Blind Deconvolution with Sparse Priors on the Deconvolution Filters

Hyung-Min Park[1], Jong-Hwan Lee[1],
Sang-Hoon Oh[2], and Soo-Young Lee[1]

[1] Brain Science Research Center and Dept. Biosystems,
Korea Advanced Institute of Science and Technology,
Daejeon, 305-701, Republic of Korea
hmpark@kaist.ac.kr
[2] Dept. Information Communication Eng.,
Mokwon University, Daejeon, 302-729, Republic of Korea

Abstract. In performing blind deconvolution to remove reverberation from speech signal, most acoustic deconvolution filters need a great many number of taps, and acoustic environments are often time-varying. Therefore, deconvolution filter coefficients should find their desired values with limited data, but conventional methods need lots of data to converge the coefficients. In this paper, we use sparse priors on the acoustic deconvolution filters to speed up the convergence and obtain better performance. In order to derive a learning algorithm which includes priors on the deconvolution filters, we discuss that a deconvolution algorithm can be obtained by the joint probability density of observed signal and the algorithm includes prior information through the posterior probability density. Simulation results show that sparseness on the acoustic deconvolution filters can be successfully used for adaptation of the filters by improving convergence and performance.

1 Introduction

Blind deconvolution has become an important topic for research and development in digital signal processing because it has high potential for broad applications in speech enhancement as well as communications. Especially, blind deconvolution in acoustic environments is a very challenging problem because natural acoustic signals are time-correlated and deconvolution for acoustic environments are very complex.

For example, let us consider the teleconferencing problem, in which people talk into a microphone located not at their mouth as in usual telephone conversation, but located some distance away. The speech is reverberated and can have interference among phonemes. In that case, the speech intelligibility is degraded. We can model the situation as a single-input-single-output (SISO) discrete-time linear system, in which the relationship between the input and the output signal is given by

$$x(n) = \sum_{k=0}^{L_m-1} h(k)s(n-k) + v(n). \tag{1}$$

J. Rosca et al. (Eds.): ICA 2006, LNCS 3889, pp. 658–665, 2006.

The goal of blind deconvolution is to recover the input signal $s(n)$ from the output $x(n)$ when the channel $h(k)$ is unknown. Typically, the noise sequence $v(n)$ is modeled by a zero-mean white Gaussian noise process.

Many researchers have studied on the problem and proposed a number of blind deconvolution algorithms [1, 2]. In most of the blind deconvolution methods, a causal finite-impulse-response (FIR) filter as a linear deconvolutive system is used to recover the input signal $s(n)$. Hence, the deconvolutive model can be formulated by

$$u(n) = \sum_{k=0}^{L_a-1} w(k)x(n-k), \tag{2}$$

where $w(k)$ is a filter coefficient of the deconvolution filter. The overall system is shown in Fig. 1. Since the blind deconvolution methods do not have a training sequence, adaptation of $w(k)$ usually makes use of some *a priori* statistical knowledge of the recovered output signal $u(n)$.

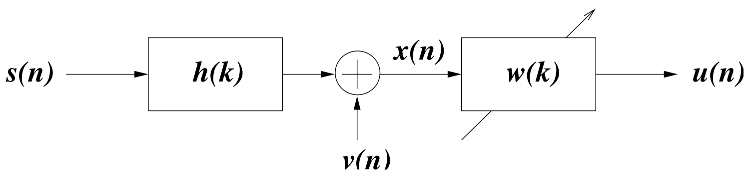

Fig. 1. Overall system for convolution and deconvolution

Among various signals, speech is so time-correlated and non-stationary that many algorithms do not work well to deconvolve it. In order to derive a robust algorithm, entropy might be a good candidate [3]. By forming the cost function as negative entropy of the output probability density function (pdf) $p(u)$ and minimizing the cost using its gradient with respect to coefficients of the deconvolution filter, the learning rule can be obtained in the frequency domain as

$$\Delta W \propto (1/W^* - \text{fft}\{g(\underline{u})\}X^*)|W|^2, \tag{3}$$

where W and X are the discrete Fourier transform of the deconvolution filter and the observed signal $x(n)$, respectively [3]. For the nonlinear function $g(u)$, we can use $-p'(u)/p(u)$.

Although some prior knowledge on the distribution of the recovered output signal is used to adapt the deconvolution filter as we mentioned in the previous paragraph, we do not usually have any assumption on the deconvolution filter $w(k)$. From a point of view, this is advantageous because the learning permits that the filter may have any kind of types. In some application fields, however, we can obtain some knowledge on the deconvolution filter or assume statistics on the filter. In those cases, the filter may be estimated more exactly or easily if we make use of the knowledge or statistics. Especially, when the deconvolution filter is too complex or the number of observed data is too limited to adapt the filter, prior information on the deconvolution filter can play an important role.

Even though assumption on a form of the mixing matrix in independent component analysis gives successful estimation of the image features [4], most of the image feature extraction problems do not suffer from lack of data by obtaining sufficient number of image patches. Thus, the assumption hardly affect the results as critical information. In this paper, we try to set priors on acoustic deconvolution filters for blind deconvolution of speech signals. It is known that most of the acoustic deconvolution filters require lots of taps. In addition, acoustic environments are often time-varying, and deconvolution algorithm should adapt the deconvolution filter in a short time period (i.e., with limited number of data). Therefore, prior information on the deconvolution filter is likely to give better estimates of the filter. The deconvolution filters actually estimate the inverse of acoustic reverberation, and the filters require very different numbers of taps according to the acoustic reverberation. In order to deconvolve various types of acoustic reverberation, we need to use a sufficient number of taps for the deconvolution filter, and a large number of taps are almost zero in most cases. Therefore, we can impose sparse priors on the acoustic deconvolution filters, and the plausible priors help us estimate more exact filters.

2 Another Derivation on the Blind Deconvolution Algorithm

For simple derivation, the deconvolutive model, Eq. (2), can be represented in the form using time-delay operator z^{-1} as

$$u(n) = W(z)x(n), \tag{4}$$

where

$$W(z) = \sum_{k=0}^{L_a-1} w(k)z^{-k}. \tag{5}$$

In order to derive the blind deconvolution algorithm, Eq. (3), let us consider the input and the output signal of the deconvolutive model over a N sample block, defined by the following vectors:

$$\begin{aligned} \mathbf{x} &= [x(0),\ x(1),\ \cdots,\ x(N-1)]^T, \\ \mathbf{u} &= [u(0),\ u(1),\ \cdots,\ u(N-1)]^T. \end{aligned} \tag{6}$$

Both the input and the output signal, $x(n)$ and $u(n)$ are zeros for $n < 0$.

Then, we can write the output signal vector $\mathbf{u}$ as

$$\mathbf{u} = \begin{bmatrix} w(0) & 0 & \cdots & 0 \\ w(1) & w(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ w(N-1) & w(N-2) & \cdots & w(0) \end{bmatrix} \mathbf{x}. \tag{7}$$

Here, $w(L_a+1) = w(L_a+2) = \cdots = w(N-1) = 0$ by assuming that the length of the channel L_a is much smaller than N.

The joint probability density of the observed signal vector $\mathbf{x}$ can be given by

$$p(\mathbf{x}|W(z)) = |w(0)^N|p(\mathbf{u}), \tag{8}$$

and $p(\mathbf{u}) = p^N(u(n))$ for an i.i.d. signal. Therefore, the log-likelihood of Eq. (8) is

$$L(W(z)) = N \log |w(0)| + N \log p(u(n)). \tag{9}$$

By maximizing the log-likelihood with respect to $w(k)$, the natural gradient algorithm [5, 6] for updating $w(k)$ is given by

$$\Delta w(k) \propto w(k) - \varphi(u(n))r_k(n), \tag{10}$$

where $\varphi(u(n))$ denotes the score function given by $-p'(u(n))/p(u(n))$, and

$$r_k(n) = \sum_{l=0}^{L_a-1} w(l)u(n-k+l). \tag{11}$$

The algorithm has the almost same form as in [7].

As an efficient way to implement the algorithm, a frequency-domain processing using the short-time Fourier transform for sample blocks can be considered. The resulting algorithm is the same as Eq. (3).

3 Imposing Sparse Priors on the Deconvolution Filters

In order to make use of priors on the deconvolution filters during adaptation of the filter coefficients, let us reconsider the joint probability density of the observed signal vector, $p(\mathbf{x}|W(z))$. In addition, we assume that the joint probability density of the deconvolution filter $p(W(z))$ is known as prior information. In that case, the logarithm of the posterior probability density can be expressed as

$$\log p(W(z)|\mathbf{x}) = \log p(\mathbf{x}|W(z)) + \log p(W(z)) - \log p(\mathbf{x}). \tag{12}$$

Maximizing $\log p(W(z)|\mathbf{x})$ with respect to $W(z)$ provides a learning algorithm for adapting $W(z)$ with priors on the deconvolution filter. Since the third term of the right side in Eq. (12) does not depend on the deconvolution filter, it does not affect the learning algorithm. Note that the first term is the same as the log-likelihood of Eq. (9). Therefore, Eq. (3) can be used for updating $W(z)$. Adding to Eq. (3), we have to maximize the second term $\log p(W(z))$ with respect to $W(z)$ as

$$\frac{d \log p(W(z))}{dW(z)} = \frac{\frac{dp(W(z))}{dW(z)}}{p(W(z))}. \tag{13}$$

As we mentioned in Section 1, a sufficient number of taps are used for the deconvolution filter, and a large number of taps are almost zero in most cases. Therefore, sparseness can be imposed on the acoustic deconvolution filters. As a simple and general pdf for sparse distribution, Laplacian distribution can be considered. For simple formulation, we also assume that coefficients of the deconvolution filter are i.i.d. Thus, equation for updating $w(k)$ from the second term in Eq. (12) is

$$\begin{aligned} \Delta w(k) &\propto \frac{d \log p(W(z))}{dw(k)} = \frac{dL_a \log p(w(k))}{dw(k)} \\ &\propto -\mathrm{sgn}(w(k)). \end{aligned} \tag{14}$$

In maximizing the posterior probability density, we obtained two equations for learning the deconvolution filter. One of them is processed in the frequency domain whereas the other is performed in the time domain. In real implementation even without prior information on the deconvolution filter, the inverse Fourier transform of the filter should be computed whenever the filter is adapted in the frequency domain. This is because one has to take the part corresponding to length of the deconvolution filter and pad zeros to the remaining part of the block. Therefore, we can easily apply Eq. (15) after the deconvolution filter in the time domain is computed.

Overall procedure for updating the deconvolution filter is as follows:

- Begin
  1. Transform the deconvolution filter into the frequency domain.
  2. Make a sample block.
  3. Transform the block into the frequency domain.
  4. Update the deconvolution filter using Eq. (3).
  5. Transform the deconvolution filter into the time domain.
  6. Update the filter using Eq. (15) and set the outside of the deconvolution filter to zero.
  7. Go to step 1.
- End

4 Experimental Results

We have performed experiments on blind deconvolution to show the effect of sparse priors on the deconvolution filters. Experimental results were compared in terms of the intersymbol interference (ISI) [3, 8], which is computed by

$$\mathrm{ISI}(dB) = 10 \log \left(\frac{\sum_k |t(k)|^2 - \max_k |t(k)|^2}{\max_k |t(k)|^2} \right), \tag{15}$$

where $t(k) = w(k) * h(k)$.

As input data to the SISO linear system of Eq. (1), some speech files from a male speaker were selected in the TIMIT database [9]. The total signal had about 27 second length, and the sampling rate was converted into $8kHz$. It is known that speech signal approximately follows Laplacian distribution. Therefore, sgn(·) was used as $g(\cdot)$ in Eq. (3). In addition, note that speech signal is not i.i.d. When one performs blind deconvolution with speech signal, the deconvolution filter learns from the signal in order to remove not only reverberation in acoustic environments but also correlation or dependence of the speech signal. To avoid the side-effect that deconvolution algorithm removes dependence of speech, a pre-whitening filter has been learned from speech and then used for whitening the signal which was reverberated by a convolutive channel.

The convolutive channel to generate the output signal $x(n)$ of the SISO linear system of Eq. (1) was a 32 tap non-minimum phase filter as shown in Fig. 2. The channel was a part of the impulse response measured in a normal office room. In order to deconvolve the channel, we have employed a 512 tap filter for $w(k)$ with tap-centering initialization. The block size was 1024 to apply the fast Fourier transform. As shown in the SISO linear system, the observed signal $x(n)$ is generally corrupted by noise which may come from various noise sources. In this paper, additive white Gaussian noise was used to corrupt the observed signal.

Fig. 3 displays the ISI for the deconvolution algorithm with sparse priors on the deconvolution filter. In this experiment, the signal-to-noise ratio (SNR) of the observed signal was 15dB. For comparison, the simulation on the algorithm

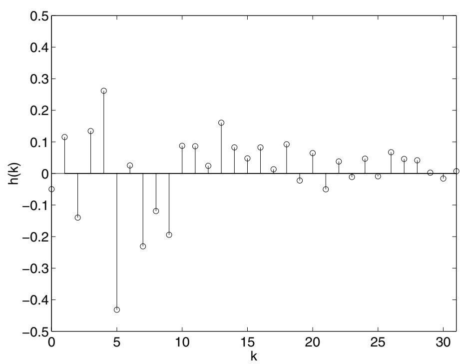

Fig. 2. A 32 tap non-minimum phase convolutive channel

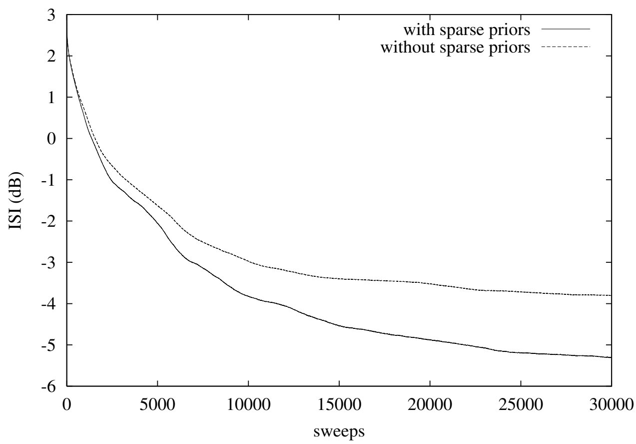

Fig. 3. The ISI for the deconvolution algorithm with the signal whose SNR is 15dB

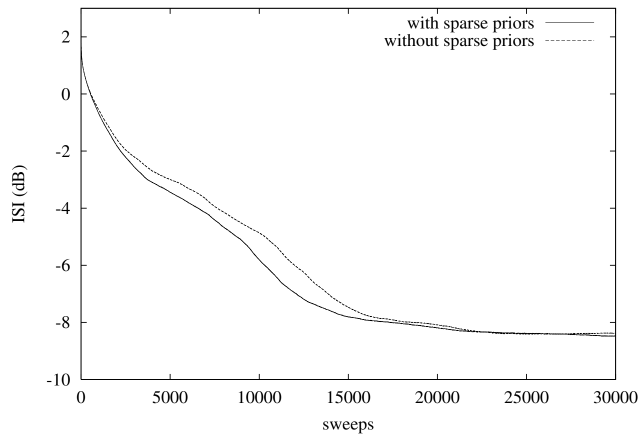

Fig. 4. The ISI for the deconvolution algorithm with the signal whose SNR is 20dB

without sparse priors has been performed with the same parameters, and the result was included. Although additional computation requirement to impose sparse priors was negligible, the deconvolution algorithm with sparse priors on the filter showed faster convergence and better performance than that without the priors. The result indicated that the deconvolution algorithm could not converge to a desired solution without sparse priors and the priors on the filter provided useful information to learn it.

In order to consider the effect of noise, we repeated the simulation for the signal whose SNR was 20dB, and Fig. 4 shows the result. The difference between the ISIs for the deconvolution algorithm with and without the priors was smaller than that in the previous experiment. In this simulation, the observed signal was contaminated with less noise. Even without resort to the priors, hence, the deconvolution algorithm could easily adapt the filter by using the observed data which contained more clear information on the convolutive channel. However, note that convolutive channels in real-world situation are not fixed as in this experiment but time-varying and the noise comes from very various sources such as measurement error and distributed noise.

Strengthening sparse priors excessively by increasing the step-size of Eq. (15) might accelerate the convergence speed in the early stage but disturb the frequency domain update algorithm of Eq. (3). Therefore, the convergence speed might slow down. In order to avoid the disturbance, we need to choose a moderate step-size, and using time-decaying step-sizes can be an appropriate strategy to give fast convergence in the beginning part and not to disturb the frequency domain update algorithm in the ending part.

5 Conclusion

In this paper, we imposed sparse priors on acoustic deconvolution filters for blind deconvolution to remove reverberation from speech signal. In order to include the sparse priors in the deconvolution algorithm, we maximized the posterior probability density of convolved signal with respect to the filters. The resulting algorithm needed negligible additional computation. Simulations indicated that sparseness imposed on the filters could provide useful information to accelerate convergence speed of the filters and provide better performance comparing with completely blind prior information.

Acknowledgments

This work was supported by the Brain Neuroinformatics Research Program sponsored by Korean Ministry of Science and Technology.

References

1. Haykin, S., ed.: Blind Deconvolution. Prentice-Hall, Englewood Cliffs, NJ (1994)
2. Ding, Z., Li, Y., eds.: Blind Equalization and Identification. Marcel Dekker, New York, NY (2001)
3. Lambert, R.H.: Multichannel Blind Deconvolution: FIR Matrix Algebra and Separation of Multipath Mixtures. PhD thesis, University of Southern California, Los Angeles (1996)
4. Hyvärinen, A., Karthikesh, R.: Imposing sparsity on the mixing matrix in independent component analysis. Neurocomputing **49**(1-4) (2002) 151–162
5. Amari, S.: Natural gradient works efficiently in learning. Neural Computation **10**(2) (1998) 251–276
6. Cardoso, J.F., Laheld, B.: Equivariant adaptive source separation. IEEE Transactions on Signal Processing **44**(12) (1996) 3017–3030
7. Amari, S., Douglas, S.C., Cichocki, A., Yang, H.H.: Novel on-line algorithms for blind deconvolution using natural gradient approach. In: IFAC Symposium on System Identification, Kitakyushu, Japan (1997) 1057–1062
8. Sklar, B., ed.: Digital Communications (Fundamentals and Applications). Prentice-Hall, Englewood Cliffs, NJ (1988)
9. DARPA TIMIT acoustic-phonetic continuous speech corpus cd-rom documentation. NIST Speech Disc 1-1.1 (1993)