MARKOV CHAIN MONTE CARLO METHODS FOR SPEECH ENHANCEMENT

Jaco Vermaak and Mahesan Niranjan

Cambridge University Engineering Department,
Trumpington Street, Cambridge CB2 1PZ, U.K.
e-mail: jv211@eng.cam.ac.uk

ABSTRACT

This paper investigates a Bayesian approach to the enhancement of speech signals corrupted by additive white Gaussian noise. Parametric models for the speech and noise processes are constructed, leading to a posterior distribution for the model parameters and uncorrupted speech samples given the observed noisy speech samples. Being analytically intractable, inferences concerning these variables are performed using Markov chain Monte Carlo (MCMC) methods. The efficiency of the sampling scheme within this framework is further improved by employing state-space techniques based on the Kalman filter.

1. SPEECH ENHANCEMENT

The model considered for the enhancement of speech corrupted by additive white Gaussian noise is the well-known AR process. It has been extensively used for this purpose [8], and has some justification from a physical point of view in that it is a simplification of the system resulting when the vocal tract is modelled as a lossless acoustic tube [4]. Its main utility, however, follows from the analytic results it facilitates. A p-th order AR process is described by the equation:

$$x_t = \sum_{i=1}^{p} \psi_i x_{t-i} + e_t, \tag{1}$$

with $\psi_i,\ i = 1, \ldots, p$ the AR coefficients, and $e_t \sim \mathcal{N}(0, \sigma_e^2)$ an i.i.d. excitation sequence. This equation can be written in matrix-vector form for a block of T samples as: $\mathbf{e} = \mathbf{\Psi x} = \mathbf{x}^{(1)} - \mathbf{G}\boldsymbol{\psi}$, where $\mathbf{e} = [e_{p+1}, \ldots, e_T]'$, $\mathbf{x} = [x_1, \ldots, x_T]'$, $\mathbf{x}^{(1)}$ is $\mathbf{x}$ with the first p elements removed, $\boldsymbol{\psi} = [\psi_1, \ldots, \psi_p]'$, and $\mathbf{\Psi}$ and $\mathbf{G}$ are constructed so as to make the expression hold. The observations are given by:

$$z_t = x_t + v_t, \tag{2}$$

where $\{v_t\}$ is a white Gaussian noise process, i.e. $v_t \sim \mathcal{N}(0, \sigma_v^2)$, with $\mathrm{cov}(v_s, v_t) = 0$ for $s \neq t$.

The estimation of the uncorrupted speech samples $\{x_t\}$ and the model parameters can be cast in a statistical framework. Defining $\boldsymbol{\theta} = [\psi_i,\ i = 1, \ldots, p, \sigma_e^2, \sigma_v^2]'$ and interpreting this together with $\mathbf{x}$ as random vectors, Bayes' rule leads to a joint posterior distribution for the uncorrupted samples and the model parameters (assumed stationary over T samples) of the form:

$$p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{z}) = \frac{p(\mathbf{z}|\mathbf{x}, \boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{z})}, \tag{3}$$

with $\mathbf{z} = [z_1, \ldots, z_T]'$ the vector of observations. In (3), $p(\mathbf{x}|\boldsymbol{\theta})$ is the *likelihood* of the parametric model generating the uncorrupted speech $\mathbf{x}$, when the model parameters $\boldsymbol{\theta}$ are known. For the AR(p) model the likelihood is Gaussian and can be approximated as [2]:

$$p(\mathbf{x}|\boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}; \mathbf{0}, \sigma_e^2(\mathbf{\Psi}'\mathbf{\Psi})^{-1}). \tag{4}$$

On the other hand, $p(\mathbf{z}|\mathbf{x}, \boldsymbol{\theta})$ is the *total likelihood* of the observed data $\mathbf{z}$ when both the model parameters $\boldsymbol{\theta}$ and the uncorrupted speech sequence $\mathbf{x}$ are known. It, thus, depends solely on the characteristics of the noise process, and for white Gaussian noise is given by:

$$p(\mathbf{z}|\mathbf{x}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{z}; \mathbf{x}, \sigma_v^2\mathbf{I}). \tag{5}$$

The distribution $p(\boldsymbol{\theta})$ is the *prior density* expressing initial beliefs about the unknown parameters $\boldsymbol{\theta}$ before the data is seen. To facilitate analytic tractability and retain flexibility, the following conjugate prior is employed:

$$p(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\psi}; \bar{\boldsymbol{\mu}}_{\boldsymbol{\psi}}, \bar{\mathbf{\Sigma}}_{\boldsymbol{\psi}})\mathrm{IG}(\sigma_e^2; \alpha_e, \beta_e)\mathrm{IG}(\sigma_v^2; \alpha_v, \beta_v), \tag{6}$$

where IG$(\cdot)$ denotes the inverted-gamma distribution. Note that the parameters are assumed to be *a priori* independent. Finally, $p(\mathbf{z})$ (sometimes called the evidence for the model) is a constant normalising factor, independent of the uncorrupted speech sequence $\mathbf{x}$ and model parameters $\boldsymbol{\theta}$, and can usually be discarded in applications where issues other than model comparison are at stake.

One commonly employed estimate for the uncorrupted speech and model parameters is the *minimum mean square error* (MMSE) estimate, which can be expressed as:

$$(\mathbf{x}, \boldsymbol{\theta})_{\mathrm{MMSE}} = \int_{\mathbf{x}} \int_{\boldsymbol{\theta}} \mathbf{x}\boldsymbol{\theta} p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{z}) d\mathbf{x} d\boldsymbol{\theta}. \tag{7}$$

If samples $(\mathbf{x}, \boldsymbol{\theta})_i, i = 1, \ldots, N$ can be drawn from the joint posterior $p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{z})$, *Monte Carlo integration* can be used to approximate this estimate as:

$$(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}})_{\text{MMSE}} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}, \boldsymbol{\theta})_i. \tag{8}$$

In general, drawing samples from $p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{z})$ is not straightforward, due to the fact that it is usually high-dimensional and non-standard. One way of doing this is by setting up a Markov chain having $p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{z})$ as its stationary distribution. The combined method is then known as *Markov chain Monte Carlo* (MCMC). A special case of these techniques, the *Gibbs sampler*, will be considered next.

2. THE GIBBS SAMPLER

The Gibbs sampler [7, 6] sets up an irreducible aperiodic Markov [10] chain that converges to the target distribution $p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{z})$ from which samples are desired. These samples are obtained by repeatedly replacing each component of the random vectors $\mathbf{x}$ and $\boldsymbol{\theta}$ with a value drawn from its distribution conditioned on the current values of all the other components. This procedure for generating $(\mathbf{x}, \boldsymbol{\theta})_{i+1}$ from $(\mathbf{x}, \boldsymbol{\theta})_i$ can be expressed more clearly as:

$$\begin{aligned}
\boldsymbol{\psi}_{(i+1)} &\sim p(\boldsymbol{\psi}|\sigma^2_{e(i)}, \mathbf{x}_{(i)}, \mathbf{z}) \\
\sigma^2_{e(i+1)} &\sim p(\sigma^2_e|\boldsymbol{\psi}_{(i+1)}, \mathbf{x}_{(i)}, \mathbf{z}) \\
\sigma^2_{v(i+1)} &\sim p(\sigma^2_v|\mathbf{x}_{(i)}, \mathbf{z}) \\
\mathbf{x}_{(i+1)} &\sim p(\mathbf{x}|\boldsymbol{\psi}_{(i+1)}, \sigma^2_{e(i+1)}, \sigma^2_{v(i+1)}, \mathbf{z}),
\end{aligned} \tag{9}$$

where the posterior conditionals are given by:

$$\begin{aligned}
p(\boldsymbol{\psi}|\sigma^2_e, \mathbf{x}, \mathbf{z}) &= \mathcal{N}(\boldsymbol{\psi}; \boldsymbol{\mu}_{\boldsymbol{\psi}}, \boldsymbol{\Sigma}_{\boldsymbol{\psi}}) \\
p(\sigma^2_e|\boldsymbol{\psi}, \mathbf{x}, \mathbf{z}) &= \text{IG}(\sigma^2_e; \alpha_e + (T-p)/2, \beta_e + \mathbf{e}'\mathbf{e}/2) \\
p(\sigma^2_v|\mathbf{x}, \mathbf{z}) &= \text{IG}(\sigma^2_v; \alpha_v + T/2, \beta_v + (\mathbf{z}-\mathbf{x})'(\mathbf{z}-\mathbf{x})/2) \\
p(\mathbf{x}|\boldsymbol{\psi}, \sigma^2_e, \sigma^2_v, \mathbf{x}, \mathbf{z}) &= \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}}),
\end{aligned} \tag{10}$$

with $\boldsymbol{\mu}_{\boldsymbol{\psi}} = \boldsymbol{\Sigma}_{\boldsymbol{\psi}}(\mathbf{G}'\mathbf{x}^{(1)}/\sigma^2_e + \bar{\boldsymbol{\Sigma}}^{-1}_{\boldsymbol{\psi}}\bar{\boldsymbol{\mu}}_{\boldsymbol{\psi}})$, $\boldsymbol{\Sigma}_{\boldsymbol{\psi}} = (\mathbf{G}'\mathbf{G}/\sigma^2_e + \bar{\boldsymbol{\Sigma}}^{-1}_{\boldsymbol{\psi}})^{-1}$, $\boldsymbol{\mu}_{\mathbf{x}} = \boldsymbol{\Sigma}_{\mathbf{x}}\mathbf{z}/\sigma^2_v$ and $\boldsymbol{\Sigma}_{\mathbf{x}} = (\boldsymbol{\Psi}'\boldsymbol{\Psi}/\sigma^2_e + \mathbf{I}/\sigma^2_v)^{-1}$.

All these conditionals are in known parametric forms and can be sampled from using standard techniques. Note, however, that the sampling step for the reconstructed speech samples requires the inversion of a $T \times T$ matrix, which is of $O(T^3)$ complexity. This is computationally very expensive, and makes the direct sampling of these values infeasible for all but unrealistically small values of T. The subsequent section describes a state-space method to sample from the full conditional distribution $p(\mathbf{x}|\psi, \sigma^2_e, \sigma^2_v, \mathbf{z})$, resulting in a computational complexity that is linear in T.

3. STATE-SPACE METHODS

The speech enhancement model can be represented in state-space form:

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{G}_t\mathbf{u}_t \\
\boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t\boldsymbol{\alpha}_t + \mathbf{H}_t\mathbf{u}_t
\end{aligned} \tag{11}$$

by defining the state, observation and disturbance vectors respectively as: $\boldsymbol{\alpha}_t = [x_t, x_{t-1}, \ldots, x_{t-p+1}]' \in \mathbb{R}^p$, $\mathbf{y}_t = [z_t] \in \mathbb{R}$ and $\mathbf{u}_t = [e_{t+1}, v_t]' \in \mathbb{R}^2$, and the system matrices as:

$$\mathbf{Z}_t = \mathbf{Z} = \begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix} \in \mathbb{R}^{1\times p}, \tag{12}$$

$$\mathbf{G}_t = \mathbf{G} = \begin{bmatrix} 0 & \sigma_v \end{bmatrix} \in \mathbb{R}^{1\times 2}, \tag{13}$$

$$\mathbf{T}_t = \mathbf{T} = \begin{bmatrix} \psi_1 & \psi_2 & \ldots & \psi_{p-1} & \psi_p \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix} \in \mathbb{R}^{p\times p} \tag{14}$$

and

$$\mathbf{H}_t = \mathbf{H} = \begin{bmatrix} \sigma_e & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{p\times 2}. \tag{15}$$

State-based sampling techniques [3, 5, 9] sample directly from the posterior conditional $p(\boldsymbol{\alpha}|\mathbf{y}^T, \boldsymbol{\theta})$, where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}'_1, \ldots, \boldsymbol{\alpha}'_T]'$ is the stacked state vector and $\mathbf{y}^t = [\mathbf{y}'_1, \ldots, \mathbf{y}'_t]'$. Using the first-order Markov nature of the state process and the causality of the system, this conditional can be decomposed according to the probability chain rule as:

$$p(\boldsymbol{\alpha}|\mathbf{y}^T, \boldsymbol{\theta}) = p(\boldsymbol{\alpha}_T|\mathbf{y}^T, \boldsymbol{\theta})\prod_{t=1}^{T-1} p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t+1}, \mathbf{y}^t, \boldsymbol{\theta}) \tag{16}$$

A draw from $p(\boldsymbol{\alpha}|\mathbf{y}^T, \boldsymbol{\theta})$ can thus be constructed by recursing backwards in time for $t = T, T-1, \ldots, 1$, provided that sub-draws from the densities on the RHS of (16) are practical. From Bayes' rule $p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t+1}, \mathbf{y}^t, \boldsymbol{\theta}) \propto p(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t, \boldsymbol{\theta})p(\boldsymbol{\alpha}_t|\mathbf{y}^t, \boldsymbol{\theta})$. Following from the state transition equation and the Gaussian nature of the disturbances $p(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{T}_t\boldsymbol{\alpha}_t, \mathbf{H}_t\mathbf{H}'_t)$. Similarly, $p(\boldsymbol{\alpha}_t|\mathbf{y}^t, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{a}_t, \mathbf{P}_t)$, with $\mathbf{a}_t$ and $\mathbf{P}_t$ given by the Kalman filter equations [1]. Consequently, $p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t+1}, \mathbf{y}^t, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{a}_{t|t+1}, \mathbf{P}_{t|t+1})$ is Gaussian, with mean vector and covariance matrix given by:

$$\mathbf{a}_{t|t+1} = \mathbf{A}_t(\mathbf{a}_t + \mathbf{B}_t\boldsymbol{\alpha}_{t+1}), \qquad \mathbf{P}_{t|t+1} = \mathbf{A}_t\mathbf{P}_t, \tag{17}$$

where $\mathbf{A}_t = (\mathbf{I} + \mathbf{B}_t\mathbf{T}_t)^{-1}$ and $\mathbf{B}_t = \mathbf{P}_t\mathbf{T}'_t(\mathbf{H}_t\mathbf{H}'_t)^{-1}$. Thus, all the densities in (16) are Gaussian, and sampling for $\boldsymbol{\alpha}$ can proceed straightforwardly.

3.1. Degeneracies

Generally, the components of the stacked state vector $\boldsymbol{\alpha}$ are not independent, and many identities link the comprising state variables $\boldsymbol{\alpha}_t$, $t = 1, \ldots, T$. These identities are problem dependent and a direct consequence of forcing the model into state-space form. They lead to degeneracies, i.e. situations where the effective number of degrees of freedom are less than the dimensionality of the problem, and consequently must be kept track of in the recursive construction of the draw.

For the speech enhancement model the definition of the state vector is such that the first $p-1$ components of $\boldsymbol{\alpha}_t$ overlap with the last $p-1$ components of $\boldsymbol{\alpha}_{t+1}$. Thus, in the recursive construction of the draw backwards in time, only the distribution for the final state $\boldsymbol{\alpha}_T$ is not degenerate. For all the other states $\boldsymbol{\alpha}_t$, $t = T-1, \ldots, 1$, the distributions are degenerate, having only one degree of freedom, with the first $p-1$ components fixed by the draw from the distribution of the successive state $\boldsymbol{\alpha}_{t+1}$. For these states, the draw for the unknown component is made from the appropriate conditional.

More precisely, if $v^{(i)}$ is the i-th component of a vector $\mathbf{v}$, and $\mathbf{v}^{(-i)}$ is the vector $\mathbf{v}$ with the i-th component removed, the state vector can be written as: $\boldsymbol{\alpha}_t = [(\boldsymbol{\alpha}_t^{(-p)})', \alpha_t^{(p)}]'$. Because of the jointly Gaussian nature of the components of the state vector $\boldsymbol{\alpha}_t$, the univariate conditional for $\alpha_t^{(p)}$ is also Gaussian. Writing the mean and covariance matrix of $p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t+1}, \mathbf{y}^t, \boldsymbol{\theta})$ as:

$$\mathbf{a}_{t|t+1} = \begin{bmatrix} \mathbf{a}_{t|t+1}^{(-p)} \\ a_{t|t+1}^{(p)} \end{bmatrix}, \qquad \mathbf{P}_{t|t+1} = \begin{bmatrix} \mathbf{P}_{t|t+1}^{(-p,-p)} & \mathbf{P}_{t|t+1}^{(-p,p)} \\ \mathbf{P}_{t|t+1}^{(p,-p)} & P_{t|t+1}^{(p,p)} \end{bmatrix}, \tag{18}$$

the univariate Gaussian distribution for the posterior conditional can be obtained as:

$$p(\alpha_t^{(p)}|\boldsymbol{\alpha}_t^{(-p)}, \boldsymbol{\alpha}_{t+1}, \mathbf{y}^t, \boldsymbol{\theta}) = \mathcal{N}(\bar{a}_{t|t+1}^{(p)}, \bar{P}_{t|t+1}^{(p)}), \tag{19}$$

with

$$\bar{a}_{t|t+1}^{(p)} = a_{t|t+1}^{(p)} + \mathbf{P}_{t|t+1}^{(p,-p)}(\mathbf{P}_{t|t+1}^{(-p,-p)})^{-1}(\boldsymbol{\alpha}_t^{(-p)} - \mathbf{a}_{t|t+1}^{(-p)}) \tag{20}$$

and

$$\bar{P}_{t|t+1}^{(p)} = P_{t|t+1}^{(p,p)} - \mathbf{P}_{t|t+1}^{(p,-p)}(\mathbf{P}_{t|t+1}^{(-p,-p)})^{-1}\mathbf{P}_{t|t+1}^{(-p,p)}. \tag{21}$$

3.2. Sampling Strategy Summary

The process of sampling from $p(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta})$ using the state-based algorithm can be summarised as follows:

(a) Cast the model into state-space form.

(b) Initialise the Kalman filter, and run it forwards in time to obtain the means $\mathbf{a}_t$ and covariance matrices $\mathbf{P}_t$ for the distributions $p(\boldsymbol{\alpha}_t|\mathbf{y}^t, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{a}_t, \mathbf{P}_t)$, $t = 1, \ldots, T$.

(c) Sample the final state as: $\boldsymbol{\alpha}_T \sim p(\boldsymbol{\alpha}_T|\mathbf{y}^T, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{a}_T, \mathbf{P}_T)$.

(d) Recurse backwards in time for $t = T-1, \ldots, 1$ and sample the unknown components of $\boldsymbol{\alpha}_t$ from the corresponding degenerate distributions, fixing the known components according to the relations between the successive state vectors. For the speech enhancement model the steps involved can be summarised as follows:

   (i) Let $\boldsymbol{\alpha}_t^{(-p)} = [\alpha_t^{(1)}, \ldots, \alpha_t^{(p-1)}]' = [\alpha_{t+1}^{(2)}, \ldots, \alpha_{t+1}^{(p)}]'$.

   (ii) Sample the p-th state component of $\boldsymbol{\alpha}_t$ as: $\alpha_t^{(p)} \sim p(\alpha_t^{(p)}|\boldsymbol{\alpha}_t^{(-p)}, \boldsymbol{\alpha}_{t+1}, \mathbf{y}^t, \boldsymbol{\theta}) = \mathcal{N}(\bar{a}_{t|t+1}^{(p)}, \bar{P}_{t|t+1}^{(p)})$, with $\bar{a}_{t|t+1}^{(p)}$ and $\bar{P}_{t|t+1}$ given by (20) and (21), respectively.

(e) Extract the vector of reconstructed samples $\mathbf{x}$ from the stacked state vector $\boldsymbol{\alpha}$ by making use of the composition of the state vector. For the speech enhancement model it amounts to setting $\mathbf{x} = [\alpha_1^{(1)}, \ldots, \alpha_T^{(1)}]'$.

Sampling from $p(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta})$ in this way reduces the computational complexity to $O(p^3 T)$, which is linear in T.

4. RESULTS AND CONCLUSIONS

The method was evaluated on the utterance: "She had your dark suit in greasy wash water all year." from the Resource Management database, spoken by a white male speaker with a New England American accent. The clean speech was corrupted by an additive noise process, with variance changing quadratically over the utterance. The resulting signal was processed in blocks of 480 samples, corresponding to a time window of 30 msec., within which the speech and noise processes were assumed to be stationary. The blocks were overlapped by p samples, and a value of $p = 10$ was found to be adequate.

Initial estimates for the AR coefficients and excitation variance for all the blocks in the sequence were obtained by maximum likelihood estimation. Also, the parameters of the priors were chosen so as to result in these densities being non-informative. To increase the computational efficiency, the number of sampling iterations per block were reduced to a maximum of 30, with the first 20 taken as the burn-in period for the Gibbs sampler.

The enhancement results are depicted in Figure 1, whereas the ability of the algorithm to track the true noise variance profile is clearly illustrated in Figure 2. An overall SNR improvement of 4.68 dB was obtained. In general,

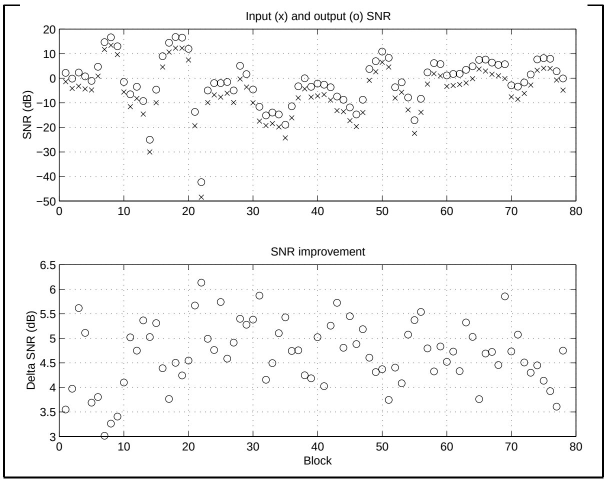

Figure 1: Input and output SNR, and SNR improvement profiles.

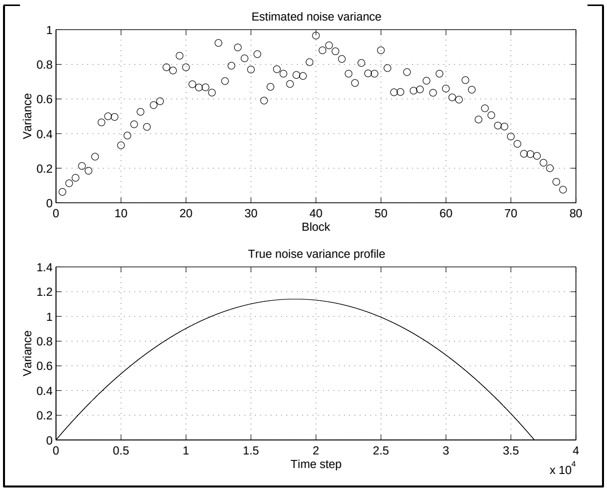

Figure 2: Estimated and true noise variance profiles.

the algorithm proved to yield satisfactory SNR improvements, even in environments where the background noise level was allowed to vary slowly over time. Furthermore, the enhancement performance was found to be relatively insensitive to the exact choices for the initial parameter estimates, prior distribution parameters, and even the AR modelling order. In most cases the Markov chain converged rapidly to a solution, so that useful MMSE estimates could be obtained after only a few iterations; a very important utility in the processing of long sequences.

[1] Brian D. O. Anderson and John B. Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, 1979.

[2] George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel. *Time Series Analysis, Forecasting and Control, Third Edition*. Prentice-Hall, Englewood Cliffs, 1994.

[3] C. K. Carter and R. Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

[4] John R. Deller, John G. Proakis, and John H. L. Hansen. *Discrete-Time Processing of Speech Signals*. Macmillan Publishing Company, Englewood Cliffs, 1993.

[5] Sylvia Fruhwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15(2):183–202, 1994.

[6] Alan E. Gelfand and Adrian F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, June 1990.

[7] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, November 1984.

[8] Jae S. Lim and Allan V. Oppenheim. All-pole modeling of degraded speech. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-26(3):197–210, June 1978.

[9] Neil Shephard. Partial non-gaussian state space. *Biometrika*, 81(1):115–131, 1994.

[10] Luke Tierney. Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, 22(4):1701–1762, 1994.