Fully Bayesian Analysis of Hidden Markov Models

Arnaud DOUCET†‡ and Patrick DUVAUT‡
†*LETI - CEA Technologies Avancées 91191 Gif sur Yvette FRANCE*
‡*ENSEA-ETIS Groupe Signal - 6, avenue du Ponceau 95014 Cergy Pontoise FRANCE*
Email: `douceta@ensea.fr` - `duvaut@ensea.fr`

ABSTRACT

In this paper, we present in an unified framework some applications of stochastic simulation techniques, the Markov chain Monte Carlo methods, to perform Bayesian inference for a very wide class of hidden Markov models. Efficient implementation of the Gibbs sampler based on finite dimensional optimal filters is described. An improved version of this algorithm is also presented. Two problems of great practical interest in signal processing are addressed: blind deconvolution of Bernoulli-Gauss processes and blind equalization of a channel. In simulations, we obtain very satisfactory results.

1 Introduction

Hidden Markov models have a very wide range of applications in statistical signal processing. In many cases, we consider linear Gaussian models or hidden Markov chains because they are well-adapted to practical problems encountered in digital communications, speech processing, biomedical signal processing and allow to perform analytical calculations. Nevertheless, in some applications, some hyperparameters of these models are unknown and must be estimated. Previous works have mainly focused on Maximum Likelihood estimation of these parameters[8]. In this paper, we assume to know a prior distribution on hyperparameters that allows to perform 'theoretically' Bayesian estimation of the hidden state process and its associated hyperparameters. As Bayesian estimation requires to evaluate high dimensional integrations, we present some efficient approximations based on Markov chain Monte Carlo techniques. Then, two applications are presented. The first one is the blind deconvolution of Bernoulli-Gauss processes and the second one the blind equalization of a channel.

2 Hidden Markov models

Let us consider a two-components random process $(X,Y)=\{x(k),y(k)\}_{k=1,\dots,N}$ where X is the hidden-state process and $x(k)$ takes its values from $\mathfrak{X}$. Y is observed and $y(k)$ takes its values from $\mathfrak{Y}$. We assume that X is a Markov process of transition probability density - dependent on a vector parameter θ - with respect to an appropriate σ-finite measure μ on $\mathfrak{X}$:

$$p(X_{1\to N}|\theta)=p(x(1)|\theta)\prod_{k=2}^{N}p(x(k)|x(k-1),\theta) \quad (1)$$

The observations are assumed conditionally independent:

$$p(Y_{1\to N}|X_{1\to N},\theta)=\prod_{k=1}^{N}p(y(k)|x(k),\theta) \quad (2)$$

where $Z_{i\to j}$ denotes $\{z(i),z(i+1),\dots,z(j)\}$. The filtering density $p(x(k)|Y_{1\to k},\theta)$ can be evaluated recursively following the standard prediction-updating formula. The joint smoothing density is equal to

$$p(X_{1\to N}|Y_{1\to N},\theta)=p(x(N)|y(N),\theta) \quad (3)$$
$$\times\prod_{k=1}^{N-1}p(x(k)|Y_{1\to k},x(k+1),\theta)$$

There are very few cases where it is possible to evaluate analytically this density. The two most well-known cases are hidden Markov chains where $\mathfrak{X}=\{a_1,\dots,a_n\}$, μ is the counting measure, and linear Gaussian state space models where $\mathfrak{X}=\mathbb{R}^n$, μ is the Lebesgue measure, the optimal filter being the Kalman filter.

As soon as θ is unknown, even in those very interesting cases, it is no more possible to estimate the optimal filter and thus the joint smoothing density. We assume here that θ is a random vector of prior density $p(\theta)$ and we want to estimate the joint smoothing density $p(X_{1\to N},\theta|Y_{1\to N})$ (or some of its features). Suboptimal analytical filters and smoothers, such as the extended Kalman filter, perform poorly in many cases of interest and it is necessary to develop more efficient approximations.

3 Bayesian estimation via stochastic simulation

In a Bayesian framework, the objective is to estimate the following a posteriori joint distribution $p(X_{1\to N},\theta|Y_{1\to N})$, given by Bayes's formula,

$$p(X_{1\to N},\theta|Y_{1\to N})=\frac{p(Y_{1\to N}|X_{1\to N},\theta)\,p(X_{1\to N},\theta)}{p(Y_{1\to N})} \quad (4)$$

and more generally some of its features such as marginals or expectations. Practically, it is impossible to perform these calculations because it requires in particular to evaluate high dimensional integrations. A Monte Carlo approach consists in estimating features of the posterior distribution using samples drawn from it. Here we describe an iterative stochastic simulation technique based on Markov chains to draw samples from the joint distribution.

3.1 Markov chain Monte Carlo techniques

Recently, Markov Chain Monte Carlo (MCMC) methods have been introduced in statistics[10]. These methods appear as powerful tools to perform Bayesian inference in high dimensional complex systems. It consists in constructing a Markov chain whose equilibrium distribution is the desired posterior distribution. We give here a brief description of one of the most interesting algorithm: the Gibbs sampler. We assume $\theta = (\theta_1, ..., \theta_M)$ where each θ_i is of arbitrary dimension. The Gibbs sampler proceeds as follow.

1. Draw an arbitrary sample $\theta^{(0)}$ and set $k = 1$.
2. At iteration k, draw a sample $X_{1\to N}^{(k)}$ from $p(X_{1\to N}|\,Y_{1\to N}, \theta^{(k-1)})$.
3. $\forall i \in \{1, ..., M\}$, draw a sample $\theta_i^{(k)}$ from $p(\theta_i|\,Y_{1\to N}, X_{1\to N}^{(k)}, \theta_{\{1,...,M\}\setminus i}^{(k)})$ where $\theta_{\{1,...,M\}\setminus i}^{(k)} = \left(\theta_1^{(k)}, ..., \theta_{i-1}^{(k)}, \theta_{i+1}^{(k-1)}, ..., \theta_M^{(k-1)}\right)$.
4. Replace k by $k+1$ and go to step 2.

Under weak regularity assumptions on the Markov chain (aperiodicity, irreducibility and Harris reccurence), $(X_{1\to N}^{(k)}, \theta^{(k)})$ is asymptotically a sample from the posterior density $p\left(X_{1\to N}, \theta|\,Y_{1\to N}\right)$, moreover the Markov chain is ergodic and thus it allows to perform statistically consistent ergodic averages.

3.2 Sampling from full conditional densities

We describe now how to sample from the so-called 'full conditional densities'. Further on, z^* denotes the current realization of a random variable z.

Sampling from the smoothing density

The decomposition of the smoothing density $p\left(X_{1\to N}|\,Y_{1\to N}, \theta^*\right)$ (3) suggests the following algorithm first presented in [1].

1. Compute the finite dimensional optimal filter for the current hyperparameter θ^* and store $\forall k$ the filtering and prediction densities $p\left(x\left(k\right)|\,Y_{1\to k}\right)$ and $p\left(x\left(k+1\right)|\,Y_{1\to k}\right)$.
2. Draw a sample $x^*\left(N\right)$ from the marginal smoothing density $p\left(x\left(N\right)|\,Y_{1\to N}\right)$ and put $k = N-1$.
3. Draw a sample $x^*\left(k\right)$ from

$$p\left(x\left(k\right)|\,Y_{1\to k}, x^*\left(k+1\right)\right) = \frac{p\left(x^*\left(k+1\right)|\,x\left(k\right)\right)p\left(x\left(k\right)|\,Y_{1\to k}\right)}{p\left(x^*\left(k+1\right)|\,Y_{1\to k}\right)}$$

If $k > 1$, set $k = k-1$ and go to step 3.

Sampling the hyperparameters

We must sample from $p(\theta_i|\,Y_{1\to N}, X_{1\to N}^*, \theta_{-i}^*)$ where θ_{-i}^* denotes $\theta_{\{1,...,M\}\setminus i}^*$. This problem should be addressed case by case. In the general case, we have

$$\begin{aligned} & p(\theta_i|\,Y_{1\to N}, X_{1\to N}^*, \theta_{-i}^*) \\ \propto\ & p(Y_{1\to N}|\,X_{1\to N}^*, \theta_{-i}^*, \theta_i)p(X_{1\to N}^*|\,\theta_{-i}^*, \theta_i)p\left(\theta_i|\,\theta_{-i}^*\right) \end{aligned}$$

If we have a conjugate prior, then sampling is generally straightforward. If it is not, one can use another stochastic simulation technique such as the accept/reject procedure or the Metropolis-Hastings algorithm[10].

3.3 An improved Gibbs sampler

Recently, Liu[6] has proposed an improvement over the random scan Gibbs sampler. His method called 'Metropolized Gibbs sampler' is more efficient in a discrete state-space framework in the sense that the asymptotic variance of ergodic averages is decreased. He reports similar improvements in simulations for mixed continous-discrete state space and deterministic scan but does not establish it theoretically. We describe it. Let x_i be a component (random variate or vector) lying in a finite state space of a random vector x with possibly continuous components. We want to sample from the joint density $\pi\left(x\right)$. The Gibbs sampler requires to sample from $\pi\left(x_i|\,x_{-i}^*\right)$. Instead, the Metropolized Gibbs sampler proceeds as follow. Draw a candidate $x_i^{'} \neq x_i$ from $\pi\left(x_i^{'}\middle|\,x_{-i}^*\right)/\left(1-\pi\left(x_i|\,x_{-i}^*\right)\right)$. Then accept it with probability $\min\left\{1, \left(1-\pi\left(x_i|\,x_{-i}^*\right)\right)/\left(1-\pi\left(x_i^{'}\middle|\,x_{-i}^*\right)\right)\right\}$, otherwise keep x_i. We have established that under weak assumptions when one component lies in a discrete state space, in a data augmentation setup (i.e. Gibbs sampler with two components), this algorithm converges geometrically, as does the Gibbs sampler[9], but faster than this latter[4].

4 Some applications

4.1 Blind Bernoulli-Gauss deconvolution

The non-blind problem is adressed in details in [3]. Let us consider an AR(p) model excited by a dynamic noise $v\left(k\right)$ which is a mixture of Gaussians, more precisely $v(k)$ is an i.i.d. sequence with $v\left(k\right) \sim \lambda\mathcal{N}\left(m, \sigma_v^2\right) + \left(1-\lambda\right)\mathcal{N}\left(0, \sigma_v^2/\alpha^2\right)$ where $1/\alpha^2 \ll 1 \neq 0$ (to ensure irreducibility of the Gibbs sampler). It models the signal at the output of a neutron sensor.

$$z\left(k\right) = \sum_{i=1}^{p} a_i z\left(k-i\right) + v\left(k\right)$$

In a state space framework

$$\mathbf{x}\left(k\right) = \begin{pmatrix} a_1 & \cdots & a_{p-1} & a_p \\ 1 & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ & & 1 & 0 \end{pmatrix} \mathbf{x}\left(k-1\right) + \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} v\left(k\right)$$

where $\mathbf{x}(k) = (z(k) \cdots z(k-p+1))^T$. We assume $\mathbf{x}(0) = (0 \cdots 0)^T$. However, we only have some noisy observations $y(k)$

$$y(k) = z(k) + w(k) = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix} \mathbf{x}^T(k) + w(k)$$

where $k = 1, ..., N$. $w(k)$ is an i.i.d. Gaussian sequence, $w(k) \sim \mathcal{N}(0, \sigma^2)$. We set $\theta = (\mathbf{a},\ \sigma_o^2,\ m,\ \sigma^2,\ \lambda,\ I_{1\to N})$ where $I_{1\to N} = (i(1), ..., i(N))$ are the missing data such that $v(k)|(i(k) = 0) \sim \mathcal{N}(0, \sigma_o^2/\alpha^2)$ and $v(k)|(i(k) = 1) \sim \mathcal{N}(m, \sigma_o^2)$, α being given. We assume the following prior density $p(\mathbf{a},\ \sigma_o^2,\ \sigma^2,\ m, \lambda, I_{1\to N}) = p(\mathbf{a})\, p(I_{1\to N}|\lambda)\, p(\lambda)\, p(\sigma_o^2)\, p(m)\, p(\sigma^2)$ where $\mathbf{a} = (a_1, ..., a_p)^T \sim \mathcal{N}(\mathbf{m}_0, \mathbf{\Sigma}_0)\, \mathbf{1}_{\mathcal{D}_{\mathbf{a},p}}(\mathbf{a})$ ($\mathcal{D}_{\mathbf{a},p}$ being the domain of stability of autoregressive models of order p), $\lambda \sim 1_{]0,1.0[}(\lambda)$, $m \sim \mathcal{N}(m_v, \sigma_v^2)$, $\sigma_o^2 \sim \text{Inv} - \chi^2(v_o; \gamma_o)$ and $\sigma^2 \sim \text{Inv} - \chi^2(v; \gamma)$, i.e. $v\gamma/\sigma^2 \sim \chi^2(v)$.

To apply the Gibbs sampler, we must be able to sample from $p(X_{1\to N}|Y_{1\to N}, \theta^*)$ and from $p(\theta_i|Y_{1\to N}, X^*_{1\to N}, \theta^*_{-i})$ that will be denoted further on $p(\theta_i|\overline{\theta_i}^*)$. To sample from $p(X_{1\to N}|Y_{1\to N}, \theta^*)$, we use the forward-backward procedure as, conditionally to θ, the optimal filter is the Kalman filter. However, in our case, we are also interested in estimating the so-called 'reflectivity sequence' $V_{1\to N} = (v(1), ..., v(N))$ instead of $X_{1\to N}$. It requires an algorithm that samples from $p(V_{1\to N}|Y_{1\to N}, \theta^*)$ which is described in [3] and is also based on a forward-backward procedure, the backward step being different. We obtain for $p(\theta_i|\overline{\theta_i}^*)$

$$\mathbf{a}|\,\overline{\mathbf{a}}^* \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma})\, \mathbf{1}_{\mathcal{D}_{\mathbf{a},p}}(\mathbf{a})$$

where

$$\begin{aligned}\mathbf{\Sigma}^{-1} = \mathbf{\Sigma}_0^{-1} &+ \sum_{\{k\}/i(k)=1} \mathbf{x}^*(k-1)\, \mathbf{x}^{*T}(k-1)/\sigma_o^{*2} \\ &+ \alpha^2 \sum_{\{k\}/i(k)=0} \mathbf{x}^*(k-1)\, \mathbf{x}^{*T}(k-1)/\sigma_o^{*2}\end{aligned}$$

$$\mathbf{m} = \left(\mathbf{\Sigma}_0^{-1}\mathbf{m}_0 + \sum_{\{k\}/i(k)=1} (z^*(k) - m^*)\, \mathbf{x}^*(k-1)/\sigma_o^{*2} \right.$$
$$\left. + \alpha^2 \sum_{\{k\}/i(k)=0} z^*(k)\, \mathbf{x}^*(k-1)/\sigma_o^{*2} \right)$$

We also have

$$\sigma^2 \left| \overline{\sigma}^{2*} \right. \sim \text{Inv} - \chi^2 \left(v + N, \frac{v\gamma + Ns}{v + N} \right)$$

where $s = \frac{1}{N}\sum_{k=1}^{N} (y(k) - z^*(k))^2$.

$$\sigma_o^2 \left| \overline{\sigma}_o^{2*} \right. \sim \text{Inv} - \chi^2 \left(v_o + N, \frac{v_o\gamma_o + n_o s_o + \alpha^2(N - n_o)s_{\overline{o}}}{v_o + N} \right)$$

where $s_o = \frac{1}{n_o}\sum_{\{k\}/i(k)=1} (v^*(k) - m^*)^2$, $s_{\overline{o}} = \frac{1}{N-n_o}\sum_{\{k\}/i(k)=0} v^{*2}(k)$ and $n_o = \text{card}\{i(k)/i(k) = 1\}$.

$$m|\,\overline{m}^* \sim \mathcal{N}(m'_v, \sigma'^2_v)$$

with

$$\begin{aligned} m'_v &= \frac{\sigma_v^2 \sum_{\{k\}/i(k)=1} v^*(k) + \sigma_o^{2*} m_v}{n_o\sigma_v^2 + \sigma_o^{2*}} \\ \sigma'^2_v &= \frac{\sigma_v^2 \sigma_o^{2*}}{n_o^*\sigma_v^2 + \sigma_o^{2*}} \end{aligned}$$

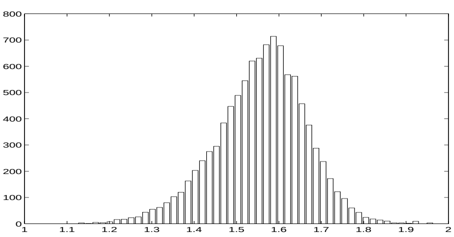
Finally

$$\lambda|\,\overline{\lambda}^* \sim \text{Beta}(n_o^* + 1, N - n_o^* + 1)$$

and

$$I_{1\to N}|\,\overline{I_{1\to N}}^* \sim \prod_{k=1}^{N} p(i(k)|\, v^*(k))$$

(Without any additional difficulty, it is possible to model $I_{1\to N}$ as a Markov chain.) One can easily sample from all the full conditional distributions, for the last one we can use a Metropolized Gibbs sampler. We have shown that this algorithm converges[4]. In the non-blind case, the duality principle, introduced by Robert et al.[9], allows to establish the geometric convergence rate of the Gibbs sampler[3], this convergence rate is improved for the Metropolized Gibbs sampler[4]. In our application we simulate a signal with the following parameters: $N = 250$, $p = 2$, $\mathbf{a} = (1.511\ \ -0.549)^T$, $\lambda = 0.03$, $\sigma = 0.3$, $m = 0.3$, $\sigma_o = 0.1$ (as $m_v \neq 0$, which is the case for a neutron sensor, it implies that there is no identifiability problem). The following known parameters for prior densities are $\mathbf{m}_0 = (\mathbf{0})$, $\mathbf{\Sigma}_0^{-1} = (\mathbf{0})$ (flat distribution over $\mathcal{D}_{\mathbf{a},2}$), $v = 2.0$, $\gamma = 0.3$ and for parameters of the dynamic noise, $m_v = 0.5$, $\sigma_v = 0.2$. Here, we assume that σ_o is known. We take $\alpha = 20$. For this example, the burn-in period of the Gibbs sampler is short (about 50 iterations to obtain good results). Then $10,000$ iterations are kept to estimate conditional expectations and standard deviation, we obtain $E(a_1|Y_{1\to N}) = 1.557$, $Std(a_1|Y_{1\to N}) = 0.107$, $E(a_2|Y_{1\to N}) = -0.595$, $Std(a_2|Y_{1\to N}) = 0.09$, $E(\lambda|Y_{1\to N}) = 0.037$, $Std(\lambda|Y_{1\to N}) = 0.017$, $E(m|Y_{1\to N}) = 0.331$, $Std(m|Y_{1\to N}) = 0.031$, $E(\sigma_o|Y_{1\to N}) = 0.122$ and $Std(\sigma_o|Y_{1\to N}) = 0.01$. We present below the estimated posterior density $p(a_1|Y_{1\to N})$.



4.2 Blind channel equalization

We present here an application of the Gibbs sampler to blind deconvolution of a linear channel. Let us consider the following linear channel

$$y(k) = \sum_{i=0}^{p} h_i x(k-i) + w(k)$$

where $k = 1, ..., N$. $x(k)$ is a finite state space Markov chain $\{a_1, ..., a_n\}$, $\mathbf{x}(k) = (x(k), x(k-1), ..., x(k-p))^T$ and $X_{1\to N} = \{\mathbf{x}(1), ..., \mathbf{x}(N)\}$. To avoid cumbersome developments we assume that $x(k)$ is i.i.d., the non-i.i.d. case is straightforward[9][2][4]. Then $\mathbf{x}(k) = (x(k), \mathbf{x}_b^T(k))^T = (\mathbf{x}_f^T(k), x(k-p))^T$ is a hidden Markov chain of probabilities transitions $p(\mathbf{x}(k)|\mathbf{x}(k-1)) = \frac{1}{n}\delta(\mathbf{x}_b(k), \mathbf{x}_f(k-1))$. We assume that $w(k)$ is a white i.i.d. Gaussian noise $w(k) \sim \mathcal{N}(0, \sigma^2)$ and the channel is assumed unknown $\mathbf{h}$. We want to estimate $X_{1\to N}$ and $\theta = (\mathbf{h}, \sigma^2)$ with prior density $p(\theta) = p(\mathbf{h})\, p(\sigma^2)$ where $\sigma^2 \sim \text{Inv}-\chi^2(v;\gamma)$ and $\mathbf{h} \sim \mathcal{N}(\mathbf{m}_0, \mathbf{\Sigma}_0) 1_{\mathcal{D}_\mathbf{h}}(\mathbf{h})$, $\mathcal{D}_\mathbf{h}$ is the domain allowing identifiability of the channel. (If one does not consider such restrictions, the posterior distribution is a mixture of distributions and ergodic averages lose their interest as there is an average effect between components of this mixture). Using the Gibbs sampler, we have to sample from $p(X_{1\to N}|Y_{1\to N}, \theta^*)$ and $p(\theta_i|\overline{\theta_i}^*)$. To sample from $p(X_{1\to N}|Y_{1\to N}, \theta^*)$, we use the forward-backward procedure, as conditionally to θ, the optimal filter is finite-dimensional. For hyperparameters, we have

$$\mathbf{h}|\overline{\mathbf{h}}^* \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma})\, 1_{\mathcal{D}_\mathbf{h}}(\mathbf{h})$$

where

$$\mathbf{\Sigma}^{-1} = \sum_{k=1}^{N} \mathbf{x}^*(k)\, \mathbf{x}^{*T}(k) / \sigma^{*2} + \mathbf{\Sigma}_0^{-1}$$

$$\mathbf{m} = \mathbf{\Sigma}\left(\sum_{k=1}^{N} y(k)\, \mathbf{x}^*(k) / \sigma^{*2} + \mathbf{\Sigma}_0^{-1}\mathbf{m}_0\right)$$

and

$$\sigma^2 \big| \overline{\sigma}^{2*} \sim \text{Inv} - \chi^2\left(v + N, \frac{v\gamma + Ns}{v + N}\right)$$

where $s = \frac{1}{N}\sum_{k=1}^{N}\left(y(k) - \mathbf{h}^{*T}\mathbf{x}^*(k)\right)^2$. Extension to Volterra channels is straightforward. The implementation of the Gibbs sampler that we present here is more efficient that than those described in [9][2] because we sample in block the hidden-state process $X_{1\to N}$ [5]. (However, if $N\text{card}(\mathfrak{X})^{p+1} \gg 1$, this method can not be used in practice). For this component, it is also possible to use the Metropolized Gibbs sampler. If σ is known, the duality principle gives the geometric convergence of the Gibbs sampler, this rate is improved for the Metropolized Gibbs sampler[4]. In our application we simulate a signal with the following parameters: $N = 250$, $p = 3$, $\sigma = 0.7$, $\mathbf{h} = (0.916; -0.183; 0.481; -0.199)^T$, $\mathfrak{X} = \{-3, -1, 1, 3\}$. The following known parameters for prior densities are $\mathbf{m}_0 = (\mathbf{0})$, $\mathbf{\Sigma}_0^{-1} = (\mathbf{0})$ and $\mathcal{D}_\mathbf{h} = \{\mathbf{h}\backslash h(0) > 0.4, h(0) > |h(i)| + 0.2 \text{ for } i \neq 0\}$ (to avoid sign and shift ambiguities), $v = 2.0$ and $\gamma = 0.3$. For this example, the burn-in period of the Gibbs sampler is very short (after less than 10 iterations we are in the area of interest) whatever the initialization is. Then $10,000$ iterations are kept to estimate conditional expectations and variance, $E(\mathbf{h}|Y_{1\to N}) = (0.879; -0.149; 0.441; -0.157)^T$, $Std(\mathbf{h}|Y_{1\to N}) = (0.028; 0.033; 0.038; 0.041)^T$, $E(\sigma|Y_{1\to N}) = 0.728$ and $Var(\sigma|Y_{1\to N}) = 0.041$.

5 Conclusion

In this article, we present in an unified framework efficient applications of stochastic simulation techniques to perform Bayesian estimation of a wide class of hidden Markov models. These powerful techniques allow to perform statistical analysis of numerous problems of interest. Simulations for blind deconvolution of Bernoulli-Gauss processes and blind channel equalization are presented for which very satisfactory results are obtained. Other applications are currently under investigations.

References

[1] C.K. Carter and R. Kohn, “On Gibbs Sampling for State Space Models”, *Biometrika*, vol. 81, no. 3, pp. 541-553, 1994.

[2] R. Chen and T. Li, “Blind Restoration of Linearly Degraded Discrete Signals by Gibbs Sampling”, *IEEE on SP*, vol. 43, no. 10, pp. 2410-2413, 1995.

[3] A. Doucet and P. Duvaut, “Fully Bayesian Analysis of Conditionally Linear Gaussian State Space Models”, in *Proc. IEEE Conf. ICASSP*, Atlanta, 1996.

[4] A. Doucet and P. Duvaut, “Bayesian Estimation of Hidden Markov Models via Markov Chain Monte Carlo Methods”, submitted.

[5] J.S. Liu, W.H. Wong and A. Kong, “Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes”, *Biometrika*, Vol. 81, No. 1, pp. 27-40, 1994.

[6] J.S. Liu, “Metropolized Gibbs Sampler: An Improvement”, technical report, 1995.

[7] R.E. McCulloch and R.S. Tsay, “Bayesian Analysis of Autoregressive Times Series via the Gibbs Sampler”, *J. Tim. Ser. Ana.*, vol. 3, no. 4, pp. 253-264, 1994.

[8] W. Qian and D.M. Titterington, “Estimation of Parameters in Hidden Markov Models”, *Phil. Trans. Roy. Soc. Lon. A*, no. 337, pp. 407-428, 1991.

[9] C.P. Robert, G. Celeux and J. Diebolt, “Bayesian Estimation of Hidden Markov Chains: a Stochastic Implementation”, *Stat. Prob. Let.*, vol. 16, pp. 77-83, 1993.

[10] A.F.M. Smith and G.O. Roberts, “Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods”, *J. Roy. Stat. Soc. B*, vol. 55, no. 1, pp. 3-23, 1993.