BLIND DECONVOLUTION USING BAYESIAN METHODS WITH APPLICATION TO THE DEREVERBERATION OF SPEECH

*Michael J. Daly, James P. Reilly**

McMaster University
Department of Electrical and Computer Engineering
Hamilton, ON, Canada
Email: mdaly@grads.ece.mcmaster.ca, reillyj@mcmaster.ca

ABSTRACT

A blind deconvolution algorithm is presented to address the problem of the dereverberation of speech. A Bayesian algorithm is developed for estimating the source, and the problem of ill-conditioning due to long tails of an acoustic impulse response (AIR) is avoided by marginalizing out the unknown channel parameters. The initial samples of the MAP estimate are determined using a stochastic MCMC technique, and these estimates are then used in a sequential procedure for estimating the remaining of the signal. A filterbank implementation is used to reduce the large deconvolution problem into several smaller independent problems. Simulation results are presented to demonstrate the performance of the algorithm applied to the dereverberation of speech.

1. INTRODUCTION

Blind deconvolution is a specific problem in the field of blind signal processing. Blind techniques attempt to determine either the unknown input signals or unknown channels from the observed signals [1],[4],[6]. Blind deconvolution deals with recovering a single input signal from multiple output signals, known as a SIMO system.

Reverberation of speech can cause significant perceptual impairment to the quality of speech received at a microphone (i.e. in a hearing aid or hands-free telephone). Previous research in this application is presented in [3]. Typical AIR can be in the order of 250ms, or 2000 samples at a sampling rate of 8kHz. The large computational cost of dealing with such a long channel is addressed in this algorithm by decomposing the problem into smaller independent blind deconvolution problems using a filterbank structure.

Another property of the AIR is that the coefficients decay smoothly towards zero ("tailed"), making the blind channel identification problem ill-conditioned. The approach taken here is to directly estimate the source itself using a Bayesian MAP algorithm that marginalizes out the unknown channel parameters. This approach offers a more computationally stable method for recovery of the source signal.

2. BAYESIAN ESTIMATION FORMULATION

2.1. Problem Model and Assumptions

The model for the jth output of a single input, J-sensor output system is

$$y_j[n] = \mathbf{s}[n] * \mathbf{h}_j + v_j[n] \quad (1)$$

For a block of Q discrete observation samples, the SIMO system written in matrix format is

$$\mathbf{y}[n] = \mathbf{T}(\mathbf{s}[n]) \cdot \mathbf{h} + \mathbf{v}[n] \quad (2)$$

The following data structures and statistical assumptions are used in the Bayesian algorithm

- $\mathbf{y}[n] = [\mathbf{y}_1^{\mathrm{T}}[n], \ldots, \mathbf{y}_J^{\mathrm{T}}[n]]^{\mathrm{T}}$ is the sensor observation vector, and $\mathbf{y}_j[n] = [y_j[n-Q+1], \ldots, y_j[n]]^{\mathrm{T}} \in \mathbb{R}^Q$, $j = 1, 2, ..., J$
- $\mathbf{v}[n] \in \mathbb{R}^{JQ}$ is the noise vector, structured analogous to $\mathbf{y}[n]$, with $v_j[n]$ i.i.d. Gaussian: $p(\mathbf{v}[n]) \sim \mathcal{N}(\mathbf{0}, \sigma_v^2 \mathbf{I}_{JQ})$
- $\mathbf{s}[n] = [s[n-Q-L+2], \ldots, s[n]]^{\mathrm{T}} \in \mathbb{R}^{(Q+L-1)}$ is the source vector with $s[n]$ i.i.d. Gaussian: $p(\mathbf{s}[n]) \sim \mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbf{I}_{Q+L-1})$. For a speech source, the simplification of using a Gaussian model allows for analytical tractability, at the expense of a degradation in performance.
- $\mathbf{h} = [\mathbf{h}_1^{\mathrm{T}}, \ldots, \mathbf{h}_J^{\mathrm{T}}]^{\mathrm{T}}$ is the FIR channel impulse response vector, and $\mathbf{h}_j = [h_j[1], \ldots, h_j[L]]^{\mathrm{T}}]$, with $\mathbf{h}$ Gaussian distributed: $p(\mathbf{h}) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_h)$ where $\mathbf{\Sigma}_h = \mathbf{I}_J \otimes diag([\sigma^2_{h[1]}, \ldots, \sigma^2_{h[L]}])$, and the variances $\sigma^2_{h[n]}$ decay smoothly towards zero. $\mathbf{I}_J$ denotes the $J \times J$ identity matrix, and $\otimes$ the Kronecker product.

*This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC).

- $\mathbf{T}(\mathbf{s}[n]) \in \mathbb{R}^{(JQ\times JL)}$ is constructed to convolve $\mathbf{s}[n]$ with the J FIR channel vectors $\mathbf{h}_j$ of length L.
- It is assumed the channel length L is known, and that the channel does not change significantly over the observed samples ("quasi-static"). All the variances $(\sigma_s^2, \sigma_v^2, \mathbf{\Sigma}_h)$ are assumed known.

2.2. Development of the Posterior Distribution

Using Bayes theorem, the posterior distribution function for the independent unknown parameters $\mathbf{s}[n]$ and $\mathbf{h}$ can be expanded within a normalizing constant as

$$p(\mathbf{s}[n], \mathbf{h}|\mathbf{y}[n]) \propto p(\mathbf{y}[n]|\mathbf{s}[n], \mathbf{h})p(\mathbf{s}[n])p(\mathbf{h}) \qquad (3)$$

In addition to the Gaussian prior distributions $p(\mathbf{s}[n])$ and $p(\mathbf{h})$ given above, the likelihood function is Gaussian distributed: $p(\mathbf{y}[n]|\mathbf{s}[n], \mathbf{h}) \sim \mathcal{N}(\mathbf{T}(\mathbf{s}[n]) \cdot \mathbf{h}, \sigma_v^2 \mathbf{I}_{JQ})$ as seen from (2). Our primary goal is to recover an estimate of the unknown source $\mathbf{s}[n]$, and therefore $\mathbf{h}$ can be treated as a nuisance parameter. The channel impulse response $\mathbf{h}$ can be marginalized out of the posterior distribution resulting in a function of only the source

$$\begin{aligned} p(\mathbf{s}|\mathbf{y}) \quad \propto \quad & \frac{|\mathbf{\Sigma}|^{1/2}}{(2\pi\sigma_v^2)^{\frac{JQ}{2}}|\mathbf{\Sigma}_h|^{1/2}} \\ & \times \exp\left[\frac{1}{2}(\mathbf{m}_h^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{m}_h - \frac{1}{\sigma_v^2}\mathbf{y}^{\mathrm{T}}\mathbf{y})\right] \\ & \times \frac{1}{(2\pi\sigma_s^2)^{\frac{(L+Q-1)}{2}}} \exp\left[-\frac{1}{2\sigma_s^2}\mathbf{s}^{\mathrm{T}}\mathbf{s}\right] \end{aligned} \qquad (4)$$

where

$$\mathbf{\Sigma}^{-1} = \frac{1}{\sigma_v^2}\mathbf{T}^{\mathrm{T}}\mathbf{T} + \mathbf{\Sigma}_h^{-1} \qquad (5)$$

$$\mathbf{m}_h = \frac{1}{\sigma_v^2}\mathbf{\Sigma}\mathbf{T}^{\mathrm{T}}\mathbf{y} \qquad (6)$$

and explicit dependence on time is ignored for simplicity.

3. OPTIMIZATION ALGORITHM FOR MAP ESTIMATION

To find the Maximum a Posterior (MAP) estimate of the source signal, the marginalized posterior distribution is maximized in terms of $\mathbf{s}$ in two stages: the initialization stage and the sequential stage.

3.1. Initialization Stage

The first stage computes estimates $\hat{s}[1], \hat{s}[2], \ldots, \hat{s}[Q_{init}]$ of the initial Q_{init} samples of the source. The algorithm assumes data has been collected from the start of the convolution of $\mathbf{s}[n]$ and $\mathbf{h}$, meaning $s[n-L+1], s[n-L], \ldots, s[n-1]$ all equal zero for $n = 1$. To find the initial estimates, the non-convex Q_{init}-dimensional posterior distribution in (4) is maximized using a stochastic Markov Chain Monte Carlo (MCMC) approach [2]. A Metropolis-Hastings (MH) technique is selected using a proposal distribution

$$d(\mathbf{s}^*) = \frac{1}{(2\pi\sigma_s^2)^{\frac{(L_{init}+Q_{init}-1)}{2}}} \exp\left[-\frac{1}{2\sigma_s^2}\mathbf{s}^{*\mathrm{T}}\mathbf{s}^*\right] \qquad (7)$$

for a candidate $\mathbf{s}^*$, assumed to be independent of the current state of the chain $\mathbf{s}^{(i)}$ at the ith iteration.

A MAP estimate of the initial source samples can then be obtained from the samples of the Markov Chain which, after a sufficient burn-in period, are distributed according to the desired posterior distribution. For the simulations, the average over a block of N MCMC samples after N_b burn-in samples is used for the estimate of each source sample

$$\hat{s}[n] = \frac{1}{N}\sum_{i=N_b+1}^{N_b+N} s^{(i)}, n = 1, 2, \ldots, Q_{init} \qquad (8)$$

It is important to note that it is not required that $Q_{init} \geq L$ since the first Q_{init} output samples only depend on the first $L_{init} = Q_{init}$ channel coefficients.

3.2. Sequential Stage

The sequential stage computes an estimate $\hat{s}[n+1]$ based on the estimates $\hat{s}[1], \ldots, \hat{s}[n]$ available at that time. The maximization of the marginalized posterior distribution is reduced to a one-dimensional problem, which can be easily optimized, by substituting previous estimates for past values. For example, given initial estimates $\hat{s}[1], \hat{s}[2]$, the structure for estimating $\mathbf{s}[3]$ for a $J = 2$ output system is

$$\mathbf{y}[3] = \mathbf{I}_2 \otimes \begin{pmatrix} \hat{s}[1] & 0 & 0 \\ \hat{s}[2] & \hat{s}[1] & 0 \\ s[3] & \hat{s}[2] & \hat{s}[1] \end{pmatrix} \mathbf{h} + \mathbf{v}[3] \qquad (9)$$

where now the only unknown is $s[3]$.

Two restrictions apply to the number of observations Q_{seq} used for the sequential procedure. First, clearly $Q_{rec} \leq n+1$ since the first sample is available at $n = 1$. The second condition is that to account for all the channel coefficients, $Q_{sec} \geq L$. In the case where $Q_{init} < L$, applying the two conditions requires that for $n+1 = Q_{init}+1, \ldots, L$, the value of $Q_{sec} = n+1$ since the observations only depend on the first $n+1$ channel coefficients.

4. COMPLEX SUBBANDING IMPLEMENTATION

For large L, the computational cost of the MAP estimation algorithm can become intractable. A filterbank implementation using the complex subband decomposition [5] is used

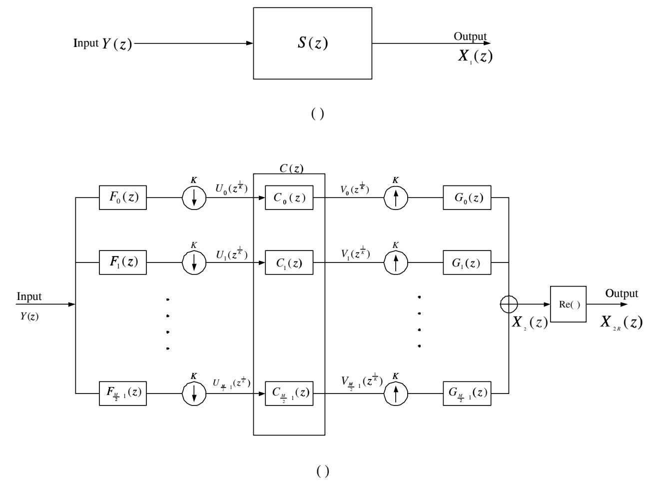

Fig. 1. (a) An arbitrary LTI system $S(z)$. (b) A subband approximation of the system $S(z)$.

to break the large blind deconvolution problem into smaller independent problems to reduce the algorithm complexity.

The complex subband decomposition describes how to decompose an arbitrary FIR system $S(z)$ of order L into M subband components of order O(L/K), where K is the downsampling/upsampling rate, using an M-channel filterbank as illustrated in Figure 1.

It is shown in [5] that the time domain versions $u_{y_j,m}[n]$ of $U_m(z^{\frac{1}{K}})$ in Figure 1(b), representing the signal $\mathbf{y}_j[n]$ in the mth subband can be written in the form

$$u_{y_j,m} \simeq f_m[n]_{\downarrow K} * c_{s,m}[n] * c_{h_j,m}[n] + f_m[n]_{\downarrow K} * c_{v_j,m}[n] \quad (10)$$

where $f_m[n]_{\downarrow K}$ is the downsampled impulse response of the mth analysis filter, $c_{s,m}[n]$ and $c_{h_j,m}[n]$ represent the signals $s[n]$ and $h_j[n]$ respectively in the mth subband, and $c_{v_j,m}[n]$ is the component of the noise in the mth subband.

Except for the presence of the $f_m[n]_{\downarrow K}$ terms, equation (10) looks like a convolution inside the subbands of the respective signals $s[n]$ and $h_j[n]$. The Bayesian blind deconvolution method described in this paper can therefore be applied to the signals $u_{y_j,m}[n]$, $j = 1, 2, \ldots, J$ in the subbands to yield estimates of the signals $f_m[n]_{\downarrow K} * c_{s,m}[n]$. These signals are then upsampled and passed through the synthesis filters as shown in Figure 1 to yield the recovered source signal $\hat{s}[n]$ at the filterbank output. The advantage of this approach is that the effective length of the channel is reduced by a factor of approximately K. The disadvantage is that the recovered signal in each subband is subjected to an unknown complex scale ambiguity.

5. SIMULATION RESULTS

Simulation results are now presented to support the developed algorithm for a $J = 2$ output system. 4000 samples of real speech sampled at 8 kHz normalized to $\sigma_s^2 = 1$ was used as the source. Two FIR channels with $L = 45$ and $W = 0.15$ were generated using the exponentially decaying channel covariance matrix

$$\mathbf{\Sigma}_h = \mathbf{I}_2 \otimes diag(e^{-\frac{0}{WL}}, \ldots, e^{-\frac{L-1}{WL}}) \quad (11)$$

The white noise was scaled to an SNR of 30dB, defined as

$$SNR(dB) = 10 \log_{10} \left(\frac{\|\mathbf{T}(\mathbf{s}[n]) \cdot \mathbf{h}\|_2^2}{\sigma_v^2} \right) \quad (12)$$

A 32-channel oversampled generalized discrete Fourier transform (GDFT) filterbank was used with a downsampling rate of $K = 20$. The complex analysis/synthesis filters using $Lf = 256$ filter coefficients were designed following the spectral factorization method specified in [5].

The first $Q_{init} = 2$ samples in each subband were estimated by running the MCMC MH algorithm for 20000 burn-in samples, and then averaging the following 10000 MCMC samples to form the estimates $\hat{u}_{s,m}[1]$ and $\hat{u}_{s,m}[2]$. For the given parameters, the length of the subband channel is $L_{c_h} = 3$. Therefore, the sequential procedure is first applied once using $Q_{seq} = L_{c_h} = 3$ to produce $\hat{u}_{s,m}[3]$. Subband 5 is arbitrarily selected to demonstrate the performance of the estimation of the initial samples in Figure 2.

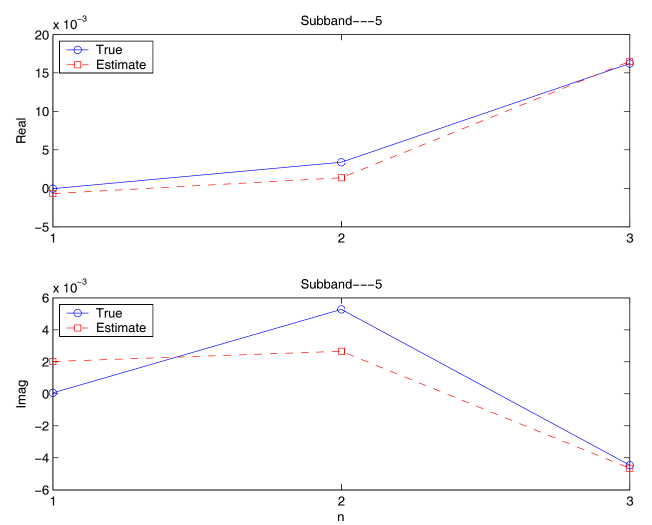

Fig. 2. Estimation of $u_{s,5}[1] \sim u_{s,5}[3]$

Now that L_{c_h} estimates have been computed, the sequential optimization procedure continues with $Q_{sec} = L_{c_h} + 1 = 4$ to estimate the remaining part of the subband signal.

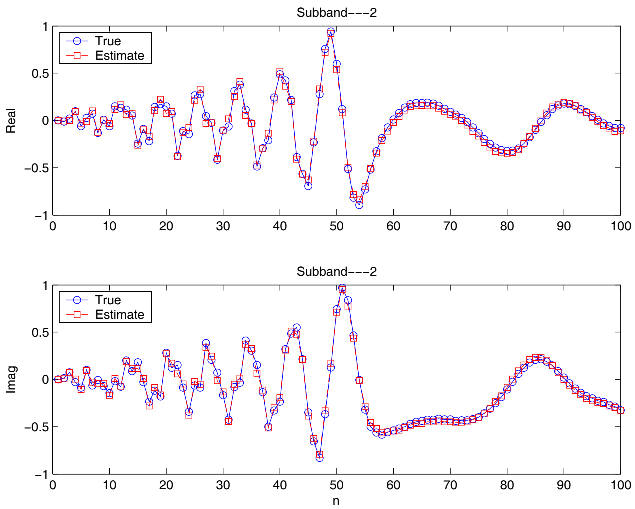

Fig. 3. Estimation of $u_{s,2}[1] \sim u_{s,2}[100]$.

Samples $n = 1$ to $n = 100$ of the signals in subband 2 are shown in Figure 3.

The synthesized speech estimate from $n = 1000$ to $n = 1200$ is compared with the true speech signal in Figure 4. The mean square error (MSE) was -14.22 dB, computed

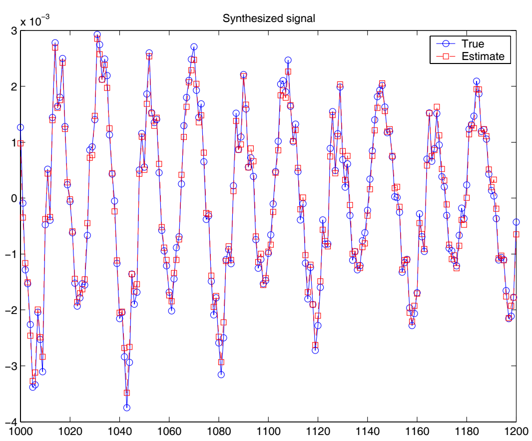

Fig. 4. Synthesized estimate of $s[1000] \sim s[1200]$

over the entire signal estimate using

$$MSE(dB) = 10\log_{10}\left(\frac{\|\mathbf{s}[n] - \hat{\mathbf{s}}[n]\|_2^2}{\|\mathbf{s}[n]\|_2^2}\right) \quad (13)$$

In order to evaluate the above MSE, it is necessary to resolve the inherent scale ambiguity in the subband signals. Based on a least squares formulation, the unknown complex scale factors a_m are estimated as follows

$$a_m = \frac{\hat{u}_{s,m}^{\mathrm{H}} u_{s,m}}{\hat{u}_{s,m}^{\mathrm{H}} \hat{u}_{s,m}} \quad (14)$$

The signal $\hat{u}_{s,m}$ in each subband is then multiplied by a_m, before upsampling and synthesis filtering. The filterbank output $\hat{s}[n]$ is then substituted into (13) for evaluation of the MSE.

6. CONCLUSIONS

A Bayesian approach to blind signal recovery has been presented. Unlike previous approaches which first estimate the channel and then form an inverse (processes which can be ill-conditioned in this application), the proposed approach treats the channel as a nuisance parameter and estimates the source directly. An efficient time-recursive procedure for the estimation of the source was proposed. Simulation results have shown the effectiveness of the method.

7. ACKNOWLEDGMENTS

The authors acknowledge helpful suggestions and insight from Dr. J.R. Larocque, currently with Dataradio Inc. in Montreal, Canada.

8. REFERENCES

[1] R. Chen and T.H. Li, "Blind restoration of linearly degraded discrete signals by Gibbs sampling", *IEEE Transactions on Signal Processing*, vol. 43, no.10, pp. 2410-2413, Oct. 1995.

[2] W. Gilks, S. Richardson, D. Spiegelhalterm, "Markov Chain Monte Carlo in Practice", Chapman and Hall, New York, 1998.

[3] J.R. Hopgood and P.J.W. Rayner, "Blind single channel deconvolution using nonstationary signal processing", *IEEE Transactions on Speech and Audio Processing*, vol.11, no.5, 2003.

[4] Y. Hua, S. An, Y. Xiang, "Blind Identification and Equalization of FIR MIMO Channels by BIDS", *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Salt Lake City,UT, May 2001.

[5] J.P. Reilly, M. Wilbur, M. Seibert, N. Ahmadvand, "The complex subband decomposition and its application to the decimation of large adaptive filtering problems", *IEEE Transactions on Signal Processing*, vol.50, no.11, 2002.

[6] L.-Q. Zhang, A. Cichocki, S. Amari, "Kalman filter and state-space approach to blind deconvolution", *Neural Networks for Signal Processing X, 2000, Proc. IEEE Signal Processing Society Workshop*, vol. 1, pp. 425-434, Dec. 2000.