[8] S. Martucci, "Signal extension and noncausal filtering for subband coding of images," in *SPIE VCIP '91: Visual Communication*, vol. 1605, Nov. 1991, pp. 137–148.
[9] H. Kiwa, K. Nishikawa, and M. Iwahashi, "A development of symmetric extension method for subband image coding," *IEEE Trans. Image Processing*, vol. 3, pp. 78–81, Jan. 1994.
[10] R. H. Bamberger, S. L. Eddins, and V. Nuri, "Generalized symmetric extensions for size-limited multirate filter banks," *IEEE Trans. Image Processing*, vol. 3, pp. 82–87, Jan. 1994.
[11] R. L. de Queiroz, "Subband processing of finite length signals without border distortions," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, vol. IV, 1992, pp. 613–616.
[12] J. N. Bradley and V. Faber, "Perfect reconstruction with critically sampled filter banks and linear boundary conditions," preprint.
[13] V. Nuri and R. H. Bamberger, "A theorectical framework for the analysis and design of size-limited multirate filter banks," *Proc. Int. Symp. Time-Frequency Time-Scale Anal.*, Victoria, Canada, Oct. 1992.
[14] R. L. de Queiroz and K. R. Rao, "Time-varying lapped transforms and wavelet packets," *IEEE Trans. Signal Processing*, vol. 41, pp. 3293–3305, Dec. 1993.
[15] C. Herley, J. Kovacevic, K. Ramchandran, and M. Vetterli, "Tilings of the time-frequency plane: Construction of arbitrary orthogonal bases and fast tiling algorithms," *IEEE Trans. Signal Processing*, vol. 41, pp. 3341–3359, Dec. 1993.

Blind Restoration of Linearly Degraded Discrete Signals by Gibbs Sampling

Rong Chen and Ta-Hsin Li

***Abstract*—This paper addresses the problem of simultaneous parameter estimation and restoration of discrete-valued signals that are blurred by an unknown FIR filter and contaminated by additive Gaussian white noise with unknown variance. Assuming that the signals are stationary Markov chains with known state space but unknown initial and transition probabilities, Bayesian inference of all unknown quantities is made from the blurred and noisy observations. A Monte Carlo procedure, called the Gibbs sampler, is employed to calculate the Bayesian estimates. Simulation results are presented to demonstrate the effectiveness of the method.**

I. Introduction

Suppose a discrete-valued (digital) signal $\{x_t\}$ is blurred by an FIR linear filter $\{\phi_i\}$ and contaminated by additive noise $\{\epsilon_t\}$, so that the observed signal $\{y_t\}$ can be written as

$$y_t = \sum_{i=0}^{q} \phi_i x_{t-i} + \epsilon_t \quad (t = 1, \cdots, n). \tag{1}$$

The so-called blind restoration problem is to *simultaneously* estimate the filter $\{\phi_i\}$ and to recover the signal $\{x_t\}$ solely from the observed data record $\{y_t\}$ along with some partial statistical information about $\{x_t\}$. This problem stems from the equalization of digital communication channels in which the signals take only discrete values (e.g., [1], [14]).

In the absence of noise, the restoration (or deconvolution) problem can be approached in many different ways under the assumption that the x_t are independent and identically distributed (i.i.d.) (e.g., [1], [4], [7]). In particular, an efficient method along the lines of inverse filtering has been proposed ([9]–[11]) that explicitly utilizes the discreteness of $\{x_t\}$ yet does not require the stationarity or other statistical information of $\{x_t\}$.

In this correspondence, we deal with the blind restoration problem under a Bayesian framework and by Gibbs sampling. The Gibbs sampling has been successfully applied to the *ordinary* image restoration problem by Geman and Geman [6] under the assumption that the filter $\{\phi_i\}$ and the statistical parameters of $\{x_t\}$ and $\{\epsilon_t\}$ are all available. In the present correspondence, we include these parameters in the list of unknowns and estimate them simultaneously with the signal $\{x_t\}$.

Manuscript received April 3, 1993; revised March 17, 1995. The associate editor coordinating the review of this paper and approving it for publication was Dr. Barry Sullivan.

The authors are with the Department of Statistics, Texas A&M University, College Station, TX 77843-3143 USA.

IEEE Log Number 9413859.

II. Formulation of the Problem

Assume that the signal $\{x_t\}$ in (1) is a stationary first-order Markov chain with known state space $\mathcal{A} := \{a_1, \cdots, a_k\}$ but unknown initial probabilities $\theta_i := \mathrm{pr}(x_{1-q} = a_i)$ and unknown transition probabilities $\theta_{ij} := \mathrm{pr}(x_t = a_j | x_{t-1} = a_i)$. It is clear that the probabilities should satisfy the constraints $\sum_{i=1}^{k} \theta_i = 1$ and $\sum_{j=1}^{k} \theta_{ij} = 1$ for $i = 1, \cdots, k$. Let $\boldsymbol{\theta}$ denote the collection of these probabilities, namely $\boldsymbol{\theta} := \{\theta_i, \theta_{ij} : i, j = 1, \cdots, k\}$. Although extensions to higher order Markov chains are quite straightforward, we restrict our effort to the first-order case for the simplicity of presentation. Assume further that $\{\epsilon_t\}$ in (1) is Gaussian white noise with zero-mean and unknown variance σ^2 and is independent of $\{x_t\}$.

Under these assumptions, the main objective of this correspondence is to simultaneously reconstruct the signal $\boldsymbol{x} := \{x_{1-q}, \cdots, x_n\}$ and estimate the FIR filter $\boldsymbol{\phi} := [\phi_0, \cdots, \phi_q]'$ along with the statistical parameters σ^2 and $\boldsymbol{\theta}$ on the basis of the data record $\boldsymbol{y} := \{y_1, \cdots, y_n\}$. Note that the values $x_{1-q}, \cdots, x_0$ (that are outside the observation interval) are also included in $\boldsymbol{x}$ for reconstruction and that the filter $\boldsymbol{\phi}$ can be minimum phase or nonminimum phase. Noncausal FIR filters can be accommodated into the problem by a transformation of time index.

III. Bayesian Approach

The problem is solved under a Bayesian framework: First, the unknown quantities $\boldsymbol{x}$, $\boldsymbol{\phi}$, σ^2, and $\boldsymbol{\theta}$ are regarded as realizations of random variables with suitable prior distributions. The Gibbs sampler, a Monte Carlo method, is then employed to calculate the minimum mean-squared error (MMSE) estimates and/or the maximum *a posteriori* (MAP) estimates of the unknowns.

A. Prior Distributions

In principle, prior distributions are used to incorporate our knowledge of the parameters, and less restrictive (or less informative) priors should be employed when such knowledge is limited. Computational complexity is another consideration that affects the selection. Conjugate priors are usually used to obtain simple analytical forms for the resulting posterior distributions (e.g., [2]). To make the Gibbs sampler more computationally efficient, the priors should also be chosen such that the conditional posterior distributions, as we shall see next, are easy to simulate.

For the restoration problem described above, the following priors are used in our procedure: to the filter $\boldsymbol{\phi}$, we impose a Gaussian distribution $p(\boldsymbol{\phi}) \sim N(\boldsymbol{\phi}_0, \Sigma_0)$, and to the noise variance σ^2 we impose an inverted chi-square distribution $p(\sigma^2) \sim \chi^{-2}(\nu; \lambda)$, i.e., $\nu\lambda/\sigma^2 \sim \chi^2(\nu)$. Note that large values of Σ_0 and small values of ν and λ correspond to less informative priors. Further, we use independent Dirichlet distributions as priors of θ_i and θ_{ij}.

More precisely, let $D(\boldsymbol{\alpha})$ denote the Dirichlet distribution with parameters $\boldsymbol{\alpha} := \{\alpha_1, \cdots, \alpha_k\}$, $(\alpha_i > -1)$; the pdf of $D(\boldsymbol{\alpha})$ is defined as $\mu(\boldsymbol{p}; \boldsymbol{\alpha}) := c \prod_{i=1}^k p_i^{\alpha_i}$ for $\boldsymbol{p} := \{p_1, \cdots, p_k\}$ with $0 < p_i < 1$ and $\sum_{i=1}^k p_i = 1$, where $c := c(\boldsymbol{\alpha})$ is the normalizing constant. Note that Jeffery's noninformative prior corresponds to $\alpha_i = -1/2$ for $i = 1, \cdots, k$ (e.g., [2]). We assume that $\boldsymbol{\theta}_0 := \{\theta_1, \cdots, \theta_k\}$ and $\boldsymbol{\theta}_i := \{\theta_{i1}, \cdots, \theta_{ik}\}$, $i = 1, \cdots, k$, are mutually independent with $\boldsymbol{\theta}_0 \sim D(\boldsymbol{\alpha})$ and $\boldsymbol{\theta}_i \sim D(\boldsymbol{\alpha}_i)$ where $\boldsymbol{\alpha}_i := \{\alpha_{i1}, \cdots, \alpha_{ik}\}$. Under this assumption, the following prior can be obtained for $\boldsymbol{\theta} = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k\}$ in the Markovian case, namely $p(\boldsymbol{\theta}) = \mu(\boldsymbol{\theta}_0; \boldsymbol{\alpha}) \prod_{i=1}^k \mu(\boldsymbol{\theta}_i; \boldsymbol{\alpha}_i)$. When $\{x_t\}$ is an i.i.d. sequence (a degenerate Markov chain), the parameter set $\boldsymbol{\theta}$ reduces to $\boldsymbol{\theta}_0$ and the prior distribution of $\boldsymbol{\theta}$ becomes $p(\boldsymbol{\theta}) = \mu(\boldsymbol{\theta}; \boldsymbol{\alpha})$.

B. Bayesian Inference

Under the Markovian assumption, the distribution of $\boldsymbol{x}$ given $\boldsymbol{\theta}$ can be written as

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = s_1(\boldsymbol{x}, \boldsymbol{\theta}) := \prod_{i=1}^k \theta_i^{\delta_i} \prod_{i,j=1}^k \theta_{ij}^{n_{ij}} \tag{2}$$

where n_{ij} is the number of pairs (x_{t-1}, x_t) in $\boldsymbol{x}$ that are equal to (a_i, a_j) and δ_i is the indicator such that $\delta_i = 1$ if $x_{1-q} = a_i$ and $\delta_i = 0$ if $x_{1-q} \neq a_i$. In the i.i.d. case, the distribution of $\boldsymbol{x}$ given $\boldsymbol{\theta} = \{\theta_1, \cdots, \theta_k\}$ reduces to

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = s_0(\boldsymbol{x}, \boldsymbol{\theta}) := \prod_{i=1}^k \theta_i^{n_i} \tag{3}$$

where n_i is the number of x_t's in $\boldsymbol{x}$ that are equal to a_i.

Since $\{\epsilon_t\}$ is white and Gaussian, the *joint posterior distribution* of the unknown quantities takes the form of

$$\begin{aligned} p(\boldsymbol{\phi}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y}) &\propto (1/\sigma^2)^{n/2} \\ &\times \exp\left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^n \left(y_t - \sum_{i=0}^q \phi_i x_{t-i} \right)^2 \right\} \\ &\times p(\boldsymbol{\phi})\, p(\sigma^2)\, p(\boldsymbol{\theta})\, s(\boldsymbol{x}, \boldsymbol{\theta}) \end{aligned} \tag{4}$$

where $s(\boldsymbol{x}, \boldsymbol{\theta}) := p(\boldsymbol{x}|\boldsymbol{\theta})$ can be either $s_1(\boldsymbol{x}, \boldsymbol{\theta})$ or $s_0(\boldsymbol{x}, \boldsymbol{\theta})$, depending on whether $\{x_t\}$ is a Markov chain or an i.i.d. sequence.

Although the joint posterior distribution (4) is given explicitly (up to a normalizing constant), direct calculation of the MMSE and/or MAP estimates of the unknowns is computationally forbidding. For example, since the MMSE estimator of $\boldsymbol{x}$ is the posterior mean $E(\boldsymbol{x}|\boldsymbol{y})$, any direct evaluation of this estimator involves the multiple integration $E(\boldsymbol{x}|\boldsymbol{y}) = \int \boldsymbol{x}\, p(\boldsymbol{\phi}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})\, d\boldsymbol{\phi}\, d\sigma^2 d\boldsymbol{\theta}\, d\boldsymbol{x}$, the computational burden of which can be enormous. A similar problem arises in any direct calculation of the MAP estimator of $\boldsymbol{x}$ that maximizes the posterior marginal density $p(\boldsymbol{x}|\boldsymbol{y}) = \int p(\boldsymbol{\phi}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})\, d\boldsymbol{\phi}\, d\sigma^2 d\boldsymbol{\theta}$.

IV. Gibbs Sampling

To avoid the direct evaluation of the Bayesian estimates that require multiple integration, we resort to a *Monte Carlo method* instead. The basic idea of the Monte Carlo method is to generate an ergodic random sample from the distribution (4) and then to average, for instance, the $\boldsymbol{x}$ components throughout the sample to obtain an approximation of $E(\boldsymbol{x}|\boldsymbol{y})$. The Gibbs sampler provides a recursive way of generating such a sample.

A. Conditional Posterior Distributions

In our problem, the implementation of Gibbs sampling requires the following conditional posterior distributions that can be easily obtained (e.g., [2]).

1) It can be shown that

$$p(\boldsymbol{\phi}|\sigma^2, \boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y}) \sim N(\boldsymbol{\phi}_*, \Sigma_*) \tag{5}$$

where

$$\Sigma_*^{-1} := \sum_{t=1}^n \frac{\boldsymbol{x}_t \boldsymbol{x}_t'}{\sigma^2} + \Sigma_0^{-1}$$

and

$$\boldsymbol{\phi}_* := \Sigma_* \left(\sum_{t=1}^n \frac{\boldsymbol{x}_t y_t}{\sigma^2} + \Sigma_0^{-1} \boldsymbol{\phi}_0 \right)$$

with $\boldsymbol{x}_t := [x_t, \cdots, x_{t-q}]'$. In fact, if $p(\boldsymbol{\phi}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})$ in (4) is regarded as a function of $\boldsymbol{\phi}$ by keeping other variables fixed, then it follows that

$$\begin{aligned} p(\boldsymbol{\phi}|\sigma^2, \boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y}) &\propto p(\boldsymbol{\phi}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y}) \\ &\propto \exp\left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - \boldsymbol{\phi}'\boldsymbol{x}_t)^2 \right. \\ &\qquad \left. - (\boldsymbol{\phi} - \boldsymbol{\phi}_0)' \Sigma_0^{-1} (\boldsymbol{\phi} - \boldsymbol{\phi}_0) \right\} \\ &\propto \exp\{ -(\boldsymbol{\phi} - \boldsymbol{\phi}_*)' \Sigma_*^{-1} (\boldsymbol{\phi} - \boldsymbol{\phi}_*) \}. \end{aligned}$$

2) It can also be shown that

$$p(\sigma^2|\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y}) \sim \chi^{-2}(\nu_*; \lambda_*) \tag{6}$$

where $\nu_* = \nu + n$, $\lambda_* := (\nu\lambda + s^2)/\nu_*$, and $s^2 := \sum_{t=1}^n (y_t - \boldsymbol{\phi}'\boldsymbol{x}_t)^2$. To verify this, it suffices to note that with fixed $\boldsymbol{\phi}$, $\boldsymbol{\theta}$, $\boldsymbol{x}$, and $\boldsymbol{y}$, $p(\sigma^2|\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y}) \propto (1/\sigma^2)^{(\nu+n+2)/2} \exp\{-(\nu\lambda + s^2)/(2\sigma^2)\}$.

3) Let $\boldsymbol{\theta}_{[-i]} := \boldsymbol{\theta} \setminus \boldsymbol{\theta}_i$ for $i = 0, 1, \cdots, k$; then, in the Markovian case

$$p(\boldsymbol{\theta}_i|\boldsymbol{\phi}, \sigma^2, \boldsymbol{\theta}_{[-i]}, \boldsymbol{x}, \boldsymbol{y}) \sim D(\boldsymbol{\alpha}_i^*) \quad (i = 1, \cdots, k) \tag{7}$$

where $\boldsymbol{\alpha}_i^* := \{\alpha_{i1}^*, \cdots, \alpha_{ik}^*\}$ with $\alpha_{ij}^* := \alpha_{ij} + n_{ij}$. This can be derived from (4) upon noting that $p(\boldsymbol{\theta}_i|\boldsymbol{\phi}, \sigma^2, \boldsymbol{\theta}_{[-i]}, \boldsymbol{x}, \boldsymbol{y}) \propto p(\boldsymbol{\theta})\, s_1(\boldsymbol{x}, \boldsymbol{\theta}) \propto \prod_{j=1}^k \theta_{ij}^{n_{ij}+\alpha_{ij}}$. Similarly, one obtains

$$p(\boldsymbol{\theta}_0|\boldsymbol{\phi}, \sigma^2, \boldsymbol{\theta}_{[-0]}, \boldsymbol{x}, \boldsymbol{y}) \sim D(\boldsymbol{\alpha}^*) \tag{8}$$

in both Markovian and i.i.d. cases, where $\boldsymbol{\alpha}^* := \{\alpha_1^*, \cdots, \alpha_k^*\}$ with $\alpha_i^* := \alpha_i + n_i$.

4) For any fixed $t^* \in \{1 - q, \cdots, n\}$, let $\boldsymbol{x}_{[-t^*]} := \boldsymbol{x} \setminus x_{t^*}$; then it follows that

$$\begin{aligned} &\mathrm{pr}(x_{t^*} = a_i|\boldsymbol{\phi}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{x}_{[-t^*]}, \boldsymbol{y}) \\ &\quad \propto \exp\left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - \boldsymbol{\phi}'\boldsymbol{x}_t^*)^2 \right\} s(\boldsymbol{x}^*, \boldsymbol{\theta}) \end{aligned} \tag{9}$$

for $i = 1, \cdots, k$, where $\boldsymbol{x}_t^* := [x_t^*, \cdots, x_{t-q}^*]'$ and $\boldsymbol{x}^* := \{x_{1-q}^*, \cdots, x_n^*\}$ with $x_{t^*}^* := a_i$ and $x_t^* := x_t$ for $t \in \{1 - q, \cdots, n\} \setminus t^*$. (Note that $\boldsymbol{x}_t^*$ and $\boldsymbol{x}^*$ can be obtained by substituting x_{t^*} with a_i in $\boldsymbol{x}_t$ and $\boldsymbol{x}$, respectively.)

B. The Gibbs Sampler

Using the conditional posterior distributions, the Gibbs sampler proceeds iteratively as follows: given initial values $\{\boldsymbol{\theta}(0), \sigma^2(0), \boldsymbol{x}(0)\}$ and for $m = 1, 2, \cdots$

1) Draw $\boldsymbol{\phi}(m)$ from $p(\boldsymbol{\phi}|\sigma^2(m-1), \boldsymbol{\theta}(m-1), \boldsymbol{x}(m-1), \boldsymbol{y})$ given by (5);
2) Draw $\sigma^2(m)$ from $p(\sigma^2|\boldsymbol{\phi}(m), \boldsymbol{\theta}(m-1), \boldsymbol{x}(m-1), \boldsymbol{y})$ given by (6);

3) For $i = 0, 1, \cdots, k$, draw $\boldsymbol{\theta}_i(m)$ from $p(\boldsymbol{\theta}_i|\boldsymbol{\phi}(m), \sigma^2(m), \boldsymbol{\theta}_{[-i]}(m-1), \boldsymbol{x}(m-1), \boldsymbol{y})$ given by (7) and (8), where $\boldsymbol{\theta}_{[-i]}(m-1) := \{\boldsymbol{\theta}_0(m), \cdots, \boldsymbol{\theta}_{i-1}(m), \boldsymbol{\theta}_{i+1}(m-1), \cdots, \boldsymbol{\theta}_k(m-1)\}$; set $\boldsymbol{\theta}(m) := \{\boldsymbol{\theta}_0(m), \boldsymbol{\theta}_1(m), \cdots, \boldsymbol{\theta}_k(m)\}$;
4) For $t^* = 1-q, \cdots, n$, draw $x_{t^*}(m)$ from $\mathrm{pr}(x_{t^*} = a_i|\boldsymbol{\phi}(m), \sigma^2(m), \boldsymbol{\theta}(m), \boldsymbol{x}_{[-t^*]}(m-1), \boldsymbol{y})$ given by (9), where $\boldsymbol{x}_{[-t^*]}(m-1) := \{x_{1-q}(m), \cdots, x_{t^*-1}(m), x_{t^*+1}(m-1), \cdots, x_n(m-1)\}$; set $\boldsymbol{x}(m) := \{x_{1-q}(m), \cdots, x_n(m)\}$;
5) Set $m = m+1$ and go to Step 1.

Standard routines can be used to generate these random samples. For example, in Step 3, $\boldsymbol{\theta}_i \sim D(\boldsymbol{\alpha}_i^*)$ can be obtained from beta random variables by first generating p_j from $\mathrm{Beta}(\alpha_{ij}^*, \alpha_{i,j+1}^* + \cdots + \alpha_{ik}^*)$ for $j = 1, \cdots, k-1$ and then setting $\theta_{i1} := p_1$, $\theta_{ij} := (1 - \sum_{v=1}^{j-1} \theta_{iv})p_j$ for $j = 2, \cdots, k-1$, and $\theta_{ik} = 1 - \sum_{j=1}^{k-1} \theta_{ij}$. Under suitable conditions, it can be shown (e.g., [5], [6]) that the random sequences $\{\boldsymbol{\phi}(m)\}$, $\{\sigma^2(m)\}$, $\{\boldsymbol{\theta}(m)\}$, and $\{\boldsymbol{x}(m)\}$ converge in distribution to the posterior marginal distributions $p(\boldsymbol{\phi}|\boldsymbol{y})$, $p(\sigma^2|\boldsymbol{y})$, $p(\boldsymbol{\theta}|\boldsymbol{y})$, and $p(\boldsymbol{x}|\boldsymbol{y})$, respectively. The sequences can also be shown to be ergodic so that the sample averages converge to the corresponding ensemble averages as the sample size grows without bound.

To ensure the convergence, the Gibbs sampler (Steps 1–5) is usually carried out $N+M$ times and the samples from the last M iterations are used to calculate the Bayesian estimates. In particular, the MMSE and MAP estimates of x_t, i.e., $E(x_t|\boldsymbol{y})$ and $\arg\max_{a\in\mathcal{A}}\{\mathrm{pr}(x_t = a|\boldsymbol{y})\}$, are approximated by $\hat{x}_t := M^{-1}\sum_{m=N+1}^{N+M} x_t(m)$ and $\tilde{x}_t := \arg\max_{a\in\mathcal{A}} M^{-1}\sum_{m=N+1}^{N+M} \delta(x_t(m) = a)$ for $t = 1-q, \cdots, n$, where $\delta(x = a) = 1$ if $x = a$ and $\delta(x = a) = 0$ if $x \neq a$. Similarly, the sample means of $\{\boldsymbol{\phi}(m)\}$, $\{\sigma^2(m)\}$, and $\{\boldsymbol{\theta}(m)\}$ can be used to approximate the corresponding MMSE estimates. Furthermore, the sample variances of $\{\boldsymbol{\phi}(m)\}$, $\{\sigma^2(m)\}$, $\{\boldsymbol{\theta}(m)\}$, and $\{\boldsymbol{x}(m)\}$ are approximations to the posterior variances $V(\boldsymbol{\phi}|\boldsymbol{y})$, $V(\sigma^2|\boldsymbol{y})$, $V(\boldsymbol{\theta}|\boldsymbol{y})$, and $V(\boldsymbol{x}|\boldsymbol{y})$, respectively, which reflect the uncertainty in estimating these unknowns on the basis of $\boldsymbol{y}$.

C. Remarks

Blind deconvolution problems in general can only be solved up to an arbitrary time delay, and sometimes also up to an arbitrary sign, if no further restrictions are imposed on the filter $\{\phi_i\}$ (e.g., [1], [9], [12]). Particularly when $\phi_i \approx 0$ for $i = q^*+1, \cdots, q$ in (1), the time delay of the input signal $\{x_t\}$ is essentially unidentifiable. In fact, the models $y_t = \sum_{i=0}^{q} \phi_{i-\tau} x_{t+\tau-i} + \epsilon_t$, for $\tau = 1, \cdots, q^*$, are all practically equivalent to (1). As a result, the posterior distribution can be a *mixture* of several distributions, each corresponding to a particular time delay. In this case, the convergence of Gibbs sampling may become very slow. If the distribution of $\{x_t\}$ is symmetric about zero, the solution is also subject to the ambiguity of sign.

To overcome this problem, one may adopt the following *constrained* Gibbs sampler along the lines of [3] and [13]. In the constrained Gibbs sampler, the coefficient ϕ_0 is restricted to be positive so that $\phi_0 \geq \eta$ for some predetermined constant $\eta > 0$. To draw samples of $\boldsymbol{\phi}$ that satisfy this condition, the so-called *rejection method* can be used: after a sample is drawn from (5) in Step 1, check to see if the constraint is satisfied; if not, the sample is rejected and a new sample is drawn from (5); the procedure continues until a sample is obtained that satisfies the constraint. If a desired sample has not been obtained after a large number of rejections, it is more appropriate to *shift* the ϕ_i's in the last sample until the first coefficient satisfies the constraint; the vacancies left at the end can be filled with zeros.

Another plausible restriction is on the location of the largest ϕ_i. For example, one can require that $\phi_{q^*} \geq |\phi_i| + \eta$ for $i \neq q^*$ and some $\eta > 0$, meaning that the largest coefficient should be positive and occur at lag $i = q^*$. A random sample of $\boldsymbol{\phi}$ that satisfies this constraint can be obtained by first generating a sample from (5) in Step 1 and then shifting the ϕ_i's in the sample, with zeros filling the vacancies, to make the largest coefficient appear at $i = q^*$.

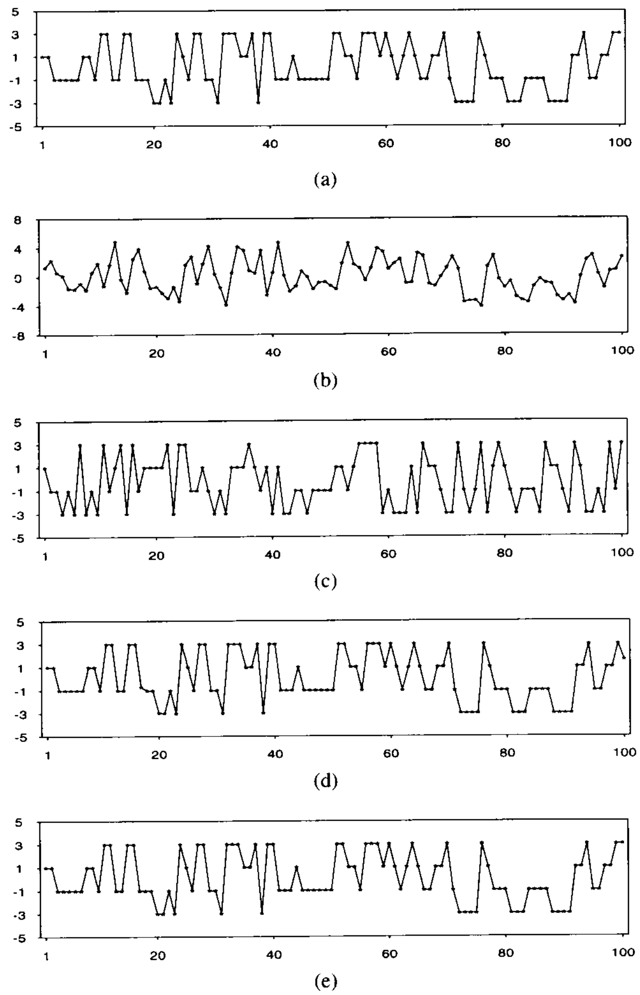

Fig. 1. Blind restoration by Gibbs sampling. (a) Markov signal $\{x_t\}$; (b) observed data $\{y_t\}$ (15 dB); (c) initial guess for $\{x_t\}$ (i.i.d. with uniform distribution in $\mathcal{A}$); (d) MMSE estimate $\{\hat{x}_t\}$from Gibbs sampling; (e) MAP estimate $\{\tilde{x}_t\}$ from Gibbs sampling.

V. Simulations

To demonstrate the performance of the method, let us consider the following examples.

Example 1: Consider an MA(3) model with $\boldsymbol{\phi} = [-0.1833, 0.9162, 0.4812, -0.1987]'$, i.e.

$$\begin{aligned} y_t = &-0.1833x_t + 0.9162x_{t-1} + 0.4812x_{t-2} \\ &- 0.1987x_{t-3} + \epsilon_t \end{aligned} \tag{10}$$

where $\{x_t\}$ is a first-order four-level Markov chain with $\mathcal{A} = \{-3, -1, 1, 3\}$, $\theta_i = 1/4$, and

$$[\theta_{ij}] = \begin{bmatrix} 0.4 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.4 & 0.2 & 0.2 \\ 0.1 & 0.3 & 0.4 & 0.2 \\ 0.1 & 0.2 & 0.3 & 0.4 \end{bmatrix}.$$

A realization of $\{x_t\}$ with $n = 100$ is shown in Fig. 1(a) and the corresponding $\{y_t\}$ shown in Fig. 1(b). The sample variance of $\{\epsilon_t\}$

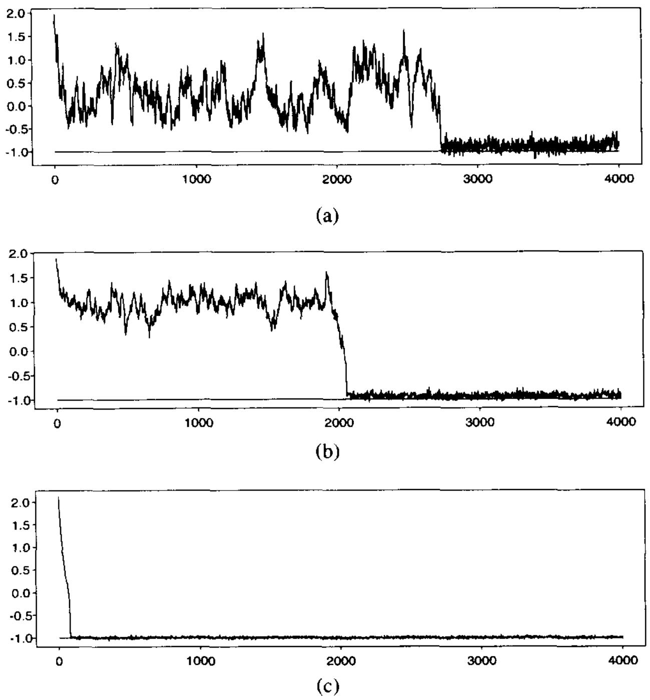

Fig. 2. Trajectory of the noise variance $\sigma^2(m)$ (in dB) obtained from a data set $\boldsymbol{y}$ of size (a) $n = 50$, (b) $n = 200$, and (c) $n = 1000$.

is adjusted so that the signal-to-noise ratio (SNR) of $\{y_t\}$ equals 15 dB. The parameters in the prior distributions are chosen as follows: $\boldsymbol{\phi}_0 = \mathbf{0}$, $\Sigma_0 = 1000 I_4$, $\nu = 2$, $\lambda = 0.3$, and $\alpha_i = \alpha_{ij} = 1$. Fig. 1(c) shows the i.i.d. uniform initial guess for $\{x_t\}$ in the Gibbs sampler. Fig. 1(d) and 1(e) present the MMSE estimate $\{\hat{x}_t\}$ and the MAP estimate $\{\tilde{x}_t\}$, respectively. These estimates are calculated from the last $M = 500$ samples of the total 1000 iterations of Gibbs sampling. The constraints $\phi_1 \geq 0.4$ and $\phi_1 \geq |\phi_i| + 0.2$ for $i \neq 1$ are used in order to remove the sign and shift ambiguities in the solution. The estimates of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are also obtained from their sample means and variances and given in the form of $E(\cdot|\boldsymbol{y}) \pm \sqrt{V(\cdot|\boldsymbol{y})}$ by, respectively, $(-0.1539, 0.8924, 0.4692, -0.1781) \pm (0.0191, 0.0195, 0.0202, 0.0211)$ and

$$\begin{bmatrix} 0.45 & 0.15 & 0.22 & 0.18 \\ 0.21 & 0.36 & 0.30 & 0.13 \\ 0.05 & 0.34 & 0.42 & 0.19 \\ 0.17 & 0.12 & 0.39 & 0.32 \end{bmatrix} \pm \begin{bmatrix} 0.11 & 0.08 & 0.09 & 0.08 \\ 0.07 & 0.08 & 0.08 & 0.06 \\ 0.03 & 0.07 & 0.08 & 0.06 \\ 0.08 & 0.06 & 0.10 & 0.09 \end{bmatrix}.$$

It is evident, by comparing Fig. 1(d) and 1(e) with Fig. 1(a), that the MAP estimate completely recovers the signal and the MMSE estimate recovers all except the last point at $t = 100$. The estimates of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are reasonably accurate given the large number of unknowns in this problem and the short length of the data record ($n = 100$).

Example 2: Consider the same MA(3) model in (10) but $\{x_t\}$ is now a 16-level i.i.d. sequence taking values uniformly in the state space $\mathcal{A} = \{\pm 1, \pm 3, \cdots, \pm 15\}$. A similar model has been used in [8] for channel equalization problems. The standard deviation of the noise is taken to be $\sigma = 0.3144$ so that SNR = 30 dB. In this example, we are interested in the convergence of Gibbs sampling for different lengths of data. Three cases are considered: they are $n = 50$, $n = 200$, and $n = 1000$. In each case, the Gibbs sampler is iterated 4000 times under the same constraints given in Example 1. The parameters in the prior distributions remain the same as in Example 1 except that the prior of $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is taken to be $D(\alpha)$ with $\alpha_i = 4$ for $n = 50$, $\alpha_i = 10$ for $n = 200$, and $\alpha_i = 40$ for $n = 1000$. Note that the growth of n necessitates the increase of α_i in order for the contribution of the resulting prior to be comparable with that of the data in the conditional posterior distribution of $\boldsymbol{\theta}$ given by (8).

As an indicator of convergence, the samples $\sigma^2(m)$ are shown in Fig. 2 against the iteration index m for the three cases. The constant line represents the true value of σ^2. Since $\sigma^2(m)$ measures the fit of the model at iteration m, convergence of the Gibbs sampler is indicated by a small variation of $\{\sigma^2(m)\}$. As we can see from Fig. 2, the Gibbs sampler needs more iterations to converge for a shorter data length than it does for a longer one. In addition, the variation of $\{\sigma^2(m)\}$ after convergence decreases as the data length grows. Another interesting feature to note is that the convergence is not achieved with a *gradually* decreasing variation of $\{\sigma^2(m)\}$; instead, it happens quite *abruptly* after a certain number of iterations.

VI. Concluding Remarks

The Gibbs sampling is applied to the blind restoration of discrete values signals when the blurring filter as well as the statistical parameters of the signal and the noise are unknown. This extends the existing methods in the literature to the simultaneous estimation of the parameters and the restoration of the signal. Simulations show that the method provides satisfactory solutions to the problem. A batch-processing-based adaptive procedure is also available (but not reported here) that can be used to track the changes of the FIR filter and/or the statistical parameters. Future research should extend the method to other signal and noise models.

References

[1] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 385–399, June 1980.

[2] G. E. P. Box and G. Tiao, *Bayesian Inference in Statistical Analysis*. New York: Wiley, 1973.

[3] R. Chen and J. S. Liu, "Predictive updating methods with applications to Bayesian classification," *J. Royal Statist. Soc. B*, in press, 1995.

[4] D. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis II*, D. F. Findley, Ed. New York: Academic, 1981.

[5] A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *J. Amer. Statist. Assoc.*, vol. 85, pp. 398–409, 1990.

[6] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, Nov. 1984.

[7] G. B. Giannakis and J. M. Mendel, "Identification of nonminimum phase systems using higher order statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 665–674, Mar. 1989.

[8] D. Hatzinakos and C. L. Nikias, "Blind equalization using a tricepstrum-based algorithm," *IEEE Trans. Commun.*, vol. 39, no. 5, May 1991.

[9] T. H. Li, "Blind identification and deconvolution of linear systems driven by binary random sequences," *IEEE Trans. Inform. Theory*, vol. 38, pp. 26–38, Jan. 1992.

[10] ——, "Blind deconvolution of linear systems with multilevel nonstationary inputs," *Ann. Statistics*, vol. 23, no. 2, pp. 690–704, 1995.

[11] T. H. Li, "Blind deconvolution of discrete-valued signals," in *Proc. IEEE 27th Asilomar Conf. Signals, Syst., Comput.*, vol. II, Pacific Grove, CA, 1993, pp. 1240–1244.

[12] K. S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes," *Ann. Statistics*, vol. 10, pp. 1195–1208, 1982.

[13] P. McCulloch and R. S. Tsay, "Statistical inference of Markov switching models with application to US GNP," Technical Report, Graduate School of Business, Univ. of Chicago, 1992.

[14] J. G. Proakis, *Digital Communications*, 2nd ed. New York: McGraw-Hill, 1989.