PAPER

Natural Gradient Learning for Spatio-temporal Decorrelation: Recurrent Network*

Seungjin CHOI[†], ***Nonmember***, **Shunichi AMARI**[††], ***and*** **Andrzej CICHOCKI**[††], ***Members***

SUMMARY Spatio-temporal decorrelation is the task of eliminating correlations between associated signals in spatial domain as well as in time domain. In this paper, we present a simple but efficient adaptive algorithm for spatio-temporal decorrelation. For the task of spatio-temporal decorrelation, we consider a dynamic recurrent network and calculate the associated natural gradient for the minimization of an appropriate optimization function. The natural gradient based spatio-temporal decorrelation algorithm is applied to the task of blind deconvolution of linear single input multiple output (SIMO) system and its performance is compared to the spatio-temporal anti-Hebbian learning rule.

1. Introduction

Spatial decorrelation is useful for numerous problems in speech processing, communications, signal and image processing. Spatial decorrelation is also known as *data sphering* or *data whitening*. The task of spatial decorrelation is to eliminate cross-correlations between associated signals. In other words, spatial decorrelation aims at finding a linear transformation from given multivariate data $\boldsymbol{x}(k) = [x_1(k) \cdots x_n(k)]^T$ to $\boldsymbol{y}(k) = [y_1(k) \cdots y_n(k)]^T$ such that the correlation matrix of $\boldsymbol{y}(k)$ becomes the identity matrix, i.e., $\boldsymbol{R}_{yy} = E\{\boldsymbol{y}(k)\boldsymbol{y}^T(k)\} = \boldsymbol{I}$. Various adaptive algorithms for spatial decorrelation have been developed. They include principal component analysis (PCA) neural networks [1], [2], the linear anti-Hebbian rule [3], the Almeida-Silva algorithm [4]. Recently local and global adaptive algorithms for spatial decorrelation was analyzed in [5].

The linear anti-Hebbian rule [3] might be one of well-known adaptive spatial decorrelation algorithms. In contrast to the Hebbian rule, the anti-Hebbian rule is known to be a energy minimizer [6]. The minimization of output energy leads to uncorrelated output variables. In [3], a linear feedback network was considered for spatial decorrelation task and the associated algorithm was derived heuristically. We revisit the anti-Hebbian rule and reveal how it is related to the energy minimizer. Moreover we employ the natural gradient adaptation method [7] and derive an efficient spatial decorrelation algorithm.

The task of temporal decorrelation is to eliminate cross-correlations between $x_i(k)$ and $x_i(k-p)$ for any $p \neq 0$. The conventional linear prediction method (for example, Levinson-Durbin algorithm) or a gradient descent based method [8] have been developed. Recently the temporal decorrelation method was applied to the task of blind equalization of SIMO channels [8], [9].

The goal of spatio-temporal decorrelation is to perform spatial decorrelation and temporal decorrelation simultaneously for multivariate data. In the task of spatio-temporal decorrelation, it is desired to design a multivariate filter such that the filter output $\boldsymbol{y}(k)$ satisfies

$$E\{\boldsymbol{y}(k)\boldsymbol{y}^T(k-p)\} = \boldsymbol{I}\delta_p, \ \forall p, \tag{1}$$

where δ_p is the Kronecker delta equal to 1 if $p = 0$, otherwise it is zero. Recently the anti-Hebbian rule was extended to the task of spatio-temporal decorrelation [10], [11]. The conventional gradient descent method was used to derive the *spatio-temporal anti-Hebbian rule* [10], [11].

The natural gradient adaptation was recently proposed by Amari [7], [12] and was shown to be efficient for on-line learning. It was shown that the natural gradient finds the steepest descent direction when the parameter space belongs to the Riemannian manifold, which is often encountered in many problems. In this paper we consider a dynamic recurrent network for the task of spatio-temporal decorrelation and calculate the associated natural gradient for the minimization of an appropriate optimization function [13]. A new spatio-temporal decorrelation algorithm based on the natural gradient is derived and its performance is compared to the spatio-temporal anti-Hebbian rule in the task of blind deconvolution of SIMO channels.

The rest of paper is organized as follows. After brief review of the linear anti-Hebbian rule, we derive the natural gradient based spatial decorrelation algorithm in Section 2. For the task of spatio-temporal decorrelation, a dynamic recurrent network is introduced and the associated natural gradient is calculated

Manuscript received November 20, 1999.
Manuscript revised August 2, 2000.
[†]The author is with the Department of Electrical Engineering, Chungbuk National University, KOREA
[††]The authors are with Brain-style Information Systems Research Group, BSI, RIKEN, JAPAN
[*]This paper was presented at International Conference on Neural Information Processing in 1997.
This research was supported in part by Korean Ministry of Science and Technology under Korean Brain Science and Engineering Research Program and by Brain Science Institute, RIKEN, JAPAN

in Section 3. As an example to illustrate the usefulness of the proposed algorithm, we consider the task of blind deconvolution of linear SIMO systems in Section 4. The performance of the natural gradient based spatio-temporal decorrelation is compared to that of the spatio-temporal anti-Hebbian rule. Finally conclusions are drawn in Section 5.

2. Spatial Decorrelation

Let us consider a linear feedback network whose output $\boldsymbol{y}(k)$ is described as

$$\begin{aligned} \boldsymbol{y}(k) &= \boldsymbol{x}(k) + \boldsymbol{W}\boldsymbol{y}(k) \\ &= [\boldsymbol{I} - \boldsymbol{W}]^{-1}\boldsymbol{x}(k), \end{aligned} \tag{2}$$

where $\boldsymbol{x}(k)$ is an n-dimensional input vector to the network and $\boldsymbol{W}$ is a synaptic weight matrix whose (i,j)-element w_{ij} represents the connection strength between $y_i(k)$ and $y_j(k)$.

Spatial decorrelation aims at finding a linear transformation from given multivariate data $\boldsymbol{x}(k)$ to $\boldsymbol{y}(k)$ such that the correlation matrix of $\boldsymbol{y}(k)$ becomes the identity matrix (or invertible diagonal matrix), i.e., $\boldsymbol{R}_{yy} = E\{\boldsymbol{y}(k)\boldsymbol{y}^T(k)\} = \boldsymbol{I}$. First we briefly review Földiák's first model and second model for the anti-Hebbian rule. Then we present a spatial decorrelation algorithm based on the natural gradient.

2.1 Linear Anti-Hebbian Rule

In the Földiák's first model, the linear feedback network (without self-feedback connections) was considered. The ith network output $y_i(k)$ is written as

$$y_i(k) = x_i(k) + \sum_{j \neq i} w_{ij} y_j(k). \tag{3}$$

The synaptic weight w_{ij} is updated in such a way that cross-correlation between $y_i(k)$ and $y_j(k)$ is minimized. Földiák [3] suggested that the synaptic weight at time $k+1$, $w_{ij}(k+1)$ is updated using the synaptic weight at time k, $w_{ij}(k)$ and signals $y_i(k)$, $y_j(k)$, i.e.,

$$w_{ij}(k+1) = w_{ij}(k) - \eta_k y_i(k) y_j(k), \tag{4}$$

where $\eta_k > 0$ is a learning rate. When the convergence of the algorithm (4) is achieved, the correlation matrix of the network output $\boldsymbol{y}(k)$ becomes diagonal.

Földiák suggested a further model (Földiák's second model) by allowing all neurons to receive their own outputs. The fully-connected linear feedback network given in (2) was considered. The updating algorithm for the self-feedback connections $\{w_{ii}\}$ was given by

$$w_{ii}(k+1) = w_{ii}(k) + \eta_k \{1 - y_i(k) y_i(k)\}. \tag{5}$$

In a compact form, the learning algorithm in Földiák's second model is given by

$$\boldsymbol{W}(k+1) = \boldsymbol{W}(k) + \eta_k \left\{\boldsymbol{I} - \boldsymbol{y}(k)\boldsymbol{y}^T(k)\right\}. \tag{6}$$

2.2 The Spatial Decorrelation Algorithm: Natural Gradient

Linear anti-Hebbian rule was shown to be an energy minimizer [6], in contrast to Hebbian rule where the objective is the maximum transfer of information , or maximization of the correlation between inputs and associated outputs. The minimization of energy can be viewed as the minimization of the Kullback-Leibler divergence between two zero-mean Gaussian distributions with covariances $E\{\boldsymbol{y}\boldsymbol{y}^T\}$ and $\boldsymbol{I}$, respectively [14]. As in [14], we consider the risk $R(\boldsymbol{W})$ given by

$$\begin{aligned} R(\boldsymbol{W}) &= E\{L(\boldsymbol{W})\} \\ &= \frac{1}{2}\sum_{i=1}^{n} E\{y_i^2\} - \log|\det[\boldsymbol{I} - \boldsymbol{W}]^{-1}|, \end{aligned} \tag{7}$$

where $L(\boldsymbol{W})$ is the loss function and some irrelevant terms (which do not depend on the parameter matrix $\boldsymbol{W}$) were omitted. The loss function $L(\boldsymbol{W})$ is

$$L(\boldsymbol{W}) = \frac{1}{2}\sum_{i=1}^{n} y_i^2 - \log|\det[\boldsymbol{I} - \boldsymbol{W}]^{-1}|. \tag{8}$$

We now derive an spatial decorrelation algorithm from the minimization of the loss function (8) using the natural gradient method. The derivation of the algorithm is just outlined here because Section 3 describes the detailed derivation of the spatio-temporal decorrelation algorithm using the natural gradient.

Simple algebraic and differential calculus yields

$$d\left\{\frac{1}{2}\sum_{i=1}^{n} y_i^2\right\} = \boldsymbol{y}^T d\boldsymbol{V}\boldsymbol{y} \tag{9}$$

$$d\left\{\log|\det[\boldsymbol{I} - \boldsymbol{W}]^{-1}|\right\} = \operatorname{tr}(d\boldsymbol{V}), \tag{10}$$

where $d\boldsymbol{V}$ is a modified differential matrix defined by

$$d\boldsymbol{V} = [\boldsymbol{I} - \boldsymbol{W}]^{-1} d\boldsymbol{W}, \tag{11}$$

and $\operatorname{tr}(\cdot)$ denotes the trace operator.

Then, the total differential $dL(\boldsymbol{W})$ is

$$dL(\boldsymbol{W}) = \boldsymbol{y}^T d\boldsymbol{V}\boldsymbol{y} - \operatorname{tr}(d\boldsymbol{V}). \tag{12}$$

The spatial decorrelation algorithm using the natural gradient has the form

$$\begin{aligned} \Delta\boldsymbol{W}(k) &= \boldsymbol{W}(k+1) - \boldsymbol{W}(k) \\ &= [\boldsymbol{I} - \boldsymbol{W}(k)]\left(-\eta_k \frac{dL(\boldsymbol{W})}{d\boldsymbol{V}}\right) \\ &= \eta_k [\boldsymbol{I} - \boldsymbol{W}(k)]\left\{\boldsymbol{I} - \boldsymbol{y}(k)\boldsymbol{y}^T(k)\right\}. \end{aligned} \tag{13}$$

Remarks

- The conventional gradient descent method for the minimization of the loss function (8) gives the following learning algorithm for $\boldsymbol{W}$

$$\Delta \boldsymbol{W}(k) = \eta_k \left[\boldsymbol{I} - \boldsymbol{W}(k)\right]^{-T} \left\{\boldsymbol{I} - \boldsymbol{y}(k)\boldsymbol{y}^T(k)\right\}. \quad (14)$$

Thus the linear anti-Hebbian rule in Földiák's second model can be viewed as the approximated version of (14) (with neglecting the term $[\boldsymbol{I} - \boldsymbol{W}(k)]^{-T}$).

- The algorithm (14) requires the matrix inversion at every iteration, which is cumbersome in the viewpoint of computation. However, the natural gradient based algorithm (13) is computationally simple, compared to (14).

3. Spatio-temporal Decorrelation

The spatio-temporal decorrelation requires spatial decorrelation as well as temporal decorrelation, simultaneously. For the task of spatio-temporal decorrelation, we consider a dynamic recurrent network (see Fig. 1) whose output $\boldsymbol{y}(k)$ is defined as

$$\begin{aligned}\boldsymbol{y}(k) &= \boldsymbol{x}(k) + \sum_{p=0}^{L} \boldsymbol{W}_p \boldsymbol{y}(k-p) \\ &= \boldsymbol{x}(k) + \boldsymbol{W}(z)\boldsymbol{y}(k) \\ &= \left[\boldsymbol{I} - \boldsymbol{W}(z)\right]^{-1} \boldsymbol{x}(k),\end{aligned} \quad (15)$$

where $\boldsymbol{W}(z)$ is a polynomial matrix in z^{-1}

$$\boldsymbol{W}(z) = \sum_{p=0}^{L} \boldsymbol{W}_p z^{-p}, \quad (16)$$

where z^{-1} is the time-shift operator, i.e., $z^{-1}\boldsymbol{y}(k) = \boldsymbol{y}(k-1)$.

In this section, we first review the spatio-temporal anti-Hebbian rule [10], [11] which is the direct extension of the linear anti-Hebbian rule in Földiák's second model. A proper optimization function for the task of spatio-temporal decorrelation is presented. Then, we calculate the natural gradient for the dynamic recurrent network and derive a new efficient spatio-temporal decorrelation algorithm.

3.1 The Spatio-temporal Anti-Hebbian Rule

Let us consider the dynamic recurrent network described in (15). A simple extension of the linear anti-Hebbian rule in Földiák's second model suggests the following learning algorithm for updating $\{\boldsymbol{W}_p\}$

$$\Delta \boldsymbol{W}_p(k) = \eta_k \left\{\boldsymbol{I}\delta_p - \boldsymbol{y}(k)\boldsymbol{y}^T(k-p)\right\}. \quad (17)$$

The algorithm (17) is the *spatio-temporal anti-Hebbian rule*. Several variants of the algorithm (17) have been applied to multichannel signal separation [15], blind deconvolution of SIMO channels [10], blind deconvolution of MIMO channels [16]. An information-theoretic derivation is given in [11].

3.2 The Spatio-temporal Decorrelation Algorithm: Natural Gradient

Let us consider n observations: $x_1(k), \ldots, x_n(k)$ over a $(N+1)$-point time block and the corresponding n outputs: $y_1(k), \ldots, y_n(k)$ over the same time block, defined by the following vectors:

$$\begin{aligned}\mathcal{X} &= [\boldsymbol{x}^T(0), \boldsymbol{x}^T(1), \ldots, \boldsymbol{x}^T(N)]^T, \\ \mathcal{Y} &= [\boldsymbol{y}^T(0), \boldsymbol{y}^T(1), \ldots, \boldsymbol{y}^T(N)]^T.\end{aligned} \quad (18)$$

Both input and output vectors, $\boldsymbol{x}(k)$ and $\boldsymbol{y}(k)$ are zeros for $k < 0$.

In terms of $\mathcal{X}$ and $\mathcal{Y}$, we can write the collection of n inputs over a $(N+1)$-point time block as

$$\mathcal{X} = \mathcal{W}\mathcal{Y}, \quad (19)$$

where

$$\mathcal{W} = \begin{bmatrix} \boldsymbol{I} - \boldsymbol{W}_0 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ -\boldsymbol{W}_1 & \boldsymbol{I} - \boldsymbol{W}_0 & \cdots & \boldsymbol{0} \\ \vdots & & & \vdots \\ -\boldsymbol{W}_N & -\boldsymbol{W}_{N-1} & \cdots & \boldsymbol{I} - \boldsymbol{W}_0 \end{bmatrix} \quad (20)$$

The length of delay, L in the dynamic recurrent network is less than N, i.e., $\boldsymbol{W}_{L+1} = \cdots = \boldsymbol{W}_N = \boldsymbol{0}$. The joint density of the observation $\mathcal{X}$ and the joint density of $\mathcal{Y}$ has the following relation

$$p(\mathcal{X}) = \frac{1}{|\det(\boldsymbol{I} - \boldsymbol{W}_0)^{(N+1)}|} p(\mathcal{Y}). \quad (21)$$

Note that $\det(\mathcal{W}) = \det(\boldsymbol{I} - \boldsymbol{W}_0)^{(N+1)}$ because of the triangular structure of the matrix $\mathcal{W}$.

As in the case of spatial decorrelation, the optimization function that we consider is the energy of the network output that results from the Kullback-Leibler divergence between jointly zero-mean Gaussian distribution and the product of zero-mean marginal Gaussian distributions with unit variance. Only difference compared to the case of spatial decorrelation is that we consider a block of data because the temporal decorrelation should be incorporated. The risk $R(\boldsymbol{W}(z))$ that we consider here is the normalized Kullback-Leibler divergence between $p(\mathcal{Y})$ and $\prod_{i=1}^{n} \prod_{k=0}^{N} p_i(y_i(k)) = \prod_{i=1}^{n} p_i^{N+1}(y_i)$, where $p(\mathcal{Y})$ is the jointly zero-mean Gaussian probability density function and $p_i(y_i(k))$ is the zero-mean Gaussian probability density function with unit variance.

Then the risk $R(\boldsymbol{W}(z))$ is given by

$$\begin{aligned}R(\boldsymbol{W}(z)) &= E\{L(\boldsymbol{W}(z))\} \\ &= \frac{1}{N+1} \int p(\mathcal{Y}) \log \frac{p(\mathcal{Y})}{\prod_{i=1}^{n} p_i^{N+1}(y_i)} d\mathcal{Y},\end{aligned} \quad (22)$$

where $L(\boldsymbol{W}(z))$ is the loss function.

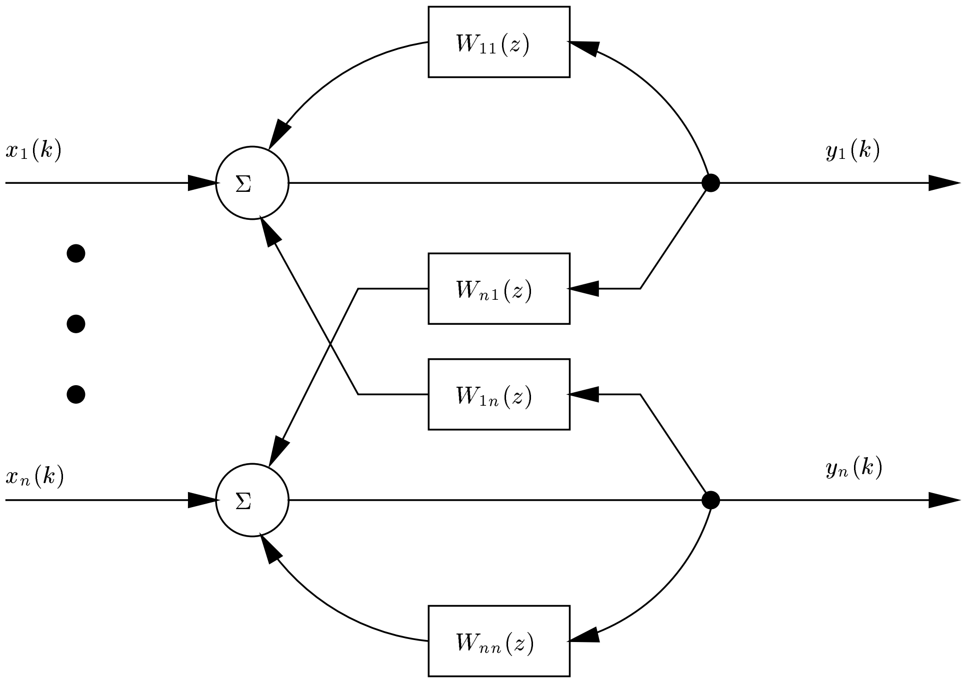


Fig. 1 The dynamic recurrent network for spatio-temporal decorrelation.

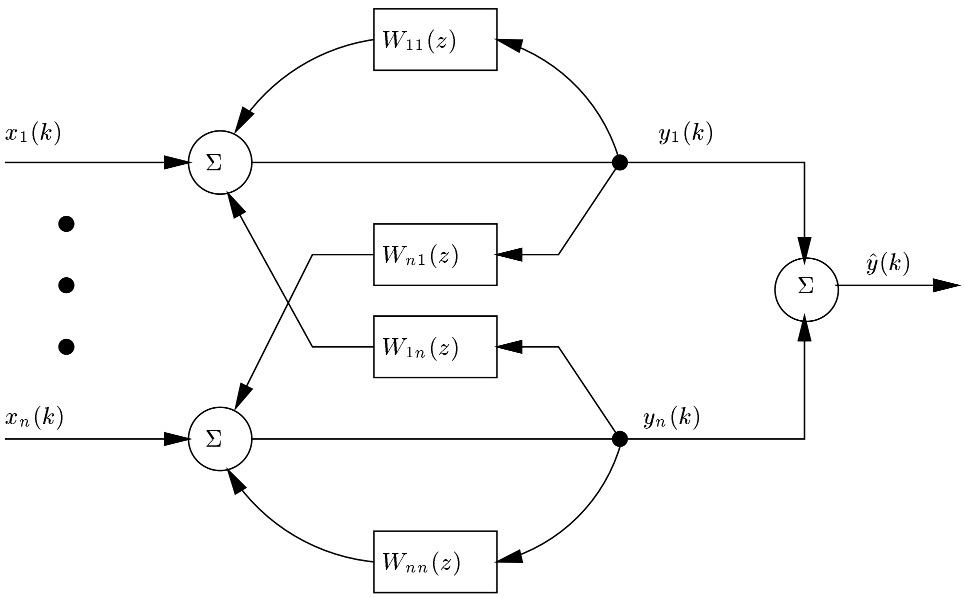


Fig. 2 The dynamic recurrent network for blind deconvolution of SIMO channel.

Thus, invoking (21), the loss function $L(\boldsymbol{W}(z))$ is

$$L(\boldsymbol{W}(z)) = -\log|\det(\boldsymbol{I}-\boldsymbol{W}_0)^{-1}| + \frac{1}{2}\sum_{i=1}^{n} y_i^2, \tag{23}$$

where an irrelevant term $\frac{1}{N+1}\log p(\mathcal{X})$ was omitted and $p_i(y_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_i^2}$ was used.

Now we derive the natural gradient algorithm for spatio-temporal decorrelation. We employ the similar technique that was used to derive the natural gradient based spatial decorrelation algorithm (13). In order to derive an adaptive algorithm for minimizing the loss function (23), we calculate an infinitesimal increment,

$$dL(\boldsymbol{W}(z)) = L(\boldsymbol{W}(z) + d\boldsymbol{W}(z)) - L(\boldsymbol{W}(z)), \tag{24}$$

corresponding to an increment $d\boldsymbol{W}(z) = d\boldsymbol{W}_0 + d\boldsymbol{W}_1 z^{-1} + \cdots$.

Simple algebraic and differential calculus yields

$$d\left\{\frac{1}{2}\sum_{i=1}^{n} y_i^2\right\} = \boldsymbol{y}^T(k)d\boldsymbol{y}(k), \tag{25}$$

where $d\boldsymbol{y}(k)$ is given by

$$\begin{aligned} d\boldsymbol{y}(k) &= [\boldsymbol{I}-\boldsymbol{W}(z)]^{-1} d\boldsymbol{W}(z) [\boldsymbol{I}-\boldsymbol{W}(z)]^{-1}\boldsymbol{x}(k) \\ &= [\boldsymbol{I}-\boldsymbol{W}(z)]^{-1} d\boldsymbol{W}(z)\boldsymbol{y}(k). \end{aligned} \tag{26}$$

Define a modified differential $d\boldsymbol{V}(z)$ as

$$d\boldsymbol{V}(z) = [\boldsymbol{I}-\boldsymbol{W}(z)]^{-1} d\boldsymbol{W}(z). \tag{27}$$

With this definition, we have

$$d\left\{\frac{1}{2}\sum_{i=1}^{n} y_i^2\right\} = \boldsymbol{y}^T(k)d\boldsymbol{V}(z)\boldsymbol{y}(k). \tag{28}$$

Similarly, it can be shown that (see Appendix for detailed derivation)

$$d\left\{\log|\det(\boldsymbol{I}-\boldsymbol{W}_0)^{-1}|\right\} = \operatorname{tr}\{d\boldsymbol{V}_0\}. \tag{29}$$

Thus, combining (28) and (29) gives

$$dL(\boldsymbol{W}(z)) = \boldsymbol{y}^T(k)d\boldsymbol{V}(z)\boldsymbol{y}(k) - \operatorname{tr}\{d\boldsymbol{V}_0\}. \tag{30}$$

This gives the following learning algorithm in terms of $d\boldsymbol{V}(z)$ for the minimization of the loss function (23),

$$\begin{aligned} \Delta\boldsymbol{V}_p(k) &= -\eta_k \frac{dL(\boldsymbol{W}(z))}{d\boldsymbol{V}_p} \\ &= \eta_k\left\{\boldsymbol{I}\delta_p - \boldsymbol{y}(k)\boldsymbol{y}^T(k-p)\right\}. \end{aligned} \tag{31}$$

Note that $d\boldsymbol{V}(z)$ is a linear combination of the differentials $dW_{ij}(z)$. Thus $d\boldsymbol{V}(z)$ represents a valid search direction to minimize (23), because $d\boldsymbol{V}(z)$ spans the same tangent space of matrices as spanned by $d\boldsymbol{W}(z)$. Then, the learning algorithm for updating $\{\boldsymbol{W}_p\}$ is given by

$$\begin{aligned} \Delta\boldsymbol{W}_p(k) &= \Delta\boldsymbol{V}_p(k) - \sum_{r=0}^{L}\boldsymbol{W}_r(k)\Delta\boldsymbol{V}_{p-r}(k) \\ &= \eta_k\left\{\boldsymbol{I}\delta_p - \boldsymbol{y}(k)\boldsymbol{y}^T(k-p) \right. \\ &\quad \left. -\boldsymbol{W}_p(k) + \boldsymbol{Y}_p(k)\right\}, \end{aligned} \tag{32}$$

where $\boldsymbol{Y}_p(k)$ ($n\times n$ matrix) is defined as

$$\boldsymbol{Y}_p(k) = \sum_{r=0}^{L}\boldsymbol{W}_r(k)\boldsymbol{y}(k)\boldsymbol{y}^T(k-p+r). \tag{33}$$

Note that the pth coefficient matrix update in (32) depends on future outputs of the network, $\boldsymbol{y}(k+r)$ through the definition of $\boldsymbol{Y}_p(k)$ in (33). Thus we delay $\boldsymbol{Y}_p(k)$ by L samples and assume that $\boldsymbol{W}_p(k)$ does not change much during L iterations. Therefore, the practical implementation of the learning algorithm takes the form

$$\begin{aligned} \Delta\boldsymbol{W}_p(k) &= \eta(k)\left\{\boldsymbol{I}\delta_p - \boldsymbol{y}(k)\boldsymbol{y}^T(k-p) \right. \\ &\quad \left. -\boldsymbol{W}_p(k) + \boldsymbol{Y}_p(k-L)\right\}, \end{aligned} \tag{34}$$

where

$$\boldsymbol{Y}_p(k-L) = \sum_{r=0}^{L}\boldsymbol{W}_{L-r}(k)\boldsymbol{y}(k-L)\boldsymbol{y}^T(k-p-r). \tag{35}$$

A slightly modified version of the algorithm given in (34) and (35) is also presented. The algorithm given in (34) and (35) decorrelate multivariate input data in spatio-temporal domain and normalize the output variance to be unity. It is often useful to decorrelate the data without normalization in overdetermined problem [17]. For this case, it was suggested in [17] to replace the identity matrix $\boldsymbol{I}$ in Eq. (31) by the diagonal matrix $\boldsymbol{\Lambda}_p(k)$ which is defined by

$$\boldsymbol{\Lambda}_p(k) = \operatorname{diag}\{y_1(k)y_1(k-p),\ldots,y_n(k)y_n(k-p)\} \tag{36}$$

Then, we have

$$\Delta\boldsymbol{V}_p(k) = \eta_k\left\{\boldsymbol{\Lambda}_p\delta_p - \boldsymbol{y}(k)\boldsymbol{y}^T(k-p)\right\}. \tag{37}$$

From this, the learning algorithm for $\{\boldsymbol{W}_p\}$ is given by

$$\begin{aligned} \Delta\boldsymbol{W}_p(k) &= \eta(k)\left\{\boldsymbol{\Lambda}_p(k)\delta_p - \boldsymbol{y}(k)\boldsymbol{y}^T(k-p) \right. \\ &\quad \left. -\boldsymbol{W}_p(k)\boldsymbol{\Lambda}_0(k) + \boldsymbol{Y}_p(k-L)\right\}. \end{aligned} \tag{38}$$

4. Illustration: Blind Deconvolution of Linear SIMO Systems

We illustrate the useful behavior and high performance of the natural gradient based spatio-temporal decorrelation algorithm for blind deconvolution of linear SIMO systems.

Blind deconvolution is a fundamental problem which is encountered in numerous applications such as digital communications, cable HDTV, global positioning system, and biomedical processing. In the context of blind deconvolution of linear SIMO systems, it is assumed that the ith sensor output $x_i(k)$ is generated from a linear time-invariant filter as

$$\begin{aligned} x_i(k) &= \sum_{j=0}^{M} h_{ij} s(k-j) \\ &= H_i(z) s(k), \text{ for } i = 1, 2, \ldots, n, \end{aligned} \tag{39}$$

where $s(k)$ is the *unobserved* zero-mean uncorrelated source signal, $H_i(z) = \sum_{j=0}^{M} h_{ij} z^{-j}$ is the transfer function of the ith channel.

The task of blind deconvolution of SIMO channel is to design a multiple input single output (MISO) system (see Fig. 2) so that its composite output $\hat{y}(k) = y_1(k) + \cdots y_n(k)$ is possibly scaled and/or delayed estimate of the original source signal $s(k)$, i.e., $\hat{y}(k) = \lambda s(k-d)$, where λ is some scaling factor and d is a delay. The term "blind" means that the problem should be solved without the knowledge of channels $\{H_i(z)\}$ and source signal $s(k)$.

In general, the blind deconvolution task resorts to higher-order statistics. However, the spatial diversity (from multiple sensors) in blind deconvolution of SIMO systems, allows us to identify the channels [18], [19] or estimate source signal [8], [11] by only second-order statistics, provided that multiple channels do not have common zeros (except for zeros at origin). In [10], [11], it was shown that spatio-temporal decorrelation of the network output can deconvolve the channels blindly. It was shown [10], [11] that if $E\{\boldsymbol{y}(k)\boldsymbol{y}^T(k-p)\} = 0$ for $p = 1, \ldots, L$, then the composite output $\hat{y}(k)$ becomes $\hat{y}(k) = \alpha s(k-d)$ for some scaling factor α and unknown delay d. The spatio-temporal anti-Hebbian learning algorithm given in Eq. (17) was employed in [10], [11]. In this section, we compare the natural gradient based spatio-temporal decorrelation algorithm to the spatio-temporal anti-Hebbian rule.

In simulation, source signal $s(k)$ is assumed to be temporally uncorrelated and to consist of random variables that are uniformly distributed over the binary set $\{+1, -1\}$. Two sensor signals $x_1(k)$ and $x_2(k)$ are assumed to be generated by

$$x_1(k) = H_1(z) s(k), \tag{40}$$
$$x_2(k) = H_2(z) s(k), \tag{41}$$

where $H_1(z) = .4983 + .2491 z^{-3} + .8305 z^{-6}$ and $H_2(z) = .3714 + .7428 z^{-3} + .5571 z^{-4}$.

Here, the synaptic weight matrix $\boldsymbol{W}_0$ was set as zero matrix since the decorrelation between $\boldsymbol{y}(k)$ and $\boldsymbol{y}(k-p)$ for $p = 1, \ldots, L$ is sufficient to deconvolve the channels (see [11] for more details). The length L in the dynamic recurrent network was chosen as $L = 10$. The learning rate $\eta_t = 0.001$ was used in the simulation. We have tested the spatio-temporal anti-Hebbian rule (17) and the natural gradient based spatio-temporal decorrelation algorithm (38). The composite output $\hat{y}(k) = y_1(k) + y_2(k)$ is shown in Fig. 3. We can easily see that the natural gradient based spatio-temporal decorrelation algorithm has faster convergence and better performance than the spatio-temporal anti-Hebbian rule.

The mean square error (MSE) was calculated by sample averaging over a 30-point rectangular window, after the arbitrary delay was removed. The time evolution of MSE is shown in Fig. 4.

5. Conclusions

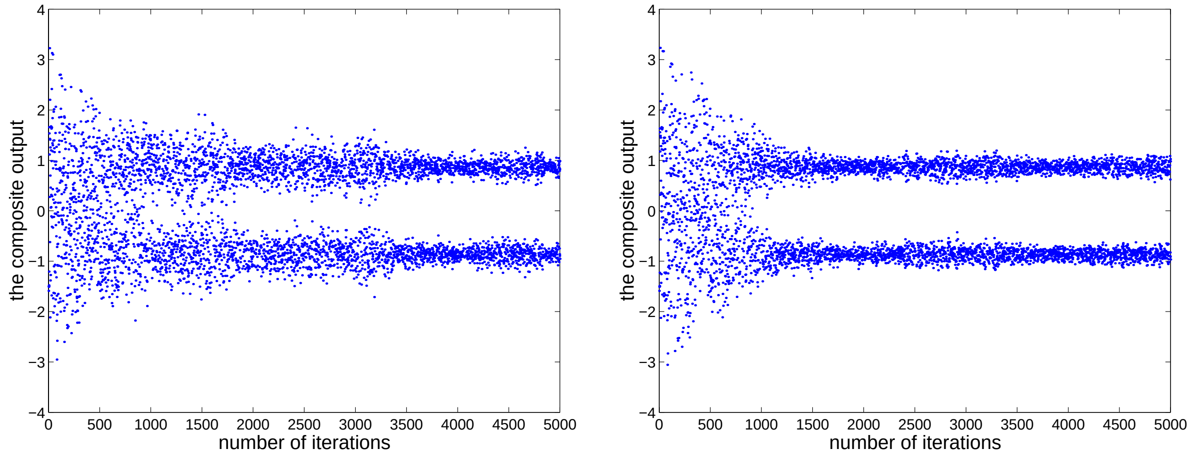
In this paper we have considered (dynamic) recurrent networks for the task of spatial decorrelation and spatio-temporal decorrelation. The natural gradient adaptation method was employed to derive efficient algorithms. This paper addressed two fundamental contributions:

(1) In the task of spatial decorrelation, we have revealed how Foldiak's anti-Hebbian rule is related to energy minimizer. Moreover, we have calculated the natural gradient for the recurrent network and have derived an efficient spatial decorrelation algorithm.
(2) In the task of spatio-temporal decorrelation, an appropriate optimization function was introduced. We have also calculated the natural gradient for the dynamic recurrent network for the minimization of the risk and have derived an efficient spatio-temporal decorrelation algorithm.

In order to illustrate the useful behavior of the natural gradient based spatio-temporal decorrelation algorithm, we have applied it to the problem of blind deconvolution of linear SIMO systems. The performance of the proposed algorithm was compared to that of the spatio-temporal anti-Hebbian rule. Rigorous derivations of the algorithms were presented and computer simulations verified the high performance of the proposed algorithms.

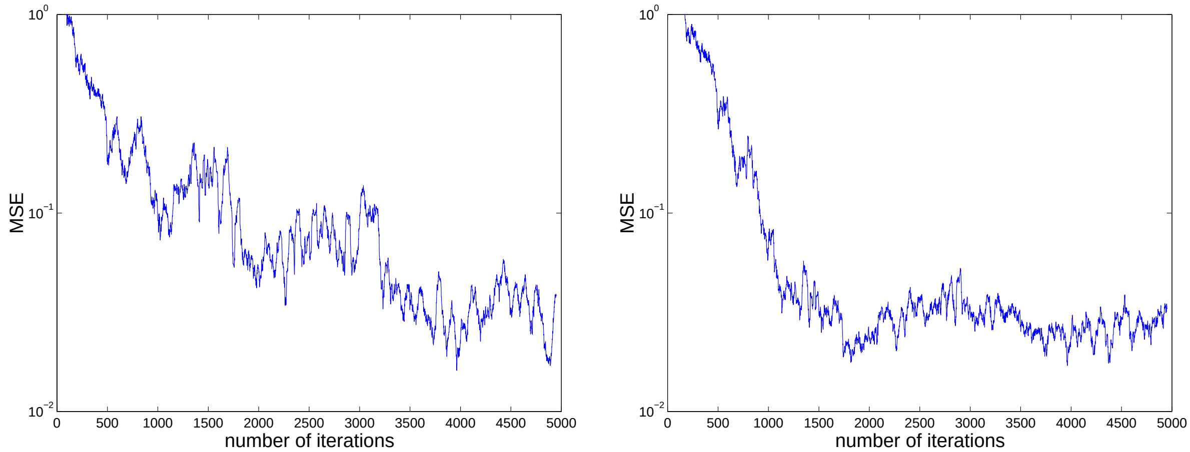
References

[1] E. Oja, "Neural networks, principal component analysis, and subspaces," *International Journal of Neural Systems*, vol. 1, pp. 61–68, 1989.
[2] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks: Theory and Applications*, John Wiely & Sons, INC, 1996.
[3] P. Földiák, "Adaptive network for optimal linear feature extraction," in *Proc. Int. Joint Conf. Neural Networks*, 1989, pp. 401–405.
[4] L. B. Almeida and F. M. Silva, "Adaptive decorrelation," in *Artificial Neural Networks*, pp. 149–156. 1992.



(a) (b)

Fig. 3 The output of the equalizer, $\hat{y}(k) = y_1(k) + y_2(k)$: (a) the spatio-temporal anti-Hebbian learning (17); (b) the natural gradient spatio-temporal decorrelation algorithm (38).



(a) (b)

Fig. 4 Mean square error (MSE): (a) the spatio-temporal anti-Hebbian learning (17); (b) the natural gradient based spatio-temporal decorrelation algorithm (38).

[5] S. C. Douglas and A. Cichocki, "Neural networks for blind decorrelation of signals," *IEEE Trans. Signal Processing*, vol. 45, no. 11, pp. 2829–2842, 1997.
[6] F. Palmieri, J. Zhu, and C. Chang, "Anti-Hebbian learning in topologically constrained linear networks: A tutorial," *IEEE Trans. Neural Networks*, vol. 5, no. 4, pp. 748–761, Sep. 1993.
[7] S. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, no. 2, pp. 251–276, 1998.
[8] S. Choi, "Adaptive blind equalization for multi-channel transmission via temporal decorrelation," *Journal of Electrical Engineering and Information Science*, vol. 4, no. 1, pp. 130–136, Feb. 1999.
[9] S. Choi, H. Luo, and R. Liu, "An adaptive system for direct multi-channel blind equalization," in *Proc. SPAWC*, Paris, France, 1997, pp. 85–88.
[10] S. Choi, A. Cichocki, and S. Amari, "Blind equalization of SIMO channels via spatio-temporal anti-Hebbian learning rule," in *Neural Networks for Signal Processing VIII*, T. Constantinides, S. Y. Kung, M. Niranjan, and E. Wilson, Eds., 1998, pp. 93–102.
[11] S. Choi, "Linear neural networks with FIR synapses for blind deconvolution and equalization," *Journal of Electrical Engineering and Information Science*, vol. 4, no. 2, pp. 224–231, Apr. 1999.
[12] S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. SPAWC*, Paris, France, 1997, pp. 101–104.
[13] S. Choi, S. Amari, and A. Cichocki, "Natural gradient learning algorithms for decorrelation," in *Proc. ICONIP*, Dunedin, New Zealand, 1997, vol. 1, pp. 645–648.
[14] J. -F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.
[15] S. Van Gerven and D. Van Compernolle, "Signal separation by symmetric adaptive decorrelation: Stability, convergence, and uniqueness," *IEEE Trans. Signal Processing*, vol. 43, no. 7, pp. 1602–1612, 1995.
[16] S. Choi and A. Cichocki, "An unsupervised hybrid network for blind separation of independent non-gaussian source signals in multipath environment," *Journal of Communications and Networks*, vol. 1, no. 1, pp. 19–25, 1999.
[17] S. Choi, S. Amari, A. Cichocki, and R. Liu, "Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels," in *Proc. ICA'99*, Aussois, France, 1999, pp. 371–376.
[18] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channel," in *Proc. Asilomar Conference*, 1991, pp. 340–349.
[19] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multipath channel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, 1995.

Appendix A: Detailed Derivation of Eq. (29)

We rewrite Eq. (27) as

$$[\boldsymbol{I} - \boldsymbol{W}(z)]\, d\boldsymbol{V}(z) = d\boldsymbol{W}(z). \tag{A·1}$$

Then we have

$$\begin{aligned}&\left[\boldsymbol{I} - \boldsymbol{W}_0 - \boldsymbol{W}_1 z^{-1} - \cdots\right]\left(d\boldsymbol{V}_0 + d\boldsymbol{V}_1 z^{-1} + \cdots\right)\\&\quad = d\boldsymbol{W}_0 + d\boldsymbol{W}_1 z^{-1} + \cdots .\end{aligned} \tag{A·2}$$

From this one can easily see

$$d\boldsymbol{V}_0 = [\boldsymbol{I} - \boldsymbol{W}_0]^{-1}\, d\boldsymbol{W}_0. \tag{A·3}$$

Now we calculate

$$\begin{aligned}& d\left\{\log|\det(\boldsymbol{I} - \boldsymbol{W}_0)^{-1}|\right\}\\&\quad = \mathrm{tr}\left\{d(\boldsymbol{I} - \boldsymbol{W}_0)^{-1}(\boldsymbol{I} - \boldsymbol{W}_0)\right\}\\&\quad = \mathrm{tr}\left\{(\boldsymbol{I} - \boldsymbol{W}_0)^{-1} d\boldsymbol{W}_0\right\}\\&\quad = \mathrm{tr}\left\{d\boldsymbol{V}_0\right\}.\end{aligned} \tag{A·4}$$

Seungjin CHOI was born in Seoul, Korea, on October 26, 1964. He received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Korea, in 1987 and 1989, respectively and the Ph.D degree in electrical engineering from the University of Notre Dame, Indiana, in 1996.
After spending the fall of 1996 as a Visiting Assistant Professor in the Department of Electrical Engineering at University of Notre Dame, Indiana, he was as Frontier Researcher with the Laboratory for Artificial Brain Systems, RIEKN in Japan. In August 1997, he joined the School of Electrical and Electronics Engineering at Chungbuk National University where he is currently an Assistant Professor. He has also been an Invited Senior Research Fellow at Advanced Brain Signal Processing Laboratory, Brain-style Information Systems Research Group in Brain Science Institute, RIKEN in Japan. His current research interests include brain information processing, statistical (blind) signal processing, independent component analysis, unsupervised learning, and multiuser communications.

Shun-ichi AMARI was born in Tokyo, Japan, on January 3, 1936. He graduated from the University of Tokyo in 1958, having majored in mathematical engineering, and he received the Dr.Eng. degree from the University of Tokyo in 1963.
He was an Associate Professor at Kyushu University, an Associate and then Full Professor at the Department of Mathematical Engineering and Information Physics, University of Tokyo, and is now Professor-Emeritus at the University of Tokyo. He is the Director of the Brain-Style Information Systems Group, RIKEN Brain Science Institute, Saitama Japan. He has been engaged in research in wide areas of mathematical engineering and applied mathematics, such as topological network theory, differential geometry of continuum mechanics, pattern recognition, mathematical foundations of neural networks, and information geometry.
Dr. Amari served as President of the International Neural Network Society, Council member of Bernoulli Society for Mathematical Statistics and Probability Theory, and Vice President of the Institute of Electrical, Information and Communication Engineers. He was founding Coeditor-in-Chief of Neural Networks.

He has been awarded the Japan Academy Award, the IEEE Neural Networks Pioneer Award, and the IEEE Emanuel R. Piore Award.

Andrzej CICHOCKI was born in Poland on August 1947. He received the M.Sc.(with honors), Ph.D., and Habilitate Doctorate (Dr.Sc.) degrees, all in electrical engineering and computer science, from Warsaw University of Technology (Poland) in 1972, 1975, and 1982, respectively.
Since 1972, he has been with the Institute of Theory of Electrical Engineering and Electrical Measurements at the Warsaw University of Technology, where he became a full Professor in 1995.
He is the co-author of two international books: MOS Switched-Capacitor and Continuous-Time Integrated Circuits and Systems (Springer-Verlag, 1989) and Neural Networks for Optimization and Signal Processing (J Wiley and Teubner Verlag,1993/94) and author or co-author of more than hundred fifty (150) scientific papers.
He spent at University Erlangen-Nuernberg (GERMANY) a few years as Alexander Humboldt Research Fellow and Guest Professor. In 1995-96 he has been working as a Team Leader of the Laboratory for Artificial Brain Systems, at the Frontier Research Program RIKEN (JAPAN), in the Brain Information Processing Group directed by professor Shun-ichi Amari. Currently he is head of the laboratory for Open Information Systems in the Brain Science Institute, Riken, Wako-schi, JAPAN.
He is reviewer of several international Journals, e.g. IEEE Trans. on Neural Networks, Signal Processing, Circuits and Systems, Biological Cybernetics, Electronics Letters, Neurocomputing, Neural Computation. He is also member of several international Scientific Committees and the associated Editor of IEEE Transaction on Neural Networks (since January 1998). His current research interests include signal and image processing (especially blind signal/image processing), neural networks and their electronic implementations, learning theory and algorithms, independent and principal component analysis, optimization problems, circuits and systems theory and their applications, artificial intelligence.