A NOVEL UNSUPERVISED DECONVOLUTION NETWORK: EFFICIENT FOR A SPARSE SOURCE

Seungjin CHOI

School of Electrical and Electronics Engineering
Chungbuk National University, KOREA
schoi@engine.chungbuk.ac.kr

ABSTRACT

This paper presents a novel neural network structure to the blind deconvolution task where the input (source) to a system is not available and the source has any type of distribution including sparse distribution. We employ multiple sensors so that spatial information plays a important role. The resulting learning algorithm is linear so that it works for both sub- and super-Gaussian source. Moreover, we can successfully deconvolve the mixture of a sparse source, while most existing algorithms [5] have difficulties in this task. Computer simulations confirm the validity and high performance of the proposed algorithm.

1. INTRODUCTION

Deconvolution is the reverse processing of convolution, in which the task is to determine the impulse response of the unknown system or the input signal (source). When the input signal to a system is available, this is a problem in system identification. In many important applications, however, the input to a system is not available, where the problem is called *blind deconvolution*. In this paper, we tackle the blind deconvolution task in unsupervised learning manner, that is why we refer it to *unsupervised deconvolution*. Unsupervised deconvolution has a great deal of applications such as digital communications, seismic deconvolution, image restoration, and biomedical signal reconstruction.

The widely-used blind deconvolution method is a family of Bussgang-type algorithms (see [5] and references therein) which has been successfully applied to digital communications where the source is typically sub-Gaussian. Recently there has been a great deal of interest in sparseness [9]. Although it was shown that a super-Gaussian source could be recovered through the maximization of kurtosis, no successful demonstration was not reported, to our knowledge. In this letter, we present a novel network structure and an associated simple algorithm for unsupervised deconvolution especially for a sparse source. One exemplary simulation result is given in Section 4 to demonstrate the successful result for a sparse source.

2. DECONVOLUTION: MULTIPLE SENSORS

When the single observation is available, we have to rely on higher-order statistics (for example, kurtosis) for deconvolution task [7]. It was shown that multiple measurements (either from multiple sensors or from oversampling) provide extra information so that blind identification can be achieved by second-order statistics only [8]. Motivated by this result, we employ multiple sensors to obtain several measurements (spatial diversity). The spatial diversity enables us to use linear learning for deconvolution task.

Let us assume that there are two measurement signals are available.[1] Two different observations $x_1(k)$ and $x_2(k)$ are assumed to be generated through

$$x_i(k) = H_i(z)s(k), \quad \text{for } i = 1, 2, \tag{1}$$

where $H_i(z) = \sum_{p=0}^{M} h_{i,p} z^{-p}$ and z^{-1} is the delay operator, i.e., $z^{-1}s(k) = s(k-1)$. We assume that $H_1(z)$ and $H_2(z)$ are co-primes. The source signal $s(k)$ is assumed to temporally uncorrelated Under these two conditions, the deconvolution task is to find an innovation sequence of observations [3, 1, 2]. The approach we take here is to construct a feedback network (see Figure 1) to minimize statistical correlation between $y_i(k)$ and $y_j(k+\tau)$ for any $\tau \neq 0$, $i, j = 1, 2$. One can easily see that

$$E\{y_i(k)y_j(k+\tau)\} = 0, \quad \forall \tau \neq 0, i, j = 1, 2 \tag{2}$$

implies that $y(k) = y_1(k) + y_2(k)$ is an innovation sequence (a white sequence). Note that an innovation

[1]Multiple sensors might improve the performance, but two sensors are sufficient.

sequence is not unique. The white sequence scaled by any constant can be an innovation sequence. This corresponds to a scaling ambiguity in blind deconvolution. Another ambiguity is an unknown delay. Thus the estimate of the source signal would be $y(k) = \alpha s(k-d)$, where α is a scaling factor and d is an unknown delay.

3. THE LEARNING ALGORITHM

For spatio-temporal decorrelation task, we consider a linear feedback network (see Figure 1) where the input-output relation is given by

$$y_i(k) = x_i(k) + \sum_{j=1}^{2}\sum_{p=1}^{L} w_{ij,p}(k) y_j(k-p), \quad (3)$$

where $w_{ij,p}(k)$ is the connection strength between $y_i(k)$ and $y_j(k-p)$. Let us consider two observations: $x_1(k)$, $x_2(k)$ over a N-point time block and the corresponding two outputs: $y_1(k), y_2(k))$ over the same time block, defined by the following vectors:

$$\mathcal{X} = [x_1(0), x_2(0), \ldots, x_1(N-1), x_2(N-1)]^T \quad (4)$$
$$\mathcal{Y} = [y_1(0), y_2(0), \cdots, y_1(N-1), y_2(N-1)]^T \quad (5)$$

As an optimization function, we choose the Kullback-Leibler divergence between the joint density of the network output $p(\mathcal{Y})$ and the factorizable density $q(\mathcal{Y}) = \prod_{i=1}^{2}\prod_{k=0}^{N-1} q_i(y_i(k))$. Thus the risk $R(\{w_{ij,p}\})$ is given by

$$\begin{aligned} R(\{w_{ij,p}\}) &= \frac{1}{N}\int p(\mathcal{Y}) \log \frac{p(\mathcal{Y})}{\prod_{i=1}^{2}\prod_{k=0}^{N-1} q_i(y_i(k))} d\mathcal{Y} \\ &= \frac{1}{N} E\{\log p(\mathcal{Y})\} - \sum_{i=1}^{2} E\{\log q_i(y_i)\}. \end{aligned} \quad (6)$$

Note that the assumption on identical distribution was used, i.e.,

$$\prod_{k=0}^{N-1} q_i(y_i(k)) = \{q_i(y_i)\}^N. \quad (7)$$

To determine the Jacobian $|\frac{d\mathcal{X}}{d\mathcal{Y}}|$, we write (3) in a compact form

$$\mathcal{X} = \mathcal{W}\mathcal{Y}, \quad (8)$$

where

$$\mathcal{W} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{W}_1 & \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & & & \vdots \\ -\mathbf{W}_{N-1} & -\mathbf{W}_{N-2} & \cdots & \mathbf{I} \end{bmatrix}, \quad (9)$$

where the matrix $\mathbf{W}_p$ is a collection of $w_{ij,p}$. One can easily see that the Jacobian $|\frac{d\mathcal{X}}{d\mathcal{Y}}| = N$, which can be viewed as a volume preserving transformation [4]. Taking this into account, the risk (6) is rewritten as

$$R(\{w_{ij,p}\}) = -\sum_{i=1}^{2} E\{\log q_i(y_i)\}. \quad (10)$$

For the minimization of (10), we employ the stochastic gradient descent. The learning algorithm for $\{w_{ij,p}(k)\}$ has the form

$$w_{ij,p}(k+1) = w_{ij,p}(k) - \eta_k \varphi_i(y_i(k)) y_j(k-p), \quad (11)$$

where $\eta_k > 0$ is the learning rate and $\varphi_i(y_i(k))$ is defined by

$$\varphi_i(y_i(k)) = -\frac{d \log q_i(y_i(k))}{dy_i(k)}. \quad (12)$$

As a special case which we are interested in this paper, we can use a Gaussian density model (zero mean and unit variance) for $q_i(k)$, which results in $\varphi_i(y_i(k)) = y_i(k)$. Since the deconvolution is possible using only linear learning, our approach does not require to know the probability distribution of source. Although the proposed method works for sub-Gaussian source, its performance is worse than the well-known Bussgang-type algorithms [5]. However, the proposed method works surprisingly well for a sparse source , whereas some existing methods do not. It was also reported [6] that the linear learning was successful in separating two speech signals which are typically super-Gaussian.

4. COMPUTER SIMULATIONS

We present one exemplary demonstration. A source signal having sparse distribution (kurtosis=273) was used. (see Figure 2) Two sensors signal $x_1(k)$ and $x_2(k)$ were generated through the following FIR filters:

$$\begin{aligned} H_1(z) &= z^{-1} + .8z^{-8} + .6z^{-16} + .4z^{-20} \\ &\quad + .3z^{-24} + .1z^{-26} \\ H_2(z) &= .9z^{-3} + .7z^{-11} + .5z^{-18} + .8z^{-22} \\ &\quad + .2z^{-26} + .1z^{-30}. \end{aligned} \quad (13)$$

The signal to noise ratio (SNR) was 20dB. The length of delay in the network (see Figure 1) was $L = 35$ The learning rate was set as $\eta_k = .0001$. All synaptic weight matrices was initially set to zeros.

We have applied the algorithm based on the maximization of kurtosis in this task, and the result was not successful. The successful result using the proposed network and algorithm are shown in Figure 2. It can be observed that the recovered signal $y(k)$ is almost identical to the source signal $s(k)$. Although one can see small distortions in $y(k)$, the point where the pulse occurs is exactly recovered.

5. CONCLUSIONS

We have presented a novel neural network structure and an associate simple learning algorithm for the unsupervised deconvolution task. Since we employ multiple sensors, it was possible to perform deconvolution using linear learning. The main contribution of this letter is summarized as follows: (1) A novel neural network structure was incorporated into the unsupervised deconvolution task; (2) A simple spatio-temporal decorrelation algorithm was derived from an information-theoretic viewpoint and was successfully applied to the case where the mixture was filtered version of a spare source.

Acknowledgement

This work was supported in part by Korea Science and Engineering Foundation under grant 981-0913-063-1.

6. REFERENCES

[1] S. Choi. Temporal decorrelation-based blind equalization (tdbe) algorithm for multichannel transmission. In *Joint Conference on Communications and Information*, volume 2, pages 641–645, 1998.

[2] S. Choi, A. Cichocki, and S. Amari. Blind equalization of SIMO channels via spatio-temporal anti-Hebbian learning rule. In *IEEE Workshop on Neural Networks for Signal Processing*, pages 93–102, Cambridge, UK, 1998.

[3] S. Choi, H. Luo, and R. Liu. An adaptive system for direct multi-channel blind equalization. In *The First Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, pages 85–88, Paris, France, 1997.

[4] G. Deco and D. Obradovic. *An Information Theoretic Approach to Neural Computing*. Springer-Verlag, 1996.

[5] S. Haykin eds. *Blind Deconvolution*. Prentice-Hall, Englewood Cliffs, NJ, 1994.

[6] M. Girolami and C. Fyfe. A temporal model of linear anti-Hebbian learning. *Neural Processing Letters*, 4(3):139–148, 1996.

[7] D. N. Godard. Self-recovering equalization and carrier tracking in two-dimensional data communication systems. *IEEE Trans. Communications*, 28:1867–1875, 1980.

[8] 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*, 43:516–525, 1995.

[9] B. Rao. Signal processing with the sparseness constraint. In *Proc ICASSP*, pages 1861–1864, 1998.

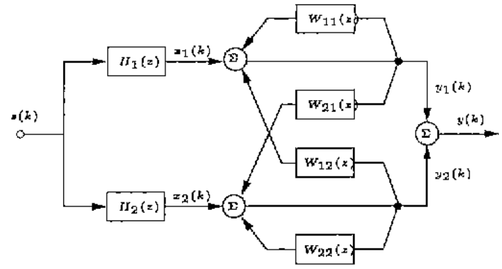

Figure 1: The structure of the neural network for unsupervised deconvolution.

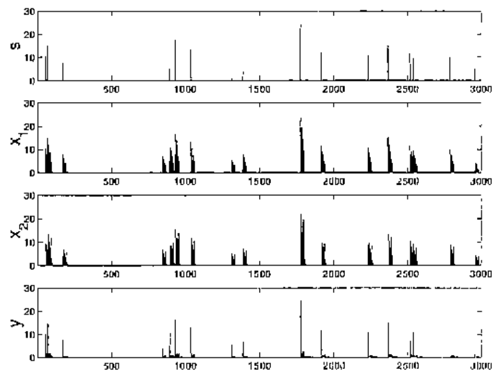

Figure 2: From top to bottom, the source signal ($s(k)$, mixtures $x_1(k)$, $x_2(k)$, and the recovered signal $y(k)$ are shown. It is plotted over the duration [27001, 30000].