}$ is defined as $\|\mathbf{X}\|\triangleq\left(\operatorname{tr}\mathbf{X}\mathbf{X}^H\right)^{1/2}$. Also the norm of transfer function matrix $\mathbf{X}(z)$ is defined as

$$\|\mathbf{X}(z)\|\triangleq\left(\sum_{k=-\infty}^{\infty}\|\mathbf{X}_k\|^2\right)^{1/2}=\left(\int_{-1/2}^{1/2}\|\tilde{\mathbf{X}}(f)\|^2\,df\right)^{1/2}.$$

- $\operatorname{diag}\{d_i\}$ represents the diagonal matrix that has diagonal entries $d_1,\ldots,d_N$. $\operatorname{diag}\mathbf{X}$ (off-diag $\mathbf{X}$) sets every off-diagonal (diagonal) entry of square matrix $\mathbf{X}$ to be zero.
- $\mathbf{0}$ and $\mathbf{I}$ represent the zero matrix and the identity matrix, respectively.
- $\delta(\tau)=1$ $(\tau=0)$, 0 $(\tau\neq0)$.

2 Mixing and Demixing Processes

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 by N sensors. It is assumed that every source signal $s_i(t)$ is a stationary random signal with mean zero, 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 $\mathbf{A}(z)\triangleq\sum_{\tau=0}^{\infty}\mathbf{A}_\tau z^{-\tau}$.

It is known that, for BSS, at most one source signal is allowed to be Gaussian. Due to statistical independence among $\{s_i(t)\}$, the auto-correlation matrix $\mathbf{R}_\tau\triangleq E[\mathbf{s}(t)\mathbf{s}^T(t-\tau)]$ and its two-sided z-transform $\mathbf{\Phi}(z)$ are diagonal. Fourier transform $\tilde{\mathbf{\Phi}}(f)$ of $\{\mathbf{R}_\tau\}$ is the power spectrum matrix of $\mathbf{s}(t)$, which is assumed to be strictly positive definite for every f. In the next section, without loss of generality we shall assume that $\tilde{\mathbf{\Phi}}(f)=\mathbf{I}$, i.e., $\mathbf{\Phi}(z)=\mathbf{I}$.

For the mixing process we assume that $\|\tilde{\mathbf{A}}(f)\|$ and $\|\tilde{\mathbf{A}}^{-1}(f)\|$ have finite values. The first condition $\|\tilde{\mathbf{A}}(f)\|<\infty$ states that the mixing process is stable. The second one $\|\tilde{\mathbf{A}}^{-1}(f)\|<\infty$ claims that $\mathbf{A}(z)$ must be invertible (though the inverse $\mathbf{A}^{-1}(z)$ may not be a causal system), but it should be noted that $\tilde{\mathbf{A}}(f)$ can be *almost* singular for some part of frequency range.

In BSS the definition of the source signals and that of the mixing process have an indeterminacy. Namely, if $s_1(t)$, …, $s_N(t)$ are source signals, their arbitrarily linear-filtered signals $e_1(z)s_1(t)$, …, $e_N(z)s_N(t)$ can also be considered source signals because they are also statistically mutually independent. The mixing process is then $\mathbf{A}(z)\operatorname{diag}\{e_i^{-1}(z)\}$. There is no way to distinguish between $\{s_i(t)\}$ and $\{e_i(z)s_i(t)\}$ (and equivalently between $\mathbf{A}(z)$ and $\mathbf{A}(z)\operatorname{diag}\{e_i^{-1}(z)\}$) because in the task of BSS the only information we are given is the fact that the sources are mutually independent.

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

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

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 $\mathbf{A}(z)$ happens to be known beforehand, the source signals can be recovered by setting as $\mathbf{W}(z)=\mathbf{A}^{-1}(z)$, of course. Essential difficulty in BSS is that $\mathbf{A}(z)$ 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., $\mathbf{W}_\tau\neq\mathbf{O}$ $(\tau<0)$. The problem of non-causality is solved approximately by designing $\mathbf{W}(z)$ such that the source signals are reproduced with a time delay.

3 Minimal Distortion Principle

The algorithm we want to propose in this paper is basically based on the minimal distortion principle (MDP), an idea for normalizing the separator [3]. Below we give a brief explanation about that.

We refer to a separator achieving source separation as a valid separator. Due to indeterminacy in the definition of the source signals, any separator of the following form is valid:

$$\mathbf{W}(z) = \mathbf{D}(z)\mathbf{A}^{-1}(z)\,, \tag{3}$$

where $\mathbf{D}(z)$ is an arbitrary nonsingular diagonal matrix, $\mathbf{D}(z) = \operatorname{diag}\{d_i(z)\}$. If the separator is valid, each source signal appears at an output terminal of the separator though it is subjected to a linear transformation $d_i(z)$. [More generally we might have to define a valid separator as $\mathbf{D}(z)\left(\mathbf{A}(z)\mathbf{P}\right)^{-1}$, where **P** is an arbitrary permutation matrix. However, to avoid immaterial complexity of description we consider only the case of **P** = **I** here.]

Indeterminacy in $d_i(z)$ is sometimes considered unsubstantial, but it cannot be ignored in view of actual implementations and applications of BSS. Among the set of all valid separators, the choice $\mathbf{D}(z) = \operatorname{diag}\mathbf{A}(z)$ has a special meaning:

$$\mathbf{W}^*(z) \triangleq \operatorname{diag}\mathbf{A}(z)\cdot\mathbf{A}^{-1}(z)\,. \tag{4}$$

We call this separator the optimal separator for the reasons described later.

It should be noted that there is no indeterminacy in the definition of the optimal separator. Namely, the optimal separator is uniquely determined independently of the indeterminacy in the definition of the source signals or the mixing process because the following holds for any diagonal matrix $\mathbf{E}(z)$:

$$\begin{aligned}&\operatorname{diag}\mathbf{A}(z)\mathbf{E}(z)\cdot\left(\mathbf{A}(z)\mathbf{E}(z)\right)^{-1}\\ &= \operatorname{diag}\mathbf{A}(z)\cdot\mathbf{A}(z)^{-1} = \mathbf{W}^*(z).\end{aligned} \tag{5}$$

The optimal separator $\mathbf{W}^*(z)$ can be characterized in some ways.

Proposition 1: The optimal separator is a valid separator that minimizes

$$Q_0(\mathbf{W}(z)) \triangleq E\left[\left\|\mathbf{y}(t)-\mathbf{x}(t)\right\|^2\right]. \tag{6}$$

Proposition 2: The optimal separator is a valid separator that satisfies

$$\operatorname{diag}E\left[\left(\mathbf{y}(t)-\mathbf{x}(t)\right)\mathbf{y}^H(t,z)\right] = \mathbf{0}\,, \tag{7}$$

where $\mathbf{y}^H(t,z)$ is formally defined as $\mathbf{y}(t,z) \triangleq \sum_{\tau=-\infty}^{\infty}\mathbf{y}(t+\tau)z^{-\tau}$ and $\mathbf{y}^H(t,z) \triangleq \mathbf{y}^T(t,z^{-1})$.

Eqn (7) is equivalent to $\operatorname{diag}E\left[\left(\mathbf{y}(t)-\mathbf{x}(t)\right)\mathbf{y}^H(t-\tau)\right] = \mathbf{0}$ for every τ. The proofs of these propositions are given in [3].

The terminology of the minimal distortion principle comes from the first proposition. Namely, the optimal separator is optimal in the sense that, among the set of valid separators, it is determined such that the observed signal is the least subjected to distortion by the separator. It should be noted that these propositions are described using the observed data $\mathbf{x}(t)$ and the separating matrix $\mathbf{W}(z)$ only, not containing the unknown matrix $\mathbf{A}(z)$.

The optimal separator has some properties favorable in actual implementation of BSS.

(1) The separator's output then becomes $\mathbf{y}(t)$ = $\operatorname{diag}\mathbf{A}(z)\cdot\mathbf{A}^{-1}(z)\mathbf{A}(z)\mathbf{s}(t) = \operatorname{diag}\mathbf{A}(z)\cdot\mathbf{s}(t)$, i.e., $y_i(t)$ = $a_{ii}(z)s_i(t)$. This implies that $y_i(t)$ becomes the *i*-th source that would be observed at the *i*-th sensor in the absence of other source signals. Namely, the source signals observed at the sensors undergo no distortion through the demixing process. This property will be convenient for evaluation of the separation result and later processing.

(2) The optimal separator does not depend on the properties of the sources; it depends only on the mixing process $\mathbf{A}(z)$. So, even for such nonstationary signals as speech, the optimal separator is invariant with time as long as the mixing process is fixed.

(3) In actual implementation the separator is usually embodied with an FIR filter. Then, it is desirable that the filter's degree (length) is as low as possible. MDP determines a valid separator such that its output becomes as close to the input as possible. So, it can be expected that the (FIR) separator will be realized with a relatively short filter length.

4 A Method for Avoiding Instability Due to Singularity of the Mixing Matrix

Although the idea of the normalization based on MDP seems to be natural, it has a serious problem in common with other conventional algorithms for BSS. Namely, when the mixing matrix is almost singular, the norm of the separating matrix becomes very large and hence it can induce some numerical instability.

Since there is an indeterminacy in the definition of the mixing process, the words “the mixing matrix is almost singular” is somewhat ambiguous. To eliminate the ambiguity we introduce a normalized form of the mixing process as follows.

(A1) The source signals are white signals with zero mean and unity variance, i.e., $\mathbf{\Phi}(z) = \mathbf{I}$ or $\tilde{\mathbf{\Phi}}(f) = \mathbf{I}$;

(A2) Sensor signal $x_i(t)$ has been scaled to be of order unity; $E\left[\|x_i(t)\|^2\right] \sim o(1)$.

Let the i-th row of $\mathbf{A}(z)$ be $\mathbf{a}_i(z)$. Then, since $E\left[x_i^2(t)\right] = \int_{-1/2}^{1/2} \tilde{\mathbf{a}}_i(f)\tilde{\mathbf{\Phi}}(f)\tilde{\mathbf{a}}_i^H(f)df = \|\mathbf{a}_i(z)\|^2$, (A2) is equivalent to $\|\mathbf{a}_i(z)\| \sim o(1)$. From this we obviously have

(A3) $\|a_{ii}(z)\| \sim o(1)$ (or less).

On these assumptions we want to give a definition about the singularity of the mixing process. Although it might not be mathematically rigorous, it is enough for the present purpose. Let the i-th row of $\mathbf{A}^{-1}(z)$ be $\mathbf{b}_i(z)$. We say that the mixing matrix $\mathbf{A}(z)$ is almost singular with respect to source i if $\|\mathbf{b}_i(z)\|$ is very large. Note that even if $\|\tilde{\mathbf{b}}_i(f)\|$ is large for a frequency f, it does not necessarily imply that $\|\mathbf{b}_i(z)\|$ is large.

As describe previously, the optimal separator based on MDP has a serious problem when $\mathbf{A}(z)$ is almost singular. Let $\mathbf{w}_i^*(z)$ be the i-th row of $\mathbf{W}^*(z)$, then we have $\mathbf{w}_i^*(z) = a_{ii}(z)\mathbf{b}_i(z)$. This implies that if $\mathbf{A}(z)$ is almost singular with respect source i, the norm of $\mathbf{w}_i^*(z)$ becomes very large. This can cause instability when we execute an algorithm for BSS.

Moreover, even if separation has been attained successfully, the separator obtained becomes very sensitive to noise. Suppose that the sensors’ signals are corrupted with noise $\mathbf{d}(t)$. Then the output of the separator becomes

$$\begin{aligned} y_i(t) &= \mathbf{w}_i^*(z)(\mathbf{A}(z)\mathbf{s}(t) + \mathbf{d}(t)) \\ &= a_{ii}(z)s_{ii}(t) + \mathbf{w}_i^*(z)\mathbf{d}(t). \end{aligned} \tag{8}$$

So, when $\|\mathbf{b}_i(z)\|$ and hence $\|\mathbf{w}_i^*(z)\|$ are very large, $y_i(t)$ becomes undesirably sensitive to noise.

To overcome these problems we extend the optimal separator as

$$\begin{aligned} \mathbf{W}^{**}(z) &\triangleq \mathbf{C}(z)\,\mathrm{diag}\mathbf{A}(z)\cdot\mathbf{A}^{-1}(z) \\ &= \mathbf{C}(z)\,\mathbf{W}^*(z) \end{aligned} \tag{9}$$

where $\mathbf{C}(z)$ is a diagonal matrix defined as

$$\begin{aligned} \mathbf{C}(z) &= \left\{\mathbf{I} + \varepsilon\ \mathrm{diag}\left(\mathbf{A}^{-1}(z)\mathbf{A}^{-H}(z)\right)\right\}^{-1} \\ &= \mathrm{diag}\left\{\frac{1}{1+\varepsilon\mathbf{b}_i(z)\mathbf{b}_i^H(z)}\right\} \end{aligned}$$

and ε is a small positive constant. We call this separator the ε-optimal separator. The ε-optimal separator is obviously valid, and the special case of $\varepsilon = 0$ reduces to the original optimal separator. Although the ε-optimal separator has lost some of the favorable properties held by the original optimal separator, it should be noted that the diagonal entries $\left\{1+\varepsilon\ \mathrm{diag}\left(\mathbf{b}_i(z)\mathbf{b}_i^H(z)\right)\right\}^{-1}$ of $\mathbf{C}(z)$ are zero-phase filters. It implies that every frequency component in the output of the ε-optimal separator receives no phase shift relative to that in the output of the optimal separator.

The i-th row of $\tilde{\mathbf{W}}^{**}(f)$ becomes

$$\tilde{\mathbf{w}}_i^{**}(f) = \frac{\tilde{a}_{ii}(f)}{1+\varepsilon\|\tilde{\mathbf{b}}_i(f)\|^2}\tilde{\mathbf{b}}_i(f). \tag{10}$$

So, we have for any value of $\|\tilde{\mathbf{b}}_i(f)\|$

$$\begin{aligned} \|\tilde{\mathbf{w}}_i^{**}(f)\|^2 &= \frac{\|\tilde{\mathbf{b}}_i(f)\|^2}{\left(1+\varepsilon\|\tilde{\mathbf{b}}_i(f)\|^2\right)^2}|\tilde{a}_{ii}(f)|^2 \\ &\le \frac{1}{4\varepsilon}|\tilde{a}_{ii}(f)|^2. \end{aligned}$$

By integrating this with respect to f over [-1/2, 1/2] we obtain

$$\|\mathbf{w}_i^{**}(z)\|^2 \le \frac{1}{4\varepsilon}\|a_{ii}(z)\|^2$$

and from (A3) we find

$$\|\mathbf{w}_i^{**}(z)\|^2 \sim \frac{1}{\varepsilon}o(1).$$

Namely, $\|\mathbf{w}_i^{**}(z)\|$ never exceeds a finite value of order $1/\sqrt{\varepsilon}$ (even for $\|\mathbf{b}_i(z)\| \to \infty$).

The ε-optimal separator can be characterized in similar ways as the original optimal separator.

Proposition 1′: The ε-optimal separator is a valid separator that minimizes

$$Q_\varepsilon(\mathbf{W}(z)) \triangleq E\left[\|\mathbf{y}(t) - \mathbf{x}(t)\|^2\right] + \varepsilon\|\mathbf{W}(z)\|^2. \tag{11}$$

Term $\varepsilon\|\mathbf{W}(z)\|^2$ is a kind of regularization term, which is often introduced in certain types of ill-posed optimization problems. We refer to the normalization based on minimization of $Q_\varepsilon(\mathbf{W}(z))$ as ε-MDP. Corresponding to Proposition 2, we have

Proposition 2′: The ε-optimal separator is a valid separator that satisfies

$$\mathrm{diag}\left\{ E\left[(\mathbf{y}(t)-\mathbf{x}(t))\mathbf{y}^H(t,z)\right] + \varepsilon\mathbf{W}(z)\mathbf{W}^H(z) \right\} = \mathbf{0}. \tag{12}$$

5 An Adaptive Algorithm Based on the ε-MDP

First we describe a rough guideline for deriving a BSS algorithm based on the ε-MDP. Let $\mathcal{V}$ and $\mathcal{N}$ be the set of valid separators and that of $\mathbf{W}(z)$ satisfying (12), respectively. Investigating the dimensions of manifolds $\mathcal{V}$ and $\mathcal{N}$, we find that the intersection of them gives an isolated point in the space of nonsingular transfer function matrices.

Consider two dynamical systems:

$$\frac{d\mathbf{W}_\tau}{du} = -f(\{\mathbf{W}_\tau\}) \; (\tau = \ldots,-1,0,1,\ldots), \tag{13}$$

$$\frac{d\mathbf{W}_\tau}{du} = -g(\{\mathbf{W}_\tau\}) \; (\tau = \ldots,-1,0,1,\ldots). \tag{14}$$

[For the moment, parameter u is irrelevant to time parameter t.] The first system is a gradient system whose stable equilibrium points form $\mathcal{V}$, while the second system is a gradient system whose stable equilibrium points form $\mathcal{N}$. The word 'gradient' means that $f(\{\mathbf{W}_\tau\})$ and $g(\{\mathbf{W}_\tau\})$ are gradient vectors derived from some scalar functions of $\{\mathbf{W}_\tau\}$. Then, the optimal separator will become a stable equilibrium of the combined system

$$\frac{d\mathbf{W}_\tau}{du} = -f(\{\mathbf{W}_\tau\}) - g(\{\mathbf{W}_\tau\}). \tag{15}$$

The choice of $f(\{\mathbf{W}_\tau\})$ is somewhat task-dependent. Here we show a candidate. We start with an approach proposed by Amari et al. [1] Define

$$I(\mathbf{W}(z)) \triangleq -\sum_{i=1}^{N} E[\log q_i(y_i(t))] - h[\mathbf{y}(t)] \tag{16}$$

where $h[\mathbf{y}(t)]$ is the entropy rate of $\mathbf{y}(t)$ and $q_i(u)$ is a pdf assumed for source signal $s_i(t)$. If the source signals are iid and $q_i(u)$ approximates well the real pdf of $s_i(t)$, then minimizing $I(\mathbf{W}(z))$ provides a valid separator. Using the idea of natural gradient learning proposed by Amari [1], we have

$$\frac{d\mathbf{W}(z)}{du} = -\alpha \frac{\partial I(\mathbf{W}(z))}{\partial \mathbf{W}(z)} \mathbf{W}^H(z)\mathbf{W}(z)$$
$$= \alpha\left(\mathbf{I} - E\left[\varphi(\mathbf{y}(t))\mathbf{y}^H(t,z)\right]\right)\mathbf{W}(z), \tag{17}$$

where $\varphi(\mathbf{y}(t))$ is defined as $\varphi(\mathbf{y}(t)) \triangleq \left[\varphi_1(y_1(t)),\ldots,\varphi_N(y_N(t))\right]^T$ and $\varphi_i(u) \triangleq -d\log q_i(u)/du$. Choi et al. [2] introduced a nonholonomic constraint to the algorithm. Let $d\mathbf{W}(z)$ be a tangent vector at $\mathbf{W}(z)$ in the space of nonsingular transfer function matrices and define $d\mathbf{U}(z) = \sum_\tau d\mathbf{U}_\tau z^{-\tau} \triangleq d\mathbf{W}(z)\mathbf{W}^{-1}(z)$. Choi et al. [2] propose a method that performs a natural gradient descent optimization under a nonholonomic constraint as $\mathrm{diag}\, d\mathbf{U}_0 = \mathbf{0}$. We here modify this as $\mathrm{diag}\, d\mathbf{U}(z) = \mathbf{0}$. Applying this constraint to (17) leads to a dynamics that will embody (13):

$$\frac{d\mathbf{W}(z)}{du} = -\alpha\, \text{off-diag}\, E\left[\varphi(\mathbf{y}(t))\mathbf{y}^H(t,z)\right]\cdot \mathbf{W}(z). \tag{18}$$

As for the second dynamics, (14), we introduce the ε-MDP (Proposition 2′) as

$$\frac{d\mathbf{W}(z)}{du} = -\beta\,\mathrm{diag}\left\{E\left[(\mathbf{y}(t)-\mathbf{x}(t))\mathbf{y}^H(t,z)\right] + \varepsilon\mathbf{W}(z)\mathbf{W}^H(z)\right\}\cdot\mathbf{W}(z) \tag{19}$$

We can prove that the equilibrium points of this system are semi-stable.

Combining (18) and (19), we have an algorithm based on ε-MD as

$$\frac{d\mathbf{W}(z)}{du} = -\alpha\,\text{off-diag}\,E\left[\varphi(\mathbf{y}(t))\mathbf{y}^H(t,z)\right]\cdot\mathbf{W}(z) - \beta\,\mathrm{diag}\left\{E\left[(\mathbf{y}(t)-\mathbf{x}(t))\mathbf{y}^H(t,z)\right] + \varepsilon\mathbf{W}(z)\mathbf{W}^H(z)\right\}\cdot\mathbf{W}(z). \tag{20}$$

When deriving an adaptive algorithm for determining the separator, parameter u is replaced by real time parameter t and the expectation $E[\cdot]$ is replaced by an instantaneous value of $[\cdot]$. Moreover, the separator must be embodied with an FIR filter as

$$\mathbf{y}(t-L) = \sum_{\tau=-L}^{L} \mathbf{W}_\tau \mathbf{x}(t-L-\tau) = \mathbf{W}(z)\mathbf{x}(t-L) \tag{21}$$

where $\mathbf{W}(z)=\sum_{\tau=-L}^{L}\mathbf{W}_\tau z^{-\tau}$. This aims to recover the source signals with a time delay of around L, taking into consideratin the possibly non-minimum phase characteristic of the mixing matrix. Thus we have the following algorithm

$$\begin{aligned}
&\Delta\mathbf{W}_\tau \\
&=-\sum_{r=-L}^{L}\Big\{\ \alpha\text{off-diag}\ \varphi(\mathbf{y}(t-3L))\mathbf{y}^T(t-3L-\tau+r) \\
&\quad+\beta\,\text{daig}\ \big(\mathbf{y}(t-3L)-\mathbf{x}(t-3L)\big)\mathbf{y}^T(t-3L-\tau+r) \\
&\quad+\beta\varepsilon\sum_{k=-L}^{L}\ \text{daig}\ \mathbf{W}_k\mathbf{W}_{k-\tau+r}^T\ \Big\}\cdot\mathbf{W}_r, \qquad (22)
\end{aligned}$$

where parameters α and β are small positive constants.

6 An Example

To demonstrate effectiveness of the proposed algorithm we show a simple example. The mixing process (and hence the demixing process) is a two-input, two-output system. The two sources are both iid binary signals taking ±1 with equal probability. The mixing matrix is given as

$$\mathbf{A}(z)=\begin{bmatrix}1 & z^{-1}\\ z^{-1} & 1\end{bmatrix}.$$

Note that $\tilde{\mathbf{A}}(f)$ is singular at f = 0, 1/2.

For φ_i, we use $\varphi_i(u) = u^3$, which corresponds to a sub-Gaussian distribution $q_i(u)\propto e^{-u^4/4}$. The length of the separator and the initial values of $\{\mathbf{W}_\tau\}$ are set as L = 20 and $\mathbf{W}_\tau=\delta(\tau)\mathbf{I}$, respectively. Parameters α, β, and ε are chosen as 1.0×10^{-5}, 1.0×10^{-5}, and 1.0×10^{-1}, respectively.

Fig. 1 shows the frequency response (gain) of the overall process, $\tilde{\mathbf{V}}(f)=\left[\tilde{v}_{ij}(f)\right]\triangleq\tilde{\mathbf{W}}(f)\tilde{\mathbf{A}}(f)$, after $\mathbf{W}(z)$ converges. In the case of MDP (a), $|\tilde{v}_{ij}(f)|$ for any i and j is similar, namely the separation is a total failure. In the case of ε-MD (b) on the other hand, $|\tilde{v}_{ij}(f)|\ (i\neq j)$ are very small compared to $\tilde{v}_{ii}(f)$. Note that $|\tilde{v}_{ii}(f)|$ take small values around f = 0, 0.5, which means $\|\tilde{\mathbf{w}}_i(f)\|$ is small in the same region.

References:

[1] S. Amari, S. C. Douglas, A. Cichocki and H. H.Yang, Multichannel blind deconvolution and equalization using the natural gradient, Proc. IEEE International Workshop on Wireless Communication, pp. 101-104, 1997.

[2] S. Choi, S. Amari, A. Cichocki, and R. Liu, Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels, Proc. International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99), pp. 371-376, 1999.

[3] K. Matsuoka and S. Nakashima, Minimal distortion principle for blind source separation, Proc. International Workshop on Independent Component Analysis and Blind Signal Separation (ICA2001), pp. 722-727, 2001.

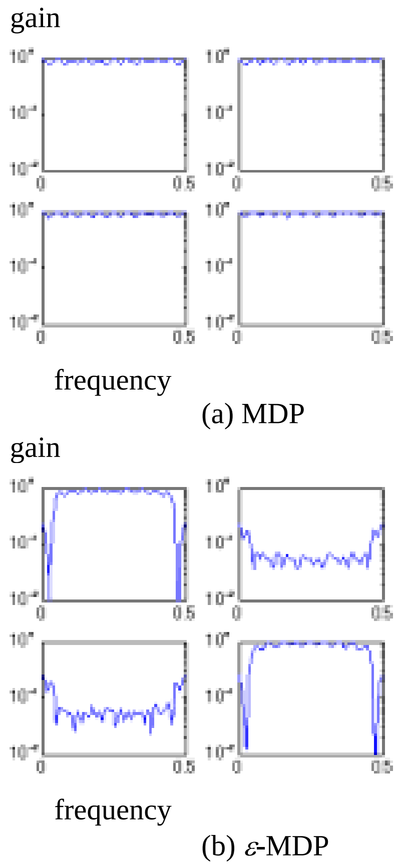

Fig. 1 Frequency response (gain) of the overall process, $\tilde{\mathbf{W}}(f)\tilde{\mathbf{A}}(f)$